



ATLAS Note

GROUP-2017-XX



31st May 2021

Draft version 0.1

1

2

3

4

5

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

6

The ATLAS Collaboration

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

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

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

© 2021 CERN for the benefit of the ATLAS Collaboration.

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

26 Contents

27 I Introduction	4
28 1 Introduction	4
29 II Theoretical Motivation	7
30 2 The Standard Model and the Higgs Boson	7
31 2.1 The Forces and Particles of the Standard Model	7
32 2.2 The Higgs Mechanism	10
33 2.2.1 The Higgs Field	10
34 2.2.2 Electroweak Symmetry Breaking	12
35 2.3 $t\bar{t}H$ Production	15
36 2.4 WZ + Heavy Flavor Production	18
37 2.5 Extensions to the Standard Model	19
38 III The LHC and the ATLAS Detector	24
39 3 The LHC	24
40 4 The ATLAS Detector	27
41 4.1 Inner Detector	28
42 4.2 Calorimeters	29
43 4.3 Muon Spectrometer	30
44 4.4 Trigger System	31
45 IV Measurement of WZ + Heavy Flavor	33
46 5 Introduction	33
47 6 Data and Monte Carlo Samples	35
48 6.1 Data Samples	36
49 6.2 Monte Carlo Samples	37
50 7 Object Reconstruction	39
51 7.1 Trigger	40
52 7.2 Light leptons	40
53 7.3 Jets	41
54 7.4 B-tagged Jets	42
55 7.5 Missing transverse energy	44

56	7.6 Overlap removal	44
57	8 Event Selection and Signal Region Definitions	45
58	8.1 Event Preselection	46
59	8.2 Fit Regions	50
60	8.3 Non-Prompt Lepton Estimation	67
61	8.3.1 $t\bar{t}$ Validation	67
62	8.3.2 Z+jets Validation	73
63	9 tZ Separation Multivariate Analysis	78
64	9.1 Top Mass Reconstruction	78
65	9.2 tZ BDT	79
66	9.3 tZ Interference Studies	84
67	10 Systematic Uncertainties	86
68	11 Results	95
69	11.1 1-jet Fit Results	97
70	11.2 2-jet Fit Results	105
71	V Differential Studies of $t\bar{t}H$ Multilepton	114
72	12 Data and Monte Carlo Samples	114
73	12.1 Data Samples	114
74	12.2 Monte Carlo Samples	114
75	13 Object Reconstruction	117
76	13.1 Trigger Requirements	117
77	13.2 Light Leptons	118
78	13.3 Jets	119
79	13.4 B-tagged Jets	119
80	13.5 Missing Transverse Energy	120
81	13.6 Overlap removal	120
82	14 Higgs Momentum Reconstruction	121
83	14.1 Physics Object Truth Matching	123
84	14.2 b-jet Identification	124
85	14.2.1 2lSS Channel	125
86	14.2.2 3l Channel	129
87	14.3 Higgs Reconstruction	132
88	14.3.1 2lSS Channel	133
89	14.3.2 3l Semi-leptonic Channel	137
90	14.3.3 3l Fully-leptonic Channel	141
91	14.4 p_T Prediction	145

92	14.4.1 2lSS Channel	146
93	14.4.2 3l Semi-leptonic Channel	149
94	14.4.3 3l Fully-leptonic Channel	151
95	14.5 3l Decay Mode	153
96	15 Signal Region Definitions	155
97	15.1 Pre-MVA Event Selection	155
98	15.2 Event MVA	159
99	15.3 Signal Region Definitions	166
100	16 Systematic Uncertainties	167
101	17 Results	171
102	17.1 Results - 79.8 fb^{-1}	172
103	17.2 Projected Results - 139 fb^{-1}	176
104	VI Conclusion	181
105	VII Appendices	189
106	Appendices	189
107	A Non-prompt lepton MVA	189
108	B Supplementary WZ + Heavy Flavor Studies	194
109	B.1 Non-prompt CR Modelling	194
110	B.2 tZ Interference Studies	199
111	B.3 Alternate tZ Inclusive Fit	203
112	B.3.1 tZ Inclusive Fit	204
113	B.3.2 Floating tZ	205
114	C Supplementary t<bar>t>H Differential Analysis Studies</bar>	206
115	C.1 Higgs Reconstruction Model Details	206
116	C.1.1 b-jet Identification Features - 2lSS	206
117	C.1.2 b-jet Identification Features - 3l	210
118	C.1.3 Higgs Reconstruction Features - 2lSS	215
119	C.1.4 Higgs Reconstruction Features - 3lS	220
120	C.1.5 Higgs Reconstruction Features - 3lF	225
121	C.2 Background Rejection MVA Details	229
122	C.2.1 Background Rejection MVA Features - 2lSS	229
123	C.2.2 Background Rejection MVA Features - 3l	233
124	C.3 Truth Level Studies	238

125	C.4 Alternate b-jet Identification Algorithm	239
126	C.5 Binary Classification of the Higgs p_T	240
127	C.6 Impact of Alternative Jet Selection	242

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

Particle physics is an attempt to describe the fundamental building blocks of the universe and their interactions. The Standard Model (SM) - our best current theory of fundamental particle physics - does a remarkable job of that. All known fundamental particles and (almost) all of the forces underlying their interactions can be explained by the SM, and the predictions from this theory agree with experiment to an incredibly precise degree. This is especially true since the Higgs Boson, the last piece of the SM predicted decades before, was finally discovered at the Large Hadron Collider (LHC) in 2012 [1].

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

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

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

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

158 The proton-proton collision data collected by the ATLAS detector at the LHC from 2015-
159 2018 provides the opportunity to make this measurement for the first time. The unprecedented
160 energy achieved by the LHC during this period greatly increase the rate at which $t\bar{t}H$ events are
161 produced, and the large amount of data collected provides the necessary statistics for a differential
162 measurement to be performed.

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

166 final state - either two same-sign leptons, or three leptons. A deep neural network is used to
167 reconstruct the momentum of the Higgs Boson in each event. This momentum spectrum is fit to
168 data for each analysis channel in order to search for evidence of these BSM effects.

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

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

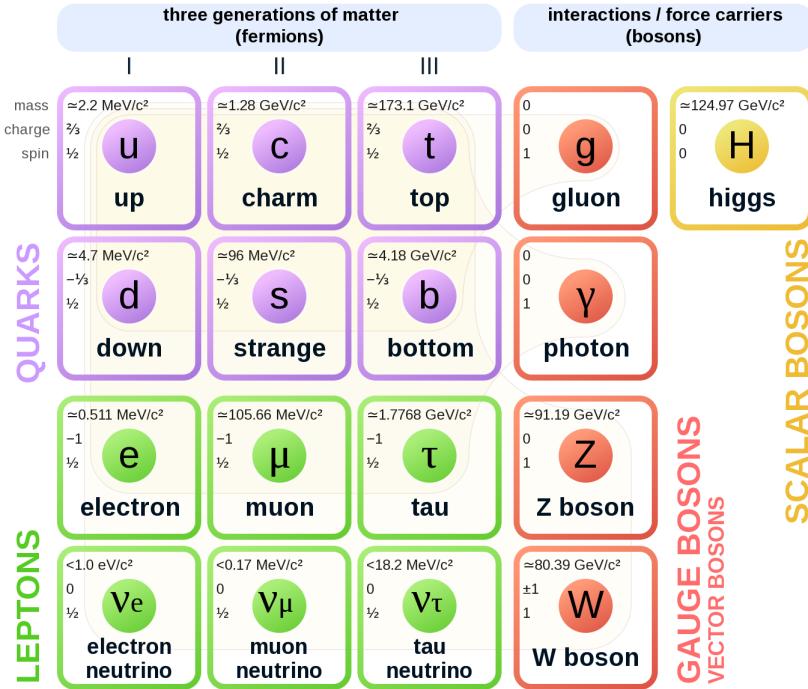


Figure 2.1: A summary of the particles of the Standard Model, including their mass, charge and spin, with the fermions listed on the left, and the bosons on the right. [2]

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

196 Leptons are fermions which interact via the electroweak force, but not the strong force.
 197 The three generation of leptons consist of the electron and electron neutrino, the muon and muon
 198 neutrino, the tau and tau neutrino. The quarks, by contrast, do interact via the strong force - which
 199 is to say they have color charge - in addition to the electroweak force. The three generations
 200 include the up and down quarks, the strange and charm quarks, and the top and bottom quarks.

201 Each of these generations form left-handed doublets invariant under SU(2) transfor-

202 mations. For the leptons these doublets are:

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

203 And for the quarks:

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

204 For both leptons and quarks, the heavier generations can decay into the lighter generation

205 of particles, while the first generation does not decay. Hence, ordinary matter generally consists

206 of this first generation of fermions - electrons, up quarks, and down quarks. Each of these

207 fermions has a corresponding anti-particle, which has an equal mass as its partner but opposite

208 charge. The fermions acquire their mass via the Higgs Mechanism, except for the neutrinos,

209 whose mass has been experimentally confirmed but is not accounted for in the SM.

210 Bosons, by contrast, have integer spin, and are therefore unconstrained by the Pauli-

211 exclusion principle. The SM includes two kinds of bosons: Gauge bosons, which are spin-1

212 particles that mediate the interactions between the fermions, and a single scalar, i.e. spin-0,

213 particle - the Higgs Boson. Of the gauge bosons, the W^+ , W^- and Z bosons - which are the

²¹⁴ mass eigenstates of the electroweak bosons - mediate the weak interaction, while the photon
²¹⁵ mediates the electric force, and the gluon mediates the strong force.

²¹⁶ **2.2 The Higgs Mechanism**

²¹⁷ A key feature of the SM is the gauge invariance of its Lagrangian. However, any terms added to
²¹⁸ the Lagrangian giving mass to the the gauge bosons would violate this underlying symmetry of
²¹⁹ the theory. This presents a clear problem with the theory: The experimental observation that the
²²⁰ W and Z bosons have mass seems to contradict the basic structure of the SM.

²²¹ Rather than abandoning gauge invariance, an alternative way for particles to acquire mass
²²² beyond adding a simple mass term to the Lagrangian was theorized by Higgs, Englert and Brout
²²³ in 1964 [3]. This procedure for introducing masses for the gauge bosons while preserving local
²²⁴ gauge invariance, known as the Higgs mechanism, was incorporated into the electroweak theory
²²⁵ by Weinberg in 1967 [4].

²²⁶ **2.2.1 The Higgs Field**

²²⁷ The Higgs mechanism introduces a complex scalar SU(2) doublet, Φ , with the form:

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

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

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

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

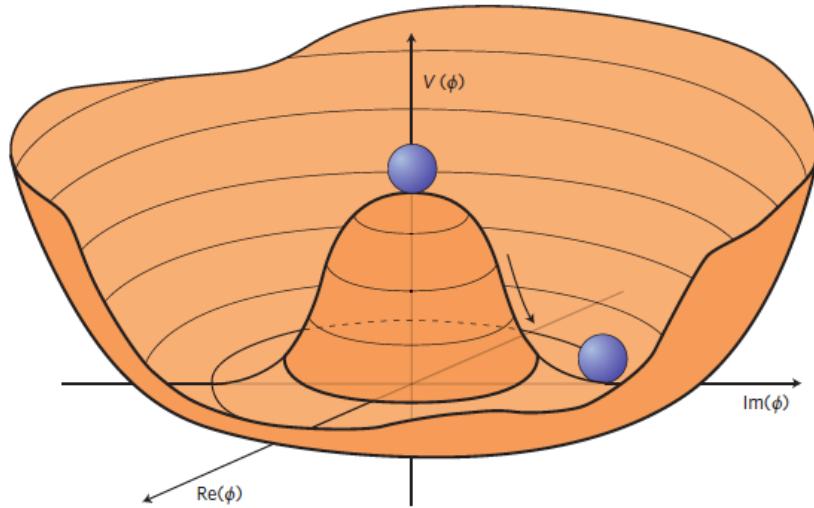


Figure 2.2: The value of the Higgs potential, $V(\Phi)$ as a function of Φ , for the case that $\mu^2 < 0$ [5].

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

²³⁶ the minimum is determined only by $\Phi^\dagger \Phi$, there is some ambiguity in the particular definition of

²³⁷ the VEV, but it is generally represented as

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

²³⁸ The full value of Φ can be written as

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

²³⁹ with v being the value of the VEV, and H being the real value of the scalar field.

²⁴⁰ 2.2.2 Electroweak Symmetry Breaking

²⁴¹ The Electroweak (EWK) interaction is described in the SM by a $SU(2)_L \otimes U(1)_Y$ gauge theory.

²⁴² 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,

²⁴³ B_μ . The couplings of these bosons to the Higgs field show up in the kinetic terms of the scalar

²⁴⁴ field Φ in the Lagrangian:

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

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

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

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

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

252 And the masses of these fields are given by:

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

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

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

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

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

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

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

²⁶⁹ free parameters of the theory, the mass of the Higgs boson and its interactions with the fermions

²⁷⁰ must be measured experimentally.

²⁷¹ **2.3 $t\bar{t}H$ Production**

²⁷² The strength of a particles interaction with the Higgs, given by its Yukawa coupling, is proportionate to its mass. The top quark - as the heaviest known particle - has the strongest interaction,
²⁷³ making this interaction particularly interesting to study. While several processes involve interactions between the Higgs and the top, some Higgs production modes include the top interaction
²⁷⁵ only as a part of a loop diagram, such as the gluon-gluon fusion diagram shown in Figure 2.3.

²⁷⁷ This process therefore only allows for an indirect probe of the Higgs-top Yukawa coupling, as
²⁷⁸ the flavor of the quark in this diagram is not unique.

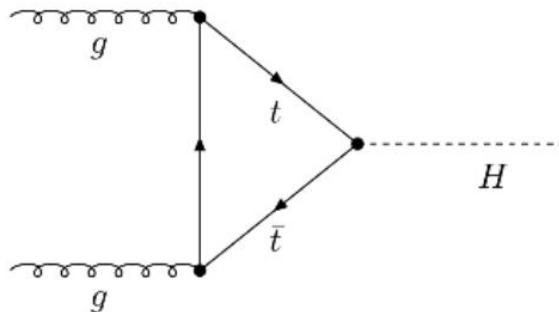


Figure 2.3: Diagram of a Higgs boson produced via gluon-gluon fusion.

²⁷⁹ Studying the Higgs produced in association with top quark pairs, $t\bar{t}H$, allows this interaction to be measured directly. This process, as shown in Figure 2.4, involves a unique coupling
²⁸⁰ between the Higgs and the top, which can be identified by the top quark pair in the final state.
²⁸¹

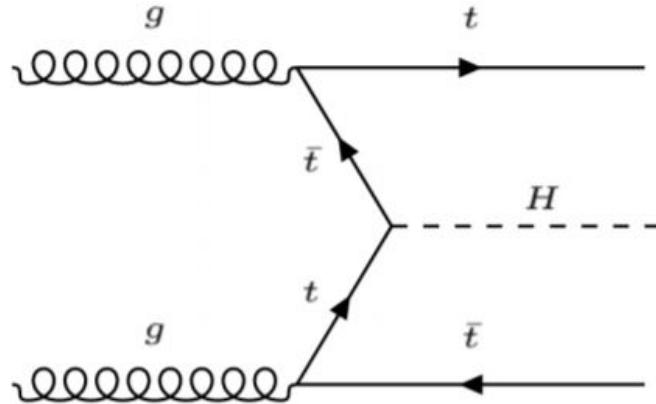


Figure 2.4: Diagram of a Higgs boson produced in association with a pair of top quarks.

282 The Higgs boson, as well as the top quarks, have very short lifetimes - on the order of
 283 10^{-22} s and 10^{-25} s respectively - meaning they can only be observed via their decay products.
 284 Measuring this process is therefore a matter of identifying events with final states consistent with
 285 $t\bar{t}H$ production.

286 Studies of $t\bar{t}H$ production have been reported by the ATLAS collaboration for $H \rightarrow b\bar{b}$,
 287 $H \rightarrow \gamma\gamma$ and multilepton (encompassing $H \rightarrow W^+W^-$, $H \rightarrow ZZ$ and $H \rightarrow \tau^-\tau^+$, with
 288 $H \rightarrow ZZ \rightarrow 4l$ as a separate analysis) decay modes. While the branching ratio of $H \rightarrow W^+W^-$
 289 is smaller than $H \rightarrow b\bar{b}$ (see Table 2.3), it produces a clearer signal, as $H \rightarrow b\bar{b}$ suffers from
 290 large $t\bar{t}$ backgrounds. On the other hand, $H \rightarrow \gamma\gamma$ produces the most easily identifiable signal,
 291 but has a much smaller branching ratio than $H \rightarrow W^+W^-$. Therefore, compared with other final
 292 states of $t\bar{t}H$, the $t\bar{t}H$ – ML channel is an attractive candidate for study, as it involves a good
 293 balance between statistical power and identifiability.

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

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

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

$$\mu_{t\bar{t}H} = \frac{\sigma_{t\bar{t}H}^{\text{obs.}}}{\sigma_{t\bar{t}H}^{\text{SM}}} \quad (2.12)$$

297 $t\bar{t}H$ production was observed by ATLAS using up to 79.8 fb^{-1} of data collected at \sqrt{s}
 298 $= 13 \text{ TeV}$, based on a combination of five Higgs decay modes: $b\bar{b}$, WW^* , $\tau^-\tau^+$, $\gamma\gamma$, and ZZ^*
 299 [6]. A significance of 5.8σ was observed, compared to a 4.9σ expected significance. Since then,
 300 two analyses have published updated results ($H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^* \rightarrow 4l$) with the full Run 2
 301 dataset, representing 139 fb^{-1} . Studies are still ongoing in the remaining channels.

302 This thesis focuses on $t\bar{t}H$ events with multiple leptons in the final state, $t\bar{t}H - ML$,
 303 specifically targeting events with two same-sign leptons (2lSS) or three leptons (3l) in the

304 final state. This includes $H \rightarrow W^+W^-$ events, where at least one of the W bosons decays
 305 leptonically.

306 2.4 WZ + Heavy Flavor Production

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

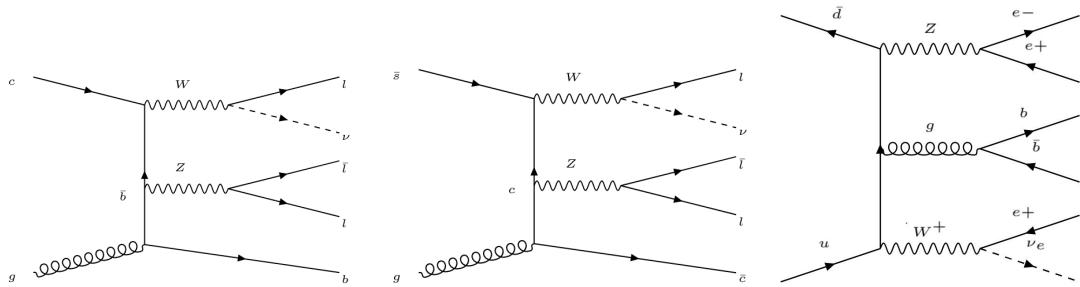


Figure 2.5: Example Feynman diagrams of WZ + heavy flavor production

312 The b -jets produced in this process can be thought of in two different ways: either as
 313 originating from the quark “sea” of the initial state hadrons, or as the result of a gluon from
 314 one the colliding protons splitting into $b\bar{b}$ pairs. However, the heavy flavor contribution to the
 315 parton distribution function (PDF) of the proton is uncertain, and simulations of this process
 316 disagree depending on which of these two approaches one considers. This makes WZ + heavy

317 flavor difficult to accurately simulate, and introduces a large uncertainty for any analysis which
318 includes it as a background, motivating a measurement of this process.

319 **2.5 Extensions to the Standard Model**

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

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

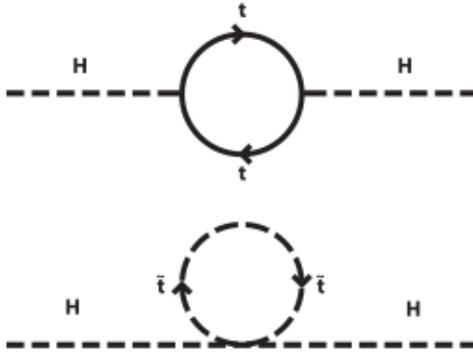


Figure 2.6: Above diagram is the leading order correction to the Higgs mass via a top quark loop, and below is the stop squark loop, coming from a supersymmetric extension of the SM, that provides a potential cancellation of the top diagram.

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

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

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

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{SM}} + \frac{f}{\Lambda} O^6 \quad (2.13)$$

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

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

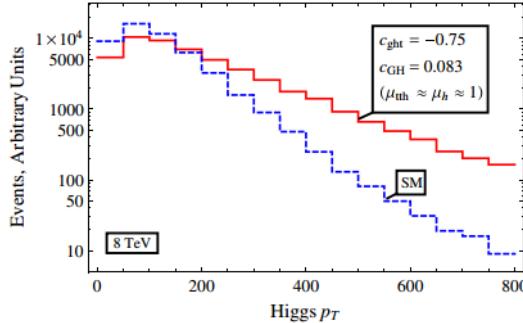


Figure 2.7: The momentum spectrum of the Higgs boson produced via top-quarks with (red) and without (blue) the presence of dimension-six operators.

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

360 Reconstructing the Higgs is a particular challenge in the multilepton channels of $t\bar{t}H$, due
 361 to an ambiguity arising from multiple sources of missing energy. In the $H \rightarrow \gamma\gamma$ channel, the
 362 kinematics of the Higgs can be fully reconstructed from the two photons. The same is true of
 363 $H \rightarrow b\bar{b}$, though with the additional challenge of identifying which two of the four b-quarks in
 364 the final state originated from the Higgs. By contrast, the two channels (2lSS and 3l) targeted
 365 by this analysis include at least one neutrino originating from the Higgs decay.

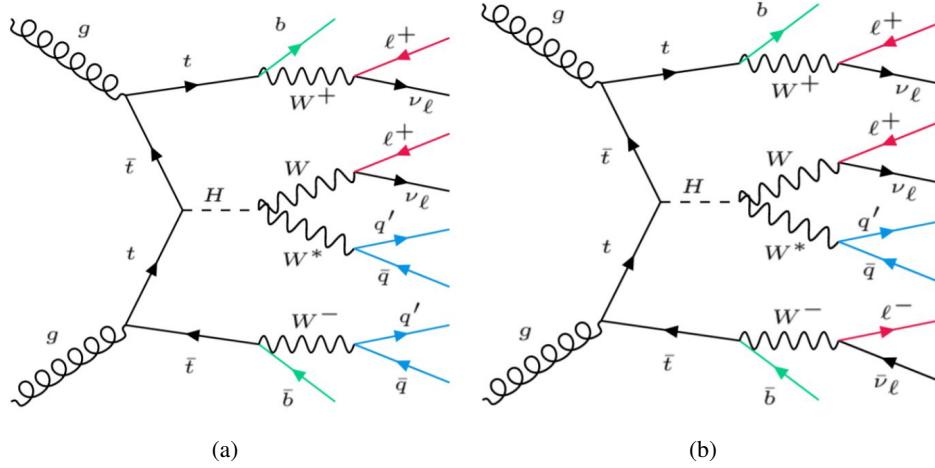


Figure 2.8: Feynman diagrams of $t\bar{t}H$ production with (a) two same-sign leptons and (b) three leptons in the final state.

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

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

374 Part III

375 The LHC and the ATLAS Detector

376 3 The LHC

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

386 The LHC first began running in 2009 at a proton-proton center of mass energy of $\sqrt{s} = 8$
387 TeV. It operated at this energy from 2009 to 2012, known as Run 1, and data collected during
388 this period was used in discovering the Higgs Boson. The LHC began running again in 2015,
389 and collected data at an increased energy of $\sqrt{s} = 13$ TeV until 2018, a period referred to as Run
390 2.

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

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

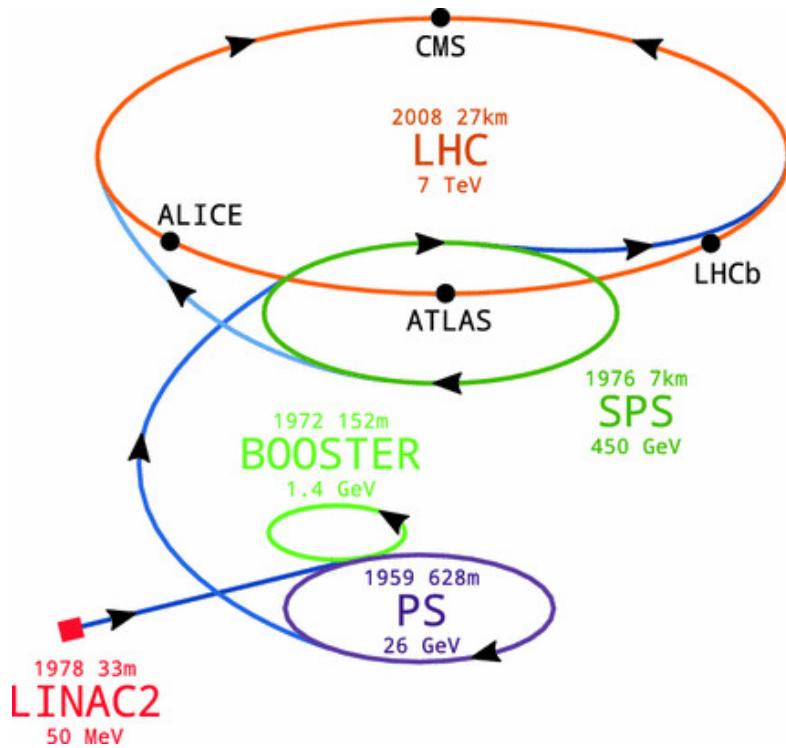


Figure 3.1: A summary of the accelerator chain used to feed protons into the LHC [11].

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

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

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

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

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

410 The integrated luminosity collected by the ATLAS detector as of the end of 2018 is around
 411 140 fb^{-1} , exceeding the expected integrated luminosity of 100 fb^{-1} .

412 4 The ATLAS Detector

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

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

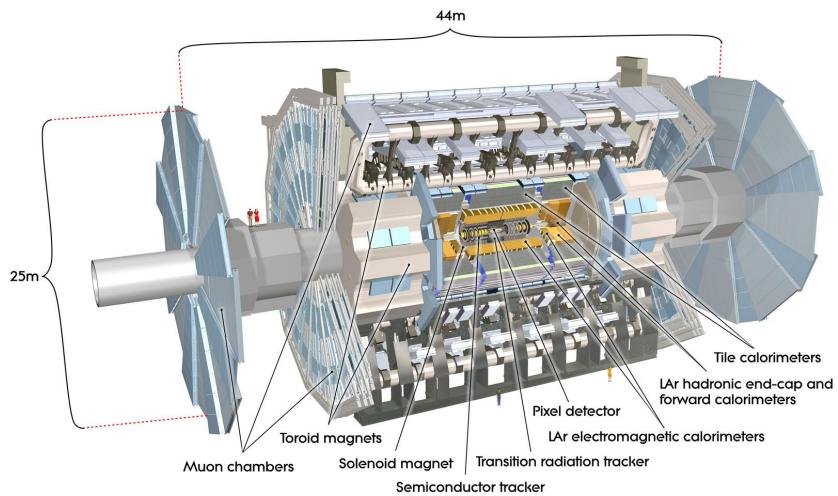


Figure 4.1: Cutaway view of the ATLAS detector, with labels of its major components [12].

421 **4.1 Inner Detector**

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

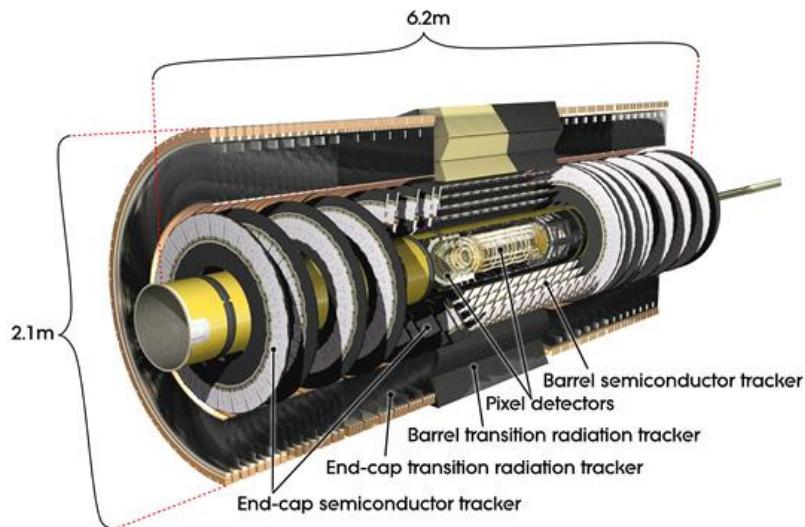


Figure 4.2: Cutaway view of the Inner Detector [13].

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

432 The Semiconductor Tracker (SCT) is similar to the Pixel detector, but uses long strips of
 433 silicon rather than small pixels to cover a larger spatial area. It includes over 6 million readout
 434 strips, allowing the position of charged particles to be measured to an accuracy of 17 μm .

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

441 4.2 Calorimeters

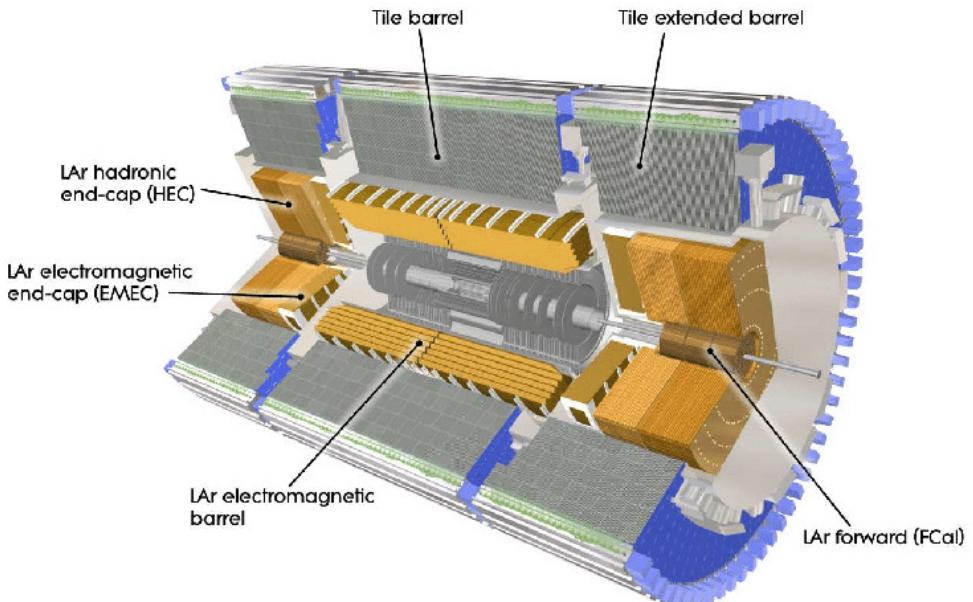


Figure 4.3: Cutaway view of the calorimeter system of the ATLAS detector [13].

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

449 The outer calorimeter measures the energy from particles that pass through the EM
450 calorimeter, and measures the energy of particles that interact via the strong force. This is
451 primarily hadrons. It is composed of steel plates designed to cause hadronic showering and
452 around 500,000 scintillating tiles as the active material. The signals from the hadronic calorimeter
453 are read out by photomultiplier tubes (PMTs).

454 **4.3 Muon Spectrometer**

455 Because muons are heavier than electrons and photons, and do not interact via the strong force,
456 they generally pass through the detector without being stopped by the calorimeters. The outermost
457 components of the detector are designed specifically to measure the energy and momentum of
458 muons produced in the LHC. The muon spectrometer consists of tracking and triggering systems.

459 The largest of the subdetectors, it extends from the outside of the calorimeter system, about a
460 4.25 m radius from the beam line, to a radius of 11 m. This large detector system is necessary

461 to accurately measure the momentum of muons, which is essential not only for measurements
462 involving the muons themselves, but also to accurately estimate the missing energy in each
463 event.

464 Two large toroidal magnets within the muon system generate a large magnetic field which
465 covers an area 26 m long with a radius of 10 m. Because the area covered by this magnet system
466 is so large, a uniform magnetic field like the one produced in the Inner Detector is impractical.
467 Instead, the magnetic field that exists in the muon spectrometer ranges between 2 T and 8 T, and
468 is much less uniform. The path of the muons passing through the spectrometer is bent by this
469 field, allowing their charge to be determined.

470 1200 tracking chambers are placed in the muon system in order to precisely measure the
471 tracks of muons with high spatial resolution. The path of the muons are tracked by Monitored
472 Drift Tubes, which are drift chambers formed by aluminum tubes and filled with ionizing gas.
473 These tubes produce a multi-layer spatial resolution on the order of 50 μm .

474 **4.4 Trigger System**

475 Because of the high collision rate and large amount of data collected by the various subdetectors,
476 ATLAS produces far more data than can actually be stored. Each event produces around 25 Mb
477 of raw data, which multiplied by the bunch crossing rate of 40 MHz, comes out to around a
478 petabyte of data every second. The information from every event cannot practically be stored,

479 therefore a sophisticated trigger system is employed in real time to determine whether events are
480 sufficiently interesting to be worth storing.

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

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

493 Part IV

494 Measurement of WZ + Heavy Flavor

495 5 Introduction

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

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

506 Events that meet a preselection criteria are sorted into regions based on the b-tagging score
 507 of their associated jets. This is done to separate WZ + b-jet events from WZ + charm and WZ +
 508 light jets. These regions are fit to data in order to make a more accurate estimate of the contribution
 509 of WZ + heavy-flavor, where heavy-flavor jets include b-jets and charm jets. The full Run-2

510 dataset collected by the ATLAS detector, representing 139 fb^{-1} of data from pp collisions at
 511 $\sqrt{s} = 13 \text{ TeV}$, is used for this study.

512 The fiducial volume at particle level is defined based on the number of stable leptons and
 513 jets in each event. Three light leptons with total charge ± 1 and one or two associated jets are
 514 required. Only leptons which do not originate from hadron or τ decays are considered. The
 515 phase space definitions use dressed kinematics of the final state particles. Leptons are dressed
 516 by summing the momentum of photons within a cone of $\Delta R < 0.1$ of the lepton to correct the
 517 leptons energy. Particle level jets are reconstructed using the anti- k_t algorithm with a radius of
 518 $R = 0.4$. The kinematic selection applied to these objects is summarized below:

- 519 • Three light leptons with total charge ± 1 , $|\eta| < 2.5$
- 520 • OS lepton with $p_T > 10 \text{ GeV}$, SS leptons with $p_T > 20 \text{ GeV}$
- 521 • One OSSF lepton pair with $|M(l\bar{l}) - 91.2 \text{ GeV}| < 10 \text{ GeV}$
- 522 • One or two associated truth jets with $p_T > 25 \text{ GeV}$, $|\eta| < 2.5$, $R < 0.4$

523 The result of the fit is used to extract the cross-section in this fiducial region for WZ + b
 524 and WZ + c with one associated jet, and WZ + b and WZ + c with two associated jets, where the
 525 number and flavor of the jets is determined at particle level. Events with both charm and b-jets
 526 are counted as WZ + b. The analysis reports a cross-section measurement of WZ + b and WZ +
 527 c, along with their correlations, for both 1-jet and 2-jet exclusive regions. Normalization factors,
 528 representing how the MC prediction differs from the observed result, are also reported.

529 Section 6 details the data and Monte Carlo (MC) samples used in the analysis. The
 530 reconstruction of various physics objects is described in Section 7. Section 8 describes the event
 531 selection applied to these samples, along the definitions of the various regions used in the fit.
 532 The multivariate analysis techniques used to separate the tZ background from WZ + heavy flavor
 533 are described in Section 9. Section 16 describes the various sources of systematic uncertainties
 534 considered in the fit. Finally, the results of the analysis are summarized in Section 17, followed
 535 by a brief conclusion in Section ??.

536 **6 Data and Monte Carlo Samples**

537 Both data and Monte Carlo samples used in this analysis were prepared in the xAOD format,
 538 which was used to produce a Dx AOD sample in the HIGG8D1 derivation framework. The HIGG8D1
 539 framework is designed for the $t\bar{t}H$ multi-lepton analysis, which targets events with multiple
 540 leptons as well as tau hadrons. This framework reduces the size of the dataset by removing
 541 events based on event topology and only keeping useful information for each event. Events are
 542 removed from the derivations that do not meet one of the following selections:

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

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

550 **6.1 Data Samples**

551 The study uses a sample of proton-proton collision data collected by the ATLAS detector from
552 2015 through 2018 at an energy of $\sqrt{s} = 13$ TeV, which represents an integrated luminosity of
553 139 fb^{-1} [15]. This data set was collected with a bunch-crossing rate of 25 ns. All data used
554 in this analysis was verified by data quality checks [16], having been included in the following
555 Good Run Lists:

- 556 • data15_13TeV.periodAllYear_DetStatus-v79-repro20-02_DQDefects-00-02-02
557 _PHYS_StandardGRL_All_Good_25ns.xml
- 558 • data16_13TeV.periodAllYear_DetStatus-v88-pro20-21_DQDefects-00-02-04
559 _PHYS_StandardGRL_All_Good_25ns.xml
- 560 • data17_13TeV.periodAllYear_DetStatus-v97-pro21-13_Unknown_PHYS_StandardGRL
561 _All_Good_25ns_Triggerno17e33prim.xml
- 562 • data18_13TeV.periodAllYear_DetStatus-v102-pro22-04_Unknown_PHYS_StandardGRL
563 _All_Good_25ns_Triggerno17e33prim.xml

564 Runs included from the AllYear period containers are included.

565 **6.2 Monte Carlo Samples**

566 Several different generators were used to produce Monte Carlo simulations of the signal and
567 background processes. For all samples, the response of the ATLAS detector is simulated using
568 GEANT4 [17]. The WZ signal samples are simulated using Sherpa 2.2.2 [18]. Signal events are
569 generated using NNPDF30NNLO PDF set with up to one parton at NLO and 2 to 3 partons at
570 LO [**Butterworth:2015oua**].

571 The tZ background is simulated at NLO with **MADGRAPH5_AMC@NLO**, with **PYTHIA8**
572 used to perform parton showering and fragmentation. The NNPDF30NNLO PDF set is used.

573 Specific information about the Monte Carlo samples being used can be found in Table 21.
574 A list of the specific samples used by data set ID is shown in Table 3.

Table 2: The configurations used for event generation of signal and background processes, including the event generator, matrix element (ME) order, parton shower algorithm, and parton distribution function (PDF).

Process	Event generator	ME order	Parton Shower	PDF
WZ, VV	SHERPA 2.2.2	MEPS NLO	SHERPA	CT10 [19]
tZ	MG5_AMC [20]	NLO	PYTHIA 8	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++) [22]	NNPDF 3.0 NLO [21] (CT10 [19])
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 [23]
t̄tW+W-	MG5_AMC	LO	PYTHIA 8	NNPDF 2.3 LO
t̄t	PowHEG-BOX v2 [24]	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 [25]	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	412063-5
tW	410013-4
WtZ	410408
VVV	364242-9
VH	342284-5
WtH	341998
t̄tγ	410389
t̄t	410470
t̄tH	345873-5, 346343-5

Table 3: List of Monte Carlo samples by data set ID used in the analysis.

575 7 Object Reconstruction

- 576 All regions defined in this analysis share a common lepton, jet, and overall event preselection.
- 577 The selection applied to each physics object is detailed here; the event preselection, and the
- 578 selection used to define the various fit regions, is described in Section 8.

579 **7.1 Trigger**

580 Events are required to be selected by dilepton triggers, as summarized in Table 4.

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

581 **7.2 Light leptons**

582 Electron candidates are reconstructed from energy clusters in the electromagnetic calorimeter
 583 that are associated with charged particle tracks reconstructed in the inner detector [26]. Electron
 584 candidates are required to have $p_T > 10$ GeV and $|\eta_{\text{cluster}}| < 2.47$. Muon candidates are
 585 reconstructed by combining inner detector tracks with track segments or full tracks in the muon
 586 spectrometer [27]. Muon candidates are required to have $p_T > 10$ GeV and $|\eta| < 2.5$. Candidates

587 in the transition region between different electromagnetic calorimeter components, $1.37 < | $\eta_{\text{cluster}}| < 1.52$$, are rejected. A multivariate likelihood discriminant combining shower shape
 588 and track information is used to distinguish real electrons from hadronic showers (fake electrons).
 589
 590 To further reduce the non-prompt electron contribution, the track is required to be consistent
 591 with originating from the primary vertex; requirements are imposed on the transverse impact
 592 parameter significance ($|d_0|/\sigma_{d_0} < 5$) and the longitudinal impact parameter ($|\Delta z_0 \sin \theta_\ell| < 0.5$
 593 mm). Electron candidates are required to pass TightLH identification.

594 Muon candidates are reconstructed by combining inner detector tracks with track segments
 595 or full tracks in the muon spectrometer [27]. Muon candidates are required to have $p_T > 10$ GeV
 596 and $|\eta| < 2.5$. The longitudinal impact parameter is the same for both electrons and muons, while
 597 muons are required to pass a slightly tighter transverse impact parameter, $|d_0|/\sigma_{d_0} < 3$. Muons
 598 are also required to pass Medium ID requirements.

599 Leptons are additionally required to pass a non-prompt BDT selection developed by the
 600 $t\bar{t}H$ multilepton/ $t\bar{t}W$ analysis group. This BDT and the WPs used are summarized in Appendix
 601 A, and described in detail in [28]. Optimized working points and scale factors for this BDT are
 602 taken from that analysis.

603 7.3 Jets

604 Jets are reconstructed from calibrated topological clusters built from energy deposits in the
 605 calorimeters [29], as well as information from the inner tracking detector, using the anti- k_t

606 algorithm with a radius parameter $R = 0.4$. Particle Flow, or PFlow, jets are used in the analysis.
607 Jets with energy contributions likely arising from noise or detector effects are removed from
608 consideration [30], and only jets satisfying $p_T > 25 \text{ GeV}$ and $|\eta| < 2.5$ are used in this analysis.
609 For jets with $p_T < 60 \text{ GeV}$ and $|\eta| < 2.4$, a jet-track association algorithm is used to confirm
610 that the jet originates from the selected primary vertex, in order to reject jets arising from pileup
611 collisions [31].

612 **7.4 B-tagged Jets**

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

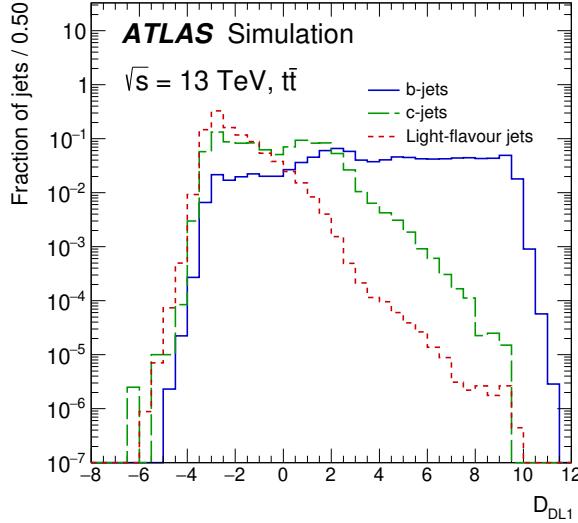


Figure 7.1: Output distribution of the DL1r algorithm for pure samples of b-jets, charm jets, and light jets, with each normalized to unity [32]

618 From the output of the BDT, working points (WPs) are developed based on the efficiency
 619 of truth b-jets at particular values of the DL1r algorithm. These WPs are taken from the March
 620 2020 CDI file, 2020-21-13TeV-MC16-CDI-2020-03-11_v2.root. The working points used in
 621 this analysis are summarized in Table 5.

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

Table 5: c-jet and light-flavor jet rejections corresponding to each b-tagging Working Point by b-jet efficiency, evaluated on $t\bar{t}$ events.

622 As shown in table 5, a tighter WP will accept fewer b-jets, but reject a higher fraction of

623 charm and light jets. Generally, analyses that include b-jets will use a fixed working point, for
 624 example, requiring that a jet pass the 70% threshold. By instead treating these working points
 625 as bins, e.g. events with jets that fall between the 85% and 77% WPs fall into one bin, while
 626 events with jets passing the 60% WP fall into another, additional information can be gained.
 627 This analysis uses each of these working points to form orthogonal regions in order to provide
 628 separation between $WZ + b$, $WZ + c$, and $WZ + \text{light}$.

629 **7.5 Missing transverse energy**

630 Missing transverse momentum (E_T^{miss}) is used as part of the event selection. The missing
 631 transverse momentum vector is defined as the negative of the vector sum of the transverse
 632 momenta of all reconstructed physics objects as well as remaining unclustered energy, the latter
 633 of which is estimated from low- p_T tracks associated with the primary vertex but not assigned
 634 to a hard object, with object definitions taken from [33]. Light leptons considered in the E_T^{miss}
 635 reconstruction are required to have $p_T > 10 \text{ GeV}$, while jets are required to have $p_T > 20 \text{ GeV}$.

636 **7.6 Overlap removal**

637 To avoid double counting objects and remove leptons originating from decays of hadrons, overlap
 638 removal is performed in the following order: any electron candidate within $\Delta R = 0.1$ of another
 639 electron candidate with higher p_T is removed; any electron candidate within $\Delta R = 0.1$ of a muon
 640 candidate is removed; any jet within $\Delta R = 0.2$ of an electron candidate is removed; if a muon

641 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
 642 is kept and the muon is removed if the jet has at least three associated tracks, otherwise the jet is
 643 removed and the muon is kept.

644 This algorithm is applied to the preselected objects. The overlap removal procedure is
 645 summarized in Table 22.

Keep	Remove	Cone size (ΔR)
electron	electron (low p_T)	0.1
muon	electron	0.1
electron	jet	0.2
jet	muon	$\min(0.4, 0.04 + 10[\text{GeV}]/p_T(\text{muon}))$, $n_{\text{track}} > 3$
muon	jet	$\min(0.4, 0.04 + 10[\text{GeV}]/p_T(\text{muon}))$, $n_{\text{track}} < 3$
electron	tau	0.2

Table 6: Summary of the overlap removal procedure between electrons, muons, and jets.

646 8 Event Selection and Signal Region Definitions

647 Event are required to pass a preselection described in Section 8.1 and summarized in Table 7.
 648 Those that pass this preselection are divided into various fit regions described in Section 8.2,
 649 based on the number of jets in the event, and the b-tag score of those jets.

650 8.1 Event Preselection

651 Events are required to include exactly three reconstructed light leptons passing the requirement
 652 described in 7.2, which have a total charge of ± 1 .

653 The leptons are ordered in the analysis code as 0, 1, and 2. Lepton 0 is the lepton whose
 654 charge is opposite the other two. Lepton 1 is the lepton closest to the opposite charge lepton, i.e.
 655 the smallest ΔR , leaving lepton 2 as the lepton further from the opposite charge lepton. Lepton 0
 656 is required to have $p_T > 10$ GeV, as it is found to be prompt the vast majority of the time, while
 657 the same sign leptons, 1 and 2, are required to have $p_T > 20$ GeV to reduce the contribution of
 658 non-prompt leptons, as non-prompt leptons tend to be soft.

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

663 An additional requirement is placed on the missing transverse energy, $E_T^{\text{miss}} > 20$ GeV, and
 664 the transverse mass of the W candidate, $m(E_T^{\text{miss}} + l_{\text{other}}) > 30$ GeV, defined as $\sqrt{2 p_T^{\text{lep}} E_T^{\text{miss}} * (1 - \cos(\phi_{\text{lep}} - \phi_{E_T^{\text{miss}}}))}$.
 665 Here E_T^{miss} is the missing transverse energy, and l_{other} is the lepton not included in the Z-
 666 candidate.

667 Events are required to have one or two reconstructed jets passing the selection described
 668 in Section 7.3. Events with more than two jets are rejected in order to reduce the contribution of
 669 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 total charge ± 1
Two same-charge leptons with $p_T > 20 \text{ GeV}$
One opposite charge lepton with $p_T > 10 \text{ GeV}$
$m(l^+l^-)$ within 10 GeV of 91.2 GeV
Transverse mass of W-candidate, $\sqrt{2p_T^{\text{lep}} E_T^{\text{miss}} * (1 - \cos(\phi_{\text{lep}} - \phi_{E_T^{\text{miss}}}))} > 30 \text{ GeV}$
Missing transverse energy, $E_T^{\text{miss}} > 20 \text{ GeV}$
One or two jets with $p_T > 25 \text{ GeV}$

Table 7: Summary of the selection applied to events for inclusion in the fit

670 The event yields in the preselection region for both data and Monte Carlo are summarized
 671 in Table 8.1, which shows good agreement between data and Monte Carlo, and demonstrates that
 672 this region consists primarily of WZ + jets events. The WZ events are split into WZ + b, WZ
 673 + c, and WZ + l based on the truth flavor of the associated jet in the event. Specifically, this
 674 determination is made based on the HadronConeExclTruthLabelID of the jet, as recommended
 675 by the b-tagging working group [34]. In this ordering b-jet supersedes charm, which supersedes
 676 light. That is, WZ + light events contain no charm and no b jets at truth level, WZ + c contain at
 677 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 8: Event yields in the preselection region at 139.0 fb^{-1}

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

680 Simulations are further validated by comparing the kinematic distributions of the Monte
 681 Carlo with data, which are shown in Figure 8.1. Here, bins with 5% or more WZ+b are blinded.

WZ Fit Region - Inclusive

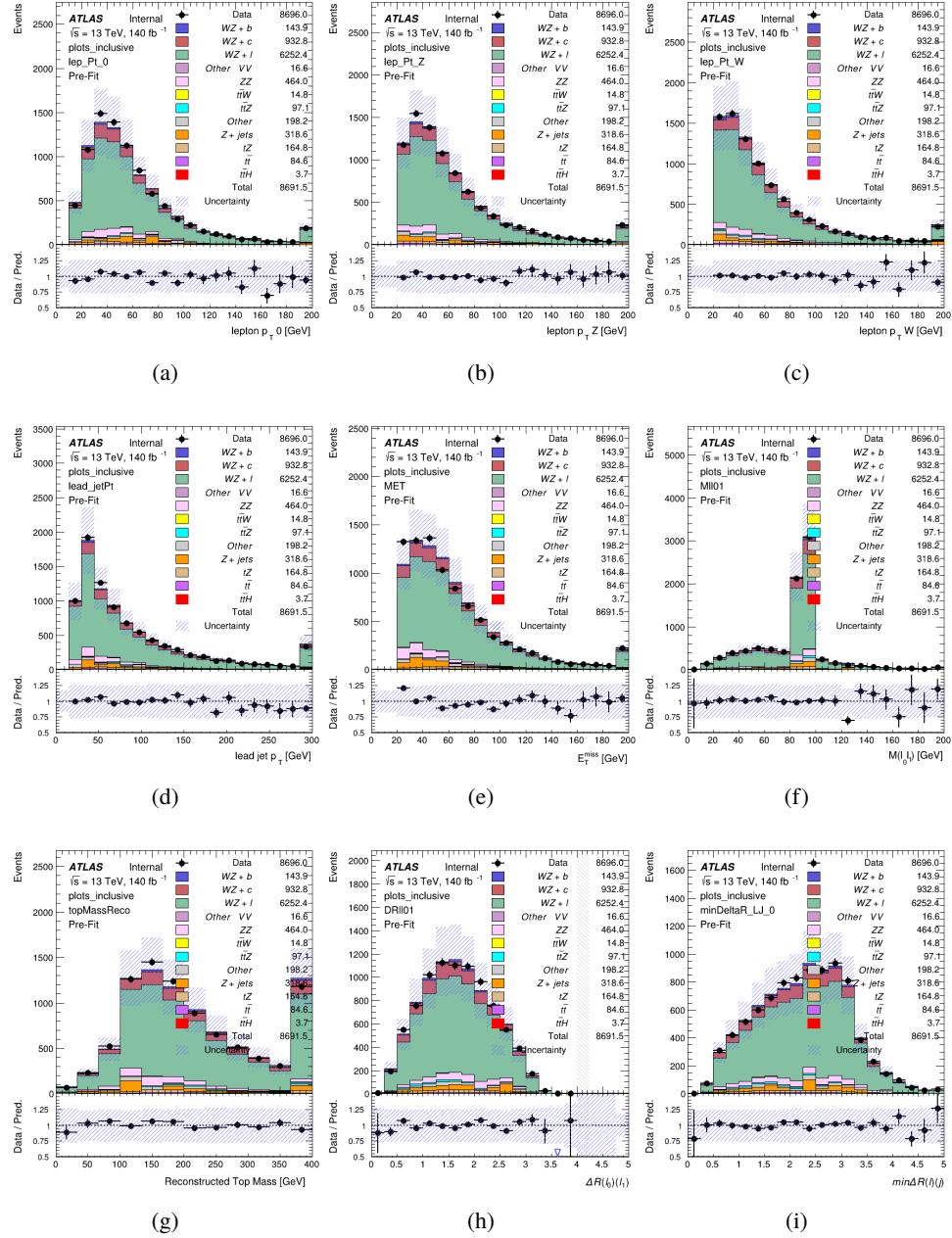


Figure 8.1: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant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

682 8.2 Fit Regions

683 Once preselection has been applied, the remaining events are categorized into one of twelve
 684 orthogonal regions. The regions used in the fit are summarized in Table 9.

Table 9: 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%	$N_{\text{jets}} = 1, n\text{Jets_DL1r_60} = 1, tZ \text{ BDT} > 0.12$
1j tZ CR	$N_{\text{jets}} = 1, n\text{Jets_DL1r_60} = 1, tZ \text{ BDT} < 0.12$
2j, <85%	$N_{\text{jets}} = 2, n\text{Jets_DL1r_85} = 0$
2j, 85%-77%	$N_{\text{jets}} = 2, n\text{Jets_DL1r_85} \geq 1, n\text{Jets_DL1r_77} = 0$
2j, 77%-70%	$N_{\text{jets}} = 2, n\text{Jets_DL1r_77} \geq 1, n\text{Jets_DL1r_70} = 0$
2j, 70%-60%	$N_{\text{jets}} = 2, n\text{Jets_DL1r_70} \geq 1, n\text{Jets_DL1r_60} = 0$
2j, >60%	$N_{\text{jets}} = 2, n\text{Jets_DL1r_60} \geq 1, tZ \text{ BDT} > 0.12$
2j tZ CR	$N_{\text{jets}} = 2, n\text{Jets_DL1r_60} \geq 1, tZ \text{ BDT} < 0.12$

685 The working points discussed in Section 7.4 are used to separate events into fit regions
 686 based on the highest working point reached by a jet in each event. Because the background
 687 composition differs significantly based on the number of b-jets, events are further subdivided
 688 into 1-jet and 2-jet regions in order to minimize the impact of background uncertainties.

689 An unfolding procedure is performed to account for differences in the number of recon-
 690 structed jets compared to the number of truth jets in each event. The `AntiKt4TruthDressedWZJets`
 691 truth jet collection is used to make this determination. In order to account for migration of WZ+1-
 692 jet and WZ+2-jet events between the 1-jet and 2-jet bins at reco level, the signal samples are
 693 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 9, 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 8.2. 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 16.

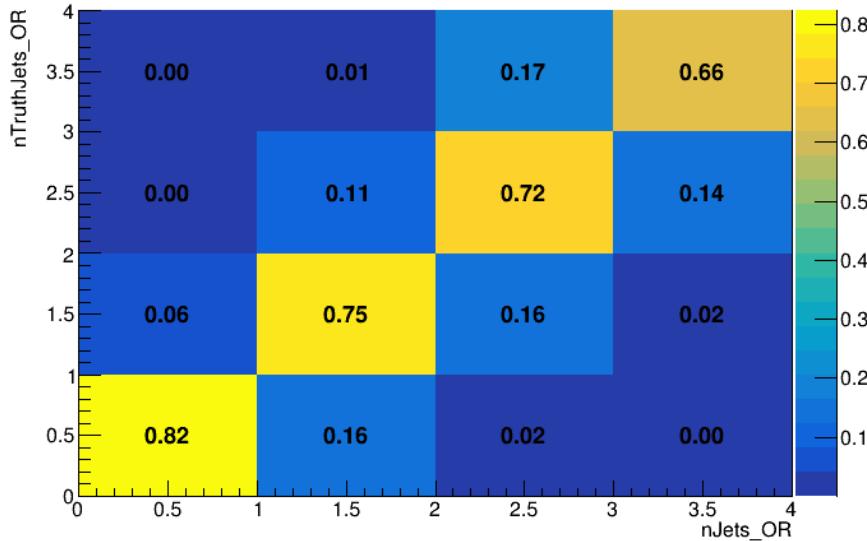


Figure 8.2: Number of truth jets compared to the number of reconstructed jets in each WZ event falling in the preselection region. Each row is normalized to unity.

An additional tZ control region is created based on the BDT described in Section 9. The region with 1-jet passing the 60% working point is split in two - a signal enriched region of events with a BDT score greater than 0.12, and a tZ control region including events with a score less than 0.12. This cutoff is arrived at by performing a fit an Asimov dataset with a variety of cutoffs, and selecting the value that produces the highest significance for the measurement of WZ + b.

704 The modeling in each region is validated by comparing data and MC predictions for
705 various kinematic distributions. Events containing 5% or more WZ + b are blinded. These plot
706 are shown in Figures 8.3-8.16.

WZ Fit Region - 1j Inclusive

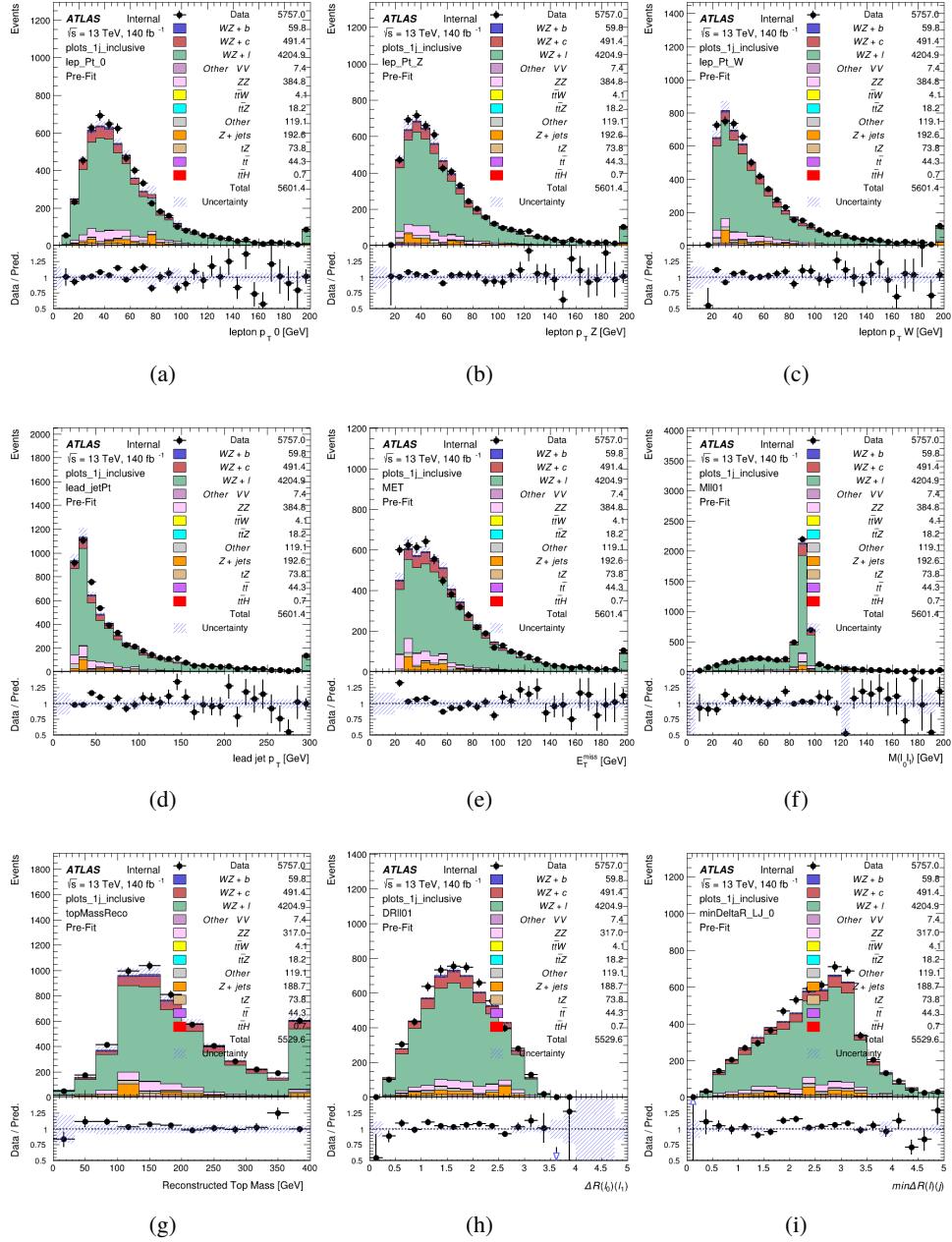


Figure 8.3: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant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

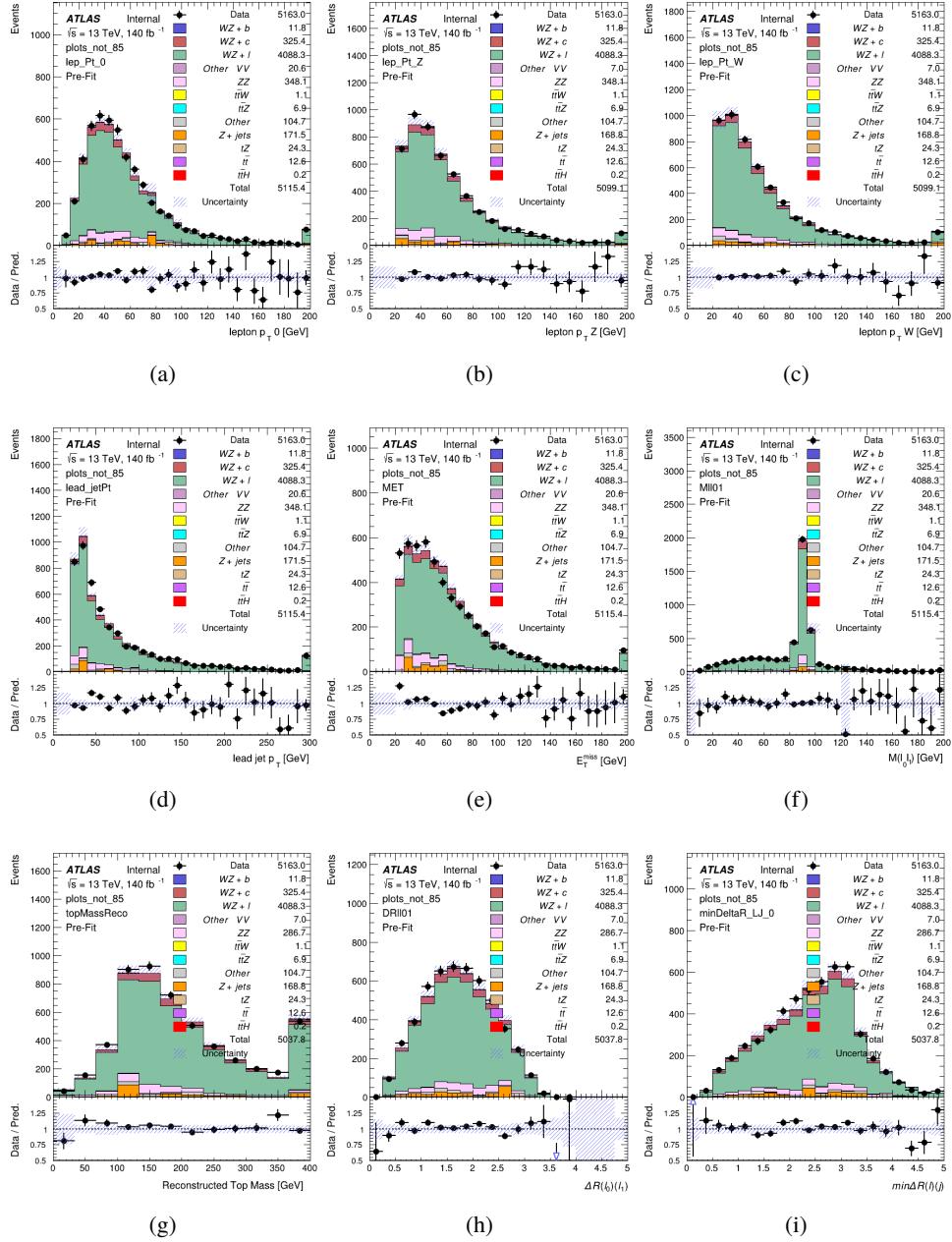


Figure 8.4: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant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

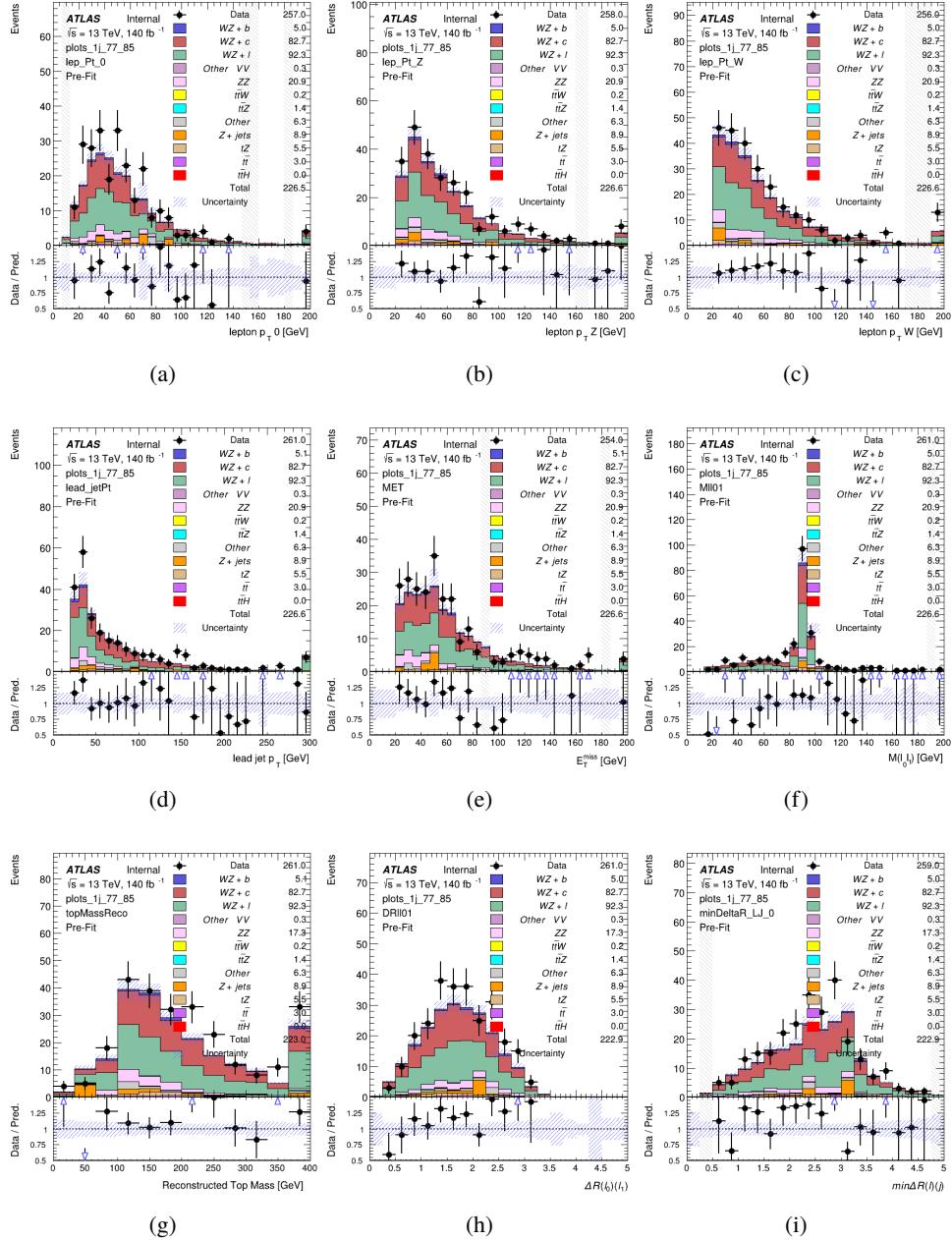


Figure 8.5: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant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

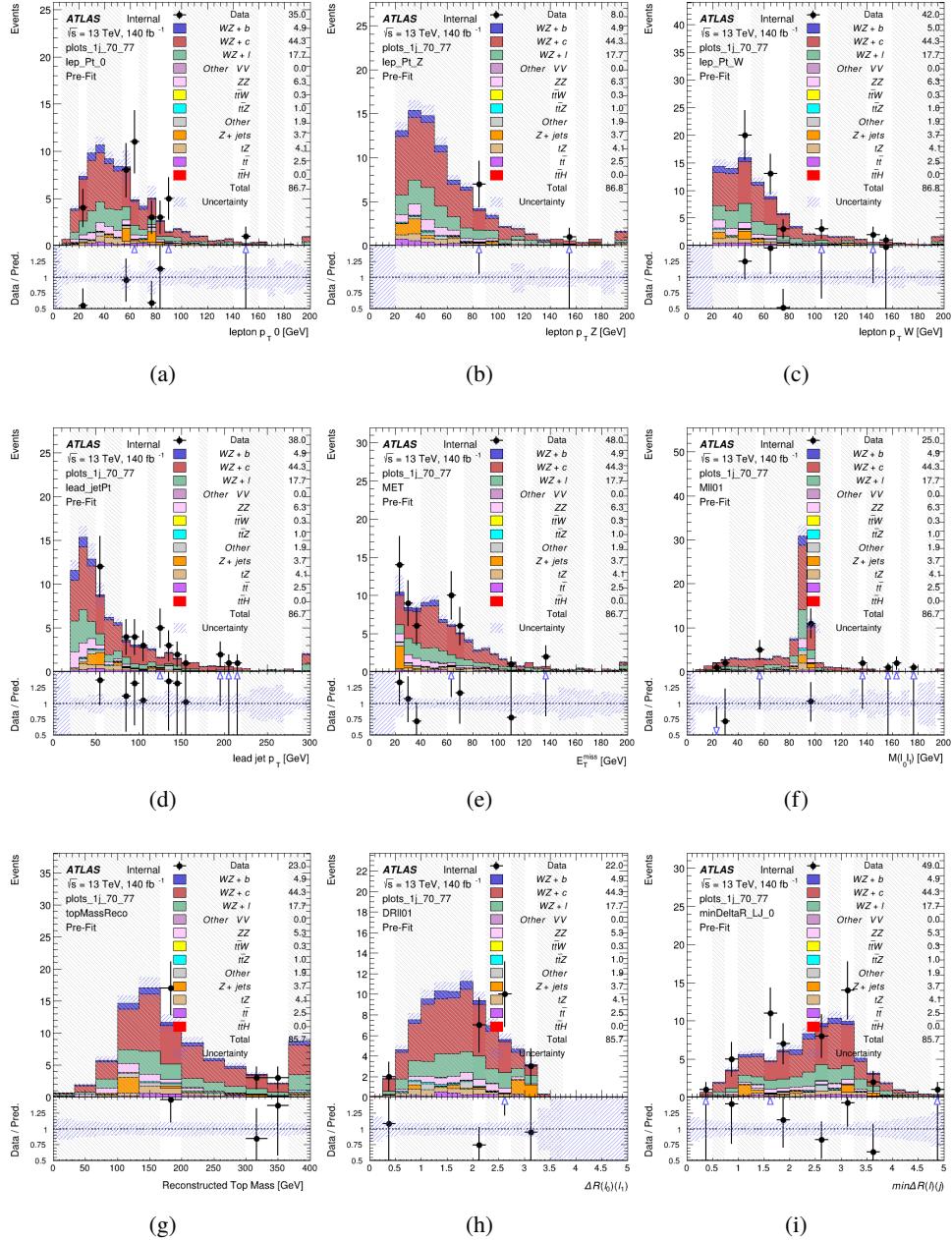


Figure 8.6: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant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

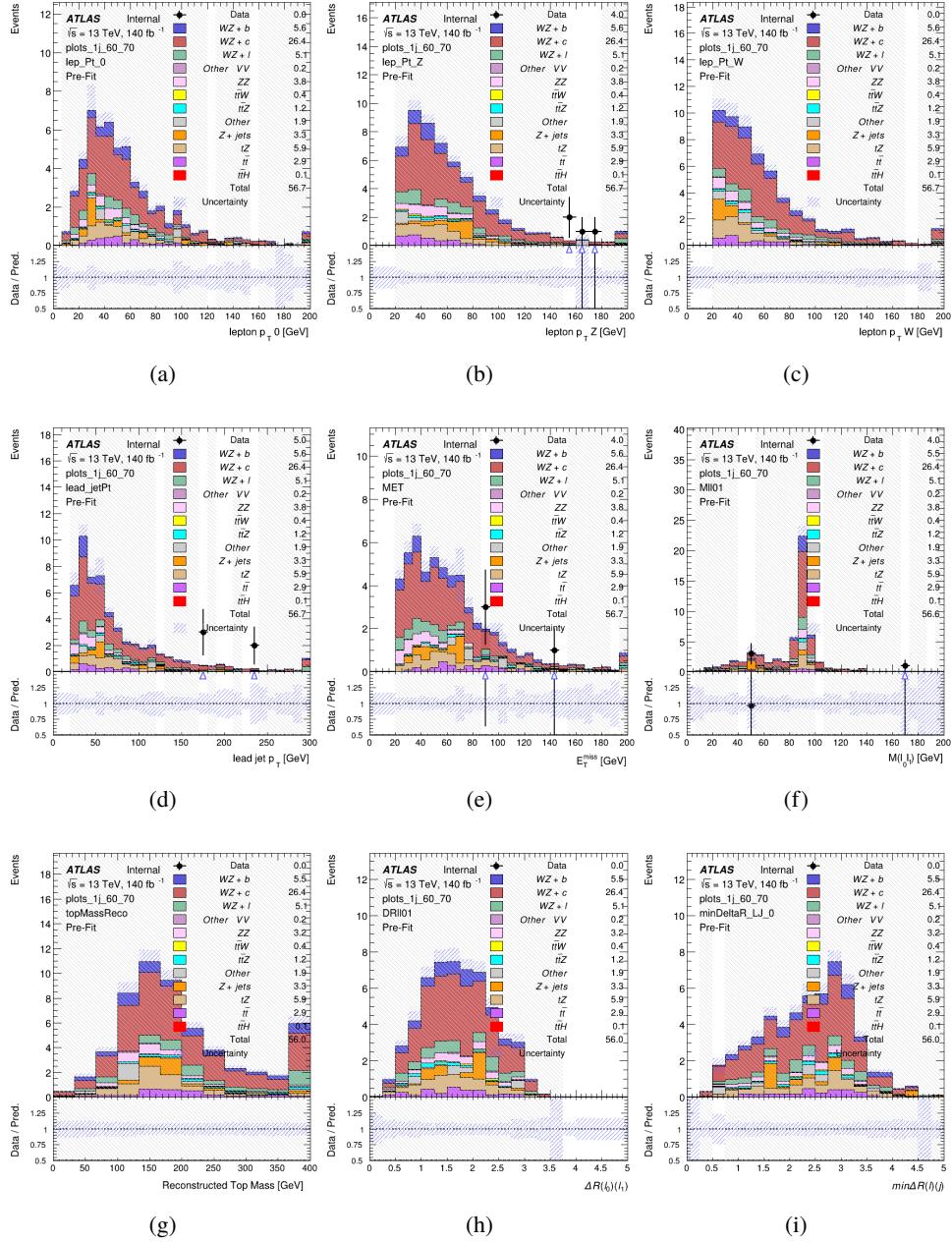


Figure 8.7: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant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

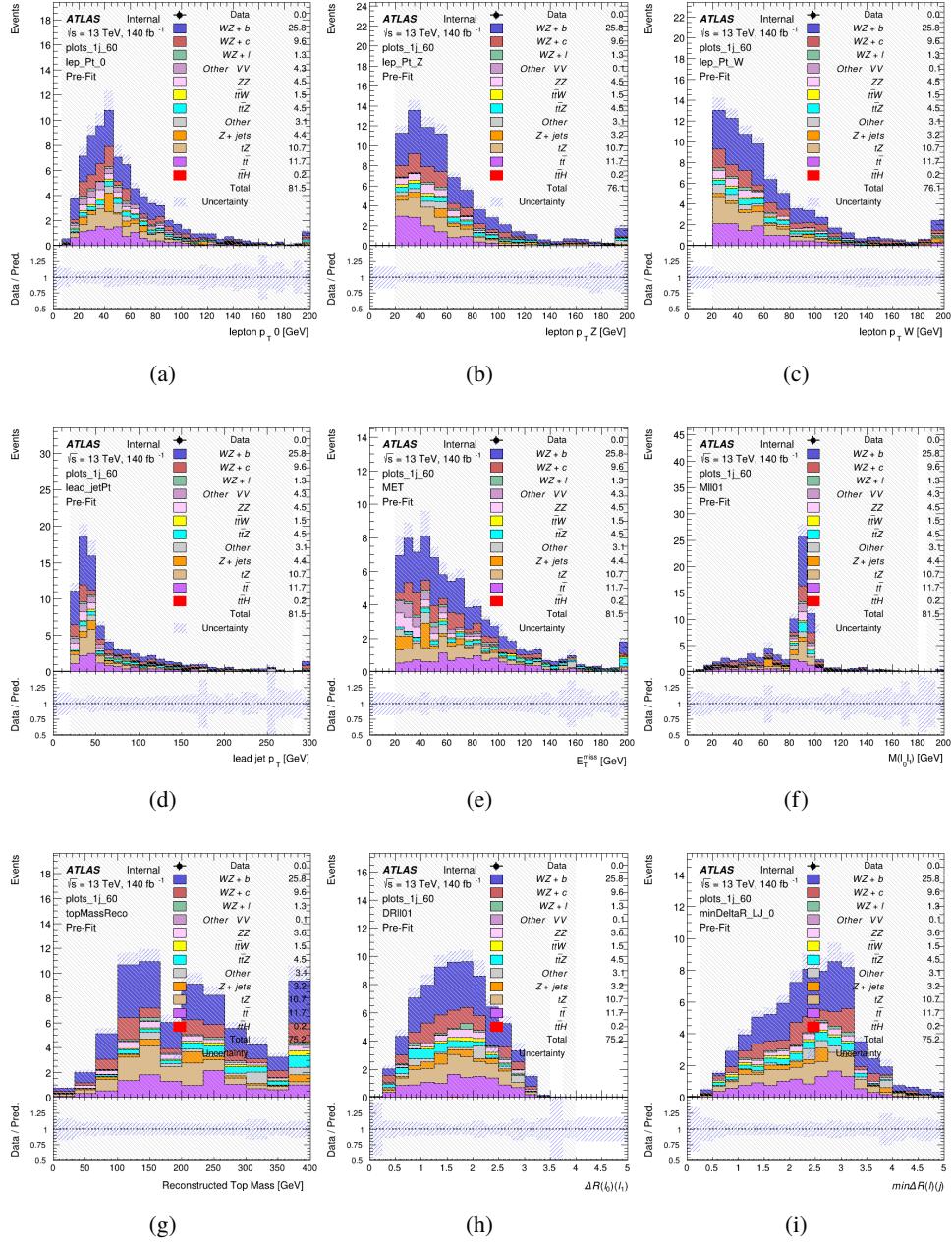


Figure 8.8: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant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

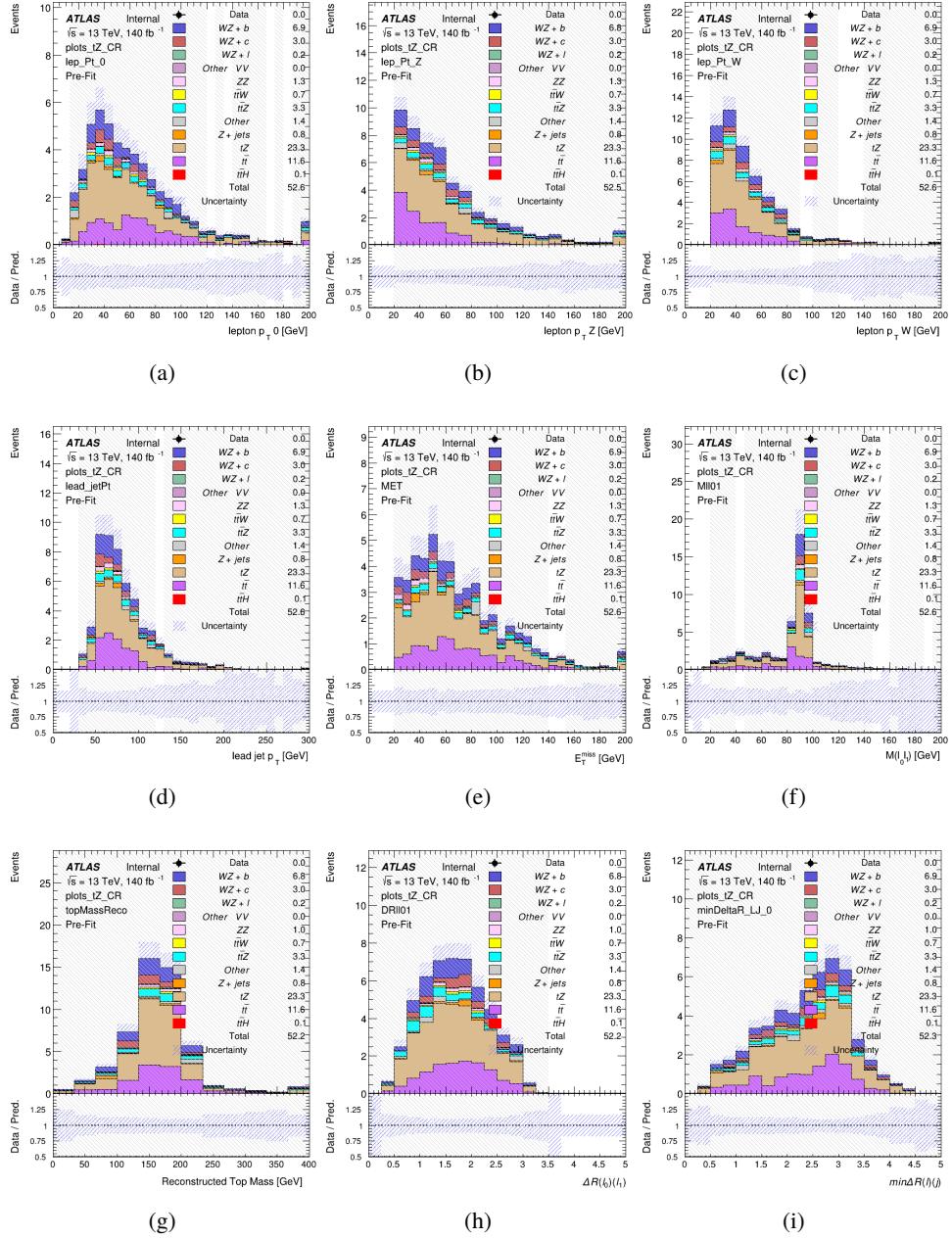


Figure 8.9: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant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

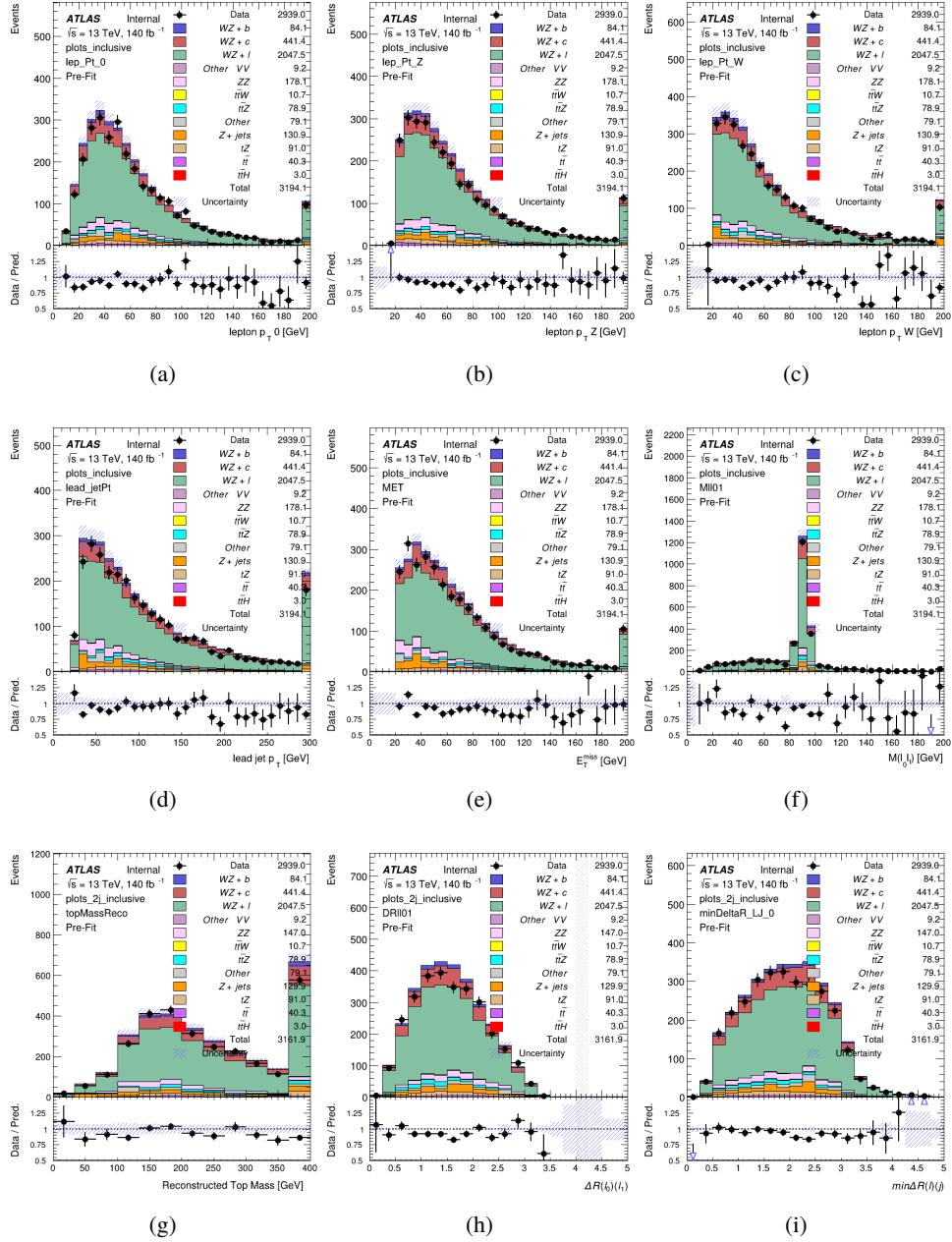


Figure 8.10: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant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

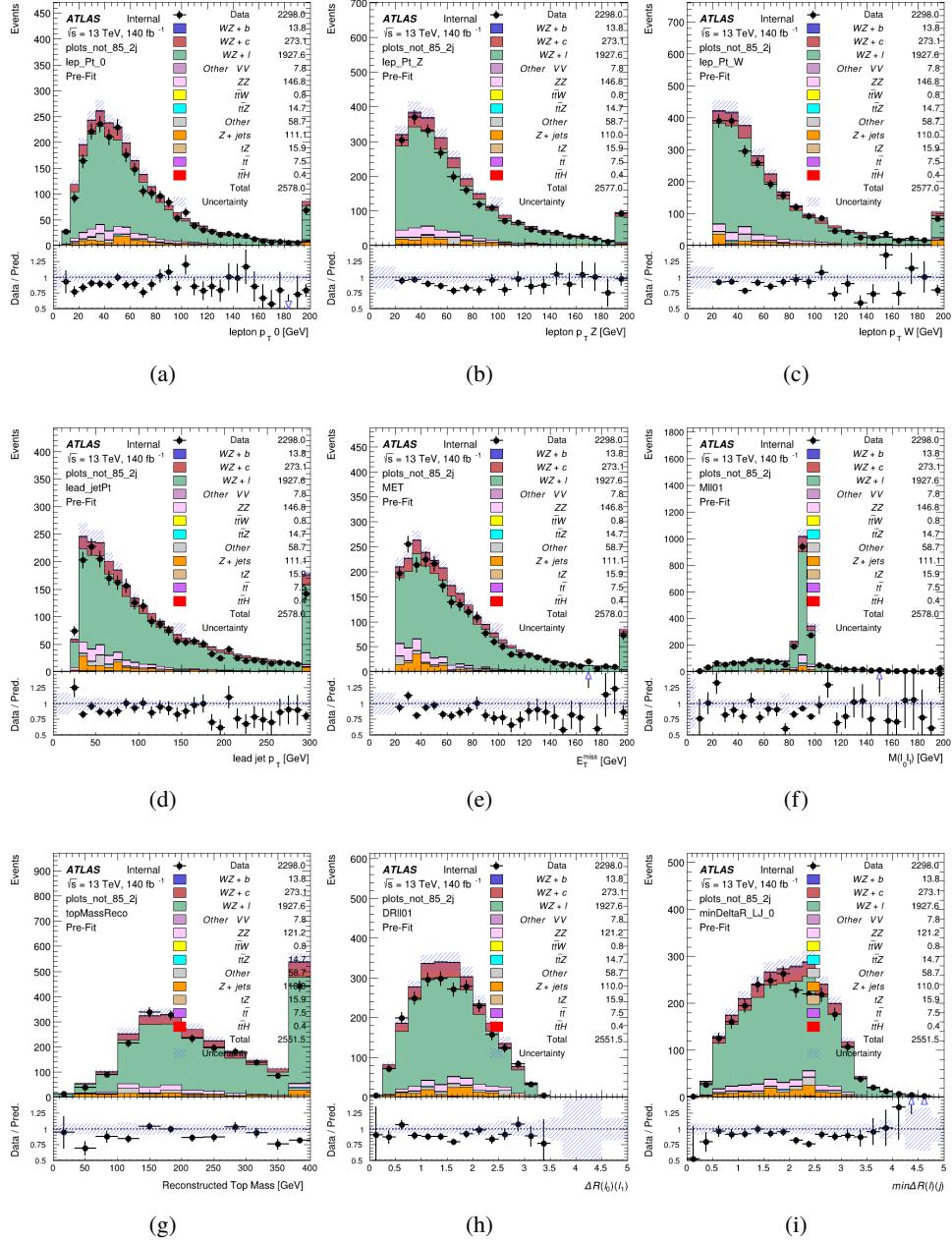
WZ Fit Region - $2j < 85\%$ WP

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

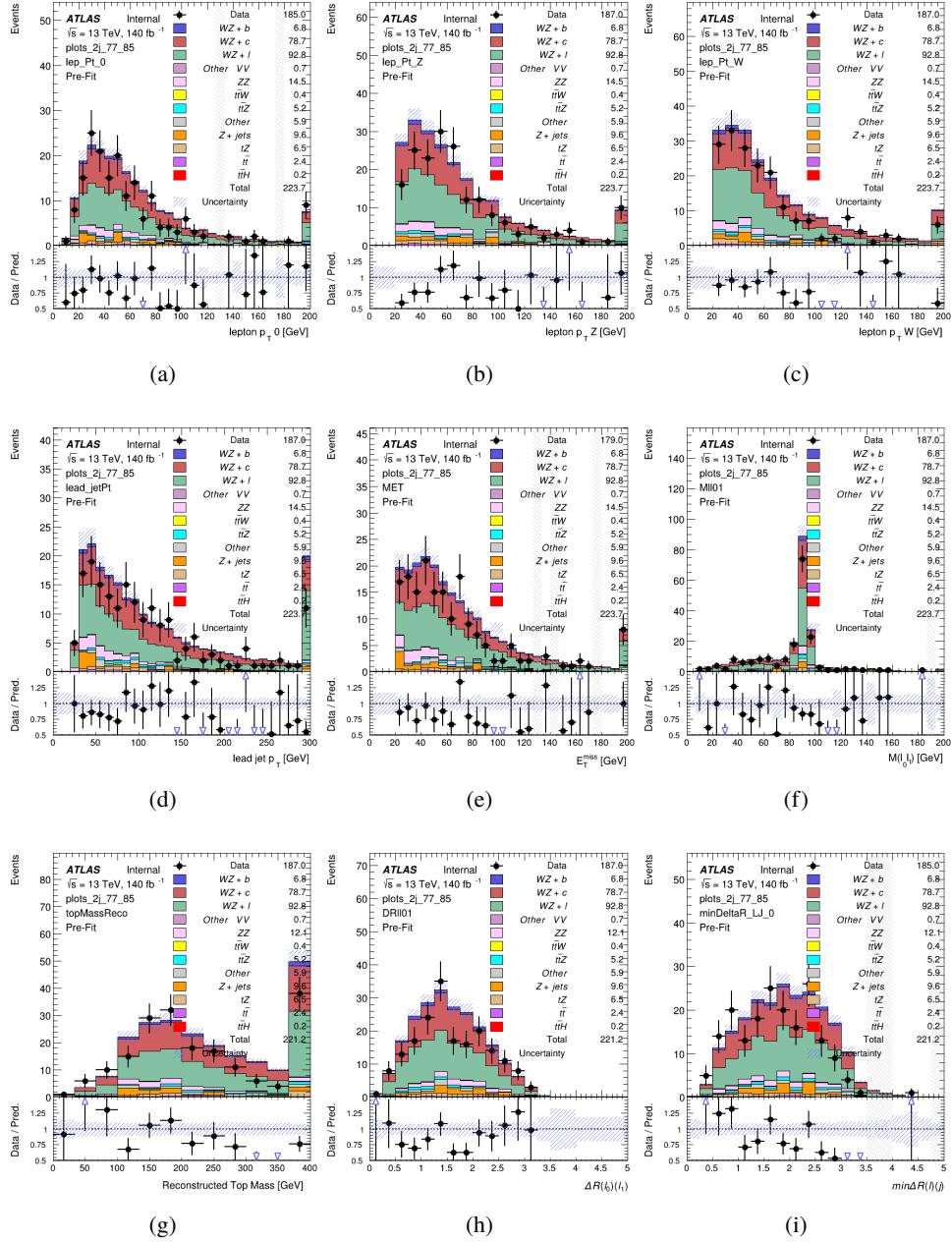


Figure 8.12: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant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

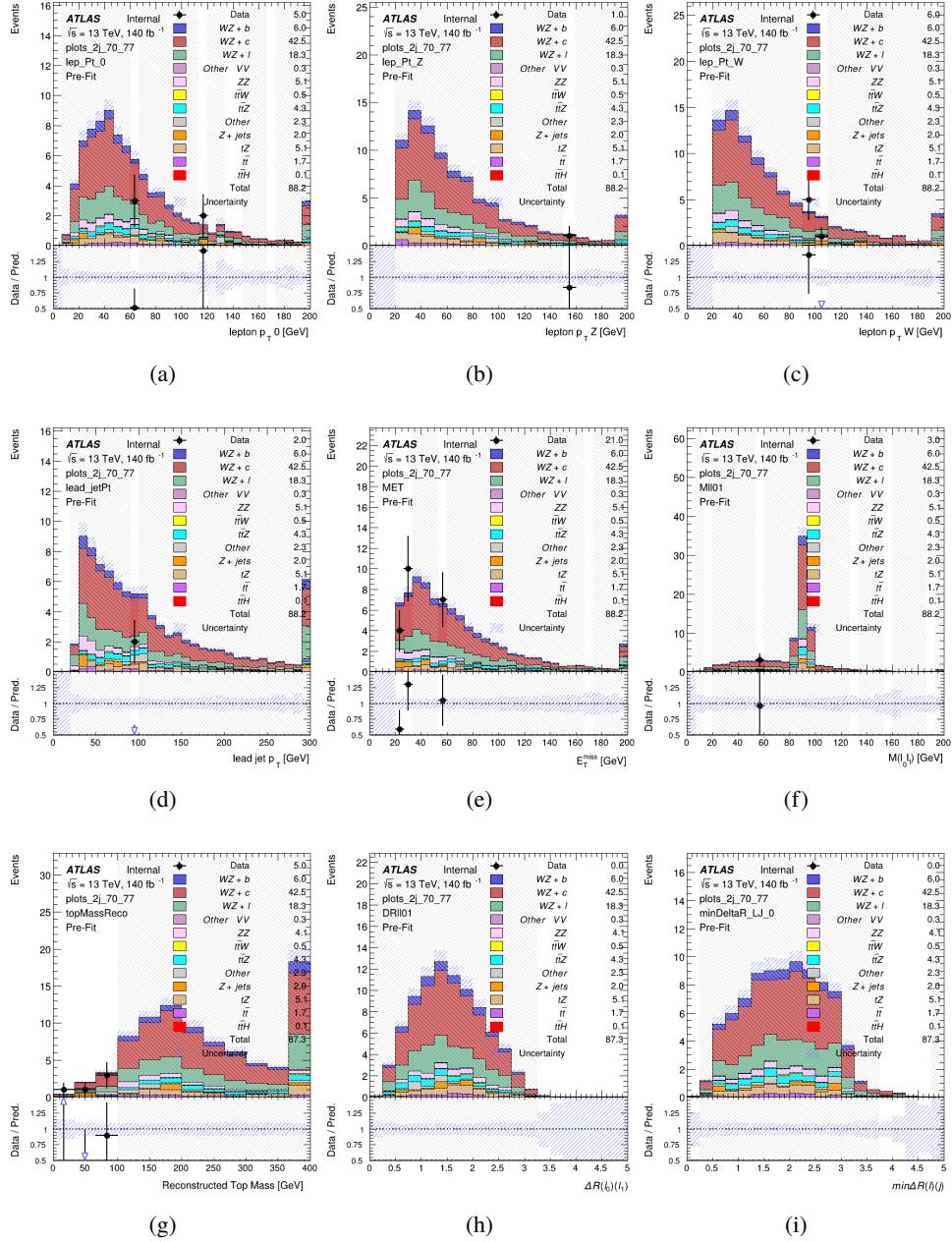


Figure 8.13: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant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

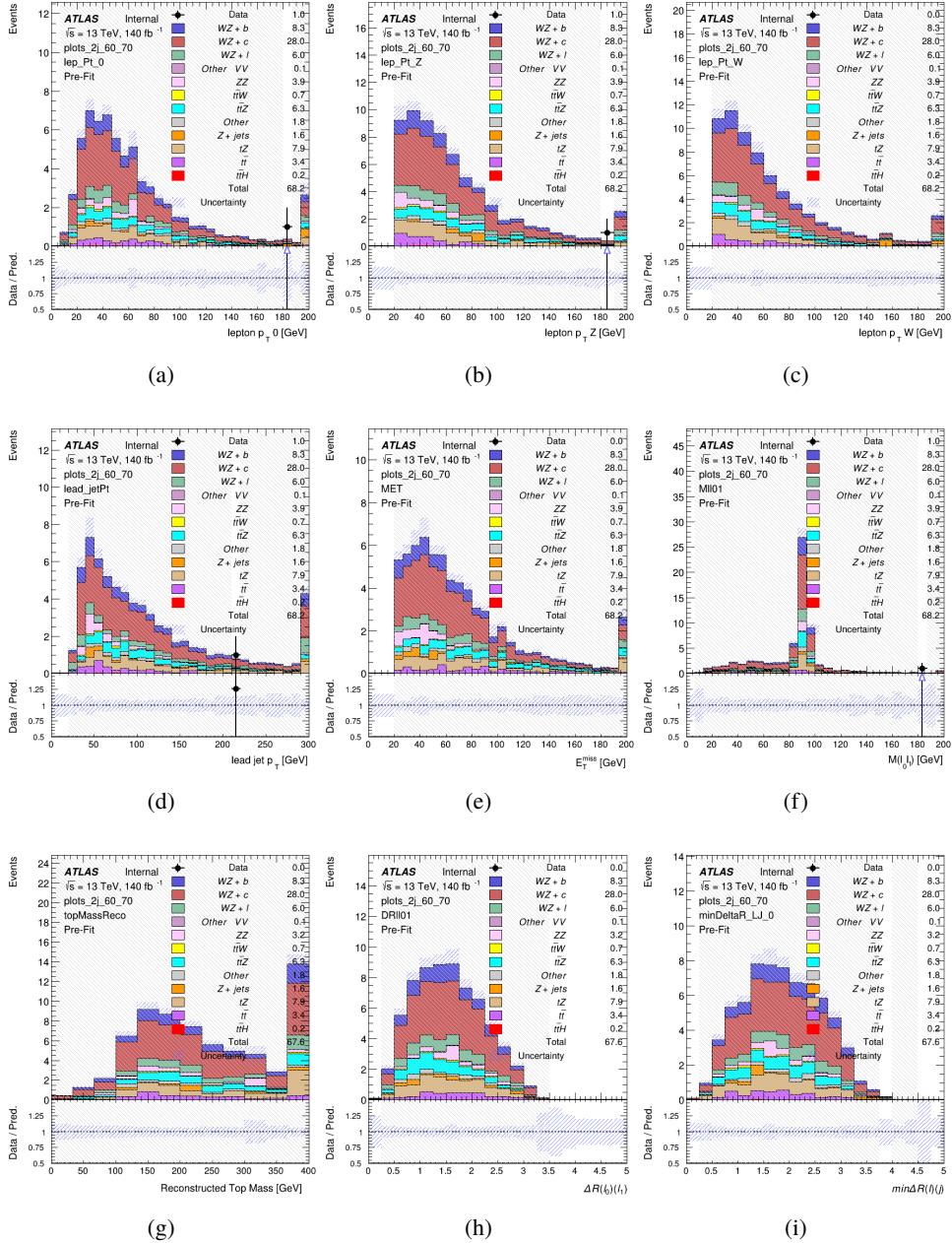


Figure 8.14: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant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

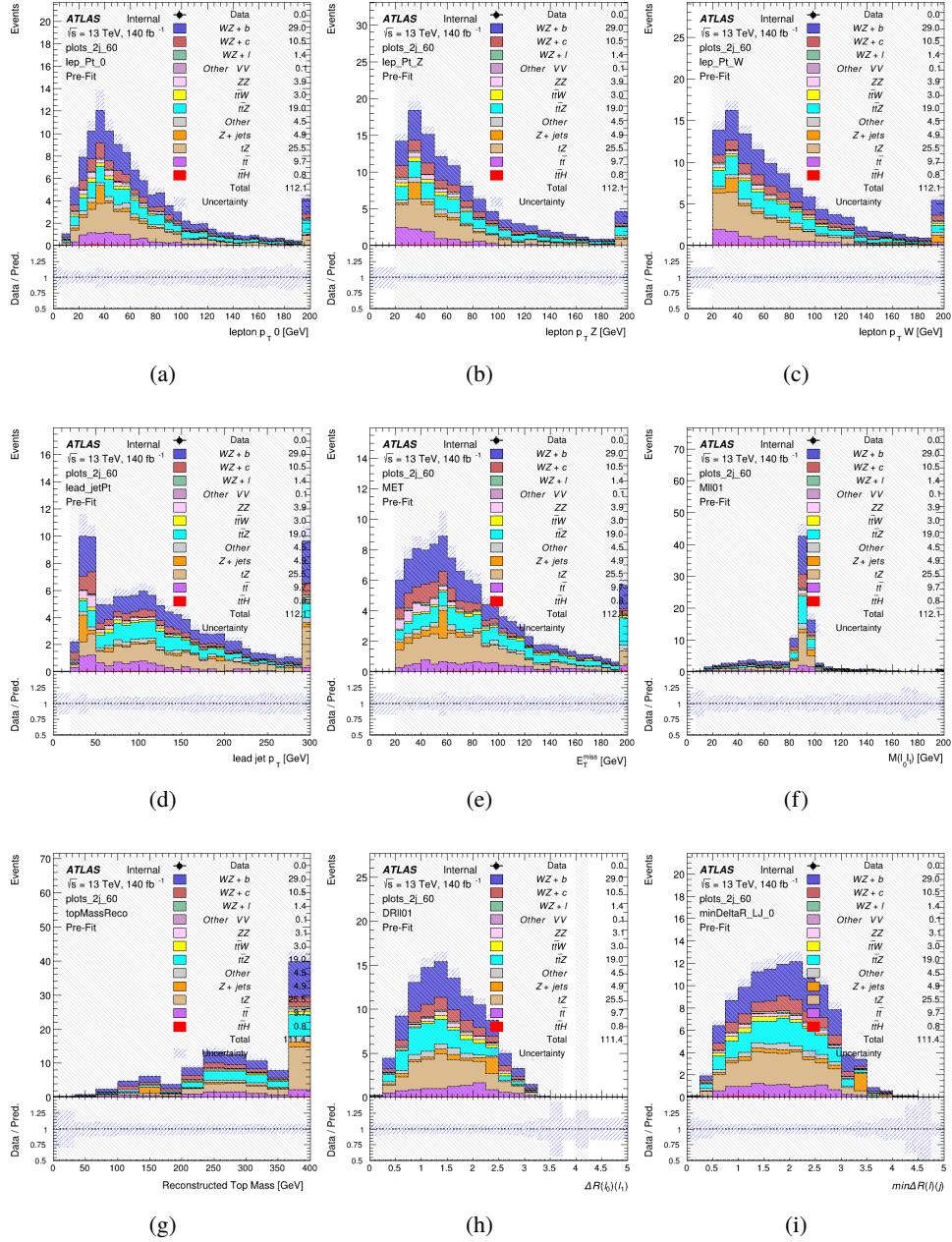


Figure 8.15: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant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

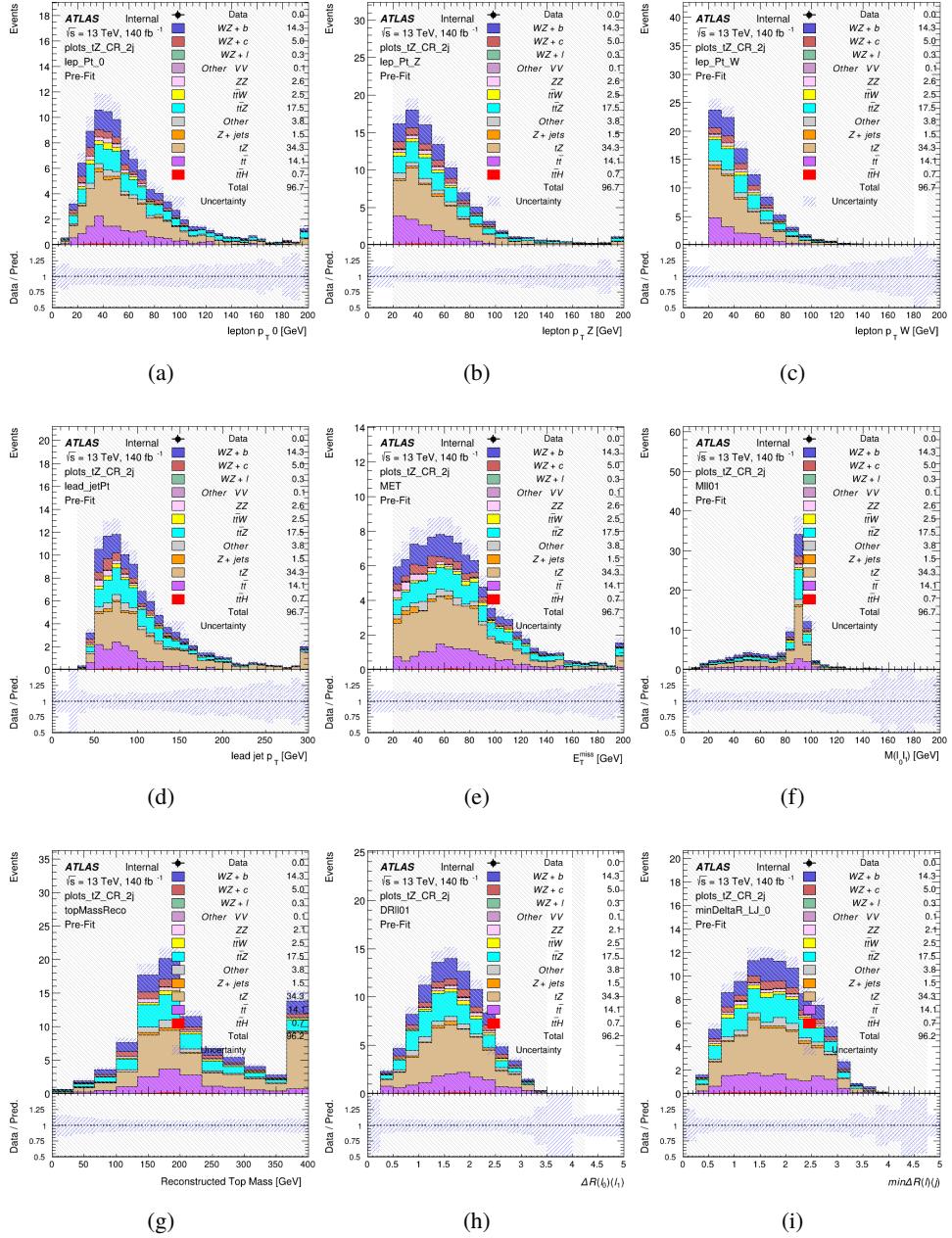


Figure 8.16: Comparisons between the data and MC distributions in the preselection region for (a) the p_T of the opposite sign lepton, (b) the p_T of the other lepton from the Z candidate, (c) the p_T of the lepton from the W candidate, (d) the leading jet p_T , (e) the E_T^{miss} , and (f) the invariant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

707 **8.3 Non-Prompt Lepton Estimation**

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

712 The modelling in the Z+jets and $t\bar{t}$ CRs is further validated for each of the pseudo-
 713 continuous b-tag regions used in the analysis. The relevant lepton p_T spectrum in each b-tag
 714 region is shown in Appendix B.1 for these CRs after the correction factors derived below have
 715 been applied.

716 **8.3.1 $t\bar{t}$ Validation**

717 $t\bar{t}$ events can produce two prompt leptons from the decay of each of the tops. These top decays
 718 produce two b-quarks, the decay of which can produce additional non-prompt leptons, which
 719 occasionally pass the event preselection. In order to validate that the Monte Carlo accurately
 720 simulates this process accurately, the MC prediction in a non-prompt $t\bar{t}$ enriched validation
 721 region is compared to data.

722 The $t\bar{t}$ validation region is similar to the preselection region - three leptons meeting the
 723 criteria described in Section 8 are required, and the requirements on E_T^{miss} remain the same.
 724 However, the selection requiring that a lepton pair form a Z-candidate are reversed. Events

725 where the invariant mass of any two opposite sign, same flavor leptons falls within 10 GeV of
726 91.2 GeV are rejected. This ensures the $t\bar{t}$ validation region is orthogonal to the preselection
727 region.

728 Further, because the jet multiplicity of $t\bar{t}$ events tends to be higher than WZ + jets, the
729 number of jets in each event is required to be greater than 1. As b-jets are almost invariably
730 produced from top decays, at least one b-tagged jet passing the 70% DL1r WP in each event is
731 required. Various kinematic plots of this region are shown in Figure 8.17.

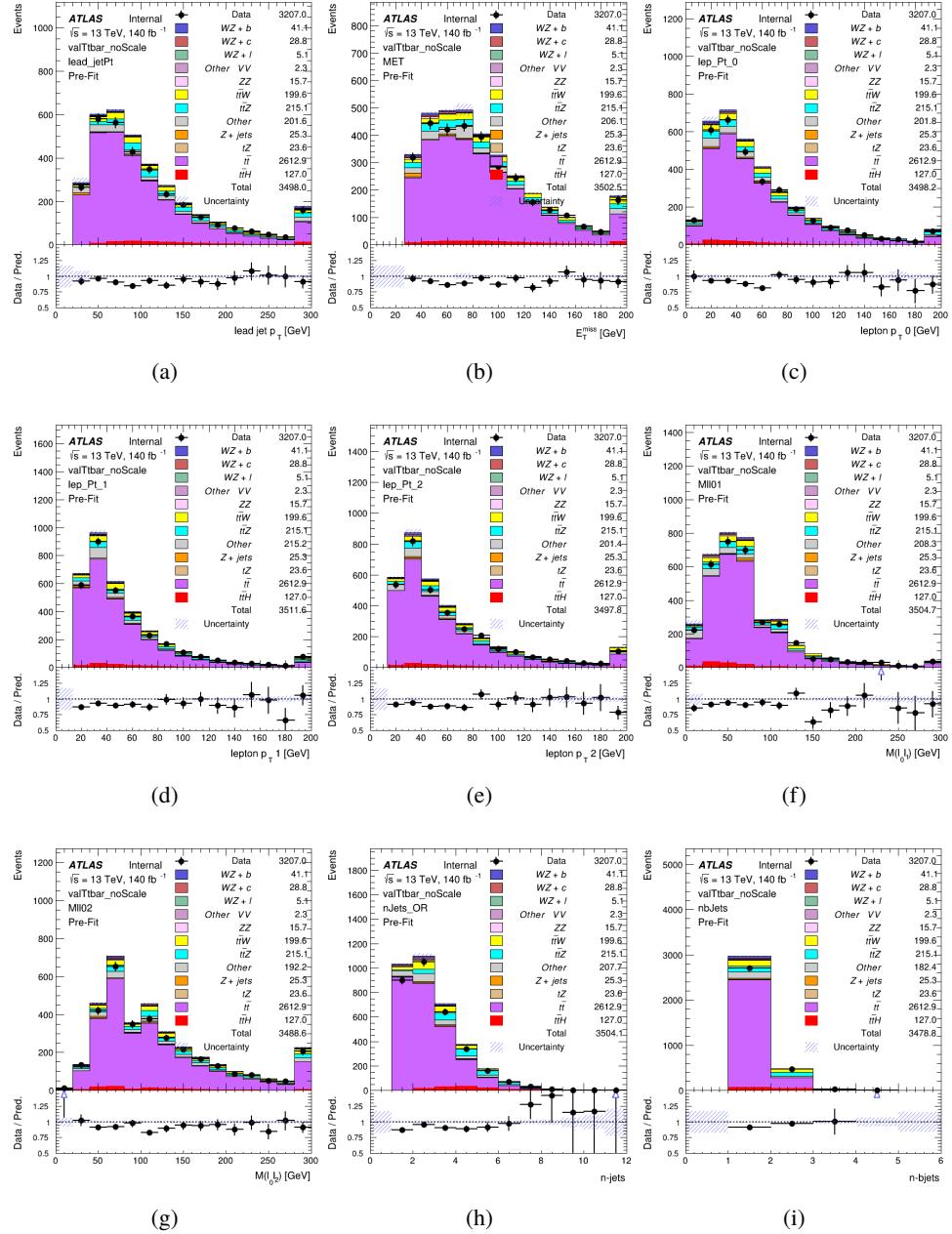


Figure 8.17: Comparisons between the data and MC distributions in the $t\bar{t}$ validation region for (a) the p_T of the leading jet, (b) the missing transverse energy, (c) the p_T of lepton 0, (d) p_T of lepton 1, (e) p_T of lepton 2, (f) the invariant mass of leptons 0 and 1, (g) the invariant mass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

732 The shape of each distribution agrees quite well between data and MC, with a constant
733 offset between the two. This is accounted for by applying a constant correction factor of 0.9 to
734 the $t\bar{t}$ MC prediction. Plots showing the kinematics of the $t\bar{t}$ VR after this correction factor has
735 been applied are shown in Figure 8.18.

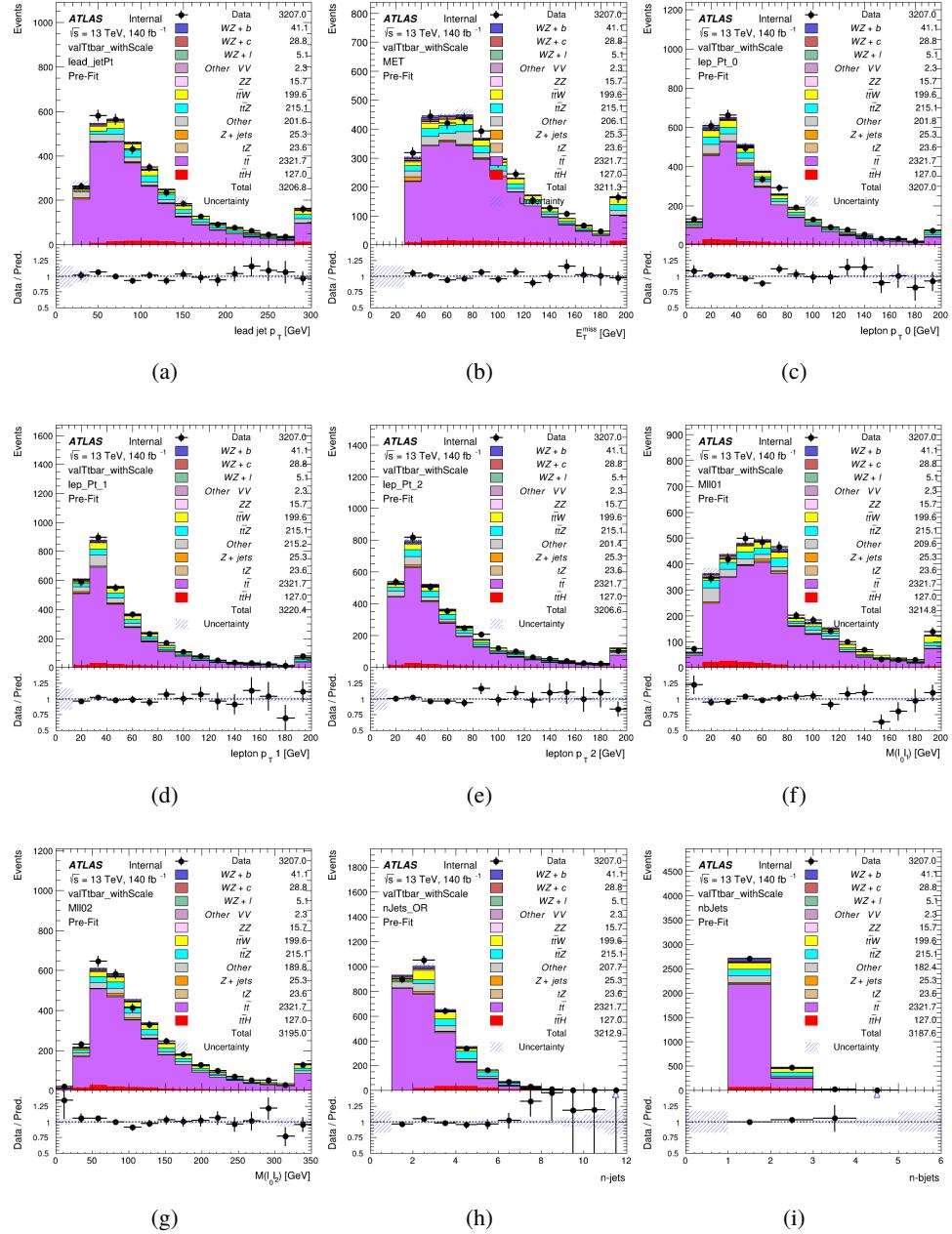


Figure 8.18: Comparisons between the data and MC distributions in the $t\bar{t}$ validation region after the correction factor has been applied for (a) the p_T of the leading jet, (b) the missing transverse energy, (c) the p_T of lepton 0, (d) p_T of lepton 1, (e) p_T of lepton 2, (f) the invariant mass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

736 The modeling is further validated by looking at the yield in the $t\bar{t}$ VR for each DL1r
 737 WP, giving a clearer correspondence to the signal regions used in the fit. For these plots, the
 738 requirement that each event contain at least one b-tagged jet is removed. Each region shown in
 739 Figure 8.19 requires one or more jets pass the listed WP, with no jets passing the next highest
 740 WP.

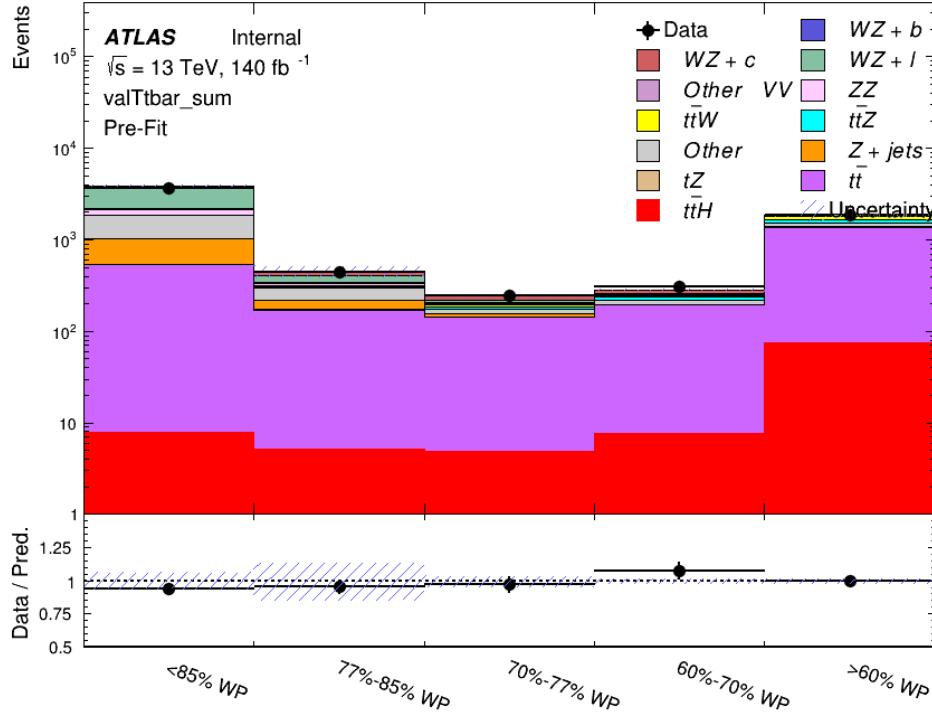


Figure 8.19: Data and MC comparisons for each DL1r WP for both 1-jet and 2-jet regions, after the $t\bar{t}$ VR selection and correction factor have been applied

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

743 8.3.2 Z+jets Validation

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

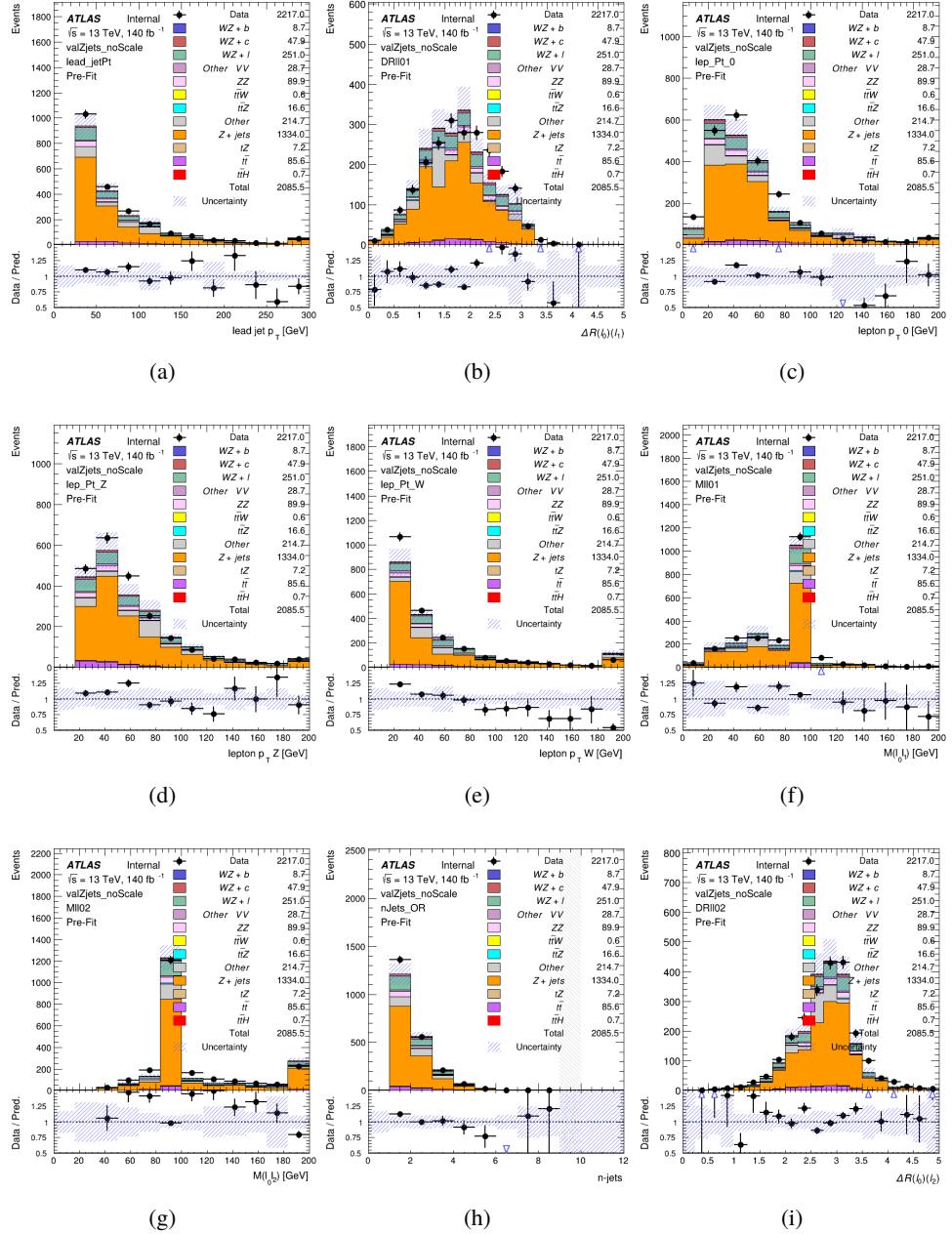


Figure 8.20: Comparisons between the data and MC distributions in the $Z+jets$ validation region for (a) the p_T of the leading jet, (b) ΔR between leptons 0 and 1, (c) the p_T of lepton 0, (d) p_T of SS lepton from the Z candidate, (e) p_T of the SS lepton from the W candidate, (f) the invariant mass of leptons 0 and 1, (g) the invariant mass of leptons 0 and 2, (h) the number of jets, (i) ΔR between leptons 0 and 2. Includes only statistical uncertainties

751 While there is general agreement between data and MC within statistical uncertainty, the
 752 shape of the p_T spectrum of the lepton from the W candidate is found to differ. As this is
 753 the lepton not included in the Z-candidate, in the case of Z+jets, this lepton is most often the
 754 non-prompt lepton. A similar effect is seen for both non-prompt muons and electrons in the
 755 Z+jets validation region, as shown in Figure 8.21.

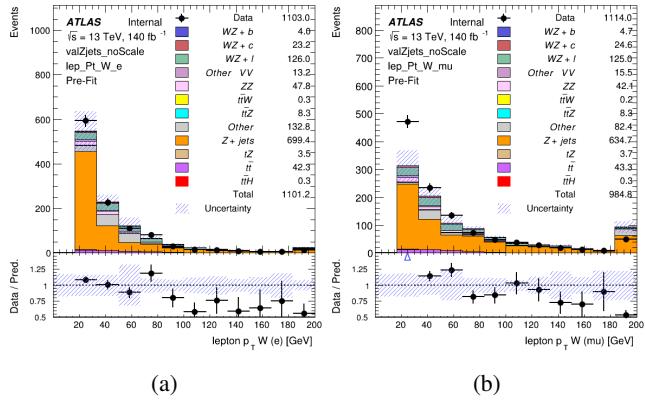


Figure 8.21: p_T spectrum of the lepton from the W candidate for (a) electrons and (b) muons

756 To account for this discrepancy, a variable correction factor is applied to Z+jets. χ^2
 757 minimization of the W lepton p_T spectrum is performed to derive a correction factor as a
 758 function of this p_T . Kinematic plots of the Z + jets validation region after this correction factor
 759 has been applied are shown in Figure 8.22.

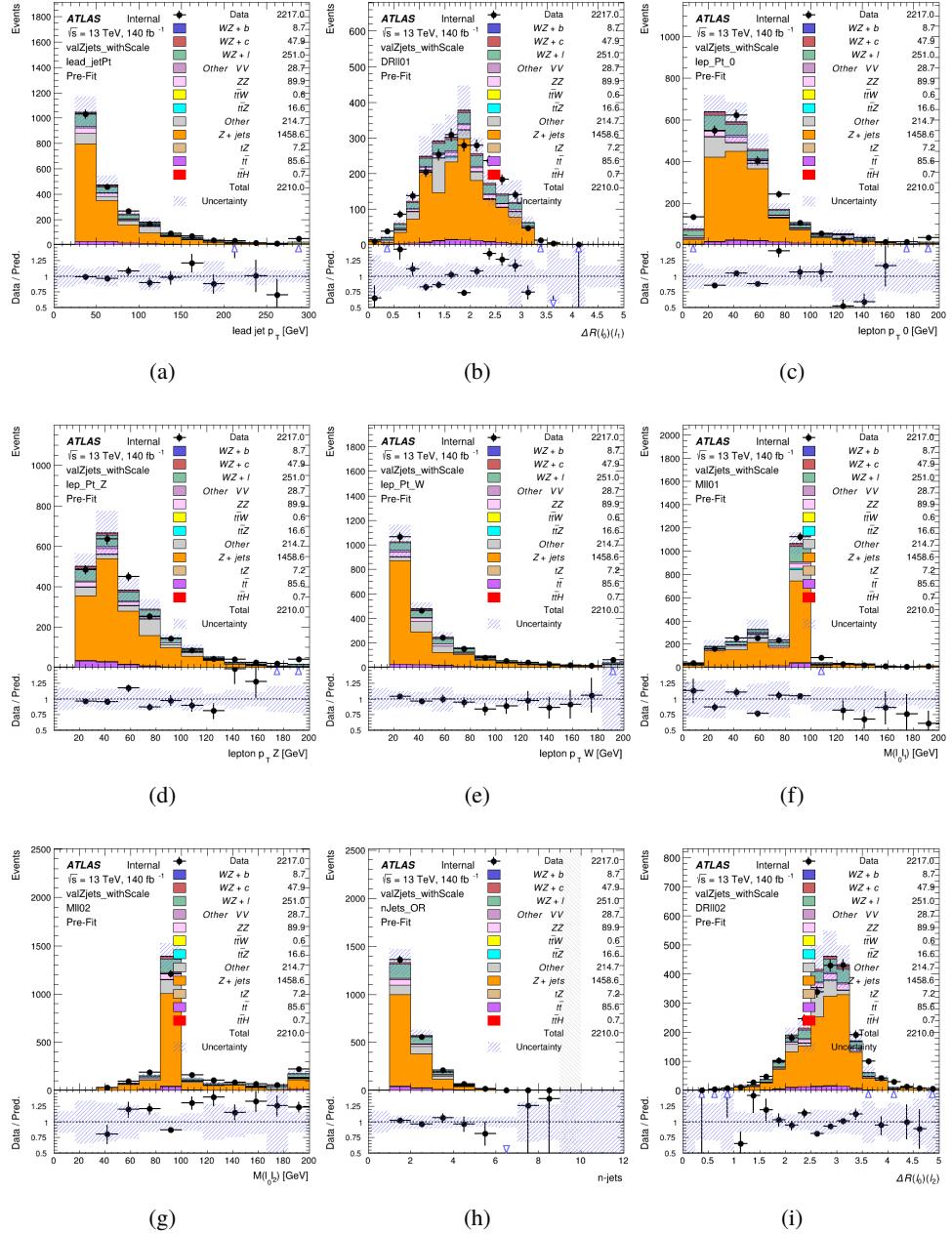


Figure 8.22: Comparisons between the data and MC distributions in the $Z + \text{jets}$ validation region after the correction factor has been applied for (a) the p_T of the leading jet, (b) ΔR between leptons 0 and 1, (c) the p_T of lepton 0, (d) p_T of SS lepton from the Z candidate, (e) p_T of the SS lepton from the W candidate, (f) the invariant mass of leptons 0 and 1, (g) the invariant mass of leptons 0 and 2, (h) the number of jets, (i) ΔR between leptons 0 and 2

760 The modeling is further validated by looking at the yield in the Z+jets VR for each DL1r
 761 WP, giving a clearer correspondence to the signal regions used in the fit. Each region shown in
 762 Figure 8.23 requires one or more jets pass the listed WP, with no jets passing the next highest
 763 WP.

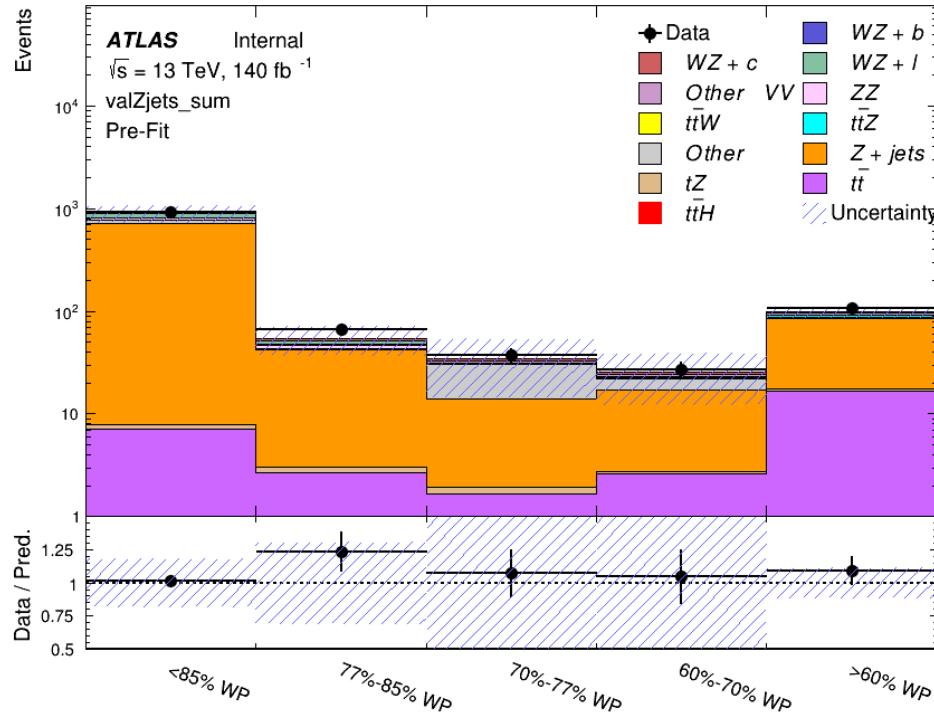


Figure 8.23: Data and MC comparisons for each DL1r WP for both 1-jet and 2-jet regions, after the Z+jets VR selection and correction factor have been applied

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

767 9 tZ Separation Multivariate Analysis

768 Because tZ produces a final state identical to signal, it represents a predominant background in
 769 the most signal enriched regions. That is, the region with one jet passing the 60% DL1r WP.
 770 Therefore, a boosted decision tree (BDT) algorithm is trained to separate WZ + heavy flavor
 771 from tZ.

772 Separation between tZ and WZ + heavy flavor is achieved in part by reconstructing the
 773 invariant mass of the top candidate, which clusters more closely to the top mass for tZ than WZ
 774 + heavy flavor. The result of this BDT is used to create a tZ enriched region in the fit, reducing
 775 its impact on the measurement of WZ + heavy flavor.

776 9.1 Top Mass Reconstruction

777 The reconstruction of the top mass follows the procedure described in detail in section 6.1 of
 778 [35]. The mass of the top quark candidate is reconstructed from the jet, the lepton not included
 779 in the Z-candidate, and a reconstructed neutrino. In the case that there is one jet in the event,
 780 there is only possible b-jet candidate. For events with two jets, the jet with the highest DL1r
 781 score is used.

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

785 This unknown is solved for by taking the combined invariant mass of the lepton and

786 neutrino to give the invariant mass of the W boson:

$$787 \quad (p_l + p_\nu)^2 = m_W^2$$

788 Expanding this out into components, this equation gives:

$$789 \quad \sqrt{p_{T\nu}^2 + p_{z\nu}^2} E_l = \frac{m_W^2 - m_l^2}{2} + p_{T\nu}(p_{lx}\cos\phi_\nu + p_{ly}\sin\phi_\nu) + p_{lz}p_{vz}$$

790 This equation gives two solutions for p_{vz} . For cases where only one of these solutions is

791 real, that is taken as the value of p_{vz} . For instances with two real solutions, the one which is

792 shown to be correct in the largest fraction of simulations is taken. For cases when no real solution

793 is found, often because of detector effects, the value of E_T^{miss} is varied in decreasing increments

794 of 100 MeV until a real solution is found.

795 The reconstructed top mass distribution for tZ and WZ + b can be seen in figure 9.1.

796 9.2 tZ BDT

797 A Boosted Decision Tree (BDT), specifically XGBoost [36], is used to provide separation between

798 tZ and WZ+b. The following kinematic variables are used as inputs:

- 799 • The invariant mass of the reconstructed top candidate

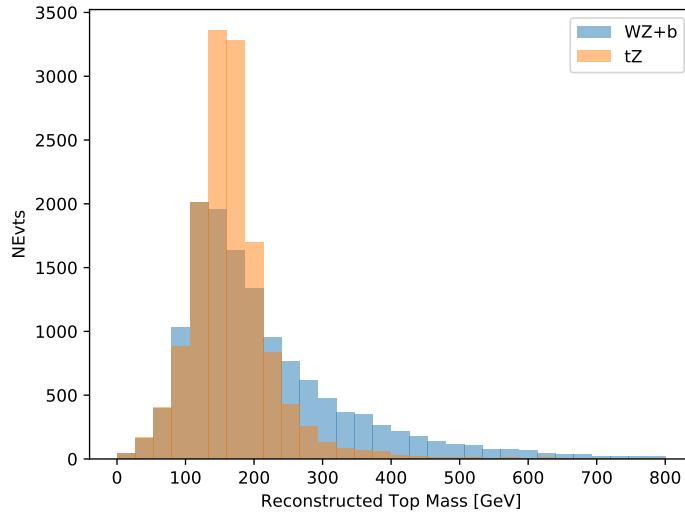


Figure 9.1: Reconstructed top mass distributions for tZ and WZ + b, measured in MeV.

- 800 • p_T of each of the leptons, jet
 - 801 • The invariant mass of each combination of lepton pairs, $M(l\bar{l})$
 - 802 • E_T^{miss}
 - 803 • Distance between each combination of leptons, $\Delta R(l\bar{l})$
 - 804 • Distance between each lepton and the jet, $\Delta R(lj)$
- 805 The training samples included only events meeting the requirements of the 1-jet, >60%
- 806 region, i.e. passing all the selection described in section 8 and having exactly one jet which
- 807 passes the tightest (60%) DL1r working point.

808 The distributions of a few of these features for both signal and background is shown in
 809 figure 9.2.

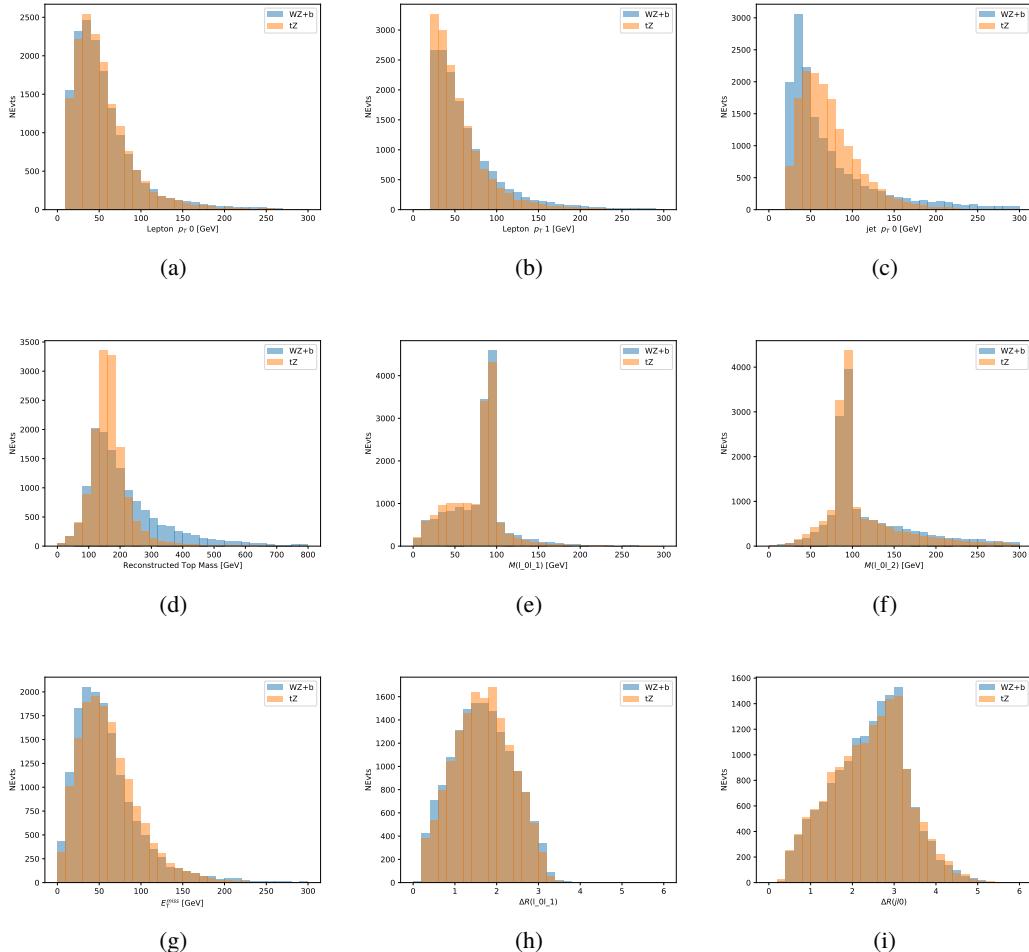


Figure 9.2: Distribution of input features of the BDT for signal (WZ) and background (tZ). Both are scaled to an equal number of events. (a), (b) and (c) show the p_T of lepton 0, lepton 1, and the jet, (d) show the reconstructed top mass, (e) and (f) show the invariant mass of leptons 0 and 1, leptons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

810 A sample of 20,000 background (tZ) and signal (WZ+b) Monte Carlo events are used to
 811 train the BDT. And additional 5,000 events are reserved for testing the model, in order to prevent

812 over-fitting. A total of 750 decision trees with a maximum depth of 6 branches are used to build
 813 the model. These parameters are chosen empirically, by training several models with different
 814 parameters and selecting the one that gave the best separation for the test sample.

815 The results of the BDT training are shown in figure 9.3. The output scores for both signal
 816 and background events is shown on the left. The right shows the receiving operating characteristic
 817 (ROC) curve that results from the MVA. The ROC curve represents the background rejection
 818 as a function of signal efficiency, where each point on the curve represents a different response
 819 score.

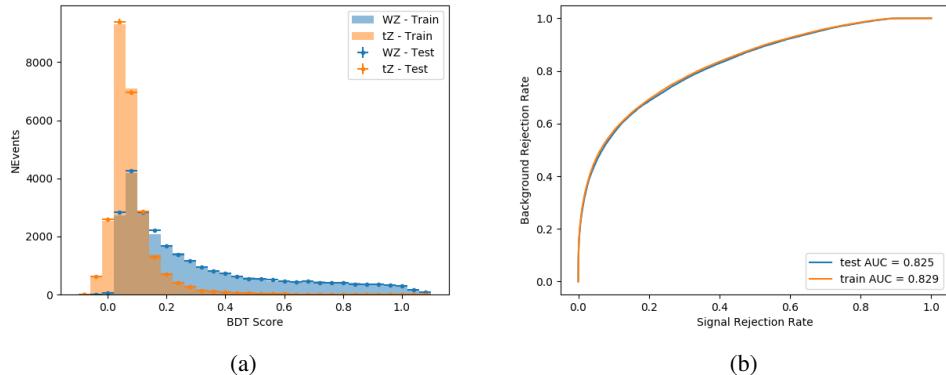


Figure 9.3: Distribution of the BDT response for signal and background events on the left, the ROC curve for the BDT on the right.

820 The relative important of each input feature in the model, measured by how often they
 821 appeared in the decision trees, is shown in figure 9.4.

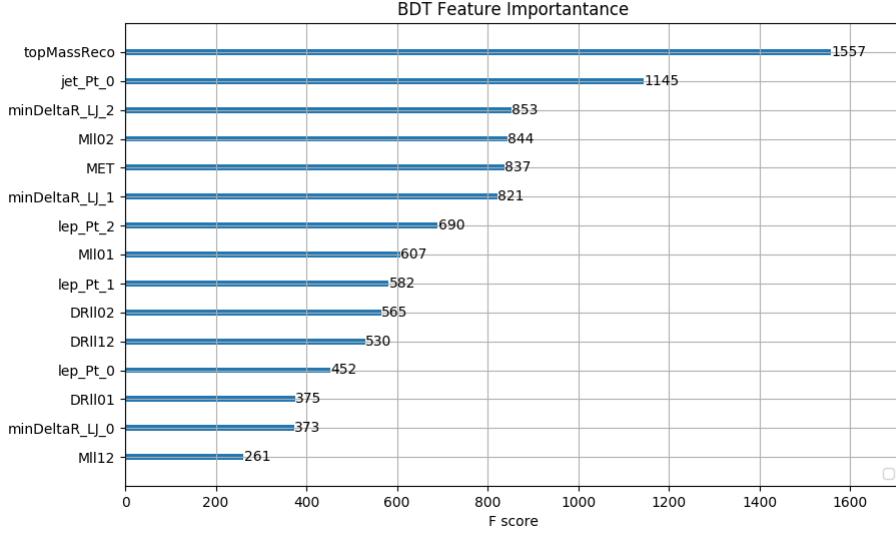


Figure 9.4: Relative importance of each input feature in the model.

822 These results suggest that some amount of separation can be achieved between these two
 823 processes, with a high BDT score selecting a set of events that is pure in $WZ + b$. A BDT score
 824 of 0.12 is selected as a cutoff, where events with scores higher than this form a signal enriched
 825 region, and events with scores lower than this form a $t\bar{Z}$ control region. This cutoff is selected by
 826 varying the value of this cutoff in stat-only Asimov fits, and selecting the value that minimizes
 827 the statistical uncertainty on $WZ + b$. A working point of 0.12 produces a background rejection
 828 rate of 74%, compared to a signal acceptance of 78%.

829 9.3 tZ Interference Studies

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

835 In order to estimate the matrix level interference effects between tZ and WZ + b-jet, two
836 different sets of simulations are produced using MadGraph 5 [37] - one which simulates these
837 two processes independently, and another where they are produced simultaneously, such that
838 interference effects are present. These two sets of samples are then compared, and the difference
839 between them can be taken to represent any interference effects.

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

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

847 The kinematics of these samples after the selection has been applied are shown below:

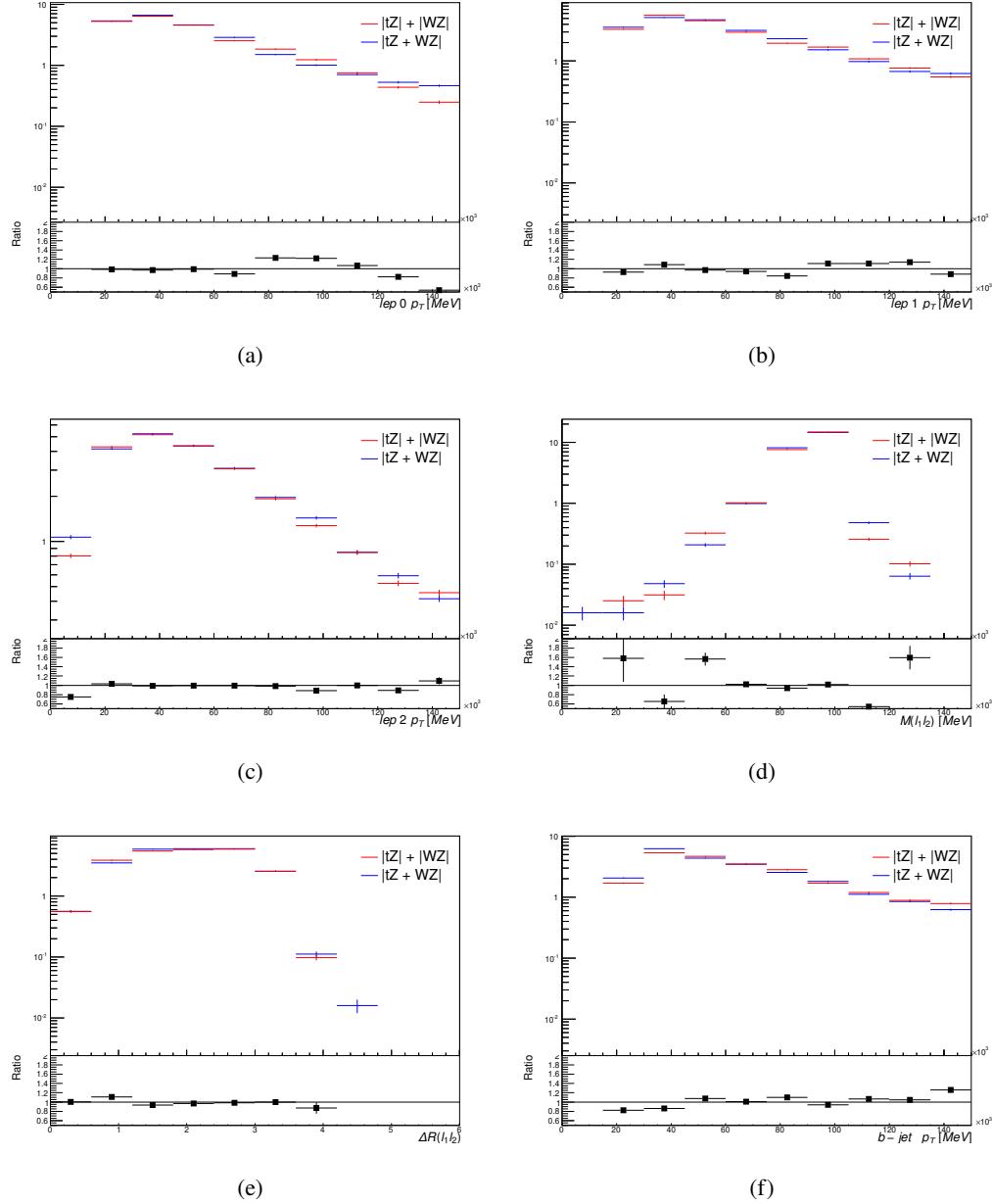


Figure 9.5: Comparisons between (a) the p_T of the lepton from the W, (b) and (c) show the p_T of the lepton from the Z, (d) the invariant mass of the Z-candidate, (e) ΔR of the leptons from the Z, and (f) the p_T of the b-jet, for WZ and tZ events generated with interference effects (blue) and without interference effects (red).

849 ences in the kinematic distributions are seen. It is therefore concluded that interference effects
850 do not significantly impact the results.

851 **10 Systematic Uncertainties**

852 The systematic uncertainties that are considered are summarized in Table 42. These are imple-
853 mented in the fit either as a normalization factors or as a shape variation or both in the signal
854 and background estimations. The numerical impact of each of these uncertainties is outlined in
855 Section 17.

Table 10: Sources of systematic uncertainty considered in the analysis.

Systematic uncertainty	Components
Luminosity	1
Pileup reweighting	1
Physics Objects	
Electron	5
Muon	14
Prompt Lepton Veto	1
Jet energy scale	28
Jet energy resolution	8
Jet vertex fraction	1
Jet flavor tagging	131
E_T^{miss}	3
Total (Experimental)	194
Signal Modeling	
Shape modelling	3
Renormalization and factorization scales	5
nJet Migration	5
Background Modeling	
Cross section	15
Renormalization and factorization scales	12
Parton shower and hadronization model	2
Shower tune	4
Total (Signal and background modeling)	41
Total (Overall)	236

856 The uncertainty in the combined 2015–2018 integrated luminosity is 1.7% [38], obtained
 857 using the LUCID-2 detector [39] for the primary luminosity measurements.

858 The experimental uncertainties are related to the reconstruction and identification of light
 859 leptons and b-tagging of jets, and to the reconstruction of E_T^{miss} . The TOTAL electron ID
 860 correlation model is used, corresponding to 1 electron ID systematic [`ele_eff`]. Electron ID is

861 found to be a subleading systematic that is unconstrained by the fit, making it an appropriate
862 choice for this analysis.

863 The sources which contribute to the uncertainty in the jet energy scale (JES) [40] are decom-
864 posed into uncorrelated components and treated as independent sources in the analysis. The
865 CategoryReduction model is used to account for JES uncertainties, which decomposes the uncer-
866 tainties into 30 nuisance parameters included in the fit. The SimpleJER model is used to account
867 for jet energy resolution (JER) uncertainties, and 8 JER uncertainty components included as NPs
868 in the fit.

869 The uncertainties in the b-tagging efficiencies measured in dedicated calibration analyses
870 [32] are also decomposed into uncorrelated components. The large number of components for
871 b-tagging is due to the calibration of the distribution of the MVA discriminant.

872 The full list of systematic uncertainties considered in the analysis is summarized in Tables
873 [11](#), [12](#) and [13](#).

874

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 11: 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 12: 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 13: 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.

875 Theoretical uncertainties applied to MC predictions, including cross section, PDF, and scale
 876 uncertainties are taken from theory calculations, with the exception of non-prompt and diboson
 877 backgrounds. The cross-section uncertainty on tZ is taken from [41]. Derivation of the non-
 878 prompt background uncertainties, Z+jets and t \bar{t} , are explained in detail in Section 8.3. These
 879 normalization uncertainties are chosen so as to account for the complete uncertainty in the
 880 non-prompt contribution, and therefore no additional modelling uncertainties are considered for
 881 Z+jets and t \bar{t} .

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

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

887 The theory uncertainties applied to the predominate background estimates are summarized
 888 in Table 44.

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

Table 14: Summary of theoretical uncertainties for MC predictions in the analysis.

889 Due to its importance as a background, additional modelling uncertainties are considered
 890 for tZ. Alternative tZ samples with variations in scale (DSID 412064-5) and shower modelling
 891 (DSID 501046) are included as systematics..

892 The fit involves varying the overall normalization of signal templates over the regions de-
 893 scribed in Section 8.2, which are defined by the flavor and number of associated jets at truth-
 894 level. The modelling of these template shapes therefore significantly impacts the final result.
 895 Additional signal uncertainties, probing the shape of the signal templates as well as the rate of

896 migrations between the number of truth-jets and reconstructed jets, are estimated by comparing
 897 estimates from the nominal Sherpa WZ samples with alternative WZ samples generated with
 898 POWHEG+PYTHIA8 (DSID 361601). Separate systematics are included in the fit for WZ + b, WZ
 899 + c and WZ + light, where the distribution among each of the fit regions is varied based on the
 900 prediction of the Powheg sample.

901 The variations in the signal templates are shown in Figures 10.1 and 10.2. Each of these
 902 plots are normalized to unity in order to capture the relevant differences in shape.

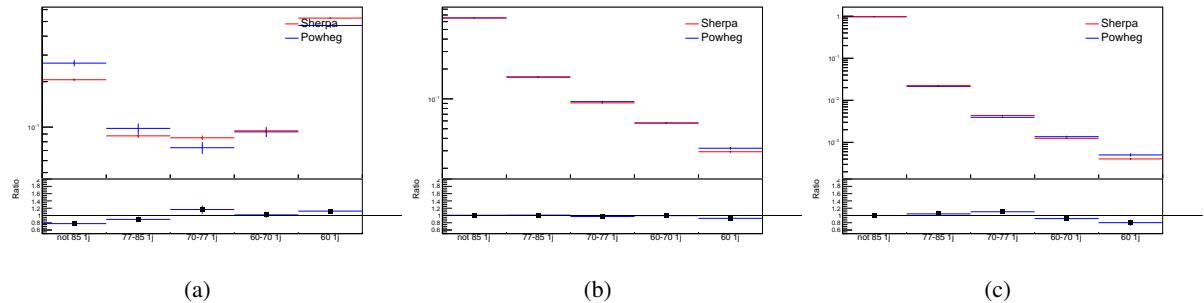


Figure 10.1: Comparison between Sherpa and Powheg predictions of the distribution of (a) WZ + b, (b) WZ + charm, and (c) WZ + light among the various b-tag WPs used in the 1-jet fit.

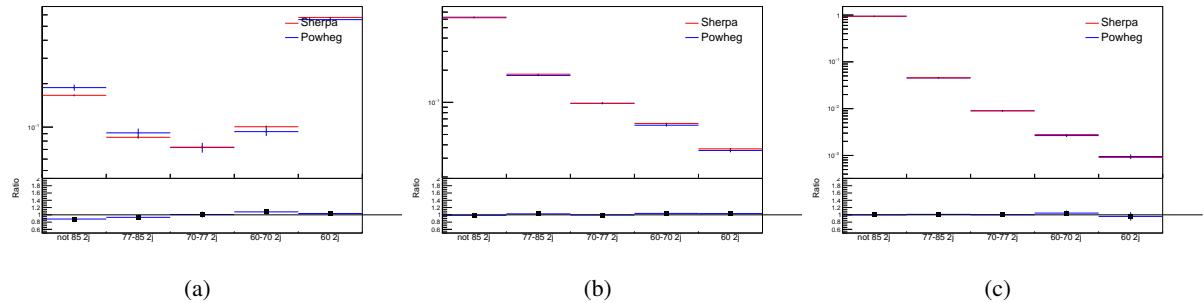


Figure 10.2: Comparison between Sherpa and Powheg predictions of the distribution of (a) WZ + b, (b) WZ + charm, and (c) WZ + light among the various b-tag WPs used in the 2-jet fit.

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

906 A similar approach is taken to account for uncertainties in migrations between the number of
 907 reco and truth jets. The fraction of events with 1 truth jet which fall into the 1 jet bin versus the
 908 2 jet bin at reco level is compared for Sherpa and Powheg. The same is done for events with 2
 909 truth jets. This comparison is shown in figure 10.3.

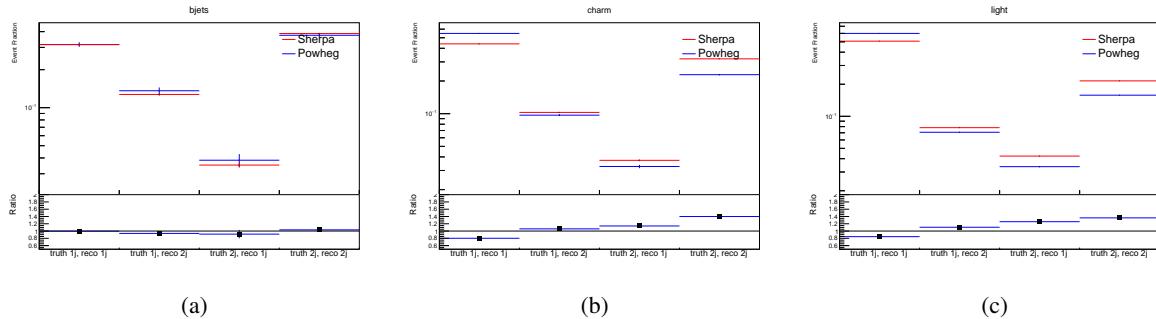


Figure 10.3: Comparison between Sherpa and Powheg predictions for truth jet migrations between the 1 and 2 jet reco bins for (a) WZ + b, (b) WZ + charm, and (c) WZ + light

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

913 Additional systematics are included to account for the uncertainty in the contamination of
 914 0 jet and 3 or more jet events (at truth level) in the 1 and 2 reco jet bins. Because these events

915 fall outside the scope of this measurement, these events are included as a background. As such,
 916 a normalization, rather than a shape, uncertainty is applied for this background.

917 The number of WZ events with 0-jets and ≥ 3 -jets in the reconstructed 1-jet and 2-jet
 918 regions are compared for Sherpa and Powheg, as seen in figure 10.4. These differences are taken
 919 as separate normalization systematics on the yield of WZ+0-jet and WZ+ ≥ 3 -jet events.

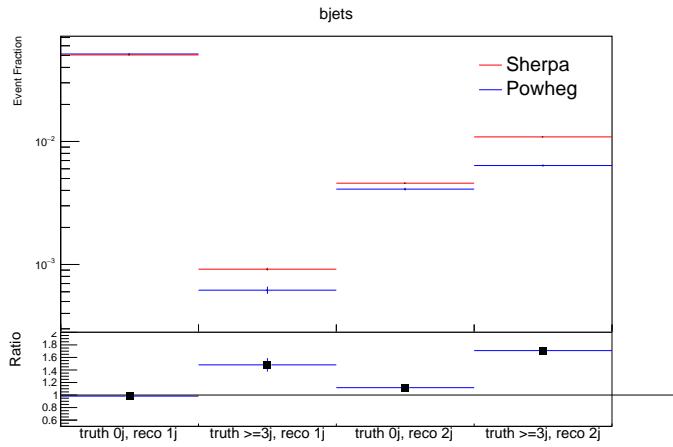


Figure 10.4: Comparison between Sherpa and Powheg predictions for 0 and ≥ 3 truth jet contributions in the 1 and 2 jet reco bins

920 11 Results

921 A separate maximum-likelihood fit is performed over the 1-jet and 2-jet fit regions in order to
 922 extract the best-fit value of the WZ + b-jet and WZ + charm jet contributions. The WZ + b,
 923 WZ + charm and WZ + light contributions are allowed to float, with the remaining background
 924 contributions are held fixed. **The current fit strategy treats the WZ + b-jet contribution as**
 925 **the parameter of interest, with the normalization of the WZ + charm and the WZ + light**

926 contributions taken as systematic uncertainties. This could however be adjusted, depending
 927 on whether it is decided the goal of the analysis should be to measure WZ+b specifically or
 928 **WZ + heavy flavor overall.** The result of the fit is used to extract the cross-section of WZ +
 929 heavy-flavor production.

930 A maximum likelihood fit to data is performed simultaneously in the regions described
 931 in Section 8. The parameters μ_{WZ+b} , $\mu_{WZ+charm}$, $\mu_{WZ+light}$, where $\mu = \sigma_{\text{observed}}/\sigma_{\text{SM}}$, are
 932 extracted from the fit.

933 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
 934 WZ + b. The fitted cross-section modifiers for WZ + charm and WZ + light are $1.00 \pm 0.17 \pm 0.15$
 935 and $1.00 \pm 0.04 \pm 0.07$, respectively.

936 The expected cross-section of WZ+b with 1 associated jet obtained from the fit is
 937 $1.74^{+0.65}_{-0.61}(\text{stat})^{+0.42}_{-0.37}(\text{sys})$ fb with an expected significance of 2.2σ . The expected cross-section
 938 of WZ + charm is measured to be $14.6 \pm 2.5(\text{stat}) \pm 2.1(\text{sys})$ fb, with a correlation of -0.23.

939 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 +
 940 b. The fitted cross-section modifiers for WZ + charm and WZ + light are $1.00 \pm 0.17 \pm 0.15$ and
 941 $1.00 \pm 0.04 \pm 0.08$, respectively.

942 The expected WZ + b cross-section in the 2-jet region is $2.46^{+0.83}_{-0.81}(\text{stat})^{+0.073}_{-0.68}(\text{sys})$ fb
 943 with an expected significance of 2.6σ . The 2-jet expected cross-section of WZ + charm is
 944 $12.7 \pm 2.2(\text{stat}) \pm 2.0(\text{sys})$ fb, and the correlation between WZ + charm and WZ + b is -0.26.

⁹⁴⁵ **11.1 1-jet Fit Results**

⁹⁴⁶ **The results of the fit are currently blinded.**

⁹⁴⁷ The pre-fit yields in each of the regions used in the fit are shown in Table 11.1, and

⁹⁴⁸ summarized in Figure 11.1.

⁹⁴⁹

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

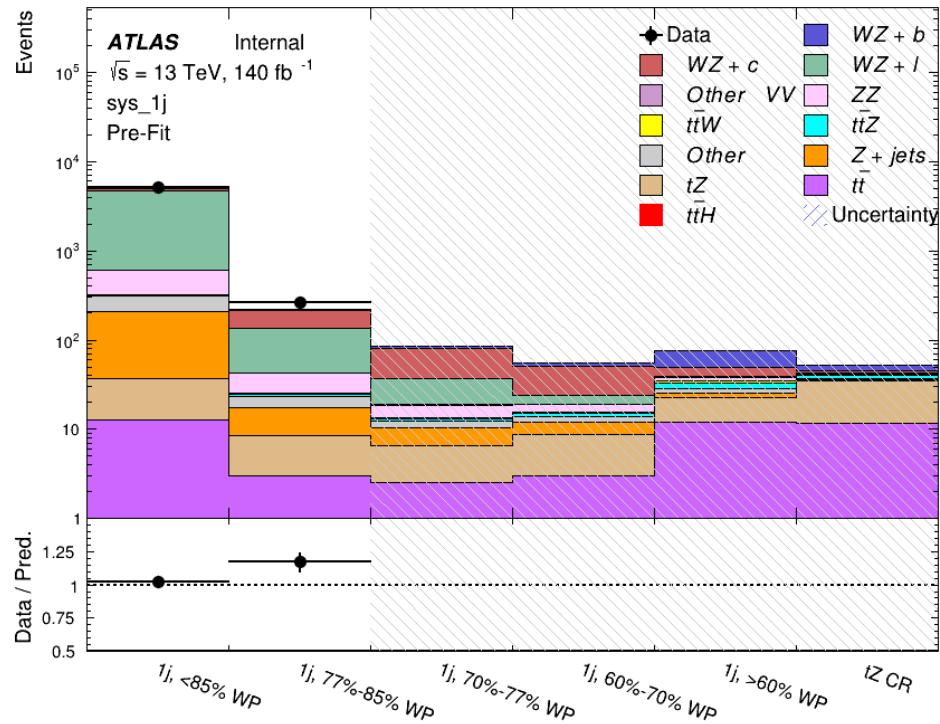


Figure 11.1: Pre-fit summary of the 1-jet fit regions.

950

The post-fit yields in each region are summarized in Figure 11.1.

951

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

952

A post-fit summary plot of the 1-jet fitted regions is shown in Figure 11.2:

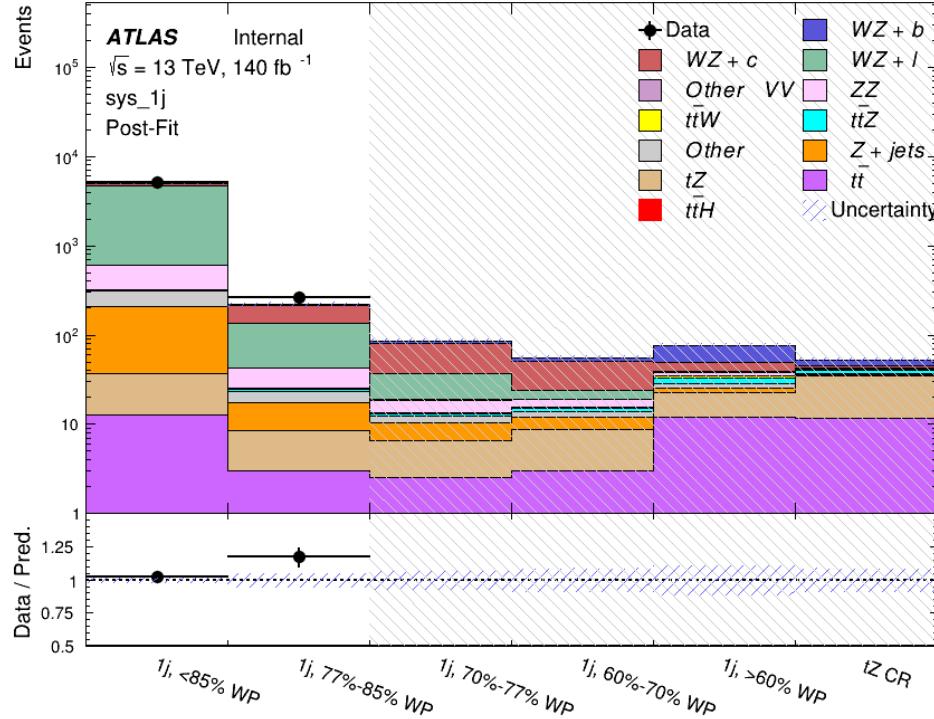


Figure 11.2: Post-fit summary of the 1-jet fit regions.

953 As described in Section 16, there are 226 systematic uncertainties that are considered
 954 as NPs in the fit. These NPs are constrained by Gaussian or log-normal probability density
 955 functions. The latter are used for normalisation factors to ensure that they are always positive.
 956 The expected number of signal and background events are functions of the likelihood. The prior
 957 for each NP is added as a penalty term, decreasing the likelihood as it is shifted away from its
 958 nominal value.

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

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

962 [17](#).

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} cross-section</bar>	-0.05	0.05
WZ cross-section - QCD scale	-0.04	0.03
Total Systematic Uncertainty	0.28	0.32

Table 17: Summary of the most significant sources of systematic uncertainty on the measurement of WZ + b with exactly one associated jet.

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

964 overall uncertainty is shown in Figure [11.3](#).

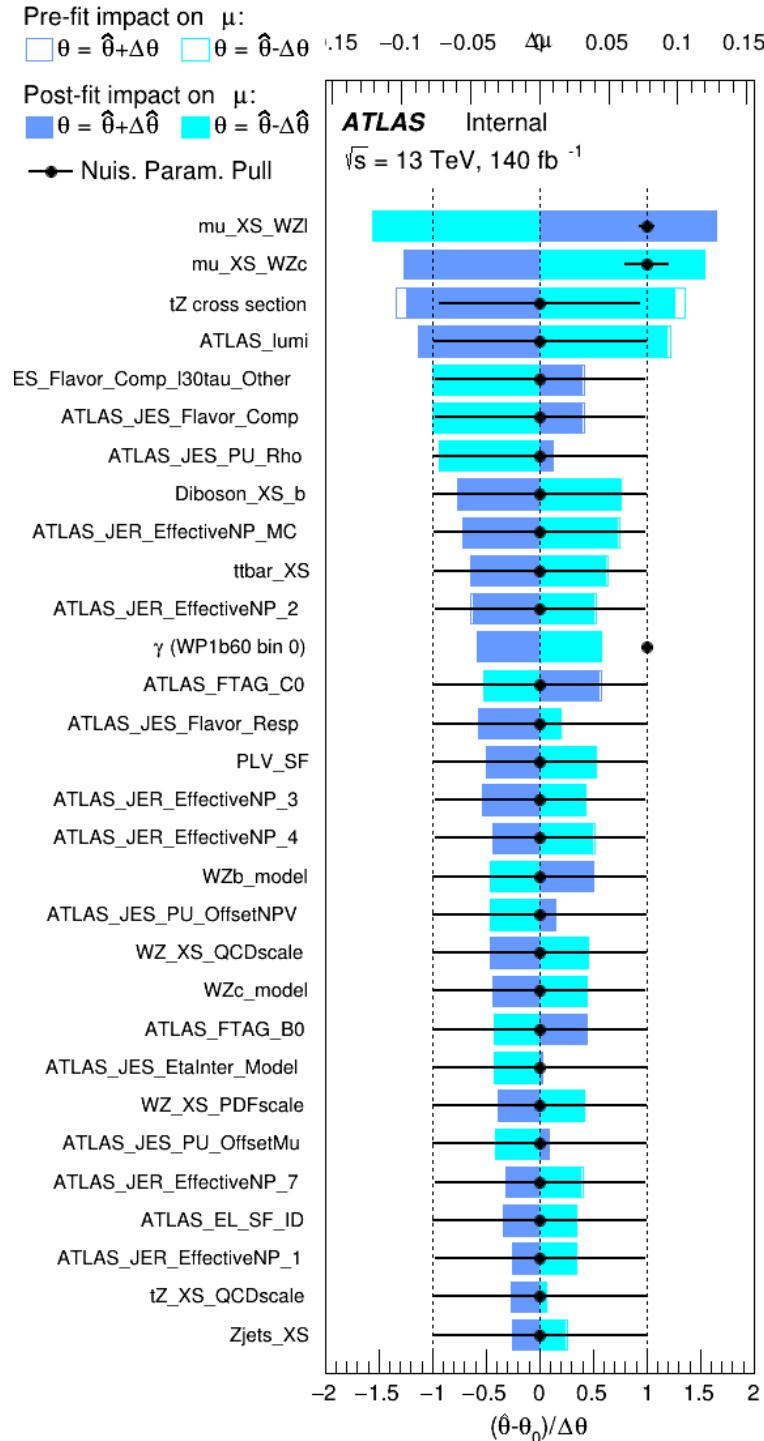


Figure 11.3: Impact of systematic uncertainties on the signal-strength of $WZ + b$ for events with exactly one jet

The large impact of the Jet Energy Scale and Jet Flavor Tagging is unsurprising, as the shape of the fit regions depends heavily on the modeling of the jets. The other major sources of uncertainty come from background modelling and cross-section uncertainty. The pie charts in Figure 11.4 show that for the modelling uncertainties that contribute most correspond to the most significant backgrounds.

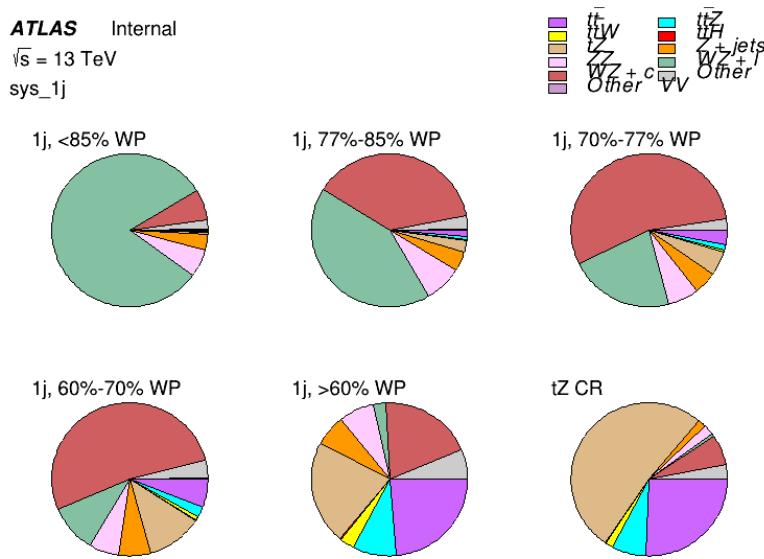


Figure 11.4: Post-fit background composition of the fit regions.

The correlations between these nuisance parameters are summarized in Figure 11.5.

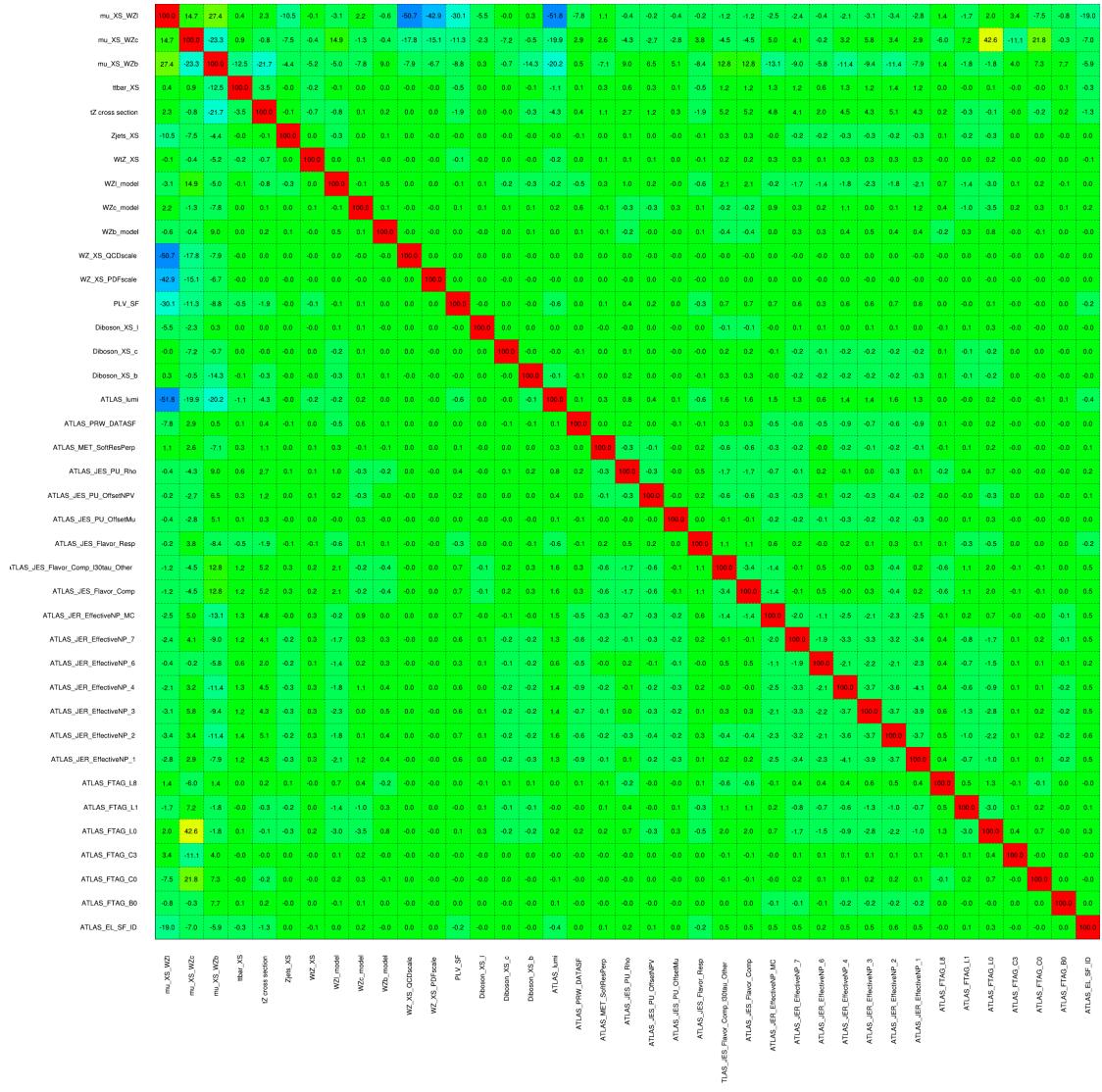


Figure 11.5: Correlations between nuisance parameters

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

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

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

982 **11.2 2-jet Fit Results**

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

984 Pre-fit yields in each of the 2-jet fit regions are shown in Figure 11.2.

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

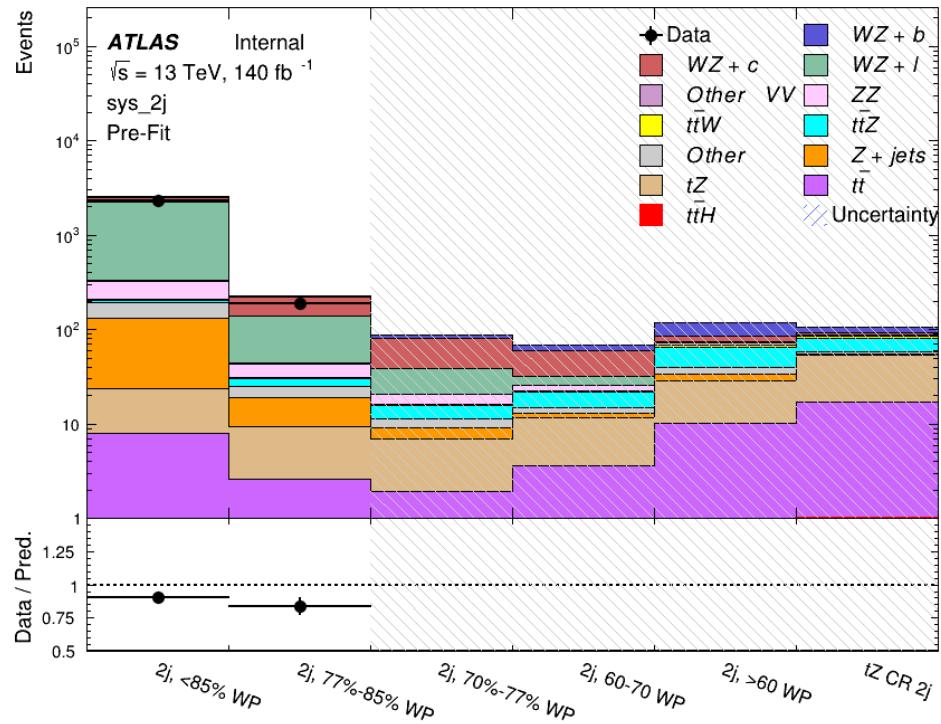


Figure 11.6: Pre-fit summary of the 2-jet fit regions.

985

The post-fit yields in each region are summarized in Figure 11.2.

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

986

A post-fit summary of the fitted regions is shown in Figure 11.7:

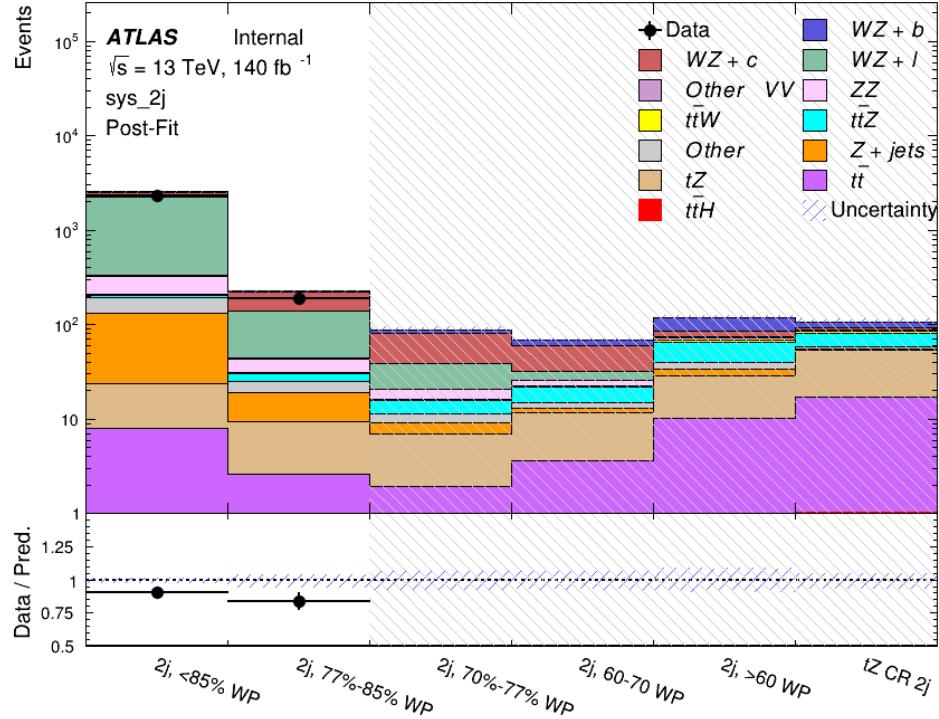


Figure 11.7: Post-fit summary of the fit over 2-jet regions.

987 The same set of systematic uncertainties consider for the 1-jet fit are included in the 2-jet
 988 fit as well. The impact of the most significant systematic uncertainties is summarized in Table
 989 20.

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

990 The ranking and impact of those nuisance parameters with the largest contribution to the
 991 overall uncertainty is shown in Figure 11.8.

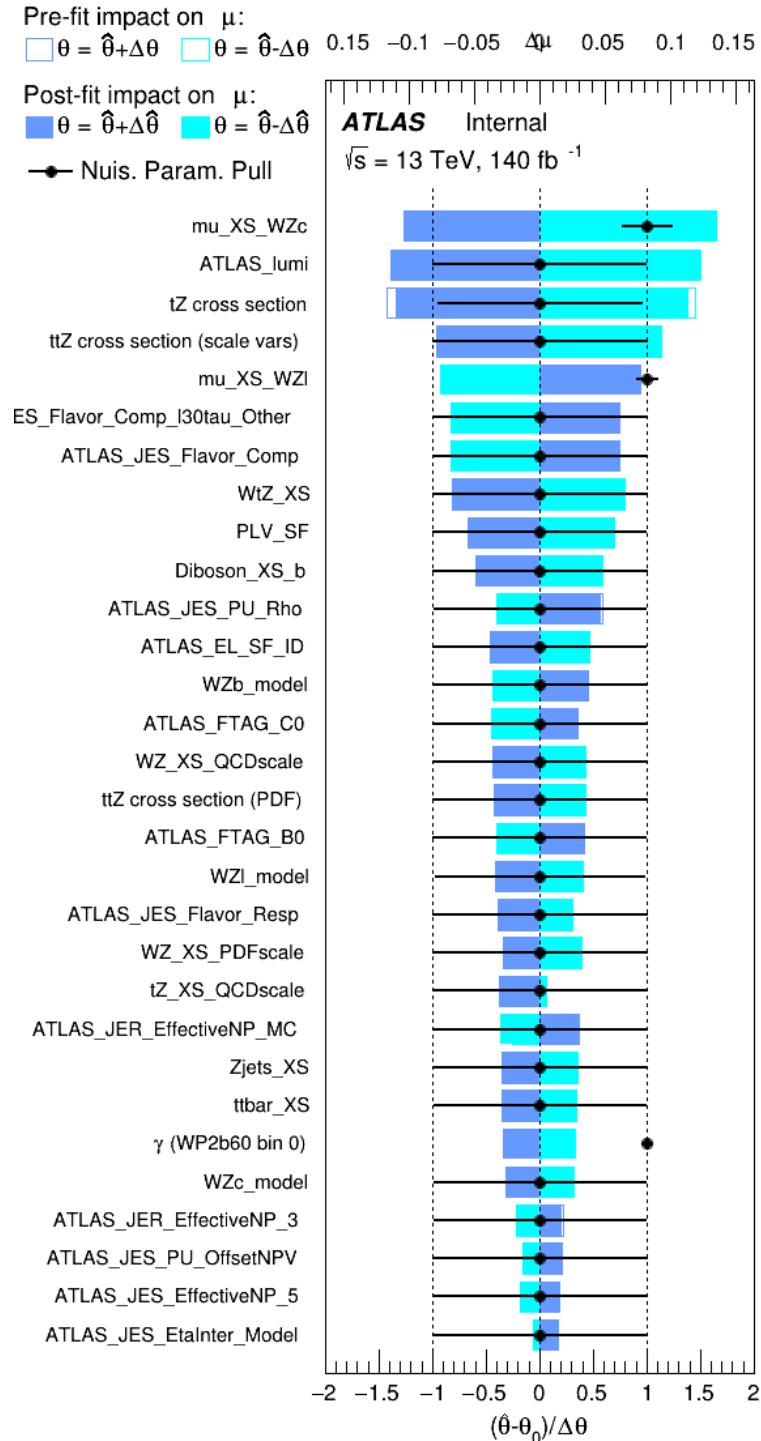


Figure 11.8: Impact of systematic uncertainties on the signal-strength of $WZ + b$ in 2-jet events.

992 The large impact of the Jet Energy Scale and Jet Flavor Tagging is unsurprising, as the
 993 shape of the fit regions depends heavily on the modeling of the jets. The other major sources
 994 of uncertainty come from background modelling and cross-section uncertainty. The pie charts
 995 in Figure 11.9 show that for the modelling uncertainties that contribute most correspond to the
 996 most significant backgrounds.

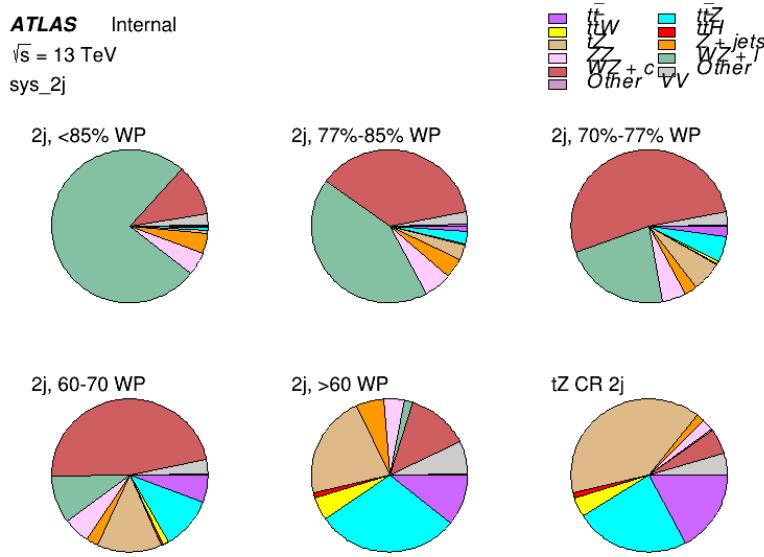


Figure 11.9: Post-fit background composition of the 2-jet fit regions.

997 The correlations between these nuisance parameters are summarized in Figure 11.10.

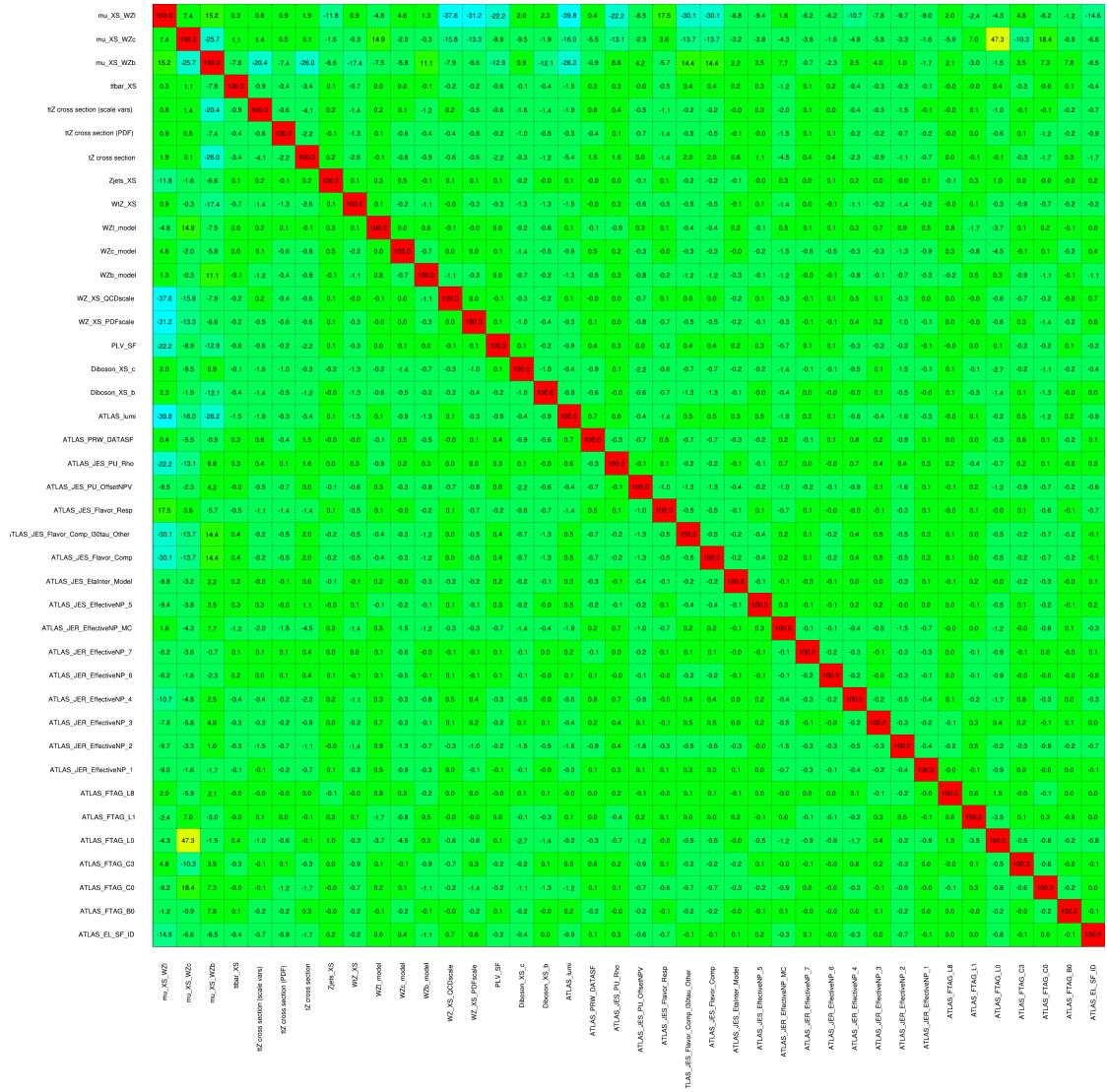


Figure 11.10: Correlations between nuisance parameters in the 2-jet fit

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

Part V**Differential Studies of $t\bar{t}H$ Multilepton****12 Data and Monte Carlo Samples****12.1 Data Samples**

1004 The study uses proton-proton collision data collected by the ATLAS detector from 2015 through
1005 2018, which represents an integrated luminosity of 139 fb^{-1} and an energy of $\sqrt{s} = 13 \text{ TeV}$. All
1006 data used in this analysis was included in one of the Good Run Lists verified by Data Quality
1007 checks.

12.2 Monte Carlo Samples

1008 Several Monte Carlo (MC) generators were used to simulate both signal and background pro-
1009 cesses. For all of these, the effects of the ATLAS detector are simulated in Geant4. The specific
1010 event generator used for each of these MC samples is listed in Table 21. A Higgs mass of 125
1011 GeV is assumed in all simulations.

Table 21: 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 [21] (CT10 [19])
t̄W	MG5_AMC (SHERPA 2.1.1)	NLO (LO multileg)	PYTHIA 8 (SHERPA)	NNPDF 3.0 NLO (NNPDF 3.0 NLO)
t̄(Z/γ* → ll)	MG5_AMC	NLO	PYTHIA 8	NNPDF 3.0 NLO
VV	SHERPA 2.2.2	MEPS NLO	SHERPA	CT10
t̄t	POWHEG-BOX v2 [24]	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, t̄tt̄t	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 [25]	NLO	PYTHIA 6	CT10
qqVV, VVV				
Z → l+l-	SHERPA 2.2.1	MEPS NLO	SHERPA	NNPDF 3.0 NLO

1013 The signal sample ($t\bar{t}H$) is modelled at NLO with POWHEG-BOX v2 using the NNPDF2.0
 1014 parton distribution function (PDF) [42]. Parton showering and hadronisation were modelled
 1015 with PYTHIA 8.2 [43]. The $t\bar{t}H$ sample is normalized to a cross-section of 507^{+35}_{-50} fb based on
 1016 NLO calculations. Uncertainties are based on varying the QCD factorisation and renormalisation
 1017 scale, as well as uncertainties in the PDF and α_s .

1018 The $t\bar{t}W$ background is simulated using Sherpa 2.2.1 with the NNPDF3.0 NLO PDF. The
 1019 matrix element is calculated with up to one additional parton at NLO, and up to two at LO. As
 1020 explained in detail in [7], the $t\bar{t}W$ contribution predicted by MC is found disagree significantly

1021 with what is observed in data. While an effort is currently being undertaken to measure $t\bar{t}W$ more
 1022 accurately, the approach used by the 79.8 fb^{-1} $t\bar{t}H$ analysis is used here: A normalization
 1023 factor of 1.68 is applied to the MC estimate of $t\bar{t}W$ and additional systematic uncertainties on
 1024 $t\bar{t}W$ are included to account for this modelling discrepancy, as outlined in Section 16.

1025 The $t\bar{t}(Z/\gamma^*)$ process is simulated with the **MADGRAPH5_AMC@NLO** generator, using
 1026 NNPDF3.0. Diboson processes are generated with **SHERPA 2.2.2** at NLO precision for one extra
 1027 parton, and at LO for up to three extra partons.

1028 The “fake”, or non-prompt, background comes primarily from leptons originating from
 1029 hadron decays, leptons with missidentified charge, and photon conversions. While the main
 1030 $t\bar{t}H$ analysis is currently refining a data-driven approach for estimating the contribution of
 1031 events with non-prompt leptons, at the time of this note this strategy has not been completely
 1032 developed for the full Run-2 dataset. Therefore, the non-prompt contribution is estimated with
 1033 MC, while applying normalization corrections and systematic uncertainties derived from data
 1034 driven techniques developed for the 79.8 fb^{-1} $t\bar{t}H/t\bar{t}W$ analysis [7].

1035 The primary contribution to the non-prompt lepton background is from $t\bar{t}$ production, with
 1036 V+jets and single-top as much smaller sources. Estimation of this background is done primarily
 1037 using an inclusive $t\bar{t}$ sample, with corrections applied based on data driven methods. This sample
 1038 is generated using PowHEG, with Pythia8 performing the parton shower and fragmentation.
 1039 Likelihood fits over several control regions enriched with these non-prompt backgrounds are fit
 1040 to data in order to derive normalization factors for these backgrounds. The specific normalization

1041 factors and uncertainties applied to the non-prompt contributions are listed in Section 16.

1042 Other processes, such as tH , tZ , $t\bar{t}WW$ and $t\bar{t}t\bar{t}$, are expected to make minor contributions
1043 to the total background. The generators and setting used for these backgrounds are summarized
1044 in Table 21.

1045 **13 Object Reconstruction**

1046 All analysis channels considered in this note share a common object selection for leptons and
1047 jets, as well as a shared trigger selection.

1048 **13.1 Trigger Requirements**

1049 Events are required to be selected by dilepton triggers. The p_T thresholds of the dilepton trigger
1050 on two electrons were 12 GeV in 2015, 17 GeV in 2016, and 24 GeV in 2017 and 2018, while for
1051 the dimuon triggers the p_T thresholds on the leading (sub-leading) muon were 18 GeV (8 GeV)
1052 in 2015, and 22 GeV (8 GeV) in 2016-2018. For the electron+muon triggers, the p_T thresholds
1053 on the electron (muon) were 17 GeV (14 GeV) for all datasets.

1054 **13.2 Light Leptons**

1055 Electron candidates are reconstructed from energy clusters in the electromagnetic calorimeter
 1056 that are associated with charged particle tracks reconstructed in the inner detector [26]. Electron
 1057 candidates are required to have $p_T > 10 \text{ GeV}$ and $|\eta_{\text{cluster}}| < 2.47$. Candidates in the transition
 1058 region between different electromagnetic calorimeter components, $1.37 < |\eta_{\text{cluster}}| < 1.52$, are
 1059 rejected. A multivariate likelihood discriminant combining shower shape and track information
 1060 is used to distinguish prompt electrons from nonprompt leptons, such as those originating from
 1061 hadronic showers.

1062 To further reduce the non-prompt contribution, the track of each electron is required to
 1063 originate from the primary vertex; requirements are imposed on the transverse impact parameter
 1064 significance ($|d_0|/\sigma_{d_0}$) and the longitudinal impact parameter ($|\Delta z_0 \sin \theta_\ell|$).

1065 Muon candidates are reconstructed by combining inner detector tracks with track segments
 1066 or full tracks in the muon spectrometer [27]. Muon candidates are required to have $p_T > 10 \text{ GeV}$
 1067 and $|\eta| < 2.5$.

1068 All leptons are required to pass a non-prompt BDT selection developed by the main
 1069 $t\bar{t}H/t\bar{t}W$ analysis, described in detail in [7]. Optimized working points and scale factors for this
 1070 BDT are taken from that analysis. This BDT and the WPs used are summarized in Appendix
 1071 [A](#),

1072 13.3 Jets

1073 Jets are reconstructed from calibrated topological clusters built from energy deposits in the
1074 calorimeters [29], using the anti- k_t algorithm with a radius parameter $R = 0.4$. Particle Flow, or
1075 PFlow, jets are used in the analysis, which are hadronic objects reconstructed using information
1076 from both the tracker and the calorimeter. Jets with energy contributions likely arising from noise
1077 or detector effects are removed from consideration [30], and only jets satisfying $p_T > 25$ GeV
1078 and $|\eta| < 2.5$ are used in this analysis. For jets with $p_T < 60$ GeV and $|\eta| < 2.4$, a jet-track
1079 association algorithm is used to confirm that the jet originates from the selected primary vertex,
1080 in order to reject jets arising from pileup collisions [31].

1081 13.4 B-tagged Jets

1082 Each analysis channel used in this analysis includes b-jets in the final state. These are identified
1083 using the DL1r b-tagging algorithm, which uses jet vertex and kinematic information to distin-
1084 guish heavy and light flavored jets. These features are used as inputs to a neural network, the
1085 output of which is used to form calibrated working points (WPs) based on how likely a jet is to
1086 have originated from a b-quark. This analysis uses the 70% DL1r WP - implying an efficiency of
1087 70% for truth b-jets - for selecting b-tagged jets.

1088 13.5 Missing Transverse Energy

1089 Because all $t\bar{t}H - ML$ channels considered include multiple neutrinos, missing transverse
1090 energy (E_T^{miss}) is present in each event. The missing transverse momentum vector is defined as
1091 the inverse of the sum of the transverse momenta of all reconstructed physics objects as well
1092 as remaining unclustered energy, the latter of which is estimated from low- p_T tracks associated
1093 with the primary vertex but not assigned to a hard object [44].

1094 13.6 Overlap removal

1095 To avoid double counting objects and remove leptons originating from decays of hadrons, overlap
1096 removal is performed in the following order: any electron candidate within $\Delta R = 0.1$ of another
1097 electron candidate with higher p_T is removed; any electron candidate within $\Delta R = 0.1$ of a muon
1098 candidate is removed; any jet within $\Delta R = 0.3$ of an electron candidate is removed; if a muon
1099 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
1100 is kept and the muon is removed.

1101 This algorithm is applied to the preselected objects. The overlap removal procedure is
1102 summarized in Table 22.

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

1103 14 Higgs Momentum Reconstruction

1104 Reconstructing the momentum of the Higgs boson is a particular challenge for channels with
 1105 leptons in the final state: Because all channels include at least two neutrinos in the final state, the
 1106 Higgs can never be fully reconstructed. However, the momentum spectrum can be well predicted
 1107 by a neural network when provided with the kinematics of the Higgs Boson decay products - as
 1108 verified by studies detailed in Appendix C.3. With this in mind, several layers of MVAs are used
 1109 to reconstruction the Higgs momentum:

1110 The first layer is a model designed to select which jets are most likely to be the b-jets that
 1111 came from the top decay, detailed in Section 14.2. As described in Section 14.3, the kinematics
 1112 of these jets and possible Higgs decay products are fed into the second layer, which is designed to
 1113 identify the decay products of the Higgs Boson itself. The kinematics of the particles this layer
 1114 identifies as most likely to have originated from the Higgs decay are then fed into yet another
 1115 neural-network, which predicts the momentum of the Higgs (14.4). For the 3l channel, because
 1116 the Higgs can decay into either one lepton and two jets or two leptons, an additional MVA is
 1117 used to determine the decay mode of the Higgs boson in the 3l channel (14.5).

1118 Models are trained on Monte Carlo simulations of $t\bar{t}H$ events generated using MG5_AMC.

1119 Specifically, events DSIDs 346343-5, 345873-5, and 345672-4 are used for training in order to

1120 increase the statistics of the training sample.

1121 For all of these models, the Keras neural network framework, with Tensorflow 2.0 as

1122 the backend [**tensorflow**], is used, and the number of hidden layers and nodes are determined

1123 using grid search optimization. Each neural network uses the LeakyReLU activation function,

1124 a learning rate of 0.01, and the Adam optimization algorithm, as alternatives are found to either

1125 decrease or have no impact on performance. Batch normalization is applied after each layer in

1126 order to stabilize the model and decrease training time. For the classification algorithms (b-jet

1127 matching, Higgs reconstruction, and 3l decay identification) binary-cross entropy is used as the

1128 loss function, while the p_T reconstruction algorithm uses MSE.

1129 The specific inputs features used for each model are arrived at through a process of trial

1130 and error - features considered potentially useful are tried, and those that are found to increase

1131 performance are included. While each model includes a relatively large number of features, some

1132 using upwards of 30, this inclusive approach is found to maximize the performance of each model

1133 while decreasing the variance compared to a reduced number of inputs. Each input feature is

1134 validated by comparing MC simulations to 79.8 fb^{-1} of data, with the full set of features shown

1135 in Section C..

1136 **14.1 Physics Object Truth Matching**

1137 Machine Learning algorithms are trained to identify the decay products of the Higgs Boson using
1138 MC simulations of $t\bar{t}H$ events. The kinematics of the reconstructed physics objects, as well as
1139 event level variables such the jet multiplicity and missing energy, used as inputs, with the parent
1140 ID taken from the truth record used to label the data. The objects considered include light leptons
1141 and jets.

1142 Reconstructed physics objects are matched to particle level objects, in order to identify the
1143 parent particle of these reconstructed objects. Reconstructed jets are matched to truth jets based
1144 on the requirements that the reco jet and truth jet fall within $\Delta R < 0.4$, and the two objects have
1145 a p_T that agrees within 10%. Truth level and reco level leptons are required to have the same
1146 flavor, a $\Delta R < 0.1$, and p_T that agree within 10%. Events where no match can be found between
1147 the particle level decay products and the reconstructed objects are not included in training.

1148 Leptons considered as possible Higgs and top decay candidates are required to pass the
1149 selection described in Section 13.2. For jets, however, it is found that a large fraction that
1150 originate from either the top decay or the Higgs decay fall outside the selection described in
1151 Section 13.3. Specifically, jets from the Higgs decay tend to be soft, with 32% having $p_T <$
1152 25 GeV. Therefore jets with $p_T < 15$ GeV are considered as possible candidates in the models
1153 described below. By contrast, less than 5% of the jets originating from the Higgs fall below this
1154 p_T threshold. The jets are found to be well modeled even down to this low p_T threshold, as shown
1155 in Section 15.1. The impact of using different p_T selection for the jet candidates is considered

1156 in detail in Section C.6. As they are expected to originate from the primary vertex, jets are also
1157 required to pass a JVT cut. The overlap removal selection is not applied to the objects considered
1158 in the models.

1159 **14.2 b-jet Identification**

1160 Including the kinematics of the b-jets that originate from the top decay is found to improve the
1161 identification of the Higgs decay products, and improve the accuracy with which the Higgs
1162 momentum can be reconstructed. Because these b-jets are reconstructed by the detector with
1163 high efficiency (just over 90% of the time), and can be identified relatively consistently, the first
1164 step in reconstructing the Higgs is selecting the b-jets from the top decay.

1165 Exactly two b-jets are expected in the final state of $t\bar{t}H - \text{ML}$ events. However, in both
1166 the 3l and 2LSS channels, only one or more b-tagged jets are required (where the 70% DL1r
1167 b-tag working point is used). Therefore, for events which have exactly one, or more than two,
1168 b-tagged jets, deciding which combination of jets correspond to the top decay is non-trivial.
1169 Further, events with 1 b-tagged jet represent just over half of all $t\bar{t}H - \text{ML}$ events. Of those,
1170 both b-jets are reconstructed by the detector 75% of the time. Therefore, rather than adjusting
1171 the selection to require exactly 2 b-tagged jets, and losing more than half of the signal events, a
1172 neural network is used to predict which pair of jets is most likely to correspond to truth b-jets.

1173 Once the network is trained, kinematic variables for all possible pairings of jets are fed into
1174 the model, and the pair of jets with the highest output score are taken to be b-jets in successive

1175 steps of the analysis.

1176 An alternate approach is considered, where information about all jets in each event are
 1177 used as the feature set, and the model is tasked with identifying which two originated from the
 1178 top decay. While this approach is found to underperform the nominal approach, and therefore
 1179 not used in the analysis, the results are documented in Appendix C.4.

1180 14.2.1 2lSS Channel

1181 For the 2lSS channel, the input features shown in Table 23 are used for training. Here j_0 and j_1
 1182 are the two jet candidates, while l_0 and l_1 are the two leptons in the event, both ordered by p_T . jet
 1183 DL1r is an integer corresponding to the calibrated b-tagging working points reached by each jet,
 1184 where 5 represents the tightest working point and 1 represents the loosest. The variables nJets
 1185 DL1r 60% and nJets DL1r 85% represent the number of jets in the event passing the 60% and
 1186 85% b-tag working points, respectively.

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

Table 23: Input features used in the b-jet identification algorithm for the 2lSS channel

1187 As there are far more incorrect combinations than correct ones, by a factor of more than
 1188 20:1, the training set is resampled to reduce the fraction of incorrect combinations. A random
 1189 sample of 5 million incorrect entries are used for training, along with around 1 million correct
 1190 entries. 10% of the dataset is set aside for testing, leaving around 5 million datapoints for
 1191 training.

1192 The difference between the distributions for a few of these features for the "correct" (i.e.
 1193 both jets are truth b-jets), and "incorrect" combinations are shown in Figure 14.1. The correct and
 1194 incorrect contributions are scaled to the same integral, so as to better demonstrate the differences
 1195 in the distributions.

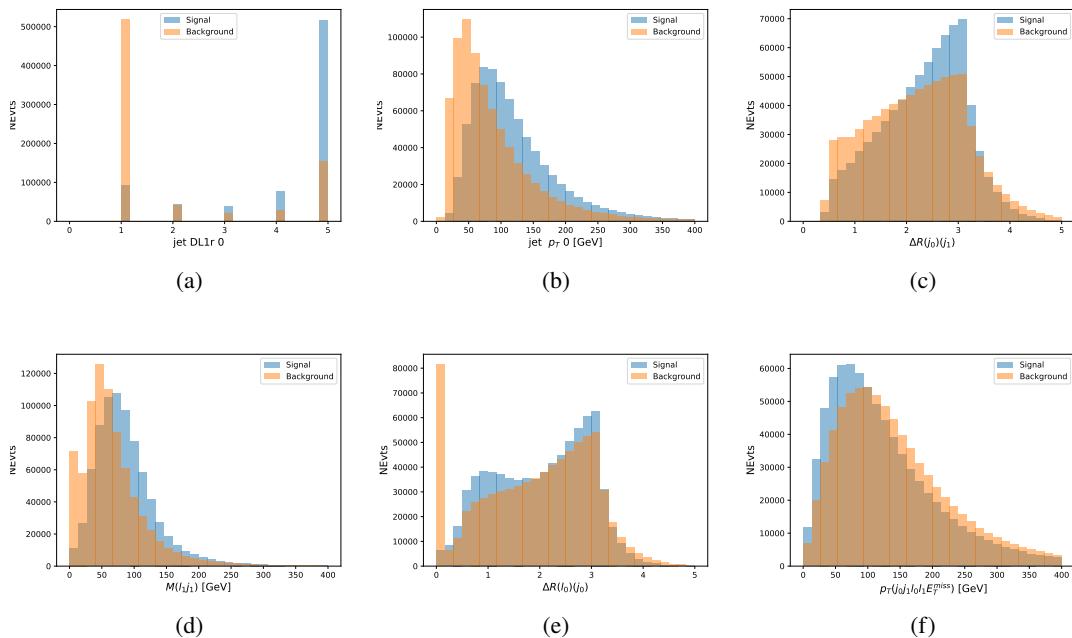


Figure 14.1: Input features for top2ISS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Each are scaled to the same number of events. (a) shows the DL1r working point of leading jet, (b) shows the p_T of the leading jet, (c) shows the ΔR of the two jets, (d) the invariant mass of lepton 1 and jet 1, (e) the ΔR of lepton 0 and jet 0, and (f) the p_T of both jets, both leptons, and the E_T^{miss} .

1196 The modeling of these inputs is validated against data, with Figure 14.2 showing good
 1197 general agreement between data and MC. Plots for the complete list of features can found in
 1198 Appendix C.

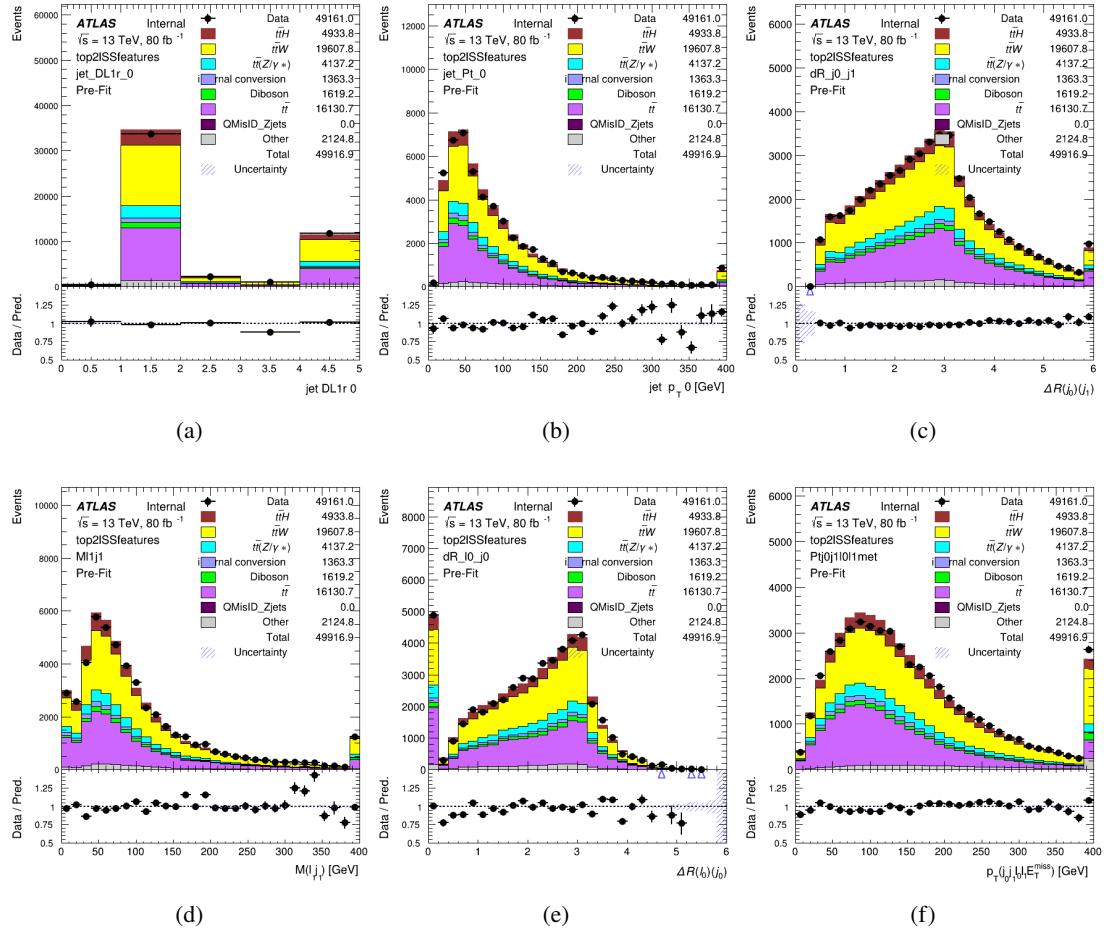


Figure 14.2: Data/MC comparisons of input features for top2ISS training for 79.8 fb^{-1} of data. (a) shows the DL1r working point of leading jet, (b) shows the p_T of the leading jet, (c) shows the ΔR of the two jets, (d) the invariant mass of lepton 1 and jet 1, (e) the ΔR of lepton 0 and jet 0, and (f) the p_T of both jets, both leptons, and the E_T^{miss} .

1199 Based on the results of grid search evaluation, the optimal architecture is found to include
 1200 5 hidden layers with 40 nodes each. No regularizer or dropout is added to the network, as

1201 overfitting is found to not be an issue. The output score distribution as well as the ROC curve for
 1202 the trained model are shown in Figure 14.2.1. The model is found to identify the correct pairing
 1203 of jets for 73% of 2lSS signal events on test data.

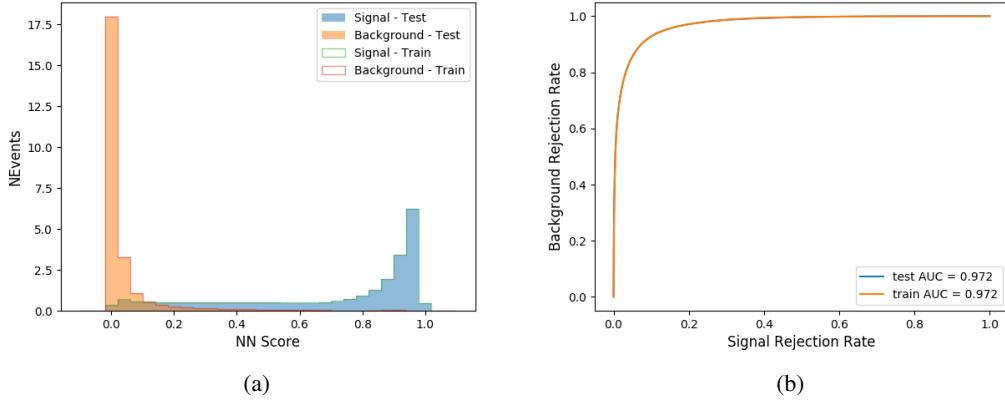


Figure 14.3: Results of the b-jet identification algorithm for the 2lSS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

1204 For point of comparison, a "naive" approach to identifying b-jets is used as well: The two
 1205 jets which pass the highest DL1r b-tag working point are assumed to be the b-jets from the top
 1206 decay. In the case that multiple jets meet the same b-tag working point, the jet with higher p_T is
 1207 used. This method identifies the correct jet pair 65% of the time.

1208 The accuracy of the model for different b-tagged jet multiplicities, compared to this naive
 1209 approach, is shown in Table 24.

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

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

14.2.2 3l Channel

The input features used in the 3l channel are listed in Table 25, 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 25: Input features for the b-jet identification algorithm in the 3l channel.

1216 A few of these features are shown in Figure 14.4, comparing the distributions for correct
 1217 and incorrect combinations of jets.

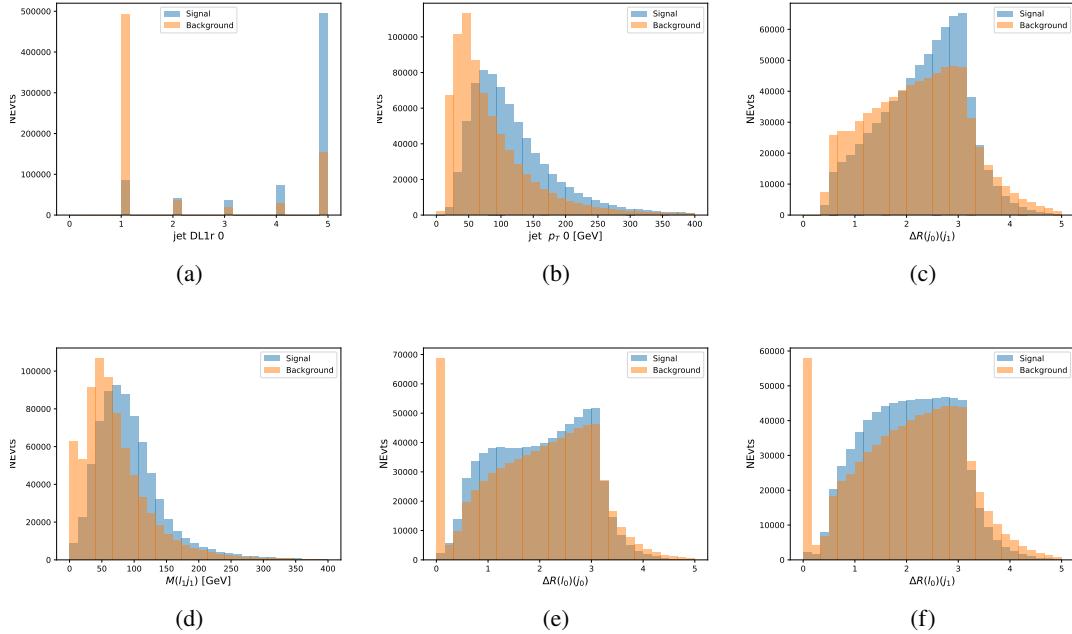


Figure 14.4: Input features for top3l training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events. (a) show the DL1r WP of jet 0, (b) the p_T of jet 0, (c) ΔR between jet 0 and 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

1218 The modeling of these inputs is validated against data, with Figure 14.5 showing good
 1219 general agreement between data and MC.

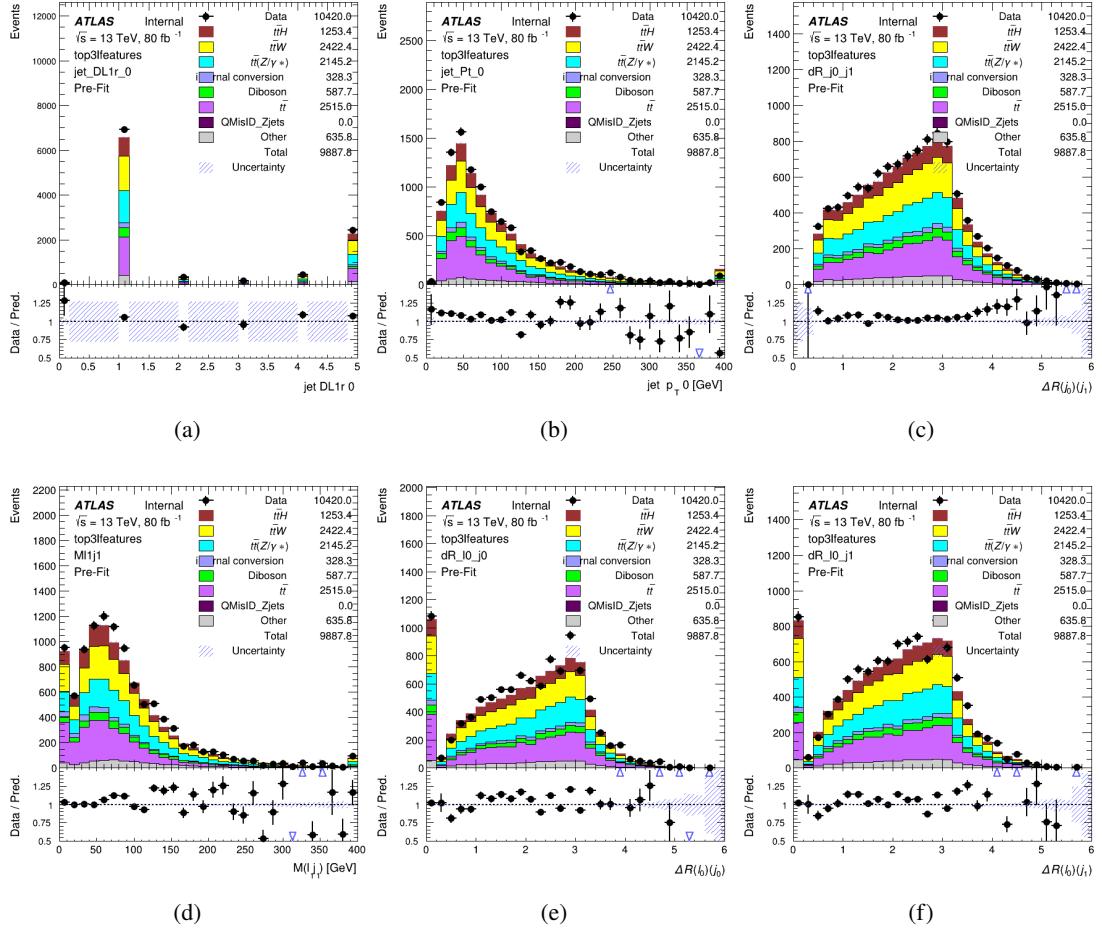


Figure 14.5: Data/MC comparisons of input features for top3l training for 79.8 fb^{-1} of data. (a) show the DL1r WP of jet 0, (b) the p_T of jet 0, (c) ΔR between jet 0 and jet 1, (d) the invariant mass of lepton 1 and jet 1, (e) ΔR between lepton 0 and jet 0, and (f) ΔR between lepton 0 and jet 1

Again, the dataset is downsized to reduce the ratio of correct and incorrect combination from 20:1, to 5:1. Around 7 million events are used for training, with 10% set aside for testing. Based on the results of grid search evaluation, the optimal architecture is found to include 5 hidden layers with 60 nodes each. The output score distribution as well as the ROC curve for the trained model are shown in Figure 14.2.2.

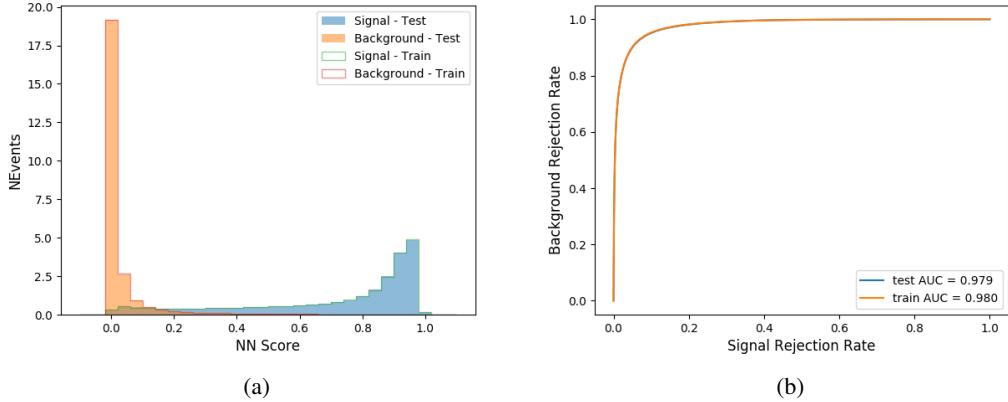


Figure 14.6: Results of the b-jet identification algorithm for the 3l channel, showing (a) the output score of the NN for correct and incorrect combinations of jets. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

1225 This procedure is found to identify the correct pairing of jets for nearly 80% of 3l signal
 1226 events. The accuracy of the model is summarized in Table 26, once again compared to the naive
 1227 approach described above.

Table 26: 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%

14.3 Higgs Reconstruction

1228 Techniques similar to the b-jet identification algorithms are employed to select the decay products
 1229 of the Higgs: kinematics of all possible combinations of reconstructed objects are fed into a neural
 1230

1231 network to determine which of those is most mostly to be the decay products of the Higgs.

1232 Again separate models are used for the 2lSS and 3l channels, while the 3l channel has
1233 now been split into two: $t\bar{t}H$ events with three leptons in the final state include both instances
1234 where the Higgs decays into a lepton (and a neutrino) and a pair of jets, and instances where the
1235 Higgs decays to two leptons (and two neutrinos which are not reconstructed).

1236 3l events are therefore categorized as either semi-leptonic (3lS) or fully-leptonic (3lF). In
1237 the semi-leptonic case the reconstructed decay products consist of two jets and a single lepton.
1238 For the fully-leptonic case, the decay products include 2 of the three leptons associated with
1239 the event. For training these models, events are separated into these two categories using truth
1240 level information. A separate MVA, described in Section 14.5, is used to make this distinction
1241 at reconstructed level, and determine which model to use.

1242 For all channels, the models described in Section 14.2 are used to identify b-jet candidates,
1243 whose kinematics are used as additional input features to help identify the Higgs decay products.
1244 These jets are not considered as possible candidates for the Higgs decay, justified by the fact that
1245 these models are found to misidentify jets from the Higgs decay as jets from the top decay less
1246 than 1% of the time.

1247 **14.3.1 2lSS Channel**

1248 For the 2lSS channel, the Higgs decay products include one light lepton and two jets. The neural
1249 network is trained on the kinematics of different combinations of leptons and jets, as well as the

1250 b-jets identified in Section 14.2, with the specific input features listed in Table 27.

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

Table 27: Input features used to identify the Higgs decay products in 2LSS events

1251 Here j_0 and j_1 , and l_H are the jet and lepton decay candidates, respectively. The other
1252 lepton in the event is labeled l_T , as it is assumed to have come from the decay of one of the top
1253 quarks. b_0 and b_1 are the two b-jets identified by the b-jet identification algorithm. The b-jet
1254 Reco Score is the output of the b-jet reconstruction algorithm.

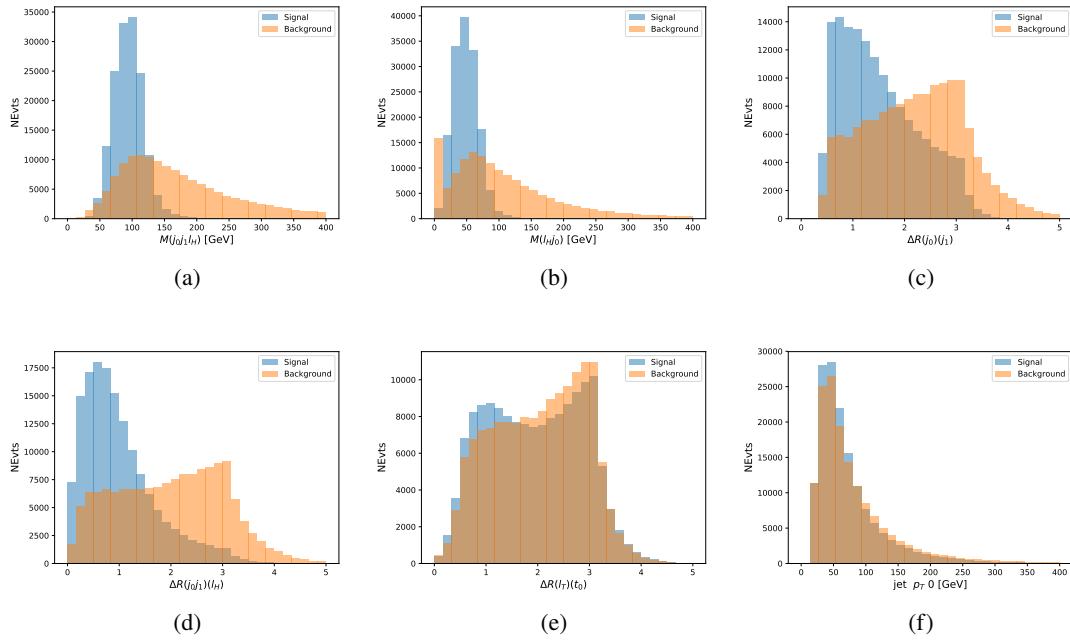


Figure 14.7: Input features for higgs2lSS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events. (a) shows the invariant mass of the two jet candidates and the lepton candidate, (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.

1255 The modeling of these inputs is validated against data, with Figure 14.2 showing good
 1256 general agreement between data and MC. Plots for the complete list of features can found in
 1257 Section C.

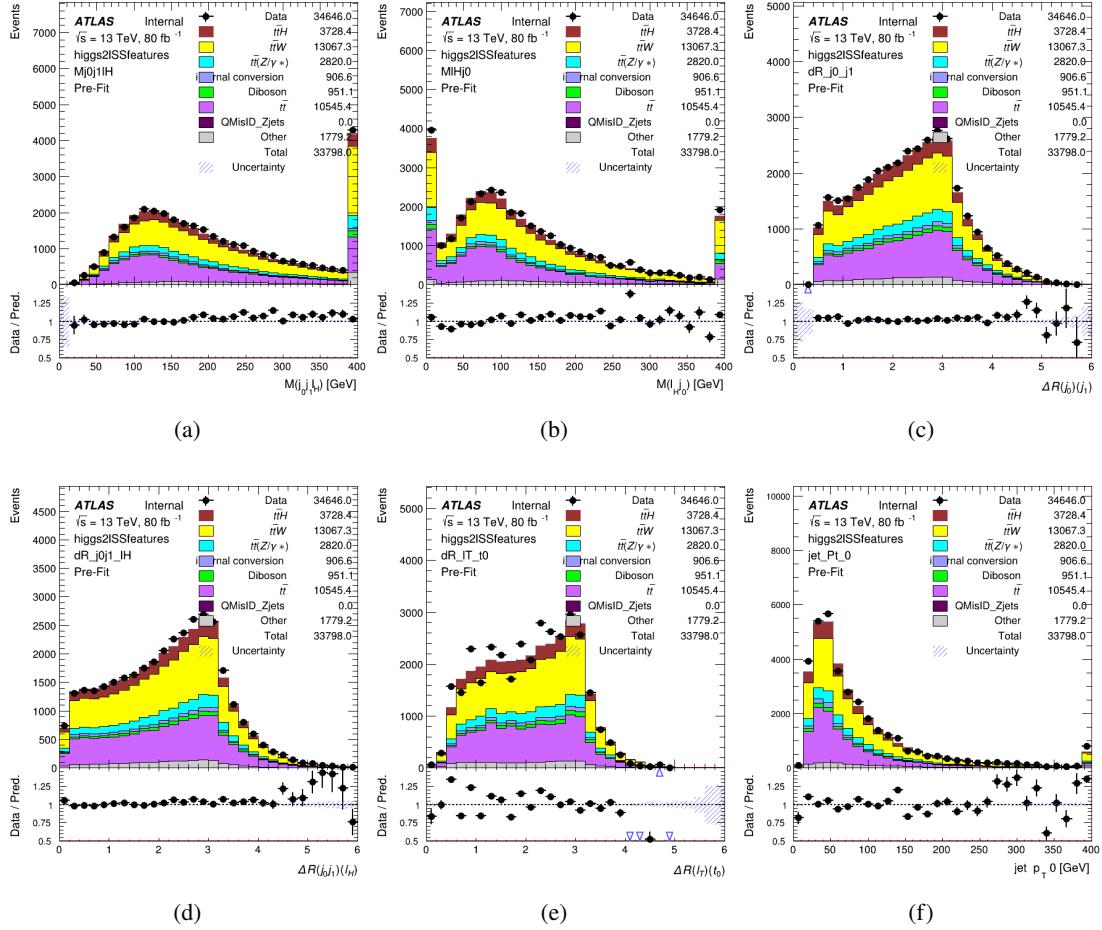


Figure 14.8: Data/MC comparisons of input features for higgs2ISS training for 79.8 fb^{-1} of data. (a) shows the invariant mass of the two jet candidates and the lepton candidate, (b) the invariant mass of jet 0 and the lepton candidate, (c) ΔR between the jet candidates, (d) ΔR between jet 0 + jet 1 and the lepton candidate, (e) ΔR between the lepton from the top and the leading b-jet, (f) the p_T of jet 0.

1258 A neural network consisting of 7 hidden layers with 60 nodes each is trained on around 2
 1259 million events, with an additional 200,000 reserved for testing the model. In order to compensate
 1260 for the large number of incorrect combinations, these have been downsampled such that the correct
 1261 combinations represent over 10% of the training set. The output of the NN is summarized in
 1262 Figure 14.3.1.

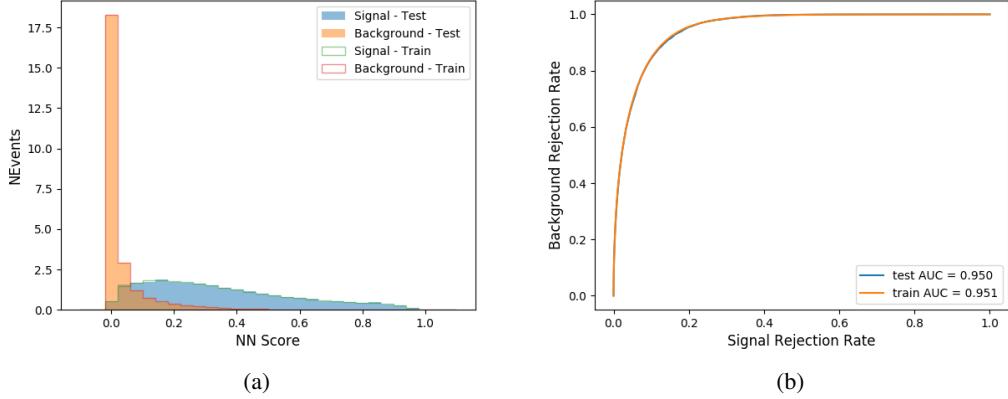


Figure 14.9: Result of the Higgs reconstruction algorithm in the 2lSS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets scaled to an equal number of events, and (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

1263 The neural network identifies the correct combination 55% of the time. It identifies the
 1264 correct lepton 85% of the time, and selects the correct lepton and at least one of the correct jets
 1265 81% of the time.

1266 14.3.2 3l Semi-leptonic Channel

1267 For 3l $t\bar{t}H$ where the Higgs decay semi-leptonically, the decay products include one of the three
 1268 leptons and two jets. In this case, the other two leptons originated from the decay of the tops,
 1269 meaning the opposite-sign (OS) lepton cannot have come from the Higgs. This leaves only the two
 1270 same-sign (SS) leptons as possible Higgs decay products.

Lepton p_T H	Lepton p_T T_0	Lepton p_T T_1
jet p_T 0	jet p_T 1	top p_T 0
top p_T 1	jet η 0	jet η 1
jet ϕ 0	jet ϕ 1	$\Delta R(j_0)(j_1)$
$M(j_0j_1)$	$\Delta R(l_H)(j_0)$	$\Delta R(l_H)(j_1)$
$\Delta R(j_0j_1)(l_H)$	$\Delta R(j_0j_1)(l_{T_0})$	$\Delta R(l_{T_0})(l_{T_1})$
$\Delta R(l_H)(l_{T_1})$	$M(j_0j_1l_{T_0})$	$M(j_0j_1l_{T_1})$
$M(j_0j_1l_H)$	$\Delta R(j_0j_1l_H)(l_{T_0})$	$\Delta R(j_0j_1l_H)(l_{T_1})$
$\Delta\phi(j_0j_1l_H)(E_T^{\text{miss}})$	$p_T(j_0j_1l_Hl_{T_0}l_{T_1}b_0b_1E_T^{\text{miss}})$	$M(j_0j_1b_0)$
$M(j_0j_1b_1)$	$\Delta R(l_{T_0})(b_0)$	$\Delta R(l_{T_0})(b_1)$
$\Delta R(l_{T_1})(b_0)$	$\Delta R(l_{T_1})(b_1)$	$\Delta R(j_0)(b_0)$
$\Delta R(j_0)(b_1)$	$\Delta R(j_1)(b_0)$	$\Delta R(j_1)(b_1)$
b-jet Reco Score	E_T^{miss}	HT jets
nJets		

Table 28: Input features used to identify the Higgs decay products in 3l semi-leptonic events

1271 Here j_0 and j_1 , and l_H are the jet and lepton decay candidates, respectively. The other
 1272 two leptons in the event are labeled as l_{T_0} and l_{T_1} . b_0 and b_1 are the two b-jets identified by
 1273 the b-jet identification algorithm. The b-jet Reco Score is the output of the b-jet identification
 1274 algorithm.

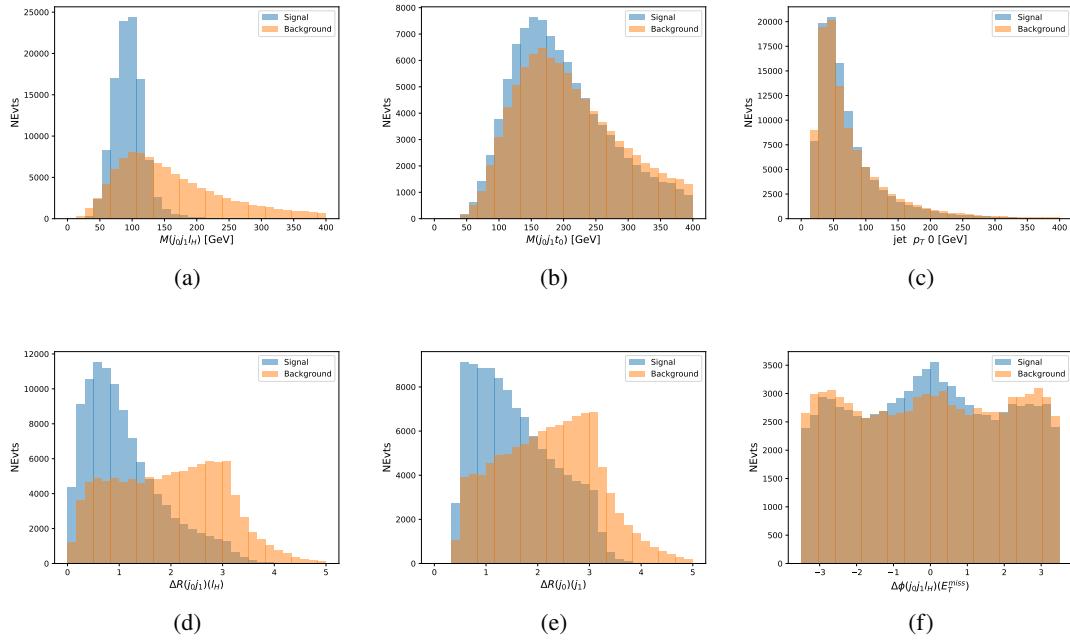


Figure 14.10: Input features for higgs3lS training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

1275 The modeling of these inputs is validated against data, with Figure 14.11 showing good
 1276 general agreement between data and MC. Plots for the complete list of features can found in
 1277 appendix C.1.

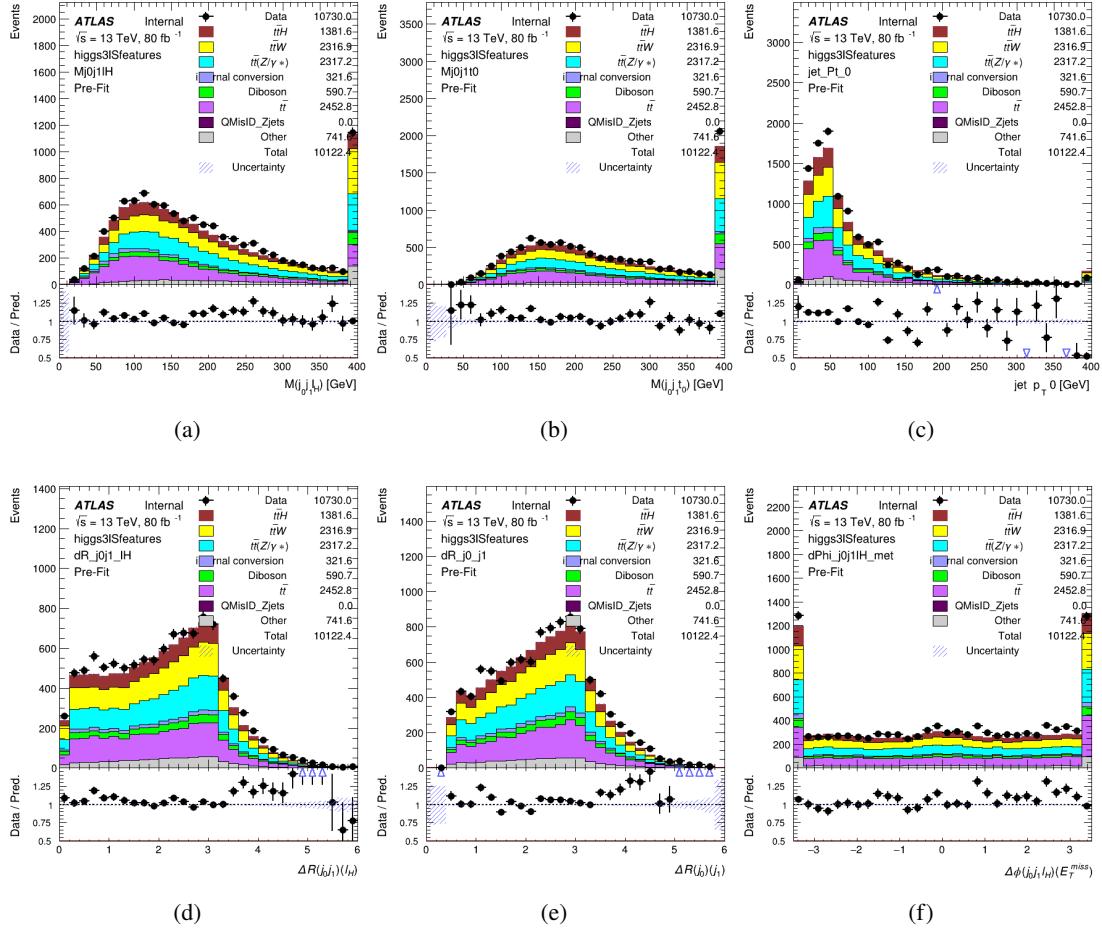


Figure 14.11: Data/MC comparisons of input features for higgs3lS training for 79.8 fb^{-1} of data.

1278 A neural network of 7 hidden layers with 70 nodes each is trained on 1.8 million events.
 1279 Once again, incorrect combinations are downsampled, such that the correct combinations are
 1280 around 10% of the training set. 10% of the dataset is reserved for testing. The output of the NN
 1281 is summarized in Figure 14.3.2.

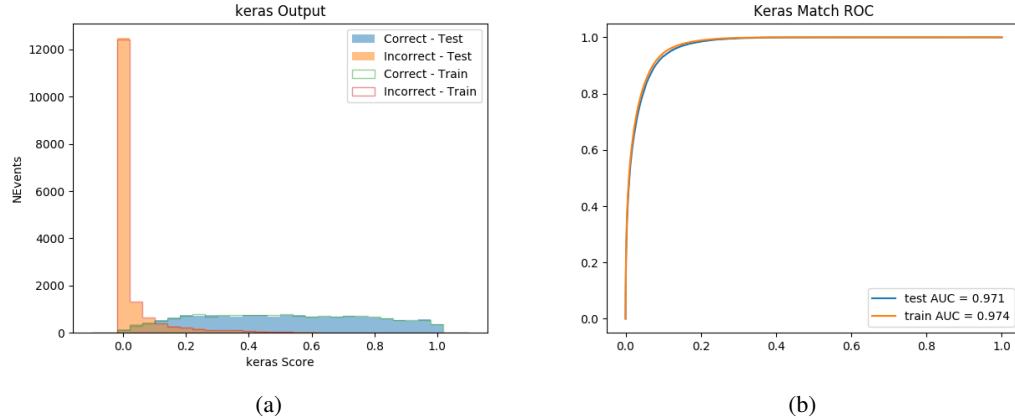


Figure 14.12: Results of the Higgs reconstruction algorithm in the 3lS channel, showing (a) the output score of the NN for correct and incorrect combinations of jets, scaled to an equal number of entries.,. (b) the ROC curve of the output, showing background rejection as a function of signal efficiency

1282 The neural network identifies the correct combination 64% of the time. It identifies the
 1283 correct lepton 85% of the time, and selects the correct lepton and at least one of the correct jets
 1284 83% of the time.

1285 14.3.3 3l Fully-leptonic Channel

1286 In the fully-leptonic 3l case, the goal is identify which two of the three leptons originated from
 1287 the Higgs decay. Since one of these two must be the OS lepton, this problem is reduced to
 1288 determining which of the two SS leptons originated from the Higgs. The kinematics of both
 1289 possibilities are used for training, one where the SS lepton from the Higgs is correctly labeled,
 1290 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}	b-jet Reco Score

Table 29: Input features used to identify the Higgs decay products in 3lF events

Table 30: Input features used to identify the Higgs decay products in 3l fully leptonic events

1291 Here l_{H_0} and l_{H_1} are the Higgs decay candidates. The other lepton in the event is labeled
 1292 l_T . b_0 and b_1 are the two b-jets identified by the b-jet identification algorithm. The b-jet Reco
 1293 Score is the output of the Higgs reconstruction algorithm.

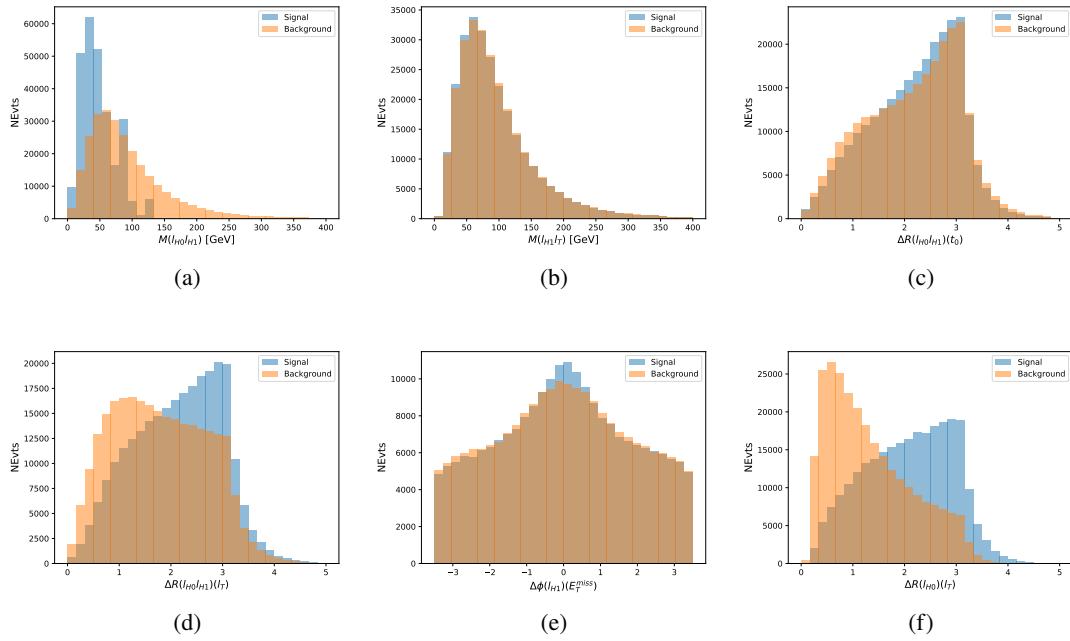


Figure 14.13: Input features for higgs3lF training. The signal in blue represents events where both jet candidates are truth b-jets from top decays, and the orange is all other combinations. Scaled to the same number of events.

1294

The modeling of these inputs is validated against data, with Figure 14.14 showing good

1295

general agreement between data and MC. Plots for the complete list of features can found in

1296

Section C.

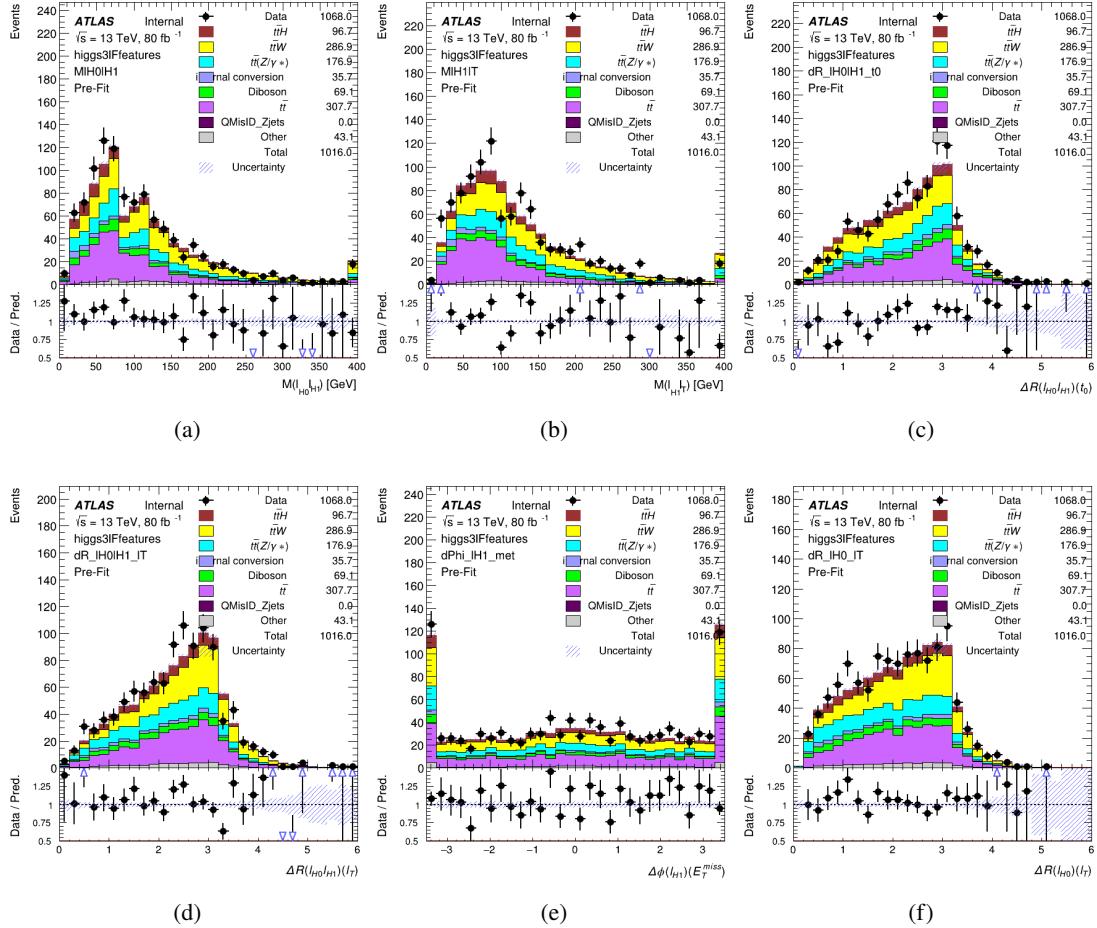


Figure 14.14: Data/MC comparisons of input features for higgs3lF training for 79.8 fb^{-1} of data.

1297 A neural network consisting of 5 nodes and 60 hidden units is trained on 800,000 events,
1298 with 10% of the dataset reserved for testing. The output of the model is summarized in Figure
1299 [14.3.3.](#).

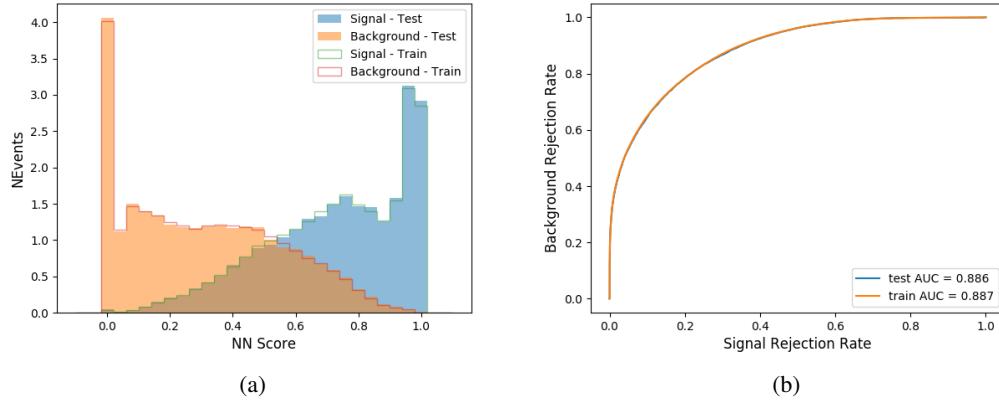


Figure 14.15: (a) the output score of the NN for correct and incorrect combinations of jets. (a) the ROC curve of the output, showing background rejection as a function of signal efficiency

1300 The correct lepton is identified by the model for 80% of events in the testing data set.

1301 14.4 p_T Prediction

1302 Once the most probable decay products have been identified, their kinematics are used as inputs
 1303 to a regression model which attempts to predict the momentum of the Higgs Boson. Once again,
 1304 a DNN is used. Input variables representing the b-jets and leptons not from the Higgs decay
 1305 are included as well, as these are found to improve performance. The truth p_T of the Higgs,
 1306 as predicted by MC, are used as labels. Separate models are built for each channel - 2lSS, 3l
 1307 Semi-leptonic and 3l Fully-leptonic.

1308 As a two-bin fit is targeted for the final result, some metrics evaluating the performance
 1309 of the models aim to show how well it distinguished between "high p_T " and "low p_T " events. A

1310 cutoff point of 150 GeV is used to define these two categories.

1311 Because the analysis uses a two bin fit of the Higgs p_T , the momentum reconstruction
1312 could be treated as a binary classification problem, rather than a regression problem. This
1313 approach is explored in detail in Section C.5, and is found not to provide any significant increase
1314 in sensitivity. The regression approach is used because it provides more flexibility for future
1315 analyses, as it is independent of the cutoff between high and low p_T , as well as the number of
1316 bins. Further, a regression allows the output of the neural network to be more clearly understood,
1317 as it can be directly compared to a physics observable.

1318 **14.4.1 2lSS Channel**

1319 The input variables listed in Table 31 are used to predict the Higgs p_T in the 2lSS channel. Here
1320 j_0 and j_1 are the two jets identified as Higgs decay products. The lepton identified as originating
1321 from the Higgs is labeled l_H , while the other lepton is labeled l_T , as it is assumed to have come
1322 from the decay of one of the top quarks. b_0 and b_1 are the two b-jets identified by the b-jet
1323 identification algorithm. The Higgs Reco Score and b-jet Reco Score are the outputs of the Higgs
1324 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 31: Input features for reconstructing the Higgs p_T spectrum for 2lSS events

1325 The optimal neural network architecture for this channel is found to consist of 7 hidden
 1326 layers with 60 nodes each. The input data set includes 1.2 million events, 10% of which is used
 1327 for testing, the other 90% for training. Training is found to converge after around 150 epochs.

1328 To evaluate the performance of the model, the predicted p_T spectrum is compared to the
 1329 truth Higgs p_T in Figure 14.16. In order to visualize the model performance more clearly, in (a)
 1330 of that figure, the color of each point is determined by Kernel Density Estimation (KDE). The
 1331 color shown represents the logarithm of the output from KDE, to counteract the large number of
 1332 low p_T events. For that same reason, each column of the histogram shown in (b) of Figure 14.16
 1333 is normalized to unity. This plot therefore demonstrates what the model predicts for each slice
 1334 of truth p_T .

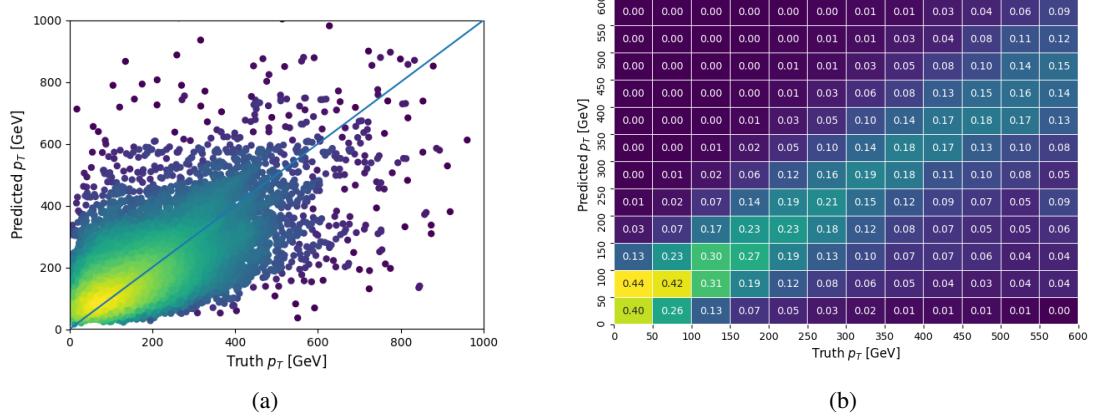


Figure 14.16: The regressed Higgs momentum spectrum as a function of the truth p_T for 2lSS $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized to one.

1335 We are also interested in how well the model distinguishes between events with $p_T < 150$
 1336 GeV and > 150 GeV. Figure 14.17 demonstrates the NN output for high and low p_T events based
 1337 on this cutoff.

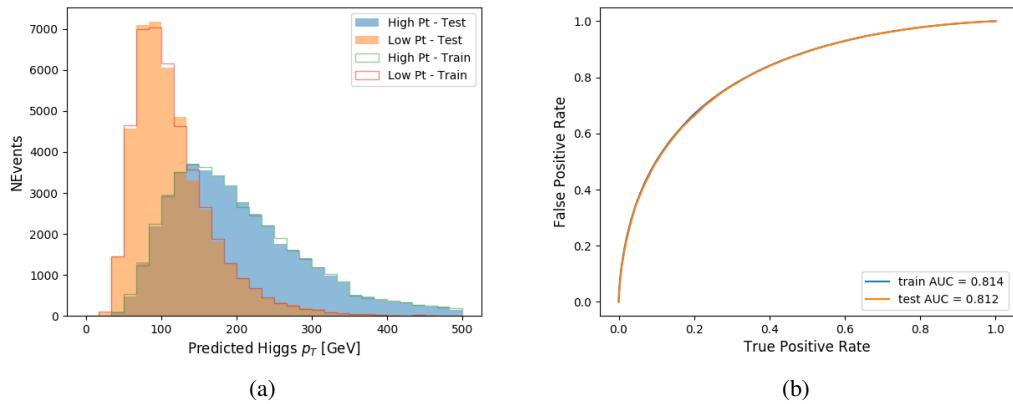


Figure 14.17: (a) shows the reconstructed Higgs p_T for 2lSS events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

¹³³⁸ **14.4.2 3l Semi-leptonic Channel**

¹³³⁹ The following input features are used to predict the Higgs p_T for events in the 3lS channel:

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

Table 32: Input features for reconstructing the Higgs p_T spectrum for 3lS events

¹³⁴⁰ Again, j_0 and j_1 are the two jets identified as Higgs decay products, ordered by p_T . The

¹³⁴¹ lepton identified as originating from the Higgs is labeled l_H , while the other two leptons are

¹³⁴² labeled l_{T0} and l_{T1} . b_0 and b_1 are the two b-jets identified by the b-jet identification algorithm.

¹³⁴³ The Higgs Reco Score and b-jet Reco Score are the outputs of the Higgs reconstruction algorithm,

¹³⁴⁴ and the b-jet identification algorithm, respectively.

¹³⁴⁵ The optimal neural network architecture for this channel is found to consist of 7 hidden

¹³⁴⁶ layers with 80 nodes each. The inputdata set includes one million events, 10% of which is used

¹³⁴⁷ for testing, the other 90% for training. Training is found to converge after around 150 epochs.

1348 To evaluate the performance of the model, the predicted p_T spectrum is compared to the
 1349 truth Higgs p_T in Figure 14.18. Once again, (a) of 14.18 shows a scatterplots of predicted vs
 1350 truth p_T , where the color of each point corresponds to the log of the relative KDE at that point.
 1351 Each column of the the histogram in (b) is normalized to unity, to better demonstrate the output
 1352 of the NN for each slice of truth p_T .

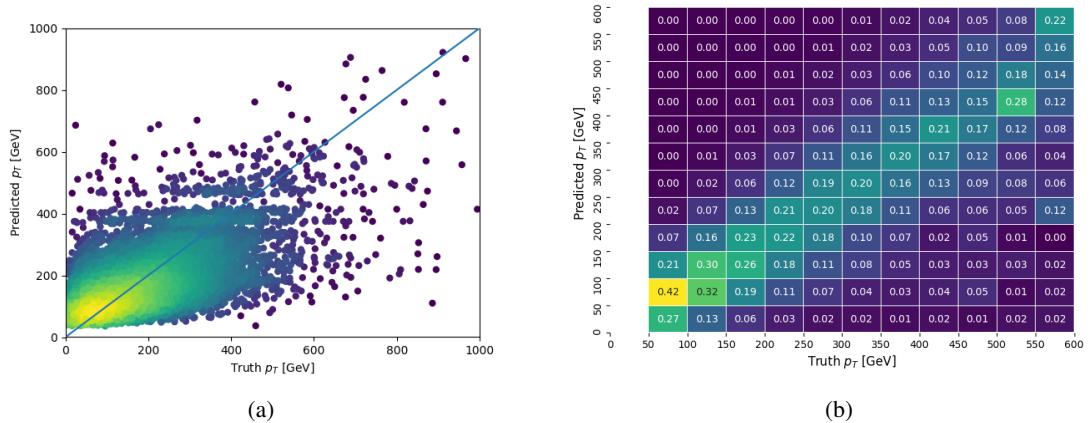


Figure 14.18: The regressed Higgs momentum spectrum as a function of the truth p_T for 3lS $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized to one.

1353 Figure 14.19 shows (a) the output of the NN for events with truth p_T less than and greater
 1354 than 150 GeV and (b) the ROC curve demonstrating how well the NN distinguishes high and low
 1355 p_T events.

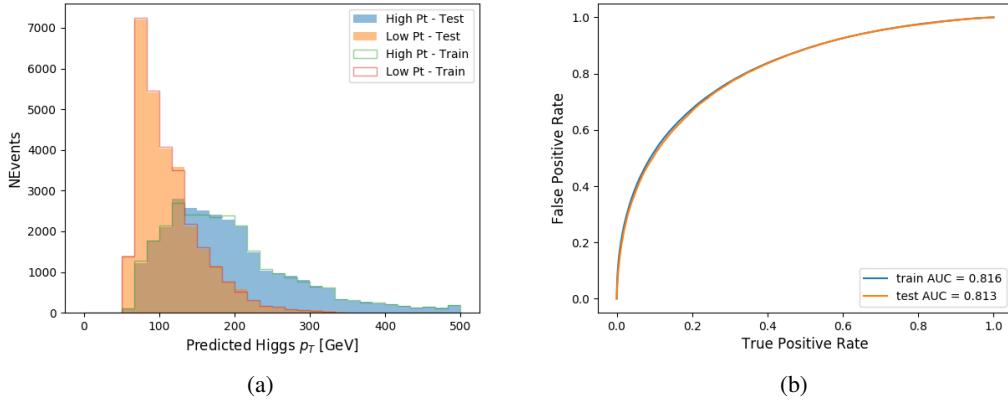


Figure 14.19: (

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.

1356

14.4.3 3l Fully-leptonic Channel

1357

The features listed in 33 are used to construct a model for predictin the Higgs p_T for 3lF events.

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

Table 33: Input features for reconstructing the Higgs p_T spectrum for 3lF events

1358

l_{H0} and l_{H1} respresent the two leptons identified by the Higgs reconstruction model as

1359 originating from the Higgs, while ℓ_T is the other lepton in the event. The Higgs Reco Score and
 1360 b-jet Reco Score are the outputs of the Higgs reconstruction algorithm, and b-jet identification
 1361 algorithm, respectively.

1362 The optimal neural network architecture for this channel is found to consist of 5 hidden
 1363 layers with 40 nodes each. The input data set includes 400,000 events, 10% of which is used for
 1364 testing, the other 90% for training. Training is found to converge after around 150 epochs.

1365 The predicted transverse momentum, as a function of the truth p_T , is shown in Figure
 1366 [14.20](#).

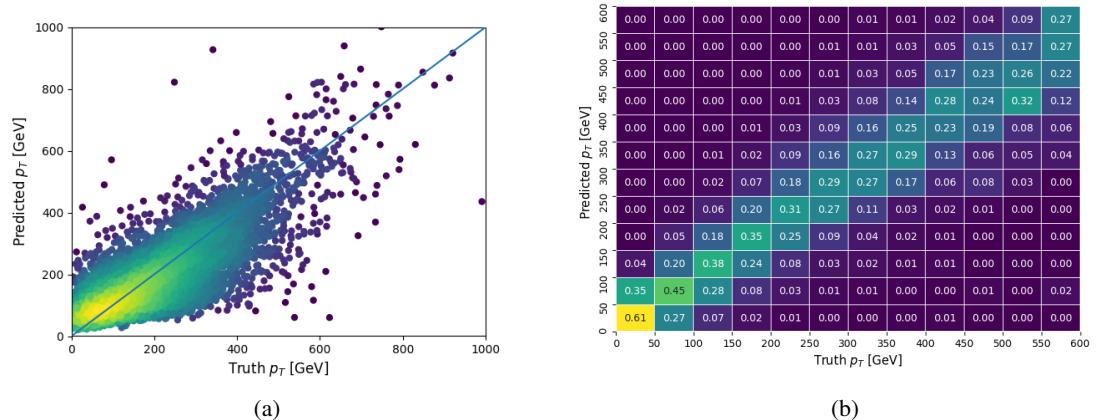


Figure 14.20: The regressed Higgs momentum spectrum as a function of the truth p_T for 3lF $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a histogram where each column has been normalized to one.

1367 When split into high and low p_T , based on a cutoff of 150 GeV, the

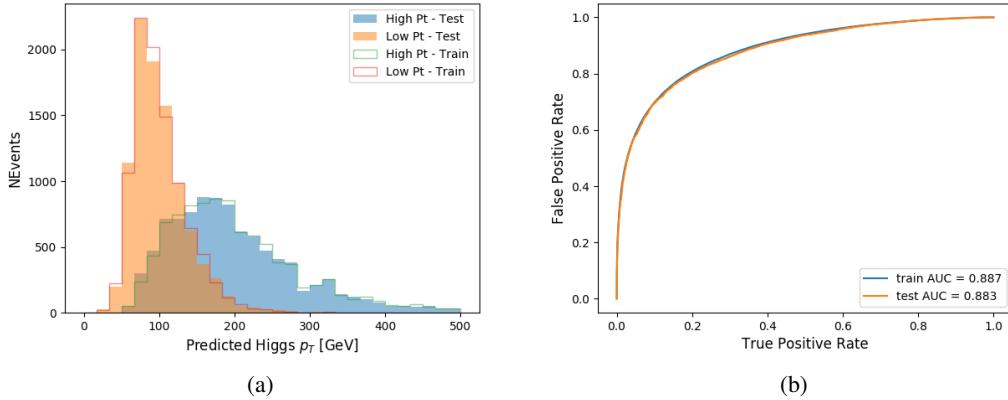


Figure 14.21: (a) shows the reconstructed Higgs p_T for 3lF events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

1368 14.5 3l Decay Mode

1369 In the 3l channel, there are two possible ways for the Higgs to decay, both involving intermediate
 1370 W boson pairs: Either both W bosons decay leptonically, in which case the reconstructed decay
 1371 consists of two leptons (referred as the fully-leptonic 3l channel), or one W decays leptonically
 1372 and the other hadronically, giving two jets and one lepton in the final state (referred to as the
 1373 semi-leptonic 3l channel). In order to accurately reconstruct the Higgs, it is necessary to identify
 1374 which of these decays took place for each 3l event.

1375 The kinematics of each event, along with the output scores of the Higgs and top recon-
 1376 struction algorithms, are used to distinguish these two possible decay modes. The particular
 1377 inputs used are listed in Table 34.

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

1378 Here l_0 is the opposite charge lepton, l_1 and l_2 are the two SS leptons order by ΔR
 1379 from lepton 0. score3lF and score3lS are the outputs of the 3lS and 3lF Higgs reconstruction
 1380 algorithms, while topScore is the output of the b-jet identification algorithm.

1381 A neural network with 5 hidden layers, each with 50 nodes, is trained to distinguish these
 1382 two decay modes. The output of the model is summarized in Figure 14.22.

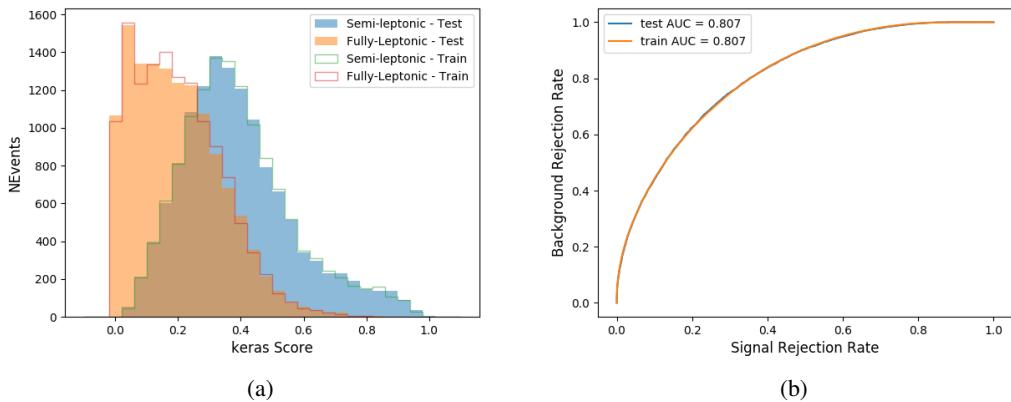


Figure 14.22: (a) shows the output of the decay separation NN for Semi-leptonic (blue) and Fully-leptonic (orange) 3l events, scaled to equal area. (b) shows the ROC curve for those two sets of events.

1383 A cutoff of 0.23 is determined to be optimal for separating 3lS and 3lF in the fit.

1384 15 Signal Region Definitions

1385 Events are divided into two channels based on the number of leptons in the final state: one
1386 with two same-sign leptons, the other with three leptons. The 3l channel includes events where
1387 two leptons originated from the Higgs boson as well as events where only one of the leptons
1388 originated from the Higgs. This motivates splitting the 3l channel into semi-leptonic, and fully
1389 leptonic channels, after an event preselection has been applied.

1390 15.1 Pre-MVA Event Selection

1391 A preselection is applied to define orthogonal analysis channels based on the number of leptons
1392 in each event. For the 2lSS channel, the following preselection is used:

- 1393 • Two very tight, same-charge, light leptons with $p_T > 20 \text{ GeV}$
- 1394 • ≥ 4 reconstructed jets, ≥ 1 b-tagged jets
- 1395 • No reconstructed tau candidates

1396 The event yield after the 2lSS preselection has been applied, for MC and data at 79.8 fb^{-1} ,
1397 is shown in Table 15.1.

Process	Yield
t̄tH high p _T	41 ± 5
t̄tH low p _T	71 ± 8
t̄tW	450 ± 70
t̄t(Z/γ*)	91 ± 11
t̄tll low mass	10 ± 6
Rare Top	20 ± 12
VV	42 ± 22
tZ	10 ± 5
QMisID	44.7 ± 2.7
Fakes int. conv	47 ± 26
Fakes ext. conv	46 ± 44
Fakes HF e	45 ± 23
Fakes HF μ	250 ± 50
Three top	2.2 ± 1.1
Four top	5.64 ± 0.31
t̄tWW	10.9 ± 0.6
tW	0.0 ± 0.0
WtZ	9.1 ± 0.8
VVV	0.30 ± 0.05
VH	0.6 ± 1.0
Total	1170 ± 120
Data	1108

Table 35: Yields of the 2lSS preselection region

1398

Figure 15.1. Good general agreement is found.

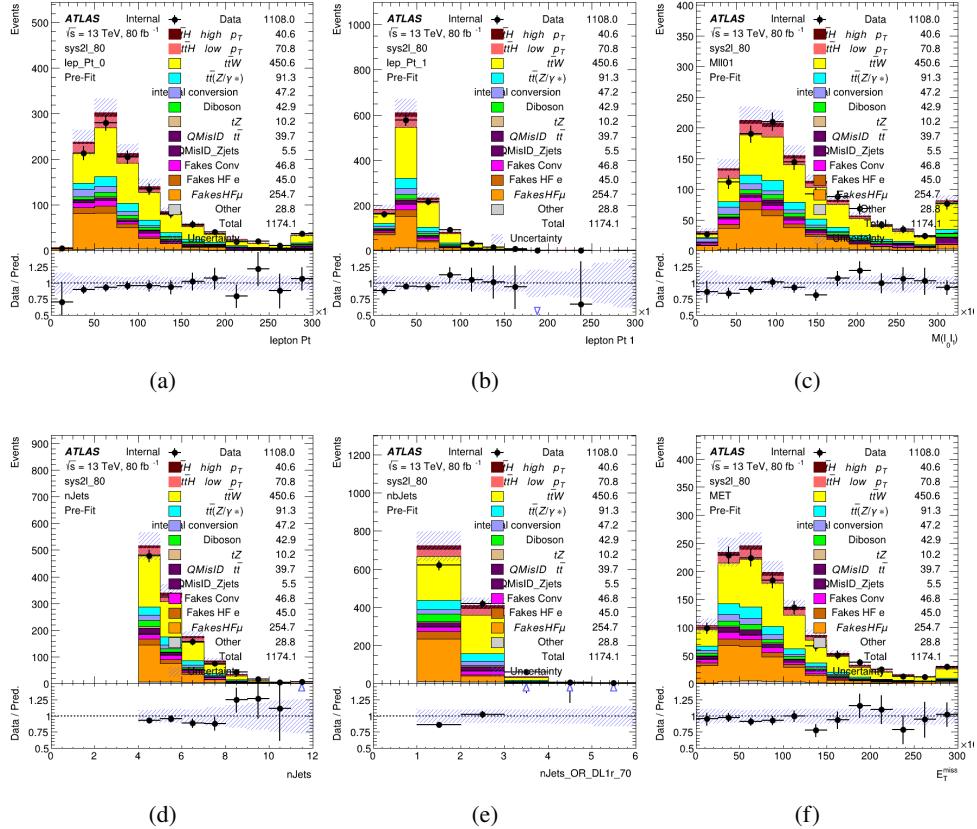


Figure 15.1: Data/MC comparisons of the 2LSS pre-selection region. (a) and (b) show the p_T of leptons 0 and 1, (c) shows the invariant mass of lepton 0 and 1, (d) shows the jet multiplicity, (e) the b-tagged jet multiplicity, and (f) the missing transverse energy.

1399

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

1400

- Three light leptons with total charge ± 1
- Same charge leptons are required to be very tight, with $p_T > 20$ GeV
- Opposite charge lepton must be loose, with $p_T > 10$ GeV
- ≥ 2 reconstructed jets, ≥ 1 b-tagged jets

1401

- 1404 • No reconstructed tau candidates
- 1405 • $|M(l^+l^-) - 91.2 \text{ GeV}| > 10 \text{ GeV}$ for all opposite-charge, same-flavor lepton pairs
- 1406 The event yield after the 3l preselection has been applied, for MC and data at 79.8 fb^{-1} ,
- 1407 is shown in Table 15.1.

Process	Yield
t̄tH high p _T	20.5 ± 2.3
t̄tH low p _T	33.6 ± 3.8
t̄tW	138 ± 18
t̄tZ/γ	80 ± 9
t̄tlllowmass	3.5 ± 2.0
rareTop	22 ± 12
VV	39 ± 19
tZ	9.2 ± 4.5
QMisID	1.8 ± 0.6
Fakes int. conv	31 ± 17
Fakes ext. conv	14 ± 11
Fakes HF e	20 ± 10
Fakes HF μ	102 ± 22
Three top	0.96 ± 0.48
Four top	6.17 ± 0.35
t̄tWW	5.46 ± 0.33
tW	0.0 ± 0.0
WtZ	8.7 ± 0.6
VVV	0.81 ± 0.11
VH	0.0 ± 0.0
Total	512 ± 48
Data	535

Table 36: Yields of the 3l preselection region.

Table 37: Yields of the 3l preselection region.

1408 Comparisons of kinematic distributions for data and MC in this region are shown in Figure

1409 15.2.

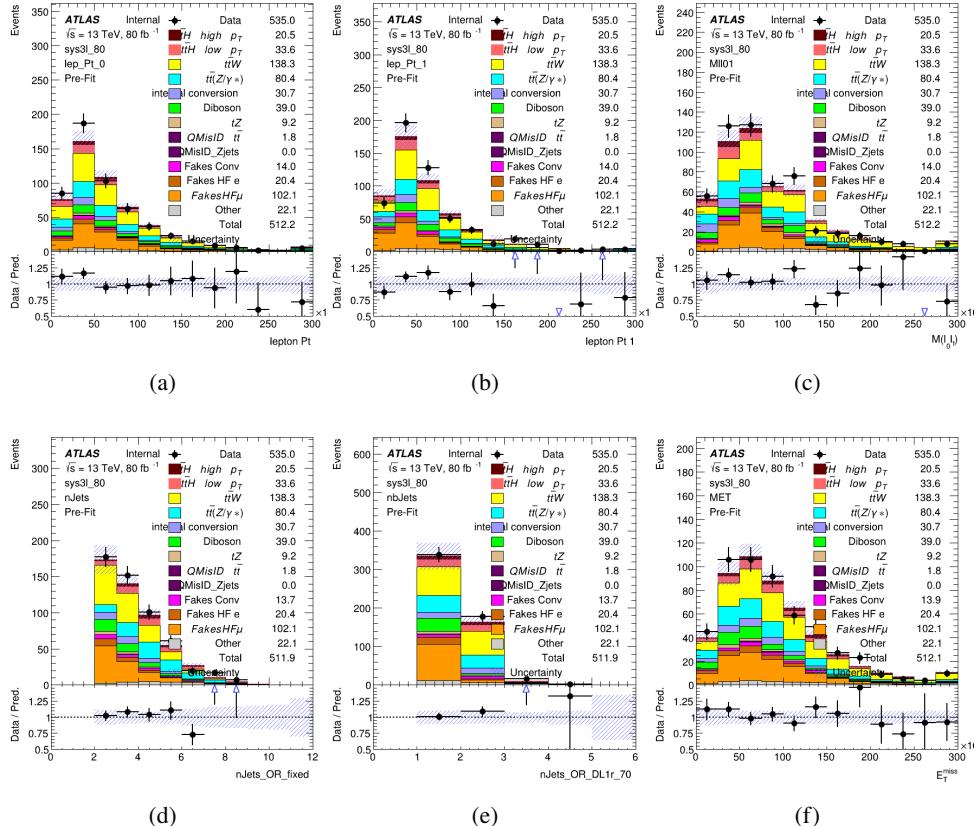


Figure 15.2: Data/MC comparisons of the 3l pre-selection region. (a) and (b) show the p_T of leptons 0 and 1, (c) shows the invariant mass of lepton 0 and 1, (d) shows the jet multiplicity, (e) the b-tagged jet multiplicity, and (f) the missing transverse energy.

1410 15.2 Event MVA

1411 Separate multi-variate analysis techniques (MVAs) are used in order to distinguish signal events
 1412 from background for each analysis channel - 2lSS, 3l semi-leptonic (3lS), and 3l fully leptonic
 1413 (3lF). Here events with three leptons are split into 3lS and 3lF based on the model described
 1414 in 14.5. In particular, Boosted Decision Tree (BDT) algorithms are produced with XGBoost

¹⁴¹⁵ [36] are trained using the kinematics of signal and background events derived from Monte Carlo
¹⁴¹⁶ simulations. Events are weighted in the BDT training by the weight of each Monte Carlo event.

¹⁴¹⁷ Because the background composition differs for events with a high reconstructed Higgs p_T
¹⁴¹⁸ compared to events with low reconstructed Higgs p_T , separate MVAs are produced for high and
¹⁴¹⁹ low p_T regions. This is found to provide better significance than attempting to build an inclusive
¹⁴²⁰ model, as demonstrated in appendix C.2. A cutoff of 150 GeV is used. This gives a total of 6
¹⁴²¹ background rejection MVAs - explicitly, 2lSS high p_T , 2lSS low p_T , 3lS high p_T , 3lS low p_T ,
¹⁴²² 3lF high p_T , and 3lF low p_T .

¹⁴²³ The following features are used in both the high and low p_T 2lSS BDTs:

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

Table 38: Input features used to distinguish signal and background events in the 2lSS channel.

¹⁴²⁴ While for each of the 3l 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	$\text{jet } \eta \ 0$	$\text{jet } \eta \ 1$
	$\text{jet } \phi \ 0$	$\text{jet } p_T \ 0$
	$\text{jet } p_T \ 1$	$\text{Lepton } \eta \ 1$
$\text{Lepton } \eta \ 2$	$\text{Lepton } \phi \ 0$	$\text{Lepton } \phi \ 1$
$\text{Lepton } \phi \ 2$	$\text{Lepton } p_T \ 0$	$\text{Lepton } p_T \ 1$
$\text{Lepton } p_T \ 2$	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 39: Input features used to distinguish signal and background events in the 3l channel.

1425 Modelling of each of these input features is verified in Appendix C.2 by comparing data
 1426 and MC for 79.8 fb^{-1} . The BDTs are produced with a maximum tree depth of 6, using AUC as
 1427 the target loss function. The BDT response distribution and ROC curve for each model is shown
 1428 in Figures 15.3-15.5.

2lSS

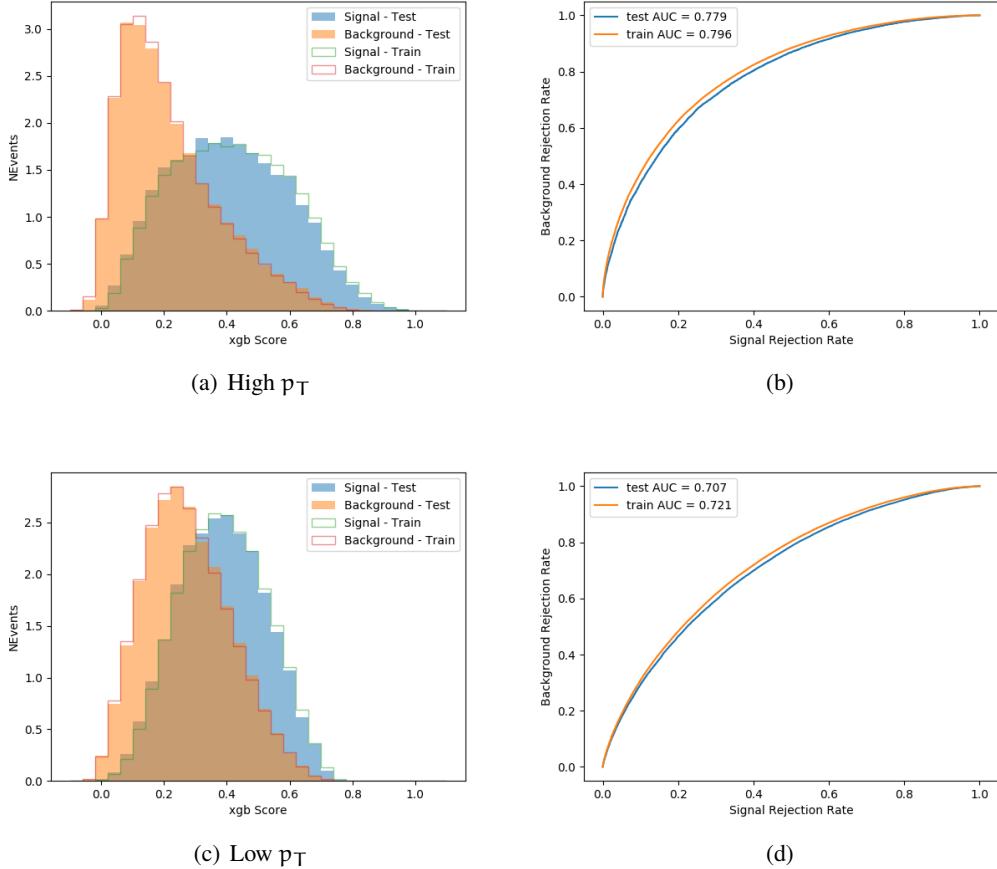


Figure 15.3: Output BDT scores of training and testing data for signal (blue) and background (orange) for 2lSS events with (a) high regressed Higgs p_T and (b) low regressed Higgs p_T . (b) and (d) show the ROC curve for the 2lSS high and low p_T models, respectively.

3l - Semileptonic

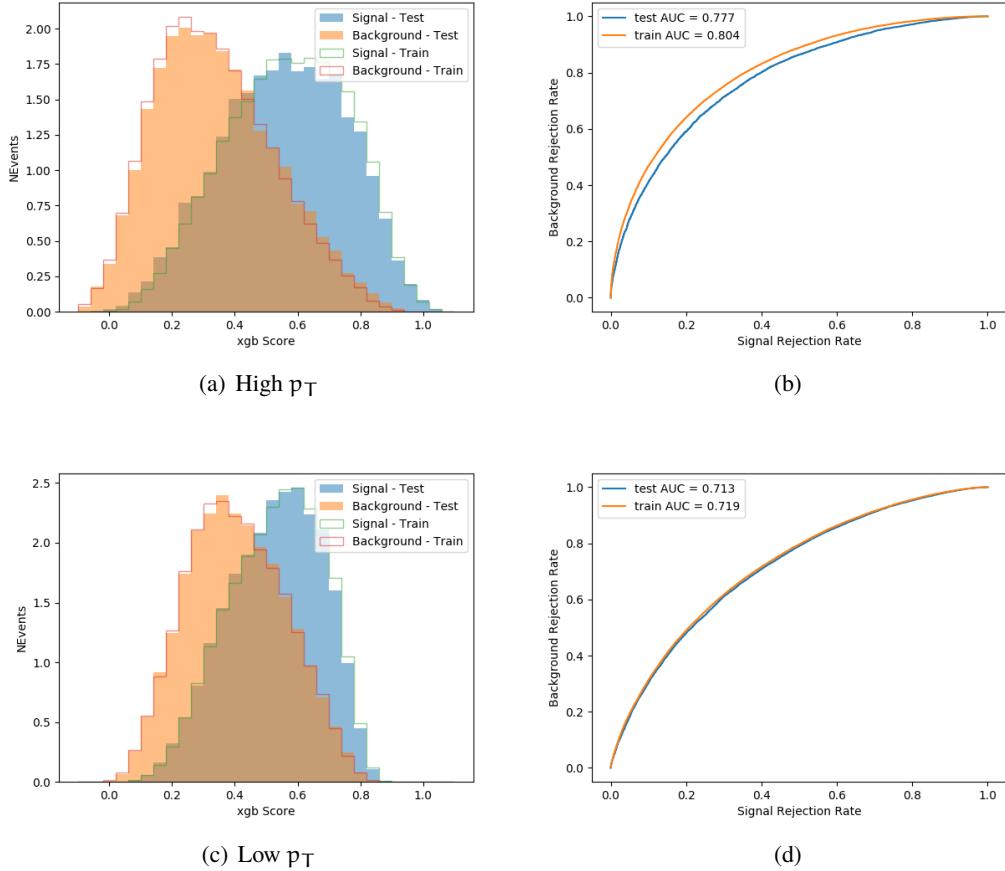


Figure 15.4: Output BDT scores of training and testing data for signal (blue) and background (orange) for 3lS events with (a) high regressed Higgs p_T and (b) low regressed Higgs p_T . (b) and (d) show the ROC curve for the 3lS high and low p_T models, respectively.

3l - Fully Leptonic

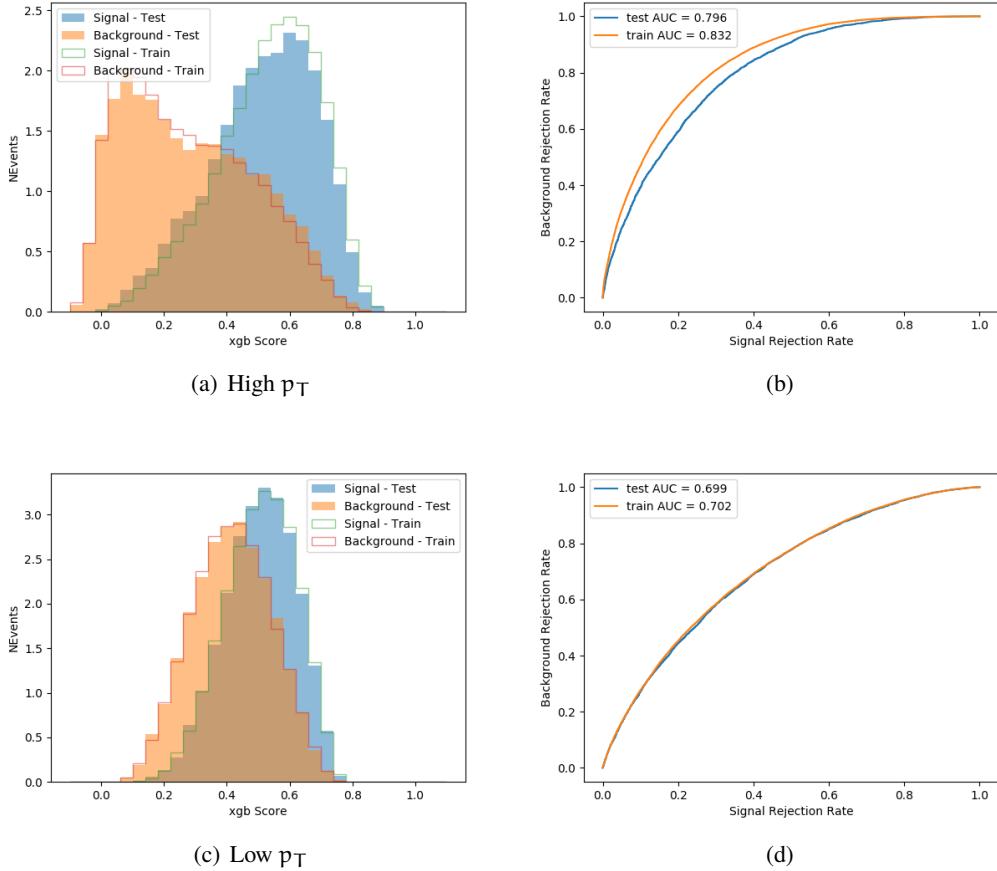


Figure 15.5: Output BDT scores of training and testing data for signal (blue) and background (orange) for 3lF events with (a) high regressed Higgs p_T and (b) low regressed Higgs p_T . (b) and (d) show the ROC curve for the 3lF high and low p_T models, respectively.

1429 Output distributions of each MVA, comparing MC predictions to data at 79.8 fb^{-1} are
 1430 shown in figures 15.6-15.2.

High p_T Background Rejection BDTs

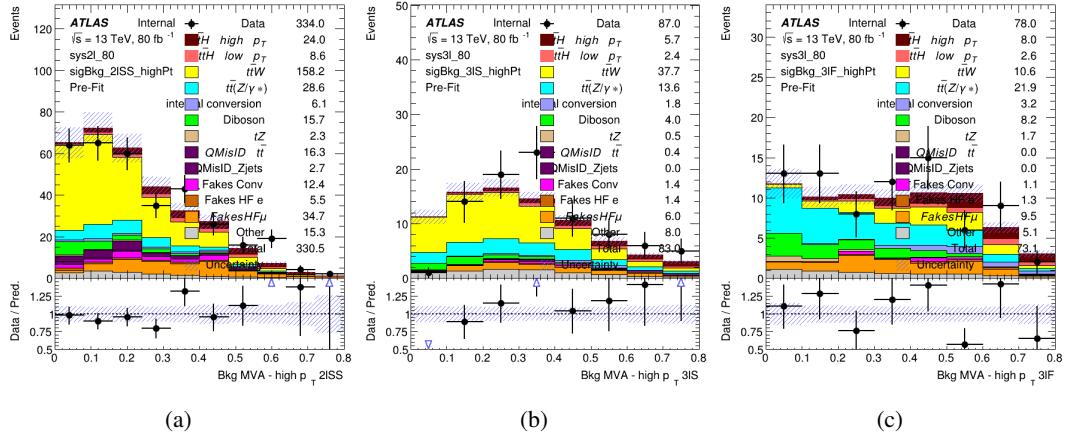


Figure 15.6: Output score of the high p_T BDTs in the (a) 2lSS, (b) 3lS, and (c) 3lF channels

Low p_T Background Rejection BDTs

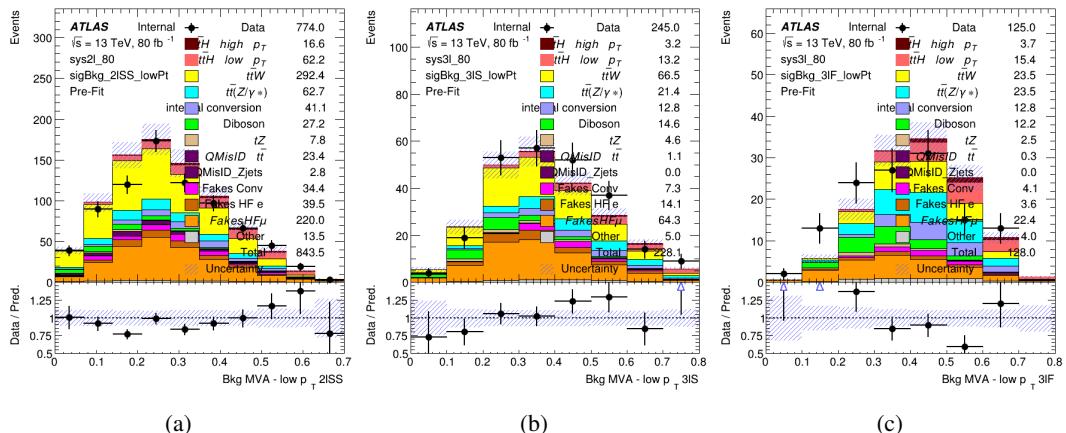


Figure 15.7: Output score of the low p_T BDTs in the (a) 2lSS, (b) 3lS, and (c) 3lF channels

1431 15.3 Signal Region Definitions

1432 Once pre-selection has been applied, channels are further refined based on the MVAs described
 1433 above. The output of the model described in Section 14.5 is used to separate the three channel
 1434 into two - Semi-leptonic and Fully-leptonic - based on the predicted decay mode of the Higgs
 1435 boson. This leaves three orthogonal signal regions - 2lSS, 3lS, and 3lF.

1436 For each event, depending on the number of leptons as well as whether the p_T of the Higgs
 1437 is predicted to be high (> 150 GeV) or low (< 150 GeV), a cut on the appropriate background
 1438 rejection MVA is applied. The particular cut values, listed in Table 40, are determined by
 1439 maximizing S/\sqrt{B} in each region.

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

1440 The event preselection and MVA selection listed in Table 40 are used define the three
 1441 signal regions used in the fit. These signal region definitions are summarized in Table 41.

Region	Selection
2lSS	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
3lS	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
3lF	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 41: Selection applied to define the three signal regions used in the fit.

1442 16 Systematic Uncertainties

1443 The systematic uncertainties that are considered are summarized in Table 42. These are imple-
 1444 mented in the fit either as a normalization factors or as a shape variation or both in the signal
 1445 and background estimations. The numerical impact of each of these uncertainties is outlined in
 1446 section 17.

Table 42: 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

¹⁴⁴⁷ The uncertainty in the combined integrated luminosity is derived from a calibration of the
¹⁴⁴⁸ luminosity scale using x-y beam-separation scans performed for 13 TeV proton-proton data [15],
¹⁴⁴⁹ [39].

¹⁴⁵⁰ The experimental uncertainties are related to the reconstruction and identification of light
¹⁴⁵¹ leptons and b-tagging of jets, and to the reconstruction of E_T^{miss} .

1452 The sources which contribute to the uncertainty in the jet energy scale [45] are decomposed
 1453 into uncorrelated components and treated as independent sources in the analysis. This method
 1454 decomposes the uncertainties into 30 nuisance parameters included in the fit. A similar method
 1455 is used to account for jet energy resolution (JER) uncertainties, and 8 JER uncertainty components
 1456 are unclued as NPs in the fit.

1457 The uncertainties in the b-tagging efficiencies measured in dedicated calibration analyses
 1458 [32] are also decomposed into uncorrelated components. The large number of components for
 1459 b-tagging is due to the calibration of the distribution of the BDT discriminant.

1460 As mentioned in Section 12.2, a normalization corrections and uncertainties on the estim-
 1461 ates of non-prompt leptons backgrounds are derived using data driven techniques, decribed in
 1462 detail in [7]. These are derived from a likelihood fit over various non-prompt enriched control
 1463 regions, targeting several sources of non-prompt light leptons separately: external conversion
 1464 electrons, internal conversion electrons, electrons from heavy flavor decays, and muons from
 1465 heavy flavor decays.

1466 The normalization factor and uncertainty applied to each source of non-prompt leptons is
 1467 summarized in Table 16

Processss	Normalization Factor
NF_e^{ExtCO}	1.70 ± 0.51
NF_e^{IntCO}	0.75 ± 0.26
NF_e^{HF}	1.09 ± 0.32
NF_{μ}^{HF}	1.28 ± 0.17

Table 43: Normalization factors - with statistical and systematic uncertainties - derived from the fit over fake control regions for each source of non-prompt leptons considered.

1468 In addition to those derived from the control regions, several additional uncertainties are
1469 assigned to the non-prompt lepton background. An additional 25% uncertainty on material
1470 conversions is assigned, based on the comparison between data and MC in a region where a
1471 loose electron fails the photon conversion veto. A shape uncertainty of 15% (6%) is assigned to
1472 the HF non-prompt electron (muon) background based on a comparison between data and MC
1473 where the second leading electron (muon) is only required to be loose. As the contribution from
1474 light non-prompt leptons is small, about 10% percent of the contribution from HF non-prompt
1475 leptons, it is derived from the agreement between data and simulation in a LF enriched region at
1476 low values of the non-prompt lepton BDT. The resulting uncertainty is 100%, and is taken to be
1477 uncorrelated between internal and material conversions.

1478 Theoretical uncertainties applied to MC predictions, including cross section, PDF, and
1479 scale uncertainties are taken from theory calculations for the predominate prompt backgrounds.
1480 Following the nominal $t\bar{t}H - ML$ analysis, a 50% uncertainty is applied to Diboson to account
1481 for the large uncertainty in estimating $VV +$ heavy flavor. The other “rare” background processes
1482 - including tZ , rare top processes, $ttWW$, WtZ , VVV , $tHjb$ and WtH - are assigned an overall
1483 50% normalization uncertainty as well. The theory uncertainties applied to the MC estimates
1484 are summarized in Table 44.

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

Table 44: Summary of theoretical uncertainties for MC predictions in the analysis.

1485 Additional uncertainties to account for t̄W mismodelling are also applied. These include
 1486 a “Generator” uncertainty, based on a comparison between the nominal Sherpa 2.2.5 sample,
 1487 and the formerly used aMC@NLO sample, and an “Extra radiation” uncertainty, which includes
 1488 renormalisation and factorisation scale variations of the Sherpa 2.2.5 sample.

1489 17 Results

1490 A maximum likelihood fit is performed simultaneously over the reconstructed Higgs p_T spectrum
 1491 in the three signal regions, 2lSS, 3lS, and 3lF. The signal is split into high and low p_T samples,
 1492 based on whether the truth p_T of the Higgs is above or below 150 GeV. The parameters $\mu_{t\bar{t}H \text{ high } p_T}$
 1493 and $\mu_{t\bar{t}H \text{ low } p_T}$, where $\mu = \sigma_{\text{observed}}/\sigma_{\text{SM}}$, are extracted from the fit, signifying the difference
 1494 between the observed value and the theory prediction. Unblinded results are shown for the 79.8
 1495 fb⁻¹ data set, as well as MC only projections of results using the full Run-2, 139 fb⁻¹ dataset.

1496 As described in Section 16, there are 229 systematic uncertainties that are considered
 1497 as NPs in the fit. These NP s are constrained by Gaussian or log-normal probability density
 1498 functions. The latter are used for normalisation factors to ensure that they are always positive.
 1499 The expected number of signal and background events are functions of the likelihood. The prior
 1500 for each NP is added as a penalty term, decreasing the likelihood as it is shifted away from its
 1501 nominal value.

1502 17.1 Results - 79.8 fb^{-1}

1503 As the data collected from 2015-2017 has been unblinded for $t\bar{t}H$ – ML channels, represent-
 1504 ing 79.8 fb^{-1} , those events are unblinded. The predicted Higgs p_T spectrum is fit to data
 1505 simultaneously in each of the three signal regions shown in Figure 17.1.

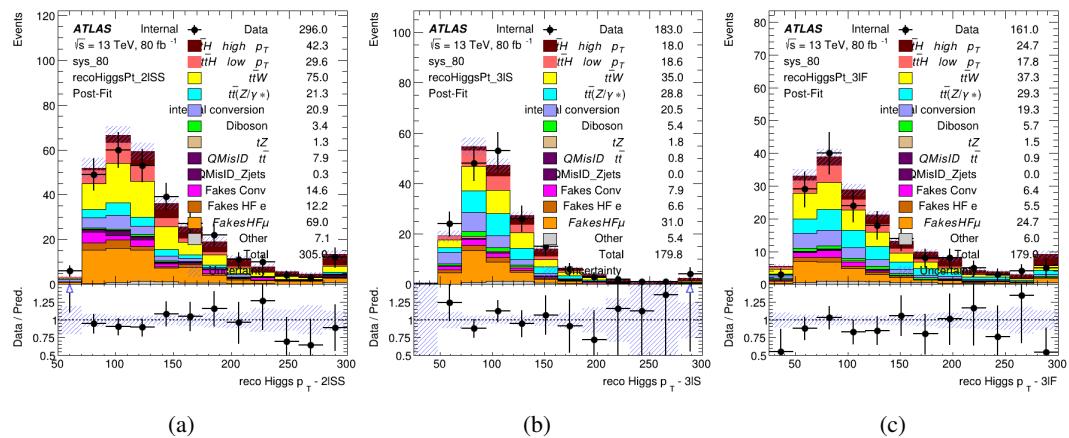


Figure 17.1: Post-fit distributions of the reconstructed Higgs p_T in the three signal regions, (a) 2lSS, (b) 3lS, and (c) 3lF, for 79.8 fb^{-1} of MC

1506 A post-fit summary of the fitted regions is shown in Figure 17.2.

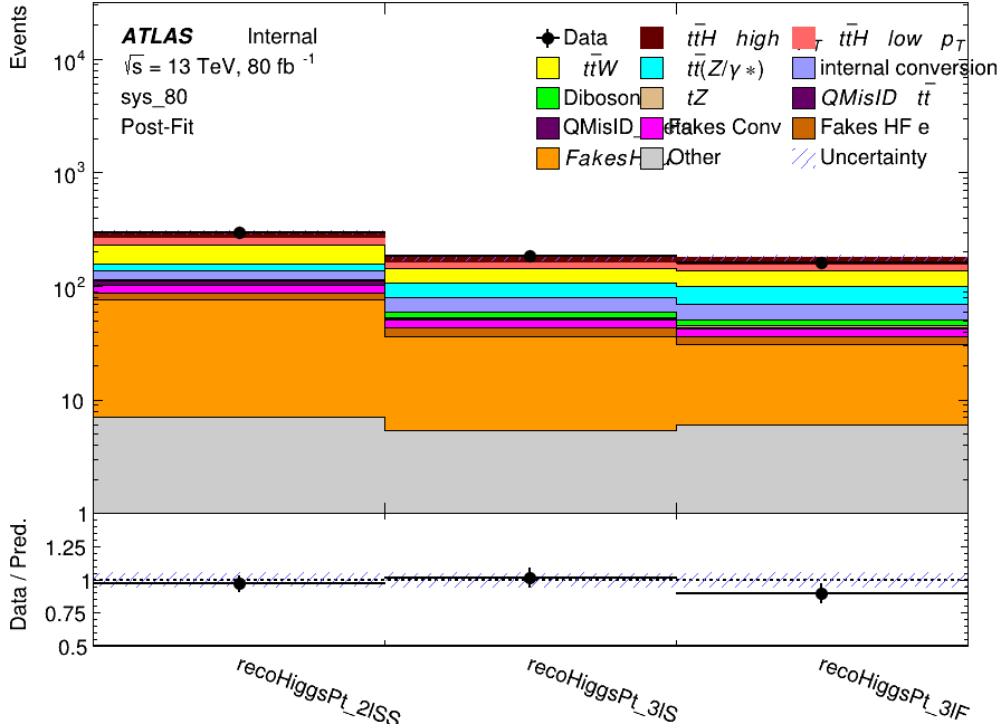


Figure 17.2: Post-fit summary of the yields in each signal region.

1507 The measured μ values for high and low p_T Higgs production obtained from the fit
 1508 are shown in 45. A significance of 1.7σ is observed for $t\bar{t}H$ high p_T , and 2.1σ is measured for
 1509 $t\bar{t}H$ low p_T .

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

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

Table 45: Best fit μ values for $t\bar{t}H$ high p_T and $t\bar{t}H$ low p_T , where $\mu = \sigma_{\text{obs}}/\sigma_{\text{pred}}$

1510 The most prominent sources of systematic uncertainty, as measured by their impact on
 1511 $\mu_{t\bar{t}H \text{ high } p_T}$, are summarized in Table 46.

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

Table 46: Summary of the most significant sources of systematic uncertainty on the measurement of t̄H high p_T.

1512 The most significant sources of uncertainty on the measurement of t̄H - low p_T are shown
 1513 in Table 47.

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

Table 47: Summary of the most significant sources of systematic uncertainty on the measurement of t̄H low p_T.

1514 The ranking and impact of those nuisance parameters with the largest contribution to the

1515 overall uncertainty is shown in Figure 17.3.

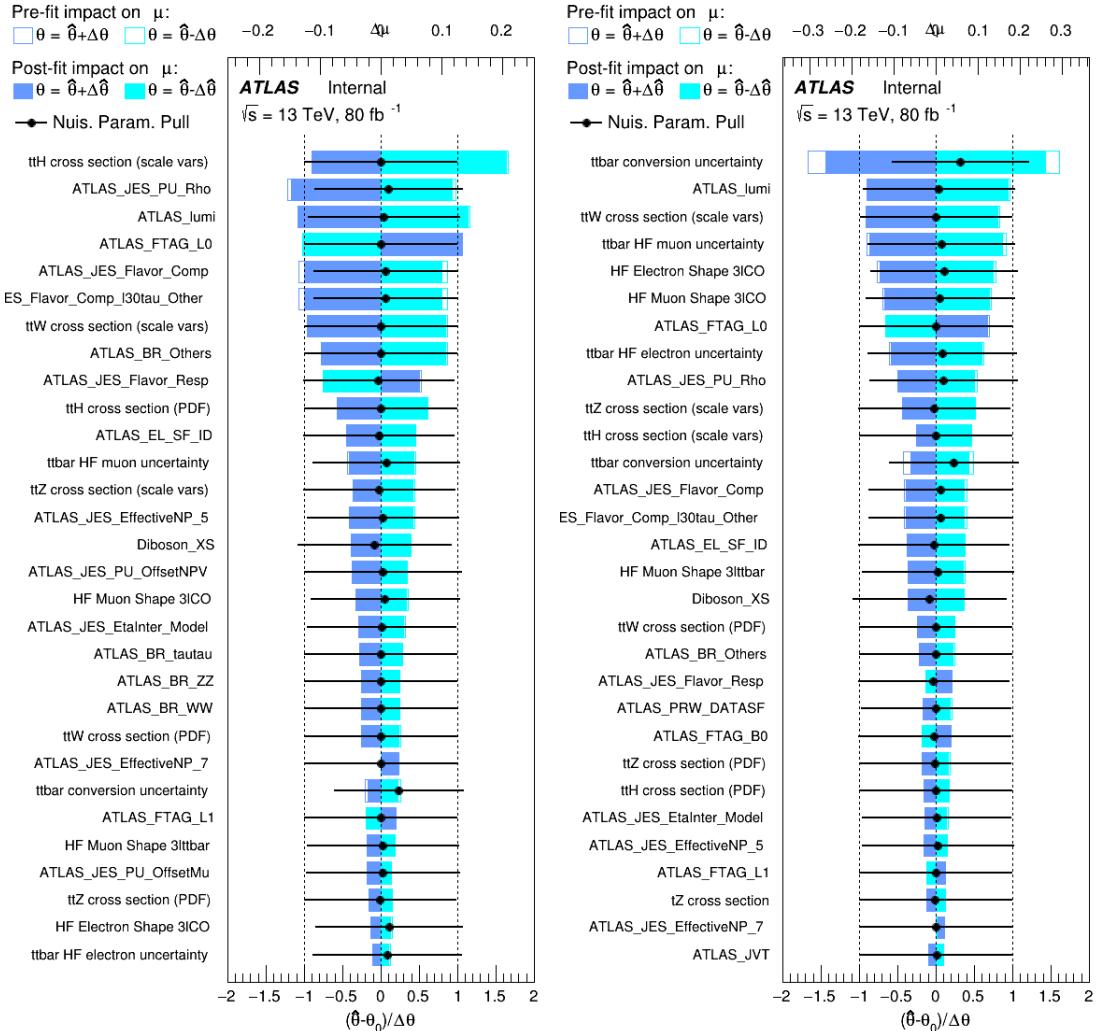


Figure 17.3: Impact of systematic uncertainties on the measurement of high p_T (left) and low p_T (right) $t\bar{t}H$ events

1516

The background composition of each of the fit regions is shown in Figure 17.4.

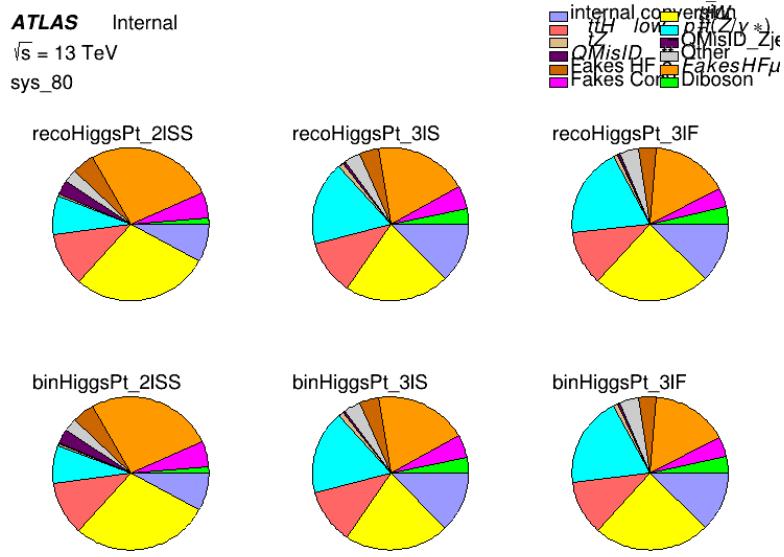


Figure 17.4: Background composition of the fit regions.

1517 17.2 Projected Results - 139 fb^{-1}

1518 As data collected in 2018 has not yet been unblinded for $t\bar{t}H$ – ML at the time of this note, data
 1519 from that year remains blinded. Instead, an Asimov fit is performed - with the MC prediction
 1520 being used both as the SM prediction as well as the data in the fit - in order to give expected
 1521 results.

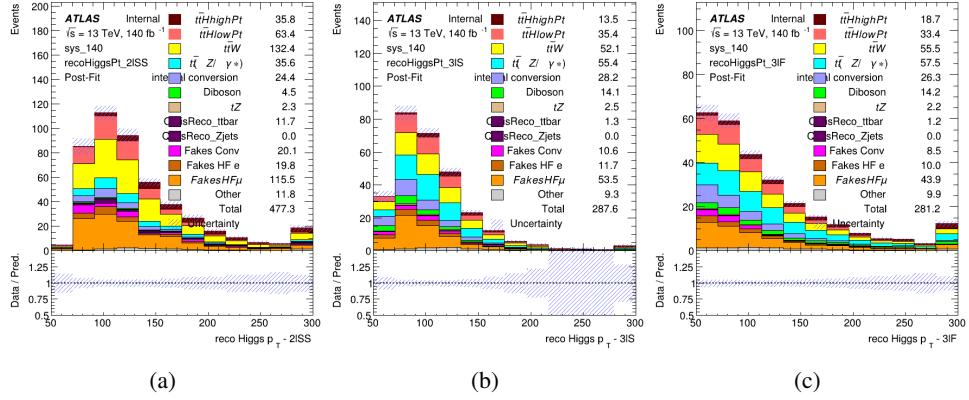


Figure 17.5: Blinded post-fit distributions of the reconstructed Higgs p_T in the three signal regions, (a) 2lSS, (b) 3lS, and (c) 3lF, for 139 fb^{-1} of data

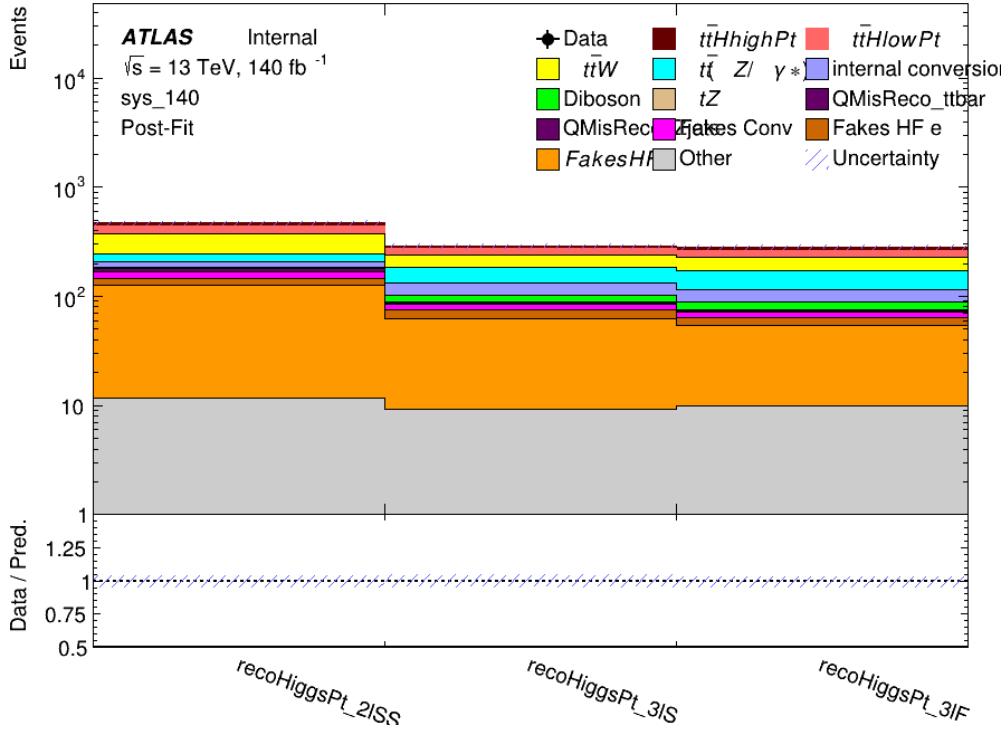


Figure 17.6: Post-fit summary of fit.

1523 shown in 48. A significance of 2.0σ is expected for $t\bar{t}H$ high p_T , and a projected significance

1524 2.3σ is extracted for $t\bar{t}H$ low p_T .

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

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

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

1525 The most prominent sources of systematic uncertainty, as measured by their impact on

1526 $\mu_{t\bar{t}H \text{ high } p_T}$, are summarized in Table 49.

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

Table 49: Summary of the most significant sources of systematic uncertainty on the measurement of $t\bar{t}H$ high p_T .

1527 The most significant sources of systematic uncertainty on $t\bar{t}H$ low p_T are summarized in

1528 Table 50.

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

Table 50: Summary of the most significant sources of systematic uncertainty on the measurement of t̄H low p_T.

1529 The ranking and impact of those nuisance parameters with the largest contribution to the
 1530 overall uncertainty is shown in Figure 17.7.

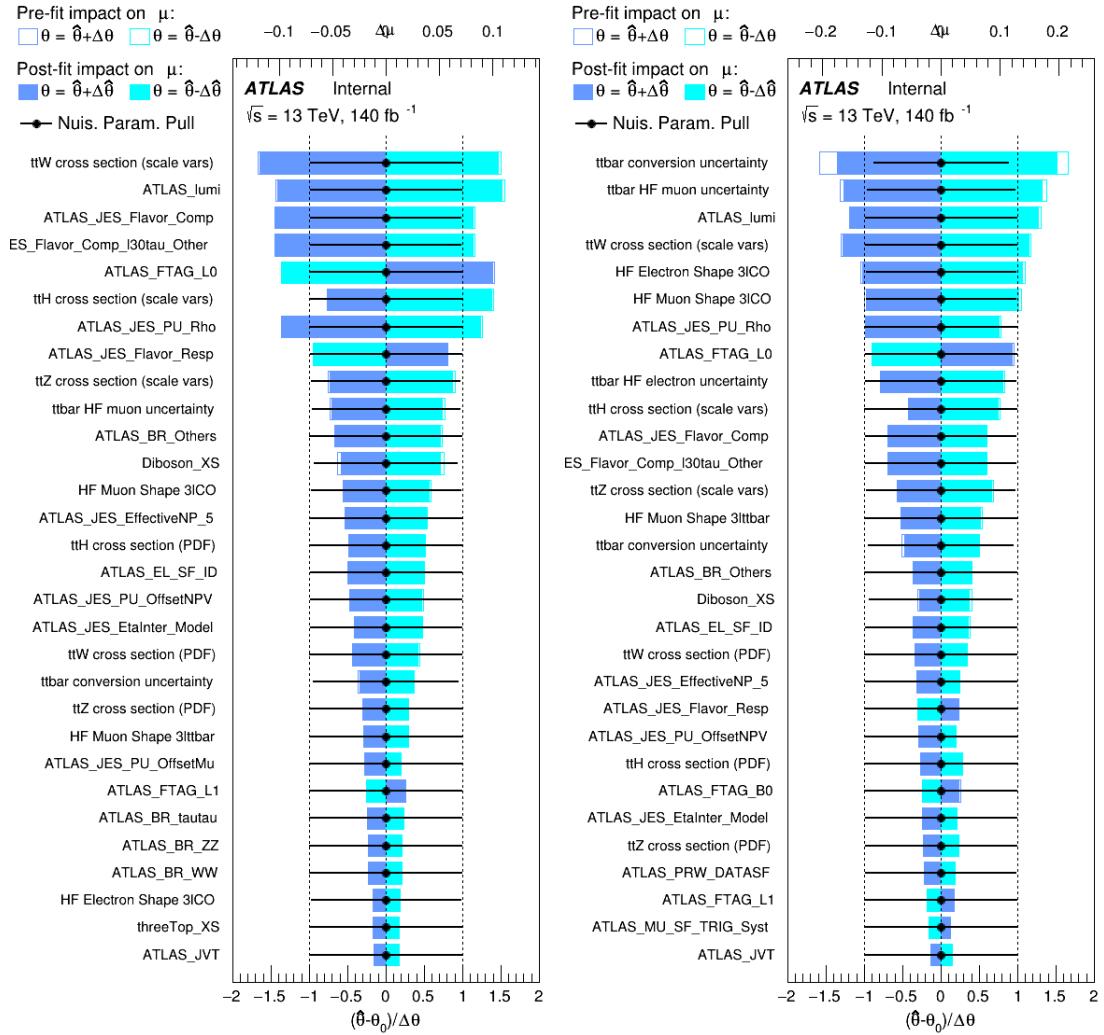


Figure 17.7: Impact of systematic uncertainties on the measurement of high p_T (left) and low p_T (right) $t\bar{t}H$ events

1531

The background composition of each of the fit regions is shown in Figure 17.8.

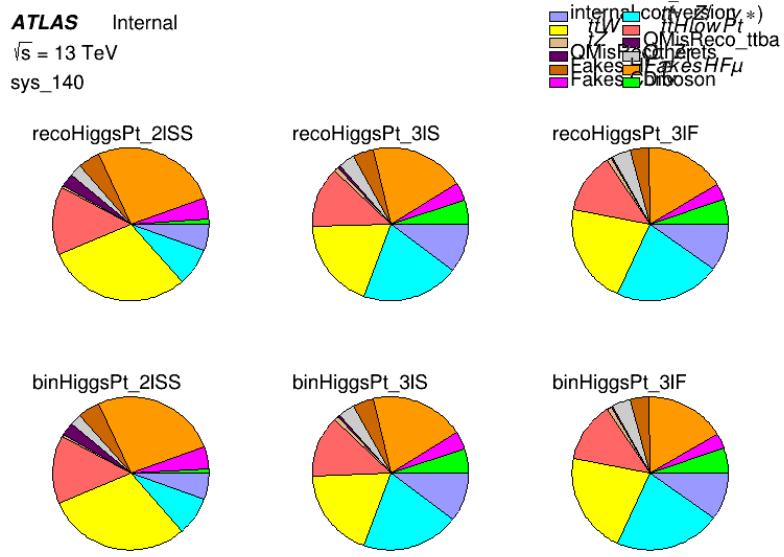


Figure 17.8: Background composition of the fit regions.

1532 Part VI

1533 Conclusion

1534 As search for the effects of dimension-six operators on $t\bar{t}H$ production is performed. An effective
 1535 field theory approached is used to parameterize the effects of high energy physics on the Higgs
 1536 momentum spectrum. The momentum spectrum is reconstructed using various MVA techniques,
 1537 and the limits on dimension-six operators are limited to X.

References

- 1538 [1] ATLAS Collaboration, *Observation of a new particle in the search for the Standard*
1539 *Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716** (2012) 1,
1540 arXiv: [1207.7214 \[hep-ex\]](https://arxiv.org/abs/1207.7214).
- 1541 [2] R. Oerter, *The theory of almost everything : the Standard Model, the unsung triumph of*
1542 *modern physics*, Pi Press, 2006, ISBN: 978-0-13-236678-6.
- 1543 [3] P. W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*,
1544 Phys. Rev. Lett. 13 (Oct, 1964) 508-509 ().
- 1545 [4] J. Goldstone, A. Salam and S. Weinberg, *Broken Symmetries*,
1546 Phys. Rev. **127** (3 1962) 965,
1547 URL: <https://link.aps.org/doi/10.1103/PhysRev.127.965>.
- 1548 [5] J. Ellis, *Higgs Physics*,
1549 (2013) 117, 52 pages, 45 figures, Lectures presented at the ESHEP 2013 School of
1550 High-Energy Physics, to appear as part of the proceedings in a CERN Yellow Report,
1551 URL: <https://cds.cern.ch/record/1638469>.
- 1552 [6] T. A. Collaboration, *Observation of Higgs boson production in association with a top*
1553 *quark pair at the LHC with the ATLAS detector*, Physics Letters B **784** (2018) 173,
1554 ISSN: 0370-2693, URL:
1555 <https://www.sciencedirect.com/science/article/pii/S0370269318305732>.

- 1557 [7] *Evidence for the associated production of the Higgs boson and a top quark pair with the*
1558 *ATLAS detector*, tech. rep. ATLAS-CONF-2017-077, CERN, 2017,
1559 URL: <https://cds.cern.ch/record/2291405>.
- 1560 [8] A. M. Sirunyan et al., *Observation of $t\bar{t}H$ Production*,
1561 *Physical Review Letters* **120** (2018), ISSN: 1079-7114,
1562 URL: <http://dx.doi.org/10.1103/PhysRevLett.120.231801>.
- 1563 [9] B. Dumont, S. Fichet and G. von Gersdorff,
1564 *A Bayesian view of the Higgs sector with higher dimensional operators*,
1565 *Journal of High Energy Physics* **2013** (2013), ISSN: 1029-8479,
1566 URL: [http://dx.doi.org/10.1007/JHEP07\(2013\)065](http://dx.doi.org/10.1007/JHEP07(2013)065).
- 1567 [10] S. Banerjee, S. Mukhopadhyay and B. Mukhopadhyaya,
1568 *Higher dimensional operators and the LHC Higgs data: The role of modified kinematics*,
1569 *Physical Review D* **89** (2014), ISSN: 1550-2368,
1570 URL: <http://dx.doi.org/10.1103/PhysRevD.89.053010>.
- 1571 [11] M. Zinser, ‘The Large Hadron Collider’, *Search for New Heavy Charged Bosons and*
1572 *Measurement of High-Mass Drell-Yan Production in Proton—Proton Collisions*,
1573 Springer International Publishing, 2018 47, ISBN: 978-3-030-00650-1,
1574 URL: https://doi.org/10.1007/978-3-030-00650-1_4.
- 1575 [12] *Detector and Technology*, URL: <https://atlas.cern/discover/detector>.
- 1576 [13] M. Marjanovic, *ATLAS Tile calorimeter calibration and monitoring systems*, 2018.

- 1577 [14] ATLAS Collaboration, *Observation of electroweak $W^\pm Z$ boson pair production in*
 1578 *association with two jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector,*
 1579 *Phys. Lett. **B793** (2019) 469*, arXiv: [1812.09740 \[hep-ex\]](#).
- 1580 [15] ATLAS Collaboration, *Luminosity determination in pp collisions at $\sqrt{s} = 7$ TeV using*
 1581 *the ATLAS detector at the LHC*, *Eur. Phys. J. C **71** (2011) 1630*,
 1582 arXiv: [1101.2185 \[hep-ex\]](#).
- 1583 [16] ATLAS Collaboration, *Performance of the ATLAS detector using first collision data*,
 1584 *JHEP **09** (2010) 056*, arXiv: [1005.5254 \[hep-ex\]](#).
- 1585 [17] S. Agostinelli et al., *GEANT4: A Simulation toolkit*,
 1586 *Nucl. Instrum. Meth. **A506** (2003) 250*.
- 1587 [18] T. Gleisberg et al., *Event generation with SHERPA 1.1*, *JHEP **02** (2009) 007*,
 1588 arXiv: [0811.4622 \[hep-ph\]](#).
- 1589 [19] H.-L. Lai et al., *New parton distributions for collider physics*,
 1590 *Phys. Rev. D **82** (2010) 074024*, arXiv: [1007.2241 \[hep-ph\]](#).
- 1591 [20] J. Alwall et al., *The automated computation of tree-level and next-to-leading order*
 1592 *differential cross sections, and their matching to parton shower simulations*,
 1593 *JHEP **07** (2014) 079*, arXiv: [1405.0301 \[hep-ph\]](#).
- 1594 [21] R. D. Ball et al., *Parton distributions for the LHC Run II*, *JHEP **04** (2015) 040*,
 1595 arXiv: [1410.8849 \[hep-ph\]](#).
- 1596 [22] M. Bahr et al., *Herwig++ Physics and Manual*, *Eur. Phys. J. C **58** (2008) 639*,
 1597 arXiv: [0803.0883 \[hep-ph\]](#).

- 1598 [23] R. D. Ball et al., *Parton distributions with LHC data*, Nucl. Phys. B **867** (2013) 244,
 1599 arXiv: [1207.1303 \[hep-ph\]](https://arxiv.org/abs/1207.1303).
- 1600 [24] S. Frixione, G. Ridolfi and P. Nason,
 1601 *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*,
 1602 JHEP **09** (2007) 126, arXiv: [0707.3088 \[hep-ph\]](https://arxiv.org/abs/0707.3088).
- 1603 [25] E. Re, *Single-top Wt-channel production matched with parton showers using the*
 1604 *POWHEG method*, Eur. Phys. J. C **71** (2011) 1547, arXiv: [1009.2450 \[hep-ph\]](https://arxiv.org/abs/1009.2450).
- 1605 [26] ATLAS Collaboration, *Electron efficiency measurements with the ATLAS detector using*
 1606 *the 2015 LHC proton–proton collision data*, ATLAS-CONF-2016-024, 2016,
 1607 URL: <https://cds.cern.ch/record/2157687>.
- 1608 [27] ATLAS Collaboration, *Measurement of the muon reconstruction performance of the*
 1609 *ATLAS detector using 2011 and 2012 LHC proton–proton collision data*,
 1610 Eur. Phys. J. C **74** (2014) 3130, arXiv: [1407.3935 \[hep-ex\]](https://arxiv.org/abs/1407.3935).
- 1611 [28] R. Narayan et al., *Measurement of the total and differential cross sections of a*
 1612 *top-quark-antiquark pair in association with a W boson in proton-proton collisions at a*
 1613 *centre-of-mass energy of 13 TeV with ATLAS detector at the Large Hadron Collider*,
 1614 tech. rep. ATL-COM-PHYS-2020-217, CERN, 2020,
 1615 URL: <https://cds.cern.ch/record/2712986>.
- 1616 [29] ATLAS Collaboration, *Jet Calibration and Systematic Uncertainties for Jets*
 1617 *Reconstructed in the ATLAS Detector at $\sqrt{s} = 13 \text{ TeV}$* , ATL-PHYS-PUB-2015-015,
 1618 2015, URL: <https://cds.cern.ch/record/2037613>.

- 1619 [30] ATLAS Collaboration,
1620 *Selection of jets produced in 13 TeV proton–proton collisions with the ATLAS detector,*
1621 ATLAS-CONF-2015-029, 2015, URL: <https://cds.cern.ch/record/2037702>.
- 1622 [31] ATLAS Collaboration, *Performance of pile-up mitigation techniques for jets in pp*
1623 *collisions at $\sqrt{s} = 8$ TeV using the ATLAS detector*, *Eur. Phys. J. C* **76** (2016) 581,
1624 arXiv: [1510.03823 \[hep-ex\]](https://arxiv.org/abs/1510.03823).
- 1625 [32] A. Collaboration, *Performance of b -jet identification in the ATLAS experiment*,
1626 *Journal of Instrumentation* **11** (2016) P04008,
1627 URL: <http://stacks.iop.org/1748-0221/11/i=04/a=P04008>.
- 1628 [33] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction*
1629 *with the ATLAS detector using proton–proton collisions at $\sqrt{s} = 13$ TeV*,
1630 *The European Physical Journal C* **78** (2018), ISSN: 1434-6052,
1631 URL: <http://dx.doi.org/10.1140/epjc/s10052-018-6288-9>.
- 1632 [34] 2021, URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/BTagCalibrationRecommendationsRelease21#Tools_for_Flavor_Tagging_Calibra.
- 1635 [35] ATLAS Collaboration, *Measurement of the fiducial and differential cross-section of a*
1636 *top quark pair in association with a Z boson at 13 TeV with the ATLAS detector*, (2019),
1637 URL: <https://cds.cern.ch/record/2672207>.
- 1638 [36] T. Chen and C. Guestrin, ‘XGBoost: A Scalable Tree Boosting System’,
1639 *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge*

- 1640 *Discovery and Data Mining*, KDD '16, ACM, 2016 785, ISBN: 978-1-4503-4232-2,
- 1641 URL: <http://doi.acm.org/10.1145/2939672.2939785>.
- 1642 [37] R. Frederix et al., *The automation of next-to-leading order electroweak calculations*,
- 1643 *Journal of High Energy Physics* **2018** (2018), ISSN: 1029-8479,
- 1644 URL: [http://dx.doi.org/10.1007/JHEP07\(2018\)185](http://dx.doi.org/10.1007/JHEP07(2018)185).
- 1645 [38] *Luminosity determination in pp collisions at $\sqrt{s} = 13$ TeV using the ATLAS detector at*
- 1646 *the LHC*, (2019).
- 1647 [39] ATLAS Collaboration,
- 1648 *The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS*,
- 1649 *JINST* **13** (2018) P07017.
- 1650 [40] ATLAS Collaboration, *Jet energy scale measurements and their systematic uncertainties*
- 1651 *in proton–proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*,
- 1652 *Phys. Rev. D* **96** (2017) 072002, arXiv: [1703.09665 \[hep-ex\]](https://arxiv.org/abs/1703.09665).
- 1653 [41] ATLAS Collaboration, *Observation of the associated production of a top quark and a Z*
- 1654 *boson in pp collisions at $\sqrt{s}= 13$ TeV with the ATLAS detector*,
- 1655 *Journal of High Energy Physics* **2020** (2020), ISSN: 1029-8479,
- 1656 URL: [http://dx.doi.org/10.1007/JHEP07\(2020\)124](http://dx.doi.org/10.1007/JHEP07(2020)124).
- 1657 [42] ATLAS Collaboration, *Comparison of Monte Carlo generator predictions from Powheg*
- 1658 *and Sherpa to ATLAS measurements of top pair production at 7 TeV*,
- 1659 ATL-PHYS-PUB-2015-011, 2015, URL: <https://cds.cern.ch/record/2020602>.

- 1660 [43] ATLAS Collaboration, *ATLAS tunes of PYTHIA 6 and Pythia 8 for MC11*,
1661 ATL-PHYS-PUB-2011-009, 2011, URL: <https://cds.cern.ch/record/1363300>.
- 1662 [44] ATLAS Collaboration, *Performance of missing transverse momentum reconstruction*
1663 *with the ATLAS detector in the first proton–proton collisions at $\sqrt{s} = 13 \text{ TeV}$,*
1664 ATL-PHYS-PUB-2015-027, 2015, URL: <https://cds.cern.ch/record/2037904>.
- 1665 [45] G. Aad, T. Abajyan, B. Abbott and etal, *Jet energy resolution in proton-proton collisions*
1666 *at $\sqrt{s} = 7 \text{ TeV}$ recorded in 2010 with the ATLAS detector,*
1667 The European Physical Journal C **73** (2013) 2306, ISSN: 1434-6052,
1668 URL: <https://doi.org/10.1140/epjc/s10052-013-2306-0>.

List of contributions

1669

1670

¹⁶⁷¹ **Part VII**

¹⁶⁷² **Appendices**

¹⁶⁷³ **A Non-prompt lepton MVA**

1674 A lepton MVA has been developed to better reject non-prompt leptons than standard cut
1675 based selections based upon impact parameter, isolation and PID. The name of this MVA is
1676 **PromptLeptonVeto**. The full set of studies and detailed explanation can be found in [28].

1677 The decays of W and Z bosons are commonly selected by the identification of one or two
1678 electrons or muons. The negligible lifetimes of these bosons mean that the leptons produced in the
1679 decay originate from the interaction vertex and are thus labelled “prompt”. Analyses using these
1680 light leptons impose strict reconstruction quality, isolation and impact parameter requirements
1681 to remove “fake” leptons. A significant source of the fake light leptons are non-prompt leptons
1682 produced in decays of hadrons that contain bottom (b) or charm (c) quarks. Such hadrons
1683 typically have microscopically significant lifetimes that can be detected experimentally.

1684 These non-prompt leptons can also pass the tight selection criteria. In analyses that
1685 involve top (t) quarks, which decay almost exclusively into a W boson and a b quark, non-
1686 prompt leptons from the semileptonic decay of bottom and charm hadrons can be a significant
1687 source of background events. This is particularly the case in the selection of same-sign dilepton
1688 and multilepton final states.

1689 The main idea is to identify non-prompt light leptons using lifetime information associated
1690 with a track jet that matches the selected light lepton. This lifetime information is computed
1691 using tracks contained within the jet. Typically, lepton lifetime is determined using the impact
1692 parameter of the track reconstructed by the inner tracking detector which is matched to the recon-
1693 structed lepton. Using additional reconstructed charged particle tracks increases the precision of

1694 identifying the displaced decay vertex of bottom or charm hadrons that produced a non-prompt
 1695 light lepton. The MVA also includes information related to the isolation of the lepton to reject
 1696 non-prompt leptons.

1697 **PromptLeptonVeto** is a gradient boosted BDT. The training of the BDT is performed on
 1698 leptons selected from the PowHEG+PYTHIA6 non-allhad $t\bar{t}$ MC sample. Eight variables are used
 1699 to train the BDT in order to discriminate between prompt and non-prompt leptons. The track
 1700 jets that are matched to the non-prompt leptons correspond to jets initiated by b or c quarks, and
 1701 may contain a displaced vertex. Consequently, three of the selected variables are used to identify
 1702 b-tag jets by standard ATLAS flavour tagging algorithms. Two variables use the relationship
 1703 between the track jet and lepton: the ratio of the lepton p_T with respect to the track jet p_T and
 1704 ΔR between the lepton and the track jet axis. Finally three additional variables test whether the
 1705 reconstructed lepton is isolated: the number of tracks collected by the track jet and the lepton
 1706 track and calorimeter isolation variables. Table 51 describes the variables used to train the BDT
 1707 algorithm. The choice of input variables has been extensively discussed with Egamma, Muon,
 1708 Tracking, and Flavour Tagging CP groups.

Variable	Description
N_{track} in track jet	Number of tracks collected by the track jet
$\text{IP2 log}(P_b/P_{\text{light}})$	Log-likelihood ratio between the b and light jet hypotheses with the IP2D algorithm
$\text{IP3 log}(P_b/P_{\text{light}})$	Log-likelihood ratio between the b and light jet hypotheses with the IP3D algorithm
N_{TrkAtVtx} SV + JF	Number of tracks used in the secondary vertex found by the SV1 algorithm in addition to the number of tracks from secondary vertices found by the JetFitter algorithm with at least two tracks
$p_T^{\text{lepton}}/p_T^{\text{track jet}}$	The ratio of the lepton p_T and the track jet p_T
$\Delta R(\text{lepton}, \text{track jet})$	ΔR between the lepton and the track jet axis
$p_T^{\text{VarCone30}}/p_T$	Lepton track isolation, with track collecting radius of $\Delta R < 0.3$
$E_T^{\text{TopoCone30}}/p_T$	Lepton calorimeter isolation, with topological cluster collecting radius of $\Delta R < 0.3$

Table 51: A table of the variables used in the training of **PromptLeptonVeto**.

1709 The output distribution of the BDT is shown in Figure A.

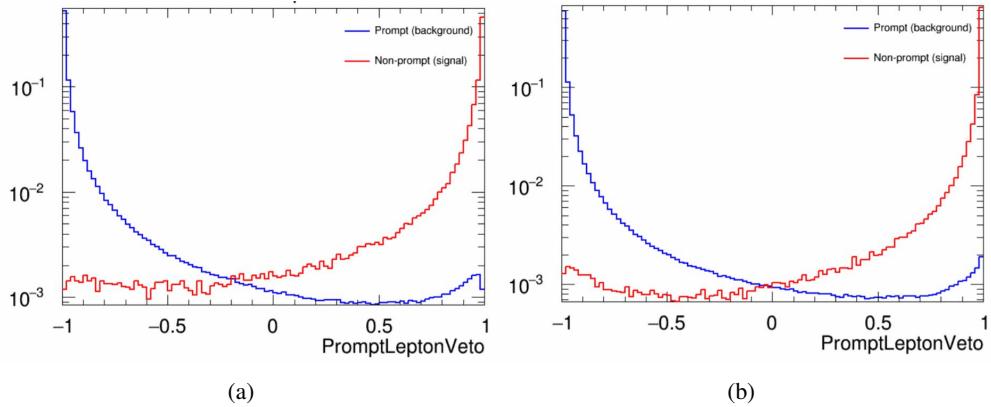


Figure A.1: Distribution of the PLV BDT discriminant for (a) electrons and (b) muons.

1710 The ROC curve for the BDT response, compared to the standard `FixedCutTight` WP, is
1711 shown in figure A, which shows a clear improvement when using this alternative training.

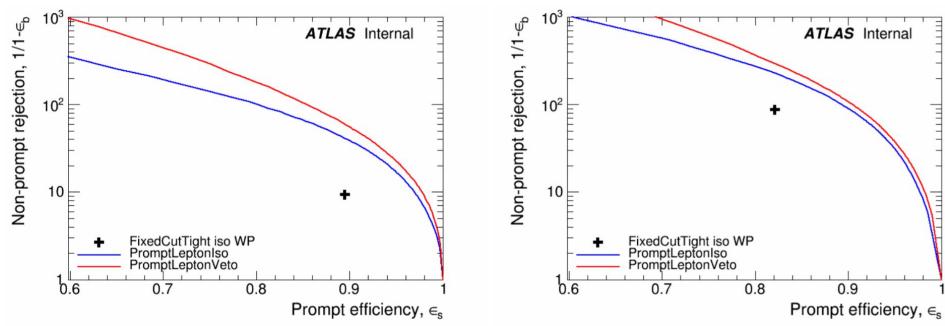


Figure A.2: ROC curves for the PLV as well as the performance of the standard FixedCutTight WP for (left) electrons and (right) muons. Plot also includes curves for PROMPTLEPTONIso, which is not considered in this analysis.

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

1715 The efficiency of the tight PromptLeptonVeto working point is measured using the tag
1716 and probe method with $Z \rightarrow \ell^+ \ell^-$ events. Such calibration are performed by analysers from
1717 this analysis in communication with the Egamma and Muon combined performance groups. The
1718 scale factor are approximately 0.92 for $10 < p_T < 15$ GeV, and averaging at 0.98 to 0.99 for
1719 higher p_T leptons. An extra systematic is applied to muons within $\Delta R < 0.6$ of a calorimeter
1720 jet, since there is a strong dependence on the scale factor due to the presence of these jets. For
1721 electrons, the dominant systematics is coming from pile-up dependence. Overall the systematics
1722 are a maximum of 3% at low p_T and decreasing at a function of p_T .

¹⁷²³ **B Supplementary WZ + Heavy Flavor Studies**

¹⁷²⁴ **B.1 Non-prompt CR Modelling**

1725 In order to further validate the modeling in each of the non-prompt CRs, additional
 1726 kinematic plots are made in the Z+jets CR and $t\bar{t}$ CR in each of the continuous b-tag regions,
 1727 after the correction factors detailed in Section 8.3 have been applied.

1728 In the case of the Z+jets CR, the p_T spectrum of the lepton originating from the W
 1729 candidate is shown, as this is the distribution used to extract the scale factor applied to Z+jets.

1730 These plots are shown in Figures B.1 and B.2.

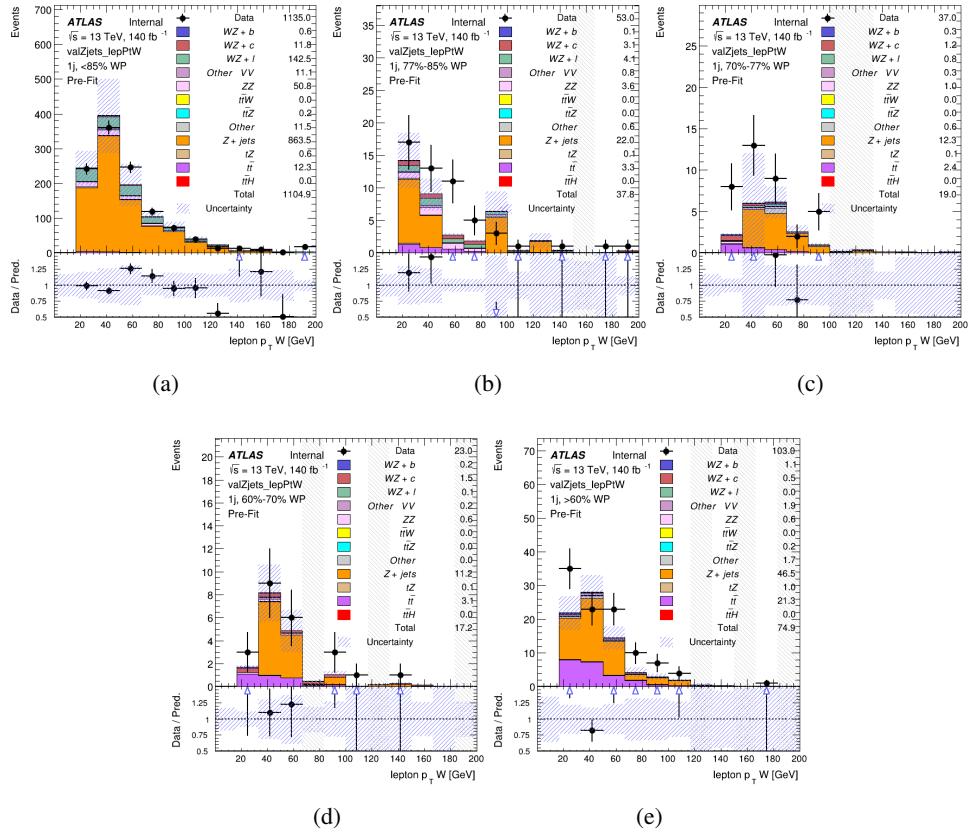


Figure B.1: Comparisons between the data and MC distributions of the p_T of the lepton originating from the W-candidate in the Z+jets CR for each of the 1-jet b-tag working point regions

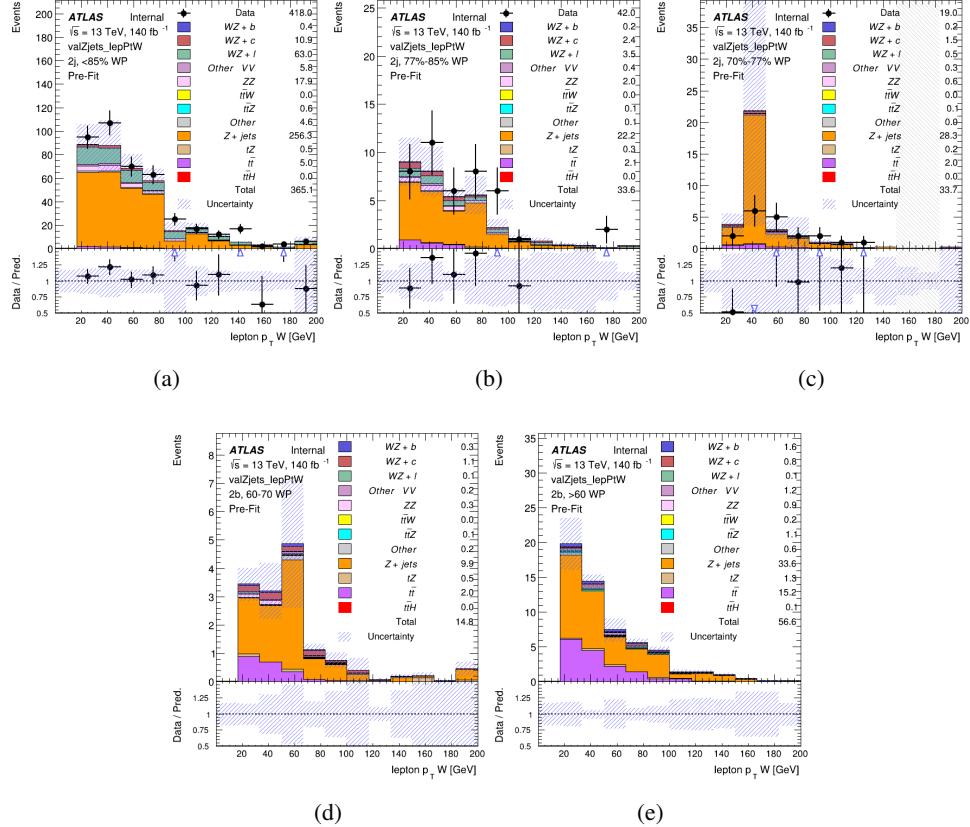


Figure B.2: Comparisons between the data and MC distributions of the p_T of the lepton originating from the W-candidate in the Z+jets CR for each of the 2-jet b-tag working point regions

1731 The same is shown for the $t\bar{t}$ CR, but the p_T of the OS lepton is used instead as a
 1732 representation of the modeling, as the lepton from the W is not well defined for $t\bar{t}$ events. These
 1733 plots are shown in Figures B.3 and B.4.

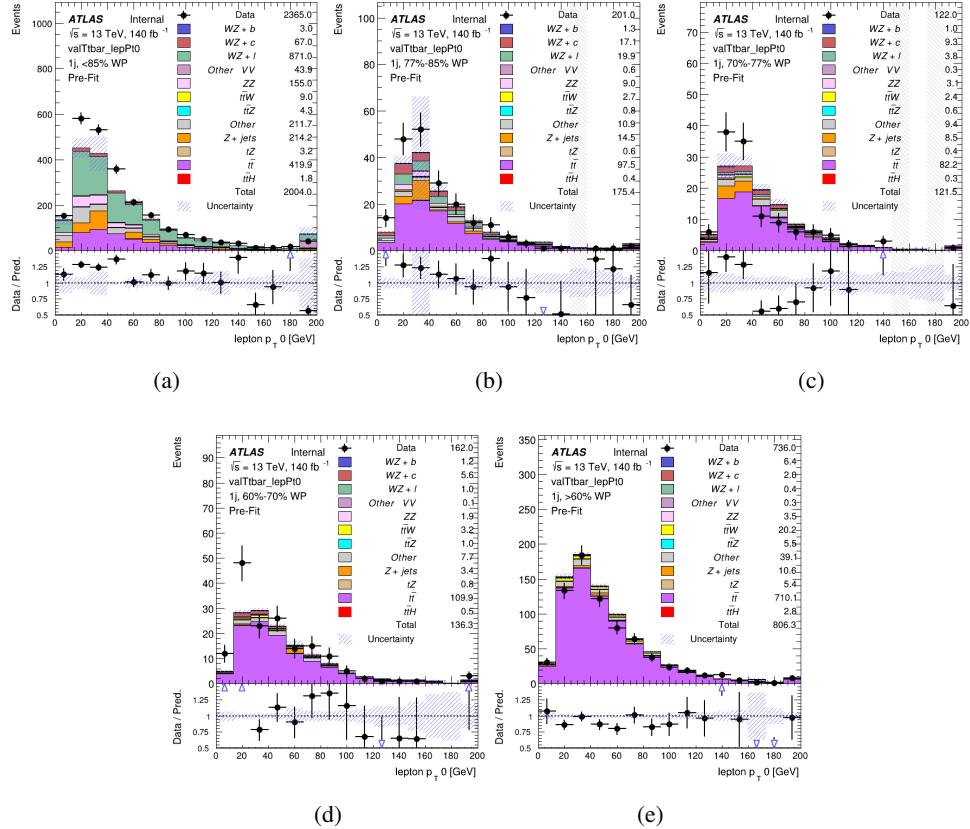


Figure B.3: Comparisons between the data and MC distributions of the p_T of the OS lepton in the $t\bar{t}$ CR for each of the 1-jet b-tag working point regions

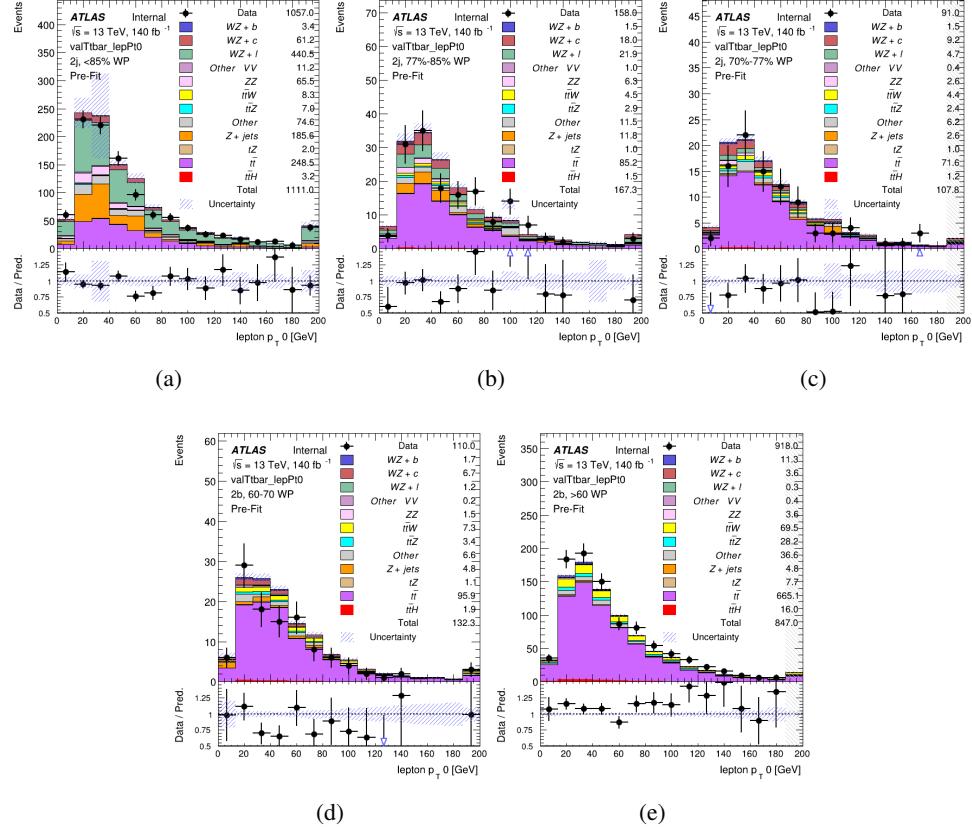


Figure B.4: Comparisons between the data and MC distributions of the p_T of the OS lepton in the $t\bar{t}$ CR for each of the 2-jet b-tag working point regions

¹⁷³⁴ **B.2 tZ Interference Studies**

1735 Because it includes an on-shell Z boson as well as a b-jet and W from the top decay, tZ
1736 production represents an identical final state to WZ + b-jet. This implies the possibility of matrix
1737 level interference between these two processes not accounted for in the Monte Carlo simulations,
1738 which consider the two processes independently. Truth level studies are performed in order to
1739 estimate the impact of these interference effects.

1740 In order to estimate the matrix level interference effects between tZ and WZ + b-jet, two
1741 different sets of simulations are produced using MadGraph 5 [37] - one which simulates these
1742 two processes independently, and another where they are produced simultaneously, such that
1743 interference effects are present. These two sets of samples are then compared, and the difference
1744 between them can be taken to represent any interference effects.

1745 MadGraph simulations of 10,000 tZ and 10,000 WZ + b-jet events are produced, along
1746 with 20,000 events where both are present, in the fiducial region where three leptons and at least
1747 one jet are produced.

1748 A selection mimicking the preselection used in the main analysis is applied to the samples:
1749 The SS leptons are required to have $p_T > 20$ GeV, and > 10 GeV is required for the OS lepton.
1750 The associated b-jet is required to have $p_T > 25$ GeV, and all physics objects are required to fall
1751 in a range of $|\eta| < 2.5$.

1752 The kinematics of these samples after the selection has been applied are shown below:

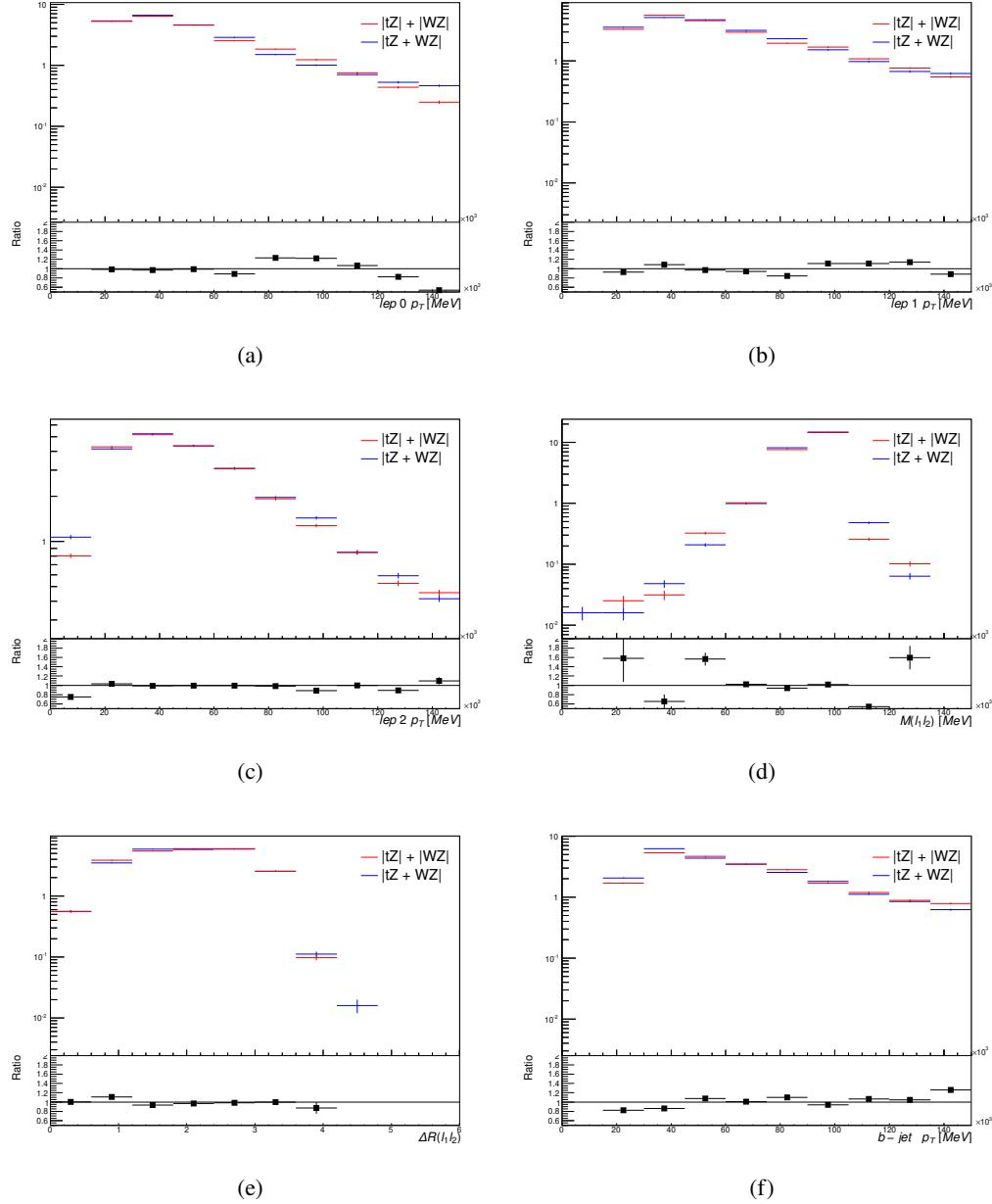


Figure B.5: Comparisons between (a) the p_T of the lepton from the W, (b) and (c) show the p_T of the lepton from the Z, (d) the invariant mass of the Z-candidate, (e) ΔR of the leptons from the Z, and (f) the p_T of the b-jet, for WZ and tZ events generated with interference effects (blue) and without interference effects (red).

1753 The overall cross-section of the two methods agree within error, and no significant differ-
1754 ences in the kinematic distributions are seen. It is therefore concluded that interference effects
1755 do not significantly impact the results.

1756 **B.3 Alternate tZ Inclusive Fit**

1757 **B.3.1 tZ Inclusive Fit**

1758 While tZ is often considered as a distinct process from WZ + b, this could also be considered part
 1759 of the signal. Alternate studies are performed where, using the same framework as the nominal
 1760 analysis, a measurement of WZ + b is performed that includes tZ as part of WZ+b.

1761 Because of this change, the tZ CR is no longer necessary, and only the five pseudo-
 1762 continuous b-tag regions are used in the fit. Further, systematics related to the tZ cross-section
 1763 are removed from the fit, as they are now encompassed by the normalization measurement of
 1764 WZ + b. All other systematic uncertainties are carried over from the nominal analysis.

1765 An expected WZ + b cross-section of $4.1^{+0.78}_{-0.74}(\text{stat})^{+0.53}_{-0.52}(\text{sys}) \text{ fb}$ is extracted from the fit,
 1766 with an expected significance of 4.0σ .

1767 The impact of the predominate systematics are summarized in Table 52.

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

Table 52: Summary of the most significant sources of systematic uncertainty on the measurement of WZ + b with exactly one associated jet.

1768 **B.3.2 Floating tZ**

1769 In order to quantify the impact of the tZ uncertainty on the fit, an alternative fit strategy is
1770 used where the tZ normalization is allowed to float. This normalization factor replaces the
1771 cross-section uncertainty on tZ, and all other parameters of the fit remain the same.

1772 An uncertainty of 17% on the normalization of tZ is extracted from the fit, compared to a
1773 theory uncertainty of 15% applied to the tZ cross-section. The measured uncertainties on WZ
1774 remain the same.

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

1775 The following section provides details of the various MVAs as well as a few studies performed
1777 in support of this analysis, exploring alternate decisions and strategies.

1778 **C.1 Higgs Reconstruction Model Details**

1779 **C.1.1 b-jet Identification Features - 2lSS**

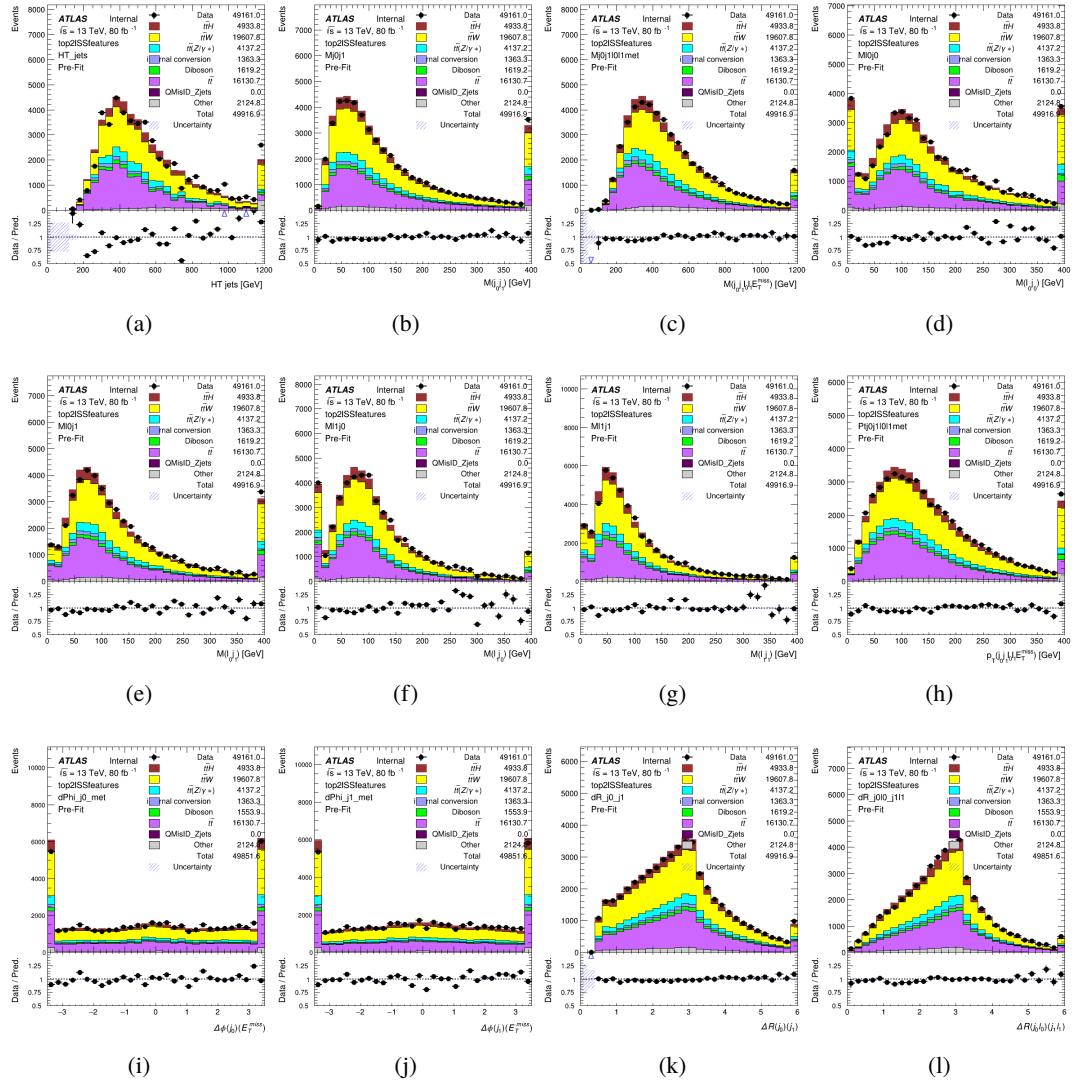


Figure C.1: Input features for top21SS

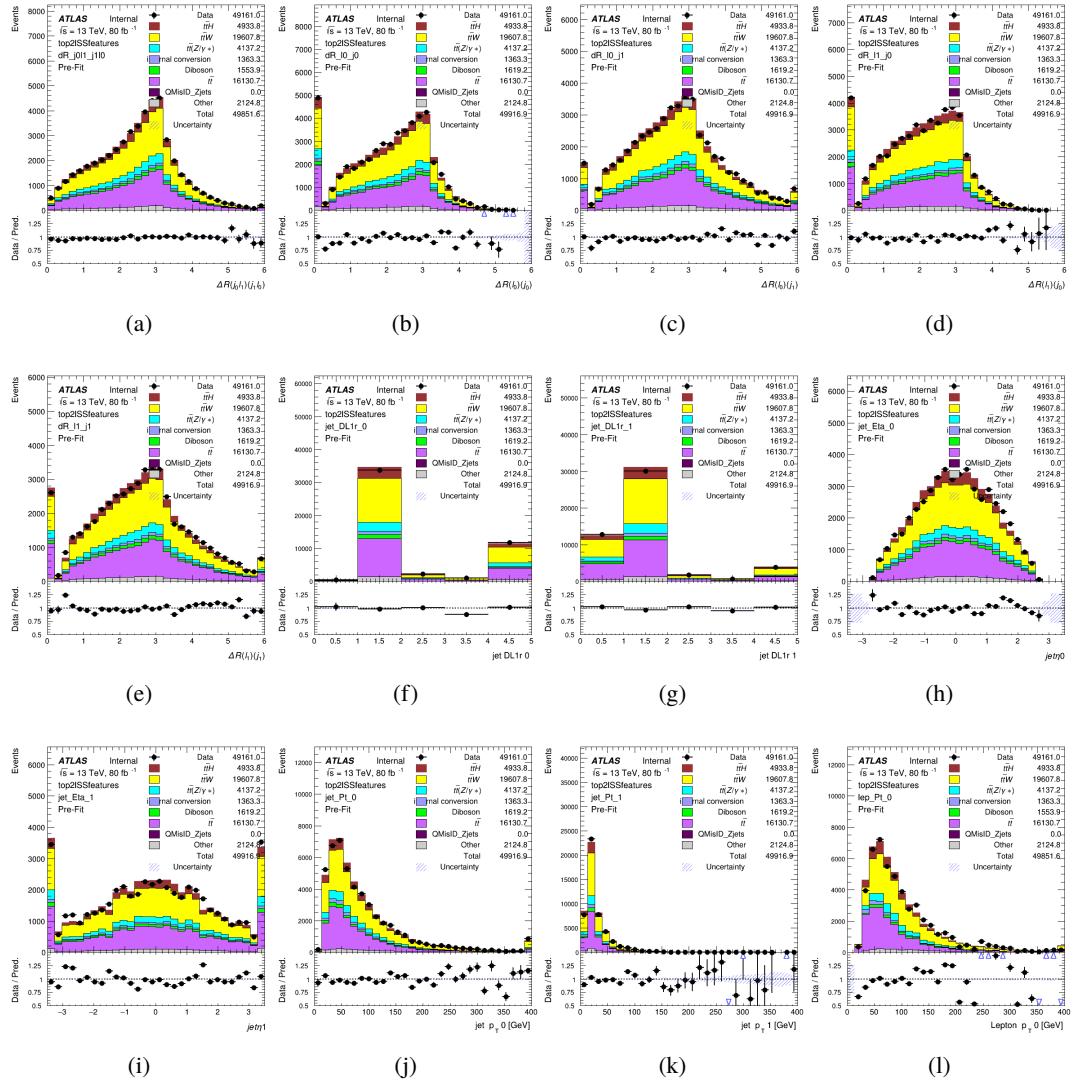


Figure C.2: Input features for top2lSS

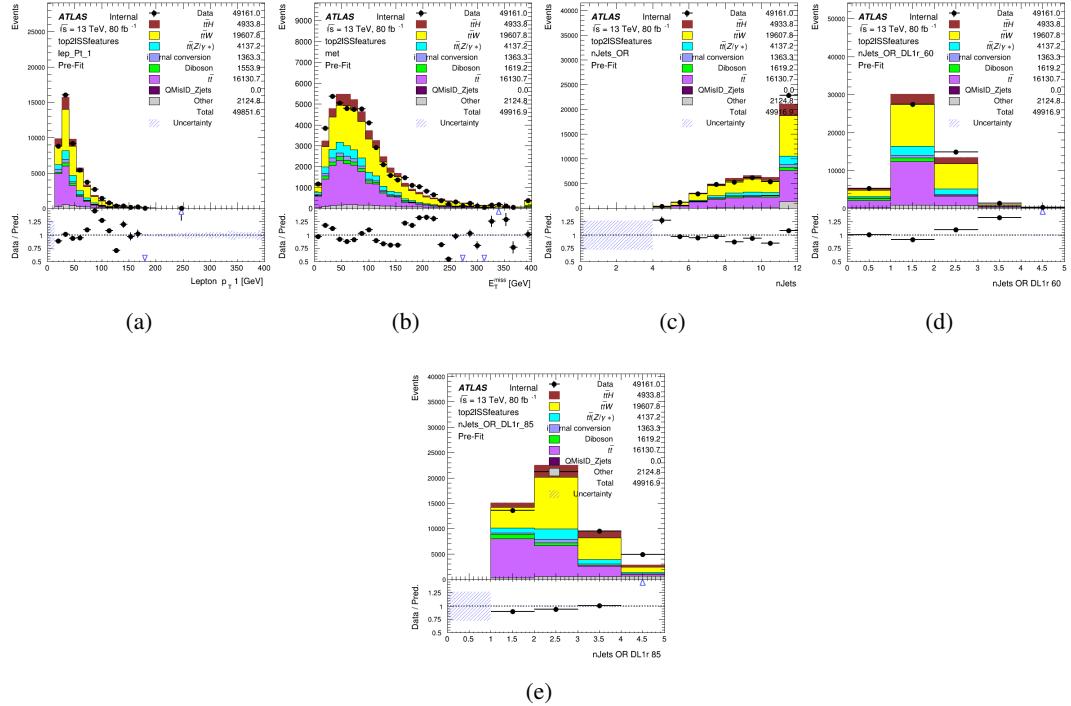


Figure C.3: Input features for top2ISS

1780 **C.1.2 b-jet Identification Features - 3l**

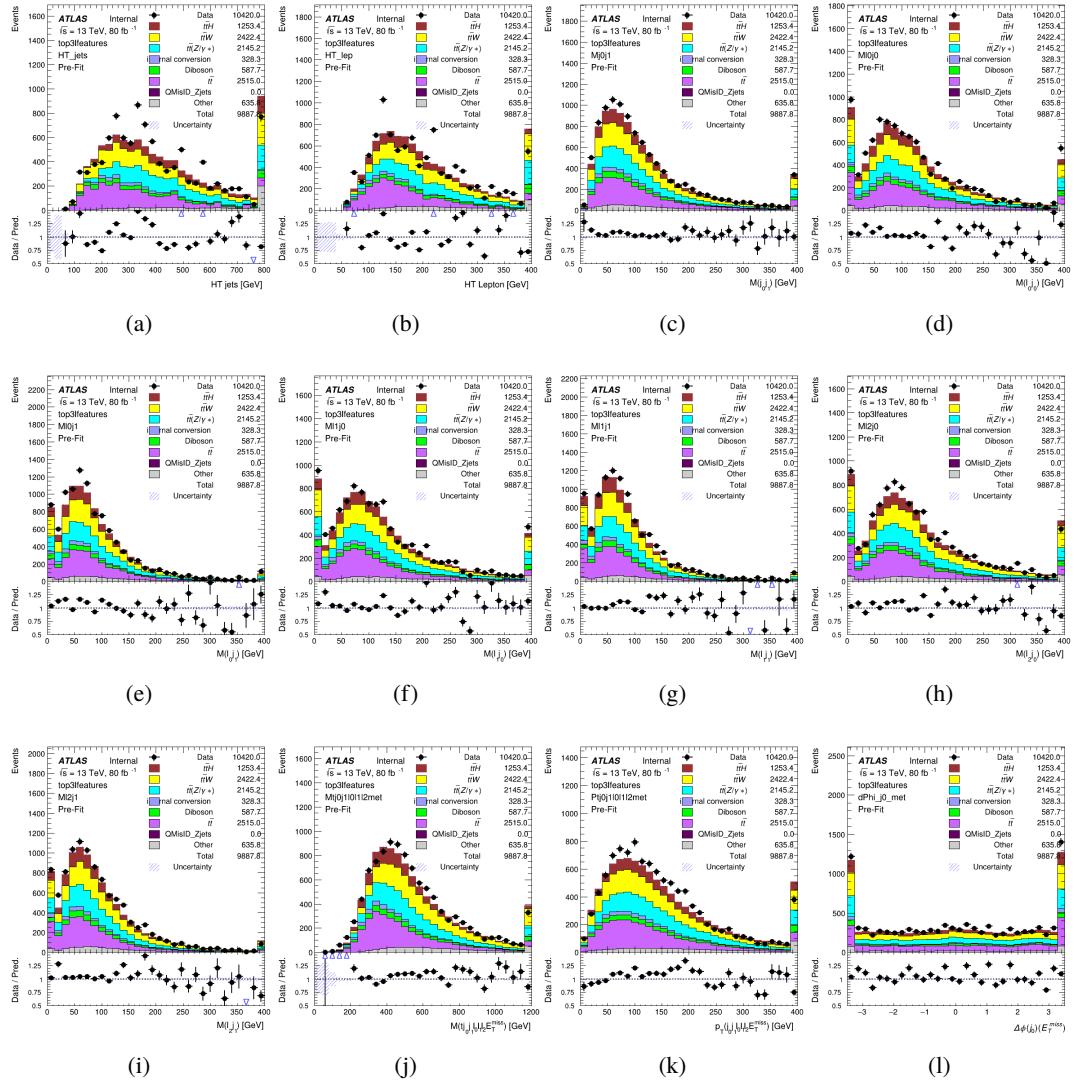


Figure C.4: Input features for top3l

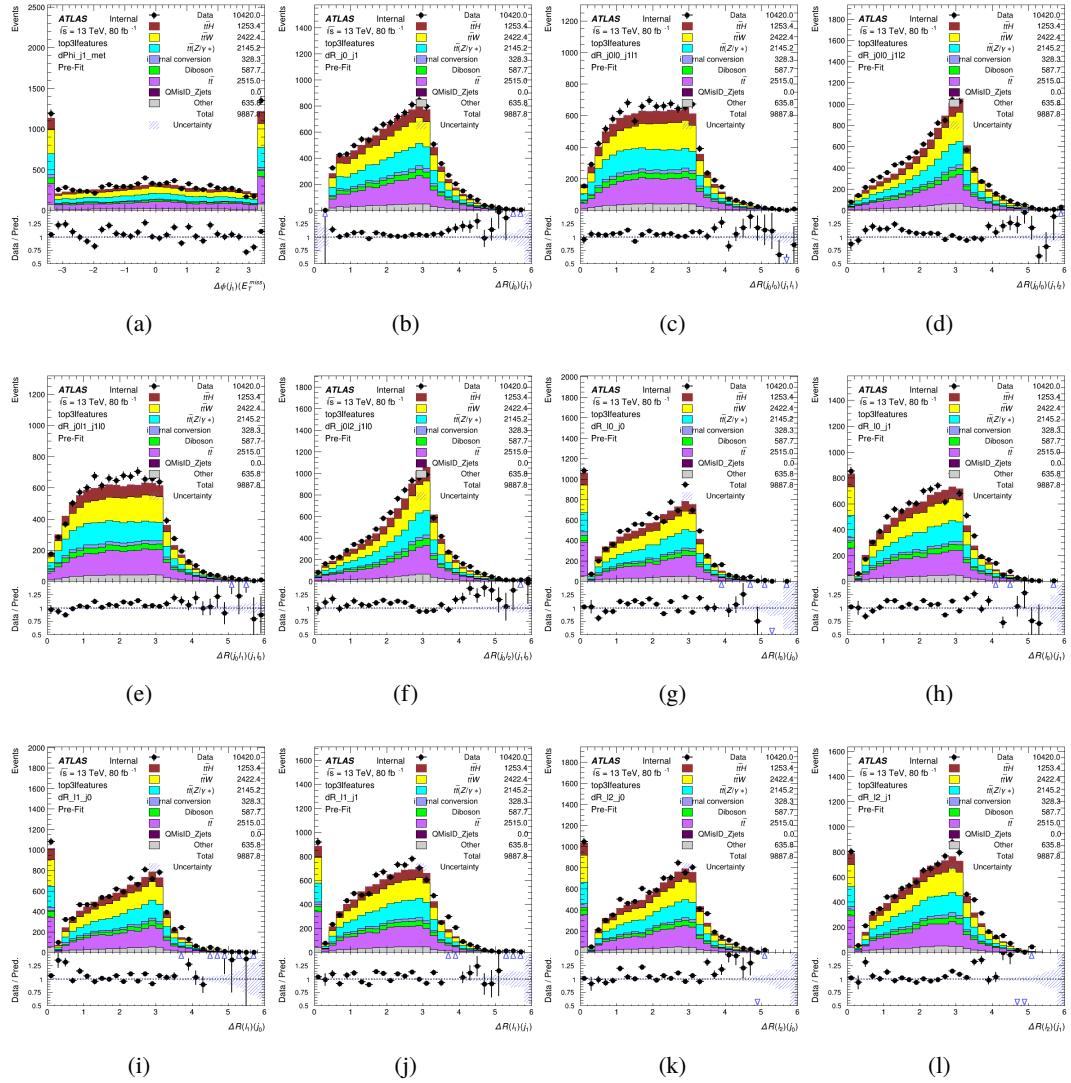


Figure C.5: Input features for top31

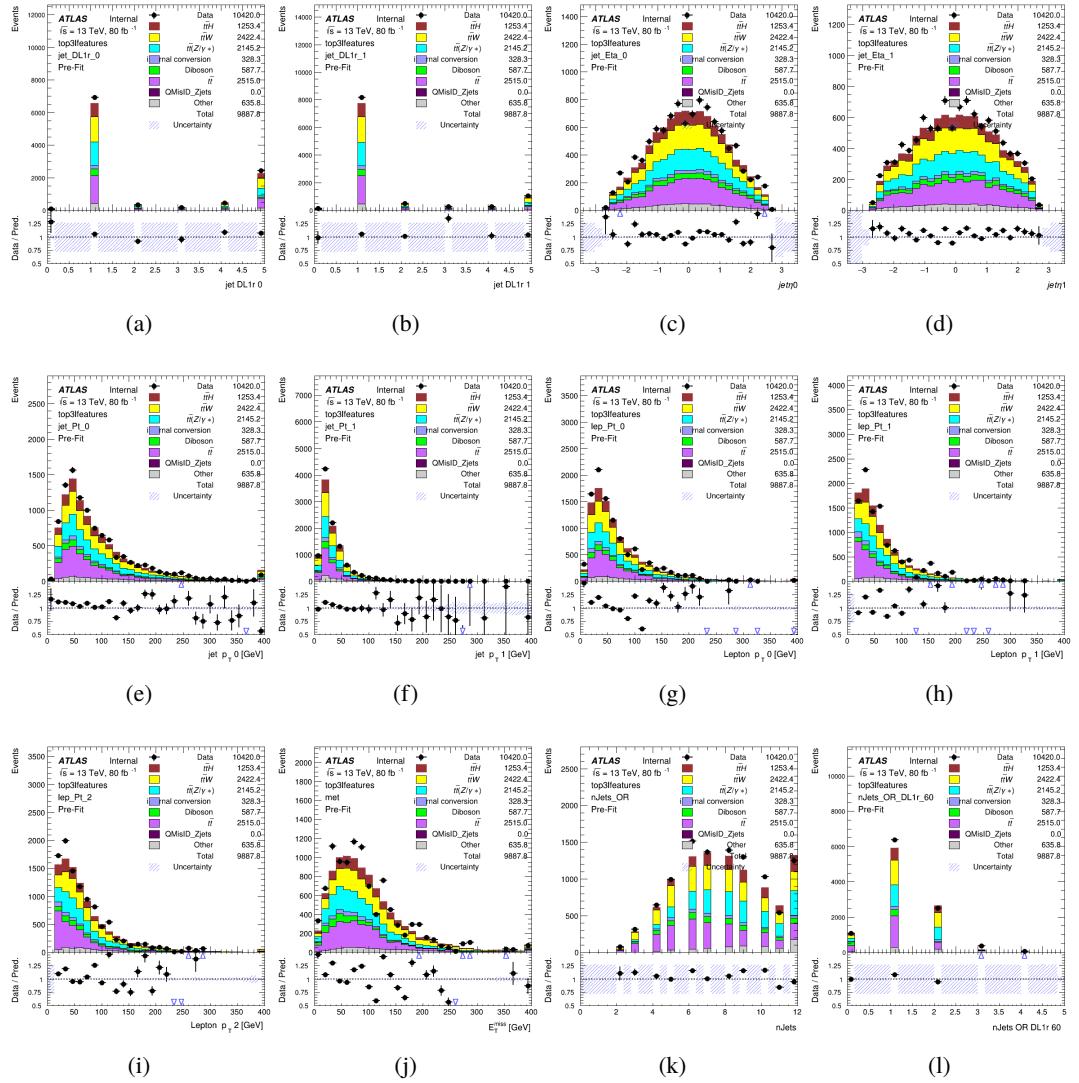
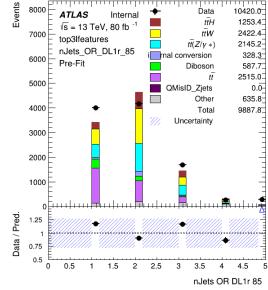


Figure C.6: Input features for top3l



(a)

Figure C.7: Input features for top3l

¹⁷⁸¹ **C.1.3 Higgs Reconstruction Features - 2lSS**

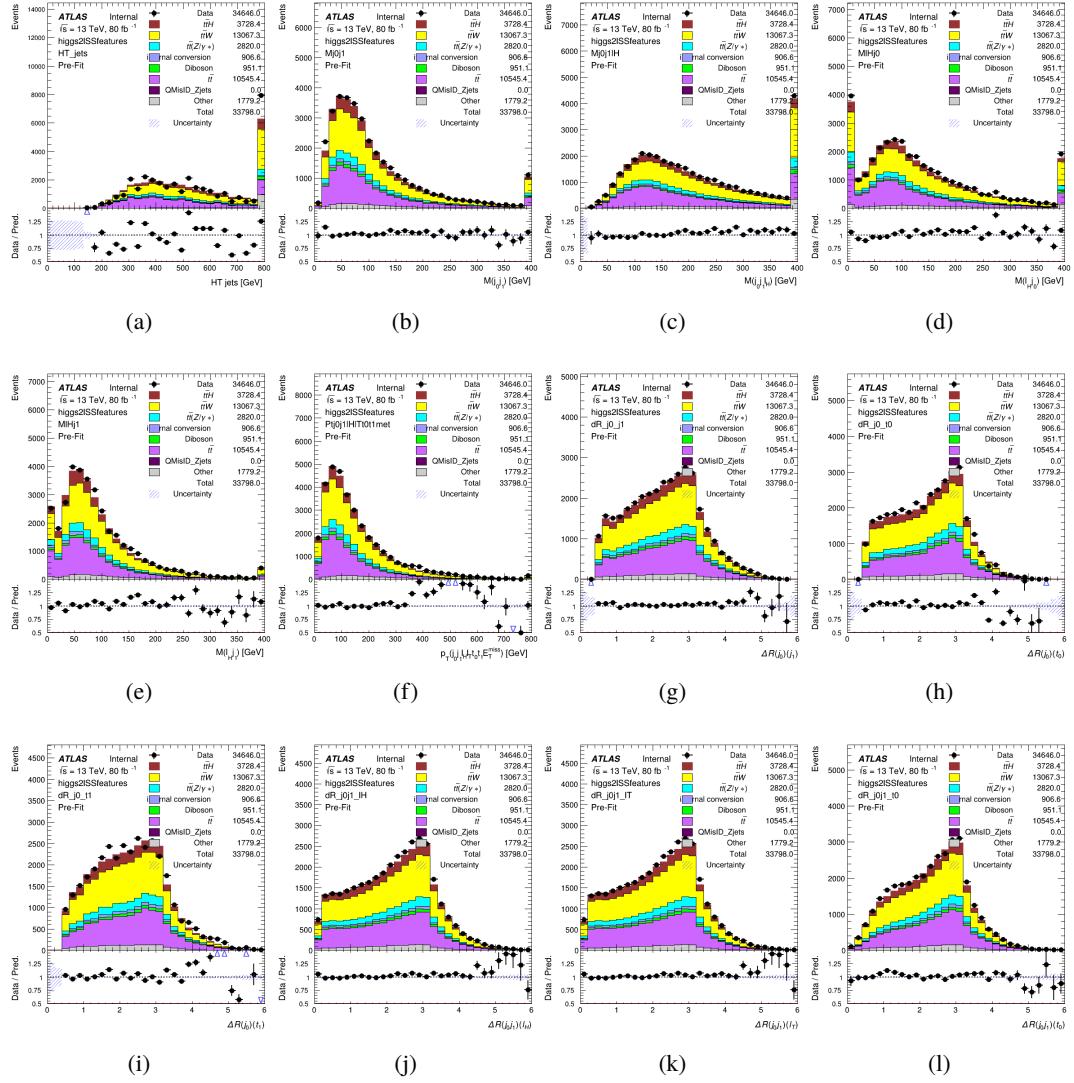


Figure C.8: Input features for higgs2ISST

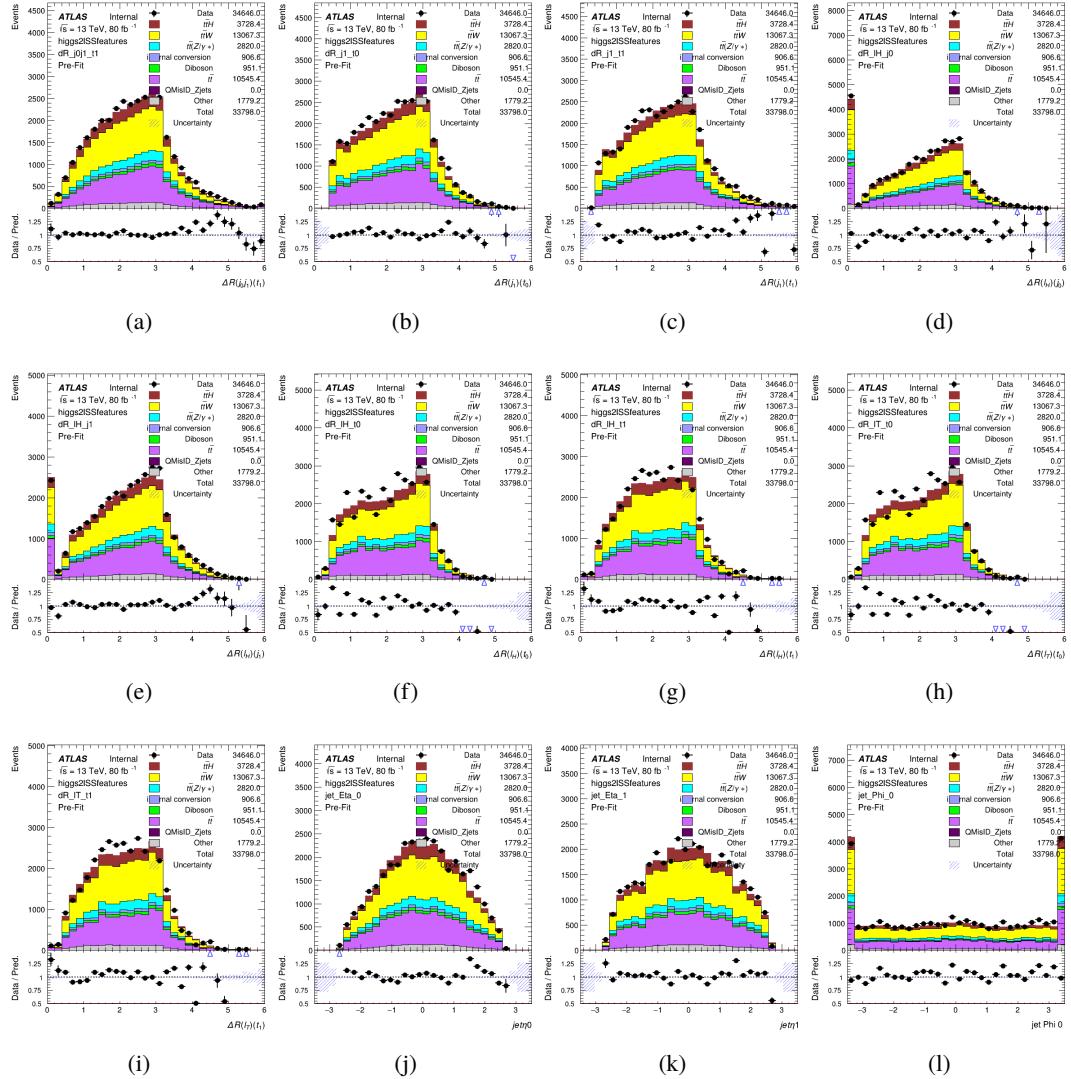


Figure C.9: Input features for higgs2ISSTest

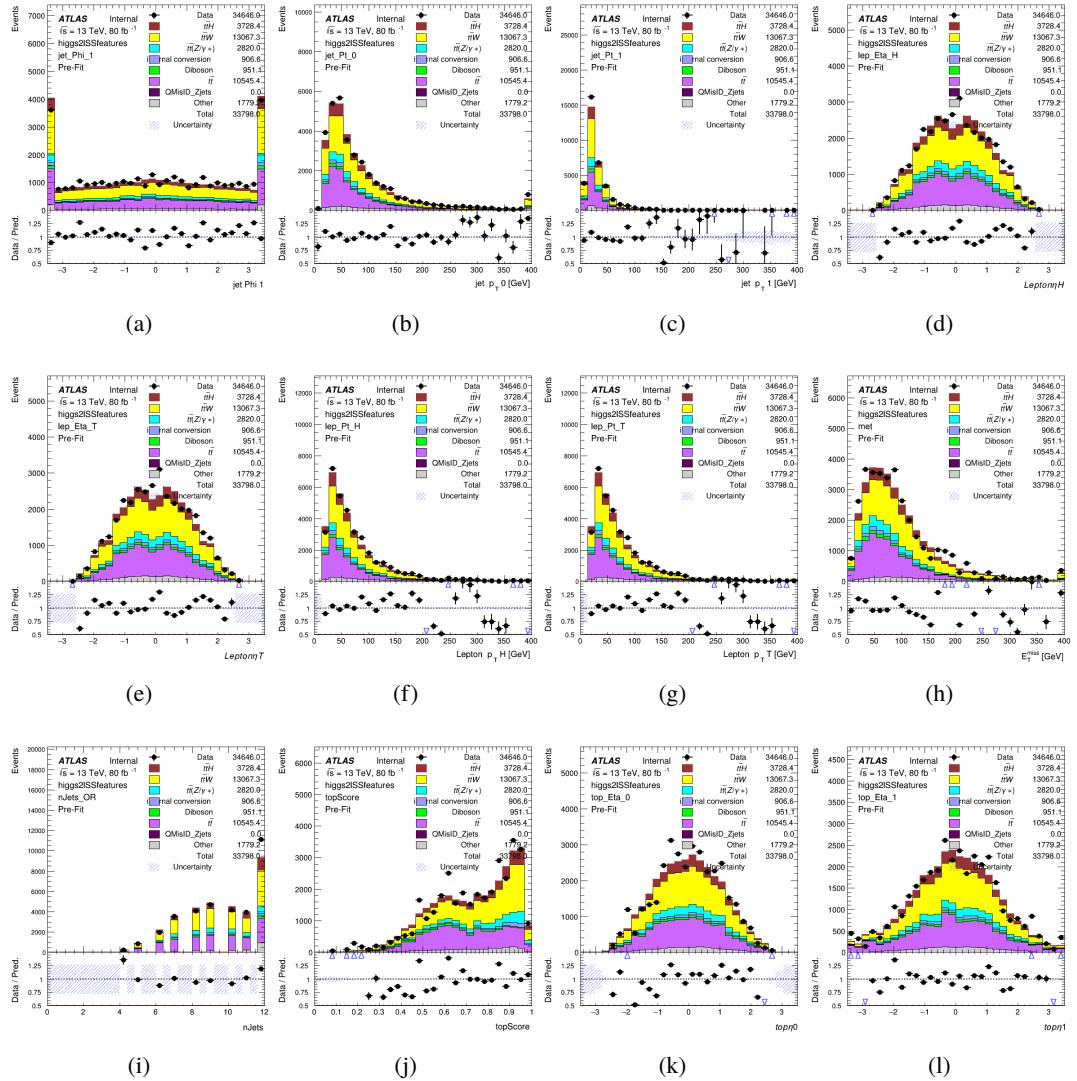


Figure C.10: Input features for higgs2lSS

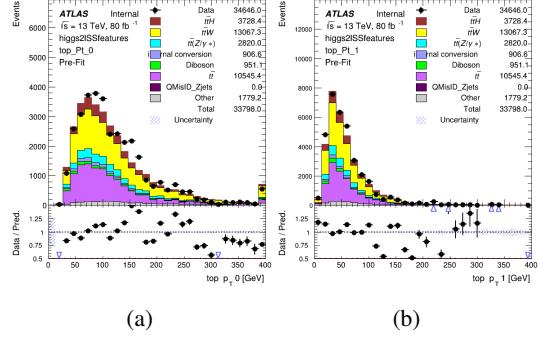


Figure C.11: Input features for higgs2lSS

₁₇₈₂ **C.1.4 Higgs Reconstruction Features - 3lS**

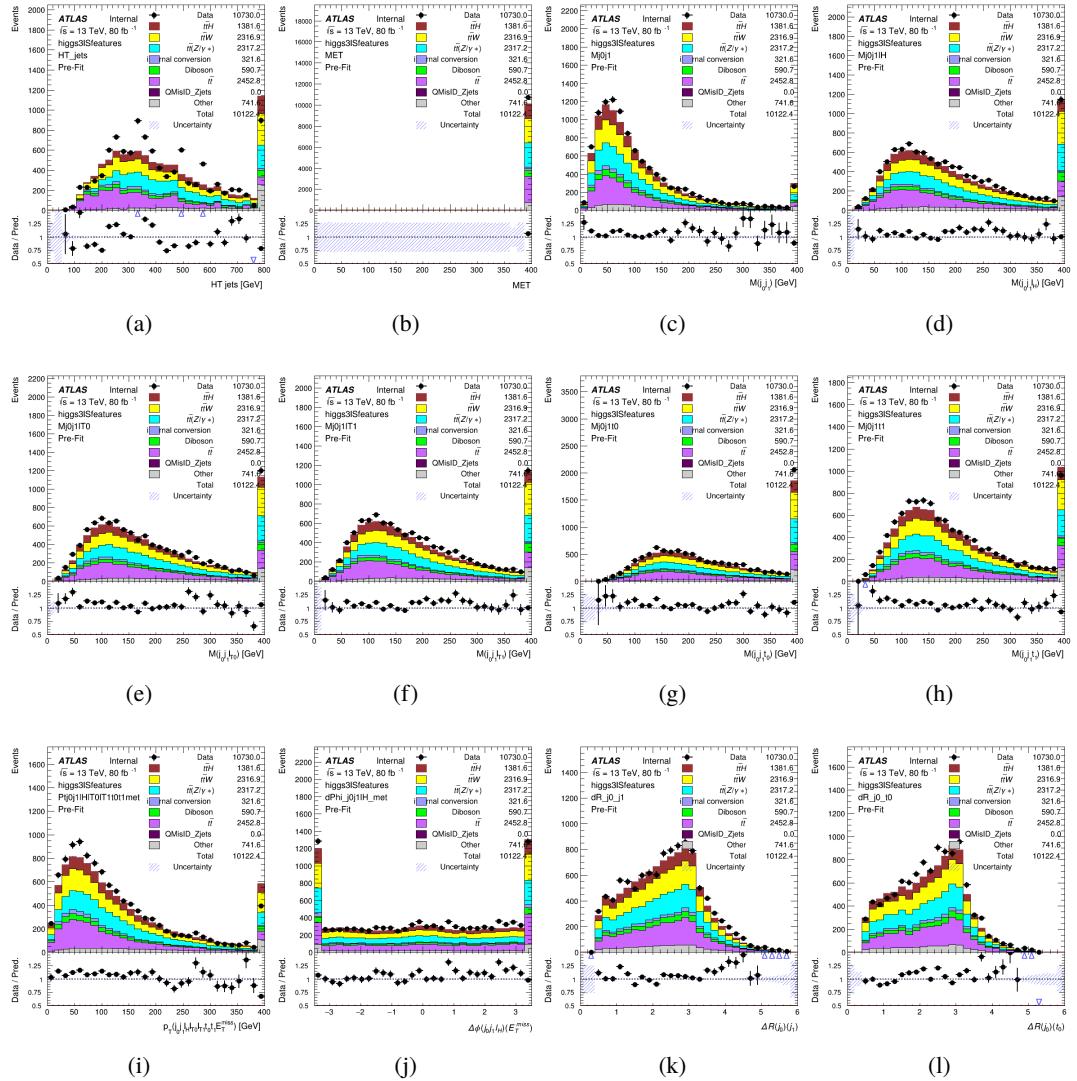


Figure C.12: Input features for higgs3IS

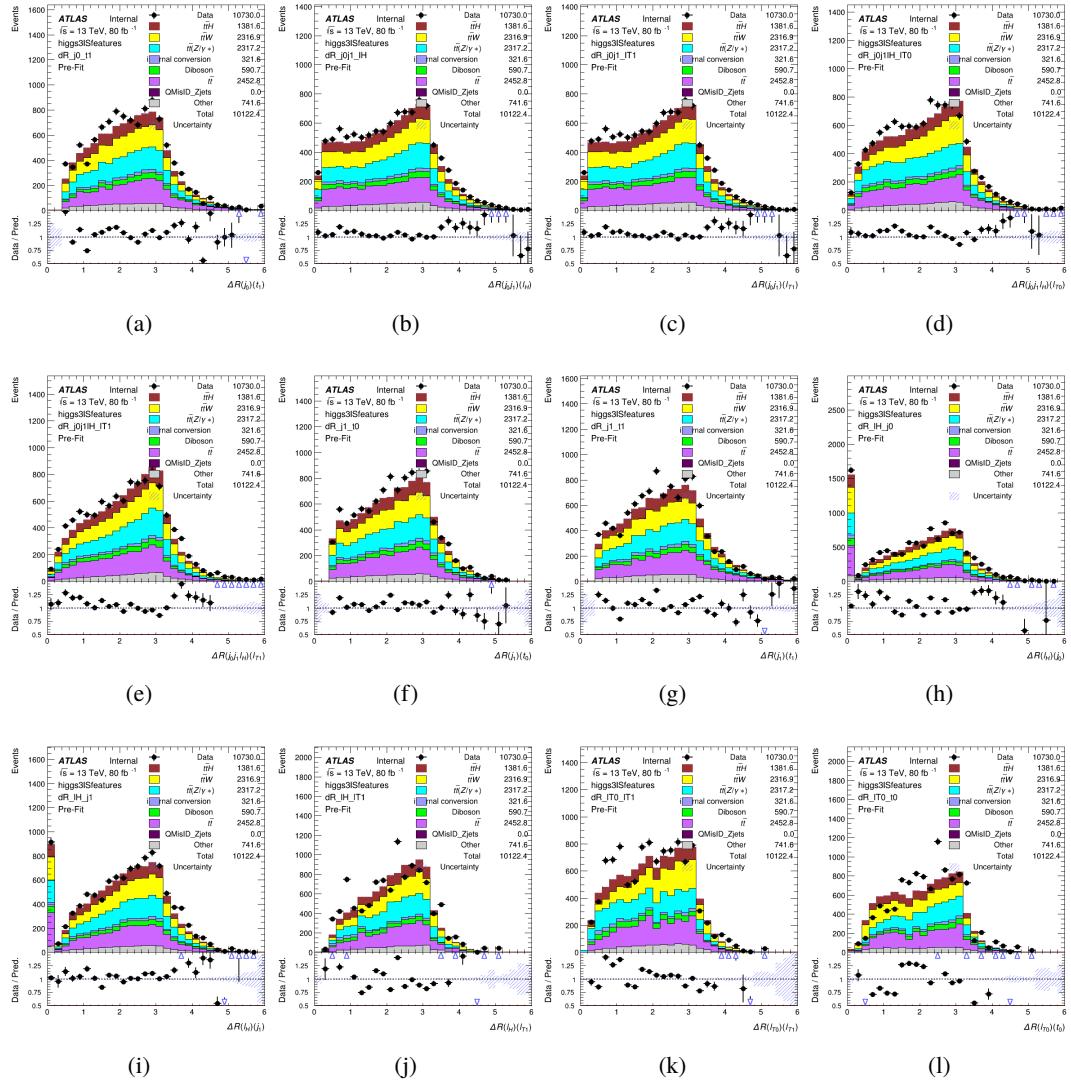


Figure C.13: Input features for higgs3lS

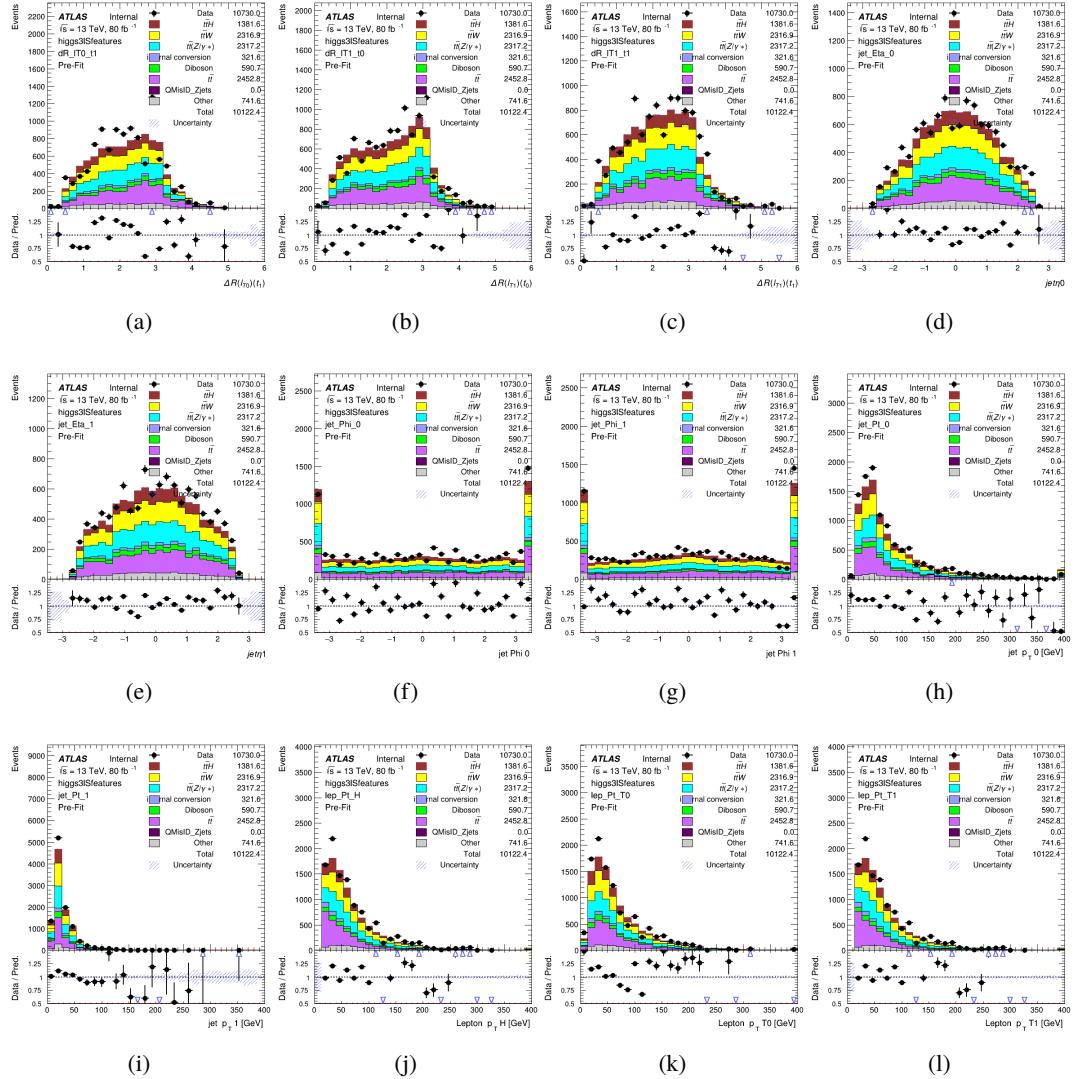


Figure C.14: Input features for higgs3IS

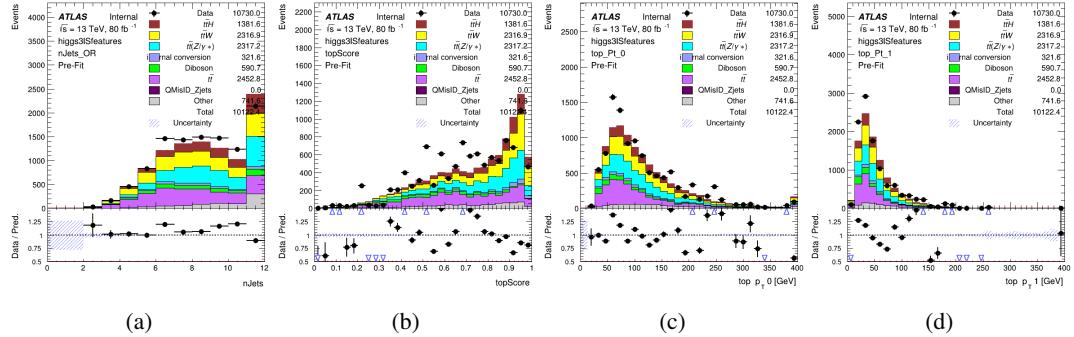


Figure C.15: Input features for higgs3IS

₁₇₈₃ **C.1.5 Higgs Reconstruction Features - 3lF**

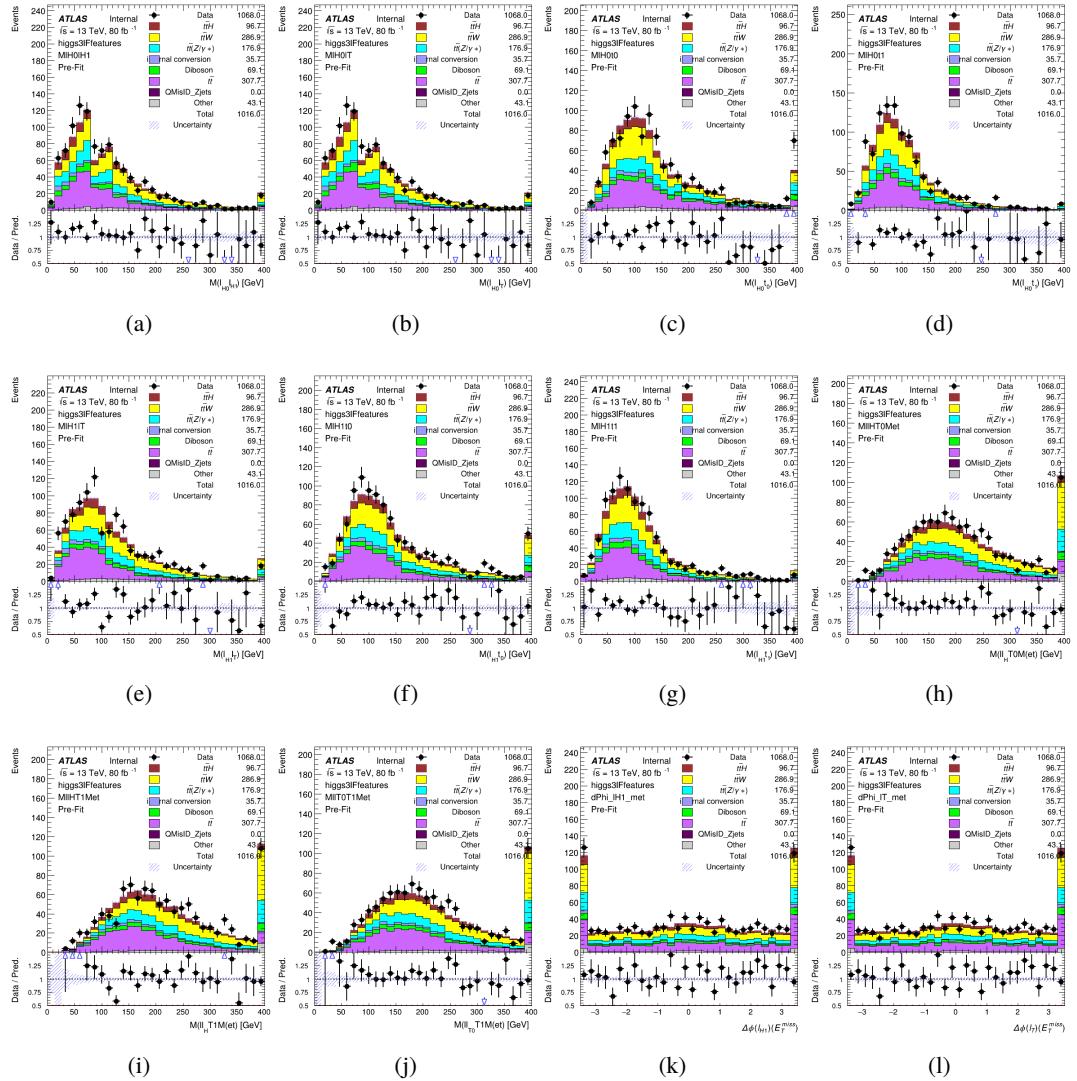


Figure C.16: Input features for higgs3lF

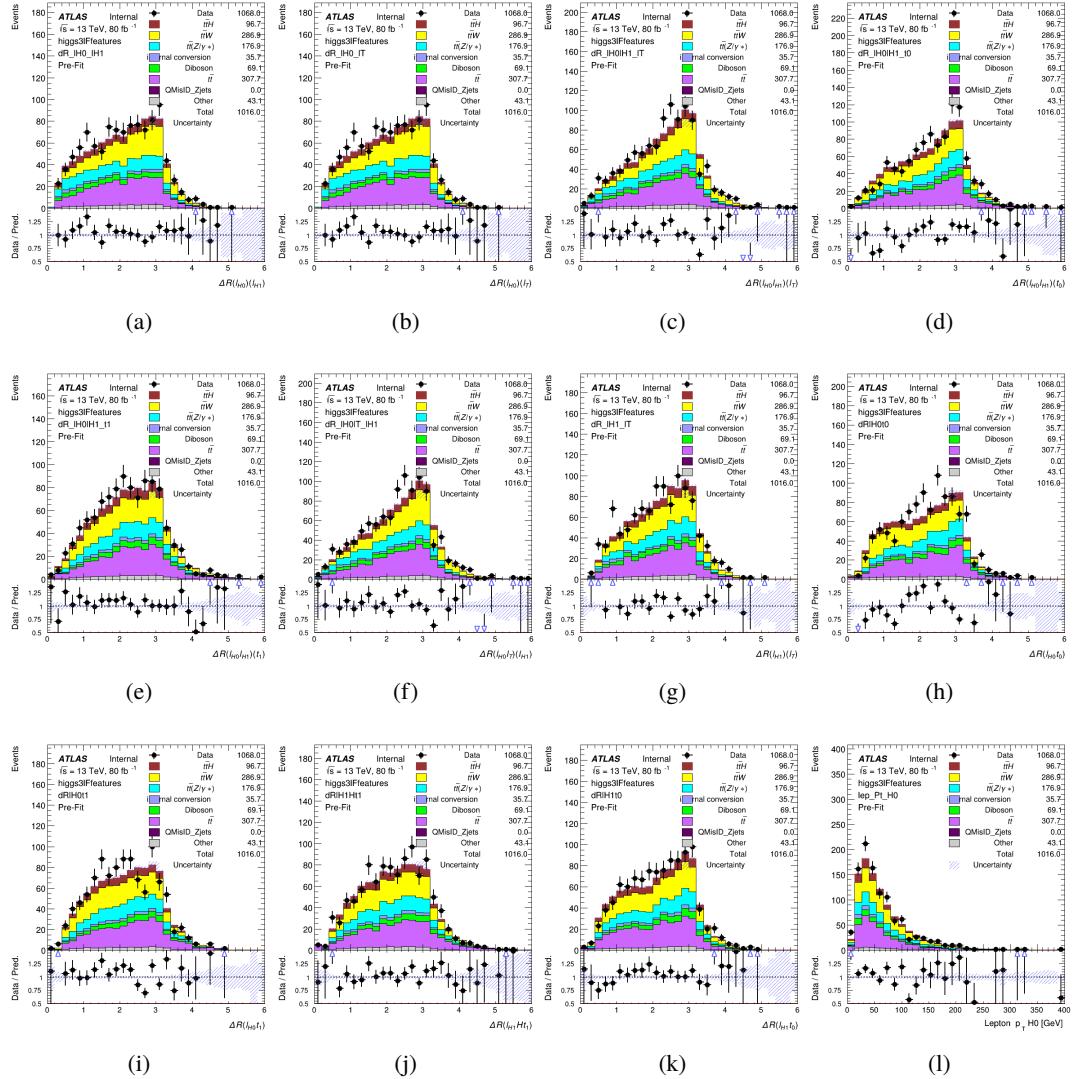


Figure C.17: Input features for higgs3IF

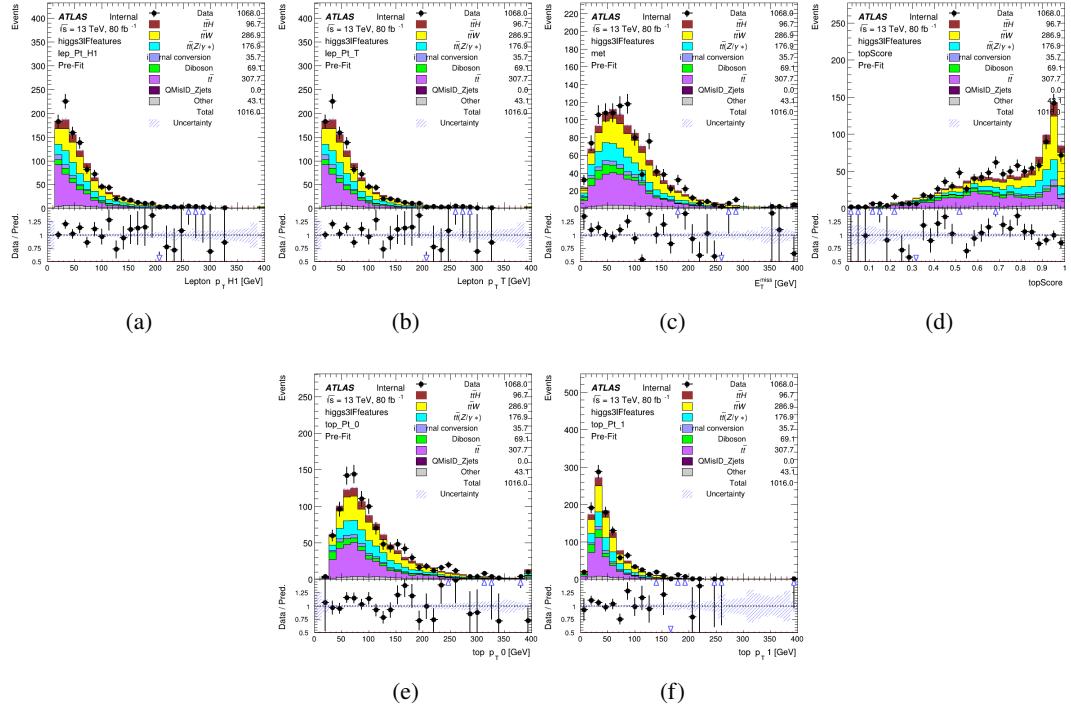


Figure C.18: Input features for higgs3lF

¹⁷⁸⁴ **C.2 Background Rejection MVA Details**

¹⁷⁸⁵ **C.2.1 Background Rejection MVA Features - 2ISS**

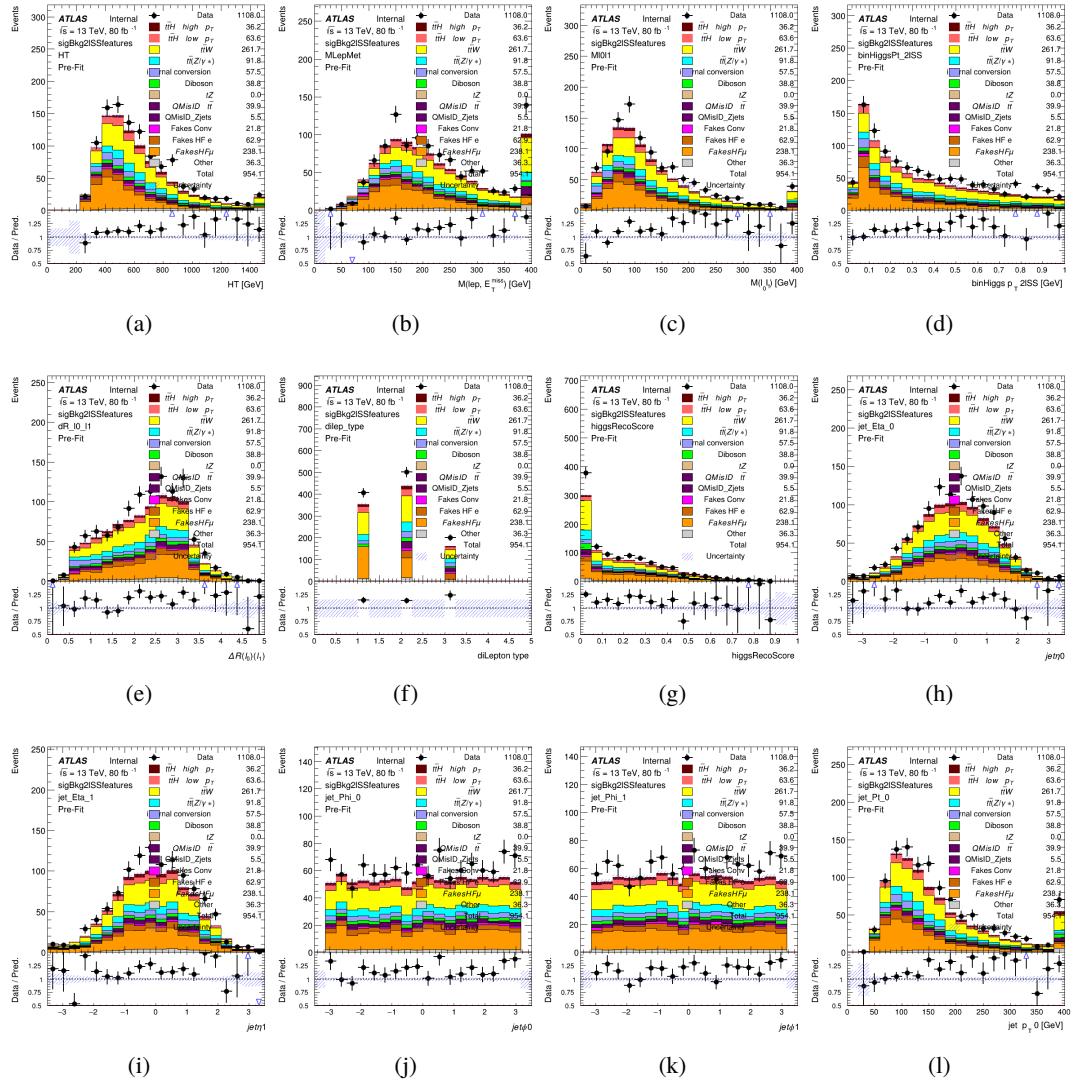


Figure C.19: Input features for sigBkg2ISS

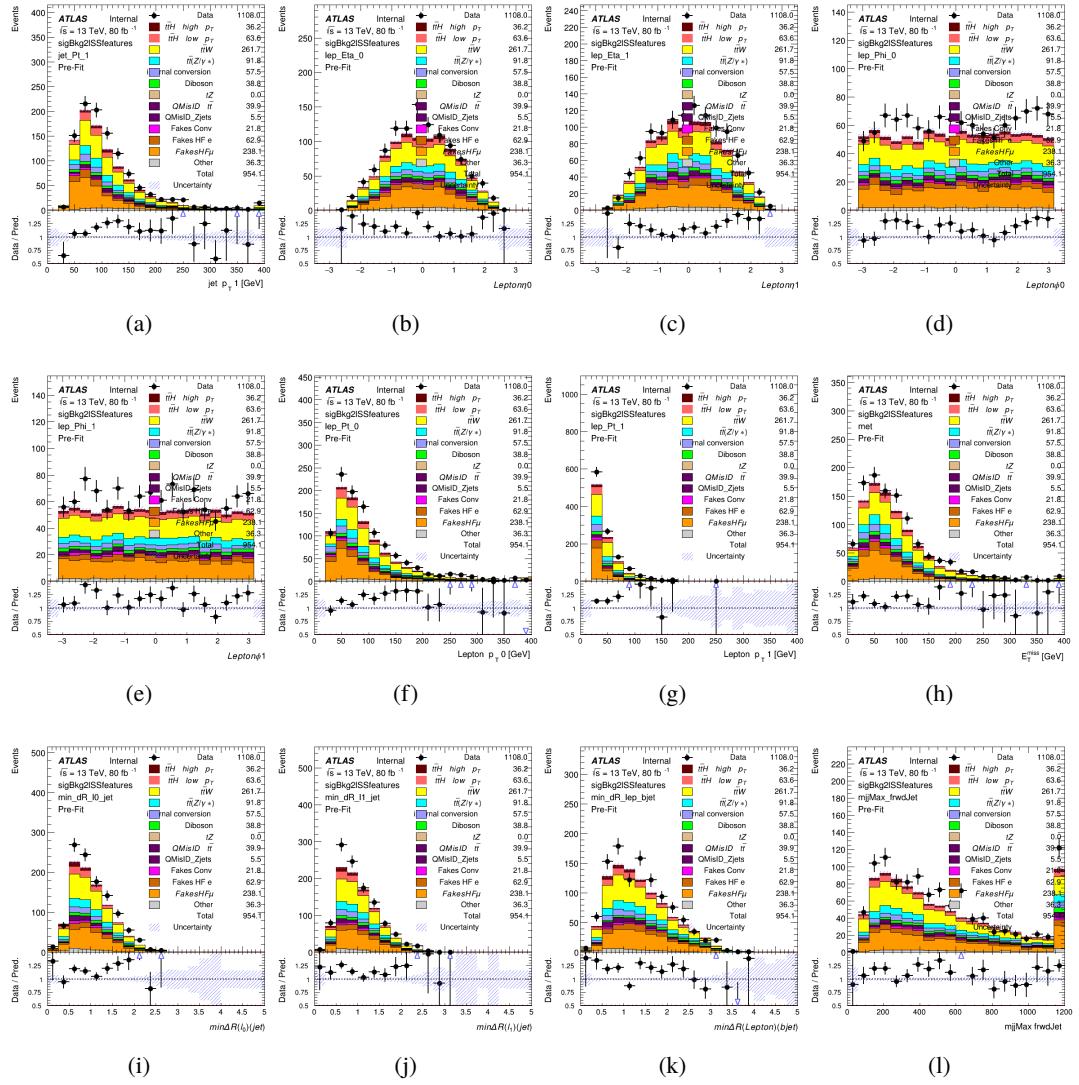


Figure C.20: Input features for sigBkg2lSS

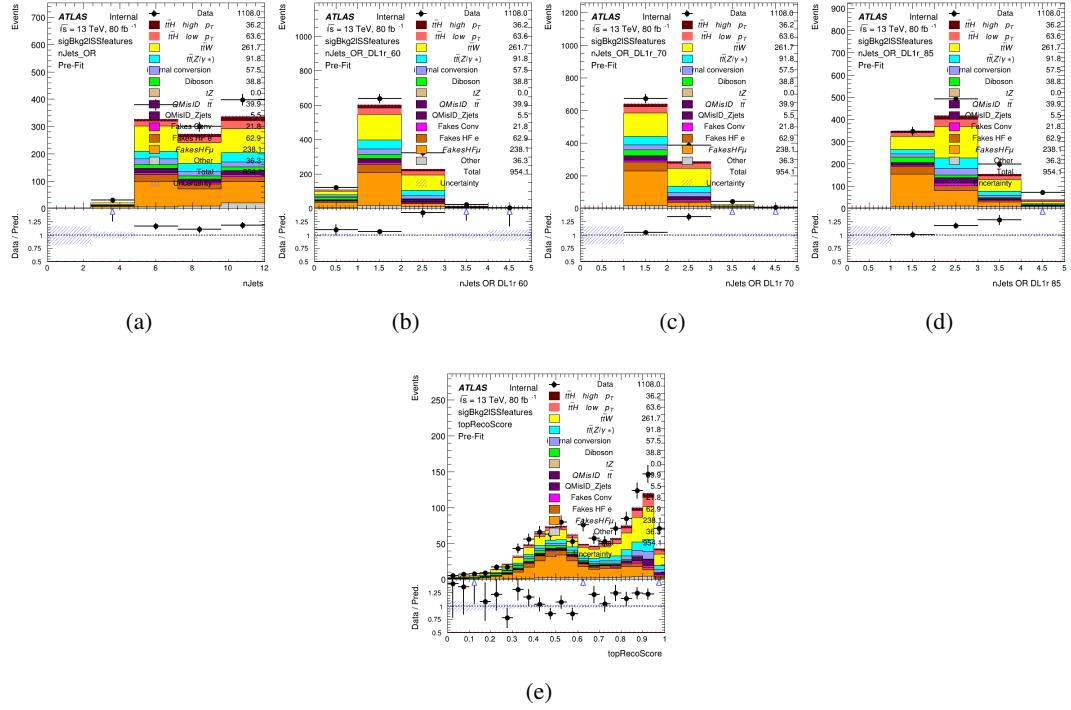


Figure C.21: Input features for sigBkg2lSS

¹⁷⁸⁶ **C.2.2 Background Rejection MVA Features - 3l**

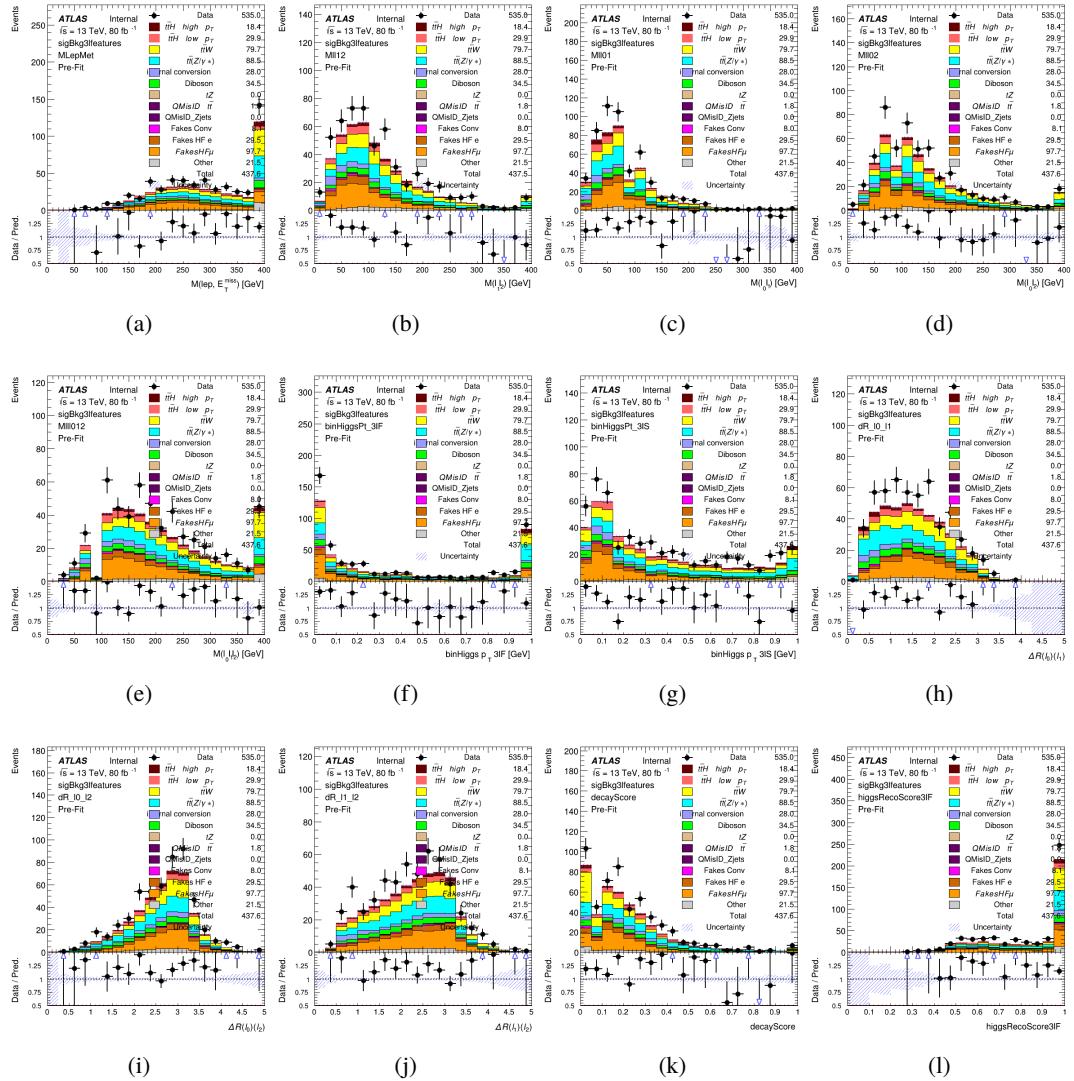


Figure C.22: Input features for sigBkg3l

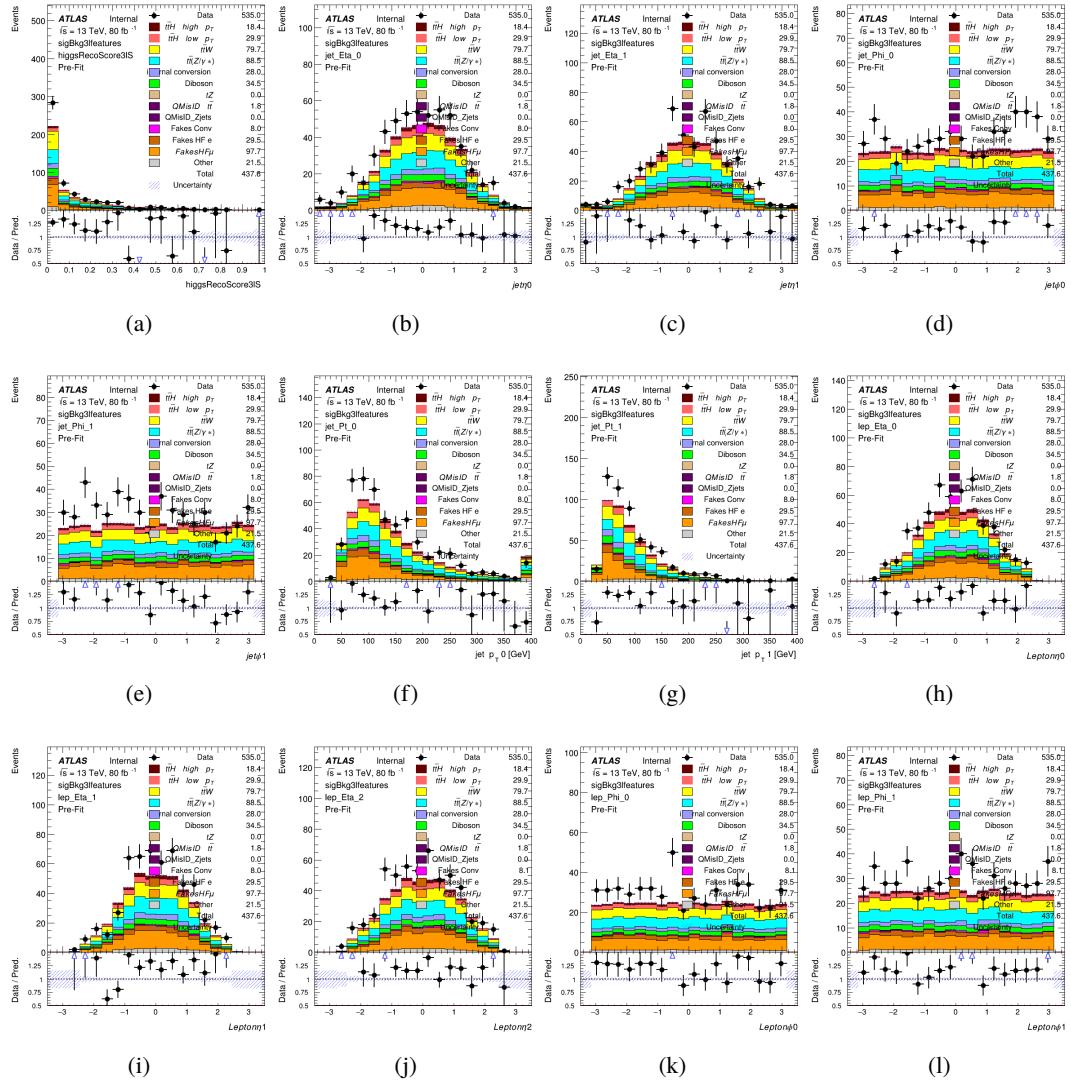
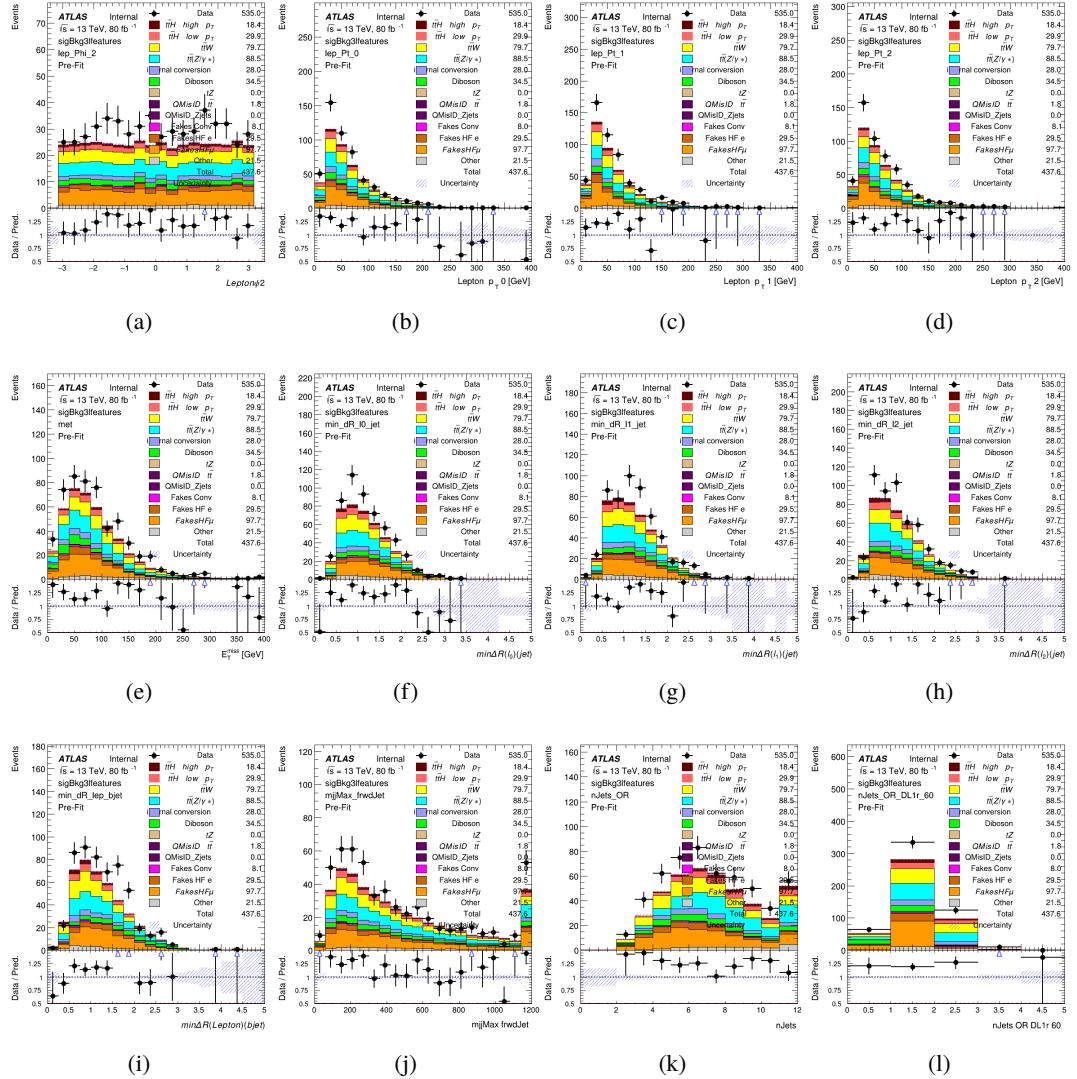


Figure C.23: Input features for sigBkg3l

Figure C.24: Input features for `sigBkg3l`

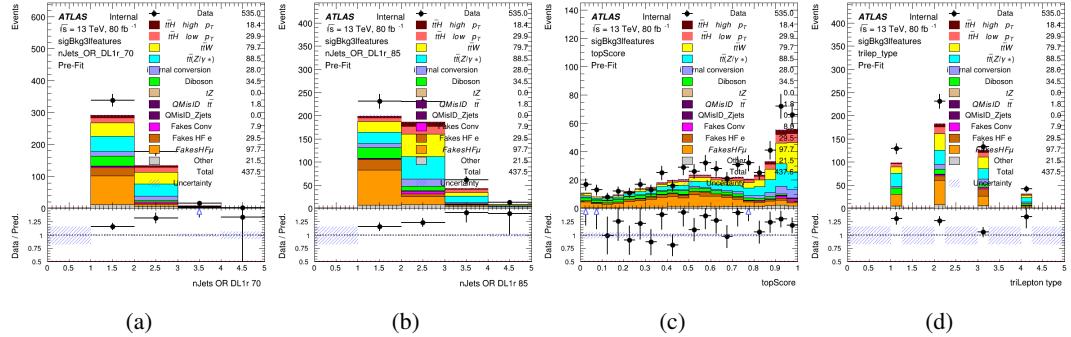


Figure C.25: Input features for sigBkg3l

1787 C.3 Truth Level Studies

1788 Attempts to identify the decay products of the Higgs are motivated by the ability to reconstruct
 1789 the Higgs momentum based on their kinematics. In order to demonstrate that this is case, the
 1790 kinematics of reconstructed objects that are truth matched to the Higgs decay are used as inputs
 1791 to a neural network which is designed to predict of the momentum of the Higgs. This is done in
 1792 the 2lSS channel, as it proves to be the most challenging for p_T reconstruction.

1793 Only leptons and jets which are truth matched to the Higgs are used as inputs for the
 1794 model; events where the lepton and both jets are not reconstructed are not included. The model
 1795 uses the same feature set and network architecture as the p_T prediction model used in the main
 1796 analysis, as described in Section 14.4.1.

1797 The results of the model are summarized below:

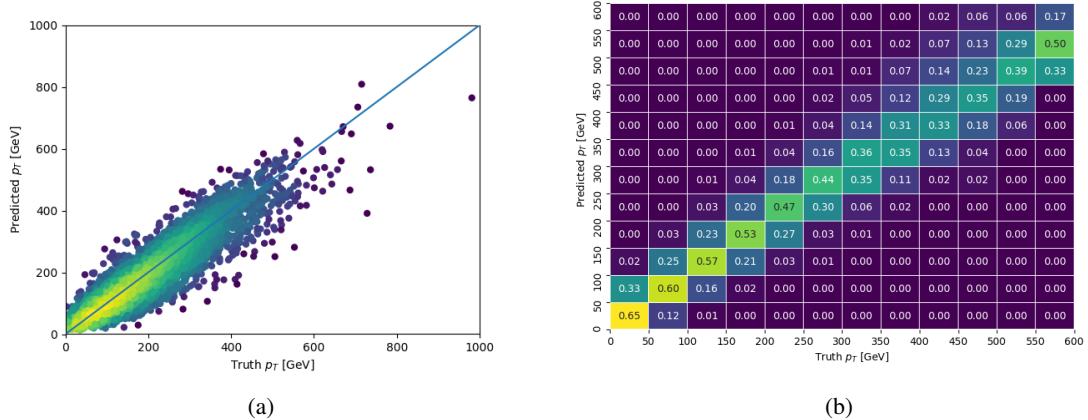


Figure C.26: The regressed Higgs momentum spectrum as a function of the truth p_T for 2lSS $t\bar{t}H$ events in (a) a scatterplot, where the color of each point represents the log of the point density, based on Gaussian Kernel Density Estimation, and (b) a heatmap histogram where each column has been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

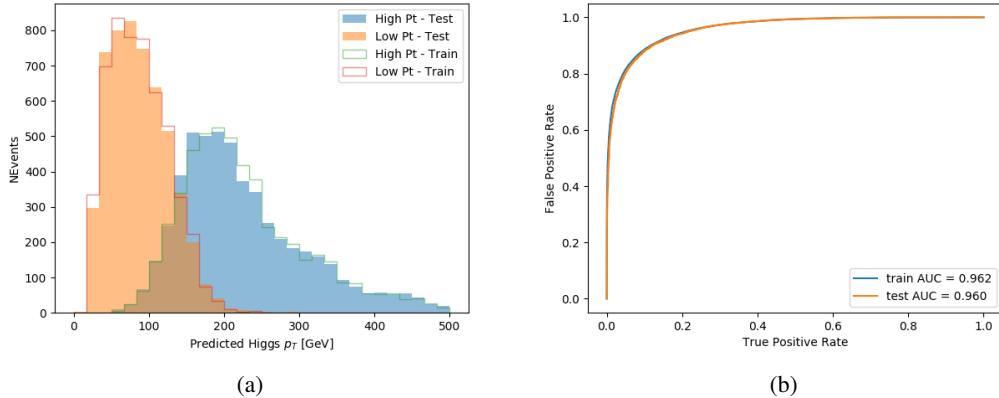


Figure C.27: (a) shows the reconstructed Higgs p_T for 2lSS events with truth $p_T > 150$ GeV and < 150 GeV, while (b) shows the ROC curve for those two sets of events.

1798 Based on the performance of the model, as shown Figures C.26 and C.27, the Higgs
 1799 momentum can be reconstructed with fairly high precision when its decay products are correctly
 1800 identified.

1801 C.4 Alternate b-jet Identification Algorithm

1802 The nominal analysis reconstructs the b-jets by considering different combinations of jets, and
 1803 asking a neural network to determine whether each combination consists of b-jets from top quark
 1804 decays. An alternate approach would be to give the neural network about all of the jets in an event
 1805 at once, and train it to select which two are most likely to be the b-jets from top decay. It was
 1806 hypothesized that this could perform better than considering each combination independently, as
 1807 the neural network could consider the event as a whole. While this is not found to be the case,
 1808 these studies are documented here as a point of interest and comparison.

1809 For these studies, the kinematics of the 10 highest p_T jets in each event are used for
 1810 training. This includes the vast majority of truth b-jets. Specifically the p_T , η , ϕ , E , and DL1r
 1811 score of each jet are used. For events with fewer than 10 jets, these values are substituted with 0.
 1812 The p_T , η , ϕ , and E of the leptons and E_T^{miss} are included as well. Categorical cross entropy is
 1813 used as the loss function.

Table 53: 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%

1814 C.5 Binary Classification of the Higgs p_T

1815 A two bin fit of the Higgs momentum is used because statistics are insufficient for any finer
 1816 resolution. This means separating high and low p_T events is sufficient for this analysis. As
 1817 such, rather than attempting to reconstruct the full Higgs p_T spectrum, a binary classification
 1818 approach is explored.

1819 A model is built to determine whether $t\bar{t}H$ events include a high p_T (>150 GeV) or low
 1820 p_T (<150 GeV) Higgs Boson. While this is now a classification model, it uses the same input
 1821 features described in section 14.4. Binary crossentropy is used as the loss function.

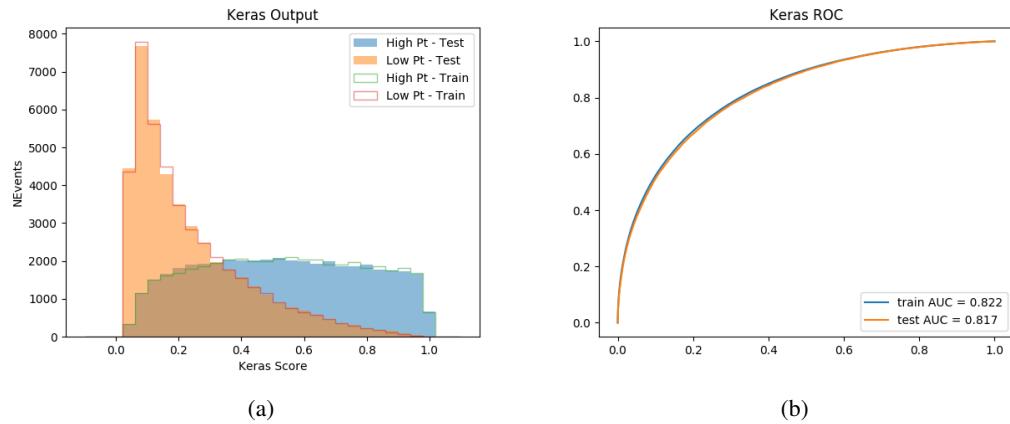


Figure C.28: Output distribution of the NN score for the binary high/low p_T separation model in the 2lSS channel.

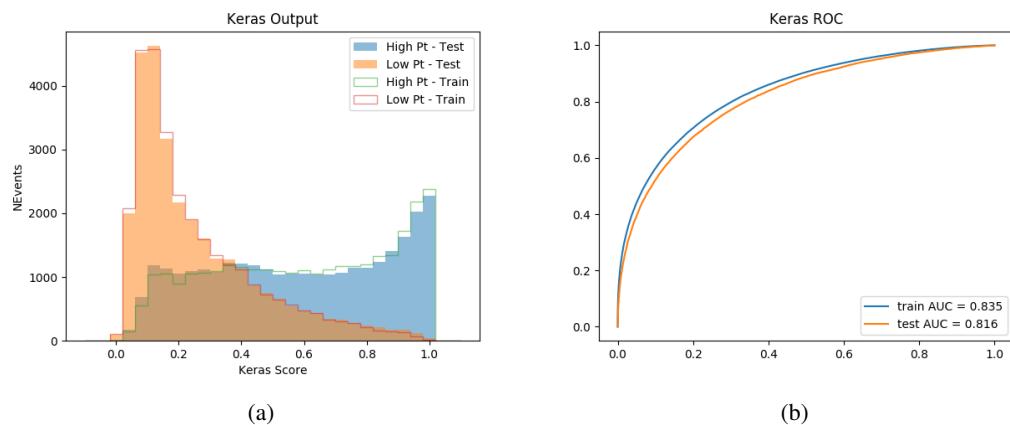


Figure C.29: Output distribution of the NN score for the binary high/low p_T separation model in the 3lS channel.

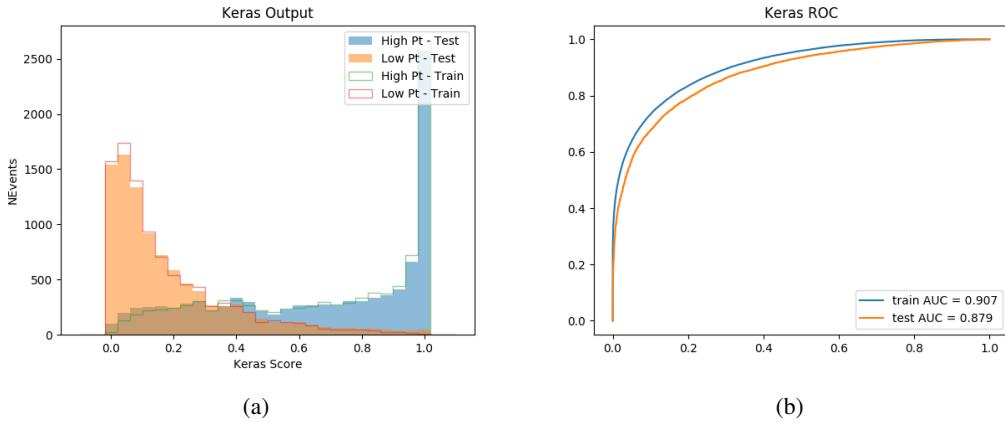


Figure C.30: Output distribution of the NN score for the binary high/low p_T separation model in the 3LS channel.

C.6 Impact of Alternative Jet Selection

1823 A relatively low p_T threshold of 15 GeV is used to determine jet candidates, as the jets originating
 1824 from the Higgs decay are found to fall between 15 and 25 GeV a large fraction of the time. The
 1825 impact of different jet p_T cuts on our ability to reconstruct the Higgs p_T is explored here.

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

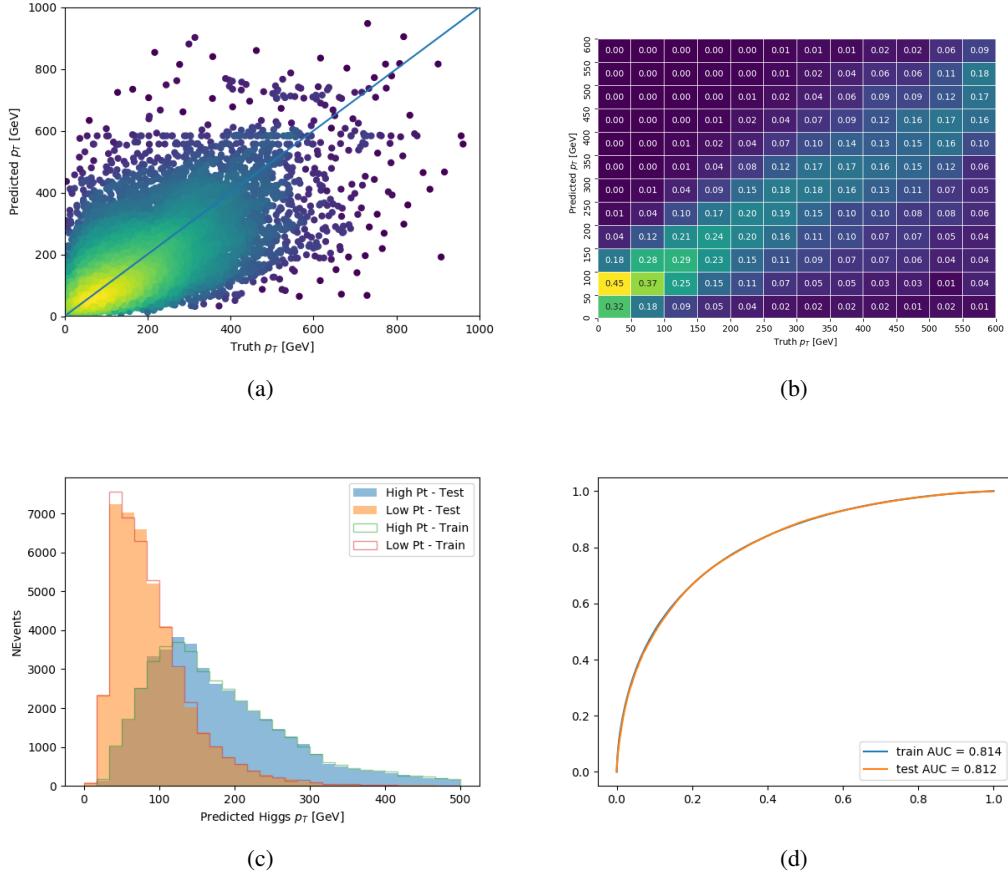
Jet $p_T > 20$ GeV

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

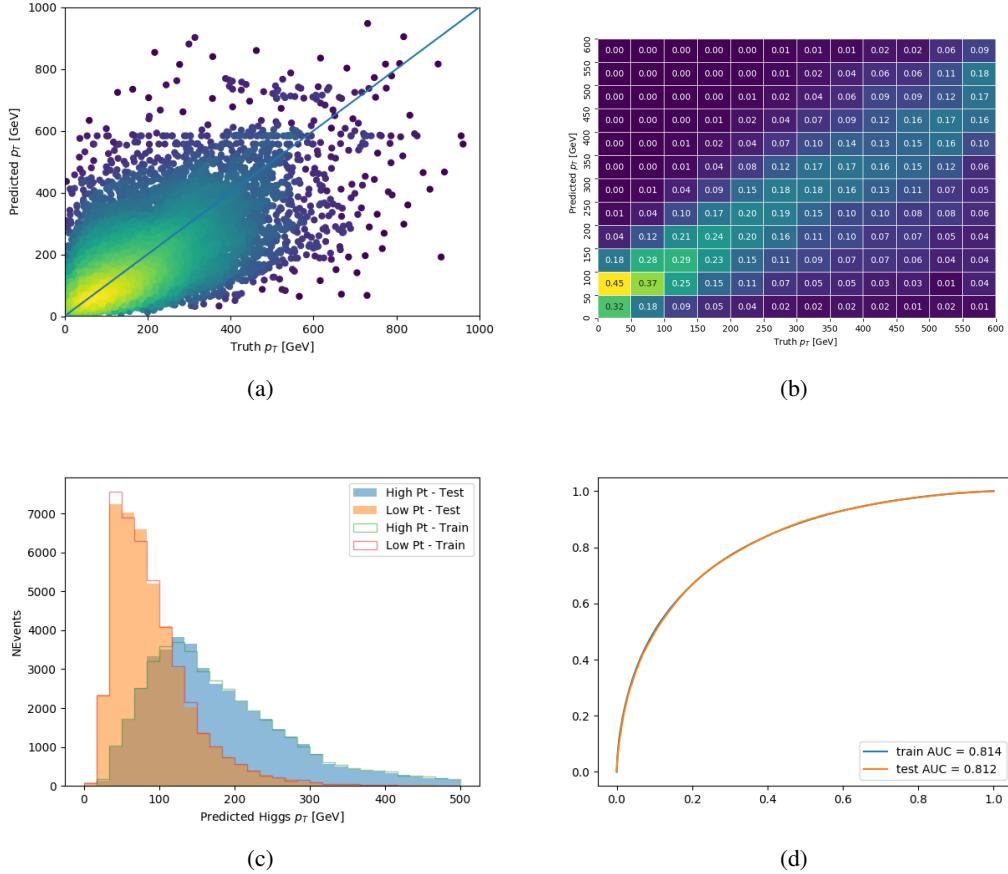
Jet $p_T > 25 \text{ GeV}$ 

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

1829 **D**