



DEPARTMENT OF PHYSICS
UNIVERSITY OF CAPE TOWN
IYUNIVESITHI YASEKAPA • UNIVERSITEIT VAN KAAPSTAD

Search for tWZ production in the Full Run 2 ATLAS
dataset using events with four leptons

Jake Reich

¹

Student Number: RCHJAK001

Supervisor: Dr. James Keaveney

Co-Supervisor: Dr. Sahal Yacoob

A thesis submitted in fulfilment
of the requirements for the degree of
Master in Science

August 2021

Abstract

3 The search for tWZ production using 139 fb^{-1} of pp collision data at a centre-of-mass energy of $\sqrt{s} = 13$
4 TeV, recorded by the ATLAS experiment at CERN, is presented. Events that contain exactly four
5 leptons (electrons or muons) are selected. Additional criteria are applied based on the number of jets,
6 the number of b -tagged jets and the number of Z boson candidates to define signal and control regions.
7 The large $t\bar{t}Z$ and ZZ backgrounds are distinguished from signal by a BDT-based algorithm. Inputs to
8 the BDT-based algorithm include, a kinematic reconstruction algorithm which reconstructs leptonically
9 decaying top quarks and a variable constructed from the output of a BDT-based algorithm which aims to
10 classify ℓb systems originating from top quarks. The expected signal strength is extracted via a blinded
11 maximum-likelihood fit to multiple signal and control regions. The expected signal strength was measured
12 as, $\mu(tWZ) = 1.91^{+0.95}_{-0.82}$, leading to an expected significance of 1.44σ . An expected upper limit on the
13 signal strength is set to $\mu_{up}^{exp} = 1.61^{+2.35}_{-1.16}$. Furthermore, a combined blinded maximum-likelihood fit was
14 performed across the tetralepton channel and an independent analysis of the trilepton channel, to further
15 increase the sensitivity this analysis to the tWZ cross section. The expected signal strength was measured
16 as, $\mu(tWZ) = 1.80^{+0.70}_{-0.65}$, leading to an expected significance of 1.61σ . An expected upper limit on the
17 signal strength is set to $\mu_{up}^{exp} = 1.43^{+2.04}_{-1.03}$.

Declaration

19 *I certify that this assignment/report is my own work, based on my personal study and/or research and
20 that I have acknowledged all material and sources used in its preparation, whether they be books, articles,
21 reports, lecture notes, and any other kind of document, electronic or personal communication. I also
22 certify that this assignment/report has not previously been submitted for assessment in any other unit,
23 except where specific permission has been granted from all unit coordinators involved, or at any other time
24 in this unit, and that I have not copied in part or whole or otherwise plagiarised the work of other students
25 and/or persons.*

Acknowledgements

Contents

27

28	1 Introduction	7
29	2 Theory	8
30	2.1 Standard Model of Particle Physics	8
31	2.1.1 The Top Quark	10
32	2.2 tWZ	12
33	2.2.1 Tetralepton Channel	12
34	2.2.2 Comparison to Trilepton Channel	13
35	3 The ATLAS Experiment and Detector	15
36	3.1 The ATLAS Experiment	15
37	3.1.1 Large Hadron Collider (LHC)	15
38	3.2 The ATLAS Detector	17
39	3.2.1 Coordinate System and Kinematics	18
40	3.2.2 Inner Detector	18
41	3.2.3 Electromagnetic and Hadronic Calorimeters	19
42	3.2.4 Muon Spectrometer	19
43	3.2.5 Trigger and Data Acquisition System	19
44	3.2.6 Particle Identification and Object Reconstruction	19
45	4 The search for tWZ production using events with four leptons	22
46	4.1 Data and Monte Carlo Simulation	22
47	4.1.1 Data Samples	22
48	4.1.2 Monte Carlo Samples	23
49	4.1.3 Trigger Strategy	24
50	4.2 Physics objects	26
51	4.2.1 Leptons	26
52	4.2.2 Jets	27
53	4.2.3 b -tagging	27
54	4.2.4 Overlap Removal Procedure	28
55	4.3 Kinematic cuts	28
56	4.4 Regions and Event Selection	28
57	4.4.1 Optimization studies for event selection	29

58	4.5	Signal and Control Regions	31
59	4.5.1	tWZ OF SR	33
60	4.5.2	tWZ SF SR	35
61	4.5.3	$t\bar{t}Z$ CR	35
62	4.5.4	ZZb CR	37
63	4.5.5	$(tWZ)_{\text{fake}}$ CR	39
64	4.6	Fake Lepton Estimation	39
65	4.7	Improving signal vs background discrimination	42
66	4.7.1	Two Neutrino Scanning Method ($2\nu\text{SM}$) Algorithm	42
67	4.7.2	Boosted Decision Trees	48
68	4.7.3	Object-level BDT	49
69	4.7.4	Event-level BDT	56
70	4.8	Systematic Uncertainties	62
71	4.8.1	Experimental uncertainties	62
72	4.8.2	Theoretical uncertainties	63
73	4.8.3	Generic shape systematic uncertainties	64
74	4.9	Analysis Pipeline and TRexFitter	64
75	4.9.1	Fitting Procedure	65
76	4.10	Results	67
77	4.10.1	Tetralepton Channel	68
78	4.10.2	Trilepton and Tetralepton Channels	71
79	5	Summary and Conclusions	76

80

Chapter 1

81

Introduction

82 The production of a single top quark in association with a W^\pm and Z boson (tWZ) at the CERN LHC is
83 sensitive to both the neutral and charged electroweak couplings of the top quark as the process involves
84 the simultaneous production of a W boson and a Z boson in association with the top quark. Due to the
85 very large coupling of the top quark to the Higgs boson, the electroweak couplings of the top quark are a
86 theoretically well-motivated area in which to search for the first signs of new physics [75, 50] that could
87 offer a resolution to the Hierarchy Problem [58, 29]. The recent lack of signs of new physics from LHC
88 data tells us that new physics is either very heavy, or is very weakly coupled to Standard Model particles,
89 therefore signs of new physics might only be observed in anomalous rates of well-chosen processes. A
90 prime example of such a process is tWZ . This has an extremely low production cross section (≈ 160 fb
91 for $\sqrt{s} = 13$ TeV [24]), meaning that it is an extremely rare process and subsequently, it has never been
92 observed by any particle physics experiment.

93

94 The latest datasets recorded by the ATLAS experiment at the CERN LHC are sufficiently large to
95 potentially allow for an observation of tWZ production. In this analysis, the Full Run 2 dataset recorded
96 by ATLAS is used to search for tWZ production in the tetralepton channel (tWZ with exactly four
97 final state leptons). A kinematic reconstruction technique is used which aims to discriminate between
98 tWZ and our most prominent background process, $t\bar{t}Z$. In addition to this, Machine Learning techniques
99 are implemented to further isolate our tWZ signal. Backgrounds from Standard Model (SM) processes
100 including those in which one or more leptons originate from the semileptonic decay of a heavy hadron
101 or a photon conversion are estimated by fitting predictions from simulation to data in dedicated control
102 regions. As this work forms the basis of an official ATLAS analysis, only blinded results are shown.
103 A maximum likelihood fit is performed over our two tWZ signal regions and three control regions, to
104 measure the cross section of tWZ in the tetralepton channel. In this thesis only the tetralepton channel
105 is explored, however, the results of a combined fit that includes an entirely independent analysis of the
106 trilepton channel is presented with no overlap of events between the regions defined for these channels.

107

Chapter 2

108

Theory

109 2.1 Standard Model of Particle Physics

110 The SM is a model based on Quantum Field Theory (QFT) which classifies all known elementary particles
 111 and describes their interactions. It is a well-tested model and has shown to be hugely successful in
 112 describing experimental data to great precision [8, 7]. For example, in the top quark sector, the $t\bar{t}$ cross
 113 section predictions have been confirmed to 3.9% accuracy [39, 67]. It incorporates three of the four
 114 fundamental forces of nature: the electromagnetic, the weak and the strong forces. In Figure 1, all known
 115 elementary particles described by the SM, are shown.

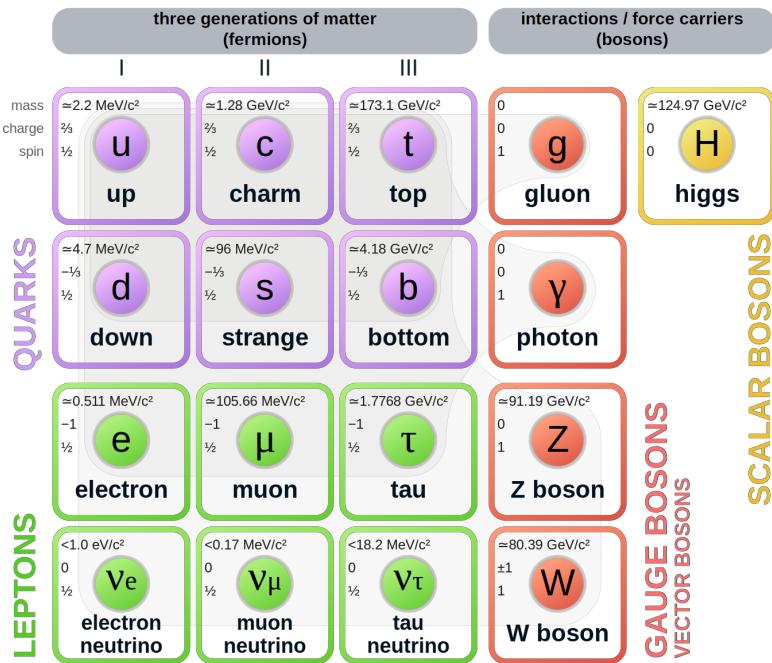


Figure 1: A summary of the elementary particles described by the SM [87] is shown. Fermions are shown on the left, with quarks shown in purple and leptons shown in green. Bosons are shown on the right, with gauge bosons shown in red and the Higgs boson shown in yellow. The mass, electric charge and spin of each particle is shown on the top left of each particle's block.

116 Particles in the SM are uniquely described by their quantum numbers: electric charge and spin. The
117 SM particles are split into main two classes, based off their spin quantum numbers. Particles which
118 have half-integer spin are called fermions, and those which have integer spin are called bosons. Fermions
119 are further divided into three generations, each comprising of two quarks, one charged lepton and one
120 neutrino. In a generation, the more massive quark has an electric charge of $+2/3$ (up-type) and the less
121 massive quark has an electric charge of $-1/3$ (down-type). All charged leptons have an electric charge
122 of -1 and all neutrinos are electrically neutral. The masses of the particles in a generation increase
123 with increasing generation number, with generation 1 particles being the least massive and generation 3
124 particles being the most massive. Quarks carry electric and colour charge, and can therefore interact via
125 the electromagnetic, weak and strong forces. Colour charge can take on three values: red, green and blue.
126 It is important to note that colour charge is completely unrelated to the everyday meaning of colour, and
127 it just represents the quantum state of the particle. Due to colour confinement [96], quarks cannot be
128 isolated from one another. They exist in colourless bound states, called hadrons, consisting of two or more
129 quarks. Hadrons consisting of an even number of quarks are known as mesons and those consisting of an
130 odd number of quarks are known as baryons. On the other hand, charged leptons (electron (e), muon (μ)
131 and tau (τ)) only carry electric charge and can therefore interact electromagnetically and weakly, but not
132 through the strong interaction. The electric and colour neutral fermions, neutrinos, can only interact via
133 the weak force.

134
135 Particles are able to interact with one-another via the exchange of a gauge boson (boson with spin-1).
136 Photons are massless, spin-1 gauge bosons which mediate electromagnetic interactions between particles
137 which carry electric charge, such as quarks and charged leptons (e , μ and τ). The weak interaction is
138 mediated by three massive gauge bosons, the electrically charged W^+ and W^- bosons and the electrically
139 neutral Z boson. Gluons are massless, spin-1 gauge bosons which mediate strong interactions between
140 particles which carry colour charge, such as quarks. Since gluons carry colour charge, they interact with
141 themselves. The massive, spin-0, electrically neutral Higgs boson mediates the Higgs field which gives
142 mass to the W^\pm and Z bosons via the so-called Brout-Englert-Higgs mechanism [44, 55, 54]. The Brout-
143 Englert-Higgs mechanism induces spontaneous electroweak symmetry breaking to provide mass terms for
144 the W^\pm and Z bosons in the electroweak Lagrangian of the SM. All particles described in the SM have
145 their own antiparticle, with the same mass, but opposite charges. Some particles, such as the photon, are
146 their own antiparticle.

147
148 Although the SM is hugely successful, it is incomplete and fails to describe certain observed phenomena.
149 The most notable example being the absence of gravity. The gravitational force is $\approx 10^{29}$ [88] weaker than
150 the weak force, therefore quantum gravitational effects are expected to only become significant at energies
151 much larger than that currently accessible by the LHC (known as the Planck scale $\approx 10^9$ GeV) [58]. This
152 large difference in strength between the weak force and gravity is known as the Hierarchy Problem.
153 Cosmological observations infer that around 84% of the matter in the universe consists of gravitationally
154 interacting matter known as dark matter [57]. None of the particles described in the SM are good dark
155 matter candidates, therefore the SM only accounts for a small fraction of the total matter of the universe.
156 The large discrepancy between the observed amount of matter and antimatter in the universe, sometimes

referred to as the matter-antimatter asymmetry, is not fully explained by the SM. Neutrinos in the SM are assumed to be massless, however observations of neutrino oscillations (neutrinos undergoing flavour change as they travel through space) imply that neutrinos do have mass [47]. Beyond the Standard Model (BSM) theories attempt to explain the phenomena which the SM cannot. For example, a popular extension to the SM, Supersymmetry (SUSY) introduces new particles to the SM which are counterparts to the existing SM particles with the same quantum numbers, except for their spins [60]. It provides elegant explanations to many shortcomings of the SM, however none of the supersymmetric particles described by SUSY have been observed experimentally [33].

2.1.1 The Top Quark

The top quark is the heaviest particle in the SM, with a mass of 172.76 ± 0.30 GeV [52]. According to the SM, since the coupling to the Higgs boson is proportional to the mass of the interacting particle, the top quark is strongly coupled to the Higgs boson. Therefore, physics processes involving top quarks is a theoretically well-motivated area to search for new physics, since top quarks are the most likely particles to couple to new physics theories at the TeV scale. Its large mass also makes it highly unstable, with a mean lifetime of $\approx 0.5 \times 10^{-24}$ s [52]. The top quark's lifetime is shorter than that of the hadronisation process, and it therefore decays before hadronising. The top quark can therefore be measured indirectly via its decay products. Top quarks almost always decay to a W boson and a b -quark ($\frac{\Gamma(Wb)}{\Gamma(Wq(q=b,s,d))} = 0.957 \pm 0.034$ [52]). The b -quark is the second heaviest quark in the SM, however its lifetime is still longer than the hadronisation time scale [52]. In hadron collider experiments, b -quarks travel a short distance in the detector before hadronising to form jets. In Table 1, the dominant final state branching fractions of the top quark are shown.

Decay Mode	Branching Fraction ($\frac{\Gamma_i}{\Gamma}$)
$t \rightarrow Wb \rightarrow e\nu_e b$	$(11.10 \pm 0.30)\%$
$t \rightarrow Wb \rightarrow \mu\nu_\mu b$	$(11.40 \pm 0.20)\%$
$t \rightarrow Wb \rightarrow \tau\nu_\tau b$	$(10.70 \pm 0.50)\%$
$t \rightarrow Wb \rightarrow q\bar{q}b$	$(66.50 \pm 1.40)\%$

Table 1: The dominant final state branching fractions of the top quark [52] are shown.

Top quark decays with hadronic final states are more than twice as likely than those with leptonic final states. Final state decays to different lepton flavours are roughly equally probable.

180

Top quark production can be placed into two main categories: pair production ($t\bar{t}$) and single-top production (t) [83]. In the LHC, top quarks are mainly produced in pairs via strong interactions in gluon-gluon fusion ($gg \rightarrow t\bar{t}$) or quark annihilation ($q\bar{q} \rightarrow t\bar{t}$). Top quark production via gluon-gluon fusion is the dominating process [22]. The production cross section for $t\bar{t}$ (leptonic final state) in pp collisions with $\sqrt{s} = 13$ TeV was measured by ATLAS with a value of $830 \pm 0.4(\text{stat}) \pm 36(\text{syst}) \pm 14(\text{lumi})$ pb [4], with good agreement between measurement and theoretical prediction.

187

188 Single top production occurs via the weak interaction. The most abundant production mechanisms leading
 189 to single top production are the s -, t - and Wt - channels [52]. In the s -channel, an initial quark annihilates
 190 with an anti-quark of different flavour, producing a virtual W boson which decays to a top quark and
 191 anti-bottom quark. In the t -channel, an initial b quark interacts with a different flavour quark via the
 192 exchange of a W boson. This interaction produces a top quark and another quark. In the Wt -channel,
 193 an initial gluon interacts with a b quark to produce a top quark and a W boson, either via the absorption
 194 of the gluon by the b quark or via the exchange of a top quark. In Table 2, single top production cross
 195 sections in pp collisions at $\sqrt{s} = 13$ TeV for the s -, t - and Wt -channels, are shown.

Channel	Process	Total Cross Section [pb]
s	$q\bar{q}' \rightarrow W \rightarrow \bar{b}t$	$10.32^{+0.40}_{-0.36}$
t	$bq' \rightarrow W \rightarrow tq$	$216.99^{+9.04}_{-7.71}$
Wt	$bg \rightarrow b/t \rightarrow Wt$	71.7 ± 3.85

Table 2: Single top production cross sections in pp collisions at $\sqrt{s} = 13$ TeV for the s -, t - and Wt - channels [34] are shown. The prime superscript on q' indicates that the quark has a different flavour to q .

196 Single top production is suppressed compared to pair produced top production, with $t\bar{t}$ production (lepto-
 197 nomic final state) being around three times as likely to occur than single top production across all decay
 198 channels.

199 2.1.1.1 Motivation for the search for tWZ production in the tetralepton channel

200 The recent lack of signs of new physics from LHC data [86] tells us that new physics is either very heavy,
 201 or is very weakly coupled to SM particles. We therefore might only observe signs of new physics in
 202 anomalous rates of well-chosen processes. The tWZ process is a prime example of such a process. It
 203 has an extremely low production cross section of ≈ 160 fb for $\sqrt{s} = 13$ TeV [24], and has subsequently
 204 never been observed by any particle physics experiment. Since tWZ involves a charged W boson and
 205 neutral Z boson, its cross section is sensitive to the charged and neutral couplings to the top quark. In
 206 turn, the top quark is strongly coupled to the Higgs boson, due to the top quark's large mass. Due to
 207 the large coupling of the top quark to the Higgs boson, corrections to the Higgs boson mass diverge in
 208 the SM. The top quark's couplings are modified, in order to remove this divergence, in many scenarios
 209 of new physics that aim to resolve the Hierarchy Problem. Since the Z boson may be radiated from the
 210 initial-state b -quark, the final-state top quark, or the final-state Z boson, the tWZ process embeds the
 211 $b - Z$, $t - Z$ and $W - Z$ electroweak couplings which are often modified in BSM physics. Therefore tWZ
 212 is an important process in the search for signs of new physics and BSM physics.

213

214 One such BSM theory which is sensitive to tWZ production [62, 69] is the Standard Model Effective Field
 215 Theory (SMEFT) [26]. The SMEFT attempts to describe physics at large energy scales which we have
 216 not yet been able to probe experimentally. The SMEFT inherits the same QFT framework as the SM,
 217 and adds terms to the SM Lagrangian which describe the interactions of SM particles at higher energy
 218 scales. Analogous to the coupling constants found in the SM Lagrangian, which indicate the interaction

strengths between different particles, SMEFT contains scalar coefficients which operate in the same way. These scalar coefficients are known as Wilson coefficients. It has been shown that the cross section of tWZ is sensitive to many Wilson coefficients. An experimental constraint on the cross section of tWZ is therefore expected to be impactful on a global fit on all the Wilson coefficients in SMEFT.

223

Prior to this analysis, only three experimental studies of tWZ in ATLAS have been performed. Two of the studies utilised the trilepton channel to search for tWZ production, whereas the third study utilised both the tri- and tetralepton channels. The first search utilised 36 fb^{-1} of ATLAS data and an upper limit on the cross section of tWZ was set at a value of ≈ 6 times the SM cross section [74]. The second search in the trilepton channel utilised 139 fb^{-1} of ATLAS data and an expected upper limit on the cross section of tWZ was set at a value of ≈ 2.6 times the SM cross section [95]. In Section 4.10.2, the aforementioned analysis will be used in combination with this analysis, in order to further increase the sensitivity of this analysis to the cross section of tWZ . The third study investigated the feasibility of a cross section measurement of tWZ production with CMS Run 3 data (300 fb^{-1}) [94], by utilising the tri- and tetralepton channels. The study showed that it is possible to exclude $\mu(tWZ)$ at the 7σ significance level using 300 fb^{-1} of data. This study needs to be further investigated, since its findings seem improbable given the results obtained in this thesis.

236 2.2 tWZ

237 2.2.1 Tetralepton Channel

In Figure 2, the Leading Order (LO) Feynman diagram for tWZ in the tetralepton channel, is shown.

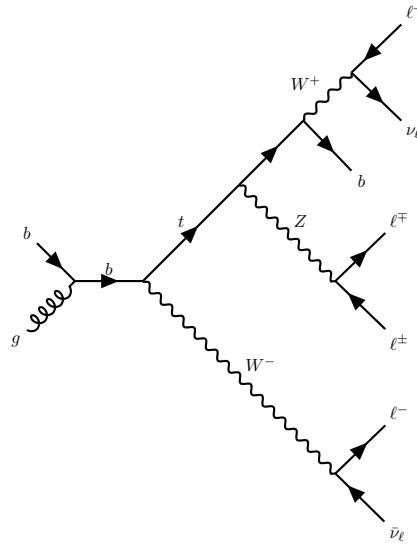
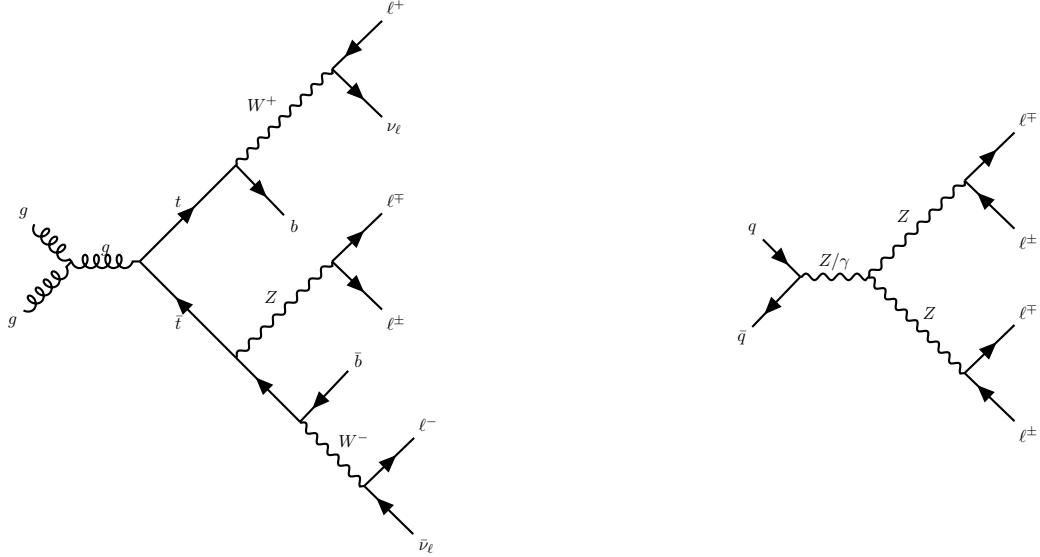


Figure 2: The LO Feynman diagram of tWZ production in the tetralepton channel is shown.

239 **2.2.1.1 Backgrounds**

240 The main backgrounds for tWZ (tetralepton channel) are the production of a two tops, both in the $\ell\nu b$ ¹
 241 final state channel, together with a Z boson ($t\bar{t}Z$) and diboson production with fully leptonic final states
 242 (ZZ). In Figure 2.2.1.1, LO Feynman diagrams for $t\bar{t}Z$ and ZZ in the tetralepton channel, are shown.



243

244

Figure 3: The LO Feynman diagrams for $t\bar{t}Z$ (left) and ZZ (right) in the tetralepton channel are shown.

245 The $t\bar{t}Z$ process contains fours leptons and two b -quarks in its final state ($\sigma_{(t\bar{t}Z).Br(4\ell)}^{\text{NLO}} = 0.95 \pm 0.08_{\text{stat}} \pm$
 246 $0.10_{\text{syst}} \text{ pb}$ at $\sqrt{s} = 13 \text{ TeV}$ [2]) and can easily mimic the tWZ signal process, for instance, by one of its
 247 b -jets being undetected or mis-identified in the detector. The ZZ process contains four leptons and zero
 248 b -quarks in its final state ($\sigma_{(ZZ).Br(4\ell)}^{\text{NNLO}} = 14.6^{+1.9}_{-1.8}(\text{stat})^{+0.5}_{-0.3}(\text{syst}) \pm 0.2(\text{theo}) \pm 0.4(\text{lumi}) \text{ pb}$ at $\sqrt{s} = 13$
 249 TeV [59]). One way in which ZZ can mimic the tWZ signal process is by reconstruction of a non-prompt
 250 b -jet.

251 **2.2.2 Comparison to Trilepton Channel**

252 The most apparent difference between the tri and tetralepton channels is the number of events present,
 253 with the tetralepton channel having far less events in its phase space than that of the tri-lepton channel.
 254 The lack of statistics in the tetralepton channel can be attributed to its low cross section times branching
 255 ratio of $\sigma_{(tW^\pm Z).Br(4\ell)}^{\text{NLO}} = 0.7 \text{ fb}$ [24]. The tri-lepton channel has a cross section times branching ratio of
 256 $\sigma_{(tW^\pm Z).Br(3\ell)}^{\text{NLO}} = 3.9 \text{ fb}$ [24], which is around a factor of four larger than that of the tetralepton channel.
 257 This difference between the production cross section of the two decay channels can be largely attributed
 258 to the difference in branching ratios ($\frac{\Gamma_i}{\Gamma}$) between a hadronically decaying W boson, $\frac{\Gamma_{W \rightarrow \text{had}}}{\Gamma_W} = (67.41 \pm$
 259 $0.27)\%$ [52], present in the tri-lepton channel and a leptonically decaying W boson, $\frac{\Gamma_{W \rightarrow \ell\nu}}{\Gamma_W} = (10.86 \pm$
 260 $0.09)\%$ [52], present in the tetralepton channel. Despite the tetralepton channel's low statistics, it is not
 261 subject to the large WZ background present in the trilepton channel [95]. The tetralepton channel has

¹In this thesis, ℓ refers to an electron or muon, ν refers to a neutrino or anti-neutrino and b refers to a bottom quark or anti-bottom quark

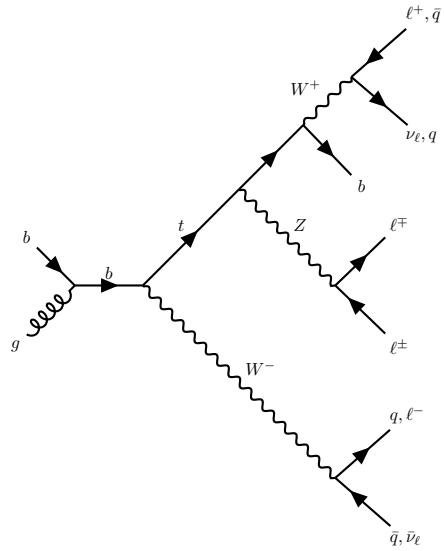


Figure 4: The LO Feynman diagram of tWZ production in the tri-lepton channel.

262 a substantial amount of ZZ background (not present in the trilepton channel), fortunately this can be
263 easily suppressed due to the full reconstructability of the two leptonically decaying Z -bosons.

264

Chapter 3

265 The ATLAS Experiment and Detector

266 3.1 The ATLAS Experiment

267 The ATLAS (A Toroidal LHC ApparatuS) detector is one of two general purpose detectors at CERN
 268 (the European Organization for Nuclear Research) near Geneva in Switzerland. These detectors collect
 269 data from the collisions provided by the worlds highest energy particle accelerator [28], the Large Hadron
 270 Collider (LHC) situated at CERN.

271

272 In this section, information about the LHC and the ATLAS detector are given. This includes technical
 273 aspects of the ATLAS detector and the processing of data into meaningful physics objects¹ to be used
 274 in analyses. The following chapter consists of information from "The LHC Design Report" [28], "LHC
 275 Machine" [46] and "The ATLAS Experiment at the CERN Large Hadron Collider" [40] unless otherwise
 276 stated.

277 3.1.1 Large Hadron Collider (LHC)

278 The LHC is a circular 27 km particle accelerator located in an underground tunnel on the border between
 279 France and Switzerland. The accelerator consists of supercooled, superconducting magnets which
 280 accelerate and collide beams of protons at centre-of-mass energies up to $\sqrt{s} = 13$ TeV at instantaneous
 281 luminosities of $\mathcal{L} \sim 10^{34}$ cm $^{-2}$ s $^{-1}$. In the LHC, pp beams consist of bunches of protons which collide
 282 every 25 ns, corresponding to a frequency of 40 MHz. Several accelerator systems are used to accelerate
 283 protons and heavy ions to such high energies. Protons are extracted from a tank of ionised hydrogen gas
 284 and are injected into the Linear Accelerator 2 (LINAC), where they are linearly accelerated to momenta
 285 of 50 MeV. The proton bunches are then sequentially accelerated by a chain of circular accelerators. The
 286 chain starts with the Booster which accelerates the protons to momenta of up to 1.4 GeV. The proton
 287 bunches are then fed through to the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS)
 288 which accelerate the protons to momenta of up to 25 GeV and 450 GeV respectively. The protons are
 289 then transferred to two beam pipes of the LHC where they travel in opposite directions. Both proton
 290 beams are accelerated to their final momenta of 6.5 TeV, resulting in a centre-of-mass energy of 13 TeV.
 291 These proton beams then collide at one of the four main interaction points (positions along the beam

¹a detector signal (or a combination of detector signals) that represent a candidate for a particle (e.g leptons and jets)

292 pipe where collisions occur) situated along the LHC.

293

294 The four main experiments located at the interaction points are ATLAS, the Compact Muon Solenoid
 295 (CMS), Large Hadron Collider Beauty (LHCb) Experiment and A Large Ion Collider Experiment (AL-
 296 ICE). The ATLAS and CMS detectors are general-purpose detectors which investigate a wide range of
 297 physics processes. Since both ATLAS and CMS can measure the same processes, they are able to cross-
 298 check and validate measurements taken by one another. The LHCb detector is specifically designed to
 299 study decays of particles containing b -quarks. The ALICE detector is designed to study the strongly
 300 interacting quark-gluon plasma which is formed at extremely high energy densities. At the interaction
 301 points, the two proton beams which consist of protons in closely packed bunches, travel in opposite direc-
 302 tions to one another and collide. Many hard pp collisions (events) can occur per bunch crossing, however
 303 it is the most energetic collision in the bunch crossing that is interesting for discovery potential. The
 304 most energetic collision is therefore chosen to be studied and any additional collisions are aimed to be
 305 rejected. These additional collisions are referred to as *pile-up*. Pileup complicates the reconstruction of
 306 the particles originating from the hard collision of interest.

307 3.1.1.1 Luminosity

308 This section consists of information from "Modern Particle Physics" [88], unless otherwise stated. The
 309 event production rate at the LHC, $R(t)$, for a certain process of interest is given by,

$$R(t) = \mathcal{L}(t)\sigma \quad (3.1)$$

310 where $\mathcal{L}(t)$ is the instantaneous luminosity and σ and is the cross section of the process of interest. The
 311 instantaneous luminosity, $\mathcal{L}(t)$, is independent on the process of interest, and depends on various collider
 312 and beam parameters. $\mathcal{L}(t)$ can be written in terms of these parameters as,

$$\mathcal{L}(t) = f \frac{N n_1 n_2}{4\pi \sigma_x \sigma_y} \quad (3.2)$$

313 where f is the beam revolution frequency, N is the number of proton bunches colliding per second, n_1
 314 and n_2 are the number of protons in the colliding bunches, σ_x and σ_y are the beam spread in the x and y
 315 directions respectively. The average number of visible (particles that are measurable by the detector) pp
 316 interactions per bunch crossing, $\langle \mu \rangle$, can be written as,

$$\langle \mu \rangle = \frac{\sigma_{inel} \mathcal{L}}{N f} \quad (3.3)$$

317 where σ_{inel} is the inelastic pp interaction cross section. The total integrated luminosity, L , across some
 318 time interval, is given by,

$$L = \int \mathcal{L} dt. \quad (3.4)$$

319 The units of L are inverse area, and are given by fb^{-1} at the LHC and the ATLAS detector. In Figure 5,
 320 the total integrated luminosity vs time, recorded by ATLAS for $\sqrt{s} = 13$ TeV pp collisions at the LHC is
 321 shown [90].

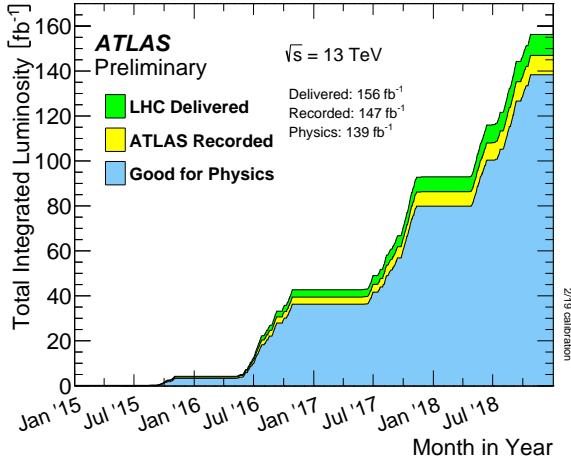


Figure 5: The total integrated luminosity vs time, recorded by ATLAS for $\sqrt{s} = 13$ TeV pp collisions at the LHC is shown [90]. The total integrated luminosity delivered by the LHC, recorded by ATLAS and certified to be good quality data are shown by the green, yellow and blue histograms respectively. The month and year of data taking is shown on the x-axis and the total integrated luminosity (in fb^{-1}) is shown on the y-axis.

322 A total integrated luminosity of 139 fb^{-1} of data certified as good for physics was recorded by ATLAS
 323 between 2015 and 2018. This data taking period is referred to as Run 2. It occurred after the Run 1 data
 324 taking period (2011 and 2012) and the Long Shutdown 1 LHC upgrade period (2013 and 2014). In this
 325 analysis, we use the Full Run 2 dataset.

326 3.2 The ATLAS Detector

327 In Figure 6, the schematic of the ATLAS detector, is shown.

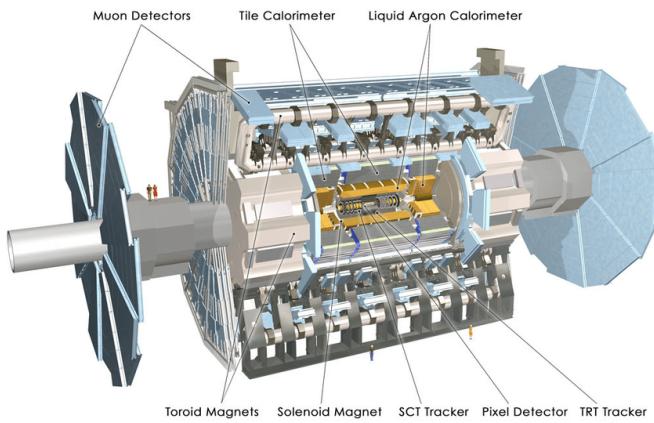


Figure 6: A schematic of the ATLAS detector is shown. [40]

328 The detector is cylindrically shaped which covers close to 4π in solid angle. It has a length of 44 m,
 329 a diameter of 25 m and a mass of 7000 tons. The ATLAS detector consists of four main sub-detectors
 330 arranged in concentric cylindrical layers around the beam pipe. These include the inner detector, the
 331 electromagnetic calorimeter, the hadronic calorimeters and the muon spectrometer. The sub-detectors

332 record the momenta, energies and trajectories of different particles produced in the collider, allowing for
 333 the reconstruction and identification of these particles to be used in physics analyses.

334 3.2.1 Coordinate System and Kinematics

335 The ATLAS detector adopts a right-handed coordinate system. The origin is at the nominal interaction
 336 point with the z -axis defined to be counter-clockwise along the beam line. The $x - y$ plane (or transverse
 337 plane) is perpendicular to the beam line, with the x -axis pointing towards the centre of the LHC ring and
 338 the y -axis pointing upwards towards the Earth's surface. The azimuthal angle, $\phi \in [-\pi, \pi]$, is measured
 339 in the transverse plane with respect to the positive x -axis. The polar angle, $\theta \in [0, \pi]$, is measured in
 340 the $z - y$ plane with respect to the positive y -axis. A quantity called the pseudorapidity, $\eta \in [0, \infty]$ is
 341 defined as,

$$\eta = -\ln \tan\left(\frac{\theta}{2}\right) \quad (3.5)$$

342 The polar angle, η , is often used instead of θ , since the difference in η between two particles, $\Delta\eta$, is
 343 invariant under a Lorentz boost in the z -direction [91]. The angular distance between two physics
 344 objects, ΔR , can be written as,

$$\Delta R = \sqrt{(\Delta\phi)^2 + (\Delta\eta)^2} \quad (3.6)$$

345 where $\Delta\phi$ is the difference in ϕ between the two physics objects of interest. Quantities defined in the
 346 transverse plane are often used to describe the kinematics of physics objects in hadron collider experiments.
 347 The transverse momentum, p_T , is defined as,

$$p_T = \sqrt{(p_x)^2 + (p_y)^2} \quad (3.7)$$

348 where p_x and p_y are the x and y components of the physics object's momenta, respectively. The transverse
 349 energy, E_T , is defined as,

$$E_T = \sqrt{m^2 + p_T^2} \quad (3.8)$$

350 where m is the invariant mass of the physics object.

351 3.2.2 Inner Detector

352 The inner detector is the first layer of concentric cylindrical sub-detector layers in the ATLAS detector.
 353 It is used to identify charged particles and reconstruct the trajectories of charged particles produced in
 354 the collisions via energy deposition in semiconductor material (hits) and the ionisation of gas. It consists
 355 of three complementary sub-detectors (in order from nearest to farthest from the beam pipe): the Pixel
 356 Detector, the Semiconductor Tracker (SCT) and the Transition Radiation Detector (TRT). The Pixel
 357 Detector and SCT are based on semiconductor technology and have the highest granularity of any sub-
 358 detector in ATLAS, in order to cope with the high frequency of collisions near the interaction point. The
 359 TRT consists of drift tubes (straws) containing a gas mixture, which allows measurement of the energy
 360 deposited by charged particles through the ionisation of the gas. Solenoid magnets surround the inner

361 detector and bend the trajectories of charged particles. The charges and transverse momenta of particles
362 can be inferred from their bent trajectories, which are reconstructed by the hits produced via energy
363 deposition in the Inner Detector.

364 **3.2.3 Electromagnetic and Hadronic Calorimeters**

365 The Electromagnetic Calorimeter (ECAL) and Hadronic Calorimeter (HCAL) surround the Inner Detec-
366 tor, with the ECAL nearer to the beam line. The ECAL and HCAL provide accurate measurements of
367 the energy of particles which interact electromagnetically (e.g. photons and electrons) and hadronically
368 (e.g. jets), respectively. Particles entering the calorimeters interact with the detector material and create
369 either a electromagnetic shower (in the ECAL) or a hadronic shower (in the HCAL), depositing all their
370 energy in the calorimeter cells. The primary mechanism of energy deposition in the ECAL is through
371 bremsstrahlung (for electrons) and pair production (photons). Hadrons usually deposit a small amount
372 of their energy in the ECAL, and interact via inelastic scattering with the nuclei of the detector material.
373 The hadronic showers (jets) produced in these nuclear interactions travel much further than an electro-
374 magnetic shower, and for that reason, the volume of the HCAL is designed to occupy a much larger space
375 than that of the ECAL.

376 **3.2.4 Muon Spectrometer**

377 The Muon Spectrometer (MS) is the outermost sub-detector of ATLAS and surrounds the HCAL. Muons
378 traverse through the inner detector and calorimeters, with minimal energy loss, before reaching the MS.
379 The MS consists of trigger and high-precision tracking systems. Large superconducting toroid shaped
380 magnets deflect the incoming muons to measure their trajectories and subsequently their momenta via
381 the curvature of the trajectories. The MS measures muon trajectories as they ionize gas (filled with Ar
382 and CO₂ gas) in the MS drift chambers.

383 **3.2.5 Trigger and Data Acquisition System**

384 The Trigger and Data Acquisition System (TDAQ) manages and handles the large amount of data pro-
385 duced within the ATLAS detector. In Run 2, pp bunch crossings occur every 25 ns, corresponding to
386 an event rate of 40 MHz. The TDAQ system performs a fast preliminary reconstruction to select events
387 with signatures which are interesting for physics analyses. The information collected from these events
388 are permanently stored for offline reconstruction and analysis, and the rest (the vast majority of events)
389 are discarded. The trigger system reduces the 40 MHz data rate to around 1 kHz.

390 **3.2.6 Particle Identification and Object Reconstruction**

391 Particles originating from pp collisions, or from their subsequent decays, traverse through the ATLAS
392 detector and interact with its different sub-detectors, producing characteristic electronic signals. These
393 signals are then processed by various algorithms to reconstruct and identify the physics objects (e.g.
394 electrons, muons, jets) in the event. This section outlines the procedures used to define these physics
395 objects.

396 3.2.6.1 Tracks and primary vertices

397 The trajectories of charged particles, or tracks, are reconstructed in the ID. First, energy is deposited by
 398 charged particles (hits) in pixels or strips, in the Pixel and SCT detectors respectively. Adjacent pixels
 399 or strips are grouped together in *energy clusters*. Energy clusters define 3D space-points indicating the
 400 location where the charged particle traversed. Track seeds are then defined as sets of three space-points,
 401 in either the Pixel or SCT detectors. A Kalman filter [11] is then used to build track candidates from the
 402 track seeds. Often, multiple track candidates are built per track seed, therefore an ambiguity solver [37]
 403 is needed for finding the track which best represents the traversal of the charged particle. The ambiguity
 404 solver ranks each track from a given seed based on, the number of associated hits, the number of holes
 405 (expected hits which are absent), track momenta and the χ^2 of the track fit. Low ranked tracks are then
 406 discarded. High ranked tracks are refitted, introducing information from the TRT.

407
 408 The primary vertex is the location of the pp collision of interest (i.e. from the hard scatter). The primary
 409 vertex from the hard scatter needs to be identified, to isolate the event of interest from unwanted pile-
 410 up events. In the event reconstruction procedure [68], the primary vertex is defined as the vertex of
 411 the event with the largest sum of $(p_T)^2$ (corresponding to the measured $(p_T)^2$ of the particle from its
 412 reconstructed track) of its associated tracks. Furthermore, the primary vertex is required to have at least
 413 two associated tracks. To reduce contamination from fake tracks used in primary vertex reconstruction,
 414 only tracks which pass certain tight selection criteria are used in the reconstruction procedure. An iterative
 415 fitting procedure is then used to reconstruct the primary vertex by finding a set of reconstructed tracks
 416 which have a common vertex.

417 3.2.6.2 Electrons

418 Since electrons are charged particles, they give rise to tracks in the Inner Detector. They also deposit
 419 energy in the ECAL via electromagnetic showering. Electrons are therefore reconstructed and identified
 420 from signals in the Inner Detector and ECAL. Electrons are reconstructed using a dynamic clustering
 421 algorithm [1] which matches electron candidate tracks in the Inner Detector to energy clusters in the
 422 ECAL. The dynamic clustering algorithm matches tracks to energy clusters which have local maxima, to
 423 form electron candidates. A likelihood discriminant is used to identify electrons. Quantities measured
 424 in the Inner Detector and ECAL are used as input, such that they discriminate well between prompt
 425 isolated electrons and other physics objects (e.g. jets, electron from a photon conversion, electron from a
 426 semi-leptonically decaying hadron). Important input variables include the shape of the electromagnetic
 427 shower, track quality in the Inner Detector and information from the TRT.

428 3.2.6.3 Muons

429 Muons leave tracks in the Inner Detector and the MS. They traverse the ECAL and HCAL with no
 430 significant energy loss. Muons are therefore reconstructed and identified from information in the Inner
 431 Detector and MS. Tracks are reconstructed [5] in the Inner Detector and MS independently. Both tracks
 432 are combined, using a global χ^2 fit, resulting in reconstructed muon candidates. Similar to electron
 433 identification, muons use a likelihood discriminant to identify prompt muons and suppress background

⁴³⁴ contamination (mainly from pion and kaon decays).

⁴³⁵ 3.2.6.4 Jets and b -tagging

⁴³⁶ Coloured particles emerging from the interaction point result in collimated streams of colourless particles,
⁴³⁷ known as jets. Jets can deposit energy in the Inner Detector and in the HCAL. Jets in ATLAS are
⁴³⁸ reconstructed from topological clusters using the anti- k_t algorithm [30]. Topological clusters are groups
⁴³⁹ of adjacent calorimeter cells which contain energy deposition above the average amount of noise expected
⁴⁴⁰ in the cell. Adjacent cells are grouped together under certain criteria to form topological clusters which
⁴⁴¹ form jets.

⁴⁴²

⁴⁴³ Different tagging algorithms are used to identify the quark flavour which initiated a jet. Tagging of b -
⁴⁴⁴ quarks is used extensively in top physics, due to the b -quark present in the top quark's dominant decay
⁴⁴⁵ channel (See Table 1). Hadrons arising from b -quark hadronisation have mean lifetimes ~ 1.5 ps and
⁴⁴⁶ travel (on average) a few millimetres before decaying. This creates a secondary vertex within the jet
⁴⁴⁷ (See Figure 7). This characteristic decay signature, along with several other unique features of b -jets,
⁴⁴⁸ are exploited in b -tagging algorithms to distinguish b -jets from c - or light flavour jets. In Figure 7, an
⁴⁴⁹ illustration of the production of a b -jet, is shown.

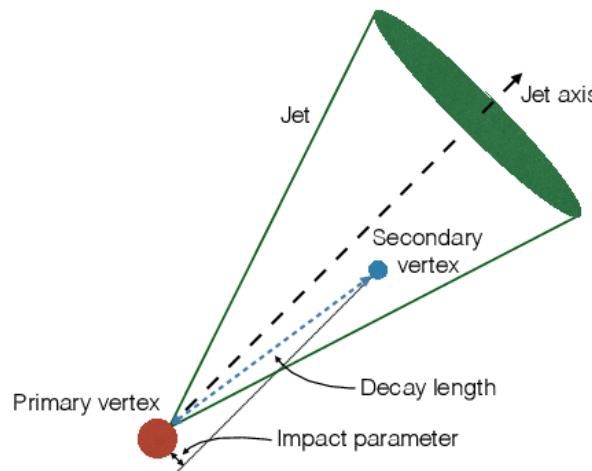


Figure 7: An illustration [Connelly2017PerformanceAC] of the production of a b -jet is shown. The jet is represented by the cone and axis of the jet is represented by the dotted black line drawn at a right angle to the circular base of the cone. The primary vertex is represented by the red circle and the secondary vertex is represented by the blue circle. The dotted blue line represents the path travelled by the hadron before decaying at its secondary vertex. The impact parameter is shown by a black line drawn from the secondary vertex.

⁴⁵⁰ This illustrates the presence of a secondary vertex within a b -jet. In this analysis, we use the recommended
⁴⁵¹ DL1r (Deep-Learning Flavour Tagger) tagging algorithm [17]. The DL1r algorithm combines outputs from
⁴⁵² several low-level tagging algorithms using a Deep Neural Network and outputs the probability that a given
⁴⁵³ input jet is identified as a b , c or light flavoured jet.

454

Chapter 4

The search for tWZ production using events with four leptons

457 In this chapter, the full analysis procedure is presented for tWZ production in the tetralepton channel
458 based on an integrated luminosity of 139 fb^{-1} of data recorded by ATLAS. This includes a description
459 of the data and simulation, definitions of physics objects, event selection and the subsequent definition of
460 signal and control regions used in this analysis. Furthermore, estimation of the fake lepton component
461 using the MC template method is described. A kinematic reconstruction algorithm and various Machine
462 Learning techniques, used to discriminate between signal and background events, are described. An
463 outline of the systematic uncertainties affecting the measurement are presented. Finally, the results of
464 the analysis and a study of the analysis projected to higher luminosities is presented.

465 4.1 Data and Monte Carlo Simulation

466 4.1.1 Data Samples

467 The pp collision data used in this analysis was collected with the ATLAS detector at the LHC from 2015
468 to 2018. This data period of data taking is referred to as Run 2. During this period, pp collisions at
469 $\sqrt{s} = 13 \text{ TeV}$, corresponding to an integrated luminosity of 156 fb^{-1} , were delivered by the LHC. The
470 ATLAS detector recorded 147 fb^{-1} of this total delivered data. 139 fb^{-1} of the data recorded by ATLAS
471 is considered to be good enough for physics analyses (the data passes certain quality control criteria)
472 and placed into the *Good Runs List* [51]. Specific time-intervals where the luminosity is assumed to be
473 constant are known as *lumi-blocks*. Only events from lumi-blocks in which LHC beams were stable and all
474 ATLAS detectors were operational are selected. The list of suitable run and lumi-blocks is summarised
475 in the official Good Runs Lists for 2015-2018 data, as specified in Ref. [51]. The integrated luminosities
476 corresponding to the individual datasets for years 2015, 2016, 2017 and 2018 are 3.2 fb^{-1} , 33.0 fb^{-1} , 44.3
477 fb^{-1} and 58.5 fb^{-1} respectively.

478 **4.1.2 Monte Carlo Samples**

479 Simulated Monte Carlo (MC) samples were generated and used to model the SM tWZ signal and its
480 backgrounds. The following background processes are considered:

481 • **$t\bar{t}Z$:** $t\bar{t}$ with an associated Z -boson, in the tetralepton final state. Therefore, both top quarks
482 decay leptonically (e.g. $t \rightarrow W^+ b \rightarrow \ell^+ \nu b$) and of these top quarks emits a Z -boson which decays
483 leptonically ($Z \rightarrow \ell^\pm \ell^\mp$ (OSSF lepton pair)). This results in a final state with 4 leptons and 2
484 b-quarks.

485 • **ZZ :** Diboson production with a tetralepton final state, therefore both Z -bosons decay leptonically
486 ($Z \rightarrow \ell^\pm \ell^\mp$ (OSSF lepton pair)).

487 • **other:** Processes with a relatively minimal, but non-negligible background contribution

488 - $VVV(V = W/Z)$

489 - $t\bar{t}$

490 - $t\bar{t}W$

491 - $t\bar{t}WW$

492 - $t\bar{t}H$

493 - WZ

494 - $t\bar{t}t$

495 - $t\bar{t}t\bar{t}$

496 - tZq

497 The MC simulations are achieved via the use of event generators and parton shower generators. The
498 parton shower generators simulate any incoming or outgoing particles from the hard process, which carry
499 QCD color charge and can therefore lead to parton showers. The production of tWZ events is simulated
500 with the **MADGRAPH5_AMC@NLO 2.3.3** [9] generator providing matrix element (ME) calculations at NLO
501 (next-to-leading order) in QCD. The events are interfaced with **PYTHIA 8.235** for the parton shower. The
502 production of $t\bar{t}Z$ and $t\bar{t}W$ events are simulated with the **MADGRAPH5_AMC@NLO 2.3.3** generator providing
503 ME calculations at NLO in QCD. The events are interfaced with **PYTHIA 8.210** [85] for the parton
504 shower. Event generation of tWZ and $t\bar{t}Z$ results in diagrams which overlap with one another, that is,
505 these diagrams contain the same initial and final state particles. Several methods exist in order to separate
506 between the two processes, by removing the overlap, therefore avoiding double counting. There are two
507 different diagram removal procedures, diagram removal procedure 1 (DR1) [42] and diagram removal
508 procedure 2 (DR2). The DR1 scheme is used to remove the overlap between tWZ and $t\bar{t}Z$. A comparison
509 to the DR2 scheme is used to estimate part of the theoretical systematic on the modelling of the tWZ
510 signal (See Section 4.8.2). Diboson processes which feature the three charged leptons and one neutrino
511 or four charged lepton in their final states, such as WZ and ZZ , are simulated using **SHERPA 2.2.2** [25]
512 at NLO in QCD precision. The events are interfaced with **SHERPA** for the parton shower. Triboson
513 processes such as WWW , WWZ , WZZ , and ZZZ containing up to six leptons in their final states are

514 simulated using **SHERPA 2.2.2** at NLO in QCD precision. The events are interfaced with **SHERPA** for the
 515 parton shower. The production of $t\bar{t}$ events are simulated with the **POWHEG** [49] generator providing ME
 516 calculations at NLO in QCD. The events are interfaced with **PYTHIA 8.210** for the parton shower. The
 517 production of $t\bar{t}t\bar{t}$, $t\bar{t}WW$ and $t\bar{t}H$ are simulated using the **MADGRAPH5_AMC@NLO 2.2.2** generator at LO in
 518 QCD precision. The events are interfaced with **PYTHIA 8.186** [84] for the parton shower. The production
 519 of $t\bar{t}$ with an associated Higgs boson, $t\bar{t}H$, are generated using the **MADGRAPH5_AMC@NLO 2.6.0** generator
 520 at NLO in QCD precision. The events are showered using **PYTHIA 8.230** [85]. The production of $t\bar{t}$
 521 events are simulated with the **POWHEG** generator providing ME calculations at NLO in QCD. The events
 522 are showered using **PYTHIA 8.230**. The production of a single top quark in association with a Z -boson
 523 and an extra parton, tZq , is simulated using **MADGRAPH5_AMC@NLO 2.3.3** at NLO in QCD precision. The
 524 events are interfaced with **PYTHIA 8.230** for the parton shower. In Table 3, the event generator and
 525 parton shower used for each process's sample are shown.

Process	Event Generator	Cross section calculation	Parton Shower
tWZ	MADGRAPH5_AMC@NLO 2.3.3	NLO	PYTHIA 8.235
$t\bar{t}Z$	MADGRAPH5_AMC@NLO 2.3.3	NLO	PYTHIA 8.210
ZZ, WZ	SHERPA 2.2.2	NLO	SHERPA
$VVV(V = W/Z)$	SHERPA 2.2.2	NLO	SHERPA
$t\bar{t}$	POWHEG	NLO	PYTHIA 8.230
$t\bar{t}W$	MADGRAPH5_AMC@NLO 2.3.3	NLO	PYTHIA 8.210
$t\bar{t}WW$	MADGRAPH5_AMC@NLO 2.2.2	LO	PYTHIA 8.186
$t\bar{t}H$	MADGRAPH5_AMC@NLO 2.6.0	NLO	PYTHIA 8.230
$t\bar{t}t, t\bar{t}t\bar{t}$	MADGRAPH5_AMC@NLO 2.2.2	LO	PYTHIA 8.186
tZq	MADGRAPH5_AMC@NLO 2.3.3	NLO	PYTHIA 8.230

Table 3: The event generator and parton shower used for the signal and background process's MC samples is shown.

526 4.1.3 Trigger Strategy

527 Events in data and simulation are selected via either single lepton or dilepton (electron or muon) triggers.
 528 To selected events, a logical **OR** between the triggers is applied. This means that events have to be
 529 selected by at least one of the single or dilepton triggers. Electrons and muons may be surrounded by
 530 other particles which could lead to problems in the reconstruction and identification process. They are
 531 therefore required to be sufficiently isolated from other particles. This is done by defining an isolation
 532 variable which ensures that the summed energies of the particles in a cone defined around the muon, with
 533 radius ΔR , are small. The isolation variable for a cone of radius x , $I_{\Delta R=x}$, is given by,

$$I_{\Delta R=x} = \frac{\sum_i E_i}{p_T^{e/m}} \quad (4.1)$$

534 where E_i is the energy of the i^{th} particle within the cone and $p_T^{e/m}$ is the p_T of the electron or muon.
 535 In order to identify an electron or muon, various properties and detector signals related to them are fed
 536 into different Machine Learning classifiers to determine the likelihood of the particle being an electron
 537 or a muon. Cuts are applied to the output of the classifiers to define working points which correspond

to different selection efficiencies. The names commonly given to different working points, in order of decreasing efficiency, are very loose, loose, medium and tight. The single-muon trigger used to select muons in 2015 data and simulation requires a muon with $p_T > 20$ GeV, an identification criteria corresponding to a loose working point and $I_{\Delta R=0.2} < 0.12$. Two single-muon triggers are used to select muons in 2016-2018 data and simulation. The first requires a muon with $p_T > 26$ GeV, an identification criteria corresponding to a medium working point and $I_{\Delta R=0.3} < 0.06$. The second requires a muon with $p_T > 50$ GeV. Two single-electron triggers are used to select electrons in 2015 data and simulation. The first requires an electron with $p_T > 24$ GeV and an identification criteria corresponding to a medium working point. The second requires an electron with $p_T > 120$ GeV and an identification criteria corresponding to a loose working point. Three single leptons triggers are used to select electrons in 2016-2018 data and simulation. The first requires an electron with $p_T > 26$ GeV, an identification criteria corresponding to a tight working point and $I_{\Delta R=0.2} < 0.1$. The second requires an electron with $p_T > 60$ GeV and an identification criteria corresponding to a medium working point. The third requires an electron with $p_T > 140$ GeV and an identification criteria corresponding to a loose working point. The di-muon trigger used to select muons in 2015 data and simulation requires that one muon have $p_T > 18$ GeV and another muon to have $p_T > 8$ GeV. The di-muon trigger used to select muons in 2016-2018 data and simulation requires that one muon have $p_T > 22$ GeV and another muon to have $p_T > 8$ GeV. The di-electron trigger used to select electrons in 2015 data and simulation requires two electrons, each with $p_T > 12$ GeV, and identification criteria corresponding to loose working points. The di-electron trigger used to select electrons in 2016-2018 data and simulation requires two electrons, each with $p_T > 17$ GeV, and identification criteria corresponding to very loose working points. A final electron-muon trigger is used to select electrons and muons in 2016-2018 data and simulation that requires an electron with $p_T > 17$ GeV and a muon with $p_T > 14$ GeV, and identification criteria corresponding to loose working points. In Table 4, a summary of the triggers used to select muons and electrons in data and simulation are shown.

Trigger	Selection [GeV]	Working Point
Single-muon*	$p_T > 20$	loose
Single-muon	$p_T > 26$	medium
Single-muon	$p_T > 50$	-
Single-electron*	$p_T > 24$	medium
Single-electron*	$p_T > 120$	loose
Single-electron	$p_T > 26$	loose
Single-electron	$p_T > 60$	medium
Single-electron	$p_T > 140$	loose
Di-muon*	$p_T > 18, p_T > 8$	-
Di-muon	$p_T > 18, p_T > 8$	-
Di-electron*	$p_T > 12$	loose
Di-electron*	$p_T > 12$	loose
Di-electron	$p_T > 17$	very loose
Electron-Muon	electron $p_T > 17$, muon $p_T > 14$	loose

Table 4: Single lepton and dilepton (electron and muon) triggers used to select events in data and simulation. Triggers labelled with * are only used for 2015 data and simulation.

562 4.2 Physics objects

563 In this section the physics objects (leptons, jets and b -tagged jets) used in this analysis are outlined.

564 4.2.1 Leptons

565 In this analysis only e and μ leptons are considered, since τ leptons are difficult to detect in the AT-
 566 LAS detector. They are challenging to detect since they have an extremely short lifetime (290.3 ± 0.5
 567 fs [52]) which causes them to decay before reaching any detector components and therefore can only be
 568 reconstructed via their decay products. In addition to our selection criteria of exactly four leptons, it is
 569 required that the Leading (L), Next-to-Leading (NL), Next-to-Next-to-Leading (NNL) and Next-to-Next-
 570 to-Next-to-Leading (NNNL) leptons have p_T greater than 28, 18, 10 and 10 GeV respectively. Relativity
 571 loose object-level cuts are chosen in an attempt to maximize our signal statistics, since the analysis is
 572 heavily statistically limited. Reconstructed electrons are required to be within $|\eta| < 2.47$ and excluding
 573 the transition region between the barrel and end-cap calorimeters at $1.37 < |\eta| < 1.52$. Reconstructed
 574 muons are required to be within $|\eta| < 2.5$. The transverse impact parameter, d_0 , is defined as the minimal
 575 spacial distance between the object's (referring to leptons) trajectory and the primary vertex (the vertex
 576 associated with the pp hard scatter). The longitudinal impact parameter, z_0 , is defined as the value of
 577 z of the point on the object's trajectory which determines d_0 . To ensure consistency between the lepton
 578 and the primary vertex, it is required that $|\frac{d_0}{\sigma(d_0)}| < 5$, $|z_0 \sin \theta| < 0.5$ mm for electrons and $|\frac{d_0}{\sigma(d_0)}| < 3$,
 579 $|z_0 \sin \theta| < 0.5$ mm for muons, following the current recommendations [89]. To avoid instances where one
 580 detector signal can result in multiple different reconstructed objects, an overlap removal procedure is ap-
 581 plied which ignores all but one of these objects (See Section 4.2.4). Electrons are selected using a likelihood
 582 based discriminant [1] which takes measurements from the tracking system, calorimeter system and quan-
 583 tities derived from both the tracking and calorimeter system as input. Muons are selected using the Muon
 584 Selection Tool [73]. Loose electrons are defined with the criteria above, using the `LooseAndBLayerLH`
 585 ($\sim 91\%$ selection efficiency for electrons with $E_T > 30$ GeV [43]) identification algorithm (which has a
 586 certain cut applied). Similarly, tight electrons are defined with the criteria above, using the `TightLH` (\sim
 587 80% selection efficiency for electrons with $E_T > 30$ GeV [43]) algorithm (which has a certain cut applied).
 588 Both loose and tight muons use the `Medium` ($\sim 95\%$ selection efficiency [5]) algorithm (which has a certain
 589 cut applied). Tight leptons additionally require that they are sufficiently isolated from other particles
 590 produced in the collision. This is done by defining a cone of radius $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$ around the parti-
 591 cle of interest and summing the p_T of all the reconstructed particles surrounding the particle of interest,
 592 situated within the cone. A quantity, I_{rel} , is then defined as, $I_{rel} = \frac{\sum p_T(\text{surrounding candidate})}{p_T(\text{candidate})}$, the ratio of
 593 this sum to the p_T of the lepton candidate. If this value is large, is it likely that the particle of interest
 594 originated from a jet (together with many other particles), whereas a prompt decay product resulting from
 595 the hard scatter will have little to no energy surrounding it ($I_{rel} \ll 1$). The `IsolationSelectionTool`
 596 with the `PLVTight` ($\sim 70\%$ efficiency [6]) and `PLVTight` ($\sim 70\%$ efficiency at $p_T = 30$ GeV [36]) algorithm
 597 are used for tight electrons and tight muons respectively (following the current recommendations [80]).
 598 In Table 5, a summary of the selection criteria for leptons is shown.

	Electrons		Muons	
	Tight	Loose	Tight	Loose
p_T cuts	$p_T(\ell_1, \ell_2, \ell_3, \ell_4) > (28, 18, 10, 10)$ GeV			
Overlap Removal	Described in Section 4.2.4			
η cuts	$ \eta(\ell_e) < 2.47$ excluding $1.37 < \eta(\ell_e) < 1.52$		$ \eta(\ell_\mu) < 2.5$	
Impact Parameters	$ \frac{d_0}{\sigma(d_0)} < 5$, $ z_0 \sin \theta < 0.5$ mm		$ \frac{d_0}{\sigma(d_0)} < 3$, $ z_0 \sin \theta < 0.5$ mm	
Identification WP	TightLH	LooseAndBLayerLH	Medium	Medium
Isolation WP	PLVTight	Not Used	PLVTight	Not Used

Table 5: A summary of the requirements applied for selecting tight and loose leptons (e, μ) is shown.

599 4.2.2 Jets

600 Jets are reconstructed using the anti- k_t algorithm (See Section 3.2.6.4). The `AntiKt4EMPFflowjets` (\sim
 601 97% average efficiency with JVT (outlined in the subsequent paragraph) > 0.2 [89]) algorithm (which
 602 has a certain cut applied) is used, following the current recommendations [89]. The jet-vertex-tagger
 603 (JVT) and the forward jet-vertex-tagger (fJVT) [48] are likelihood discriminant which aim to suppress
 604 pile-up jets. The Medium algorithm (which has a certain cut applied) is used for the JVT and the fJVT
 605 (following the current recommendations [78]). Additionally, a requirement that jets have a JVT value
 606 greater than 0.5 is applied. In the same way as with leptons, ambiguities are removed where one detector
 607 signal can result in multiple different reconstructed objects, via overlap removal (See Section 4.2.4). Jets
 608 are required to be within $p_T(\text{jet}) > 20$ GeV. These relatively loose p_T cuts are applied in an attempt to
 609 increase our limited signal statistics. A forward jet is a signature of single top quark production, jets are
 610 therefore required to have $|\eta| < 4.5$ in order to include these forward jets.

611 4.2.3 b -tagging

612 The DL1r b -tagger [71] was used to identify jets as b -jets (See Section 3.2.6.4). Different DL1r working
 613 points are used to identify b -jets in our event selection (See Section 4.4). The working points are defined
 614 as different cuts on the DL1r score corresponding to a b -jet tagging efficiency of 60%, 70%, 77% and
 615 85%. The efficiency of the DL1r b -tagger is measured using control samples in data and in simulation.
 616 From these measurements, correction factors are derived to correct the tagging rates in the simulation.
 617 In the case of b -tagged jets, the correction factors and their uncertainties are estimated from data using
 618 dileptonic $t\bar{t}$ events [15, 12]. Sources of uncertainty affecting the b -tagging efficiencies are evaluated as
 619 a function of jet p_T , including bin-to-bin correlations. Since this analysis is heavily statistically limited,
 620 the amount of statistics in our regions are aimed to be maximized. In an attempt to achieve this goal in
 621 the $t\bar{t}Z$ CR, b -tagged jets were placed under *tight* and *loose* definitions. A tight b -tagged jet is defined
 622 as a jet which passes the 77%, 70%, 65% or 60% DL1r b -tagger working point. A loose b -tagged jet is
 623 defined as a jet which passes 85% DL1r b -tagger working point, but not the 77%, 70%, 65% or 60% DL1r
 624 b -tagger working points. Different numbers and definitions of tight and loose b -tagged jets were tried in
 625 each region, with the final selection criteria being chosen which maximised the expected significance of
 626 $\sigma(tWZ)$ (See Section 4.4.1).

4.2.4 Overlap Removal Procedure

An overlap removal procedure is performed to avoid instances where one detector signal can result in multiple different reconstructed objects. The overlap removal procedure is used on pre-selected leptons and jets. It is performed sequentially, in the following steps:

1. If the separation between a pre-selected electron and pre-selected muon is within $\Delta R < 0.01$, or they share a track, the pre-selected electron is discarded.
2. If the separation between a jet and a pre-selected electron is within $\Delta R < 0.2$, the jet is discarded.
3. Any remaining electron or muon closer than $\Delta R = 0.4$ to a jet, is discarded.
4. If the distance between a jet and a pre-selected muon is $\Delta R < 0.4$ and the jet has more than two associated tracks, then the muon is discarded, otherwise the jet is discarded.

4.3 Kinematic cuts

In order to suppress potential fakes and quarkonia (low mass resonances such as J/ψ and upsilon) a requirement that all OSSF lepton pairs have an invariant mass, m_{OSSF} , greater than 10 GeV is applied. The final state lepton charges must sum to zero. Therefore a requirement of $\sum_{i=1}^4 \text{charge}(\ell_i) = 0$ is applied. The invariant mass of the OSSF lepton pair coming from the Z boson must equal the invariant mass of the Z boson, and noting that e,μ reconstruction and identification in the ATLAS detector has a high efficiency [63], these OSSF leptons are used to reconstruct Z bosons with relatively high confidence. A Z candidate is defined in this analysis as an OSSF lepton pair with an invariant mass, m_{OSSF} , satisfying the condition, $|m_{\text{OSSF}} - m_Z| < 30$ GeV, where $m(Z)$ is the nominal Z boson mass (91.1876 GeV [52]). This wider mass window is used in order to cover the full range of the $m(Z)$ distribution, in an attempt to increase the number of events which pass our baseline selections. Multiple Z candidates can be present in certain decay channels (e.g. $eeee$, $\mu\mu ee$, $\mu\mu\mu\mu$). In these cases, the Z candidate which has an invariant mass closest to the nominal Z boson mass is chosen.

4.4 Regions and Event Selection

Two tWZ SRs are defined in an attempt to suppress and constrain the ZZ background. Both tWZ SRs are required to have exactly four tight leptons, exactly one Z -boson candidate, exactly one tight b -tagged jet (from the decay of the top quark) and greater than or equal to one jet. The two tWZ SR's differ by the flavours of their leptons which don't originate from the decay of a Z -boson (non- Z leptons). The ZZ background has two Z -bosons which decay into a pair of OSSF lepton pairs, in order to mimic the tWZ signal. This is taken advantage of, to define a tWZ region enrich in ZZ background and one with a minimal ZZ background component. This is done by requiring that one of the tWZ SRs has its two non- Z leptons to have opposite flavour and the other tWZ SR is required to have its non- Z leptons to have the same flavour. These two disjoint tWZ SRs are named tWZ OF SR and tWZ SF SR respectively. It is therefore expected that the tWZ SF SR contains the majority of the ZZ background events across both tWZ SRs. In order to check the modelling of the most dominant

background components in our signal region, $t\bar{t}Z$ and ZZb control regions are defined. The $t\bar{t}Z$ control region has the same requirement on the number of reconstructed Z boson candidates in the signal region (due to a commonality on the number of Z bosons present in both processes), however it is required that there are at least two jets and that exactly two of these jets are b -tagged (corresponding to the b -quark jets originating from the two top quark decays). A ZZb region is defined, as opposed to a ZZ region, since the ZZ background present in the tWZ signal region contains exactly one b -tagged jet. Therefore defining a region with ZZ plus exactly one b -jet more closely resembles the ZZ background present in the signal region. In addition to this, mis-modelling of ZZ has been seen in other analyses [3, 38], further motivating the use of a ZZb control region over a ZZ CR. The ZZb CR requires exactly two Z boson candidates and exactly one b -tagged jet, resulting in an implicit requirement on the number of jets ($N_{jet} \geq 1$).

Fake leptons are objects reconstructed as leptons, but do not correspond to the leptons that are of interest in our analysis. Fake leptons can be split up into two main categories, irreducible fakes and reducible fakes. Irreducible fakes are true leptons which do not come from the process of interest. Reducible fakes are objects which are mis-identified or incorrectly reconstructed as leptons. In the ATLAS detector, the probability for a fake to occur is very low. In order to constrain the fake lepton component contained within the $t\bar{t}Z$ sample, a $(tWZ)_{fake}$ CR is defined which is as similar as possible to the tWZ SRs but is enhanced in fakes. This is achieved by defining the $(tWZ)_{fake}$ CR to inherit the same selection criteria as the tWZ SRs however, in this case, a requirement of exactly 3 tight leptons and exactly 1 loose (and NOT tight) lepton is applied. Loose leptons are required in this region, since looser leptons are more likely to be fakes compared to tighter leptons. A $(t\bar{t}Z)_{fake}$ CR, requiring exactly 3 tight leptons and exactly 1 loose (and NOT tight) lepton, was tried as an alternative to the $(tWZ)_{fake}$ CR, however a much larger suppression of fakes were observed in this region compared to the $(tWZ)_{fake}$ CR. This suppression of fakes can be explained by the extra b -tagged jet requirement (exactly two b -tagged jets are required in the $t\bar{t}Z$ region, compared to exactly one in the tWZ SRs (See Table 6)) which causes suppression of fakes via the overlap removal procedure (See Section 4.2.4). In Table 6, a summary of the final selection criteria and region definitions is shown.

4.4.1 Optimization studies for event selection

In order to find the selection criteria for jets and leptons which maximized sensitivity to the tWZ signal, studies were performed by plotting the expected significance (Z_μ^{exp}) and expected upper limit (μ_{up}^{exp}) for different selection criteria. The fitting procedure as described in Section 4.9.1 was used to calculate the expected upper limits and expected significances in this study (these metrics are fully described later in Section 4.9.1). The same selection criteria and regions defined in Table 6 was used (unless otherwise specified), except for the selection(s) which were being optimised in each case. In Figure 8 the expected significance (Z_μ^{exp}) and expected upper limits (μ_{up}^{exp}) for different $\eta(jet)$ cuts are shown. From Figure 8, it can be seen that the $\eta(jet)$ cut which maximises the sensitivity of tWZ in the tetralepton channel is requiring that $\eta(jet) < 4.5$. This selection criteria was set for the $\eta(jet)$ across all regions. In Figure 9 the expected significance (Z_μ^{exp}) and expected upper limits (μ_{up}^{exp}) for different $p_T(jet)$ cuts are shown. From Figure 9, it can be seen that the $p_T(jet)$ cut which maximises the sensitivity of tWZ is requiring that $p_T(jet) > 20$ GeV. This selection criteria was set for the $p_T(jet)$ across all regions. In Figure 10 the

Baseline selections				
$N_\ell = 4$ $p_T(\ell_1, \ell_2, \ell_3, \ell_4) > (28, 10, 10, 10)$ GeV $p_T(\text{jet}) > 20$ GeV, $ \eta(\text{jet}) < 4.5$, $\text{jvt} > 0.5$ $ \eta(\ell_e) < 2.47$ excluding $1.37 < \eta(\ell_e) < 1.52$ $ \eta(\ell_\mu) < 2.5$ $\sum_{i=1}^4 \text{charge}(\ell_i) = 0$ All OSSF lepton pairs require $m_{\text{OSSF}} > 10$ GeV				
Regions				
tWZ OF SR	tWZ SF SR	$t\bar{t}Z$ CR	ZZb CR	$(tWZ)_{\text{fake}}$ CR
$N_\ell(\text{tight}) = 4$	$N_\ell(\text{tight}) = 4$	$N_\ell(\text{tight}) = 4$	$N_\ell(\text{tight}) = 4$	$N_\ell(\text{tight}) = 3$ $N_\ell(\text{loose and NOT tight}) = 1$
N_Z candidate = 1	N_Z candidate = 1	N_Z candidate = 1	N_Z candidate = 2	N_Z candidate = 1
$N_{\text{jet}} \geq 1$	$N_{\text{jet}} \geq 1$	$N_{\text{jet}} \geq 2$	$N_{\text{jet}} \geq 1$	$N_{\text{jet}} \geq 1$
$N_{b\text{-jet}}(\text{tight}) = 1$	$N_{b\text{-jet}}(\text{tight}) = 1$	$N_{b\text{-jet}}(\text{tight}) \geq 1$ $N_{b\text{-jet}}(\text{loose}) \geq 0$ $N_{b\text{-jet}}(\text{tight}) + N_{b\text{-jet}}(\text{loose}) = 2$	$N_{b\text{-jet}}(\text{tight}) = 1$	$N_{b\text{-jet}}(\text{tight}) = 1$
Opp. Flavour non-Z leptons	Same Flavour non-Z leptons	-	-	-

Table 6: A summary of the requirements applied for selecting events in the signal and control regions is shown.

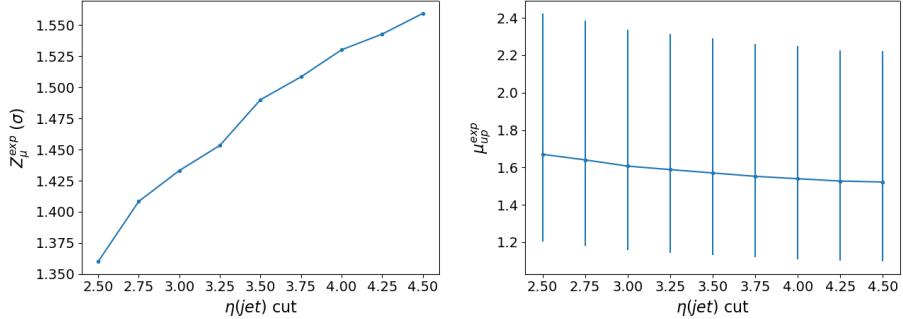


Figure 8: **Left:** The expected significance (Z_μ^{exp}) for different $\eta(\text{jet})$ cuts is shown. The cuts applied on the $\eta(\text{jet})$ are shown on the x-axis and the corresponding expected significance from the likelihood fit is shown on the y-axis. **Right:** Expected upper limit ($\mu_{\text{up}}^{\text{exp}}$) for different $\eta(\text{jet})$ cuts is shown. The cuts applied on the $\eta(\text{jet})$ are shown on the x-axis and corresponding expected upper limits are shown on the y-axis. Error bars representing the total uncertainty on the expected upper limits are shown as vertical lines.

expected significance (Z_μ^{exp}) and expected upper limits ($\mu_{\text{up}}^{\text{exp}}$) for a range of different configurations of DL1r b -tagged jet working points across different regions. From Figure 10, it can be seen that requiring that b -tagged jets pass the 77% DL1r WP in the tWZ SR, $(tWZ)_{\text{fake}}$ CR and the ZZb CR and that at least one b -tagged jet in the $t\bar{t}Z$ SR passes the 77% DL1r WP (the other jet is just required to pass the 85% DL1r WP) maximises the sensitivity overall (compared to the other investigated configurations). This configuration was chosen b -tagged jets. The $p_T(\text{L Lepton})$ is constrained by the single lepton triggers 4. A cut was chosen to be applied on the $p_T(\text{NL Lepton})$ slightly tighter than the tightest single lepton p_T cut in the trigger. The $p_T(\text{NL Lepton})$ cut can be optimized by comparing the expected significance and limit for a range of $p_T(\text{NL Lepton})$ cuts to determine the cut which maximizes sensitivity. In Figure

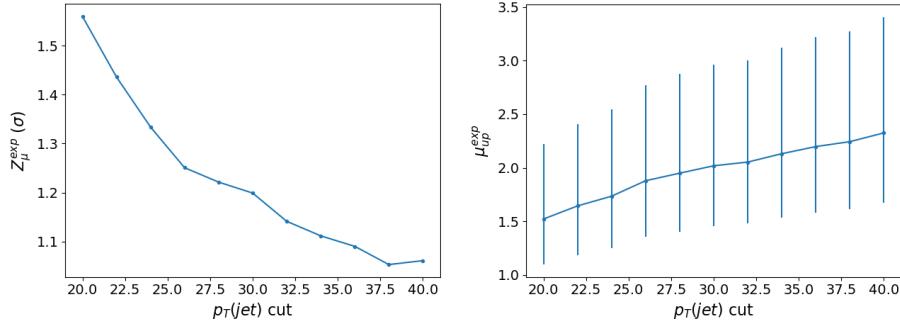


Figure 9: **Left:** The expected significance (Z_μ^{exp}) for different $p_T(jet)$ cuts is shown. The cuts applied on the $p_T(jet)$ are shown on the x-axis and the corresponding expected significance from the likelihood fit is shown on the y-axis. **Right:** Expected upper limit (μ_{up}^{exp}) for different $p_T(jet)$ cuts is shown. The cuts applied on the $p_T(jet)$ are shown on the x-axis and corresponding expected upper limits are shown on the y-axis. Error bars representing the total uncertainty on the expected upper limits are shown as vertical lines.

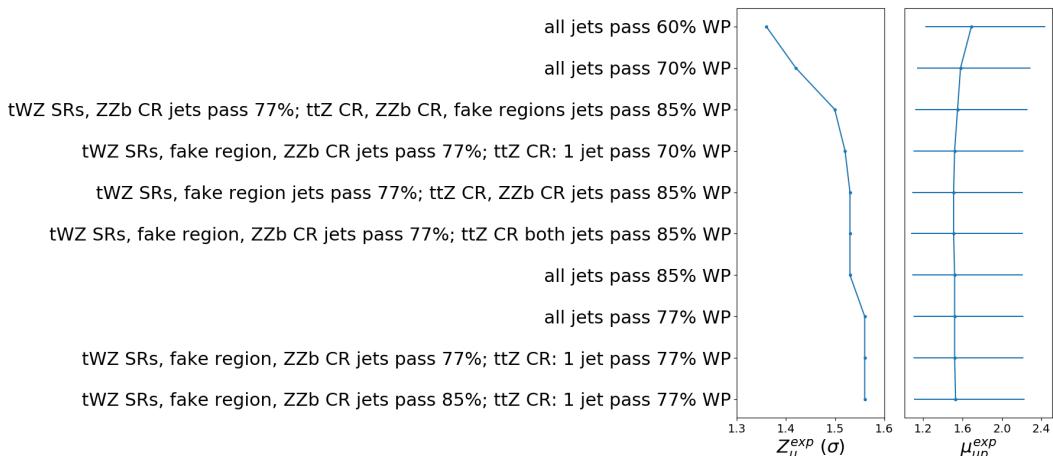


Figure 10: Expected significance (Z_μ^{exp}) and expected upper limit (μ_{up}^{exp}) for different configurations of DL1r b -tagged jet working points is shown. The common y-axis shows the different configurations of DL1r b -tagged jet working points. On the left panel, the expected significance from the likelihood fit is shown on the x-axis. On the right panel, the expected upper limit from the likelihood fit is shown on the x-axis (with the corresponding total uncertainty represented by horizontal lines).

11 the expected significance (Z_μ^{exp}) and expected upper limits (μ_{up}^{exp}) for different p_T (NL Lepton) cuts is
 712 shown. Since there is a very small change between the different p_T (NL Lepton) cuts on the sensitivity
 713 of tWZ , a p_T (NL Lepton) cut is applied at 18 GeV (avoiding a p_T cut near the sharp drop in expected
 714 significance after 28 GeV), therefore applying a cut above the tightest, looser dilepton trigger p_T cut (17
 715 GeV) to suppress any systematic from the modelling of the trigger efficiency.

716 4.5 Signal and Control Regions

717 In this section, expected number of events of variables in each region are shown. For each figure in
 718 this section, the data is given by the black points and the MC predictions for each process are given by

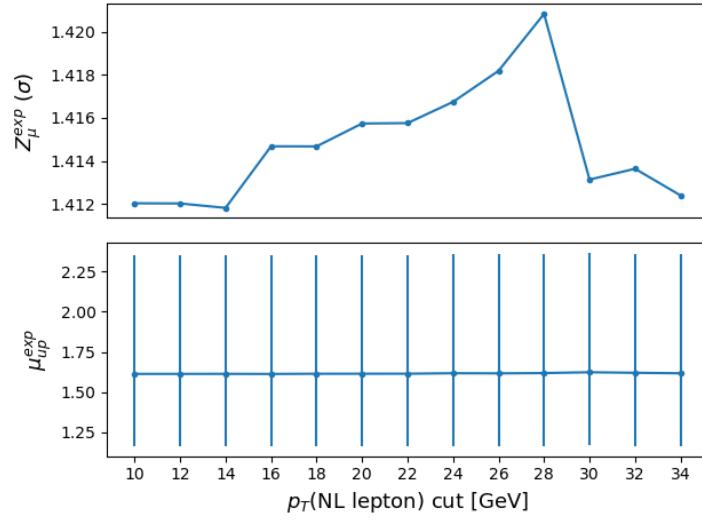


Figure 11: Expected significance (Z_μ^{exp}) and expected upper limit ($\mu_{\text{up}}^{\text{exp}}$) for different p_T (NL Lepton) cuts is shown. The common x-axis shows cut applied to the p_T of the next-to-leading lepton. On the top panel, the expected significance from the likelihood fit is shown on the y-axis. On the bottom panel, the expected upper limit from the likelihood fit is shown on the y-axis (with the corresponding total uncertainty represented by vertical lines).

the filled histograms. The vertical lines on the data points represent the statistical uncertainty in the data and the diagonally-lined bands represent the total (statistical and systematic added in quadrature) uncertainty. The lower panel in each plot shows the ratios of the data to the theoretical predictions. In order to suppress a bias towards large signal observations in the development of the analysis, data has not been analysed in bins where the expected $\frac{\text{signal}}{\text{background}}$ exceeds 0.1. This is known as blinding. Blinded bins are shaded with black diagonal lines and their data points are omitted. In Table 7, the expected number of events for each sample in each region are shown. The finite number of events expected to

		tWZ OF SR	tWZ SF SR	$t\bar{t}Z$ CR	ZZb CR	$(tWZ)_{\text{fake}}$ CR
$t\bar{t}Z$		13.9 ± 1.8	10.1 ± 1.4	31.7 ± 4.5	5.3 ± 0.7	19.1 ± 2.5
$t\bar{t}Z$ fakes		0.068 ± 0.048	0.032 ± 0.026	0.07 ± 0.04	0.05 ± 0.03	5.0 ± 2.5
tWZ		3.8 ± 0.4	2.6 ± 0.3	2.6 ± 0.9	1.4 ± 0.2	5.0 ± 0.7
ZZ		0.5 ± 0.2	8.8 ± 2.7	1.2 ± 0.4	46 ± 14	7.8 ± 2.4
other	$t\bar{t}$	$6e-06 \pm 3e-06$	0.25 ± 0.44	0.27 ± 0.22	$6e-06 \pm 3e-06$	2.4 ± 0.9
	tZq	0.08 ± 0.04	0.08 ± 0.04	0.06 ± 0.03	0.06 ± 0.02	4.9 ± 0.8
	$t\bar{t}W$	0.007 ± 0.007	0.003 ± 0.003	$6e-06 \pm 3e-06$	0.002 ± 0.006	1.0 ± 0.3
	WZ	0.04 ± 0.02	0.04 ± 0.02	0.013 ± 0.013	0.05 ± 0.03	1.8 ± 0.4
	$t\bar{t}t$	0.0010 ± 0.0008	0.002 ± 0.001	0.014 ± 0.004	$6e-06 \pm 3e-06$	0.010 ± 0.004
	$t\bar{t}\bar{t}$	0.0093 ± 0.0081	0.011 ± 0.009	0.057 ± 0.021	$6e-06 \pm 3e-06$	0.02 ± 0.01
	$t\bar{t}WW$	0.029 ± 0.026	0.03 ± 0.02	0.26 ± 0.10	0.01 ± 0.03	0.20 ± 0.06
	$VVV (V = W/Z)$	0.28 ± 0.09	0.20 ± 0.06	0.07 ± 0.02	0.20 ± 0.05	0.3 ± 0.1
	$t\bar{t}H$	0.85 ± 0.18	0.67 ± 0.14	2.0 ± 0.4	0.15 ± 0.04	2.2 ± 0.5
		Total	19.7 ± 2.0	22.9 ± 3.1	38.4 ± 4.6	53.2 ± 14.0
		data	-	-	36	49
						57

Table 7: The expected number of events for each sample in each region is shown.

725

be observed in data (MC simulation) carries an associated statistical uncertainty. To first order, this

727 uncertainty can be written as the square root of the expected number of events to be observed in data.
 728 In contrast to this, predictions based on MC simulation carry uncertainties due to the finite number of
 729 simulated events utilised. This uncertainty can be quantified by the Number of Equivalent Events [45],
 730 N_{equiv} , which relates the sample of N events (weighted by MC event weights) to N_{equiv} events with all
 731 MC event weights equal to 1, that would have the same relative statistical fluctuation. The Number of
 732 Equivalent Events, N_{equiv} , can be written as,

$$N_{equiv} = \frac{(\sum_i^N w_i)^2}{\sum_i^N w_i^2} \quad (4.2)$$

733 where w_i is the MC event weight for event i . The standard uncertainty of N_{equiv} is given by $u(N_{equiv}) =$
 734 $\sqrt{N_{equiv}}$. The Number of Equivalent Events for each sample in each region can be studied in order to
 735 ensure that the number of events simulated for a given process is large in comparison to the number of
 736 events expected for that process in data, thereby ensuring that uncertainties from MC statistics will be
 737 small (or sub-leading). In Table 8, the number of equivalent events, N_{equiv} , is shown for each sample in
 each region. N_{equiv} is much larger compared to the number of expected events (See Table 7) for the signal

	tWZ OF SR	tWZ SF SR	$t\bar{t}Z$ CR	ZZb CR	(tWZ) fake CR
	N_{equiv}	N_{equiv}	N_{equiv}	N_{equiv}	N_{equiv}
tWZ	6463 ± 80	4153 ± 64	4800 ± 69	2497 ± 50	8645 ± 93
$t\bar{t}Z$	1364 ± 37	1031 ± 32	3237 ± 57	561 ± 24	1923 ± 44
ZZ	51 ± 7	975 ± 31	268 ± 16	7023 ± 84	969 ± 31
other	748 ± 27	2.5 ± 1.6	4.2 ± 2.1	255 ± 16	21.5 ± 4.6
$t\bar{t}Z$ fakes	6.7 ± 2.6	1.3 ± 1.1	16.1 ± 4.0	7.2 ± 2.7	484 ± 22
Total	8633 ± 93	6163 ± 79	8326 ± 91	10344 ± 102	12044 ± 110

Table 8: The number of equivalent events, N_{equiv} , is shown for each sample in each region.

738 and background processes in all regions. This tells us that there is a large number of simulated events
 739 for these samples. Therefore ensuring that uncertainties resulting from MC statistics will be small (or
 740 sub-leading).

742 4.5.1 tWZ OF SR

743 In this section, comparisons of simulation and data for different variables in the tWZ OF SR are shown.
 744 In Figure 12, comparisons of simulation and data for p_T , η and ϕ for leading (L) leptons and leading (NL)
 745 jets in the tWZ OF SR are shown. The bins in all of the plots in Figure 12 have $\frac{\text{signal}}{\text{background}}$ exceeding 0.1.
 746 This region is enriched in tWZ signal events. In Figure 13, comparisons of simulation and data of H_T
 747 (scalar sum of Jet p_T), the Number of jets, the scalar sum of b -tagged jet p_T and the number of b -tagged
 748 jets in the tWZ OF SR are shown. All bins for each plot in Figure 13 have $\frac{\text{signal}}{\text{background}}$ exceeding 0.1 and
 749 are therefore blinded. This region is therefore enriched in tWZ signal events.

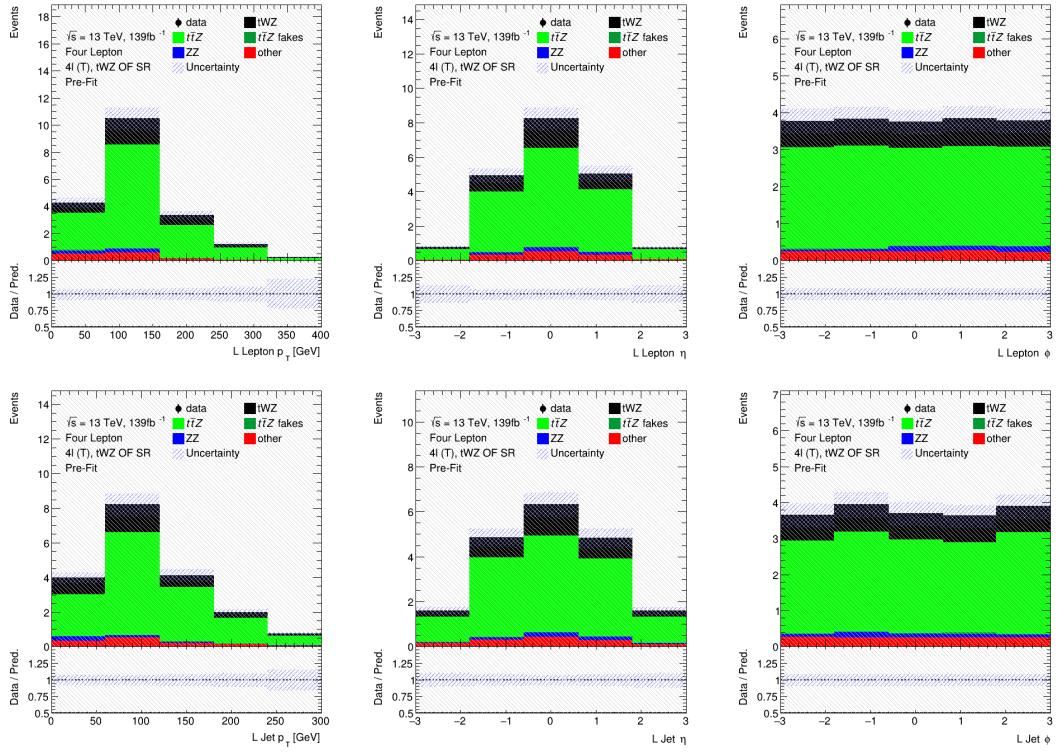


Figure 12: Comparisons of simulation and data of p_T , η and ϕ for leading (L) leptons (top row) and leading (NL) jets (bottom row) in the tWZ OF SR are shown.

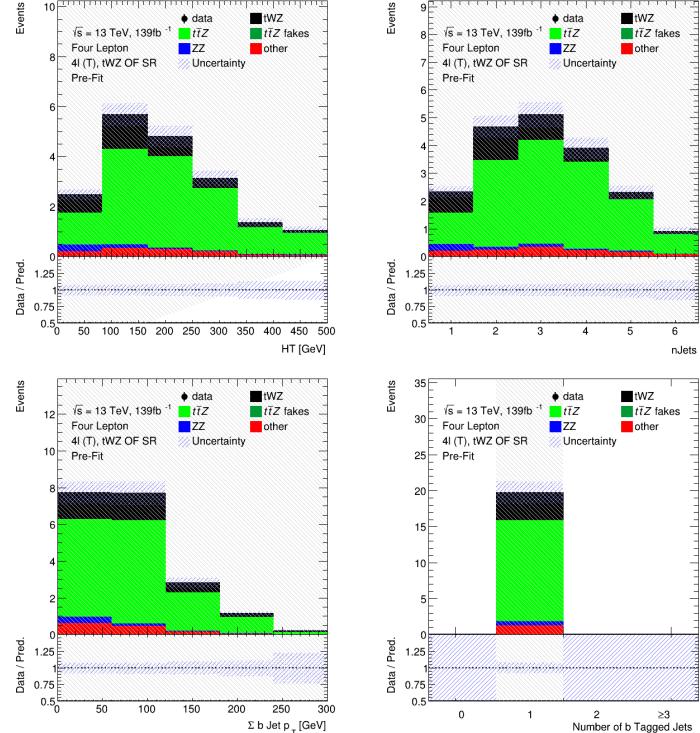


Figure 13: Comparisons of simulation and data of H_T (scalar sum of Jet p_T), the Number of jets, the scalar sum of b -tagged jet p_T and the number of b -tagged jets (top left to bottom right) in the tWZ OF SR are shown.

4.5.2 tWZ SF SR

In this section, expected number of events of variables in the tWZ SF SR are shown. In Figure 14, comparisons of simulation and data of p_T , η and ϕ for leading (L) leptons and leading (NL) jets in the tWZ SF SR are shown. The vast majority of bins for each plot in Figure 14 have $\frac{\text{signal}}{\text{background}}$ exceeding

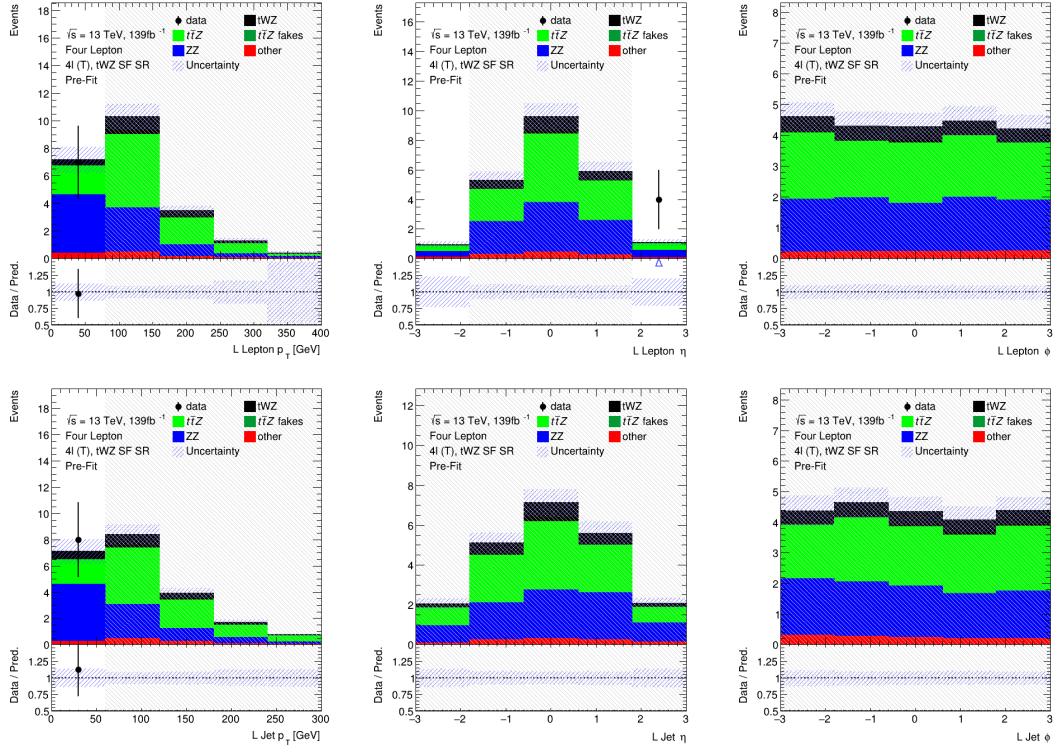


Figure 14: Comparisons of simulation and data of p_T , η and ϕ for leading (L) leptons (top row) and leading (NL) jets (bottom row) in the tWZ SF SR are shown.

0.1 and are therefore blinded. This region is therefore enriched in tWZ signal events. In Figure 15, comparisons of simulation and data of H_T (scalar sum of Jet p_T), the Number of jets, the scalar sum of b -tagged jet p_T and the number of b -tagged jets in the tWZ SF SR are shown. The vast majority of bins in each plot in Figure 15 have $\frac{\text{signal}}{\text{background}}$ exceeding 0.1 and are therefore blinded. This region is therefore enriched in tWZ signal events. The deviations in data and simulation in the two bins (in the HT and σb jet p_T distributions) which are not blinded, are within the expected uncertainties.

4.5.3 $t\bar{t}Z$ CR

In this section, expected number of events of variables in the $t\bar{t}Z$ CR are shown. In Figure 16, comparisons of simulation and data of p_T , η and ϕ for leading (L) leptons and leading (NL) jets in the $t\bar{t}Z$ CR are shown. The majority of the deviations in data and simulation for each plot in Figure 16 are within the expected uncertainties. The few plots which have bins where there is a disagreement between data and simulation are either within 2σ (L Jet ϕ) or 3σ (L Jet η) standard uncertainties from one another, or are show more than a 3σ (L Lepton p_T) disagreement. The disagreement in the L Lepton p_T distribution could be due to statistical fluctuations in data or simulation, since there are so few events in these bins.

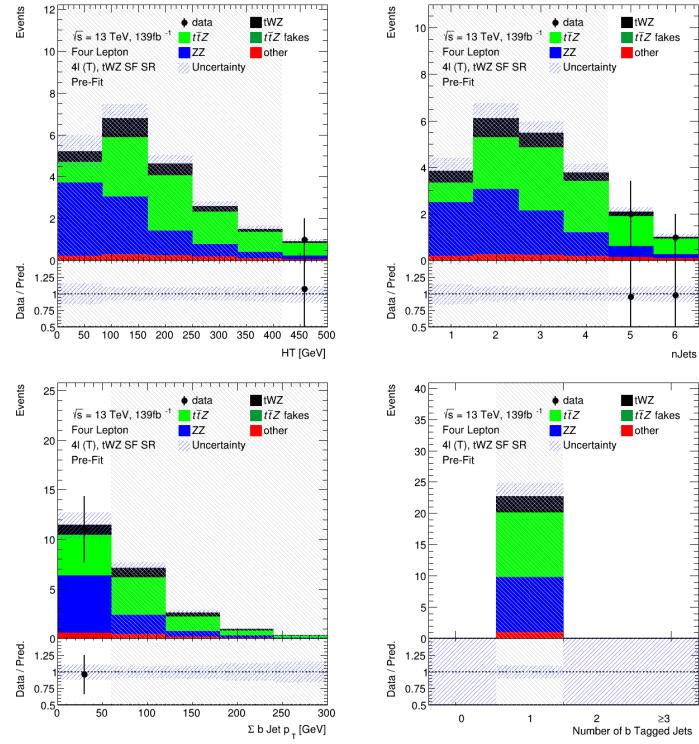


Figure 15: Comparisons of simulation and data of H_T (scalar sum of Jet p_T), the Number of jets, the scalar sum of b -tagged jet p_T and the number of b -tagged jets (top left to bottom right) in the tWZ SF SR are shown.

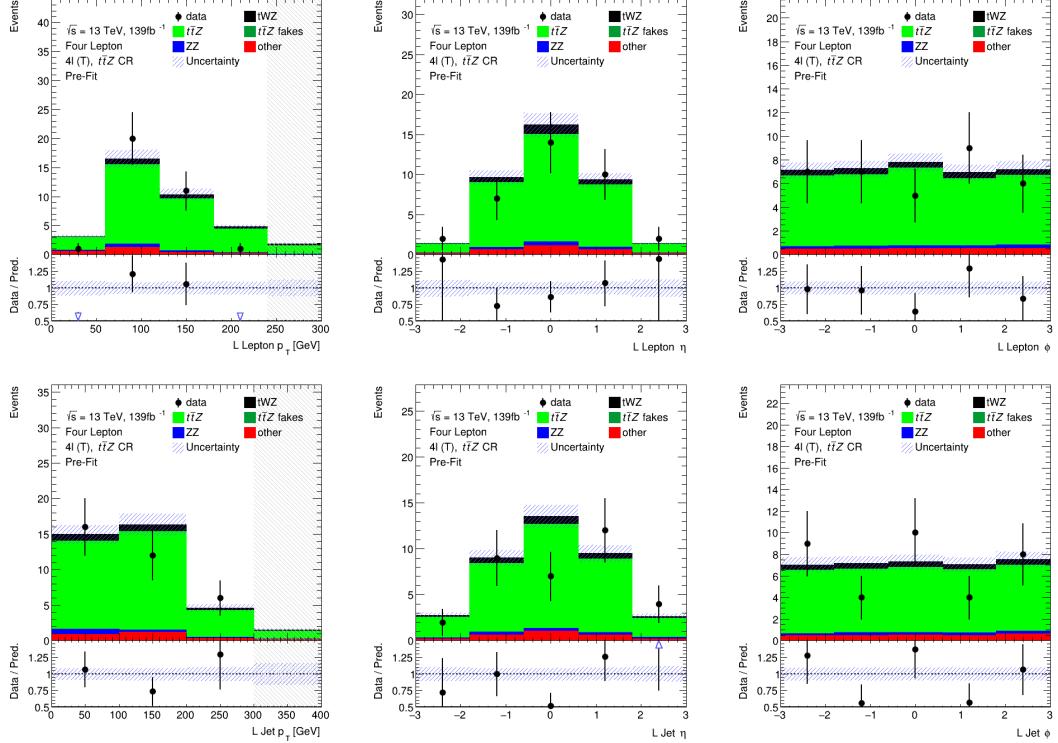


Figure 16: Comparisons of simulation and data of p_T , η and ϕ for leading (L) leptons (top row) and leading (NL) jets (bottom row) in the $t\bar{t}Z$ CR are shown.

- 768 In Figure 17, comparisons of simulation and data of H_T (scalar sum of Jet p_T), the Number of jets, the scalar sum of b -tagged jet p_T and the number of b -tagged jets in the $t\bar{t}Z$ CR are shown. Almost all of

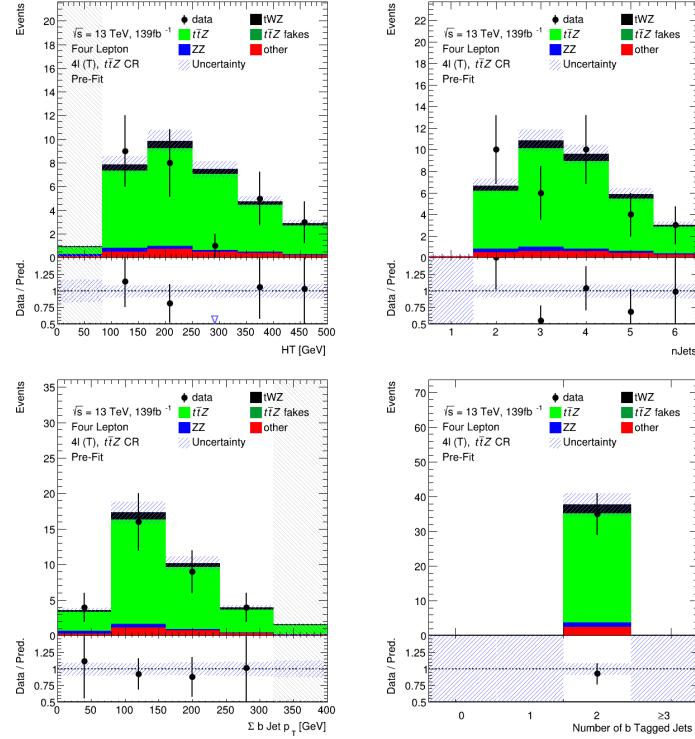


Figure 17: Comparisons of simulation and data of H_T (scalar sum of Jet p_T), the Number of jets, the scalar sum of b -tagged jet p_T and the number of b -tagged jets (top left to bottom right) in the $t\bar{t}Z$ CR are shown.

- 769
 770 the deviations in data and simulation for each plot in Figure 17 are within the expected uncertainties.
 771 There is a 2σ disagreement in one of the bins in the $n\text{Jets}$ distribution and a large disagreement ($> 5\sigma$)
 772 in one of the bins in the H_T distribution. The large disagreement between data and simulation in the
 773 H_T distribution is surprising since all other bins in the distribution agree within 1σ uncertainties, and it
 774 is therefore not fully understood.

775 4.5.4 ZZb CR

- 776 In this section, expected number of events of variables in the ZZb CR are shown. In Figure 18, comparisons
 777 of simulation and data of p_T , η and ϕ for leading (L) leptons and leading (NL) jets in the ZZb CR are
 778 shown. Most of the deviations in data and simulation for each plot in Figure 18 are within the expected
 779 uncertainties. There are a few bins with 2σ and $> 2\sigma$ disagreements between data and simulation in the
 780 L Lepton p_T , L Lepton η and L Jet p_T distributions, with the disagreement being much more noticeable
 781 in the L Lepton distributions. This could suggest some mis-modelling for L Leptons in this region. In
 782 Figure 19, comparisons of simulation and data of H_T (scalar sum of Jet p_T), the Number of jets, the
 783 scalar sum of b -tagged jet p_T and the number of b -tagged jets in the ZZb CR are shown. Most of the
 784 deviations in data and simulation for each plot in Figure 18 are within the expected uncertainties.

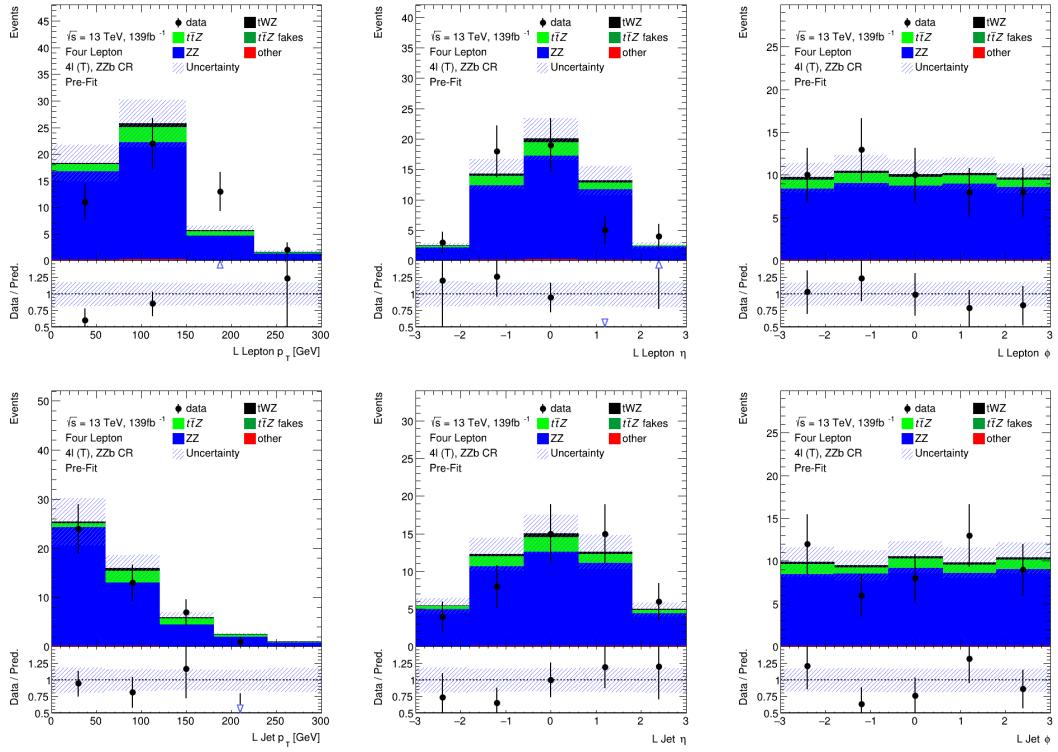


Figure 18: Comparisons of simulation and data of p_T , η and ϕ for leading (L) leptons (top row) and leading (NL) jets (bottom row) in the ZZb CR are shown.

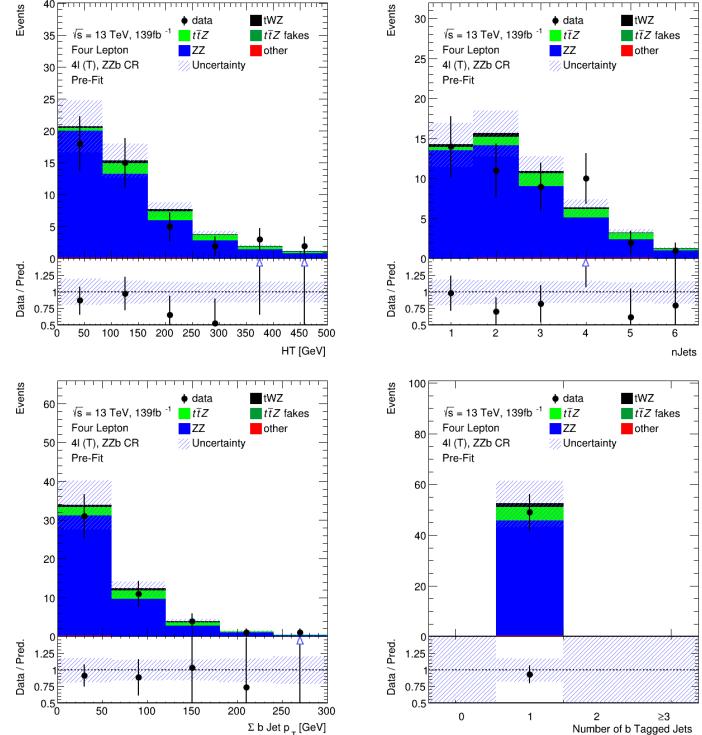


Figure 19: Comparisons of simulation and data of H_T (scalar sum of Jet p_T), the Number of jets, the scalar sum of b -tagged jet p_T and the number of b -tagged jets (top left to bottom right) in the ZZb CR are shown.

785 **4.5.5 $(tWZ)_{\text{fake}}$ CR**

786 In this section, expected number of events of variables in the $(tWZ)_{\text{fake}}$ CR are shown. In Figure 20,
 787 comparisons of simulation and data of p_T , η and ϕ for leading (L) leptons and leading (NL) jets in the
 788 $(tWZ)_{\text{fake}}$ CR are shown. The vast majority of bins in each plot in Figure 20 have $\frac{\text{signal}}{\text{background}}$ exceeding

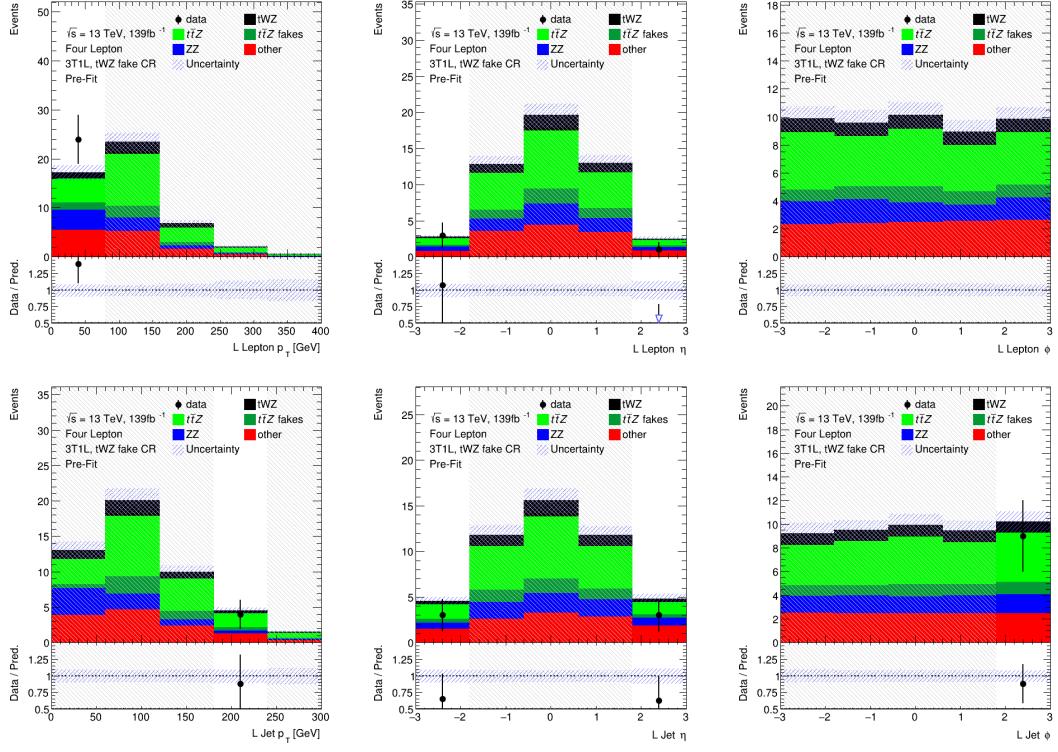


Figure 20: Comparisons of simulation and data of p_T , η and ϕ for leading (L) leptons (top row) and leading (NL) jets (bottom row) in the $(tWZ)_{\text{fake}}$ CR are shown.

788 0.1 and are therefore blinded. This region is therefore enriched in tWZ signal events. Most deviations in
 789 data and simulation in the bins which are not blinded, are within the expected uncertainties. Only two
 790 out of seven unblinded bins are not within expected uncertainties and are within a 2σ uncertainty. In
 791 Figure 21, comparisons of simulation and data of H_T (scalar sum of Jet p_T), the Number of jets, the scalar
 792 sum of b -tagged jet p_T and the number of b -tagged jets in the $(tWZ)_{\text{fake}}$ CR are shown. The majority
 793 of bins in each plot in Figure 21 have $\frac{\text{signal}}{\text{background}}$ exceeding 0.1 and are therefore blinded. This region is
 794 therefore enriched in tWZ signal events. Most deviations in data and simulation in the bins which are
 795 not blinded, are within the expected uncertainties. Only two out of seven unblinded bins are not within
 796 expected uncertainties and are within a 2σ uncertainty.
 797

798 **4.6 Fake Lepton Estimation**

799 Fake leptons are physics objects reconstructed as leptons, but do not correspond to the leptons which
 800 originate from the hard scatter process or those physics objects that are mis-identified as leptons.
 801 The sources of fake leptons include those originating from heavy hadron decays, light hadron de-
 802 cays or via the conversion of a photon to a lepton. In the ATLAS detector, the probability for a fake

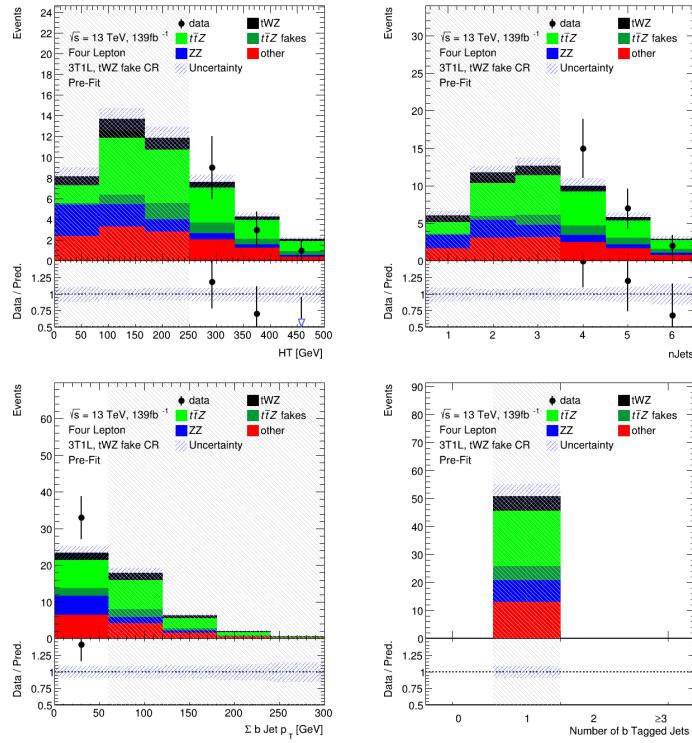


Figure 21: Comparisons of simulation and data of H_T (scalar sum of Jet p_T), the Number of jets, the scalar sum of b -tagged jet p_T and the number of b -tagged jets (top left to bottom right) in the $(tWZ)_{\text{fake}}$ CR are shown.

803 to occur is very low. In this section, the method used to estimate the fake lepton contribution is described.

804

805 As $t\bar{t}Z$ is the dominant background process ($\sim 75\%$ of the total background contribution), it is assumed
 806 that $t\bar{t}Z$ will also dominate the events containing fake leptons. The fake lepton efficiency, ϵ , can be
 807 written as $\epsilon = \frac{N_{\text{fake}}^{\text{tight}}}{N_{\text{fake}}^{\text{loose}}}$, where $N_{\text{fake}}^{\text{tight}}$ is the number of fake leptons which pass the tight lepton selection
 808 (See Section 4.2.1) and $N_{\text{fake}}^{\text{loose}}$ is the number of fake leptons which pass the loose lepton selection (See
 809 Section 4.2.1). The probability of one fake lepton to occur, $P(\text{one fake } \ell)$, is proportional to $\epsilon_1 \ll 1$ [61,
 810 77] and the probability for two fakes to occur is, $P(\text{two fakes } \ell)$, is proportional to $\epsilon_2 < \epsilon_1 \ll 1$. In this
 811 analysis, an estimation of the fake lepton component to the highest order is investigated and therefore
 812 the case where at least one fake lepton occurs in a $t\bar{t}Z$ event is considered.

813

814 Firstly, the dominant $t\bar{t}Z$ background is split up into $t\bar{t}Z$ and $(t\bar{t}Z)_{\text{fake}}$ components. Secondly, a
 815 $(tWZ)_{\text{fake}}$ CR (See Section 4.4) is defined which is enhanced in fakes and aims to constrain the $(t\bar{t}Z)_{\text{fake}}$
 816 background in the SR. All events which contribute to the $(t\bar{t}Z)_{\text{fake}}$ background are determined by the
 817 IFF Truth Classifier [56]. The IFF Truth Classifier is a tool which aims to classify leptons based
 818 off their truth information. It uses the more general MCTruthClassifier [66] tool's output as input and
 819 returns one of the following lepton categories: Unknown, KnownUnknown (leptons which can (in principle) be
 820 classified, but the MCTruthClassifier fails to classify the lepton's truth type or origin), IsoElectron,
 821 ChargeFlipIsoElectron, PromptMuon, PromptPhotonConversion, ElectronFromMuon, TauDecay,
 822 BHadronDecay, CHadronDecay or LightFlavorDecay (More details [56]). Given these categories, leptons

are considered as fake if they are classified as `PromptPhotonConversion`, `BHadronDecay`, `CHadronDecay` or `LightFlavorDecay` (i.e. a lepton originating from the decay of a b -Hadron, c -Hadron or light-flavour jet). Events which contribute to the $(t\bar{t}Z)_{\text{fake}}$ background are those where at least one lepton from the $t\bar{t}Z$ sample are classified by the IFF Truth Classifier with one of the four aforementioned categories.

827

The $(tWZ)_{\text{fake}}$ CR aims to be as similar as possible to the tWZ SRs, but enhanced in fakes. This CR can then be used to constrain the normalisation of the $(t\bar{t}Z)_{\text{fake}}$ template. To ensure that this region is enhanced in fakes, it is required that it contains 3 tight leptons and 1 loose lepton, since loose leptons are more likely to be fakes. Leptons from heavy decays are produced in jets and are typically surrounded by other energetic particles. Since the loose lepton definition relaxes the isolation requirement, leptons satisfying the loose criteria are more enhanced in these fake leptons. By using the p_T of the loose lepton ($p_T(\text{Loose Lepton})$) in this region as the variable used in the fit, the shape (and normalisation) of the $(t\bar{t}Z)_{\text{fake}}$ template can be constrained. In Figure 22, the number of leptons classified as fake and the relative dominance of the different classifications for fake leptons, split up by their IFF Truth classification, in each region are shown.

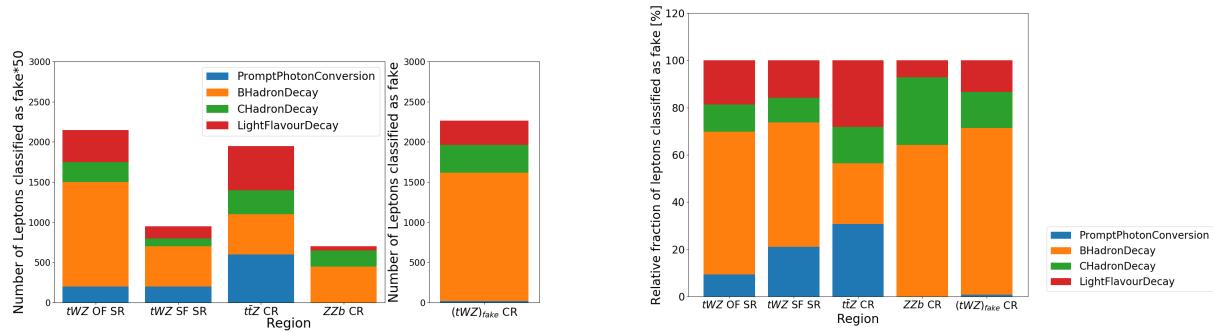


Figure 22: **Left:** The number of leptons classified as fake, split up by their IFF Truth classification, in each region is shown. The left panel shows the number of leptons classified as fakes, scaled by a factor of 50, on the y-axis. The right panel shows the number of leptons classified as fakes (unscaled), on the y-axis. The different signal and control regions are shown on the x-axes of the left and right panels. The different coloured stacked histograms correspond to the IFF truth classification of the leptons, as shown in the legend. **Right:** The relative dominance of the different classifications for fake leptons (classified by the IFF truth classifier) in each region, is shown. The relative dominance of leptons classified as fakes, as a fraction of the total number of fake leptons (in each region), is shown on the y-axis. The different signal and control regions are shown on the x-axis. The different coloured stacked histograms correspond to the IFF truth classification of the leptons, as shown in the legend.

The plot on the left illustrates that there is a large amount of fake leptons which pass our selection criteria for the $(tWZ)_{\text{fake}}$ CR, compared to remaining four regions. Therefore there is a significant amount of fake leptons present in the $(tWZ)_{\text{fake}}$ CR which allow the fake lepton component to be sufficiently constrained. The plot on the right illustrates that the majority of fake leptons which pass our selection criteria originate from the decay of b -hadrons, in all regions but the $t\bar{t}Z$ CR. The smaller proportion of fake leptons originating from b -hadron decays in the $t\bar{t}Z$ CR could possibly be due to statistical fluctuations resulting from the low number of fake leptons which pass our selection criteria in this region (~ 40 fake leptons). In Figure 23, the amount of fake and real $t\bar{t}Z$ events which pass our selection criteria, in each

region, is shown. Around 20% of all $t\bar{t}Z$ events are classified as fake events (having one or more of its

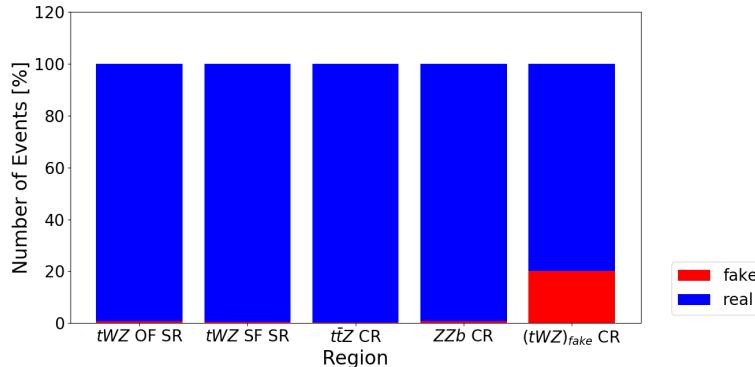


Figure 23: The percentage of fake and real $t\bar{t}Z$ events which pass our selection criteria, in each region, is shown. The relative number of fake and real events (in % of the total number of events in the nominal and fake $t\bar{t}Z$ background samples) is shown on the y-axis. The different signal and control regions are shown on the x-axis. The blue and red histograms represent the percentage of real and fake events (out of the total number of events in the nominal and fake $t\bar{t}Z$ background samples), respectively.

leptons being classified as fake) in the $(tWZ)_{fake}$ CR. The tWZ OF SR, tWZ SF SR, ttZ CR and ZZb CR have less than 1% of their total $t\bar{t}Z$ events being fake. The significant fraction of fake $t\bar{t}Z$ events present in the $(tWZ)_{fake}$ CR allows the $t\bar{t}Z$ fake background to be sufficiently constrained by the $(tWZ)_{fake}$ CR.

4.7 Improving signal vs background discrimination

The presence of different numbers of top quarks is a key discriminator between signal and the dominant background process, $t\bar{t}Z$. This information is aimed to be exploited by reconstructing ℓb systems as a proxy for top quarks (since, $t \rightarrow W(\rightarrow \ell\nu)b$). This is done in two ways, firstly, by implementation of a kinematic reconstruction algorithm (Two Neutrino Scanning Method) which aims to determine the likelihood of an event containing two top quarks and secondly, by implementing a Boosted Decision Tree (BDT) which is used to distinguish between ℓb systems that originate from top quarks and ℓb systems which do not originate from top quarks. In this thesis, this BDT is referred to as an *object-level* BDT. Certain variables constructed from event information show discrimination between signal and background events. This information can be exploited to discriminate between signal and background events by constructing a BDT which uses these discriminating variables for training. A BDT is implemented and is used to discriminate between tWZ events and its major backgrounds, $t\bar{t}Z$ and ZZ . Furthermore, this BDT takes information from the kinematic reconstruction algorithm and the object-level BDT in order to maximize its discriminating power. In this thesis, this BDT is referred to as an *event-level* BDT. The discriminator output from the object-level BDT can be converted to a variable which can then be used as an input to the event-level BDT.

4.7.1 Two Neutrino Scanning Method (2ν SM) Algorithm

The difference in the number of resonant top quarks in the tWZ signal and the dominant background, $t\bar{t}Z$, is a key feature which can be exploited in order to discriminate between these two processes. In

Section 4.7.3, a BDT was implemented which exploits this information by aiming to identify ℓb systems originating from top quarks. In this section, a kinematic reconstruction algorithm (Two Neutrino Scanning Method) is implemented which exploits the same feature. The Two Neutrino Scanning Method (2ν SM) algorithm¹ [65, 64] aims to reconstruct $t\bar{t}$ systems in the 2ℓ , 3ℓ and 4ℓ final states (e.g. 2ℓ case: $t\bar{t} \rightarrow \ell^+\nu_\ell b\ell^-\bar{\nu}_\ell\bar{b}$). The 2ν SM algorithm aims to reconstruct a $t\bar{t}$ system by finding two neutrinos (ν_1 and ν_2) which are most likely to correspond to the neutrinos that originate from the decay of a $t\bar{t}$ system. This algorithm can be used in our analysis to discriminate between tWZ and $t\bar{t}Z$, since the OSSF leptons which decay from the Z boson can be easily reconstructed and removed before inputting the event into the algorithm. The removal of the Z boson results in tWZ events that don't resemble $t\bar{t}$ systems and $t\bar{t}Z$ events that do resemble $t\bar{t}$ systems, which the algorithm is designed to distinguish between. It would then be expected that the 2ν SM algorithm returns a higher score from a $t\bar{t}Z$ event (~ 1 , i.e. it resembles a $t\bar{t}$ event after removal of the Z boson) and a lower score from a tWZ event (~ 0 , i.e. it does not resemble a $t\bar{t}$ event after removal of the Z boson). The first step in the 2ν SM algorithm involves stating four equations which correspond to the invariant masses of the top quark ($m(t)$) and W boson ($m(W)$) for the two top quark decays (i.e. $t \rightarrow W^+b \rightarrow \ell^+\nu_\ell$) in a dileptonic $t\bar{t}$ event. These can be written as,

$$(\ell_1 + \nu_1)^2 = m(W)^2 = (80.385 \text{ GeV})^2 \quad (4.3)$$

$$(\ell_1 + \nu_1 + b_{1,2})^2 = m(t)^2 = (172.5 \text{ GeV})^2 \quad (4.4)$$

$$(\ell_2 + \nu_2)^2 = m(W)^2 = (80.385 \text{ GeV})^2 \quad (4.5)$$

$$(\ell_2 + \nu_2 + b_{2,1})^2 = m(t)^2 = (172.5 \text{ GeV})^2 \quad (4.6)$$

where the subscripts indicate that these particles originate from the decay of two different top quarks in a $t\bar{t}$ system. An assumption is made such that the mass of the neutrinos (ν_1 and ν_2) are exactly zero, which leaves us with 6 unknowns, p_{T,ν_1} , ϕ_{ν_1} , η_{ν_1} , p_{T,ν_2} , ϕ_{ν_2} and η_{ν_2} (components of the two neutrino's 4-vectors). The 4-vectors of the two reconstructed leptons (not from the Z boson) and the two jets with the highest DL1r b -tagger score are used as input to the algorithm. For each neutrino (ν_1 and ν_2), a scan over a range of possible η and ϕ values is performed. These values were chosen to be $\phi_{\nu_1}, \phi_{\nu_2} \in [-\pi, \pi]$ with a step size of ≈ 0.25 and $\eta_{\nu_1}, \eta_{\nu_2} \in [-5, 5]$ with a step size of ≈ 0.31 . These ranges were chosen to maximize accuracy and minimize computation time. For each of these possible η and ϕ values, the corresponding p_T for each neutrino is calculated (p_{T,ν_1} and p_{T,ν_2}) via,

$$p_{T,\nu} = \frac{\frac{1}{2}(m(W)^2 - m(\ell)^2)}{E_\ell \cosh \eta_\nu - p_{\ell,z} \sinh \eta_\nu - p_{\ell,x} \cos \phi_\nu - p_{\ell,y} \sin \phi_\nu} \quad (4.7)$$

where E_ℓ is the energy of the lepton, $m(\ell)$ is the invariant mass of the lepton and $p_{\ell,z}, p_{\ell,x}, p_{\ell,y}$ are the z , x and y components of lepton's momentum. After computing $p_{T,\nu}$ for both neutrinos for each possible η and ϕ combination in the defined ranges, a collection of possible 4-vectors for ν_1 and ν_2 can be reconstructed.

¹software tool and weights provided by Thomas McCarthy ($t\bar{t}Z$ analysis group - Max Planck Institute)

Using ν_1 and ν_2 , two possible $t\bar{t}$ systems are reconstructed,

$$t_1 = \ell_1 + b_1 + \nu_1 \text{ and } t_2 = \ell_2 + b_2 + \nu_2 \quad (4.8)$$

OR

$$t_1 = \ell_1 + b_2 + \nu_1 \text{ and } t_2 = \ell_2 + b_1 + \nu_2 \quad (4.9)$$

893 The 2ν SM algorithm is extremely computationally intensive. The computation time depends on the
 894 number step size of the ϕ and η ranges which are scanned over to reconstruct the neutrinos. For example,
 895 consider the step sizes chosen in this analysis, $\Delta\eta \approx 0.31$ and $\Delta\phi \approx 0.25$ which corresponds to 32 values
 896 for η and 25 values for ϕ . There will be $(32)(32)(25) = 640\,000$ possible pairs of neutrinos (ν_1 and
 897 ν_2) to consider per event. Since two possible $t\bar{t}$ systems (See Equations 4.8 and 4.9) are considered, this
 898 number effectively increases to $(2)(640000) = 128\,000$ iterations per event. In order to reduce the
 899 number of $t\bar{t}$ systems needed to be considered, therefore decreasing computation time, distributions of
 900 variables from $t\bar{t}$ events are studied to apply a veto to a possible reconstructed $t\bar{t}$ system if the variable
 901 in question is improbable or unlikely to be observed in a $t\bar{t}$ event. To achieve this, an allowed range is
 902 defined for these variables (See Figure 25 and Figure 26), and if the possible reconstructed $t\bar{t}$ system's
 903 corresponding value for this variable lies outside this range, it is vetoed and the algorithm continues with
 904 the next iteration. The first variable which is considered, is the difference between average mass of the
 905 two possible ℓb system combinations, $\Delta\langle m(\ell b) \rangle$. The two possible ℓb system combinations are,

$$(\ell_1 b_1) = \ell_1 + b_1 \text{ and } (\ell_2 b_2) = \ell_2 + b_2 \quad (4.10)$$

OR

$$(\ell_1 b_2) = \ell_1 + b_2 \text{ and } (\ell_2 b_1) = \ell_2 + b_1 \quad (4.11)$$

906 $\Delta\langle m(\ell b) \rangle$ is therefore defined as,

$$\Delta\langle m(\ell b) \rangle = \frac{1}{2} |[(m(\ell_1 b_1) + m(\ell_2 b_2)) - (m(\ell_1 b_2) + m(\ell_2 b_1))]| \quad (4.12)$$

907 The idea here is that, in events where the average masses of the two possible ℓb system combina-
 908 tions differ greatly, the correct combination is likely to be given by the combination with the smaller
 909 average mass. Furthermore, it was shown in Ref. [65] that reconstructed top quarks in a $t\bar{t}$ sys-
 910 tem that contain b -tagged jets in opposite hemispheres² ($\eta(b_1) \times \eta(b_2) < 0$) of the ATLAS detector
 911 are easier to determine the correct ℓb system combination than reconstructed $t\bar{t}$ systems that con-
 912 tain b -tagged jets in the same hemispheres ($\eta(b_1) \times \eta(b_2)$). To illustrate this, the distributions (con-
 913 structed from $t\bar{t}$ events) of the probability of choosing the correct ℓb system combination, given that
 914 the one with the minimum $\langle m(\ell b) \rangle$ chosen ($P(\text{Correct combination of } \ell b \text{ systems} | \text{minimum} \langle m(\ell b) \rangle)$) vs
 915 $\Delta\langle m(\ell b) \rangle$ for b -tagged jets in the same and opposite hemispheres are investigated. In Figure 24, the
 916 $P(\text{Correct combination of } \ell b \text{ systems} | \text{minimum} \langle m(\ell b) \rangle)$ vs $\Delta\langle m(\ell b) \rangle$, for b -tagged jets in the same and
 917 opposite hemispheres, constructed from $t\bar{t}$ events are shown. From Figure 24, for both cases where the
 918 b -tagged jets are in the same and opposite hemispheres, the probability for a correct ℓb system being

²The ATLAS detector can be split into two regions or *hemispheres*, defined where $z > 0$ and $z < 0$

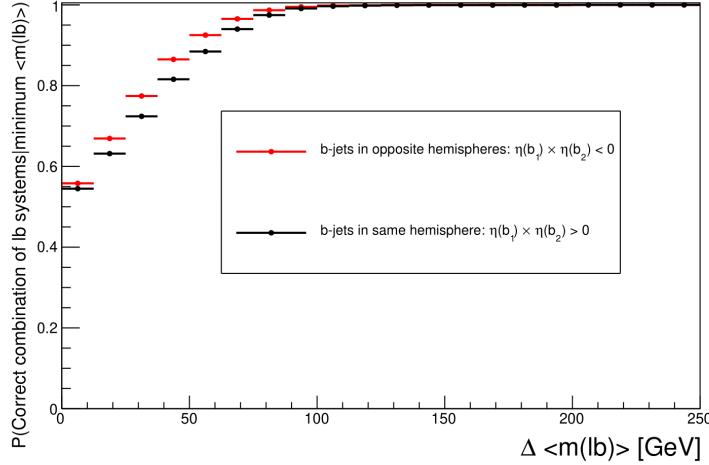


Figure 24: $P(\text{Correct combination of } lb \text{ systems}|\text{minimum}\langle m(lb) \rangle)$ vs $\Delta\langle m(lb) \rangle$, for b -tagged jets in the same and opposite hemispheres, constructed from $t\bar{t}$ events is shown. The horizontal red lines represent the distribution in the case when the two b -jets are in opposite hemispheres. The dot in the middle of the line represents the midpoint of the line. The horizontal black lines represent the distribution in the case when the two b -jets are in the same hemispheres. The dot in the middle of the line represents the midpoint of the line. The average $m(lb)$ is shown on the x-axis. The $P(\text{Correct combination of } lb \text{ systems}|\text{minimum}\langle m(lb) \rangle)$ is shown on the y-axis.

chosen, given that the lb system with the minimum average mass is under consideration, is an increasing function which plateaus to 1 at ~ 90 GeV. One of these two distributions are used (depending on whether or not the two b -tagged jets are in the same or opposite hemispheres) to interpolate the $P(\text{Correct combination of } lb \text{ systems}|\text{minimum}\langle m(lb) \rangle)$ from $\Delta\langle m(lb) \rangle$. A veto is applied to the lb combination with the maximum $\Delta\langle m(lb) \rangle$ if $P(\text{Correct combination of } lb \text{ systems}|\text{minimum}\langle m(lb) \rangle) > 0.8$, indicating that there is at least an 80% certainty that the lb combination with the minimum $\langle m(lb) \rangle$ is the correct combination. If $P(\text{Correct combination of } lb \text{ systems}|\text{minimum}\langle m(lb) \rangle) < 0.8$, both possible lb system combinations need to be considered. The η of the $b\bar{b}\ell\ell$ system, $\eta(b\bar{b}\ell\ell)$, to veto improbable $\eta(\nu_1)$ and $\eta(\nu_2)$ values is then considered. In the same way as for $\Delta\langle m(lb) \rangle$, a distribution is generated to determine values $\eta(\nu)$ which are improbable for a $t\bar{t}$ event. In this case, a 2D histogram from simulated $t\bar{t}$ events (dileptonic final state) at generator-level of $\eta(\nu)$ vs $\eta(b\bar{b}\ell\ell)$ is generated. Using this histogram, a veto region (where a $t\bar{t}$ event is extremely unlikely to occur) is defined which contains 95% of events. A veto is applied if either possible neutrino lies within this region. In Figure 25, a heatmap of occupancy for $\eta(\nu)$ vs $\eta(b\bar{b}\ell\ell)$ (produced from simulated $t\bar{t}$ events) and its corresponding veto region are shown. The final kinematic constraint which is considered is the scalar sum of lepton p_T , $L_T = p_T(\ell_1) + p_T(\ell_2)$, which is used to veto certain possible neutrinos, ν_1 and ν_2 . Again, a distribution is generated to determine (and veto) improbable possible neutrinos in simulated $t\bar{t}$ events (dilepton final state). Using the same method as described for Figure 25, a veto region is defined where a veto is applied if either possible neutrino lies within this region. In Figure 26, a heatmap of occupancy for $\Delta R(\ell, \nu)$ vs L_T (produced from simulated $t\bar{t}$ events) and its corresponding veto region are shown. In order to choose the solution which best represents the two top quarks in a $t\bar{t}$ system, the likelihood of each solution is evaluated in the

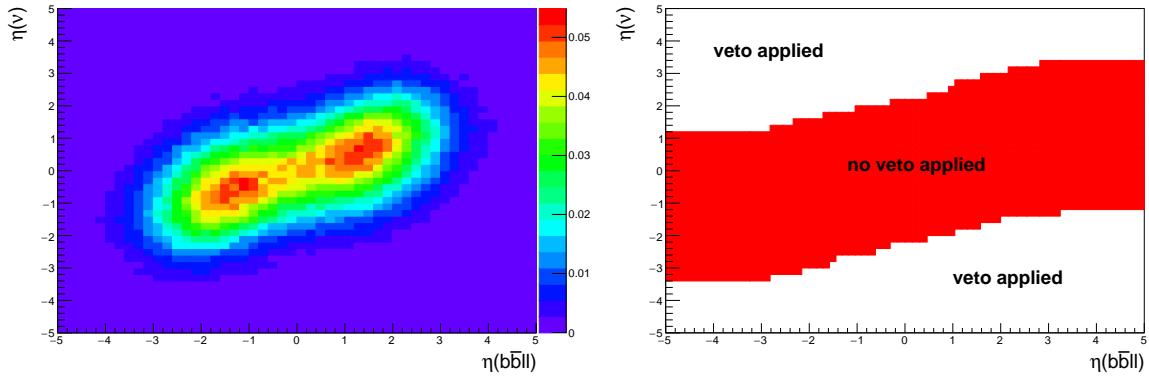


Figure 25: **Left:** Heatmap of occupancy for $\eta(\nu)$ vs $\eta(b\bar{b}ll)$ produced from simulated $t\bar{t}$ events (dileptonic final state) at generator-level is shown. η of the $b\bar{b}ll$ system is shown on the x-axis. η of the neutrino is shown on the y-axis. The colourbar on the right represents the fraction of events in the phase space. **Right:** The regions where vetoes are applied for the $\eta(b_1 b_2 \ell_1 \ell_2)$ constraint is shown. η of the $b\bar{b}ll$ system is shown on the x-axis. η of the neutrino is shown on the y-axis. The red band shows the region where the neutrino would not be vetoed. The white areas (above and below the red band) are regions where the neutrino is vetoed.

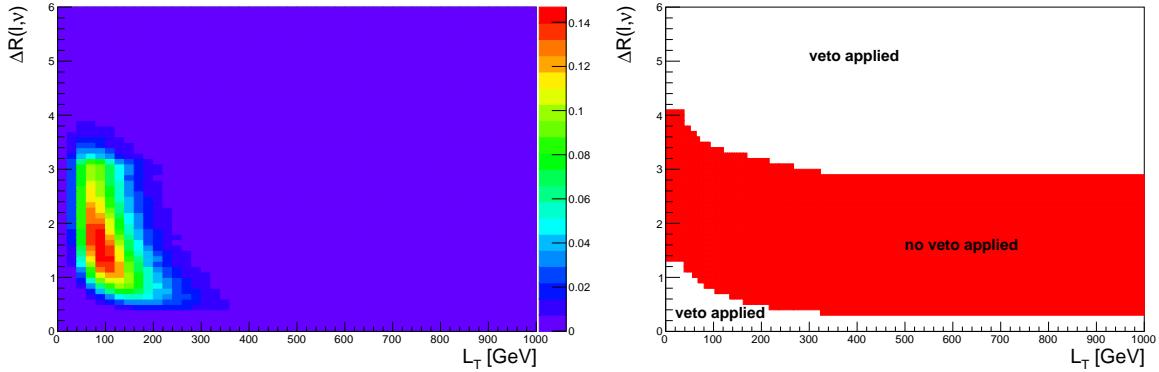


Figure 26: **Left:** A heatmap of occupancy for $\Delta R(\ell, \nu)$ vs L_T produced from simulated $t\bar{t}$ events (dileptonic final state) at generator-level is shown. ΔR between leptons and neutrinos is shown on the x-axis. L_T (scalar sum of lepton p_T) is shown on the y-axis. The colourbar on the right represents the fraction of events in the phase space. **Right:** The regions where vetoes are applied for the L_T constraint is shown. ΔR between leptons and neutrinos is shown on the x-axis. L_T (scalar sum of lepton p_T) is shown on the y-axis. The red band shows the region where the neutrino would not be vetoed. The white areas (above and below the red band) are regions where the neutrino is vetoed.

940 SM $t\bar{t}$ hypothesis. This is performed using the product of probabilities derived from certain distributions
 941 of variables from simulated $t\bar{t}$ events. The events in these distributions are obtained from an ATLAS
 942 simulation of generated $t\bar{t}$ events in the dileptonic final state. A normalised distribution of the mass of
 943 reconstructed top quarks, $m_{b\ell\nu}$, from a $t\bar{t}$ sample is generated to determine the probabilities $P_{m_{t_1}}$ and
 944 $P_{m_{t_2}}$ which correspond to the likelihood of the reconstructed top quarks under the SM $t\bar{t}$ hypothesis. The
 945 distribution is generated from reco-level leptons, generator-level neutrinos and reco-level jets matched in
 946 ΔR to generator-level b -quarks, therefore only filling the distribution with correct detector-level objects.
 947 For both reconstructed top quarks, $m(b\ell\nu)$ is calculated and interpolated (i.e. estimate the value of

948 the distribution for some value of the independent variable), via linear interpolation based on the two
 949 nearest bin centres, against the $m_{b\ell\nu}$ distribution which returns a weight value from 0 to 1, with higher
 950 values corresponding to a reconstructed top quark which has a mass close to that of a top quark from
 951 a $t\bar{t}$ system. This interpolation is done for both reconstructed top quarks, t_1 and t_2 , corresponding to
 952 probabilities $P_{m_{t_1}}$ and $P_{m_{t_2}}$. A similar method is used to determine $P_{\Delta E_x}$ and $P_{\Delta E_y}$, which corresponds
 953 to the likelihood of the reconstructed top quarks under the SM $t\bar{t}$ hypothesis. In this case, a weight
 954 distribution of $\Delta E_x = (p_{T,\nu_1})_x + (p_{T,\nu_2})_x - (E_T^{\text{miss}})_x$ based off simulated $t\bar{t}$ events is generated. In partic-
 955 ular, this distribution is generated using reco-level E_T^{miss} and generator-level neutrinos. The use of this
 956 distribution lies under the assumption that neutrinos are the dominant source of E_T^{miss} , and therefore,
 957 $(E_T^{\text{miss}})_x \approx (p_{T,\nu_1})_x + (p_{T,\nu_2})_x$ and $(E_T^{\text{miss}})_y \approx (p_{T,\nu_1})_y + (p_{T,\nu_2})_y$. This distribution is then used to inter-
 958 polate the value of ΔE_x and ΔE_y from our reconstructed neutrinos. This returns a weight value from 0
 959 to 1, with higher values corresponding to ΔE_x and ΔE_y (and in turn our reconstructed neutrino's p_T)
 960 closer to those observed in a $t\bar{t}$ event. It is expected that the ΔE_x and ΔE_y distributions have the same
 961 shapes, therefore only one is needed to be generated. In this case the the ΔE_x distribution was chosen.
 962 In Figure 27, the $m_{b\ell\nu}$ and ΔE_x distributions (generated from simulated $t\bar{t}$ events) are shown.

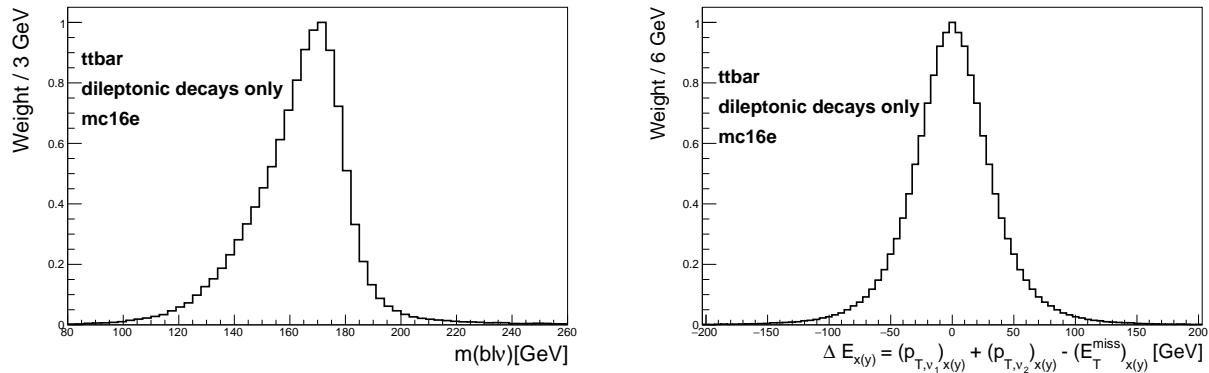


Figure 27: **Left:** $m_{b\ell\nu}$ distribution generated from simulated $t\bar{t}$ events, used to calculate $P_{m_{t_1}}$ and $P_{m_{t_2}}$ is shown. The $m_{b\ell\nu}$ distribution is shown by the black lined histogram. The mass of the $b\ell\nu$ system is shown on the x-axis. The corresponding weight of the $m_{b\ell\nu}$ distribution is shown on the y-axis. **Right:** ΔE_x distribution generated from simulated $t\bar{t}$ events, used to calculate $P_{\Delta E_x}$ and $P_{\Delta E_y}$ is shown. The ΔE_x distribution is shown by the black lined histogram. ΔE_x is shown on the x-axis. The corresponding weight of ΔE_x distribution is shown on the y-axis.

963 A final weight, $w_{2\nu SM} \in [0, 1]$, is then calculated by multiplying the four probabilities ($P_{m_{t_1}}$, $P_{m_{t_2}}$, $P_{\Delta E_x}$
 964 and $P_{\Delta E_y}$) described above. This final weight represents a total probability of the reconstructed top
 965 quarks under the SM $t\bar{t}$ hypothesis, and can be written as,

$$w_{2\nu SM} = P_{m_{t_1}} \times P_{m_{t_2}} \times P_{\Delta E_x} \times P_{\Delta E_y} \quad (4.13)$$

966 The $w_{2\nu SM}$ is calculated for each pair of reconstructed neutrinos (or reconstructed $t\bar{t}$ systems), with the
 967 maximum value being chosen as the final value for the event. In Figure 28, a flow chart that summarises
 968 the steps taken to calculate the $w_{2\nu SM}$ for an event is shown.

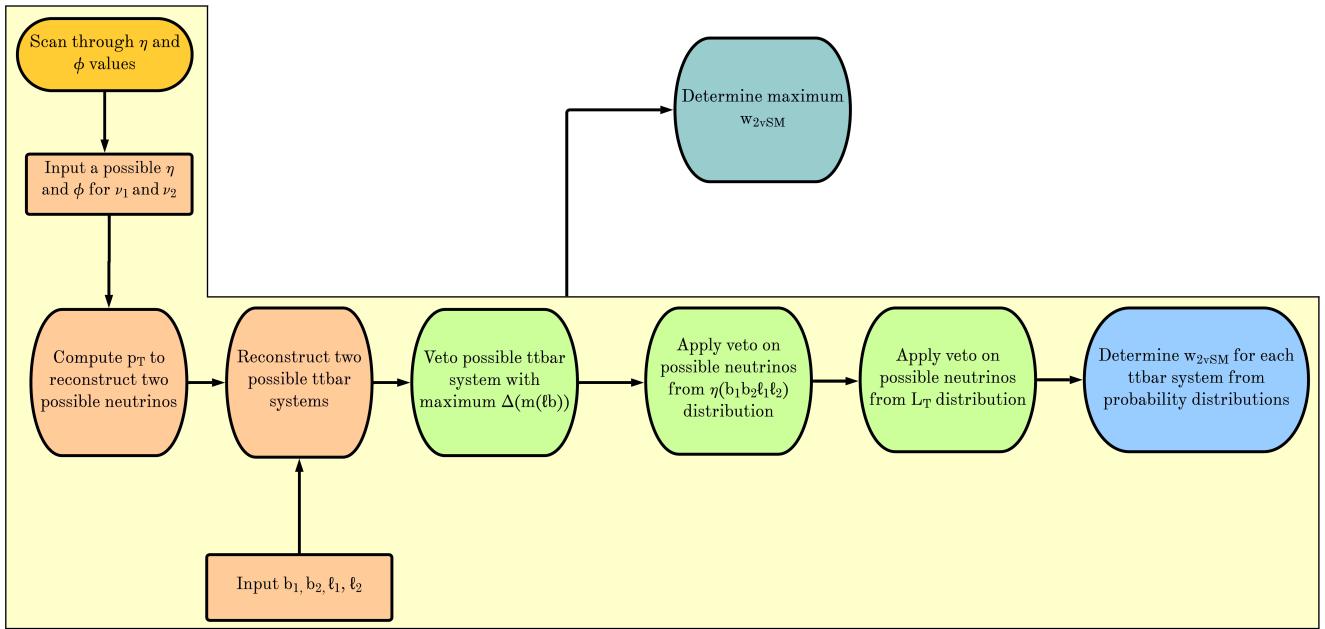


Figure 28: A flow chart that summarises the steps taken to calculate the $w_{2\nu SM}$ for an event is shown.

969 4.7.2 Boosted Decision Trees

970 Machine Learning techniques can be used to build multivariate discriminators that exploits information
 971 from many discriminators to form one discriminator with more discriminating power than the combination
 972 of those which it is built from. A BDT is a Machine Learning technique which classifies data in a dataset
 973 into different categories by iteratively applying binary cuts on features of the data (input variables, in
 974 the context of this analysis) [97]. The method in which a BDT combines many discriminators to build a
 975 single discriminator is called *boosting*. In boosting, weak discriminators are sequentially combined, where
 976 each classifier iteration is fitted to the difference between the observed and predicted values (residuals)
 977 of the training set from the previous step, such that the classifier performance improves [53]. A few
 978 concepts related to Machine Learning and BDTs that are used in this analysis, are described briefly
 979 in the following text. Performance metrics can be used to evaluate how well a classifier performs in a
 980 classification problem [70]. A performance metric used extensively in this analysis is the *accuracy* of a
 981 classifier. The accuracy is defined as the percentage of correct predictions for the test dataset (accuracy =
 982 $\frac{\text{correct number of predictions}}{\text{total number of predictions}}$). Machine Learning classifiers can be susceptible to learning a training dataset
 983 too well, in such a way as to negatively affect its performance on unseen data. This is known as *over-*
 984 *training*. Over-training occurs when noise or random fluctuations in the training dataset are learnt by
 985 the classifier [27]. Cross Validation [72] is a procedure used to evaluate a Machine Learning classifier.
 986 Cross validation gives an estimate on how the classifier is expected to perform on unseen data and it can
 987 be useful tool to protect against over-training. In this analysis we use a type of cross validation called,
 988 *k-fold* cross validation. In *k*-fold cross validation, the training dataset is randomly split up into *k* subsets,
 989 or folds, of approximately equal size. A fold is defined as a test dataset and the remaining *k*-1 folds are
 990 used to train the classifier. The classifier is then evaluated on the test set and a performance metric (or
 991 multiple) is evaluated. This procedure is performed once on each fold. Hyper-parameters are user-defined

parameters of a classifier that are govern the entire training process. Typical examples of hyper-parameters include the learning rate, the number of discriminators and the type of loss function to be minimised. The learning rate determines the step size at each iteration in determining the minimum of the loss function. Hyper-parameter optimisation is a process which aims to determine the best hyper-parameters for a classifier, based off some performance metric. In this analysis hyper-parameter optimisation is performed using a *grid search*. In a grid search, a user-defined list of hyper-parameter values are chosen for each hyper-parameter that one aims to optimise. The classifier is then trained using each permutation of hyper-parameters and determines the set of hyper-parameters in which the performance metric is maximised. BDTs are chosen to be used in this analysis, since they are not prone to over-training and perform well with minimal optimisation or tweaking of the hyper-parameters. A multi-layered sequential neutral network was tried, however, it was out-performed by a BDT. More specifically, Scikit-Learn’s `GradientBoostingClassifier` [81] was used.

4.7.3 Object-level BDT

The object-level BDT was trained on an $t\bar{t}$ sample simulated using the same generator, parton shower and to the same order of QCD as the $t\bar{t}$ sample described in Section 4.1.2 but with an orthogonal baseline selection of exactly 1 tight lepton with $p_T > 28$ GeV such that there is no overlap between this sample and the nominal $t\bar{t}$ sample used in the analysis. Additionally, jets in this sample are required to have $p_T > 20$ GeV. Jets are identified as b -tagged jets by the 77% DL1r working point. These baseline selections were chosen to mimic those used in the event selection of the analysis (outlined in Table 6). The leptons and b -jets used for training the object-level BDT are required to pass the aforementioned baseline selections. This $t\bar{t}$ sample was utilised in training the BDT to avoid using a subset of events from the MC samples used in the rest of the analysis, therefore maximizing the amount of generated events available to use in other parts of the analysis. The signal class is defined to consist of reconstructed ℓb systems (defined as the sum of the 4-vectors of a lepton and a b -tagged jet) originating from top quarks which are well matched to their truth counterparts. All possible combinations of ℓ and b -tagged jets are selected from the events. In particular, it is required that ΔR between the reconstructed and truth ℓb system is less than 0.05. An additional requirement is implemented such that the reconstructed lepton and the truth top quark have charges with the same sign (since $t \rightarrow b\ell^+\bar{\nu}_\ell$ and $\bar{t} \rightarrow \bar{b}\ell^-\nu_\ell$). The background class is defined to consist of all reconstructed ℓb systems which fail to pass the criteria for ℓb systems which are labelled as signal. These definitions for the signal and background classes ensure that the signal class consists of mostly ℓb systems originating from top quarks and the background class consists of mostly ℓb systems which do not originate from top quarks. The input variables chosen to be used in the object-level BDT are related to measurable quantities of ℓb systems. The optimum values for the hyper-parameters used were determined via the use of a grid-search (See Section 4.7.2) that determined the set of hyper-parameters which maximized the mean accuracy (based off 5 fold kfold cross-validation). After hyper-parameter optimisation, the mean accuracy of each fold increased from 0.76 to 0.77 ($\sim 1\%$ increase). Input variables can be assigned a score called *variable importance*, based on their usefulness on predicting a target variable (in this case, a signal or background event). The variable importance for any given input variable was obtained by computing the mean accuracy of the classifier, removing the input variable from training, retraining the classifier and computing the mean accuracy of this new classifier. The difference between

mean accuracies of the unaltered classifier and the retrained classifier (after removal of the input variable) gives us the variable importance of the given input variable. This method returns positive values for input variables which increase the mean accuracy of the classifier and negative values for input variables which decrease the mean accuracy of the classifier. Input variables with negative variable importances were completely removed from training. In Table 9, the input variables used for training the object-level BDT are shown. In Figure 29, normalised distributions of the input variables used in the object-level BDT, for

Input Variable	Description	Variable Importance
$m(\ell b)$	Invariant mass of the ℓb system	0.0025
$p_T(\ell b)$	p_T of the ℓb system	0.0005
$\Delta\eta(\ell, b)$	$\Delta\eta$ between the ℓ and b -tagged jet	0.0003
$\Delta\phi(\ell, b)$	$\Delta\phi$ between the ℓ and b -tagged jet	0.0003
$\Delta R(\ell, b)$	ΔR between the ℓ and b -tagged jet	0.0001

Table 9: A list of the input variables used in the object-level BDT, ordered by variable importance (descending, top to bottom) is shown.

1037

the signal and background classes, are shown. The input variables used in the object-level BDT show a

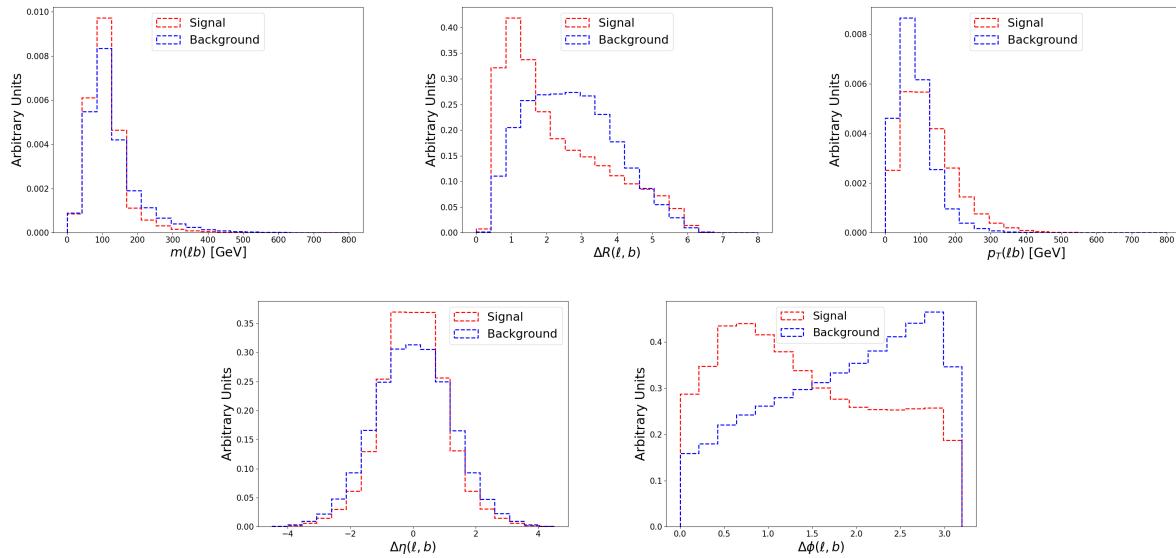


Figure 29: Normalised distributions of the input variables used in the object-level BDT (ordered from top left to bottom right via decreasing variable importance), for the signal and background classes are shown. **From top left to bottom right:** The invariant mass of the ℓb system, ΔR between the ℓ and b -tagged jet, the p_T of the ℓb system, $\Delta\eta$ between the ℓ and b -tagged jet, and $\Delta\phi$ between the ℓ and b -tagged jet. The red and blue dotted lined histograms represent the signal and background classes events (from the training set), respectively. These histograms are normalised to an area of 1. The input variable used in training is shown on the x-axis. The y-axis shows the relative number of events for the signal and background classes in arbitrary units.

1038

clear distinction between signal and background ℓb systems. The modelling of the input variables used in the object-level BDT can be checked by studying the agreement between data and simulation in the

1041 $t\bar{t}Z$ CR. In Figure 30, MC predictions for the input variables used in the object-level BDT in the $t\bar{t}Z$ CR are shown. Overall, there is good agreement between data and simulation for the input variables used in

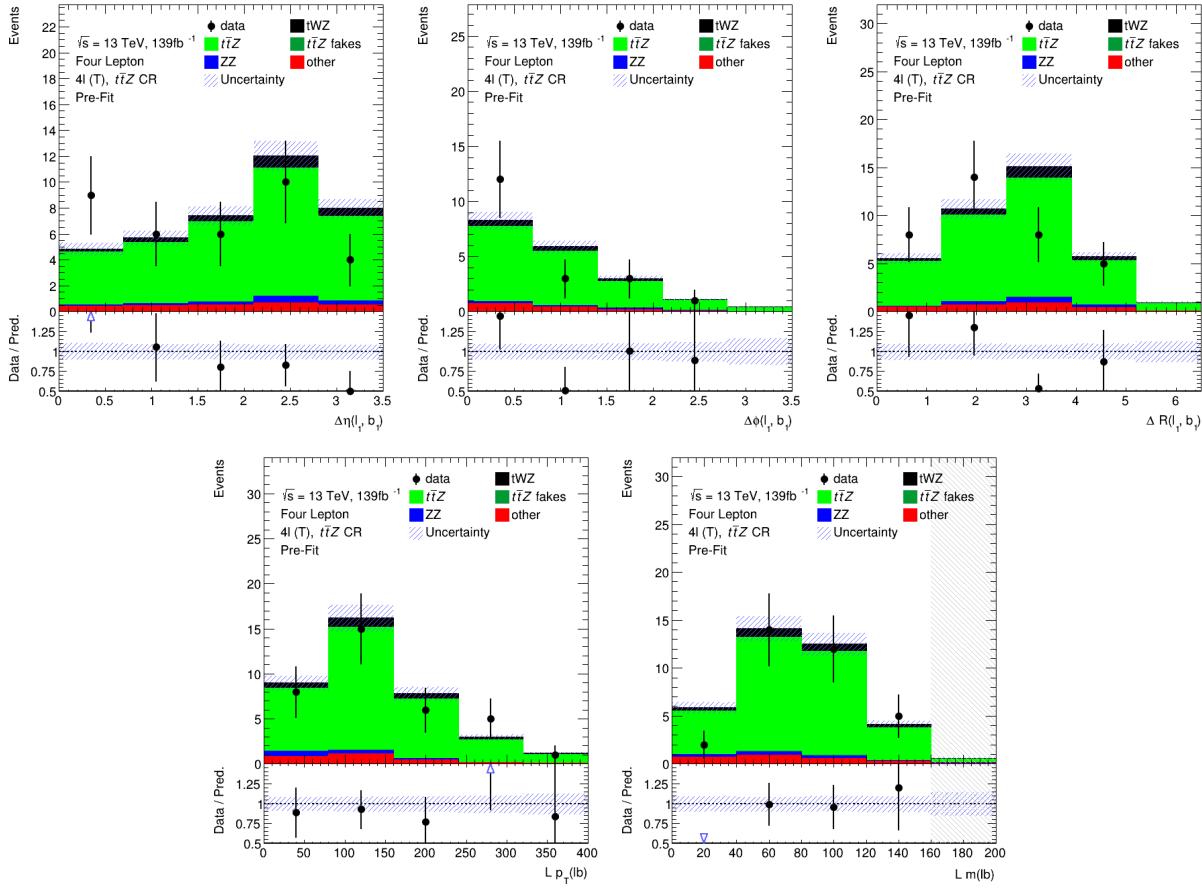


Figure 30: The expected number of events for the object-level BDT input variables (ordered from top left to bottom right via decreasing variable importance), in the $t\bar{t}Z$ CR, are shown. **From top left to bottom right:** $\Delta\eta$ between the lepton and b -jet of the leading ℓb system, $\Delta\phi$ between the lepton and b -jet of the leading ℓb system, ΔR between the lepton and b -jet of the leading ℓb system, p_T of the leading ℓb system, and the mass of the leading ℓb system. The data is given by the black points and the MC predictions for each process are given by the filled histograms. The vertical lines on the data points represent the statistical uncertainty in the data and the blue diagonally-lined bands represent the total (statistical and systematic added in quadrature) uncertainty. The lower panel in each plot shows the ratios of the data to the theoretical predictions. Bins with $\frac{\text{signal}}{\text{background}}$ greater than 0.1 are kept blinded. Blinded bins are shaded with black diagonal lines and their data points are omitted.

1042
 1043 the object-level BDT, in the $t\bar{t}Z$ CR. This suggests that the input variables used in the object-level BDT
 1044 are well-modelled and are reasonable to include as inputs to the object-level BDT. A final check can be
 1045 done to study the similarity of the ℓb systems present in the $t\bar{t}$ sample which are used for training the
 1046 object-level BDT, and the ℓb systems which are aimed to be identified using the object-level BDT. More
 1047 specifically, the study is done to ensure that the modelling of the ℓb systems in the $t\bar{t}$ sample are sufficiently
 1048 similar to those in the tWZ and $t\bar{t}Z$ samples (see Table 3). This is done to understand how well the
 1049 BDT (trained on ℓb systems in the $t\bar{t}$ sample) generalises to classifying ℓb systems in the analysis (tWZ
 1050 and $t\bar{t}Z$ samples). In Figure 31, normalised distributions of the input variables used in the object-level

BDT for the $t\bar{t}$, tWZ and $t\bar{t}Z$ samples, are shown. The distributions of the signal events from all three

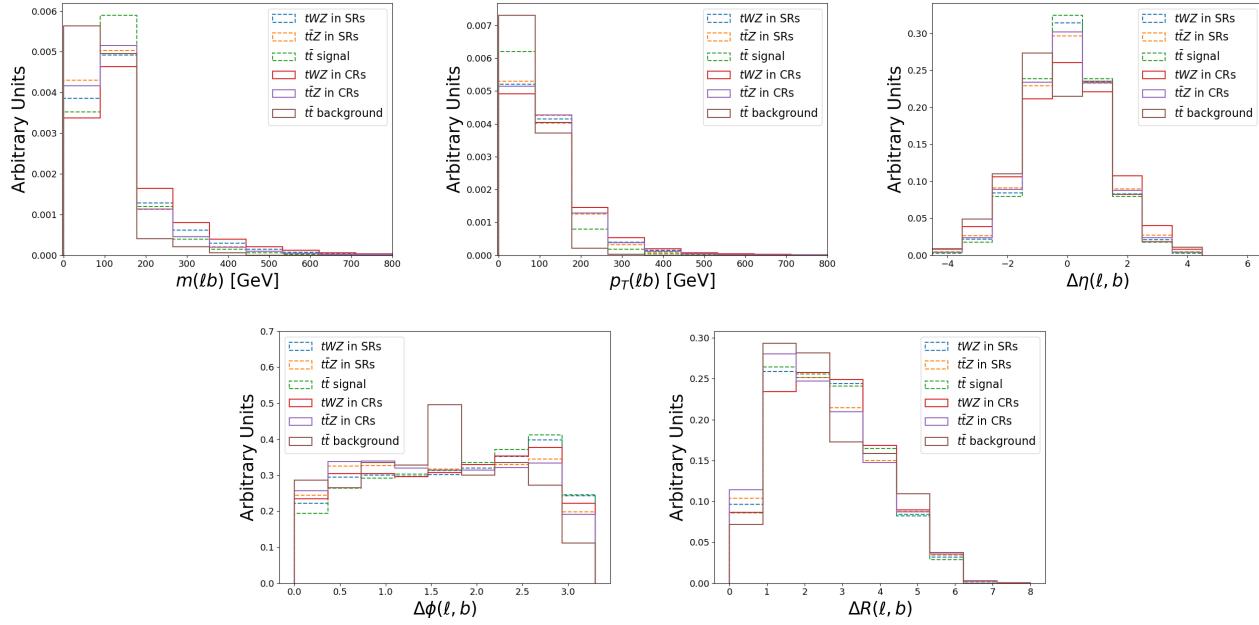


Figure 31: Normalised distributions of the input variables (ordered from top left to bottom right via decreasing variable importance) used in the object-level BDT for the $t\bar{t}$, tWZ and $t\bar{t}Z$ samples, are shown. **From top left to bottom right:** The invariant mass of the ℓb system, ΔR between the ℓ and b -tagged jet, the p_T of the ℓb system, $\Delta\eta$ between the ℓ and b -tagged jet, $\Delta\phi$ between the ℓ and b -tagged jet. The blue and orange dotted lined histograms represent events from the tWZ and $t\bar{t}Z$ samples in the SRs, respectively. The red and purple solid lined histograms represent events from the tWZ and $t\bar{t}Z$ samples in the CRs, respectively. The dotted lined green histograms and the solid lined brown histograms represents signal and background events from the $t\bar{t}$ sample, respectively. These histograms are normalised to an area of 1. The input variable used in training is shown on the x-axis. The y-axis shows the relative number of events in arbitrary units.

1051
 1052 processes for all of the input variables show minimal deviation between one another. This suggests that
 1053 the ℓb systems, that are classified as signal in training, are similar to those used in the analysis and are
 1054 therefore sufficient to include in training. There are substantially larger deviations in the distributions
 1055 of the background events (compared to the signal events) from all three processes for all of the input
 1056 variables. The deviations are especially noticeable in the $\Delta\phi(\ell, b)$ distribution, with a large excess of $t\bar{t}$
 1057 background events over the remaining processes. These deviations suggest that the use of the $t\bar{t}$ sample
 1058 in training the object-level BDT may be sub-optimal in classifying ℓb systems which do not originate
 1059 from top quarks. However, it still represents the best option available, since our other options involve
 1060 utilising of a subset of generated events used in the other parts of the analysis. This would result in a
 1061 smaller number of generated events used in the background prediction, leading to larger MC statistical
 1062 uncertainties. In Table 10, the hyper-parameters used in the object-level BDT is shown. The number
 1063 of events used in training for the signal and background classes were 49871 and 384152 respectively.
 1064 Imbalanced datasets can cause ML classifiers to ignore small classes while concentrating on classifying
 1065 large classes more accurately, which may result in the trained BDT performing sub-optimally. In order
 1066 to correct this dataset imbalance, it is ensured that the relative weighting of each event is such that the

Hyper-parameter	Value	Description
loss	deviance	The loss function to be optimised
criterion	friedman_mse	The function used to measure the quality of a split
n_estimators	200	The number of boosting stages to perform
learning_rate	0.1	The step size at each iteration during optimisation
max_depth	6	The maximum depth of the individual regression estimators
min_samples_split	2	The minimum number of samples (events) required to split an internal node
min_samples_leaf	1	The minimum number of samples (events) required to be at a leaf node
validation_fraction	0.1	The proportion of training data to set aside as validation set for early stopping
n_iter_no_change	20	Training terminates when the validation score does not improve in all of the previous

Table 10: A list of the hyper-parameters used in the object-level BDT is shown. Hyper-parameters not listed in this table use the default values as stated in the Scikit-learn Documentation[82].

1067 sum of the signal weights is equal to the sum of the background weights. In order to avoid over-training,
 1068 the BDT outputs to the training set and a test set can be studied. If over-training occurs, the BDT will
 1069 fit the data in the training set too closely, resulting in the BDT outputs of the training and test sets to
 1070 differ. In Figure 32 the normalised histograms of the training and test sets (extracted from fold 5 from a
 1071 5 fold kfold cross validation) for signal and background is shown.

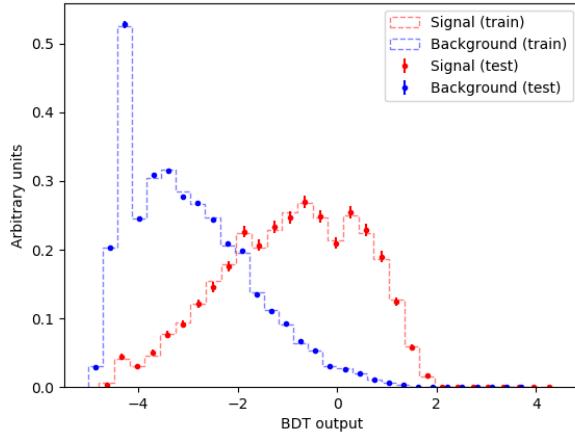


Figure 32: Normalised histograms of the object-level BDT discriminator output from the signal and background classes for the training and test sets from the 5th fold in a 5 fold kfold cross validation is shown. The output of the object-level BDT is shown on the x-axis and the relative number of events in arbitrary units, is shown on the y-axis. The training set for the signal class is shown by the red dotted histogram. The test set for the signal class is shown by the red points, with the total uncertainty represented by the vertical error bars. The training set for the background class is shown by the blue dotted histogram. The test set for the background class is shown by the blue points, with the total uncertainty represented by the vertical error bars.

1072 The shapes of the training and test sets for both signal and background agree within uncertainties in
 1073 the vast majority of bins. This is a good indicator that no over-training occurred, since it indicates
 1074 that statistical fluctuations (or noise) present in the training set was not learnt during training. Another
 1075 over-training check is performed using 5 fold kfold cross validation. To ensure that the BDT is not over-
 1076 training, it is ensured that the variance of the mean accuracy of each folds' test set in cross validation
 1077 is substantially small. This tells us that the BDT does not perform better on different subsets of the

1078 training set over another and it is therefore not prone to learning statistical fluctuations of a dataset
 1079 in training, which would result in the BDT not being able to generalise well to unseen datasets. For
 1080 the object-level BDT, a variance of 3.24×10^{-7} was calculated for the mean accuracies of each folds' test
 1081 set in cross validation. This small variance therefore provides further evidence that no over-training
 1082 occurred. The output of the object-level BDT is converted to an event-level variable to be used in the
 1083 event-level BDT. This variable, $\text{BDTScore}\left(\frac{\text{Best}}{\text{2nd Best}}\right)$, is defined as the ratios of the score of the best scoring
 1084 ℓb system to the 2nd best scoring ℓb system. The 2nd best scoring ℓb system in a tWZ event is expected
 1085 to be low, since there is only one ℓb system originating from a top quark. Thus $\text{BDTScore}\left(\frac{\text{Best}}{\text{2nd Best}}\right)$ is
 1086 expected to be large for tWZ events and closer to one for $t\bar{t}Z$ events, therefore providing discrimination
 1087 between tWZ and $t\bar{t}Z$. In Figure 33, normalised distributions of the signal and total background of the
 1088 $\text{BDTScore}\left(\frac{\text{Best}}{\text{2nd Best}}\right)$ variable in the tWZ OF SR, tWZ SF SR and $t\bar{t}Z$ CR are shown. The amount of

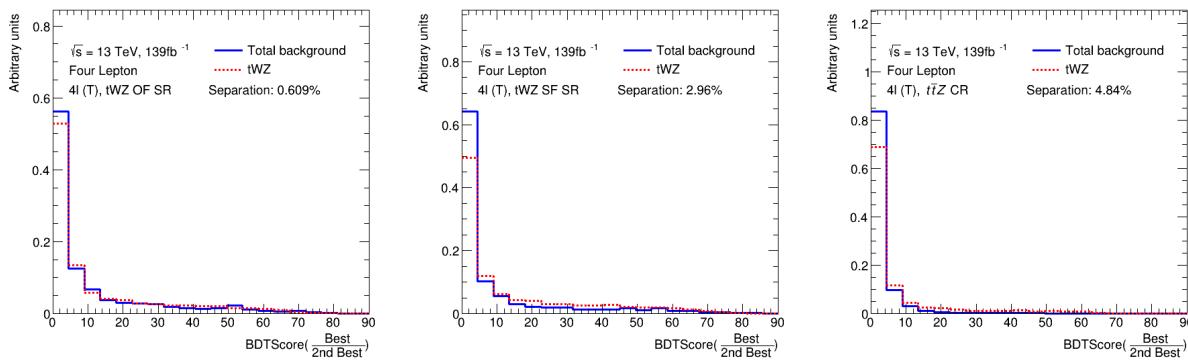


Figure 33: Normalised distributions of the signal and total background of the $\text{BDTScore}\left(\frac{\text{Best}}{\text{2nd Best}}\right)$ variable in the tWZ OF SR, tWZ SF SR and $t\bar{t}Z$ CR (left to right) are shown. The dotted red and solid blue lines represent the distributions of the signal and total background events respectively. These histograms are normalised to an area of 1. The x-axis shows the $\text{BDTScore}\left(\frac{\text{Best}}{\text{2nd Best}}\right)$ and the y-axis show the relative number of events in arbitrary units.

1088 discrimination can be quantified by the separation metric, which gives the percentage of the total area of
 1089 the distributions which do not overlap. A value of 1 indicates that the distributions are fully separated
 1090 (no overlap) and a value of 0 indicates that the distributions have no separation (fully overlapped). The
 1091 separation between signal and background for $\text{BDTScore}\left(\frac{\text{Best}}{\text{2nd Best}}\right)$ in the tWZ OF SR, tWZ SF SR and
 1092 $t\bar{t}Z$ CR are 0.609%, 2.96% and 4.84% respectively. The larger separation in the $t\bar{t}Z$ CR, compared
 1093 to the tWZ SRs, can be explained since there is a larger proportion of $t\bar{t}Z$ events (events with two
 1094 b -tagged jets) in this region, due to the baseline selection requirement of exactly two b -tagged jets. In a
 1095 similar way, the smaller separation in the two tWZ SRs can be explained by the tighter selection on the
 1096 number of b -tagged jets (exactly one) leading to regions which are enriched in only one ℓb system which
 1097 originates from a top quark. Using the $\text{BDTScore}\left(\frac{\text{Best}}{\text{2nd Best}}\right)$ variable in training in the event-level BDT
 1098 (see Section 4.7.4) improves the mean accuracy of the BDT. This tells us that the event-level BDT is
 1099 taking advantage of the discrimination between signal and background present in the $\text{BDTScore}\left(\frac{\text{Best}}{\text{2nd Best}}\right)$
 1100 variable. In an attempt to optimise the performance of the object-level BDT, signal events which are
 1101 pure in ℓb systems originating from top quarks are targeted for training the BDT. Similarly, background
 1102 events which are pure in ℓb systems which do not originate from top quarks are targeted for training
 1103

the BDT. This is done by studying the distribution of ΔR between the reconstructed ℓb system and the truth ℓb system ($\Delta R((lb)_{reco}, (lb)_{truth})$), and excluding ℓb systems from training which are moderately matched in ΔR to their truth counterparts, leaving well matched ℓb systems being labelled as signal and badly matched ℓb systems labelled as background. The ΔR range where ℓb systems are excluded from training is referred to as the exclusion region. In Figure 34, the distribution of ΔR between the reconstructed ℓb system and the truth ℓb system ($\Delta R((lb)_{reco}, (lb)_{truth})$) in the $t\bar{t}$ sample, along with the exclusion region, is shown. A large number of reconstructed ℓb systems have $\Delta R((lb)_{reco}, (lb)_{truth})$ at

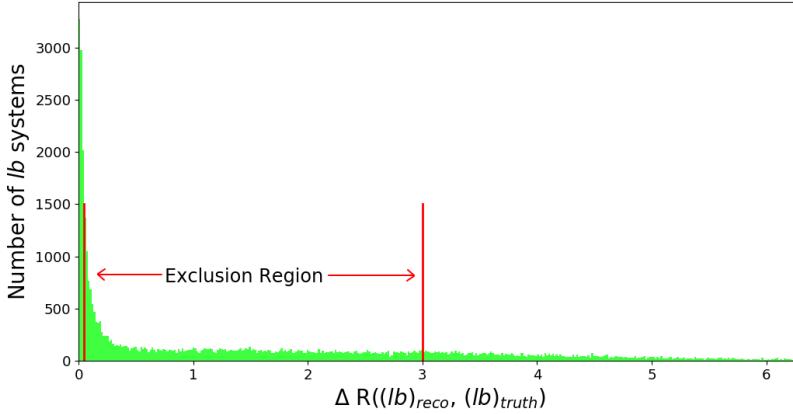


Figure 34: The distribution of ΔR between the reconstructed ℓb system and the truth ℓb system ($\Delta R((lb)_{reco}, (lb)_{truth})$) in the $t\bar{t}$ sample, along with the exclusion region, is shown. The ΔR distribution is shown in green. ΔR between the reconstructed ℓb system and the truth ℓb system ($\Delta R((lb)_{reco}, (lb)_{truth})$) is shown on the x-axis. The number of ℓb systems is shown on the y-axis. The exclusion region is shown between the vertical red lines situated at $\Delta R((lb)_{reco}, (lb)_{truth}) = 0.05$ and $\Delta R((lb)_{reco}, (lb)_{truth}) = 3.0$. Reconstructed ℓb systems with $\Delta R((lb)_{reco}, (lb)_{truth}) \leq 0.05$ are labelled as signal and reconstructed ℓb systems with $\Delta R((lb)_{reco}, (lb)_{truth}) \geq 3.0$ are labelled as background.

values near 0. These are matched (in ΔR) extremely well to truth ℓb systems originating from top quarks. Therefore our exclusion region is defined to be between $0.05 < \Delta R((lb)_{reco}, (lb)_{truth}) < 3.0$, such that all reconstructed ℓb systems with $\Delta R((lb)_{reco}, (lb)_{truth}) \leq 0.05$ are labelled as signal and reconstructed ℓb systems with $\Delta R((lb)_{reco}, (lb)_{truth}) \geq 3.0$ are labelled as background. All reconstructed ℓb systems with $0.05 < \Delta R((lb)_{reco}, (lb)_{truth}) < 3.0$ are excluded from training. The performance of the object-level BDT with and without the exclusion region can be compared by studying the discrimination between signal and background events in the $BDTScore(\frac{\text{Best}}{2nd \text{ Best}})$ variable (object-level output converted to an event-level variable to be used in the event-level BDT) for both object-level BDTs. In Figure 35, normalised distributions of $BDTScore(\frac{\text{Best}}{2nd \text{ Best}})$ using the object-level BDT without the exclusion region (see Figure 34) for the tWZ OF SR, tWZ SF SR and $t\bar{t}Z$ CR are shown. The separation metrics can be compared between the $BDTScore(\frac{\text{Best}}{2nd \text{ Best}})$ variable in the tWZ OF SR, tWZ SF SR and $t\bar{t}Z$ CR for the object-level BDT with (Figure 33) and without (Figure 35) the exclusion region by taking the absolute difference between the two values in each region. The differences are 0.31%, 0.37% and 0.36% for the tWZ OF SR, tWZ SF SR and $t\bar{t}Z$ CR, respectively. These differences are minimal and the object-level BDT with the exclusion region outperforms the object-level BDT without the exclusion region in the tWZ SF SR. Due to the small differences in performance between the two BDTs, the BDT with the exclusion region was

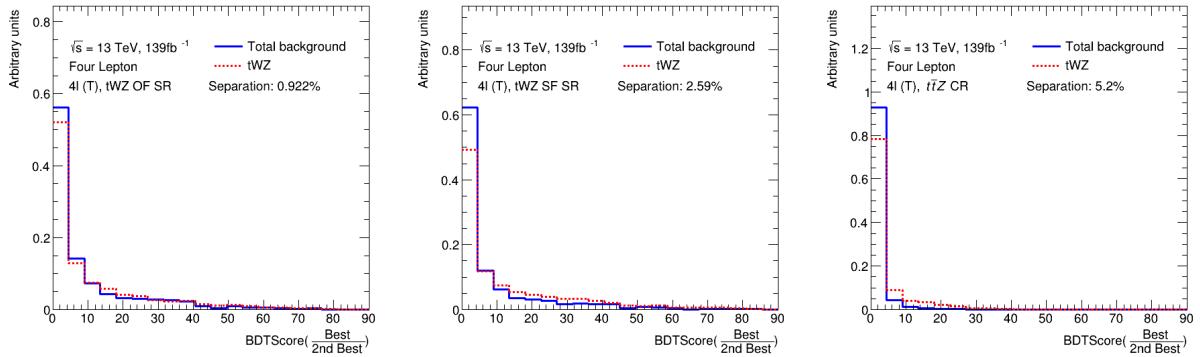


Figure 35: Normalised distributions of $\text{BDTScore}(\frac{\text{Best}}{\text{2nd Best}})$ using the object-level BDT without the exclusion region (see Figure 34) for the tWZ OF SR, tWZ SF SR and $t\bar{t}Z$ CR (left to right) are shown. The dotted red and solid blue lines represent the distributions of the signal and total background events respectively. These histograms are normalised to an area of 1. The x-axis shows the $\text{BDTScore}(\frac{\text{Best}}{\text{2nd Best}})$ and the y-axis show the relative number of events in arbitrary units.

1127 chosen to be kept.

1128 4.7.4 Event-level BDT

1129 The event-level BDT is used to distinguish between signal and its major background events, $t\bar{t}Z$ and
 1130 ZZ . The key difference between the object-level BDT and the event-level BDT is that while the former
 1131 exploits information associated with ℓb systems and thus distinguishes between ℓb systems, the event-level
 1132 BDT exploits information based on the entirety of the event and thus distinguishes between events. The
 1133 event-level BDT was trained on 50% of the tWZ MC sample's events for the signal class and similarly,
 1134 50% of the $t\bar{t}Z$ and ZZ MC sample's events were used for the background class. The input variables used
 1135 to train the BDT are chosen on the basis that they are somewhat uncorrelated from one another and
 1136 show some discrimination between tWZ and $t\bar{t}Z$. Similarly to the object-level BDT, the optimum values
 1137 for the hyper-parameters used were determined via a grid-search (See Section 4.7.2) that determined the
 1138 set of hyper-parameters which maximized the mean accuracy (based off 5 fold kfold cross-validation).
 1139 After hyper-parameter optimisation, the mean accuracy of each fold (determined from 5 fold kfold cross
 1140 validation) increased from 0.72 to 0.74 ($\sim 3\%$ increase). The variable importance of each input variable
 1141 was computed in the same way as described for the object-level BDT (See Section 4.7.3). In Table 11, the
 input variables used for training the event-level BDT are shown. In Figure 36, normalised distributions

Input Variable	Description	Variable Importance
$2\nu\text{SM}$	Maximum weight from the $2\nu\text{SM}$ algorithm (See Section 4.7.1)	0.029
HT	Scalar sum of jet p_T	0.016
LT	Scalar sum of lepton p_T	0.011
$\sum p_T(b-jet)$	Scalar sum of b -tagged jet p_T	0.006
$\text{BDTScore}(\frac{\text{Best}}{\text{2nd Best}})$	Ratio of the best to the 2nd best scoring ℓb system from the object-level BDT	0.006
$\Delta\eta(\ell_{1,\text{non}-Z}, \ell_{2,\text{non}-Z})$	$\Delta\eta$ between the two leptons, not coming from a Z candidate	0.005

Table 11: A list of the input variables used in the event-level BDT, ordered by variable importance (descending, top to bottom) is shown.

1142

1143 of the input variables used in the event-level BDT, for the signal and background classes are shown.

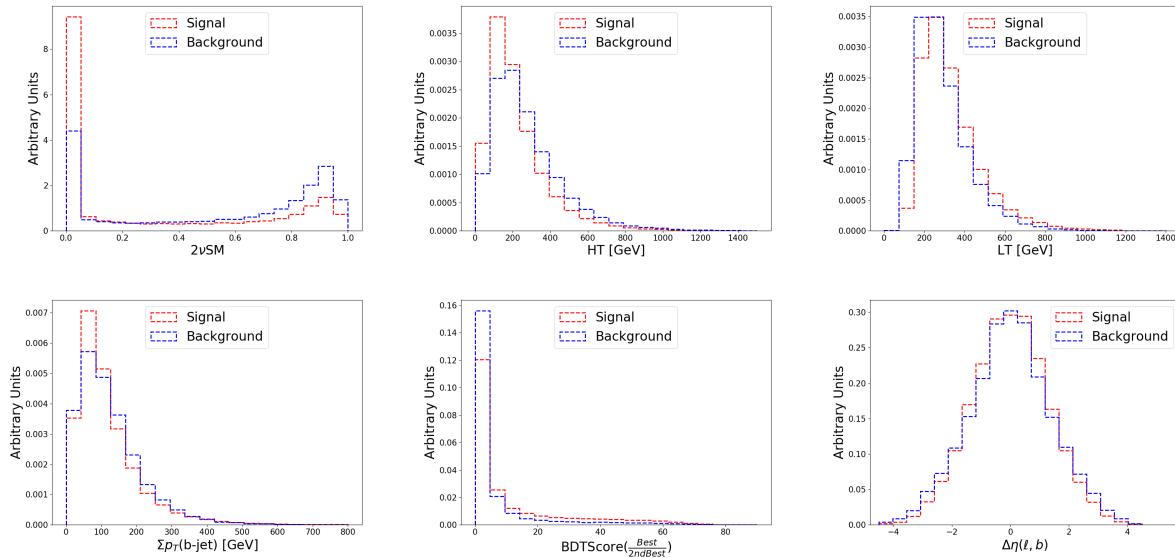


Figure 36: Normalised distributions of the input variables used in the event-level BDT (ordered from top left to bottom right via decreasing variable importance), for the signal and background classes are shown. **From top left to bottom right:** The output weight from the $2\nu\text{SM}$ algorithm, the scalar sum of jet p_T , the scalar sum of lepton p_T , the ratio of the best scoring ℓb system to the 2nd best scoring ℓb system from the output of the object-level BDT, and $\Delta\eta$ between the two leptons, not coming from a Z candidate. The red and blue dotted lined histograms represent the signal and background classes events, respectively. These histograms are normalised to an area of 1. The input variable used in training is shown on the x-axis. The y-axis shows the relative number of events for the signal and background classes in arbitrary units.

1144 The input variables used in the event-level BDT show a clear distinction between signal and background
 1145 events. In particular the output weight from the $2\nu\text{SM}$ algorithm shows the most discrimination. When
 1146 determining which input variables to use to train the event-level BDT, the output weight from $2\nu\text{SM}$ was
 1147 shown to provide the most sizeable boost in performance of the BDT. Surprisingly, the least important
 1148 input variable, $\Delta\phi$ between the non- Z lepton system (leptons not originating from a Z -candidate) and the
 1149 leading b -tagged jet, seem to discriminate well between signal and background. A possible explanation
 1150 for its low ranking variable importance is due to it being relatively highly correlated with many of the
 1151 other input variables used in the BDT. The modelling of the input variables used in the event-level BDT
 1152 can be checked by referring to the expected number of events of data and simulation in control regions
 1153 where they are defined. Note that certain input variables which are ill-defined in certain regions (e.g.
 1154 $\Delta\eta(\ell_{1,\text{non-}Z}, \ell_{2,\text{non-}Z})$) in the ZZb CR, as all leptons originate from a b -jet in this region) will not be
 1155 shown. In Figure 37, MC predictions for the input variables used in the event-level BDT in the $t\bar{t}Z$ CR are
 1156 shown. The deviations between data and simulation, across all input variables used in the $t\bar{t}Z$ CR, in all
 1157 but three bins are within expected uncertainties. In Figure 38, MC predictions for the input variables used
 1158 in the event-level BDT in the ZZb CR are shown. The deviations between data and simulation, across all
 1159 input variables used in the ZZb CR, are within expected uncertainties. In Figure 39, MC predictions for the input variables used
 1160 in the event-level BDT in the $(tWZ)_{\text{fake}}$ CR are shown. The deviations between data and simulation, across all input variables used in the $(tWZ)_{\text{fake}}$ CR, in all but one bin are within

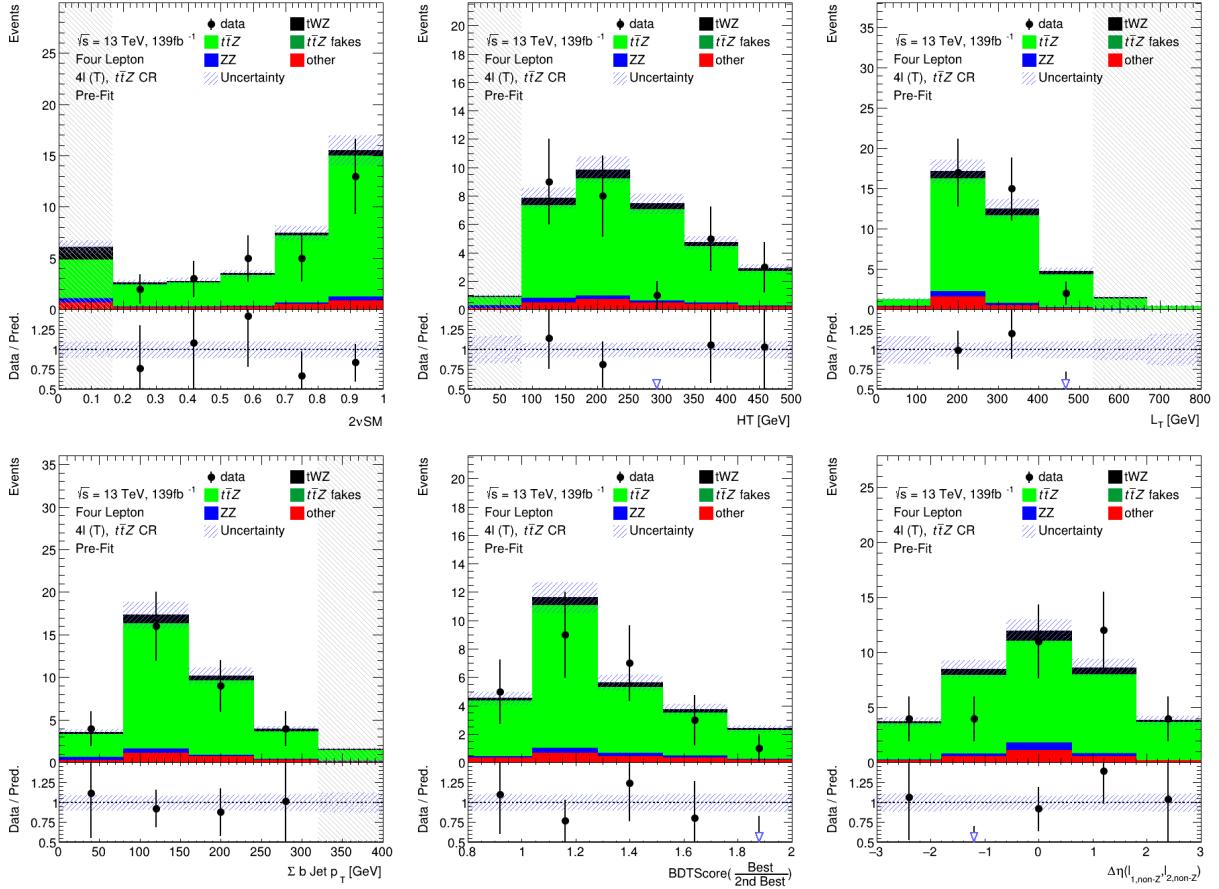


Figure 37: The expected number of events for the event-level BDT input variables (ordered from top left to bottom right via decreasing variable importance), in the $t\bar{t}Z$ CR, are shown. **From top left to bottom right:** The output weight from the 2ν SM algorithm, the scalar sum of jet p_T , the scalar sum of lepton p_T , the sum of b -tagged jet p_T , ratio of the best scoring ℓb system to the 2nd best scoring ℓb system from the output of the object-level BDT, and $\Delta\eta$ between the two leptons, not coming from a Z candidate. The data is given by the black points and the MC predictions for each process are given by the filled histograms. The vertical lines on the data points represent the statistical uncertainty in the data and the blue diagonally-lined bands represent the total (statistical and systematic added in quadrature) uncertainty. The lower panel in each plot shows the ratios of the data to the theoretical predictions. Bins with $\frac{\text{signal}}{\text{background}}$ greater than 0.1 are kept blinded. Blinded bins are shaded with black diagonal lines and their data points are omitted.

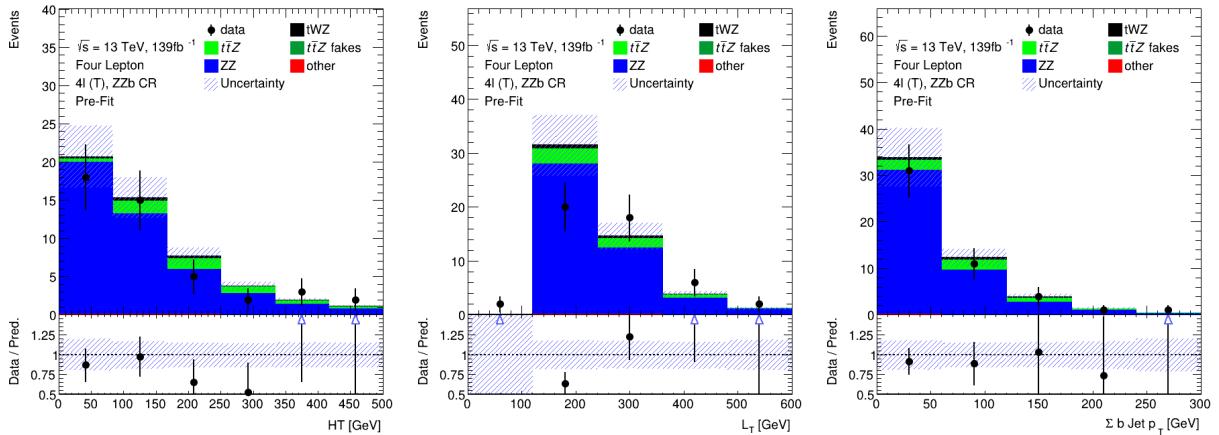


Figure 38: The expected number of events for the event-level BDT input variables (ordered from top left to bottom right via decreasing variable importance), in the ZZb CR, are shown. **From left to right:** The scalar sum of jet p_T , the scalar sum of lepton p_T , and the sum of b -tagged jet p_T . The data is given by the black points and the MC predictions for each process are given by the filled histograms. The vertical lines on the data points represent the statistical uncertainty in the data and the blue diagonally-lined bands represent the total (statistical and systematic added in quadrature) uncertainty. The lower panel in each plot shows the ratios of the data to the theoretical predictions. Bins with $\frac{\text{signal}}{\text{background}}$ greater than 0.1 are kept blinded. Blinded bins are shaded with black diagonal lines and their data points are omitted.

1162 expected uncertainties. Overall, the vast majority of predictions between data and simulation in the bins
 1163 of the event-level BDT distributions in the $t\bar{t}Z$ CR, ZZb CR and $(tWZ)_\text{fake}$ CR, are within the expected
 1164 uncertainties. Therefore, these input variables are well-modelled and reasonable to include as inputs to
 the event-level BDT. In Table 12, the hyper-parameters used in the event-level BDT are shown. Since

Hyper-parameter	Value	Description
loss	deviance	The loss function to be optimised
criterion	friedman_mse	The function used to measure the quality of a split
n_estimators	200	The number of boosting stages to perform
learning_rate	0.1	The step size at each iteration during optimisation
max_depth	6	The maximum depth of the individual regression estimators
min_samples_split	2	The minimum number of samples (events) required to split an internal node
min_samples_leaf	1	The minimum number of samples (events) required to be at a leaf node
validation_fraction	0.1	The proportion of training data to set aside as validation set for early stopping
n_iter_no_change	20	Training terminates when the validation score does not improve in all of the previous

Table 12: A list of the hyper-parameters used in the event-level BDT is shown. Hyper-parameters not listed in this table use the default values as stated in the Scikit-learn Documentation[82].

1165
 1166 the training is performed on $t\bar{t}Z$ and ZZ events for the background class, it is ensured that the relative
 1167 weighting of these events are such that they mimic the amount of $t\bar{t}Z$ and ZZ expected to be present in
 1168 the regions where the BDT discriminator is used. This is done by applying normalization weights to each
 1169 event, defined as,

$$W = \frac{\sigma \mathcal{L} \text{weight(MC)}}{\text{totalWeight(MC)}} \quad (4.14)$$

1170 where σ is the cross section of the process, \mathcal{L} is the integrated luminosity, weight(MC) is the weight
 1171 assigned to the event by the MC event generator and totalWeight(MC) is the sum of those weights for

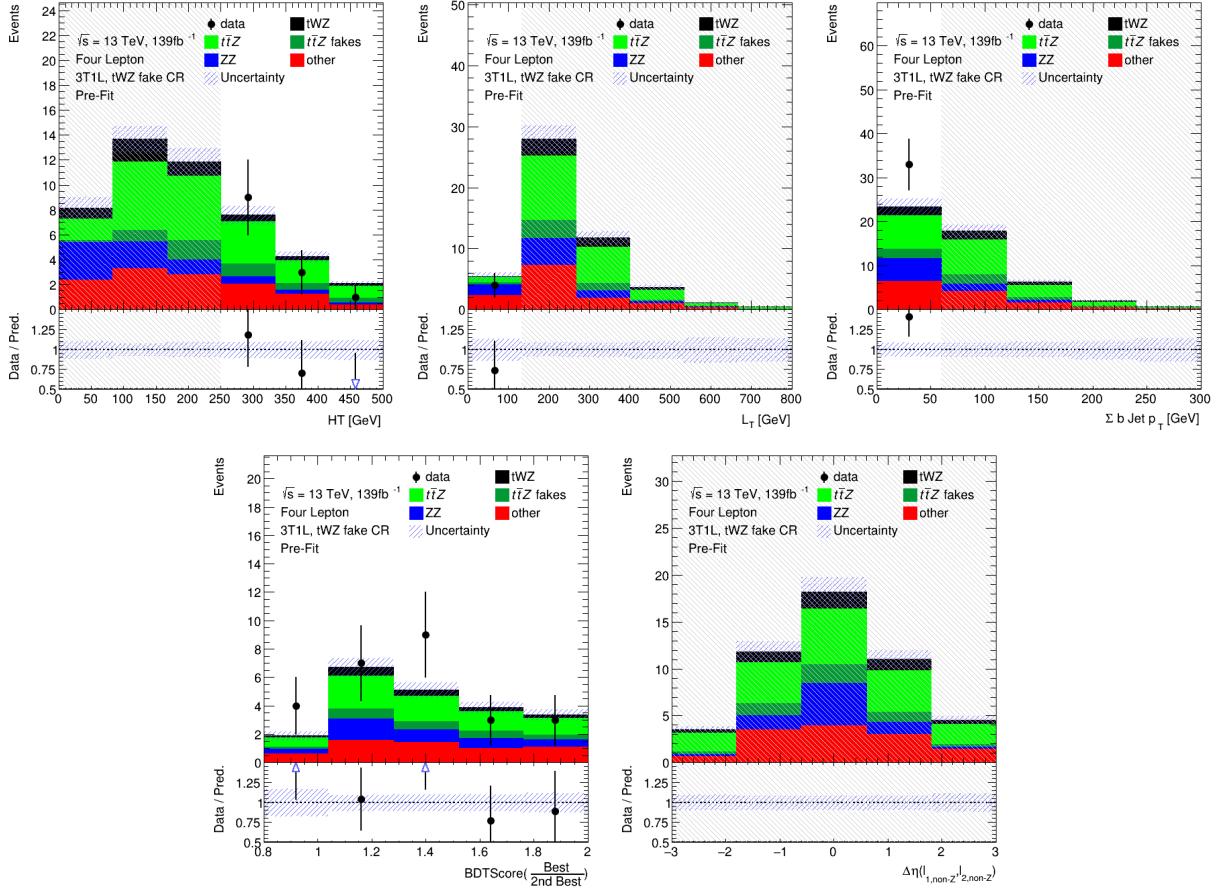


Figure 39: The expected number of events for the event-level BDT input variables (ordered from top left to bottom right via decreasing variable importance), in the $(tWZ)_{\text{fake}} \text{CR}$, are shown. **From top left to bottom right:** The output weight from the $2\nu\text{SM}$ algorithm (See Section 4.7.1), the scalar sum of jet p_T , the scalar sum of lepton p_T , the sum of b -tagged jet p_T , ratio of the best scoring ℓb system to the 2nd best scoring ℓb system from the output of the object-level BDT (See Section 4.7.3), and $\Delta\eta$ between the two leptons, not coming from a Z candidate. The data is given by the black points and the MC predictions for each process are given by the filled histograms. The vertical lines on the data points represent the statistical uncertainty in the data and the blue diagonally-lined bands represent the total (statistical and systematic added in quadrature) uncertainty. The lower panel in each plot shows the ratios of the data to the theoretical predictions. Bins with $\frac{\text{signal}}{\text{background}}$ greater than 0.1 are kept blinded. Blinded bins are shaded with black diagonal lines and their data points are omitted.

1172 all the generated events. The number of events used in training for the signal and background classes
 1173 were 41066 and 22608 respectively. Similarly to the object-level BDT, there is a dataset imbalance. This
 1174 imbalance is corrected for, in the same way as before with the object-level BDT, by ensuring that the
 1175 relative weighting of each event is such that the sum of the signal weights is equal to the sum of the
 1176 background weights. In order to avoid over-training, in the same way that was performed for the object-
 1177 level BDT (See Section 4.7.3), the BDT outputs to the training set and a test set can be studied. In
 1178 Figure 40 the normalised histograms of the training and test sets (extracted from fold 5 from a 5 fold kfold
 cross validation) for signal and background is shown. The shapes of the training and test sets for both

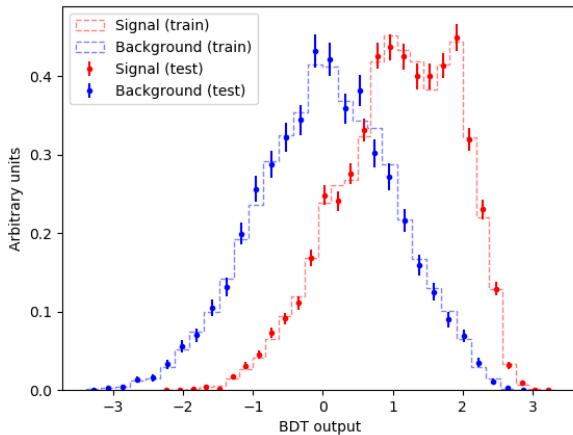


Figure 40: Normalised histograms of the event-level BDT discriminator output from the signal and background classes for the training and test sets from the 5th fold in a 5 fold kfold cross validation are shown. The output of the event-level BDT is shown on the x-axis and the relative number of events (normalised to have an area of 1, in arbitrary units) is shown on the y-axis. The training set for the signal class is shown by the red dotted histogram. The test set for the signal class is shown by the red points, with the total uncertainty represented by the vertical error bars. The training set for the background class is shown by the blue dotted histogram. The test set for the background class is shown by the blue points, with the total uncertainty represented by the vertical error bars.

1179 signal and background agree within uncertainties in the vast majority of bins. This is a good indicator
 1180 that no over-training occurred, since it indicates that statistical fluctuations (or noise) present in the
 1181 training set was not learnt during training. As with the object-level BDT, another over-training check
 1182 is performed, by ensuring that the variance of the mean accuracy of each folds' test set in a 5 fold kfold
 1183 cross validation is sufficiently small. This indicates that fluctuations in features from different training
 1184 sets are not learnt by the BDT. For the event-level BDT, a variance of 0.00026 was calculated for the
 1185 mean accuracies of each folds' test set in cross validation, providing further evidence that no over-training
 1186 occurred. In Figure 41, normalised distributions of the signal and total background of the event-level
 1187 BDT discriminator output in the tWZ OF SR, tWZ SF SR and $t\bar{t}Z$ CR, are shown. The event-level
 1188 BDT discriminates well between signal and background events in the tWZ OF SR, tWZ SF SR and $t\bar{t}Z$
 1189 CR, with separations of 9.14%, 9.45% and 19.5%, respectively.
 1190

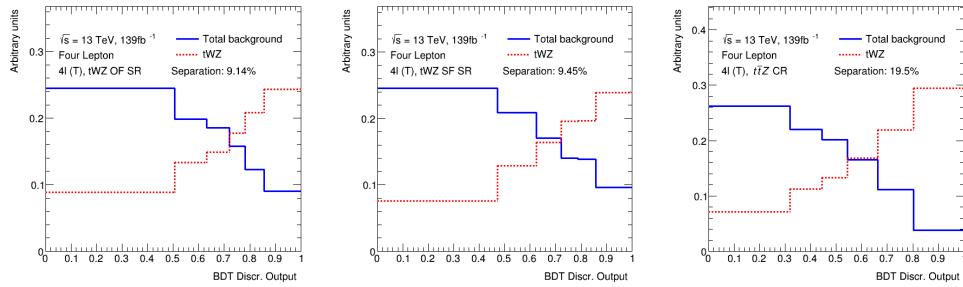


Figure 41: Normalised distributions of the signal and total background of the event-level BDT discriminator output in the tWZ OF SR, tWZ SF SR and $t\bar{t}Z$ CR (left to right) are shown. The dotted red and solid blue lines represent the distributions of the signal and total background events respectively. These histograms are normalised to an area of 1. The x-axis shows the event-level BDT discriminator output and the y-axis shows the relative number of events in arbitrary units.

1191 4.8 Systematic Uncertainties

1192 The final results are obtained with a maximum likelihood fit (described in Section 4.9.1) in which normal-
 1193 isations and shapes of distributions are allowed to vary to account for uncertainties of both experimental
 1194 and theoretical origin. Systematic uncertainties are allowed to vary within a pre-determined envelope
 1195 (consisting of two MC templates) which defines the bounds within which the systematic can vary. In this
 1196 section, the methods used to determine the variations that form the inputs to the fit are described.

1197 4.8.1 Experimental uncertainties

1198 In this section, the experimental systematic uncertainties are outlined.

- 1199 • **Luminosity:**

1200 The 2015–2018 luminosity estimate of 139 fb^{-1} has a relative uncertainty of 3%. The uncertainty
 1201 associated with the luminosity is obtained using the LUCID-2 detector [21], which is the primary
 1202 luminosity monitor for ATLAS. This systematic uncertainty affects all processes modelled using MC
 1203 simulations.

- 1204 • **Pile-up modelling:**

1205 An uncertainty related to the SFs used for MC to account for differences in pile-up distributions
 1206 between MC and data is applied. This uncertainty is obtained by re-scaling the $\langle\mu\rangle$ value in data
 1207 by 1.00 and 1/1.18 corrections are only applied to MC.

- 1208 • **Jet vertex tagger:**

1209 Uncertainties associated to the JVT are applied via the `JetJvtEfficiency` package [20] which
 1210 account for the residual contamination from pile-up jets after pile-up suppression and the MC
 1211 generator choice [14].

- 1212 • **Heavy- and light-flavour tagging:**

1213 An additional uncertainty is assigned to account for the extrapolation of the b -tagging efficiency
 1214 measurement from the p_T region used to determine the correction factors to regions with higher p_T .

1215 • **Electron identification efficiency:**

1216 Uncertainties associated with the electron efficiency SFs arise from the reconstruction, ID, isolation
1217 and trigger efficiencies [18]. They correct for the efficiency difference between data and MC [13] and
1218 are measured with a "tag-and-probe" method in $Z \rightarrow e^+e^-$ and $J/\psi \rightarrow e^+e^-$ events. The infor-
1219 mation on the correlation of the different components of the systematic uncertainties are provided
1220 for all efficiency measurements. The default correlation model for the uncertainties is used, which
1221 provides one up/down variation for each of the SF components separately [18, 19].

1222 • **Muon identification efficiency:**

1223 As for electrons, SFs obtained from $Z \rightarrow \mu^+\mu^-$ and $J/\psi \rightarrow \mu^+\mu^-$ events are applied to correct for the
1224 differences between data and MC in the muon ID, isolation and trigger efficiencies [16]. Uncertainties
1225 on these SFs are applied as up/down variations of the nominal SFs for each component.

1226 **4.8.2 Theoretical uncertainties**

1227 In this section, the theoretical systematic uncertainties are outlined.

1228 • **$t\bar{t}Z$ background:**

1229 An overall normalization uncertainty of 10% is considered for the $t\bar{t}Z$ background. Two generic
1230 shape systematic uncertainties are considered for the $t\bar{t}Z$ background. They are constructed (see
1231 Section 4.8.3) by either applying a linear or triangular interpolation to up and down variations which
1232 are defined to be $\pm 20\%$ from the nominal $t\bar{t}Z$ background. These linear and triangular variations
1233 are applied to distributions used in the maximum likelihood fit (described in Section 4.9.1).

1234 • **ZZ background:**

1235 An overall normalization uncertainty of 30% is considered for the ZZ background.

1236 • **$t\bar{t}H$ background:**

1237 An overall normalization uncertainty of 20% is considered for the $t\bar{t}H$ background.

1238 • **tZq background:**

1239 An overall normalization uncertainty of 14% is considered for the tZq background.

1240 • **$t\bar{t}Z$ fake background:**

1241 An overall normalization uncertainty of 50% is considered for the $t\bar{t}Z$ fake background.

1242 • **other background processes:**

1243 The 'other' background consists of many processes which have minimal but non-negligible contribu-
1244 tion in the signal regions (See Table 7). An overall normalization uncertainty of 30% is considered
1245 for the 'other' background processes.

1246 • **tWZ :**

1247 A modelling uncertainty on tWZ is considered by comparing the nominal sample (using the DR1

1248 scheme) and a DR2 sample. Two generic shape systematic uncertainties are applied to the tWZ
1249 signal. They are constructed (see Section 4.8.3) by either applying a linear or triangular interpolation
1250 to up and down variations which are defined to be $\pm 20\%$ from the nominal tWZ signal.

1251 4.8.3 Generic shape systematic uncertainties

1252 The quantity which is measured from the maximum likelihood fit, the expected signal strength of tWZ
1253 ($\mu(tWZ)$), is defined as the ratio of the cross section of tWZ , $\sigma(tWZ)$, to the SM cross section of tWZ ,
1254 $\sigma_{SM}(tWZ)$ (described in more detail in Section 4.9.1). It is evident that the tetralepton channel is sta-
1255 tistically limited, as seen by the small number of expected events in the control plots shown previously
1256 in Section 4.5. Therefore it is expected that the uncertainty on $\mu(tWZ)$ is dominated by statistical
1257 uncertainty and that the impact of shape systematic uncertainties will be negligible in comparison. In
1258 this analysis, not all of the systematic uncertainties have been estimated, as this involves time consum-
1259 ing analysis of extra event samples, especially those related to the modelling of background processes.
1260 However, as described above, these missing systematic uncertainties are likely to have a negligible im-
1261 pact. To check this assumption, generic systematic uncertainties are generated as a proxy for the missing
1262 systematic uncertainties for the modelling of processes. If these generic shape systematic uncertainties
1263 have negligible impact on the measurement of the expected $\mu(tWZ)$, it would give us confidence that the
1264 missing systematic uncertainties will not greatly alter our results when they are eventually added. An
1265 envelope consisting of two MC templates is constructed that represents the bounds in which the system-
1266 atic can vary in the fit. One with the nominal MC template increased by 20% on its normalisation and
1267 the other with the nominal MC template decreased by 20% on its normalisation. The templates are then
1268 modified from their original shape either by doing linear interpolation (from the leftmost-up variation to
1269 the rightmost-down variation) or triangular interpolation (shape is set to zero at the higher and lower
1270 parts and extends towards the envelope in the middle). The linear and triangular interpolation is done
1271 using TRF’s `ForceShape` option [76], which alter the original templates (as described above). In Figure
1272 42 the envelope before and after the shape change, for both the linear and triangular interpolations, for
1273 the $t\bar{t}Z$ background in the $t\bar{t}Z$ CR is shown.

1274 Both a linear and a triangular generic systematic uncertainty are considered for tWZ and the most
1275 dominant background process, $t\bar{t}Z$.

1276 4.9 Analysis Pipeline and **TRExFitter**

1277 For this analysis, we make use of CERN’s high energy data analysis framework, `ROOT`, which is written in
1278 `C++`, through `Python` using `PyROOT`. `Python` is used extensively in many fields of science (not limited to
1279 physics and data science) due to its simplicity and ongoing support by the communities which utilize it.
1280 In `PyROOT` users are able to access the full `ROOT` functionality within `Python`. More specifically, `PyROOT`
1281 provides `Python` bindings for `ROOT`. To produce all control plots (including fit statistics, e.g. limit,
1282 significance, μ_{tWZ}), we used `TRExFitter` (tag: `TRExFitter-00-04-13`), which is a framework for binned
1283 template profile likelihood fits [92].

1284

1285 The analysis pipeline starts with sample derivations (data and simulation from ATLAS with certain

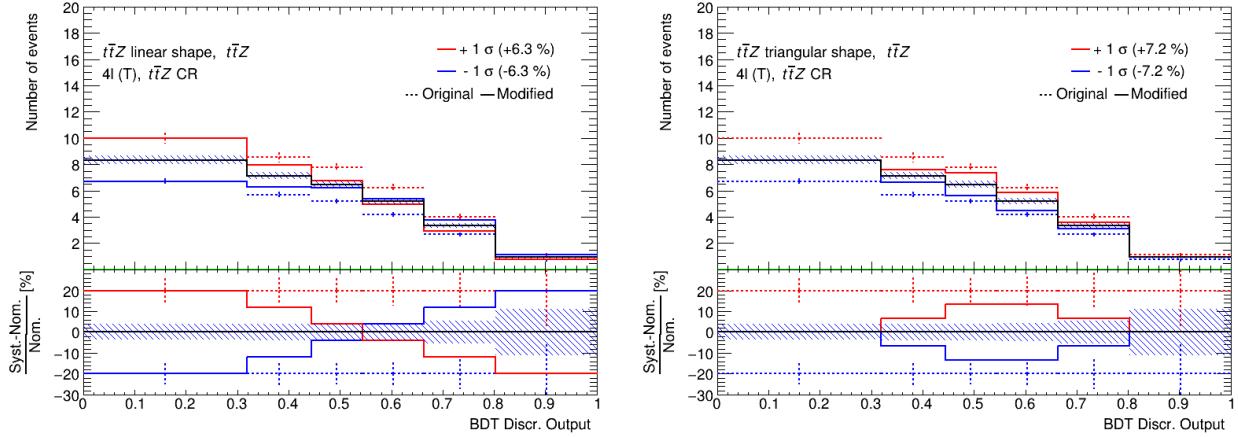


Figure 42: The $t\bar{t}Z$ generic systematic uncertainty before and after linear (left) and triangular (right) interpolation in the $t\bar{t}Z$ CR is shown. The output from the event-level BDT shown on the x-axis. In the upper panel, the number of events is shown on the y-axis. In the lower panel, the difference between the systematic variation and the nominal template, divided by the nominal template, is shown on the y-axis. The nominal $t\bar{t}Z$ template is shown by the solid black lined histogram, with the diagonal lined bands representing its total uncertainty. The templates of the upper and lower envelopes, before modification, is given by the dotted red and blue lined histograms respectively. The templates of the upper and lower envelopes, after modification, is given by the solid red and blue lined histograms respectively, with the vertical dotted lines representing its total uncertainty.

selection criteria applied in order to reduce its file size such that it is more manageable to work with) being submitted to the grid for dataset production. This applies cuts and selections to the already reduced derivations and produces datasets with trees containing variables (e.g. scale factors, variables, MC truth flags) that will be used at future stages in the analysis. These datasets are then read by PyROOT where the events are looped over, before being written to ROOT files as input to **TRExFitter**. The Python scripts are used to define the different regions and apply the final cuts and selections outlined in Table 6. In addition to this purpose, they are used to train the two BDTs and to produce the output from these trained BDTs. As each event is looped over, the cuts and selection criteria are checked for the given event and is either discarded (if the event does not pass the selection criteria), or gets written to a ROOT file (if the event passes the selection criteria) corresponding to the MC sample and Run 2 data-set which it belongs to. These ROOT files contain all variables, weights and scale factors (corresponding to an event) which we wish to use in **TRExFitter**. **TRExFitter** then takes these files as input, runs a maximum likelihood fit and produces relevant plots (e.g. expected number of events, pull plots) and statistical parameters (e.g. limit, significance, μ_{tWZ}).

4.9.1 Fitting Procedure

Maximum likelihood fits can be used to calculate the probability of data being consistent with a given hypothesis. In this analysis, the hypothesis is represented by the signal strength of tWZ production, $\mu(tWZ)$, defined as,

$$\mu(tWZ) = \frac{\sigma(tWZ)}{\sigma_{SM}(tWZ)} \quad (4.15)$$

where $\sigma_{(tWZ)}$ expected cross section of tWZ and $\sigma_{SM}(tWZ)$ is the SM cross section of tWZ . The maximum likelihood fit is performed by finding the set of values for the parameters which maximize a likelihood function or equivalently by minimizing the negative logarithm of the likelihood function. In this context, the likelihood function is constructed from probability distributions of the expected number of events in each bin of a distribution and the nuisance parameters. The data in a given bin of a distribution is expected to follow a Poisson probability distribution representing its expected number of events. Nuisance parameters represent the systematic uncertainties that could affect the estimation of signal and background but are not directly related to our final measurement. The likelihood function, $L(\mu, \theta)$, used in analyses involving binned histograms of signal and background events is therefore constructed as a product of Poisson probability terms for all bins (N_{bins}) and probability density functions that define the nuisance parameters, and is given by [23],

$$L(\mu, \theta) = \prod_i^{N_{bins}} \frac{(\mu s_i(\vec{\theta}) + b_i(\vec{\theta}))^{n_i}}{n_i!} e^{-(\mu s_i(\vec{\theta}) + b_i(\vec{\theta}))} \prod_{\theta \in \vec{\theta}} \rho(\theta) \quad (4.16)$$

where μ is the signal strength of the process of interest, s_i and b_i represent the expected number of signal and background events in bin i of the distribution respectively, θ represents a nuisance parameter contained in the set of all nuisance parameters ($\vec{\theta}$) and $\rho(\theta)$ represents the probability density functions which define the nuisance parameter's distribution. In analyses with multiple regions, the likelihood function contains all bins across the regions. A fit across all regions in the tetralepton channel is performed to determine the sensitivity of tWZ in this channel. Plots shown prior to the fit are referred to as *pre-fit* and those shown after the fit are referred to as *post-fit*. In this analysis, the *mixed data and MC* fit setup [93] is used. This is done to obtain the most accurate prediction of the expected results while keeping the signal regions blinded. For this setup, first a background-only fit to the control regions using data is performed to estimate the nuisance parameters. Then these estimates are used to construct a modified Asimov dataset in the signal regions. Finally, the fit is performed using data in the control regions and the aforementioned modified Asimov data-set in the signal regions. In these fits, the POI is $\mu(tWZ)$. The POI is ultimately the quantity to be measured and it is set as a *free parameter* in the fit. This means that during the fitting procedure, $\mu(tWZ)$ is unconstrained and can take any value in the fit. The nuisance parameters are assigned to the systematic uncertainties outlined in Section 4.8. Furthermore, a gamma (γ) nuisance parameter for a bin is added to the likelihood function if the statistical uncertainty in the bin exceeds 0.1% of its nominal value. A sample's shape and normalisation nuisance parameter is pruned (removed from the limit/fit) if the impact of the systematic uncertainty on the POI is less than 0.01. Pruning is done per sample and per region on the shape and normalisation uncertainties for samples. An auto-binning algorithm, `TransfoD` [31, 32], was used to define the binning. This aims to maximise $\frac{\text{signal}}{\text{background}}$ in each bin. Furthermore, it aims to avoid defining bins with a low number of events. The auto-binning algorithm works by scanning through the bins of the given distribution with user-defined bins of equal width, and merges bins until a certain fraction of $\frac{\text{signal}}{\text{background}}$ events is obtained. The merging threshold, MT , is defined as,

$$MT = z_s \frac{n_s}{N_s} + z_b \frac{n_b}{N_b} \quad (4.17)$$

1339 where n_s and n_b are the number of signal and background events in the merging bin, respectively. The total
1340 number of signal and background events is given by N_s and N_b , respectively. User-defined parameters,
1341 z_s and z_b , control the maximum fraction of signal and background events in each bin with the condition,
1342 $z_s + z_b = \text{number of bins}$. A bin is formed when $MT \geq 1$. To characterise the sensitivity and associated
1343 uncertainty of our measurement of $\mu(tWZ)$, we compute two metrics: the expected significance (Z_μ^{\exp})
1344 and the expected upper limit (μ_{up}^{\exp}). In this context, the expected significance can be interpreted as,
1345 the probability that the expected signal is due to a background fluctuation. The expected significance,
1346 Z_μ^{\exp} , is related to the p -value. The p -value can be defined as the probability, under the assumption of a
1347 given hypothesis (in the context of this analysis, this would refer to the background-only hypothesis), of
1348 finding data of equal or greater incompatibility with the predictions of the hypothesis [41]. The expected
1349 significance can be written in terms of the p -value (p) by,

$$Z_\mu^{\exp} = \Phi^{-1}(1 - p) \quad (4.18)$$

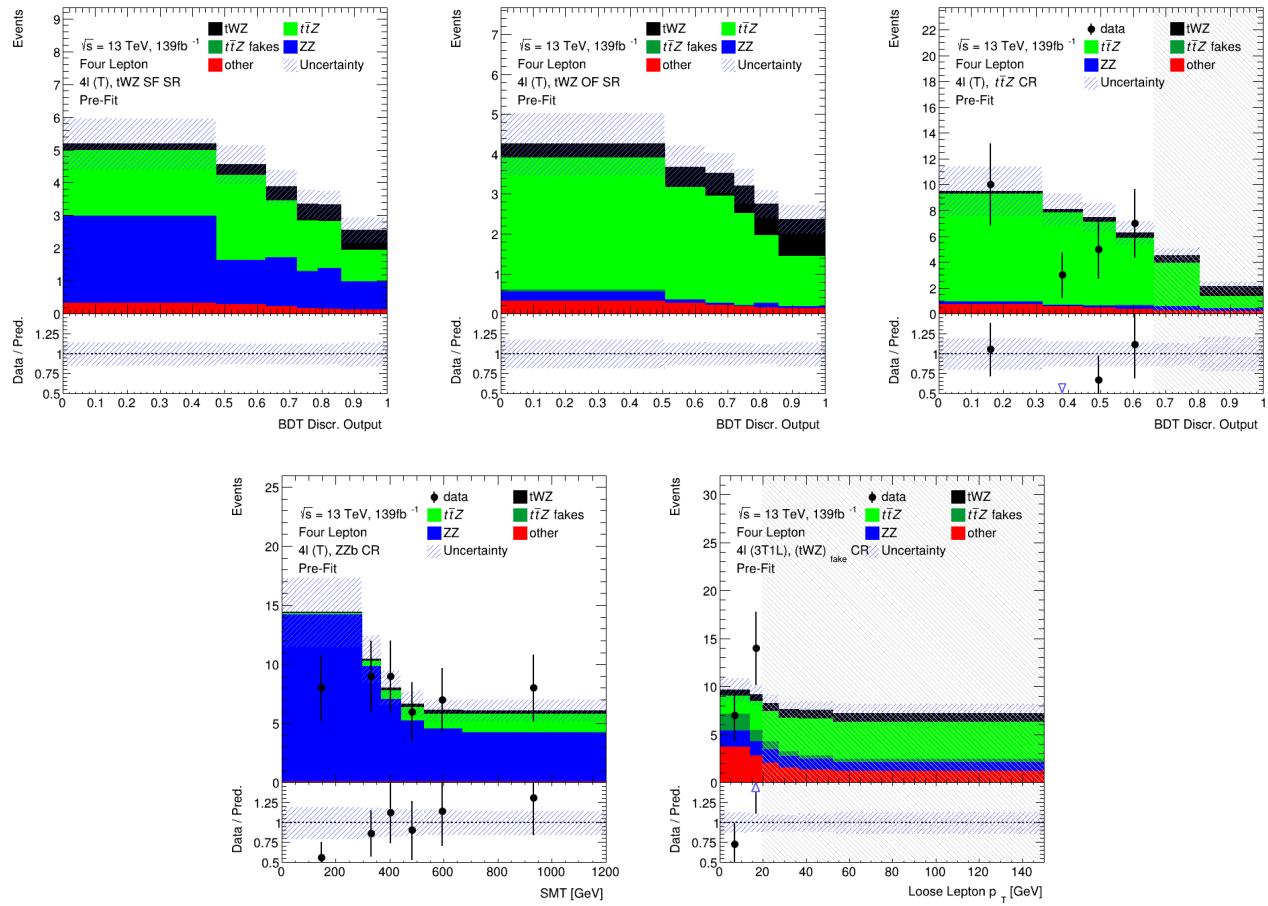
1350 where Φ^{-1} is the inverse of the cumulative function of the standard Gaussian distribution. This is defined
1351 such that a Gaussian distributed variable found Z_μ^{\exp} standard deviations above its mean has an area
1352 under its rightmost tail equal to p . Larger values indicate lower probabilities and smaller values indicate
1353 higher probabilities. Particle physicists have adopted a standard to define the significance necessary
1354 for evidence and discovery of a particular particle or phenomena. A 3σ (corresponding to a background
1355 fluctuation probability of $\approx 10^{-3}$) significance is considered to be evidence and a 5σ (corresponding to a
1356 background fluctuation probability of $\approx 10^{-7}$) is considered to be a discovery. The expected upper limit
1357 is a single-sided interval test statistic, associated with the parameter of interest (POI) in the maximum-
1358 likelihood fit ($\mu(tWZ)$, in our case). In this context, the expected upper limit can be understood in the
1359 following way: consider running an ensemble of MC toy experiments, each with their own confidence
1360 interval (a range of possible values for $\mu(tWZ)$). An expected upper limit, at some fixed percentage $x\%$
1361 (or *confidence level*), can be determined from this ensemble. The expected upper limit tells us that, $x\%$
1362 of the toy MC experiment's confidence intervals will contain the true value of $\mu(tWZ)$. A commonly used
1363 percentage in particle physics is 95%, which we adopt for this analysis. In particle physics, this is referred
1364 to as the *CLs Method* [41]. The test statistic derived from the CLs method can be calculated 'brute
1365 force' by running these MC toy experiments, however this is very CPU intensive. Asymptotic formulae
1366 are able to describe the underlying CLs test statistic distributions under certain approximations [10].
1367 Instead of running toy MC experiments, we use asymptotic formulae to perform the CLs method, which
1368 considerably reduces computation time. The significance and upper limits which are calculated in this
1369 analysis are given a prefix of 'expected' in order to indicate that these are results from a blinded analysis.

1370 4.10 Results

1371 In the section, an expected upper limit and an expected significance are set on the signal strength of tWZ .
1372 This is performed for the current analysis in the tetralepton channel as well as for a combined analysis
1373 across the trilepton and tetralepton channels. The trilepton analysis was performed as an independent
1374 study by Benjamin Warren (UCT) [95]. Note that throughout this section, all signal regions remain
1375 blinded.

1376 4.10.1 Tetralepton Channel

1377 In Figure 43, the pre-fit distributions of the expected number of events as a function of the variables used in the likelihood fit in each region are shown. In Figure 44, the post-fit distributions of the expected



1378
1379 Figure 43: Pre-fit distributions (blinded) of variables used in the fit are shown. **From top left to bottom right:** The event-level BDT Disc. Output in the tWZ SF SR, the event-level BDT Disc. Output in the tWZ OF SR, the event-level BDT Disc. Output in the $t\bar{Z}$ CR, $SMT = \sum p_T(\ell) + \sum p_T(jet) + E_T^{\text{miss}}$ in the ZZb CR, and p_T (loose lepton) in the $(tWZ)_\text{fakes}$ (3T1L) CR. The data is given by the black points and the MC predictions for each process are given by the filled histograms. The vertical lines on the data points represent the statistical uncertainty in the data and the diagonally-lined bands represent the total (statistical and systematic added in quadrature) uncertainty. The lower panel in each plot shows the ratios of the data to the theoretical predictions. Bins in the CRs with $\frac{\text{signal}}{\text{background}}$ greater than 0.1 are kept blinded. Blinded bins in the CRs are shaded with black diagonal lines and their data points are omitted. The plots in the tWZ OF SR and tWZ SF SR are kept blinded by omitting the data points.

1380 number of events as a function of the variables used in the likelihood fit in each region are shown. In
1381 Table 13, the expected number of events (after the fit) for each sample in each region is shown. The
1382 expected upper limit of tWZ in the tetralepton channel is measured as, $\mu_{up}^{\text{exp}} = 1.61^{+2.35}_{-1.16}$. The expected
1383 significance of tWZ in the tetralepton channel is measured as, $Z_\mu^{\text{exp}} = 1.44\sigma$. The best-fit value of the
1384 signal strength, $\mu(tWZ) = \frac{\sigma(tWZ)}{\sigma(tWZ)_{SM}}$, from the likelihood fit is measured as, $\mu(tWZ) = 1.91^{+0.95}_{-0.82}$. The
1385 best fit value of the signal strength is inconsistent with 1 by 9%. This indicates a disagreement between data and the post-fit model within post-fit model uncertainties. This may suggest that there is some

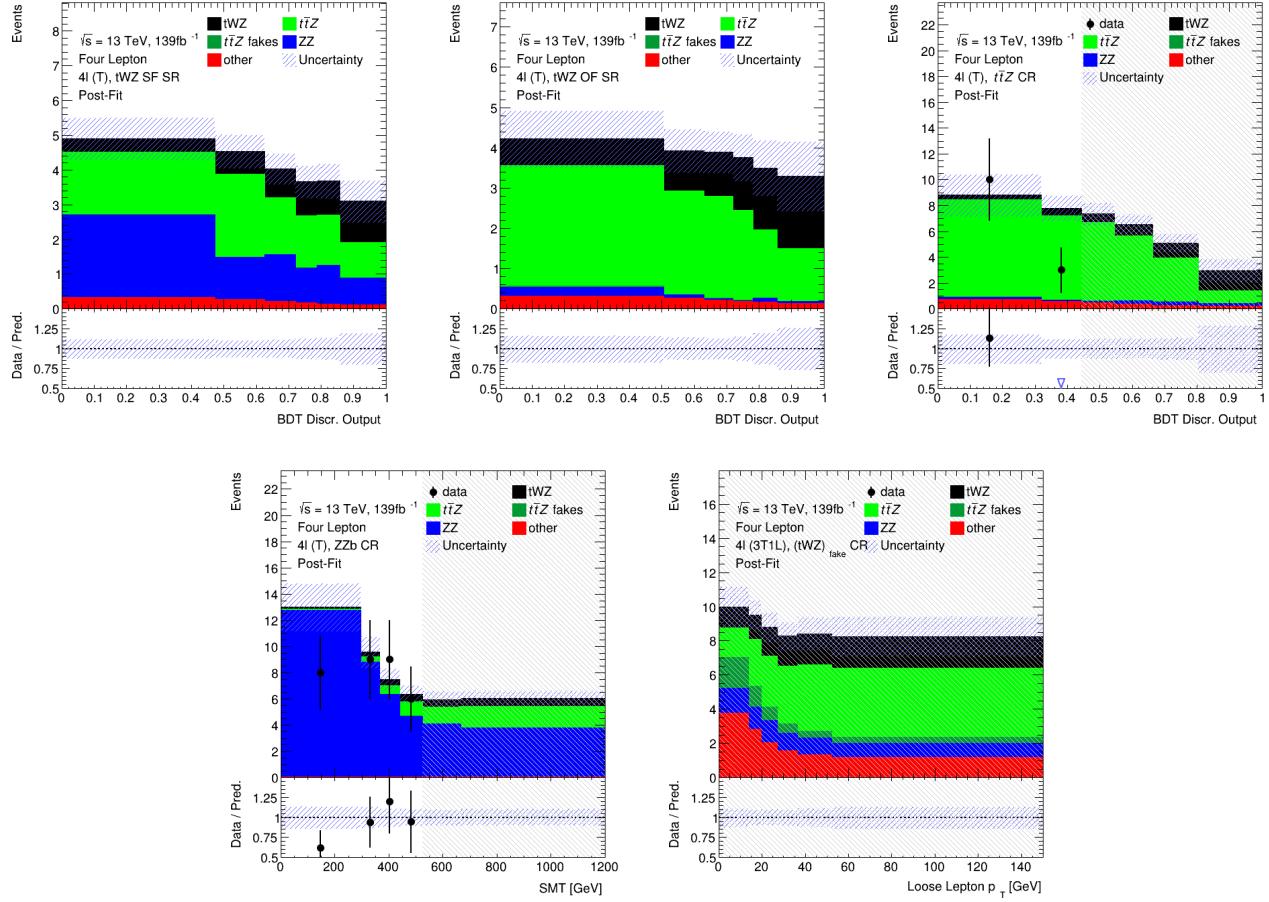


Figure 44: Post-fit distributions (blinded) of variables used in the fit are shown. **From top left to bottom right:** The event-level BDT Disc. Output in the tWZ SF SR, the event-level BDT Disc. Output in the tWZ OF SR, the event-level BDT Disc. Output in the $t\bar{Z}$ CR, $SMT = \sum p_T(\ell) + \sum p_T(jet) + E_T^{\text{miss}}$ in the ZZb CR, and p_T (loose lepton) in the $(tWZ)_{\text{fake}}$ (3T1L) CR. The data is given by the black points and the MC predictions for each process are given by the filled histograms. The vertical lines on the data points represent the statistical uncertainty in the data and the diagonally-lined bands represent the total (statistical and systematic added in quadrature) uncertainty. The lower panel in each plot shows the ratios of the data to the theoretical predictions. Bins in the CRs with $\frac{\text{signal}}{\text{background}}$ greater than 0.1 are kept blinded. Blinded bins in the CRs are shaded with black diagonal lines and their data points are omitted. The plots in the tWZ OF SR and tWZ SF SR are kept blinded by omitting the data points.

		tWZ OF SR	tWZ SF SR	$t\bar{t}Z$ CR	ZZb CR	$(tWZ)_{\text{fake}}$ CR
$t\bar{t}Z$		13.2 ± 1.5	9.6 ± 1.1	29.9 ± 3.6	5.1 ± 0.6	18.5 ± 2.2
$t\bar{t}Z$ fakes		0.070 ± 0.047	0.033 ± 0.026	0.072 ± 0.042	0.052 ± 0.021	5.1 ± 2.3
tWZ		7.8 ± 3.3	5.3 ± 2.2	5.7 ± 2.6	2.9 ± 1.2	10.2 ± 4.3
ZZ		0.48 ± 0.12	7.7 ± 1.2	1.1 ± 0.2	40.6 ± 6.3	6.9 ± 1.1
other	$t\bar{t}$	$6e-06 \pm 3e-06$	0.25 ± 0.44	0.27 ± 0.22	$6e-06 \pm 3e-06$	2.4 ± 0.9
	tZq	0.083 ± 0.040	0.076 ± 0.035	0.064 ± 0.030	0.060 ± 0.024	4.9 ± 0.7
	$t\bar{t}tW$	0.0067 ± 0.0079	0.0028 ± 0.0028	$6e-06 \pm 3e-06$	0.0023 ± 0.0056	0.94 ± 0.29
	WZ	0.04 ± 0.02	0.04 ± 0.02	0.013 ± 0.013	0.047 ± 0.033	1.8 ± 0.4
	$t\bar{t}t$	0.0010 ± 0.0007	0.002 ± 0.001	0.014 ± 0.005	$6e-06 \pm 3e-06$	0.010 ± 0.003
	$t\bar{t}\bar{t}$	0.009 ± 0.008	0.011 ± 0.008	0.06 ± 0.02	$6e-06 \pm 3e-06$	0.02 ± 0.01
	$t\bar{t}WW$	0.029 ± 0.026	0.03 ± 0.02	0.3 ± 0.1	0.01 ± 0.03	0.2 ± 0.1
	$VVV (V = W/Z)$	0.3 ± 0.1	0.2 ± 0.1	0.10 ± 0.02	0.17 ± 0.05	0.26 ± 0.08
	$t\bar{t}H$	0.9 ± 0.2	0.7 ± 0.1	2.0 ± 0.4	0.20 ± 0.03	2.2 ± 0.5
Total		22.9 ± 3.0	24.0 ± 2.1	39.4 ± 3.4	49.1 ± 6.0	53.4 ± 4.3
data		-	-	36	49	57

Table 13: The expected number of events (after the fit) for each sample in each region is shown.

mis-modelling in the backgrounds of the post-fit model. Since there is a small number of events in the regions of tetralepton channel, the disagreement may be due to statistical fluctuations in data or simulation.

The effect of each systematic uncertainty on the POI, $\mu(tWZ)$, in the fit can be studied from a *ranking plot* [79]. A ranking plot lists the nuisance parameters used in the fit in order of their *impact* on the POI. The impact of a nuisance parameter describes how much the nuisance parameter effects the extracted value of the POI from the fit. The impact of a nuisance parameter, θ , is calculated by first fixing the pre-fit and post-fit uncertainties, given by $\Delta\theta$ and $\Delta\hat{\theta}$ respectively, of all the nuisance parameters to ± 1 , performing the fit, and extracting the nominal value of the POI. The impact of a given nuisance parameter is then defined by the difference in this reference POI value to a POI value extracted from a fit performed by removing the given nuisance parameter. In Figure 45, a ranking plot showing the impact of the systematic uncertainties on the POI, $\mu(tWZ)$ is shown. The systematic uncertainties with largest impacts are the cross sections of $t\bar{t}Z$ and ZZ , and shape modelling on $t\bar{t}Z$ ($t\bar{t}Z$ triangular shape) and tWZ (tWZ -DR2 and tWZ triangular shape). The cross section of ZZ is slightly shifted down in the fit. The $t\bar{t}Z$ cross section, tWZ -DR2 and $t\bar{t}Z$ triangular shape nuisance parameters are similarly shifted down in the fit from their nominal values, but to a much lesser degree than the cross section of ZZ . These pulls are all within 1σ uncertainty and are thus relatively small. It is expected that the modelling uncertainties (shape and normalisations) of the most dominant backgrounds (e.g. $t\bar{t}Z$, ZZ) have relatively large impacts on $\mu(tWZ)$, since the uncertainty of the analysis is dominated by statistical uncertainty.

4.10.2 Trilepton and Tetralepton Channels

In the section, an expected upper limit and an expected significance are set on the signal strength of tWZ ($\mu(tWZ)$) from the combined fit across all regions of tWZ in the tetralepton and trilepton channels. The trilepton channel is an entirely independent analysis [95] that uses separate data and simulated events compared to the tetralepton channel. Although the trilepton channel uses separate data and simulated

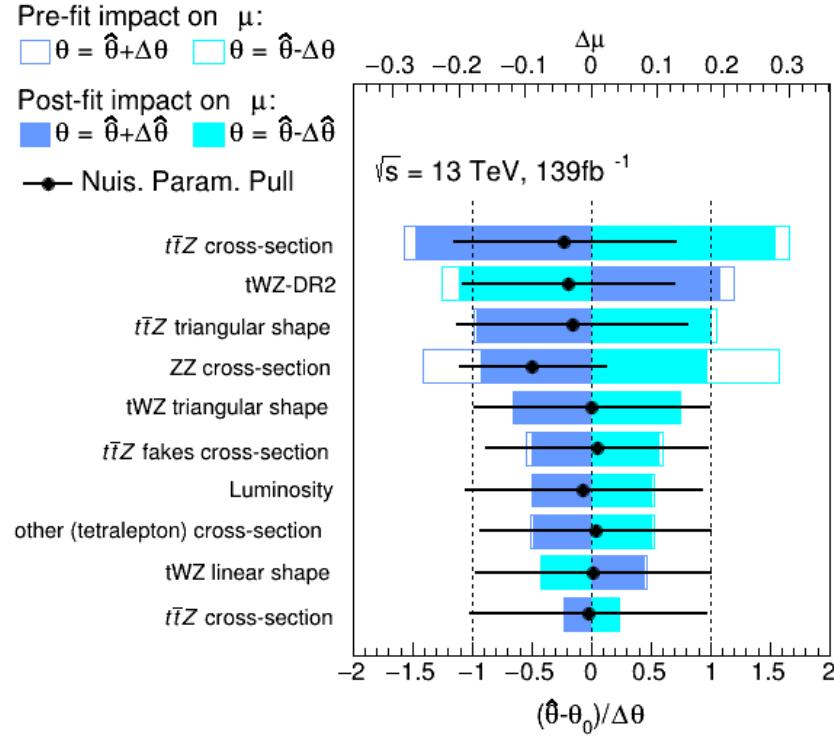


Figure 45: A ranking plot showing the impact (ordered from top to bottom via decreasing impact) of the systematic uncertainties (top 10) on the POI, $\mu(tWZ)$, in the tetralepton channel is shown. The best-fit value of the nuisance parameter is given by $\hat{\theta}$. The post-fit and pre-fit uncertainties are given by $\Delta\hat{\theta}$ and $\Delta\theta$ respectively. The post-fit and pre-fit impact of each nuisance parameter on $\mu(tWZ)$ are shown with the solid and lined rectangles respectively. The empty and solid blue rectangles correspond to the pre-fit and post-fit impacts on $\mu(tWZ)$ respectively. These impacts are shown on the upper axis ($\Delta\mu$). On the lower axis, the nuisance parameter pull, $\frac{\hat{\theta} - \theta_0}{\Delta\theta}$, is shown (θ_0 is the nominal pre-fit value of the nuisance parameter). The nuisance parameter pull is indicated by the black points, with their relative post-fit errors ($\frac{\Delta\hat{\theta}}{\Delta\theta}$) shown by the black horizontal error bars.

events, the treatment of the statistical and systematic uncertainties (and therefore the statistical model) in the combined fit is entirely coherent with the tetralepton channel. Therefore the inclusion of the trilepton channel in the combined fit is only a matter of adding the regions in the trilepton channel to the likelihood definition. The trilepton analysis follows a similar analysis strategy to that of the tetralepton analysis. It includes an event-level BDT which aims to discriminate between the tWZ and all background as well as an object-level BDT which aims to identify hadronically decaying W bosons to discriminate between tWZ and the large WZ background. One tWZ SR is defined and five CRs are defined. The WZ and $t\bar{t}Z$ CRs are defined to constrain the dominant WZ and $t\bar{t}Z$ backgrounds. Three CRs which require that one of the three selected leptons are loose, are defined for WZ , $t\bar{t}Z$ and tWZ in order to constraint the fake lepton component (using the MC template method - similar to the method used in Section 4.6 to estimate the fake lepton component). The expected upper limit of tWZ in the trilepton channel is measured as, $\mu_{up}^{exp} = 2.65^{+3.67}_{-1.91}$. The expected significance of tWZ in the trilepton channel is measured as, $Z_\mu^{exp} = 0.75\sigma$. The best-fit value of the signal strength, $\mu(tWZ) = \frac{\sigma(tWZ)}{\sigma(tWZ)_{SM}}$, from the likelihood fit is measured as, $\mu(tWZ) = 1.16^{+1.33}_{-1.30}$. The best fit value of the signal strength is consistent with 1. This indicates an agreement between data and the post-fit model within post-fit model uncertainties.

Therefore, this demonstrates that the model is able to sufficiently model the background processes. In Table 14, the nuisance parameters that are included (or excluded) in terms of the likelihood function corresponding to a certain channel, used in the fit, are shown. Note that the $\sigma(tWZ - DR2)$ systematic

Nuisance Parameter	Channel	
	trilepton	tetralepton
$\sigma(t\bar{t}H)$	✓	✓
$\sigma(t\bar{t}Z)$	✓	✓
$\sigma(WZ)$	✓	✓
$\sigma(tZq)$	✓	✓
$\sigma(ZZ)$	✓	✓
$\sigma(\text{other(trilepton)})$	✓	✗
$\sigma(\text{other(tetralepton)})$	✗	✓
$\sigma(t\bar{t}Z)_{fakes}$	✗	✓
$\sigma(t\bar{t})_{fakes}$	✓	✗
$\sigma(Z + \text{jets})_{fakes}$	✓	✗
Luminosity	✓	✓
jvt	✓	✓
pileup	✓	✓
DL1r SF (b jets)	✓	✓
DL1r SF (light jets)	✓	✓
$\sigma(tWZ - DR2)$	✗	✓
lepton SF	✓	✓
$t\bar{t}Z$ triangular shape	✓	✓
$t\bar{t}Z$ linear shape	✓	✓
$WZ + b$ triangular shape	✓	✗
$WZ + b$ linear shape	✓	✗
$WZ + c$ triangular shape	✓	✗
$WZ + c$ linear shape	✓	✗

Table 14: A summary of the nuisance parameters used in the combined fit is shown. Nuisance parameters that are included in terms of the likelihood function corresponding to a certain channel are indicated with a ✓ and those which are excluded are indicated with a ✗.

uncertainty was not included in the trilepton analysis, and the trilepton analysis was unaltered when used in this analysis for the combined fit. The expected upper limit of tWZ across both channels is measured as, $\mu_{up}^{exp} = 1.43^{+2.04}_{-1.03}$. The expected significance of tWZ across both channels is measured as, $Z_\mu^{exp} = 1.61\sigma$. The best-fit value of the signal strength, $\mu(tWZ) = \frac{\sigma(tWZ)}{\sigma(tWZ)_{SM}}$, from the likelihood fit is measured as, $\mu(tWZ) = 1.80^{+0.70}_{-0.65}$. In Figure 46, the expected upper limits and the best-fit values of $\mu(tWZ)$ in the trilepton channel, tetralepton channel and both channels combined are shown. It can be seen that the sensitivity of tWZ is mostly driven by the tetralepton analysis, with the trilepton analysis attributing a small decrease in the expected upper limit of the combined analysis, and its associated uncertainty. The best fit value for the signal strength on tWZ , $\mu(tWZ)$, and the expected limits for the tri- and tetralepton channels are consistent with one-another (their uncertainties overlap). Therefore it is appropriate to combine these two analyses. In Figure 47, a ranking plot showing the impact of the systematic uncertainties on the POI, $\mu(tWZ)$, in the combined fit across both the tri- and tetralepton channels is shown. Some nuisance parameters are pulled down from their nominal pre-fit values, however these are all within 1σ uncertainty and are thus relatively small. The systematic uncertainties with largest impacts are the cross sections of $t\bar{t}Z$, ZZ and $WZ + b$, and shape modelling on $t\bar{t}Z$ ($t\bar{t}Z$ triangular shape). The aforementioned nuisance parameters are slightly shifted down in the fit from their nominal value. The lower ranked systematic uncertainties are pulled in the fit from their nominal values, but to

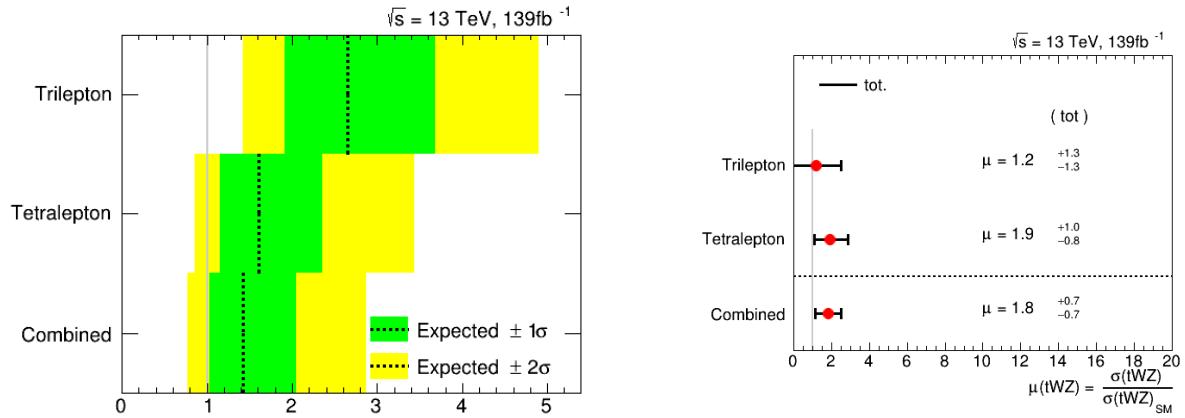


Figure 46: **Left:** The expected upper limits of the trilepton channel, tetralepton channel and both channels combine are shown. The y-axis shows the channels in which the fitting procedure was performed. The expected limits are represented by the vertical dotted line. One- and two- σ uncertainty bands are shown in green and yellow respectively. The vertical grey line indicates when $\mu(tWZ) = 1$. **Right:** The best-fit values of $\mu(tWZ)$ from the fit for the trilepton channel, tetralepton channel and both channels combined are shown. The y-axis shows the channels in which the fitting procedure was performed. The signal strength $\mu(tWZ)$ is shown on the x-axis. The nominal signal strengths are represented by the red dots. The total uncertainty associated with the best-fit $\mu(tWZ)$ value is shown by the black error bars.

1446 a much lesser degree than the cross section of those mentioned above. It is expected that the modelling
 1447 uncertainties of the most dominant backgrounds (e.g. $t\bar{t}Z$, ZZ and $WZ + b$) have relatively large impacts
 1448 on $\mu(tWZ)$, since the uncertainty of the analysis is dominated by statistical uncertainty.

1449 4.10.2.1 Projection to Higher Luminosity

1450 The expected significance of tWZ across both channels was measured as 1.61σ and it is likely that the
 1451 analysis is statistically limited. Therefore, the question naturally arises whether or not it is possible to
 1452 observe tWZ with a 3σ or 5σ significance and how much data one would need to achieve this. In this
 1453 section we apply a fully blinded fit to the Asimov dataset for integrated luminosities larger than the
 1454 139fb^{-1} currently available from the ATLAS Full Run 2 dataset. This study gives us insight into the
 1455 sensitivity of this analysis to the cross section of tWZ which we could expect if we were to replicate the
 1456 current analysis, given more data. Given the upgrades planned for the LHC and the ATLAS detector,
 1457 we will soon expect a large increase in available pp collision data (reaching $\mathcal{L} = 3000\text{fb}^{-1}$ [35]). This
 1458 study therefore gives us an idea of the sensitivity of this analysis to the cross section of tWZ which we
 1459 can expect in the future. It would also provide a clear indication whether or not the uncertainties in this
 1460 analysis is dominated by statistical uncertainties. In Figure 48, the expected upper limit and significance
 1461 for combined fully blinded fits to the Asimov dataset across both the tri- and tetralepton channels for a
 1462 range integrated luminosities are shown. As we increase the total integrated luminosity, the sensitivity
 1463 of this analysis to the tWZ process increases. The increase in sensitivity is shown by the rapid decrease
 1464 of the expected upper limit and the rapid increase of the expected significance with increasing integrated
 1465 luminosity. It can be noted that a 3σ expected significance is achieved at $\mathcal{L} \sim 400\text{ fb}^{-1}$, indicating evidence
 1466 for observation of tWZ . The 5σ significance 'discovery' standard is above our 1000fb^{-1} luminosity range
 1467 of study. Furthermore, these plots show that the sensitivity of tWZ is hindered by the low amount of

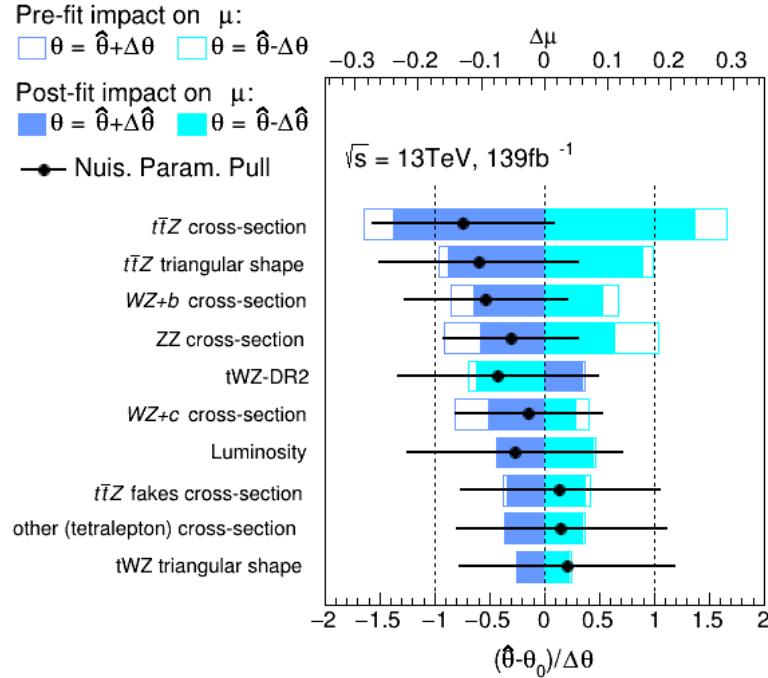


Figure 47: A ranking plot showing the impact of the systematic uncertainties (top 10) on the POI, $\mu(tWZ)$, in the combined fit across both the tri- and tetralepton channels is shown. The best-fit value of the nuisance parameter is given by $\hat{\theta}$. The post-fit and pre-fit uncertainties are given by $\Delta\hat{\theta}$ and $\Delta\theta$ respectively. The post-fit and pre-fit impact of each nuisance parameter on $\mu(tWZ)$ are shown with the solid and lined rectangles respectively. The empty and solid blue rectangles correspond to the pre-fit and post-fit impacts respectively. These impacts are shown on the upper axis ($\Delta\mu$). On the lower axis, the nuisance parameter pull, $\frac{\hat{\theta} - \theta_0}{\Delta\theta}$, is shown (θ_0 is the nominal pre-fit value of the nuisance parameter). The nuisance parameter pull is indicated by the black points, with their relative post-fit errors ($\frac{\Delta\hat{\theta}}{\Delta\theta}$) shown by the black horizontal error bars.

1468 events we observe, rather than the systematic effects. This is evident since increasing the luminosity,
 1469 therefore increasing the number of events we see in the detector, directly causes a steady increase in
 1470 sensitivity.

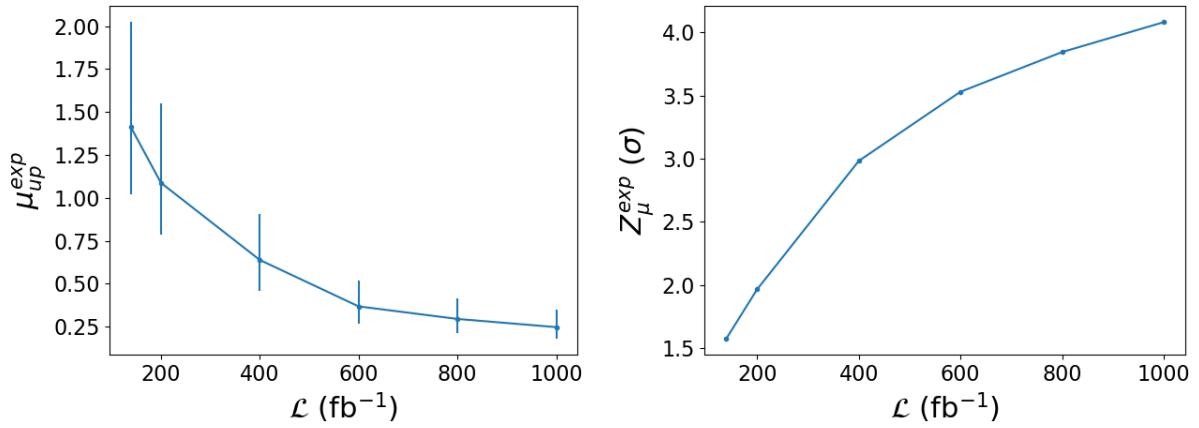


Figure 48: The expected upper limit (left) and significance (right) for combined fully blinded fits to the Asimov dataset across both the tri- and tetralepton channels for a range integrated luminosities are shown. **Left:** The integrated luminosity, \mathcal{L} , is shown on the x-axis. The expected upper limit, μ_{up}^{exp} , is shown on the y-axis. The vertical lines represent the total uncertainty ($\pm 1\sigma$) on the expected upper limit. **Right:** The integrated luminosity, \mathcal{L} , is shown on the x-axis. The expected significance, Z_μ^{exp} , is shown on the y-axis. The vertical lines represent the total uncertainty ($\pm 1\sigma$) on the expected significance.

1471

Chapter 5

1472

Summary and Conclusions

1473 The tWZ process is an important process in the search for new physics since its cross section is sensitive to
 1474 the charged and neutral couplings to the top quark, which is strongly coupled to the Higgs boson. The top
 1475 quark's couplings are modified in many scenarios of new physics that aim to resolve the Hierarchy Problem,
 1476 therefore a constraint placed on the cross section of tWZ is expected to be impactful in constraining such
 1477 BSM models. A search for tWZ production using 139 fb^{-1} of pp collision data at a centre-of-mass energy
 1478 of $\sqrt{s} = 13 \text{ TeV}$, recorded by the ATLAS experiment at CERN, has been presented. This thesis targeted
 1479 the tetralepton final state channel. The dominant background processes, $t\bar{t}Z$ and ZZ were constrained
 1480 by the definition of $t\bar{t}Z$ and ZZ CRs, respectively. The dominant source of fake leptons, originating
 1481 from the $t\bar{t}Z$ background, was constrained by the $(tWZ)_{\text{fake}}$ CR, using the MC template method. Two
 1482 BDTs were implemented: an object-level BDT which aims to classify between ℓb systems coming from
 1483 top quarks and an event-level BDT which aims to discriminate between tWZ and our major backgrounds,
 1484 $t\bar{t}Z$ and ZZ . The output from the object-level BDT was converted to an event-level variable to be used
 1485 as input to the event-level BDT. A kinematic reconstruction algorithm, $2\nu\text{SM}$, was used to reconstruct
 1486 top quarks in order to discriminate between tWZ and $t\bar{t}Z$. The output from this algorithm was used as
 1487 an input variable to the event-level BDT. The BDT was shown to discriminate well between signal and
 1488 background events. Using a modified Asimov dataset in the SRs and real data in the CRs, a blinded
 1489 maximum-likelihood fit was performed across all regions in the tetralepton channel. The best-fit value of
 1490 the signal strength in the tetralepton channel was,

$$\mu(tWZ) = 1.91^{+0.95}_{-0.82} \quad (5.1)$$

1491 with an expected significance of 1.44σ . The expected upper limit on the signal strength of tWZ in the
 1492 tetralepton channel was,

$$\mu_{up}^{\exp} = 1.61^{+2.35}_{-1.16} \quad (5.2)$$

1493 To further increase the sensitivity of this analysis to the tWZ process, a blinded maximum-likelihood fit
 1494 was performed across all regions across the trilepton (studied in an independent analysis by Benjamin
 1495 Warren (UCT) [95]) and tetralepton channels. The best-fit value of the signal strength across both the
 1496 trilepton and tetralepton channels were,

$$\mu(tWZ) = 1.80^{+0.70}_{-0.65} \quad (5.3)$$

1497 with an expected significance of 1.61σ . The expected upper limit on the signal strength of tWZ across
1498 both the trilepton and tetralepton channels were,

$$\mu_{up}^{exp} = 1.43^{+2.04}_{-1.03} \quad (5.4)$$

1499 The results in this analysis are heavily statistically limited, it is therefore expected that future analyses
1500 of this process, using larger datasets (such as that from the HL-LHC), would significantly improve the
1501 results.

1502

Bibliography

- 1503 [1] M. Aaboud et al. “Electron reconstruction and identification in the ATLAS experiment using the
 1504 2015 and 2016 LHC proton–proton collision data at $\sqrt{s} = 13$ TeV”. In: *The European Physical
 1505 Journal C* 79.8 (Aug. 2019). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-7140-6. URL:
 1506 <http://dx.doi.org/10.1140/epjc/s10052-019-7140-6>.
- 1507 [2] M. Aaboud et al. “Measurement of the $t\bar{t}Z$ and $t\bar{t}W$ cross sections in proton-proton collisions at
 1508 $\sqrt{s} = 13$ TeV with the ATLAS detector”. In: *Phys. Rev. D* 99 (7 Apr. 2019), p. 072009. DOI: 10.
 1509 1103/PhysRevD.99.072009. URL: <https://link.aps.org/doi/10.1103/PhysRevD.99.072009>.
- 1510 [3] Aaboud, M. and Aad, G. and Abbott, B. and Abbott, D. C. and Abdinov, O. and Abed Abud, A. and
 1511 Abhayasinghe, D. K. and Abidi, S. H. and AbouZeid, O. S. and et al. “Measurement of ZZ production
 1512 in the $\ell\ell\nu\nu$ final state with the ATLAS detector in pp collisions at $s \sqrt{s} = 13$ TeV”. In: *Journal
 1513 of High Energy Physics* 2019.10 (Oct. 2019). ISSN: 1029-8479. DOI: {10.1007/jhep10(2019)127}.
 1514 URL: %7B[http://dx.doi.org/10.1007/JHEP10\(2019\)127](http://dx.doi.org/10.1007/JHEP10(2019)127)%7D.
- 1515 [4] G. Aad et al. “Measurement of the tt^- production cross-section in the lepton+jets channel at $s=13$
 1516 TeV with the ATLAS experiment”. In: *Physics Letters B* 810 (2020), p. 135797. ISSN: 0370-2693.
 1517 DOI: <https://doi.org/10.1016/j.physletb.2020.135797>. URL: <https://www.sciencedirect.com/science/article/pii/S0370269320306006>.
- 1519 [5] G. Aad et al. “Muon reconstruction performance of the ATLAS detector in proton–proton collision
 1520 data at $\sqrt{s} = 13$ TeV”. In: *The European Physical Journal C* 76.5 (May 2016). ISSN: 1434-6052.
 1521 DOI: 10.1140/epjc/s10052-016-4120-y. URL: <http://dx.doi.org/10.1140/epjc/s10052-016-4120-y>.
- 1523 [6] Shunichi Akatsuka and Shion Chen. *Isolation WPs summary: PLV + LowPtPLV*. Oct. 2019. URL:
 1524 https://indico.cern.ch/event/854783/contributions/3595486/attachments/1929380/3195230/PLV_Summary.pdf.
- 1526 [7] G. ALTARELLI, R. BARBIERI, and F. CARAVAGLIOS. “ELECTROWEAK PRECISION TEST:
 1527 A CONCISE REVIEW”. In: *International Journal of Modern Physics A* 13.07 (Mar. 1998),
 1528 pp. 1031–1058. ISSN: 1793-656X. DOI: 10.1142/s0217751x98000469. URL: <http://dx.doi.org/10.1142/S0217751X98000469>.
- 1530 [8] Guido Altarelli. *The Higgs and the Excessive Success of the Standard Model*. 2014. arXiv: 1407.2122
 1531 [hep-ph].

- [9] J. Alwall et al. “The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations”. In: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].
- [10] Aaron Armbruster. *Asymptotic Formulae*. Feb. 2013. URL: https://indico.cern.ch/event/233551/contributions/493678/attachments/389871/542293/asymptotics_armbruster.pdf.
- [11] Pierre Astier et al. “Kalman filter track fits and track breakpoint analysis”. In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 450.1 (2000), pp. 138–154. ISSN: 0168-9002. DOI: [https://doi.org/10.1016/S0168-9002\(00\)00154-6](https://doi.org/10.1016/S0168-9002(00)00154-6). URL: <https://www.sciencedirect.com/science/article/pii/S0168900200001546>.
- [12] ATLAS Collaboration. “ATLAS b -jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13\text{TeV}$ ”. In: *Eur. Phys. J. C* 79 (2019), p. 970. DOI: 10.1140/epjc/s10052-019-7450-8. arXiv: 1907.05120 [hep-ex].
- [13] ATLAS Collaboration. “Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data”. In: *JINST* 14 (2019), P12006. DOI: 10.1088/1748-0221/14/12/P12006. arXiv: 1908.00005 [hep-ex].
- [14] ATLAS Collaboration. “Identification and rejection of pile-up jets at high pseudorapidity with the ATLAS detector”. In: *Eur. Phys. J. C* 77 (2017), p. 580. DOI: 10.1140/epjc/s10052-017-5081-5. arXiv: 1705.02211 [hep-ex].
- [15] ATLAS Collaboration. “Measurements of b -jet tagging efficiency with the ATLAS detector using $t\bar{t}$ events at $\sqrt{s} = 13\text{TeV}, \text{TeV}$ ”. In: *JHEP* 08 (2018), p. 089. DOI: 10.1007/JHEP08(2018)089. arXiv: 1805.01845 [hep-ex].
- [16] ATLAS Collaboration. “Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at $\sqrt{s} = 13\text{TeV}$ ”. In: (2020). arXiv: 2012.00578 [hep-ex].
- [17] Manuel Guth on behalf of the ATLAS collaboration. *Deep-Neural-Network-based b -Tagging as Basis for Improvements in Top Analyses*. URL: <https://cds.cern.ch/record/2693088/files/ATL-PHYS-SLIDE-2019-751.pdf>.
- [18] ATLAS Internal. *Electron Efficiencies for Analyses*. 2021. URL: <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/ElectronEfficienciesForAnalysis>.
- [19] ATLAS Internal. *Electron Efficiency Correlation Model*. 2021. URL: <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/ElectronEfficiencyCorrelationModel>.
- [20] ATLAS Internal. *Jet Vertex Tagger for Run 2 in reco and analysis*. 2021. URL: <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/JetVertexTaggerTool>.
- [21] G. Avoni et al. “The new LUCID-2 detector for luminosity measurement and monitoring in ATLAS”. In: *JINST* 13.07 (2018), P07017. DOI: 10.1088/1748-0221/13/07/P07017.
- [22] Richard D. Ball et al. “Parton distributions for the LHC run II”. In: *Journal of High Energy Physics* 2015.4 (Apr. 2015). ISSN: 1029-8479. DOI: 10.1007/jhep04(2015)040. URL: [http://dx.doi.org/10.1007/JHEP04\(2015\)040](http://dx.doi.org/10.1007/JHEP04(2015)040).

- 1570 [23] Roger Barlow and Christine Beeston. “Fitting using finite Monte Carlo samples”. In: *Computer*
 1571 *Physics Communications* 77.2 (1993), pp. 219–228. ISSN: 0010-4655. DOI: [https://doi.org/10.1016/0010-4655\(93\)90005-W](https://doi.org/10.1016/0010-4655(93)90005-W). URL: <https://www.sciencedirect.com/science/article/pii/001046559390005W>.
- 1574 [24] Olga Bessidskaia Bylund. *Measurement of ttZ and ttW production at ATLAS in 13 TeV data, using trilepton and same charge dimuon final states*. Tech. rep. Geneva: CERN, Aug. 2016. DOI:
 1575 10.22323/1.276.0237. URL: <http://cds.cern.ch/record/2211022>.
- 1577 [25] Enrico Bothmann et al. “Event generation with Sherpa 2.2”. In: *SciPost Physics* 7.3 (Sept. 2019).
 1578 ISSN: 2542-4653. DOI: 10.21468/scipostphys.7.3.034. URL: <http://dx.doi.org/10.21468/SciPostPhys.7.3.034>.
- 1580 [26] Ilaria Brivio and Michael Trott. “The standard model as an effective field theory”. In: *Physics*
 1581 *Reports* 793 (Feb. 2019), pp. 1–98. ISSN: 0370-1573. DOI: 10.1016/j.physrep.2018.11.002. URL:
 1582 <http://dx.doi.org/10.1016/j.physrep.2018.11.002>.
- 1583 [27] Jason Brownlee. *Overfitting and Underfitting With Machine Learning Algorithms*. URL: <https://machinelearningmastery.com/overfitting-and-underfitting-with-machine-learning-algorithms/>.
- 1586 [28] Oliver Sim Brüning et al. *LHC Design Report*. CERN Yellow Reports: Monographs. Geneva: CERN,
 1587 2004. DOI: 10.5170/CERN-2004-003-V-1. URL: <https://cds.cern.ch/record/782076>.
- 1588 [29] Gustavo Burdman et al. “Colorless top partners, a 125 GeV Higgs boson, and the limits on natural-
 1589 ness”. In: *Physical Review D* 91.5 (Mar. 2015). ISSN: 1550-2368. DOI: 10.1103/physrevd.91.055007.
 1590 URL: <http://dx.doi.org/10.1103/PhysRevD.91.055007>.
- 1591 [30] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. “The anti- k_t jet clustering algorithm”. In:
 1592 *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].
- 1593 [31] Thomas Calvet. *Automatic binning implementation in TTHFitter - Htop(bb)*. URL: https://indico.cern.ch/event/455289/contributions/1953694/attachments/1209081/1762963/Calvet_binning_Htop_160108.pdf.
- 1596 [32] Thomas Philippe Calvet. “Search for the production of a Higgs boson in association with top quarks
 1597 and decaying into a b-quark pair and b-jet identification with the ATLAS experiment at LHC”.
 1598 Presented 08 Nov 2017. Dec. 2017. URL: <https://cds.cern.ch/record/2296985>.
- 1599 [33] Anadi Canepa. “Searches for supersymmetry at the Large Hadron Collider”. In: *Reviews in Physics*
 1600 4 (2019), p. 100033. ISSN: 2405-4283. DOI: <https://doi.org/10.1016/j.revip.2019.100033>.
 1601 URL: <https://www.sciencedirect.com/science/article/pii/S2405428318300091>.
- 1602 [34] *CERN Twiki - TOP WG Summary Plots*. URL: <https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCTopWGSummaryPlots>.
- 1604 [35] CERN Yellow Reports: Monographs. *CERN Yellow Reports: Monographs, Vol. 10 (2020): High-
 1605 Luminosity Large Hadron Collider (HL-LHC): Technical design report*. en. 2020. DOI: 10.23731/
 1606 CYRM-2020-0010. URL: <https://e-publishing.cern.ch/index.php/CYRM/issue/view/127>.

- 1607 [36] Shion Chen. *Track isolation variable for the PFlow WPs*. Oct. 2019. URL: https://indico.cern.ch/event/854783/contributions/3595529/attachments/1926980/3190772/IFF_20191003-PflowWPs.pdf.
- 1608
- 1609
- 1610 [37] KyungEon Choi. “Tracking and Vertexing with the ATLAS Inner Detector in the LHC Run-2”. In: *Springer Proc. Phys.* 213 (2018). Ed. by Zhen-An Liu, pp. 400–403. DOI: 10.1007/978-981-13-1316-5_75.
- 1611
- 1612
- 1613 [38] CMS Collaboration. *Measurements of $pp \rightarrow ZZ$ production cross sections and constraints on anomalous triple gauge couplings at $\sqrt{s} = 13$ TeV*. 2020. arXiv: {2009.01186} (hep-ex).
- 1614
- 1615 [39] CMS Collaboration. “Measurement of the ttbar production cross section using events with one lepton and at least one jet in pp collisions at $\sqrt{s}=13$ TeV”. In: (Jan. 2017).
- 1616
- 1617 [40] The ATLAS Collaboration et al. “The ATLAS Experiment at the CERN Large Hadron Collider”. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08003–S08003. DOI: 10.1088/1748-0221/3/08/s08003. URL: <https://doi.org/10.1088/1748-0221/3/08/s08003>.
- 1618
- 1619
- 1620 [41] Glen Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”. In: *The European Physical Journal C* 71.2 (Feb. 2011). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-011-1554-0. URL: <http://dx.doi.org/10.1140/epjc/s10052-011-1554-0>.
- 1621
- 1622
- 1623 [42] Federico Demartin et al. “tWH associated production at the LHC”. In: *Eur. Phys. J. C* 77.1 (2017), p. 34. DOI: 10.1140/epjc/s10052-017-4601-7. arXiv: 1607.05862 [hep-ph].
- 1624
- 1625 [43] *Electron identification efficiency in data for electrons with $E_T > 30\text{GeV}$* . URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/EGAM-2018-01/fig_16.png.
- 1626
- 1627 [44] F. Englert and R. Brout. “Broken Symmetry and the Mass of Gauge Vector Mesons”. In: *Phys. Rev. Lett.* 13 (9 Aug. 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321. URL: <https://link.aps.org/doi/10.1103/PhysRevLett.13.321>.
- 1628
- 1629
- 1630 [45] *Errors in weighted histograms*. URL: <https://www.zeuthen.desy.de/~wischnew/amanda/discussion/wgterror/working.html>.
- 1631
- 1632 [46] Lyndon Evans and Philip Bryant. “LHC Machine”. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08001–S08001. DOI: 10.1088/1748-0221/3/08/s08001. URL: <https://doi.org/10.1088/1748-0221/3/08/s08001>.
- 1633
- 1634
- 1635 [47] Guido Fantini et al. *The formalism of neutrino oscillations: an introduction*. 2020. arXiv: 1802.05781 [hep-ph].
- 1636
- 1637 [48] *Forward Jet Vertex Tagging: A new technique for the identification and rejection of forward pileup jets*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2015-034>. Geneva: CERN, Aug. 2015. URL: <https://cds.cern.ch/record/2042098>.
- 1638
- 1639
- 1640
- 1641 [49] Stefano Frixione, Paolo Nason, and Carlo Oleari. “Matching NLO QCD computations with parton shower simulations: the POWHEG method”. In: *JHEP* 11 (2007), p. 070. DOI: 10.1088/1126-6708/2007/11/070. arXiv: 0709.2092 [hep-ph].
- 1642
- 1643

- [50] Michele Gallinaro. “Top quark physics: A tool for discoveries”. In: *Journal of Physics Conference Series* 447 (July 2013), pp. 2012–. DOI: 10.1088/1742-6596/447/1/012012.
- [51] *Good Run Lists for Analysis Run 2 - ATLAS Twiki*. URL: <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/GoodRunListsForAnalysisRun2>.
- [52] Particle Data Group et al. “Review of Particle Physics”. In: *Progress of Theoretical and Experimental Physics* 2020.8 (Aug. 2020). 083C01. ISSN: 2050-3911. DOI: 10.1093/ptep/ptaa104. eprint: <https://academic.oup.com/ptep/article-pdf/2020/8/083C01/33653179/ptaa104.pdf>. URL: <https://doi.org/10.1093/ptep/ptaa104>.
- [53] Trevor Hastie. *The elements of statistical learning : data mining, inference, and prediction*. New York: Springer, 2009. ISBN: 978-0-387-84858-7.
- [54] Peter W. Higgs. “Broken Symmetries and the Masses of Gauge Bosons”. In: *Phys. Rev. Lett.* 13 (1964). Ed. by J. C. Taylor, pp. 508–509. DOI: 10.1103/PhysRevLett.13.508.
- [55] Peter W. Higgs. “Spontaneous Symmetry Breakdown without Massless Bosons”. In: *Phys. Rev.* 145 (1966), pp. 1156–1163. DOI: 10.1103/PhysRev.145.1156.
- [56] *IFF Truth Classifier GitLab Repository*. URL: <https://gitlab.cern.ch/ATLAS-IFF/IFFTruthClassifier/-/tree/master>.
- [57] N. Jarosik et al. “SEVEN-YEAR WILKINSON MICROWAVE ANISOTROPY PROBE (WMAP) OBSERVATIONS: SKY MAPS, SYSTEMATIC ERRORS, AND BASIC RESULTS”. In: *The Astrophysical Journal Supplement Series* 192.2 (Jan. 2011), p. 14. ISSN: 1538-4365. DOI: 10.1088/0067-0049/192/2/14. URL: <http://dx.doi.org/10.1088/0067-0049/192/2/14>.
- [58] F. Jegerlehner. “The hierarchy problem of the electroweak Standard Model revisited”. In: (May 2013).
- [59] V. Khachatryan et al. “Measurement of the ZZ production cross section and $Z \rightarrow \ell^+ \ell^- \ell^+ \ell^-$ branching fraction in pp collisions at $\sqrt{s} = 13\text{TeV}$ ”. In: *Physics Letters B* 763 (2016), pp. 280–303. ISSN: 0370-2693. DOI: <https://doi.org/10.1016/j.physletb.2016.10.054>. URL: <https://www.sciencedirect.com/science/article/pii/S0370269316306256>.
- [60] H. J. W. Kirsten. *Introduction to supersymmetry*. Singapore Hackensack, NJ: World Scientific, 2010. ISBN: 978-9814293426.
- [61] Arthur Alexis Jules Lesage. *Lepton and photon performance at ATLAS and CMS*. 2017. arXiv: 1709.02598 [hep-ex].
- [62] Fabio Maltoni, Luca Mantani, and Ken Mimasu. “Top-quark electroweak interactions at high energy”. In: *Journal of High Energy Physics* 2019.10 (Oct. 2019). ISSN: 1029-8479. DOI: 10.1007/jhep10(2019)004. URL: [http://dx.doi.org/10.1007/JHEP10\(2019\)004](http://dx.doi.org/10.1007/JHEP10(2019)004).
- [63] Luigi Marchese. *Lepton and photon reconstruction and identification performance in ATLAS and CMS*. Tech. rep. Geneva: CERN, Sept. 2019. DOI: 10.22323/1.350.0237. URL: <https://cds.cern.ch/record/2688452>.

- 1680 [64] Thomas McCarthy. *Macro developed to compare t/W/Z reconstruction performance (2ℓ, 3ℓ, 4ℓ)*.
 1681 URL: https://indico.cern.ch/event/986357/contributions/4172907/attachments/2169451/3666801/reco_performance_macro_20210112.pdf.
- 1683 [65] Thomas McCarthy and Florian Fischer. “Exploiting full/partial $t\bar{t}$ reconstruction for background suppression in 2ℓ ”. In: (). URL: https://indico.cern.ch/event/955360/contributions/4016465/attachments/2102418/3534816/top_reco_bkgd_suppression_2L_20200915.pdf.
- 1686 [66] “MCTruthClassifier - ATLAS Twiki”. In: (). URL: <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/MCTruthClassifier>.
- 1688 [67] *Measurement of the $t\bar{t}$ production cross-section and lepton differential distributions in $e\mu$ dilepton events from pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*. Tech. rep. All figures including auxiliary figures are available at <https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2019-041>. Geneva: CERN, Aug. 2019. URL: <https://cds.cern.ch/record/2686255>.
- 1693 [68] F. Meloni. “Primary vertex reconstruction with the ATLAS detector”. In: *Journal of Instrumentation* 11.12 (Dec. 2016), pp. C12060–C12060. DOI: 10.1088/1748-0221/11/12/c12060. URL: <https://doi.org/10.1088/1748-0221/11/12/c12060>.
- 1696 [69] Ken Mimasu. *Top quark electroweak interactions*. 2021. arXiv: 2105.10261 [hep-ph].
- 1697 [70] Aditya Mishra. *Metrics to Evaluate your Machine Learning Algorithm*. URL: <https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234>.
- 1700 [71] *Monte Carlo to Monte Carlo scale factors for flavour tagging efficiency calibration*. Tech. rep. ATL-PHYS-PUB-2020-009. Geneva: CERN, May 2020. URL: <https://cds.cern.ch/record/2718610>.
- 1702 [72] Hussain Mujtaba. *What is Cross Validation in Machine learning? Types of Cross Validation*. URL: <https://www.mygreatlearning.com/blog/cross-validation/>.
- 1704 [73] “Muon Selection Tool, ATLAS TWiki”. In: (). URL: <https://twiki.cern.ch/twiki/bin/view/Atlas/MuonSelectionTool>.
- 1706 [74] O. Ogul Oncel. “Search for Single Top Quark Production in Association with a W and a Z Boson in the 3 Lepton Final State with the ATLAS Experiment at 13 TeV”. Presented 04 Jun 2018. May 2018. URL: <http://cds.cern.ch/record/2625170>.
- 1709 [75] António Onofre. “Top Quark Couplings and Search for New Physics at the LHC”. In: *Journal of Physics Conference Series* 447 (July 2013), pp. 2030–. DOI: 10.1088/1742-6596/447/1/012030.
- 1711 [76] *Option to force a shape withing an error band by hand - TRF documentation*. URL: https://trexfitter-docs.web.cern.ch/trexfitter-docs/model_building/shape/.
- 1713 [77] *Performance of the ATLAS Electron and Photon Trigger in p-p Collisions at $\sqrt{s} = 7$ TeV in 2011*. Tech. rep. Geneva: CERN, May 2012. URL: <https://cds.cern.ch/record/1450089>.
- 1715 [78] *Pileup jet recommendations*. URL: <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/PileupJetRecommendations>.

- 1717 [79] Michele Pinamonti. *Statistical methods at ATLAS and CMS*. URL: https://indico.cern.ch/event/727396/contributions/3021899/attachments/1657532/2654085/Statistical_methods_at_ATLAS_and_CMS_2.pdf.
- 1718
- 1719
- 1720 [80] *Recommended isolation working points - ATLAS Twiki*. URL: <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/RecommendedIsolationWPs>.
- 1721
- 1722 [81] *scikit-learn Documentation - Gradient Boosting Classifier*. URL: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html>.
- 1723
- 1724 [82] *Scikit-Learn GradientBoostingClassifier Documentation*. URL: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.GradientBoostingClassifier.html>.
- 1725
- 1726 [83] Pedro Ferreira da Silva. *Top quark production at the LHC*. 2016. arXiv: 1605.05343 [hep-ex].
- 1727 [84] T. Sjöstrand, S. Mrenna, and P. Skands. “A brief introduction to PYTHIA 8.1”. In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph].
- 1728
- 1729
- 1730 [85] Torbjörn Sjöstrand et al. “An introduction to PYTHIA 8.2”. In: *Comput. Phys. Commun.* 191 (2015), p. 159. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].
- 1731
- 1732 [86] Jory Sonneveld. *Searches for physics beyond the standard model at the LHC*. 2019. arXiv: 1905.06239 [hep-ex].
- 1733
- 1734 [87] *Standard model of elementary particles - Wikimedia Commons*. URL: https://en.wikipedia.org/wiki/File:Standard_Model_of_Elementary_Particles.svg.
- 1735
- 1736 [88] Mark Thomson. *Modern particle physics*. Cambridge, United Kingdom New York: Cambridge University Press, 2013. ISBN: 9781107034266.
- 1737
- 1738 [89] *TopRecoObjTwikiModel*. URL: <https://twiki.cern.ch/twiki/bin/view/AtlasProtected/TopRecoObjTwikiModel>.
- 1739
- 1740 [90] *Total Integrated Luminosity and Data Quality in 2015-2018 - LuminosityPublicResultsRun2*. URL: <https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2>.
- 1741
- 1742 [91] Daniel R. Tovey. “Transformation properties of the transverse mass under transverse Lorentz boosts at hadron colliders”. In: *The European Physical Journal C* 79.4 (Apr. 2019). ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-6813-5. URL: <http://dx.doi.org/10.1140/epjc/s10052-019-6813-5>.
- 1743
- 1744
- 1745
- 1746 [92] *TRExFitter Documentation*. URL: <https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/TtHFitter>.
- 1747
- 1748 [93] *TRExFitter: Mixed data and MC fit*. URL: <https://trexfitter-docs.web.cern.ch/trexfitter-docs/AdvancedTutorial2020/Mixed/>.
- 1749
- 1750 [94] Tommy Tschida. “Cross section measurements of processes with a single top quark and two vector bosons with the CMS experiment”. MA thesis. Vienna, Tech. U., 2020. DOI: 10.34726/hss.2020.79420.
- 1751
- 1752
- 1753 [95] Benjamin Warren. “A search for tWZ production in the trilepton channel using Run 2 data from the ATLAS experiment.” In: () .
- 1754

- 1755 [96] Kenneth G. Wilson. “Confinement of quarks”. In: *Phys. Rev. D* 10 (8 Oct. 1974), pp. 2445–2459. DOI:
1756 10.1103/PhysRevD.10.2445. URL: <https://link.aps.org/doi/10.1103/PhysRevD.10.2445>.
- 1757 [97] Katherine Woodruff. *Introduction to boosted decision trees*. URL: <https://indico.fnal.gov/event/15356/contributions/31377/attachments/19671/24560/DecisionTrees.pdf>.
1758