



UNIVERSITY OF CAPE TOWN

MASTER'S THESIS

---

A search for  $tWZ$  production at the  
ATLAS experiment

---

*Author:*

Alex Veltman

*Supervisor:*

Dr. James Keaveney

1 August 2023

# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
<b>2</b>	<b>LHC and the ATLAS detector</b>	<b>3</b>
2.1	The Large Hadron Collider . . . . .	3
2.1.1	Luminosity at the LHC . . . . .	4
2.2	The ATLAS detector . . . . .	4
2.2.1	Coordinate system at ATLAS . . . . .	4
2.2.2	Inner Detector . . . . .	5
2.2.3	Calorimeters . . . . .	6
2.2.4	Muon Spectrometer . . . . .	6
2.2.5	Trigger and data acquisition system . . . . .	7
<b>3</b>	<b>Theoretical Background</b>	<b>8</b>
3.1	The Standard Model . . . . .	8
3.1.1	Fermions . . . . .	8
3.1.2	Bosons and Interactions . . . . .	9
3.1.3	Shortcomings of the Standard Model . . . . .	9
3.2	The Top quark . . . . .	10
3.3	The $tWZ$ process . . . . .	11
3.3.1	Diagram Removal Scheme . . . . .	12
<b>4</b>	<b>Graph Neural Networks</b>	<b>14</b>
4.1	Overview of Machine Learning . . . . .	14
4.2	Graph Neural Networks in Particle Physics . . . . .	15
4.2.1	Outline of a graph network . . . . .	16
4.2.2	GNN Architecture . . . . .	17
4.2.3	Application and limitations of graphs in HEP . . . . .	18
<b>5</b>	<b>Search for <math>tWZ</math> production</b>	<b>20</b>
5.1	Data and Monte Carlo Samples . . . . .	21
5.2	Physics object definitions . . . . .	24

5.2.1	Leptons . . . . .	25
5.2.2	Jets . . . . .	26
5.2.3	Missing Transverse Momentum . . . . .	27
5.2.4	Overlap Removal . . . . .	28
5.3	Systematic Uncertainties . . . . .	28
5.3.1	Experimental Uncertainties . . . . .	28
5.3.2	Modelling Uncertainties . . . . .	29
5.4	Definition of regions . . . . .	31
5.5	Signal and background discrimination in regions . . . . .	32
5.6	Statistical methods . . . . .	33
5.6.1	Definition of likelihood function . . . . .	33
5.6.2	Statistical tests for signal discovery . . . . .	35
5.6.3	Statistical test for upper limits . . . . .	36
5.6.4	Asimov data and experimental sensitivity . . . . .	37
5.6.5	Blinding of regions . . . . .	37
<b>6</b>	<b>Analysis of the <math>tWZ</math> trilepton channel</b>	<b>39</b>
6.1	Baseline selections and region definitions . . . . .	40
6.1.1	Optimisation of requirement on $E_T^{\text{miss}}$ . . . . .	43
6.1.2	Modelling of photon conversions . . . . .	44
6.2	Identification of hadronically-decaying $W$ bosons . . . . .	45
6.2.1	Dataset of hadronically-decaying $W$ bosons . . . . .	46
6.2.2	Features of dijet systems . . . . .	46
6.2.3	Dijet selection using $D_{jj}$ metric . . . . .	48
6.3	Graph neural network in the trilepton channel . . . . .	52
6.3.1	Dataset of graph events . . . . .	52
6.3.2	Graph Construction . . . . .	52
6.3.3	Performance of trilepton GNN . . . . .	54
6.4	Trilepton Asimov Fit Results . . . . .	55
6.5	Signal injection test . . . . .	59
<b>7</b>	<b>Analysis of the <math>tWZ</math> tetralepton channel</b>	<b>61</b>
7.1	Baseline selections and region definitions . . . . .	62
7.2	Two Neutrino Scanning Method . . . . .	65
7.2.1	Estimating observable distributions . . . . .	66
7.2.2	Algorithm Procedure . . . . .	66
7.2.3	Performance of Two Neutrino Scanning Method . . . . .	67
7.3	GNN in tetralepton . . . . .	68
7.3.1	Graph dataset . . . . .	68
7.3.2	Graph features . . . . .	69

7.3.3	Performance of tetralepton GNN	71
7.4	Tetralepton Asimov Fit Results	73
7.5	Signal injection test	75
<b>8</b>	<b>Combined analysis</b>	<b>76</b>
8.1	Combined trilepton and tetralepton Asimov fit results	76
8.2	High luminosity estimations of $tWZ$ measurements	78
<b>9</b>	<b>Conclusion</b>	<b>80</b>

## Abstract

The production of a single top quark in association with a  $W$  boson and a  $Z$  boson ( $tWZ$ ) is a rare Standard Model process that has never before been observed. The  $tWZ$  process will be a useful input into global Standard Model Effective Field Theory (SMEFT) fits due to its sensitivity to top-electroweak SMEFT coefficients in regions of high  $W$  boson and  $Z$  boson transverse momenta. The process is also a relevant background in other top quark related measurements, such as the  $t\bar{t}Z$  cross section measurement. In this thesis, a search was performed for  $tWZ$  production using  $140\text{ fb}^{-1}$  of  $\sqrt{s}=13\text{ TeV}$  proton-proton collision data measured using the ATLAS detector. The search targets the trilepton and tetralepton final states, where regions were defined using physics object multiplicities. Graph neural networks were employed to perform signal-background discrimination. The signal strength of  $tWZ$  production  $\mu_{tWZ}$  was extracted using a profile likelihood fit with a full systematic model describing experimental and modelling uncertainties. Blinded measurements of  $\mu_{tWZ}$  are shown for each decay channel using an Asimov dataset as well as a combined signal extraction across both channels. The expected significance for the measurement of  $tWZ$  was determined to be  $1.34\sigma$  where the precision of the measurement was limited by both systematic and statistical sources of uncertainty. Estimations for the measurements of  $\mu_{tWZ}$  for higher luminosity datasets were also investigated.

# Chapter 1

## Introduction

The field of experimental particle physics measures data from large high energy particle colliders and make comparisons to theoretical model of physics to create a better understanding of the fundamental building blocks of the universe. The field depends enormously not just on the experimentalist but on the labour of the specialists who make physics discoveries possible. No data can be taken unless the engineer and the construction worker can successfully build a high energy particle accelerator and a world-class detector. No testable physics model would exist without the theorist grinding away at a chalkboard or a Mathematica notebook. This thesis presents the tail end of a long journey that has allowed for the probing of unexplored regions of the Standard Model of particle physics.

The production of a single top quark with an associated  $W$  boson and a  $Z$  boson ( $tWZ$ ) is an unobserved top electro-weak process. The production of  $tWZ$  is predicted by the Standard Model of particle physics, but no successful observation of the process has been made. A search for the  $tWZ$  production has only been possible in the last half decade due to the availability of high energy proton-proton collisions provided by the Large Hadron Collider (LHC). This dissertation identifies the trilepton and tetralepton decay channels of  $tWZ$  as areas of interest with the analysis in each channel inspired by the previous work by fellow University of Cape Town students [1, 2]. The work included in this dissertation combines the measurements in both decay channels and presents a blinded measurement of  $tWZ$  production performed at the ATLAS experiment.

The measurement of  $tWZ$  production is preceded by additional background context before the experimental analysis and results are presented. The background context includes a short overview of the Large Hadron Collider and the ATLAS experiment in Chapter 2. The Standard Model of particle physics and the motivations for studying  $tWZ$  production are detailed in Chapter 3. Graph neural networks are discussed in Chapter 4

as these types of models are employed to provide signal-background discrimination. The portions of the dissertation associated with the measurement include an overview of the analysis strategy in Chapter 5, the details of the trilepton decay channel in Chapter 6, the details of the tetralepton decay channel in Chapter 7 and a combined measurement in Chapter 8.

# Chapter 2

## LHC and the ATLAS detector

### 2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) [3] is a 27 km long circular particle accelerator and collider located at the European Organization for Nuclear Research (CERN) in Geneva, Switzerland. Two rings of counter-rotating beams of hadrons are accelerated using a superconducting magnet system and collide in multiple experiments at different points throughout the LHC. At the LHC, both proton-proton and heavy ions collisions are performed, allowing for a wide physics program. From the initial commissioning of the LHC in 2008, the LHC has experienced many upgrades which have drastically increased the energy of the colliding beams. The initial Run 1 data taking period ran from 2010 to 2012 where proton-proton collisions had a centre of mass energy of 7 TeV in 2010 and 2011 and a centre of mass energy of 8 TeV in 2012. After upgrades to the LHC magnets, the proton beams during the Run 2 operation period between the years of 2015 and 2018 had an operational centre-of-mass energy of 13 TeV. Presently, the Run 3 operational period is underway where the LHC produces proton-proton collisions with higher energies of 13.6 TeV and a higher instantaneous luminosity.

The high energies at the LHC are achieved using an injection chain comprised of other lower energy accelerators. The hadrons are first accelerated in the linear accelerator Linac4 and then passed through a chain of circular accelerators: the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS). Each accelerator increases the energy of the hadrons before injecting them into the next accelerator.

The LHC is the home of 4 large detector experiments with varying physics programs. The ATLAS experiments and CMS experiments are general experiments that measure interactions at the highest possible luminosity to explore many areas of the Standard

Model and to search for new physics. The other two more specialised detectors are the LHCb experiment which is focused on investigating physics associated with  $B$  hadrons and the ALICE experiment which is dedicated to measuring heavy ion collisions.

### 2.1.1 Luminosity at the LHC

At the LHC, proton-proton collisions can be performed with an instantaneous luminosity of  $10^{34} \text{ cm}^{-1}\text{s}^{-1}$  at collision points inside the ATLAS and CMS experiments. Protons are placed into bunches of  $10^{11}$  protons where bunches can be collided up to 40 million times per second. A simplified expression for the instantaneous luminosity at the LHC is

$$L = \frac{N_b^2 n_b f_{\text{rev}} \gamma}{4\pi \epsilon_n \beta^*} F \quad (2.1)$$

where  $N_b$  is the number of particles per bunch,  $n_b$  is the number of bunches per beam,  $f_{\text{rev}}$  is the revolution frequency,  $\epsilon_n$  is the normalised transverse beam emittance and  $\beta^*$  is the beta function at the collision point. An additional geometric reduction factor of  $F$  is included which is dependent on the crossing angle of the beams. The physical units of instantaneous luminosity are derived from  $f_{\text{rev}}$  which has units of frequency and the  $b^*$  function which is proportional to the variance of the beam width in the transverse plane.

The total amount of data collected at the LHC is expressed in terms of integrated luminosity  $\mathcal{L} = \int L(t) dt$  which has dimensions of inverse area. For a scattering process of  $pp \rightarrow X$  with a cross section of  $\sigma$ , the total number of expected events can be determined by  $N = \sigma \cdot \mathcal{L}$ .

## 2.2 The ATLAS detector

The ATLAS (A Large Toroidal ApparatuS) detector is located around interaction point 1 at the LHC and detects both proton-proton collisions and heavy ion collisions. Much of the design for the ATLAS detector is detailed in the ATLAS design report [4] but an overview will be provided in this section. The detector comprises multiple sub-detectors for measuring different physics objects. High precision measurements can be performed by combining the signatures of energetic tracks through these various systems. The components of the detector are labelled in a cutaway of the detector in Figure 2.1.

### 2.2.1 Coordinate system at ATLAS

The coordinate system used by the ATLAS detector is introduced in this section to orient the locations of reconstructed physics objects with respect to the interaction point

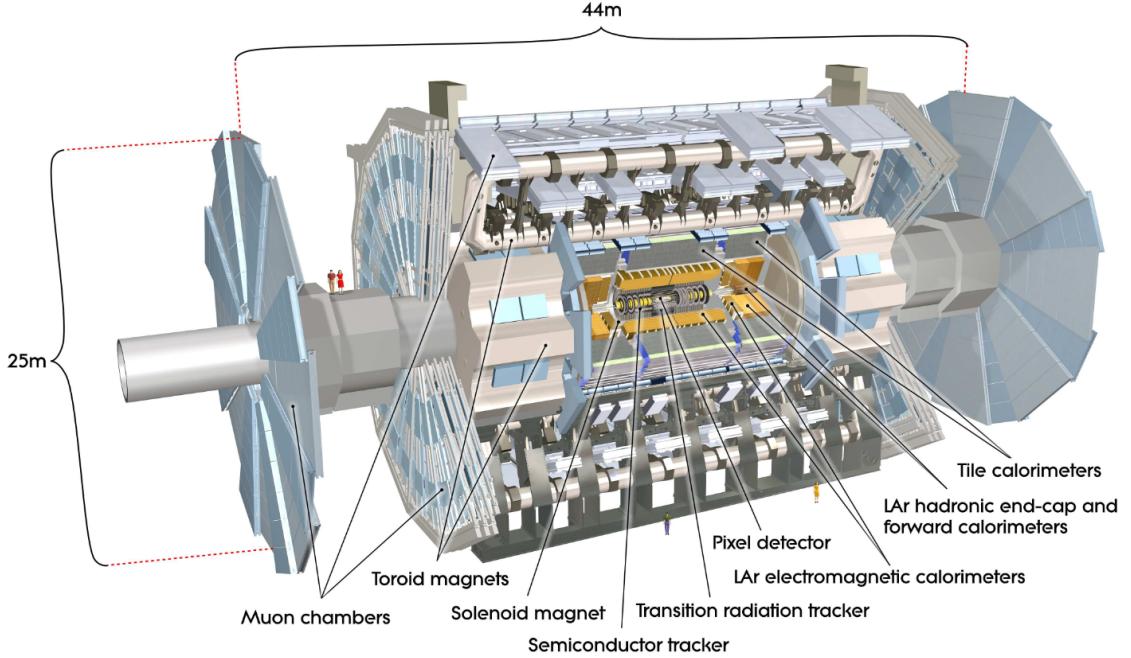


Figure 2.1: A cutaway of the ATLAS detector with various sub-detector components labelled [4].

as well as describe the coverage of the various ATLAS subdetectors. The origin of the coordinate system is the nominal interaction point between the two  $pp$  beams where the beamline defines the  $z$ -axis and the transverse  $x$ - $y$  plane. The positive  $x$ -axis points from the interaction point towards the centre of the LHC and the positive  $y$ -axis is upwards with respect to the plane formed by the LHC ring. The azimuthal angle  $\phi$  in the transverse plane is defined as the angle around the beamline with respect to the positive  $x$  axis. The polar angle  $\theta$  is the angle with respect to the beamline in the positive  $z$  direction. The pseudorapidity  $\eta$  coordinate is defined as

$$\eta = -\ln \left( \tan \left( \frac{\theta}{2} \right) \right) \quad (2.2)$$

which is preferred over the polar angle in hadron colliders as the distribution of jet production with respect to pseudorapidity is approximately constant [5]. Since jet production is approximately constant in  $\phi$  and  $\eta$ , the metric  $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$  in the pseudorapidity-azimuthal space is commonly used to describe the angular distances between objects in the detectors coordinate system.

### 2.2.2 Inner Detector

The Inner Detector (ID) comprises three subcomponents that contribute to the momentum measurements of charged particles as well as the primary and secondary vertex reconstruction for charged particles. The three components are the Pixel detector, the

silicon microstrip trackers (SCT) and the Transition Radiation Tracker (TRT) which are layered in the transverse direction. The pixel and SCT provide precision tracking at small radii closer to the beamline with a coverage of  $|\eta| < 2.5$ . The TRT is placed the furthest away from the beamline and provides tracking at large radii with a coverage of  $|\eta| < 2.0$ . The tracking provided by the ID allows for particle identification and reconstruction when matched with deposits in the calorimeters and the muon systems.

### 2.2.3 Calorimeters

The calorimetry system includes the electromagnetic (EM) and hadronic calorimeters that differ in their coverage and targeted physics objects. The EM calorimeter is designed to measure electrons and photons using a high granularity lead/liquid-argon scintillating detector which consists of a barrel component that covers the region of  $|\eta| < 1.475$  and two end-cap components that cover the region of  $1.375 < |\eta| < 3.2$ . The parts of the EM that are covered by the ID in the range of  $|\eta| < 2.5$  can provide precision measurements of electrons and photons by using the tracks in the ID.

The hadronic calorimeters are placed behind the EM calorimeter and include the tile calorimeter and the Hadronic End-cap Calorimeter (HEC) as well as the forward calorimeters (FCal). The lower granularity of the hadronic calorimeters allows for jet and missing transverse momentum reconstruction. The tile calorimeter covers the barrel region of  $|\eta| < 1.7$  using scintillating tiles as the active material. The HEC covers the end-cap region of  $1.5 < |\eta| < 3.2$ . The FCal modules are placed close to the beamline in the range of  $3.1 < |\eta| < 4.9$  to cover forward measurements.

### 2.2.4 Muon Spectrometer

The Muon Spectrometer (MS) is placed on the outer radius of the ATLAS detector as muons typically pass through the other subdetectors. The muon system comprises a magnet system, components for precision measurement and components for triggering. The MS has three toroidal magnets which generate the magnetic field required to bend particles in the various muon chambers. The Monitored Drift Tube (MDT) supplies precision measurements of track coordinates over the range of  $|\eta| < 2.7$  while the Cathode Drift Tubes (CDT) has a higher granularity and covers the region  $2.0 < |\eta| < 2.7$ . The Resistive Plate Chambers (RPC's) in the barrel region and the Thin Gap Chambers (TGC's) in the end-cap regions provide triggering for muon measurements. The triggering system covers  $|\eta| < 2.4$ .

### **2.2.5 Trigger and data acquisition system**

The trigger system is divided into three levels, the Level-1 (L1), Level-2 (L2) and event filter where the L2 and event filter are known as the High Level Trigger (HLT). The L1 trigger identifies interesting physics signatures such as leptons and jets as well as events with high missing transverse energy by using information from the various sub-detectors. The L1 system uses custom electronics and allows for a readout with an accept rate of 100 kHz. The L2 trigger limits the amount of data that is read out by looking at trigger objects identified by the L1 trigger. The event filter then considers fully built events and reduces the overall rate such that the events can be recorded for offline physics analysis. The L2 and event filter use regular computing and network hardware due to the more involved processing and the lower throughput requirement.

# Chapter 3

## Theoretical Background

### 3.1 The Standard Model

The Standard Model is an extremely successful model of particle physics that describes a set of elementary particles and the interactions between these particles [5]. The Standard Model combines several quantum field theories to produce a mathematical description of fundamental particles. The Standard Model describes two families of particles known as fermions and bosons each defined by their spin.

#### 3.1.1 Fermions

Fermions are half-integer spin particles which make up most matter in the universe. Each fermion is a member of one of three generations which only differ by a change in the particle's mass. All fermions have an associated antiparticle which has the same physical properties as the original particle, except it has an oppositely signed electric charge. Antiparticles are denoted with a bar over the original particle's symbol. Fermions are further categorised into leptons and quarks.

Leptons consist of three generations of charged leptons, the electron  $e$ , muon  $\mu$  and tauon  $\tau$ . There are also three generations of neutral leptons, the electron neutrino  $\nu_e$ , muon neutrino  $\nu_\mu$  and tauon neutrino  $\nu_\tau$ . The charged leptons have a negative electrical charge of  $-e$  or  $-1.6 \times 10^{-19}$  C and interact with other particles through the electromagnetic or weak force. The neutrinos are electrically neutral and only interact via the weak force.

Quarks are the only fermions that possess a colour charge, allowing them to interact via the strong force. Quarks are classified as up-type or down-type quarks. The class of up-type quarks includes three generations, the up quark  $u$ , charm quark  $c$  and top quark  $t$ . The up-type quarks have an electrical charge of  $+\frac{2}{3}$ . The down-type quarks include the down quark  $d$ , strange quark  $s$  and bottom quark  $b$  with an electrical charge of  $-\frac{1}{3}$ . As

quarks have a colour charge, they cannot exist in isolation due to colour confinement and must form composite particles known as hadrons. These hadrons are colour neutral and have different quark configurations which can be classified as baryons  $qqq$ , anti-baryons  $\bar{q}\bar{q}\bar{q}$  and mesons  $q\bar{q}$ .

### 3.1.2 Bosons and Interactions

Interactions between fundamental particles in the Standard Model are facilitated by the exchange of gauge boson particles. Gauge bosons are particles with integer spin and are each associated with a force. The photon  $\gamma$  mediates the electromagnetic force, the  $W^\pm$  and  $Z$  bosons mediate the weak force and the gluon  $g$  mediates the strong force. The interactions between bosons and fermions have their associated coupling constants which determine the rate at which these interactions occur.

The transition from an initial set of particles to another set of particles through various interactions is known as a quantum scattering process. Each quantum scattering process has an associated matrix element  $\mathcal{M}_{fi}$  which describes the probability of the initial quantum state  $i$  transitioning to a given final state  $f$ . The matrix elements can be encoded as Feynman diagrams. The matrix element for a process cannot be measured experimentally and therefore must be related to the cross section  $\sigma$  of the process.

### 3.1.3 Shortcomings of the Standard Model

The Standard Model generates many definite predictions that are in agreement with the precision measurements performed at modern particle accelerators. However, there are phenomena unexplained by the model that motivate studying physics beyond the Standard Model (BSM). The Standard Model does not explain the matter-antimatter asymmetry observed in the universe. In the Standard Model, matter and antimatter are mostly treated equally except for minor instances of Charge-Parity violation, but this would not account for the disparity seen in nature. There may be larger CP-violating effects introduced by BSM physics. The existence of dark matter and dark energy is another dilemma. Cosmologists have identified that the universe comprises ordinary Standard Model matter, dark matter and dark energy according to the  $\Lambda$ CDM model [6]. The dark matter and dark energy could be explained with extensions to the Standard Model such as supersymmetric theories, but no deviations from the Standard Model have been observed. The Standard Model also does not describe gravity.

## 3.2 The Top quark

The top quark is the most massive particle in the Standard Model and has a mass of  $m_t = 172.76 \pm 0.30$  GeV [7]. The top quark has a very short mean lifetime, meaning it decays before it experiences colour confinement, forming hadrons. The top quark almost always decays into a  $W$  boson and a  $b$  quark, allowing for a consistent experimental signature as the  $V_{tb}$  term is almost unity in the CKM matrix.

Top quarks are produced at leading orders of interactions through  $t\bar{t}$  pair production and single top production. In proton-proton collisions at the LHC,  $t\bar{t}$  pair production is more common. The leading order processes of  $t\bar{t}$  pair production are seen in Figure 3.1 which include two QCD vertices. The leading order single top production modes are shown in Figure 3.2 which require one or two electro-weak vertices. The electro-weak vertices in single top production introduce electro-weak coupling constants, reducing the cross section of this production mode compared to  $t\bar{t}$  pair production.



Figure 3.1: Examples of leading order Feynman diagrams for top quark pair production through  $gg \rightarrow t\bar{t}$  and  $q\bar{q} \rightarrow t\bar{t}$ .

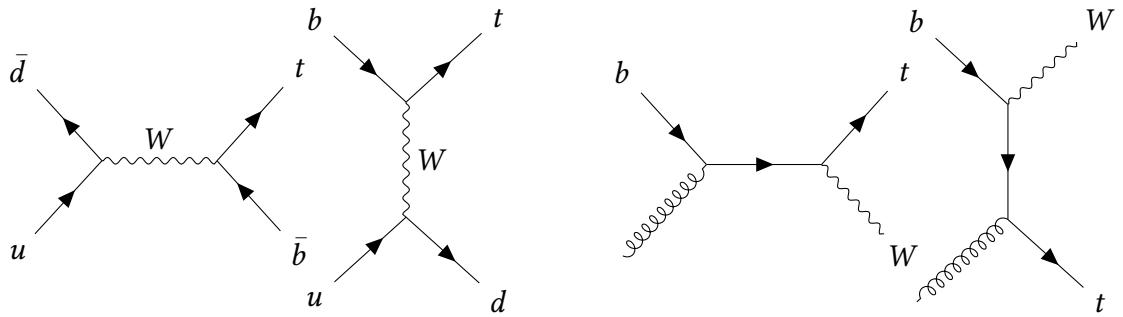


Figure 3.2: Leading order Feynman diagrams for single top quark production. The  $s$ -channel and  $t$ -channel single top production are shown on the left and single top production modes through  $tW$  production are shown on the right.

The LHC is a top quark factory as it can produce large amounts of top quarks due to the high centre-of-mass energy  $\sqrt{s}$  in proton-proton collisions. The cross sections for the various top production modes are proportional to  $\sqrt{s}$ . The large amount of data has allowed for precision measurements of the properties of the top quark as well as the observation of multitop production and other SM top-related processes. The high mass

of the top quark and its availability at the LHC have made the top quark an interesting area to search for physics beyond the Standard Model.

### 3.3 The $tWZ$ process

The production of a single top in association with a  $W$  boson and a  $Z$  boson ( $tWZ$ ) is an extremely rare Standard Model process and has never been observed. The process requires the production of a single top quark as well as two additional weak vertices to produce a  $W$  boson and a  $Z$  boson. Due to the introduction of these weak couplings, the expected cross section of the  $tWZ$  final state is heavily suppressed. The cross section when the  $Z$  boson is required to decay into two leptons is approximately  $(16.10 \pm 0.98) \text{ fb}$  [8] at next-to-leading order in QCD [9].

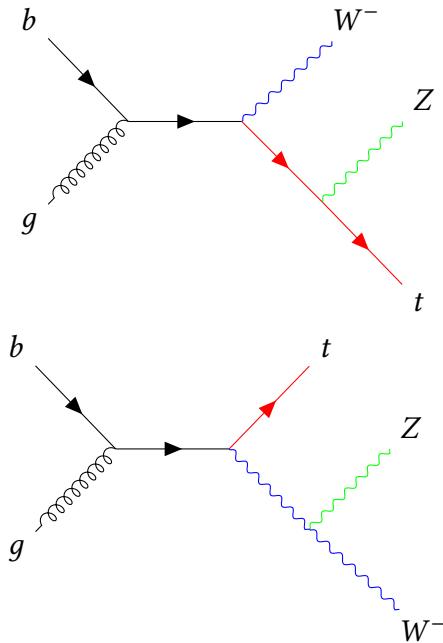


Figure 3.3: Leading order Feynman diagram of  $tWZ$  production. The diagrams are related to leading order  $tW$  production diagrams except with a final state emission of a  $Z$  boson.

A major difficulty in measuring  $tWZ$  is the similarity between the final state of  $tWZ$  and  $t\bar{t}Z$  production. The  $tWZ$  and  $t\bar{t}Z$  processes have similar decay signatures at leading-order, only differing by a single  $b$ -quark. These challenges are enhanced at higher orders as will be discussed in Section 3.3.1. Though the  $tWZ$  process is rare, other rarer top processes have been observed using the currently available dataset of  $pp$  collisions. For example, the production of  $t\bar{t}t\bar{t}$  has a SM cross section of only 12.0 fb but has recently been observed by the CMS collaboration [10] and the ATLAS collaboration [11].

The  $tWZ$  process may provide insight into BSM physics, where the  $tWZ$  process is seen as an ideal location to identify deviations from SM predictions in top-electroweak in-

teractions at high energies [12]. These deviations are typically phrased in terms of the framework of Standard Model Effective Field Theory (SMEFT) [13]. The framework introduces additional BSM higher-dimensional operators into the SM Lagrangian, each with associated Wilson coefficients. The production of  $tWZ$  is particularly sensitive to certain combinations of SMEFT operators at high energies [14]. These combinations are unique compared to the effects of other top-electroweak production modes such as  $t\bar{t}Z$ . Therefore, the ability to simultaneously constrain multiple operators makes the measurement of  $tWZ$  a promising input into global fits for SMEFT coefficients. The  $tWZ$  process is also a source of background contributions in other top quark-related measurements, particularly the measurement of the cross section of  $t\bar{t}Z$  production [15].

### 3.3.1 Diagram Removal Scheme

A hurdle in the measurement of the cross section of  $tWZ$  production is the similarity between the diagrams of  $tWZ$  and  $t\bar{t}Z$  at high orders of QCD interactions [16, 17]. If the additional  $Z$  emission is ignored, then the differences in the diagrams of  $tW$  and  $t\bar{t}$  production at next-to-leading orders of QCD [18] are considered. The overlap between  $tW$  and  $t\bar{t}$  production becomes challenging when considering the production of  $tW$  with a  $b$  emission ( $tWb$ ) as both  $tWb$  and  $t\bar{t}$  produce the same  $WbWb$  final state. However,  $tWb$  and  $t\bar{t}$  production differ in the number of resonant top quarks produced. The  $tWb$  diagram contributes a component with a single resonant top quark to the  $WbWb$  final state while the  $t\bar{t}$  diagram contributes a doubly resonant component. Examples of these diagrams are shown in Figure 3.4 where both processes produce the same  $WbWb$  final state but with different numbers of resonant top quarks.

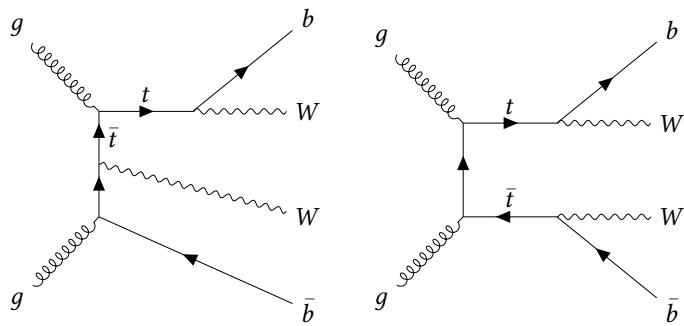


Figure 3.4: Examples of a diagram showing  $gg \rightarrow WbWb$  with a single resonant top quark (left) denoted by  $\mathcal{M}_{1t}$  and a doubly resonant diagram (right) denoted by  $\mathcal{M}_{2t}$ .

To account for the diagram overlap, the DR1 and DR2 diagram removal schemes are used to modify which diagrams contribute to the final state. The Diagram Subtraction (DS) scheme [19] can also handle this overlap by modifying the cross section of the production, but this scheme was not considered for this thesis. The amplitude of the

final state appears as,

$$\begin{aligned} |\mathcal{M}_{WbWb}|^2 &= |\mathcal{M}_{1t} + \mathcal{M}_{2t}| \\ &= |\mathcal{M}_{1t}|^2 + 2 \operatorname{Re}(\mathcal{M}_{1t} \mathcal{M}_{2t}^*) + |\mathcal{M}_{2t}|^2 \end{aligned}$$

where  $\mathcal{M}_{1t}$  is the amplitude of the diagrams with a single resonant top quark and  $\mathcal{M}_{2t}$  is the amplitude of the diagrams with two resonant top quarks. The overlap term emerges with both a  $\mathcal{M}_{1t}$  and  $\mathcal{M}_{2t}$  contribution. The DR1 diagram removal scheme ignores the overlap term

$$|\mathcal{M}_{DR1}|^2 = |\mathcal{M}_{1t}|^2, \quad (3.1)$$

and the DR2 diagram removal scheme opts to include the term

$$|\mathcal{M}_{DR2}|^2 = |\mathcal{M}_{1t}|^2 + 2 \operatorname{Re}(\mathcal{M}_{1t} \mathcal{M}_{2t}^*). \quad (3.2)$$

In the analysis, both options are considered. The nominal DR1 scheme is compared to the alternative DR2 scheme and the variation between them is introduced as a source of modelling uncertainty.

# Chapter 4

## Graph Neural Networks

A graph neural network (GNN) is a type of multivariate machine learning model whose inputs are in the form of a graph data structure. In particle physics, a graph can be used as a viable abstraction of a set of particles in an event. This chapter provides a short overview of machine learning models in Section 4.1. The details of graph neural networks, their applications in high energy physics and the model architecture that was employed in this thesis are discussed in Section 4.2.

### 4.1 Overview of Machine Learning

Machine learning is a classification of statistical models which use data to educate the model output. These models include least-squares regression models, Boosted Decision Trees (BDTs) and Neural Networks (NN). All these models are constructed to try to solve an empirical risk minimization problem [20]. Assume the problem has a set of input data  $\mathbf{x}$  and output data  $\mathbf{y}$  with some dimensionality, where a function  $f$  is required in which  $f(\mathbf{x}) = \mathbf{y}$ . The problem requires finding the function  $f$  which minimises the differences between the predictions of the model  $\hat{\mathbf{y}} = f(\mathbf{x})$  and the true target  $\mathbf{y}$ . This is expressed with a loss function  $L(f(\mathbf{x}), \mathbf{y})$  which approaches 0 as  $f(\mathbf{x})$  approaches  $\mathbf{y}$ . The loss function is adjusted to match the problem.

An optimal solution to this problem would be a function  $\Phi'$  which minimises the average loss,

$$\Phi' = \operatorname{argmin}_{\Phi} \int L(\Phi(\mathbf{x}), \mathbf{y}) dp(\mathbf{x}, \mathbf{y}) \quad (4.1)$$

where  $p(\mathbf{x}, \mathbf{y})$  is the probability density function from which the input data  $\mathbf{x}$  and targets  $\mathbf{y}$  are sampled. This shows an integral not over just the dataset on hand but over the probability density function of all possible  $\mathbf{x}$  and their true associated outputs  $\mathbf{y}$ .

Since the entire probability density function could be difficult to determine, the problem

is adapted to the available dataset and the weighted average empirical loss is used,

$$\Phi^* = \operatorname{argmin}_{\Phi} \frac{1}{\sum w_i} \sum_{i=1}^N w_i L(\Phi(\mathbf{x}_i), \mathbf{y}_i) \quad (4.2)$$

where  $N$  is the number of datapoints in the dataset and  $w_i$  is the associated weighting of the  $i^{\text{th}}$  datapoint. If the search is instead limited to a specific class of solution with a parametrisation  $\alpha$ , the model can be rewritten as  $\Phi(\mathbf{x}) = \phi(\mathbf{x}, A)$ . The problem can be rephrased to instead search for a parametrisation  $A$  of the function  $\phi$  which minimises the empirical loss,

$$\Phi^* = \operatorname{argmin}_A \frac{1}{\sum w_i} \sum_{i=1}^N w_i \cdot L(\phi(\mathbf{x}_i, A), \mathbf{y}_i) \quad (4.3)$$

There is freedom in the choice of the loss function  $L$  and the model  $\phi$  with a parametrisation  $A$ . The loss function is dictated by the original problem, but there is freedom in choosing the model. Once a model architecture is chosen, the parametrisation or weights of the model can be optimised to minimise to find  $\Phi^*$ . The next section describes graph neural network models in general as well as the model architecture of the machine learning models in the analysis.

## 4.2 Graph Neural Networks in Particle Physics

Neural networks have been successful in many areas of particle physics analysis [21] from their implementation in trigger systems to the discrimination of physics objects. Unfortunately, the structure of low-level information of particle physics events does not map onto the traditional data structures used in neural networks. Fully connected layers rely on vectors of features, convolutional layers use grids and recurrent layers require sequences of features. All these methods require placing some ordering on the input information which typically has some experimental underpinning but not a physical one. Since a particle physics event can be seen as an unordered set of physics objects with relations between them, this structure suggests that events can be seen as a set of graphs.

The benefits of GNNs in particle physics applications stem from the model's equivariance and transferability properties [22]. Equivariance allows for the model output to be invariant under the permutation of nodes as no ordering is placed on the vertices. Transferability allows the GNN model to generalise across a variety of graph sizes which allows for graphs to be constructed out of arbitrary numbers of particles.

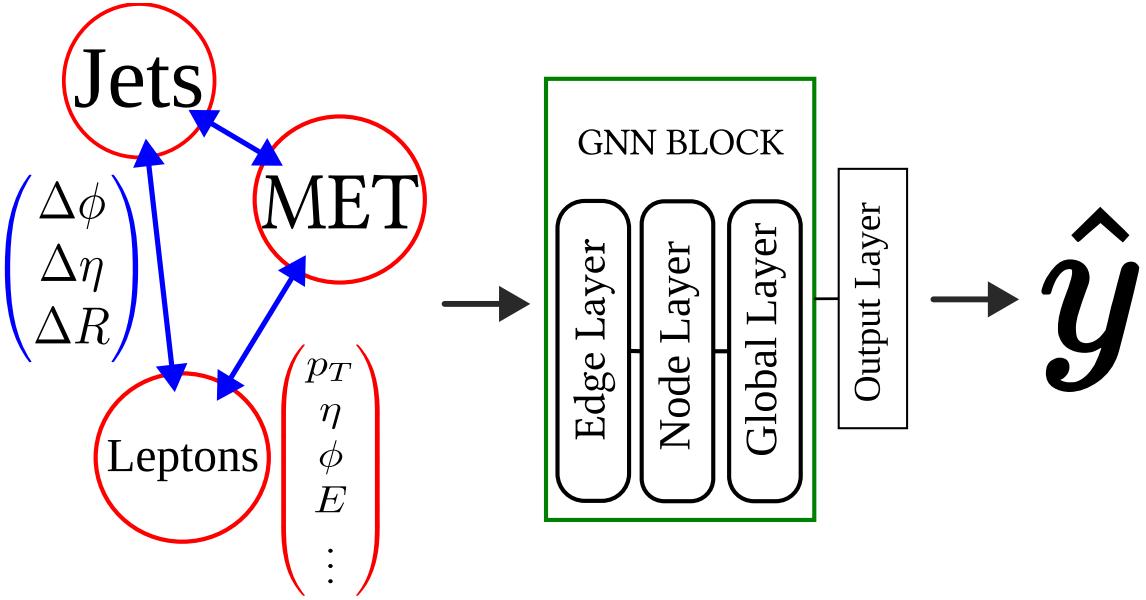


Figure 4.1: Diagram showing the classification of an event through the graph neural network.

#### 4.2.1 Outline of a graph network

The graph neural network will be explained using the *graph network* framework [23] that provides a generalised description of many possible graph neural network architectures. A graph is defined as  $G = (\mathbf{u}, V, E)$  where  $V$  is a set of vertices,  $E$  is a set of edges and  $\mathbf{u}$  is a vector of global features. The set of  $N_v$  vertices of the graph  $V = \{\mathbf{v}_{i=1,\dots,N_v}\}$  is a collection of feature vectors that describe each node. The edges between the nodes of the graph  $E = \{(\mathbf{e}_k, r_k, s_k)\}_{k=1,\dots,N_e}$  comprises a feature vector for the edge  $\mathbf{e}_k$ , the index of the node which is receiving the edge  $r_k$  and the index of the node which is the source of the edge  $s_k$ . The global attributes  $\mathbf{u}$  of the graph are a vector of properties associated with the graph itself, which will contain the graph-level features.

With the graph data structure defined, a *graph network block* can be established which includes an edge, a node and a global block where each block passes information into the next. A depiction of a full graph network block is shown in Figure 4.2a. The edge block contains the edge update function  $\phi^e$  that can use the information for the edges, the nodes and the global feature vector to update the edge features. These updated edges  $E'$  can then be passed to the node block and/or the global block. The node block first aggregates the updated edges on a per-node basis with the aggregation function  $\rho^{e \rightarrow v}(E')$ . The nodes are then updated to  $V'$  with the node update function  $\phi^v$  using the aggregated edge information and the current node vectors. The final block is the global block which first aggregates the edge and node information with the aggregations functions  $\rho^{e \rightarrow u}$  and  $\rho^{v \rightarrow u}$ . The global feature vector is updated via  $\phi^u$  using the current feature vector and the aggregating information. The final output of the block will be a new graph

$G' = (\mathbf{u}', V', E')$  with updated components. The graph network block is defined by the update  $\phi$  and aggregation  $\rho$  functions but not all graph network blocks require all these to be defined. The graph network block provides a general framework to describe graph neural networks.

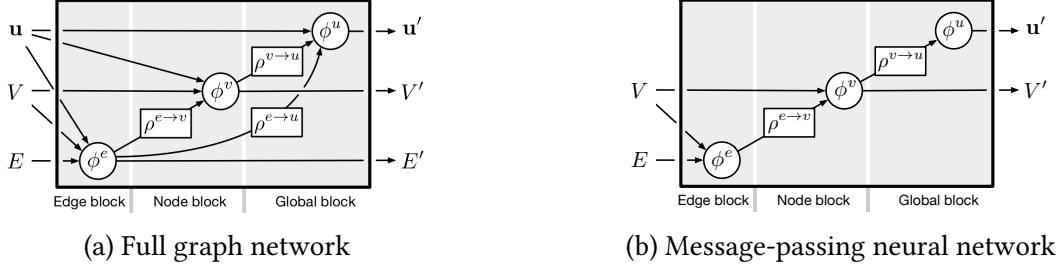


Figure 4.2: Subfigures showing different blocks of a graph network blocks. The details of the edge, node, and global blocks are depicted [23].

### 4.2.2 GNN Architecture

The architecture of the graph neural networks used in this thesis is structured around a modified message passing neural network architecture [24]. The graph network block for a message-passing neural network is shown in Figure 4.2b. The update functions were implemented as fully connected neural networks (denoted by  $NN$ ) with an output dimensionality that matches the required dimensions for the block. The blocks will be discussed using the formalism detailed in Section 4.2.1.

The edge block has the edge update function  $\phi^e(\mathbf{e}_k, \mathbf{v}_{r_k}, \mathbf{v}_{s_k}, \mathbf{u}) = NN_e([\mathbf{e}_k, \mathbf{v}_{r_k}, \mathbf{v}_{s_k}, \mathbf{u}])$  which uses the edge features, the global features as well as the node vectors at both ends of the edge. The updated features for edge  $k$  is then  $\mathbf{e}'_k = NN(\mathbf{e}_k, \mathbf{v}_{r_k}, \mathbf{v}_{s_k}, \mathbf{u})$ . The node block aggregates the updated edges using  $\rho^{e \rightarrow v}(E'_i) = \frac{1}{N_i} \sum_{\{k: s_k=i\}} \mathbf{e}'_k$  which takes the average of the edges which terminate at the node  $i$ . The aggregated value for the node  $i$  is then denoted by  $\bar{\mathbf{e}}'_i$  which is passed to the update function  $\phi^v(\bar{\mathbf{e}}_i, \mathbf{v}_i, \mathbf{u}) = NN_v([\bar{\mathbf{e}}_i, \mathbf{v}_i, \mathbf{u}])$ . The global block aggregates the updated node vectors in a single vector  $\bar{\mathbf{v}}'$  using  $\rho^{v \rightarrow u}(V') = \frac{1}{N_i} \sum \mathbf{v}_i$  which takes the average of the node vectors. Finally, the global block updates the global feature vector using  $\phi^u(\bar{\mathbf{v}}', \mathbf{u}) = NN_u([\bar{\mathbf{v}}', \mathbf{u}])$ .

Multiple graph blocks can be stacked to produce a deep graph network, where each layer produces a new graph with updated features. An output layer may be required to adjust the final updated global vector  $\mathbf{u}'$  to match the dimensionality of the target vector. Since the relevant models in this thesis are binary classification models for signal-background discrimination, the output dimensionality is a single number representing the probability of an event to be a signal event. A classification with multiple classes is possible by adjusting the final output layer.

For a classification task such as signal-background discrimination, it is common to use cross entropy as a loss function. The binary cross entropy for two classes is defined as,

$$L(\hat{y}_i, y_i) = -w_i \cdot [y_i \ln(\hat{y}_i) + (1 - y_i) \ln(1 - \hat{y}_i)] \quad (4.4)$$

where  $\hat{y}_i = \phi(x_i, A) \in [0, 1]$  is the model output for the input  $x_i$  and  $y_i \in \{0, 1\}$  is the target associated with the input  $x_i$ . The model prediction is commonly restricted between 0 and 1 by a sigmoid function.

### 4.2.3 Application and limitations of graphs in HEP

The applications of GNNs typically replace existing machine learning strategies to allow for a more natural data structure. A few examples will be mentioned in this section but a more extensive overview is available [25].

In this thesis, graph neural networks were produced to discriminate between signal and background events. Other analyses have also applied GNNs in this way where physics objects in an event are treated as nodes in a graph. The notable application was the observation of four top production with the ATLAS detector [11] which involved the use of a graph neural network to perform signal-background classification. The GNN was used to separate the signal  $t\bar{t}t\bar{t}$  events from other tetralepton background contributions. A GNN was also employed as a proof of concept for performing signal-background discrimination in a probe for stop pair production [26]. The GNN was able to outperform a deep neural network and allow for an enhanced mass reach of the supersymmetric top mass. At the IceCube experiment, a graph convolution network was designed to perform signal identification in order to identify neutrinos from high energy cosmic muons [27]. The model was evaluated using simulated data and it outperformed a traditional convolutional neural network.

The identification of jet flavours is seen as an ideal classification task for GNNs. A state-of-the-art jet tagging algorithm built using a GNN is planned to be implemented in ATLAS analyses [28] and shows improvement over current jet tagging models. The ParticleNet [29] is another GNN that has been applied to jet tagging problems.

Point cloud graph neural networks were also shown to improve the reconstruction of pions in the ATLAS calorimeters over other baselines [30]. The reconstruction involved the classification of  $\Pi^0$  and  $\Pi^\pm$  as well as pion energy regression.

A novel use of graph neural networks is the topological reconstruction of particles [31]. The approach was applied to the reconstruction of  $t\bar{t}$  pairs in the all-hadronic final state and performed better than the current ATLAS reconstruction methods and similarly to the state-of-the-art machine learning techniques.

Graph networks can provide better model performance over other machine learning methods but this comes at the cost of a more complex data structure. Compared to vectors, grids or sequences, graphs require more effort to create and handle. Additional processing time is introduced when applying the model, as a graph must be generated from the information of a collection of particles. The graph data structure is also larger, putting more strain on system memory requirements. If an analysis team does not have access to suitable computing resources, a graph network approach may not be feasible and hinder progress in the analysis.

# Chapter 5

## Search for $tWZ$ production

The search for the production of a single top quark in association with a  $W$  boson and a  $Z$  boson at the ATLAS experiment was performed using  $140\text{ fb}^{-1}$  of  $\sqrt{s}=13\text{ TeV}$  proton-proton collisions. The aim of this search is to measure the signal strength of  $tWZ$  production  $\mu_{tWZ}$  which is defined as

$$\mu_{tWZ} = \frac{\sigma_{\text{obs}}(tWZ)}{\sigma_{\text{SM}}(tWZ)}, \quad (5.1)$$

where  $\sigma_{\text{obs}}(tWZ)$  is the measured cross section of  $tWZ$  production and  $\sigma_{\text{SM}}(tWZ)$  is the theoretical prediction of the cross section according to the Standard Model. A signal strength of  $\mu_{tWZ} = 0$  corresponds to the BSM scenario with no  $tWZ$  production and  $\mu_{tWZ} = 1$  corresponds to  $tWZ$  production at the rate predicted in the Standard Model. A signal strength of  $\mu_{tWZ} > 1$  would indicate a BSM scenario of a larger  $tWZ$  production rate.

Various decay channels can be used to search for  $tWZ$  production but experimental signatures with less than three leptons have larger relative backgrounds than the channels with three or more leptons. Tau leptons are ignored in the search due to the poor reconstruction resolution of taus at the ATLAS detector. Therefore, the prompt  $Z$  boson must decay to either an electron-positron or a muon-antimuon lepton pair. The requirement of a leptonically-decaying  $Z$  boson rejects many trilepton and tetralepton background processes which do not produce a  $Z$  boson. The trilepton and tetralepton decay modes are shown in Figure 5.1 using an example of a leading order  $tWZ$  diagram with additional particle decays.

This chapter provides an overview of the analysis strategy employed in the measuring of signal strength  $\mu_{tWZ}$  in the trilepton and tetralepton channels.

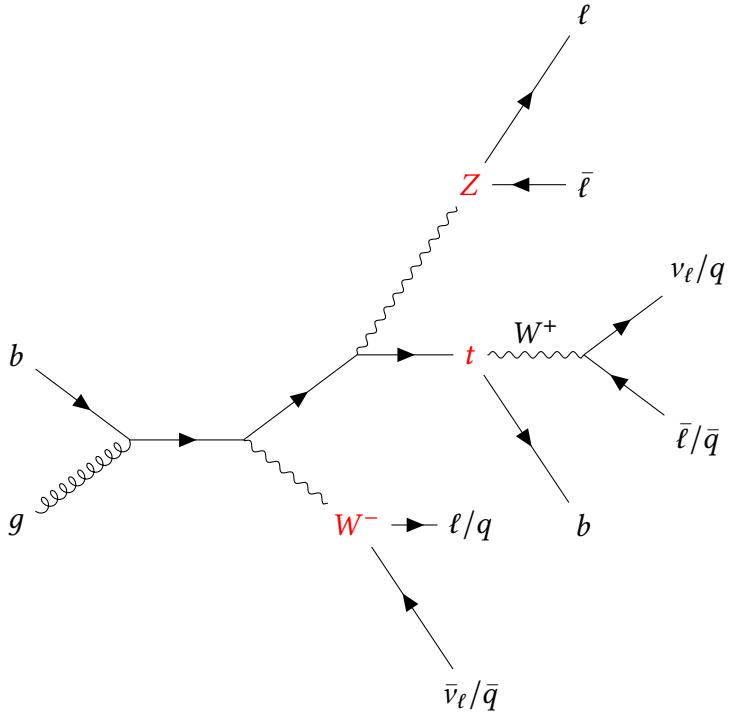


Figure 5.1: Diagram showing the various decay schemes of  $tWZ$  production using an example of  $gb \rightarrow tWZ$  production. The process that produces the  $tWZ$  final state in this example is  $tW$  production with a final-state radiation of a  $Z$  boson where the red labels highlight the final state particles. The  $Z$  boson is required to decay leptonically into an electron-positron or muon-antimuon pair. The prompt  $W$  boson and the  $W$  boson originating from the top decay can decay leptonically or hadronically.

## 5.1 Data and Monte Carlo Samples

The data used in this analysis were obtained from proton-proton collisions using the ATLAS detector at the Large Hadron Collider. The data was measured between the years of 2015 and 2018 at a centre-of-mass energy of  $\sqrt{s}=13$  TeV and will be referred to as the ATLAS Run 2 dataset. Only events from runs included in the official ATLAS Good Runs Lists (GRL) for 2015-2018 were considered. Runs in the GRL used all components of the ATLAS detector, had stable LHC beam conditions and passed ATLAS data quality standards. The integrated luminosity across the entire Run 2 data-taking period was  $140.1\text{ fb}^{-1}$  [32] with the values for each of the individual years detailed in Table 5.1.

In the analysis, simulated events were used to estimate contributions from background processes and provide comparisons between ATLAS data and Standard Model predictions. Monte Carlo (MC) events of the  $tWZ$  process and the various background processes present in the trilepton and tetralepton channel were simulated. More simulated events can be produced than there are observed events in the dataset to minimise statistical errors from the simulation. The simulated events for a process are weighted such that each event has an individual weight and the sum of all event weights is equal to the

Year	Int. Lumi. ( $\text{fb}^{-1}$ )
2015	3.2
2016	33.4
2017	44.6
2018	58.8
Total	140.1

Table 5.1: Table showing the amount of data recorded during each year of the ATLAS Run 2 data-taking period

cross section of the process. All event weights are scaled to an integrated luminosity of  $\mathcal{L}=140 \text{ fb}^{-1}$  so the simulation can be compared to the observed dataset. The effects of the detector on simulated events were estimated by passing the events through a full ATLAS detector simulation [33]. Pileup effects due to additional  $p\bar{p}$  collision within the same bunch crossing must be overlaid in the simulated events as these effects may differ between the observed data and the MC samples. A reweighting was performed to correct the differences between the pileup distributions of the simulated samples and data [34].

Samples were produced for modelling relevant top and multi-boson processes which produce a final state of two, three or four leptons. A full summary of the simulated samples used in this analysis can be found in Table 5.2. In all samples that involve processes with top quark production, the mass of the top quark was assumed to be 172.5 GeV. Where required, the mass of the  $W$  boson was set to 80.2 GeV and the mass of the  $Z$  boson was set to 91.2 GeV.

The samples for modelling  $tWZ$  production were generated using the `MADGRAPH5_AMC` event generator [9, 35] at next-to-leading order (NLO) accuracy in QCD. The parton showering and hadronisation was simulated by interfacing the samples with `PYTHIA8`. The A14 set of tuned MC parameters was used [36] and the Parton Distribution Function (PDF) calculations for the simulation were provided by the NNPDF3.0NNLO PDF set. Since the  $tWZ$  diagram at NLO in QCD has diagram overlaps with  $t\bar{t}Z$ , a diagram removal scheme is introduced as described in Section 3.3.1. The nominal  $tWZ$  sample uses the DR1 diagram removal scheme while an additional sample was produced which uses the DR2 diagram removal scheme.

Simulated samples of background processes were required for use in both the trilepton and tetralepton channel analyses. The background  $t\bar{t}Z$  process was simulated at NLO precision in QCD using `MADGRAPH5_AMC` as the event generator where parton showering and hadronisation was performed using `PYTHIA8`. The NNPDF3.0NLO PDF set and the A14 tune were used. A separate sample was produced for each leptonic decay scheme of the  $Z$  boson which includes the decay to an electron-positron pair, the decay to a

muon-antimuon pair and the decay to a tau-antitau pair. An additional alternative  $t\bar{t}Z$  sample was produced using the SHERPA event generator where parton showering calculations were handled by SHERPA using default SHERPA MC tune and the NNPDF3.0NNLO PDF set. The events in the  $t\bar{t}Z$  sample which decay via the trilepton channel also have the possibility of including a fake/non-prompt lepton such that the event is included in the tetralepton channel and will be labelled as the  $t\bar{t}Z$  fake sample in the tetralepton channel.

A sample of events involving single top quark production in association with a  $Z$  boson and an additional parton ( $tZq$ ) was generated using the MADGRAPH5\_AMC event generator at NLO with the NNPDF3.0NLO PDF set and the A14 MC tune. The sample includes off-shell  $Z$  bosons which decay into a dilepton pair with a mass of  $m_{\ell\ell} > 5$  GeV. Samples of diboson processes with three charged leptons and a neutrino ( $WZ+\text{jets}$ ) or four charged leptons ( $ZZ+\text{jets}$ ) were simulated using the same SHERPA configuration. The additional jets in the diboson events were simulated at NLO if one extra parton was present and at LO if two or three partons were present.

The  $WZ+\text{jets}$  samples are split into  $WZ+l$ ,  $WZ+c$  and  $WZ+b$  components depending on if the parton which produced the additional jets was a light ( $u$ ,  $d$  or  $s$ ) quark,  $c$  quark or  $b$  quark. The flavour information is obtained using the truth record from the event generator. If a  $WZ+\text{jets}$  event contains a  $b$ -jet in the truth record, it is considered a  $WZ+b$  event. If no  $b$ -jets are present and a  $c$ -jet is present, the event is considered to be a  $WZ+c$ . Otherwise, the event is a  $WZ+l$  event. However, this partitioning scheme allows for misclassification in the trilepton channel if a heavy-flavoured jet produces a non-prompt lepton and the prompt  $W$  boson decays to produce light jets. It is assumed that this misclassification will have a negligible effect on the overall rates of each component. The  $ZZ+\text{jets}$  samples are similarly partitioned into  $ZZ+l$ ,  $ZZ+c$  and  $ZZ+b$  components.

The samples for  $Z+\text{jets}$  production were simulated using SHERPA2.2.11 and the default SHERPA parton shower with the NNPDF3.0NNLO PDF set. The simulation could produce up to two jets at NLO accuracy and three more additional jets at LO accuracy. Simulated events of  $Z$  boson production with an associated photon ( $Z\gamma$ ) were produced using the SHERPA event generator at NLO level. The default SHERPA parton showering setup was used with the NNPDF3.0NNLO PDF set.

A set of samples to describe processes involving three vector bosons ( $WWW$ ,  $WWZ$ ,  $WZZ$  and  $ZZZ$ ) were generated using SHERPA2.2.2 with the NNPDF3.0NLO PDF set. The samples were simulated at NLO but if additional partons were present in the final state, this component was calculated at LO.

The production of a  $t\bar{t}$  pair in association with a  $W$  boson ( $t\bar{t}W$ ) was modelled at NLO using the SHERPA2.2.10 event generator and the default SHERPA parton showering where

the NNPDF3.0NNLOPDF set was used. The samples for  $t\bar{t}$ ,  $t\bar{t}H$  and  $tW$  were all simulated at NLO using the Powheg[37, 38, 39] generator and PYTHIA8 using the NNPDF3.0NLO PDF set and the A14 MC tune. The  $h_{\text{damp}}$  parameter in Powheg which regulates high- $p_T$  radiation is set to be 1.5 times the nominal top quark mass of 172.5 GeV.

The production of three top ( $t\bar{t}t$ ) and  $t\bar{t}$  pair with two  $W$  bosons ( $t\bar{t}WW$ ) was modelled at NLO using MADGRAPH5\_AMC interfaced with PYTHIA8. A sample of production of four top production ( $t\bar{t}t\bar{t}$ ) was simulated at NLO using SHERPA with the NNPDF3.0NNLO PDF set. The production of  $t\bar{t}$  events in association with photon ( $t\bar{t}\gamma$ ) was simulated at LO using MADGRAPH5\_AMC and then interfaced to PYTHIA8. The NNPDF3.0NNLO PDF set and the A14 MC tune were used.

Process	Generator	CS calc.	PS/Hadronisation	PDF	MC tune	$\sigma$ (pb)
$tWZ$ (DR1)	MADGRAPH5_AMC	NLO	PYTHIA8.235	NNPDF3.0NNLO	A14	0.0161
$tWZ$ (DR2)	MADGRAPH5_AMC	NLO	PYTHIA8.235	NNPDF3.0NNLO	A14	0.0147
$t\bar{t}Z$	MADGRAPH5_AMC	NLO	PYTHIA8.244	NNPDF3.0NLO	A14	0.12363
$t\bar{t}Z$ (Alt.)	SHERPA 2.2.11	NLO	SHERPA 2.2.11	NNPDF3.0NNLO	SHERPA	0.190081
$WZ+\text{jets}$	SHERPA 2.2.2	NLO	SHERPA 2.2.2	NNPDF3.0NNLO	SHERPA	7.5464
$ZZ+\text{jets}$	SHERPA 2.2.2	NLO	SHERPA 2.2.2	NNPDF3.0NNLO	SHERPA	2.7146
$tZq$	MADGRAPH5_AMC	NLO	PYTHIA8.230	NNPDF3.0NNLO	A14	0.038720
$t\bar{t}$ (dilepton)	POWHEG	NLO	PYTHIA8.230	NNPDF3.0NLO	A14	87.70761
$Z + \text{jets}$	POWHEG	NLO	PYTHIA8.186	NNPDF3.0NNLO	AZNLO	13880.72
$t\bar{t}W$	SHERPA 2.2.10	NLO	SHERPA 2.2.10	NNPDF3.0NNLO	SHERPA	0.597
$t\bar{t}\gamma$	MADGRAPH5_AMC	NLO	PYTHIA8.212	NNPDF3.0NLO	A14	5.36
$t\bar{t}H$	POWHEG	NLO	PYTHIA8.230	NNPDF3.0NLO	A14	0.50701
$VVV$	SHERPA 2.2.2	NLO	SHERPA 2.2.2	NNPDF3.0NNLO	SHERPA	0.01374
$t\bar{t}WW$	MADGRAPH5_AMC	LO	PYTHIA8.186	NNPDF2.3LO	A14	0.0099
$ZH$	POWHEG	LO	PYTHIA8.230	NNPDF3.0NLO	AZNLO	0.7610
$WH$	POWHEG	LO	PYTHIA8.230	NNPDF3.0NLO	AZNLO	1.401
$t\bar{t}t$	MADGRAPH5_AMC	LO	PYTHIA8.186	NNPDF2.3LO	A14	0.00164
$t\bar{t}t\bar{t}$	SHERPA 2.2.11	NLO	SHERPA 2.2.11	NNPDF3.0NNLO	SHERPA	0.011873
$tW$	POWHEG	NLO	PYTHIA8.230	NNPDF3.0NLO	A14	7.5553
$Z\gamma$	SHERPA 2.2.11	NLO	SHERPA 2.2.11	NNPDF3.0NNLO	SHERPA	298.74

Table 5.2: List of simulated processes with their associated event generators and parton showering/hadronisation programs. The PDF set and Monte Carlo tuning used during simulation are included as well as the theoretical cross section of the simulated process.

## 5.2 Physics object definitions

A physics object is a set of experimental signatures in the various subdetectors that aim to represent true objects produced in the proton-proton collision. The physics objects defined in this analysis are missing transverse momentum  $E_T^{\text{miss}}$ , jets and leptons of which

only electrons and muons are considered.

### 5.2.1 Leptons

The Electrons and muons are measured using different components of the ATLAS detector as discussed in Section 2.2. The lepton momenta is reconstructed through tracks found in the ID as well as clusters of energy deposits in the EM calorimeter for electrons and tracks in the MS for muons. Lepton objects can be defined as loose or tight leptons where tight leptons pass all the criteria required to be a loose lepton as well as some additional requirements. The scale and resolution of the electron energy measurements are calibrated in the simulated samples through corrections determined in an independent study [40]. The muon momentum scale and resolution calibrations are motivated in a similar manner [41].

Loose electrons must satisfy the conditions that  $p_T > 10 \text{ GeV}$  which avoids the misclassification of charged hadrons and the object must lie within  $|\eta| < 2.47$ . The object also cannot lie between  $1.37 < |\eta| < 1.52$  since this is the transition region between the barrel and the end-cap of the EM calorimeter which provides a poor energy resolution. Loose electrons must meet the Tight likelihood identification criteria [42, 43]. The track through the inner detector associated with the loose electron must have a longitudinal impact parameter  $z_0$  and transverse impact parameter  $d_0$  in which  $z_0 \sin(\theta) < 0.5 \text{ mm}$  and  $\frac{|d_0|}{\sigma(d_0)} < 5$  where  $\theta$  is the angle between the track and the  $+z$  axis and  $\sigma(d_0)$  is the uncertainty on the measurement of  $d_0$ . This condition reduces pileup signals from secondary vertices that are away from the primary interaction point.

Tight electrons are defined as loose electrons that have met an additional isolation requirement. The tight electron must satisfy the PromptLeptonImprovedVeto (PLIV) isolation condition which discriminates between prompt electrons and non-prompt/fake electrons by combining electromagnetic shower shapes and track information from the ID. Electrons must pass an electron/photon ambiguity requirement which reduces the selection of electrons that originate from photon conversions. The requirement evaluates if tracks are found close to the energy deposits from an electron candidate in the EM calorimeter and if there is any evidence of photon conversion vertices.

Loose muons must satisfy  $p_T > 10 \text{ GeV}$ ,  $|\eta| < 2.5$  and the medium identification selection in [44] which defines requirements on the number of hits in the MS and the charge to momentum ratio of the object. The impact parameters of the pre-selected muon track are required to have  $z_0 \sin(\theta) < 0.5 \text{ mm}$  and  $\frac{|d_0|}{\sigma(d_0)} < 3$ . Tight muons meet the loose requirements, as well as the same isolation condition as used for tight electrons. The definitions of lepton objects are summarised in Table 5.3.

Efficiency scale factors are calculated per lepton in an event to account for efficiency

differences between data and simulation. Electrons [43] have separate scale factors in order to apply corrections for electron reconstruction, identification and isolation. For muons [41], separate scale factors are used for muon identification, isolation and track-to-vertex association. The scale factors for the whole event are calculated as the product of the scale factors of each lepton in the event. All simulated events are reweighted by multiplying the Monte Carlo weight of the event by the scale factors.

Non-prompt leptons or some non-leptonic particles can pass the loose or tight lepton criteria and mistakenly be reconstructed as lepton objects. In simulation, these lepton objects can be compared to their associated object in the truth record and identified as non-prompt or fake leptons. The reconstruction of fake leptons can change the total number of lepton objects in an event, which can impact the event selection. In signal-sensitive regions, the requirement of tight leptons is preferred since fake leptons have a lower rate of being reconstructed as tight leptons.

	Loose Electron	Loose Muon
Identification Criteria	TightLH	Medium
Acceptance	$p_T > 10 \text{ GeV}$ $ \eta^{\text{clust}}  < 2.47$ except $1.37 <  \eta^{\text{clust}}  < 1.52$	$p_T > 10 \text{ GeV}$ $ \eta  < 2.5$
Impact Parameter	$ d_0 /\sigma(d_0) < 5$ $ z_0 \sin(\theta)  < 0.5 \text{ mm}$	$ d_0 /\sigma(d_0) < 3$ $ z_0 \sin(\theta)  < 0.5 \text{ mm}$
	Tight Electron	Tight Muon
Isolation Criteria	PLImprovedTight	PLImprovedTight
Additional	e/ $\gamma$ ambiguity-cuts	

Table 5.3: Summary of the object definitions for electrons and muons in the analysis. Objects which are defined as preselected leptons met the upper set of conditions. Leptons meet the lower set of criteria as well as the conditions of a preselected lepton.

### 5.2.2 Jets

Jets are collimated collection of particles that are produced during high energy collisions [45]. A hard parton (a quark or gluon) produced in the interaction radiates gluons that emit additional gluons or quark-antiquark pairs. These additional coloured particles hadronise into hadrons which then decay into more stable particles that interact with the detector. This shower of soft particles is used to reconstruct a jet which can provide information regarding the momentum vector and the flavour of the hard parton.

Jet reconstruction is performed using the anti- $k_t$  clustering algorithm [46] with a radius parameter of  $R = 0.4$ . The  $R$  parameter dictates the radius around a hard higher energy particle in which soft lower energy particles will be considered as part of the jet. If no

other hard particles are within a radius of  $2R$ , the soft particles are included in the jet. If there is a hard particle within a radius of  $2R$ , the soft particles are divided between two different jets formed by the two hard particles. A jet is reconstructed using the four-momenta of the hard and soft particles. Jets must have  $p_T > 20\text{ GeV}$  and must lie in the central region with a pseudorapidity range of  $|\eta| < 2.5$ . In order to remove jets that originate from pileup collisions, a multivariate discriminant known as the Jet Vertex Tagger (JVT) [47] is used to select prompt jets. Jets that have  $p_T < 60\text{ GeV}$  and  $|\eta| < 2.4$  must also have a  $\text{JVT} > 0.2$ .

Some jets are created due to the hadronisation of heavy flavoured  $b$  and  $c$  quarks which produce  $b$  and  $c$  hadrons respectively. The labelling of the flavour of the original parton is known as flavour-tagging and is performed using the DL1r [48, 49] flavour-tagging algorithm. The DL1r algorithm combines the output quantities of low-level algorithms that use track and vertex information of the jet. The DL1r neural network outputs the probability for a jet to be a  $b$ -jet, a  $c$ -jet or a light jet. A light jet is a jet originating from a gluon or a quark which is not a  $b$  or a  $c$  quark. Only the identification of jets from  $b$  quarks is relevant for the search. Instead of directly using the DL1r  $b$ -tagging discriminant, efficiency operating points are used. Efficiency operating points provide a tagging efficiency by seeing if the score of a jet is above a certain proportion of  $b$ -jets from the  $t\bar{t}$  training sample. For example, the 77% efficiency DL1r operating point is a value in which 77% of  $b$ -jets in the sample have a score higher than the value. In this search, a jet is considered a  $b$ -tagged jet if the jet passes the 77% efficiency DL1r operating point.

The jet energies are calibrated such that the reconstructed jets match the expected energies of jets at particle level [50]. The jet energy scale (JES) is calculated by comparing the response of a dedicated sample of simulated dijet events at the detector level and at a particle level. The jet response of simulated jets is also compared to a dedicated observed data sample. The jet energy resolution (JER) is determined using in-situ techniques and is calculated separately for data and Monte Carlo samples. Efficiency scale factors are applied to each event to account for the mis-modelling of the JVT and flavour-tagging requirements between data and simulated samples. A separate scale factor exists for each flavour category in the DL1r probability output. Weights are applied to each event to correct for these efficiency differences.

### 5.2.3 Missing Transverse Momentum

Some particles produced in the  $pp$  collision may not be detected. Due to the conservation of momentum in the transverse plane, the net  $p_T$  of the undetected particles and the detected particles should be zero. Therefore, the missing transverse momentum  $\mathbf{E}_T^{\text{miss}}$  [51] is defined as the negative  $p_T$  of the detected particles and is used a proxy for the  $p_T$  of

the undetected particles. The undetected particles commonly include neutrinos which cannot be measured at the ATLAS detector. The  $\mathbf{E}_T^{\text{miss}}$  is calculated by,

$$\mathbf{E}_T^{\text{miss}} = -\mathbf{E}_T^{\text{soft}} + \sum_{\ell \in \text{leptons}} -\mathbf{p}_T^\ell + \sum_{j \in \text{jets}} -\mathbf{p}_T^j, \quad (5.2)$$

which is the negative vectorial sum of the  $p_T$  of the leptons and jets as well as the missing transverse momentum soft term  $\mathbf{E}_T^{\text{soft}}$ . The missing transverse momentum soft term  $\mathbf{E}_T^{\text{soft}}$  is calculated using momentum from tracks that are not associated with any reconstructed lepton or jet objects. The missing transverse momentum can be written as the magnitude  $E_T^{\text{miss}} = |\mathbf{E}_T^{\text{miss}}|$  and the polar angle in the transverse plane  $\phi^{\text{miss}}$ .

### 5.2.4 Overlap Removal

Once all leptons and jets have been reconstructed, some spatial overlaps may occur in the regions used to reconstruct these objects. Additional procedures are conducted to remove ambiguities when two objects are reconstructed at a close angular distance from each other. If an electron and a muon were constructed from the same track, only the muon is considered. Jets which are within a distance of  $\Delta R_{y,\phi} = \sqrt{(\Delta y)^2 + (\Delta \phi)^2} < 0.2$  of an electron are removed. If  $0.2 < \Delta R_{y,\phi} < 0.4$ , the electron is discarded instead. Muons within  $\Delta R_{y,\phi} < 0.4$  of a jet are removed unless the jet has more than two associated tracks. In this case, the jet is removed.

## 5.3 Systematic Uncertainties

The systematic uncertainties involved in the following analysis include both experimental and modelling uncertainties. This thesis includes an expanded number of systematic uncertainties compared to the previous work associated with the measurement of  $tWZ$  production [1, 2]. The experimental uncertainties associated with data-taking at the ATLAS detector are detailed in Section 5.3.1. In Section 5.3.2, the uncertainties associated with modelling the production of SM processes are discussed.

### 5.3.1 Experimental Uncertainties

The measurements of physical observables at colliders have a number of associated systematic uncertainties due to variations in data-taking conditions and errors introduced during object reconstruction. The following experimental uncertainties were included under recommendation from the ATLAS combined performance (CP) groups.

The expected luminosity of the combined ATLAS Run 2 dataset has an estimated uncertainty of 0.83% [52, 32]. The luminosity calibration was obtained with the use of the van

Der Meer method at the LHC with the primary measurements being performed by the LUCID-2 subdetector. The uncertainties associated with the reweighting of the pileup distributions of simulated events are treated as a systematic variable.

There are various uncertainties associated with the reconstruction of jets. The uncertainty associated with the measurement of the JES is included. The uncertainty of the JER is determined by comparing the difference in jet energy resolution between data and the Monte Carlo samples. The jet scale factors provide an uncertainty by varying the scale factors and accounting for how much the overall sample varies. The uncertainty of the jet flavour-tagging efficiency scale factor is included for each flavour type. An additional uncertainty is included for jets with a higher  $p_T$  than the  $p_T$  region in which the scale factors were originally determined. The variations in the choice of JVT working point efficiency are also applied by varying the JVT efficiency scale factor.

The uncertainties associated with the electron and muon scale factors are considered for each efficiency correction. The uncertainty in the scale and resolution calibration of the electron energy and muon momentum are also introduced. The uncertainty related to the scale and resolution of the calibration of the  $E_T^{\text{miss}}$  soft term is included as a systematic uncertainty.

### 5.3.2 Modelling Uncertainties

This section outlines the systematic uncertainties associated with the modelling of the signal and background processes relevant to the search for  $tWZ$  production. These simulated samples are discussed in Section 5.1. Additional calculations are required to model the variations within the matrix element calculations as well as modelling parton showering and hadronisation.

#### $tWZ$ signal process

The  $tWZ$  signal process has the uncertainty associated with the choice of diagram removal scheme, as described in Section 3.3.1. The uncertainty is calculated by comparing the distribution of the nominal  $tWZ$  sample which uses the DR1 diagram removal scheme and the distribution of the alternative  $tWZ$  DR2 sample. The uncertainty associated with the PDF calculations was determined according to the PDF4LHC recommendations for Run 2 [53]. The PDF sets comprise an ensemble of PDFs with different parametrisations. The  $tWZ$  prediction of the PDF for each event is assigned as the mean value of the predictions from the ensemble of NNPDF PDFs. The standard deviation of the predictions from the ensemble is used as the uncertainty associated with the PDF calculation. The uncertainties associated with the variations of  $\alpha_s$  in the PDF calculations and the choice of PDF set are not included.

The matrix element calculations in event generators use perturbative methods which are truncated at NLO in pQCD and are dependent on a renormalisation scale factor  $\mu_R$  and a factorisation scale factor  $\mu_F$ . To account for this dependence, an uncertainty is derived by varying  $\mu_R$  and  $\mu_F$  from the nominal value equal to 1 to a down variation of 0.5 and an up variation of 2. The  $\mu_R$  and  $\mu_F$  scale factors are varied individually as well as together where both scale factors are set to the down or up variations.

### **$t\bar{t}Z$ background process**

The uncertainty associated with the choice of event generator can be estimated by comparing the simulated samples produced by different generators. An uncertainty associated with the  $t\bar{t}Z$  process was introduced due to variations between the nominal sample and an alternative SHERPA sample.

The uncertainty associated with the scale variations of  $\mu_R$  and  $\mu_F$  were considered for the  $t\bar{t}Z$  sample. An uncertainty of 6% is applied on the normalisation cross section of  $t\bar{t}Z$  produced and this uncertainty is considered across both the trilepton and tetralepton channels. The uncertainty is motivated by the latest ATLAS  $t\bar{t}Z$  cross section measurement [54]. In the tetralepton channel, the contribution of  $t\bar{t}Z$  events with a fake lepton have a larger normalisation uncertainty of 50%.

### **Diboson background processes**

An uncertainty was introduced to account for the scale variations of  $\mu_R$  and  $\mu_F$  in the simulation of diboson processes  $WZ+\text{jets}$  and  $ZZ+\text{jets}$ . An uncertainty of 30% is applied to the normalisation of the  $WZ + l$  and  $WZ + c$  components while a larger uncertainty of 50% is applied to the normalisation of the  $WZ + b$  component. The cross section normalisation uncertainty for the  $ZZ + l$  sample was set to 10% and the uncertainty for the  $ZZ + c/b$  samples was set to 30%. In the tetralepton channel,  $WZ+\text{jets}$  events had a normalisation uncertainty of 50% as a fake lepton would be required in the event.

### **$tZq$ background process**

The scale variations of the  $tZq$  sample were considered in both channels. In the trilepton channel, a normalisation uncertainty of 14% was applied. The uncertainty is motivated by the latest ATLAS  $tZq$  cross section measurement [55]. In the tetralepton channel, a larger normalisation uncertainty of 50% is applied due to the requirement of a fake lepton.

## Other background processes

Additional samples were introduced for modelling other minor background processes in the trilepton and tetralepton channels. The contribution of the trilepton processes of  $t\bar{t}W$ ,  $t\bar{t}t$  and  $H + W$  carry an associated 30% normalisation uncertainty in the trilepton channel. The normalisation uncertainty for these processes was increased to 50% in the tetralepton channel, as an additional fake lepton is required.

The contribution of the tetralepton processes of  $t\bar{t}t\bar{t}$ ,  $t\bar{t}H$ ,  $H+Z$  and  $t\bar{t}WW$  have a normalisation uncertainty of 30% in both the trilepton and tetralepton channels. The triboson processes  $VVV$  were assigned an overall normalisation uncertainty of 30% in the trilepton and tetralepton channels. Some processes which produce a maximum of two prompt lepton ( $t\bar{t}$ ,  $Z+jets$ ,  $tW$  and  $Z+\gamma$ ) require one or more fake leptons to appear in the trilepton or tetralepton channel. Therefore, these processes have a normalisation uncertainty of 50% in both the trilepton and tetralepton channels.

## MC statistical uncertainty

The simulated samples discussed contain a finite number of events and provide only an estimation of the underlying sample distributions provided by the event generator. The larger the total number of MC events in a sample, the closer the sample will describe the distributions. Nuisance parameters known as  $\gamma$  parameters [56] are introduced to estimate the differences between true rate of production of the processes and the number of events found in the MC samples. The uncertainties on the  $\gamma$  parameters are Poisson and decrease as the number of MC events increases.

## 5.4 Definition of regions

Regions are used to separate events into orthogonal categories to isolate important processes. In this search, regions can be signal regions (SR), control regions (CR) or validation region (VR). These regions are defined by the number of leptons, jets and  $b$ -tagged jets in an event.

Signal regions are defined to maximise the proportion of signal events as opposed to background events. In this search, the signal regions have a larger proportion of  $tWZ$  events than the other regions. Therefore, the measurement of  $\mu_{tWZ}$  should be most sensitive to the signal regions.

Control regions are designed to have a high sensitivity to a background process. The control region should provide a more precise estimate of the rate of production of the background processes than if only the signal regions were considered. The estimated rate of production of the background process can then be extrapolated to estimate the

background contribution in the signal region. In this analysis, control regions were established for each major background process in the trilepton and tetralepton channels.

Validation regions are regions not considered during the fitting procedure and do not affect the measurement. Validation regions can be used to explore simulated events in a phase space separate from the signal or control regions to verify the modelling of events. Analysis techniques such as background estimation and discrimination models can be verified using a validation region before applying it to signal regions.

## 5.5 Signal and background discrimination in regions

The events in a given region from data and simulation are analysed as a set of histograms that display the distribution of these events according to an observable measured for each event. Depending on the choice of observable, the distribution of the signal events may differ from the distribution of the background events. The difference between the distributions of signal and background events causes some bins to have higher proportions of signal events. The signal enriched bins are more sensitive to changes in  $\mu_{tWZ}$  which provide a greater estimation of  $\mu_{tWZ}$  during the fitting procedure.

The discriminating variable within a region is usually chosen as an observable which maximally separates the signal and background distributions. The separation can be quantified with the separation  $\langle S^2 \rangle$  metric [57] defined by the integral

$$\langle S^2 \rangle = \frac{1}{2} \int \frac{(\hat{s}_x(x') - \hat{b}_x(x'))^2}{\hat{s}_x(x') + \hat{b}_x(x')} dx', \quad (5.3)$$

where  $x$  is an event observable and  $\hat{s}_x$  and  $\hat{b}_x$  are the probability density functions for the signal and background processes. It is assumed that  $\hat{s}_x$  and  $\hat{b}_x$  are defined on the same domain and the integration is performed over this domain. The metric allows for the comparison of the separation power of different observables. A larger  $\langle S^2 \rangle$  indicates that the signal process is better isolated and therefore, the fitting procedure can perform a better estimation of the signal strength. The observable distributions are approximated by histograms and 5.3 is discretised as a sum over each bin of the histogram.

The discriminating variable can be a single event observable or a higher level variable which is constructed by combining multiple event observables. The high-level variables are commonly a product of multivariate machine learning models such as boosted decision trees and neural networks. These models use multiple event observables as model inputs. In this analysis, graph neural networks (GNN) are employed to construct discriminating variables for use in the signal regions and some control regions. These models are trained using simulated samples but are applied to both data and simulated events.

The GNNs used are binary classification models for discriminating between signal and background events. A binary classification model outputs a prediction  $\hat{p} \in [0, 1]$  which can be interpreted as the probability for the input event to be a signal event.

## 5.6 Statistical methods

This section will outline the statistical tools and techniques involved in the measurement of  $\mu_{tWZ}$ . The likelihood model is defined and motivated in Section 5.6.1. The statistical tests for the discovery of a signal process and the setting of upper limits on  $\mu_{tWZ}$  are detailed in Section 5.6.2 and section 5.6.3. Asimov datasets are defined in Section 5.6.4 as well as how Asimov datasets are used to determine experimental sensitivity. The blinding of regions with the use of Asimov datasets is explained in Section 5.6.5. Much of the content in the following section is adapted from the lecture notes of Kyle Crammer [58].

### 5.6.1 Definition of likelihood function

The extraction of the signal strength of  $tWZ$  production  $\mu_{tWZ}$  is performed using the method of maximum likelihood [59]. The best-fit values of the model parameters for a given dataset can be estimated by maximising the likelihood function with respect to the model parameters. The model parameters include the parameter of interest  $\mu_{tWZ}$  as well as a set of parameters  $\boldsymbol{\alpha}$  which describe various systematic uncertainties. Any model parameter which is not a parameter of interest is referred to as a nuisance parameter. The likelihood function is constructed out of many terms which allows for the probability model to describe binned data while modelling the effects of systematic uncertainties.

#### Observable distributions

The manner in which events are distributed with respect to an observable  $x$  is determined by an underlying probability density function  $f(x)$ . The probability density function will also depend on the parameters of the model, such that the function is written as  $f(x | \mu, \boldsymbol{\alpha})$ . The probability of obtaining a set of measurements of  $x$  for a given set of parameters can be written as

$$\prod_{r \in \text{Regions}} \prod_e^{N_r} f_r(x_e | \mu, \boldsymbol{\alpha}), \quad (5.4)$$

where  $N_r$  is the number of events in a region  $r$ ,  $x_e$  is the measurement of  $x$  in event  $e$  and  $f_r$  is the choice of observable distribution in the region  $r$  assuming the model parameters  $\mu$  and  $\boldsymbol{\alpha}$ . In the fitting procedure, the observable distributions  $f_r$  are constructed using the Monte Carlo samples and are approximated as histograms.

## Extended likelihood

The statistical fluctuation of the number of events measured in each bin of a histogram is modelled by a Poisson distribution. Therefore, the term describing the statistical fluctuations for all bins across all regions is

$$\prod_{r \in \text{Regions}} \prod_i^{N_r} \frac{\nu_i^{n_i}}{n_i!} e^{-\nu_i}, \quad (5.5)$$

where  $n_i$  is the number of counts in a bin  $i$  and  $\nu_i$  is the expected number of counts in that bin. In the final likelihood function, the expected bin counts will vary with the parameters of the model and can be written as  $\nu_i(\mu, \boldsymbol{\alpha})$ . The observed counts  $n_i$  are dependent on the measured dataset and are fixed. The introduction of these Poisson terms expands the likelihood function into an extended likelihood function.

## Constraint terms

The final contributions to the probability model are the constraint terms for the nuisance parameters. Constraint terms are additional factors in the likelihood function that encode prior knowledge of some model parameters. The prior knowledge may be experimentally determined, such as the calibrations of JES and JER, or theoretically motivated, such as the choice of diagram removal scheme in the  $tWZ$  simulated sample. The effect of the constraint terms is to discourage model parameters that differ significantly from their pre-fit values that correspond to the prior knowledge.

The constraint terms for the nuisance parameters are introduced into the likelihood function as,

$$\prod_p f_p(a_p | \alpha_p), \quad (5.6)$$

where  $p$  is a systematic variable,  $a_p$  is the nuisance parameter associated with  $p$  and  $\alpha_p$  is the pre-fit value motivated by prior knowledge of  $p$ . In this analysis, the parameter of interest  $\mu_{tWZ}$  is left unconstrained and all nuisance parameters have a constraint term. The constraint terms are typically chosen to be Gaussian terms  $\mathcal{N}(a_p | \alpha_p, \sigma_p)$  where  $\sigma_p$  is the uncertainty associated prior knowledge associated with the value of  $\alpha_p$ . This choice of constraint assumes all nuisance parameters are independent.

## Complete likelihood function

The complete likelihood function can be stated by taking the product of the individual probability terms discussed above. The full likelihood model can be evaluated for an

observed dataset  $\mathcal{D} = \{x_{r,e}\}$  and a set of pre-fit values  $\boldsymbol{a} = \{a_p\}$  as

$$\mathcal{L}(\mathcal{D}, \boldsymbol{a} | \mu, \boldsymbol{\alpha}) = \left[ \prod_{r \in \text{Regions}} \left( \prod_i^{N_r} \frac{v_i^{n_i}}{n_i!} e^{-v_i} \right) \prod_e^{n_r} f_r(x_{r,e} | \mu, \boldsymbol{\alpha}) \right] \prod_p f_p(a_p | \alpha_p). \quad (5.7)$$

The signal strength  $\mu$  and the nuisance parameters  $\boldsymbol{\alpha}$  can be estimated by maximising the likelihood function with respect to these parameters.

The likelihood function was constructed for the data in the trilepton channel  $\mathcal{L}(\mathcal{D}_{3\ell})$  and the data in the tetralepton channel  $\mathcal{L}(\mathcal{D}_{4\ell})$ . A combined maximum likelihood estimation across both channels is performed by using a combined likelihood function  $\mathcal{L}(\mathcal{D}_{3\ell} \cup \mathcal{D}_{4\ell})$ . Model parameters in the combined likelihood function are the union of the parameters of the trilepton and tetralepton likelihood functions, with the common parameters being merged

### 5.6.2 Statistical tests for signal discovery

The discovery of a signal process requires a measurement that would allow for the rejection of the hypothesis where no events are contributed by the signal process. This null hypothesis is known as the background-only hypothesis and corresponds to the case where  $\mu_{tWZ} = 0$ . To formally define signal discovery, the test statistic  $q_0$  is first defined.

The test statistic  $q_0$  requires the profile likelihood ratio  $\lambda$  which is defined for a given dataset as

$$\lambda(\mu) = \frac{\mathcal{L}(\mu, \hat{\boldsymbol{\theta}}_\mu)}{\mathcal{L}(\hat{\mu}, \hat{\boldsymbol{\theta}})} \quad (5.8)$$

where  $\mu$  is the parameter of interest and  $\boldsymbol{\theta}$  are the nuisance parameters. Once the maximum likelihood estimation has been performed,  $\hat{\mu}$  and  $\hat{\boldsymbol{\theta}}$  are the estimators for  $\mu$  and  $\boldsymbol{\theta}$  and  $\hat{\boldsymbol{\theta}}_\mu$  is the estimator for  $\boldsymbol{\theta}$  with a fixed value for  $\mu$ . The values can range from  $0 \leq \lambda \leq 1$  with a value of  $\lambda = 1$  shows the best agreement between the data and a hypothetical value of  $\mu$ .

Discovery of a signal process requires strong disagreement between the  $\hat{\mu}$  observed in the dataset and the null hypothesis value of  $\mu = 0$  as defined by the following test. Using 5.8, a test statistic for the discovery of a signal process can be determined

$$q_0 = \begin{cases} -2 \ln \lambda(0) & \text{if } \hat{\mu} \geq 0 \\ 0 & \text{if } \hat{\mu} < 0 \end{cases}. \quad (5.9)$$

Since only positive signal contributions are of interest, only positive values of the estimator  $\hat{\mu}$  should be considered. The corresponding  $p$ -value for the observed  $q_0$  statistic

can be determined with

$$p_0 = \int_{q_{0,\text{obs}}}^{\infty} f(q_0|\mu = 0) dq_0, \quad (5.10)$$

which integrates over the distribution of  $q_0$  given the background-only hypothesis. In particle physics, the measured  $p$ -value is typically presented as the statistical significance of the measurement,

$$Z_0 = \Omega^{-1}(1 - p_0), \quad (5.11)$$

where  $\Omega^{-1}$  is the inverse cumulative distribution of a standard Gaussian distribution. For the discovery of a signal process, a benchmark significance of  $Z_0 = 5$  must be exceeded.

### 5.6.3 Statistical test for upper limits

The setting of upper limits on the parameter of interest is another useful statistical test. When performing a search for a signal process, upper limits allow for the exclusion of areas of parameter space. In a search for a SM process, the upper limit can be phrased as the theory with the largest  $\mu$  value which excludes the standard model hypothesis of  $\mu = 1$ .

The establishment of an upper limit on  $\mu$  requires the test statistic

$$q_\mu = \begin{cases} -2 \ln \lambda(\mu) & \text{if } \hat{\mu} < \mu \\ 0 & \text{if } \hat{\mu} > \mu \end{cases} \quad (5.12)$$

where the estimator for signal strength  $\hat{\mu}$  is required to be below the upper limit  $\mu$ . The corresponding  $p$ -value can be determined by

$$p_\mu = \int_{q_{\mu,\text{obs}}}^{\infty} f(q_\mu|\mu) dq_\mu. \quad (5.13)$$

Analyses at the ATLAS collaboration opt to use the  $\text{CL}_s$  method [60] for setting upper limits. The  $\text{CL}_s$  method establishes the upper limit using a ratio of  $p$ -values,

$$p'_\mu = \frac{p_\mu}{1 - p_b} \quad (5.14)$$

where  $p_b$  is the  $p$ -value of the test statistic  $q_{\mu,\text{obs}}$  under the background-only hypothesis. A signal model is excluded with a confidence interval of 95% if

$$p'_\mu < \alpha \quad (5.15)$$

where  $\alpha = 0.05$ . The upper limit on the signal strength  $\mu_{\text{up}}$  is defined as the largest value

of  $\mu$  which is excluded.

#### 5.6.4 Asimov data and experimental sensitivity

An Asimov dataset [61] is a constructed dataset in which the fitted values of the model parameters are the same as the expected values of the new hypothesis being tested. For this analysis, the Asimov dataset is constructed from simulated data with the signal strength  $\mu = 1$  and all nuisance parameters  $\alpha = 0$ . The Asimov dataset can be used to estimate experimental sensitivity.

The experimental sensitivity suggests the ability of the analysis to measure the new hypothesis if it is assumed that the data follows the new hypothesis. Particularly for signal extraction, can the background-only hypothesis be rejected if the observed data matches the signal-plus-background hypothesis? Experimental sensitivity is evaluated by determining the expected significance and expected median upper limit.

The expected (or median) significance  $Z^{\text{exp}}$  for an assumed signal strength of  $\mu' = 1$  is calculated with

$$Z_0^{\text{exp}} = \text{med} [Z_0 | \mu' = 1], \quad (5.16)$$

which is the median of the distribution of  $Z_0$  as defined in 5.11. This can be approximated as  $Z_0^{\text{exp}} = \sqrt{q_{0,A}}$  where the test statistic  $q_{0,A}$  is similar to 5.9 but evaluated using the Asimov dataset instead of the observed dataset.

The expected (or median) upper limit  $\mu_{\text{up}}^{\text{exp}}$  at a confidence level of 95% can be approximated using the Asimov dataset with

$$\mu_{\text{up}}^{\text{exp}} = \text{med} [\mu_{\text{up}} | \mu' = 1] \approx \mu' + \sigma_A \Phi^{-1}(0.95), \quad (5.17)$$

where  $\sigma_A$  is the uncertainty of the  $\hat{\mu}$  estimator when fitted to the Asimov dataset.

#### 5.6.5 Blinding of regions

Using 5.16 and 5.17, the experimental sensitivity of the analysis can be determined using only the Asimov dataset. The analysis can then be optimised to maximise experimental sensitivity without using the ATLAS data. This separates the development of the analysis strategy from the final signal extraction, which reduces the influence of bias on the construction of the analysis. Observed data can be used to evaluate the modelling of background processes but is otherwise blinded during the signal extraction.

The blinding of data can be performed either in figures or in the fitting procedure. If a region is blinded in figures, only the Asimov dataset is shown in that region and not the observed data. In this analysis, all signal regions are blinded in the figures. Individual

bins with  $S/B \geq 10\%$  in the control regions are also blinded.

The results presented in this thesis include only blinded results which are obtained by performing the fitting procedure on an Asimov dataset. The results still allow for the evaluation of the experimental sensitivity of the analysis.

# Chapter 6

## Analysis of the $tWZ$ trilepton channel

This chapter will discuss the analysis strategy and the resulting measurements specific to the  $tWZ$  trilepton decay channel. The final state products of  $tWZ$  production can decay to an experimental signature involving three leptons. In the trilepton decay channel, two leptons are produced by the  $Z$  boson decaying into a lepton-antilepton pair. The third lepton originates from a  $W$  boson which decays into leptons via  $W \rightarrow \ell \nu_\ell$ . The leptonically decaying  $W$  boson could be the prompt  $W$  boson from the primary interaction or a  $W$  boson from the top quark decay.

The background process which contributes the largest number of events in the trilepton channel is the production of a  $t\bar{t}$  pair in association with a  $Z$  boson ( $t\bar{t}Z$ ). Like the  $tWZ$  process, the  $t\bar{t}Z$  process can decay via three leptons where exactly one of the top quark decays produces a leptonically decaying  $W$  boson. The production of a  $W$  boson and  $Z$  boson with additional jets ( $WZ+\text{jets}$ ) is the background process that contributes the next largest number of events. Both the  $W$  boson and the  $Z$  boson in the  $WZ+\text{jets}$  process must decay leptonically to produce three prompt leptons.

In this chapter, many figures will include the expected rates of the various simulated samples detailed in Section 5.1. The  $tWZ$  sample and the samples associated with the major background processes in the trilepton channel will be individually labelled, while some of the non-major background samples will be grouped. The label *Other* represents the group of non-major background processes which produce three or more prompt leptons. The label *Fakes* represents the group of background processes which produce less than three prompt leptons, not including the  $Z+\text{jets}$  sample.

The definitions of the signal and control regions used in the trilepton channel are discussed in Section 6.1. An observable is defined in Section 6.2 to discriminate between  $tWZ$  events and background  $WZ+\text{jets}$  events by identifying the presence of a hadronically-decaying  $W$  boson. In Section 6.3, a GNN is employed to perform signal-background

discrimination. Extraction of the signal strength of  $tWZ$  in the trilepton channel using an Asimov dataset is performed in Section 6.4.

## 6.1 Baseline selections and region definitions

All lepton, jet and  $b$ -tagged jet objects must meet their respective object definitions as stated in Section 5.2 as well as some region-specific criteria which will be introduced in this section. Events in both data and simulation were only selected if the single lepton or dilepton (including di-electron, di-muon and electron-muon) triggers were applied [62, 63]. These triggers have a minimum  $p_T$  threshold between 12 GeV and 26 GeV which depends on the lepton flavour and the data-taking period.

Events in the trilepton channel are required to contain exactly 3 tight lepton objects where the highest energy lepton has  $p_T > 30$  GeV, the next highest energy lepton has  $p_T > 20$  GeV and the third highest energy lepton has  $p_T > 14$  GeV. To reduce the contribution from leptons produced by decaying neutral mesons, a restriction on the reconstructed mass of oppositely signed same flavour (OSSF) lepton pairs is introduced. All OSSF pairs are required to have  $m_{\text{OSSF}} > 10$  GeV since  $J/\psi$  and  $\Upsilon$  mesons have a mass lower than this threshold. The mass of the  $Z$  boson is above the threshold so OSSF pairs originating from decaying  $Z$  bosons will not be affected. All regions in the trilepton channel require events to have a single  $Z$  boson. Therefore, the events require that exactly one OSSF lepton pair has a reconstructed mass within 10 GeV of mass of the  $Z$  boson  $m_Z = 91.19$  GeV [7]. The reconstructed OSSF lepton pair will be referred to as the  $Z$  candidate of the event. In the trilepton region, a selection on the  $p_T$  of all jets was imposed where  $p_T > 25$  GeV. This is a stricter requirement than was used in the tetralepton channel.

Jets are also required to have a  $\text{JVT} > 0.5$  to suppress contributions from pileup jets.

The regions in the trilepton channel are defined based on the number of physics objects that are present in an event. The trilepton channel has one signal region and two control regions where the regions differ by the required number of jets and  $b$ -tagged jets. The signal region  $tWZ$  SR was designed to isolate the  $tWZ$  signal sample and provide the best sensitivity to  $\mu_{tWZ}$ . In the  $tWZ$  trilepton final state, two jets are produced from the hadronically-decaying  $W$  boson and a single  $b$ -jet is produced by the decaying top quark. Therefore, events in the  $tWZ$  SR are required to have more than or equal to 3 jets, with exactly 1 of these jets being a  $b$ -tagged jet.

The  $t\bar{t}Z$  control region ( $t\bar{t}Z$  CR) is intended to estimate the background contributions from  $t\bar{t}Z$  production. The production of  $t\bar{t}Z$  via the trilepton decay scheme produces 2 quarks from the single hadronically-decaying  $W$  boson as well as 2  $b$  quarks being

produced due to the decay of the 2 top quarks. To capture the contribution from  $t\bar{t}Z$ , the events in the  $t\bar{t}Z$  CR are required to contain more than or equal to 4 jets where 2 or more of those jets are  $b$ -tagged jets.

The  $WZ$  control region ( $WZ$  CR) captures the contributions from  $WZ$ +jets production to estimate the contributions due to additional heavy  $b$ - and  $c$ -jets. The  $WZ$  CR requires each event to contain 1 or 2 jets, where at least one jet must be  $b$ -tagged.

An additional  $E_T^{\text{miss}} > 20 \text{ GeV}$  requirement is placed on events in the  $tWZ$  SR and the  $WZ$  CR to reduce contributions from events with non-prompt/fake leptons. The choice of  $E_T^{\text{miss}}$  value is motivated and explored in Section 6.1.1. The baseline selections placed on the physics objects in the trilepton channel as well as the definition of the regions are summarised in Table 6.1.

Baseline selections		
$N_\ell = 3$		
$p_T(\ell_1, \ell_2, \ell_3) > (30, 20, 14) \text{ GeV}$		
$p_T(\text{jet}) > 25 \text{ GeV},  \eta(\text{jet})  < 2.5, \text{JVT} > 0.5$		
$ \eta(\ell_e)  < 2.47$ excluding $1.37 <  \eta(\ell_e)  < 1.52$		
$ \eta(\ell_\mu)  < 2.5$		
All OSSF lepton pairs require $m_{\text{OSSF}} > 10 \text{ GeV}$		
1 Z Candidate		
Regions		
$tWZ$ SR	$t\bar{t}Z$ CR	$WZ$ CR
$\geq 3$ jets	$\geq 4$ jets	1 or 2 jets
1 $b$ -tagged jet	$\geq 2$ $b$ -tagged jet	1 $b$ -tagged jet
$E_T^{\text{miss}} > 20 \text{ GeV}$		$E_T^{\text{miss}} > 20 \text{ GeV}$

Table 6.1: Summary of the trilepton channel specific selections for physics object definitions and the requirements for the signal and control regions.

The number of expected SM events in the regions of the trilepton channel are shown in Figure 6.1 as well as the observed data rates in the control regions. Each of the control regions is dominated by their respective process of interest where  $t\bar{t}Z$  production contributes the majority of events in the  $t\bar{t}Z$  CR and  $WZ$ +jets production contributes the majority of events in the  $WZ$  CR. Notable contributions from  $tZq$  production are seen in the  $tWZ$  SR as well as in the  $WZ$  CR. The overall number of observed events in the control regions appears to agree with the expectation from simulation. The uncertainty associated with simulated event counts in the  $WZ$  CR is large with a relative uncertainty of 17% which is primarily due to the conservative pre-fit cross section normalisation uncertainties associated with  $WZ$ +jets events. The expected number of events from each simulated sample is shown in Table 6.2. The individual counts show  $t\bar{t}W$  provides a relatively low number of events in the signal region which implies that many of the  $t\bar{t}W$

events are excluded due to the  $Z$  candidate requirement. The contributions from processes with less than 3 prompt leptons are primarily from dilepton  $t\bar{t}$  events and  $Z + \text{jets}$  events.

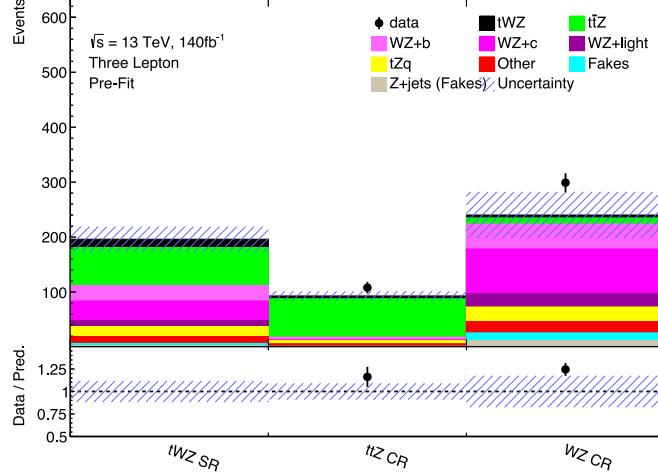


Figure 6.1: The total number of expected SM events for the signal process and background processes in the  $tWZ$  SR region,  $t\bar{t}Z$  CR region and  $WZ$  CR region. A different colour represents each simulated sample. The overall systematic uncertainty associated with the SM processes is shown as blue diagonal lines. The observed number of data events is included in the control regions as black points. Data events are not shown in the  $tWZ$  SR to keep the region blinded. The bottom panel shows the ratio between the number of data events and the predicted number of SM events.

Sample	$tWZ$ SR	$t\bar{t}Z$ CR	$WZ$ CR
$tWZ$	15.2 $\pm$ 2.0	5.6 $\pm$ 1.9	5.9 $\pm$ 0.9
$t\bar{t}Z$	69 $\pm$ 7	70 $\pm$ 7	11.1 $\pm$ 2.0
$WZ + b$	28 $\pm$ 14	4.8 $\pm$ 2.4	45 $\pm$ 23
$WZ + c$	35 $\pm$ 11	0.83 $\pm$ 0.32	81 $\pm$ 26
$WZ + l$	11 $\pm$ 4	0.045 $\pm$ 0.031	25 $\pm$ 8
$tZq$	18.5 $\pm$ 2.9	6.5 $\pm$ 1.1	26 $\pm$ 5
$ZZ+jets$	6.2 $\pm$ 1.3	0.69 $\pm$ 0.21	14.6 $\pm$ 3.3
$t\bar{t}W$	1.9 $\pm$ 0.6	1.1 $\pm$ 0.4	3.1 $\pm$ 1.0
$t\bar{t}t$	$0.0039 \pm 0.0027$	$0.039 \pm 0.014$	$0.0004 \pm 0.0012$
$VVV$	0.45 $\pm$ 0.14	0.040 $\pm$ 0.019	0.66 $\pm$ 0.22
$WH$			0.005 $\pm$ 0.015
$Z + jets$	1.8 $\pm$ 1.2	0.24 $\pm$ 0.23	11 $\pm$ 6
$t\bar{t}\gamma$	0.6 $\pm$ 0.4	0.28 $\pm$ 0.19	1.4 $\pm$ 0.8
$Z\gamma$	0.6 $\pm$ 0.4	0.11 $\pm$ 0.07	2.1 $\pm$ 1.2
$t\bar{t}$	3.7 $\pm$ 2.0	0.7 $\pm$ 0.4	11 $\pm$ 6
$tW$	0.4 $\pm$ 0.4		1.5 $\pm$ 0.5
$t\bar{t}t\bar{t}$	$0.033 \pm 0.012$	0.33 $\pm$ 0.10	$0.0007 \pm 0.0006$
$t\bar{t}WW$	0.16 $\pm$ 0.07	0.17 $\pm$ 0.08	
$t\bar{t}H$	2.0 $\pm$ 0.6	1.9 $\pm$ 0.6	0.47 $\pm$ 0.16
$ZH$	1.5 $\pm$ 1.0		2.3 $\pm$ 1.3
Total	197 $\pm$ 23	93 $\pm$ 8	240 $\pm$ 40

Table 6.2: Expected number of events according to simulation in each of the trilepton regions for each process or group of processes. Empty cells indicate no simulated events are present in the region. The uncertainty associated with each yield is shown and comprises both statistical and systematic uncertainties.

### 6.1.1 Optimisation of requirement on $E_T^{\text{miss}}$

This section aims to motivate the additional  $E_T^{\text{miss}} > 20 \text{ GeV}$  requirement on events in the  $tWZ$  SR and the  $WZ$  CR. The  $WZ$  CR has a larger contribution from events with non-prompt/fake leptons than the other regions due to the lower jet multiplicity requirements. In this section, a control region (3L-1or2j-1b CR) which requires 3 tight leptons and one or two jets where one jet is a  $b$ -tagged jet is considered. The distribution of  $E_T^{\text{miss}}$  in the 3L-1or2j-1b CR is shown in Figure 6.2. In Figure 6.2, the majority of the non-prompt/fake contributions have  $E_T^{\text{miss}}$  below 30 GeV. In the four bins lowest in  $E_T^{\text{miss}}$ , a low proportion of events are contributed by  $WZ+jets$  processes. The large contributions from events with fake leptons and the low proportion of  $WZ+jets$  events motivate placing a lower limit on the  $E_T^{\text{miss}}$  of events. There is also poor agreement between observed data and the simulated events in the region of low  $E_T^{\text{miss}}$ . The poor agreement could be attributed to poor modelling of processes with low transverse momenta. Poor modelling in the  $WZ$  CR suggests a similar poor modelling may be present in the  $tWZ$  SR.

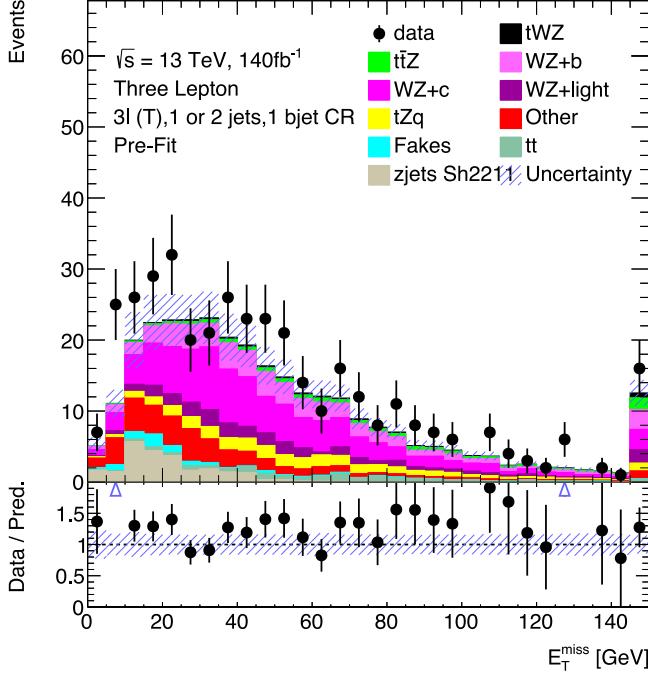


Figure 6.2: Distribution of  $E_T^{\text{miss}}$  for the various simulated samples in the control region which requires one or two jets, where one jet must be a  $b$ -tagged jet. A different colour represents each simulated sample. The overall systematic uncertainty associated with the SM processes is shown as blue diagonal lines. The observed number of data events is included in the control regions as black points. The bottom panel shows the ratio between the number of data events and the predicted number of SM events.

A decision was made to apply a requirement of  $E_T^{\text{miss}} > 20 \text{ GeV}$  on events in the  $WZ$  CR. Since the  $tWZ$  SR and the  $WZ$  CR differ by only a single jet, the same requirement of  $E_T^{\text{miss}} > 20 \text{ GeV}$  was applied to events in the  $tWZ$  SR. The additional  $E_T^{\text{miss}}$  requirement reduces the total number of events in the  $tWZ$  SR and the  $WZ$  CR. The  $E_T^{\text{miss}}$  requirement improves the agreement between data and the simulated prediction in the  $WZ$  CR and removes events with non-prompt/fake contributions that are difficult to estimate. The effect of the introduction of the  $E_T^{\text{miss}}$  requirement on the measurement of  $\mu_{tWZ}$  is discussed in Section 6.4.

### 6.1.2 Modelling of photon conversions

Some leptons which are selected may be the product of photon conversions. A photon can produce an electron-antielectron pair when interacting with the detector. The introduction of additional leptons complicates the reconstruction of the  $Z$  boson in an event where one of the leptons from the photon may be selected. Photon conversions can be investigated by considering when one of the OSSF leptons produced by the  $Z$  boson emits a photon. If the emitted photon produces a lepton which is selected and the OSSF leptons from the  $Z$  boson are selected, the mass of the three selected leptons will

be approximately the mass of the  $Z$  boson.

To investigate the modelling of photon conversions, a region was defined with the same baseline selections as described in Section 6.1 except no  $Z$  boson candidates are required. If one of the OSSF leptons emits a photon, the mass of the OSSF pair will have a lower energy than is required to reconstruct a  $Z$  boson. Therefore, the  $Z$  boson may not be successfully reconstructed as the mass of the OSSF pair will be too far from the mass of the  $Z$  boson. In Figure 6.3, the distribution of the trilepton mass in the region specified above is shown. Figure 6.3 shows a peak around the mass of the  $Z$  boson. The peak is the result of events which produce an additional third lepton from photon conversions. The simulated Monte Carlo samples appear to agree with the observed data around the  $Z$  boson mass and therefore, photon conversions are modelled fairly well.

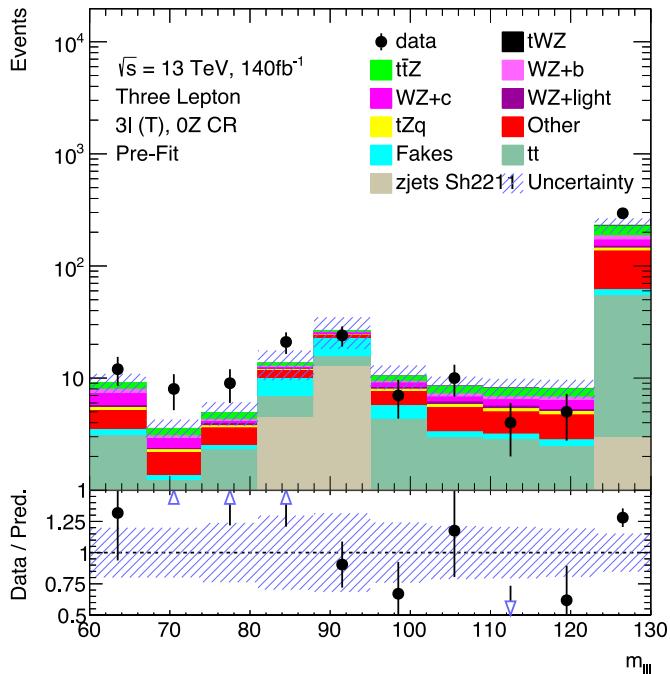


Figure 6.3: Distribution of reconstructed mass of the three leptons in the validation region which requires three tight leptons and zero  $Z$  boson candidates. A logarithmic scale is applied on the vertical axis. A different colour represents each simulated sample. The overall systematic uncertainty associated with the SM processes is shown as blue diagonal lines. The observed number of data events is included in the control regions as black points. The bottom panel shows the ratio between the number of data events and the predicted number of SM events.

## 6.2 Identification of hadronically-decaying $W$ bosons

A distinguishing feature between  $tWZ$  production and  $WZ+jets$  production is the presence of a hadronically-decaying  $W$  boson in  $tWZ$  production. The trileptonic decay

channel of  $WZ$  does not contain a hadronically-decaying  $W$  boson which provides an area of discrimination between the signal and background process.

A hadronically-decaying  $W$  boson will appear as a dijet system in the detector. An algorithm was introduced to identify dijets systems that originated from hadronically-decaying  $W$  bosons. This section includes introducing the production of a dataset of dijets, the specifications of a new discriminating variable and the performance of the variable in separating  $tWZ$  and  $WZ$  production.

### 6.2.1 Dataset of hadronically-decaying $W$ bosons

A dataset of dijets was required to evaluate the discriminating power of a new variable. The dataset included dijets that were produced from hadronically-decaying  $W$  bosons and dijets that were not. The dataset was created using a separate Monte Carlo sample of  $t\bar{t}$  pair events which was simulated at NLO with PowHEG using the NNPDF3.0NLO PDF set and the A14 MC tune. Parton showering and hadronisation of the events was performed using PYTHIA8.230. Only events with a single lepton were considered. The true kinematics of reconstructed  $W$  bosons can be identified by matching them to  $W$  bosons at the truth-level.

The dataset is split into dijets which originate from hadronically-decaying  $W$  bosons ( $W_{had}$  dijets) and dijets which did not (arbitrary dijets). The labelling of dijet systems as  $W_{had}$  dijets or arbitrary dijets is determined by considering the angular separation  $\Delta R_{had}$  between the true  $W$  boson and the reconstructed dijet,

$$\Delta R_{had} = \sqrt{(\phi_W - \phi_{dijet})^2 + (\eta_W - \eta_{dijet})^2} \quad (6.1)$$

where  $\phi_W$  and  $\eta_W$  are the angular coordinates of the true  $W$  boson and  $\phi_{dijet}$  and  $\eta_{dijet}$  are the angular coordinates of the reconstructed dijet. A dijet that is labelled as a  $W_{had}$  dijet must have the smallest  $\Delta R_{had}$  out of all dijets in an event and must be below a threshold of  $\Delta R_{had} < 0.05$ . All other dijets in the event are considered as background arbitrary dijets. If no  $W_{had}$  dijet is found in an event, all dijets in the event are ignored and not added to the dataset. The resulting dataset produced by the specified  $t\bar{t}$  sample included 23 421  $W_{had}$  dijet systems and 215 808 arbitrary dijet systems.

### 6.2.2 Features of dijet systems

Multiple observables associated with a dijet can be used to discriminate between dijets originating from hadronically-decaying  $W$  bosons and those that do not. The following observables will be combined to form the  $D_{jj}$  metric in the next section.

The angular separation between the jets in the dijet is defined as

$$\Delta R_{jj} = \sqrt{\Delta\phi(j_1, j_2)^2 + \Delta\eta(j_1, j_2)^2} \quad (6.2)$$

where  $j_1$  and  $j_2$  are the jets in the dijet. The  $W_{had}$  dijets will have a lower angular separation than arbitrary dijets as the large mass of the  $W$  bosons will result in the dijets being more collimated. In Figure 6.4a, the distribution of angular separation  $\Delta R_{jj}$  for the  $W_{had}$  and arbitrary dijets is shown. The  $W_{had}$  dijets are shown to have a lower angular separation than the arbitrary dijets. The angular separation between jets has a lower limit of  $\Delta R_{jj} > 0.4$  since jets are reconstructed with the anti- $k_t$  clustering algorithm with  $R = 0.4$ .

The mass of the dijet  $m_{jj}$  is calculated by combining the two jet 4-vectors in the dijet. If a dijet is produced by the decay of an on-shell  $W$  boson, the mass of the combined system will be similar to that of a  $W$  boson. In Figure 6.4b, the distribution of the dijet mass  $m_{jj}$  for the  $W_{had}$  and arbitrary dijets shows a resonance in the mass distribution of  $W_{had}$  dijets which is centred around the mass of the  $W$  boson. The other arbitrary dijets mass distribution peaks at a higher mass and has a tail continuing to a higher mass range.

A hadronically-decaying  $W$  boson is unlikely to produce a  $b$  quark in the SM. Therefore, any jet from a  $W$  boson decay which is a  $b$ -tagged jet has been misclassified. A dijet is said to have a  $b$ -tagged jet if at least one of the jets passes the 77% efficiency DL1r  $b$ -tagging operating point. In Figure 6.4c, the proportion of  $W_{had}$  and arbitrary dijets which have a  $b$ -tagged jet is shown. The majority of  $W_{had}$  dijets do not contain a  $b$ -tagged jet while the majority of arbitrary dijets have at least one  $b$ -tagged jet.

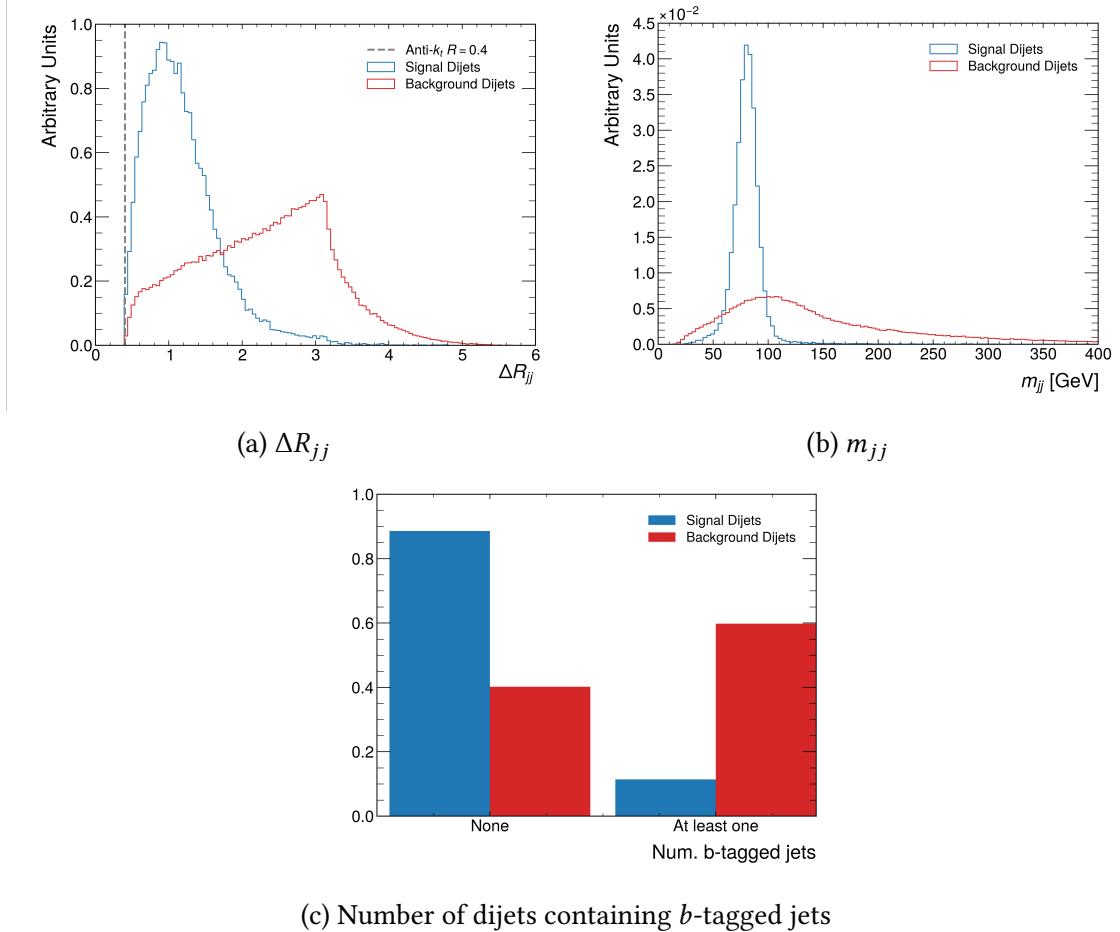


Figure 6.4: The normalised distributions of various dijet observables for the  $W_{had}$  and background dijets. The  $W_{had}$  or signal dijets are shown in blue and the arbitrary or background dijets are shown in red. The distributions include (a) the angular separation between the two jets  $\Delta R_{jj}$ , (b) the mass of the dijet  $m_{jj}$  and (c) the number of dijets containing  $b$ -tagged jets.

### 6.2.3 Dijet selection using $D_{jj}$ metric

The two-dimensional distributions of  $\Delta R_{jj}$  and  $m_{jj}$  are shown in Figure 6.5 for the the  $W_{had}$  dijet and arbitrary dijet datasets. There are differences between the  $W_{had}$  dijets in blue and arbitrary dijets in red. The  $W_{had}$  dijet distribution has a single peak and a low correlation between  $\Delta R_{jj}$  and  $m_{jj}$ . The distribution of arbitrary dijets has a larger spread and a positive correlation is present between  $\Delta R_{jj}$  and  $m_{jj}$ .

The metric  $D_{jj}$  was used to select the dijet in events that originated from a hadronically-decaying  $W$  boson. The  $D_{jj}$  metric for a dijet is defined as

$$D_{jj}(\Delta R'_{jj}, m'_{jj}) = \sqrt{\frac{(\Delta R'_{jj} - \mu(\Delta R_{jj}))^2}{\sigma^2(\Delta R_{jj})} + \frac{(m'_{jj} - \mu(m_{jj}))^2}{\sigma^2(m_{jj})}}. \quad (6.3)$$

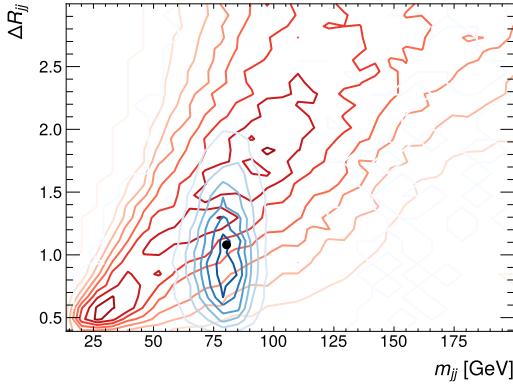


Figure 6.5: The normalised distributions of  $\Delta R_{jj}$  and  $m_{jj}$  of the  $W_{had}$  dijets and arbitrary dijet samples. The  $W_{had}$  dijets are represented in blue and the arbitrary dijets are represented in red. The darker contours represent a larger number of dijets. A black dot represents the central values of the signal  $W_{had}$  distribution with  $\mu(\Delta R_{jj}) = 1.1$  and  $\mu(m_{jj}) = 80.4\text{GeV}$ .

where  $\Delta R'$  and  $m'_{jj}$  are the observables of the dijets. The metric  $D_{jj}$  has the parameters  $\mu(\Delta R_{jj})$ ,  $\sigma(\Delta R_{jj})$ ,  $\mu(m_{jj})$  and  $\sigma(m_{jj})$ . The parameters  $\mu(\Delta R_{jj})$  and  $\sigma(\Delta R_{jj})$  are the central value and standard deviation of the distribution of  $\Delta R_{jj}$  for the signal  $W_{had}$  dijets. Similarly, the parameters  $\mu(m_{jj})$  and  $\sigma(m_{jj})$  are the central value and standard deviation of the distribution of  $m_{jj}$  for the  $W_{had}$  dijets.

The  $D_{jj}$  metric represents the distance between the measured observables of a dijet ( $\Delta R', m'_{jj}$ ) and the central values of the  $W_{had}$  distributions of those observables. The distances with respect to  $\Delta R$  and  $m_{jj}$  are standardised using  $\sigma(\Delta R_{jj})$  and  $\sigma(m_{jj})$ . The parameters of the metric were calculated to be  $\mu(\Delta R_{jj}) = 1.08$ ,  $\sigma(\Delta R_{jj}) = 0.47$ ,  $\mu(m_{jj}) = 80.4\text{ GeV}$  and  $\sigma(m_{jj}) = 12.8\text{ GeV}$ . These parameters are kept fixed, which allows the  $D$  metric to be evaluated for any dijet system.

The distribution of the  $D_{jj}$  metric for the  $W_{had}$  dijets and the arbitrary dijets are shown in Figure 6.6. In Figure 6.6, the distribution shapes are different between the  $W_{had}$  dijets and the arbitrary dijets. Lower values of  $D_{jj}$  have a higher signal-to-background ratio than larger values of  $D_{jj}$ .

The dijet in an event that was selected as the  $W$  boson candidate was the dijet with no  $b$ -tagged jets that has the lowest  $D_{jj}$  value. The selected  $W$  boson candidates were used to discriminate between samples with and without a hadronically-decaying  $W$  boson. The distributions of the  $D_{jj}$  metric value of the  $W$  boson candidate in the  $tWZ$  SR and the  $t\bar{t}Z$  CR are shown in Figure 6.7. The proportion of background events without a hadronically-decaying  $W$  boson ( $WZ+\text{jets}$  and  $tZq$ ) decreases as the  $D_{jj}$  decreases. The relative signal contribution is largest at low  $D_{jj}$  values which indicates that the  $D_{jj}$  metric is a successful discriminator between the  $tWZ$  and  $WZ+\text{jets}$  processes. In the  $t\bar{t}Z$

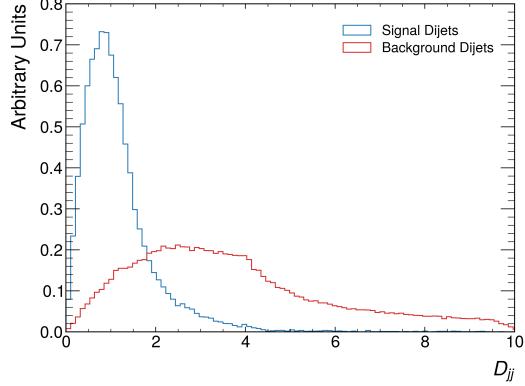


Figure 6.6: The normalised  $D_{jj}$  distributions of the  $W_{had}$  dijets and the arbitrary dijets. The  $W_{had}$  dijets are represented in blue and the arbitrary dijets are represented in red.

CR, the distribution of  $D_{jj}$  shows reasonable modelling of data by the simulated events.

The mass of the  $W$  boson candidate is a more easily interpretable observable than the  $D_{jj}$  value since dijets from hadronically-decaying  $W$  bosons are expected to peak around the mass of the  $W$  boson of 80 GeV. The distributions of the mass of the  $W$  boson candidates in the  $tWZ$  SR and  $t\bar{t}Z$  CR are shown in Figure 6.8. The contributions from processes without a hadronically-decaying  $W$  boson ( $WZ+jets$  and  $tZq$ ) are reduced around the  $W$  boson mass. Therefore, a higher purity of  $tWZ$  and  $t\bar{t}Z$  events is available around this region. The relative signal contribution in the  $tWZ$  SR is peaked around the  $W$  boson mass as expected since  $tWZ$  events contain a hadronically-decaying  $W$  boson. The distribution in the  $t\bar{t}Z$  CR shows adequate agreement between data and simulated events.

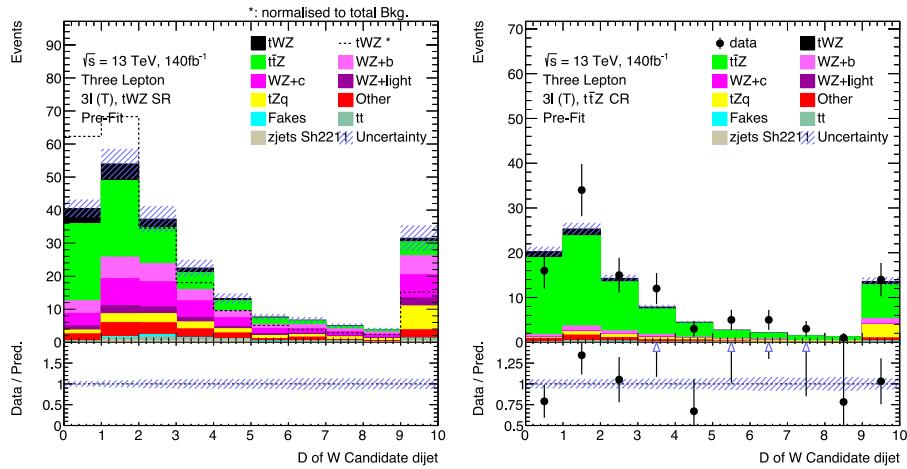


Figure 6.7: The distribution of the  $D$  value of the  $W$  boson candidate for the events in the trilepton  $tWZ$  SR (left) and  $t\bar{t}Z$  CR (right). A dotted line in the  $tWZ$  SR shows the relative contribution of  $tWZ$  when normalised with respect to the total background events.

The separation between the  $tWZ$  events and all other background events in the  $tWZ$  SR

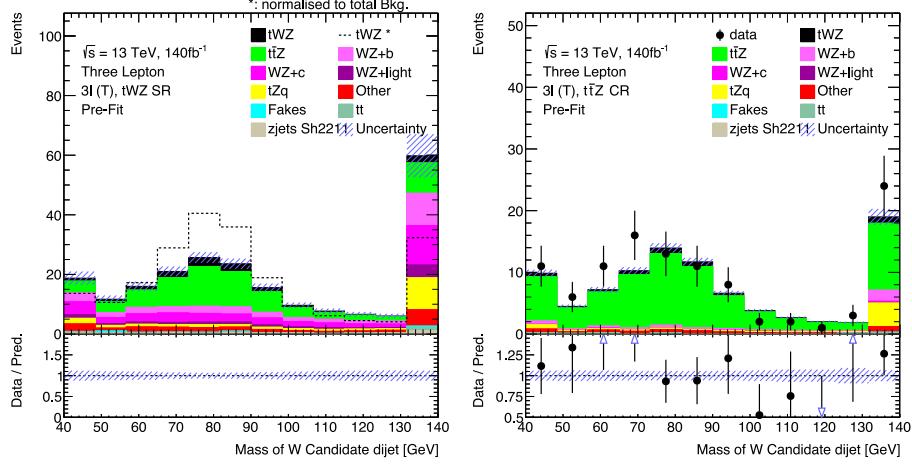


Figure 6.8: The distribution of the mass of the  $W$  boson candidate for the events in the trilepton  $tWZ$  SR (left) and  $t\bar{t}Z$  CR (right). A dotted line in the  $tWZ$  SR shows the relative contribution of  $tWZ$  when normalised with respect to the total background events.

is shown in Figure 6.9. Both the  $D_{jj}$  metric value of the  $W$  candidate and the mass of the  $W$  candidate separate the relevant backgrounds and provide an observable that can be used in a multivariate signal-background discrimination model. The distribution of the mass of the  $W$  candidate for the  $tWZ$  signal sample is peaked around the  $W$  boson mass. In Figure 6.9, the separation with respect to the mass of the  $W$  candidate ( $\langle S^2 \rangle = 4.58\%$ ) is larger than the separation with respect to the  $D_{jj}$  metric value of the  $W$  candidate ( $\langle S^2 \rangle = 4.03\%$ ). Since the mass of the  $W$  candidate shows a larger separation  $\langle S^2 \rangle$  and has a more natural physical interpretation, this observable was used as a global feature in a graph neural network, which will be discussed in Section 6.3.

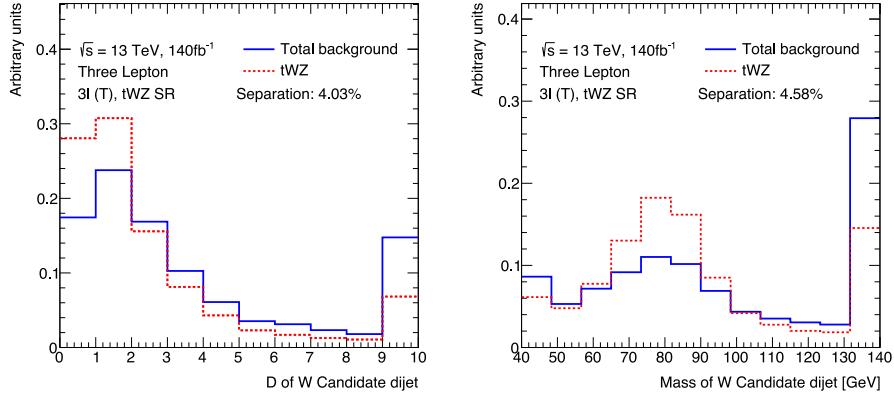


Figure 6.9: Distributions of the  $D$  metric and mass of the  $W$  boson candidates for the  $tWZ$  sample (dotted red line) and all other background samples (blue line) in the trilepton  $tWZ$  SR. The separation  $\langle S^2 \rangle$  between the signal sample and the background samples is included.

## 6.3 Graph neural network in the trilepton channel

To discriminate between signal and background events, a GNN was developed to perform binary classification in the trilepton channel. The objective of the model was to produce a discriminating variable that can separate the  $tWZ$  signal sample from the major background contributions. The discriminating score produced by the GNN in the trilepton channel will be referred to as  $\alpha_{3\ell}$  where  $\alpha_{3\ell}$  can be interpreted as a probability between 0 and 1. If an event produces a score close to  $\alpha_{3\ell} \approx 1$ , the model has estimated that the event is a signal event. An output of  $\alpha_{3\ell} \approx 0$  implies that the model estimates that the event is not a signal event. The model used in the trilepton channel follows the message-passing neural network architecture described in Section 4.2.2.

### 6.3.1 Dataset of graph events

The training and testing of the graph neural network was conducted using simulated events from the  $tWZ$  SR. The events from the nominal  $tWZ$  sample were assigned a target output of  $y = 1$ . The events from the major background samples of  $t\bar{t}Z$  and  $WZ+\text{jets}$  were assigned a target output of  $y = 0$ . To improve numerical stability during the training procedure, events with negative weights were excluded from the dataset.

A class imbalance exists in the dataset since the total sum of weights of the background events is larger than the sum of weights of the signal events. A class imbalance can negatively affect the training performance of classification models. To remove the imbalance, the signal events were reweighted before training by

$$w_{s,i}^{\star} = \frac{\sum_b w_b}{\sum_s w_s} w_{s,i}, \quad (6.4)$$

which  $\sum_b w_b$  is the sum of weights of the background events and  $\sum_s w_s$  is the sum of weights of the signal events and  $w_{s,i}$  is the event weight for the  $i$ th signal event. The reweighting procedure ensures that the signal and background datasets have the same sum of weights. The weighting of events is left unchanged during model evaluation.

### 6.3.2 Graph Construction

Each event can be converted into a graph which encodes all the required information associated with the event. The graph neural network is applied to these graph structures to generate scores. The features of an event are included as node features, edge features and global features of the graph.

The physics objects in an event can be represented as a set of nodes. A node will be placed in the graph for each lepton and jet object in the event as well as a single node

representing the  $E_T^{\text{miss}}$  object. Each node has an associated vector containing the features of the corresponding physics object. The features in this node vector include the transverse momentum  $p_T$ , psuedorapidity  $\eta$  and energy  $E$  of the object. The  $\phi$  component of the object is also included as  $\cos(\phi)$  and  $\sin(\phi)$  for better numerical stability.

The classification of the type of the object is encoded in the node vector using a one-hot encoding scheme. The seven possible physics object types in the analysis are missing transverse momentum, jet,  $b$ -tagged jet, electron, positron, muon and antimuon. The type of the physics object can be encoded in a 7-dimensional vector with one entry set to 1 and all others set to 0. For example, the feature vector for an electron would be

$$\mathbf{x}_e = \begin{pmatrix} p_T & \eta & \sin(\phi) & \cos(\phi) & E & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{pmatrix}^T. \quad (6.5)$$

A node equipped with an 12-dimensional feature vector can describe the kinematics and the type of the physics object.

Since physics objects are represented as nodes, the edges between the nodes are described by the angular differences between objects in the event. The edge weights between each node are taken to be the  $\Delta R$  between the corresponding physics objects. Event-level observables can be included as global graph information in the form of a global feature vector. The global features for each event include the number of jets and the mass of the  $W$ -candidate selected using the  $D_{jj}$  metric discussed in Section 6.2. Control plots for each of the global variables in each region can be seen in Figure 6.10. The features of the input graphs are summarised in Table 6.3.

Block Type	Features
Node	$p_T, \eta, E, \cos \phi, \sin \phi$ , object encoding
Edge	$\Delta R$
Global	$N_{\text{jets}}$ , Mass of the $W$ -candidate dijet

Table 6.3: Table detailing the features of the input graphs used in the trilepton GNN model. The features are separated into the separate node, edge and global features which are applied in the node, edge and global blocks of the model.

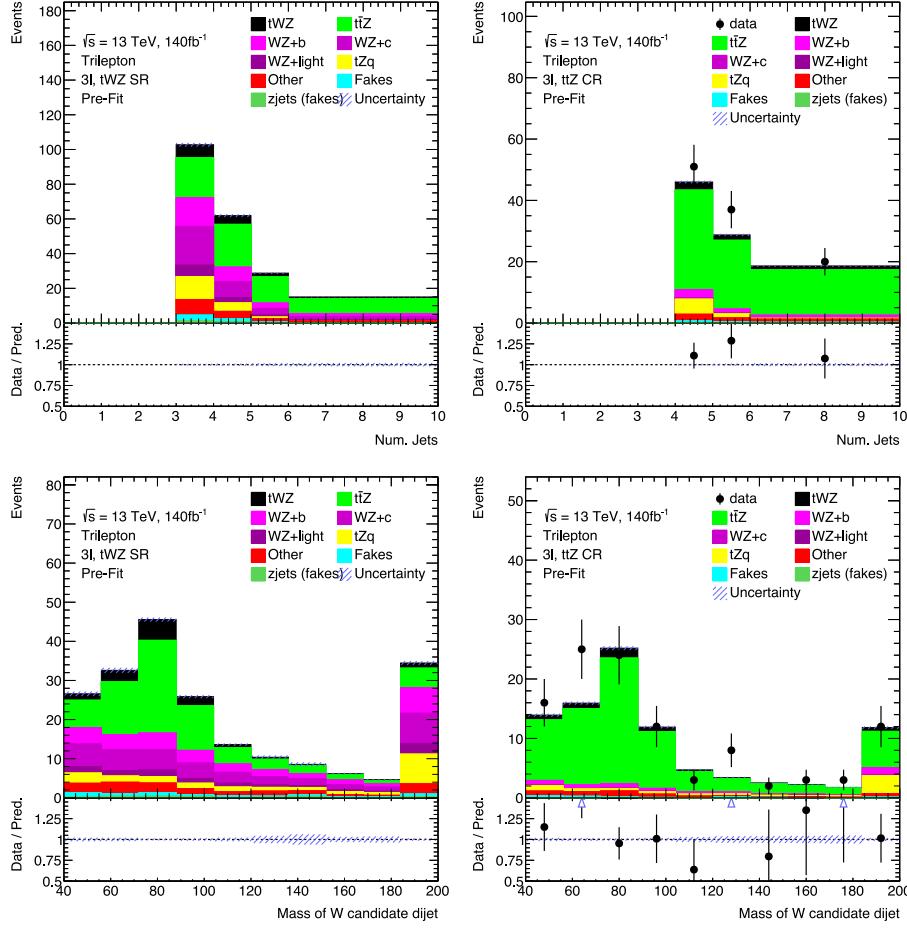


Figure 6.10: Control plots which show the distribution of event observables which were implemented as global features in the trilepton GNN. The top row shows the number of jets in the  $tWZ$  SR and the  $t\bar{t}Z$  CR. The bottom row shows the mass of the  $W$  boson candidate in each region.

### 6.3.3 Performance of trilepton GNN

The following results were obtained with a model that was trained using the ADAM [64] optimiser with a learning rate of  $\gamma = 0.001$ . The dataset of graphs was split such that 80% of events were used for training purposes and 20% were used for testing purposes. The classification of events in the testing dataset can be seen in Figure 6.11. In Figure 6.11, the model appears to discriminate the signal and background events in the  $tWZ$  SR where the distribution of the output scores for the signal events skew towards higher scores than the background events. The receiver operating characteristic (ROC) curve for the testing dataset can be seen in Figure 6.12 with an area under the ROC curve (AUC) score of 0.64. An AUC value of above 0.5 indicates that the predictor performs better than a random classification.

In both Figure 6.11 and Figure 6.12, the performance of the model when applied to the training dataset is also included. The performance on the training dataset allows for

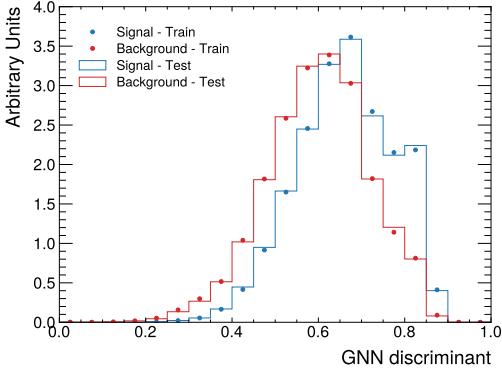


Figure 6.11: Normalised distribution of trilepton GNN output for the  $tWZ$  and background events in the testing and training dataset.

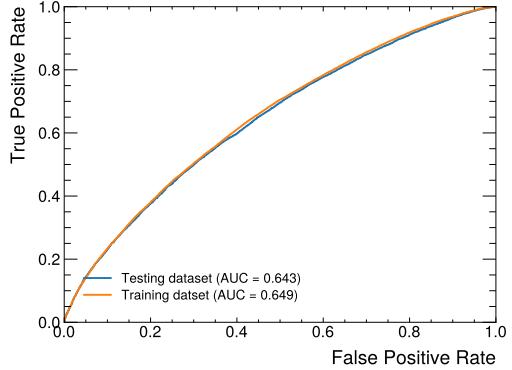


Figure 6.12: Receiver operating characteristic (ROC) curve for the trilepton GNN model using the testing and training dataset. The associated area under ROC curve (AUC) score.

possible overtraining of the model to be identified. The normalised distribution of GNN output scores for the training dataset is similar to the testing data. The AUC of the ROC curve for the training dataset is only slightly larger than that of the testing dataset which indicates no major overtraining of the model.

The distribution of the trilepton GNN output in the  $tWZ$  SR and the  $t\bar{t}Z$  CR is shown in Figure 6.13a and Figure 6.13b respectively. In both the  $tWZ$  SR and the  $t\bar{t}Z$  CR, the relative contribution per bin from the  $WZ + \text{jets}$  and  $tZq$  samples decreases as the classification score increases. This relationship motivates the use of the GNN classifier in the  $t\bar{t}Z$  CR to improve the estimation of the  $t\bar{t}Z$  normalisation uncertainty. In the  $tWZ$  SR, the relative contribution per bin of the  $tWZ$  sample increases as the classification score increases. The value of the separation metric between the signal and all background events in the  $tWZ$  SR was found to be  $\langle S^2 \rangle = 7.06\%$ . This is larger than the separation with respect to the mass of the  $W$  boson candidate which implies that the GNN model can separate signal and background better than using the mass of the  $W$  boson candidate.

## 6.4 Trilepton Asimov Fit Results

The signal strength of  $tWZ$  production is estimated by a maximum likelihood estimation using the likelihood function defined in Section 5.6.1. The signal extraction was performed on Asimov datasets in the signal and control regions to determine the experimental sensitivity of the analysis. The fitting procedure was performed on the distributions of the trilepton GNN output in both the  $tWZ$  SR and the  $t\bar{t}Z$  CR. The distribution of  $E_T^{\text{miss}}$  was considered in  $WZ$  CR in order to isolate the contributions from events with fake leptons. The differential distribution of the discriminating variables in each region can be found in Figure 6.13. The binning of the distributions in each region is deter-

mined using an automated binning algorithm known as TransfoD [65]. The TransfoD algorithm systematically merges bins in the distribution to maximise the proportion of signal over background events in certain bins.

The systematic uncertainties detailed in Section 5.3 are implemented as nuisance parameters which each introduce an additional Gaussian constraint term in the likelihood function. The nuisance parameters associated with the cross-section normalisation of background processes are implemented as an overall normalisation effect on the relevant sample while all other parameters associated with the modelling of background processes are treated as a shape effect.

During the fitting procedure, systematic variables in which the upward and downward variations do not vary sufficiently from the nominal value are pruned. Pruning simplifies the model by removing the normalisation or shape effects of the nuisance parameters and is performed per sample and per region. The normalisation component of a systematic variable is pruned for a sample in a region if all bins do not differ by more than 1% when the variable is varied. The pruning of the shape component follows the same condition.

The uncertainty associated with the measurement of  $\mu_{tWZ}$  due to statistical fluctuations is estimated by performing a *stat-only* fit. A stat-only fit uses a saturated model where all nuisance parameters are fixed to their respective post-fit estimations. The likelihood maximisation is then repeated with only the parameter of interest being allowed to vary.

The results of the maximum likelihood estimation using the trilepton Asimov dataset yielded an estimated signal strength of  $\mu_{tWZ} = 1.00^{+0.72}_{-0.68}(\text{stat.})^{+0.83}_{-0.71}(\text{syst.})$ . The statistical uncertainty associated with the measurement is the dominating uncertainty contribution. The large statistical uncertainty is attributed to the difficulty in separating signal and background events in the signal region as well as the low number of expected  $tWZ$  events at a luminosity of  $140 \text{ fb}^{-1}$ . The expected significance associated with the measurement was found to be  $Z^{\text{exp}} = 1.02\sigma$ . The expected significance of the measurement suggests that a rejection of the background-only hypothesis would not be possible if the observed data agreed with the simulated samples. When the  $E_T^{\text{miss}}$  requirement is applied, the expected significance  $Z_0^{\text{exp}}$  was reduced from  $1.03\sigma$  to  $1.02\sigma$ . The  $E_T^{\text{miss}}$  requirement results in a small decrease in expected significance but provides improvements to the comparison between observed data events and simulated events. Using the  $\text{CL}_s$  method, the expected upper limit with a confidence interval of 95% was measured as  $\mu_{up}^{\text{exp}} = 2.09^{+2.27}_{-0.97}$ .

The correlation matrix associated with the fitting procedure is shown in Figure 6.14. The correlation matrix shows that the cross-section normalisations of the primary backgrounds are anti-correlated with the parameter of interest  $\mu_{tWZ}$ . An anti-correlated relationship is expected as an increase in the number of background events implies a lower

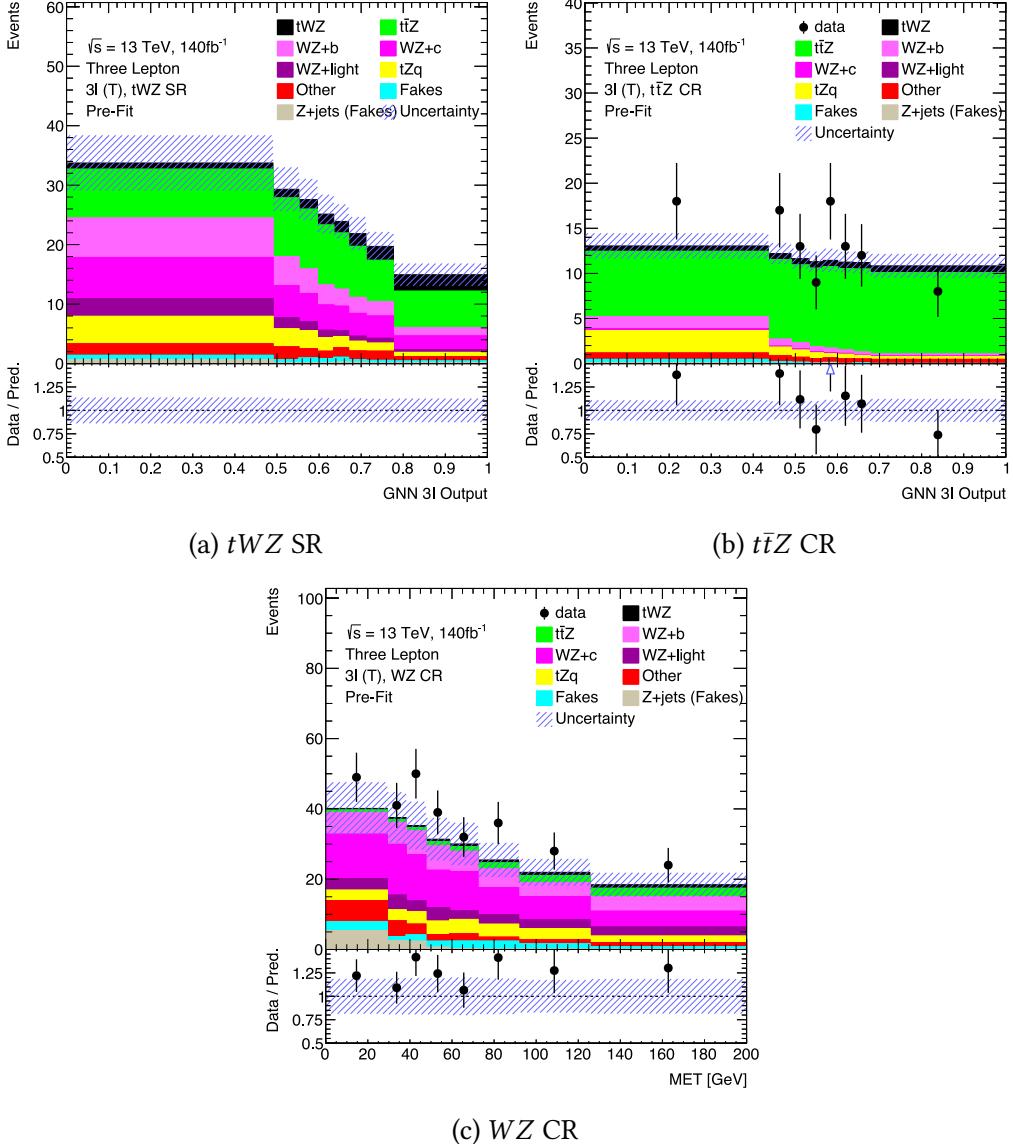


Figure 6.13: Distributions of the discriminating variable for each region in the trilepton channel. The classification score of the trilepton GNN model is shown for (a)  $tWZ$  SR and (b)  $t\bar{t}Z$  CR and the  $E_T^{\text{miss}}$  is shown in (c)  $WZ$  CR. The observed data is shown in the control regions.

proportion of the observed data is due to the production of  $tWZ$ . The cross-section normalisation nuisance parameters associated with  $WZ+\text{jets}$  production are all anti-correlated with each other.

The impact of a nuisance parameter  $\Delta\mu$  indicates how changing the value of a parameter will change the extracted signal strength. The pre-fit impact of a nuisance parameter  $\theta$  is defined by the change in  $\mu_{tWZ}$  between the nominal fit and a fit where  $\theta$  is fixed as  $\hat{\theta} \pm 1$ . The value  $\hat{\theta}$  is the estimated value of  $\theta$  in the nominal fit. The post-fit impact is calculated similarly except  $\theta$  is fixed as  $\hat{\theta} \pm \Delta\theta$  where  $\Delta\theta$  is the uncertainty of the estimated value of  $\theta$  in the nominal fit.

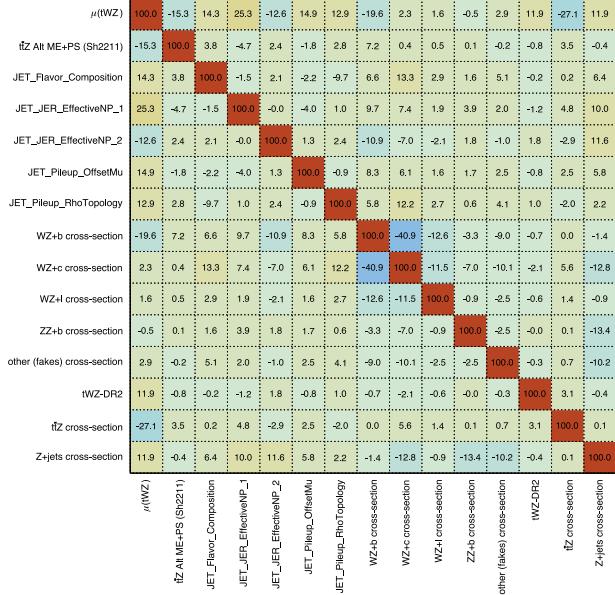


Figure 6.14: The correlation matrix associated with the Asimov fit in the trilepton channel with select nuisance parameters. Only nuisance parameters with a correlation higher than 10% are included. Each correlation is shown in units of percentage.

The nuisance parameters with the highest impact are shown as a ranking in Figure 7.8. The nuisance parameters associated with the cross-section normalisations of the  $WZ + \text{jets}$  and  $t\bar{t}Z$  background processes have the highest impact on the fitting procedure. As a result, the modelling of the background contributions has the largest effect on the systematic uncertainty associated with estimating  $\mu_{tWZ}$ . The post-fit impact of the  $WZ+b$  and  $WZ+c$  normalisation uncertainties is smaller than their pre-fit impact which indicates that the fit is constraining these systematic uncertainties.

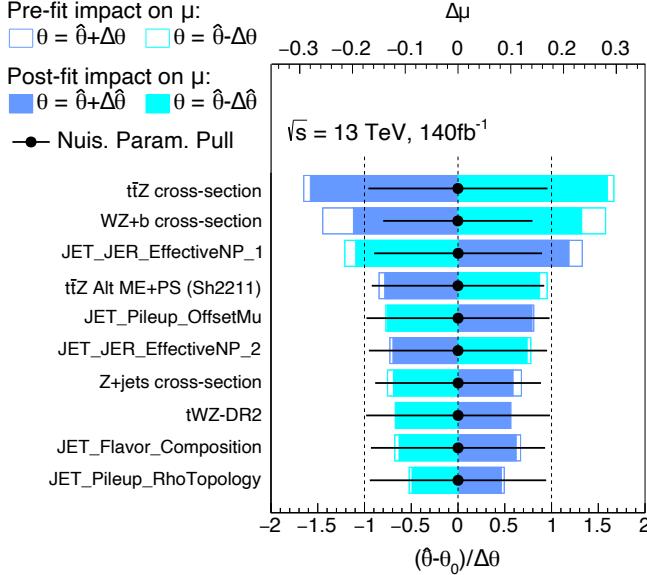


Figure 6.15: Ranking of the pre-fit and post-fit impact values of each nuisance parameter in the fit to the trilepton Asimov dataset. Only the top ten highest-impact nuisance parameters are shown. The pre-fit impacts are shown as unfilled rectangles and the post-fit impacts are shown as filled rectangles. The value and uncertainty of each nuisance parameter estimated by the fitting procedure are included as black points. The upper axis is in terms of impact  $\Delta\mu$  and the lower axis is in terms of the relative change in the nuisance parameter during the fit.

## 6.5 Signal injection test

To evaluate the reliability of the fitting procedure described in Section 6.4, a signal injection test was used. Signal injection tests adjust the size of the signal contribution in the fitting procedure to determine if the procedure can correctly extract this new signal strength. The test begins by constructing toy datasets which have the same background components as the trilepton Asimov dataset but have an adjusted signal component with a different signal strength. For this analysis, the injected signal strengths used in the creation of the toy datasets were 0.5, 1.5, 2 and 3. The fitting procedure is then performed on each toy dataset and the extracted signal strength is compared to the true signal strength of the dataset. A reliable fitting procedure should extract a signal strength that is the same as the injected signal strength.

The relationship between the injected and extracted signal strength was estimated by performing an unweighted linear regression to obtain a gradient and an intercept. The uncertainty associated with the extracted signal strengths is ignored in this test as very little of the uncertainty is associated with the fitting procedure itself. For a reliable fitting procedure, the gradient should be in agreement with 1 and the intercept should be negligible. Any deviation from these expectations indicates a bias in the fitting procedure.

A signal injection test was performed using the trilepton Asimov dataset with a comparison between the injected and extracted signal strengths shown in Figure 6.16. The estimated gradient is in agreement with the expected value of 1. The estimated intercept agrees with 0 within 2 standard uncertainties but the relative size of the intercept is small. Therefore, the fitting procedure can reliably extract the signal strength from Asimov datasets.

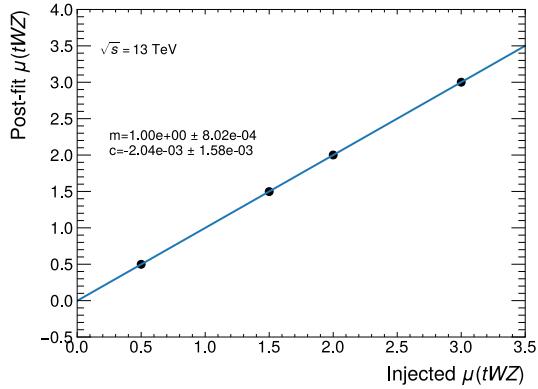


Figure 6.16: Comparison between the extracted signal strength and the injected single strength for the trilepton toy datasets. A line of best fit is shown that was obtained through linear regression. The fitted gradient  $m$  and intercept  $c$  parameters of the line are included.

# Chapter 7

## Analysis of the $tWZ$ tetralepton channel

The following chapter will detail the search for  $tWZ$  production in the tetralepton channel. The final state products of  $tWZ$  production can decay via the tetralepton decay channel. The tetralepton channel requires that both the  $W$  boson from the primary interaction and the  $W$  boson from the decaying top quark decay leptonically. The requirement of two leptonically-decaying  $W$  bosons drastically reduces the number of events in the tetralepton channel as  $W \rightarrow \ell\nu_\ell$  has a branching ratio of 10.86% [7]. Therefore, fewer total  $tWZ$  events are available when performing the signal extraction in the tetralepton channel than in the trilepton channel. The major background process in the tetralepton channel is  $t\bar{t}Z$  production via the tetralepton decay mode in which the  $W$  bosons from each top quark decay leptonically. The production of  $ZZ$  dibosons with additional jets also provides a background contribution since both  $Z$  bosons can decay leptonically to produce a tetraleptonic decay signature.

Similarly to the figures in Chapter 6, some of the non-major background samples are grouped in this chapter's figures. The label *Other 4l* represents the group of non-major background processes which produce four prompt leptons. The label *Other fakes* represents the group of non-major background processes which produce less than four prompt leptons.

The region definitions are introduced in Section 7.1. Similarly to the model utilised in the trilepton channel, a GNN is developed as a signal-background discriminator in Section 7.3. The blinded results for the measurement of  $tWZ$  signal strength in the tetralepton channel are shown in Section 7.4.

## 7.1 Baseline selections and region definitions

The trigger requirements for the event selection in tetralepton use the same triggering strategy as detailed in Section 6.1. Each event is required to have four leptons where the highest energy lepton has  $p_T > 28 \text{ GeV}$ , the second highest energy lepton has  $p_T > 18 \text{ GeV}$  and the third and fourth highest energy leptons have  $p_T > 10 \text{ GeV}$ . All jets must have  $\text{JVT} > 0.5$  in order to suppress pileup jets. In the tetralepton channel, two leptons are produced due to the decay of a neutral  $Z$  boson and two leptons are produced due to the decay of oppositely signed  $W$  bosons. Therefore, the sum of the lepton charges must be neutral. All possible pairs of opposite signed same flavoured (OSSF) leptons must have a mass of  $m_{\text{OSSF}} > 10 \text{ GeV}$ . A  $Z$  candidate is defined as any OSSF lepton pair that has a reconstructed mass within 30 GeV of the mass of the  $Z$  boson. Multiple  $Z$  candidates are possible in the tetralepton channel since the  $e^+e^-e^+e^-$ ,  $e^+e^-\mu^+\mu^-$  and  $\mu^+\mu^-\mu^+\mu^-$  lepton configurations are present in events.

Three signal regions are defined for the tetralepton channel. All signal regions must include 1  $Z$  candidate,  $\geq 1$  jet and exactly 1  $b$ -tagged jet. Two of the signal regions ( $tWZ$  OF SR and  $tWZ$  SF SR) require events to have 4 tight leptons. The only difference between these two tight signal regions is the flavours of the 2 leptons which are not associated with the reconstruction of the  $Z$  candidate. If an event has 4 tight leptons and the two non- $Z$  leptons have opposite flavours, the event is placed in the  $tWZ$  OF SR. If an event has 4 tight leptons and the two non- $Z$  leptons have the same flavour, the event is placed in the  $tWZ$  SF SR. The motivation for this distinction is to isolate the majority of the  $ZZ$ +jets contributions into a single signal region. If two leptons from a  $ZZ$  event are used to reconstruct the  $Z$  candidate, the other two leptons will have the same flavour.

A signal region with looser requirements was defined where 3 of the leptons must be tight and 1 of the leptons be loose and not tight. The looser selection increases the overall number of  $tWZ$  events but introduces additional contributions from background events that contain fake leptons. The region is referred to as the  $tWZ$  3T1L SR where the non-lepton object multiplicity requirements are the same as  $tWZ$  OF SR and  $tWZ$  SF SR.

Control regions were defined for each major background process. The  $t\bar{t}Z$  control region ( $t\bar{t}Z$  CR) requires the event to have 4 tight leptons, 1  $Z$  candidate,  $\geq 2$  jet and exactly 2  $b$ -tagged jet. The requirement of 2  $b$ -tagged jets is intended to select the  $b$  quarks produced during the decays of the top quarks. The  $ZZb$  control region ( $ZZb$  CR) requires the event to have 4 tight leptons, 2  $Z$  candidates,  $\geq 1$  jet and exactly 1  $b$ -tagged jet. The channel-specific selection and the region definitions are summarised in Table 7.1

<b>Baseline selections</b>				
$p_T(\ell_1, \ell_2, \ell_3, \ell_4) > (28, 18, 10, 10)\text{GeV}$ $p_T(\text{jet}) > 20\text{ GeV},  \eta(\text{jet})  < 2.5, \text{JVT} > 0.5$ $ \eta(\ell_e)  < 2.47 \text{ excluding } 1.37 <  \eta(\ell_e)  < 1.52$ $ \eta(\ell_\mu)  < 2.5$ $\sum_\ell q_\ell = 0$ All OSSF lepton pairs require $m_{\text{OSSF}} > 10\text{ GeV}$				
<b>Regions</b>				
$tWZ$ OF SR	$tWZ$ SF SR	$tWZ$ 3T1L SR	$t\bar{t}Z$ CR	$ZZb$ CR
$N_\ell = 4$ 1 Z Candidate $\geq 1$ jet $1 b$ -jet Opp. Flavour Non-Z leptons	$N_\ell = 4$ 1 Z Candidate $\geq 1$ jet $1 b$ -jet Same Flavour Non-Z leptons	$N_\ell = 3, N_{\ell, \text{not tight}} = 1$ 1 Z Candidate $\geq 1$ jet $1 b$ -jet	$N_\ell = 4$ 1 Z Candidate $\geq 2$ jet $2 b$ -jet	$N_\ell = 4$ 2 Z Candidate $\geq 1$ jet $1 b$ -jet

Table 7.1: A summary of the tetralepton channel-specific baseline selections for object definitions and the requirements for the signal and control regions.

The expected number of events for each sample is shown in Figure 7.1 with the detailed yields shown in Table 7.2. The low number of expected events available in the tetralepton channel produces large statistical uncertainties associated with each bin height. The observed ATLAS data in the control regions is in agreement with the simulation events which shows that the overall expected rates in the control regions are well modelled by the Monte Carlo samples. The  $t\bar{t}Z$  CR region is dominated by  $t\bar{t}Z$  production with a purity of about 84% which allows for a precise estimation of the tetralepton  $t\bar{t}Z$  production. The  $ZZb$  CR region is dominated by  $ZZ + b$  and  $ZZ + c$  production which allows for a better estimation of cross section normalisation on each of the heavy flavoured components compared to the conservative pre-fit uncertainty. The control regions in the tetralepton channel have a far higher sensitivity to their respective background process of interest compared to the trilepton channel but this sensitivity is harmed by the low overall number of events in each region.

The selection placed on the two non- $Z$  leptons in the signal region definitions leads to the  $tWZ$  SF SR having a large contribution from  $ZZ + \text{jets}$  events while the  $tWZ$  OF SR contains negligible contributions from  $ZZ + \text{jets}$  events. The selection enables  $t\bar{t}Z$  production to be the only major background contribution in  $tWZ$  OF SR. Due to the looser lepton requirement in the  $tWZ$  3T1L SR region, the region includes background contributions from processes with an additional fake/non-prompt lepton. The number of  $t\bar{t}Z$  events which include a fake lepton is explicitly labelled in Figure 7.1. The simulated samples show that few events with fake/non-prompt leptons are present outside the

$tWZ$  3T1L SR. The additional isolation criteria applied to each tight lepton appear to successfully remove these fake contributions.

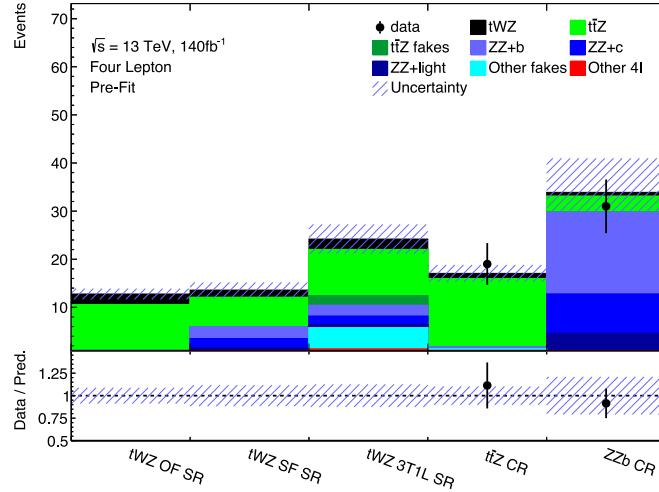


Figure 7.1: Number of simulated events for all regions in the tetralepton channel. The amount of signal and background Monte Carlo events are shown in different colours on the upper panel. The ratio between the number of data and simulated events in each region is plotted on the lower panel. The observed ATLAS data is only shown in the control regions as all signal regions are blinded.

	$tWZ$ OF SR	$tWZ$ SF SR	$tWZ$ 3T1L SR	$t\bar{t}Z$ CR	$ZZb$ CR
$tWZ$	$2.24 \pm 0.21$	$1.52 \pm 0.27$	$2.17 \pm 0.33$	$1.1 \pm 0.4$	$0.81 \pm 0.09$
$t\bar{t}Z$	$9.5 \pm 1.0$	$6.1 \pm 0.7$	$9.6 \pm 1.1$	$14.1 \pm 1.5$	$3.3 \pm 0.4$
$t\bar{t}Z$ (fake)	$0.033 \pm 0.020$	$0.014 \pm 0.012$	$2.0 \pm 1.0$	$0.025 \pm 0.019$	$0.017 \pm 0.013$
$ZZ + b$	$0.17 \pm 0.07$	$2.5 \pm 0.8$	$2.3 \pm 0.7$	$0.40 \pm 0.13$	$17 \pm 5$
$ZZ + c$	$0.09 \pm 0.05$	$1.9 \pm 0.6$	$1.8 \pm 0.6$	$0.047 \pm 0.020$	$8.2 \pm 2.6$
$ZZ + l$	$0.031 \pm 0.020$	$0.73 \pm 0.17$	$0.59 \pm 0.14$	$0.012 \pm 0.009$	$4.3 \pm 0.8$
$t\bar{t}t\bar{t}$	$0.18 \pm 0.06$	$0.14 \pm 0.05$	$0.14 \pm 0.05$	$0.072 \pm 0.024$	$0.11 \pm 0.04$
$t\bar{t}WW$	$0.010 \pm 0.027$	$0.024 \pm 0.019$	$0.07 \pm 0.04$	$0.11 \pm 0.06$	
$t\bar{t}H$	$0.50 \pm 0.06$	$0.37 \pm 0.05$	$1.09 \pm 0.13$	$0.87 \pm 0.10$	$0.102 \pm 0.014$
$ZH$		$0.2 \pm 0.5$			
$tZq$	$0.012 \pm 0.009$	$0.005 \pm 0.005$	$0.36 \pm 0.19$	$0.005 \pm 0.005$	$0.003 \pm 0.004$
$WZ+jets$	$0.004 \pm 0.006$	$0.010 \pm 0.011$	$0.9 \pm 0.5$		$0.002 \pm 0.007$
$t\bar{t}W$	$0.009 \pm 0.009$	$0.002 \pm 0.006$	$0.42 \pm 0.22$		$0.006 \pm 0.007$
$t\bar{t}t$	$0.0017 \pm 0.0013$	$0.0020 \pm 0.0018$	$0.7 \pm 0.4$	$0.07 \pm 0.07$	$0.03 \pm 0.07$
$Z + jets$			$0.05 \pm 0.28$		$0.3 \pm 0.7$
$t\bar{t}\gamma$	$0.04 \pm 0.08$	$0.07 \pm 0.14$	$0.6 \pm 0.4$	$0.24 \pm 0.16$	$0.04 \pm 0.10$
$WH$					
$tW$					
$Z\gamma$			$0.24 \pm 0.25$		$0.004 \pm 0.012$
Total	$12.8 \pm 1.3$	$13.6 \pm 1.6$	$23 \pm 3$	$17.1 \pm 1.8$	$34 \pm 6$

Table 7.2: Expected number of events according to simulation in each of the tetralepton regions for each process or group of processes. Empty cells indicate no simulated events are present in the region. The uncertainty associated with each yield is shown and comprises both statistical and systematic uncertainties.

## 7.2 Two Neutrino Scanning Method

The primary difference between the experimental signature of the tetraleptonic decay of  $tWZ$  and  $t\bar{t}Z$  is the presence of a  $t\bar{t}$  pair. A  $t\bar{t}$  system can be reconstructed from the kinematics of top decay products which includes the leptons,  $b$ -tagged jets and neutrinos. However, the information regarding the kinematics of each individual neutrino is unknown as the ATLAS detector cannot measure neutrinos. The two neutrino scanning method ( $2\nu SM$ ) algorithm attempts to reconstruct a viable  $t\bar{t}$  system by scanning over possible neutrino kinematics. The success of the  $t\bar{t}$  reconstruction is determined by comparing observables of the reconstructed system to a simulated dilepton  $t\bar{t}$  sample. The  $2\nu SM$  algorithm was developed by the latest ATLAS Run 2  $t\bar{t}Z$  analysis [54] where it was used to separate backgrounds in the dilepton and tetralepton channels.

The inputs of the algorithm are the kinematics of two leptons and the kinematics of two jets. When considering the lepton inputs for the algorithm, the two non- $Z$  leptons are used. The two jets with the highest DL1r  $b$ -tagging score are chosen as inputs and the jets are matched to the closest lepton with respect to  $\Delta R$ .

### 7.2.1 Estimating observable distributions

The algorithm requires estimating the probability distributions of kinematic observables of top quarks and their decay products. The kinematic observables of interest are the mass of the tops  $m_{t_1}$  and  $m_{t_2}$  and the amount of transverse momentum not attributed to the neutrinos  $\Delta E_x$  and  $\Delta E_y$ . The observable probability distributions are calculated using a dedicated simulated  $t\bar{t}$  sample where reconstructed objects and event generator-level information are available. The relevant observables are the mass of the top quarks  $m_t$  and the difference between the transverse momenta of the neutrinos and the missing transverse momentum  $\Delta \mathbf{E}_T$ .

The probability distribution of the mass of the top quarks  $\text{Pr}_{m_t}$  is found using kinematic information from the simulated sample. The mass of a top  $m_t$  for a simulated top decay is calculated by

$$m_t^2 = \ell^2 + b^2 + v^2$$

where  $\ell$  are lepton 4-vectors,  $b$  are the jet 4-vectors and  $v$  are the neutrino 4-vector. The lepton and jet kinematics are obtained from detector-level information while the neutrino information is extracted from the event generator. Each simulated  $t\bar{t}$  event provides two entries to the probability distribution. The difference between the transverse momenta of the neutrinos and the missing transverse momentum  $\Delta \mathbf{E}_T = (\Delta E_x \ \Delta E_y)^T$  is calculated with

$$\begin{aligned}\Delta E_x &= (p_{T,v_1})_x + (p_{T,v_2})_x - (E_T^{\text{miss}})_x \\ \Delta E_y &= (p_{T,v_1})_y + (p_{T,v_2})_y - (E_T^{\text{miss}})_y\end{aligned}$$

where the  $x$  and  $y$  components are considered separately. Due to azimuthal symmetry in the transverse plane, both  $\Delta E_x$  and  $\Delta E_y$  of  $t\bar{t}$  events will have the same probability distribution  $\text{Pr}_{\Delta E}$ . The  $E_T^{\text{miss}}$  is determined from detector-level information and the neutrino kinematics are determined by generator-level information. The probability distributions of  $m_t$  and  $\Delta E$  for the simulated  $t\bar{t}$  sample are shown in Figure 7.2. The  $\Delta E_T$  of the  $t\bar{t}$  events peaks at 0 GeV as most of the transverse momentum is due to the undetected neutrinos.

### 7.2.2 Algorithm Procedure

The goal of the procedure is to find a set of suitable neutrino kinematics  $v_1$  and  $v_2$  which produces the  $t\bar{t}$  system which is most similar to the dedicated sample. The algorithm assumes different values of  $\phi_\nu$  and  $\eta_\nu$  to calculate the  $p_T$  of the neutrinos. The  $p_T$  of a

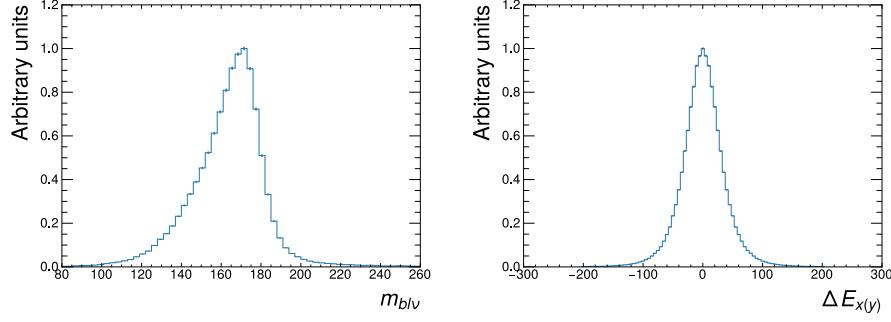


Figure 7.2: Observable distributions determined from simulated  $t\bar{t}$  events. The combined mass of the decay products of top quarks  $m_{b\ell\nu}$  (left) is calculated using the detector-level jet and lepton and event generator-level neutrino information. The difference in transverse momentum between the generator-level neutrinos and the detector-level missing transverse momentum  $\Delta E_{x(y)}$  (right) is shown.

neutrino from a decaying  $W$  boson can be determined by

$$p_{T,\nu} = \frac{\frac{1}{2} (m_W^2 - m_\ell^2)}{E_\ell \cosh(\eta_\nu) - p_{\ell,z} \sinh(\eta_\nu) - p_{\ell,x} \cos(\phi_\nu) - p_{\ell,y} \sin(\phi_\nu)}. \quad (7.1)$$

where  $m_W$  is the mass of the  $W$  boson and  $m_\ell$  is the mass of the associated charged lepton. Using 7.1 to determine  $p_{T,\nu_1}$  and  $p_{T,\nu_2}$ , the required observables  $m_{t_1}$ ,  $m_{t_2}$ ,  $\Delta E_x$ , and  $\Delta E_y$  can be calculated for an event. The observables are evaluated on the probability distribution functions in Figure 7.2 and the 2 $\nu$ SM score is calculated by,

$$\omega_{2\nu SM} = \Pr_{m_t}(m_{t_1}) \Pr_{m_t}(m_{t_2}) \Pr_{\Delta E}(\Delta E_x) \Pr_{\Delta E}(\Delta E_y). \quad (7.2)$$

The  $\omega_{2\nu SM}$  score for an event can vary between 0 and 1 where a higher score implies that the kinematic configuration is similar to a  $t\bar{t}$  system. For a given event, the algorithm scans over different values of  $\nu_\eta$  and  $\nu_\phi$  for each neutrino to find the configuration that produces the maximum  $\omega_{2\nu SM}$  score. The values of  $\nu_\eta$  are varied between -5 and 5 and the values of  $\nu_\phi$  are varied between  $-\pi$  and  $\pi$ .

In some scenarios, the algorithm can fail if the required input particles are unavailable in an event. Some regions in the tetralepton channel allow for events with less than two jets, yet the algorithm requires at least two jets. The maximum  $\omega_{2\nu SM}$  score for events with less than two jets is set to 0 as no valid  $t\bar{t}$  system can be reconstructed.

### 7.2.3 Performance of Two Neutrino Scanning Method

The  $t\bar{t}Z$  events are expected to have a higher maximum  $\omega_{2\nu SM}$  than the  $tWZ$  events as a resonant  $t\bar{t}$  system is produced during  $t\bar{t}Z$  production. The event distribution of the maximum  $\omega_{2\nu SM}$  scores are shown for the  $tWZ$  OF SR and the  $t\bar{t}Z$  CR in Figure 7.3. The

$tWZ$  OF SR shows an increase in the ratio between the number of signal and background events as the maximum  $\omega_{2vSM}$  score decreases. In the  $t\bar{t}Z$  CR, there is reasonable agreement between the number of data and simulated events as well as an increase in the number of  $t\bar{t}Z$  events at higher maximum  $\omega_{2vSM}$  scores. Therefore, the maximum  $2vSM$  score discriminates between  $tWZ$  and  $t\bar{t}Z$  events and can be introduced as an input to a multivariate machine learning algorithm.

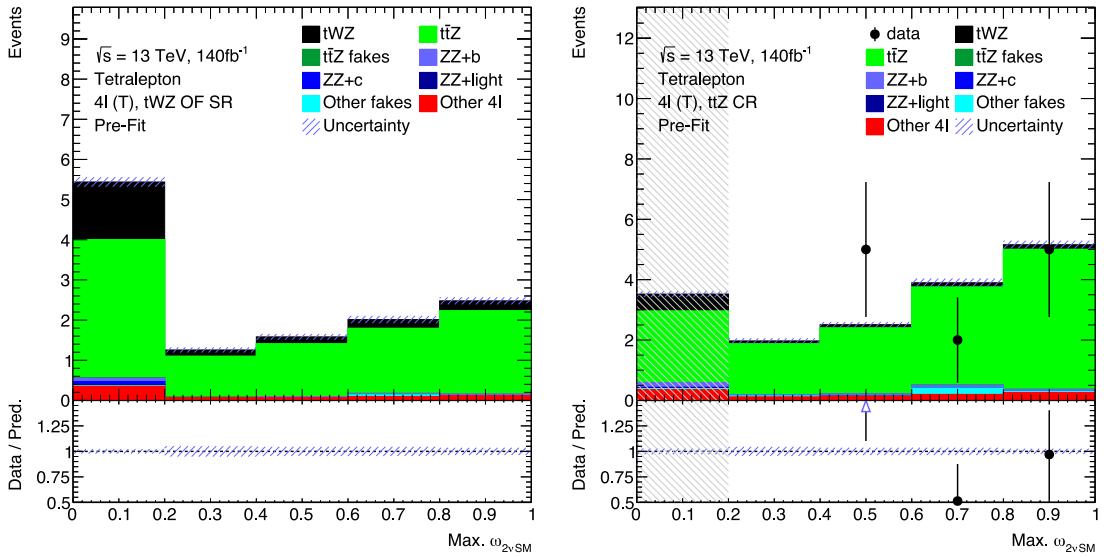


Figure 7.3: The distribution of events with respect to the maximum  $\omega_{2vSM}$  scores in the  $tWZ$  OF SR (left) and  $t\bar{t}Z$  CR (right). A different colour represents each simulated sample. Data is included in the  $t\bar{t}Z$  CR as black points except in the first bin which is blinded due to a high  $S/B$  ratio.

## 7.3 GNN in tetralepton

A GNN was developed to perform signal-background discrimination in the tetralepton channel to separate  $tWZ$  events from the contributions due to the production of  $t\bar{t}Z$  and  $ZZ+jets$ . The tetralepton model is a separate discriminator from the model applied in the trilepton channel which was detailed in Section 6.3. The model architecture used is specified in Section 4.2.2.

### 7.3.1 Graph dataset

The dataset of graphs used for training and testing was created using the events from the  $tWZ$ ,  $t\bar{t}Z$  and  $ZZ+jets$  samples in the  $tWZ$  OF signal region, the  $tWZ$  SF signal region and the  $t\bar{t}Z$  control region. The  $tWZ$  events were assigned a target output of  $y = 1$  while the background events from the  $t\bar{t}Z$  and  $ZZ+jets$  samples were assigned a target output of  $y = 0$ . Events were reweighted during training using the same procedure described

in 6.4 for the trilepton GNN. When evaluating the model’s performance, events were weighted according to their regular event weights.

### 7.3.2 Graph features

The input graphs for the tetralepton GNN were constructed similarly to the trilepton GNN in Section 6.3.2. The features of each node in the graph are determined by the properties of the associated physics object. The node vectors consist of the  $p_T$ ,  $\eta$ ,  $E$ ,  $\cos(\phi)$ ,  $\sin(\phi)$  and the object encoding for the physics object in the event. The edge values between nodes are defined by the  $\Delta R$  between the physics objects associated with the nodes.

Block Type	Features
Node	$p_T$ , $\eta$ , $E$ , $\cos \phi$ , $\sin \phi$ , object encoding
Edge	$\Delta R$
Global	$H_T$ , $H_{T,b}$ , $N_{jets}$ , $L_T$ , $S_T$ , Maximum $2\nu SM$ score

Table 7.3: Table of features for the input graphs used in the tetralepton GNN model. The features are separated into the separate node, edge and global features which are applied in the node, edge and global blocks of the model.

The graphs used in the tetralepton model also have global features. The global features associated with the entire event include the sum of jet  $p_T$  ( $H_T$ ), the sum of  $b$ -tagged jet  $p_T$  ( $H_{T,b}$ ), the sum of lepton  $p_T$  ( $L_T$ ) and the number of jets. The scalar sum of the  $p_T$  of all physics objects in the event  $S_T = H_T + L_T + E_T^{miss}$  is also considered as a global variable. The maximum  $\omega_{2\nu SM}$  score for the event as defined in Section 7.2 is also attached to each event graph. The graph features are summarised in Table 7.3. The distribution of global graph features in the  $tWZ$  SF SR, the  $tWZ$  OF SR and the  $t\bar{t}Z$  CR are shown in Figure 7.4. Examining the modelling in the  $t\bar{t}Z$  CR, there is reasonable agreement between the observed data and simulated events for the event observables.

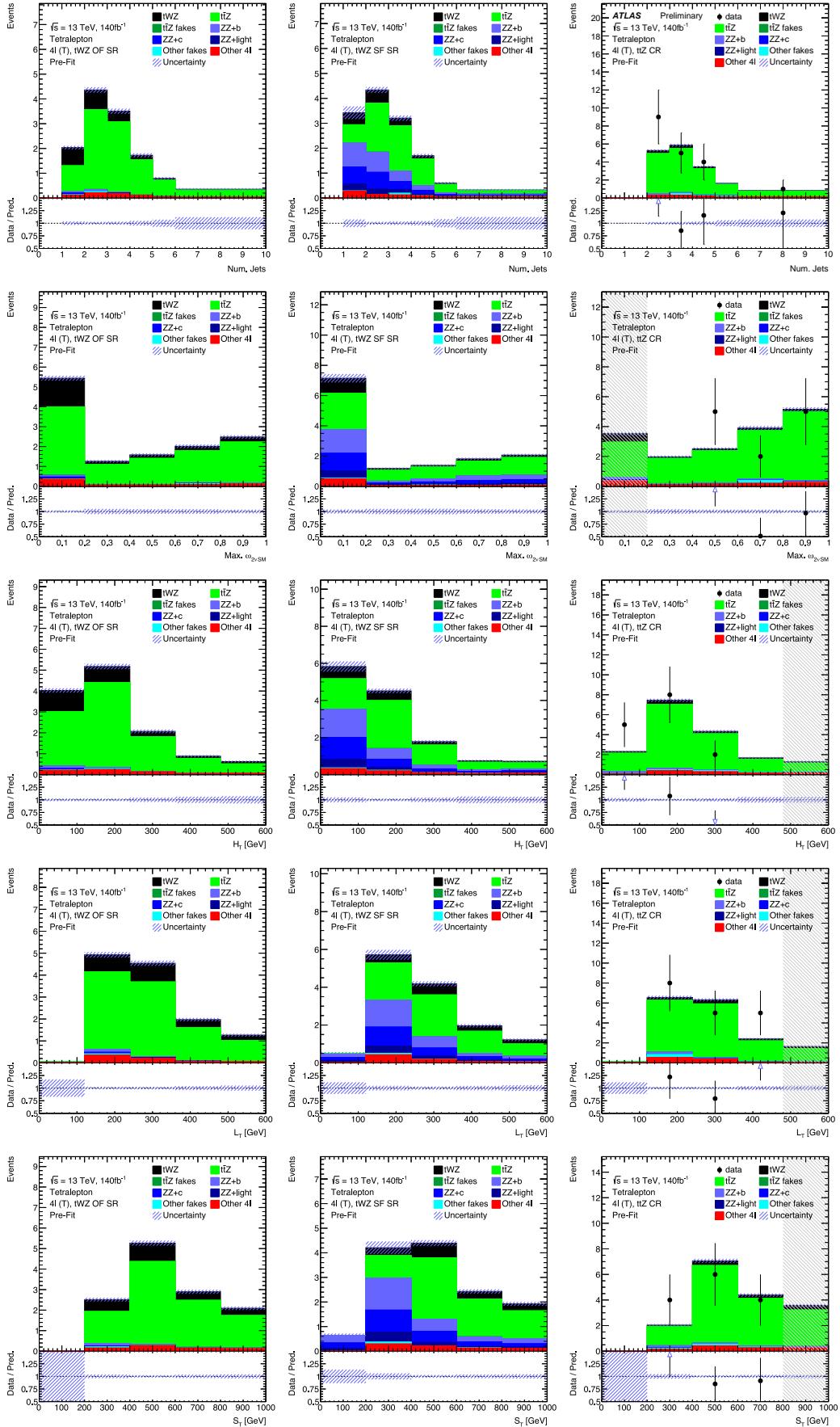


Figure 7.4: Control plots which show the distribution of event observables which were implemented as global features in the tetralepton GNN.

### 7.3.3 Performance of tetralepton GNN

The model was trained using the ADAM [64] optimiser with a learning rate of  $\gamma = 0.001$ . The graph dataset had a test-train split of 80% for training purposes and 20% for model evaluation. The normalised distribution of the output scores for the signal and background events are shown in Figure 7.5. In Figure 7.6, the ROC curve associated with the tetralepton model is shown with an AUC score of 0.74 for the testing dataset. The difference in the output score distributions and the ROC AUC score between the testing and training datasets is not notable, which shows no major overtraining has occurred.

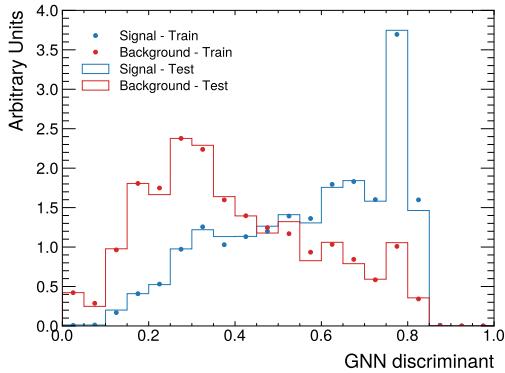


Figure 7.5: Comparison between the normalised distribution of tetralepton GNN output for the  $tWZ$  and background events in the testing and training dataset.

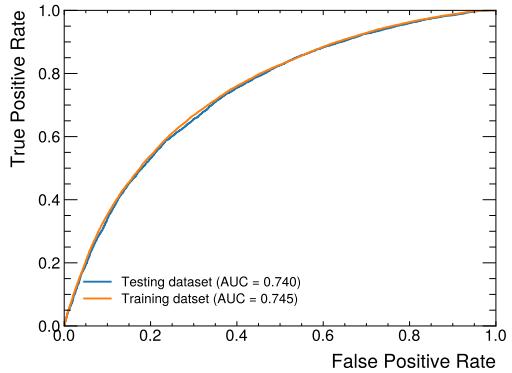


Figure 7.6: Receiver operating characteristic (ROC) curve for the tetralepton GNN model using the testing and training dataset. The associated area under ROC curve (AUC) score.

The tetralepton GNN output scores for each of the relevant regions are shown in Figure 7.7. All regions show that the relative contribution from  $tWZ$  production increases as the GNN output increases. In the regions  $tWZ$  SF SR and  $tWZ$  3T1L SR which have notable  $ZZ+jets$  contributions, the model can better discriminate  $ZZ+jets$  events by assigning low scores to these events.

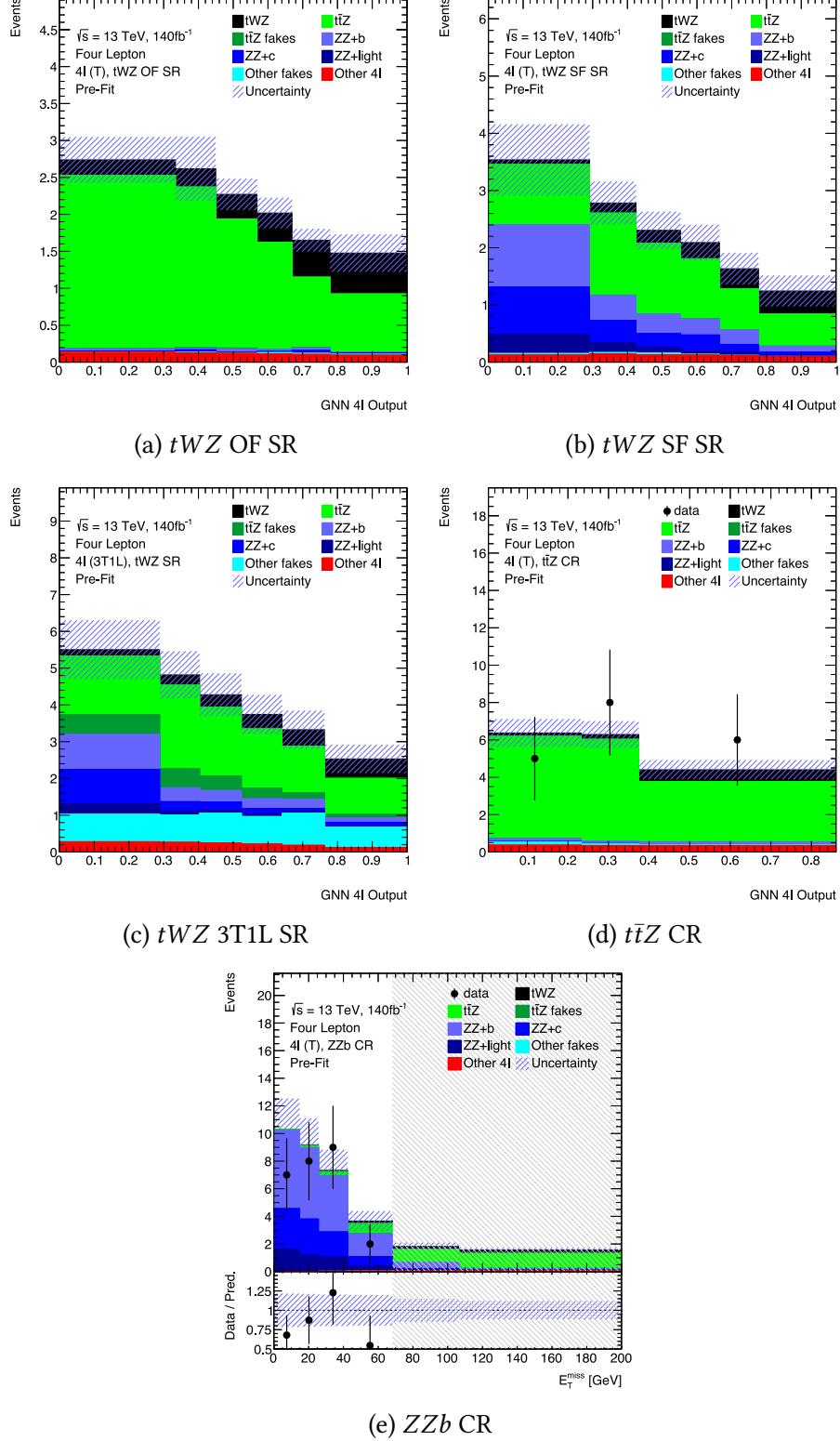


Figure 7.7: The distribution of the events in each region which are used in the tetralepton fitting procedure. A different colour represents each simulated sample. The events in the signal regions and the  $t\bar{Z}$  control region are shown in terms of their tetralepton GNN output value. The distribution of  $E_T^{\text{miss}}$  is shown in the  $ZZb$  control region. Observed data is included in the  $t\bar{Z}$  CR and  $ZZb$  CR.

## 7.4 Tetralepton Asimov Fit Results

The signal strength of  $tWZ$  production  $\mu_{tWZ}$  in the tetralepton channel was estimated using an Asimov dataset in the signal and control region. The signal extraction was performed using the tetralepton GNN output distributions in the signal regions and  $t\bar{t}Z$  CR as well as using the distribution of  $E_T^{\text{miss}}$  in the  $ZZb$  CR. The  $E_T^{\text{miss}}$  variable discriminates between  $ZZ + \text{jets}$  and  $t\bar{t}Z$  due to the presence of two energetic neutrinos in the tetraleptonic decay of  $t\bar{t}Z$ . The binning of distribution and the handling of systematic variables were implemented in the same manner as the fit in the trilepton channel.

The Asimov fit in the tetralepton channel estimated a signal strength of  $tWZ$  of  $\mu_{tWZ} = 1.00^{+0.95}_{-0.83}(\text{stat.})^{+0.51}_{-0.32}(\text{syst.})$ . The statistical uncertainty is the dominant uncertainty due to the low number of expected  $tWZ$  events in the tetralepton channel. The expected significance associated with the tetralepton Asimov fit was found to be  $Z^{\text{exp}} = 1.13\sigma$ . The expected upper limit with a confidence interval of 95% was estimated to be  $\mu_{up.}^{\text{exp}} = 2.00^{+2.47}_{-0.92}$ .

The impact of each nuisance parameter on the measurement of  $\mu_{tWZ}$  is ranked in Figure 7.8. The uncertainties associated with the reconstruction of jets appear to have the largest impact on the extraction in the tetralepton channel. The highest-ranking systematic variation is due to the uncertainty associated with the flavour composition of jets when calculating JES. The gluon fraction of the jets is estimated to be 50% with a conservative up and down variation of 50%. The flavour composition could be better estimated using truth information associated with the jets in each sample rather than the conservative estimation.

The next highest impact nuisance parameters are associated with the modelling of the signal and background processes. Due to the similarity between the  $tWZ$  and  $t\bar{t}Z$  processes, the cross-section normalisation of  $t\bar{t}Z$  process has a large impact on the signal extraction. The systematic variation between the nominal  $t\bar{t}Z$  sample and the alternative  $t\bar{t}Z$  sample also impacts the measurement due to the presence of  $t\bar{t}Z$  events in the signal regions.

Nuisance parameters which have high correlations in the tetralepton fit are shown in Figure 7.9. Similarly to the correlation matrix in the trilepton Asimov fit, the nuisance parameters associated with the cross-section normalisation of background processes are anti-correlated with the signal strength. The  $ZZ + b$  and  $ZZ + c$  cross-section normalisations are highly anti-correlated which is expected due to the similarities between the heavy flavour components of the  $ZZ + \text{jets}$  contribution.

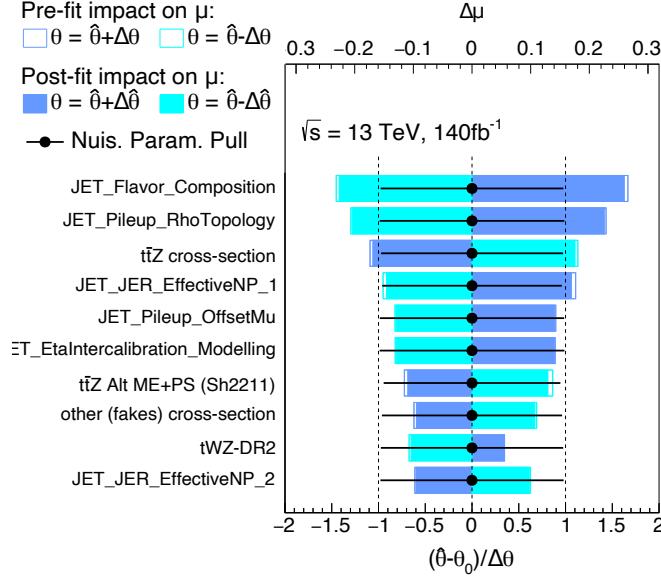


Figure 7.8: Ranking of the pre-fit and post-fit impact values for each nuisance parameter in the tetralepton Asimov fit. Only the top ten highest-impact nuisance parameters are shown. The pre-fit impacts are shown as unfilled rectangles and the post-fit impacts are shown as filled rectangles. The value and uncertainty of each nuisance parameter estimated by the fitting procedure are included as black points. The upper axis is in terms of impact  $\Delta\mu$  and the lower axis is in terms of the relative change in the nuisance parameter during the fit.

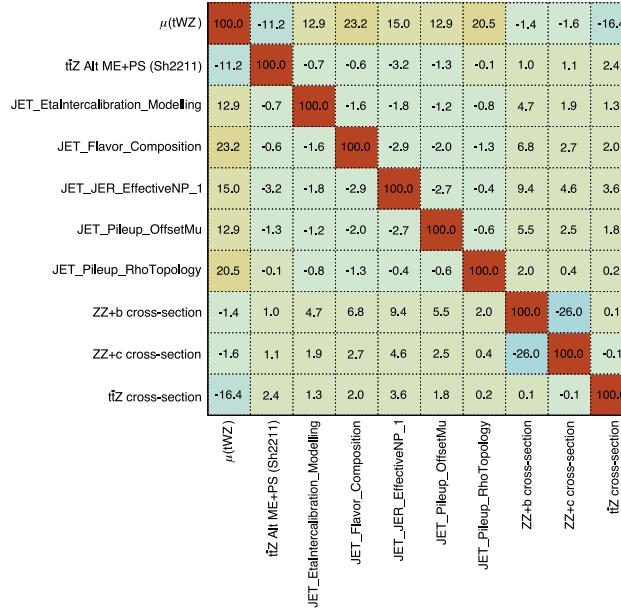


Figure 7.9: The correlation matrix of the nuisance parameters associated with the Asimov fit in the tetralepton channel. Each correlation is shown in units of percentage. Only nuisance parameters with a correlation higher than 10% are included.

## 7.5 Signal injection test

A signal injection test was performed on the tetralepton Asimov dataset using the procedure described in Section 6.5. The results of the signal injection test are shown in Figure 7.10. An unweighted linear regression was performed to determine the linear relationship between the extracted and the injected signal strengths. The gradient and intercept parameters of the fit agree with the expected values. Therefore, the fitting procedure in the tetralepton channel produces reliable extractions of the signal strength parameter.

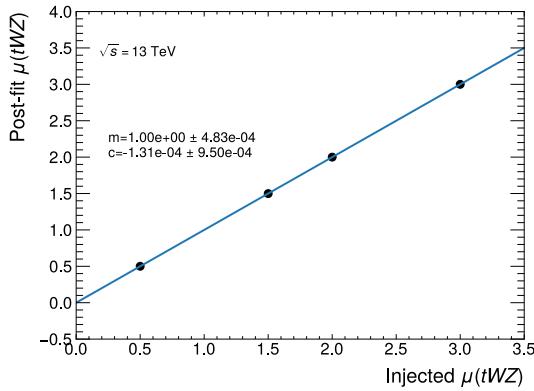


Figure 7.10: Comparison between the extracted signal strength and the injected single strength for the tetralepton toy datasets. A line of best fit is shown that was obtained through linear regression. The fitted gradient  $m$  and intercept  $c$  parameters of the line are included.

# Chapter 8

## Combined analysis

A search for a signal process can be performed in multiple channels, providing several measurements for the signal strength. For statistically limited searches such as the search for  $tWZ$  production, several measurements can be combined to improve the estimation of  $\mu_{tWZ}$ . The following chapter details the combined measurement of  $\mu_{tWZ}$  in both the trilepton and tetralepton channels.

### 8.1 Combined trilepton and tetralepton Asimov fit results

The fitting procedure for the combined fit used the regions from the trilepton and tetralepton channels when performing the maximum likelihood estimation. Any common systematic uncertainties in the trilepton and tetralepton channels were treated by a single nuisance parameter. Implementation and pruning of systematic uncertainties follow the same treatment detailed in the trilepton and tetralepton fitting procedures. The simultaneous fit to the trilepton and tetralepton Asimov datasets yielded a signal strength of  $\mu_{tWZ} = 1.00^{+0.56}_{-0.53}(\text{stat.})^{+0.54}_{-0.41}(\text{syst.})$ . The expected significance associated with the combined Asimov fit was  $Z^{\text{exp}} = 1.34\sigma$ . Therefore, evidence or observation of  $tWZ$  production is not expected to be seen when the ATLAS dataset is considered. The expected upper limit with a confidence interval of 95% was estimated to be  $\mu_{up.}^{\text{exp}} = 1.57^{+1.70}_{-0.73}$ .

The impact of the nuisance parameters in the combined fitting procedure is ranked in Figure 8.1. The highest impact nuisance parameters in the combined fit were also highly ranked in either the trilepton or tetralepton fits. The measured value of  $\mu_{tWZ}$  is most sensitive to the cross-section normalisation of  $t\bar{t}Z$  production as  $t\bar{t}Z$  events are a major contributing background in both the trilepton and tetralepton channels.

The fit results in the trilepton and tetralepton channels are compared in Figure 8.2 as well

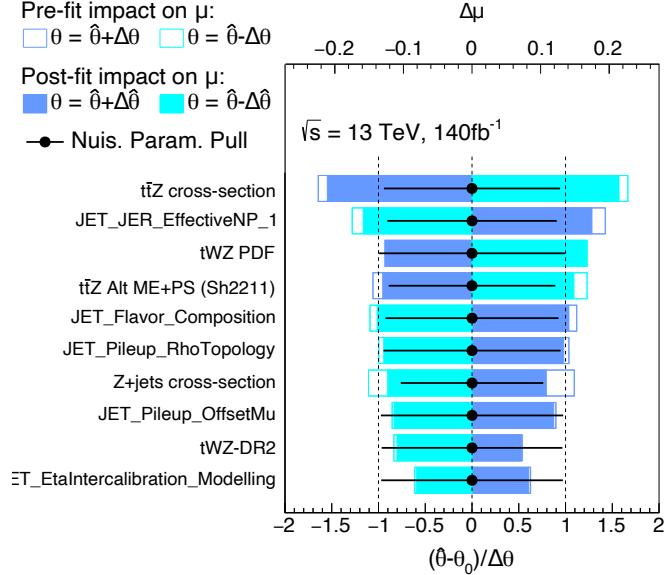


Figure 8.1: Ranking of the pre-fit and post-fit impact values of each nuisance parameter in the combined fit. Only the top ten highest-impact nuisance parameters are shown. The pre-fit impacts are shown as unfilled rectangles and the post-fit impacts are shown as filled rectangles. The value and uncertainty of each nuisance parameter estimated by the fitting procedure are included as black points. The upper axis is in terms of impact  $\Delta\mu$  and the lower axis is in terms of the relative change in the nuisance parameter during the fit.

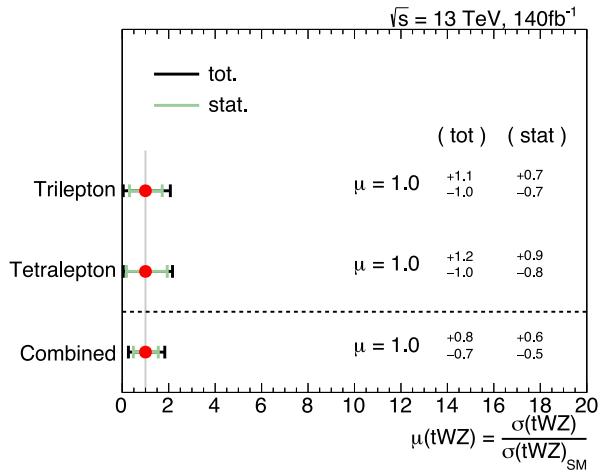


Figure 8.2: Comparison between the estimated value of  $\mu_{tWZ}$  and its uncertainty for various fits to different Asimov datasets. This includes the fit results from the trilepton-only, tetralepton-only and combined fits.

as the combined fit results. The trilepton and tetralepton results show similar overall uncertainties in their signal strength measurements. The statistical uncertainty associated with the measurement in the tetralepton channel is higher than in the trilepton channel due to the lower number of expected  $tWZ$  events. The trilepton channel results show a higher systematic uncertainty due to the poor modelling of the  $WZ+jets$  process

which does not impact the trilepton channel. As seen in Figure 8.2, the combined signal strength measurement across both the trilepton and tetralepton channels resulted in lower overall uncertainty than the individual measurements in each channel.

## 8.2 High luminosity estimations of $tWZ$ measurements

An observation of  $tWZ$  production using the full ATLAS Run 2 dataset was not possible using an Asimov dataset in the combined trilepton and tetralepton channel analysis. The leading contribution to the uncertainty on the estimation of  $\mu_{tWZ}$  is the associated statistical uncertainty due to the low estimated cross section of  $tWZ$  production. A higher luminosity dataset would allow for a tighter constraint of  $\mu_{tWZ}$  by decreasing the overall statistical uncertainty. The LHC has recently begun the Run 3 data-taking period that will increase the availability of high energy proton-proton collisions.

In the following section, the extraction of the signal strength  $\mu_{tWZ}$  was repeated for estimations of high luminosity datasets. The luminosity of the estimated datasets ranges from the expected size of the ATLAS Run 3 dataset ( $300 \text{ fb}^{-1}$ ) and the size of the dataset to be produced by the HL-LHC ( $3 \text{ ab}^{-1}$ ). The higher luminosity datasets were estimated by normalising both the Asimov dataset and Monte Carlo samples by the Run 2 luminosity of  $140 \text{ fb}^{-1}$  and then scaling the data by the high luminosity. The high luminosity targets of  $300 \text{ fb}^{-1}$ ,  $1000 \text{ fb}^{-1}$  and  $1000 \text{ fb}^{-1}$  were estimated. The estimated datasets do not account for the decrease in the estimations of systematics uncertainties that will occur as the understanding of modelling and detector uncertainties improves.

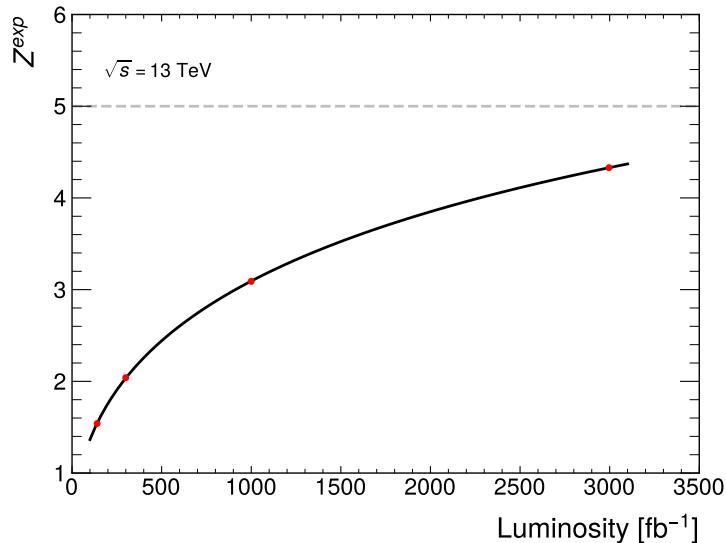


Figure 8.3: Expected significance associated with the measurement of  $\mu_{tWZ}$  for estimated toy datasets with different luminosities. A grey dashed line is included to represent the  $5\sigma$  value required to achieve signal observation.

The combined trilepton and tetralepton fitting procedure was applied to each of the estimated datasets, where the expected significance associated with each fit is shown in Figure 8.3. The function  $f(x, a, b, c) = \frac{a}{\sqrt{x+c}} + b$  was fitted to the data shown in Figure 8.3 to show the expected relationship between luminosity and expected significance. The expected significance plateaus at higher luminosities where the signal strength  $\mu_{tWZ}$  becomes more sensitive to systematic uncertainties.

# Chapter 9

## Conclusion

A search was conducted for the production of a single top quark in association with a  $W$  boson and a  $Z$  boson in the trilepton and tetralepton decay channels. The  $Z$  boson was required to decay into an electron-positron or a muon-antimuon pair. The trilepton decay channel requires one leptonically-decaying  $W$  boson and one hadronically-decaying  $W$  boson. The dominant backgrounds in the trilepton channel were the trilepton production of a top quark pair in association with a  $Z$  boson ( $t\bar{t}Z$ ) and the production of a fully leptonically-decaying  $WZ$  boson pair with an additional  $b$ -tagged jets. A graph neural network was applied to discriminate signal and background events in the trilepton channel, but only a marginal separation was achieved. The results of the profile likelihood fit using the Asimov trilepton dataset estimated a signal strength of  $\mu_{tWZ} = 1.00^{+0.72}_{-0.68}(\text{stat.})^{+0.83}_{-0.71}(\text{syst.})$ . The trilepton measurement had an expected significance of  $Z^{\text{exp}} = 1.02\sigma$ . The measurement was both statistically and systematically limited with the leading systematic uncertainties being the estimated cross-section normalisations for the major background processes.

The tetralapton decay channel requires the prompt  $W$  boson and the  $W$  boson from the top quark decay to decay leptonically. The major backgrounds were the tetralepton production of  $t\bar{t}Z$  and the  $ZZ$  diboson production with an additional  $b$ -tagged jet. A graph neural network was developed for the tetralepton channel to discriminate signal and background events. The tetralepton graph neural network succeeded in identifying diboson  $ZZ+\text{jets}$  events as background events. The results of the profile likelihood fit using the Asimov tetralepton dataset estimated a signal strength of  $\mu_{tWZ} = 1.00^{+0.95}_{-0.83}(\text{stat.})^{+0.51}_{-0.32}(\text{syst.})$ . The measurement in the tetralepton channel had an associated expected significance of  $Z^{\text{exp}} = 1.13\sigma$ . The measurement in the tetralepton channel was more statistically limited than the trilepton channel but was similarly sensitive to the signal strength of  $tWZ$ . The leading systematic uncertainties were found to be related to jet reconstruction and the precision of the estimated  $t\bar{t}Z$  cross section.

A combined fit to the Asimov datasets in the trilepton and tetralepton channels estimated a combined signal strength of  $\mu_{tWZ} = 1.00^{+0.56}_{-0.53}$ (stat.) $^{+0.54}_{-0.41}$ (syst.). The measurement had an expected significance of  $Z^{\text{exp}} = 1.34\sigma$  and therefore, an observation of  $tWZ$  production is not expected. The final extraction was shown to be sensitive to both the trilepton and tetralepton channels and combining these regions reduced the overall uncertainty. The leading systematic uncertainties were associated with the estimated  $t\bar{t}Z$  cross section, jet reconstruction and PDF calculations. High luminosity estimations suggest that an observation of  $tWZ$  production is not possible with a significant increase in the luminosity of high energy  $pp$  collisions. Both improved estimations of the systematic uncertainties and an improvement in the analysis method will be required for an observation to be possible.

# Bibliography

- [1] Benjamin Warren. “A search for tWZ production in the trilepton channel using Run 2 data from the ATLAS experiment”. MA thesis. 2021. URL: <http://hdl.handle.net/11427/36228>.
- [2] Jake Reich. “A search for tWZ production in the Full Run 2 ATLAS dataset using events with four leptons”. MA thesis. 2021. URL: <http://hdl.handle.net/11427/36050>.
- [3] Lyndon Evans and Philip Bryant. “LHC Machine”. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08001. doi: [10.1088/1748-0221/3/08/S08001](https://doi.org/10.1088/1748-0221/3/08/S08001). URL: <https://dx.doi.org/10.1088/1748-0221/3/08/S08001>.
- [4] ATLAS Collaboration. “The ATLAS Experiment at the CERN Large Hadron Collider”. In: *Journal of Instrumentation* 3.08 (Aug. 2008), S08003. doi: [10.1088/1748-0221/3/08/S08003](https://doi.org/10.1088/1748-0221/3/08/S08003). URL: <https://dx.doi.org/10.1088/1748-0221/3/08/S08003>.
- [5] M. Thomson. *Modern Particle Physics*. Modern Particle Physics. Cambridge University Press, 2013. ISBN: 9781107034266.
- [6] Antonio Masiero. “Dark Matter: the Particle Physics View”. In: *Dark Matter and Dark Energy: A Challenge for Modern Cosmology*. Ed. by Sabino Matarrese et al. Dordrecht: Springer Netherlands, 2011, pp. 273–293. ISBN: 978-90-481-8685-3. doi: [10.1007/978-90-481-8685-3\\_6](https://doi.org/10.1007/978-90-481-8685-3_6). URL: [https://doi.org/10.1007/978-90-481-8685-3\\_6](https://doi.org/10.1007/978-90-481-8685-3_6).
- [7] Particle Data Group. “Review of Particle Physics”. In: *Progress of Theoretical and Experimental Physics* 2022.8 (Aug. 2022). 083C01. ISSN: 2050-3911. doi: [10.1093/ptep/ptac097](https://doi.org/10.1093/ptep/ptac097). eprint: <https://academic.oup.com/ptep/article-pdf/2022/8/083C01/49175539/ptac097.pdf>. URL: <https://doi.org/10.1093/ptep/ptac097>.
- [8] ATLAS Collaboration. *Modelling of rare top quark processes at  $\sqrt{s} = 13$  TeV in ATLAS*. Tech. rep. Geneva: CERN, 2020. URL: <https://cds.cern.ch/record/2730584>.

- [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: *Journal of High Energy Physics* 2014.7 (July 2014). doi: [10.1007/jhep07\(2014\)079](https://doi.org/10.1007/jhep07(2014)079). URL: [https://doi.org/10.1007/JHEP07\(2014\)079](https://doi.org/10.1007/JHEP07(2014)079).
- [10] The CMS Collaboration. *Observation of four top quark production in proton-proton collisions at  $\sqrt{s} = 13$  TeV*. Tech. rep. Geneva: CERN, 2023. URL: <https://cds.cern.ch/record/2853304>.
- [11] ATLAS Collaboration. “Observation of four-top-quark production in the multilepton final state with the ATLAS detector”. In: (Mar. 2023). arXiv: [2303.15061 \[hep-ex\]](https://arxiv.org/abs/2303.15061).
- [12] Fabio Maltoni, Luca Mantani, and Ken Mimasu. “Top-quark electroweak interactions at high energy”. In: *JHEP* 10 (2019), p. 004. doi: [10.1007/JHEP10\(2019\)004](https://doi.org/10.1007/JHEP10(2019)004). arXiv: [1904.05637 \[hep-ph\]](https://arxiv.org/abs/1904.05637).
- [13] B. Grzadkowski et al. “Dimension-Six Terms in the Standard Model Lagrangian”. In: *JHEP* 10 (2010), p. 085. doi: [10.1007/JHEP10\(2010\)085](https://doi.org/10.1007/JHEP10(2010)085). arXiv: [1008.4884 \[hep-ph\]](https://arxiv.org/abs/1008.4884).
- [14] James Keaveney. “Constraining the SMEFT with a differential cross section measurement of  $tWZ$  production at the HL-LHC”. In: *Phys. Rev. D* 107 (3 Feb. 2023), p. 036021. doi: [10.1103/PhysRevD.107.036021](https://doi.org/10.1103/PhysRevD.107.036021). URL: <https://link.aps.org/doi/10.1103/PhysRevD.107.036021>.
- [15] ATLAS Collaboration. “Measurements of the inclusive and differential production cross sections of a top-quark–antiquark pair in association with a Z boson at  $\sqrt{s} = 13$  TeV with the ATLAS detector”. In: *Eur. Phys. J. C* 81.8 (2021), p. 737. doi: [10.1140/epjc/s10052-021-09439-4](https://doi.org/10.1140/epjc/s10052-021-09439-4). arXiv: [2103.12603 \[hep-ex\]](https://arxiv.org/abs/2103.12603).
- [16] Olga Bessidskaia Bylund. “Modelling Wt and tWZ production at NLO for ATLAS analyses”. In: *9th International Workshop on Top Quark Physics*. Nov. 2016. arXiv: [1612.00440 \[hep-ph\]](https://arxiv.org/abs/1612.00440).
- [17] *Studies on top-quark Monte Carlo modelling for Top2016*. Tech. rep. Geneva: CERN, 2016. URL: [http://cds.cern.ch/record/2216168](https://cds.cern.ch/record/2216168).
- [18] Federico Demartin et al. “tWH associated production at the LHC”. In: *The European Physical Journal C* 77.1 (Jan. 2017). doi: [10.1140/epjc/s10052-017-4601-7](https://doi.org/10.1140/epjc/s10052-017-4601-7). URL: <https://doi.org/10.1140/epjc/s10052-017-4601-7>.
- [19] Stefano Frixione et al. “Single-top hadroproduction in association with a W boson”. In: *Journal of High Energy Physics* 2008.07 (July 2008), pp. 029–029. doi: [10.1088/1126-6708/2008/07/029](https://doi.org/10.1088/1126-6708/2008/07/029). URL: <https://doi.org/10.1088/1126-6708/2008/07/029>.

- [20] V. Vapnik. “Principles of Risk Minimization for Learning Theory”. In: *Proceedings of the 4th International Conference on Neural Information Processing Systems*. NIPS’91. Denver, Colorado: Morgan Kaufmann Publishers Inc., 1991, pp. 831–838. ISBN: 1558602224.
- [21] HEP ML Community. *A Living Review of Machine Learning for Particle Physics*. URL: <https://iml-wg.github.io/HEPML-LivingReview/>.
- [22] Luana Ruiz, Fernando Gama, and Alejandro Ribeiro. *Graph Neural Networks: Architectures, Stability and Transferability*. 2020. doi: [10.48550/ARXIV.2008.01767](https://doi.org/10.48550/ARXIV.2008.01767). URL: <https://arxiv.org/abs/2008.01767>.
- [23] Peter W. Battaglia et al. *Relational inductive biases, deep learning, and graph networks*. 2018. arXiv: [1806.01261 \[cs.LG\]](https://arxiv.org/abs/1806.01261).
- [24] Justin Gilmer et al. *Neural Message Passing for Quantum Chemistry*. 2017. arXiv: [1704.01212 \[cs.LG\]](https://arxiv.org/abs/1704.01212).
- [25] Jonathan Shlomi, Peter Battaglia, and Jean-Roch Vlimant. “Graph neural networks in particle physics”. In: *Machine Learning: Science and Technology* 2 (2 Jan. 2021), p. 021001. ISSN: 2632-2153. doi: [10.1088/2632-2153/abbf9a](https://doi.org/10.1088/2632-2153/abbf9a). URL: <https://iopscience.iop.org/article/10.1088/2632-2153/abbf9a>.
- [26] Murat Abdughani et al. “Probing stop pair production at the LHC with graph neural networks”. In: *Journal of High Energy Physics* 2019.8 (Aug. 2019). doi: [10.1007/jhep08\(2019\)055](https://doi.org/10.1007/jhep08(2019)055). URL: [https://doi.org/10.1007/jhep08\(2019\)055](https://doi.org/10.1007/jhep08(2019)055).
- [27] Nicholas Choma et al. *Graph Neural Networks for IceCube Signal Classification*. 2018. arXiv: [1809.06166 \[cs.LG\]](https://arxiv.org/abs/1809.06166).
- [28] ATLAS Collaboration. *Graph Neural Network Jet Flavour Tagging with the ATLAS Detector*. Tech. rep. Geneva: CERN, 2022. URL: <https://cds.cern.ch/record/2811135>.
- [29] Huilin Qu and Loukas Gouskos. “Jet tagging via particle clouds”. In: *Physical Review D* 101.5 (Mar. 2020). doi: [10.1103/physrevd.101.056019](https://doi.org/10.1103/physrevd.101.056019). URL: <https://doi.org/10.1103/physrevd.101.056019>.
- [30] ATLAS Collaboration. *Point Cloud Deep Learning Methods for Pion Reconstruction in the ATLAS Experiment*. Tech. rep. Geneva: CERN, 2022. URL: <https://cds.cern.ch/record/2825379>.
- [31] Lukas Ehrke et al. *Topological Reconstruction of Particle Physics Processes using Graph Neural Networks*. 2023. arXiv: [2303.13937 \[hep-ph\]](https://arxiv.org/abs/2303.13937).

- [32] ATLAS Collaboration. *Luminosity determination in  $pp$  collisions at  $\sqrt{s} = 13$  TeV using the ATLAS detector at the LHC*. Tech. rep. Geneva: CERN, 2022. arXiv: [2212.09379](#). URL: <http://cds.cern.ch/record/2844887>.
- [33] ATLAS Collaboration. “The ATLAS Simulation Infrastructure”. In: *The European Physical Journal C* 70.3 (Sept. 2010), pp. 823–874. doi: [10.1140/epjc/s10052-010-1429-9](#). URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-010-1429-9>.
- [34] William Buttinger. *Using Event Weights to account for differences in Instantaneous Luminosity and Trigger Prescale in Monte Carlo and Data*. Tech. rep. Geneva: CERN, 2015. URL: <https://cds.cern.ch/record/2014726>.
- [35] R. Frederix et al. “The automation of next-to-leading order electroweak calculations”. In: *Journal of High Energy Physics* 2018.7 (July 2018). doi: [10.1007/jhep07\(2018\)185](#). URL: [https://doi.org/10.1007/jhep07\(2018\)185](https://doi.org/10.1007/jhep07(2018)185).
- [36] ATLAS Collaboration. *ATLAS Pythia 8 tunes to 7 TeV data*. Tech. rep. Geneva: CERN, 2014. URL: <https://cds.cern.ch/record/1966419>.
- [37] Paolo Nason. “A New method for combining NLO QCD with shower Monte Carlo algorithms”. In: *JHEP* 11 (2004), p. 040. doi: [10.1088/1126-6708/2004/11/040](#). arXiv: [hep-ph/0409146](#).
- [38] Simone Alioli et al. “A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX”. In: *JHEP* 06 (2010), p. 043. doi: [10.1007/JHEP06\(2010\)043](#). arXiv: [1002.2581 \[hep-ph\]](#).
- [39] 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 \[hep-ph\]](#).
- [40] ATLAS Collaboration. “Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data”. In: *Journal of Instrumentation* 14.12 (Dec. 2019), P12006–P12006. doi: [10.1088/1748-0221/14/12/p12006](#). URL: <https://doi.org/10.1088/1748-0221/14/12/p12006>.
- [41] ATLAS Collaboration. “Muon reconstruction and identification efficiency in ATLAS using the full Run 2  $pp$  collision data set at  $\sqrt{s} = 13$  TeV”. In: *Eur. Phys. J. C* 81.7 (2021), p. 578. doi: [10.1140/epjc/s10052-021-09233-2](#). arXiv: [2012.00578 \[hep-ex\]](#).

- [42] ATLAS Collaboration. “Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton–proton collision data at  $\sqrt{s} = 13\text{TeV}$ ”. In: *The European Physical Journal C* 79.8 (Aug. 2019). doi: [10.1140/epjc/s10052-019-7140-6](https://doi.org/10.1140/epjc/s10052-019-7140-6). URL: <https://doi.org/10.1140/epjc/s10052-019-7140-6>.
- [43] ATLAS Collaboration. “Electron and photon performance measurements with the ATLAS detector using the 2015–2017 LHC proton-proton collision data”. In: *Journal of Instrumentation* 14.12 (Dec. 2019), P12006–P12006. doi: [10.1088/1748-0221/14/12/p12006](https://doi.org/10.1088/1748-0221/14/12/p12006). URL: <https://doi.org/10.1088/1748-0221/14/12/p12006>.
- [44] ATLAS Collaboration. “Muon reconstruction and identification efficiency in ATLAS using the full Run 2 pp collision data set at  $\sqrt{s} = 14\text{ TeV}$ ”. In: *The European Physical Journal C* 81.7 (July 2021), p. 578. ISSN: 1434-6052. doi: [10.1140/epjc/s10052-021-09233-2](https://doi.org/10.1140/epjc/s10052-021-09233-2). URL: <https://doi.org/10.1140/epjc/s10052-021-09233-2>.
- [45] M.D. Schwartz. *Quantum Field Theory and the Standard Model*. Quantum Field Theory and the Standard Model. Cambridge University Press, 2014. ISBN: 9781107034730.
- [46] Matteo Cacciari, Gavin P Salam, and Gregory Soyez. “The anti- $k_t$  jet clustering algorithm”. In: *Journal of High Energy Physics* 2008.04 (Apr. 2008), pp. 063–063. doi: [10.1088/1126-6708/2008/04/063](https://doi.org/10.1088/1126-6708/2008/04/063). URL: <https://doi.org/10.1088/1126-6708/2008/04/063>.
- [47] ATLAS Collaboration. *Tagging and suppression of pileup jets with the ATLAS detector*. Tech. rep. Geneva: CERN, 2014. URL: <https://cds.cern.ch/record/1700870>.
- [48] 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: *The European Physical Journal C* 79.11 (Nov. 2019). doi: [10.1140/epjc/s10052-019-7450-8](https://doi.org/10.1140/epjc/s10052-019-7450-8). URL: <https://doi.org/10.1140/epjc/s10052-019-7450-8>.
- [49] ATLAS Collaboration. “ATLAS flavour-tagging algorithms for the LHC Run 2  $pp$  collision dataset”. In: (Nov. 2022). arXiv: [2211.16345 \[physics.data-an\]](https://arxiv.org/abs/2211.16345).
- [50] ATLAS Collaboration. “Jet energy scale and resolution measured in proton–proton collisions at  $\sqrt{s} = 13\text{ TeV}$  with the ATLAS detector”. In: *Eur. Phys. J. C* 81.8 (2021), p. 689. doi: [10.1140/epjc/s10052-021-09402-3](https://doi.org/10.1140/epjc/s10052-021-09402-3). arXiv: [2007.02645 \[hep-ex\]](https://arxiv.org/abs/2007.02645).

- [51] ATLAS Collaboration. “Performance of missing transverse momentum reconstruction with the ATLAS detector using proton-proton collisions at  $\sqrt{s} = 13$  TeV”. In: *Eur. Phys. J. C* 78.11 (2018), p. 903. doi: [10.1140/epjc/s10052-018-6288-9](https://doi.org/10.1140/epjc/s10052-018-6288-9). arXiv: [1802.08168 \[hep-ex\]](https://arxiv.org/abs/1802.08168).
- [52] ATLAS Collaboration. *Luminosity determination in pp collisions at  $\sqrt{s} = 13$  TeV using the ATLAS detector at the LHC*. Tech. rep. Geneva: CERN, 2019. URL: <https://cds.cern.ch/record/2677054>.
- [53] Jon Butterworth et al. “PDF4LHC recommendations for LHC Run II”. In: *Journal of Physics G: Nuclear and Particle Physics* 43.2 (Jan. 2016), p. 023001. doi: [10.1088/0954-3899/43/2/023001](https://doi.org/10.1088/0954-3899/43/2/023001). URL: <https://doi.org/10.1088/0954-3899/43/2/023001>.
- [54] ATLAS Collaboration. “Inclusive and differential cross-section measurements of  $t\bar{t}Z$  production in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector, including EFT and spin-correlation interpretations”. In: (2023). arXiv: [2312.04450](https://arxiv.org/abs/2312.04450). URL: <https://cds.cern.ch/record/2883457>.
- [55] ATLAS Collaboration. “Observation of the associated production of a top quark and a Z boson in  $pp$  collisions at  $\sqrt{s} = 13$  TeV with the ATLAS detector”. In: *JHEP* 07 (2020), p. 124. doi: [10.1007/JHEP07\(2020\)124](https://doi.org/10.1007/JHEP07(2020)124). arXiv: [2002.07546 \[hep-ex\]](https://arxiv.org/abs/2002.07546).
- [56] Kyle Cranmer et al. *HistFactory: A tool for creating statistical models for use with RooFit and RooStats*. Tech. rep. New York: New York U., 2012. URL: <https://cds.cern.ch/record/1456844>.
- [57] A. Hoecker et al. *TMVA - Toolkit for Multivariate Data Analysis*. 2007. doi: [10.48550/ARXIV.PHYSICS/0703039](https://doi.org/10.48550/ARXIV.PHYSICS/0703039). URL: <https://arxiv.org/abs/physics/0703039>.
- [58] Kyle Cranmer. “Practical Statistics for the LHC”. In: (2015). doi: [10.48550/ARXIV.1503.07622](https://doi.org/10.48550/ARXIV.1503.07622). URL: <https://arxiv.org/abs/1503.07622>.
- [59] G. Cowan. *Statistical Data Analysis*. Oxford science publications. Clarendon Press, 1998. ISBN: 9780198501558.
- [60] A L Read. “Presentation of search results: the CLs technique”. In: *Journal of Physics G: Nuclear and Particle Physics* 28.10 (Sept. 2002), p. 2693. doi: [10.1088/0954-3899/28/10/313](https://doi.org/10.1088/0954-3899/28/10/313). URL: <https://dx.doi.org/10.1088/0954-3899/28/10/313>.

- [61] Glen Cowan et al. “Asymptotic formulae for likelihood-based tests of new physics”. In: *The European Physical Journal C* 71.2 (Feb. 2011). doi: [10.1140/epjc/s10052-011-1554-0](https://doi.org/10.1140/epjc/s10052-011-1554-0). URL: <https://doi.org/10.1140%2Fepjc%2Fs10052-011-1554-0>.
- [62] ATLAS Collaboration. “Performance of the ATLAS muon triggers in Run 2”. In: *Journal of Instrumentation* 15.09 (Sept. 2020), P09015–P09015. doi: [10.1088/1748-0221/15/09/p09015](https://doi.org/10.1088/1748-0221/15/09/p09015). URL: <https://doi.org/10.1088/1748-0221/15/09/p09015>.
- [63] ATLAS Collaboration. “Performance of electron and photon triggers in ATLAS during LHC Run 2”. In: *The European Physical Journal C* 80.1 (Jan. 2020). doi: [10.1140/epjc/s10052-019-7500-2](https://doi.org/10.1140/epjc/s10052-019-7500-2). URL: <https://doi.org/10.1140/epjc/s10052-019-7500-2>.
- [64] Diederik P. Kingma and Jimmy Ba. *Adam: A Method for Stochastic Optimization*. 2014. doi: [10.48550/ARXIV.1412.6980](https://arxiv.org/abs/1412.6980). URL: <https://arxiv.org/abs/1412.6980>.
- [65] Thomas Philippe Calvet. “Search for the production of a Higgs boson in association with top quarks and decaying into a b-quark pair and b-jet identification with the ATLAS experiment at LHC”. Section 5.3.1. 2017. URL: <https://cds.cern.ch/record/2296985>.