Non-prompt Lepton Background Estimation for Associated Production of a Top Quark and a Z Boson in pp Collisions at $\sqrt{s} = 13$ TeV at ATLAS

Marius Blaut

Masterarbeit in Physik
angefertigt im Physikalischen Institut

vorgelegt der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität
Bonn

Oktober 2018

I hereby declare that this thesis was formulated by myself and that no sources or tools other than those cited were used.

Bonn, ...

Date Signature

1. Gutachter: Prof. Dr. Ian Brock
2. Gutachter: Prof. Dr. Jochen Dingfelder

Contents

CHAPTER 1

Introduction

To answer fundamental questions about the universe, physicists make an enormous effort to find and understand the particles the visible world is made of and the interactions between them. The most popular theory describing nature we see in the field of particle physics is the Standard Model (SM). The reason for the success of the SM is the fact that it offers answers regarding what elementary particles are. With the discovery of the Higgs boson in 2012 at the Large Hadron Collider (LHC), all particles which have been predicted by the SM have been experimentally detected. Due to some open questions this theory is not able to give an answer to, the investigation beyond the SM is necessary: neutrino masses are not generated, gravitation is not included and dark matter is not predicted.

By letting particles collide with each other at very high energies, new particles of the SM can be produced and studied. Postulated in 1973 by Kobayashi and Maskawa and experimentally found in 1995, the top quark with its very large mass is one of the most interesting particles to study. The determination of properties of this short lived particle is important since top quark mass measurements provide constraints on the mass of the Higgs boson, which represents a self-consistency test for the SM.

The production of top quarks at hadron colliders takes place either via the strong interaction, in top-antitop quark pairs or via the weak interaction, which is usually called single top quark production. The three possible channels for the single top quark production are the s-channel, t-channel and tW channel, in which along with the top quark an on-shell W boson is produced as well.

A single top quark produced in association with the a Z boson is a rare process for which strong evidence has recently been found. The dominant mechanism for this production channel is the t-channel. It is a process which is sensitive to the coupling of the top quark and the Z boson (tZ coupling) and the WWZ coupling. The production of a single top quark in association with a Z boson (tZq) is very interesting and important since it is a rare process predicted by the SM. In addition, the production mechanism is used for testing SM predictions, in the sense that this production is background for important processes such as tH and in the search for tZ production via flavour-changing neutral currents.

For the SM tZq process different final states are defined depending on the decay products of the Z boson and the top quark. The analysis presented in this thesis considers events with three charged leptons (electrons and/or muons) and two jets, one of which is identified as a b quark jet (trilepton final state). Although this channel has the smallest branching ratio, the presence of three charged leptons reduces the background. The remaining channels, in which at least one of the particles

decays hadronically, have larger branching ratios but also an order of magnitude higher background contribution.

Background events originate from SM processes with similar final states as tZq. Such processes are diboson, particularly WZ or ZZ, or the associated production of $t\bar{t}$ with a W, Z or H boson as well as, Z+jets and $t\bar{t}$ with an additional non-prompt lepton. Non-prompt leptons typically originate from heavy hadron decays. To estimate the contribution of the backgrounds, Monte Carlo (MC) simulations and for the non-prompt backgrounds data-driven techniques are used. This thesis investigates and quantifies the improvement that can be achieved by using a new method based on MC simulations to estimate the background for the tZq production by looking at $79.8\,\text{fb}^{-1}$ of data collected by the ATLAS detector in 2015, 2016 and 2017. Since the cross section for tZq production is very small a large amount of data is necessary in order to study this production mechanism.

The new non-prompt background estimation method presented in this thesis turns dilepton events into trilepton events by generating a lepton and solves the problem of low statistics by increasing the raw number of generated Z+jets and $t\bar{t}$ background events in the signal region. The method is called the *b-jet replacement method* because the generated lepton replaces a b quark jet.

In order to understand the theoretical and experimental basis of this analysis chapter 2 summarizes the Standard Model of particle physics, as well as introduces main properties of the top quark. Chapter 3 is dedicated to the ATLAS detector which collects the data used in this analysis. In chapter 4 the analysis strategy is presented and the description of the samples that were used. The needed selection in order to consider events with a final state corresponding to that of tZq production is discussed in detail. Chapter 5 focuses on the *b-jet replacement method* and the comparison between its results and that from data-driven techniques. In the end, chapter 6 summarises the results of the performed analysis.

CHAPTER 2

Theory

Very high energies are necessary to resolve the internal structure of particles and probe small scales. The Standard Model of elementary particles describes our knowledge about the fundamental constituents of matter and the interactions between them. The term "elementary" refers to objects for which no internal structure has been found yet and are point-like. Discoveries of particles from the SM such as the top quark in 1995 or the Higgs boson in 2012 show how effective this theory is at modeling the universe. Looking at figure 2.1 one can see the agreement of the measured cross-section with the theoretical expectations published by the ATLAS collaboration which shows the quality of the SM theory.

In this chapter the Standard Model, the top quark and its production mechanism is briefly presented.

2.1 Standard Model

2.1.1 Particles

The Standard Model of particle physics consists of 30 constituents building up the current understanding on the structure of matter and how they interact. They are characterized by their properties such as mass, spin, charge and their way of interaction. There are two types of elementary particles in the SM with different properties: fermions, which form matter (matter particles) and bosons, which are are the mediators of force (force carrier particles). Those particles predicted by the SM with their properties are shown in figure 2.2.

Matter particles

Particles with half-integer spin are called fermions. They are divided into even more subgroups called leptons and quarks. Each of these subclasses has 6 particles and they are arranged into three generations (families).

Leptons are further subdivided into charged and neutral. The electron is the lightest and best known charged lepton, discovered in 1897. Its electric charge of 1.602×10^{-19} C is used as standard unit of electric charge **e**. The muon and tau are more massive leptons with the same charge as the electron corresponding to generation two and three, respectively. Because of their masses much higher than that of the electron, their lifetime is small, especially the one of the tau.

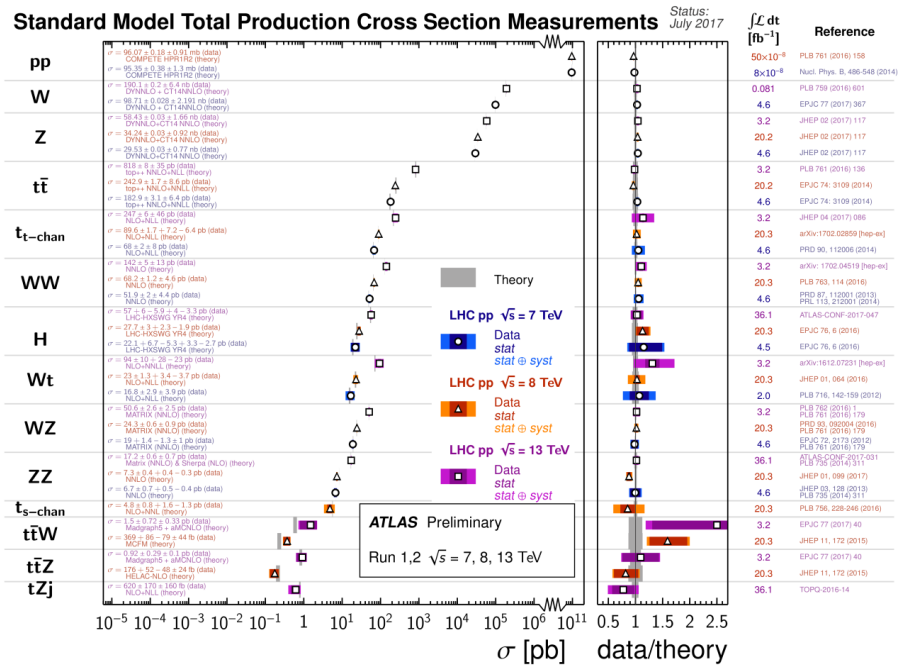


Figure 2.1: Cross-sections for different processes measured by the ATLAS collaboration and the corresponding theoretical prediction. The right panel shows the ratio of data to theory, the corresponding integrated luminosity and the reference to the measurement [1].

For every charged lepton there is a neutral counterpart called neutrino which has the same lepton family number as its charged partner. Neutrinos have been postulated by Pauli after studying the energy spectrum of β-decay. Although neutrinos are massless in the SM, neutrino oscillation has been observed, which is only possible if neutrinos have a mass.

The second subgroup of fermions, the so-called quarks, are: up, down, charm, strange, top and bottom. They are, unlike leptons, all electrically charged and organised in pairs containing two quarks with a fractional charge of the standard unit. One of these quarks is an up-type quark with a charge of $\frac{2}{3}e$ and the other a down-type quark with a charge of $-\frac{1}{3}e$. Up-type quarks are up, charm and top and down-type quarks are down, strange and bottom, listed in increasing mass. In comparison to leptons, quarks are not able to exist in a free state. They only exist in bound states called hadrons (with exception of the top quark, discussed in Section 2.2.1). This behavior is called confinement and comes from the conservation af a quantum number called colour charge. Each quark has the colour red, green or blue.

For each of these particles there is a corresponding antiparticle which has the opposite sign of charge and flavour numbers. For quarks the antiparticle, the anti quark, carries an anti-colour (anti-red, anti-green, anti-blue). The hadronic bound state formed by quarks has to be colour neutral and therefore only combinations of the type $q\bar{q}$ (mesons) and qqq (baryons) are allowed to exist.

2.1.2 Particle interactions

Bosons are particles which have an integer spin. In the SM they are associated to an interaction with exception of the Higgs boson. In order to interact with each other, particles need force carriers which

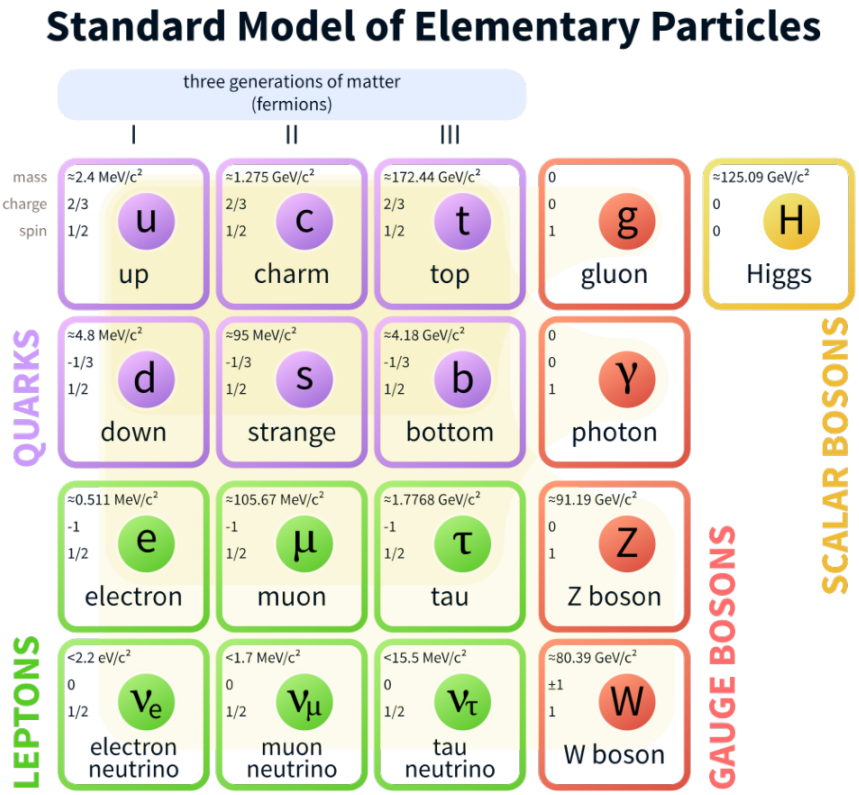


Figure 2.2: Schematic representation of the Standard Model of particle physics. Each particle is shown along with its properties like spin, mass and electric charge [2].

enable particles interactions. Each of the three fundamental forces included in the SM has one or more force carriers and acts at distinct ranges. The properties and association to an interaction of each boson is described in more details.

The SM includes three out of the four forces of nature. All interactions that occur in nature can be explained with four fundamental forces: strong, electromagnetic, weak and gravity. Each of the forces has a different strength and scale it is acting on. Gravity is not included in the SM. The effect of the gravitational force, even though its range is infinite, can be neglected due to the fact that at the energy scale of particle physics and the SM its effect is very small. The mediation of each force is described by a quantum field theory (QFT).

The electromagnetic interaction has the photon as mediator. It interacts with all particles which are electrically charged. This boson is massless, electrically neutral and acts attractive or repulsive. Those interactions are described by quantum electrodynamics (QED).

The weak force is mediated by the $W^{\pm}$ and Z bosons which are the only gauge bosons not being massless. Because of their relatively high masses of around 80 to 90 GeV and hence their very short lifetimes, the range of this force is short. The $W^{\pm}$ bosons are charged while the Z boson is neutral. The decay in which a W boson is involved causes always a flavour change whereas the Z boson leaves

the flavour unchanged. In this sense the weak force is neither acting attractive nor repulsive but rather seen as a force enabling decays. The electromagnetic and weak interaction are presented together in the SM and unified in the electroweak theory.

The strong force is mediated by gluons which do not carry any electric charge but a colour charge and hence have the ability to self-interact. They are responsible for keeping quarks in bound states. As mentioned earlier, bound quark states must be colour neutral consisting of two quarks of opposite colour charge or three quarks of all different colours. The dynamics related to these gauge bosons are described in quantum chromodynamics (QCD).

The last boson of the SM and recently discovered is the Higgs boson. This boson being the only one not having any spin, is electrically and colour neutral. The Higgs field plays a fundamental role in the SM since its mechanism explains why particles of the SM have a mass. Breaking the $SU(2) \times U(1)$ symmetry by requiring masses for the elementary particles can be understood with the Higgs mechanism. It explains the generation of masses for the fermions and the $W^{\pm}$ and Z bosons.

2.2 Top quark physics

2.2.1 Top quark properties

Postulated in 1973, a long time before its discovery in 1995, the top quark was the missing piece of puzzle to explain CP violation in kaon decays after the existence of a third generation of leptons and quarks was widely accepted. The top quark with the heaviest mass of all fermions in the SM, and even heavier than the bosons, cannot hadronise as it decays rapidly and mostly into a W and b quark. Since the lifetime of the top quark is in the order of 0.5×10^{-24} s which is smaller than the time necessary for the hadronization process, this inability of creating a bound state with other quarks, makes it possible to investigate its properties in more detail and with great precision.

The top quark forms together with its down-type partner, the bottom quark, a weak isospin doublet. Its electrical charge is $+\frac{2}{3}e$ and has a spin of $\frac{1}{2}$. With a mass of $(173.34 \pm 0.76)\,\text{GeV}$ [3], it is the heaviest elementary particle of the SM and has because of that a strong coupling to the Higgs boson which adds important constraints to.

2.2.2 Production mechanism

The top quark is produced in hadron collisions either via strong interaction where it exists in form of a pair of top-antitop quarks or via weak interaction involving the Wtb vertex in single top quark production. In order to illustrate, physicists developed a graphical representation of the mathematical description of those interactions: the so-called *Feynman diagrams*. In this representation the depiction of fermions and bosons differs. For fermions a solid line with an arrow pointing with time for particles is used and against time for antiparticles. Depending on the convention, for bosons either a wavy or dashed line is used with exception of gluons where a spring-like line shows their involvement.

Top quark pair production

The production of top quarks through strong interaction is the mechanism in which the creation of a top and anti-top quark is involved. This can be done via gluon-fusion ($gg \rightarrow t\bar{t}$) or quark annihilation ($q\bar{q} \rightarrow t\bar{t}$). Both processes are shown in figure 2.3. For the process in which gluons

fuse and generate a top quark pair, the diagrams for the *s*- and *t*-channel are shown. These are diagrams for processes at *leading order* (LO) in QCD where no higher order corrections are applied. In contrast to proton-antiproton colliders like the Tevatron, in pp collisions at the LHC the gluon induced process is the dominant one. For a centre-of-mass energy of 13 TeV roughly 90% of all $t\bar{t}$ events are produced by this process. At this centre-of-mass energy theory predicts a top quark pair cross section of $\sigma_{t\bar{t}} = 832\,\text{pb}$ which includes next-to-next-to-leading order (NNLO) and assumes the top quark mass to be $m_t = 172.5\,\text{GeV}$.

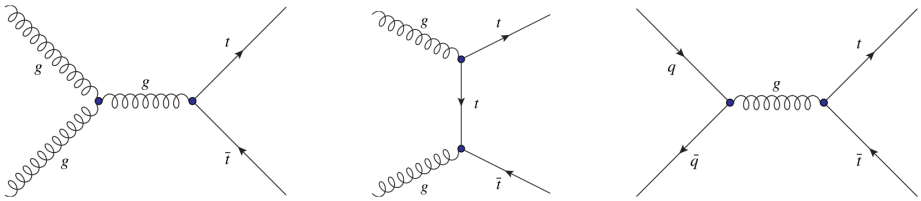


Figure 2.3: Feynman diagrams of $t\bar{t}$ production at LO QCD [4].

Single top quark production

The three different processes in which the top quark is produced through weak interaction are shown in figure 2.4. These single top quark production processes at leading order involve a Wbt vertex where in two of them, namely in the *t*- and *s*-channel, a virtual W boson is produced.

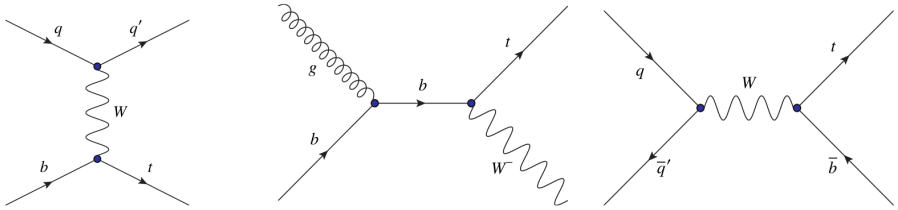


Figure 2.4: Feynman diagrams of single top quark production at LO. From left to right: *t*-, tW- and *s*-channel [4].

The *t*-channel is the dominant one accounting for approximately 70% of the total electroweak production. In this production mode the top quark is produced through the interaction of a b quark and a virtual W boson. At the LHC it is more likely to produce top quarks than antitop quarks. The theoretical predictions up to NNLO have been calculated.

The *s*-channel describes the production of the single top quark through the exchange of a virtual W boson as well. These events constitute a small fraction of all events in which a single top quark is produced. The measurement of this channel is very difficult due to the very low cross section.

The third channel of single top quark production is the tW channel in which the top quark is produced in association with a real W boson. The b quark, after an interaction with a gluon, emits a W boson whereby the b quark changes into its up-type partner, the top quark. It is a process with the second largest cross-section and the only one involving a real W boson, unlike the *t*- or *s*-channel. At NLO diagrams in which a top quark from an internal line can become on-shell contribute to the tW channel resulting in the same final state as LO $t\bar{t}$ production.

The fact that the addition of all single top quark cross-sections is not even half of the $t\bar{t}$ cross-section shows the difficulty of measuring the production of single top quark having $t\bar{t}$ as one of the main sources of backgrounds. The cross-section measurements of all single top quark processes are shown in figure 2.5. For different centre-of-mass energies, the measured values published by the ATLAS and CMS collaborations are compared to the calculated predictions. The comparison of the SM expectations and the measured cross-sections shows satisfactory results.

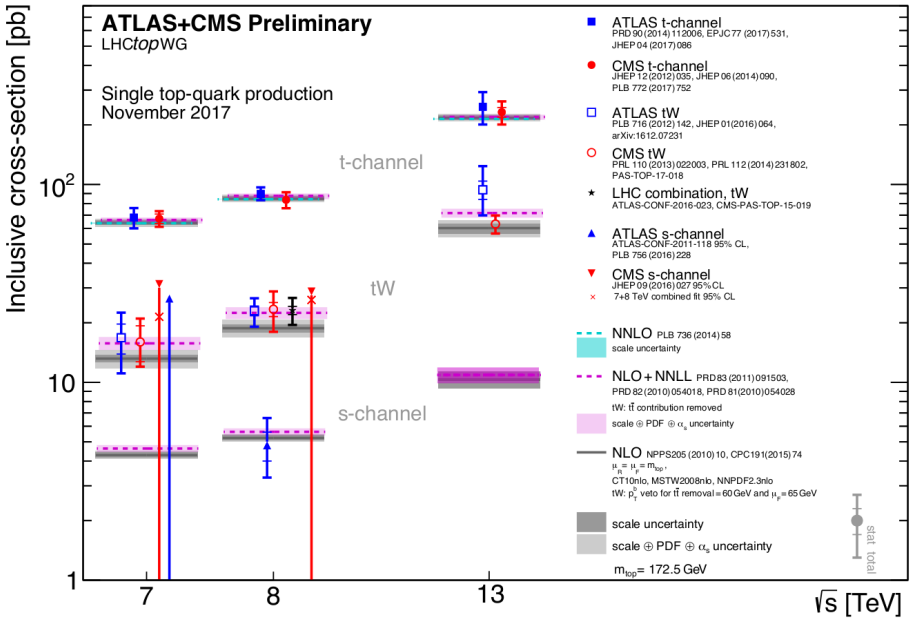


Figure 2.5: Single top quark cross-section measurements as a function of the centre-of-mass enery compared to the theoretical predictions published by ATLAS and CMS [5].

2.2.3 Rare processes involving top quarks

Higher luminosity and collision energy at the LHC enable the investigation of processes with very low cross-sections, many of them involving top quarks. The predicted NLO cross-sections at 13 TeV for the production of a top quark pair in association with a boson (Z,W or even H) are more than one order of magnitude lower than the cross-section for the production of a single top quark in the s-channel, belonging already to processes with a high difficulty of measuring. Production of single top quarks in association with a Z or H boson have even lower cross-sections. These processes can be investigated since the total luminosity collected by the ATLAS collaboration reached a value of 79.8 fb^{-1}. A direct way to investigate the top-Yukawa coupling offer $t\bar{t}H$ and tHq events.

The subject of this thesis is the tZq production which can be seen as a forerunner of a tHq production analysis. The tZq process probes different SM couplings suitable for testing SM predictions. This process will be introduced and discussed in more detail in section 4.1.

CHAPTER 3

Experimental setup

In this chapter the general overview of the LHC [6] and the ATLAS [7] detector which collected the dataset used in this analysis is presented.

3.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is a proton-proton accelerator near Geneva at the French-Swiss border. It provides the large amount of data needed to study very rare processes. The LHC is a circular

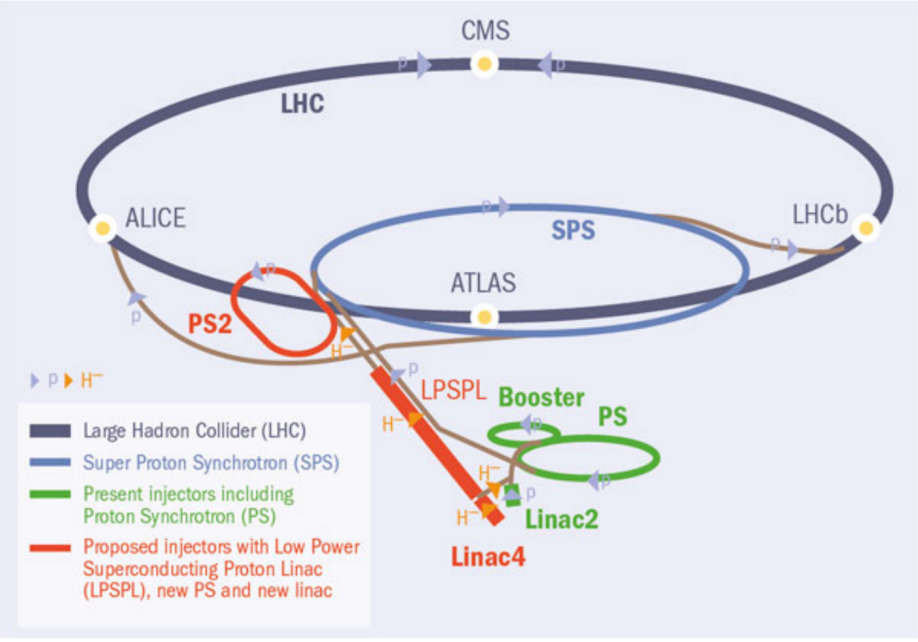


Figure 3.1: Schematic of the LHC and its preaccelerators [8].

accelerator located in a tunnel with a circumference of nearly 27 km length and located 100 m below ground. Collisions of protons at energies up to 13 TeV are reached by accelerating two proton beams in opposite direction with an energy of 6.5 TeV per beam. In order to achieve a collision energy of a

value never reached by any particle collider before, some smaller accelerators are needed to increase the energy of the particles before loading them into the main ring. After being first accelerated by a linear accelerator (LINAC2), the protons are injected with an energy of 50 MeV into the Proton Synchrotron Booster (PSB) which feeds 1.4 GeV protons to the Proton Synchrotron (PS). With the last accelerator before entering the LHC, the protons energy is increased by the Super Proton Synchrotron from 26 GeV to 450 GeV. Once this threshold is passed, the protons are injected to the main ring in bunches to finally be accelerated to the target energy of 6.5 TeV. Figure 3.1 illustrates the different components in the LHC accelerator chain.

Protons are carried by vacuum tubes in which the particles are spatially separated to prevent collisions outside of the detector. Guided by dipole magnets and focused by quadrupole magnets they follow the circular orbit while they keep being accelerated by accelerating cavities and electromagnetic resonators. A complex cooling system using liquid helium is used to ensure very low temperatures. After an acceleration time of 25 minutes the target energy is reached and the proton beams circulate in the accelerator for about ten hours. To ensure the two beams rotating in opposite directions are spatially separated and hence no collisions take place, the complex adjustment must not allow any intersection outside of the four large detectors which are the only interaction points of the beams.

The LHC data-taking period which started in 2015 and is still currently ongoing is called Run 2. In this time data of 125 fb^{-1} is expected to be collected. The dataset used in this analysis was collected by the ATLAS detector from 2015 until 2017 consisting of 79.8 fb^{-1} of collision data.

Along the LHC tunnel the four detectors are located at which the interaction points are situated. The location of these experiments along the ring are shown in figure 3.2. ATLAS and CMS [9], the two major general purpose detectors, are designed to study topics related to elementary particle physics and perform precision SM tests. The LHCb [10] studies flavour physics focusing on b physics and ALICE [11] seeks to understand interactions from heavy ion collisions. In the next section the ATLAS detector will be introduced in more detail.

3.2 The ATLAS detector

The name of the largest detector, ATLAS, stands for “A Toroidal LHC Apparatus”. The detectors structure is typical for colliding beam experiments, which has all its detector components surrounding the beam pipe. To provide an almost full azimuth angle coverage, the components are symmetrically located around the interaction point. The ATLAS detector has a diameter of 25 m, is 46 m long and a weight of 7 000 t. Figure 3.3 shows each of these detector components. Closest to the beam pipe is the Inner Detector (ID), responsible for recording and reconstructing particle tracks. Followed by the electromagnetic and hadron calorimeters (ECAL and HCAL) measuring the energy of the particles passing the detector, the Muon Spectrometer (MS) and magnet system. The following subsection describes the purpose and sensitivity of each component.

Inner detector (ID)

The inner detector has the function to track charged particles passing through the detector. Therefore the momentum, direction and electric charge of the particle has to be measured. In order to achieve this, a magnetic field surrounds the inner detector causing charged particles to have a bent trajectory. The information extracted in the ID can be used to associate particles with a vertex. All subdetectors

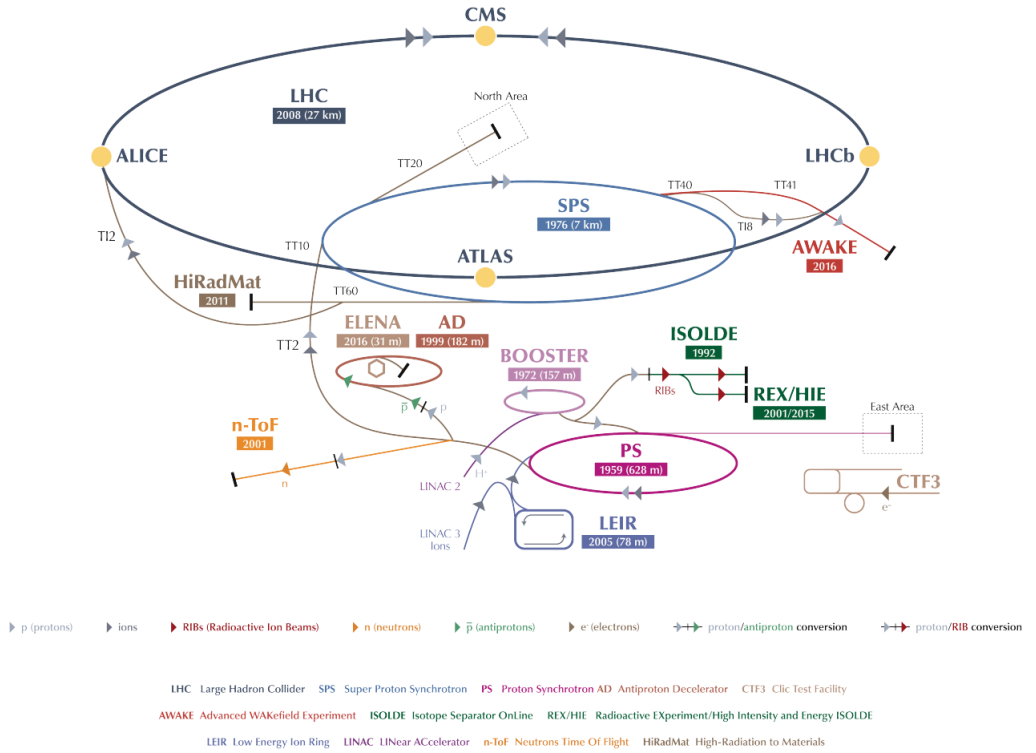


Figure 3.2: The LHC and the four detectors [12].

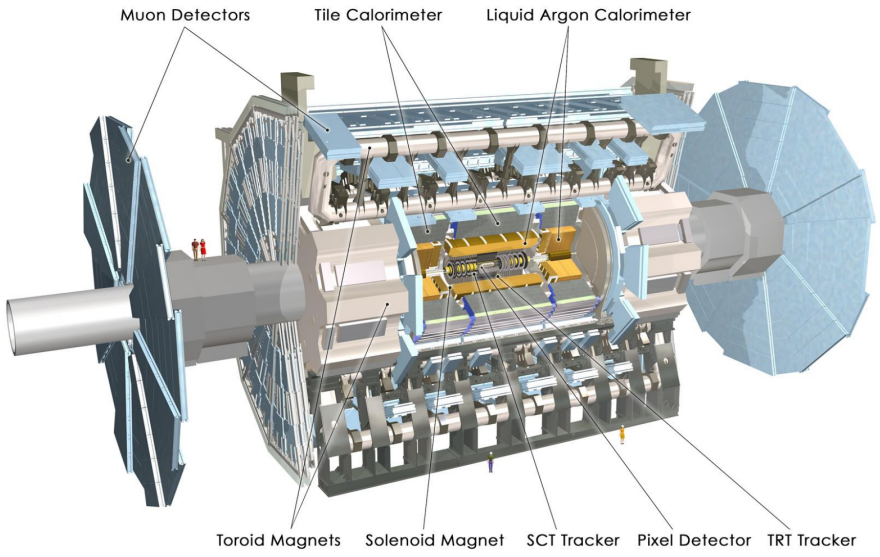


Figure 3.3: A detailed overview of the ATLAS detector with its components [13].

have a similar common structure: the central part of the detector is covered by a cylindrical component parallel to the beam line and a so-called endcap on each side which registers particles travelling at high η. The ID is composed of a pixel detector, semiconducter tracker (SCT) and a transition radiation tracker (TRT), covering a range of pseudorapidity $|\eta| \leq 2.5$.

Pixel detector

Pixel detectors are used in the innermost layers due to their quality offering high accuracy in identifying signals in a radiation flooded region as they are so close to the interaction points, used in the innermost layers of the detector. When a charged particle passes through, it creates an electron-hole pair by ionising the semiconducter medium. When charges accumulate at this position, a current is produced generating a signal in the electronics. The ATLAS pixel detector uses silicon as semiconducter medium.

The pixel detector contains four radial layers and three endcap disks with the purpose of covering the forward region. Initially only three radial layers have been used but during the shutdown in 2015 an additional layer has been installed, the Insertable B-layer (IBL), located only a few centimeters from the beam axis in order to further improve track reconstruction during Run 2 operations.

Consisting of rectangular shaped modules, each layer provides 1834 of them. These modules have an approximate size of $6 \times 2\text{cm}^2$ containing about 46 000 pixels. Having a size of $50 \times 400\mu\text{m}^2$ each and being read out independently, these pixels amount to 92 million readout channels.

By passing one of these layers, a signal is registered and the position of the particle is determined. These spatially distributed points are used to reconstruct the trajectory of a passing particle and set the position where the interaction started. The spatial resolution has a value of 8 µm in $R - \phi$ and 75 µm in z.

Semi-conducter tracker (SCT)

The SCT consists of four cylindrical barrel layers and 18 planar endcap disks, having a similar function as the pixel detector. Silicon strips placed every 80 µm on 60 m^2 of material enable the readout covering a larger area than the previous detector. The precision of the measurement of the charged particles position is done with an accuracy of up to 17 µm, in the transverse direction of the strips. A precision of 580 µm is achieved for the z resolution for the barrel and for the R resolution for the endcap region.

Transition radiation tracker (TRT)

Having a volume of 12 m^3, the TRT makes up the largest component of the inner detector. Its detector components are small drift chambers with a diameter of 4 mm called straw tubes having a wire made of gold-plated tungsten measuring 0.03 mm in diameter. The TRT has about 300 000 straws in total, consisting of 50 000 straws in the barrel part and of 250 000 straws in the endcap region. The length of the straws varies depending on the region they are placed in (144 cm in central region and 39 cm in endcap region). In a straw filled with a gas mixture, each particle passing it causes ionisation. The applied voltage accelerates the primary electrons towards the anode which is the wire. On its way towards the wire, the electron ionises other atoms creating an avalanche. The movement of all charge carriers produces a signal measured as current flow. Each straw detects this signal individually.

By construction, no information about the z component is obtained, since the tubes are aligned with the beam axis. Only information in the $R - \phi$ plane is provided with an uncertainty of 130 µm per straw.

Calorimeters

Outside of the solenoid magnet which surrounds the inner detector, the calorimeters are located. In contrast to the inner detector, this section of the ATLAS detector is meant to absorb particles passing through the detector in order to measure their deposited energy and contribute to particle identification. The system is divided into two calorimeters: the electromagnetic calorimeter (ECAL) for measurements related to electrons and photons, followed by the hadron calorimeter (HCAL) responsible for measuring the energies of hadronic shower. Unlike for the ID, a much greater range in pseudorapidity ($|\eta| < 4.9$) is covered by the calorimeters.

The electromagnetic calorimeter (ECAL)

The electromagnetic calorimeter is divided into three parts covering different η ranges. In the range of $|\eta| < 1.475$ the barrel covers the central region while the endcaps cover $1.375 < |\eta| < 3.2$ having two coaxial wheels each.

Highly energetic electrons lose energy by radiating a photon in the electromagnetic field of the nuclei (bremsstrahlung) when passing through matter, whereas photons produce an electron-positron pair provided if $E_\gamma > 1\,\text{MeV}$. Electrons coming from the electron-positron production radiate a photon and the photon is followed again by the production of electron pairs until a certain threshold is reached where the electrons energy is so low that the loss through ionisation and bremsstrahlung becomes equal. The radiation length X_0 is the average distance travelled by the particle until its energy decreases by $\frac{1}{e}$.

The barrel's energy absorbing material is lead and has liquid argon as the sampling material, located in the central region. The thickness of the barrel is approximately $22X_0$, while the endcaps have a thickness of $26X_0$.

To improve the precision required in the inner detector region ($|\eta| < 2.5$), the ECAL is divided into three sections. Having a higher granularity for layers located closest to the centre of the detector, the first layer is composed of strips with $\Delta\eta \times \Delta\phi = 0.003 \times 0.1$. The endcaps have a worse granularity but still sufficient to reconstruct jets.

The hadron calorimeter (HCAL)

All particles that pass through the ECAL enter the HCAL which is intended to measure the energy of particles losing energy via ionisation or strong interaction with nuclei. A cascade of events initiated when particles created in the initial interaction, also interact with nuclei and create secondary particles called a hadronic shower. Due to the larger size of hadronic showers in contrast to the electromagnetic ones, the HCAL makes up a larger part of the calorimeter in order to completely contain them.

The detector is composed of steel and plastic scintillating tiles, used as absorber and active material respectively. It covers the central part of the detector ($|\eta| < 1.7$) and is called Tile Calorimeter (TileCal). A hadron endcap is located in the mid-outer region and covers $1.5 < |\eta| < 3.2$. It is composed of liquid argon and uses copper as an absorber. Between $3.1 < |\eta| < 4.9$ a high-density calorimeter is used. It is made of copper in the first layers while the outer layers use tungsten.

Muon spectrometer

Since muons do not get stopped in the calorimeters, they are the only particles (excluding neutrinos) that reach the outer part of the detector. The muon spectrometer is meant to identify muons and designed to measure their momenta, providing also a muon trigger. Three toroid magnets generate a magnetic field causing a deflection of the muon tracks. The muon spectrometer consists of four subsections: Thin Gap Chambers (TGC), Resistive Plate Chambers (RPC), Monitored Drift Tubes (MDT) and Cathode Strip Chambers (CSC).

When a particle passes the TGC (forward region) or the RPC (central regions) the muon trigger is fired. Their combined coverage range is limited to $|\eta| \leq 2.4$.

In order to determine the momentum of particles in the central region, the MDT is used to measure the curvature. When a muon passes in the forward regions, the trajectory is registered by the CSC. The resolution of these drift tubes is 80 µm in the central region and 60 µm in the forward region.

Magnet system

The central solenoid magnet and the toroid system build up the magnet system consisting of superconducting magnets. Located right outside the inner detector, the first magnet provides a 2 T magnetic field necessary to bend the tracks in order to calculate the particles momentum. The sign of the electric charge of the passing particle can be determined from the tracks curvature as well.

For measurements of the momentum of muons the toroid magnet is used. It is composed of three parts: one located in the central region and the other in the endcaps. Each has eight separate coils that creates a magnetic field of 4 T. They have to be placed in a cryostat since the working point temperature for these magnets lies in a low temperature regime with only a few Kelvin in order to be superconducting.

Trigger and data acquisition

It is not possible to store and process all events which occur in the detector. The number of interactions that happen in a very short time interval is so high that a trigger system must be used to identify "interesting" events and decide which of them should be kept for offline analysis.

For Run 2 the trigger and data acquisition system (TDAQ) has been updated to improve the processing of data recorded with a five times larger event rate in contrast to Run 1. The decision which events are of interest is done in two steps: a first level trigger (L1), that is hardware based, and high level trigger (HLT), which is software based, are used to choose events to be written in different data streams.

The Level-1 trigger decreases the event rate from 40 MHz to 100 kHz. This part of the selection process defines interesting events using information from a few components of the detector. This trigger reaches a conclusion about an event to pass or fail in 2.5 µs after the interaction took place. Then these events are passed to the HLT.

The second trigger segment, the HLT, reduces the event rate further to 1 kHz. It decides if an event is worth keeping by taking the information from the L1 as an input and performing a simplified event reconstruction. This decision is made in 200 ms. All information read from the electronics into readout drivers is kept for events passing the HLT and transfered to the storage. Physics analysis uses all information of an event available while event information are partially recorded for monitoring, detector calibration or trigger level analysis.

The trigger system has a data rate of fully recorded events of 300 MB/s achieved.

3.3 Particle identification and object reconstruction

To identify a particle penetrating the detector, the interaction with the detector material must be understood since particle identification is based on the location of the interaction and the particle's trajectory. The construction of the ATLAS detector with its different layers allows the identification of neutral and charged particles. The signatures of some well known particles detected in specific layers of the ATLAS detector are shown in figure 3.4.

All electrically charged particles will leave a track in the tracking system located in the inner detector. Its curvature is used to determine the momentum of the particle and the sign of its electric charge. No track is visible for neutral particles since no ionisation in the TRT can be caused by them and because of their inability to interact with the pixel detector. In the ECAL electrons and photons will deposit their energy, while muons will reach the muon spectrometer almost without interacting with any layer of the detector. In contrast to neutral hadrdons, which do not deposit any of their energy in the ECAL but purely in the HCAL, the charged hadrons deposit a small fraction of energy in the ECAL and completely are absorbed in the HCAL. Neutrinos are the only particles not interacting with any components of the detector and can only be detected indirectly through the imbalance in energy conservation in the transverse plane ($E_{\mathrm{T}}^{\mathrm{miss}}$).

In this thesis events in which a top quark and a Z boson is produced, are analysed. The signature of such events involves leptons, like electrons and muons, jets and $E_{\mathrm{T}}^{\mathrm{miss}}$. In the following subsection the reconstruction of these objects will be discussed in detail.

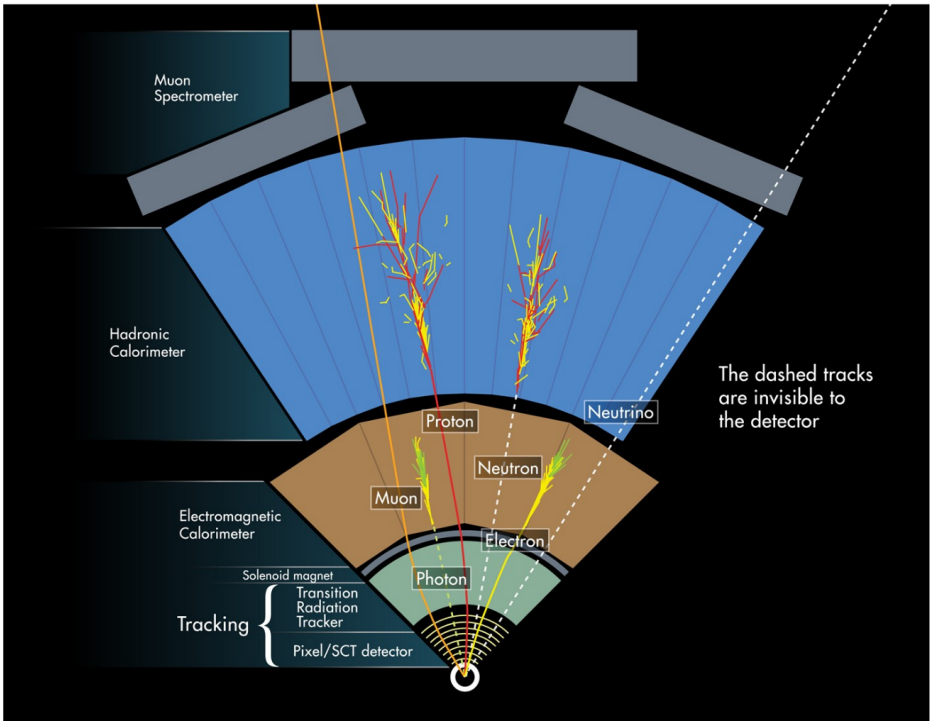


Figure 3.4: Illustration of particle detection in the subcomponents of ATLAS [14].

Electrons

Electron candidates are reconstructed from energy deposits in the ECAL. These track candidates must be associated with tracks from the ID. A transverse momentum of at least $p_{\mathrm{T}} = 25\,\mathrm{GeV}$ is required coming from the inner detector region ($|\eta| < 2.47$). Electron candidates from the transition region between the barrel and endcap electromagnetic calorimeters ($1.37 < |\eta| < 1.52$) are removed because of the large fall in reconstruction efficiency and energy resolution. For a better background rejection, other properties of the event, like the electromagnetic shower shape, calorimeter energy to tracker momentum ratio, and other discriminating requirements are taken into account and combined into a likelihood-based discriminant. Further selections are applied to reject electrons coming from photon conversion, hadronic particle decays or fake electrons.

Electrons are required to be isolated. In general the isolation of objects is related to the energy other particles deposited close to the considered object. The degree of isolation gives information about the origin of the electron. Leptons coming from W or Z boson decays are called prompt while non-prompt leptons resulting from photon conversion or heavy flavour hadron decays are called non-prompt and are not spatially well separated from other energy depositions. The isolation of an electron is defined with two variables: the first one checks how much energy is deposited in the calorimeter within a cone of $\Delta R = 0.2$ surrounding the electron and a track related variable which is defined as the transverse momentum sum of all tracks within a cone of $\Delta R = \min(0.2, 10\,\mathrm{GeV}/E_{\mathrm{T}})$ with the electron track as midpoint. For electrons with $p_{\mathrm{T}} = 25(60)\mathrm{GeV}$ an efficiency of the isolation criteria of 90(99)% is reached.

Muons

Track candidates found in the inner detector are matched with the ones from the muon spectrometer. Hits from the muon spectrometer can be either added or removed to the track candidate from the inner detector in order to improve fit quality. To minimise the probability producing a track by other particle decays, like pions or kaons, each track must fulfill several quality requirements. Additionally the selection of muons with a reliable momentum measurement is improved by these requirements. The quality of combined tracks is verified by the following variables: the normalised χ^2 of the global fit, the accordance of the charge and momentum measurements in the ID with the ones in the muon spectrometer, and the variation between the two ratios of p_{T} measurement in each subdetector to the momentum of the combined track.

The muon candidates are required to have at least a transverse momentum of $p_{\mathrm{T}} = 25\,\mathrm{GeV}$ and be in the inner detector region ($|\eta| = 2.4$). The required isolation for these objects is very similar to electron candidates.

Jets

The reconstruction of jets is done with the anti-k_t algorithm [15] with a distance parameter R of 0.4. Jets are removed if they are within $\Delta R = 0.2$ of an accepted electron candidate. They are also rejected if they are not located within the inner detector ($|\eta| < 4.5$) and have a transverse momentum smaller than 30 GeV. Energy corrections are applied to account for pile-up (additional pp collisions within the same or previous bunch crossings).

In this analysis the final state of the considered event will have a specific type of jet. This jet originates from a b quark where it has its name from, b-tagged jet. The long lifetime of hadrons containing b quarks and the large invariant mass of their decay products compared to the lighter hadrons, is used to identify a jet as a b-jet. Mis-tagging a light- or c-flavour jet as a b-jet corresponds to the rejection factor. This number indicates out of how many non-b-tagged jets one will be wrongly identified as a b-jet.

Missing transverse energy

Particles which do not interact with the detector material, escape detection and hence no reconstruction and energy measurement can be done. Neutrinos are particles passing all layers of the ATLAS detector without depositing energy in any detector component. Missing transverse energy is a quantity which can not be measured directly but is calculated by vectorially summing up the transverse momentum of all identified and reconstructed particles. It is assigned to be the energy of neutrinos.

CHAPTER 4

Analysis setup and strategy

4.1 *tZq* channel

In this chapter the analysis performed for the *tZq* final state and the different possible topologies for this process is presented and discussed. As mentioned earlier, the production of a single top quark involving a gauge boson or the Higgs boson occurs through processes with a very low cross-section requiring high luminosity and collision energy in order to be accessible for study. The production of a top quark in association with a Z boson is such a process with a predicted cross-section below 1 pb. The corresponding Feynman diagram for this production is the t-channel single-top production diagram, along with radiation of the Z boson from any of the quark lines or from the exchanged W boson, both through the electroweak interaction. Figure 4.1 shows the Feynman diagrams at LO in which a top quark in association with a Z boson is produced. For the production of an anti-top quark, charge conjugated processes have to be considered. The cross-section σ_{tZq} is expected to be approximately twice as high as for the associated production of an antitop quark and a Z boson, $\sigma_{\bar{t}Zq}$, due to the difference in the parton distribution functions of the u- and d quark.

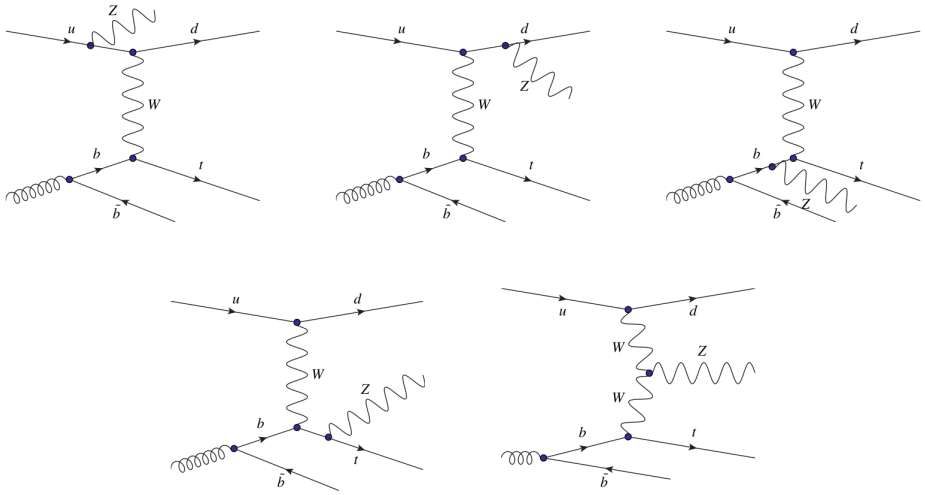


Figure 4.1: Feynman diagrams for the tZq production at LO [4].

tZq production involves two SM couplings, the *WWZ* and *tZ* coupling (see. figure 4.1). Probing two SM couplings in one process, this channel offers access for test of the SM. In addition, *tZq* production is a background in the search for FCNC *tZ* which is highly suppressed in SM since there is no vertex directly coupling neutral currents with two fermions from different generations.

ATLAS was the first collaboration reporting the evidence for top quark produced in association with a Z boson at a centre-of-mass energy of 13 TeV with a measured cross-section of 600 ± 170 stat. ± 140 syst.fb and an observed (expected) significance of $4.2\,\sigma$ ($5.4\,\sigma$). The ATLAS measurement is compatible within the uncertainties with the calculation of the SM cross-section at NLO of $800^{+6.1}_{-7.4}$% fb.

4.1.1 *tZq* final state reconstruction

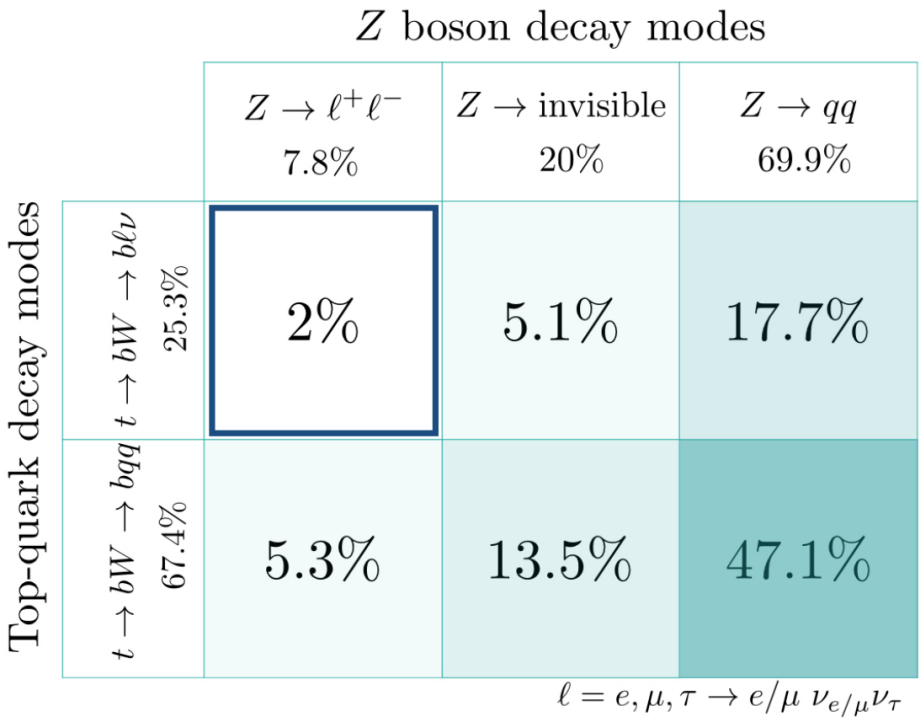


Figure 4.2: The squares show the probability for each combined final state. The channel considered is this analysis is outlined in blue. On each axis the branching ratio for the top quark and Z boson decay is given [4].

The final state of a *tZq* event has different signatures depending on the decay products of the top quark and Z boson. Both of these particles are not stable and decay into a combination of particles, giving the specific signature for each topology. In almost all cases the top quark decays into a W boson and a b quark. In turn, the W boson decays either leptonically or hadronically. In the first case its decay products are a charged lepton and the corresponding neutrino (e.g. $W \to e^- \nu_e$), while for the latter case a quark-antiquark pair is produced ($W \to q\bar{q}$). The Z boson has three ways to decay: into a pair of opposite sign, same flavour leptons, hadronically or into neutrinos, also called invisible decay. Figure 4.2 shows the the corresponding branching ratio for each mentioned final state and the probability for each combination of the W and Z decays. A final state involving leptons has equal

probability for all three lepton families (electrons, muons, taus). Due to its heavy mass, the tau decays into a final state including hadrons or, in roughly 35% of the cases, evenly through two possible leptonic decays: $\tau^- \to e^- \bar{\nu}_e \nu_\tau$ and $\tau^- \to \mu^- \bar{\nu}_\mu \nu_\tau$. The hadronic τ decays are not taken into account in this analysis and are as well not included in the branching ratio numbers shown in figure 4.2. All decays involving leptons only include as decay products electrons, muons and leptonically decaying taus.

Although the *tZq* trilepton channel has the smallest branching ratio for a process with an already very small cross-section, this channel is chosen for this analysis since the presence of three charged leptons reduces the background greatly and is difficult to be reproduced by background processes. In addition, the reconstruction and measurement of charged leptons have better precision than that of hadronic objects in the detector giving the best potential for discovery. The Feynman diagram of the *tZq* final state is shown in figure 4.3 consisting of three charged leptons, two jets, originating from a light and *b* quark, and a neutrino, only observable by the lack of energy conservation in the transverse plane.

4.2 Sources of background

Other processes can have a similar final state as the *tZq* process. These are called background processes and can have misidentified objects by the reconstruction software. Background events will be selected when applying a selection which aims at selecting *tZq* events found by the detector.

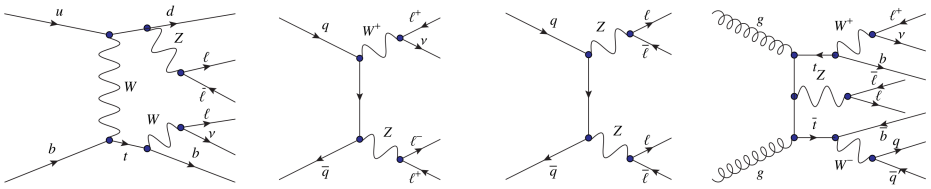
A distinction is made between prompt and non-prompt leptons. Prompt (or real) leptons have their origin in decays of a massive boson (W, Z, H) or the τ lepton. Processes including three prompt leptons, as the *tZq* process, have therefore a similar final state. Non-prompt (or fake) leptons originate from decays of bottom or charm hadrons, electrons coming from photon conversion or jets misidentified as electrons.

The background contribution for processes in which only real leptons are involved, is estimated using MC simulated samples which are normalised to their SM cross-sections. For backgrounds with non-prompt leptons different techniques are used to estimate the number of background events as precisely as possible. In the next subsections the different background processes will be described starting with processes involving three prompt leptons and followed by processes having a non-prompt lepton in the final state.

Diboson

Events in which two W bosons, two Z bosons, or the combination of both bosons are produced, represent the main contribution of backgrounds for the selection of *tZq* events. The three different combinations of gauge bosons do not contribute to the background equally. The main background contribution comes from WZ events in which both bosons decay leptonically. In addition, jets either mis-reconstructed by the reconstruction software or coming from pile-up, will lead to the same final state as a *tZq* event. Events involving two Z bosons contribute much less. This small fraction of ZZ events will have the same signature in the case of a leptonically decay for both bosons in which the fourth lepton is missed either by not being reconstructed or failing the object selection criteria. In figure 4.3 the WW and WZ process with its particles in the final state compared to a *tZq* process are shown.

A LO MC generated sample is used to estimate the background contribution of diboson events. An additional normalisation factor is applied to improve the agreement between data and MC. The determination of this normalisation factor will be discussed in more detail in the following subsection.



(a) tZq (b) WZ (c) ZZ (d) $t\bar{t}Z$

Figure 4.3: Feynman diagrams for (a) tZq), (b)/(c) diboson and (d) $t\bar{t}Z$ processes with final states involving charged leptons [4].

Top quark pair and heavy boson

The production of a top quark pair in association with a massive boson can also pass the selection applied for selecting tZq events. Processes already including a top quark and a Z boson in the final state, such as $t\bar{t}Z$, have a very similar signature. An example of a corresponding Feynman diagram for this process is shown in figure 4.3. The largest fraction of all background events produced in association with a top quark pair and a heavy boson comes from the $t\bar{t}Z$ process, due to the largest cross-section of all productions including $t\bar{t}$ and a boson, even higher than that of the signal. The contribution of $t\bar{t}H$ events is much smaller since the cross-section for this process is very low and the Higgs boson decay involving leptons has a small branching ratio.

Single top quark and gauge bosons

Events with a predicted cross-section of 12.2 fb, being so small that it has not been experimentally measured so far, are those which have a single top quark produced along with a W and Z boson. For leptonically decaying Z and W bosons the final state can be very similar to the trilepton tZq production.

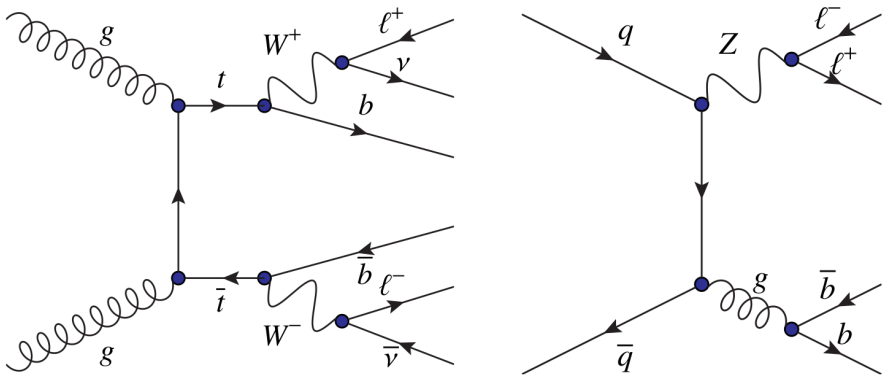
Top quark pair and Z boson in association with jets

The three leptons in the tZq final state are coming from a Z or W boson or a τ lepton, and are therefore prompt (real). In all background events the discussed leptons in the final state have been prompt so far, coming as a decay product from the above mentioned particles. Non-prompt (fake) leptons can lead in processes that already have two real leptons to similar final states with three leptons and pass the selection for such events.

Fake electrons have their origin in the decay of bottom or charm hadrons, photon conversions or a misidentification of jet depositing a large amount of energy in ECAL. Whereas fake muons come from semileptonic decays of heavy flavour objects, particles from decays of charged hadrons or high-energy particles reach the muon spectrometer not being stopped in HCAL. $t\bar{t}$ and Z+jets events are the two sources of non-prompt leptons leading to events with a similar final state to be selected as a tZq event when applying the SR cuts.

Although the presence of three charged leptons reduces the background greatly, the contribution of non-prompt lepton background coming from $t\bar{t}$ and Z+jets events is still non-negligible since both productions have a cross-section several orders of magnitude higher than the tZq process.

The $t\bar{t}$ final state looks identical if one of the b-jets is not properly tagged, both W bosons decay leptonically and an additional non-prompt lepton is reconstructed. A Z+jets event with an additional non-prompt lepton in its final state, will have the same signature as well, again under the condition of a wrong b-tagging and badly reconstructed objects in order to appear as missing energy, since no neutrinos are involved. The Feynman diagrams for both processes including subsequent decays involving charged leptons are shown in figure 4.4.



(a) $t\bar{t}$ (b) Z + jets

Figure 4.4: Feynman diagrams for LO $t\bar{t}$ and Z+jets processes [4].

The agreement of the measured cross-section mentioned in section 2 and discussed in figure 2.1 shows the specific cross-sections of all processes described above. tZq events, labeled as tZj, have the lowest cross-section in contrast to the other productions. The two processes contributing to the non-prompt lepton background, $t\bar{t}$ and Z+jets, have by far the largest cross-section compared to all discussed ones. The WZ process being the dominant source of background has a cross-section of 50.6 ± 2.6(stat.) ± 2.0(syst.) ± 0.9(th.) ± (lumi.) pb [4]. Whereas the $t\bar{t}Z$ cross-section yields 0.92 ± 0.29(stat.) ± 0.1(syst.) pb [4].

4.3 Signal region definition

To minimise the contribution of events having a similar final state as the tZq process, a region must be defined which is enriched in signal events. To achieve this, selection criteria (also called cuts) are defined and applied on the kinematic variables of objects corresponding to particles in the final state. In this signal region (SR) tZq candidate events will be selected while background events will be discarded as much as possible improving the signal to background ratio. Not only requirements on the final state particles are applied but also the reconstructed intermediate particles, like the Z boson, have to fulfill certain conditions. Other regions are defined as well to validate or derive normalisation factors. The cuts for these regions are chosen to be complementary to the SR cuts to avoid overlapping of these regions. Regions which are used for deriving factors or estimating the non-prompt lepton background are called control regions (CRs) whereas in validation regions (VRs), events coming

from a certain process are validated, being at the same time contaminated by signal events as little as possible.

For the tZq trilepton final state the physics objects on which the cuts are applied are three leptons, two jets of which one is untagged and one b-tagged, and missing transverse momentum. To reconstruct the Z boson, two opposite sign, same flavour (OSSF) leptons are selected. If out of the three selected leptons, two such pairs can be identified, the two leptons with an invariant mass closest to the mass of the Z boson are chosen. To reconstruct the top quark, the remaining lepton not used for the Z boson reconstruction, the b-tagged jet and the missing transverse momentum are used. The second and remaining jet is called the forward jet, since it is detected by the detector at high values of $|\eta|$. This jet is assumed to be coming from the hadronisation process of the light quark from the hard scattering process. In order to increase the number of selected signal events and improve the signal to background ratio, a second signal region will be defined. In this region events will be selected identically with the exception of having a second untagged jet, since given the NLO sample, a significant number of tZq events with a jet multiplicity of three is given. One of the two untagged jets, the one with the highest value of $|\eta|$ is marked as the *forward jet*. The region with two jets will be referred to as *2j1b*, whereas the second region as *3j1b*.

In table 4.1 the definitions for the SR are given, showing which cuts are applied on certain objects. Different p_{T} thresholds are required for the three selected leptons. Within a range of $|\eta| < 4.5$ exactly two jets are selected with a p_{T} of at least 30 GeV. The region is extended to this value of η in order to find jets going in the forward direction. However, the b-tagged jet which is one of the two selected jets, has to be found in the central region of the detector ($|\eta| < 2.5$) and pass the 70% b-tagging working point. In order to reduce all backgrounds which do not have a Z boson, the invariant mass of the two leptons associated to the Z boson must be within a 20 GeV window around the Z boson mass ($M_Z = 91.2\,\mathrm{GeV}$). Since a leptonically decaying W boson always produces a neutrino which is only observable in the lack of energy conservation in the transverse plane, one can use the reconstructed transverse mass of the W boson ($m_{\mathrm{T}}(\ell_W, \nu)$) and reject in this way events in which the W boson decays hadronically.

3 leptons with $\lvert\eta\rvert < 2.5$ and $p_{\mathrm{T}} > 15\,\mathrm{GeV}$
$p_{\mathrm{T}}(\ell_1) > 28\,\mathrm{GeV}$, $p_{\mathrm{T}}(\ell_2) > 25\,\mathrm{GeV}$, $p_{\mathrm{T}}(\ell_3) > 15\,\mathrm{GeV}$
≥ 1 OSSF pair
2 jets (3 jets), $\lvert\eta\rvert < 4.5$
1 b-jet, $\lvert\eta\rvert < 2.5$
$p_{\mathrm{T}}(\mathrm{jet}) > 30\,\mathrm{GeV}$
$\lvert m_{\ell\ell} - m_Z\rvert < 10\,\mathrm{GeV}$
$m_{\mathrm{T}}(\ell_W, \nu) > 20\,\mathrm{GeV}$

Table 4.1: Signal region selection criteria applied for selecting tZq events for the 2j1b (3j1b) SR.

Figure 4.5 shows the effect of each cut on the SR composition. Each pie-chart shows the composition of the SR after each cut listed above, was applied. The requirement on the number leptons being three, significantly increases the number of signal events in this region. The estimation of Z+jets events in the SR is derived using a purely data-driven-method and is therefore not shown in this representation. The improvement of the signal over background ratio after each cut is visible in this figure.

As mentioned earlier, the CRs are used for deriving a normalisation factor to get an estimation for

the contribution of diboson events in SR and performing the method to estimate the contribution of $t\bar{t}$ and Z+jets background events. VRs on the other hand are used to validate the derived normalisation factor and the number of estimated background events. Both regions have similar cuts as the SR and are described in the following subsection in more detail.

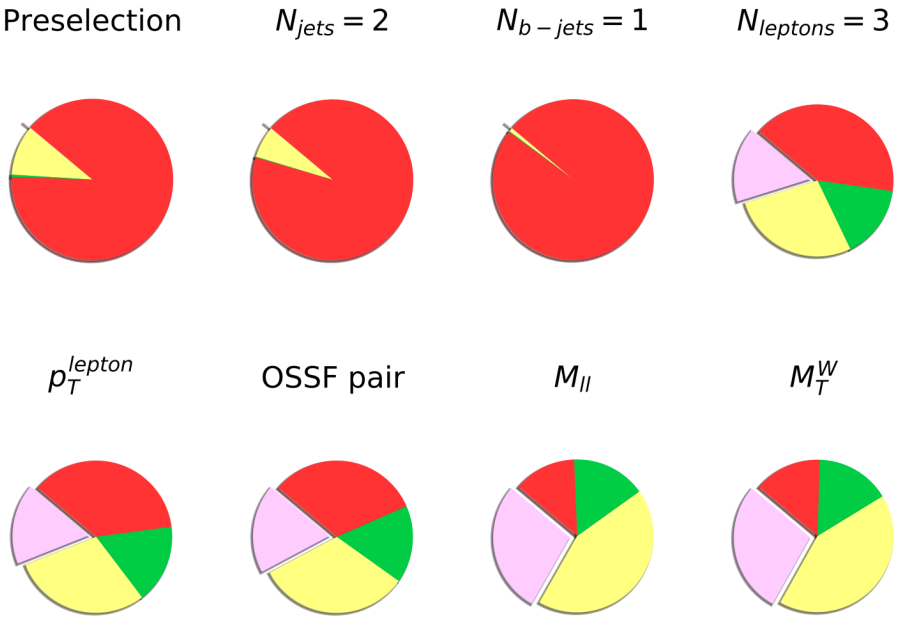


Figure 4.5: Composition of the SR after applying selection criteria [4].

4.4 Data samples and Monte Carlo simulations

In this thesis the performed analysis uses pp data collected by the ATLAS detector in the years from 2015 to 2017 at a centre-of-mass energy of 13 TeV and a total integrated luminosity of 79.8 fb^{-1}, out of which 3.2 fb^{-1}, 32.9 fb^{-1} and 43.8 fb^{-1} have been recorded in 2015, 2016 and 2017, respectively. Before the collected data can be used for physics analysis, it has to be checked first by a data quality group. A *good run list* (GRL) contains all the information about the time period in which data was taken and the corresponding luminosity for which the detector was fully operational. In this sense the amount of data collected by ATLAS is higher than actually used for physics analysis since the selection based on the GRL has to be passed.

Monte Carlo simulation

MC simulations are tools which model processes based on the knowledge of particle interaction and production. Physical parameters are used by MC generators to make predictions for physical processes. Different tests on the SM hypothesis can be done by comparing the collected and analysed data with MC simulation. Such parameters can be masses of particles, in this case, since the genereated MC samples contain top quarks, the top mass is a relevant parameter which is set to 172.5 GeV. The

branching ratio of its decay to a W boson and a b quark has value of 1. As mentioned earlier, leptons will be used to refer to electrons and muons only, since only events including leptonically decaying τ leptons will be selected. Simulated samples are very useful to make estimations about the number of predicted events or background contributions.

To properly generate pp collision events the simulation of certain processes is divided into several steps which correspond to different stages of a collision beginning with parton collision and ending with final state particles detected by the detector.

To simulate a hard scattering process the interaction of two colliding protons has to be well understood. At high centre-of-mass energies the constituents of the two protons interact with each other, not the proton in its entirety but the valence quarks, gluons and sea quarks. Therefore the PDFs of the incoming partons have to be taken into account. Available experimental data of electron-proton and proton-(anti)proton collisions are used by different collaborations to determine and calculate PDFs as precisely as possible to improve the generation of such simulated processes.

Due to the fact that partons (quarks or gluons) are not colour neutral but carry a colour charge and can therefore radiate additional gluons which again can cause emission of radiation, a shower will be produced containing mostly soft gluons. Several algorithms are used by the MC generators to reproduce such parton showering.

Since quarks con not exist in a free state and hence own the property to form bound states forming new hadrons, the process of hadronisation must be modeled and implemented in MC generators. The cluster model [16] and the Lund string model [17] are the most commonly used hadronisation models.

Events in which multiple parton interactions take place have to be considered as well. This is the case when secondary interactions occur between partons coming from the beam remnants. Included are also events in which no hard scattering processes are involved.

After the hadronisation process in which hadrons are produced, the decay products of the unstable hadrons have to be modelled based on the known decay information and branching ratios.

At this level of MC generation all information about the origin and kinematics of the stable particles in the final state and the relation between them are known. The detector simulation has to be applied in order to use the generation for analysis to make any conclusion about the result of data study.

In ATLAS several generators are available, POWHEG [18] [19] [20], MADGRAPH [21], HERWIG 7 and SHERPA are used to generate the matrix element, whereas for parton showering PYTHIA [22] [23], HERWIG 7 and SHERPA are used as generator including hadronisation processes.

Signal and background samples

MC signal events have been generated using MADGRAPH5_aMC@NLO2.6.0 at NLO in which the Z boson was not forced to be on-shell. For the parton shower PYTHIA 8 [23] was used with the A14 set of tuned parameters. For the matrix-element calculations the NNPDF30_NLO_AS_0118_NF_6 parton distribution functions have been used.

Table 4.2 includes details of all simulated background samples used for this analysis. For each process the name of the generator and the parton shower is given. As mentioned above, the estimation of Z+jets background events is derived with a purely data-driven-method and hence no Z+jets MC samples have been used which is the current approach for the analysis. However, the Z+jets samples are listed in the table since the b-jet replacement method for background estimation described in section 5 uses the Z+jets MC samples.

Process	MC Generator	Parton shower
$t\bar{t}$	POWHEG	PYTHIA 8
Z+jets	SHERPA 2.2.1	SHERPA
tW	POWHEG	PYTHIA 8
tWZ	MADGRAPH5_aMC@NLO	PYTHIA 8
$t\bar{t}V$	MADGRAPH5_aMC@NLO	PYTHIA 8
$t\bar{t}H$	POWHEG	PYTHIA 8
WZ	SHERPA 2.2.1 / SHERPA 2.2.2	SHERPA
ZZ	SHERPA 2.2.1 / SHERPA 2.2.2	SHERPA

Table 4.2: Summary of background MC generated samples.

4.5 Background estimation

4.5.1 Diboson normalisation correction

The largest contribution of background events come from the diboson samples. In order to ensure that the observed number of diboson MC events is correctly estimated, regions that are enriched in diboson events are defined as close to the SR as possible. These regions are selected with the same selection criteria as the SR with the exception of requiring a b-jet. An additional cut is applied on $m_{\mathrm{T}}(\ell_W, \nu)$ to reduce Z+jets contamination to a negligible level. A tighter cut on this quantity differentiates the CRs from the VRs. In the diboson CR, with the requirement of $m_{\mathrm{T}}(\ell_W, \nu) > 60\,\mathrm{GeV}$ to separate from Z+jets background populating the low $m_{\mathrm{T}}(\ell_W, \nu)$ region, the number of observed events is compared with the number of predicted events in order to decide whether a normalisation scale factor has to be derived and applied. The comparison of the predicted number of diboson MC events using the SHERPA generator and the number of observed events in the diboson dedicated CR shows a decent agreement demonstrating that no normalisation scale factor has to be applied.

VRs are used to validate and check the modelling of background sources and hence for checking the quality of an applied scale factor (SF). In figure 4.6 the comparison between data and MC prediction is shown for events in the diboson VR for both regions, 2j1b and 3j1b, for which no SF has been applied, based on the good agreement of data and MC in the diboson CR.

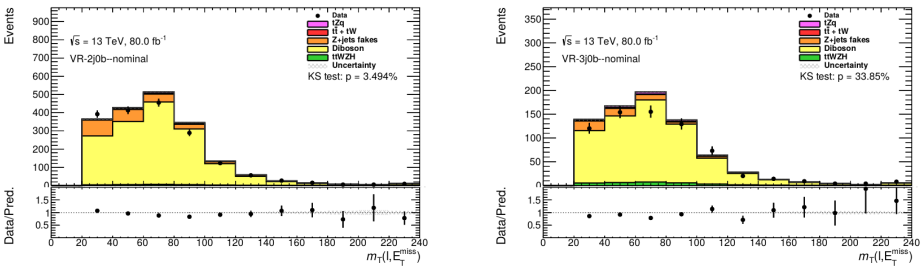


Figure 4.6: Comparison of data and MC predictions as a function of $m_{\mathrm{T}}(\ell_W, \nu)$ for events in the diboson VR. No scale factors have been applied. Only statical uncertainties are included.

4.5.2 Non-prompt lepton in trilepton final state

As explained in section 4.2 a non-prompt lepton coming from a hadron decay or a jet misidentification can cause events to pass the trilepton selection even if the considered events are not trilepton events, since the non-prompt lepton is seen as a "real" one coming from a W, Z boson or τ lepton decay.

For the tZq final state the two sources of non-prompt lepton background are $t\bar{t}$ and Z+jets events. The two contributions in the SR of each sample are evaluated separately with different methods. For Z+jets events the so-called fake-factor method was used, which is a purely data-driven technique, whereas for the $t\bar{t}$ a data/MC scale factor was applied to the $t\bar{t}$ MC predicted events in the SR which was derived in a dedicated CR enriched in $t\bar{t}$ events.

Top quark pair

To select a region having a signal contamination as small as possible and being dominated by $t\bar{t}$ events, different cuts must be applied. The selection criteria for the $t\bar{t}$ CR are kept the same as for the SR but instead of requiring an OSSF lepton pair, only events will pass the selection which have a pair of leptons with an opposite sign and opposite flavour (OSOF). In addition, all events containing an OSSF lepton pair will be discarded. This requirement ensures that no Z+jets events contribute to this dedicated $t\bar{t}$ CR. However, for the $t\bar{t}$ VR an OSSF lepton pair is again required, but this time with its invariant mass being outside of the Z mass window, $m_{\ell\ell} < 81$ GeV or $m_{\ell\ell} > 101$ GeV.

Looking at the comparison between observed data events and the predicted events shown figure 4.7, the good agreement between both samples can be seen. The MC distribution of the p_{T} of the second lepton associated to the Z boson and the p_{T} of the lepton associated to the W boson agrees for each bin within its uncertainty with the observed number of data events in the dedicated $t\bar{t}$ CR. Based on this observation no additional SF for the $t\bar{t}$ background estimation is applied.

For generated MC samples the true origin of particles can be obtained since all information about each particle in an event is known. In order to better understand events contributing to the non-prompt lepton background, a closer look is taken on the origin of such leptons. For all three regions in the $t\bar{t}$ sample (SR, CR and VR) the origin of all leptons in the final state is investigated. Figure 4.8 shows the origin of all three leptons separated for events having an electron or a muon as a non-prompt lepton in the final state. Passing the trilepton final state cuts, which means being reconstructed as a tZq event, the event reconstruction procedure associates each lepton either to the Z boson or the W boson. The highest transverse momentum lepton candidate from the OSSF lepton pair is associated as the first lepton to the Z boson ("Z1" in blue), while the second lepton associated to the Z boson has the second largest transverse momentum ("Z2" in red). The remaining third lepton is associated to the W boson ("W" in green). Although there is no Z boson in a $t\bar{t}$ event and hence the association is not meaningful, the labelling of the leptons association is kept since the same notation is performed for the Z+jets background study and the rest of this analysis.

Figure 4.8 shows that the highest p_{T} lepton is almost exclusively identified as the hardest Z lepton candidate. In turn, the non-prompt electron is identified as the softest Z lepton candidate in over 60% of the cases or as the W lepton candidate in roughly 30%. For the non-prompt muon the probability to be associated to the Z boson as the second lepton or to the W boson is equally distributed around 50%.

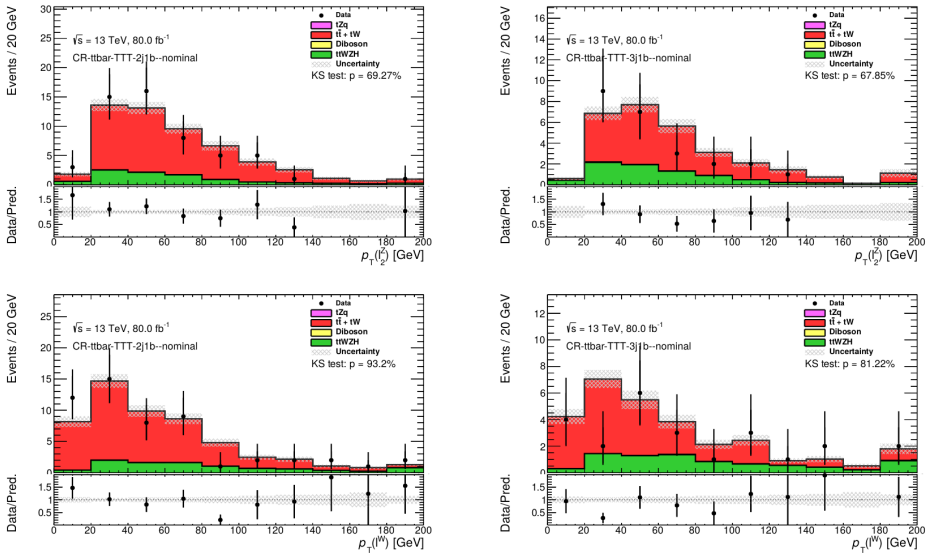


Figure 4.7: Comparison of data and MC predictions of the p_{T} of the second lepton associated to the Z boson and to the W boson for 2j1b (left) and 3j1b (right) for events in the $t\bar{t}$ CR. No scale factors have been applied. Only statistical uncertainties are included.

In addition, the plot shows the dominant source of non-prompt leptons, the decay of the B meson. Performed truth-level studies for the $t\bar{t}$ CR and VR show similar results as for the SR.

Fake factor method

The estimation of the Z+jets non-prompt lepton background is done with a technique which is purely data-driven and called *fake-factor* method. This method estimates the normalisation and shape of this background. With the method a fake factor is derived from the data which gives the probability for a fake lepton being identified as a real one. The factor is derived in a region which is enriched in Z+jets events having at least one non-prompt lepton (FR-TTT region) and applied to a region with the same selection criteria as the SR with the exception that one of the three leptons passes isolation criteria which are looser in contrast to the other two leptons (SR-LTT). With this technique a complete estimation of the non-prompt lepton contribution in the SR is obtained.

A schematic representation of each region with its definition is shown in figure 4.9. The two regions defined for the derivation of the fake-factor are the fake-factor region (FR-TTT/FR-LTT) and the dedicated control region (SR-LTT). The fake-factor region has the same selection criteria as the SR inverting the cut on the transverse mass $m_{\mathrm{T}}(\ell_W, \nu)$ of the W boson reconstructed from its associated lepton and missing transverse energy. For the SR it is required to be greater than 20 GeV while for the fake-factor region a value smaller than 20 GeV is used. The notation "TTT" and "LTT" refer to three leptons and two leptons passing the tight isolation criteria, respectively. As already discussed in section 3.3 the isolation degree of a lepton contains is related to its origin. Non-prompt leptons are not spatially well separated from other energy depositions in the detector and hence they are not as isolated as the prompt leptons. Since it is more likely to select prompt leptons by using a tight isolation working point, it is chosen to be *gradient* with the given efficiency described in the previous

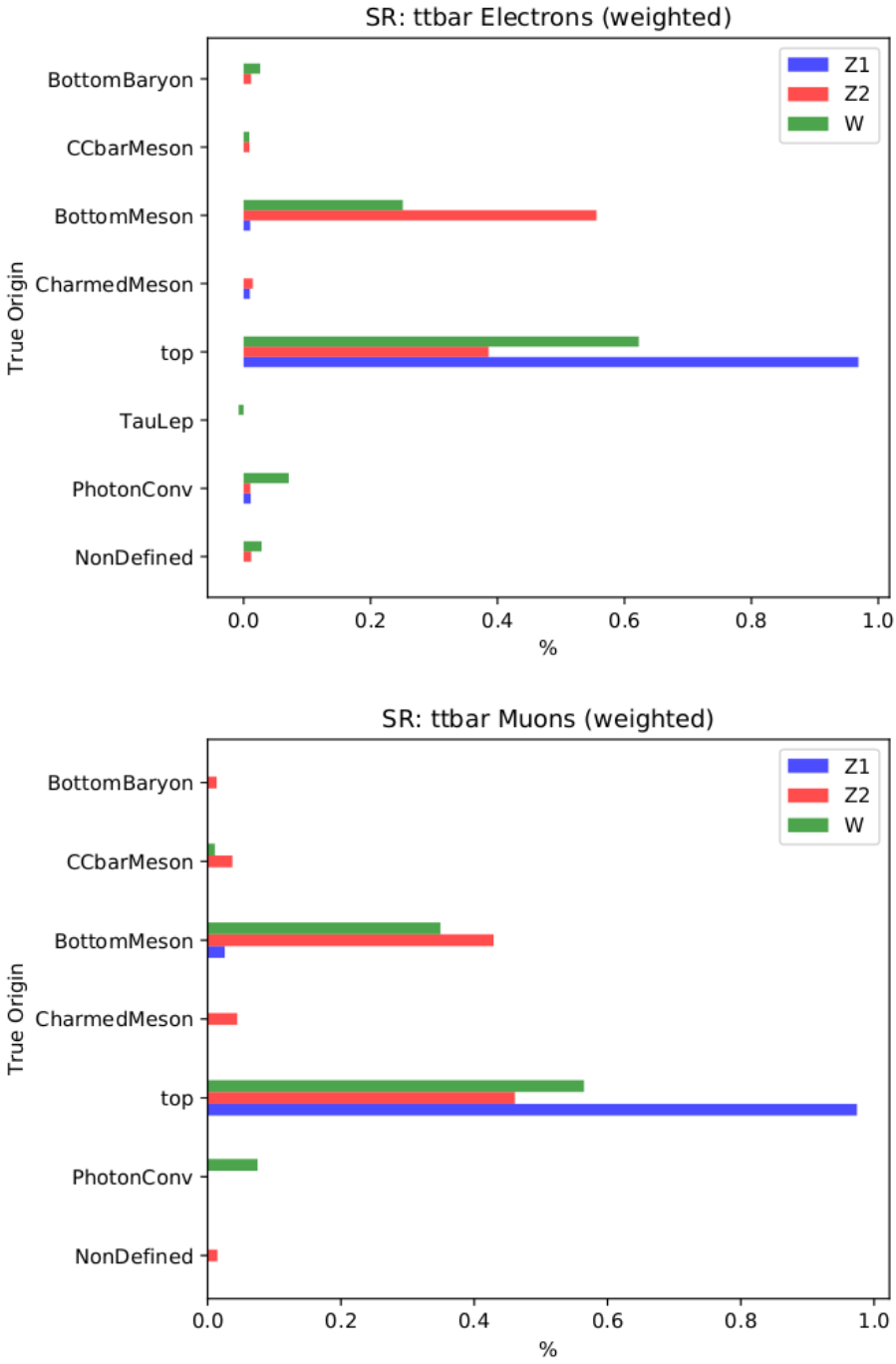


Figure 4.8: Origin of leptons in the SR in the $t\bar{t}$ sample. Fake electron events are included in the top plot and fake muon events in the bottom plot. The lepton with the highest p_{T} is associated to the Z boson as the first lepton (blue) and the lepton with second highest p_{T} as the second lepton (red). The remaining lepton is associated to the W boson (green). On the y axis the origin of the lepton is shown.

section. The requirement for a lepton of being “tight” means to fulfil the isolation criteria of a gradient working point, whereas for “loose” leptons no isolation criteria are required. This is valid for electrons and muons in the same way.

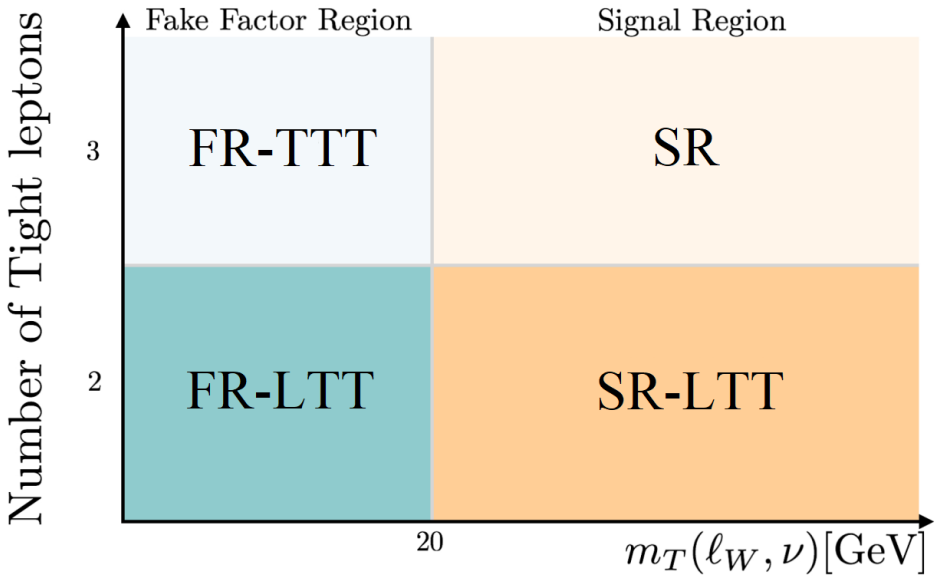


Figure 4.9: Definition of the contol region (SR-LTT) and the fake-factor regions (FR-TTT/FR-LTT) [4].

The fake-factor region is divided into two regions, the FR-TTT in which all selected events have three tight leptons and the FR-LTT region containing events with two tight leptons and one lepton only fulfilling the loose requirements and not being tight (loose-not-tight). The control region to which the fake-factor will be applied to, has the same cuts as the SR except for the requirement of selecting two tight leptons and one which is loose-not-tight (SR-LTT) in order to obtain the total number of Z+jets events in the SR.

The fake-factor (FF) is defined as

$$FF = \frac{N_{TTT}^{FR}}{N_{LTT}^{FR}}, \tag{4.1}$$

where N_{TTT}^{FR} is the number of events in which all three leptons fulfil the tight requirement and N_{LTT}^{FR} gives the number of event in which one of the leptons is loose-not-tight.

Out of the three leptons in the final state of a Z+jets event, the two leptons coming from the Z boson decay will be with very high efficiency successfully reconstructed to the Z boson, with the consequence that the remaining non-prompt lepton will be associated to the W boson. Since in almost all Z+jets events this lepton will be the non-prompt one, the FF will be derived in p_T bins of ℓ_W.

This assertion is checked again at truth-level for all three regions containing a non-prompt lepton. Figure 4.10 shows the origin and the association of the leptons from events from each region in which

the non-prompt lepton is an electron and figure 4.11 for the fake lepton being a muon. In all three regions the electrons and muons coming from the Z boson decay are correctly associated to the Z boson by the reconstruction software confirming the statement that the lepton associated to the W boson is the one being non-prompt. In addition, both figures show that the leptons origin in all regions as well for the electrons as for the muons is coming from heavy flavour decays.

In order to obtain the correct number of Z+jets events in the FR-TTT and FR-LTT region, all events coming from other processes passing the loose and tight isolation criteria, have to be subtracted from the number of data events in both regions. The FF is then calculated and after the subtraction of all background events from data in the SR-LTT, applied in this region to the remaining events to obtain the estimation of the Z+jets contribution in the SR. This method suffers from low statistics in the involved regions which is the dominant source for the large uncertainty for the Z+jets background estimation.

4.6 Event yields

With the estimation of all background events using MC samples and the data-driven method, the expected number of events of each sample, including the NLO signal sample, in the SR is obtained after the full selection criteria have been applied. The event yields for both SRs, 2j1b and 3j1b, are shown in table 4.3 and table 4.4, respectively. The two SRs show similar signal to background ratios. For the 2j1b SR the signal to background ratio has a value of $S/B = 0.23$ while the 3j1b SR has a signal to background ratio of 0.22. In order to improve the low statistics and reduce the uncertainty for the non-prompt lepton background a new method, the b-jet replacement method, to estimate the contribution of events containing a fake lepton to the SR will be introduced and discussed in the following chapter.

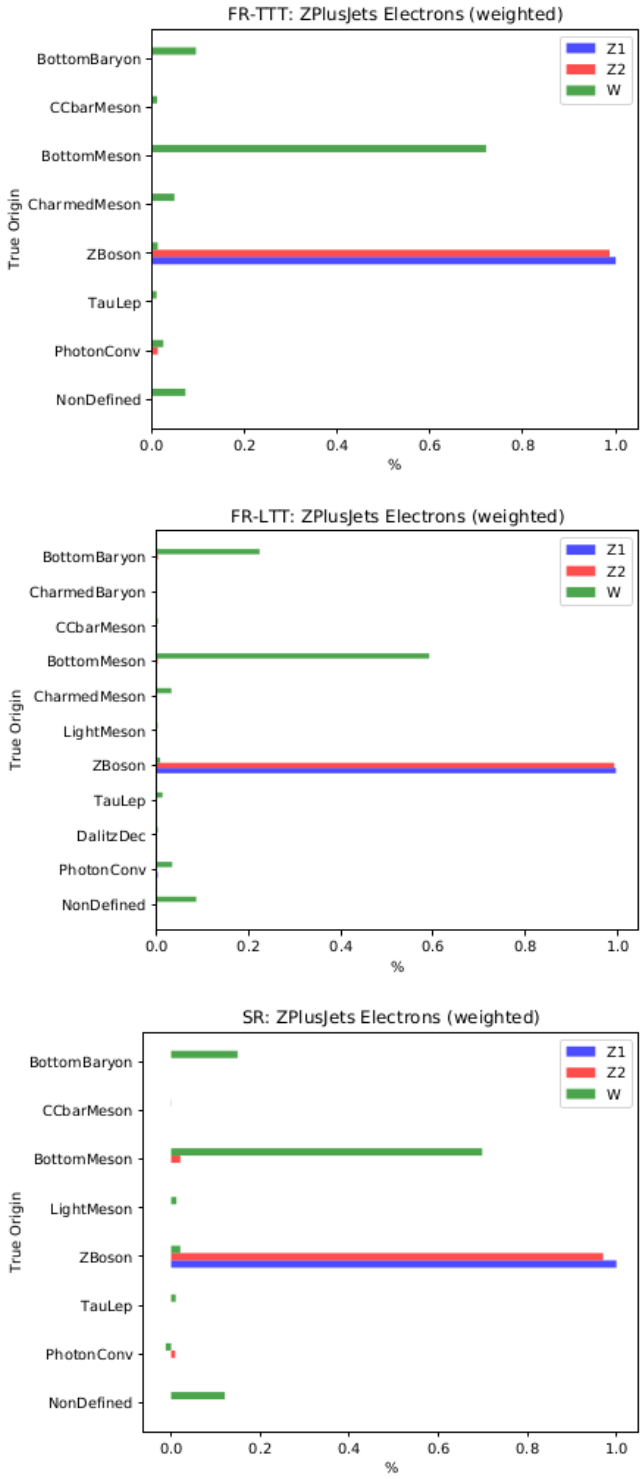


Figure 4.10: Origin of leptons in the FR-TTT (top row), FR-LTT (middle row) and SR-LTT (bottom row) in the Z+jets sample for events that have a non-prompt electron. The lepton with the highest p_{T} is associated to the Z boson as the first lepton (blue) and the lepton with second highest p_{T} as the second lepton (red). The remaining lepton is associated to the W boson (green). On the y axis the origin of the lepton is shown.

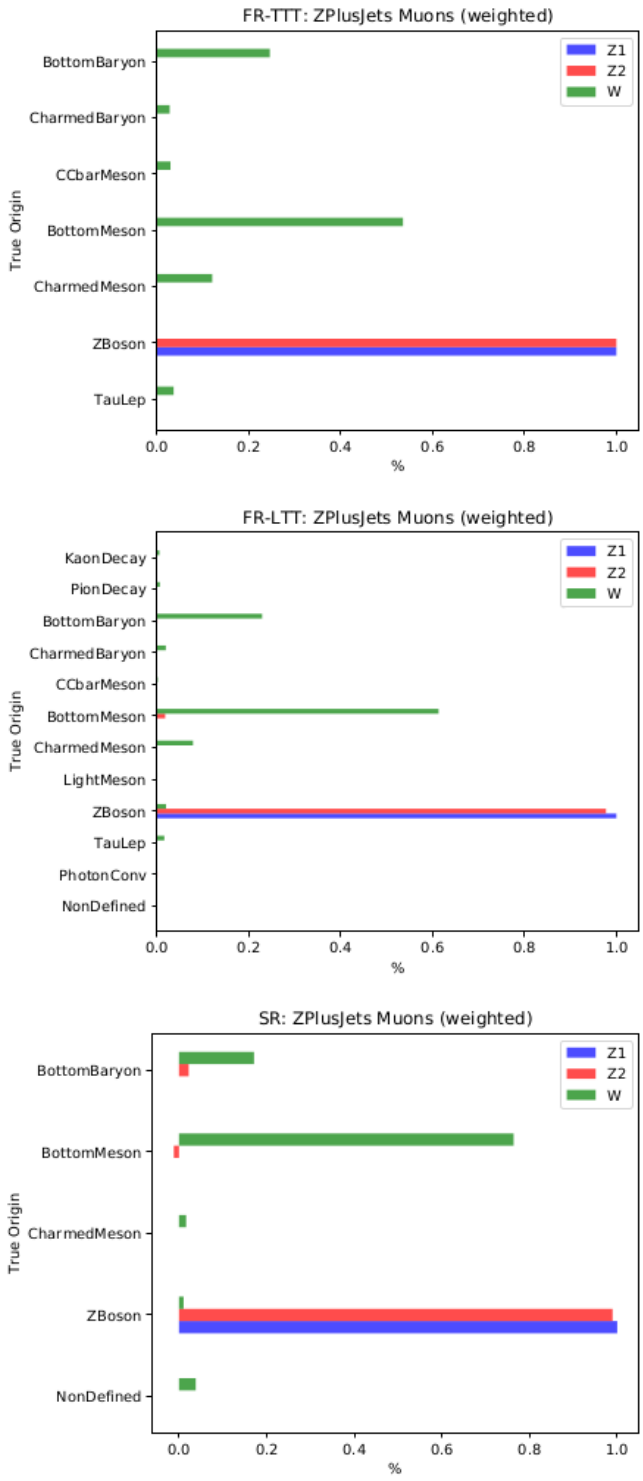


Figure 4.11: Origin of leptons in the FR-TTT (top row), FR-LTT (middle row) and SR-LTT (bottom row) in the Z+jets sample for events that have a non-prompt muon. The lepton with the highest p_{T} is associated to the Z boson as the first lepton (blue) and the lepton with second highest p_{T} as the second lepton (red). The remaining lepton is associated to the W boson (green). On the y axis the origin of the lepton is shown.

Process	Number of events		Process	Raw events
tZq	37.6 ±	0.3	tZq	200 474
$t\bar{t}+tW$	24.5 ±	1.5	$t\bar{t}+tW$	335
Z+jets	31.1 ±	1.5	Z+jets	546
Diboson	70.5 ±	0.9	Diboson	28 242
$t\bar{t}V+t\bar{t}H+tWZ$	37.5 ±	0.6	$t\bar{t}V+t\bar{t}H+tWZ$	21 776
Total expected	201.2 ±	2.4	Total expected	251 373

Table 4.3: Expected event numbers in the 2j1b SR. The left table shows the event numbers after luminosity weighting and the right table shows raw event numbers.

Process	Number of events		Process	Raw events
tZq	27.6 ±	0.3	tZq	142 057
$t\bar{t}+tW$	13.1 ±	1.1	$t\bar{t}+tW$	175
Z+jets	6.5 ±	0.8	Z+jets	266
Diboson	45.2 ±	0.6	Diboson	20 664
$t\bar{t}V+t\bar{t}H+tWZ$	63.2 ±	0.7	$t\bar{t}V+t\bar{t}H+tWZ$	41 929
Total expected	155.5 ±	1.7	Total expected	205 091

Table 4.4: Expected event numbers in the 3j1b SR. The left table shows the event numbers after luminosity weighting and the right table shows raw event numbers.

CHAPTER 5

b-jet replacement method

In order to define a region with a very high signal contribution having at the same time a background contamination as small as possible, the choice of applied selection criteria for the SR is well-considered after the investigation of the SR composition and the precise determination of each background process contribution. MC simulation samples are used to estimate the number of background events containing all information about the involved particles in the process. Since the reconstruction of a non-prompt lepton, which originates mostly from heavy hadron decays, strongly depends on detector effects and details of the reconstruction, a simple use of MC simulation is not sufficient to precisely obtain the number of non-prompt leptons in the SR and give a good estimate. The fake-factor method introduced in the previous chapter gives an estimation of events in the SR containing a non-prompt lepton. An alternative method will be described and discussed in this chapter. This aims at solving the main problem of the fake-factor method by offering high statistics.

In section 4.5.2 the investigation of the non-prompt leptons origin has been discussed for the Z+jets and the $t\bar{t}$ backgrounds. Figures 4.8, 4.10 and 4.11 show that non-prompt leptons used for the reconstruction of the top quark and the Z boson originate from B meson decays. In figure 5.1 the Feynman diagram of a semileptonic decay of a B meson is shown. Based on that, the b-jet replacement method will generate and add a fake lepton to the event record, assuming its origin from a semileptonic B meson decay. Replacing the parent b-jet by the generated non-prompt lepton by applying an overlap removal is explained in section 5.2.1. The idea of this fake estimate is to obtain the shape of the distribution for the non-prompt lepton background.

This method was developed by the analysers of the tZ production via flavour-changing neutral currents (FCNC) [24]. This process has a similar final state as the SM tZq production with similar non-prompt lepton background sources (Z+jets and $t\bar{t}$ production).

5.1 Preselection

The generated lepton has to be added to the event record and handled as any other lepton in the event and later in the subsequent analysis. In order to select dilepton events that can be turned into trilepton events by adding an additional generated lepton, some preselections have to be applied. The starting point are non-prompt lepton background events with similar selection criteria as the SR. Z+jets and $t\bar{t}$ events are selected which consist of two leptons and two b-tagged jets. The rest of the SR selection criteria remain the same. The B meson is assumed to decay exclusively into electrons or muons.

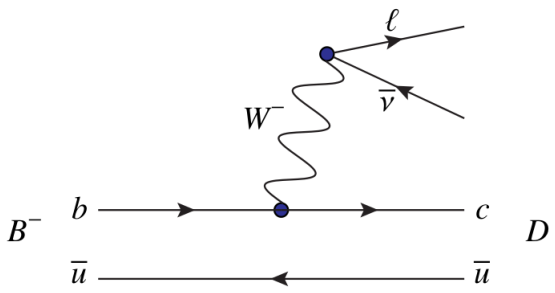


Figure 5.1: Feynman diagram of a semileptonic decay of a B meson [4].

The artificially created lepton is then boosted into one of the b-tagged jets rest frame for which the overlap removal is applied. Because of that procedure, it is required to have two b-jets since one of them is replaced by the fake lepton assuming that it is originating from the b-hadron decay. For every event which passed the preselection criteria and after applying the overlap removal, the kinematic properties of the b-jet are removed and the ones of the generated lepton are added. Selecting events with only one b-jet, none of the events would pass the SR cuts after the overlap removal due to the lack of a b-tagged jet in the final state. In this manner, dilepton events are turned into trilepton events containing a non-prompt lepton and having the same signature as a tZq event.

5.2 Non-prompt lepton generation

All necessary information about the kinematic properties of a lepton coming from a B meson decay in order to be generated, is obtained from MC events in which a non-prompt lepton occurs. The involved neutrino in the semileptonic B meson decay, seen in figure 5.1, makes it more complicated to determine the direction of the lepton, since the track of the neutrino can not be directly reconstructed in the detectors layers. To obtain the energy and the direction of the outgoing lepton, a $t\bar{t}$ MC sample was used, since the knowledge about the decay products from the B meson decay is already encoded. It should be mentioned that the type of the MC sample does not play a role since the kinematic properties of a B meson decay do not depend on the primary hard scattering process. Only events are selected in which electrons or muons are originating from B mesons. In order to associate the lepton to the corresponding b quark in the event, a simple ΔR-matching is performed, searching for the smallest ΔR between the lepton and the b quark. The assumption is made that the four-momenta of the B meson and the b quark are the same. After selecting the pair of lepton and b-tagged jet, the four-momentum of each lepton is boosted in the rest frame of the b quark, in order to calculate the energy of the lepton in the b rest frame. Furthermore a decay angle is defined to obtain information about the direction of the lepton. The leptons three-momentum in the b rest frame and the three-momentum of the b quark in the laboratory frame are used to determine the decay angle. Both quantities, the energy in the b rest frame and the decay angle of each selected lepton, are saved in a two-dimensional histogram shown in figure 5.2.

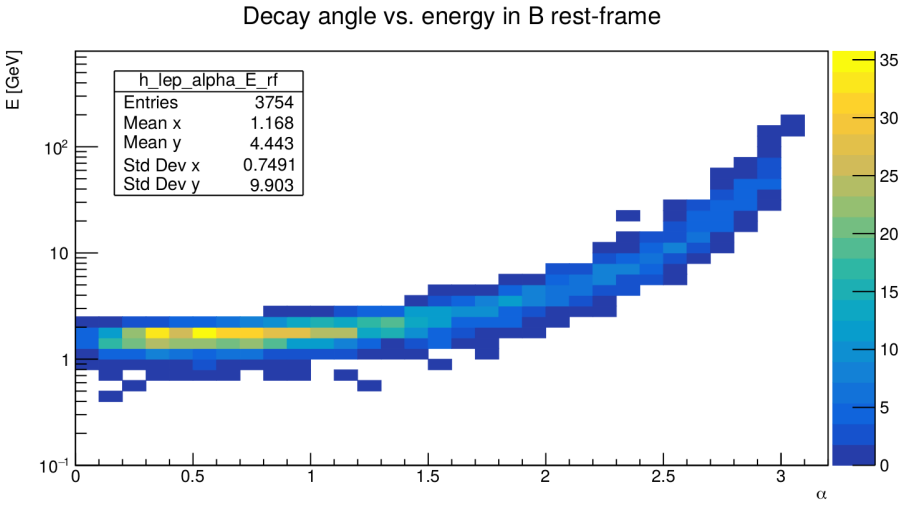


Figure 5.2: Distribution of the leptons energy in the b rest frame against the decay angle α.

After applying the preselection cuts for Z+jets and $t\bar{t}$ events, for each event passing these cuts a lepton is generated randomly with an energy and decay angle according to the two-dimensional histogram and added to the event record. The probability for each lepton flavour ($e^+, e^-, \mu^+$and μ^-) is set to be equal. The lepton is radiated isotropically in the rest frame and hence equally distributed in the range from 0 to 2π in the azimuth plane. The information about the energy and direction of the lepton allows one to construct the four-momentum in the b rest frame which needs to be boosted into the laboratory frame using the b-jet four-momentum. The rotation around the angle between the z-axis of the ATLAS system and the axis according to the jet is the last step in order to fully transform the generated lepton into the detectors coordinate system. The four-momentum of the generated, added and boosted lepton replaces after applying the overlap removal the kinematic properties of the replaced b-jet, in this manner the lepton is implemented and handled as any other lepton in the event record. Which b-jet is chosen to be replaced is discussed in section 5.2.1.

5.2.1 b-jet selection and overlap removal

The requirement of two b-jets before the non-prompt lepton generation, leads to the question which b-tagged jet will be replaced by the generated lepton. In order to choose the "right" b-jet to turn a dilepton event into a "real" trilepton event containing one fake lepton which originates from a B meson decaying semileptonically, the generated trilepton events have to be investigated. As well the fake-factor method for Z+jets events as the MC sample for $t\bar{t}$ events, both estimates give a number of such "real" trilepton fake events since they passed the SR selection criteria requiring three charged and real leptons. The decision which b-jet will be replaced is based on the comparison between trilepton events passing the SR cuts given by the fake-factor method (Z+jets) and MC simulation ($t\bar{t}$) and the dilepton events with a generated lepton replacing a b-tagged jet chosen in two different ways. The two options of selecting a b-jet are either randomly or the one with the lowest p_{T} value. Figure 5.3 shows the comparison for the Z+jets and $t\bar{t}$ samples between both methods choosing the b-tagged jet randomly and with the lowest transverse momentum to be replaced by the generated lepton. The plotted values of the Kolmogorov–Smirnov (KS) test show a significant improvement for

the agreement of the shapes for the case in which the replaced b-jet has been chosen randomly.

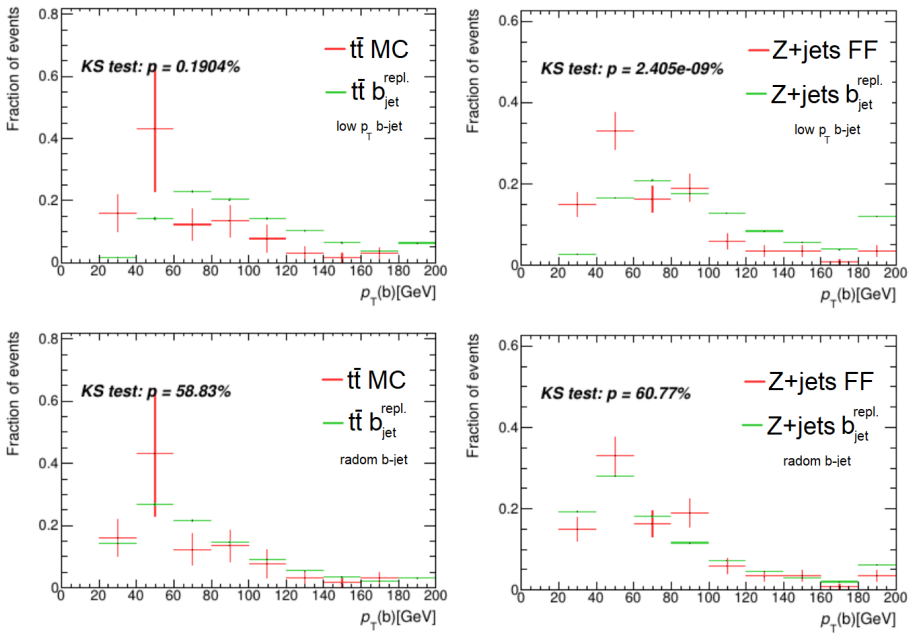


Figure 5.3: Comparison of the distributions of the transverse momentum of the b-jet for the $t\bar{t}$ sample (left) and the Z+jets sample (right) between the two methods choosing the lowest p_{T} b-jet (top) or a random b-jet (bottom) to be replaced for b-jet replacement method.

Based on this observation the four-momentum of the generated lepton is added to the event record after removing all kinematic properties of a b-tagged jet which was selected randomly and after applying the overlap removal.

The procedure of the overlap removal is applied to avoid double counting of the deposited energy when physics objects are reconstructed by the signal registered in the detector. All jets overlapping with electrons or muons within a ΔR smaller than 0.2 are removed from the event. Since this is performed at a very early stage of the analysis, the removed b-jet, if a lepton from a B meson decay overlaps, is no longer available for the offline analysis. Removing a parent b-jet in case of a semileptonic B meson decay, leads to an isolated lepton which is identified as a lepton but is actually a non-prompt lepton. In order to simulate leptons handled in the same way as any other fake leptons, this overlap removal has to be performed.

Removing a randomly chosen b-tagged jet after boosting to its rest frame a lepton generated randomly following a two-dimensional distribution and after applying the overlap removal, creates an event containing a non-prompt lepton passing the SR selection cuts in the same way as a Z+jets or $t\bar{t}$ MC event. With this procedure the raw number of non-prompt lepton background events increases by several orders of magnitude providing high statistics for both of the backgrounds. Due to this improvement the shape and uncertainty can be predicted very well.

5.3 E_T^{miss} recalculation

For events in which the b-jet is replaced by a lepton, the missing transverse energy must be recalculated since the kinematic properties of particles involved in the process and in the calculation of E_T^{miss} have changed. The replaced b-tagged jet is removed from the event record and the added lepton is taken into account as any other reconstructed lepton. The recalculation affects the two background samples, $t\bar{t}$ and Z+jets, differently. Since the $t\bar{t}$ process has in its final state already two neutrinos, (see figure 4.4), carrying a certain amount of energy that will not be detected, the effect on the $t\bar{t}$ sample is smaller compared to the Z+jets sample. For both samples the recalculated distribution of E_T^{miss} has a harder spectrum due to the fact that the generated lepton coming from the b-hadron decay carries less energy than the b-tagged jet and hence the remaining energy in this process is associated to the missing transverse energy. As already mentioned, the lack of neutrinos in the Z+jets final state causes a very soft E_T^{miss} distribution which changes and becomes much harder after the replacement by the fake lepton, since again the energy difference of the b-jet and the fake lepton is associated to E_T^{miss}. In figure 5.4 and 5.5 the comparison of the E_T^{miss} distributions before and after the recalculation for Z+jets and $t\bar{t}$ events are shown, respectively.

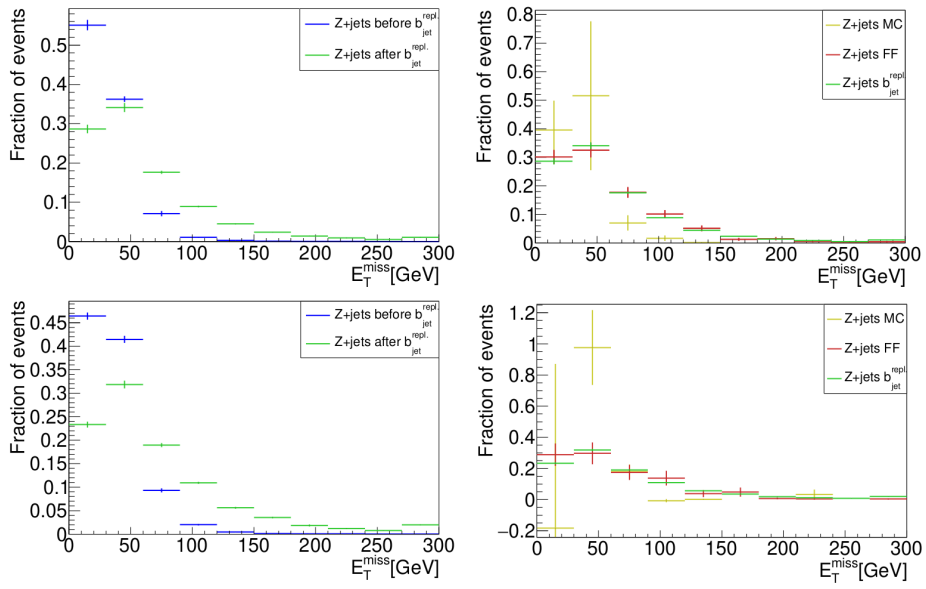


Figure 5.4: Comparison of distributions of E_T^{miss} for Z+jets events before and after the recalculation (left) and Z+jets MC, the fake-factor method and the b-jet replacement method after the recalculation (right). The upper row shows the 2j1b SR and the lower row shows the 3j1b SR.

The recalculated E_T^{miss} distribution is also compared to those of the MC for each background sample and additionally for the Z+jets sample to the fake-factor method. Figure 5.4 shows on the right-hand side for the two SRs, 2j1b and 3j1b, the comparison between the Z+jets MC, the fake-factor method and the b-jet replacement method. Both fake estimates show after the recalculation a good agreement. In contrast, the MC spectrum has a softer spectrum and large error bars indicating the discussed problem of low statistics. Considering the $t\bar{t}$ events in figure 5.5 the change in the E_T^{miss} spectrum is smaller since these events already have a higher E_T^{miss} due to the involved neutrinos in the final

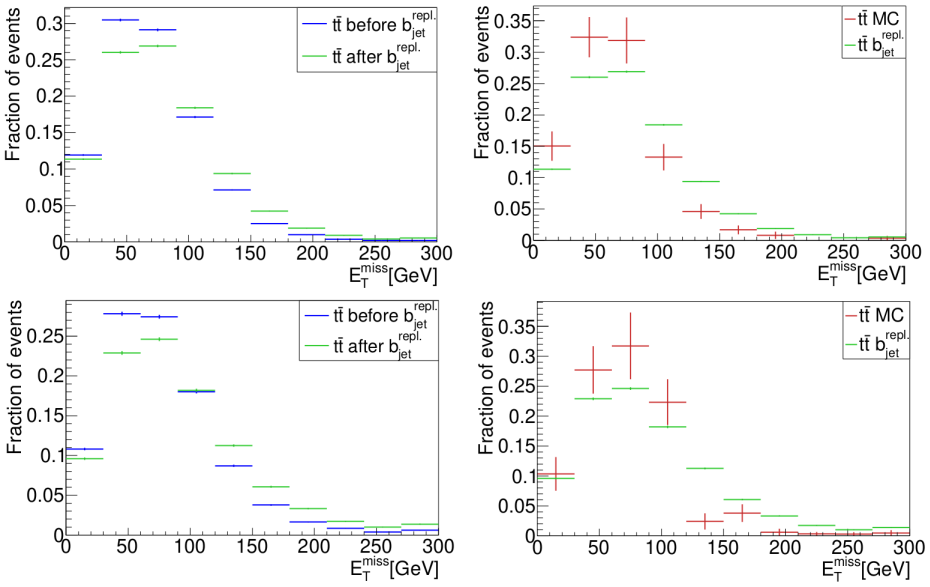


Figure 5.5: Comparison of distributions of E_T^{miss} for $t\bar{t}$ events before and after the recalculation (left) and $t\bar{t}$ MC and the b-jet replacement method (right). The upper row shows the 2j1b SR and the lower row shows the 3j1b SR.

state. The prediction coming from the MC sample agrees with the distribution obtained using the b-jet replacement method. The advantage of higher statistics is visible by comparing the error bars of both distributions.

5.4 Discussion and comparison

To evaluate the b-jet replacement method, kinematic properties of reconstructed physics objects are compared to the ones obtained with the fake-factor method or from $t\bar{t}$ MC. The focus lies especially on the non-prompt leptons associated to the Z or W boson according to truth level studies. The discussion in section 4.5.2 in which the origin of fake leptons associated to the bosons has been studied, showed that for Z+jets events in almost all cases the fake lepton is associated to the W boson, whereas for $t\bar{t}$ events the fake lepton is associated either to the Z boson as the softest lepton or to the W boson with almost equal probability.

Figure 5.6 shows the distribution of the transverse momentum of the lepton associated to the W boson for both SRs. In both regions the prediction of the Z+jets MC sample is, due to the low statistics and hence large uncertainty, not useful for the analysis since no conclusion can be drawn from that. The distribution obtained using the fake-factor method shows, compared to the one retrieved from the b-jet replacement method, a harder spectrum for the transverse momentum of the fake lepton.

In figure 5.7 the p_T distributions of the lepton associated to W boson and the softer lepton associated to the Z boson predicted by the MC and obtained with the b-jet replacement method are compared. For each distribution of the lepton associated to the W boson in both regions the predicted shape by the b-jet replacement method gets softer while the leptons p_T distribution associated to the Z boson is shifted towards higher values compared to the $t\bar{t}$ MC prediction.

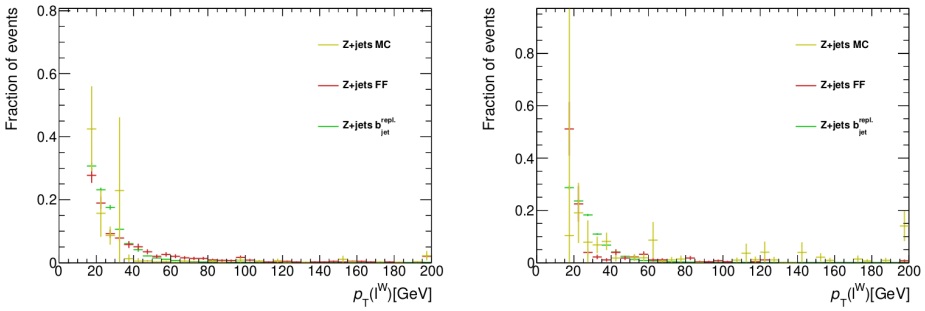


Figure 5.6: Comparison of p_{T} distributions obtained with Z+jets MC, fake-factor method and the b-jet replacement method. The transverse momentum of the lepton associated to the W boson is shown for 2j1b (left) and 3j1b (right).

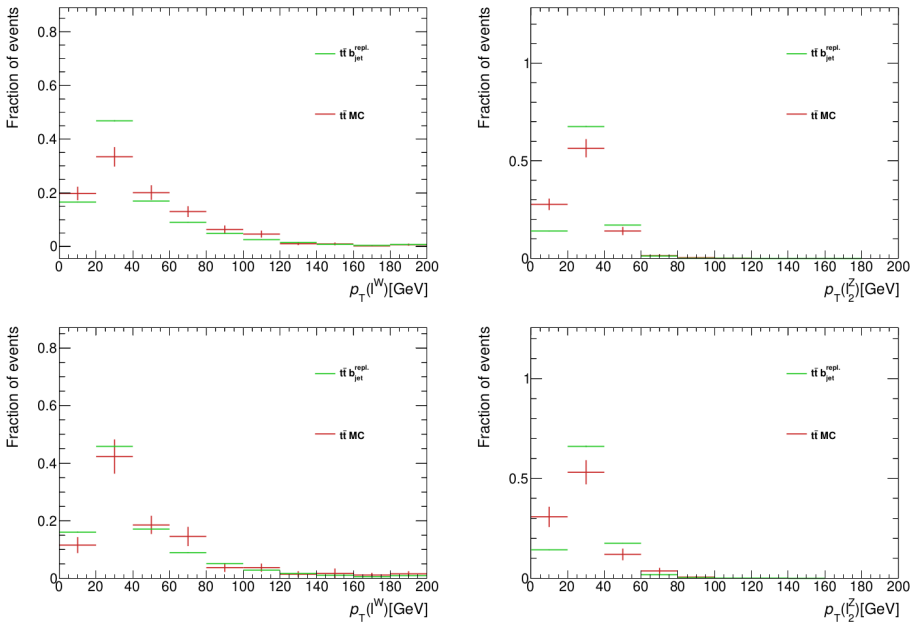


Figure 5.7: Comparison of p_{T} distributions obtained with $t\bar{t}$ MC, and the b-jet replacement method. The transverse momentum of the lepton associated to the W boson and the softer lepton associated to the Z boson is shown for 2j1b (top) and 3j1b (low).

Of particular importance is the remaining b-jet due to the fact that its kinematic properties are used to reconstruct the top quark. Figure 5.8 shows for both SRs the shapes obtained for $t\bar{t}$ and Z+jets using for both backgrounds the MC samples, the b-jet replacement method and additionally only for Z+jets the fake-factor method. A good agreement between both methods for the Z+jets sample, and between the b-jet replacement method and MC for $t\bar{t}$ events is observed which indicates a properly generation of trilpeton events containing a non-prompt lepton.

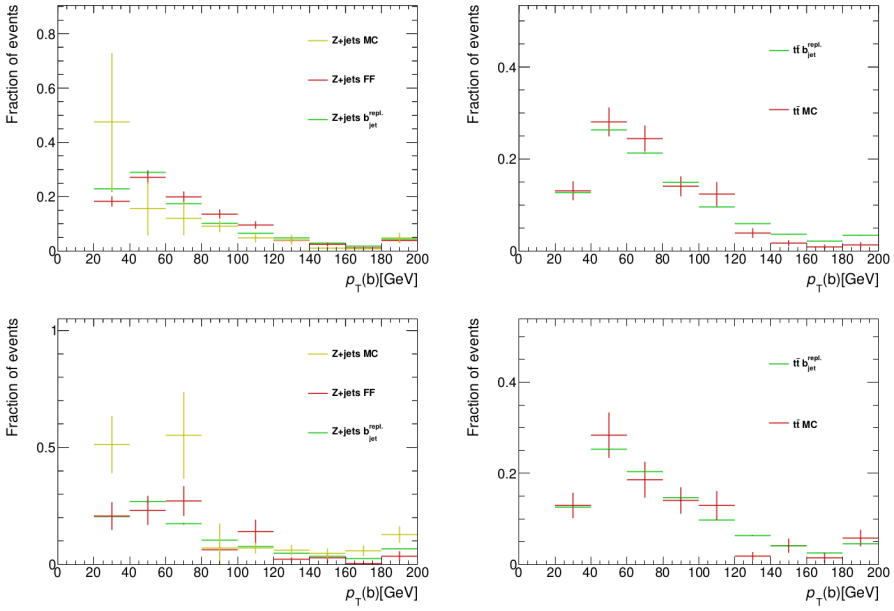


Figure 5.8: Comparison of p_{T} distributions obtained with Z+jets MC, fake-factor method and the b-jet replacement method (left) and with $t\bar{t}$ MC, and the b-jet replacement method (right). The transverse momentum of the remaining b-jet is shown for 2j1b (top) and 3j1b (low).

5.4.1 Event yields

Table 5.1 shows the number of unweighted events obtained with the b-jet replacement method. For the $t\bar{t}$ events the MC sample was used whereas for the Z+jets sample the estimate was made with the fake-factor method. The number of raw events is given for both SRs. Looking at the numbers in the table, one can see the high statistics offered by the b-jet replacement method and hence the smaller statistical errors of this method. In addition the stability of the MVA training is improved which is explained in the next subsection.

Process	MC/FF	$b_{\mathrm{jet}}^{\mathrm{repl.}}$
$t\bar{t}+tW$	335	111 150
Z+jets	546	134 686

Process	MC/FF	$b_{\mathrm{jet}}^{\mathrm{repl.}}$
$t\bar{t}+tW$	175	55 398
Z+jets	266	70 537

Table 5.1: Number of raw events obtained with $t\bar{t}$ MC and the fake-factor method (FF) for Z+jets in the 2j1b SR (left) and 3j1b SR (right) compared the number of raw events obtained with the b-jet replacement method.

5.4.2 Normalisation

Using the b-jet replacement method one can obtain the shapes of the kinematic distributions of the final state objects in events involving a non-prompt lepton. A normalisation for this method has to be obtained since it assumes for each event with two b-jets a semileptonical B meson decay. For this reason the estimated number of non-prompt lepton background events is higher in contrast to the fake-factor method. In order to correct this number a normalisation has to be applied. The b-jet replacement method is evaluated by comparing the shapes of the kinematic distributions from both methods. For this, the initial number of raw background events obtained with both methods has to be normalised to one. In order to estimate the final number of non-prompt lepton events, the normalisation scale factor is set, at this level of the analysis, to the value predicted by the fake-factor method (Z+jets) or from MC ($t\bar{t}$). A final step in the process of normalisation is the performance of a binned likelihood fit which is the ongoing investigation.

A technique which is widely used nowadays to separate signal and background, is the multivariate analysis (MVA). To solve the problem of classification whether an observed event is signal- or background-like, simple selection requirements are not sufficient to identify a region which contains only signal or only background events. Based on all available information about the kinematic properties of particles in a MC event for signal and background samples, the MVA is able to learn to correctly classify events, by using different algorithms. After the process of learning the features of such events, the MVA is "fed" with data events being able to assign a score to events which gives information about the probability for an event to be signal or background.

In this analysis a gradient boosted decision tree (GBDT) based techique was used. This method reproduces a single discriminant by combining all input event information. This type of MVA uses multiple weak learners and converts them into one strong learner. The capability of distinguishing signal from background is for one weak learner low and on its own not useful but combined with other learners into a strong one, the efficiency increases greatly. A collection of iteratively learning weak classifiers is used which results in one final score on the output classification. The weak classifiers are weighted differently depending on the outcome of an event classification. Input events being correctly classified during an iteration have a reduced weight whereas wrongly classified events get a higher weight for the next iteration.

Comparing the normalised MVA output for Z+jets using each method, allows to draw a first conclusion on the method to be drawn. In figure 5.9 and 5.10 the comparison for the SR 2j1b and 3j1b is shown, respectively. The result of this overlay shows compatible shapes for both SRs. The investigation of the MVA output for $t\bar{t}$ is in progress, but is expected to give for both SRs as well a good agreement since, in contrast to Z+jets, the $t\bar{t}$ background was already properly modeled by the MC sample.

The last step in the analysis is performing a binned likelihood fit on the MVA output in order extract the tZq cross-section. The MVA output distributions of the signal and all backgrounds are used as inputs in the fit. For each MC sample, the systematic uncertainties are taken into account and included in the fit. It is a tool which is used to measure the compatibility of the considered model and the given data. It describes the agreement between data and predictions. An Asimov dataset is used to understand the expected behaviour without using data. This dataset is produced by assuming for each parameter in the fit its expected value. This translates to having a set of data consistent

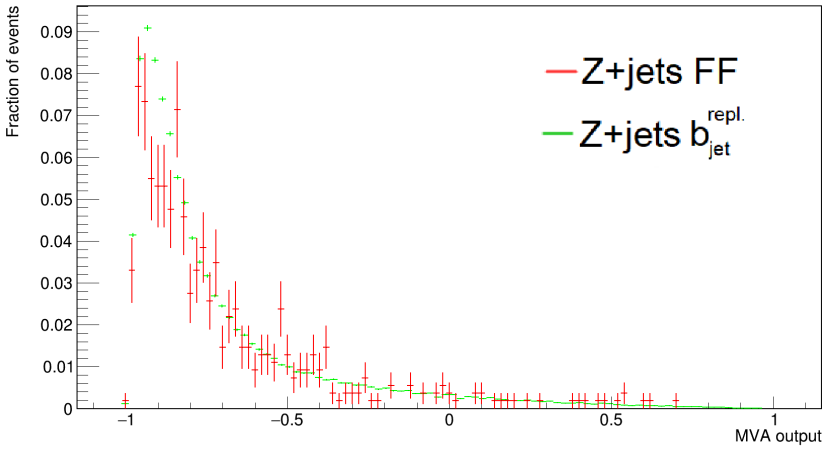


Figure 5.9: Comparison of the normalised MVA output for Z+jets in the 2j1b signal region.

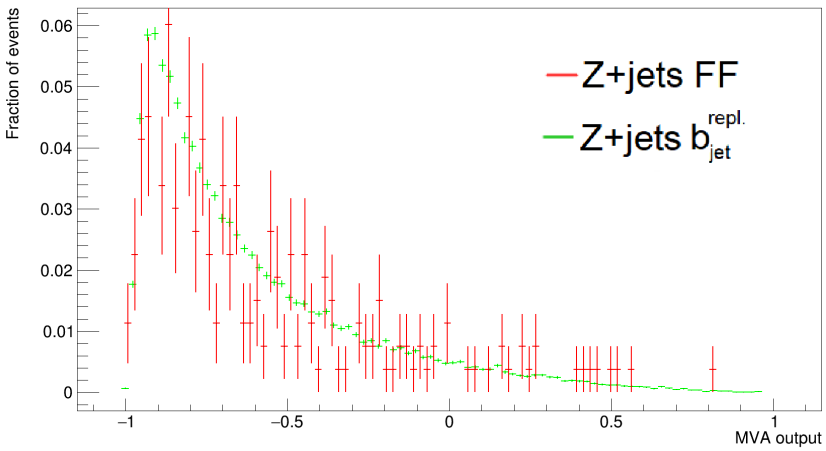


Figure 5.10: Comparison of the normalised MVA output for Z+jets in 3j1b.

with the expectations of the SM. Using the Asimov dataset the expected significance is obtained and can be used for optimising the strategy of the analysis. At the current stage of this analysis the fake-factor method is the default approach. The performance of the likelihood fit is in near future the next step in the *tZq* analysis, using the *b*-jet replacement method as the estimation of non-prompt lepton background events.

CHAPTER 6

Summary

The single top quark production in association with a Z boson is a rare process with a calculated cross-section of $800^{+6.1}_{-7.4}\%$ fb, as predicted by the SM at NLO. Data recorded in the years 2015-2017 with the ATLAS detector at a centre-of-mass energy of 13 TeV with an integrated luminosity of $79.8\,\text{fb}^{-1}$ has been used to search for tZq production.

tZq production can occur through the radiation of the Z boson from any of the quarks or from the exchanged W boson in a t-channel single top quark production diagram through electroweak interaction. Probing two SM couplings in one process, the WWZ and the tZ coupling, studies of this production offer access for testing the SM predictions and can indicate possible BSM effects.

In order to measure the tZq production, a suitable signature must be identified. This process contains several final states depending on the decay mode of the top quark and Z boson. A selection is applied which searches for events in which both particles decay leptonically since this channel offers the highest signal to background ratio and maximises the potential for discovery. For this final state three isolated leptons are required, causing a large background reduction, two jets, one of which is identified as b quark jet and missing transverse momentum.

The measurement of this process is a challenging task since events coming from other processes having many similar final states as the tZq production, pass the selection and contribute to the number of selected events. These backgrounds have much larger cross-sections and the number of such events passing the selection can be estimated using different methods. MC simulations are used to predict the number of events in which three prompt leptons are produced. Events with a dileptonic final state and an additional non-prompt lepton, resulting in the same final state as the tZq process, must be estimated with different techniques. The *b-jet replacement method* presented in this thesis estimates the number of $t\bar{t}$ and Z+jets events which are the two sources of non-prompt leptons. The current approach for estimating Z+jets events is the *fake-factor method*. The alternative method is evaluated by comparing the results of both methods.

A multivariate classification algorithm is used to achieve a separation of signal and background. Several variables are used as an input for a gradient boosted decision tree based technique in order to combine all the information and construct a discriminant to increase the quality of separation. On this discriminant a binned likelihood fit is performed to measure the cross-section for the tZq production.

The alternative non-prompt lepton background estimation method presented in this thesis, shows good agreement compared to the results of the fake-factor method. The shapes of the kinematic distributions given by each method are compatible. Offering high statistics, a precise estimation of the

background and the associated uncertainties is possible. Further developments are ongoing in order to correctly determine the normalisation and integrate the b-jet replacement method in the analysis.

The results presented in this thesis are part of the ongoing tZq analysis searching for the production of a single top quark in association with a Z boson. With more data collected at ATLAS and signal simulated samples produced at NLO, the next aim of going beyond the $5\,\sigma$ statistical significance threshold is in reach after the first strong evidence of this production.

Bibliography

[1] ATLAS Collaboration, *Summary plots from the ATLAS Standard Model physics groups*, URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CombinedSummaryPlots/SM/ (cit. on p. 4).

[2] Wikipedia, *The Standard Model - Wikipedia, The Free Encyclopedia*, URL: https://en.wikipedia.org/wiki/Standard_Model (cit. on p. 5).

[3] ATLAS, CDF, CMS and D0 Collaborations, *First combination of Tevatron and LHC measurements of the top-quark mass*, (2014), arXiv: 1403.4427 [hep-ex] (cit. on p. 6).

[4] I. A. Cioara, *Associated Production of a Top Quark and a Z Boson in pp Collisions at $\sqrt{s}$ = 13 TeV Using the ATLAS Detector*, BONN-IR-2017-XXX, PhD Thesis: University of Bonn, 2018, URL: http://hss.ulb.uni-bonn.de/2018/5194/5194.htm (cit. on pp. 7, 19, 20, 22, 23, 25, 31, 38).

[5] LHCTopWg, *LHCTopWg Summary plots*, URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCTopWGSummaryPlots (cit. on p. 8).

[6] L. Evans and P. Bryant, *LHC Machine*, JINST **3** (2008) S08001 (cit. on p. 9).

[7] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003 (cit. on p. 9).

[8] L. Evans and L. Linssen, *The Super-LHC is on the starting blocks*, CERN Courier (2008) (cit. on p. 9).

[9] CMS Collaboration, *The CMS Experiment at the CERN LHC*, JINST **3** (2008) S08004 (cit. on p. 10).

[10] LHCb Collaboration, *The LHCb Detector at the LHC*, JINST **3** (2008) S08005 (cit. on p. 10).

[11] Alice Collaboration, *The ALICE experiment at the CERN LHC*, JINST **3** (2008) S08002 (cit. on p. 10).

[12] C. D. Melis, *The CERN accelerator complex. Complexe des accélérateurs du CERN*, (2016), URL: https://cds.cern.ch/record/2197559 (cit. on p. 11).

[13] J. Pequenao, *Computer generated image of the whole ATLAS detector*, (2008), URL: http://cds.cern.ch/record/1095924 (cit. on p. 11).

[14] J. Pequenao and P. Schaffner, *A computer generated image representing how ATLAS detects particles*, (2013), URL: https://cds.cern.ch/record/1505342 (cit. on p. 15).

[15] M. Cacciari, G. P. Salam and G. Soyez, *The anti-k_t jet clustering algorithm*, JHEP **04** (2008) 063, arXiv: 0802.1189 (cit. on p. 17).

[16] J.-C. Winter, F. Krauss and G. Soff, *A Modified cluster hadronization model*, Eur. Phys. J. **C36** (2004) 381, arXiv: hep-ph/0311085 [hep-ph] (cit. on p. 26).

[17] B. Andersson et al., *Parton Fragmentation and String Dynamics*, Phys. Rept. **97** (1983) 31 (cit. on p. 26).

[18] P. Nason, *A New method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11** (2004) 040, arXiv: hep-ph/0409146 (cit. on p. 26).

[19] S. Frixione, P. Nason and C. Oleari, *Matching NLO QCD computations with parton shower simulations: the POWHEG method*, JHEP **11** (2007) 070, arXiv: 0709.2092 [hep-ph] (cit. on p. 26).

[20] S. Alioli et al., *A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX*, JHEP **06** (2010) 043, arXiv: 1002.2581 [hep-ph] (cit. on p. 26).

[21] 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*, JHEP **07** (2014) 079, arXiv: 1405.0301 [hep-ph] (cit. on p. 26).

[22] T. Sjöstrand, S. Mrenna and P. Z. Skands, *PYTHIA 6.4 physics and manual*, JHEP **05** (2006) 026, arXiv: hep-ph/0603175 (cit. on p. 26).

[23] T. Sjöstrand, S. Mrenna and P. Z. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv: 0710.3820 [hep-ph] (cit. on p. 26).

[24] N. F. Castro et al., *Search for flavour-changing t Z interactions in proton-proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector*, URL: https://cds.cern.ch/record/2222281 (cit. on p. 37).