Masterarbeit

zur Erlangung des Hochschulgrades
Master of Science
im Master-Studiengang Physik

Data-driven estimation of the non-prompt background in same charged $W^{\pm}W^{\pm}$ scattering within the ATLAS experiment

Max Vincent Stange
geboren am 06.05.1996 in Hoyerswerda

Technische Universität Dresden
Institut für Kern- und Teilchenphysik
Fachrichtung Physik
Fakultät Mathematik und Naturwissenschaften

04.11.2020

Eingereicht am 05. September 2020

1. Gutachter: Prof. Dr. Michael Kobel
2. Gutachter: Prof. Dr. Arno Straessner

Abstract

The scattering of vector bosons (VBS) offers a unique opportunity to study the electroweak sector of the Standard Model, the Higgs mechanism and furthermore physics beyond the Standard Model. A very promising channel to investigate vector boson scattering at the LHC is the scattering of same charged $W^{\pm}$ bosons due to its comparatively high cross-section of the electroweak processes. The second largest background of this $W^{\pm}W^{\pm}jj$-EW signal originates from misidentified leptons (non-prompt leptons).

This thesis aims to improve the data-driven method used so far to estimate this non-prompt background. In order to avoid large extrapolation factors from the control region used for the data-driven method to the signal region, a new control region using dilepton events is defined. This dilepton control region is kinematically closer to the signal region than the dijet control region used in the previous publications. The data-driven method is adapted to the dilepton control region and thoroughly tested with Monte Carlo simulated events. The data studied in this thesis were measured with the ATLAS experiment at a collision energy of 13 TeV with an integrated luminosity of 138.7 fb^{-1}.

Since the signal region of the analysis is blinded, the validity of the data-driven method is proven by using the low dijet invariant mass validation region, which is kinematically very close to the signal region. The data in this validation region is sufficiently well modeled by the sum of the data-driven estimated non-prompt background and the prompt and charge flip contribution estimated by Monte Carlo simulations. Therefore, the data-driven estimation of non-prompt background described in this thesis is expected to have only a small extrapolation to the signal region and thus provides a valuable contribution to the current $W^{\pm}W^{\pm}jj$-EW analysis.

Zusammenfassung

Die Streuung von Vektorbosonen (VBS) bietet eine einzigartige Gelegenheit, den elektroschwachen Sektor des Standardmodells, den Higgs-Mechanismus und darüber hinaus die Physik jenseits des Standardmodells zu untersuchen. Die Streuung von gleichgeladenen $W^{\pm}$ Bosonen ist dabei besonders vielversprechend, da sie einen vergleichsweise hohen Wirkungsquerschnitt an elektroschwachen Prozessen besitzt. Der zweitgrößte Messuntergrund dieses $W^{\pm}W^{\pm}jj$-EW-Signals stammt dabei von falsch identifizierten Leptonen ("non-prompt" Leptonen).

Die datengestützte Methode zur Abschätzung dieses Messuntergrundes soll im Rahmen dieser Arbeit verbessert werden. Um große Extrapolationsfaktoren zu vermeiden, wird dafür eine neue Kontrollregion mit zwei Leptonen pro Ereignis definiert. Diese Dilepton-Kontrollregion ist der Signalregion kinematisch ähnlicher als die Dijet-Kontrollregion, die bisher zur Abschätzung des Untergrundes verwendet wird. Die datengestützte Methode wird an diese Dilepton-Kontrollregion angepasst und gründlich mit Ereignissen aus Monte-Carlo-Simulationen getestet. Die in dieser Arbeit untersuchten Daten wurden im ATLAS-Experiment bei einer Kollisionsenergie von 13 TeV mit einer integrierten Luminosität von 138,7 fb^{-1} gemessen.

Abschließend wird die datengestützte Methode auf eine Validierungsregion mit niedriger invarianter Dijet-Masse angewendet, die kinematisch der Signalregion sehr ähnlich ist. Die Daten in dieser Validierungsregion werden dabei durch die verwendete datengestützte Methode und zusätzliche Monte Carlo Simulationen hinreichend gut modelliert. Daher wird erwartet, dass das in dieser Arbeit beschriebene Vorgehen zur Abschätzung des Untergrundes aus missidentifizierten Leptonen nur eine kleine Extrapolation zur Signalregion aufweist und somit einen wichtigen Beitrag zur aktuellen Analyse des $W^{\pm}W^{\pm}jj$-EW-Signals darstellt.

Contents

Chapter 1

Introduction

For thousands of years, humans have striven to understand what the world is made of and what rules it is subject to. This curiosity is the origin of all modern sciences and the prerequisite for technical progress. The scientific discipline that deals at the most fundamental level with the question of what our universe consists of and what laws apply to these components is particle physics. During the last century, particle physics has made enormous progress and eventually developed the Standard Model of particle physics. Besides general relativity, the Standard Model is the most thoroughly tested theory ever and represents the theoretical basis of all processes in nature apart from gravity. The Standard Model unifies the special relativity with quantum mechanic and describes all known particles and their interactions.

The most powerful experimental setup to test Standard Model predictions is the Large Hadron Collider (LHC) at CERN which is located near Geneva. In this ring accelerator, high energetic protons collide and the results of the collisions are measured by detectors, surrounding the interaction point, the ATLAS detector being one among others. One of the outstanding results of the analysis of the LHC collisions was the observation of the Higgs boson in 2012, which explains how the particles obtain their mass in the context of the Standard Model. With the Higgs boson discovery, the particle content predicted by the Standard Model is complete, nevertheless, there are still fundamental predictions of the theory which have not yet been corroborated. The existence of a quartic vertex in the scattering of the electroweak vector bosons $W^{\pm}$, Z, and γ is one of these rare processes that needed confirmation until very recently. The scattering of vector bosons is expected to proceed via several processes, including the self-interaction of four gauge bosons as well as the exchange of a Higgs boson. Without accounting for all of the processes, the calculated scattering rate grows indefinitely with energy, leading to an unphysical behavior (violation of unitarity). The sensitivity of this process offers therefore a good opportunity to study the Higgs mechanism and furthermore physics beyond the Standard Model.

One of the channels measured at the LHC that is particularly interesting to investigate the electroweak scattering of vector bosons is the production of same charged $W^{\pm}$ bosons. The peculiarity of this channel is the comparatively high cross-section of the electroweak processes. At the LHC the scattering of vector bosons can be studied when two gauge bosons are emitted by the quarks from each of the incoming proton beams. These gauge bosons subsequently scatter off each other and the bosons emerging from this interaction promptly decay. As weak gauge bosons are extremely short-lived particles, in this analysis we search for the scattering of vector bosons by looking for the production of two jets and two leptons produced by the leptonic decay of the $W^{\pm}$ boson pair.

To distinguish the leptons originating from the $W^{\pm}$ boson decay, from other possible sources of similar signals, tight selection criteria are applied to the events measured by the ATLAS detector to obtain a signal from same charged $W^{\pm}$ boson scattering which is as pure as possible. Nevertheless, one cannot fully avoid contamination of the signal region by other processes and this contamination has to be considered as the background of the actual signal. In order to extract the fraction of signal from the selected data, a proper estimation of the background contribution is indispensable.

A previous version of this analysis using 36fb^{-1} of 13 TeV proton-proton collisions gave us the first observation ever of this process [1], that the main experimental uncertainty of that publication was coming from the estimation of misidentified leptons (non-prompt leptons). The non-prompt lepton backgrounds are sources of contamination to the signal, which are generally not properly described by the simulations, and thus need to be estimated in a data-driven way. The second-largest background of the same charged $W^{\pm}$ boson scattering analysis in [1] falls in this category where non-prompt leptons are misidentified and reconstructed with a similar kinematic as the prompt leptons. The previous version of this analysis estimated this background using a data-driven method with a dijet control region. The non-prompt contribution determined in this way made up about 20% of the total background and had the highest uncertainty of all backgrounds which translates into a large impact on the possible sensitivity to new physics effects and on the measured cross-section.

A new analysis is now ongoing considering the full 13 TeV dataset, in order to improve the accuracy of the measured cross-section. It is important to reduce the large uncertainty of this non-prompt background and therefore in this thesis the data-driven estimation of the non-prompt background is performed using a dilepton control region. The advantage of using a dilepton region is that it can be defined in a way to be kinematically close to the signal region. This fact is expected to translate as reduced systematic uncertainties due to the smaller extrapolation between the control region and the signal region. This thesis shows the construction and validation of such a region and is structured as follows. In Chapter 6 the control region is defined and tested for suitability for the data-driven method. Chapter 7 then discusses how the fake factor, which is a crucial part of the data-driven method, is extracted from this region. The general setup is tested in Chapter 8 to find technical bugs and necessities for corrections. In Chapter 9 a correction of the transverse momentum of the leptons is applied, which was already motivated in the same charged $W^{\pm}$ boson scattering analysis in [2]. Finally in Chapter 10 the setup of the data-driven estimation of non-prompt background is evaluated in the low dijet invariant mass validation region.

Chapter 2

Experimental Setup

The data studied in this thesis were measured in the ATLAS experiment at CERN which is located near Geneva. Here high energy protons collide inside the ring accelerator LHC and the final states of this collision are measured by the ATLAS detector. Since both the LHC and the detector are under constant maintenance and upgrades, they are only operated for limited periods of time. This thesis studies the data taken in the Run 2 of the ATLAS experiment. In the following, a short overview of the general experimental setup of the LHC and ATLAS detector is given.

2.1 Large Hadron Collider

The Large Hadron Collider LHC is the most powerful experimental setup ever built to study particle collision. This 27 km long ring accelerator was build in the tunnel of the former electron-positron-collider LEP. At this accelerator the detectors ATLAS, ALICE, CMS and LHCb are located, each associated with a separate experiment.

The LHC was built in order to reach a center of mass collision energy of 14 TeV, but operates during the full Run 2 at 13 TeV. Due to this high energy performance the LHC was designed to discover rare Standard Model processes like the Higgs mechanism and also physic beyond the Standard Model. To achieve such a collision energy, the protons are injected into the LHC only after passing several pre-accelerators. First, the protons are accelerated with the linear accelerator LINAC2 to an energy of 50 MeV. The protons then pass through the Proton Synchrotron Booster (PSB), Proton Synchrotron (PS), and Super Proton Synchrotron (SPS) and thereby are gradually boosted to 1.4 GeV, 25 GeV and 450 GeV before they got injected in the LHC. The whole pre-acceleration takes 30 minutes and the final boost to the maximum of 6.5 TeV in the LHC requires more than 20 minutes.[3]

In order to obtain two counter-rotating proton beams, those coming from the Super Proton Synchrotron are injected into two separate beam pipes which are crossed at the collision points. The protons are kept on the circular path by superconducting magnets, which operate at fields above 8 T. The proton beams itself are not continuous beams but are divided into 2808 proton bunches each. The bunches itself consists of more than 100 billion protons with a bunch crossing every 25 ns, resulting in about 23 interactions per crossing. Due to this high frequency of proton collisions, the luminosity of the LHC is sufficient to study processes with small cross sections. The cross section σ of a process is proportional to the probability that this process will take place and the luminosity L corresponds to the number of particle encounters per time and area. The high luminosity results from the high frequency of the bunches and the exact focusing of the beam at the collision point. The integrated luminosity of the full Run 2 reaches 139 fb^{-1}.[3]

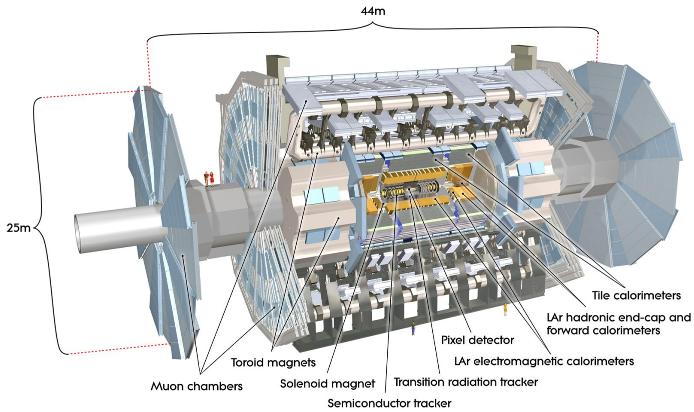


Figure 2.1: Schematic overview of the ATLAS detector with the labelling of its most important components[4]

2.2 ATLAS Detector

The ATLAS detector is a cylindrical detector system of about 45 m length and 22 m height constructed to measure the final states of high energetic proton collisions. Essentially, the ATLAS detector consists of an inner detector, an electromagnetic and hadronic calorimeter, a muon chamber and a complex magnet and cooling system.

2.2.1 Magnet System

The particle track detectors are located inside homogeneous magnetic field to reconstruct particle properties such as momentum and charge from the respective trajectories. The inner detector is surrounded by a thin superconducting solenoid which generates a magnetic field of 2T. Inside the muon chamber, there is an azimuthal symmetric system of eight large superconducting air-core toroids operating at 3.9 T. In addition to the barrel magnet system, two end-cap toroids reaching a magnetic field of 4.1 T are attached at the ends of the calorimeters. The configuration of the magnets of the muon chamber was chosen so that the field is mostly orthogonal to the muon trajectories.[5]

2.2.2 Inner Detector

The first layer starting from the collision point is the inner detector which measures the trajectories of charged particles originating from the collision. For every bunch crossing, about 1000 particles are expected to emerge from the collision point. Due to this high number of tracks, the inner detector consists of a combination of semiconductor pixel and microstrip trackers in the inner tracking volume and a drift-tube detector system in the outer part, to achieve an optimal resolution. The entire inner detector covers a length of 5.3 m and a diameter of 2.5 m.[6]

Next to the vertex region, the highest granularity is achieved by silicon pixel detectors which have a minimal size of $50 \times 400\ \mu\text{m}^2$. The pixel detector consists of three barrels at the radii of 4 cm, 10 cm and 13cm and five disks covering the radii between 11 and 20 cm on both sides.[5]

In the barrel region, the pixel detector is on both sides surrounded by eight layers of semiconductor tracking detectors made from silicon microstrips. At the end caps, nine disks of semiconductor trackers cover the region of $|\eta| \leq 2.5$ (pseudorapidity η is defined in section 2.3.1). The high granularity and vicinity of the pixel layer to the inner detector, allows excellent vertex precision and the reconstruction of secondary vertices.

The outer barrel region of the inner detector is made up by a transition radiation tracker system based on a drift-tube system. This detector section has a lower resolution than the silicon detectors but allows a large number of measurements and therefore has a significant impact on the momentum measurement. The combination of the different types of track detectors provides a very robust and accurate momentum reconstruction inside of the inner detector. The good resolution of the inner detector is also a prerequisite to distinguish between particles from the primary collision and leptons from subsequent processes, such as the decay of a b-quark.

2.2.3 Electromagnetic Calorimeter

In order to measure the energy of electrons and photons, a liquid-Argon (LAr) detector is used consisting of accordion-shaped Kapton electrodes with absorber plates. The absorber plates are made of lead to provide a high showering rate by bremsstrahlung and pair production for electrons/positrons and photons. The barrel region is covered by two identical half-barrels with a small gap of 6 mm around the collision point at $\eta = 0$. The LAr detectors at the end-caps are separated into an inner wheel covering the $2.5 \leq |\eta| \leq 3.2$ region and an outer wheel covering the region of $1.375 \leq |\eta| \leq 3.2$. To ensure that the electrons and photons store their full energy within the electromagnetic calorimeter, the barrel has a thickness of more than 24 radiation lengths and the end-cap calorimeters have a thickness of more than 26 radiation length. In addition a good resolution of the shower is required to distinguish reliably between jets and leptons.[5]

2.2.4 Hadronic Calorimeter

In order to measure the energy of jets passing the electromagnetic calorimeter, it is enclosed by the hadronic calorimeter. The barrel segment of the hadronic calorimeter consists of alternating iron absorber plates and scintillating tiles as the active material. This sampling calorimeter is made of 64 modules and is physically divided into one barrel covering the region $|\eta| \leq 1.0$ and two extending barrels covering the range $0.8 \leq |\eta| \leq 1.7$ each.[5]

The end-caps are mostly covered by two independent wheels with an outer radius of 2.03 m. Their wheels are liquid argon detectors built from parallel copper plates. These plates are 25 mm thick for the outer wheels while those on the outer wheels that are farther from interaction point have a thickness of 50 mm.[6]

The area close to the beam axis is covered by the LAr forward calorimeter which has a high-density design. It consists of one copper module for electromagnetic measurements and two tungsten modules farther from the interaction point predominantly to measure hadronic interactions.[6] The total thickness of the hadronic calorimeter is about 11 interaction lengths to minimize punch-through effects into the muon system as with the electromagnetic calorimeter, good resolution is needed to distinguish reliably between jets and leptons.

2.2.5 Muon Chamber

The muon chambers are arranged in three layers around the beam axis in the barrel region and at end-caps in three layers of plates. In order to achieve the best possible momentum reconstruction from the curvature of the muon track, the alignment of the individual chambers must be very precisely adjusted. A system of about 12000 precision-mounted alignment sensors was installed to obtain a precision of 30 μm on the relative positioning of the muon chambers.[6]

2.3 Measured Variables

The ATLAS detector measures a variety of physical properties. In the following, such properties that are relevant for this thesis and the variables that can be directly derived from them are briefly explained.

2.3.1 Spherical Coordinate System and Geometric Parameters

A spherical coordinate system centered at the collision point is employed to evaluate the geometric distribution of the outgoing particles. The coordinate system is described by the angle θ to the beam axis and the angle ϕ in the plane transverse to the beam axis. However, the pseudorapidity η is generally used rather than the azimuth angle θ because it describes the geometry just as well while having the advantage that differences in pseudorapidity are invariant under Lorentz transformation.

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right) \tag{2.1}$$

Angular separations in η and ϕ are then measured in terms of ΔR.[7]

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{2.2}$$

2.3.2 Variables Transverse to the Beamline

One of the main properties of the measured particles is the momentum p, which is determined from the curvature of the corresponding track. Of particular interest is the momentum distribution in the plane orthogonal to the beam axis called transverse momentum p_T. The advantage of this momentum component is that the sum over all transverse momenta of the particles originating from the interaction has to be zero because it can be assumed that all partons of the original protons had no significant momentum perpendicular to the beam axis. Since one expects that the transverse momenta of the resulting particles should cancel each other out, it is noticeable when high-energy particles remain undetected by the detector. This total missing transverse momentum E_T^{miss} is calculated by the sum over the measured transverse momenta. It is usually used to reconstruct neutrinos which can not be measured by the ATLAS detector.

$$E_T^{miss} = -\sum p_T \tag{2.3}$$

If a heavier particle like a $W^\pm$ boson decays into a lepton and a neutrino, its original mass can be constrained by the so called transverse mass M_T. The transverse mass is calculated from the transverse momentum of the associated lepton and the missing transverse momentum.[8]

$$M_T(\vec{p}_T^{\,lep}, \vec{p}_T^{\,miss}) = \sqrt{2p_T^{lep}p_T^{miss}\left(1 - \cos\left(\Delta\phi(\vec{p}_T^{\,lep}, \vec{p}_T^{\,miss})\right)\right)} \tag{2.4}$$

2.3.3 Invariant Dilepton Mass

The invariant dilepton mass m_{ll} is defined by the total energy of two leptons in their center of mass rest frame. It is used to reconstruct the mass of a particle which had decayed into two leptons. The mass m_{ll} is invariant under Lorentz transformation and is calculated from the four-momenta P of the two leptons.

$$m_{ll}^2 = |P_1 + P_2|^2 \tag{2.5}$$

2.3.4 Lepton Isolation

In order to distinguish particles originating from the interaction at the collision point from particles radiated by jets, the particle isolation is used. To quantify the isolation of a particle a cone around the particle and its moving direction is defined with a certain value of ΔR. The isolation $\mathfrak{I}$ is defined as the sum over all transverse momenta of the particles which are within this isolation cone. The lepton isolations are than grouped in different categories classifying how well non-isolated leptons can be rejected by the certain isolation.[9]

$$\mathfrak{I} = \sum p_T(\in cone) \tag{2.6}$$

2.3.5 Impact Parameters d_0 and z_0

In order to assess the origin of particle tracks, so called impact parameters are used. The transverse impact parameter d_0 is the shortest distance between the track and the beamline in the plane transverse to the beamline. The distance from the primary vertex to the point used to calculate d_0 is described by the longitudinal impact parameter z_0. These two parameters are used during the analysis to ensure that the studied objects originate from the same collision point.[7]

Chapter 3

Theoretical Foundation

3.1 Standard Model

The general procedure in physics is characterized by formulating preferably general theories and proving and potentially falsifying them in reproducible experiments. Besides general relativity, the standard model of particles physic is the most successful theory in modern physic which describes nearly all known processes except of gravity. In the following a very brief overview of the Standard Model is presented and for a deeper understanding e.g. [10] is recommended.

The Standard Model is based on a quantum field theory and thereby unifies special relativity with quantum mechanics. Based on the particle interactions found in nature and experiments, different gauge symmetries were assumed for the Standard Model. To keep the Lagrange density of the Standard Model invariant under the gauge transformations, additional gauge fields must be introduced. Every gauge field can be associated with a respective interaction and an excitation of the field represents a boson which transmits the corresponding interaction. In addition, according to the Noether theorem the existence of a conserved quantity follows from the assumption of a continuous symmetry. This conserved quantity represents the charge a particle must carry in order to participate in the associated interaction.

The Standard Model is characterized by three independent symmetries, a SU(2)×U(1) and a SU(3) symmetry. The SU(3) symmetry represents the strong interaction and describes eg. how protons and neutrons can be formed out of quarks. The strong interaction is transmitted by eight different types of gluons g and couples to the strong color charge $\vec{\mathrm{C}}$, which is a vectorial analogue to the electric charge Q with three components. The SU(2)×U(1) symmetry unifies the electromagnetic force and the weak interaction. Electromagnetism is the best known interaction from our everyday life and is generated by the electric charge Q and transmitted by photons γ. The effects of the weak interaction can be observed primarily in particle transformations. The charge of the weak interaction is given by the weak Isospin $\mathrm{I}_{\mathrm{w}}^{3}$ and the bosons related to this interaction are the Z, W^+ and W^- boson. In contrast to the massless photon and gluon, a mass in the order of about 85 proton masses has been measured for $W^\pm$ and Z bosons.

The particle masses in general were a huge problem for the original formulation of the Standard Model. The theory was able to describe all interactions between particles but the particles itself were massless in the context of this theory, because simple mass terms would violate the gauge invariance of the Lagrange density. This problem was solved by introducing the so called Higgs field which generates the other particle masses. Additionally, the existence of this Higgs field leads also to the existence of another boson, the so called Higgs boson. The Higgs field has also major effects on the electroweak symmetry because its vacuum state brakes the SU(2)×U(1) symmetry of the Standard Model. The Higgs boson was discovered in 2012 and was the last particle by which the Standard Model was extended. A final list of the Standard Model particles and their charges can be seen in Table 3.1. Not listed are

	fermions				bosons				
	leptons		quarks		gauge bosons				Higgs
	e,μ,τ	ν_e,ν_μ,ν_τ	u,c,t	d,s,b	γ	g	Z	$W^\pm$	H
Electric charge Q	-1	0	$+\frac{2}{3}$	$-\frac{1}{3}$	0	0	0	±1	0
Weak Isospin I_w^3	$-\frac{1}{2}$	$+\frac{1}{2}$	$+\frac{1}{2}$	$-\frac{1}{2}$	0	0	0	±1	$-\frac{1}{2}$
Color charge $\vec{\mathrm{C}}$	$\vec{0}$	$\vec{0}$	(red,blue,green)	(red,blue,green)	$\vec{0}$	$\vec{0}$	$\vec{0}$	$\vec{0}$	$\vec{0}$

Table 3.1: Overview of the particles described by the Standard Model and their corresponding charges. For fermions the weak Isospin I_w^3 is only shown for the left-chiral particles, since right-chiral fermions have a weak Isospin of 0. Also not listed are the anti-fermions which are exact duplicates of the listed fermions but with opposite signed charges.

the anti matter fermions which are exact duplicates of the listed fermions but with opposite signed charges.

3.2 Scattering of Same Charged $W^\pm$ Bosons

In order to verify the predictions of the Standard Model on the basis of measurements, especially processes which depend highly on the exact description of the theory are suitable. The scattering of vector bosons (Z, γ and $W^\pm$) for example is extremely sensitive to the exact properties of the electroweak symmetry breaking (EWSB) by the Higgs mechanism. The exchange of a $W^\pm$, γ or Z boson alone would increase the cross section of the longitudinal-polarized scattering quadratic with the center of mass energy to infinity. As a consequence of this the scattering of longitudinal-polarized vector bosons would violate perturbative unitarity after reaching a center of mass energy at the TeV scale. A violation of unitarity is comparable to a scattering probability of over 1 and thus unphysical. This problem is solved by adding the diagrams including the Higgs boson, which restores a reasonable behaviour for high energy scales. Therefore, the analysis of vector bosons is a promising opportunity to study the electroweak symmetry breaking as well as phenomena which are beyond the current Standard Model.[2]

The process of vector boson scattering (VBS) discussed in this thesis is the electroweak scattering of two same charged $W^\pm$ bosons ($W^\pm W^\pm jj$-EW). The advantage of the same charged $W^\pm W^\pm$ scattering over the opposite charged $W^\pm W^\mp$ is given by the fraction of the electroweak mediated diagrams. The scattering of opposite-sign $W^\pm W^\mp$ is dominated by QCD processes whereas in $W^\pm W^\pm$ the contribution of electroweak-mediated diagrams is in the order of the QCD-mediated ones.[2]

In order to study $W^\pm W^\pm$ scattering proton-proton collision are used in the ATLAS experiment and thereby the $W^\pm$ bosons are emitted by quarks. The ways to produce the $W^\pm$ bosons can be split into interactions with only electroweak vertices $W^\pm W^\pm jj$-EW and productions including at least one strong vertex $W^\pm W^\pm jj$-QCD. The strong production can be separated from the pure electroweak process in a gauge invariant way and becomes therefore a background of $W^\pm W^\pm jj$-EW. The electroweak process itself can be categorized into diagrams with vector boson scattering shown in Figure 3.1 and diagrams without any VBS as shown in Figure 3.2. Since there is no gauge invariant way to separate these two categories the non-VBS diagrams must be included in the electroweak VBS signal.[2]

Due to the short lifetime of the $W^\pm$ bosons, they cannot be measured directly in the detector and have to be detected via their decay products. The leptonic decays into $e + \nu_e$ and $\mu + \nu_\mu$ are highly preferred, because they are much more suitable to determine the charge than hadronic decays. Therefore the final states studied for $W^\pm W^\pm jj$-EW consist of two same charged leptons, the two corresponding neutrinos and two jets close to the beam axis originating from the initial quarks.

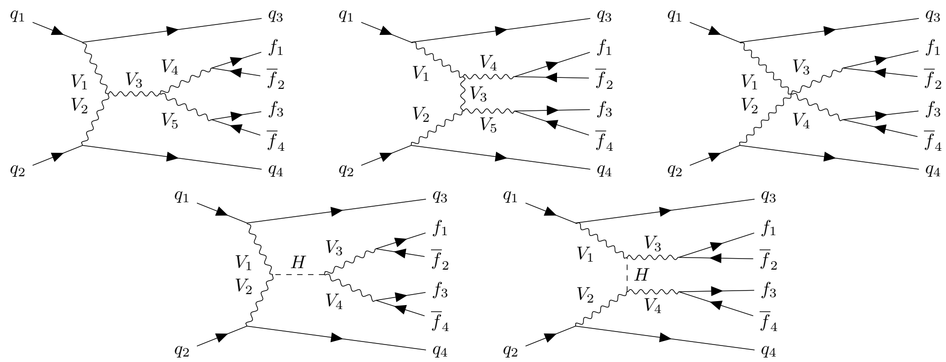


Figure 3.1: Representative feynman diagrams for $VVjj$-EW with vector boson scattering and fermionic final states. The lines are labeled with quarks (q), fermions (f), vector bosons ($V = W^{\pm}, Z, \gamma$) and the Higgs boson (H).[2]

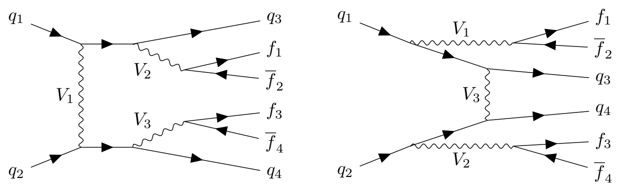


Figure 3.2: Representative feynman diagrams for $VVjj$-EW without any scattering of the vector bosons and fermionic decays of the bosons. The lines are labeled with quarks (q), fermions (f) and vector bosons ($V = W^{\pm}, Z, \gamma$).[2]

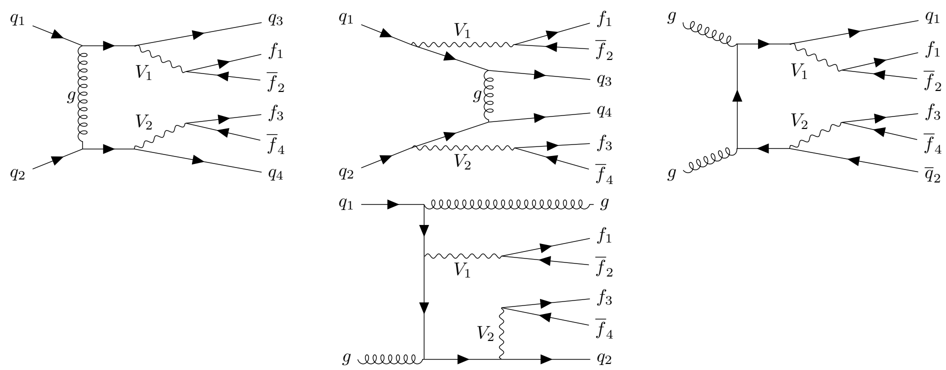


Figure 3.3: Representative feynman diagrams for $VVjj$-QCD with fermionic decays of the vector bosons. The lines are labeled with quarks (q), gluons (g), fermions (f) and vector bosons ($V = W^{\pm}, Z, \gamma$).[2]

3.3 $W^{\pm}W^{\pm}jj$-EW Signal Region

In the course of the ATLAS experiment an extremely high number of proton collisions are recorded. To find interesting events as for example the scattering of same charged $W^{\pm}$ bosons one has to sort out a large part of these data. This is done by applying selection criteria, also called cuts, on several properties measured by the ATLAS detector. In general, a distinction is made between object selection and event selection. Object selections are requirements on kinematic properties of objects measured inside of the ATLAS detector to select e.g. the leptons and jets that are of relevance for the analysis. Event selections are requirements on the whole event and discard events instead of just single objects. The following object and event selections are based on the ATLAS study for same charge $W^{\pm}$ boson scattering from 2018 in [2].

3.3.1 Object Selection

Three differenttypes of objects are involved in the analysis of the final states, electrons, muons and jets. All of them have different kinematic criteria to be selected for the analysis. The leptons of the signal region are classified into two categories. The Analysis (Ana) leptons are the particles which can be considered as decay products of the $W^{\pm}$ bosons and are further studied in the analysis. For the rejection of WZ and ZZ events there is also the class of Veto leptons that have a significant looser selection than the Ana leptons.

Electron Selection

Electrons and photons cause a shower in the electromagnetic calorimeter. Depending on whether a track in the inner detector can be matched to this cluster or not, the respective object can be classified as photon or electron. Additionally, a further classification is done based on the fraction E/p, track p_T, the presence of pixel hits and secondary vertex information of the electron candidate.[2] According to the accuracy of this classification the "author" value is decorated that gives a degree of the unambiguousness of the decision between an electron and photon candidate. Also the shape of the shower, the track quality, the matching between track and calorimeter shower and the detection of the transition radiation are considered due to a likelihood discriminant. These properties are combined in the ID criteria that can be Loose, Medium or Tight.

In addition to these criteria, which refer to the quality of the electron detection, there are also requirements on the lepton kinematics. Both the Veto and the Ana electron tracks inside of the inner detector have to pass the impact criteria of $|z_0 \cdot \sin\theta| < 0.5$mm and $d_0/\sigma_{d_0} < 5$. The absolute value of the pseudo-rapidity $|\eta|$ of Veto and Ana electrons must be smaller than 2.47 and for the Ana electrons there is an additional veto on pseudo-rapidities in the range of $1.37 < |\eta| < 1.52$ due to the crack in the detector geometry. Finally, a minimal transverse momentum selection is applied at 6 GeV for Veto electrons and at 27 GeV for Ana electrons. The full definition of Ana and Veto electrons can be found in Table 3.2.

Muon Selection

As with the electrons, an ID criterion is also used for the muon selection to ensure a sufficient quality of the object measurement. The cut $|z_0 \cdot \sin\theta| < 0.5$ mm on the impact parameter z_0 and θ is chosen as in the electron selection but the requirement on the uncertainty of the d_0 measurement is looser for muons. Another difference to the electron selection is that the same isolation is required for Ana and Veto muons. The selection for the kinematic acceptance is just like in the electron selection $p_T > 6$GeV for Veto muons and $p_T > 27$GeV for Ana muons. The geometrical acceptance of $|\eta| < 2.7$ for Veto muons is looser than for Ana muons with $|\eta| < 2.5$. The final definition of Veto and Ana muons is shown in Table 3.3

	Veto	Analysis
ID	$p_T < 25$ GeV: Medium $p_T > 25$ GeV: Tight	TightLH
p_T	$p_T > 6$ GeV	$p_T > 27$ GeV
η	$\|\eta\| < 2.47$	$\|\eta\| < 2.47$, excluding $1.37 < \|\eta\| < 1.52$
$\|z_0 \cdot \sin\theta\|$	$\|z_0 \cdot \sin\theta\| < 0.5$ mm	$\|z_0 \cdot \sin\theta\| < 0.5$ mm
$\frac{d_0}{\sigma_{d_0}}$	$\frac{d_0}{\sigma_{d_0}} < 5$	$\frac{d_0}{\sigma_{d_0}} < 5$
Isolation	Gradient	FCTight
Author	-	1
Additional requirements	Fail any Analysis selection	-

Table 3.2: Object selection for Ana and Veto electrons based on the ATLAS study for same charge $W^{\pm}$ boson scattering from 2018 in [2].

	Veto	Analysis
ID	Loose	Medium
p_T	$p_T > 6$ GeV	$p_T > 27$ GeV
η	$\|\eta\| < 2.7$	$\|\eta\| < 2.5$
$\|z_0 \cdot \sin\theta\|$	$\|z_0 \cdot \sin\theta\| < 0.5$ mm	$\|z_0 \cdot \sin\theta\| < 0.5$ mm
$\frac{d_0}{\sigma_{d_0}}$	$\frac{d_0}{\sigma_{d_0}} < 3$	$\frac{d_0}{\sigma_{d_0}} < 3$
Isolation	FCTight	FCTight
Additional requirements	Fail any Analysis selection	-

Table 3.3: Object selection for Ana and Veto muons based on the ATLAS note for same charge $W^{\pm}$ boson scattering from 2018[2]

Jet Selection

The jets are reconstructed due to their shower in the hadronic calorimeter using the anti-k_T jet finding algorithm described in [11] with a resolution parameter of $R = 0.4$. After the calibration of the jets, kinematic selections are applied according to geometrical properties of the jet. In the forward Region ($2.4 < |\eta| < 4.5$) the jet transverse momentum has to be above 30 GeV and for the central part of the detector ($|\eta| < 2.4$) a momentum of 25 GeV is sufficient.

Overlap Removal

Due to the object reconstruction the same detector entries may be identified as two different objects. An overlap removal is required to avoid a double counting of the same object. The first step is the overlap removal for two leptons followed by removal of lepton-jet-overlaps.

An electron-electron-overlap can happen if the two electrons share the same track inside the inner detector or if they have overlapping electromagnetic showers. In this case the lower p_T electron is discarded and the higher p_T electron is kept. Another form of lepton-lepton-overlap can occur between an electron and a muon. This might happen, for example, if the muon radiates a high-p_T photon via bremsstrahlung.[2] For such an overlap the muon is kept.

The lepton-jet-overlaps are removed according to their $\Delta R_{l,j}$. The jet is removed if $\Delta R_{l,j}$ is smaller than 0.2, unless it is a heavy-flavor jet. If $\Delta R_{l,j}$ is between 0.2 and 0.4 the lepton jet is kept and the lepton is removed. For muons, there is an additional check if the muon was a ghost particle associated to the respective jet. The jet is removed when it contains a ghost-associated muon and has less than three tracks.

3.3.2 Event Selection

As described in section 3.2, the $W^{\pm}W^{\pm}jj$-EW analysis studies final states with two same charged leptons, the two corresponding neutrinos and at least two jets. In the event selection, the two same charged leptons have to be Ana selected to ensure that the leptons are originating from vector boson decay. In order to avoid contamination coming from events with more than two leptons in the final state like WZ and ZZ scattering all events with additional Ana and Veto leptons are rejected. In order to reduce the background coming from Drell-Yan processes a Z-peak veto $|m_{ee} - m_Z| > 15$ GeV and a selection on the lepton track geometry $|\eta| < 1.37$ is applied in the ee channel. The Drell-Yan background are events coming from annihilation of two leptons and subsequent s-channel production of an lepton anti-lepton pair. To reduce the modelling uncertainty of low mass Drell-Yan processes events with an invariant lepton mass $m_{ll'}$ smaller than 20 GeV are rejected. Another significant background is coming from charge flip events which affect mostly just electrons and are explained in section 4.2. The impact of this background is reduced by the multi-variate ECIDS selection, described in more detail in section 4.2.1 is applied for electrons. This ECIDS selection was not applied in the same charged $W^{\pm}$ boson analysis in [1] but is recommended for the current analysis. A selection on the missing transverse energy $E_T^{miss} > 30$ GeV, motivated by the existence of the two neutrinos, further reduces the charge flip background.

The described object and event selection are thus defined in such a way that a signal of same charged $W^{\pm}$ boson scattering as pure as possible can be measured.

Exactly two Ana leptons with identical electrical charge No Veto lepton
Electrons have to pass ECIDS
ee-channel: $|\eta| < 1.37$, $|m_{ee} - m_Z| > 15$ GeV
$m_{ll'} > 20$ GeV
$E_T^{miss} > 30$ GeV
At least two jets Leading jet: $p_T > 65$ GeV Subleading jet: $p_T > 35$ GeV
b-jet veto
$|\Delta y_{jj}| > 2$
$m_{jj} > 500$ GeV

Table 3.4: The event selection for the signal region of the electroweak same charged $W^{\pm}$ boson scattering described in [2]. The event selection of this ATLAS note is extended by a ECIDS selection to suppress the charge flip background.

Chapter 4

Object and Charge Misidentification

4.1 Non-prompt Background

In the same sign $W^{\pm}$ boson analysis we are looking for events with two prompt leptons from $W^{\pm}$ decay in the final state. Prompt leptons are leptons which are originating directly from $W^{\pm}$ or Z decay and therefore they have specific kinematic properties given by the Ana selection. Unfortunately, these properties can sometimes be also fulfilled by other objects e.g. leptons coming from hadron decays. Such objects are called non-prompt leptons and result in the second largest expected background in the last ATLAS review of same charged $W^{\pm}$ boson scattering of 2019.[12] Due to insufficient Monte Carlo modeling the background originating from of non-prompt leptons was estimated by a data-driven technique.

4.1.1 Non-prompt Objects

In the case of same charged $W^{\pm}$ boson scattering the non-prompt background comes primarily from $W+jets$, $t\bar{t}$ and single top events.[2] Thereby non-prompt electrons and non-prompt muons have in general different origins and kinematic properties. For example the conversion of high energetic photons is much more likely to produce a non-prompt electron than a non-prompt muon. Non-prompt electrons are also a result of jets with a high electromagnetic energy coming from hadronisation to π^0's or from an early showering inside the calorimeter. The semileptonic heavy-flavour decay of a b or c quark can produce both non-prompt electrons and muons. Another source of non-prompt muons are decays of charged hadrons in the tracking volume or in hadronic showers and so called punch-through particles from high energetic hadronic showers, which can escape into the muon chamber.[13]

4.1.2 Non-Analysis Lepton Selection

In the object selection in section 3.3.1 the Veto and Ana were introduced. The Ana selection is chosen so that the lepton is quite likely to be prompt and the Veto selection is significant looser to detect additional leptons eg. from WZ events. In order to implement a data-driven estimation of the non-prompt background, an additional lepton classification has to be introduced. In the context of this thesis this new lepton category is referred to as Non-Analysis (Non-Ana). Non-Ana leptons should be kinematically close to the Ana leptons but more likely to be non-prompt. Therefore most of the selection criteria of Ana are kept and only a few requirements are loosened to also accept object of lower quality.

For electrons the ID criteria is set from TightLH to MediumLH and the author selection is removed. To get a significant contribution of jets faking leptons and leptons originating from jets no isolation is required for Non-Ana electrons and muons. For muons also the requirements on the impact parameters is loosened to $\frac{d_0}{\sigma_{d_0}} < 10$. The Non-Ana leptons are requested to fail at least one selection of the Ana selection to make the Non-Ana category orthogonal to the Ana selection. In order to avoid an overlap with the Veto leptons the leading and subleading Ana/Non-Ana are explicitly removed from Veto container, because these two leading leptons are studied in the non-prompt analysis and therefore cannot additionally be used as Veto leptons. The final object selections for electrons and muons can be found in Table Appendix A.3 and Appendix A.4.

Due to this new lepton category the signal region is extended by events with one Ana and one Non-Ana lepton. For reasons of simplification, this Ana+Non-Ana signal region is still referred to as signal region in the context of this thesis, although it is no longer equal to the original $W^{\pm}W^{\pm}jj$-EW signal region. Whenever the actual $W^{\pm}W^{\pm}jj$-EW signal region with two Ana leptons is mentioned, this is explicitly emphasized.

4.1.3 Data-driven Fake Factor Method

In the search for same charged $W^{\pm}$ boson scattering final states with two leptons are analyzed. If one groups the number of events according to the truth level categories of prompt (P) and non-prompt (F) leptons one will end up with four different truth level event categories:

- N_{PP}: Events with two prompt leptons
- N_{PF}, N_{FP}: Events with one non-prompt and one prompt lepton
- N_{FF}: Events with two non-prompt leptons

These categories are only accessible on Monte Carlo truth level. On detector level these truth level information are not available and one has to deal with the categories of Ana and Non-Ana leptons.

- N_{AA}: Events with two Ana leptons
- N_{AN}, N_{NA}: Events with one Non-Ana and one Ana lepton
- N_{NN}: Events with two Non-Ana leptons

In order to estimate the amount of non-prompt events which are reconstructed as events with two Ana leptons the fake factor method (also known as matrix method) is used. This data-driven approach introduces the probability e_i (f_i) that a prompt (non-prompt) with index i passes the Ana requirements. When these probabilities are given one can write down the following connection between the event categories on truth and reco level.

$$\begin{pmatrix} N_{AA} \\ N_{NA} \\ N_{AN} \\ N_{NN} \end{pmatrix} = \begin{pmatrix} e_1 e_2 & e_1 f_2 & f_1 e_2 & f_1 f_2 \\ \bar{e}_1 e_2 & \bar{e}_1 f_2 & \bar{f}_1 e_2 & \bar{f}_1 f_2 \\ e_1 \bar{e}_2 & e_1 \bar{f}_2 & f_1 \bar{e}_2 & f_1 \bar{f}_2 \\ \bar{e}_1 \bar{e}_2 & \bar{e}_1 \bar{f}_2 & \bar{f}_1 \bar{e}_2 & \bar{f}_1 \bar{f}_2 \end{pmatrix} \begin{pmatrix} N_{PP} \\ N_{FP} \\ N_{PF} \\ N_{FF} \end{pmatrix} \tag{4.1}$$

To find an expression for the number of Ana selected non-prompt events one multiplies the rows of the equation by different term of the form $\frac{f_i}{\bar{f}_i}$.

$$\begin{pmatrix} N_{AA} \\ N_{NA}\frac{f_1}{\bar{f}_1} \\ N_{AN}\frac{f_2}{\bar{f}_2} \\ N_{NN}\frac{f_1}{\bar{f}_1}\frac{f_2}{\bar{f}_2} \end{pmatrix} = \begin{pmatrix} e_1 e_2 & e_1 f_2 & f_1 e_2 & f_1 f_2 \\ \bar{e}_1 e_2 \frac{f_1}{\bar{f}_1} & \bar{e}_1 f_2 \frac{f_1}{\bar{f}_1} & f_1 e_2 & f_1 f_2 \\ e_1 \bar{e}_2 \frac{f_2}{\bar{f}_2} & e_1 f_2 & f_1 \bar{e}_2 \frac{f_2}{\bar{f}_2} & f_1 f_2 \\ \bar{e}_1 \bar{e}_2 \frac{f_1}{\bar{f}_1}\frac{f_2}{\bar{f}_2} & \bar{e}_1 f_2 \frac{f_1}{\bar{f}_1} & f_1 \bar{e}_2 \frac{f_2}{\bar{f}_2} & f_1 f_2 \end{pmatrix} \begin{pmatrix} N_{PP} \\ N_{FP} \\ N_{PF} \\ N_{FF} \end{pmatrix} \tag{4.2}$$

If one now adds the first and last row and subtracts the second and third row, one gets

$$N_{AA} - N_{NA}\frac{f_1}{\bar{f}_1} - N_{AN}\frac{f_2}{\bar{f}_2} + N_{NN}\frac{f_1}{\bar{f}_1}\frac{f_2}{\bar{f}_2} = \left(e_1 e_2 - \bar{e}_1 e_2 \frac{f_1}{\bar{f}_1} - e_1 \bar{e}_2 \frac{f_2}{\bar{f}_2} + \bar{e}_1 \bar{e}_2 \frac{f_1}{\bar{f}_1}\frac{f_2}{\bar{f}_2} \right) N_{PP} \tag{4.3}$$

Which can be rearranged into

$$N_{AA} - e_1 e_2 N_{PP} = (N_{NA} - \bar{e}_1 e_2 N_{PP})\frac{f_1}{\bar{f}_1} + (N_{AN} - e_1 \bar{e}_2 N_{PP})\frac{f_2}{\bar{f}_2} - (N_{NN} - \bar{e}_1 \bar{e}_2 N_{PP})\frac{f_1}{\bar{f}_1}\frac{f_2}{\bar{f}_2} \tag{4.4}$$

The term $N_{AA} - e_1 e_2 N_{PP}$ represents the number of events measured in the Ana signal region minus the number of events in this Ana signal region which are actually prompt. It follows that this difference corresponds to the number of non-prompt events in the $W^{\pm}W^{\pm}jj$-EW signal region N_{fake}. $\bar{e}_1 e_2 N_{PP}$, $e_1 \bar{e}_2 N_{PP}$ and $\bar{e}_1 \bar{e}_2 N_{PP}$ are the number of prompt events where at least one lepton did not pass the Ana criteria. These prompt events N_{ij}^{prompt} are well modeled in simulations and can therefore be subtracted using Monte Carlo simulated events.[14] In this way N_{AN}, N_{NA} and N_{NN} are measured in data in the signal region and N_{AN}^{prompt}, N_{NA}^{prompt} and N_{NN}^{prompt} are simulated by Monte Carlo and therefore only the fractions $\frac{f_i}{\bar{f}_i}$ are missing for the final calculation of N_{fake}. This fraction $\frac{f_i}{\bar{f}_i}$ is also known as fake factor F_i

$$F_i := \frac{f_i}{\bar{f}_i} \tag{4.5}$$

$$N_{fake} = (N_{NA} - N_{NA}^{prompt})F_1 + (N_{AN} - N_{AN}^{prompt})F_2 - (N_{NN} - N_{NN}^{prompt})F_1 F_2 \tag{4.6}$$

In the analysis for WZ scattering different fake factors were used for the leptons assigned to the $W^{\pm}$ boson and the leptons assigned to the Z boson.[14] In the same charged $W^{\pm}$ analysis all leptons are assigned to $W^{\pm}$ bosons and so we expect that the fake factors F_1 and F_2 are in general equal. This does not mean that the fake factor F is just a constant factor, it is still dependent on the object type and it's kinematic properties like p_T and η. The fake factor will be extracted from data in a separate control region as described in section 4.1.4.

In the analysis for WZ scattering additional simplifications were made due to the magnitudes of the fake factors.[14] The probability that a non-prompt lepton is Non-Ana selected is significantly higher than the probability that it is Ana selected and so one expects the fake factor to be small. This is the motivation to neglect higher orders of F and simplify the application of the fake factors. However, one has to be careful, as the number of events $(N_{NN} - N_{NN}^{prompt})$ can be large and therefore their influence could not be neglected so easily. The simplification that the fake factor is only applied on events with one Ana and one Non-Ana lepton has to be also considered when one extracts this fake factor to avoid a systematic overestimation of N_{fake}.

$$N_{fake} = (N_{NA} - N_{NA}^{prompt})F + (N_{AN} - N_{AN}^{prompt})F + \mathcal{O}(F^2) \tag{4.7}$$

This is the final fake factor method calculating the number of non-prompt events in the Ana signal region N_{fake} using following parameters from data and Monte Carlo.

- N_{NA} and N_{AN}: Number of events with one Ana and one Non-Ana lepton measured in data in the signal region
- N_{NA}^{prompt} and N_{AN}^{prompt}: Number of prompt events with one Ana and one Non-Ana lepton extracted from Monte Carlo simulated events in the signal region
- F: The fake factor extracted from a fake factor control region

4.1.4 Fake Factor Extraction

As described above, the fake factor is extracted in an additional control region. This control region consists of Ana and Non-Ana leptons and should be dominated by non-prompt events. One assumes that the probability f ($\bar{f}$) that a non-prompt lepton is Ana (Non-Ana) selected is in agreement between the signal region and the control region. These probabilities can be extracted using the total number of non-prompt leptons N^{fake} and the number of Ana (Non-Ana) selected non-prompt leptons N_A^{fake} (N_N^{fake}).

$$f = \frac{N_A^{fake}}{N^{fake}} \qquad \bar{f} = \frac{N_N^{fake}}{N^{fake}} \tag{4.8}$$

Thereby the fake factor is given by the fraction of non-prompt Ana and non-prompt Non-Ana leptons.

$$F = \frac{f}{\bar{f}} = \frac{N_A^{fake}}{N_N^{fake}} \tag{4.9}$$

Since the distinction between prompt and non-prompt objects is again only possible at truth level, the total number of Ana and Non-Ana leptons is measured and then the number of prompt leptons is subtracted using Monte Carlo simulations.

$$F = \frac{N_A - N_A^{prompt}}{N_N - N_N^{prompt}} \tag{4.10}$$

Due to this subtraction of prompt events it is advantageous to use a control region with high purity in non-prompt events. Furthermore, the region should be kinematically close to the signal region to avoid systematic biases in the fake factor method. One central assumption of this method is that the fake factors in signal and control regions are approximately the same. Therefore both regions should have a similar composition of non-prompt types/origins, since different non-prompt objects have different probabilities to be Ana or Non-Ana selected. A different approach was followed in the analysis of WZ scattering where the factors of different regions were combined.[14] However, even in this case the composition of the non-prompt objects has to be studied and the final combination of factors is far from trivial. Therefore in this thesis only one control region is used, which is kinematically optimized according to the signal region.

4.2 Charge Misidentification

One main property of a particle is its charge. In the ATLAS experiment the charge of a particle is determined by the curvature of its track measured in the inner detector and muon chamber. In most cases the charge is correctly reconstructed but sometimes a misidentification can happen. This problem typically occurs just for electrons.

One cause could be a very high energy of the electron which leads to an indeterminable curvature and thereby a charge misidentification.[15] Therefore the charge measurement of highly energetic electrons is limited by the resolution of the inner detector.[16] In this case the measured object is the same lepton originating from the collision but with a reconstructed charge with opposite sign of its true charge. Another and more dominant effect leading to charge misidentification is the process of charge flip due to bremsstrahlung. In the inner detector the electron emits via hard bremsstrahlung a photon which carries most of the momentum of the electron. This photons creates via pair production a positron with high momentum and a significantly slower electron. The two electrons get lost in the detector due to their low momentum and in the detector a positron is measured which has a similar momentum and direction as the original electron. This charge flip effect is also known as trident electron and occurs in exactly the same way for positrons.[16]

In general the effect is quite rare but especially in processes like the scattering of same charged $W^\pm$ bosons charge flip events can contribute a significant background. In the last ATLAS analysis of same charged $W^\pm$ boson scattering the background coming from $V\gamma$ and charge flip events was estimated at about 20% of the total background and is therefore comparable to the background coming from non-prompt leptons.[1] When we are looking for final states with same charged leptons originating from $W^\pm$ bosons we are also observing events originating e.g. from Z decay. When a Z boson decays in a electron positron pair one of these leptons could flip its charge and we end up with a final state with two same

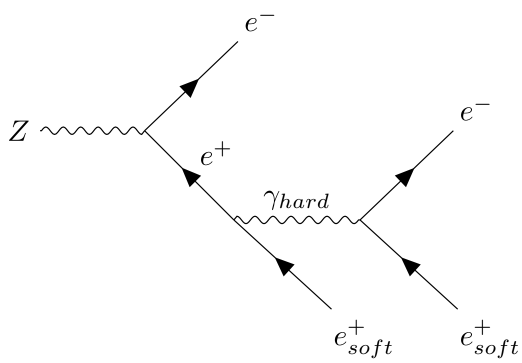


Figure 4.1: Charge flip of an positron originating from Z boson decay due to hard bremsstrahlung

charged leptons which is quite likely to contribute to the signal region. This process is shown in Figure 4.1.

In order to estimate the charge flip background a data-driven technique is used. The probability of charge misidentification ϵ_{misrec} is measured in a control region dominated by $Z \to e^+e^-$ events. To estimate the charge flip background in the signal region the misidentification rate ϵ_{misrec} is multiplied to the number of events with opposite charged leptons passing the $W^\pm W^\pm jj$-EW signal region selections except of the same charge requirement.[1]

In this thesis the charge flip electrons are not included in the non-prompt background and are therefore mostly treated just like prompt leptons.

4.2.1 Electron Charge ID Selector

In order to reduce the impact of charge misidentification on the signal region one has the choice of several possible selections. In the ATLAS review of same charged $W^\pm$ boson scattering of 2019 a Z-peak veto of $|m_{ee} - 91.2 \text{ GeV} < 15 \text{ GeV}|$ and a selection on the electron pseudorapidity $|\eta| < 2.5$ was applied on all events with two electrons in order to reduce the effects of charge flip.[1] Nevertheless a significant contribution of the charge flip background was still observed. A single variable is not sufficient to make a decision whether a lepton performed charge flip or not. Therefore a boosted decision trees (BDT) tool is used in the current analysis. This machine learning technique is used to separate signal and background events and is more advanced than the usual cut based selections. The Electron Charge ID Selector (ECIDS) was trained taking electrons with correct reconstructed events as signal and electrons with misidentified charge as background. The training was performed for $Z \to ee$ events in data-sets taken in the years 2015-2017 with a total integrated luminosity of 80fb^{-1}.[12] In the previous ECIDS version the selection of input variables for the BDT training was optimized to eight variables but the full optimization was not realistic because it was performed on a too large number of combinations.[15] The current version of the ECIDS tool was trained on six variables:[12][17]

- $q \cdot d_0$: Product of electron charge and its transverse impact parameter
- E/p: Fraction of energy deposit in calorimeter and track momentum
- SCT-weighted average charged
- $\Delta\phi_{\text{Rescaled}}$: Angle between the energy deposit in the second layer of the electromagnetic calorimeter and the electron track. The track momentum is rescaled to the deposit energy before extrapolation.
- $|\eta|$ of the lepton
- p_T of the lepton

In addition the sample dependence of the input variables was taken into account during the training since the ECIDS tool should be applicable to any sample.[15] Especially the $|\eta|$ and p_T distributions of a lepton are very dependent on the underlying physical process. For this six input variables the BDT configurations were optimized to 500 decision trees, a maximal tree depth of 5 and a minimal number of events in the last nodes of 0.5%. After the optimization of the input variables and the BDT configurations the ECIDS tool was trained for a signal efficiency of at least 97% and thereby a BDT output of -0.337671. Due to the BDT training the ECIDS tool is superior to normal cuts (like a Z peak veto) in identification of charge flip events. Therefore the ECIDS tool is used in the following analysis to classify events with charge misidentification on reconstruction level.[12][17]

Chapter 5

Monte Carlo Simulation

5.1 Monte Carlo Event Generator

One of the most important and powerful tools of modern particle physics is the simulation of processes. They are needed for the detailed planning of experiments as well as for the design of the final data analysis. Beyond that, simulation has a very high importance especially for particle physics, since most predictions of the Standard Model can not be derived analytically. Without going into the details of the event generators, this section is intended to provide a short overview of the principles of operation of Monte Carlo event generators.

In order to fully simulate a collision process at the LHC by event generators, this process has to be separated into processes of different energy regimes. This factorization is done according to the energy scales on which the particles interact with each other.[18] The highest interaction energy is given on the collision point of the two protons. Since the exact solution of the Standard Model matrix element would require an infinite number of feynman diagrams, the matrix element is computed only to a few orders of perturbative theory. Thereby various approximations must be taken into account and one has to deal with many divergence. To calculate the transition probability from the collision to this first intermediate state the relevant matrix elements has to be intergrated over the available phase space. For this integration Monte Carlo techniques are used which rely on the concept of evaluating the integral at a large set of random points in the phase space which have a density directly related to the respective physical process. In the case of a proton-proton collision, this density distribution is for example partly given by the parton distribution function of the incoming partons. Simulations of the hard process are classified by the maximal number of vertices taken into account by computing the matrix elements. Most common are generators at leading order (LO) and next leading order (NLO). If the first step of the QCD showering is also included in the simulation of the hard process this is referred to as QCD NLO.[19]

The second step of the factorization is given by the parton shower phase. Thereby QED and QCD shower processes are simulated for the objects coming from the hard process. The modeling of parton showers is done until one reaches a low momentum scale. The simulation of the showering is followed by the hadronization models. Due to their confinement the partons are forming hadron, which can be seen in the detector. In addition to the objects originating from the hard process also secondary interactions of the proton remnants are simulated. This underlying process introduces a contamination into the already simulated process. The last step of the factorization is done for unstable particles which are decaying directly after the collision. Therefore the generation of the full process is given by a factorization into the hard process, the parton shower, the hadronization, the underlying event and finally the decay of unstable particles.[19]
Since the generated event is directly given by the resulting particles and no detector effects are taken into account, these events are classified as truth level events. Various Monte Carlo generators such as Sherpa, Powheg and Madgraph are available to simulate these truth level events. Sometimes for the same process different generators are used for the hard pro-

cess and the following shower. A full list of the used Monte Carlo samples can be found in Table Appendix A.2.

In order to compare the theory prediction represented by the Monte Carlo generated truth level events with actual data, also the particle interactions with the detector have to be simulated. Thereby the expected energy deposits in the detector are simulated for all objects originating from the primary event and also from pileup events along the beam axis. Afterwards the same processing chain is performed as for real data to reconstruct the objects from the given detector entries. Finally, scale factors are applied on the events in order to reduce mismodeling. Unlike the truth level events, these reco level events are directly comparable with data. In the rest of this thesis, events simulated by Monte Carlo event generators are referred to as Monte Carlo events.

5.2 Truth Level Lepton Classification

As already mentioned, the subtraction of prompt events in the extraction of the fake factor in Equation 4.10 and its application in Equation 4.7 is done using data from Monte Carlo simulations. The truth level information itself is available for the Monte Carlo simulated events, but the objects still have to be correctly classified based on this information. The truth level splitting into the prompt, charge flip and non-prompt category is not trivial and has to be done in a consistent way with other studies performed for same charged $W^{\pm}$ boson scattering analysis. Furthermore, in order to define the control region it is necessary to be able to compare its composition of non-prompt objects with that of the signal region, which requires a more detailed analysis of the origin of these objects. In the course of this thesis the truth level analysis is done using the MCTruthClassifier and IFFTruthClassifier.

5.2.1 MCTruthClassifier

The MCTruthClassifier is an ATLAS software framework for truth level classification of a particles type, origin and outcome process. For example electrons, muons, taus and photons are mainly categorized into the type classifications Isolated, None Isolated and Background. Isolated leptons are in general the prompt leptons originating $W^{\pm}$, Z and H bosons. Particles from decays of charmed and b-flavored baryons and mesons are classified as None Isolated and the rest belongs to the Background category. If there are necessary truth information missing or the classification just fails the object is classified as Unknown. There are in total 39 categories of object types, 46 different origins and 12 particle outcomes, which make the MCTruthClassifier very detailed. Therefore it is well suited to analyze the composition and the origins of the faking objects. The decision whether a particle was charge flip or non-prompt cannot be derived trivially from the output of this tool, since the charge of the reco level particle and the origin and charges of the preceding particles must also be considered.[20]

5.2.2 IFFTruthClassifier

Based on the MCTruthClassifier the IFFTruthClassifier was developed to classify electrons and muons according to their type and origin. For this purpose the tool needs the type and origin of the particle and the type origin and PDG-ID of its first and last mother particle. In comparision to the MCTruthClassifier the IFFTruthClassifier has a reduced set of output categories, which makes the splitting into prompt, charge flip and non-prompt much easier. The classes provided by this tool are:[21]

- Unknown: The tool was not able to assign this object to any of the other categories. In a working setup no particle should be in Unknown.
- KnownUnknown: In principle the object could be classified, but some necessary information about this object is missing.

- IsoElectron: Prompt electron directly from $W^{\pm}$ or Z decay or emitted by a so called FSR photon or bremsstrahlung
- ChargeFlipIsoElectron: Electrons, which in principle could be categorized as prompt, but are affected by charge flip effects. Electrons originating from FSR photons are not included in this class.
- PromptMuon: Prompt muon
- PromptPhotonConversion: Electrons originating from the conversion of prompt photons without any distinction of external and internal conversion
- ElectronFromMuon: Electrons emitted by muons
- TauDecay: This class contains electrons and muons originating from tau decays and also electron-positron pairs created by photons which are radiated by a tau lepton.
- BHadronDecay: Leptons originating from heavy-flavor decay of b baryons
- CHadronDecay: Leptons originating from heavy-flavor decay of c baryons
- LightFlavorDecay: Leptons directly or indirectly produced by light-flavor jets

The classes given by the IFFTruthClassifier are now used to categorize reco level electrons and muons into prompt, charge flip and non-prompt:

- Prompt: IsoElectron, PromptMuon, PromptPhotonConversion. Also the Unknown and KnownUnknown are included in the prompt category in the context of this thesis to exclude them from the study of non-prompt background. The category Unknown was proven to be empty and the influence of KnownUnknown has been found to be extremely small
- Charge flip: ChargeFlipIsoElectron
- Non-prompt: ElectronFromMuon, ElectronFromMuon, TauDecay, BHadronDecay, CHadronDecay, LightFlavorDecay

Chapter 6

Dilepton Control Region

The Run 2 data and Monte Carlo samples studied in this thesis are listed in Table Appendix A.1 and Appendix A.2. No systematic variations are used and thus all presented uncertainties result only from the statistical uncertainty of the data and Monte Carlo simulated events.

6.1 The Non-prompt Control Region

In order to calculate the fake factors a control region enriched with non-prompt lepton is needed. In this thesis a control region with one Ana selected lepton and at least one additional Ana/Non-Ana lepton is used. The advantage of this multi-lepton control region in contrast with the single-lepton control region used in the previous publication [1], is that it can be adjusted kinematically closer to the signal region. The requirement of at least one Ana lepton additionally offers consistency with the neglection of events with two Non-Ana leptons in the fake factor application described in section 4.1.3. To define the multi-lepton control region, the following criteria were applied:

- Statistical significance: As shown in Equation 4.10, to calculate a fake factor the number of prompt leptons is considered as contamination and subtracted from the total number of Ana/Non-Ana leptons. To achieve minimal uncertainty, the control region must contain a large number of total events and have a sufficient purity in non-prompt events.

- Composition of the non-prompt leptons: In order to ensure that the fake factors in both regions are consistent, the non-prompt composition of the control region will be compared with that one of the signal region. One way to evaluate this is to compare the respective non-prompt proportions of the used Monte Carlo samples. Another option is to check the non-prompt composition directly with the results of the truth level lepton classifiers.

- Orthogonality to the signal region: To ensure that extraction and application of the factor are independent of each other, there must be no overlap between the control region and the signal region.

- Data modeling: Due to the subtraction of the prompt leptons it is quite important that the data in the control region are sufficiently well modeled by the Monte Carlo events.

Based on these criteria, the final selection flow of the control region is going to be motivated step by step. After each selection criteria was introduced a number of further possible selections were examined for their suitability. In order to present a compact and clear overview of the control region definition, only the selection that was finally used will be presented in the following.

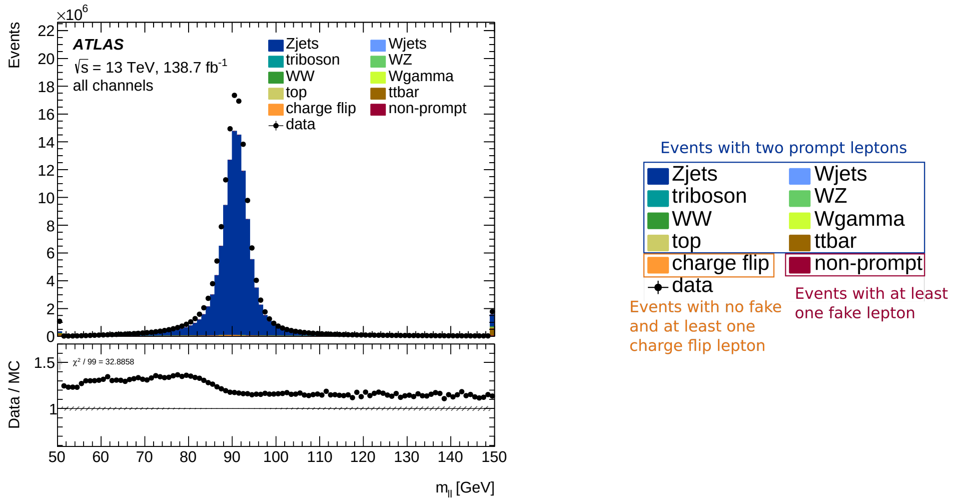


Figure 6.1: Invariant dilepton mass m_{ll} of the two leading leptons in a region with one Ana and one additional Ana/Non-Ana lepton and no further selections. As it is shown on the right hand side, the data is compared to Monte Carlo events splitted into non-prompt, charge flip and prompt events. The prompt events are also splitted into the Monte Carlo samples they originate from.

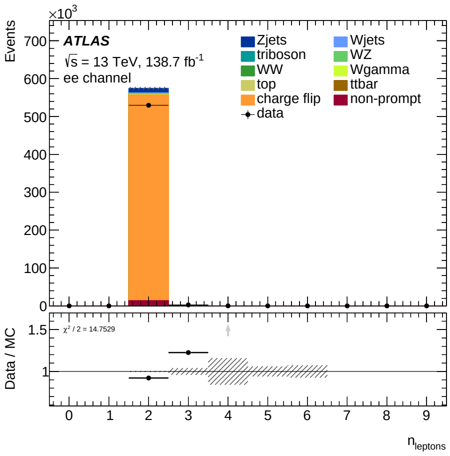
6.1.1 Selection Based on Lepton Charge

In order to find suitable selections for the control region optimization, the contribution of non-prompt events in different distributions is studied. Therefore the events are splitted in different categories. Events with at least one non-prompt lepton are defined as non-prompt events and analogously events with no non-prompt but at least one charge flip lepton are classified as charge flip events. Prompt events exclusively consist of prompt leptons and are splitted into the individual processes to show their physical origin. In Figure 6.1 the m_{ll} distribution is shown for the multi-lepton region with one Ana lepton and at least one additional lepton. This histogram indicates that this region is absolutely dominated by prompt events coming from $Z+jets$ processes. The Z boson decays into an opposite charged pair of electrons or muons which are very likely to be Ana or Non-Ana selected due to their Z boson origin. To suppress this prompt contribution a same charge selection is applied on the two leading leptons.

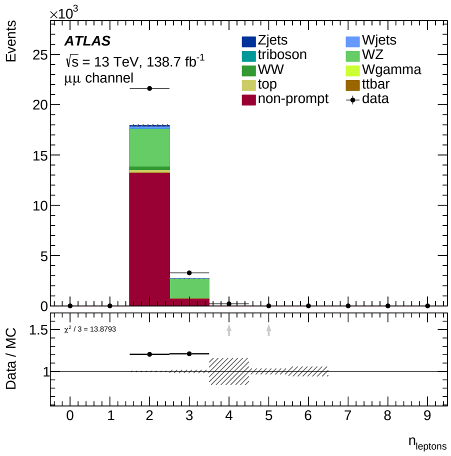
6.1.2 Selection Based on Number of Ana/Non-Ana Leptons

The number of Ana/Non-Ana leptons are shown in Figure 6.2 after the same charge selection splitted in channels. The channel splitting is done according to the flavor of the leading and subleading lepton. E.g. the $e\mu$ channel are all events where the leading lepton was an electron and the subleading lepton was a muon. Due to the same charge selection the relative contribution of non-prompt and charge flip events increased a lot. The ee channel is still dominated by events coming from $Z+jets$ processes but now in the form of charge flip events. The problem of charge flip events will be handled on a later step of the control region definition.

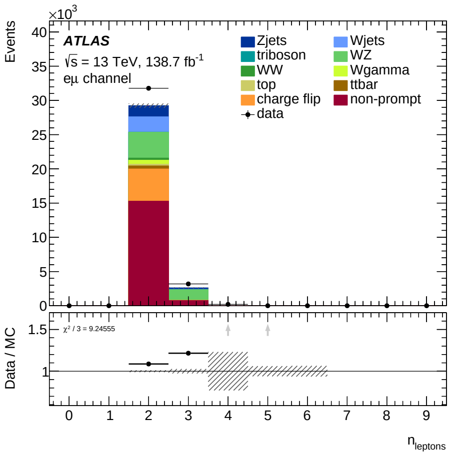
At this point a cut to the number of leptons n_{lep} is appropriate to reduce the impact of prompt WZ events in all channels and facilitates the further analysis of the control region, because then only the leading and subleading lepton have to be considered and all events have the same number of leptons.



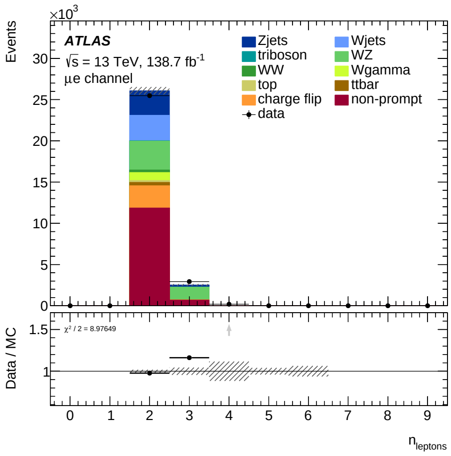
(a) n_{lep} in the ee channel



(b) n_{lep} in the $\mu\mu$ channel



(c) n_{lep} in the $e\mu$ channel



(d) n_{lep} in the μe channel

Figure 6.2: Number of Ana/Non-Ana selected leptons splitted in the lepton channels after the same charge selection was applied on the two leading leptons. The data is compared to Monte Carlo events splitted into non-prompt, charge flip and prompt events. The prompt events are also splitted into the Monte Carlo samples they originate from.

6.1.3 Orthogonality to the Signal Region and Low Dijet Invariant Mass Control Region

The next step of my analysis is to make the region for the fake factor extraction orthogonal to the signal region and the LowMjj control region already used by the analysis. This LowMjj region is used as the validation region for the non-prompt study. Therefore, the control region must be disjoint with this LowMjj region defined in Table Appendix A.5, which differs only in the $m_{jj} < 200$ GeV) selection to the signal region.

One way to achieve this is to invert a single common cut of the signal and LowMjj region. The selection of suitable cuts was narrowed down with a closer look at the respective distributions. E.g. the inversion of the cut on the dilepton mass $m_{ll} > 20$GeV would discard nearly all events and the control region would become unusable. Only the selections on the jet kinematic seemed to be viable to become disjoint to the signal and LowMjj region and these selections can also be easily combined with a cut on the number of jets. Thus the remaining orthogonal cuts to be tested are $(n_{jet} < 2)\|(n_{bjet} > 0)$, $(n_{jet} < 2)\|(|\Delta y_{jj}| < 2)$ and $(n_{jet} < 2)\|(m_{jj} < 200$ GeV$)$.

Another possibility to invert a specific selection is, to explicitly request that the events are not within the signal and LowMjj region. Thereby all events are discarded, which pass all of the selection criteria of the signal region or of the LowMjj region. By defining the control region in this way, the number of events kept is maximized and the region is orthogonal to all signal and control regions used by the analysis.

These four possible selections are now compared based on the number of passing events, the purity in non-prompt events and the expected relative statistical uncertainty of the non-prompt contribution $\Delta(N - N^{prompt})/(N - N^{prompt})$.

$$\frac{\Delta(N - N^{prompt})}{N - N^{prompt}} = \frac{\sqrt{N + N^{prompt}}}{N - N^{prompt}} \tag{6.1}$$

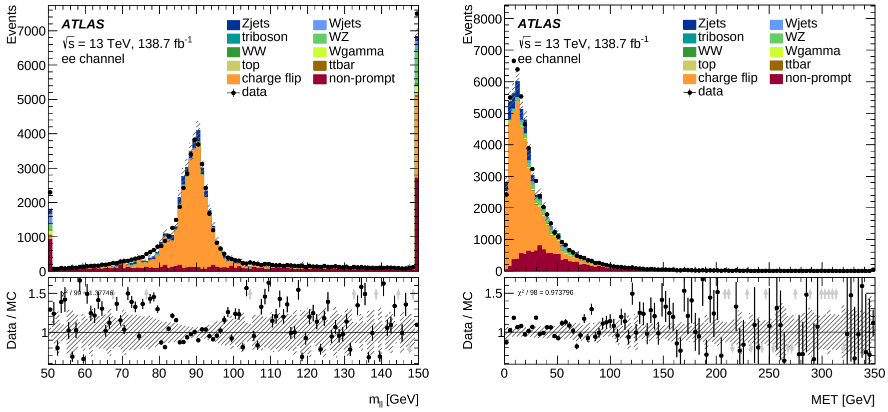
The results of this comparison can be found in Table Appendix A.6. All of the selections on the jet kinematic discards at least 6% of the total number of events without any positive effect on the contribution of non-prompt events. On the other hand, explicitly demanding all events which pass selection of the signal or LowMjj region has nearly no effect on the current region. Therefore, this explicit selection is preferred because it allows maximum further optimization of the control region and the other orthogonal cuts could be applied later without any additional losses.

6.1.4 ECIDS Selection for Electrons

The control region as defined with the selections described in the previous sections is now orthogonal to the signal and LowMjj region but is still dominated by a huge amount of charge flip events. In order to reduce the impact of charge flip electrons the ECIDS selection is applied to all electrons on event level. If one electron fails the ECIDS criteria the whole event is going to be rejected. More information about the BDT trained ECIDS selection can be found in section 4.2.1.

6.1.5 Invariant Dilepton Mass and Transverse Missing Momentum in the ee Channel

Although the ECIDS selection is very effective in detecting charge flip electrons, it only reduces the number of these events in the ee channel to about a half. Since the charge flip electrons are primarily originating from Z decays, it is expected that these events will have a peak in their dilepton mass m_{ll} close to the Z boson mass. Unlike leptonic $W^{\pm}$ decays, Z decays are not expected to produce a neutrino in the final state, so these events have also a lower missing transverse momentum E_T^{miss}. Therefore, these charge flip events are sensitive to both the invariant dilepton mass and the missing transverse momentum as shown in the Figure 6.3. To achieve the optimal non-prompt contribution, the cuts on these two variables were not chosen independently of each other. The two distributions were studied due to a



(a) m_{ll} in the ee channel

(b) E_T^{miss} in the ee channel

Figure 6.3: Invariant dilepton mass m_{ll} and missing transverse momentum E_T^{miss} after the ECIDS selection was applied on events with electrons. The data is compared to Monte Carlo events splitted into non-prompt, charge flip and prompt events. The prompt events are also splitted into the Monte Carlo samples they originate from.

$m_{ll} \times E_T^{miss}$ matrix in the ee channel to find the best cut combination on m_{ll} and E_T^{miss}. This optimization was done only for events with two Ana leptons in the final state, because especially these events have a poor non-prompt statistics due to the tight selection of both leptons.

This study concluded that a minimum of the relative statistical uncertainty $\Delta(N - N^{prompt})/(N - N^{prompt})$ is achieved in the ee canal with a Z peak veto between 79 GeV and 98 GeV and a selection on the missing transverse momentum at $E_T^{miss} > 19$ GeV.

6.1.6 Rejection of Events with Veto Leptons

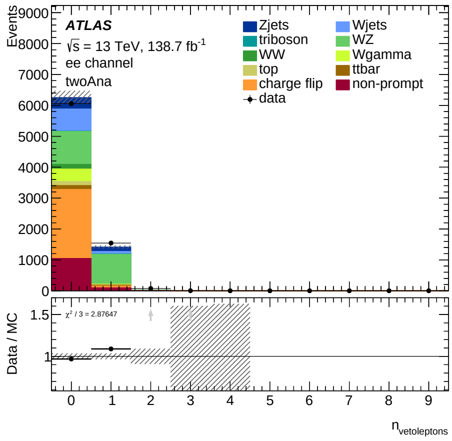
Due to the previous optimizations, the current region with a non-prompt fraction of 59.2% seems sufficient for the extraction of the fake factors. However, cases with two Ana leptons are still problematic because they are dominated by prompt events. In the case of two Ana leptons the main cause for the huge prompt contribution are WZ events where two of the three prompt leptons are Ana selected and the third one is lost in the detector or just fails the Ana/Non-Ana selection. As can be seen in Figure 6.4, this third lepton is often classified as a Veto lepton. Therefore, a cut on the number of Veto leptons would have a positive effect on the non-prompt statistics, especially for the cases with two Ana leptons.

6.1.7 Jet Multiplicity Selection in the $\mu\mu$ Channel

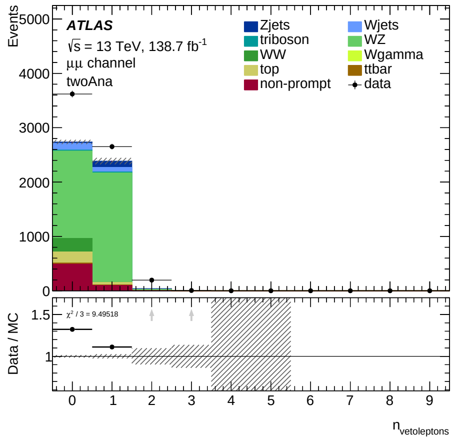
A data/Monte Carlo comparison of the control region defined up to now shows a relatively good agreement, nevertheless in the $\mu\mu$ channel a significant deviation was found. Figure 6.5a shows the jet multiplicity for the $\mu\mu$ channel, the mismodeling corresponds mainly to events with a low jet multiplicity. In order to improve the agreement of data and Monte Carlo all events without a jet are rejected in the $\mu\mu$ channel. Another advantage of this selection is that, as shown in Figure 6.5b, the significance of the non-prompt contribution in events with two Ana selected muons is also increased.

6.1.8 Final Dilepton Control Region

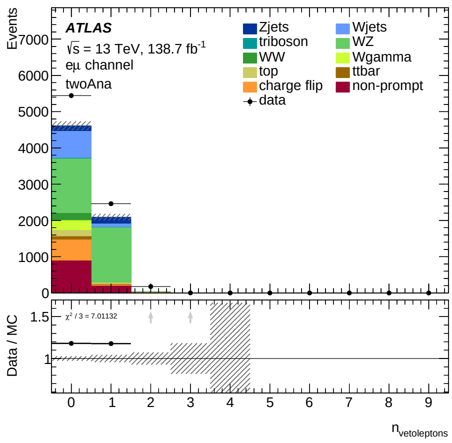
The selection described so far define the control region as it will be used in the rest of the thesis. Another promising candidate was also a cut on the missing transverse momentum also beyond the ee channel. This selection was optimized and tested for events with two Ana selected leptons in each channel separately, but had no significant effect on



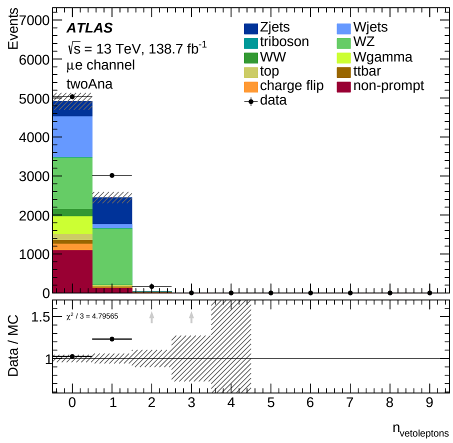
(a) $n_{vetolep}$ in the ee channel



(b) $n_{vetoLep}$ in the $\mu\mu$ channel

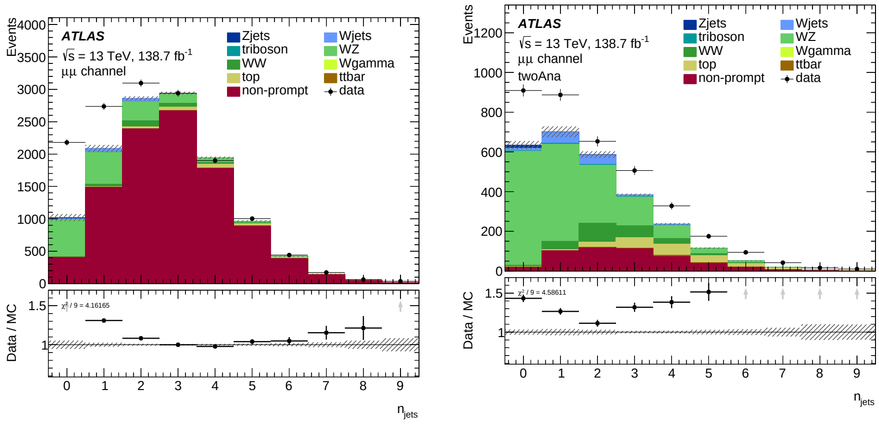


(c) $n_{vetoLep}$ in the $e\mu$ channel



(d) $n_{vetoLep}$ in the μe channel

Figure 6.4: Number of Veto leptons for events with two Ana leptons after the cut on m_{ll} and E_T^{miss} was applied in the ee channel. The histograms are splitted in the lepton channels. The data is compared to Monte Carlo events splitted into non-prompt, charge flip and prompt events. The prompt events are also splitted into the Monte Carlo samples they originate from.



(a) n_{jet} in the $\mu\mu$ channel

(b) n_{jet} in the $\mu\mu$ channel for events with two Ana leptons

Figure 6.5: Number of jets in the $\mu\mu$ channel after the cut on the number of Veto leptons was applied. On the right hand side the distribution is shown only for events with two Ana selected leptons. The data is compared to Monte Carlo events splitted into non-prompt, charge flip and prompt events. The prompt events are also splitted into the Monte Carlo samples they originate from.

Exactly two Ana/Non-Ana leptons with identical electrical charge One Ana lepton No Veto leptons
Electrons has to pass ECIDS
Fail the selection of the $W^{\pm}W^{\pm}jj$-EW signal region and the LowMjj region
In ee channel: $(m_{ll} < 79\text{ GeV})\|(98\text{ GeV} < m_{ll})$ and $E_T^{miss} > 19$ GeV
In $\mu\mu$ channel: $n_{jet} > 0$

Table 6.1: The event selection of the dilepton control region, which is used in this thesis to extract the fake factor

the non-prompt contribution. E.g. in the μe channel the optimal selection would be at $E_T^{miss} > 6$ GeV but for events with two Ana leptons this selection has just a very small improvement in the purity of non-prompt events from 21.86% to 22.6% and the uncertainty $\Delta(N - N^{prompt})/(N - N^{prompt})$ decreases just from 8.67% to 8.53%. Since this improvements are extremely small, an additional cut on the missing momentum was not reasonable.

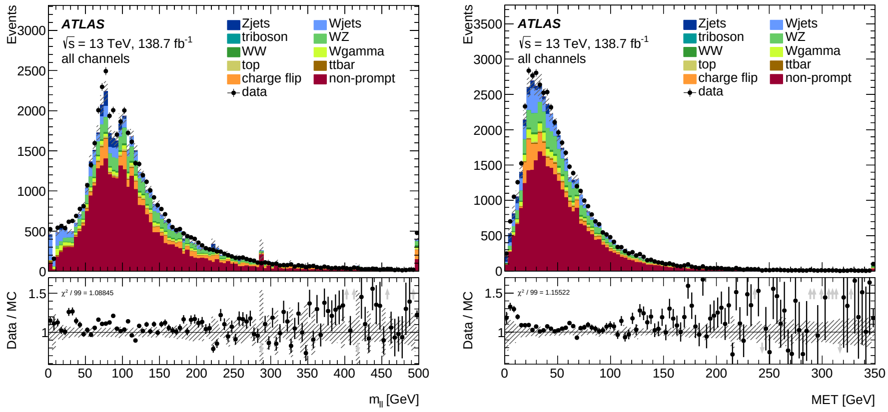
Other distributions were also examined for possible selections, but no other suitable candidate was found. Therefore the final definition of the control region is summarized in table 6.1.

6.2 Validation of the Control Region

In order to evaluate the control region, the criteria defined at the beginning of this chapter are now taken up again. The control region has to contain a significant contribution of non-prompt leptons and the composition of this leptons should be as close as possible to the non-prompt composition of the signal region. Since the prompt leptons are subtracted by Monte Carlo one needs a sufficient data modeling in the control region.

6.2.1 Statistical Significance

The selection criteria of the control region were chosen to minimize the statistical uncertainties on the number of non-prompt events. The selection flow of the control region is shown in



(a) m_{ll} distribution

(b) E_T^{miss} distribution

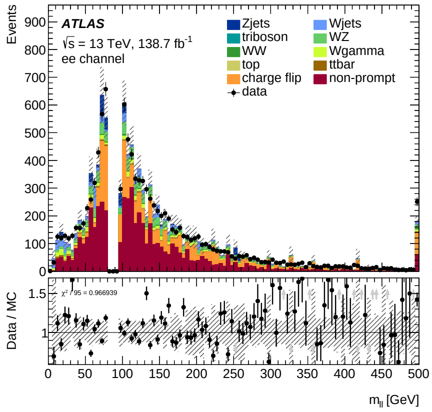
Figure 6.6: Invariant dilepton mass m_{ll} and missing transverse momentum E_T^{miss} in the control region. The data is compared to Monte Carlo events splitted into non-prompt, charge flip and prompt events. The prompt events are also splitted into the Monte Carlo samples they originate from.

Table Appendix A.7 together with the respective non-prompt ratio and the relative statistical uncertainty of the non-prompt contribution, which is described in Equation 6.1. Although the impact of the individual selections on the non-prompt uncertainty is rather small, the final non-prompt ratio in each channel is in the range of 50% to 85% and the respective uncertainty is about 1% to 2%. This suggests at first that the choice of cuts leads to a very small final uncertainty on the fake factor. For events with two Ana leptons the non-prompt ratio is much lower, which is also shown in the selection flow in Table Appendix A.8. For events with two Ana leptons the uncertainties are about 10% after applying all selections, which shows that the uncertainty of the fake factor will be dominated by its numerator. That this uncertainty can be reduced by further selections is assumed to be quite unlikely, since the number of non-prompt events with two Ana leptons is already quite low and thus further selections must be very successful in rejecting prompt events to improve the non-prompt contribution. Therefore, the selection of cuts is considered to be sufficient to increase the statistical significance of the non-prompt events.

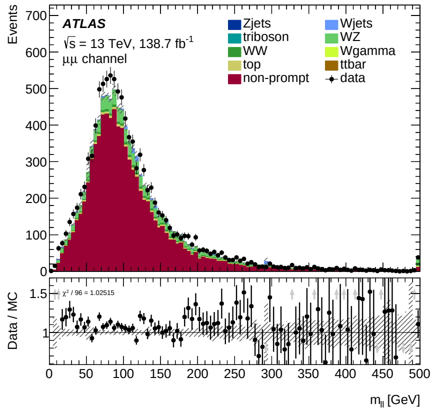
6.2.2 Data Modeling

Since the control region still has a significant prompt contribution, it is of great importance that the data modeling is as accurate as possible. The agreement between the Monte Carlo simulation and the data was analyzed in several distributions and found to be sufficiently good, as shown in Figure 6.6.

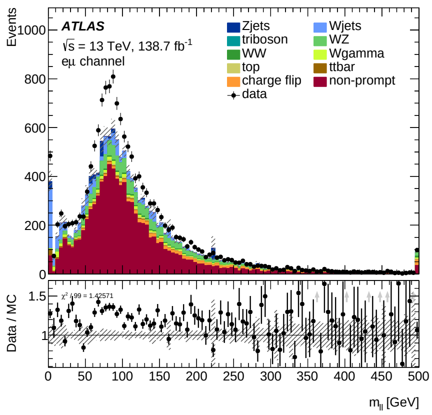
However, Figure 6.7 indicates that the agreement between Monte Carlo data and actual data differs significantly between the lepton channels. In the $\mu\mu$ channel is a slight deviation of m_{ll} next to the Z mass and in the $e\mu$ channel the data distribution in m_{ll} differs significantly from the Monte Carlo data. Finding the exact origin of this underestimation would exceed the scope of this thesis, but it can be legitimately assumed that it could be a result of the mismodeling of the non-prompt contribution. Therefore, the agreement between the Monte Carlo simulation and the actual data seems acceptable even despite the clear underestimation in the μe channel.



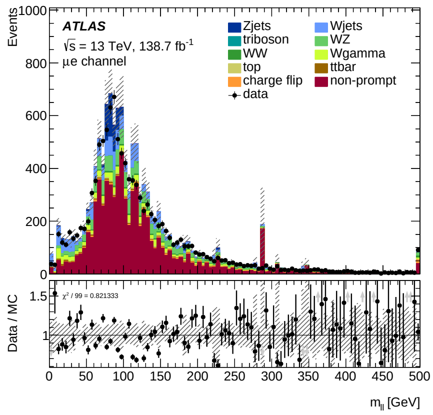
(a) m_{ll} distribution in ee channel



(b) m_{ll} distribution in $\mu\mu$ channel



(c) m_{ll} distribution in $e\mu$ channel



(d) m_{ll} distribution in μe channel

Figure 6.7: Invariant dilepton mass m_{ll} in the control region splitted in lepton channels. The data is compared to Monte Carlo events splitted into non-prompt, charge flip and prompt events. The prompt events are also splitted into the Monte Carlo samples they originate from.

6.2.3 Non-prompt Composition

The next step to corroborate the validity of the control region is to compare the non-prompt composition of the control region with the signal region, there are several approaches that could be taken. One simple way to analyze the composition is to check the non-prompt contribution of the Monte Carlo samples, which represents different physical processes. As can be seen in Table 6.2, the non-prompt leptons of the signal and control region originate primarily from $W + jets$ and $t\bar{t}$ events. The sample composition of the non-prompt events is in all channels in good agreement between the two regions within their Monte Carlo uncertainties. Such a comparison was also made separately for the events with two Ana leptons and events with one Ana and one Non-Ana lepton, which is shown in Table Appendix A.9 and A.10. This separate evaluation is of interest, because for this method to be valid, the fake factor in signal and control region should be equivalent and thus especially the non-prompt composition of the numerator and denominator should agree. As can be seen in Table Appendix A.10, the sample composition for the events with a Non-Ana lepton is in a very good agreement. For events with two Ana leptons there are significant disagreements in the μe and $e\mu$ channel, which are still within a acceptable range. The comparison of the sample composition is a good first approach to compare the non-prompt composition, but it only gives hints about the concrete origin and type of the non-prompt leptons.

A closer look at the non-prompt composition can be done on truth level using the truth

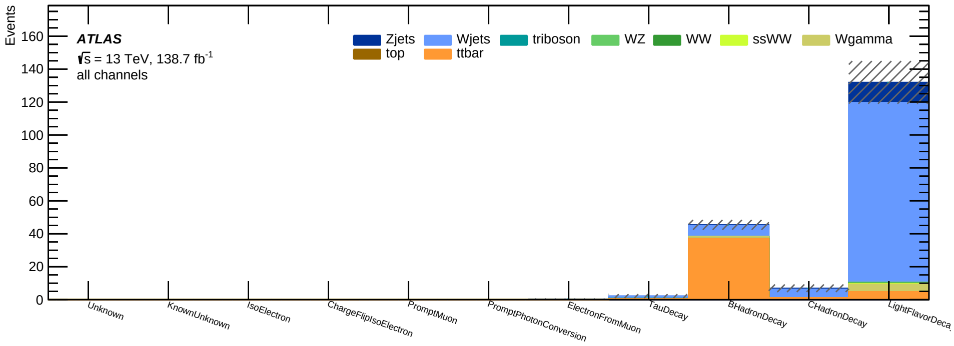
	ee	eμ	μe	$\mu\mu$	all
Signal Region:					
Zjets	$9.2 \pm 3.2\%$	$5.3 \pm 1.2\%$	$4.5 \pm 2.1\%$	$2.6 \pm 1.6\%$	$5.1 \pm 0.9\%$
Wjets	$62.6 \pm 4.4\%$	$39.0 \pm 2.8\%$	$53.3 \pm 4.0\%$	$19.8 \pm 2.5\%$	$39.8 \pm 1.7\%$
WZ	$0.4 \pm 0.1\%$	$0.1 \pm 0.0\%$	$0.1 \pm 0.0\%$	$0.0 \pm 0.0\%$	$0.1 \pm 0.0\%$
Wgamma	$0.0 \pm 1.8\%$	$2.6 \pm 1.3\%$	$3.4 \pm 2.1\%$	$0.0 \pm 0.0\%$	$1.7 \pm 0.7\%$
WZ_EW	$0.2 \pm 0.0\%$	$0.1 \pm 0.0\%$	$0.1 \pm 0.0\%$	$0.1 \pm 0.0\%$	$0.1 \pm 0.0\%$
Wt	$0.0 \pm 0.0\%$	$0.0 \pm 0.0\%$	$0.0 \pm 0.0\%$	$0.2 \pm 0.2\%$	$0.1 \pm 0.1\%$
ttbar	$27.6 \pm 3.0\%$	$52.9 \pm 1.7\%$	$38.5 \pm 2.3\%$	$77.3 \pm 1.4\%$	$53.1 \pm 1.0\%$
Other	$0.6 \pm 0.1\%$	$0.3 \pm 0.0\%$	$0.3 \pm 0.1\%$	$0.2 \pm 0.0\%$	$0.3 \pm 0.0\%$
Control Region:					
Zjets	$8.9 \pm 0.9\%$	$5.6 \pm 0.3\%$	$8.0 \pm 1.7\%$	$2.9 \pm 0.3\%$	$6.1 \pm 0.5\%$
Wjets	$57.3 \pm 1.1\%$	$30.4 \pm 0.7\%$	$50.1 \pm 1.2\%$	$13.0 \pm 0.3\%$	$35.5 \pm 0.5\%$
WZ	$0.1 \pm 0.0\%$	$0.0 \pm 0.0\%$	$0.1 \pm 0.0\%$	$0.0 \pm 0.0\%$	$0.0 \pm 0.0\%$
Wgamma	$1.3 \pm 0.1\%$	$0.8 \pm 0.1\%$	$0.9 \pm 0.1\%$	$0.2 \pm 0.0\%$	$0.8 \pm 0.0\%$
WZ_EW	$0.0 \pm 0.0\%$	$0.0 \pm 0.0\%$	$0.0 \pm 0.0\%$	$0.0 \pm 0.0\%$	$0.0 \pm 0.0\%$
Wt	$0.2 \pm 0.0\%$	$0.3 \pm 0.0\%$	$0.2 \pm 0.0\%$	$0.3 \pm 0.0\%$	$0.3 \pm 0.0\%$
ttbar	$32.2 \pm 0.6\%$	$62.7 \pm 0.3\%$	$40.6 \pm 0.7\%$	$83.6 \pm 0.2\%$	$57.3 \pm 0.2\%$
Other	$0.2 \pm 0.0\%$	$0.2 \pm 0.0\%$	$0.1 \pm 0.0\%$	$0.1 \pm 0.0\%$	$0.1 \pm 0.0\%$

Table 6.2: Comparison of signal region and control region in the composition of Monte Carlo samples the non-prompt leptons are originating from. The shown uncertainties result from the statistical uncertainties of the Monte Carlo samples.

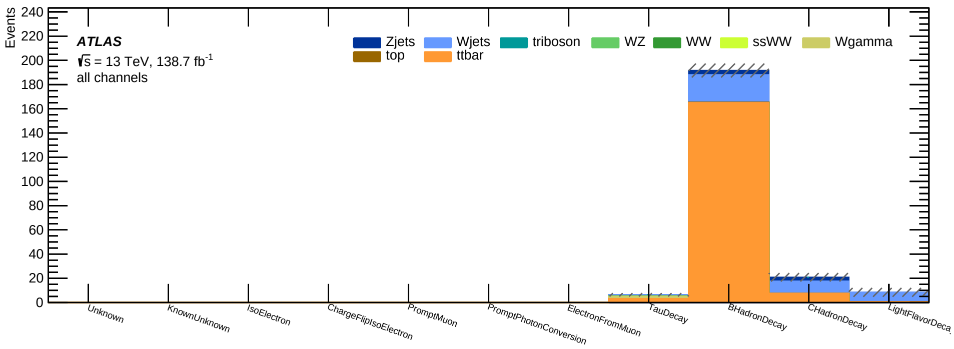
classifier. The IFFTruthClassifier is used for a first analysis, since it offers a simple but nevertheless useful classification. The composition of faking electrons and muons splitted in the samples they are originating from is shown in Figure 6.8. According to the IFFTruth-Classifier, the compositions of the non-prompt leptons within the signal region and control region are hardly distinguishable. The faking electrons originate primarily from light flavour jets of $W + jets$ events and another significant contribution comes from $t\bar{t}$ events, where a top quark decays into a bottom quark faking an electron. The muons, on the other hand, come almost exclusively from the bottom quarks originating from $t\bar{t}$ events. This analysis was again performed separately for events with two Ana leptons and one Non-Ana and one Ana lepton, as shown in Figure Appendix A.1 and A.2. It was observed that the signal and control region have a good agreement in the non-prompt composition regardless of whether the leptons were Ana or Non-Ana selected. The results of the IFFTruthClassifier thus confirm that the selected control region is kinematically very close to the signal region for non-prompt leptons.

Finally, the results of the IFFTruthClassifier will be compared with the MCTruth-Classifier. The non-prompt type and origin for events with two Ana leptons and events with one Non-Ana and one Ana lepton are shown in Figure Appendix A.3, A.4, A.5 and A.6. As expected, the output of the MCTruthClassifier also shows a very good agreement between signal and control region. However, the MCTruthClassifier is not suitable for a deeper evaluation of the physical origin of non-prompt leptons beyond that of the IFFTruthClassifier. This is illustrated by the fact that almost all non-prompt muons are just classified as NonIsoMuon and a significant amount of the origins of non-prompt electrons are in the NonDefined category. Therefore, a more precise evaluation of the output of MCTruthClassifier is not carried out.

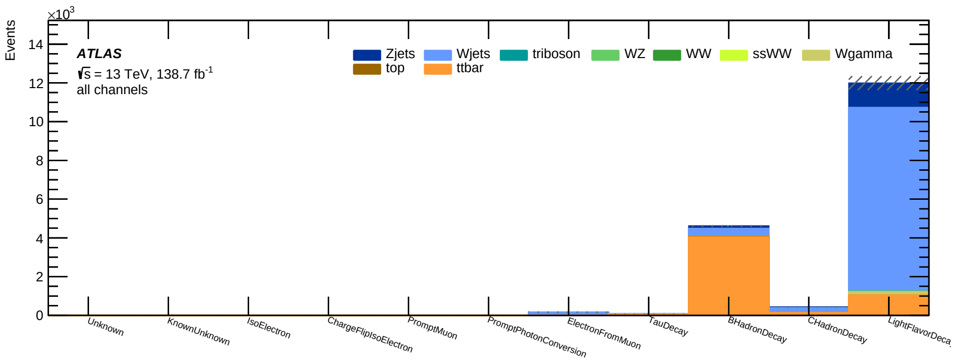
In summary, the building process of a control region was shown, which fulfills all the requirements to be suitable for the fake factor extraction. The data in this control region appears to be reasonably modeled by the Monte Carlo simulation and the composition in terms of origin of non-prompt leptons is in very good agreement between the signal and control region. The statistic of non-prompt events with two Ana leptons could become a limiting factor, since the low statistic will lead to a high uncertainty in the fake factor numerator that cannot be reduced with this control region definition.



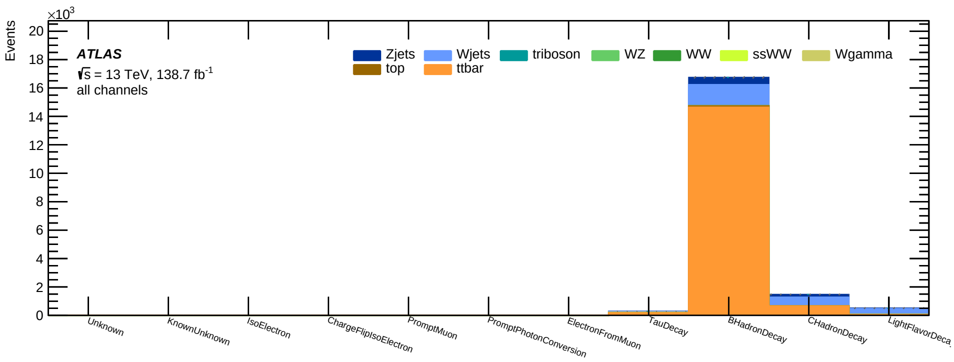
(a) Classification of non-prompt electrons in the signal region according to the IFFTruthClassifier



(b) Classification of non-prompt muons in the signal region according to the IFFTruthClassifier



(c) Classification of non-prompt electrons in the control region according to the IFFTruthClassifier



(d) Classification of non-prompt muons in the control region according to the IFFTruthClassifier

Figure 6.8: Comparison of the signal and control region in the classification of non-prompt electrons and muons according to the IFFTruthClassifier. The number of leptons is splitted in the Monte Carlo samples they originate from.

Chapter 7

Extraction of the Fake Factor

Since the chosen selection criteria are suitable for a non-prompt control region the next step is to extract the fake factor from this control region. The extraction of the fake factor was defined in Equation 4.10 as the ratio of Ana and Non-Ana selected leptons.

Unfortunately, this extraction is only trivial for control regions with just one lepton, which must be taken into account. Since the control region used in this thesis consists of two leptons, it is not clear according to which lepton the Ana/Non-Ana splitting has to be done, how one should perform the prompt subtraction and which lepton p_T is used for the fake factor binning. In this chapter, different approaches to solve this problem are presented and compared in order to ensure an optimal fake factor extraction. It is important to point out that when discussing prompt subtraction, the charge flip leptons are also subtracted, since they are not included into this study of non-prompt events.

For the p_T binning of the fake factor the binning into [27 GeV, 35 GeV), [35 GeV, 55 GeV) and [55 GeV, inf) is chosen where the last bin contains all events with a transverse momentum higher than 55 GeV. This p_T binning is taken from the current fake studies in the $W^{\pm}W^{\pm}jj$-EW analysis.[22] For compatibility with the other studies no optimization of the binning is performed.

7.1 Unbiased Fake Factor Extraction

The problem with selecting one of the leptons to extract the fake factor is that this can lead to a strong bias of this factor. To quantify the degree of bias a technical Monte Carlo closure test is performed for the control region. Thereby the fake factor is extracted and also applied for Monte Carlo simulated events in the control region. To emphasize that the data-driven method was not applied to actual data but only to Monte Carlo simulated events this is called ”MC-data-driven” in this thesis. More details about the technical Monte Carlo closure test can be found in Chapter 8. If the fake factor method is unbiased the MC-data-drivenly estimated non-prompt background has to match exactly with the number of non-prompt events with two Ana leptons. Using this criterion, a fully unbiased fake factor extraction method can be defined in the following.

Just as in the data-driven method presented in section 4.1.3, for the technical Monte Carlo closure test the fake factor is applied on events with one Non-Ana lepton considering the transverse momentum of the Non-Ana lepton. When the unbiased fake factor F_{unb} is applied in the control region, the results must exactly match with the number of non-prompt events with two Ana leptons.

$$N_{AA}^{fake}(p_T) \stackrel{!}{=} (N_{AN}(p_{T;\text{Non-Ana}}) - N_{AN}^{prompt}(p_{T;\text{Non-Ana}}))F_{unb}(p_T) \tag{7.1}$$

Events with two non-prompt leptons are not considered in this thesis, since the probability that one object is faking an Ana lepton and at the time another object is faking an Ana or Non-Ana lepton is expected to be quite small. Therefore, the prompt subtraction is done in the way that all events are subtracted except the events where the lepton considered for the

	ee	eμ	μe	$\mu\mu$	all
MC PF	533.8	241.9	785.0	305.7	1866.4
MC FP	400.6	554.7	91.8	114.8	1161.9
MC FF	0.1	0.2	0	0	0.4
Total MC	934.5	796.9	876.9	420.5	3028.7
MC-data-driven	907.9	793.5	934.7	392.2	3028.4

Table 7.1: Technical Monte Carlo closure test in the control region for the $W + jets$ and $t\bar{t}$ sample. The number of non-prompt events is splitted into the events with two Ana leptons where the leading lepton was prompt/charge flip and the subleading lepton was non-prompt (MC PF), the leading lepton was non-prompt and the subleading lepton was prompt/charge flip (MC FP) and both leptons were non-prompt (MC FF). A perfect closure is reached by the MC-data-driven approach with the unbiased fake factor in the sum over all lepton channels if one ignores the events with two non-prompt leptons.

fake factor extraction was non-prompt and at the same time the other lepton was prompt or charge flip. Hence, only events are kept that include exactly one non-prompt lepton and this lepton is used for the fake factor extraction.

The number of non-prompt events with two Ana leptons N_{AA}^{fake} can be splitted in the number of events where the leading lepton was non-prompt and where the subleading lepton was non-prompt.

$$N_{AA}^{fake}(p_T) = N_{AA}^{leadFake}(p_{T;lead}) + N_{AA}^{sublFake}(p_{T;subl}) \tag{7.2}$$

$$N_{AA}^{fake}(p_T) = (N_{AA}^{lead} - N_{AA}^{leadPrompt}) + (N_{AA}^{subl} - N_{AA}^{sublPrompt}) \tag{7.3}$$

Based on the Equations 7.1 and 7.3 the definition of the unbiased fake factor is given by:

$$F_{unb}(p_T) := \frac{(N_{AA}^{lead} - N_{AA}^{leadPrompt}) + (N_{AA}^{subl} - N_{AA}^{sublPrompt})}{(N_{AN} - N_{AN}^{prompt})} \tag{7.4}$$

The fact that the fake factor defined by Equation 7.4 leads to a perfect closure in the total number of events was confirmed for the defined control region and the selected technical setup. In Table 7.1 a perfect closure is confirmed in the all channel if one ignores events with two non-prompt leptons. The perfect closure is not required in the each lepton channel, since the perfect closure is only possible in the total number of non-prompt electrons and non-prompt muons.

Although this fake factor is unbiased, this does not mean that this is the optimal method to extract the fake factor. The double counting of events with two Ana leptons increases the statistics in the numerator of the factor but also more prompt/charge flip events have to be subtracted. For this reason, it should be technically possible to define a different extraction method, which leads to a smaller fake factor uncertainty. In the following sections different approaches will be analyzed and optimized to find an extraction method with a minimal bias combined with a minimal uncertainty.

7.2 Assignment of a Tag and Probe Lepton

In order to find an extraction method which is preferable to the unbiased fake factor, suitable criteria must be chosen to select the respective lepton for calculating the fake factor. Therefore, two new lepton categories are introduced to simplify the discussion of lepton selection.

- Tag lepton: Assumed to be the prompt/charge flip lepton in this event
- Probe lepton: Assumed to be the non-prompt lepton in this event used for the fake factor calculation

When the assignment of the tag and probe lepton is done, the Equation 4.10 can be directly used to extract the fake factor. Thereby the number of events is splitted in Ana (N_A) and Non-Ana (N_N) according the probe lepton and for the p_T binning of the fake factor also the probe lepton p_T is used. The events to be subtracted N^{prompt} are all events where the probe was prompt/charge flip or the tag lepton was non-prompt.

The assignment of the tag and probe lepton must be done with care, since an unsuitable assignment can lead to a very strong bias of the fake factor. In order to clarify this difficulty a general procedure is presented in the next section, which provides highly biased fake factors.

7.3 Highly Biased Fake Factor

The main problem in selecting a single lepton for the fake factor extraction arises when the assignment is significantly differently successful for Ana and Non-Ana leptons. Let the probability that the lepton was correctly selected be p_A for Ana leptons and p_N for Non-Ana leptons. Then the fake factor calculated from this results in a deviation from the unbiased fake factor by the factor p_A/p_N.[1]

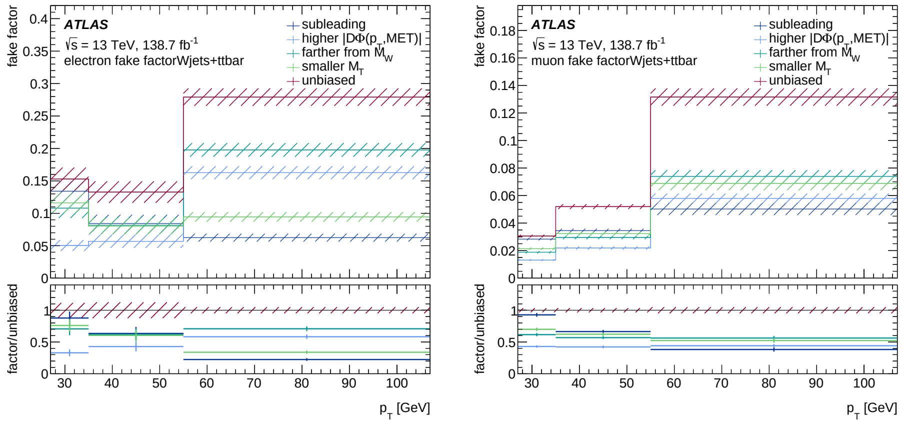
$$F = \frac{p_A \cdot N_A^{fake}}{p_N \cdot N_N^{fake}} \approx \frac{p_A}{p_N} \cdot F_{unb} \tag{7.5}$$

If the probabilities differ significantly from each other, the respective method cannot be used. This is the case, for example, if one takes for events with one Non-Ana and one Ana lepton always the Non-Ana lepton as probe and for events with two Ana leptons a physical criterion is chosen.

In order to find suitable physical criteria for the tag and probe assignment for events with two Ana leptons the origin of the prompt/charge flip leptons has to be considered. Since the non-prompt events consist mainly of $W + jets$ and $t\bar{t}$ events, it can be assumed that the prompt lepton originates from the decay of a $W^\pm$ boson. This is also confirmed by the results of the MCTruthClassifier for the prompt leptons in non-prompt events shown in Figure Appendix A.7, whereby the category "top" only additionally indicates that the $W^\pm$ boson originates from a top quark decay. Therefore, priority should be given to criteria which indicate that the tag lepton is compatible with the origin in a $W^\pm$ boson. In the following four different assignment options are listed.

- On way would be to just take the lower p_T lepton as probe. This is a very simple example of an expectedly improper assignment, since the fake factor itself is binned in p_T and so this assignment would tend to shift the events in the numerator to lower p_T values.
- A physical criterion to test whether a lepton originates from a $W^\pm$ boson would be the angle ϕ between the lepton and the missing transverse momentum. Depending on whether the $W^\pm$ boson was at rest or had a high momentum, the prompt lepton is expected to be in the opposite direction of the neutrino or close to the neutrino. At this working point we assume a high energetic $W^\pm$ boson, but in the final analysis both possibilities are tested. Therefore the lepton with smaller $|\Delta\phi(\vec{p}_T^{\,lep}, \vec{p}_T^{\,miss})|$ is chosen as probe lepton.

[1] Because of effects like double non-prompts and unlikely events where the prompt lepton is Non-Ana selected and at the same time the non-prompt lepton is Ana selected, the unbiased fake factor defined in Equation 7.4 does not perfectly match the ratio N_A^{fake}/N_N^{fake} explained here. However, since these effects are only very small, this can be neglected for the following explanations.



(a) Fake factors for electrons **(b)** Fake factors for muons

Figure 7.1: Electron and muon fake factors extracted from the $W + jets$ and $t\bar{t}$ sample. The highly biased extraction methods are compared to the unbiased fake factor. The shown uncertainties result only from the statistical uncertainty of the Monte Carlo samples.

- Since the prompt lepton and the neutrino with the corresponding missing transverse momentum originating from the $W^{\pm}$ boson one expect their transverse mass defined in Equation 2.4 to be close to the $W^{\pm}$ mass. For this reason the transverse mass of the neutrino and the probe lepton $M_T(\vec{p}_T^{lep}, \vec{p}_T^{miss})$ is farther from the mass of a $W^{\pm}$ boson.
- The last criterion refers again to the transverse mass $M_T(\vec{p}_T^{lep}, \vec{p}_T^{miss})$. This time the transverse mass M_T calculated with the probe lepton should be smaller than the transverse mass M_T calculated with the tag lepton.

For this first test it is sufficient to only use the $W+jets$ and $t\bar{t}$ samples for the fake factor extraction. Since in all channels more than 90% of the non-prompt leptons originates from these two samples, it can be assumed that this simplification has no significant impact on the results of this test.

For each of these four methods the factor is extracted for electrons and muons and is compared to the unbiased fake factor. The results of this comparison can be found in Figure 7.1. The shown uncertainties are only the statistical uncertainties of the Monte Carlo sample and so the final fake factors extracted from data will have a much higher uncertainty. As expected the fake factors calculated by the tag and probe assignment are significantly smaller than the unbiased fake factor.

If one chooses for events with only one Ana lepton always the Non-Ana lepton is probe, the probability of doing the right decision p_N is very close to one, because it is quite unlikely that the non-prompt lepton is passing the Ana selection and at the same time the prompt lepton is failing the Ana selection. For events with two Ana leptons the given physical criterion are expected to have a p_A significantly smaller than one. Therefore the fake factors are all drawn downwards with the factor of p_A/p_N and none of the four procedures is usable for a fake factor extraction.

7.4 Lepton Assignment Independent of the Ana/Non-Ana Selection

The main problem with the fake factor extraction is given by the systematic difference in the probabilities p_A and p_N. This can be solved by using the same assignment procedure for all events regardless of whether two Ana leptons or one Ana and one Non-Ana lepton were detected. Thereby the physical assignment criterion is applied to all events to select the probe lepton. As a consequence, the probability p_N goes down to p_A and thus also the statics in the fake factor denominator decreases. This effect can be tolerated, since the main statistical uncertainty of the fake factor originates from its low number of Ana leptons in the numerator. As soon as the physical selection criterion is well enough defined, the resulting fake factor can be competitive to the unbiased fake factor.

In order to implement this, the same four criteria as in the last section will be applied. In addition they are extended by their inversions, which means that the assignment of the tag and probe lepton is simply swapped. This is especially useful for the selection based on the angle to the neutrino $|\Delta\phi(\vec{p}_T^{lep}, \vec{p}_T^{miss})|$, but as it is easy to implement and increases the selection of evaluable examples, it is done for all criteria. This finally gives us eight different assignments of tag and probe leptons.
The probe lepton is the lepton with:

- Lower/higher p_T,
- Lower/higher $|\Delta\phi(\vec{p}_T^{lep}, \vec{p}_T^{miss})|$
- $M_T(\vec{p}_T^{lep}, \vec{p}_T^{miss})$ closer to/farther from the $W^\pm$ mass,
- Lower/higher $M_T(\vec{p}_T^{lep}, \vec{p}_T^{miss})$.

Since this method could really be suitable for the final analysis, it was decided to run over all Monte Carlo samples, although this is significantly more computationally intensive. This is the only way to ensure that no significant bias is introduced. In addition, the expected statistical uncertainties of the fake factors extracted from data are going to be estimated in the following, which makes the use of the complete Monte Carlo simulated events indispensable.

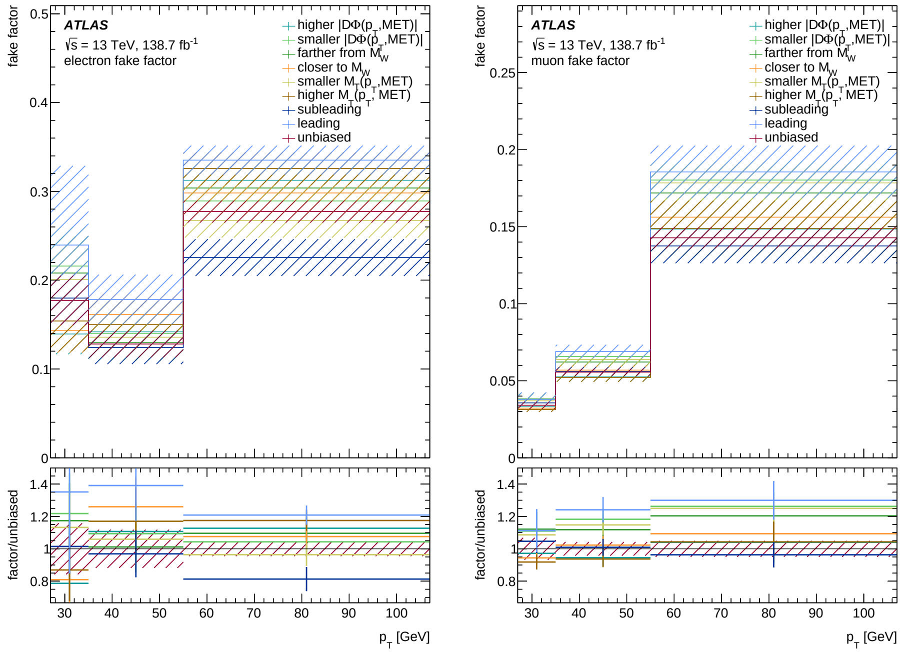
As can be found in Figure 7.2, the resulting fake factors are again compared with the unbiased fake factor. The overall bias of all assignment methods was significantly improved due to the independence to the Ana and Non-Ana selection. Nevertheless, there is a significant overestimation of the fake factor for most of the methods and surprisingly assigning the lepton with lower p_T as probe lepton seems to introduce the lowest bias in all bins except of the high p_T bin of the electron fake factor. The general overestimation of the fake factor indicates a systematic problem in the assignment process, which will be discussed in the next section, but it is worth checking whether these methods are generally competitive to the unbiased fake factor.

An assignment method is only superior to the unbiased fake factor if its result agrees with the unbiased factor within their Monte Carlo uncertainties and if it also leads to a smaller fake factor uncertainty in the final extraction from data. Since the unbiased fake factor in Equation 7.4 is written in terms of event numbers and the fake factors based on a tag and probe assignment are written in terms of numbers of probe leptons, also the final calculations of their uncertainties in Equation 7.6 and 7.7 differ. However, the two uncertainty calculations are similar, except that the double counting of events with two Ana leptons must be taken into account for the unbiased fake factor. In some cases, this double counting offers a statistical advantage as it doubles the number of leptons used for extraction. In the cases of purely prompt/charge flip events, this double counting is a clear disadvantage, since twice as much leptons have to be subtracted by Monte Carlo.

$$\frac{\Delta F_{unb}}{F_{unb}} = \sqrt{\frac{N_{AA}^{lead} + N_{AA}^{leadPrompt} + N_{AA}^{subl} + N_{AA}^{sublPrompt}}{\left(N_{AA}^{lead} - N_{AA}^{leadPrompt} + N_{AA}^{subl} - N_{AA}^{sublPrompt}\right)^2} + \frac{N_{AN} + N_{AN}^{prompt}}{\left(N_{AN} - N_{AN}^{prompt}\right)^2}} \quad (7.6)$$

$$\frac{\Delta F}{F} = \sqrt{\frac{N_A + N_A^{prompt}}{\left(N_A - N_A^{prompt}\right)^2} + \frac{N_N + N_N^{prompt}}{\left(N_N - N_N^{prompt}\right)^2}} \quad (7.7)$$

The expected relative fake factor uncertainties of the different extraction methods are compared in Figure 7.3. It is undeniable that the unbiased fake factor would have the lowest



(a) Fake factors for electrons

(b) Fake factors for muons

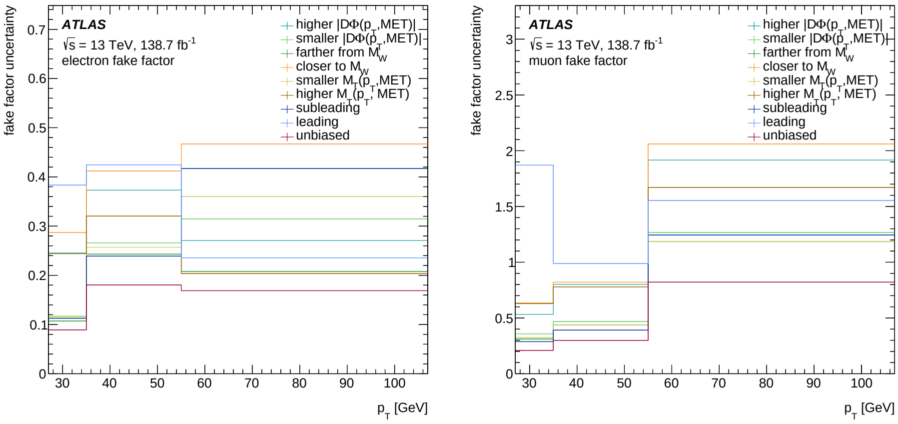
Figure 7.2: Electron and muon fake factors extracted by different tag and probe assignments in comparison to the unbiased fake factor. The shown uncertainties result only from the statistical uncertainty of the Monte Carlo samples.

statistical uncertainty in the final extraction from data. The advantage of the double counting of leptons seems to outweigh the disadvantage of the double subtraction in the case of pure prompt/charge flip events. Therefore further improvements of the assignment method are required to become competitive to the unbiased fake factor.

7.5 Improving the Assignment with an Additional Selection Criterion

The assignment of a tag and probe lepton allows the extraction of a fake factor, which has a low bias but nevertheless a huge statistical uncertainty. The lower bias was achieved by an assignment which is independent of the Ana and Non-Ana selection of the respective lepton, but this is also the reason for the huge uncertainty. For events with one Ana and one Non-Ana lepton it is very likely that the Ana lepton was prompt/charge flip and the Non-Ana lepton was non-prompt. If the assignment of the tag and probe is not extremely successful in choosing the non-prompt lepton as probe, a lot of Non-Ana events will be wrongly chosen as tag lepton for events with one Ana and one Non-Ana lepton. Since the number of events with one Non-Ana lepton is much higher than the number of events with two Ana leptons, this effect dominates the uncertainty in the fake factor numerator. Even if the probability is low that the Ana lepton was non-prompt and at the same time the Non-Ana lepton was prompt/charge flip, this can happen due to the high number of events with Non-Ana leptons. Therefore there is a small migration of correctly assigned Non-Ana events with one non-prompt Ana and one prompt/charge flip Non-Ana lepton into the numerator of the fake factor and thereby the systematic overestimation of the fake factor.

This problem can be solved by an additional cut which discards all events where the tag lepton was not Ana selected. Since the tag and probe assignment is done before this selection is applied the assignment is still independent of the Ana and Non-Ana selection. The only difference to the previous extraction method is the assumption, that nearly all events where



(a) Relative statistical uncertainty for electron fake factors

(b) Relative statistical uncertainty for muon fake factors

Figure 7.3: Relative statistical uncertainties of the fake factors, which are expected for the final extraction from data for electrons and muons. The shown uncertainties result only from the statistical uncertainty of the Monte Carlo samples.

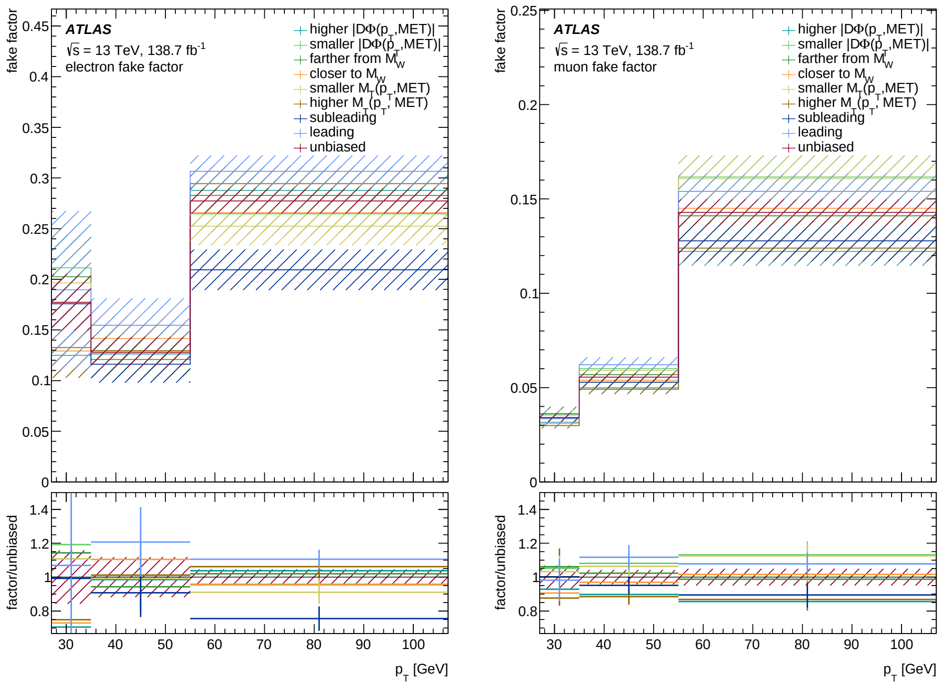
the probe lepton is Ana selected and at the same time the tag lepton is Non-Ana selected are wrongly assigned. Therefore these events can be directly discarded and do not have to be subtracted by Monte Carlo. If the assumption that these events are mostly wrongly assigned is correct then this selection should have no major effects on the value of the fake factor except of a small downward correction. This assumption is confirmed when one has a look on these fake factors in Figure 7.4. The Ana cut on the tag lepton further improves the general bias of the assignment method and most of the fake factors agree very well with the unbiased fake factor within their Monte Carlo uncertainty. Only the assignments according to lepton p_T introduce a significant bias to the fake factor, but this was expected in advance. The remaining methods seem to be rather well suited to select a lepton for the fake factor extraction, especially the assignments according to the difference between $M_T(\vec{p}_T^{lep}, \vec{p}_T^{miss})$ and the $W^\pm$ boson mass has a minimal bias in most of the bins.

The relative statistical uncertainties, which are expected for the final extraction from data, are shown in Figure 7.5 and have improved significantly in comparison to the previous uncertainties in Figure 7.3. The lowest uncertainties of the assignment procedures are given by choosing the lepton with higher $M_T(\vec{p}_T^{lep}, \vec{p}_T^{miss})$ as probe lepton and choosing the lepton with greater deviation of $M_T(\vec{p}_T^{lep}, \vec{p}_T^{miss})$ from the $W^\pm$ mass as probe.

Therefore, the best assignment method so far is selecting the lepton which provides the larger difference between M_T and the $W^\pm$ boson mass as probe lepton, since this methods provides a minimal bias and a minimal uncertainty. Although this method seems to give quite good results, the smallest uncertainty in most of the bins is still provided by the unbiased fake factor. Only in the high p_T bin of the electron fake factor, choosing the lepton with higher $M_T(\vec{p}_T^{lep}, \vec{p}_T^{miss})$ as probe lepton provides a lower uncertainty than the unbiased fake factor.

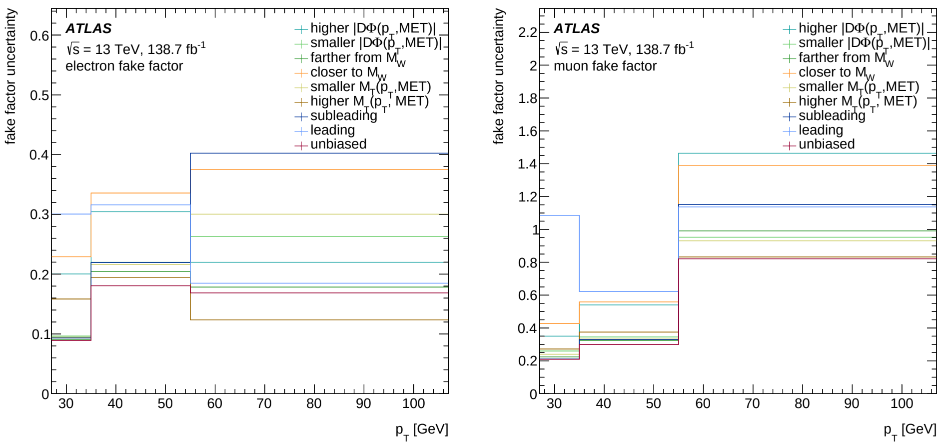
As a result of this study it can be summarized that the unbiased fake factor is still the most suitable candidate for fake factor extraction. It leads per definition to a perfect technical closure within the control region and provides in general the lowest statistical uncertainty. Therefore, the unbiased fake factor is used in the remaining parts of the thesis and any further analysis of the fake factor always refers to the unbiased fake factor.

However, this study has also shown that an assignment of a probe and a tag lepton can also be a suitable method for fake factor extraction. As soon as there is a physical assignment option which is quite successful in choosing the correct lepton, the assignment extraction method can become superior to the unbiased fake factor.



(a) Fake factors for electrons **(b)** Fake factors for muons

Figure 7.4: Electron and muon fake factors extracted by different tag and probe assignments in comparison to the unbiased fake factor. The shown uncertainties result only from the statistical uncertainty of the Monte Carlo samples.



(a) Relative statistical uncertainty for electron fake factors **(b)** Relative statistical uncertainty for muon fake factors

Figure 7.5: Relative statistical uncertainties of the fake factors, which are expected for the final extraction from data for electrons and muons. The shown uncertainties result only from the statistical uncertainty of the Monte Carlo samples.

Chapter 8

Technical Monte Carlo Closure Test within the Control Region

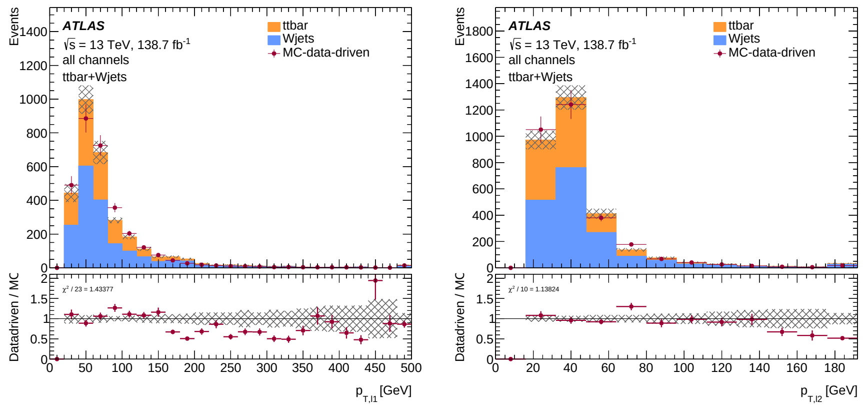
Since the control region is defined and a suitable method to extract the fake factor was developed, the next step is to test the setup. Therefore a technical Monte Carlo closure test within the control region is performed. For this test the fake factor is extracted from Monte Carlo simulated events in the control region and also applied on Monte Carlo events in the control region. Thereby the application of the fake factor is done just like for the signal region in Equation 4.7. In order to emphasize that the data-driven method was not applied to actual data but only to Monte Carlo simulated events this is called "MC-data-driven" in this thesis. The number of non-prompt events with two Ana leptons estimated by the MC-data-driven method is then compared to the number of events with two Ana lepton which are actually non-prompt according to the Monte Carlo truth level information. This test is essential to find technical bugs, discrepancies of the procedure and the setup and further dependencies of the fake factor.

Since the unbiased fake factor is used, one expects a perfect closure in the total number of events. This was already proven in Table 7.1. The perfect agreement in the total number of events confirms that there is no major technical bug in the chosen setup. In order to find more hidden discrepancies one has to take a deeper look into the individual kinematic distributions after applying the fake factor in the control region. This was done for several distributions. In the following the most important distributions will be discussed in more detail. Since in all channels more than 90% of the non-prompt leptons originates from $W + jets$ and $t\bar{t}$ events, the test is applied only for this two samples in order to reduce the computation effort. This significant smaller amount of Monte Carlo events is considered to be sufficient for this test, because all purely prompt events would anyway be subtracted in the extraction and application of the fake factor.

The non-prompt events with two Ana leptons taken directly from Monte Carlo simulated events are splitted in the processes they originate from. The shown uncertainty is only the statistical uncertainty of the Monte Carlo simulation. The uncertainty of the MC-data-driven background is given by the statistical uncertainty of the Monte Carlo simulated events with one Non-Ana lepton and the uncertainty of the fake factor. Thereby, the uncertainty of the fake factor also results from the statistical uncertainty of the Monte Carlo simulated events the fake factor was calculated from. Therefore, all of the uncertainties shown in this chapter are resulting only from the statistical uncertainty of the Monte Carlo events.

8.1 Transverse Momentum

Since the fake factor is already binned in the transverse lepton momentum p_T one would expect a very good closure within this variable. As can be seen in Figure 8.1 the MC-data-driven background is compatible with the non-prompt distribution given by Monte Carlo. Despite the general agreement, however, there are also systematic deviations between the



(a) Leading lepton p_T **(b)** Subleading lepton p_T

Figure 8.1: Technical Monte Carlo closure test within the control region for the $W+jets$ and $t\bar{t}$ sample. The MC-data-driven estimate of the non-prompt leading and subleading lepton distribution in p_T is compared to the non-prompt Monte Carlo events with two Ana leptons. The shown uncertainties result only from the statistical uncertainty of the Monte Carlo samples.

MC-data-driven background and the Monte Carlo non-prompt background. The cause of this discrepancy is the broad p_T binning of the fake factor ([27 GeV, 35 GeV), [35 GeV, 55 GeV) and [55 GeV, inf)). Especially for the leading lepton most of the range in p_T is covered by just one bin. Thus, in this higher p_T range, systematic over- and underestimations cannot be avoided without choosing a finer binning. However, this finer binning would decrease the statistic in the last bin for the fake factor extraction and thereby increase the uncertainty which is expected to be already quite high especially for muons as shown in Figure 7.5. Therefore this discrepancy can only be solved with a significantly higher statistic than given by the Run 2 data.

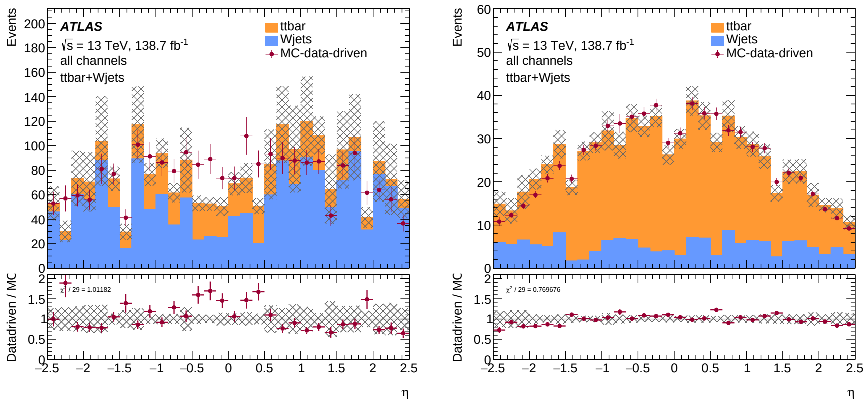
8.2 Geometrical Distribution

As already shown the fake factor is highly dependent on the transverse momentum of the non-prompt object. However, it can also be expected that the distribution of non-prompt leptons also depends on the general geometry of the detector. To check whether the geometric distribution of the non-prompt leptons has to be taken into account when one extracts the fake factor the closure test was also performed for the η distribution. In order to be sensitive to different η dependencies for electrons and muons the lepton η distribution in Figure 8.2 is splitted into the distribution of non-prompt electrons and non-prompt muons. For non-prompt muons there are no major differences between the MC-data-driven estimate and the Monte Carlo simulated non-prompt events. Therefore, a purely p_T dependent fake factor seems to be sufficient for muons. For electrons there is a significant overestimation of non-prompt electrons with η close to zero.

The only way to handle this effect is to extract the fake factor for electrons also in terms of η. Due to the limited statistic of the control region a two dimensional binning of the fake factor is not possible. In order to take the geometric distribution of the non-prompt leptons into account one could multiply a η dependent correction factor $c(\eta)$ to the extracted fake factor.

$$F'(p_T, \eta) = F(p_T) \cdot c(\eta) \tag{8.1}$$

This approach was already followed for the current control region in [23]. An analytic function for $c(\eta)$ was fitted to the Monte Carlo simulated events of the $W+jets$ and $t\bar{t}$ sample and it was proven that this η correction would improve the agreement in the technical Monte



(a) Non-prompt electron η **(b)** Non-prompt muon η

Figure 8.2: Technical Monte Carlo closure test within the control region for the $W + jets$ and $t\bar{t}$ sample. The MC-data-driven estimate of the non-prompt electron and muon distribution in η is compared to the non-prompt Monte Carlo events with two Ana leptons. The shown uncertainties result only from the statistical uncertainty of the Monte Carlo samples.

Carlo closure test. Since the results of this study have only recently become available, this correction is not yet applied in this thesis. It is not expected that the η correction would have any major effects on the results of this thesis, because there is no η dependent analysis in the remaining chapters. Nevertheless, this closure test once again confirmed that a η is required for electrons, but thanks to the study in [23] a suitable setup is already available.

8.3 Invariant Dilepton Mass

Another distribution sensitive to the exact implementation of the data-driven non-prompt estimation is the invariant mass of the two reconstructed leptons. In the non-prompt analysis for WZ scattering in [24] a shift of the Z peak towards lower energies was observed in the m_{ll} distribution if no further corrections were applied. This shift is caused by the effect that a non-prompt object is more likely to have a lower reconstructed momentum if it is Non-Ana selected. More details about this effect will follow in Chapter 9. Because of this lower momentum of Non-Ana non-prompt leptons the m_{ll} peak is for events with one Non-Ana lepton at a lower energy than the same m_{ll} peak for events with two Ana leptons. Since the fake factor has to be applied on event numbers and not on single leptons this peak shift would effect the data-driven non-prompt estimation of the m_{ll} distribution. However, this shift is not observed in Figure 8.3, since the m_{ll} is way to broad to recognize small shifts. Therefore, the technical closure test does not demand a further correction on the lepton momentum, but a deeper discussion will follow in Chapter 9.

8.4 Jet Kinematic

A very good closure is also reached in the jet kinematics. The closures in the total number of jets and the invariant mass of the two leading jets are shown in Figure 8.4. Especially the closure in the total number of jets is a significant indication that the technical setup is working since it directly depends on the overlap removal of jets and leptons. If the overlap between Non-Ana leptons and jets is ignored, this would lead to significant discrepancies in the number of jets. Since the overlap removal was already implemented in the $W^{\pm}W^{\pm}jj$-EW analysis for Ana leptons, the same procedure is also applied for Non-Ana leptons in this thesis. The number of jets estimated by the MC-data-driven method agrees with the Monte Carlo simulated events within their uncertainties, but nevertheless one could assume that there is a

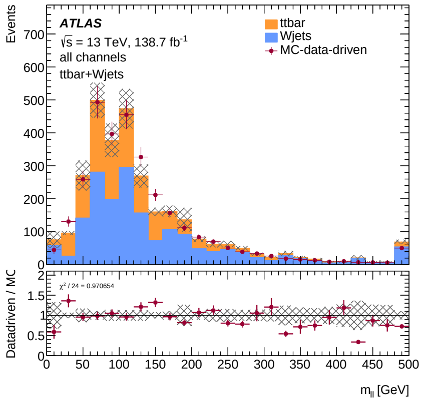
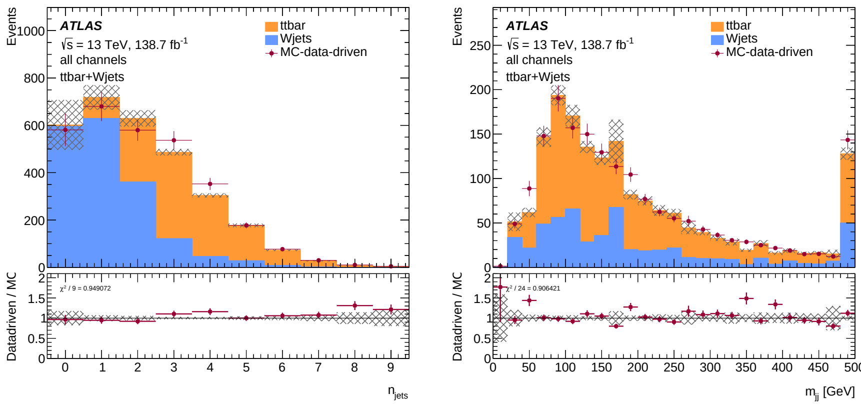


Figure 8.3: Technical Monte Carlo closure test within the control region for the $W + jets$ and $t\bar{t}$ sample. The MC-data-driven estimate of the m_{ll} distribution is compared to the non-prompt Monte Carlo events with two Ana leptons. The shown uncertainties result only from the statistical uncertainty of the Monte Carlo samples.



(a) Number of jets

(b) m_{jj} of the two leading jets

Figure 8.4: Technical Monte Carlo closure test within the control region for the $W + jets$ and $t\bar{t}$ sample. The MC-data-driven estimates of the number of jets and the m_{jj} distribution of the two leading jets are compared to the non-prompt Monte Carlo events with two Ana leptons.

small shift towards higher jet numbers. A small shift could be caused by the assumption that the overlap removal can be applied independently of the Ana and Non-Ana selection. For example, in [2] a slightly different overlap removal was applied for Non-Ana leptons. Non-Ana leptons within $0.2 < \Delta R_{l,j} < 0.4$ of a jet were kept while the jet was removed. Since the results of the jet overlap removal, which was directly taken from the overlap removal for Ana leptons, are already very good, the application of this modified overlap removal was not studied in the context of this thesis. Therefore, a study of the overlap removal between jets and leptons is still pending.

Chapter 9

Correction of the Transverse Momentum of Non-Ana Leptons

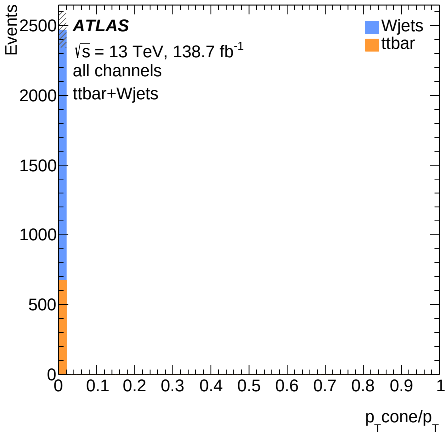
The previous analysis of same charged $W^{\pm}$ boson scattering in [2] already mentioned the necessity of a correction on the transverse momentum. Since a deeper analysis of the p_T correction can be very extensive, this thesis is limited to checking whether such a correction is necessary in principle for the used control region and to apply this correction in the same way as it was done in [24]. The p_T correction will be applied unless significant issues are found, because it has significant effects on the fake factor and the scope of this thesis is to produce a fake factor which is compatible with alternative approaches followed in the current analysis.

As in the previous chapter about the Monte Carlo closure test all the histograms and event numbers presented in this chapter are only based on the Monte Carlo events of the $W + jets$ and $t\bar{t}$ samples. This simplification is again justified by the fact that nearly all non-prompt events originate from these two samples.

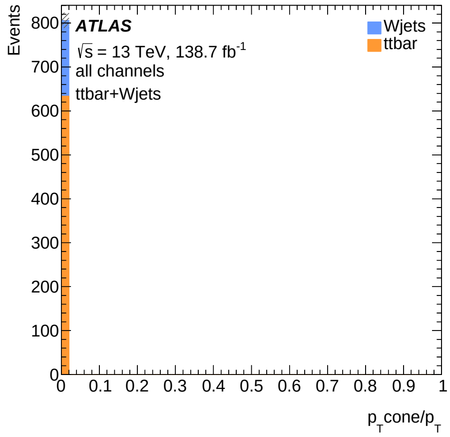
9.1 Motivation and Necessity of a Transverse Momentum Correction

Since the fake factor is binned in terms of the transverse momentum, it is appropriate to take a closer look at the way how the transverse momentum is reconstructed. If eg. a jet faking an lepton passes the Ana selection the non-prompt object needs to be well isolated. Therefore, nearly all of the momentum of the jet is reconstructed as the non-prompt leptons momentum. If the isolation criterion is failed than the faking object is spread out and only a fraction of the momentum will be used to reconstruct the non-prompt lepton p_T. This effect can be seen if one compares the remaining momentum p_T^{cone} in a cone around the track of reconstructed Ana and Non-Ana leptons, which was not used for the object p_T reconstruction. As shown in Figure 9.1 there is no momentum remaining in the cone after the reconstruction of an Ana selected lepton while poorly isolated Non-Ana leptons carry just a fraction of the whole momentum inside the cone. Especially non-prompt muons are very likely to have a high momentum p_T^{cone} which is not used for the p_T reconstruction. Therefore, the isolation criteria introduces a bias into the reconstruction of the lepton momentum. The significant lower isolation of Non-Ana muons can be caused by the different object selection of electrons and muons, which could lead to a higher probability for muons to fail the isolation selection instead one of the other cuts of the Ana selection. Furthermore the differences in the origin of non-prompt electrons and non-prompt muons and their different measurement in the detector can effect the lepton isolation.

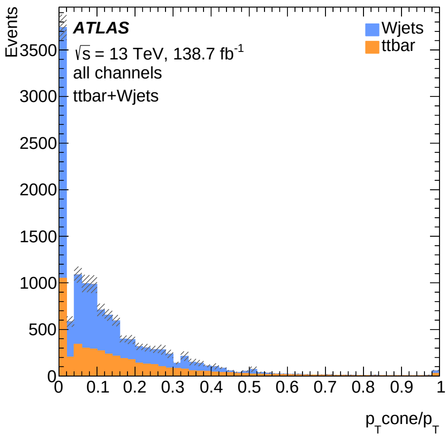
The technical Monte Carlo closure test for the m_{ll} distribution was already a good opportunity to evaluate the influence of this bias. A smaller transverse momentum of Non-Ana leptons would lead to a shift of m_{ll} towards lower energies. However, the test has shown a



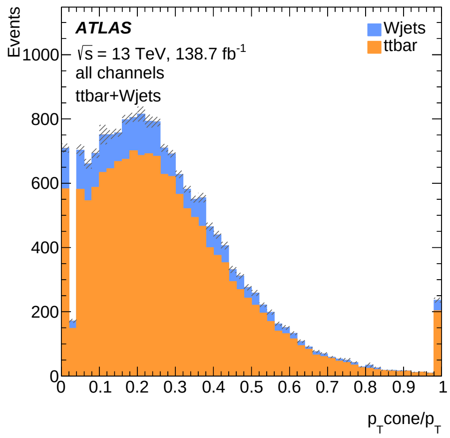
(a) p_T^{cone}/p_T for Ana selected non-prompt electrons



(b) p_T^{cone}/p_T for Ana selected non-prompt muons



(c) p_T^{cone}/p_T for Non-Ana selected non-prompt electrons



(d) p_T^{cone}/p_T for Non-Ana selected non-prompt muons

Figure 9.1: p_T^{cone}/p_T ratio for non-prompt electrons and non-prompt muons originating from the $W + jets$ and $t\bar{t}$ sample. The leptons in figure (a) and (b) are Ana selected and in figure (c) and (d) the leptons are Non-Ana selected. The shown uncertainties result only from the statistical uncertainty of the Monte Carlo samples.

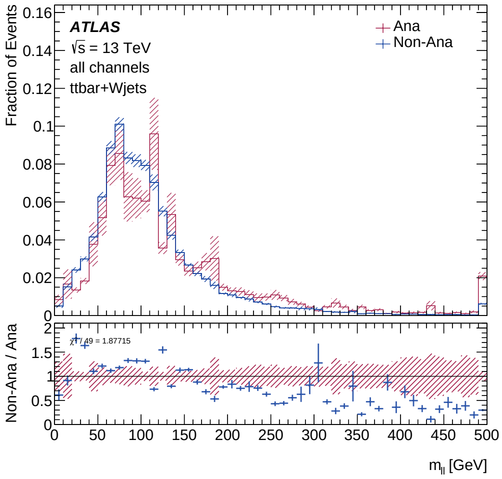


Figure 9.2: m_{ll} distribution of $W + jets$ and $t\bar{t}$ Monte Carlo simulated events for events with two Ana leptons compared to events with one Ana and one Non-Ana lepton. The shown uncertainties result only from the statistical uncertainty of the Monte Carlo samples.

good agreement and thereby no shift of the distribution was observed.

In order to verify that there is no significant shift the normalized m_{ll} distributions are compared in Figure 9.2 for events with two Ana leptons and events with one Ana and one Non-Ana lepton. The agreement between the distributions is good, but nevertheless there is a small shift towards lower energies visible. If this small shift is caused by the expected bias of the p_T reconstruction for Ana and Non-Ana leptons or just by other differences in the object selection of Ana and Non-Ana leptons has to be proven. Therefore one has to emphasize that a clear motivation was not shown by the simple analysis of the invariant mass, as it would probably have been the case with a sharp m_{ll} peak.

9.2 Procedure of the Transverse Momentum Correction

As already mentioned before, the following procedure for the p_T correction is based on the correction presented in [24]. For Non-Ana leptons the p_T^{cone} is only added to its reconstructed transverse momentum if the lepton fails the isolation criteria, because failing one of the other Ana criteria is not expected to introduce a bias on the momentum. Since the necessity of a p_T correction could not fully motivated in this work, there is no easy way to study whether only fraction of the momentum p_T^{cone} should be added. Therefore, as in the same charged $W^\pm$ boson analysis in [2], the whole p_T^{cone} is added to the transverse momentum for leptons failing the isolation criteria. The p_T correction is consistent for the extraction and application of the fake factor as illustrated in Equation 9.1 and 9.2 which are simplified to a one lepton problem for demonstration purpose.

$$F(p_T) := \frac{N_A(p_T) - N_A^{prompt}(p_T)}{N_N(p_T + p_T^{cone}) - N_N^{prompt}(p_T + p_T^{cone})} \tag{9.1}$$

$$N_{fake}(p_T) = (N_N(p_T + p_T^{cone}) - N_N^{prompt}(p_T + p_T^{cone}))F(p_T) \tag{9.2}$$

Due to the object selection on p_T for Non-Ana leptons it is required to add this correction before the object selection is applied. Therefore, a significant larger number of leptons is expected to pass the p_T selection at 27 GeV. Since a preselection, the so called *skimming*, was already applied to the used data samples on derivation level, the samples only consists of events with a certain number of leptons with a minimum criteria on their kinematic.[24] To take this preselection into account the p_T correction should be only applied on leptons with a minimal transverse momentum. In order to avoid the introduction of a bias, a maximal

	Ana selected leptons	Non-Ana selected leptons
without p_T correction:		
non-prompt electrons	2274	13102
non-prompt muons	754	17244
with p_T correction:		
non-prompt electrons	2287	17402
non-prompt muons	801	26719

Table 9.1: Number of non-prompt electrons and muons originating from $W + jets$ and $t\bar{t}$ Monte Carlo simulated events before and after the p_T correction. The number of leptons is splitted into Ana and Non-Ana selected leptons

correction on the transverse momentum is defined instead of just discard all leptons with low p_T before the correction. Therefore, all objects with p_T^{cone}/p_T higher than 0.5 are rejected and thereby only leptons with p_T higher than 18 GeV and thus also passing the derivation level *skimming* are able to pass the selection at $p_T = 27$ GeV. This minimal isolation also prevents a too large correction of a lepton's transverse momentum.

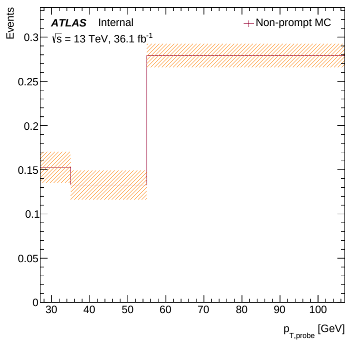
9.3 Effects of the Transverse Momentum Correction

Due to the new object selection at $p_T^{cone}/p_T = 0.5$ some Non-Ana leptons are discarded, especially for muons, but due to the p_T correction significant more leptons are expected to pass the $p_T = 27$ GeV selection criterion of Non-Ana selection. The statistic of non-prompt leptons in the $W + jets$ and $t\bar{t}$ sample is given in Table 9.1. Surprisingly, a small increase in Ana selected non-prompt leptons is observed, although they are not directly affected by the p_T correction. One possible explanation is an effect on the third lepton veto of the control region selection. Since Non-Ana leptons with a very loose isolation are now rejected in the object selection, events with two Ana leptons seem to have a lower probability to have an additional Non-Ana lepton. Why the third lepton veto in combination with the significant migration into the Non-Ana selection does not primarily lead to a lower statistics in non-prompt Ana leptons was not clarified. Nevertheless, the overall effect on the statistic of Ana selected non-prompt leptons is quite small.

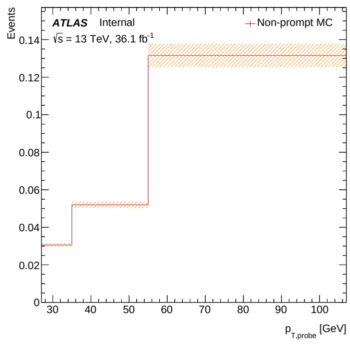
For Non-Ana selected non-prompt leptons the effects on the statistic are as expected much more significant. The number of Non-Ana selected non-prompt leptons is increased by 33% for electrons and by 55% for muons. Since the number of Non-Ana leptons was already high enough compared to the number of Ana leptons in the control region the increase of this number is not expected to have major effects on the fake factor uncertainty. However, also the number of Non-Ana selected leptons in the signal region is going to increase due to the p_T correction. Since this number of leptons is typically rather small, the total uncertainty of the data-driven estimation thereby will decrease. Therefore, the p_T correction is expected to have positive effects on the statistical uncertainty of the estimated non-prompt background.

Because of the higher amount of Non-Ana selected leptons, the fake factors for muons and electrons become significant smaller. This effect can be seen in Figure 9.3 where the fake factors without and with a p_T correction are compared. Despite the fact that the factors have decreased significantly, the effects on their general shape is rather small. Thereby the decrease of the fake factor by over 20% is in agreement with the over 30% increase in Non-Ana selected leptons. The shape of the electron fake factor has hardly changed, except for a small reduction of the value in the [25 GeV, 55 GeV) fake factor bin in relation to the other two bins. For muons the relative differences between the bin values has significantly decreased and thereby the fake factor became more flat. This change on the muon fake factors shape is expected to have positive effects on the data-driven estimate, since it could reduce the steps on bin edges in the p_T modeling which has been seen in Figure 8.1.

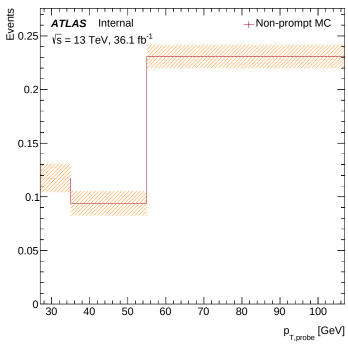
The reason why the fake factor shape has changed primarily only for muons can be explained by their much higher p_T^{cone} in Figure 9.1. Since the electron p_T^{cone} is in general



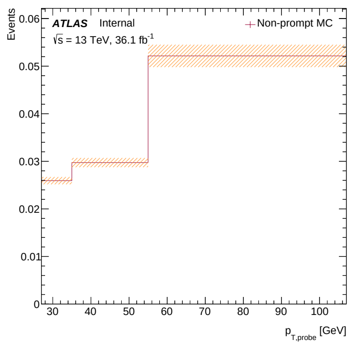
(a) electron fake factor without p_T correction



(b) muon fake factor without p_T correction



(c) electron fake factor with p_T correction



(d) muon fake factor with p_T correction

Figure 9.3: Electron and muon fake factors extracted from $W + jets$ and $t\bar{t}$ Monte Carlo simulated events with and without the p_T correction. The shown uncertainties result only from the statistical uncertainty of the Monte Carlo samples.

quite small one expects only lepton migrations next to the bin thresholds. For muons the fraction of p_T^{cone}/p_T is much higher on average and therefore a lot of low p_T Non-Ana muons are migrating into higher p_T bins, which finally reduces the values in these high p_T bins more significantly.

Finally, one has to test whether the applied p_T correction eliminates the small shift in the m_{ll} distribution. As can be seen in Figure 9.4, no significant difference is found in the m_{ll} distribution and the chi-square value between the Ana and Non-Ana distribution, and thus a further evaluation such as a Kolmogorov Smirnov test is not necessary. However, this correction will be applied in the context of this thesis in order stay consistent.

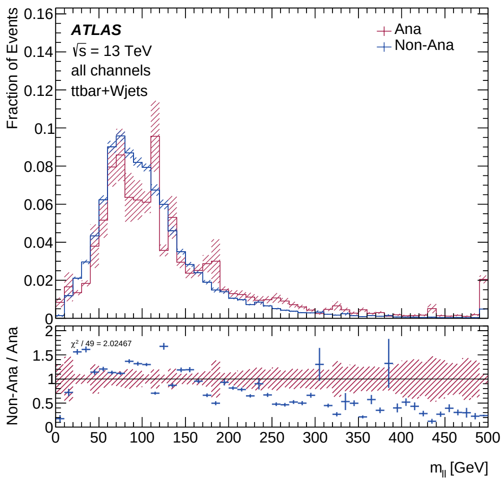


Figure 9.4: p_T corrected m_{ll} distribution of $W + jets$ and $t\bar{t}$ Monte Carlo simulated events for events with two Ana leptons compared to events with one Ana and one Non-Ana lepton. The shown uncertainties result only from the statistical uncertainty of the Monte Carlo samples.

Chapter 10

Closure Test within the Low Dijet Invariant Mass Validation Region

Within the previous chapters a suitable control region was defined, the fake factor extraction was adapted to it and the resulting setup was tested and optimized using simulation. Finally, in order to probe that the method is usable it is necessary to test whether the extracted factor is really suitable for the $W^{\pm}W^{\pm}jj$-EW signal region. Since $W^{\pm}W^{\pm}jj$-EW is at the moment still a blinded analysis, we cannot look yet at the signal region, therefore, a the low dijet invariant mass region is used as validation region. This LowMjj region is particularly suitable for validation of the fake factors, because in spite of being orthogonal to the signal region is kinematically very close to the signal region and has a large fake background contribution. As a first step to validate the data-driven background estimation the closure test is applied for Monte Carlo events to ensure a working setup and finally a full closure test is be performed using collision data.

10.1 Monte Carlo Closure Test

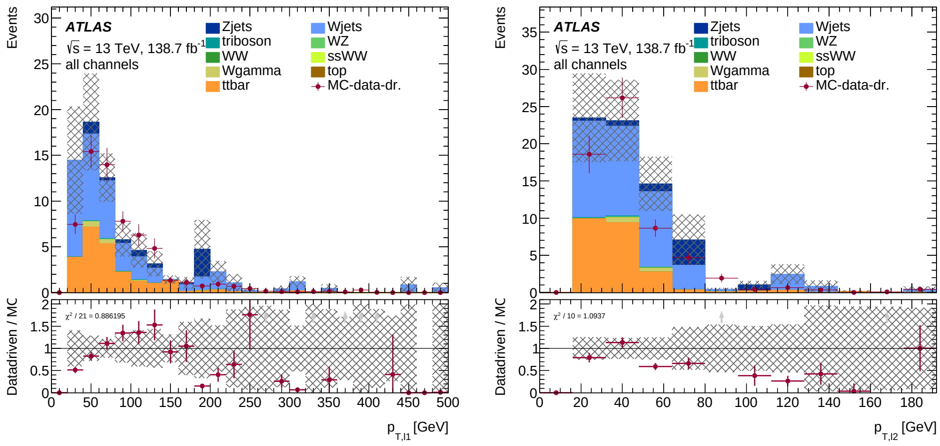
In order to be as thorough as possible the full set of Monte Carlo samples is used for the Monte Carlo closure test even if they are not expected to introduce a significant non-prompt contribution. The procedure of the Monte Carlo closure test is the same as in the technical Monte Carlo closure test in Chapter 8, except this time the fake factors are applied on Monte Carlo events in the LowMjj region.

In Table 10.1 the total numbers of non-prompt Monte Carlo events with two Ana leptons is compared to the MC-data-driven estimate of the non-prompt background. The uncertainties are splitted into the uncertainty of Monte Carlo statistics in the LowMjj region and the uncertainty introduced by the fake factor. The fake factor uncertainty at the moment only includes the Monte Carlo statistical uncertainty in the control region where the fake factor was extracted. The result of the MC-data-driven estimate is slightly below the number of non-prompt events given by Monte Carlo, but the agreement is still within the uncertainties. Therefore, the chosen setup seems to be suitable for the data-driven estimation of the non-prompt background in the $W^{\pm}W^{\pm}jj$-EW signal region.

Since a closure was observed in the total numbers of Monte Carlo events, this closure has to be confirmed in several kinematic distributions, as in the technical Monte Carlo closure test in Chapter 8, in order to avoid systematic misestimations of the non-prompt background. Therefore the Monte Carlo closure test was performed for the p_T and η for the leading and subleading lepton, the invariant mass m_{ll}, the missing transverse momentum E_T^{miss}, the

	ee	eμ	μe	$\mu\mu$	all
Non-prompt MC	19 ± 7	24 ± 4	19 ± 5	10.9 ± 2.5	74 ± 10
MC-data-driven bkg	$15.0 \pm 1.5 \pm 1.5$	$19.7 \pm 1.5 \pm 1.2$	$16.7 \pm 1.7 \pm 1.8$	$10.5 \pm 0.3 \pm 0.5$	$62 \pm 3 \pm 5$
MC-data-driven / MC	$0.8 \pm 0.3 \pm 0.1$	$0.81 \pm 0.16 \pm 0.05$	$0.87 \pm 0.28 \pm 0.10$	$0.96 \pm 0.23 \pm 0.05$	$0.84 \pm 0.13 \pm 0.07$

Table 10.1: Monte Carlo closure test within the LowMjj region splitted in lepton channels. The number of non-prompt Monte Carlo events with two Ana leptons is compared to the MC-data-driven estimate. The first uncertainty results from the statistical uncertainty of the Monte Carlo simulated events in the LowMjj region. The second uncertainty is caused by the uncertainty of the fake factor which was extracted from the control region.



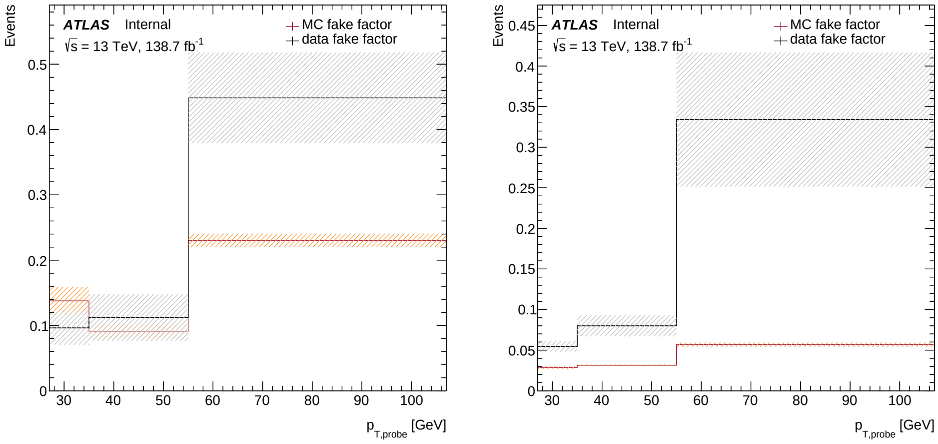
(a) Leading lepton p_T

(b) Subleading lepton p_T

Figure 10.1: Monte Carlo closure test within the LowMjj region. The MC-data-driven estimate of the non-prompt leading and subleading lepton distribution in p_T is compared to the non-prompt Monte Carlo events with two Ana leptons. The shown uncertainties result only from the statistical uncertainty of the Monte Carlo samples.

number of jets and the invariant mass m_{jj} of the two leading jets. The modeling of the non-prompt lepton momentum in Figure 10.1 indicates the underestimation which was already visible in the total number of events, but the agreement is still sufficient due to the high uncertainties in all bins. Also the MC-data-driven estimate of the non-prompt lepton η in Figure Appendix A.8 is compatible with the non-prompt Monte Carlo events. The necessity of a η correction as it was studied in Figure 8.2 was not verified because of the significant higher uncertainties. The same agreement was also observed for the invariant mass m_{ll} and the missing transverse momentum E_T^{miss} in Figure Appendix A.9 and for the number of jets and the invariant mass m_{jj} of the two leading jets in Figure Appendix A.10.

Since the data-driven procedure applied on Monte Carlo agrees with the non-prompt Monte Carlo events within their uncertainties, the assumption that the fake factor extracted from the control region can be directly applied in the signal region was confirmed for Monte Carlo simulated events. The good agreement for the Monte Carlo closure test within the LowMjj region can be explained by the fact that the selection criteria of the chosen control region are very close to the selection criteria of the LowMjj and $W^{\pm}W^{\pm}jj$-EW signal region. If similar good results are obtained in a closure test with actual data, it can be assumed that the chosen data-driven method will not introduce significant systematic uncertainties into the signal region of the same charged $W^{\pm}$ boson scattering.



(a) Electron fake factor

(b) Muon fake factor

Figure 10.2: Electron and muon fake factor extracted from actual data in the non-prompt control region compared to the Monte Carlo fake factor. The shown uncertainty of the Monte Carlo fake factor is derived from the statistical uncertainty of the non-prompt Monte Carlo events and the uncertainty of the fake factor extracted from data results from the statistical uncertainty of the data and Monte Carlo simulated events.

10.2 Data Closure Test

The application of the data-driven method to actual data in the LowMjj validation region represents the final test in this thesis. For this purpose, it is examined how well the data in the LowMjj region corresponds to the sum of Monte Carlo predictions and the non-prompt background estimated with the data-driven method. In Figure 10.2 the fake factors extracted from data in the non-prompt control region are shown. The fake factors extracted from data differ significantly from the Monte Carlo fake factors. Especially for high-energetic leptons the data fake factors are much higher than expected from the Monte Carlo simulation. For electrons the data fake factor is well modeled in the low p_T bins but for muons the fake factor in these bins is underestimated by a factor of two. In Figure Appendix A.11 and A.12 the event numbers in the denominator and numerator of the fake factor are compared for data and Monte Carlo simulation. These figures confirm that the data for the electron fake factor extraction are quite well modeled by Monte Carlo simulation and the data for the muon fake factor extraction are significantly underestimated in both the numerator and denominator. It can be assumed that this deviation is mainly due to the mismodeling in the $e\mu$ channel, which was already observed in Figure Appendix 6.7 during the definition of the control region. Since it was assumed that the non-prompt leptons are not sufficiently well simulated by Monte Carlo events, no conclusion of the data-driven method can yet be derived from the significant deviation between the data fake factor and the Monte Carlo fake factor.

In order to check whether the fake factors extracted from data are applicable to the final same charged $W^{\pm}$ scattering analysis, they are now applied to model the data in LowMjj region. If there are large deviations with the data, it must be assumed that either technical problems have occurred in the setup or the procedure described in this thesis is not suitable for the data-driven estimation of non-prompt background. The resulting total event numbers of the data closure test within the LowMjj are listed in Table 10.2. It can be clearly seen that about 25% of the data in the LowMjj region is attributed to the non-prompt background, which once again illustrates the importance of a correct estimation of this background. In general, a systematic overestimation of the expected yields is visible, but this is still within the scope of the uncertainties. The uncertainty is splitted into the statistical uncertainty of the data and Monte Carlo samples in the LowMjj region and the uncertainty due to the fake factor uncertainty. The high uncertainty of 19% on the total number of non-prompt

	ee	eμ	μe	μμ	all
Data-driven bkg	$25 \pm 3 \pm 5$	$38 \pm 3 \pm 6$	$14 \pm 2 \pm 3$	$30 \pm 2 \pm 5$	$106 \pm 5 \pm 20$
Prompt MC	55 ± 4	81 ± 4	92 ± 7	66.3 ± 2.8	295 ± 10
Charge flip MC	18 ± 4	6.2 ± 0.9	3.6 ± 1.6	0 ± 0	27 ± 4
Data	77 ± 9	125 ± 11	99 ± 10	86 ± 9	387 ± 20
Datadriven + MC	$97 \pm 6 \pm 5$	$126 \pm 6 \pm 6$	$109 \pm 8 \pm 3$	$96 \pm 4 \pm 5$	$429 \pm 12 \pm 20$
Data / (data-driven + MC)	$0.79 \pm 0.12 \pm 0.04$	$1.00 \pm 0.10 \pm 0.05$	$0.91 \pm 0.12 \pm 0.03$	$0.89 \pm 0.11 \pm 0.05$	$0.90 \pm 0.05 \pm 0.04$

Table 10.2: Data closure test within the LowMjj region splitted in lepton channels. The number of data events with two Ana leptons is compared to the prompt and charge flip events estimated by Monte Carlo simulation and the data-driven non-prompt background estimate. The first uncertainty results from the statistical uncertainty of the Monte Carlo simulated events and the statistical data uncertainty in the LowMjj region. The second uncertainty is caused by the statistical uncertainty of the fake factor which was extracted from the control region.

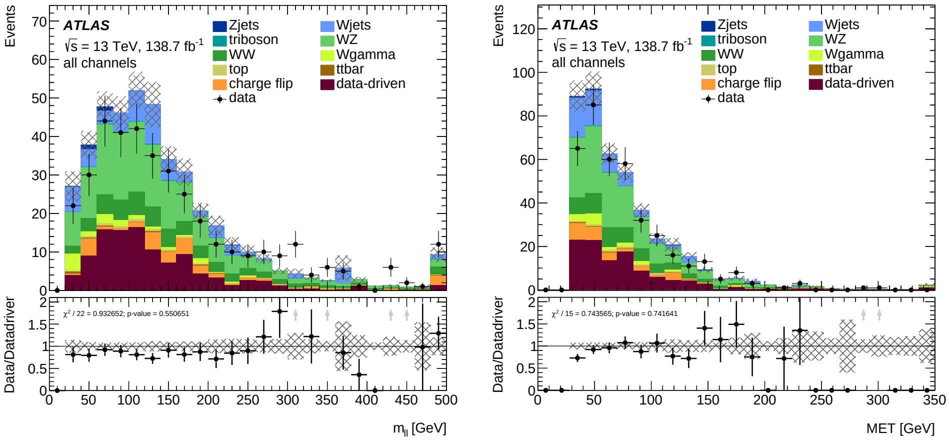
	ee	eμ	μe	μμ	all
Data - MC	4 ± 10	37 ± 12	3 ± 13	20 ± 10	65 ± 22
Data-driven bkg	$25 \pm 3 \pm 5$	$38 \pm 3 \pm 6$	$14 \pm 2 \pm 3$	$30 \pm 2 \pm 5$	$106 \pm 5 \pm 20$
Data-driven / (data - MC)	$6 \pm 6 \pm 1$	$1.0 \pm 0.3 \pm 0.2$	$4 \pm 8 \pm 1$	$1.5 \pm 0.6 \pm 0.3$	$1.6 \pm 0.5 \pm 0.3$

Table 10.3: Data closure test within the LowMjj region splitted in lepton channels. The data-driven estimate of the non-prompt background is compared to the difference between the data and the prompt and charge flip events estimated by Monte Carlo simulation. The first uncertainty results from the statistical uncertainty of the Monte Carlo simulated events and the statistical data uncertainty in the LowMjj region. The second uncertainty is caused by the statistical uncertainty of the fake factor which was extracted from the control region.

events, which originates directly from the statistical fake factor uncertainty, results from the low number of non-prompt Ana leptons in the control region and thus cannot be reduced by simple adjustments. Therefore, this high statistical uncertainty of the fake factor is a limiting factor for the studied data-driven method.

To better assess the estimated non-prompt contribution, Table 10.3 compares the data-driven background with the difference between the data and the prompt and charge flip events estimated by Monte Carlo simulation. A significant significant deviation seems to occur only in the *ee* channel, since in the other channels the deviations are sufficiently covered by the uncertainties. However, the observed differences do not necessarily point to a problem with the data-driven background method, since missing corrections on the Monte Carlo simulations can also be responsible for it. For example the charge flip events are known to be poorly modeled by Monte Carlo simulations and thus also estimated by a data-driven method in the final $W^{\pm}W^{\pm}jj$-EW analysis. Since in this thesis the charge flip background was estimated by Monte Carlo simulated events, the deviation in the *ee* seems to be reasonable. The results of the data closure test within the m_{ll} and E_T^{miss} distribution are shown in Figure 10.3. Even if within uncertainties both distributions show an overestimation of the expected background when compared with data.

Part of the systematic shift of the in Figure 10.3 could be attributed to the WZ background. In the publication of the same charged $W^{\pm}$ boson scattering from 2019 in [1] it was observed that the normalization of the WZ background predicted by Sherpa needed to be scaled by a factor of 0.86. The same simulation is used for this version of the analysis and preliminary studies show the same correction factor can be used for the actual $W^{\pm}W^{\pm}jj$-EW analysis. Since the WZ background is the main source of prompt Monte Carlo events in the LowMjj region, this scaling is expected to increase the agreement with the data. Therefore,



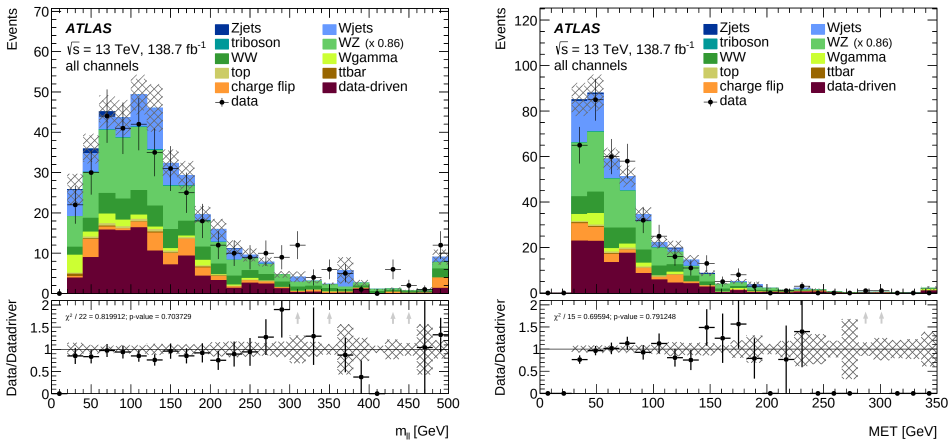
(a) m_{ll} distribution

(b) E_T^{miss} distribution

Figure 10.3: The data m_{ll} and E_T^{miss} distribution in the LowMjj region compared to the sum of the data-driven non-prompt background and the prompt and charge flip contributions estimated by Monte Carlo simulations. The shown uncertainties result from the statistical uncertainty of the Monte Carlo samples and the statistical uncertainty of the data within the dilepton control region and the LowMjj validation region.

this scaling is applied on the WZ contribution for the data closure tests within the m_{ll} and E_T^{miss} distribution in Figure 10.4, the p_T and η distribution of the leading and subleading lepton in Figure Appendix A.13 and A.14 and the number of jets and the invariant mass m_{jj} of the two leading jets in Figure Appendix A.15. In all these distributions the good agreement with the data was confirmed and thus the data-driven method seems to model the nonprompt proportion in the data sufficiently well. The data-driven method presented in this thesis has successfully passed the data closure test within the LowMjj region and therefore seems suitable for estimating the non-prompt background in the $W^{\pm}W^{\pm}jj$-EW region. For a more extensive evaluation of the extracted fake factors, it is recommended to additionally test them for data in a region with an even higher non-prompt contribution and the full set of systematic uncertainties need to be calculated.

The results of the data closure test in the LowMjj validation region seems to validate the data-driven method and control regions described in this thesis. It was also corroborated that some of the used Monte Carlo simulations significantly mismodel the collision data. It is therefore essential to apply the relevant corrections and scale factors such as the WZ scale factor in the control region before the fake factor is extracted.



(a) m_{ll} distribution

(b) E_T^{miss} distribution

Figure 10.4: The data m_{ll} and E_T^{miss} distribution in the LowMjj region compared to the sum of the data-driven non-prompt background and the prompt and charge flip contributions estimated by Monte Carlo simulations. The number of prompt WZ events is scaled by a factor of 0.86. The shown uncertainties result from the statistical uncertainty of the Monte Carlo samples and the statistical uncertainty of the data within the dilepton control region and the LowMjj validation region.

Chapter 11

Conclusion and Outlook

The main idea of this thesis is to develop a data-driven estimation of the non-prompt background for the analysis targeting the measurement of the electroweak production of same charged $W^{\pm}W^{\pm}$ bosons. A new control region using dilepton events is introduced to estimate the fake factors. By being kinematically closer to the $W^{\pm}W^{\pm}jj$-EW signal region than the previously used dijet control region, large extrapolation factors are avoided, and this is expected to lead to a smaller systematic uncertainty. The data studied in this thesis were measured in the ATLAS experiment at a collision energy of 13 TeV with an integrated luminosity of 138.7 fb^{-1}.

The final definition of the dilepton control region is given in Table 6.1. Within this non-prompt control region the data are sufficiently well modeled by the Monte Carlo simulations used in this thesis. Only in the $e\mu$ channel is a significant deviation between data and the simulated events. Finding the origin of this deviation exceeds the scope of this thesis, but it can be assumed that it originates from the mismodeling of the non-prompt events. In order to ensure that the control region is suitable for the estimation of the non-prompt background in the signal region, the composition of non-prompt leptons was extensively compared between the dilepton control region and the signal region. In both regions, the main contribution of non-prompt leptons originates from $W + jets$ and $t\bar{t}$ events. Using the IFFTruthClassifier, MCTruthClassifier and the non-prompt contribution of the Monte Carlo samples it is confirmed that the origins of non-prompt leptons within the dilepton control region correspond very well with the composition of non-prompt leptons in the signal region. Although the selection criteria of the dilepton control region were chosen to maximize the proportion of non-prompt leptons, the numerator of the fake factor is dominated by prompt and charge flip leptons, which results in a large statistical uncertainty of the fake factor. This uncertainty could not be further reduced by additional selection criteria and is therefore a limiting factor of the data-driven method presented in this thesis.

Since, in contrast to the dijet control region, two leptons must be considered in the dilepton control region, the extraction of the fake factor is not trivial. An extraction method is derived for the dilepton control region, which does not lead to a systematic bias of the fake factor. In order to reduce the statistical uncertainty of the fake factor, other extraction methods are also tested, but none of these methods is found to be superior to the unbiased extraction method. However, this study has also shown that as soon as there is a physical criteria to sufficiently distinguish prompt/charge flip and non-prompt leptons, the alternative extraction procedure provides a smaller statistical uncertainty of the fake factor.

The chosen setup is tested by applying the fake factor within the control region, using only Monte Carlo simulated events for both extraction and application of the fake factor. Since there is a consistency with the non-prompt background predicted by Monte Carlo simulations, serious systematic complications of the setup are very unlikely. According to this closure test the electron fake factor is expected to introduce a small but significant overestimation of the non-prompt electron background for η close to zero. A correction of this effect was already studied in [23] for the dilepton control region but since the results of this study have only recently become available, this correction is not yet applied in this thesis. In addition, also a study of the overlap removal between jets and leptons is missing for the data-driven method used in this thesis. The jet distributions for events with non-prompt leptons were tested to be well modeled by the data-driven method applied for Monte Carlo simulated events, but the same charged $W^{\pm}$ scattering analysis in [2] recommends a slightly adjusted overlap removal for non-prompt studies.

The only additional correction applied in this thesis is a correction of the lepton transverse momentum. The necessity of this correction is not confirmed in this thesis but nevertheless the p_T correction is applied in order to stay consistent with the study in the dijet non-prompt control region. For this purpose, the momentum p_T^{cone} in a cone around the track of the lepton is added to it's reconstructed transverse momentum.

As final test, the data-driven method is applied in the low dijet invariant mass validation region, which is kinematically very close to the signal region but also orthogonal to the signal region and the dilepton control region. At first the data-driven method is successfully performed only for Monte Carlo simulated events to ensure a working setup. Then, for the first time within this thesis, the fake factor is extracted from actual data of the dilepton control region defined in this thesis. When applying the data-driven method in the low dijet invariant mass validation region, it is shown that the data is sufficiently well modeled by the sum of the data-drivenly estimated non-prompt background and the prompt and charge flip contribution estimated by Monte Carlo simulations. This good agreement confirms that the data-driven method described in this thesis is suitable for the estimation of the non-prompt background within the signal region.

However, a final evaluation of the data-driven method using the dilepton control region is still pending. So far, only the statistical uncertainty is considered and thus the systematic uncertainty of the fake factor has to be estimated to get the full uncertainty of the estimated non-prompt background. The results of the data-driven method presented in this thesis must then be compared with the current study of the data-driven estimation using the dijet control region. The decisive factors in this comparison are the extrapolation between the respective control region and the signal region and the final uncertainty of the estimated non-prompt background. Although the final comparison is still pending, it can be assumed, as confirmed by the closure test in the low dijet invariant mass validation region, that the extrapolation between the dilepton control region defined in this thesis and the signal region is relatively small. Thus, it is expected that the results of this thesis will at least provide a valuable crosscheck to the current analysis of electroweak same charged $W^{\pm}$ scattering.

Appendix A

Tables and Figures

Data set
data15__13TeV.periodD.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data15__13TeV.periodE.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data15__13TeV.periodF.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data15__13TeV.periodG.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data15__13TeV.periodH.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data15__13TeV.periodJ.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data16__13TeV.periodA.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data16__13TeV.periodB.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data16__13TeV.periodC.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data16__13TeV.periodD.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data16__13TeV.periodE.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data16__13TeV.periodF.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data16__13TeV.periodG.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data16__13TeV.periodI.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data16__13TeV.periodK.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data16__13TeV.periodL.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data17__13TeV.periodB.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data17__13TeV.periodC.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data17__13TeV.periodD.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data17__13TeV.periodE.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data17__13TeV.periodF.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data17__13TeV.periodH.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data17__13TeV.periodI.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data17__13TeV.periodK.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data18__13TeV.periodB.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data18__13TeV.periodC.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data18__13TeV.periodD.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data18__13TeV.periodF.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data18__13TeV.periodI.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data18__13TeV.periodK.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data18__13TeV.periodL.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data18__13TeV.periodM.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data18__13TeV.periodO.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984
data18__13TeV.periodQ.physics__Main.PhysCont.DAOD__STDM3.grp23__v01__p3984

Table A.1: List of all Run 2 data samples used in this thesis

Monte Carlo sample
mc16__13TeV.361292.MGaMcAtNloPy8EG__NNPDF30LO__A14NNPDF23LO__WZ__lvll__FxFx.deriv.DAOD__STDM3.e7551__s3126__r10201__p4097
mc16__13TeV.361292.MGaMcAtNloPy8EG__NNPDF30LO__A14NNPDF23LO__WZ__lvll__FxFx.deriv.DAOD__STDM3.e7551__s3126__r10724__p4097
mc16__13TeV.361292.MGaMcAtNloPy8EG__NNPDF30LO__A14NNPDF23LO__WZ__lvll__FxFx.deriv.DAOD__STDM3.e7551__s3126__r9364__p4097
mc16__13TeV.361293.MGaMcAtNloPy8EG__NNPDF30LO__A14NNPDF23LO__WZ__lvll__FxFx.deriv.DAOD__STDM3.e7958__e5984__s3126__r9364__r9315__p4097
mc16__13TeV.361510.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__Np0.deriv.DAOD__STDM3.e3898__s3126__r10201__p3975
mc16__13TeV.361510.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__Np0.deriv.DAOD__STDM3.e3898__s3126__r10724__p3975
mc16__13TeV.361510.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__Np0.deriv.DAOD__STDM3.e3898__s3126__r9364__p3975
mc16__13TeV.361511.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__Np1.deriv.DAOD__STDM3.e3898__s3126__r10201__p3975
mc16__13TeV.361511.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__Np1.deriv.DAOD__STDM3.e3898__s3126__r10724__p3975
mc16__13TeV.361511.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__Np1.deriv.DAOD__STDM3.e3898__s3126__r9364__p3975
mc16__13TeV.361512.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__Np2.deriv.DAOD__STDM3.e3898__s3126__r10201__p3975
mc16__13TeV.361512.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__Np2.deriv.DAOD__STDM3.e3898__s3126__r10724__p3975
mc16__13TeV.361512.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__Np2.deriv.DAOD__STDM3.e3898__s3126__r9364__p3975
mc16__13TeV.361513.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__Np3.deriv.DAOD__STDM3.e3898__s3126__r10201__p3975
mc16__13TeV.361513.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__Np3.deriv.DAOD__STDM3.e3898__s3126__r10724__p3975
mc16__13TeV.361513.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__Np3.deriv.DAOD__STDM3.e3898__s3126__r9364__p3975
mc16__13TeV.361514.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__Np4.deriv.DAOD__STDM3.e3898__s3126__r10201__p3975
mc16__13TeV.361514.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__Np4.deriv.DAOD__STDM3.e3898__s3126__r10724__p3975
mc16__13TeV.361514.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__Np4.deriv.DAOD__STDM3.e3898__s3126__r9364__p3975
mc16__13TeV.361638.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__lowMll__Np0.deriv.DAOD__STDM3.e7684__s3126__r10201__p3975
mc16__13TeV.361638.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__lowMll__Np0.deriv.DAOD__STDM3.e7684__s3126__r10724__p3975
mc16__13TeV.361638.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__lowMll__Np0.deriv.DAOD__STDM3.e7684__s3126__r9364__p3975
mc16__13TeV.361639.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__lowMll__Np1.deriv.DAOD__STDM3.e7684__s3126__r10201__p3975
mc16__13TeV.361639.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__lowMll__Np1.deriv.DAOD__STDM3.e7684__s3126__r10724__p3975
mc16__13TeV.361639.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__lowMll__Np1.deriv.DAOD__STDM3.e7684__s3126__r9364__p3975
mc16__13TeV.361640.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__lowMll__Np2.deriv.DAOD__STDM3.e7684__s3126__r10201__p3975
mc16__13TeV.361640.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__lowMll__Np2.deriv.DAOD__STDM3.e7684__s3126__r10724__p3975
mc16__13TeV.361640.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__lowMll__Np2.deriv.DAOD__STDM3.e7684__s3126__r9364__p3975
mc16__13TeV.361641.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__lowMll__Np3.deriv.DAOD__STDM3.e4442__s3126__r10201__p3975
mc16__13TeV.361641.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__lowMll__Np3.deriv.DAOD__STDM3.e4442__s3126__r10724__p3975
mc16__13TeV.361641.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__lowMll__Np3.deriv.DAOD__STDM3.e4442__s3126__r9364__p3975
mc16__13TeV.361642.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__lowMll__Np4.deriv.DAOD__STDM3.e4442__s3126__r10201__p3975
mc16__13TeV.361642.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__lowMll__Np4.deriv.DAOD__STDM3.e4442__s3126__r10724__p3975
mc16__13TeV.361642.MadGraphPythia8EvtGen__A14NNPDF23LO__Ztautau__lowMll__Np4.deriv.DAOD__STDM3.e4442__s3126__r9364__p3975
mc16__13TeV.363144.MGPy8EG__N30NLO__Zmumu__Ht[sliced]__CMS__CVetoBVeto.deriv.DAOD__STDM3.e4649__s3126__r10201__p3975
mc16__13TeV.363144.MGPy8EG__N30NLO__Zmumu__Ht[sliced]__CMS__CVetoBVeto.deriv.DAOD__STDM3.e4649__s3126__r10724__p3975
mc16__13TeV.363144.MGPy8EG__N30NLO__Zmumu__Ht[sliced]__CMS__CVetoBVeto.deriv.DAOD__STDM3.e4649__s3126__r9364__p3975
mc16__13TeV.363145.MGPy8EG__N30NLO__Zmumu__Ht[sliced]__CMS__CFilterBVeto.deriv.DAOD__STDM3.e4649__s3126__r10201__p3975
mc16__13TeV.363145.MGPy8EG__N30NLO__Zmumu__Ht[sliced]__CMS__CFilterBVeto.deriv.DAOD__STDM3.e4649__s3126__r10724__p3975
mc16__13TeV.363145.MGPy8EG__N30NLO__Zmumu__Ht[sliced]__CMS__CFilterBVeto.deriv.DAOD__STDM3.e4649__s3126__r9364__p3975
mc16__13TeV.363146.MGPy8EG__N30NLO__Zmumu__Ht[sliced]__CMS__BFilter.deriv.DAOD__STDM3.e4649__s3126__r10201__p3975

Monte Carlo sample
mc16_13TeV.363146.MGPy8EG_N30NLO_Zmumu_Ht[sliced]_CMS_BFilter.deriv.DAOD_STDM3.e4649_s3126_r10724_p3975
mc16_13TeV.363146.MGPy8EG_N30NLO_Zmumu_Ht[sliced]_CMS_BFilter.deriv.DAOD_STDM3.e4649_s3126_r9364_p3975
mc16_13TeV.363168.MGPy8EG_N30NLO_Zee_Ht[sliced]_CMS_CVetoBVeto.deriv.DAOD_STDM3.e4866_s3126_r10201_p3975
mc16_13TeV.363168.MGPy8EG_N30NLO_Zee_Ht[sliced]_CMS_CVetoBVeto.deriv.DAOD_STDM3.e4866_s3126_r10724_p3975
mc16_13TeV.363168.MGPy8EG_N30NLO_Zee_Ht[sliced]_CMS_CVetoBVeto.deriv.DAOD_STDM3.e4866_s3126_r9364_p3975
mc16_13TeV.363169.MGPy8EG_N30NLO_Zee_Ht[sliced]_CMS_CFilterBVeto.deriv.DAOD_STDM3.e4866_s3126_r10201_p3975
mc16_13TeV.363169.MGPy8EG_N30NLO_Zee_Ht[sliced]_CMS_CFilterBVeto.deriv.DAOD_STDM3.e4866_s3126_r10724_p3975
mc16_13TeV.363169.MGPy8EG_N30NLO_Zee_Ht[sliced]_CMS_CFilterBVeto.deriv.DAOD_STDM3.e4866_s3126_r9364_p3975
mc16_13TeV.363170.MGPy8EG_N30NLO_Zee_Ht[sliced]_CMS_BFilter.deriv.DAOD_STDM3.e4866_s3126_r10201_p3975
mc16_13TeV.363170.MGPy8EG_N30NLO_Zee_Ht[sliced]_CMS_BFilter.deriv.DAOD_STDM3.e4866_s3126_r10724_p3975
mc16_13TeV.363170.MGPy8EG_N30NLO_Zee_Ht[sliced]_CMS_BFilter.deriv.DAOD_STDM3.e4866_s3126_r9364_p3975
mc16_13TeV.363507.Sherpa_222_NNPDF30NNLO_WWZ_3l1v2j_EW6.deriv.DAOD_STDM3.e6379_s3126_r10201_p3975
mc16_13TeV.363507.Sherpa_222_NNPDF30NNLO_WWZ_3l1v2j_EW6.deriv.DAOD_STDM3.e6379_s3126_r9364_p3975
mc16_13TeV.363508.Sherpa_222_NNPDF30NNLO_WZZ_4l2j_EW6.deriv.DAOD_STDM3.e6374_s3126_r10201_p3975
mc16_13TeV.363508.Sherpa_222_NNPDF30NNLO_WZZ_4l2j_EW6.deriv.DAOD_STDM3.e6374_s3126_r9364_p3975
mc16_13TeV.363509.Sherpa_222_NNPDF30NNLO_WZZ_3l1v2j_EW6.deriv.DAOD_STDM3.e6374_s3126_r10201_p3975
mc16_13TeV.363509.Sherpa_222_NNPDF30NNLO_WZZ_3l1v2j_EW6.deriv.DAOD_STDM3.e6374_s3126_r9364_p3975
mc16_13TeV.363621.MGPy8EG_N30NLO_Wenu_Ht[sliced]_CMS_CVetoBVeto.deriv.DAOD_STDM3.e5136_s3126_r10201_p3975
mc16_13TeV.363621.MGPy8EG_N30NLO_Wenu_Ht[sliced]_CMS_CVetoBVeto.deriv.DAOD_STDM3.e5136_s3126_r10724_p3975
mc16_13TeV.363621.MGPy8EG_N30NLO_Wenu_Ht[sliced]_CMS_CVetoBVeto.deriv.DAOD_STDM3.e5136_s3126_r9364_p3975
mc16_13TeV.363622.MGPy8EG_N30NLO_Wenu_Ht[sliced]_CMS_CFilterBVeto.deriv.DAOD_STDM3.e5136_s3126_r10201_p3975
mc16_13TeV.363622.MGPy8EG_N30NLO_Wenu_Ht[sliced]_CMS_CFilterBVeto.deriv.DAOD_STDM3.e5136_s3126_r10724_p3975
mc16_13TeV.363622.MGPy8EG_N30NLO_Wenu_Ht[sliced]_CMS_CFilterBVeto.deriv.DAOD_STDM3.e5136_s3126_r9364_p3975
mc16_13TeV.363623.MGPy8EG_N30NLO_Wenu_Ht[sliced]_CMS_BFilter.deriv.DAOD_STDM3.e4835_s3126_r10201_p3975
mc16_13TeV.363623.MGPy8EG_N30NLO_Wenu_Ht[sliced]_CMS_BFilter.deriv.DAOD_STDM3.e4835_s3126_r10724_p3975
mc16_13TeV.363623.MGPy8EG_N30NLO_Wenu_Ht[sliced]_CMS_BFilter.deriv.DAOD_STDM3.e4835_s3126_r9364_p3975
mc16_13TeV.363645.MGPy8EG_N30NLO_Wmunu_Ht[sliced]_CMS_CVetoBVeto.deriv.DAOD_STDM3.e5136_s3126_r10201_p3975
mc16_13TeV.363645.MGPy8EG_N30NLO_Wmunu_Ht[sliced]_CMS_CVetoBVeto.deriv.DAOD_STDM3.e5136_s3126_r10724_p3975
mc16_13TeV.363645.MGPy8EG_N30NLO_Wmunu_Ht[sliced]_CMS_CVetoBVeto.deriv.DAOD_STDM3.e5136_s3126_r9364_p3975
mc16_13TeV.363646.MGPy8EG_N30NLO_Wmunu_Ht[sliced]_CMS_CFilterBVeto.deriv.DAOD_STDM3.e5136_s3126_r10201_p3975
mc16_13TeV.363646.MGPy8EG_N30NLO_Wmunu_Ht[sliced]_CMS_CFilterBVeto.deriv.DAOD_STDM3.e5136_s3126_r10724_p3975
mc16_13TeV.363646.MGPy8EG_N30NLO_Wmunu_Ht[sliced]_CMS_CFilterBVeto.deriv.DAOD_STDM3.e5136_s3126_r9364_p3975
mc16_13TeV.363647.MGPy8EG_N30NLO_Wmunu_Ht[sliced]_CMS_BFilter.deriv.DAOD_STDM3.e5136_s3126_r10201_p3975
mc16_13TeV.363647.MGPy8EG_N30NLO_Wmunu_Ht[sliced]_CMS_BFilter.deriv.DAOD_STDM3.e5136_s3126_r10724_p3975
mc16_13TeV.363647.MGPy8EG_N30NLO_Wmunu_Ht[sliced]_CMS_BFilter.deriv.DAOD_STDM3.e5136_s3126_r9364_p3975
mc16_13TeV.363669.MGPy8EG_N30NLO_Wtaunu_Ht[sliced]_CMS_CVetoBVeto.deriv.DAOD_STDM3.e5136_s3126_r10201_p3975
mc16_13TeV.363669.MGPy8EG_N30NLO_Wtaunu_Ht[sliced]_CMS_CVetoBVeto.deriv.DAOD_STDM3.e5136_s3126_r10724_p3975
mc16_13TeV.363669.MGPy8EG_N30NLO_Wtaunu_Ht[sliced]_CMS_CVetoBVeto.deriv.DAOD_STDM3.e5136_s3126_r9364_p3975
mc16_13TeV.363670.MGPy8EG_N30NLO_Wtaunu_Ht[sliced]_CMS_CFilterBVeto.deriv.DAOD_STDM3.e5136_s3126_r10201_p3975
mc16_13TeV.363670.MGPy8EG_N30NLO_Wtaunu_Ht[sliced]_CMS_CFilterBVeto.deriv.DAOD_STDM3.e5136_s3126_r10724_p3975
mc16_13TeV.363670.MGPy8EG_N30NLO_Wtaunu_Ht[sliced]_CMS_CFilterBVeto.deriv.DAOD_STDM3.e5136_s3126_r9364_p3975
mc16_13TeV.363671.MGPy8EG_N30NLO_Wtaunu_Ht[sliced]_CMS_BFilter.deriv.DAOD_STDM3.e5136_s3126_r10201_p3975
mc16_13TeV.363671.MGPy8EG_N30NLO_Wtaunu_Ht[sliced]_CMS_BFilter.deriv.DAOD_STDM3.e5136_s3126_r10724_p3975
mc16_13TeV.363671.MGPy8EG_N30NLO_Wtaunu_Ht[sliced]_CMS_BFilter.deriv.DAOD_STDM3.e5136_s3126_r9364_p3975
mc16_13TeV.364242.Sherpa_222_NNPDF30NNLO_WWW_3l3v_EW6.deriv.DAOD_STDM3.e5887_s3126_r10201_p3975
mc16_13TeV.364242.Sherpa_222_NNPDF30NNLO_WWW_3l3v_EW6.deriv.DAOD_STDM3.e5887_s3126_r10724_p3975
mc16_13TeV.364242.Sherpa_222_NNPDF30NNLO_WWW_3l3v_EW6.deriv.DAOD_STDM3.e5887_s3126_r9364_p3975
mc16_13TeV.364243.Sherpa_222_NNPDF30NNLO_WWZ_4l2v_EW6.deriv.DAOD_STDM3.e5887_s3126_r10201_p3975
mc16_13TeV.364243.Sherpa_222_NNPDF30NNLO_WWZ_4l2v_EW6.deriv.DAOD_STDM3.e5887_s3126_r10724_p3975
mc16_13TeV.364243.Sherpa_222_NNPDF30NNLO_WWZ_4l2v_EW6.deriv.DAOD_STDM3.e5887_s3126_r9364_p3975
mc16_13TeV.364244.Sherpa_222_NNPDF30NNLO_WWZ_2l4v_EW6.deriv.DAOD_STDM3.e5887_s3126_r10201_p3975
mc16_13TeV.364244.Sherpa_222_NNPDF30NNLO_WWZ_2l4v_EW6.deriv.DAOD_STDM3.e5887_s3126_r10724_p3975
mc16_13TeV.364244.Sherpa_222_NNPDF30NNLO_WWZ_2l4v_EW6.deriv.DAOD_STDM3.e5887_s3126_r9364_p3975
mc16_13TeV.364245.Sherpa_222_NNPDF30NNLO_WZZ_5l1v_EW6.deriv.DAOD_STDM3.e5887_s3126_r10201_p3975
mc16_13TeV.364245.Sherpa_222_NNPDF30NNLO_WZZ_5l1v_EW6.deriv.DAOD_STDM3.e5887_s3126_r10724_p3975
mc16_13TeV.364245.Sherpa_222_NNPDF30NNLO_WZZ_5l1v_EW6.deriv.DAOD_STDM3.e5887_s3126_r9364_p3975
mc16_13TeV.364246.Sherpa_222_NNPDF30NNLO_WZZ_3l3v_EW6.deriv.DAOD_STDM3.e5887_s3126_r10201_p3975
mc16_13TeV.364246.Sherpa_222_NNPDF30NNLO_WZZ_3l3v_EW6.deriv.DAOD_STDM3.e5887_s3126_r10724_p3975
mc16_13TeV.364246.Sherpa_222_NNPDF30NNLO_WZZ_3l3v_EW6.deriv.DAOD_STDM3.e5887_s3126_r9364_p3975
mc16_13TeV.364247.Sherpa_222_NNPDF30NNLO_ZZZ_6l0v_EW6.deriv.DAOD_STDM3.e5887_s3126_r10201_p3975
mc16_13TeV.364247.Sherpa_222_NNPDF30NNLO_ZZZ_6l0v_EW6.deriv.DAOD_STDM3.e5887_s3126_r10724_p3975
mc16_13TeV.364247.Sherpa_222_NNPDF30NNLO_ZZZ_6l0v_EW6.deriv.DAOD_STDM3.e5887_s3126_r9364_p3975
mc16_13TeV.364248.Sherpa_222_NNPDF30NNLO_ZZZ_4l2v_EW6.deriv.DAOD_STDM3.e5887_s3126_r10201_p3975
mc16_13TeV.364248.Sherpa_222_NNPDF30NNLO_ZZZ_4l2v_EW6.deriv.DAOD_STDM3.e5887_s3126_r10724_p3975
mc16_13TeV.364248.Sherpa_222_NNPDF30NNLO_ZZZ_4l2v_EW6.deriv.DAOD_STDM3.e5887_s3126_r9364_p3975
mc16_13TeV.364249.Sherpa_222_NNPDF30NNLO_ZZZ_2l4v_EW6.deriv.DAOD_STDM3.e5887_s3126_r10201_p3975
mc16_13TeV.364249.Sherpa_222_NNPDF30NNLO_ZZZ_2l4v_EW6.deriv.DAOD_STDM3.e5887_s3126_r10724_p3975
mc16_13TeV.364249.Sherpa_222_NNPDF30NNLO_ZZZ_2l4v_EW6.deriv.DAOD_STDM3.e5887_s3126_r9364_p3975
mc16_13TeV.364250.Sherpa_222_NNPDF30NNLO_llll.deriv.DAOD_STDM3.e5894_s3126_r9364_p3975
mc16_13TeV.364253.Sherpa_222_NNPDF30NNLO_lllv.deriv.DAOD_STDM3.e5916_s3126_r10201_p3975
mc16_13TeV.364253.Sherpa_222_NNPDF30NNLO_lllv.deriv.DAOD_STDM3.e5916_s3126_r10724_p3975
mc16_13TeV.364253.Sherpa_222_NNPDF30NNLO_lllv.deriv.DAOD_STDM3.e5916_s3126_r9364_p3975
mc16_13TeV.364254.Sherpa_222_NNPDF30NNLO_llvv.deriv.DAOD_STDM3.e5916_s3126_r10201_p3975
mc16_13TeV.364254.Sherpa_222_NNPDF30NNLO_llvv.deriv.DAOD_STDM3.e5916_s3126_r10724_p3975
mc16_13TeV.364254.Sherpa_222_NNPDF30NNLO_llvv.deriv.DAOD_STDM3.e5916_s3126_r9364_p3975
mc16_13TeV.364286.Sherpa_222_NNPDF30NNLO_llvvjj_ss_EW4.deriv.DAOD_STDM3.e6055_s3126_r10201_p3975
mc16_13TeV.364286.Sherpa_222_NNPDF30NNLO_llvvjj_ss_EW4.deriv.DAOD_STDM3.e6055_s3126_r10724_p3975
mc16_13TeV.364286.Sherpa_222_NNPDF30NNLO_llvvjj_ss_EW4.deriv.DAOD_STDM3.e6055_s3126_r9364_p3975
mc16_13TeV.364287.Sherpa_222_NNPDF30NNLO_llvvjj_ss_EW6.deriv.DAOD_STDM3.e6055_s3126_r10201_p3975
mc16_13TeV.364287.Sherpa_222_NNPDF30NNLO_llvvjj_ss_EW6.deriv.DAOD_STDM3.e6055_s3126_r10724_p3975
mc16_13TeV.364287.Sherpa_222_NNPDF30NNLO_llvvjj_ss_EW6.deriv.DAOD_STDM3.e6055_s3126_r9364_p3975
mc16_13TeV.364525.Sherpa_222_NNPDF30NNLO_enugamma_pty_[sliced]_CMS.deriv.DAOD_STDM3.e5928_s3126_r10201_p3975
mc16_13TeV.364525.Sherpa_222_NNPDF30NNLO_enugamma_pty_[sliced]_CMS.deriv.DAOD_STDM3.e5928_s3126_r10724_p3975
mc16_13TeV.364525.Sherpa_222_NNPDF30NNLO_enugamma_pty_[sliced]_CMS.deriv.DAOD_STDM3.e5928_s3126_r9364_p3975
mc16_13TeV.364530.Sherpa_222_NNPDF30NNLO_munugamma_pty_[sliced]_CMS.deriv.DAOD_STDM3.e5928_s3126_r10201_p3975
mc16_13TeV.364530.Sherpa_222_NNPDF30NNLO_munugamma_pty_[sliced]_CMS.deriv.DAOD_STDM3.e5928_s3126_r10724_p3975
mc16_13TeV.364530.Sherpa_222_NNPDF30NNLO_munugamma_pty_[sliced]_CMS.deriv.DAOD_STDM3.e5928_s3126_r9364_p3975
mc16_13TeV.364535.Sherpa_222_NNPDF30NNLO_taunugamma_pty_[sliced]_CMS.deriv.DAOD_STDM3.e5928_s3126_r10201_p3975
mc16_13TeV.364535.Sherpa_222_NNPDF30NNLO_taunugamma_pty_[sliced]_CMS.deriv.DAOD_STDM3.e5928_s3126_r10724_p3975
mc16_13TeV.364535.Sherpa_222_NNPDF30NNLO_taunugamma_pty_[sliced]_CMS.deriv.DAOD_STDM3.e5928_s3126_r9364_p3975
mc16_13TeV.364739.MGPy8EG_NNPDF30NLO_A14NNPDF23LO_lvlljjEW6_OFMinus.deriv.DAOD_STDM3.e7421_s3126_r10201_p3975
mc16_13TeV.364739.MGPy8EG_NNPDF30NLO_A14NNPDF23LO_lvlljjEW6_OFMinus.deriv.DAOD_STDM3.e7421_s3126_r10724_p3975
mc16_13TeV.364739.MGPy8EG_NNPDF30NLO_A14NNPDF23LO_lvlljjEW6_OFMinus.deriv.DAOD_STDM3.e7421_s3126_r9364_p3975
mc16_13TeV.364740.MGPy8EG_NNPDF30NLO_A14NNPDF23LO_lvlljjEW6_OFPlus.deriv.DAOD_STDM3.e7421_s3126_r10201_p3975
mc16_13TeV.364740.MGPy8EG_NNPDF30NLO_A14NNPDF23LO_lvlljjEW6_OFPlus.deriv.DAOD_STDM3.e7421_s3126_r10724_p3975
mc16_13TeV.364740.MGPy8EG_NNPDF30NLO_A14NNPDF23LO_lvlljjEW6_OFPlus.deriv.DAOD_STDM3.e7421_s3126_r9364_p3975
mc16_13TeV.364741.MGPy8EG_NNPDF30NLO_A14NNPDF23LO_lvlljjEW6_SFMinus.deriv.DAOD_STDM3.e7421_s3126_r10201_p3975
mc16_13TeV.364741.MGPy8EG_NNPDF30NLO_A14NNPDF23LO_lvlljjEW6_SFMinus.deriv.DAOD_STDM3.e7421_s3126_r10724_p3975
mc16_13TeV.364741.MGPy8EG_NNPDF30NLO_A14NNPDF23LO_lvlljjEW6_SFMinus.deriv.DAOD_STDM3.e7421_s3126_r9364_p3975
mc16_13TeV.364742.MGPy8EG_NNPDF30NLO_A14NNPDF23LO_lvlljjEW6_SFPlus.deriv.DAOD_STDM3.e7421_s3126_r10201_p3975
mc16_13TeV.364742.MGPy8EG_NNPDF30NLO_A14NNPDF23LO_lvlljjEW6_SFPlus.deriv.DAOD_STDM3.e7421_s3126_r10724_p3975
mc16_13TeV.364742.MGPy8EG_NNPDF30NLO_A14NNPDF23LO_lvlljjEW6_SFPlus.deriv.DAOD_STDM3.e7421_s3126_r9364_p3975
mc16_13TeV.410015.PowhegPythiaEvtGen_P2012_Wt_dilepton_top.deriv.DAOD_STDM3.e3753_s3126_r10201_p3975
mc16_13TeV.410015.PowhegPythiaEvtGen_P2012_Wt_dilepton_top.deriv.DAOD_STDM3.e3753_s3126_r9364_p3975
mc16_13TeV.410016.PowhegPythiaEvtGen_P2012_Wt_dilepton_antitop.deriv.DAOD_STDM3.e3753_s3126_r10201_p3975
mc16_13TeV.410016.PowhegPythiaEvtGen_P2012_Wt_dilepton_antitop.deriv.DAOD_STDM3.e3753_s3126_r9364_p3975
mc16_13TeV.410155.aMcAtNloPythia8EvtGen_MEN30NLO_A14N23LO_ttW.deriv.DAOD_STDM3.e5070_s3126_r10201_p3975
mc16_13TeV.410155.aMcAtNloPythia8EvtGen_MEN30NLO_A14N23LO_ttW.deriv.DAOD_STDM3.e5070_s3126_r10724_p3975
mc16_13TeV.410155.aMcAtNloPythia8EvtGen_MEN30NLO_A14N23LO_ttW.deriv.DAOD_STDM3.e5070_s3126_r9364_p3975
mc16_13TeV.410218.aMcAtNloPythia8EvtGen_MEN30NLO_A14N23LO_ttee.deriv.DAOD_STDM3.e5070_s3126_r10201_p3975
mc16_13TeV.410218.aMcAtNloPythia8EvtGen_MEN30NLO_A14N23LO_ttee.deriv.DAOD_STDM3.e5070_s3126_r10724_p3975
mc16_13TeV.410218.aMcAtNloPythia8EvtGen_MEN30NLO_A14N23LO_ttee.deriv.DAOD_STDM3.e5070_s3126_r9364_p3975
mc16_13TeV.410219.aMcAtNloPythia8EvtGen_MEN30NLO_A14N23LO_ttmumu.deriv.DAOD_STDM3.e5070_s3126_r10201_p3975
mc16_13TeV.410219.aMcAtNloPythia8EvtGen_MEN30NLO_A14N23LO_ttmumu.deriv.DAOD_STDM3.e5070_s3126_r10724_p3975
mc16_13TeV.410219.aMcAtNloPythia8EvtGen_MEN30NLO_A14N23LO_ttmumu.deriv.DAOD_STDM3.e5070_s3126_r9364_p3975
mc16_13TeV.410220.aMcAtNloPythia8EvtGen_MEN30NLO_A14N23LO_tttautau.deriv.DAOD_STDM3.e5070_s3126_r10201_p3975
mc16_13TeV.410220.aMcAtNloPythia8EvtGen_MEN30NLO_A14N23LO_tttautau.deriv.DAOD_STDM3.e5070_s3126_r10724_p3975
mc16_13TeV.410220.aMcAtNloPythia8EvtGen_MEN30NLO_A14N23LO_tttautau.deriv.DAOD_STDM3.e5070_s3126_r9364_p3975
mc16_13TeV.410470.PhPy8EG_A14_ttbar_hdamp258p75_nonallhad.deriv.DAOD_STDM3.e6337_s3126_r10201_p3975
mc16_13TeV.410470.PhPy8EG_A14_ttbar_hdamp258p75_nonallhad.deriv.DAOD_STDM3.e6337_s3126_r10724_p3975
mc16_13TeV.410470.PhPy8EG_A14_ttbar_hdamp258p75_nonallhad.deriv.DAOD_STDM3.e6337_s3126_r9364_p3975
mc16_13TeV.410471.PhPy8EG_A14_ttbar_hdamp258p75_allhad.deriv.DAOD_STDM3.e6337_s3126_r10201_p3975
mc16_13TeV.410471.PhPy8EG_A14_ttbar_hdamp258p75_allhad.deriv.DAOD_STDM3.e6337_s3126_r10724_p3975
mc16_13TeV.410471.PhPy8EG_A14_ttbar_hdamp258p75_allhad.deriv.DAOD_STDM3.e6337_s3126_r9364_p3975
mc16_13TeV.410560.MadGraphPythia8EvtGen_A14_tZ_4fl_tchan_noAllHad.deriv.DAOD_STDM3.e5803_s3126_r10201_p3975
mc16_13TeV.410560.MadGraphPythia8EvtGen_A14_tZ_4fl_tchan_noAllHad.deriv.DAOD_STDM3.e5803_s3126_r10724_p3975
mc16_13TeV.410560.MadGraphPythia8EvtGen_A14_tZ_4fl_tchan_noAllHad.deriv.DAOD_STDM3.e5803_s3126_r9364_p3975

Table A.2: List of all Monte Carlo samples used in this thesis. The slicing range of samples is replaced by "[sliced]".

	Veto	Non-Analysis	Analysis
ID	$p_T < 25$ GeV: Medium $p_T > 25$ GeV: Tight	MediumLH	TightLH
p_T	$p_T > 6$ GeV	$p_T > 27$ GeV	$p_T > 27$ GeV
η	$\lvert\eta\rvert < 2.47$	$\lvert\eta\rvert < 2.47$, excluding $1.37 < \lvert\eta\rvert < 1.52$	$\lvert\eta\rvert < 2.47$, excluding $1.37 < \lvert\eta\rvert < 1.52$
$\lvert z_0 \cdot \sin\theta\rvert$	$\lvert z_0 \cdot \sin\theta\rvert < 0.5$ mm	$\lvert z_0 \cdot \sin\theta\rvert < 0.5$ mm	$\lvert z_0 \cdot \sin\theta\rvert < 0.5$ mm
$\frac{d_0}{\sigma_{d_0}}$	$\frac{d_0}{\sigma_{d_0}} < 5$	$\frac{d_0}{\sigma_{d_0}} < 5$	$\frac{d_0}{\sigma_{d_0}} < 5$
Isolation	Gradient	-	FCTight
Author	-	-	1
Additional requirements	Not the leading or subleading Ana/Non-Ana lepton	Fail any Analysis selection	-

Table A.3: Object selection for Ana, Non-Ana and Veto electrons[2]. The selection of Ana and Veto electrons is based on the ATLAS note for same charge $W^{\pm}$ boson scattering from 2018[2]

	Veto	Non-Analysis	Analysis
ID	Loose	Medium	Medium
p_T	$p_T > 6$ GeV	$p_T > 27$ GeV	$p_T > 27$ GeV
η	$\lvert\eta\rvert < 2.7$	$\lvert\eta\rvert < 2.5$	$\lvert\eta\rvert < 2.5$
$\lvert z_0 \cdot \sin\theta\rvert$	$\lvert z_0 \cdot \sin\theta\rvert < 0.5$ mm	$\lvert z_0 \cdot \sin\theta\rvert < 0.5$ mm	$\lvert z_0 \cdot \sin\theta\rvert < 0.5$ mm
$\frac{d_0}{\sigma_{d_0}}$	$\frac{d_0}{\sigma_{d_0}} < 3$	$\frac{d_0}{\sigma_{d_0}} < 10$	$\frac{d_0}{\sigma_{d_0}} < 3$
Isolation	FCTight	-	FCTight
Additional requirements	Not the leading or subleading Ana/Non-Ana lepton	Fail any Analysis selection	-

Table A.4: Object selection for Ana, Non-Ana and Veto muons[2]. The selection of Ana and Veto muons is based on the ATLAS note for same charge $W^{\pm}$ boson scattering from 2018[2]

Exactly two Ana leptons with identical electrical charge No Veto lepton
Electrons have to pass ECIDS
ee-channel: $\lvert\eta\rvert < 1.37$, $\lvert m_{ee} - m_Z\rvert > 15$ GeV
$m_{ll'} > 20$ GeV
$E_T^{miss} > 30$ GeV
At least two jets Leading jet: $p_T > 65$ GeV Subleading jet: $p_T > 35$ GeV
b-jet veto
$\lvert\Delta y_{jj}\rvert > 2$
500 GeV $> m_{jj} >$ 200 GeV

Table A.5: The event selection for the LowMjj control region of the electroweak same charged W boson scattering described in [2]. The event selection of this ATLAS note is extended by a ECIDS selection to suppress the charge flip background.

	ee	eμ	μe	$\mu\mu$	all
Without orthogonal cut:					
MC	574664	29240	26074	17944	647922
(non-prompt MC)/MC	2.4%	52.2%	45.4%	73.4%	8.3%
$\Delta(N - N^{prompt})/(N - N^{prompt})$	7.7%	1.4%	1.7%	1.1%	2.1%
$(n_{jet} < 2) \| (n_{bjet} > 0)$:					
MC	515911	22500	20361	13532	572305
(non-prompt MC)/MC	2.1%	51.6%	46.5%	75.6%	7.4%
$\Delta(N - N^{prompt})/(N - N^{prompt})$	9.4%	1.6%	1.9%	1.3%	2.5%
$(n_{jet} < 2) \| (\lvert\Delta y_{jj}\rvert < 2)$:					
MC	544492	23905	22127	14414	604938
(non-prompt MC)/MC	2.2%	51.9%	45.6%	72.1%	7.4%
$\Delta(N - N^{prompt})/(N - N^{prompt})$	8.8%	1.5%	1.8%	1.3%	2.4%
$(n_{jet} < 2) \| (m_{jj} < 200$ GeV$)$:					
MC	545011	21849	20998	13196	601054
(non-prompt MC)/MC	2.1%	51.4%	45.9%	71.4%	6.9%
$\Delta(N - N^{prompt})/(N - N^{prompt})$	9.2%	1.6%	1.9%	1.4%	2.6%
Fail the signal and LowMjj selection:					
MC	574296	28486	25491	17357	645630
(non-prompt MC)/MC	2.4%	52.0%	45.4%	73.6%	8.2%
$\Delta(N - N^{prompt})/(N - N^{prompt})$	7.8%	1.4%	1.7%	1.2%	2.1%

Table A.6: Comparison of different orthogonal selections for the definition of the dilepton control region. Shown are the total number of Monte Carlo simulated events passing the respective selection, the non-prompt fraction within these events and the relative statistical uncertainty in the total number of non-prompt events given by equation 6.1.

	ee	eμ	μe	$\mu\mu$	all
No selections:					
MC	4.329e+07	638696	600398	7.562e+07	1.201e+08
(non-prompt MC)/MC	0.1%	7.2%	6.6%	0.0%	0.1%
$\Delta(N - N^{prompt})/(N - N^{prompt})$	22.1%	2.4%	2.7%	34.2%	9.5%
sameCharge:					
MC	576735	31870	28598	20657	657860
(non-prompt MC)/MC	2.4%	48.5%	41.8%	64.4%	8.3%
$\Delta(N - N^{prompt})/(N - N^{prompt})$	7.6%	1.4%	1.8%	1.3%	2.1%
$n_{lep} < 3$:					
MC	574664	29240	26074	17944	647922
(non-prompt MC)/MC	2.4%	52.2%	45.4%	73.4%	8.3%
$\Delta(N - N^{prompt})/(N - N^{prompt})$	7.7%	1.4%	1.7%	1.1%	2.1%
Fail the signal and LowMjj selection:					
MC	574296	28486	25491	17357	645630
(non-prompt MC)/MC	2.4%	52.0%	45.4%	73.6%	8.2%
$\Delta(N - N^{prompt})/(N - N^{prompt})$	7.8%	1.4%	1.7%	1.2%	2.1%
ECIDS:					
MC	56468	21935	20168	17357	115927
(non-prompt MC)/MC	19.7%	61.2%	50.9%	73.6%	41.0%
$\Delta(N - N^{prompt})/(N - N^{prompt})$	2.9%	1.3%	1.7%	1.2%	0.9%
79 GeV $< m_{ll} <$ 98 GeV in ee:					
MC	20950	21935	20168	17357	80409
(non-prompt MC)/MC	42.9%	61.2%	50.9%	73.6%	56.5%
$\Delta(N - N^{prompt})/(N - N^{prompt})$	2.0%	1.3%	1.7%	1.2%	0.7%
$E_T^{miss} >$ 19 GeV in ee:					
MC	15096	21935	20168	17357	74555
(non-prompt MC)/MC	50.9%	61.2%	50.9%	73.6%	59.2%
$\Delta(N - N^{prompt})/(N - N^{prompt})$	2.0%	1.3%	1.7%	1.2%	0.7%
$n_{vetolep} = 0$:					
MC	12888	16425	14669	12455	56437
(non-prompt MC)/MC	53.3%	67.1%	58.5%	81.9%	65.0%
$\Delta(N - N^{prompt})/(N - N^{prompt})$	2.0%	1.3%	1.7%	1.2%	0.8%
$n_{jet} > 0$ in $\mu\mu$:					
MC	12888	16425	14669	11436	55417
(non-prompt MC)/MC	53.3%	67.1%	58.5%	85.7%	65.5%
$\Delta(N - N^{prompt})/(N - N^{prompt})$	2.0%	1.3%	1.7%	1.2%	0.8%

Table A.7: Selection flow of the non-prompt control region splitted in the lepton channels. Listed are the total numbers of Monte Carlo events, the purity in non-prompt events and the relative statistical uncertainty in the total number of non-prompt leptons given by equation 6.1

	ee	eμ	μe	$\mu\mu$	all
No selections:					
MC	4.3e+07	504100	471315	7.558e+07	1.196e+08
(non-prompt MC)/MC	0.0%	1.3%	1.6%	0.0%	0.0%
$\Delta(N - N^{prompt})/(N - N^{prompt})$	120.2%	14.8%	13.1%	424.1%	62.5%
sameCharge:					
MC	309912	12746	12054	7972	342684
(non-prompt MC)/MC	0.6%	11.2%	11.5%	7.8%	1.6%
$\Delta(N - N^{prompt})/(N - N^{prompt})$	39.6%	10.9%	10.8%	19.9%	15.2%
$n_{lep} < 3$:					
MC	308010	10403	9781	5375	333569
(non-prompt MC)/MC	0.6%	13.0%	14.0%	11.3%	1.6%
$\Delta(N - N^{prompt})/(N - N^{prompt})$	41.9%	10.3%	9.8%	16.5%	15.6%
Fail the signal and LowMjj selection:					
MC	307800	10101	9499	5164	332565
(non-prompt MC)/MC	0.6%	12.9%	14.1%	11.4%	1.5%
$\Delta(N - N^{prompt})/(N - N^{prompt})$	42.5%	10.5%	9.9%	16.8%	16.0%
ECIDS:					
MC	34726	6736	7401	5164	54027
(non-prompt MC)/MC	4.5%	16.0%	16.3%	11.4%	8.2%
$\Delta(N - N^{prompt})/(N - N^{prompt})$	16.8%	10.3%	9.6%	16.8%	7.3%
79 GeV $< m_{ll} <$ 98 GeV in ee:					
MC	11354	6736	7401	5164	30655
(non-prompt MC)/MC	11.7%	16.0%	16.3%	11.4%	13.7%
$\Delta(N - N^{prompt})/(N - N^{prompt})$	11.0%	10.3%	9.6%	16.8%	5.7%
$E_T^{miss} >$ 19 GeV in ee:					
MC	7703	6736	7401	5164	27004
(non-prompt MC)/MC	14.9%	16.0%	16.3%	11.4%	14.9%
$\Delta(N - N^{prompt})/(N - N^{prompt})$	10.4%	10.3%	9.6%	16.8%	5.6%
$n_{vetolep} = 0$:					
MC	6256	4611	4914	2738	18519
(non-prompt MC)/MC	16.7%	19.2%	22.1%	17.9%	18.9%
$\Delta(N - N^{prompt})/(N - N^{prompt})$	10.3%	10.3%	8.6%	14.4%	5.2%
$n_{jet} > 0$ in $\mu\mu$:					
MC	6256	4611	4914	2103	17885
(non-prompt MC)/MC	16.7%	19.2%	22.1%	22.3%	19.5%
$\Delta(N - N^{prompt})/(N - N^{prompt})$	10.3%	10.3%	8.6%	13.0%	5.2%

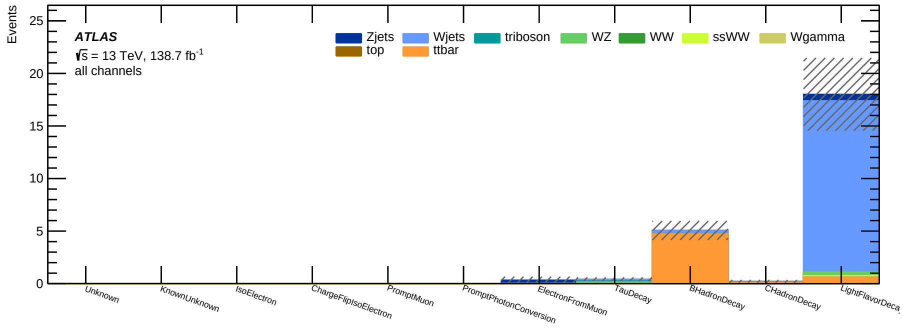
Table A.8: Selection flow of the non-prompt control region splitted in the lepton channels for events with two Ana leptons. Listed are the total numbers of Monte Carlo events, the purity in non-prompt events and the relative statistical uncertainty in the total number of non-prompt leptons given by equation 6.1

	ee	eμ	μe	μμ	all
Signal Region:					
Zjets	3.4 ± 3.0%	6.9 ± 3.2%	2.0 ± 2.0%	0.6 ± 0.6%	4.3 ± 1.9%
Wjets	53.3 ± 11.2%	69.0 ± 6.2%	50.6 ± 11.4%	16.9 ± 7.1%	54.6 ± 4.9%
WZ	2.5 ± 0.9%	0.5 ± 0.2%	0.5 ± 0.2%	0.0 ± 0.0%	0.6 ± 0.1%
Wgamma	2.5 ± 2.0%	3.4 ± 2.7%	0.0 ± 0.0%	0.0 ± 0.0%	2.0 ± 1.4%
WZ_EW	1.8 ± 0.6%	0.6 ± 0.1%	1.3 ± 0.4%	2.0 ± 0.4%	1.1 ± 0.1%
Wt	0.0 ± 0.0%	0.0 ± 0.0%	0.0 ± 0.0%	0.0 ± 0.0%	0.0 ± 0.0%
ttbar	36.5 ± 9.0%	19.5 ± 3.7%	45.6 ± 7.3%	80.5 ± 5.1%	37.4 ± 2.8%
Other	3.5 ± 1.2%	0.8 ± 0.2%	1.2 ± 0.4%	1.4 ± 0.4%	1.2 ± 0.2%
Control Region:					
Zjets	7.1 ± 1.4%	7.2 ± 1.3%	17.5 ± 9.4%	9.4 ± 3.9%	10.7 ± 3.3%
Wjets	64.7 ± 2.9%	53.4 ± 2.4%	51.6 ± 4.3%	18.8 ± 1.5%	51.6 ± 1.8%
WZ	0.3 ± 0.0%	0.2 ± 0.0%	0.2 ± 0.0%	0.0 ± 0.0%	0.2 ± 0.0%
Wgamma	2.5 ± 0.5%	2.0 ± 0.5%	1.0 ± 0.4%	0.4 ± 0.2%	1.6 ± 0.2%
WZ_EW	0.0 ± 0.0%	0.0 ± 0.0%	0.0 ± 0.0%	0.1 ± 0.0%	0.0 ± 0.0%
Wt	0.1 ± 0.1%	0.2 ± 0.1%	0.1 ± 0.1%	0.1 ± 0.1%	0.1 ± 0.0%
ttbar	25.2 ± 1.3%	36.9 ± 1.2%	29.5 ± 2.6%	71.1 ± 1.3%	35.7 ± 1.1%
Other	0.4 ± 0.0%	0.5 ± 0.1%	0.3 ± 0.0%	0.4 ± 0.0%	0.4 ± 0.0%

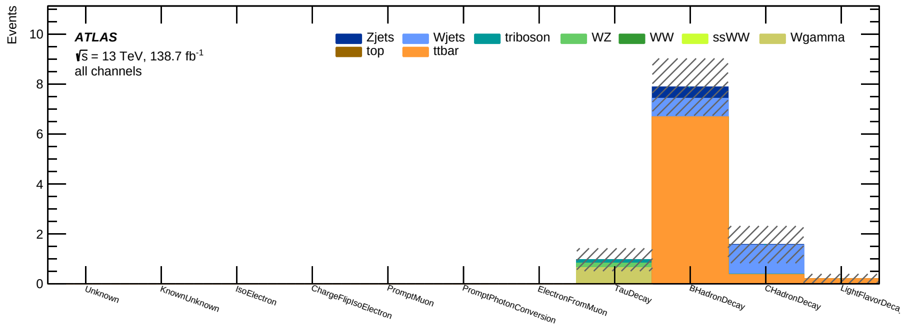
Table A.9: Comparison of signal region and control region in the composition of Monte Carlo samples. Only non-prompt events with two Ana leptons are considered. The shown uncertainties result from the statistical uncertainties of the Monte Carlo samples.

	ee	eμ	μe	μμ	all
Signal Region:					
Zjets	10.0 ± 3.5%	5.2 ± 2.0%	4.9 ± 2.3%	2.7 ± 1.7%	5.0 ± 1.0%
Wjets	63.8 ± 4.7%	35.0 ± 3.4%	54.3 ± 4.3%	19.8 ± 2.6%	38.5 ± 2.0%
WZ	0.2 ± 0.1%	0.1 ± 0.0%	0.1 ± 0.0%	0.0 ± 0.0%	0.1 ± 0.0%
Wgamma	0.0 ± 2.0%	2.2 ± 1.4%	3.7 ± 2.3%	0.0 ± 0.0%	1.6 ± 0.8%
WZ_EW	0.1 ± 0.0%	0.0 ± 0.0%	0.0 ± 0.0%	0.0 ± 0.0%	0.0 ± 0.0%
Wt	0.0 ± 0.0%	0.0 ± 0.0%	0.0 ± 0.0%	0.3 ± 0.3%	0.1 ± 0.1%
ttbar	25.9 ± 3.3%	57.4 ± 1.9%	37.0 ± 2.4%	77.3 ± 1.5%	54.7 ± 1.1%
Other	0.4 ± 0.1%	0.2 ± 0.1%	0.2 ± 0.1%	0.1 ± 0.0%	0.2 ± 0.0%
Control Region:					
Zjets	9.3 ± 1.2%	5.5 ± 0.4%	6.7 ± 2.6%	2.6 ± 0.4%	5.6 ± 0.7%
Wjets	55.9 ± 1.5%	28.4 ± 0.8%	49.9 ± 1.5%	12.7 ± 0.4%	33.8 ± 0.6%
WZ	0.0 ± 0.0%	0.0 ± 0.0%	0.0 ± 0.0%	0.0 ± 0.0%	0.0 ± 0.0%
Wgamma	1.1 ± 0.2%	0.7 ± 0.1%	0.9 ± 0.1%	0.2 ± 0.0%	0.7 ± 0.1%
WZ_EW	0.0 ± 0.0%	0.0 ± 0.0%	0.0 ± 0.0%	0.0 ± 0.0%	0.0 ± 0.0%
Wt	0.2 ± 0.0%	0.3 ± 0.0%	0.2 ± 0.0%	0.4 ± 0.0%	0.3 ± 0.0%
ttbar	33.5 ± 0.7%	64.9 ± 0.4%	42.3 ± 0.9%	84.2 ± 0.2%	59.6 ± 0.3%
Other	0.1 ± 0.0%	0.1 ± 0.0%	0.1 ± 0.0%	0.1 ± 0.0%	0.1 ± 0.0%

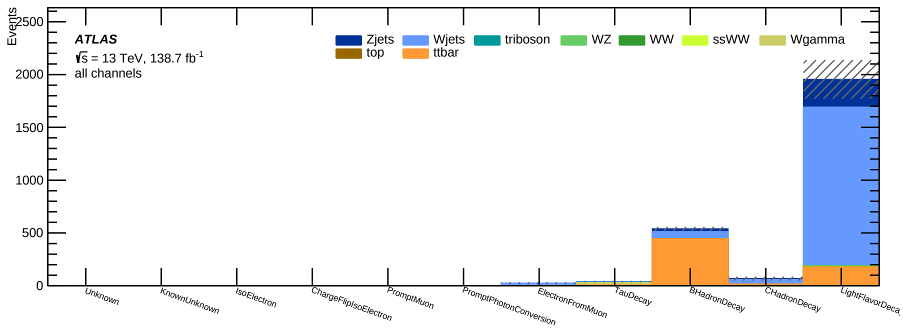
Table A.10: Comparison of signal region and control region in the composition of Monte Carlo samples. Only non-prompt events with one Ana and one Non-Ana lepton are considered. The shown uncertainties result from the statistical uncertainties of the Monte Carlo samples.



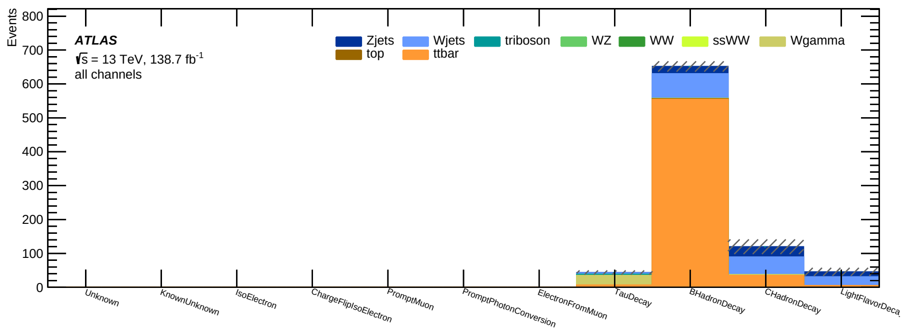
(a) Classification of non-prompt electrons in the signal region according to the IFFTruthClassifier for events with two Ana leptons



(b) Classification of non-prompt muons in the signal region according to the IFFTruthClassifier for events with two Ana leptons

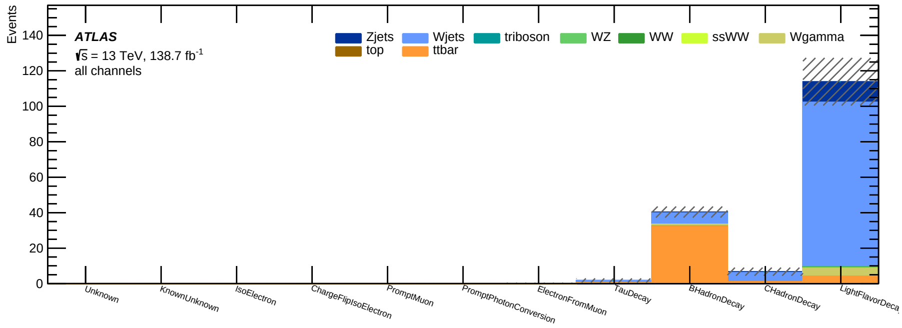


(c) Classification of non-prompt electrons in the control region according to the IFFTruthClassifier for events with two Ana leptons

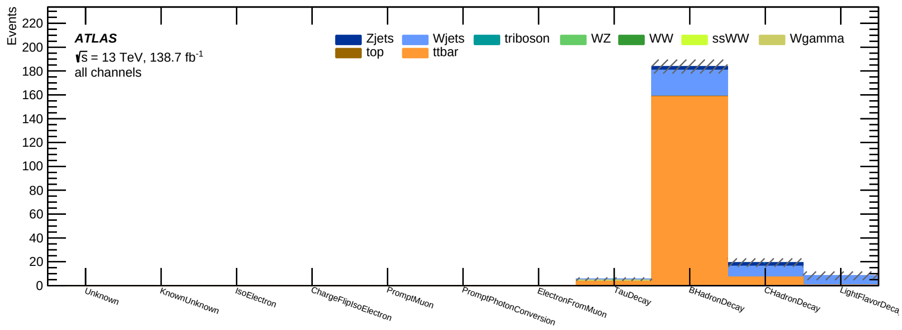


(d) Classification of non-prompt muons in the control region according to the IFFTruthClassifier for events with two Ana leptons

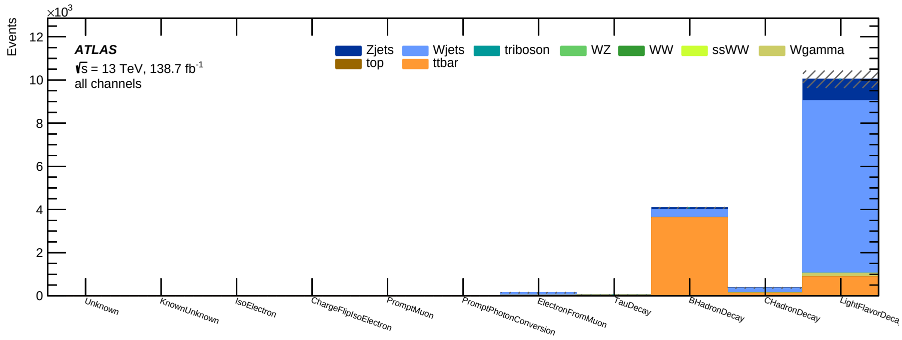
Figure A.1: Comparison of the signal and control region in the classification of non-prompt electrons and muons according to the IFFTruthClassifier for events with two Ana leptons. The number of leptons is splitted in the Monte Carlo samples they originate from.



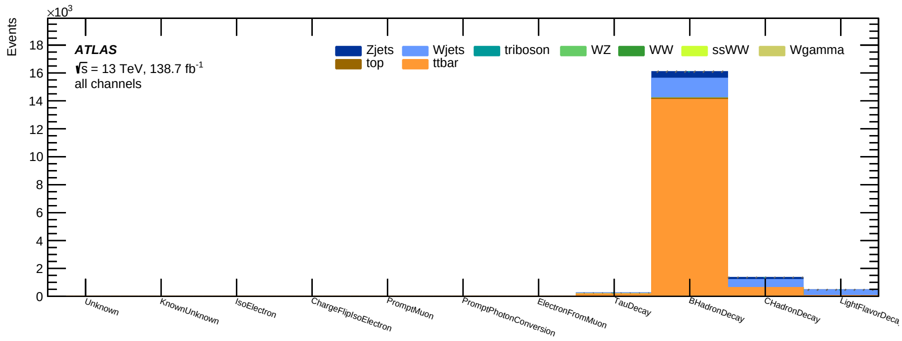
(a) Classification of non-prompt electrons in the signal region according to the IFFTruthClassifier for events with one Ana and one Non-Ana leptons



(b) Classification of non-prompt muons in the signal region according to the IFFTruthClassifier for events with one Ana and one Non-Ana leptons

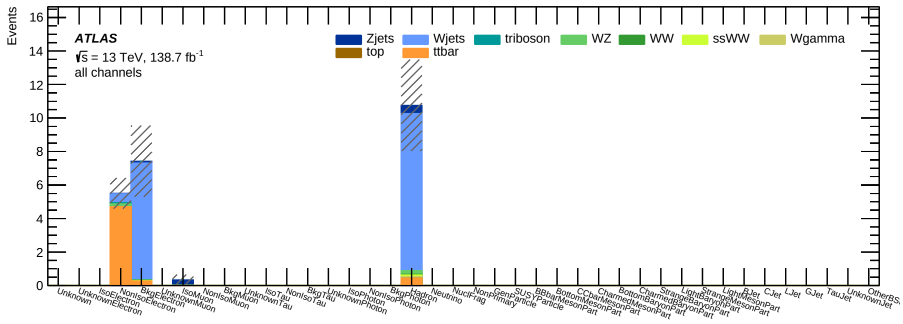


(c) Classification of non-prompt electrons in the control region according to the IFFTruthClassifier for events with one Ana and one Non-Ana leptons

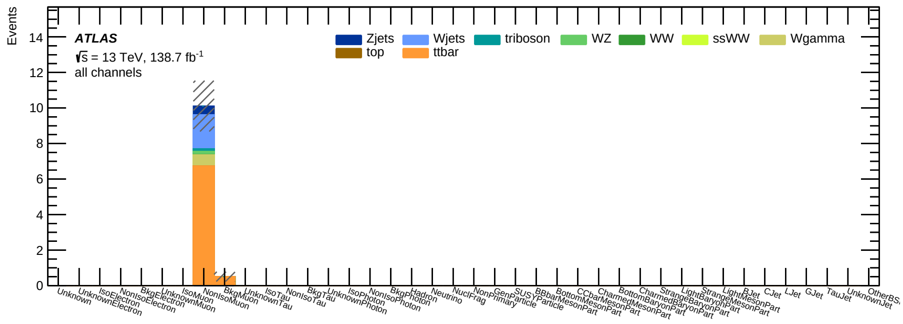


(d) Classification of non-prompt muons in the control region according to the IFFTruthClassifier for events with one Ana and one Non-Ana leptons

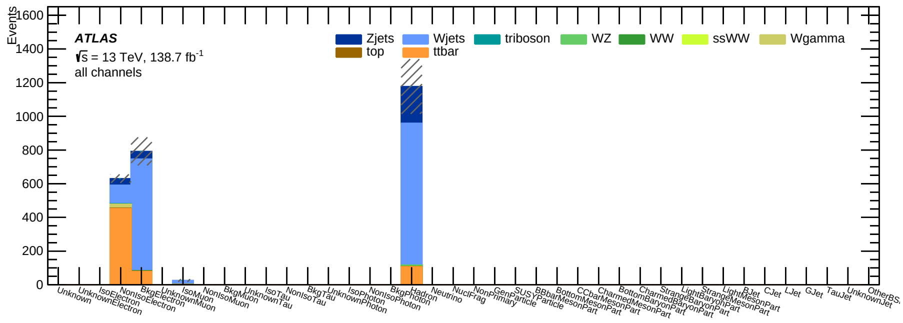
Figure A.2: Comparison of the signal and control region in the classification of non-prompt electrons and muons according to the IFFTruthClassifier for events with one Ana and one Non-Ana lepton. The number of leptons is splitted in the Monte Carlo samples they originate from.



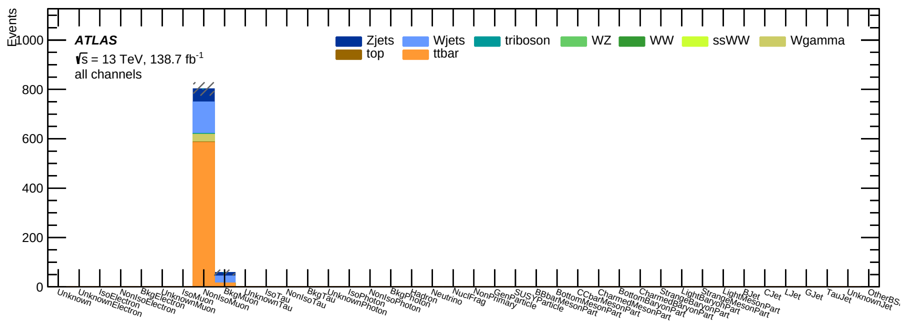
(a) Type of non-prompt electrons in the signal region according to the MCTruthClassifier for events with two Ana leptons



(b) Type of non-prompt muons in the signal region according to the MCTruthClassifier for events with two Ana leptons

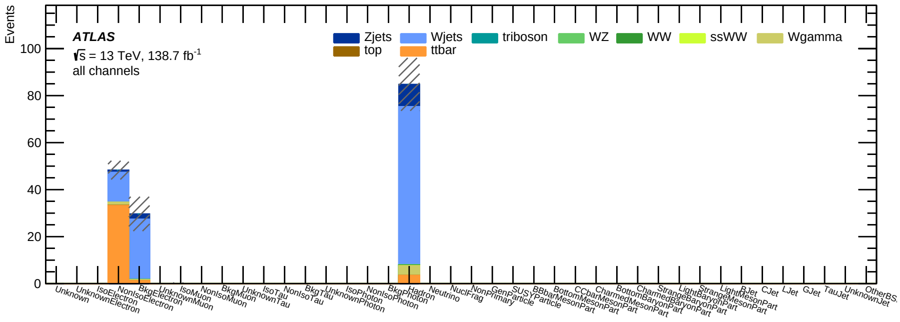


(c) Type of non-prompt electrons in the control region according to the MCTruthClassifier for events with two Ana leptons

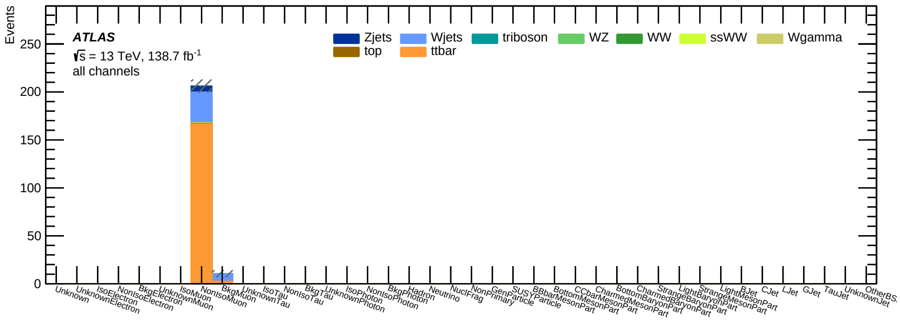


(d) Type of non-prompt muons in the control region according to the MCTruthClassifier for events with two Ana leptons

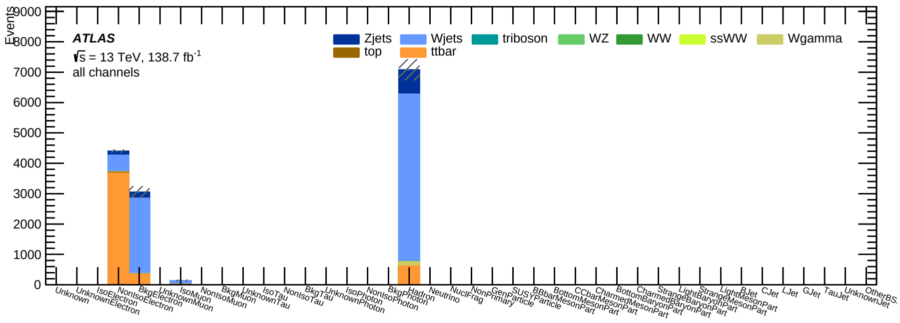
Figure A.3: Comparison of the signal and control region in the type of non-prompt electrons and muons according to the MCTruthClassifier for events with two Ana leptons. The number of leptons is splitted in the Monte Carlo samples they originate from.



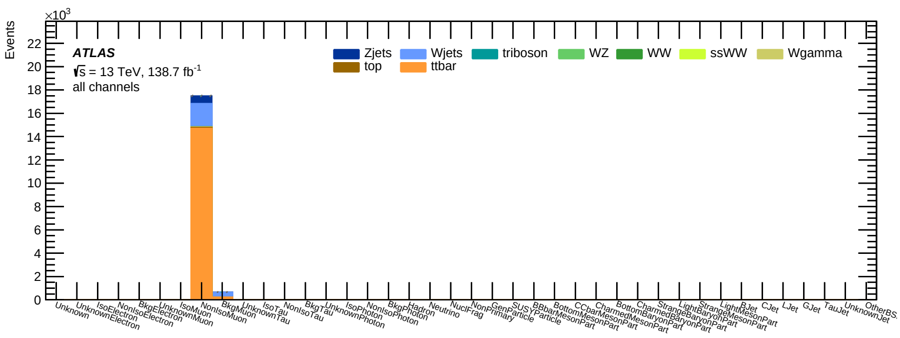
(a) Type of non-prompt electrons in the signal region according to the MCTruthClassifier for events with one Ana and one Non-Ana lepton



(b) Type of non-prompt muons in the signal region according to the MCTruthClassifier for events with one Ana and one Non-Ana lepton

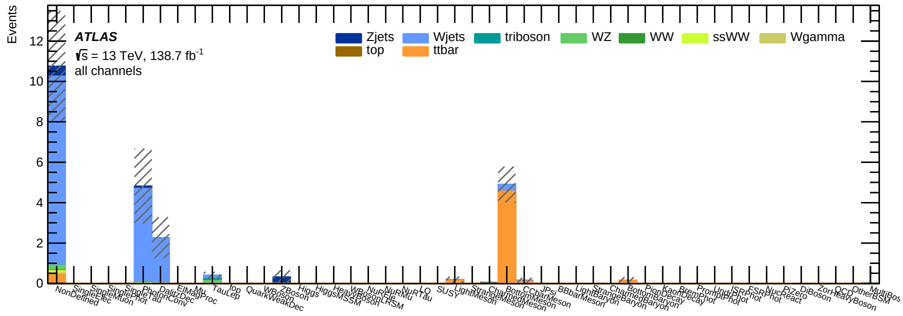


(c) Type of non-prompt electrons in the control region according to the MCTruthClassifier for events with one Ana and one Non-Ana lepton

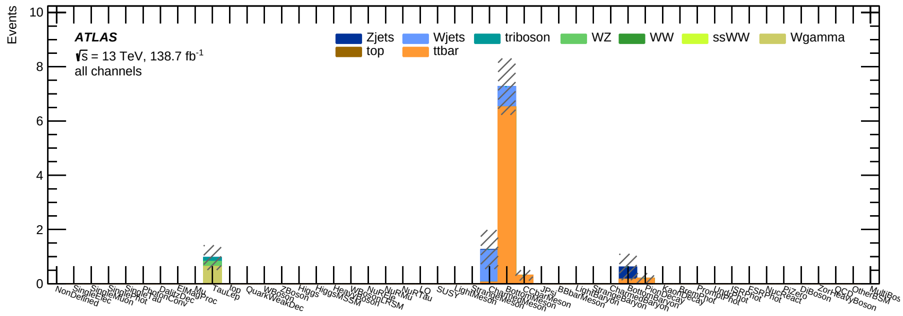


(d) Type of non-prompt muons in the control region according to the MCTruthClassifier for events with one Ana and one Non-Ana lepton

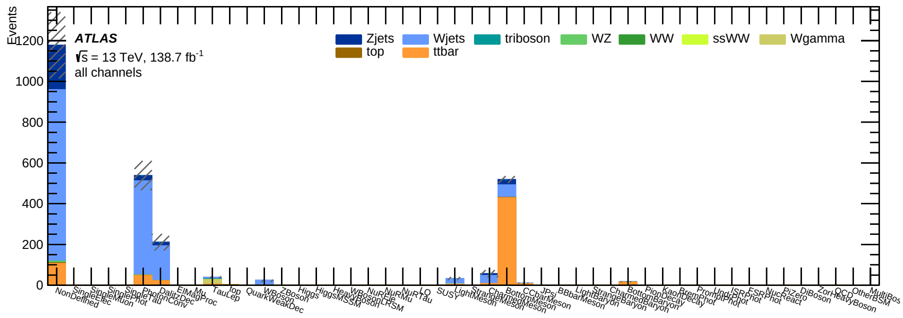
Figure A.4: Comparison of the signal and control region in the type of non-prompt electrons and muons according to the MCTruthClassifier for events with one Ana and one Non-Ana lepton. The number of leptons is splitted in the Monte Carlo samples they originate from.



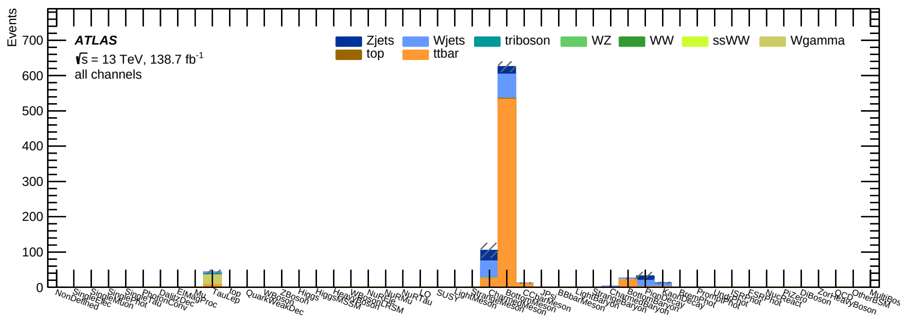
(a) Origin of non-prompt electrons in the signal region according to the MCTruthClassifier for events with two Ana leptons



(b) Origin of non-prompt muons in the signal region according to the MCTruthClassifier for events with two Ana leptons

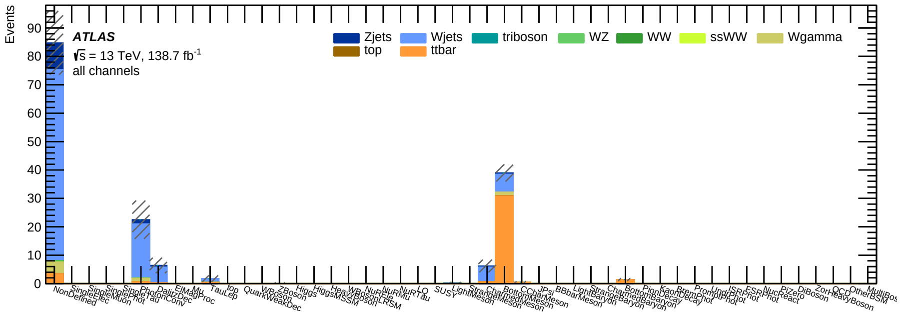


(c) Origin of non-prompt electrons in the control region according to the MCTruthClassifier for events with two Ana leptons

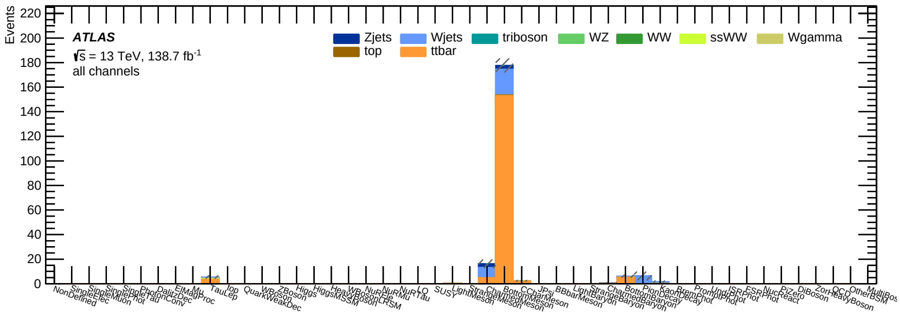


(d) Origin of non-prompt muons in the control region according to the MCTruthClassifier for events with two Ana leptons

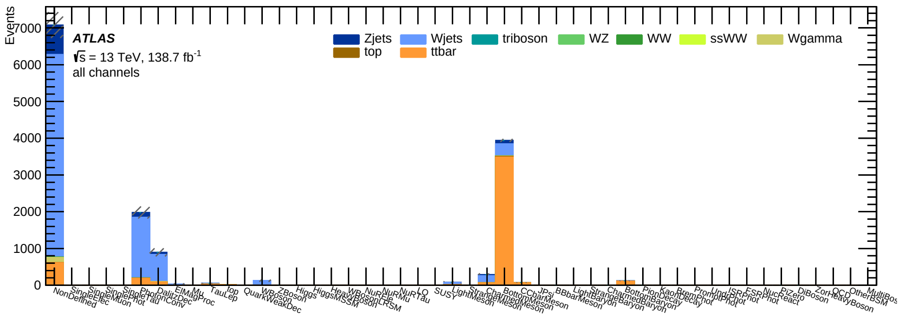
Figure A.5: Comparison of the signal and control region in the origin of non-prompt electrons and muons according to the MCTruthClassifier for events with two Ana leptons. The number of leptons is splitted in the Monte Carlo samples they originate from.



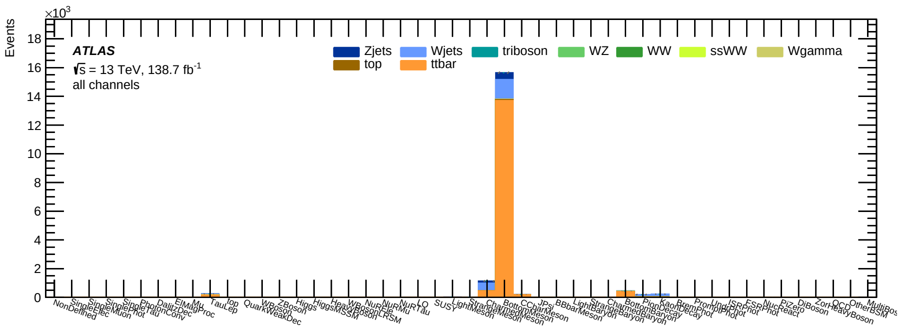
(a) Origin of non-prompt electrons in the signal region according to the MCTruthClassifier for events with one Ana and one Non-Ana lepton



(b) Origin of non-prompt muons in the signal region according to the MCTruthClassifier for events with one Ana and one Non-Ana lepton

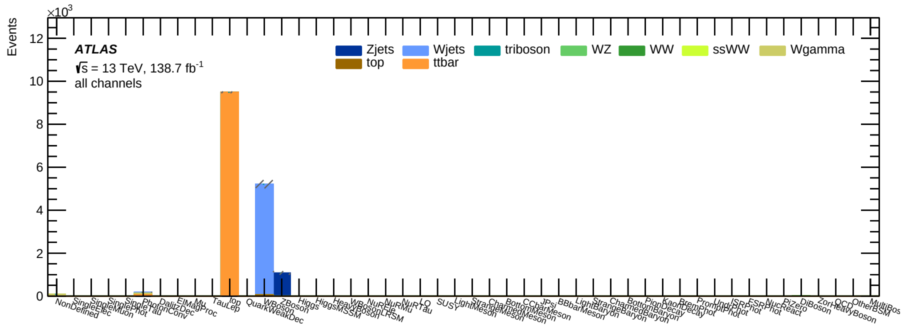


(c) Origin of non-prompt electrons in the control region according to the MCTruthClassifier for events with one Ana and one Non-Ana lepton

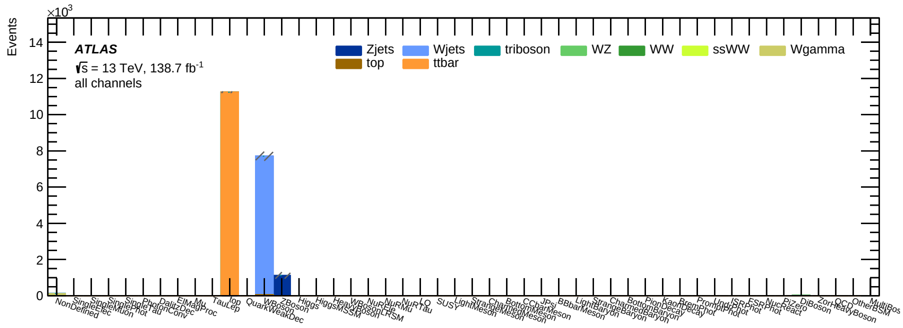


(d) Origin of non-prompt muons in the control region according to the MCTruthClassifier for events with one Ana and one Non-Ana lepton

Figure A.6: Comparison of the signal and control region in the origin of non-prompt electrons and muons according to the MCTruthClassifier for events with one Ana and one Non-Ana lepton. The number of leptons is splitted in the Monte Carlo samples they originate from.

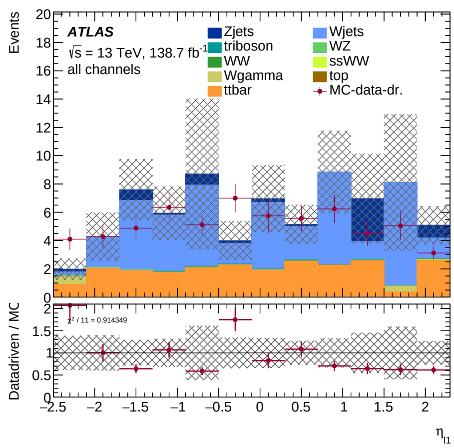


(a) Origin of prompt/charge flip electrons in the control region according to the MCTruthClassifier for events with one non-prompt lepton

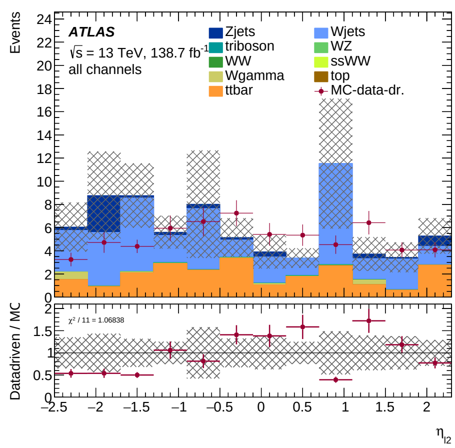


(b) Origin of prompt/charge flip muons in the control region according to the MCTruthClassifier for events with one non-prompt lepton

Figure A.7: Origin of prompt/charge flip leptons according to the MCTruthClassifier for events with one non-prompt lepton in the control region. The lepton numbers are splitted into the Monte Carlo samples they originate from.

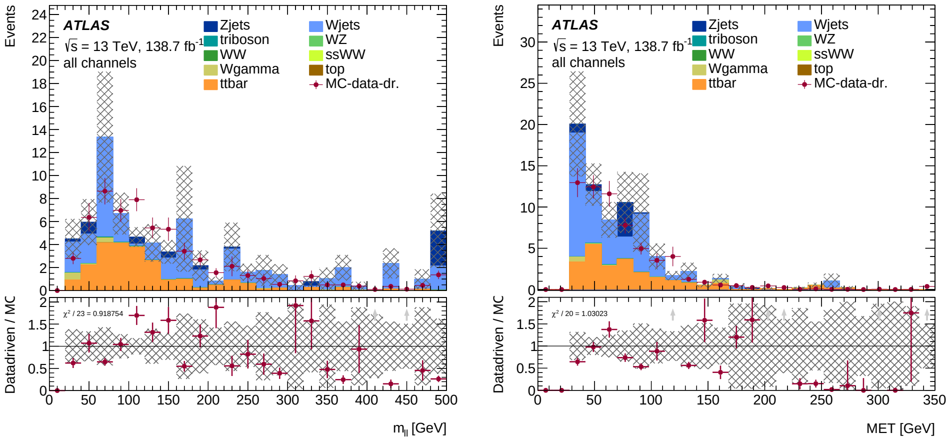


(a) Leading lepton η



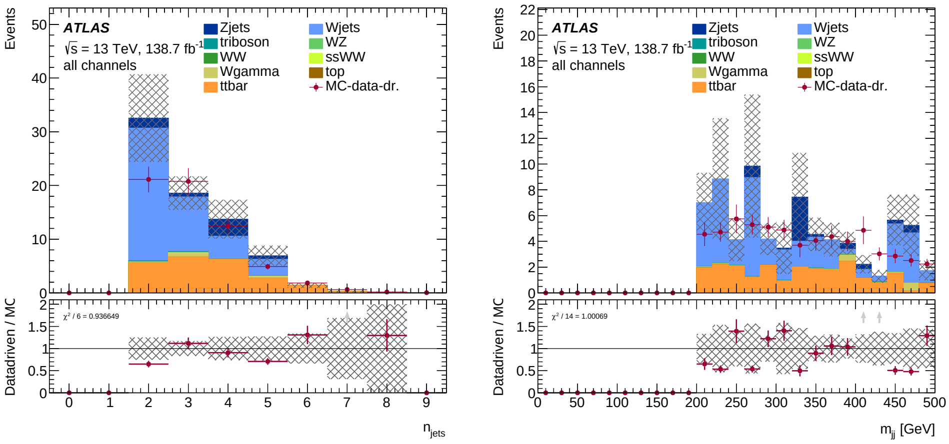
(b) Subleading lepton η

Figure A.8: Monte Carlo closure test within the LowMjj region. The MC-data-driven estimate of the non-prompt leading and subleading lepton distribution in η is compared to the non-prompt Monte Carlo events with two Ana leptons. The shown uncertainties result only from the statistical uncertainty of the Monte Carlo samples.



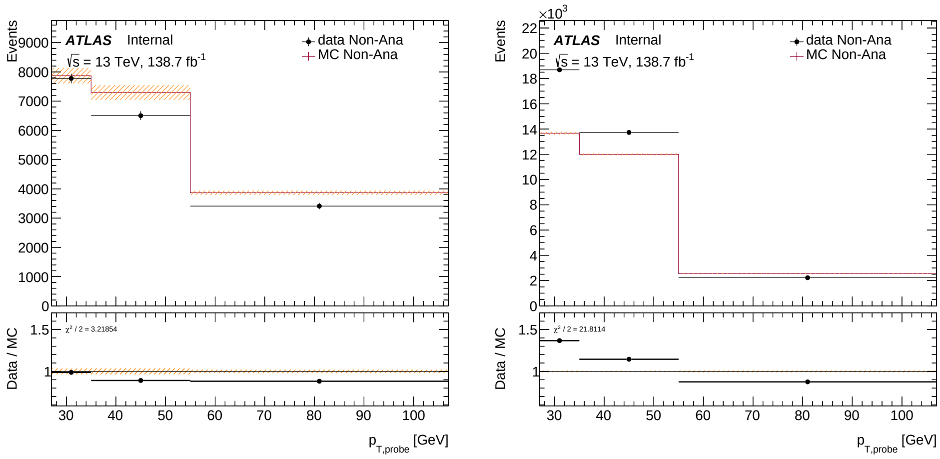
(a) m_{ll} distribution **(b)** E_T^{miss} distribution

Figure A.9: Monte Carlo closure test within the LowMjj region. The MC-data-driven estimates of the m_{ll} and E_T^{miss} distributions are compared to the non-prompt Monte Carlo events with two Ana leptons. The shown uncertainties result only from the statistical uncertainty of the Monte Carlo samples.



(a) Number of jets **(b)** m_{jj} of the two leading jets

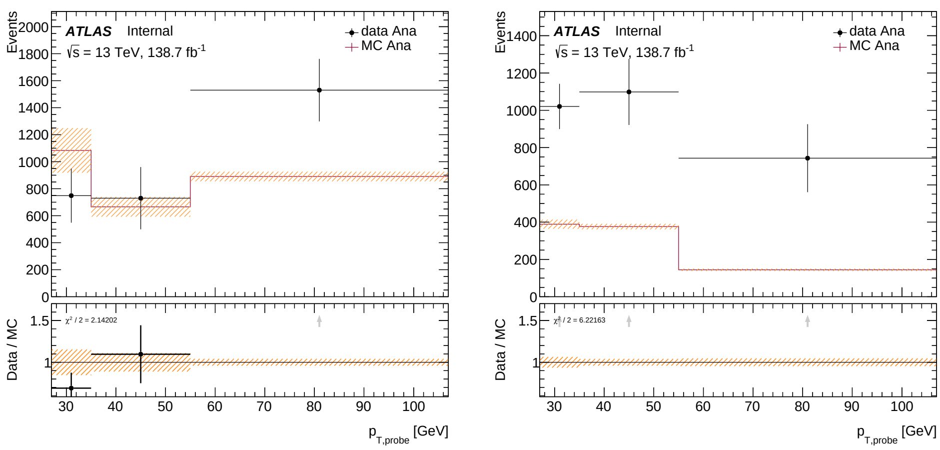
Figure A.10: Monte Carlo closure test within the LowMjj region. The MC-data-driven estimates of the number of jets and the m_{jj} distribution of the two leading jets are compared to the non-prompt Monte Carlo events with two Ana leptons. The shown uncertainties result only from the statistical uncertainty of the Monte Carlo samples.



(a) Denominator of electron fake factor

(b) Denominator of muon fake factor

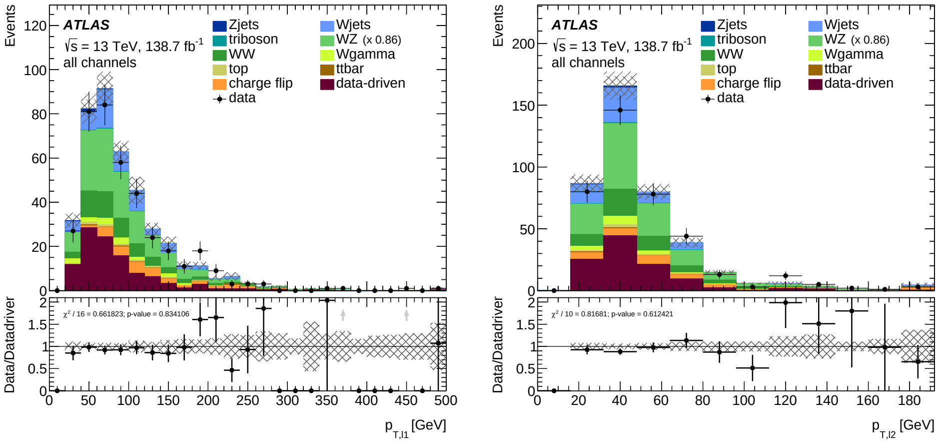
Figure A.11: Comparison of total event numbers in the fake factor denominator between the Monte Carlo simulation and actual data for electrons and muons. The shown uncertainty of the Monte Carlo fake factor denominator is derived from the statistical uncertainty of the non-prompt Monte Carlo events and the uncertainty of the fake factor denominator extracted from data results from the statistical uncertainty of the data and Monte Carlo simulated events.



(a) Numerator of electron fake factor

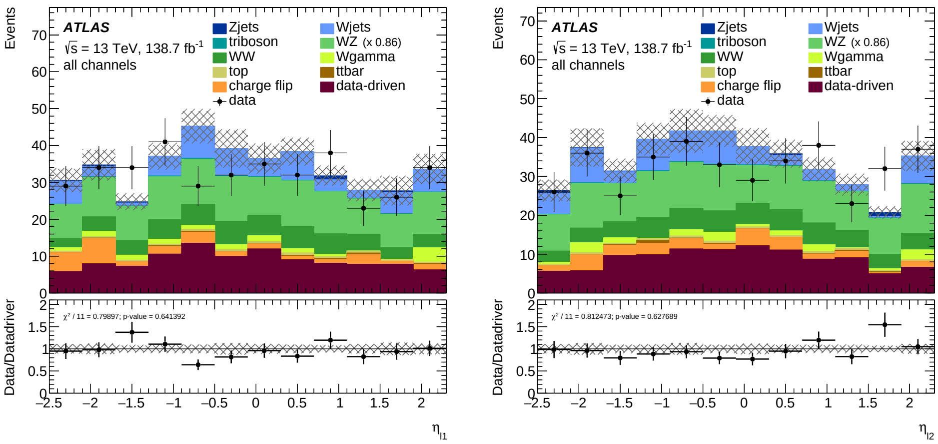
(b) Numerator of muon fake factor

Figure A.12: Comparison of total event numbers in the fake factor numerator between the Monte Carlo simulation and actual data for electrons and muons. The shown uncertainty of the Monte Carlo fake factor numerator is derived from the statistical uncertainty of the non-prompt Monte Carlo events and the uncertainty of the fake factor numerator extracted from data results from the statistical uncertainty of the data and Monte Carlo simulated events.



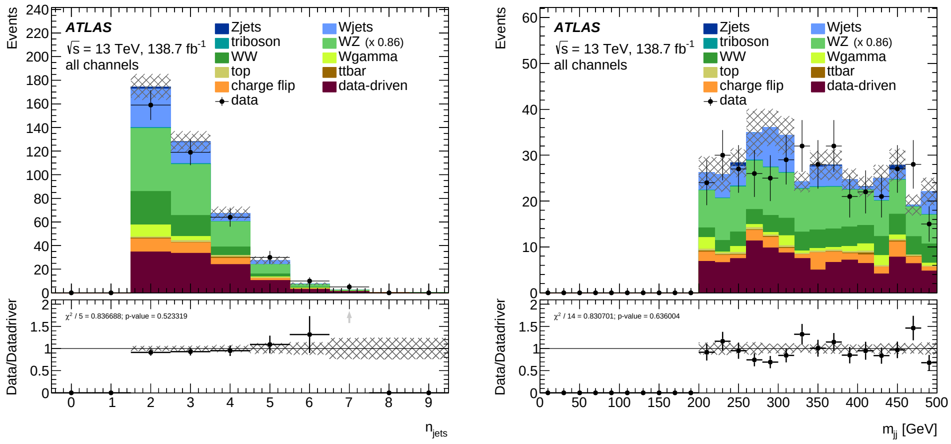
(a) Leading lepton p_T **(b)** Subleading lepton p_T

Figure A.13: The data p_T distribution of the leading and subleading lepton in the LowMjj region compared to the sum of the data-driven non-prompt background and the prompt and charge flip contributions estimated by Monte Carlo simulations. The number of prompt WZ events is scaled by a factor of 0.86. The shown uncertainties result from the statistical uncertainty of the Monte Carlo samples and the statistical uncertainty of the data within the dilepton control region and the LowMjj validation region.



(a) Leading lepton η **(b)** Subleading lepton η

Figure A.14: The data η distribution of the leading and subleading lepton in the LowMjj region compared to the sum of the data-driven non-prompt background and the prompt and charge flip contributions estimated by Monte Carlo simulations. The number of prompt WZ events is scaled by a factor of 0.86. The shown uncertainties result from the statistical uncertainty of the Monte Carlo samples and the statistical uncertainty of the data within the dilepton control region and the LowMjj validation region.



(a) Number of jets

(b) m_{jj} of the two leading jets

Figure A.15: The data n_{Jet} and m_{jj} distribution of the leading and subleading lepton in the LowMjj region compared to the sum of the data-driven non-prompt background and the prompt and charge flip contributions estimated by Monte Carlo simulations. The number of prompt WZ events is scaled by a factor of 0.86. The shown uncertainties result from the statistical uncertainty of the Monte Carlo samples and the statistical uncertainty of the data within the dilepton control region and the LowMjj validation region.

Bibliography

[1] M. Aaboud et al., *Observation of Electroweak Production of a Same-Sign W Boson Pair in Association with Two Jets in pp Collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector*, Phys. Rev. Lett. **123** (16 2019) 161801, URL: https://link.aps.org/doi/10.1103/PhysRevLett.123.161801 (cit. on pp. 2, 14, 18, 19, 25, 58).

[2] Carsten Bittrich et al., *Support note for measurement of electroweak $W^{\pm}W^{\pm}jj$ production at $\sqrt{s} = 13$ TeV*, tech. rep. ATL-COM-PHYS-2018-252, CERN, 2018, URL: https://cds.cern.ch/record/2309552 (cit. on pp. 2, 10–15, 48, 49, 51, 62, 65).

[3] Lyndon Evans and Philip Bryant, *LHC Machine*, Journal of Instrumentation **3** (2008) S08001, URL: https://doi.org/10.1088%5C%2F1748-0221%5C%2F3%5C%2F08%5C%2Fs08001 (cit. on p. 3).

[4] Joao Pequenao, "Computer generated image of the whole ATLAS detector", 2008, URL: https://cds.cern.ch/record/1095924 (cit. on p. 4).

[5] A. Airapetian et al., *ATLAS detector and physics performance: Technical Design Report, 1*, Technical Design Report ATLAS, CERN, 1999, URL: http://cds.cern.ch/record/391176 (cit. on pp. 4, 5).

[6] The ATLAS Collaboration, Journal of Instrumentation **3** (2008) S08003, URL: https://doi.org/10.1088%5C%2F1748-0221%5C%2F3%5C%2F08%5C%2Fs08003 (cit. on pp. 4, 5).

[7] M. Aaboud et al., *Study of the material of the ATLAS inner detector for Run 2 of the LHC*, Journal of Instrumentation **12** (2017) P12009, ISSN: 1748-0221, URL: http://dx.doi.org/10.1088/1748-0221/12/12/P12009 (cit. on pp. 6, 7).

[8] Particle Data Group and other, *Review of Particle Physics*, Progress of Theoretical and Experimental Physics **2020** (2020), 083C01, ISSN: 2050-3911, eprint: https://academic.oup.com/ptep/article-pdf/2020/8/083C01/33653179/ptaa104.pdf, URL: https://doi.org/10.1093/ptep/ptaa104 (cit. on p. 6).

[9] Seth Moortgat, *Lepton Isolation Using Particle Flow Objects for the ATLAS Detector*, (2014), URL: http://cds.cern.ch/record/1756841 (cit. on p. 7).

[10] Gordon Kane, *Modern elementary particle physics: explaining and extending the standard model; 2nd ed.* Cambridge University Press, 2017, URL: https://cds.cern.ch/record/2244793 (cit. on p. 9).

[11] Matteo Cacciari, Gavin P Salam, and Gregory Soyez, *The anti-ktjet clustering algorithm*, Journal of High Energy Physics **2008** (2008) 063, URL: https://doi.org/10.1088/2F1126-6708/2F2008/2F04/2F063 (cit. on p. 13).

[12] Christos Anastopoulos et al., *Electron identification and efficiency measurements in 2017 data*, tech. rep. ATL-COM-PHYS-2018-1727, CERN, 2019, URL: https://cds.cern.ch/record/2652163 (cit. on pp. 15, 19, 20).

[13] *Estimation of non-prompt and fake lepton backgrounds in final states with top quarks produced in proton-proton collisions at $\sqrt{s} = 8$ TeV with the ATLAS detector*, tech. rep. ATLAS-CONF-2014-058, CERN, 2014, URL: https://cds.cern.ch/record/1951336 (cit. on p. 15).

[14] Carsten Bittrich et al., *Study of electroweak $W^{\pm}Z$ boson pair production in association with two jets in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector*, tech. rep. ATL-COM-PHYS-2018-025, CERN, 2018, URL: https://cds.cern.ch/record/2300058 (cit. on pp. 17, 18).

[15] Jean-Francois Arguin et al., *Support Note for Electron ID: charge mis-identification*, tech. rep. ATL-COM-PHYS-2017-1014, CERN, 2017, URL: https://cds.cern.ch/record/2273365 (cit. on pp. 18–20).

[16] A Alonso and B Meirose, *New data-driven methods for lepton charge mis-identification*, tech. rep. ATL-COM-PHYS-2012-264, This note details the charge mis-identification data driven methods developed for arXiv: 1202.5520 [hep-ph], ATL-COM-PHYS-2011-1442: CERN, 2012, URL: https://cds.cern.ch/record/1430701 (cit. on p. 18).

[17] Philip Sommer, *Electron Charge ID Selector Tool*, URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/ElectronChargeFlipTaggerTool, (accessed: 05.08.2020) (cit. on pp. 19, 20).

[18] Andy Buckley et al., *General-purpose event generators for LHC physics*, Physics Reports **504** (2011) 145, ISSN: 0370-1573, URL: http://dx.doi.org/10.1016/j.physrep.2011.03.005 (cit. on p. 21).

[19] Michael H. Seymour and Marilyn Marx, *Monte Carlo Event Generators*, 2013, arXiv: 1304.6677 [hep-ph] (cit. on p. 21).

[20] Daniel Guest, *MCTruthClassifier*, URL: https://twiki.cern.ch/twiki/bin/view/AtlasProtected/MCTruthClassifier, (accessed: 10.08.2020) (cit. on p. 22).

[21] *IFFTruthClassifier*, URL: https://gitlab.cern.ch/ATLAS-IFF/IFFTruthClassifier, (accessed: 10.08.2020) (cit. on p. 22).

[22] Shalu Solomon, Internal discussion, 2020 (cit. on p. 37).

[23] Christoph Klimanek, "Eta-Correction for Data-Driven Estimate of Non-Prompt Background for Same Charged WW Scattering in the ATLAS Experiment", 2020 (cit. on pp. 46, 47, 61).

[24] Carsten Bittrich, "Evidence for Scattering of Electroweak Gauge Bosons in the $W^{\pm}Z$ Channel with the ATLAS Detector at the Large Hadron Collider", Presented 14 Jul 2020, 2020, URL: http://cds.cern.ch/record/2719126 (cit. on pp. 47, 49, 51).

Danksagung

An dieser Stelle möchte ich mich bei all den Menschen bedanken, die mich sowohl fachlich als auch persöhnlich in meinem Studium begleitet und unterstützt haben.

Im Besonderen möchte ich mich bei Prof. Michael Kobel bedanken für die herzliche Aufnahme in seine Arbeitsgruppe, für die zahlreichen Weiterbildungsmöglichkeiten und dafür, dass Sie einem immer mit offenem Ohr und auf Augenhöhe begegnet.

I would like to thank the entire VBS Group for the pleasant working atmosphere. The everyday life with you goes clearly beyond a normal work together.

Insbesondere möchte ich mich noch einmal bei Carsten Bittrich bedanken, dessen intensive und kompetente Betreuung mich während der gesamten Masterarbeit begleitet hat.

Additionally I would like to thank Joany Manjarres, Carsten Bittrich, Stefanie Todt, Tim Herrmann and Christoph Klimanek for proofreading my thesis and all the helpful comments.

Des Weiteren bedanke ich mich bei meinen Eltern und Großeltern, die mich groß gezogen und gefördert haben und mich auch während meines Studiums weiter unterstützten.

Ich bedanke mich bei meinem Schwesterherz für unsere harmonische gemeinsame Kindheit und dafür, dass du bis heute immer für mich als große Schwester da ist.

Abschließend möchte ich mich noch bei all meinen Freunden bedanken. Die gemeinsamen Tage und Nächte mit euch stellen für mich den Inbegriff von Lebensqualität dar und lassen mich jeden Tag mit einem Lächeln beginnen und beenden.

Selbständigkeitserklärung

Hiermit erkläre ich, dass ich diese Arbeit im Rahmen der Betreuung am `Institut für Kern- und Teilchenphysik` ohne unzulässige Hilfe Dritter verfasst und alle Quellen als solche gekennzeichnet habe.

Max Vincent Stange
Dresden, 04.11.2020