TECHNISCHE UNIVERSITÄT DRESDEN

Study of the non-prompt lepton background in the same charge WW analysis performed in ATLAS at 13 TeV

Bachelor-Arbeit
zur Erlangung des Hochschulgrades
Bachelor of Science
im Bachelor-Studiengang Physik

vorgelegt von

Moritz Poguntke
geboren am 23.11.1997 in Schlema

Institut für Kern- und Teilchenphysik
Fakultät Physik
Bereich Mathematik und Naturwissenschaften
Technische Universität Dresden
2020

Eingereicht am 13. Januar 2020

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

Abstract

Vector Boson Scattering (VBS) is a promising process to study the electroweak symmetry breaking mechanism of the Standard Model (SM). A consequence of this symmetry breaking is an additional Boson in the SM: the Higgs Boson. The VBS cross-section is regularized by cancellations of different contributions including the Higgs Boson. Investigations of the energy dependence of the VBS cross-section will show if the SM Higgs Boson fully unitarizes this process as predicted. Compared to other VBS processes, the scattering of two same-sign $W^{\pm}$ Bosons (ssWW) has the highest electroweak over strong production cross-section ratio and is therefore well suited to study VBS. Electroweak production of ssWW was first observed last year, with 36.1 fb^{-1} of the LHC data. The second-largest background and the one with the largest uncertainties to this process is the non-prompt lepton background, which is the focus of this study. Non-prompt leptons are misidentified objects and leptons originating in Photon Conversions or Hadron decays. Due to the characteristics of this background, MC simulations are not accurate. Data-driven techniques are commonly used to estimate the non-prompt background.

In this thesis, the 'fake factor method' is used as a starting point to obtain a data-driven assessment of the non-prompt lepton background. Monte Carlo simulations are used to probe the consistency of the method and to study the impact of the event selection and the origin of the non-prompt lepton on the background estimation. A dependence on the event selection and especially on the origin of the non-prompt lepton is observed. Several adjustments to the basic fake factor method are needed in order to obtain reliable predictions. The MC investigations lead to unexpected findings; first, an unusually high amount of non-prompt leptons originating from Photon Conversions is found. Additionally, an issue with the charge compositions of the leptons is noticed. Both issues still need to be understood or confirmed with alternative MC generators and data.

Zusammenfassung

Vektorbosonenstreuung (VBS) ist ein vielversprechender Prozess um die elektroschwache Symmetriebrechung des Standardmodells (SM) zu untersuchen. Eine Konsequenz jener Symmetriebrechung ist ein zusätzliches Boson im SM: das Higgs Boson. Der Wirkungsquerschnitt der VBS wird durch die Auslöschung verschiedener Beiträge des Higgs Bosons regularisiert. Untersuchungen der Energieabhängigkeit des Wirkungsquerschnittes der VBS werden zeigen, ob das SM Higgs Boson diesen Prozess wie erwartet vollständig unitarisiert. Im Vergleich zu anderen Prozessen der VBS hat die Streuung von zwei gleich geladenen $W^{\pm}$-Bosonen ein besonders hohes Verhältnis von elektroschwachem zu starkem Wirkungsquerschnitt und eignet sich daher gut, um VBS zu untersuchen. Die elektroschwache Produktion von zwei gleich geladenen

$W^{\pm}$-Bosonen durch VBS wurde letztes Jahr das erste Mal beobachtet, mit 36.1 fb^{-1} an Daten des LHC. Der zweitgrößte Untergrund zu diesem Prozess, jener mit den größten Unsicherheiten, ist der Untergrund aus nicht-prompten Leptonen, welcher der Fokus dieser Arbeit ist. Nicht-prompte Leptonen stammen dabei aus falsch identifizierten Objekten, Hadron zerfällen und Paarerzeugung. Aufgrund der Charakteristiken dieses Untergrundes liefern Monte-Carlo Simulationen keine genauen vorhersagen. Daher werden für die Abschätzung des Untergrundes häufig datengetriebene Methoden genutzt.
In dieser Arbeit wird die 'Fake Faktor Methode' als Startpunkt für eine datengetriebene Abschätzung des nicht-prompten Leptonen Untergrundes verwendet. Monte-Carlo Simulationen werden genutzt, um die Konsistenz der Methode zu testen und den Einfluss von Event Selektion und dem Ursprung der nicht-prompten Leptonen auf den Fake Faktor zu analysieren. Dabei kann eine Abhängigkeit des Fake Faktors von der Event Selektion und insbesondere von dem Ursprung der nicht-prompten Leptonen festgestellt werden. Um eine verlässliche Untergrundabschätzung zu erhalten, muss die allgemeine Fake Faktor Methode angepasst werden. Die MC Untersuchungen führen zu unvorhergesehenen Ergebnissen; zunächst kann eine ungewöhnlich hohe Anzahl an nicht-prompten Leptonen beobachtet werden die aus Paarerzeugungen stammen. Zusätzlich lässt sich ein Problem mit der Ladungszusammensetzung der Leptonen erkennen. Die beiden unerwarteten Ergebnisse müssen noch verstanden oder mit alternativen MC Generatoren und Daten bestätigt werden.

Contents

1 Introduction

One of the essential objectives of physics is to understand nature on the most fundamental level possible. Our current understanding of this fundamental level is based on the Standard Model of particle physics, which has been tested extensively using collider experiments such as the `Large Hadron Collider`. At the `LHC`, protons are colliding at center-of-mass energies of up to $\sqrt{s} = 13$ TeV.
The Standard-Model has proven to be very successful at describing elementary particles, and the interactions between them, as the last missing particle predicted by the Standard-Model, the Higgs-Boson, has been confirmed experimentally in 2012 with the `LHC` - 50 years after its postulation. With the increasing center-of-mass energies of colliders, predictions by the Standard-Model can be measured more and more precisely. A large amount of data collected at such high energy has opened the door to the study of the production of two vector bosons through Vector-Boson Scattering (VBS). In this study, the focus will be in particular on the VBS production of two same charge $W^{\pm}$-bosons. The Standard-Model predicts this process, and despite its small cross-section, it can be measured at the `LHC`. It has been observed by the ATLAS- and CMS collaborations with significances of $6.5\,\sigma$ [1] and $5.7\,\sigma$ [13] respectively. Vector-Boson Scattering is closely related to the electroweak symmetry-breaking mechanism. The cross-section for VBS processes at high energies is regularized by cancellations of different contributions including the Higgs boson. By looking at these cross-sections at high energy, it will be seen if the Higgs Boson fully unitarizes the cross-section as expected [5]. Deviations from the Standard-Model are expected to appear in the high-energy region.

Standard Model processes with higher cross-sections can mimic the same detector signature as the production of two same-sign $W^{\pm}$-Bosons through Vector-Boson Scattering, resulting in a low signal to background ratio. The background coming from misidentified objects (also called non-prompt) in the detector is one of the main backgrounds for this physics process. Since the available Monte Carlo simulations are not able to accurately describe this background, data-driven techniques are commonly used. The data-driven methods utilized so far led to large systematics uncertainties making the measurement less precise. Therefore a precise background estimation is necessary in order to improve the understanding of the non-prompt lepton background, which will be the focus of this thesis. The goal is to improve the modeling and to lay the foundation for a better estimation of this background.

The second chapter of this thesis will introduce the Standard Model as well as the fundamentals of Vector-Boson Scattering. Furthermore, the method used for estimating the non-prompt lepton background, a data-driven technique called the 'Fake factor method', will be explained. In the third chapter, the experiment at the LHC and the ATLAS detector will be outlined. Chapter four will probe the implementation of the fake factor method using Monte Carlo generated events and discuss the influence of event cuts and origins of the non-prompt leptons on the fake factor estimation.

2 Theoretical Background

2.1 The Standard Model

The Standard Model describes three of the four fundamental interactions between the elementary particles we and everything surrounding us are made of. Despite being called a model, it is the most mathematically profound theory in science to this date [12]. It has been developed mainly between 1961-1973 and is based on the mathematics of Noether's Theorem (Emmy Noether, 1918). Mathematically, the Standard Model is described through a Lagrangian density. A fundamental principle is a demand for the Lagrangian to be invariant under specific gauge transformations. For this, gauge fields are introduced to preserve this invariance. They can be interpreted as the gauge bosons responsible for the interaction of particles. The number of existing gauge bosons to a corresponding symmetry depends on the actual symmetry group. There is one gauge boson for each generator of the group. This leads to a different number of gauge bosons for the three interactions. As formulated in Noether's Theorem, there is a conserved quantity corresponding to each symmetry. These conserved quantities are the charges associated with each of the fundamental interactions since each interaction underlays a certain local gauge symmetry.
Gravity, the fourth interaction, is not described by the Standard Model.

Table 2.1: The three basic interactions in the Standard Model.

Interaction	**Symmetry Group**	**Gauge Boson**	**# Gauge Bosons**	**Charge**
electromagnetic	U(1)	Photon γ	1	electric charge Z
strong	SU(3)	Gluon g	8	color charge $\vec{C}$
weak	SU(2)	$W^{\pm}/Z$ Boson	2+1	weak isospin $\vec{I}^W$

Elementary particles of the Standard Model can be divided into two groups, depending on their spin: Fermions and Bosons. Fermions have half-integer spin, whereas bosons are particles with integer spins. The bosons can be grouped further regarding their spin. For each of the three fundamental interactions between fermions exists a certain number of gauge bosons (Table 2.1), which all have a spin of 1 and are therefore also called vector-bosons. There is also a spin 0 (scalar) boson within the Standard Model, the Higgs Boson, which is an essential part of the mechanism responsible for the particle masses.

Table 2.2: The Standard Model matter particles. Each fermion has a corresponding antiparticle with opposite charges and identical masses.

	Charge		**Generations**		
	Z	I_W^3	1.	2.	3.
Quarks	+2/3	+1/2	u	c	t
	-1/3	-1/2	d	s	b
Leptons	-1	-1/2	e	μ	τ
	0	+1/2	ν_e	ν_μ	ν_τ

The fermions are the constituents of matter. They can be divided further (Table 2.2) into quarks, which carry an electric charge of multiples of 1/3 and can take part in all three interactions, and leptons which have an electric charge of ±1 or 0 and can only interact through weak- and electromagnetic interaction since they do not carry color-charge. Quarks can be split into 'up-type' quarks (up, charm, top) with a charge of $Z = +2/3$ and 'down-type' quarks (down, strange, bottom) with a charge of $Z = -1/3$. As seen in the table, leptons can also be ordered further: there are electrically charged leptons (electron, muon, tau) and neutral leptons, which are also called neutrinos (electron-neutrino, muon-neutrino, tau-neutrino). Both leptons and quarks can be grouped into generations. The first generation contains the three elementary particles forming stable matter (electrons e, up-quarks u, and down-quarks d) and electron-neutrinos ν_e. The two other generations contain particles that are identical in all quantum numbers to their first-generation counterparts, but with increasing masses for increasing generation numbers.

2.2 Vector-Boson Scattering

Vector-Boson scattering (VBS) is a process of the type: $VV \rightarrow VV$ with vector-bosons $V = (W^{\pm},\ Z)$. The event topology consists of two quarks from the collision of protons, each radiating an electroweak vector-boson, which then scatter and decay (Figure 2.2). Such processes are predicted by the Standard-Model Lagrangian as a result of the SU(2) and SU(3) symmetry groups being non-abelian, which allows for gauge-boson self-coupling. The VBS process is closely related to the electroweak symmetry breaking mechanism [5]. Violation of unitarity can be observed without a Higgs Boson in the scattering of longitudinally polarized $W^{\pm}$-Bosons [5, 14]. Violation of unitarity means that above those energies, the WW production rate would be higher than the pp collision rate, which can not be possible. Investigation of the VBS energy-dependance at high energies will clarify if the predicted cross-section of this process is fully unitarized by the Higgs Boson contribution (see Figure 2.2, Higgs contribution in the two bottom Feynman diagrams) or only partially [5]. Therefore it could also give a handle to new physics beyond the Standard Model. Hence it is interesting to study vector-boson scattering. Due to the high center of mass energies and a vast amount of data needed, this is only possible at the `LHC` with the run 2 data.

Out of the possible VBS processes, the $W^{\pm}W^{\pm}jj$ (jj stands for two jets) final state has the highest ratio of electroweak production cross-section over strong production cross-section [1], making it a promising process for studying VBS. The production of two same-sign $W^{\pm}$-Bosons through the primary production mechanisms discussed in Section 2.3 will be the process of interest in this thesis. Only leptonic decays of the $W^{\pm}$-Bosons into electrons or muons will be considered, despite the approximately three times larger branching ratios for hadronic decays [9], which would, however, be more challenging to study using a hadron collider like the `LHC` [3]. Tau leptons are not considered, to maintain a cleaner signature, since their hadronic decays would lead to additional jets.

2.3 $W^{\pm}W^{\pm}jj - EW$ Signal Process

For $W^{\pm}W^{\pm}jj - EW$ final states, the event topology described in Section 2.2 can be specified further: the vector-bosons that get radiated off the initial state quarks are two same sign $W^{\pm}$-Bosons. The outgoing quarks have a high invariant mass as well as large angular separation and are close to the direction of the beam (also see [5]). Since the two $W^{\pm}$-Bosons have the same electric charge, the two leptons are also required to have the same charge. Neutrinos that are produced in the $W^{\pm}$-Boson decay can not be detected directly but are noticeable through missing transverse momentum (sometimes also referred to as *missing transverse energy* or *MET*) in the reconstruction. Because of a principle called *confinement* (which essentially means that particles which carry color charge may not be isolated), the outgoing quarks will be detected as hadron-jets. In the following, the detector signature of the signal will be shortened as $ll' + E_T^{miss} + 2$ jets.

For a particular final state, there is not one single Feynman diagram that represents the process but an infinite amount of diagrams that contribute. However, by neglecting higher orders than leading order, one can define a set of diagrams as the signal. The $W^{\pm}W^{\pm}jj - EW$ VBS process is in the order of $\mathcal{O}(\alpha_W^6)$. Non-VBS production (Figure 2.1) of the $W^{\pm}W^{\pm}jj - EW$ final state in $\mathcal{O}(\alpha_W^6)$ can not be separated from the VBS processes in a gauge-invariant way [1], so it has to be included in the signal definition.

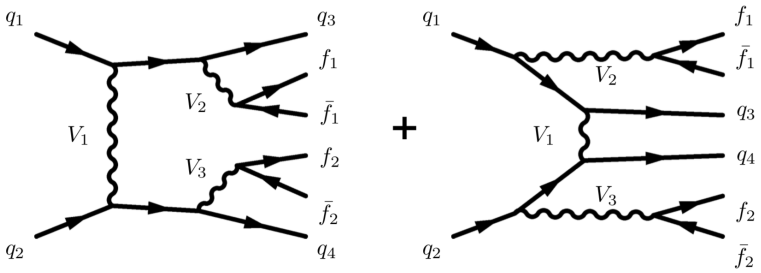


Figure 2.1: $VVjj$ non VBS diagrams

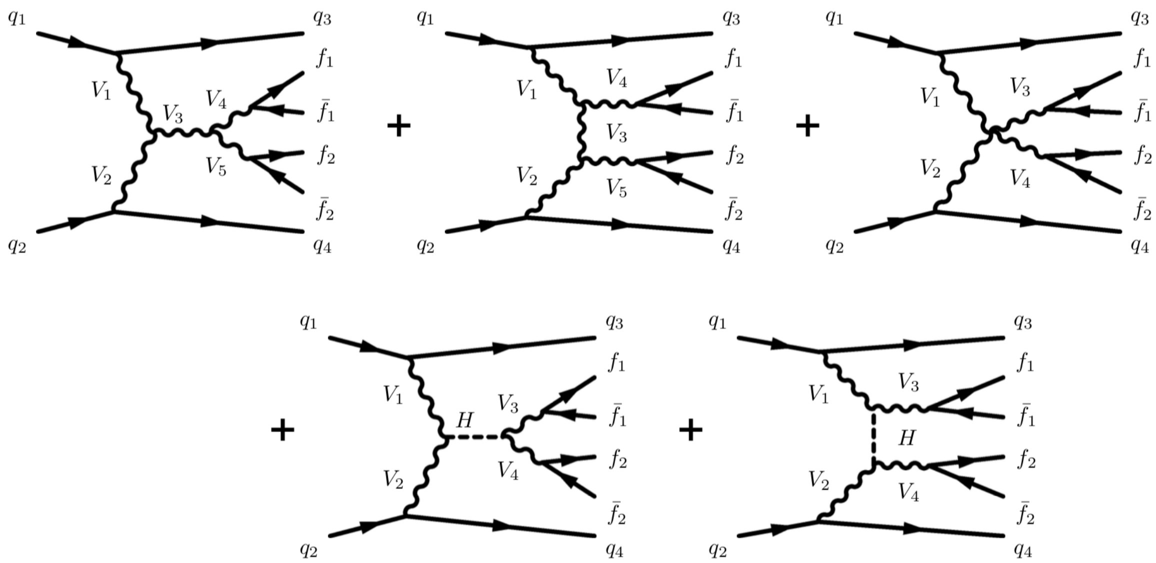


Figure 2.2: $VVjj$ VBS diagrams

2.4 $W^{\pm}W^{\pm}jj$ Background

Processes that can produce the same $ll' + E_T^{miss} + 2$ jets final state as the signal and can be separated in a gauge-invariant way are regarded as background. The major background contributions are summarized in Table 2.3. Two different groups of background need to be considered: irreducible and reducible backgrounds. The irreducible background consists of processes that produce (at least) two Prompt leptons and have the same signature in the detector as the signal. For the $W^{\pm}W^{\pm}jj - EWK$ signal, the irreducible background is dominated by the $WZjj - QCD$ (largest contribution) and $W^{\pm}W^{\pm}jj - QCD$ processes in the order of $\mathcal{O}(\alpha_W^4\alpha_S^2)$, which are similar to the non-VBS diagrams but with gluon exchange. Other sources of irreducible background are $WZjj - EWK$ processes (VBS and non-VBS diagrams

are contributing), and tri-boson processes (included in 'Other Prompt' in Table 2.3) where one $W^{\pm}$-Boson decays hadronically. $WZjj$ processes can lead to the same detector signature as the signal (described in Section 2.3) when one of the leptons from the Z-decay is lost, resulting in two same-sign leptons being detected.
Reducible backgrounds include processes that can mimic the detector signature of the signal due to jet-faked leptons, non-prompt leptons, or leptons with misidentified charge (included in e/γ conversions, account for $\approx 70\%$ of the e/γ conversions). The first two types of reducible background, can be grouped as the non-prompt lepton background.

The background estimation for processes that are relatively well modeled in *Monte Carlo* Simulations (MC) such as $W^{\pm}W^{\pm}jj - QCD$, as well as small backgrounds like tri-boson production can be taken from simulations [11]. For the WZ background, shapes are taken from MC, and a normalization correction is derived from data [5]. Other backgrounds not modeled well in simulations are resulting in non-reliable background predictions when taken from MC.

Table 2.3: Event yields for "36.1 fb^{-1} of proton–proton collision data recorded at a center-of-mass energy of $\sqrt{s} = 13$ TeV" [1]. Data, expected signal and backgrounds are displayed, taken from [1].

	e^+e^+	e^-e^-	$e^+\mu^+$	$e^-\mu^-$	$\mu^+\mu^+$	$\mu^-\mu^-$	Combined
WZ	1.48 ± 0.32	1.09 ± 0.27	11.6 ± 1.9	7.9 ± 1.4	5.0 ± 0.7	3.4 ± 0.6	30 ± 4
Nonprompt	2.2 ± 1.1	1.2 ± 0.6	5.9 ± 2.5	4.7 ± 1.6	0.56 ± 0.05	0.68 ± 0.13	15 ± 5
e/γ conversions	1.6 ± 0.4	1.6 ± 0.4	6.3 ± 1.6	4.3 ± 1.1	$\cdots$	$\cdots$	13.9 ± 2.9
Other prompt	0.16 ± 0.04	0.14 ± 0.04	0.90 ± 0.20	0.63 ± 0.14	0.39 ± 0.09	0.22 ± 0.05	2.4 ± 0.5
$W^{\pm}W^{\pm}jj$ strong	0.35 ± 0.13	0.15 ± 0.05	2.9 ± 1.0	1.2 ± 0.4	1.8 ± 0.6	0.76 ± 0.25	7.2 ± 2.3
Expected background	5.8 ± 1.4	4.1 ± 1.1	28 ± 4	18.8 ± 2.6	7.7 ± 0.9	5.1 ± 0.6	69 ± 7
$W^{\pm}W^{\pm}jj$ electroweak	5.6 ± 1.0	2.2 ± 0.4	24 ± 5	9.4 ± 1.8	13.4 ± 2.5	5.1 ± 1.0	60 ± 11
Data	10	4	44	28	25	11	122

2.4.1 Non-Prompt Lepton Background

The second-largest background (Table 2.3) of the $W^{\pm}W^{\pm}jj - EWK$ signal is the non-prompt lepton background (sometimes also referred as 'fake background'), mainly coming from $W + jets$, $t\bar{t}$, *single* t or QCD multijet processes, which have up to five orders of magnitude higher cross-sections as the signal (see Figure A.1, Appendix). Non-prompt leptons are either originating in hadron decays or in jets that are misidentified as leptons. Due to the jet fragmentation being extremely difficult to model, the probability of a jet being detected as a lepton is not modeled well in MC. Therefore, the background prediction can not be taken from MC simulations. [5]
How a background estimation can still be obtained using a data-driven technique will be outlined in the following section. Leptons originating in Photon Conversions (which are a part

of e/γ conversions in Table 2.3) are included in the non-prompt lepton background in this thesis.

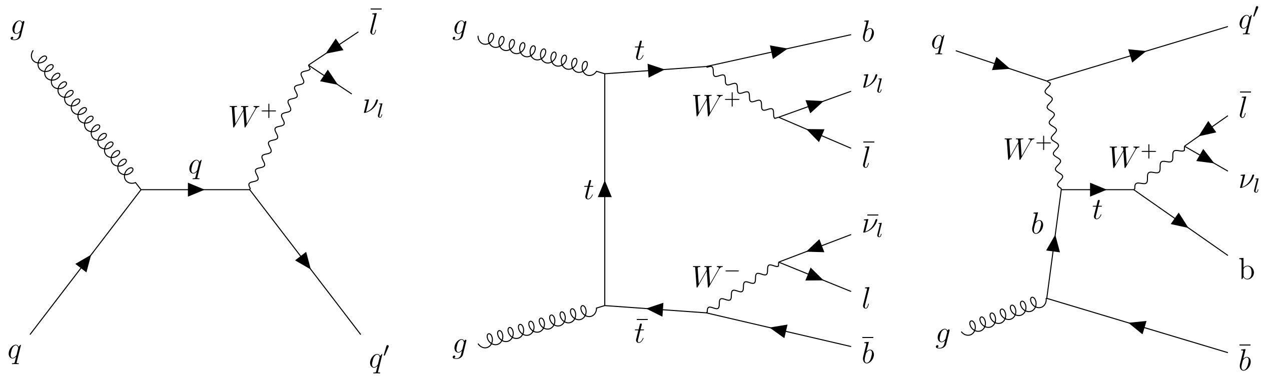


Figure 2.3: Example feynman diagrams for $W + jets$, $t\bar{t}$ and *single* t processes.

2.5 Datadriven Background Estimation - Fake Faktor Method

For reasons motivated in Section 2.4, some backgrounds can only be estimated from data reliably. This is the case for the non-prompt lepton background. A background estimation can be obtained using the 'fake factor method' [4]. For this, it is necessary to introduce two different object quality criteria on *reconstruction-level*: 'Analysis leptons' (also referred to as *Ana* in the following) and 'None-Analysis leptons' (referred to as *Noneana* in the following). *Ana* leptons pass all of the lepton selection criteria, while *Noneana* leptons are failing some of the *Ana* selection criteria, but still pass a loosened selection. The criteria for these two object quality levels are summarized in detail in Section 4.3.

'Fake' leptons are originating in Hadron decays, Photon Conversions, or misidentified objects. Leptons with other origins (*not* Fake) are classified as 'Prompt'; typically, they are included in the Matrix-Element calculation (technically: they are classified as *Isolated* by the `MCTruthClassifier` [10]). This classification is done on *truth-level* and is only possible in MC samples. An equation for the number of events with fake leptons will be derived in the following. For this, it is necessary to define four probabilities (i is the index of the lepton):

e_i – Probability that a *Prompt* (denoted as P) lepton is identified as a *Ana* lepton (A)
$\bar{e_i}$ – Probability that a *Prompt* lepton is identified as a *Noneana* lepton (N)
f_i – Probability that a *Fake* lepton is identified as a *Ana* lepton
$\bar{f_i}$ – Probability that a *Fake* lepton is identified as a *Noneana* lepton

By using those probabilities, it is possible to express the number of events on *reco-level* through the number of events on *truth-level* (N_{AA} is the number of events with two *Ana* selected leptons):

$$\begin{pmatrix} N_{AA} \\ N_{AN} \\ N_{NA} \\ N_{NN} \end{pmatrix} = \begin{pmatrix} e_1e_2 & e_1f_2 & f_1e_2 & f_1f_2 \\ e_1\bar{e_2} & e_1\bar{f_2} & f_1\bar{e_2} & f_1\bar{f_2} \\ \bar{e_1}e_2 & \bar{e_1}f_2 & \bar{f_1}e_2 & \bar{f_1}f_2 \\ \bar{e_1}\bar{e_2} & \bar{e_1}\bar{e_2} & \bar{f_1}\bar{e_2} & \bar{f_1}\bar{f_2} \end{pmatrix} \cdot \begin{pmatrix} N_{PP} \\ N_{PF} \\ N_{FP} \\ N_{FF} \end{pmatrix}. \tag{2.1}$$

By multiplying each row with a certain term one obtains:

$$\begin{pmatrix} N_{AA} \\ -N_{AN}\frac{f_2}{\bar{f_2}} \\ -N_{NA}\frac{f_1}{\bar{f_1}} \\ N_{NN}\frac{f_1f_2}{\bar{f_1}\bar{f_2}} \end{pmatrix} = \begin{pmatrix} e_1e_2 & e_1f_2 & f_1e_2 & f_1f_2 \\ -e_1\bar{e_2}\frac{f_2}{\bar{f_2}} & -e_1\bar{f_2}\frac{f_2}{\bar{f_2}} & -f_1\bar{e_2}\frac{f_2}{\bar{f_2}} & -f_1\bar{f_2}\frac{f_2}{\bar{f_2}} \\ -\bar{e_1}e_2\frac{f_1}{\bar{f_1}} & -\bar{e_1}f_2\frac{f_1}{\bar{f_1}} & -\bar{f_1}e_2\frac{f_1}{\bar{f_1}} & -\bar{f_1}f_2\frac{f_1}{\bar{f_1}} \\ \bar{e_1}\bar{e_2}\frac{f_1f_2}{\bar{f_1}\bar{f_2}} & \bar{e_1}\bar{f_2}\frac{f_1f_2}{\bar{f_1}\bar{f_2}} & \bar{f_1}\bar{e_2}\frac{f_1f_2}{\bar{f_1}\bar{f_2}} & \bar{f_1}\bar{f_2}\frac{f_1f_2}{\bar{f_1}\bar{f_2}} \end{pmatrix} \cdot \begin{pmatrix} N_{PP} \\ N_{PF} \\ N_{FP} \\ N_{FF} \end{pmatrix}. \tag{2.2}$$

Furthermore, by adding rows (1), (2), (3), and (4), all the terms in the highlighted submatrix cancel out, and the following equation can be obtained:

$$N_{AA} - N_{AN}\frac{f_2}{\bar{f_2}} - N_{NA}\frac{f_1}{\bar{f_1}} + N_{NN}\frac{f_1f_2}{\bar{f_1}\bar{f_2}} = \left(e_1e_2 - e_1\bar{e_2}\frac{f_2}{\bar{f_2}} - \bar{e_1}e_2\frac{f_1}{\bar{f_1}} + \bar{e_1}\bar{e_2}\frac{f_1f_2}{\bar{f_1}\bar{f_2}}\right) \cdot N_{PP}. \tag{2.3}$$

The number of events with fake leptons can be calculated by subtracting the number of prompt leptons that pass the *Ana* selection from the total number of leptons that pass the *Ana* selection. Through reordering the terms of Equation 2.3, this can be expressed as:

$$N_{fake} = N_{AA} - e_1e_2N_{PP} = \left(N_{AN}\frac{f_2}{\bar{f_2}} - e_1\bar{e_2}\frac{f_2}{\bar{f_2}} \cdot N_{PP}\right) + \left(N_{NA}\frac{f_1}{\bar{f_1}} - \bar{e_1}e_2\frac{f_1}{\bar{f_1}} \cdot N_{PP}\right) \tag{2.4}$$

$$- \left(N_{NN}\frac{f_1f_2}{\bar{f_1}\bar{f_2}} - \bar{e_1}\bar{e_2}\frac{f_1f_2}{\bar{f_1}\bar{f_2}} \cdot N_{PP}\right). \tag{2.5}$$

This equation can be simplified by factorizing the fractions:

$$N_{fake} = (N_{AN} - e_1\bar{e_2} \cdot N_{PP}) \cdot \frac{f_2}{\bar{f_2}} + (N_{NA} - \bar{e_1}e_2 \cdot N_{PP}) \cdot \frac{f_1}{\bar{f_1}} \tag{2.6}$$

$$- (N_{NN} - \bar{e_1}\bar{e_2} \cdot N_{PP}) \cdot \frac{f_1 f_2}{\bar{f_1}\bar{f_2}}. \tag{2.7}$$

Finally, by identifying the terms in front of the factorials as the reducible background contributions ($N_{ij}^{red} = N_{ij} - e \cdot N_{PP}$, with e as a product of two probabilities, which depend on the selection criteria $i, j \in \{A, N\}$) and defining the fake factor as $F_i \equiv \frac{f_i}{\bar{f_i}}$, this leads to:

$$N_{fake} = N_{AN}^{red} \cdot F_2 + N_{NA}^{red} \cdot F_1 - N_{NN}^{red} \cdot F_1 \cdot F_2 \approx N_{AN}^{red} \cdot F_2 + N_{NA}^{red} \cdot F_1 + \mathcal{O}(F^2). \tag{2.8}$$

The last equation is obtained by neglecting double fakes. It still has contributions for both possibilities to assign the leptons. By introducing a new way to enumerate the leptons, this equation can be simplified further: the *Tag* lepton is the lepton expected to be *Prompt*, while the *Probe* lepton is the lepton expected to be *Fake*. In cases with one *Ana* lepton, the *Ana* lepton is regarded as the *Tag*, while the *Noneana* lepton is considered to be the *Probe*. This leads to a problem when considering the assignment of the *Tag* lepton in cases where there are two *Ana* (or two *Noneana*) leptons, which will not be discussed further here since it is beyond the scope of this study. For events with two *Ana* leptons, the *Tag* is the *Prompt* lepton, while the *Probe* is the highest p_T *Fake* lepton. In the case of events with two *Prompt* leptons, *Tag* is the highest p_T *Prompt* lepton, whereas the *Probe* is the subleading p_T *Prompt* lepton. In the following, the first index is the *Tag* lepton, and the second index is the *Probe* lepton unless stated otherwise. With this way of enumeration, Equation 2.8 can be simplified more, since $N_{NA} = 0$. Because the underlying effects that produce non-prompt leptons differ for electrons and muons, it is necessary to distinguish between electron fakes and muon fakes and to calculate their fake factors separately (so $N_{fake,\mu}+N_{fake,e}$ is the actual background estimation, using Equation 2.8 twice).

The fake factor is determined in data, in a control-region containing mainly fake leptons. Two leptons are required in the control-region, while one needs to satisfy the *Ana* selection criteria. A detailed definition of the control-region can be found in Section 4.2, Paragraph two. It is possible to define the fake factor as the ratio of the probability to select a fake lepton as *Ana* over the probability of selecting a fake lepton as *Noneana*, as seen in Equation 2.8. Generally, it is binned in p_T (*transversal momentum*). Therefore Equation 2.8 is a sum over the p_T bins of the fake factor and the number of events.

Another commonly used way of enumerating the leptons is to order them by their p_T (first index is the leading p_T lepton, second index is the subleading p_T lepton). This has the benefit that the two fake factors in Equation 2.8 are the same: $F_1 = F_2$, since the fake factor is binned in p_T and both leptons are required to satisfy the same selection criteria.

$$F = \frac{f}{\bar{f}} = \frac{N_{A\underline{A}} - N_{A\underline{A}}^{Prompt,MC}}{N_{ctrl}} \left(\frac{N_{A\underline{N}} - N_{A\underline{N}}^{Prompt,MC}}{N_{ctrl}} \right)^{-1} = \frac{N_{A\underline{A}} - N_{A\underline{A}}^{Prompt,MC}}{N_{A\underline{N}} - N_{A\underline{N}}^{Prompt,MC}} \tag{2.9}$$

$N^{Prompt,MC}$ is the number of Prompt leptons, taken from MC simulation. Each of the variables used in Equation 2.9 is binned in the fake leptons p_T, visualized by the underlines. As a result, the fake factor is also binned in the fake leptons p_T. Equation 2.9 is still dependent on the assignment (here, *Tag/Probe* is used to enumerate the leptons) of the leptons. For the above formula, a perfect assignment (on *truth-level*) for the *Tag/Probe* enumeration is regarded, which means that the *Tag* lepton is always Prompt. Any assignment for *Tag* and *Probe* leptons in data will not be perfect, so there is always a certain deviation from the actual perfect fake factor. Assigning the leptons using *truth* information is only possible in MC.

As mentioned, the fake factor is defined in a control-region, but the fake estimation is needed in the signal-region. The estimation is obtained by applying the fake factor calculated in the control-region to the signal-region under the assumption that $F_{CR} \approx F_{SR}$. This conjecture is only valid when the fake origins are similar in both regions and the control-region is disjoint from the signal-region.

3 Experiment

3.1 LHC and the ATLAS Detector

The `Large Hadron Collider (LHC)` is the most powerful particle collider in the world, capable of colliding proton-beams at center-of-mass energies of up to 13 TeV. The acceleration of the particles takes part in the 27 km diameter ring of the `LHC` and its three primary pre-accelerators. The protons beams consist of bunches, 2808 per beam, with a bunch-spacing of 25 ns and each bunch containing 1.1×10^{11} protons [8]. In order to idedentify the approximately 1000 [2] particles that emerge per bunch from the collision-point, large detectors are needed. ATLAS (**A** **T**oroidal **L**HC **A**pparatu**S**) is, together with CMS (**C**ompact **M**uon **S**olenoid), one of the general-purpose detectors of the `LHC`, on whose interaction-point it is centered. It has been built to analyze and identify the particles that are emerging in the pp-collisions. [2, 8, 6]

3.1.1 Coordinate System

The origin of the coordinate system that is used to describe the detector is defined on one of the interaction-points of the `LHC`. Incoming proton-beams are parallel or anti-parallel to the z-axis of the coordinate system. The x-axis is parallel to a vector that points from the interaction-point inside the ATLAS detector towards the center of the `LHC`-ring. The positive y-axis is aligned such that a right-handed cartesian coordinate system is constructed. It is useful to define two commonly used angles: the polar angle θ is the angle towards the beam axis and the azimuthal angle ϕ, which is defined around the axis of the beam (in the x-y plane). The proton-collisions take place in the center-of-mass frame of the two protons, which is different colliding partons center-of-mass frame, and since the fraction of momentum of the proton that each parton carries is unknown, the total momentum along the beamline is also unknown. Therefore it is necessary to describe the physics using variables that are invariant under Lorentz-Transformations along the z-axis (beam direction). Instead of the total momentum, the transverse momentum p_T is the variable used, which is a projection of the momentum in the x-y plane. Other commonly used variables are the Rapidity y and the Pseudo-Rapidity η (which are not invariant themselves, but differences in rapidity are) instead of the angle θ:

$$\eta = -ln\left(\tan\frac{\theta}{2}\right) \tag{3.1}$$

$$y = \frac{1}{2}ln\left(\frac{E+p_z}{E-p_z}\right) \tag{3.2}$$

In Equation 3.2, E is the energy of a particle, and p_z is its momentum along the beam axis. The definition of the pseudo-rapidity is constructed in a way that $\eta \approx y$ (η converges to y) in cases where the mass of an object can be neglected ($m \ll p$). It is also possible to define an angle separation $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$ in the $\eta - \phi$ space, which is also invariant under Lorentz-boosts along the beam axis. [2, 14]

3.1.2 Subdetectors

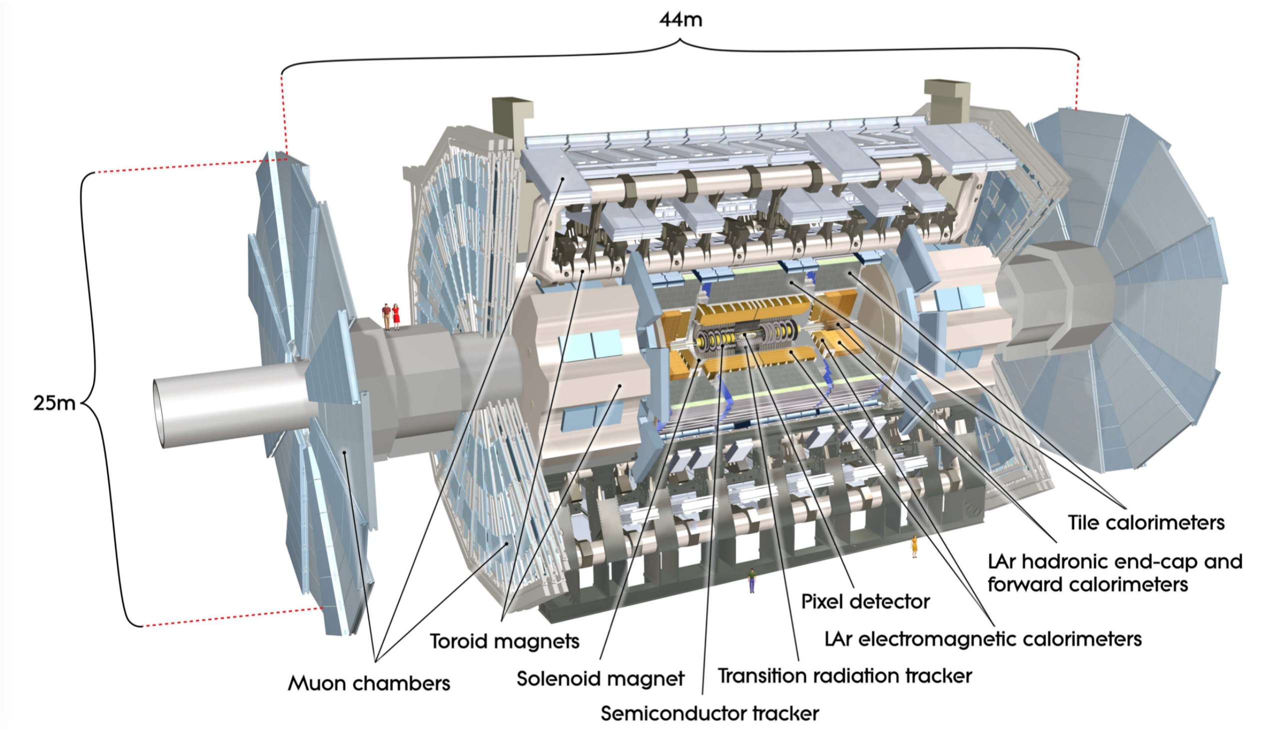


Figure 3.1: The ATLAS detector [2].

The ATLAS detector follows an onion-skin like layout and can be divided into four main parts (Figure 3.1): the inner detector, the electromagnetic calorimeter, the hadronic calorimeter, and the muon spectrometer. The inner detector is surrounding the main interaction point with coverage of $|\eta| < 2.7$, containing semiconductor pixel- and strip-detectors in the inner tracking volume and straw-tube tracking detectors in the outer part. It is used to measure momenta and trajectories of particles and is therefore penetrated by a magnetic field of 2 T, generated by the solenoid magnet. This causes the deflection of particles in the magnetic field, depending on their momentum and charge. Electromagnetic- and hadronic calorimeters are embracing the inner detector, with coverages of $|\eta| < 3.2$, and $|\eta| < 4.9$, respectively, to measure the energy of particles. The electromagnetic calorimeter consists of Lead-liquid-argon calorimeters arranged in a barrel around the inner detector, as well as Lead-liquid-argon calorimeter as the end caps of the barrel. The hadronic calorimeter, consisting of scintillator-tile (in a barrel arrangement) and lead-liquid-argon calorimeters (as the end caps), is surrounding the electromagnetic calorimeter. It measures the energies of particles not stopped by the electromagnetic calorimeter, which are primarily hadrons (and muons, which are not stopped by the hadronic calorimeter either, meaning they do not deposit much of their energy).

All particles, apart from muons, that can be detected are not able to pass the calorimeter system. Therefore, a muon spectrometer is embracing the calorimeters, to accurately measure muon momenta and trajectories in regions of $|\eta| < 2.7$. The only particles known so far that can not be detected with the ATLAS detector are Neutrinos. They can only be identified through the missing fraction of the total p_T (which should be 0 if all particles were detected) that they carry (*missing transverse energy, MET*). [2]

4 Fake Studies in Generated Events

Monte Carlo samples used for the following studies can be found in Appendix A.2. The evaluation is based on the `CommonAnalysisFramework`. Throughout this thesis, the `W+jets Powheg` Monte Carlo samples are used to extract the fake factors. In these samples, *pp*-collisions at center-of-mass energies of 13 TeV are simulated, as performed at the LHC during Run 2 in 2015-2016. These simulated events are scaled to a luminosity of $36.1\,\text{fb}^{-1}$. All Monte Carlo samples pass a full detector simulation. The leptons are enumerated as Tag (first index) and Probe (second index). A perfect assignment is used, which means that the Tag lepton is always the Prompt lepton, while the Probe lepton is always the highest p_T non-prompt lepton. Cases with no Prompt lepton are neglected. The perfect assignment has to be used due to technical reasons.

In the following, the reconstruction, selection criteria for objects, and event selection criteria will be summarized. These apply to data as well as Monte Carlo samples, but only Monte Carlo samples are used in this thesis. Furthermore, a Monte Carlo Closure Test will be described, which probes the implementation of the fake factor method. The influence of fake origins and cuts on the fake factors will be investigated, as well as the charge composition of the leptons.

4.1 Reconstruction and Object Selection

As described in Section 3.2, multiple *pp*-collisions can occur in one bunch. Therefore, several vertices can be reconstructed for each event. By regarding the sum of the squared transverse momentum of the associated tracks and choosing the vertex with the maximum sum of p_T^2, a primary vertex is selected. Events must have a minimum of three tracks associated with the primary vertex in order to be considered. The several tracks and energy deposits have to be assigned to actual physics objects, in order to be analyzed further.

An object is identified as a muon when it passes several kinematic criteria as well as criteria for the longitudinal and transversal impact parameters, to ensure that the track in the inner detector is emerging from the primary vertex. In order to reduce contributions from *b*-jets or in-flight decays, *Analysis* muons are required to fulfill an isolation criterion. *Noneana* Muons used in the fake factor method must fail the isolation criterion or the requirement on the transversal impact parameter for *Analysis* Muons.

Table 4.1: Muon selection criteria.

	Preselection	**None Analysis**	**Analysis**
Identification	`Loose`	`Medium`	`Medium`
Kinematic Acceptance	$p_T > 6\,\text{GeV}$	$p_T > 27\,\text{GeV}$	$p_T > 27\,\text{GeV}$
Geometrical Acceptacnce	$\|\eta\| < 2.7$	$\|\eta\| < 2.5$	$\|\eta\| < 2.5$
Longitudinal Impact Parameter Requirement	$\|z_0 \times \sin\theta\| < 0.5\,\text{mm}$	$\|z_0 \times \sin\theta\| < 0.5\,\text{mm}$	$\|z_0 \times \sin\theta\| < 0.5\,\text{mm}$
Transversal Impact Parameter Requirement	$\frac{d_0}{\sigma_{d_0}} < 10$	$\frac{d_0}{\sigma_{d_0}} < 10$	$\frac{d_0}{\sigma_{d_0}} < 3$
Isolation requirement	`none`	`none`	`FCTight`
Additional requirements	-	Fail `Gradient` Isolation or $\frac{d_0}{\sigma_{d_0}} < 3$	-

Table 4.2: Electron selection criteria.

	Preselection	**None Analysis**	**Analysis**
Identification	`LooseLH`, hit in b-layer	`MediumLH`	`TightLH`
Kinematic Acceptance	$p_T > 6\,\text{GeV}$	$p_T > 27\,\text{GeV}$	$p_T > 27\,\text{GeV}$
Geometrical Acceptacnce	$\|\eta\| < 2.47$	$\|\eta\| < 2.47$, excluding $1.37 \leq \|\eta\| \leq 1.52$	$\|\eta\| < 2.47$, excluding $1.37 \leq \|\eta\| \leq 1.52$
Longitudinal Impact Parameter Requirement	$\|z_0 \times \sin\theta\| < 0.5\,\text{mm}$	$\|z_0 \times \sin\theta\| < 0.5\,\text{mm}$	$\|z_0 \times \sin\theta\| < 0.5\,\text{mm}$
Transversal Impact Parameter Requirement	$\frac{d_0}{\sigma_{d_0}} < 5$	$\frac{d_0}{\sigma_{d_0}} < 5$	$\frac{d_0}{\sigma_{d_0}} < 5$
Isolation requirement	`none`	`none`	`Gradient`
Author	-	-	`1`
Additional requirements	-	Fail `Gradient` Isolation or `TightLH` Identification	-

Energy deposits in the electromagnetic calorimeter can be associated with electrons (`Author` = 1), and photons (`Author` = 4) (or both, `Author` = 16), depending on the characteristics of the track in the inner detector [11]. If no photon-conversion vertex can be reconstructed, the object is classified as an electron. *Analysis* electrons are required to be isolated, have `Author` = 1 and pass particular kinetic cuts as well as cuts on the impact parameters. The *Noneana* electrons must fail the `Gradient` isolation criterion or the `TightLH` Identification criterion. All muon and electron selection criteria for different object qualities can be found in Tables 4.1 and 4.2.[5]

A particle might be reconstructed by multiple reconstruction algorithms, which could lead to double counting. To account for this, an overlap removal is applied to the objects satisfying the preselection criteria. [5]

4.2 Event Selection

In order to reduce the background contributions described in Section 2.4 and to increase the ratio of signal to background, several event cuts are applied: Reconstructed events are required to have exactly two leptons, which both need to be *Ana* selected. The purpose of this is to suppress backgrounds coming from WZ or ZZ processes, which have more than two final state leptons. Aiming to reduce the background due to charge misidentification caused by the Drell-Yann process, a Z-mass veto and a $|\eta|$ cut are applied in the ee-channel. Since the $WWjj - EWK$ final state contains two neutrinos, events need a missing transverse momentum of 30 GeV, which also reduces the charge misidentification background. A b-jet veto is applied to reduce the background from processes including top quarks (also see Figure 2.3). Additionally, events are expected to have at least two jets, with $p_T > 65\,\text{GeV}$ and $p_T > 35\,\text{GeV}$. The two largest p_T-jets are identified as tagging jets, which are then used for the calculation of $|\Delta y_{jj}|$ and m_{jj}. To separate the $WWjj - EWK$ signal from the irreducible QCD background mentioned in Section 2.4, a cut on the VBS signature is applied, requiring two high invariant mass jets with a large angular separation in the event. All event cuts applied are summarized in Table 4.3. [5]
A very loose control-region is chosen (Table 4.4) for the fake estimation, to be less limited by statistics and to have a region containing primarily events with fake leptons. As shown in Section 2.5, Equations 2.8 and 2.9, it is necessary to have at least one *Ana* lepton. The other leptons are required to pass the *Noneana* selection. In order to enrich the control-region with events containing a Prompt lepton originating in the $W^{\pm}$-decay, the missing transverse energy is required to be greater than 30 GeV.

For the fake factor method, control- and signal-region need to be disjoint in order to be statistically independent. Since this needs to be done by inverting a particular cut of the signal-region and applying it in the control-region (since signal-region cuts should not be changed), this is not satisfied yet. For this, it is necessary to regard all processes contributing

Table 4.3: VBS Signal Region

Exactly two Analysis leptons with identical electrical charge
ee-chanel: $\lvert\eta\rvert < 1.37$
$m_{ll'} > 20\,\text{GeV}$
ee-chanel: $\lvert m_{ee} - m_Z\rvert > 15\,\text{GeV}$
$E_T^{miss} > 30\,\text{GeV}$
Atleast two jets Leading jet: $p_T > 65\,\text{GeV}$, Subleading jet: $p_T > 35\,\text{GeV}$
b-jet veto
$\lvert\Delta y_{jj}\rvert > 2$
$m_{jj} > 500\,\text{GeV}$

Table 4.4: The control-region used to extract the fake factors.

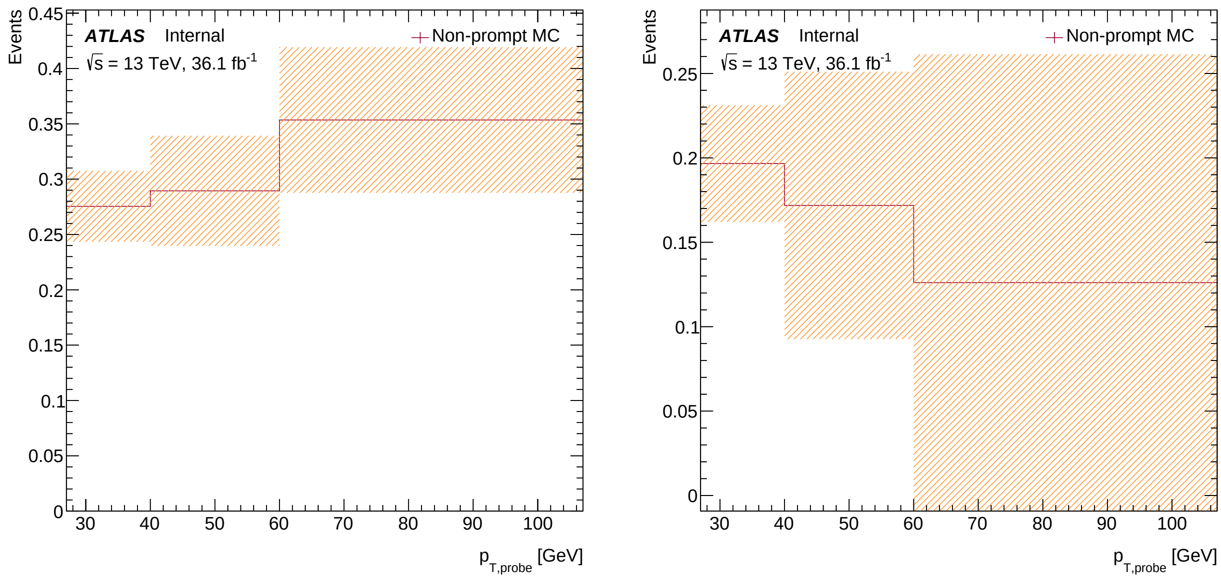
Atleast two leptons (None Analysis criteria as baseline) One Analysis lepton
$E_T^{miss} > 30\,\text{GeV}$

to the non-prompt lepton background and their fake origins in control- and signal region, when defining the control-region, which will not be done in this thesis.
Section 4.7 will still give a prospect on how different inverted signal-region cuts can influence the fake factor.

4.3 Integrated Monte Carlo Closure Test

A Monte Carlo Closure Test is performed to test if the technical implementation of the fake factor calculation is working as intended. For this, the fake factor is extracted from a Monte Carlo sample (`W+jets Powheg` in the control-region, see Appendix A.2). To obtain the fake factor binned in p_T of the Probe lepton, the Probe leptons p_T distribution under the condition that Tag and Probe leptons are both Ana selected is integrated over a specific p_T interval. The same is done under the condition that the Tag lepton is Ana selected, and the Probe lepton is Noneana selected. By dividing the two, the fake factor in the p_T interval used for the integration is calculated. Subtracting the Prompt contribution is not necessary since only events where the Tag is Prompt, and the Probe is Fake are considered due to the perfect assignment and by neglecting double fakes. Repeating this for all p_T intervals results in a p_T binned fake factor. This calculation is done for electron fakes, and muon fakes separately, leading to separate fake factors for the two lepton flavors. Figure 4.1 shows the fake factors for electrons and muons. A broad binning is chosen in order to lower the statistical uncertainties on the fake factors.

This fake factor is then applied in the control-region as an event weight to events where the Tag lepton is Ana selected, and the Probe is Noneana selected, to get an estimation of the fake leptons that pass the Ana selection.



(a) Electron fakes **(b)** Muon fakes

Figure 4.1: Fake factors extracted in the control-region.

Table 4.5: Event yields for the Monte Carlo prediction for Prompt, Fake (both Ana selected) compared to the data-driven prediction, separated for the different channels. The number of events is not rounded, because, for this technical test, the agreement should be perfect, independent of the uncertainties.

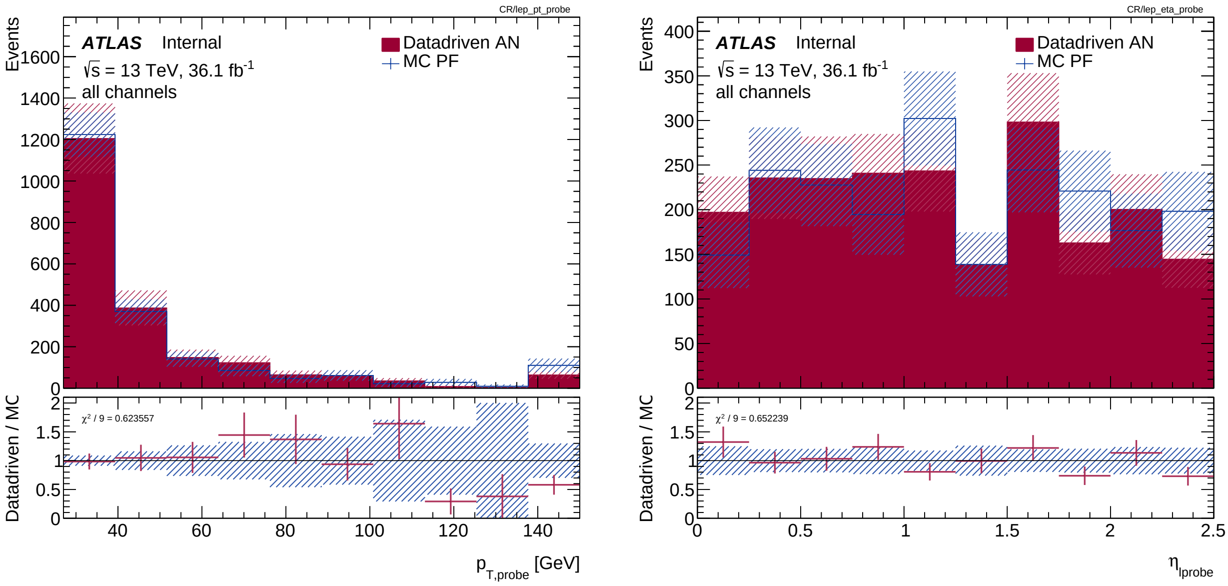
	ee	$e\mu$	μe	$\mu\mu$	all
MC PF	758.4 ± 80 ± 0	476.2 ± 70 ± 0	656.8 ± 80 ± 0	205.0 ± 50 ± 0	2096.4 ± 140 ± 0
MC FP	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
DD AN	799.6 ± 40 ± 120	437.5 ± 30 ± 80	629.6 ± 40 ± 90	229.7 ± 20 ± 50	2096.4 ± 70 ± 340
DD NA	0 ± 0	0 ± 0	0 ± 0	0 ± 0	0 ± 0
total MC Wjets	758.4 ± 80 ± 0	476.2 ± 70 ± 0	656.8 ± 80 ± 0	205.0 ± 50 ± 0	2096.4 ± 140 ± 0
data-driven Wjets	799.6 ± 40 ± 120	437.5 ± 30 ± 80	629.6 ± 40 ± 90	229.7 ± 20 ± 50	2096.4 ± 70 ± 34

The agreement between the Monte Carlo prediction (Prompt, Fake, both leptons are Ana selected) and the fake factor applied to AN events (shortened as data-driven in the following) in the same control-region as used for the extraction is perfect, which is expected since a perfect assignment is used. Therefore this only shows that the technical implementation is working correctly. Closure still needs to be verified when using a different assignment. With the `W+jets Powheg` sample, it is not possible to test how good the closure is when applying the fake factor extracted in the control-region to the signal region since there are no events passing the signal region cuts. This closure is only expected to be valid after the control-region is adjusted such that it is disjoint from the signal region and the fake origins in signal and control-region match since the fake origins are closely related to the fake factors.

4.4 Differential Monte Carlo Closure Test

The distribution of the probe leptons p_T for AN (Ana, Noneana) events scaled with the fake factor ('data-driven') is expected to match the Monte Carlo AA (Ana, Ana) prediction very well since the fake factor is binned in the Probe leptons p_T. Small deviations can be expected since the binning of the fake factor is broader than the binning of the p_T distribution of the probe lepton.

As seen in Figure 4.2, the difference between the Monte Carlo prediction and the data-driven method is higher for p_T values in the highest p_T bin of the fake factor, which matches the expectation described above. The deviations are within the respective uncertainties, as seen in the ratio plot.



(a) Probe lepton p_T distributon **(b)** Probe lepton $|\eta|$ distributon

Figure 4.2: p_T and $|\eta|$ distributions of the Probe lepton. The Monte Carlo prediction for Prompt, Fake (both Ana selected) is compared to the result of the data-driven fake factor method. Systematic uncertainties of the fake factor method are included.

It could be expected that the data-driven η distribution of the probe lepton is not matching the Monte Carlo AA prediction very well since the fake factor generally also depends on the detector region in which the fake object is reconstructed [4]. A typical correction applied to the fake factor is a η-correction. This mismatching can not be observed here since the data-driven result is compatible with the Monte Carlo prediction within the respective statistical uncertainties.

Distributions for other observables (number of jets and dilepton invariant mass) can be found in Appendix A.3. As the ratio plot shows, the Monte Carlo prediction for AA events and the data-driven method are compatible within their uncertainties.

4.5 Fake Origins

The origins of the fake leptons are classified using the `MCTruthClassifier`. All possible origins can be found in Figure A.4 (a)-(c). For studying the fake background, it is useful to group the fake origins into six categories, as seen in Table 4.6:

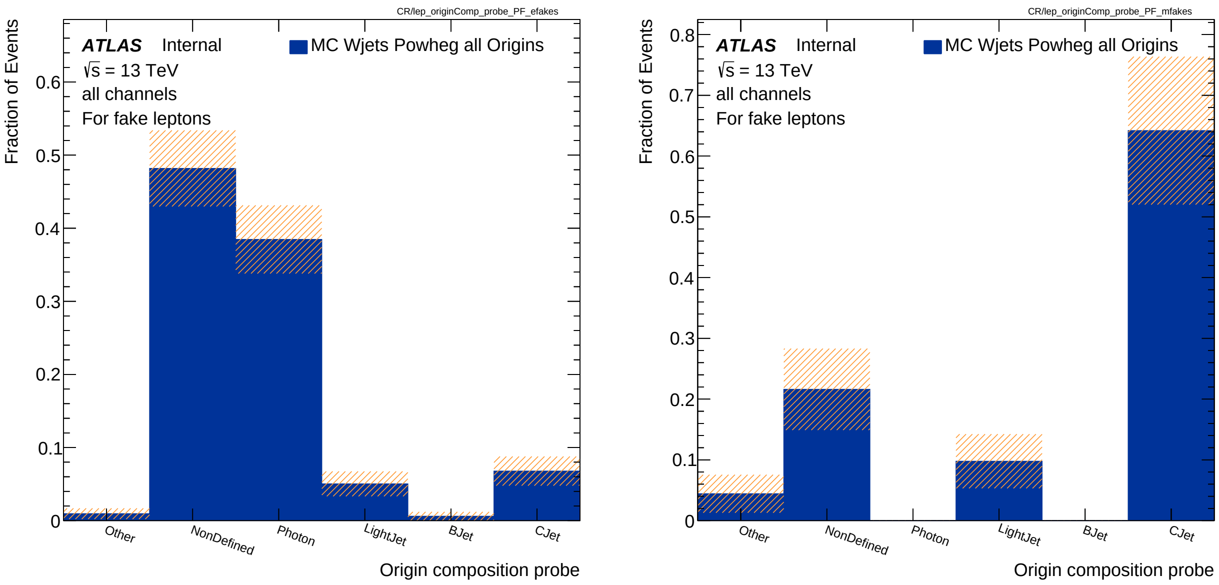
Table 4.6: The different origin categories.

Category	**Origin**
Light Jets	Hadrons containing u-, d- or s-quarks
c-Jets	Hadrons containing c-quarks
b-Jets	Hadrons containing b-quarks
Photon Conversions	Photon Conversions including ISR, FSR, Bremsstrahlung
Non Defined	no truth information found or unclear assignment
other	any other origin

Figures 4.3-4.5 show the fake origins for the `W+jets Powheg` and $\mathtt{t\bar{t}}$ `Powheg` samples in the control-region as well as in the VBS signal region ($\mathtt{t\bar{t}}$ only). Fake origins for electron and muon fakes combined can be found in Appendix A.4, Figure A.3. For the `W+jets` sample, the origin for the majority of electron fakes can not be identified and is therefore classified as Non-Defined. Fake electrons with an origin that can be identified predominantly originate in Photon Conversions ($\approx 40\%$), c-Jets, and Light Jets (both $< 10\%$). Regarding the muon fakes, c-Jets account for the largest contribution ($\approx 65\%$), followed by Non-Defined fakes ($\approx 20\%$) and fakes coming from Light Jets ($\approx 10\%$). 'Other' fake leptons are produced through Tau decays, as seen in Figure A.4 (a). For the $\mathtt{t\bar{t}}$ sample, electron fakes are mainly caused by Photon Conversions ($\approx 90\%$) and b-Jets. Muon fakes mainly originate in heavy-jets such as b- ($\approx 80\%$) or c-jets ($\approx 10\%$). Photon Conversions are the primary source of fake electrons in $\mathtt{t\bar{t}}$, which is unexpected considering the two b-Jets that are produced in the decay of the t-quarks. By comparing with the fake origins in $\mathtt{t\bar{t}}$ in [4], it can be seen that this should not be the case. For electrons, according to [4] the major source of fake leptons in $\mathtt{t\bar{t}}$ is expected to be coming from heavy-jets, not photon conversions. Fake origins described in Reference [4]

are for a WZ analysis, however, since the selection criteria are similar [1] to the criteria used in this analysis, especially for electrons fakes associated to the $W^{\pm}$-Boson, the fake origins should not differ substantially.
The reason for this vast amount of fake leptons originating in Photon Conversions is unknown but has to be investigated further.

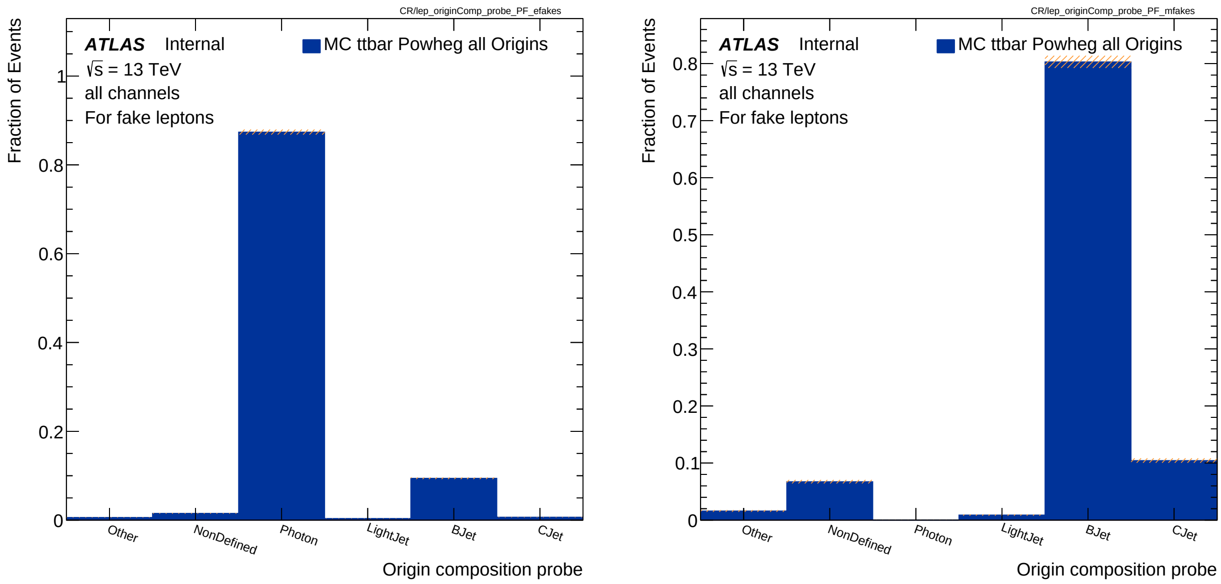
By comparing the fake origins of `tt̄` in control- and signal-region, as seen in Figures 4.4 and 4.5, it can be observed that they do not differ significantly. The dominant contributions (those with the highest fraction of the total amount of fakes) agree with the fake origins from the control-region within the statistical uncertainties, which are large, especially for muons. It also is unexpected to see a substantial fraction of fake leptons originating in b-Jets in the signal-region (dominant source of muon fakes), since a b-Jet veto is applied in the event selection of the VBS signal-region. For the `W+jets Powheg` sample, it is not possible to make such a comparison between control- and signal-region, since no events pass the VBS signal-region cuts.



(a) Electron fakes **(b)** Muon fakes

Figure 4.3: Origins of the fake lepton in the control-region for the `W+jets Powheg` sample.

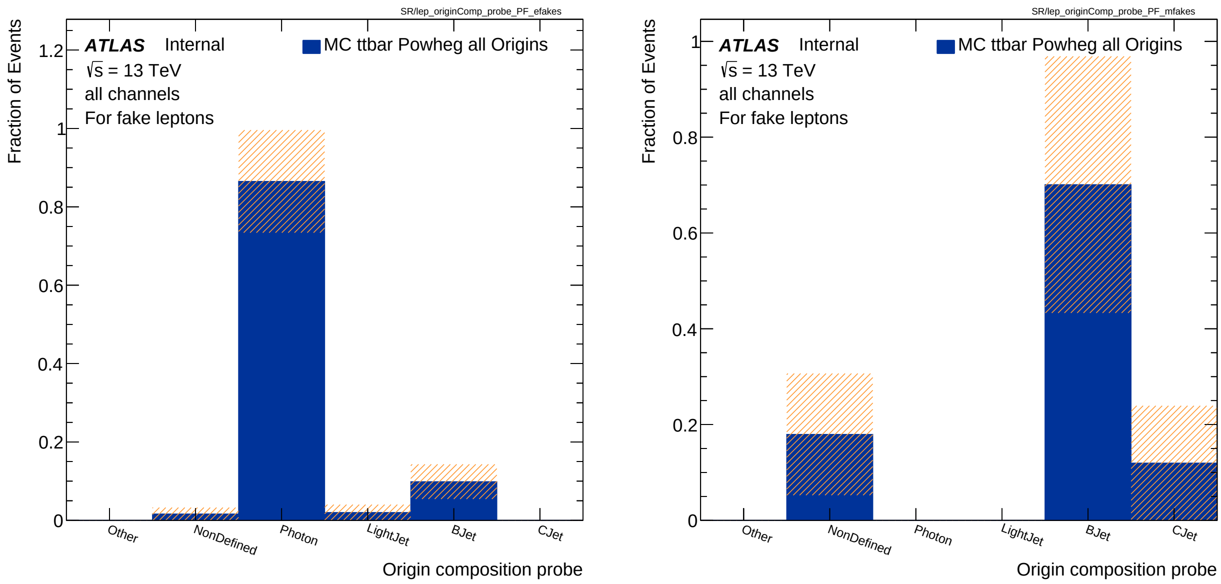
[1] In the WZ-Analysis, electron fakes associated to the Z-Boson need to satisfy `MediumLH` identification and `Gradient` isolation criteria. Fake electrons associated to the $W^{\pm}$-Boson are required to fulfill the `TightLH` identification, `Gradient` isolation, and `Author` $\neq 16$ criteria.



(a) Electron fakes

(b) Muon fakes

Figure 4.4: Origins of the fake lepton in the control-region for the `tt̄ Powheg` sample.



(a) Electron fakes

(b) Muon fakes

Figure 4.5: Origins of the fake lepton in the VBS signal-region for the `tt̄ Powheg` sample.

4.6 Influence of Fake Origins and Cuts on the Fake Factor

When using the fake factor method to get a background estimation, it is essential to know how fake factors can be influenced. Therefore, the effect of cuts and fake origins on the fake factors will be discussed in the following. Due to the high statistical uncertainties on the muon fake factors, as seen in the nominal fake factor in Figure 4.1 or the Figures A.5-A.10 in the Appendix, only the electron fake factors are compared for different origins of the fake lepton. In Figure 4.6, the electron fake factors for the two major fractions of the fake origins in the `W+jets Powheg` sample can be seen. Since they have the largest fractions of the fake origins, their fake factors have lower statistical uncertainties than the fake factors of other origin categories and are therefore suitable for comparison. Especially in the low p_T region, the electron fake factors of Photon Conversions and Non Defined origins differ significantly. The Photon Conversion fake factor for electrons is significantly higher than the nominal fake factor in the lowest p_T bin. It is decreasing for increasing p_T values. The electron fake factor for Non Defined origins is slightly lower than the nominal fake factor in the lowest p_T bin, and increasing with increasing p_T values. For higher p_T bins, the Non Defined electron fake factor is compatible with the nominal fake factor within the statistical uncertainties. For electrons originating in c-Jets (Figure 4.7), the fake factor is lower than the nominal fake factor in the first two p_T bins. Due to the substantial statistical uncertainties on the c-Jet fake factor, it is compatible with the other mentioned fake factors in the highest p_T bin.

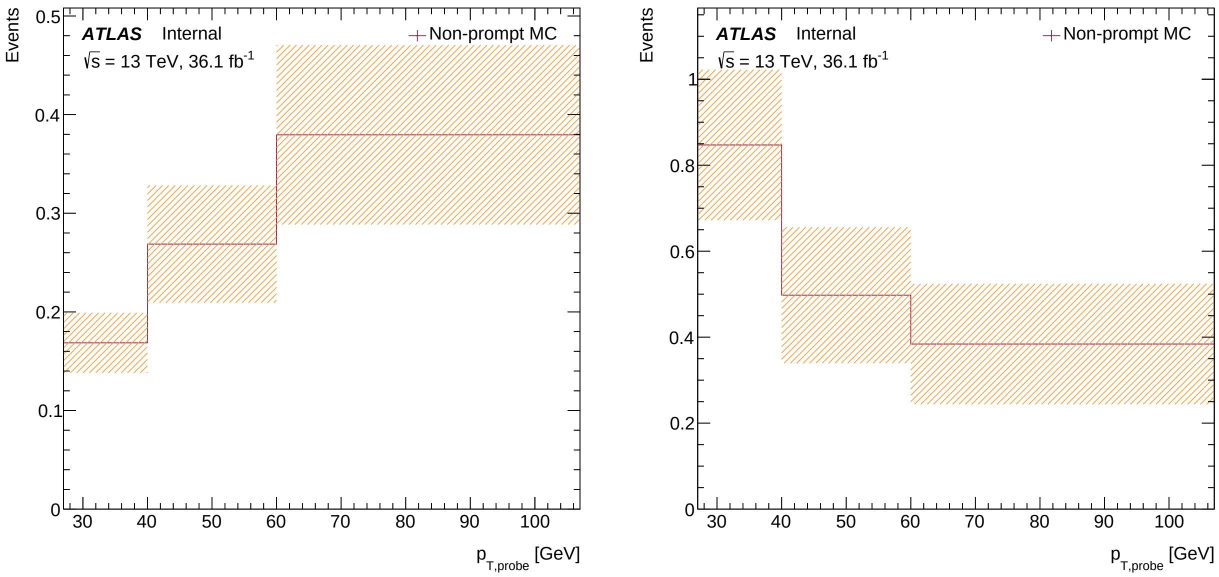
Since differences between the fake factors for different origins can be observed despite the high statistical uncertainties, it is possible to conclude that the fake origins do influence the fake factor.

The fake factors for the different origins are expected to be dependent on the nature of the fake object. Therefore, fake factors for origins producing non-prompt leptons mainly through decay should have a different fake factor than origins, which produce fakes through a misidentification by the detector (e.g., if a jet is misidentified as a lepton, included in 'Non Defined' origins) because said misidentification is very unlikely (when assuming equal fractions of the total number of fakes). This expected difference in the fake factor would follow purely from its definition (probability to identify a fake as Ana/probability to identify fake as $Noneana$). Such a difference can be observed here: despite the lower fraction of the total number of fakes (see Figure 4.3 (a)), the fake factor for Photon Conversions is higher than the fake factor for Non Defined fakes. Consequently, Photons can fake a lepton easier than objects with a Non Defined origin. However, this can not be explained from a physical standpoint, because it is not specified further (by the `MCTruthClassifier`) what causes Non Defined fakes. Due to those being the only fake origins where the fake factors (for electrons) have sufficient statistics, it is not possible to observe this when comparing the fake factors of other origins.

A summary table with the cuts applied to investigate their influence on the fake factor and potential cuts to make control- and signal region disjoint can be found in the Appendix, Table A.1. The resulting fake factors after applying those cuts can be seen in the Figures A.11 to A.22. Due to the high statistical uncertainties of the muon fake factors, only the electron fake factors will be used for the comparison.

By comparing the fake factor for electrons after applying a $m_{jj} > 200\,\text{GeV}$ (Figure A.17, Appendix) cut from the VBS signal with the nominal fake factor shown in Figure 4.1, it can be seen that the fake factor after applying the cut is higher, at least in the first bin. For the other bins, the fake factors match within their respective uncertainties. When regarding the fake factors after deploying $m_{jj} > 500\,\text{GeV}$ (Figure A.18, Appendix) and $m_{ll} \leq 20\,\text{GeV}$ (Figure A.22, Appendix) cuts, it can be seen that the fake factors are zero below (for the former cut) and above (for the latter cut) a specific p_T value. It is possible to conclude that cuts do influence the fake factors; however, with such large uncertainties, this is not very significant. For the remaining cuts displayed in Table A.1, the corresponding fake factors match the nominal fake factor within their uncertainties. For this reason, the influence of the cuts is not visible. This is a problem with the Monto-Carlo statistics and should not occur when using the `tt̄` sample or a different `W+jets` sample with better statistics to extract the fake factors.

To make the signal- and control-region disjoint, a specific signal-region cut has to be inverted (Table A.1). The used cut should not change the fake factor or lower statistics significantly. With those requirements, cuts to consider are the same sign veto, the second jet veto, as well as the $m_{jj} \leq 200\,\text{GeV}$ and $m_{jj} \leq 500\,\text{GeV}$ cuts. As seen in Table A.1, a second jet veto is cutting away the majority of the events in the `tt̄` sample and is therefore not suitable for the described purpose. A same-sign veto cut would increase the contribution of Prompt leptons from `Z+jets` processes in the control-region significantly. This leaves the two m_{jj} cuts as possible candidate cuts. The $m_{jj} \leq 500\,\text{GeV}$ does lower the statistics by 1% in `W+jets` and 7% in `tt̄`, which is rather small compared to the other cuts. Additionally, the fake factor after utilizing the said cut is still similar to the nominal fake factor. Therefore the $m_{jj} \leq 500\,\text{GeV}$ cut could be used to make the signal- and control-region disjoint. Since no fake factor for `tt̄` (or other processes in general) is calculated, the effect on the `tt̄` fake factor is unknown, but the cut does not lower the statistics significantly (Table A.1), so it would be expected that the effect is not very large. The $m_{jj} \leq 200\,\text{GeV}$ cut would lower the statistics by 4% (`W+jets`) and 32% (`tt̄`), respectively, while still conforming with the nominal fake factor. Using the latter for the definition of the control-region would have the benefit of also making the control-region disjoint to a *QCD* dominated region in $200\,\text{GeV} < m_{jj} \leq 500\,\text{GeV}$. However, the effect on the `tt̄` fake factor needs to be investigated for this. Verification of this is needed once all



(a) Non Defined origin

(b) Photon Conversion origin

Figure 4.6: Fake factors for electrons originating in Photon Conversions or with a Non Defined origin.

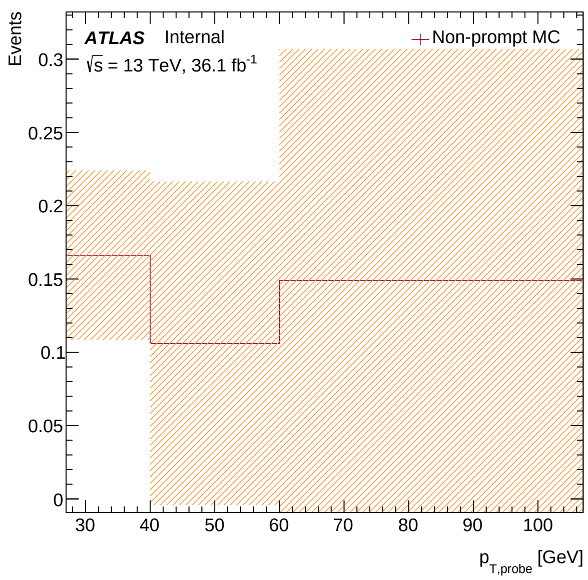


Figure 4.7: Fake factor for electrons originating in c-Jets.

processes contributing to the non-prompt lepton background are available, and especially after the vast fraction of Photon fakes in `tt̄` are explained or fixed, in case of a technical problem.

Since the fake factor is strongly dependent on the fake leptons origin as shown above (or seen in [4]), this means that if the fake origins in the signal- and control-region are compatible, the assumption $F_{CR} = F_{SR}$ is reasonable. In order to get a reliable background prediction, this needs to be ensured, e.g., by adjusting the control-region by applying specific cuts to influence the fake origins. However, for this, it is necessary to regard all processes contributing to the non-prompt lepton background.

4.7 Charge Composition of Fake Leptons

Table A.1 in the Appendix shows that the opposite sign ('SameSign Inverted') cut discards 3/4 of the events in the `W+jets Powheg` sample and 3/5 of the events in the `tt̄ Powheg` sample. This is unexpected; hence it is interesting to investigate the charge compositions of the leptons for different fake origins further.
For this, it is necessary to distinguish between the two flavors of the fake lepton. The charge composition for electron and muon fakes combined as well as the charge compositions for all fake origins combined can be found in the Appendix, Figures A.23-26. A Table form of Figures 4.9, 4.10, A.23, and A.24 can be found in the Tables A.2-A.7.

Electron fakes originating in Photon Conversions account for the largest fraction of fake leptons in the `tt̄ Powheg` sample (Figure 4.4 (a)). Combined with a large number of events in the sample, this leads to small statistic uncertainties on the charge composition of Photon Conversion fakes. The expectation is that fake leptons originating in Photon Conversions have an evenly distributed charge. However, this can not be observed. As seen in Figure 4.10 (a), the Probe lepton is more likely (almost by a factor of 2) to carry the opposite charge of the Tag lepton. Due to the small statistical uncertainties, this can not be a statistical fluctuation. The reason for this unexpected charge distribution is not known.
For different origins of the fake leptons in the `tt̄` sample, the charge composition agrees with the expectation. An example of this are b-Jets, which are expected to have an evenly distributed charge composition across all four possible combinations of the Tag and Probe leptons charge. This expectation can be deducted from the Feynman diagram (see Figure 2.3): the process is symmetric, and whether the Prompt lepton (which is the Tag lepton due to the assignment) originates in the W^+- or W^--decay should have the same probability. The same applies to the charge of the non-prompt lepton (Probe) originating in b-jets, since there is an equal amount of b and $\bar{b}$ quarks produced in the process, resulting in equal fractions for all of the possible charge combinations as seen in Figure 4.10.

For the `W+jets` sample, the statistical uncertainties on the charge composition histograms are higher (Figure 4.9). Therefore only for muon fakes caused by c-Jets, a slight discrepancy from the predicted charge composition can be observed. The prediction is again derived from the dominant Feynman diagram (Figure 4.8). As seen in the Figure, d- and s-quarks can occur in the initial state, but since $|V_{cd}|^2 \approx 0.04$ and $|V_{cs}|^2 \approx 1$, the s-quark diagram is dominant. The Prompt lepton originates in the W^--decay, while the non-prompt lepton originates in the c-decay and has the opposite charge of the Prompt lepton. To be precise, the non-prompt lepton also originates in a $W^\pm$-decay (produced in the c-decay), as seen in the Feynman diagram. However, it is still regarded as non-prompt since it is not included in the Matrix-Element calculation of the Monte Carlo sample (consequently, in order to be exact, the definition of a Prompt lepton depends on the regarded process). Since s-quarks are sea-quarks, s-and $\bar{s}$-quarks are produced with the same probability. Therefore, an equal distribution between the two bins with opposite charges for Tag and Probe leptons is expected, which is not observed in Table A.4 or Figure 4.9 (b). However, this can be a statistical fluctuation, since a 68% confidence level is used for the statistical uncertainties.

The charge composition of fake leptons needs further investigation. Especially the unexpected behavior of fake leptons originating in Photon Conversions in the `tt̄` sample needs to be understood or confirmed with a different event generator.

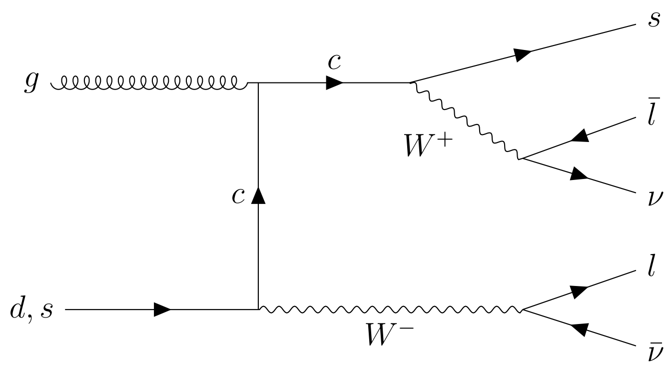
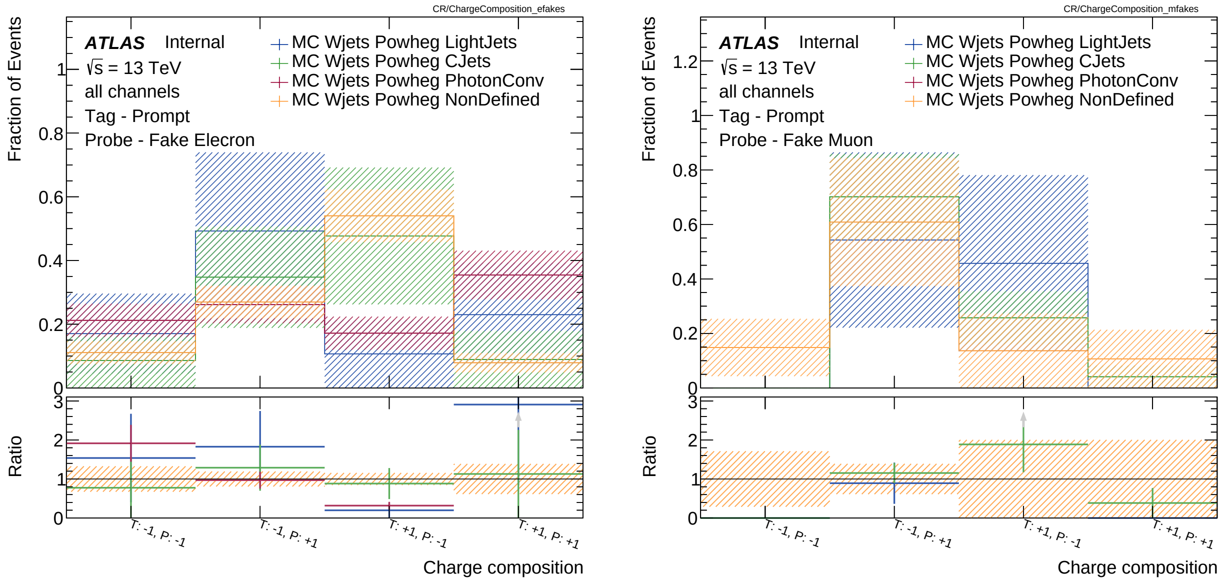
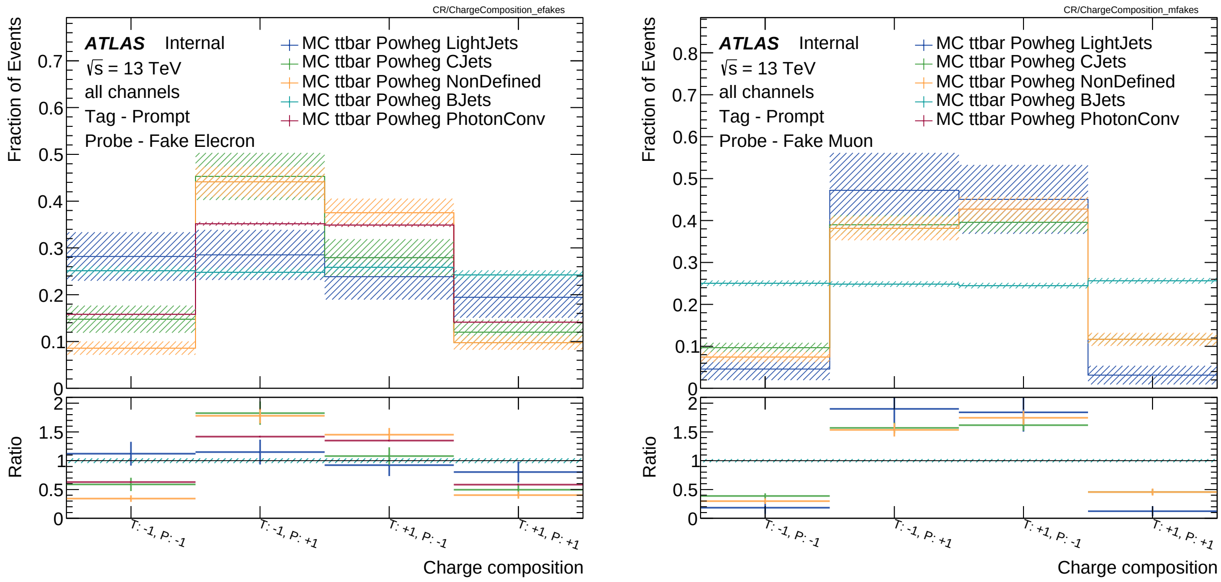


Figure 4.8: Dominant Feynman diagram for a `W+jets` process producing non-prompt leptons originating in c-Jets.



(a) Electron fakes　　**(b)** Muon fakes

Figure 4.9: Charge Composition of the Tag and Probe lepton for the `W+jets Powheg` sample in the control-region. Additionally, the ratio over the Non Defined distribution is displayed. A table form of this diagram can be found in Appendix A.6



(a) Electron fakes　　**(b)** Muon fakes

Figure 4.10: Charge Composition of the Tag and Probe lepton for the $\mathtt{t\bar{t}}$ `Powheg` sample in the control-region. Additionally, the ratio over the b-jet distribution is displayed. A table form of this diagram can be found in Appendix A.6

5 Summary and Outlook

The aim of this thesis is to improve the modeling of the non-prompt lepton background and to lay a foundation for its estimation. A data-driven technique, the 'fake factor method', is deployed for this, which is implemented and tested in a Closure Test. Monte Carlo generated samples are used to perform the test. The fake factors are extracted for a `W+jets` process, generated with the `Powheg` event generator (Appendix A.2). A perfect assignment for the Tag and Probe lepton is used. The agreement between the Monte Carlo prediction and the data-driven method applied to Monte Carlo events in the control-region, where the fake factor is extracted, is investigated. For an integrated Monte Carlo Closure Test, done by comparing the total event-yields in the control-region, this leads to a perfect agreement.
Within the uncertainties, an agreement is achieved in a differential Monte Carlo Closure Test, by comparing the Monte Carlo prediction and the data-driven method for different distributions. However, the closure still needs to be verified using a different assignment for the Tag and Probe lepton.
Fake origins of $\mathtt{t\bar{t}}$ and `W+jets` processes in the control-region (and in the signal-region for $\mathtt{t\bar{t}}$) are studied. An unexpectedly high number of non-prompt leptons originating in Photon Conversions is observed in the $\mathtt{t\bar{t}}$ process when generated with the `Powheg` event generator. The cause of this can not be identified.
Furthermore, the influence of cuts and fake origins on the fake factors is studied. It can be concluded that fake origins do influence the fake factors since they are dependent on the underlying effects that cause the non-prompt lepton. An influence of cuts on the fake factor is also observed; however, conclusions are limited by statistics. In order to improve the significance of the obtained results, a sample with higher Monte Carlo statistics could be used to extract the fake factors.
Investigations of the charge composition of the non-prompt leptons show that for $\mathtt{t\bar{t}}$, leptons originating in Photon Conversions have predominantly the opposite charge of the Tag (Prompt) lepton, which is not plausible. Since Photon Conversions produce Electron-Positron pairs, the charge of the Probe lepton should be independent on the Tag leptons charge. The reason for this unexpected result still needs further investigation.

Towards the aim of obtaining an estimation of the non-prompt lepton background in the signal region in data, the control-region used to extract the fake factors needs further adjustments.

For this, all processes contributing to the background have to be considered, the control-region needs to be changed by applying cuts, such that it is disjoint from the signal region, as well as that fake origins in the control-region and signal-region are in agreement. Alternatively, another way to obtain a reliable background estimation is to define multiple control-regions, each for a particular process contributing to the background, and to merge the corresponding fake factors while requiring the fake origins in the control-regions and the signal-region to accord. A Closure Test should be performed again, to probe the method using the final control-region as well as all processes. Corrections to the fake factor, such as a η correction or an isolation correction (using PTVarCone30), should be considered to enhance the (differential) closure further. Finally, it might be necessary to reduce the systematic uncertainties of the fake factor method, which could be done by changing control-region definitions.

A Appendix

A.1 Theoretical Background

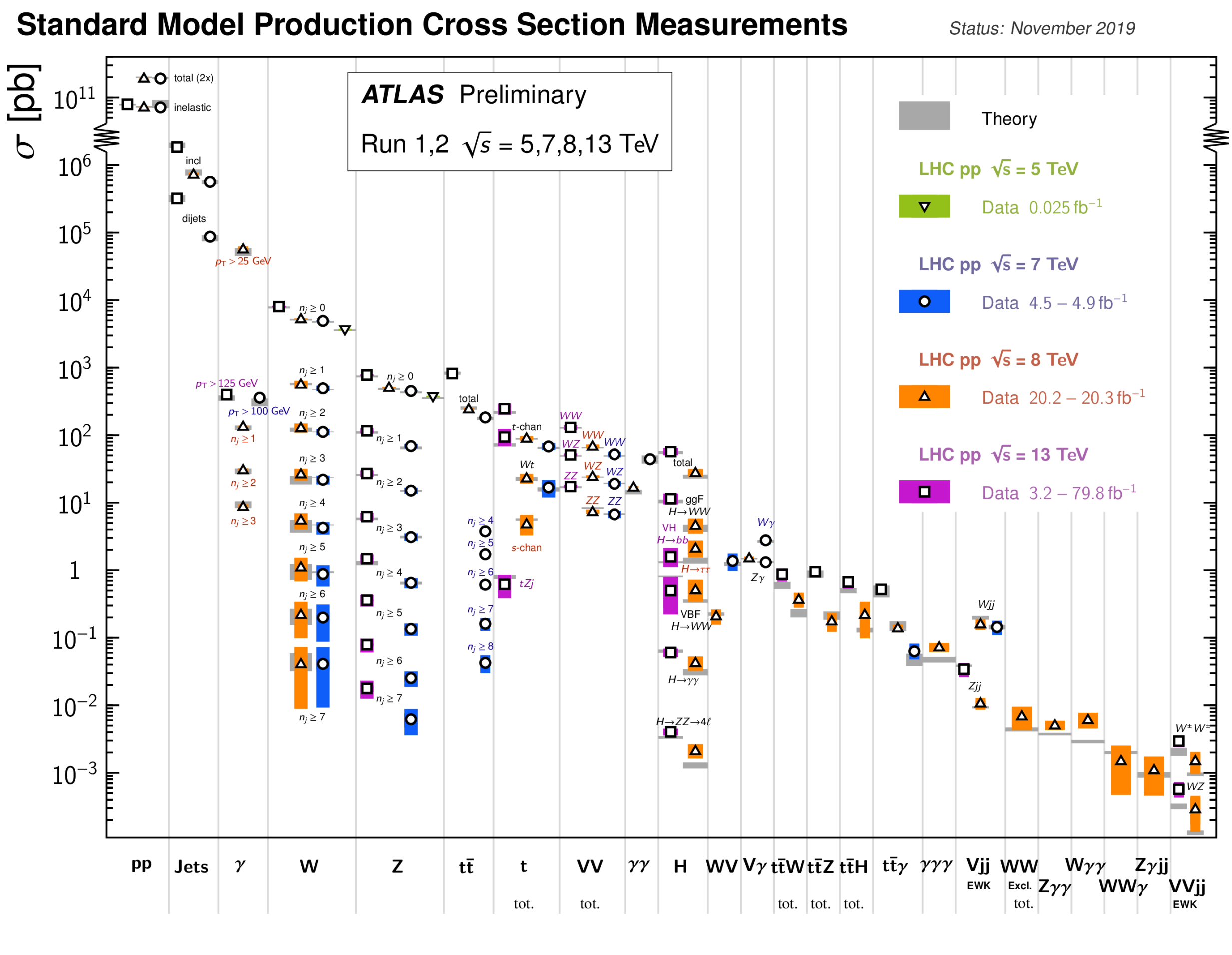


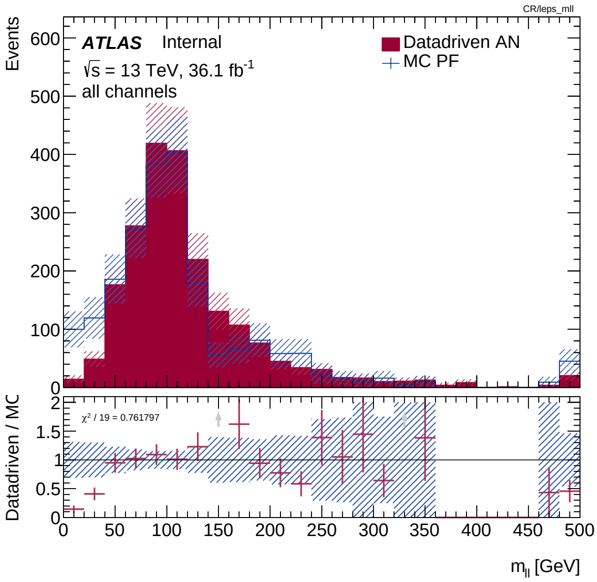
Figure A.1: "Standard Model Production Cross Section Measurements". Taken from [7]

A.2 Monte Carlo Samples

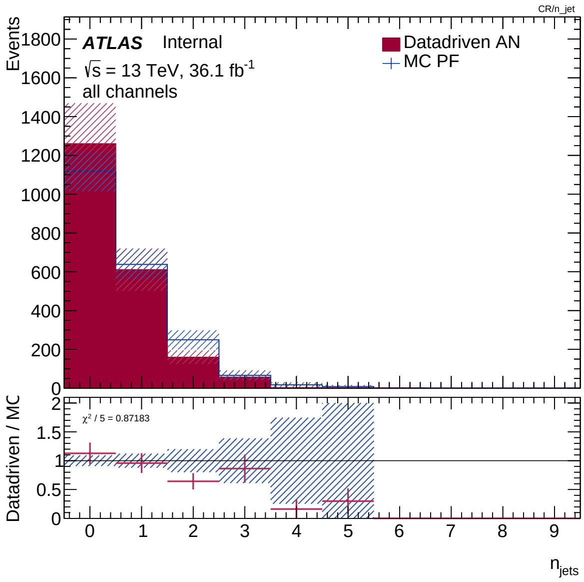
The following Monto-Carlo generated samples are used:

```
hist-mc16_13TeV.361100.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Wplusenu.deriv.DAOD_STDM3.e3601_s3126_r9364_p3916.root
hist-mc16_13TeV.361101.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Wplusmunu.deriv.DAOD_STDM3.e3601_s3126_r9364_p3916.root
hist-mc16_13TeV.361103.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Wminusenu.deriv.DAOD_STDM3.e3601_s3126_r9364_p3916.root
hist-mc16_13TeV.361104.PowhegPythia8EvtGen_AZNLOCTEQ6L1_Wminusmunu.deriv.DAOD_STDM3.e3601_s3126_r9364_p3916.root
hist-mc16_13TeV.410470.PhPy8EG_A14_ttbar_hdamp258p75_nonallhad.deriv.DAOD_STDM3.e6337_e5984_a875_r9364_r9315_p3931.root
```

A.3 Differential Monte Carlo Closure Test



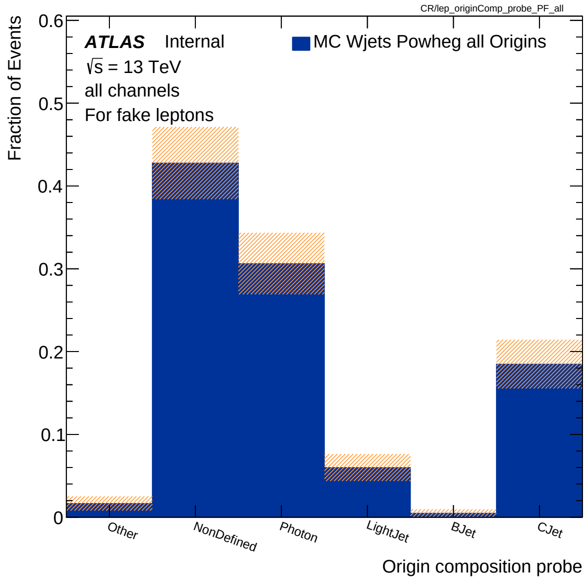
(a) m_{ll} distribution



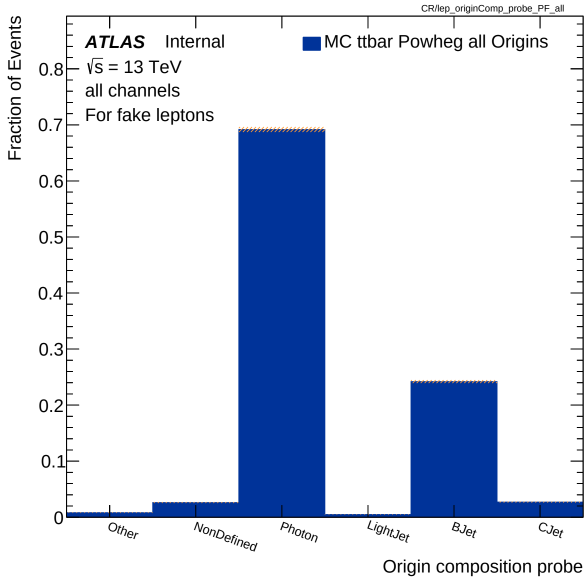
(b) n_{jets} distribution

Figure A.2: m_{ll} and n_{jets} distributions. The Monte Carlo prediction for Prompt, Fake (both Ana selected) is compared to the result of the data-driven fake factor method. Systematic uncertainties of the fake factor method are included.

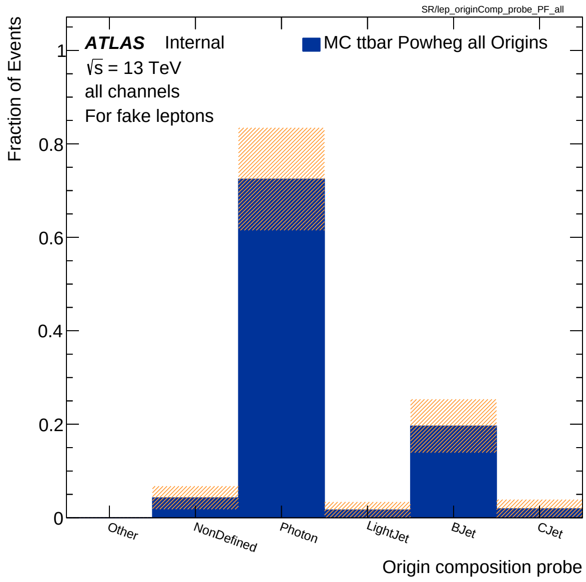
A.4 Fake Origins



(a) W+jets Powheg in the control-region

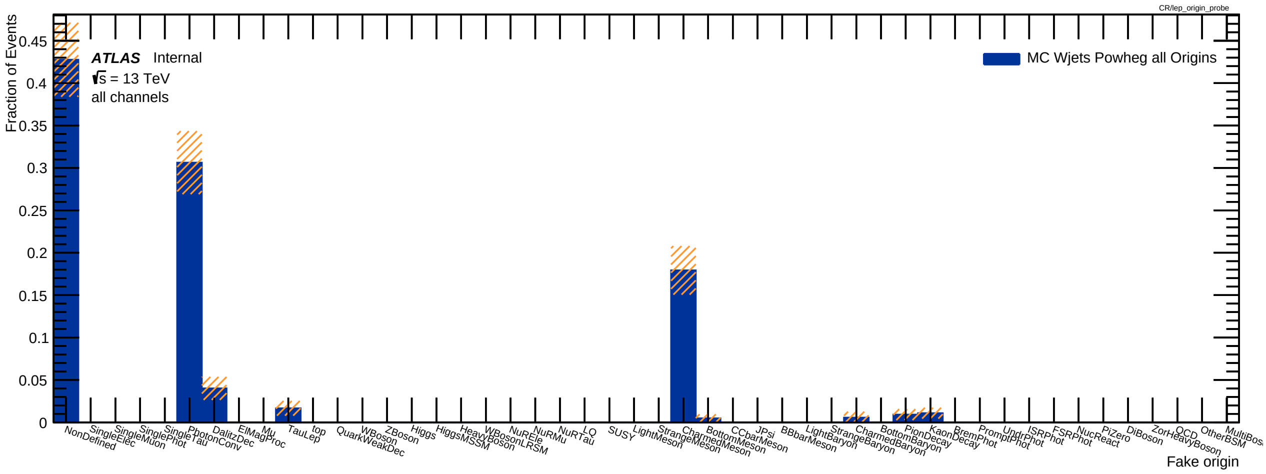


(b) $t\bar{t}$ Powheg in the control-region

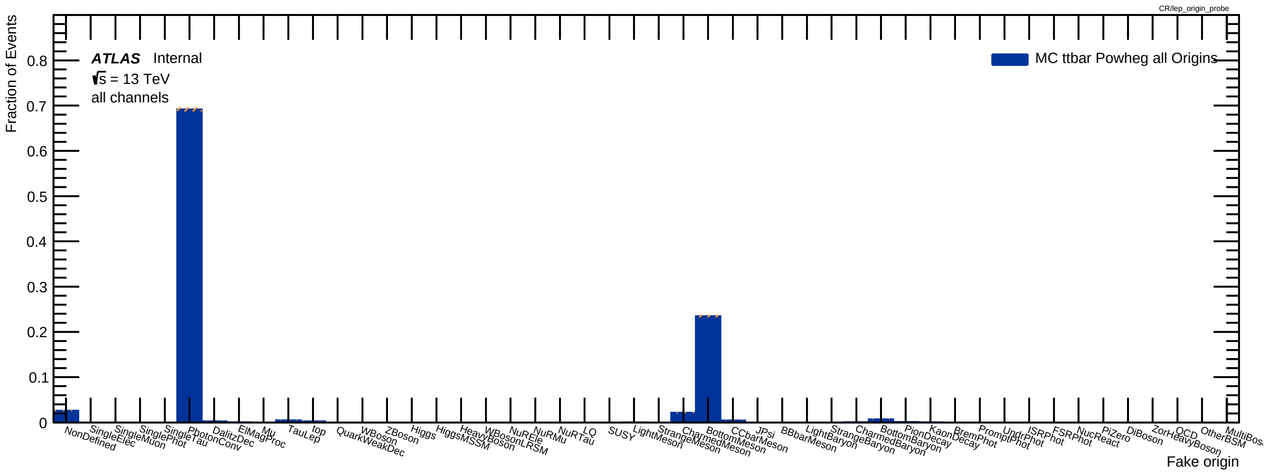


(c) $t\bar{t}$ Powheg in the VBS signal-region

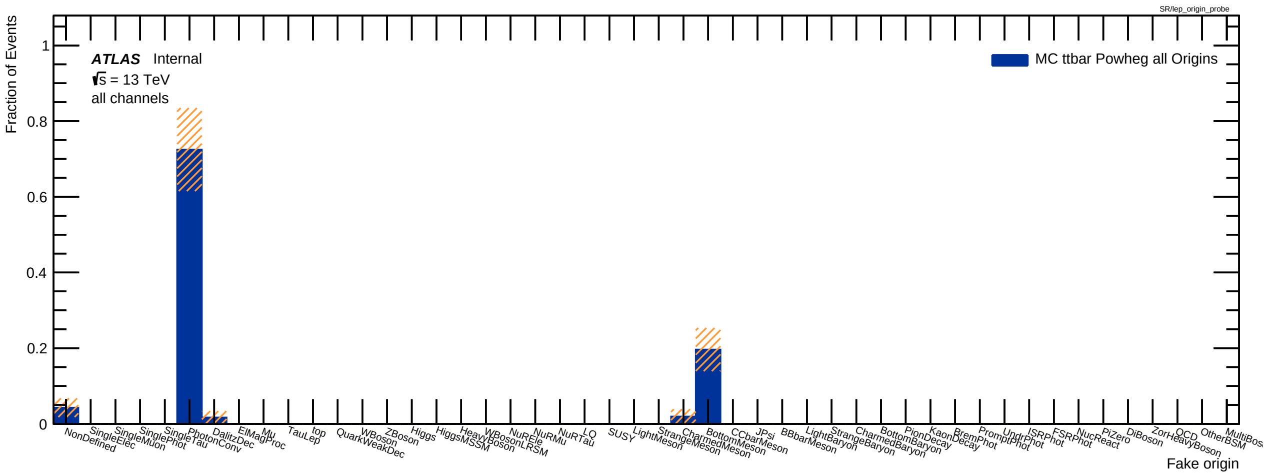
Figure A.3: Origins of the fake lepton in control- and VBS signal-region.



(a) W+jets Powheg in the control-region



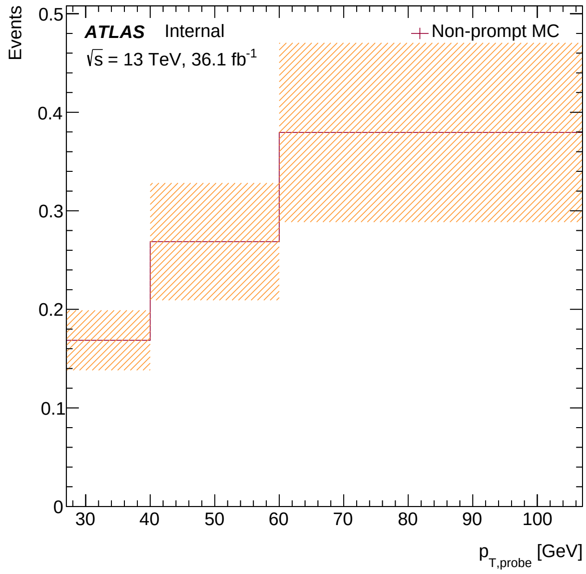
(b) $t\bar{t}$ Powheg in the control-region



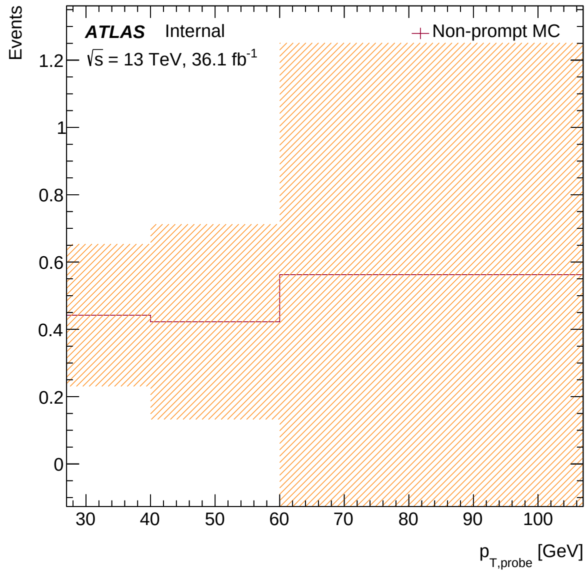
(c) $t\bar{t}$ Powheg in the VBS signal-region

Figure A.4: Detailed origins of the fake lepton in control- and VBS signal-region.

A.5 Influence of Fake Origins and Cuts on the Fake Factor

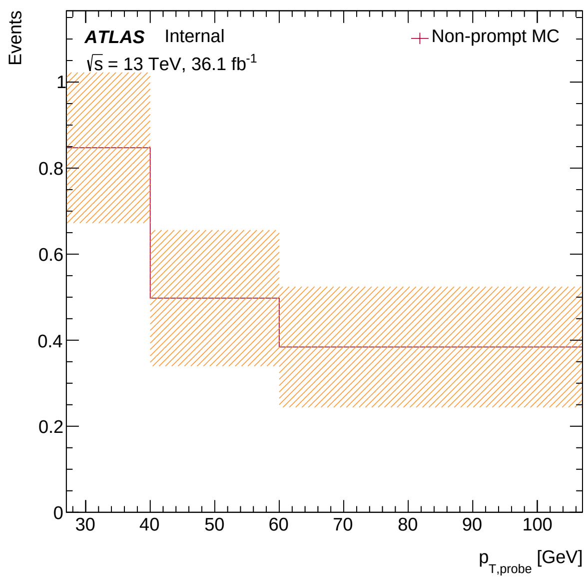


(a) Electron fakes

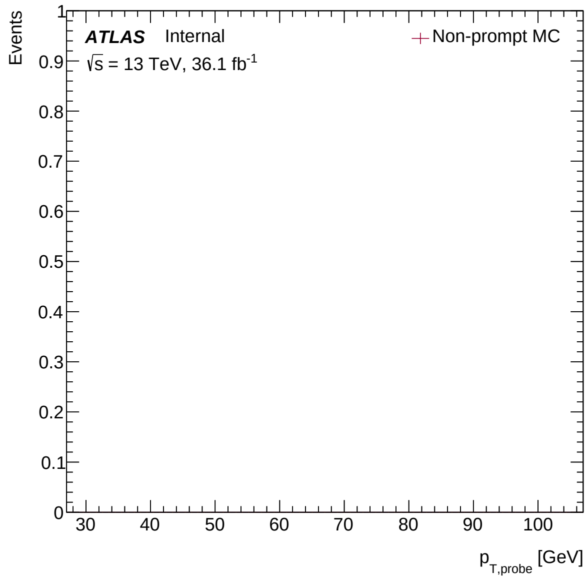


(b) Muon fakes

Figure A.5: Fake factors for leptons with Non-Defined origins.

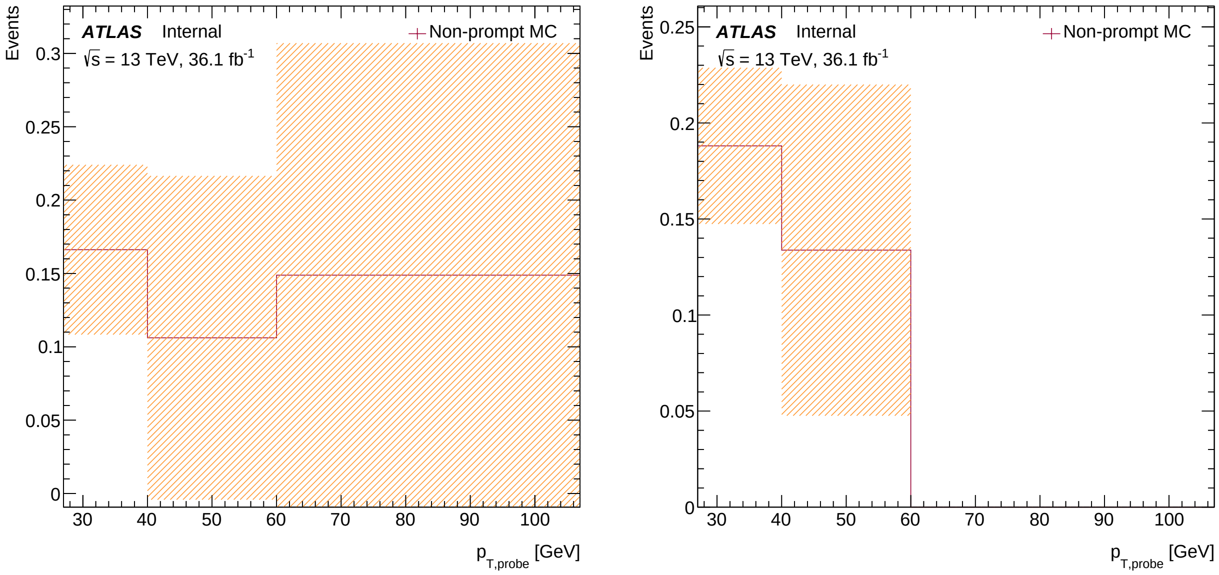


(a) Electron fakes



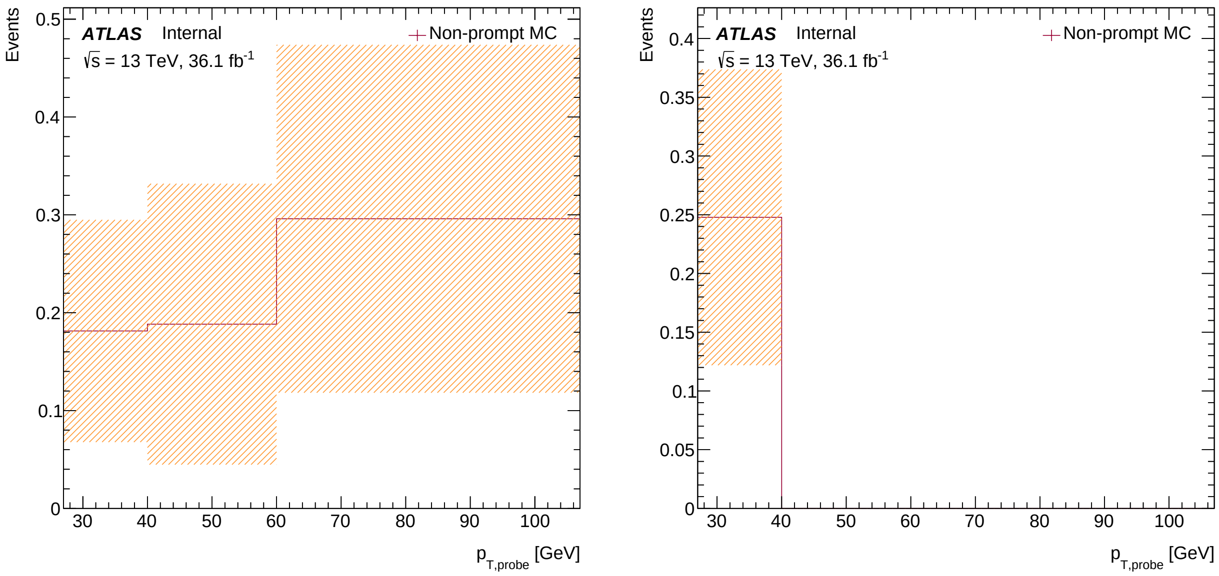
(b) Muon fakes

Figure A.6: Fake factors for leptons originating in Photon Conversions. Muons do not undergo Photon Conversion, therefore the histogram is empty for muon fakes.



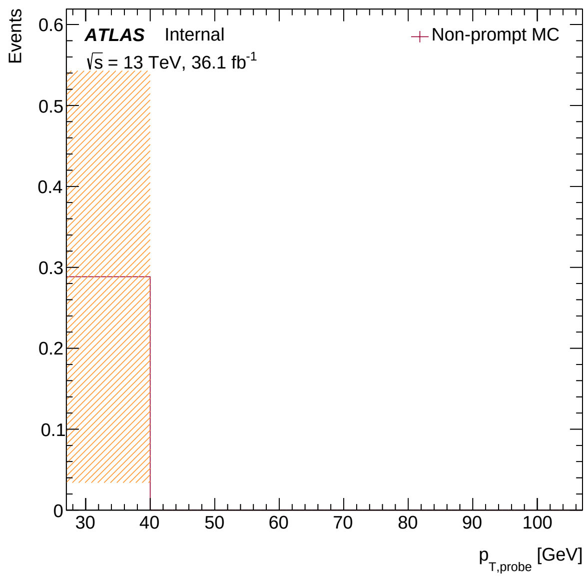
(a) Electron fakes **(b)** Muon fakes

Figure A.7: Fake factors for leptons originating in *c*-Jets.

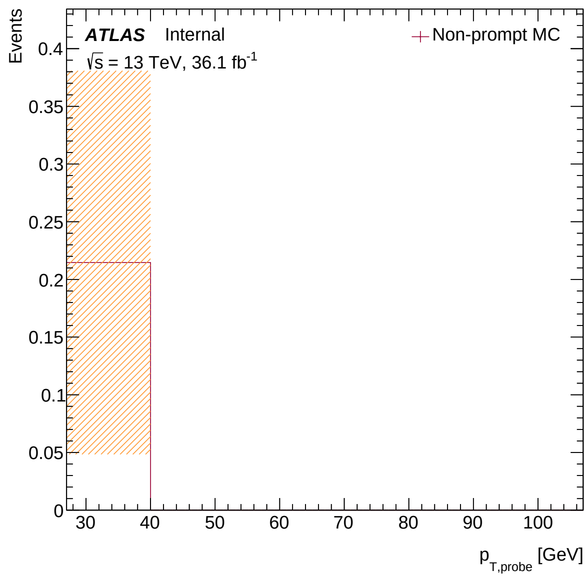


(a) Electron fakes **(b)** Muon fakes

Figure A.8: Fake factors for leptons originating in Light-Jets.

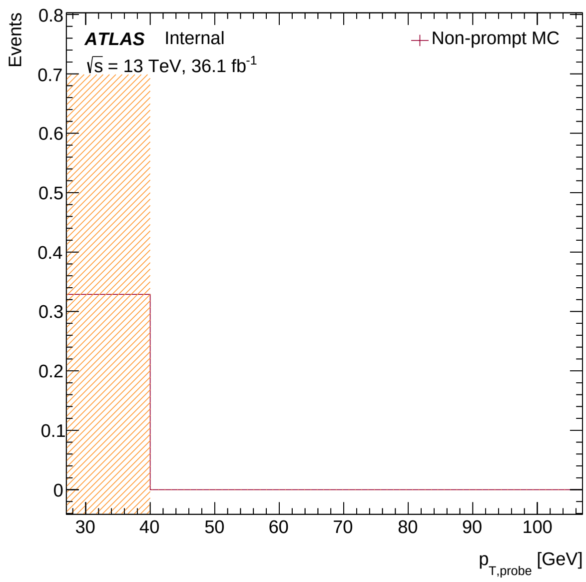


(a) Electron fakes

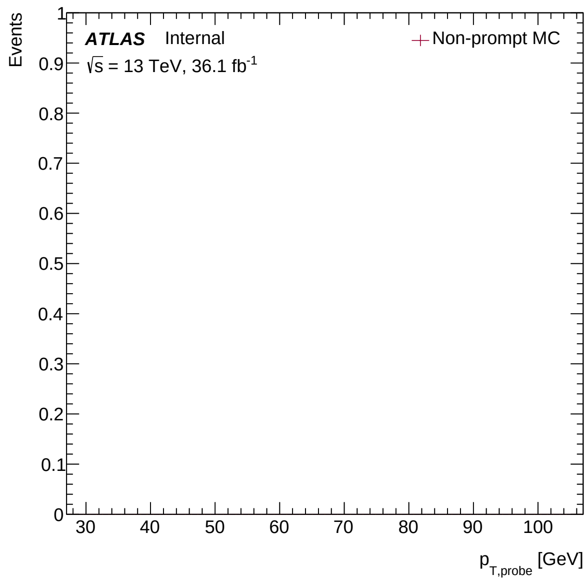


(b) Muon fakes

Figure A.9: Fake factors for leptons with other origins.



(a) Electron fakes

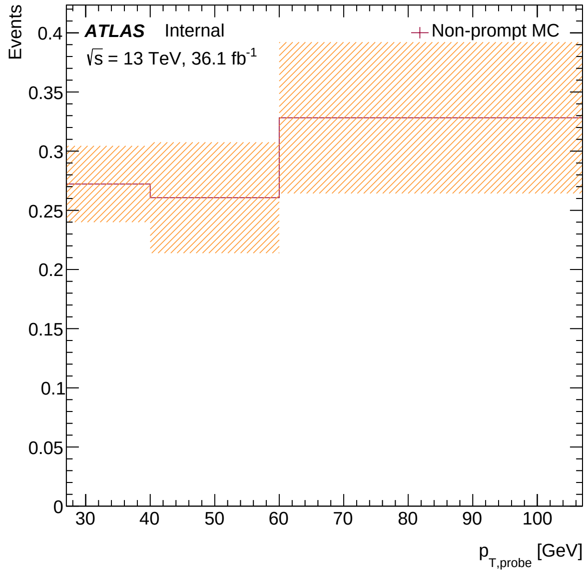


(b) Muon fakes

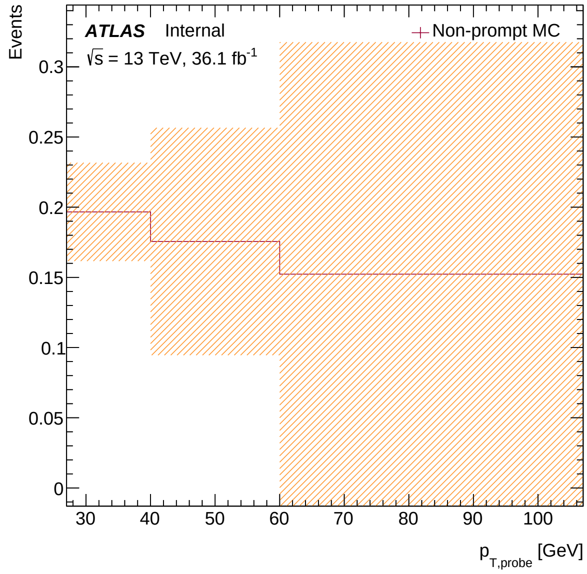
Figure A.10: Fake factors for leptons originating in b-Jets.

Table A.1: Cutflow for `W+jets Powheg` and `tt̄ Powheg` samples. Different cuts from the VBS signal-region, as well as their inverses, are displayed. *Cutname* Inverted means that this is a signal-region cut but inverted, as used to make control- and signal-region disjoint. All cuts are applied after the CR cut unless specified otherwise in the cut name.

	Wjets all	Wjets Fake e	Wjets Fake μ	ttbar all	ttbar Fake e	ttbar Fake μ
All	1.721	1.261	0.460	30.66	13.89	16.77
MC_weight	16414	12204	4210	22388	10130	12258
PU_weight	16451	12224	4227	22433	10151	12281
LepSF	15804	11754	4050	22433	10151	12281
Lep2SF	15668	11637	4031	22433	10151	12281
JetSF	15660	11623	4037	22288	1007	12211
MET	10486	7797	2690	18397	8459	9937
lep_isAna	10025	7356	2669	17820	7993	9827
CR	10025	7356	2669	17820	7993	9827
BJetVeto SR	9631	7042	2589	4292	1488	2805
BJetVeto Inverted	393.5	313.9	79.65	13528	6505	7022
TwoJets Inverted	8869	6425	2444	2614	1168	1446
SameSign Inverted	7579	5226	2353	10454	4983	5471
m_{jj} Low Inverted	9636	7016	2620	12200	5238	6962
m_{jj} High Inverted	9923	7273	2650	16589	7402	9187
m_{jj} Low SR	389.1	340.0	49.06	5620	2756	2865
m_{jj} High SR	101.9	82.99	18.91	1231	590.8	640.1
OppositeSign $\rightarrow m_{jj}$ Low SR	240.8	200.9	39.96	3187	1701	1486
OppositeSign $\rightarrow m_{jj}$ High SR	76.47	57.56	18.91	735.5	366.7	368.8
m_{ll} SR	9871	7209	2662	17530	7891	9638
m_{ll} Inverted	154.0	146.9	7.097	290.1	101.9	188.2

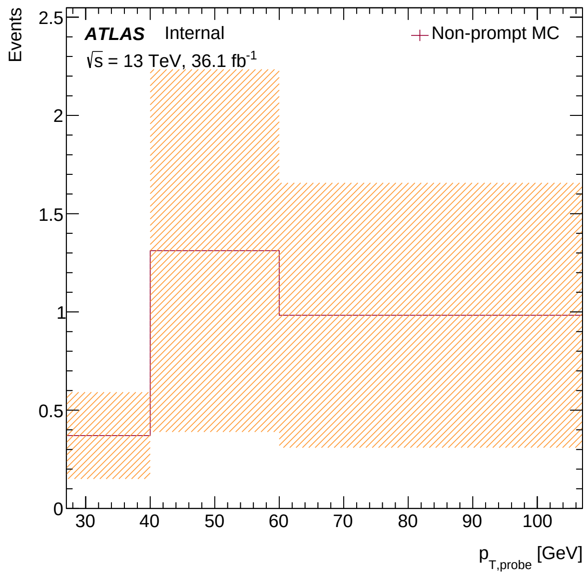


(a) Electron fakes

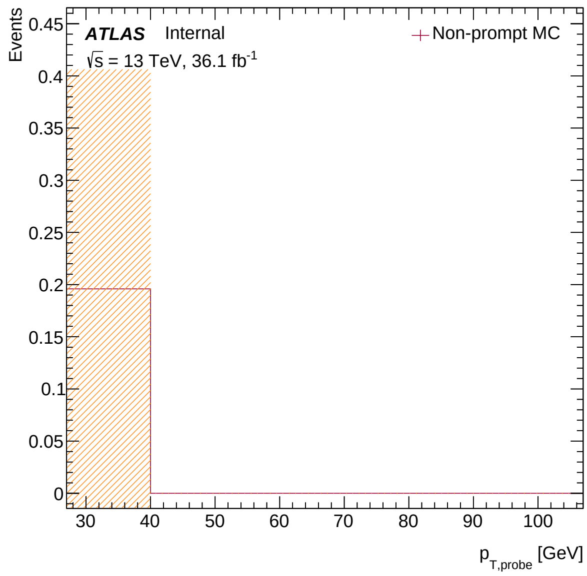


(b) Muon fakes

Figure A.11: Fake factors after applying the b-jet veto cut from the VBS signal-region.

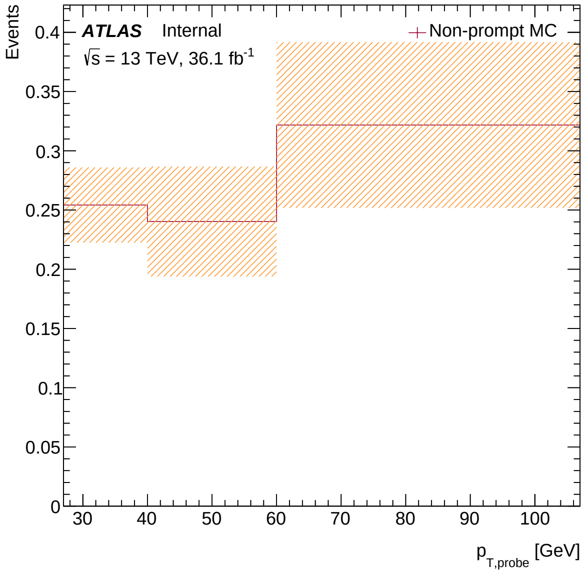


(a) Electron fakes

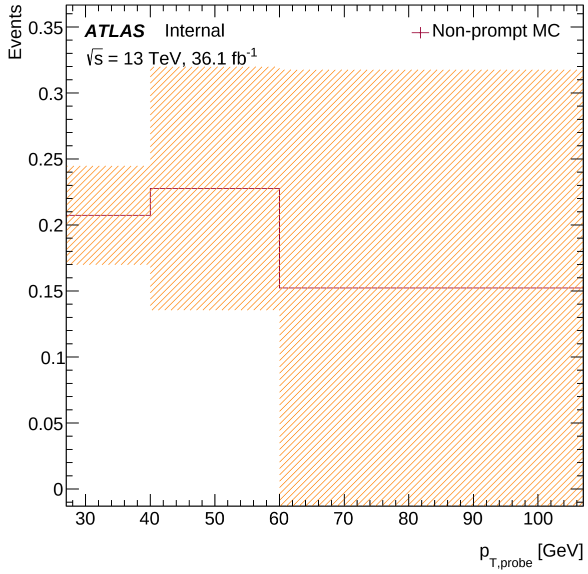


(b) Muon fakes

Figure A.12: Fake factors after requireing atleast one b-jet.

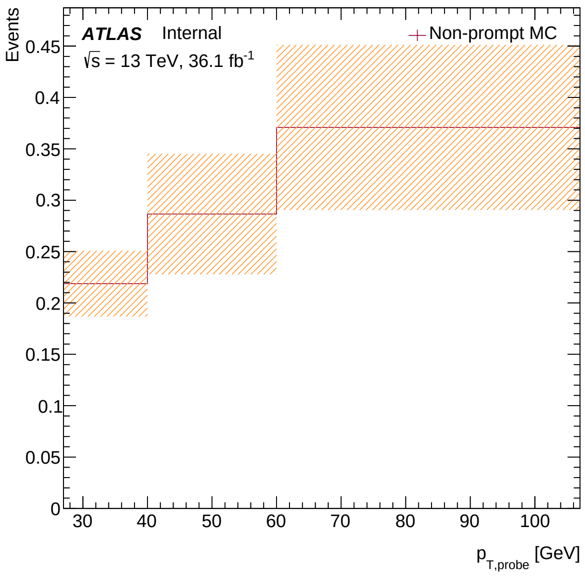


(a) Electron fakes

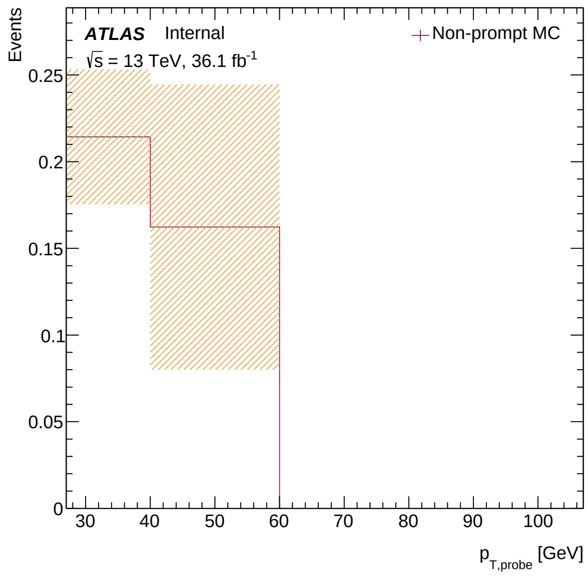


(b) Muon fakes

Figure A.13: Fake factors after applying a second jet veto cut.

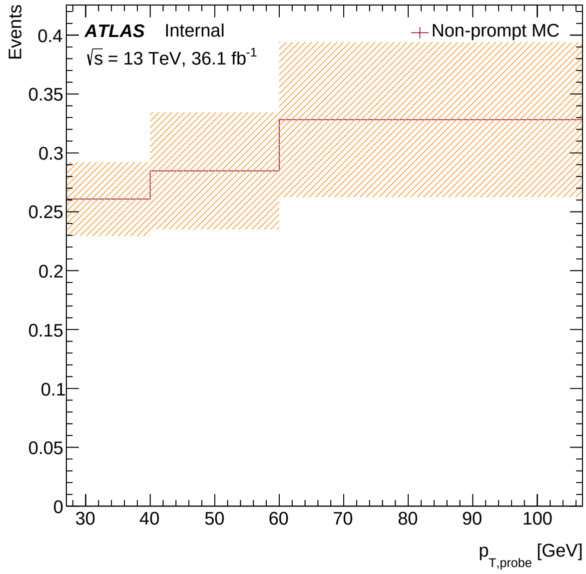


(a) Electron fakes

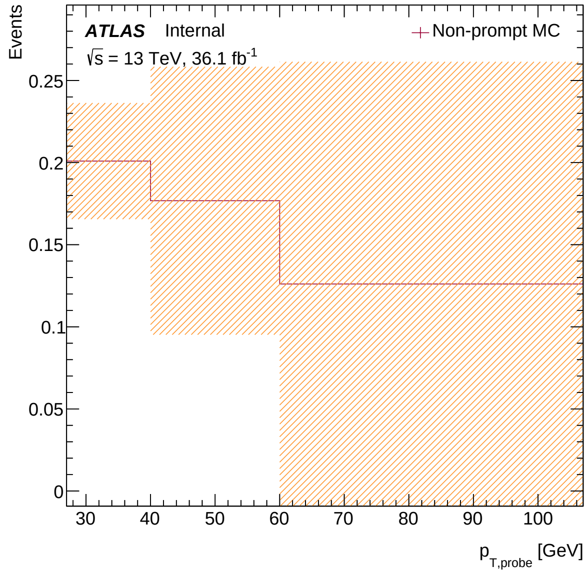


(b) Muon fakes

Figure A.14: Fake factors after applying a same sign veto cut.

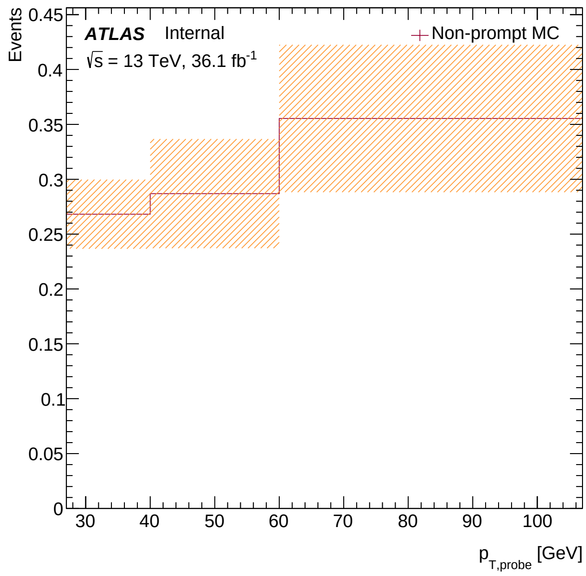


(a) Electron fakes

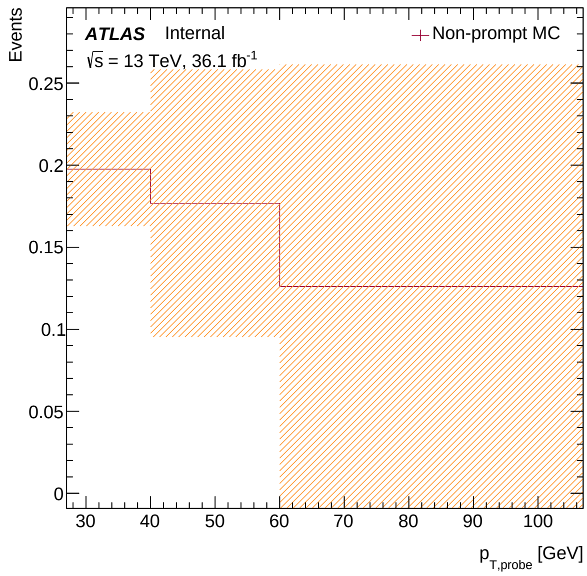


(b) Muon fakes

Figure A.15: Fake factors after applying a $m_{jj} \leq 200\,\text{GeV}$ cut.

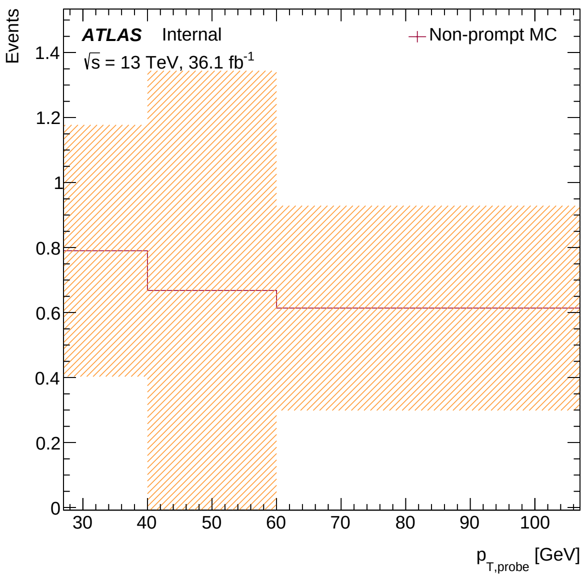


(a) Electron fakes

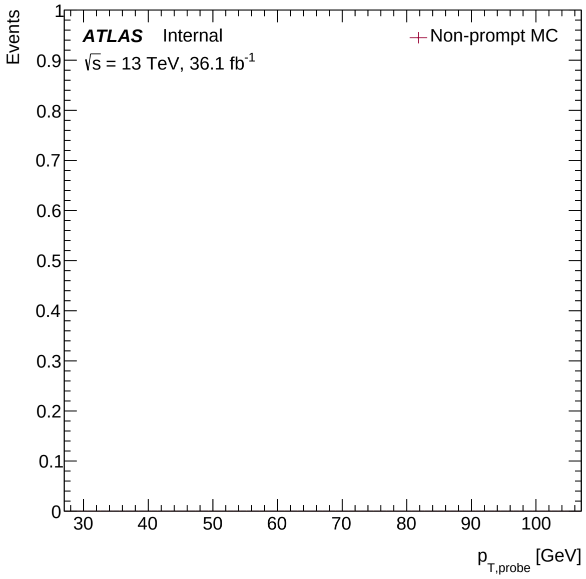


(b) Muon fakes

Figure A.16: Fake factors after applying a $m_{jj} \leq 500\,\text{GeV}$ cut.

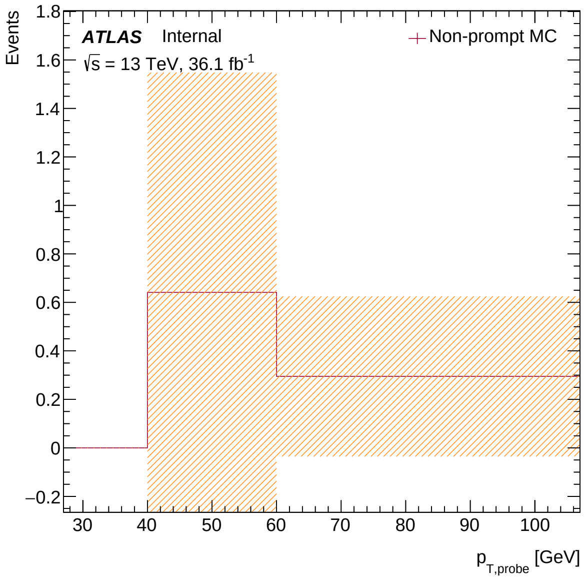


(a) Electron fakes

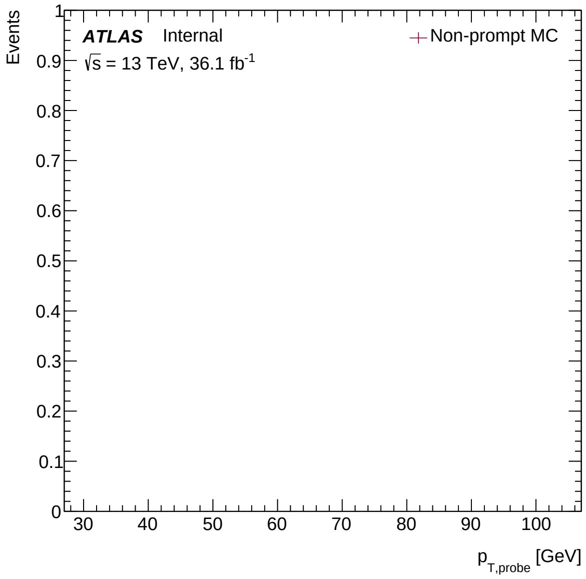


(b) Muon fakes

Figure A.17: Fake factors after applying the $m_{jj} > 200\,\text{GeV}$ VBS signal-region cut.

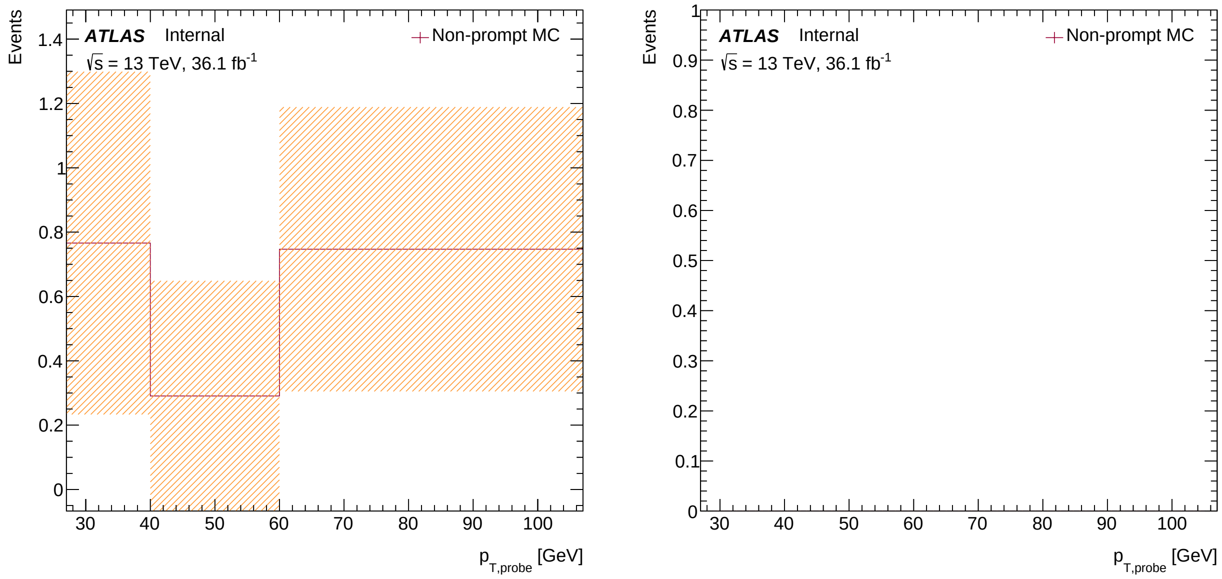


(a) Electron fakes



(b) Muon fakes

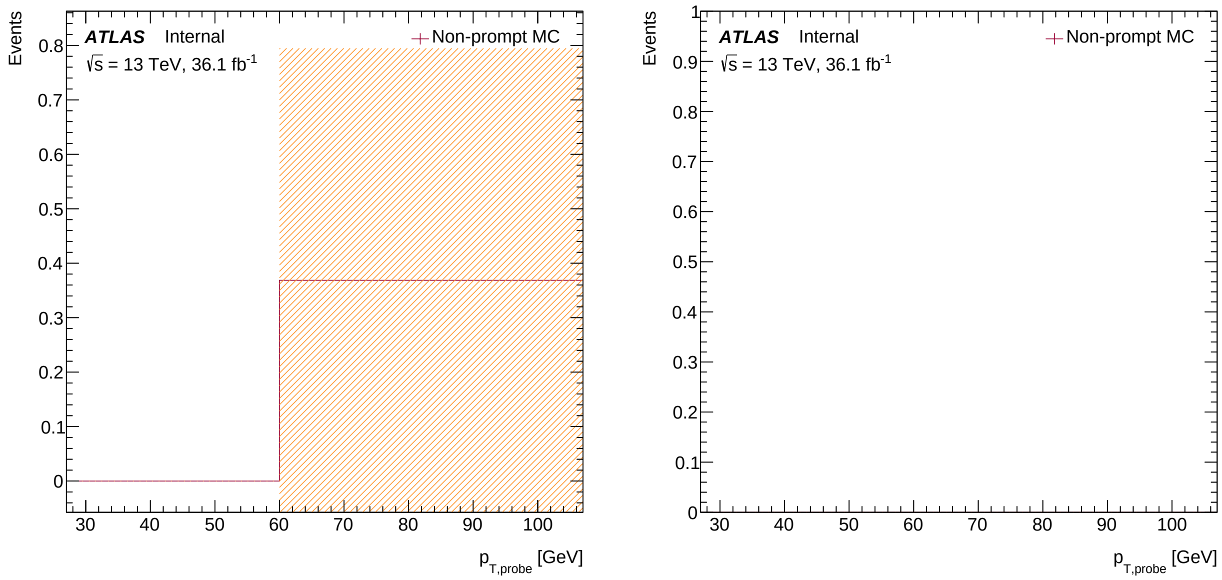
Figure A.18: Fake factors after applying the $m_{jj} > 500\,\text{GeV}$ VBS signal-region cut.



(a) Electron fakes

(b) Muon fakes

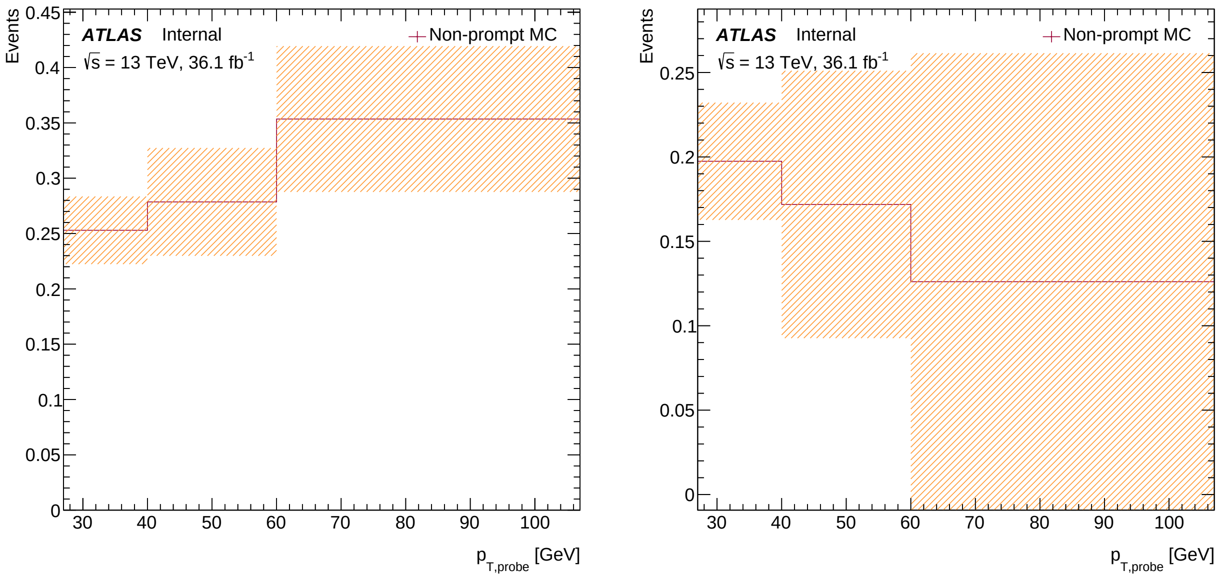
Figure A.19: Fake factors after applying an opposite sign cut and the $m_{jj} > 200\,\text{GeV}$ VBS signal-region cut.



(a) Electron fakes

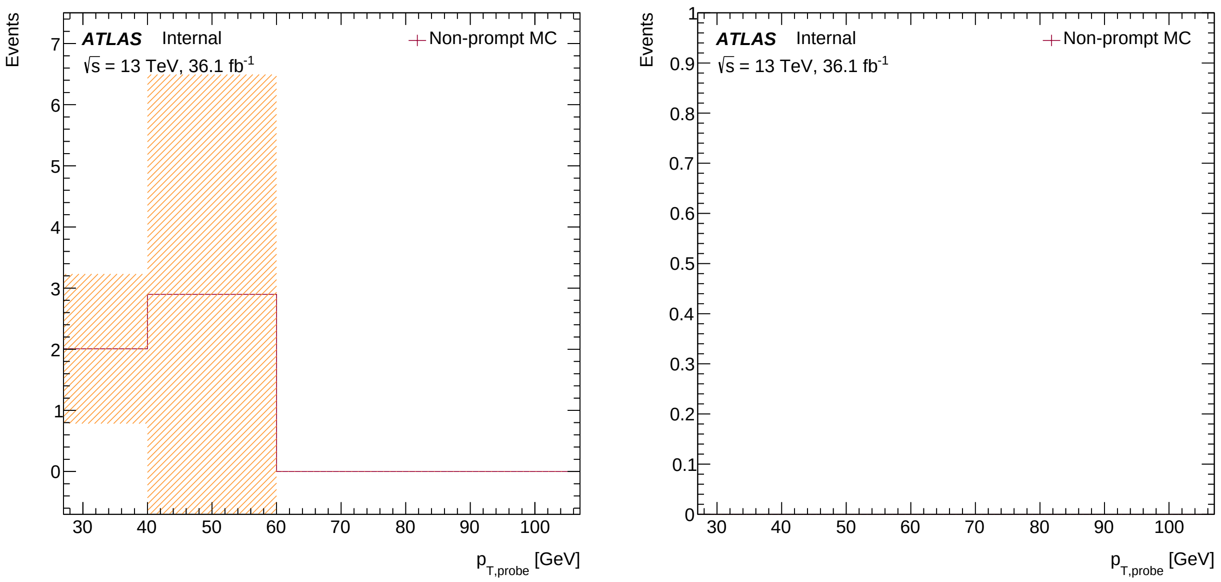
(b) Muon fakes

Figure A.20: Fake factors after applying an opposite sign cut and the $m_{jj} > 500\,\text{GeV}$ VBS signal-region cut.



(a) Electron fakes **(b)** Muon fakes

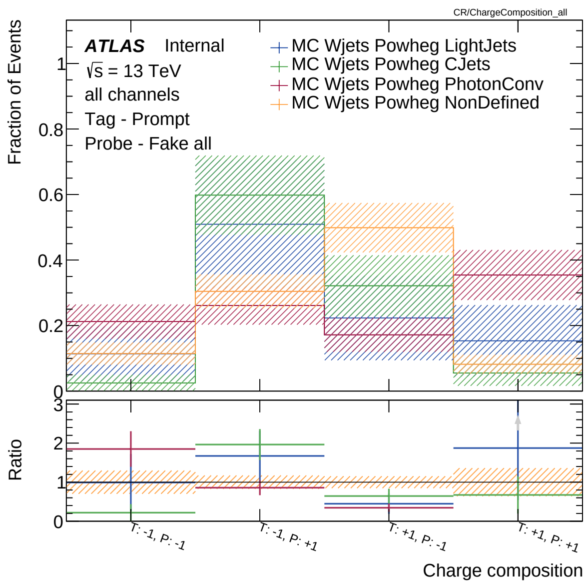
Figure A.21: Fake factors after applying the $m_{ll} > 20\,\mathrm{GeV}$ VBS signal-region cut.



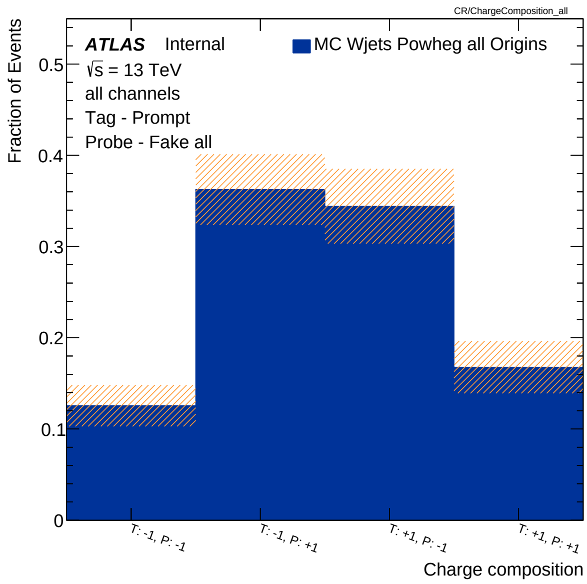
(a) Electron fakes **(b)** Muon fakes

Figure A.22: Fake factors after applying a $m_{ll} \leq 20\,\mathrm{GeV}$ cut.

A.6 Charge Composition of Fake Leptons

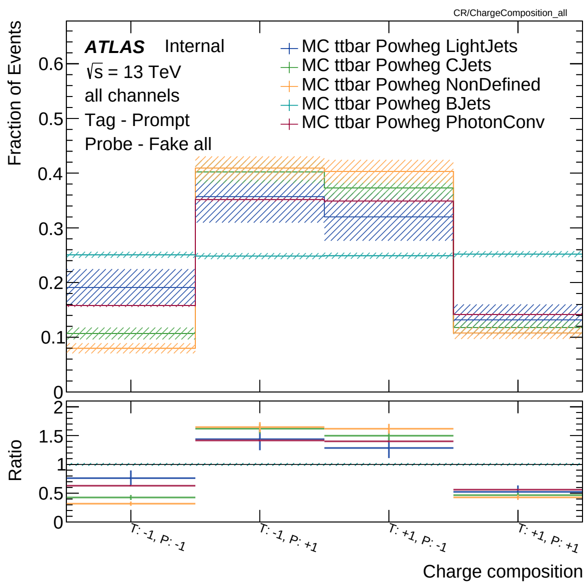


(a) Splitted fake origins, `W+jets Powheg`

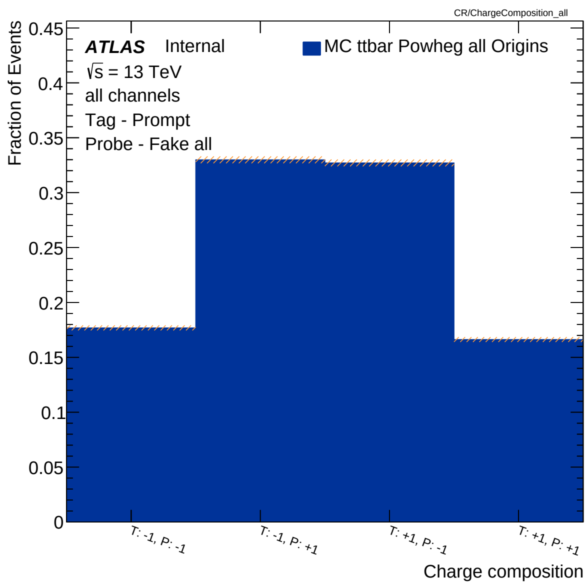


(b) All fake origins, `W+jets Powheg`

Figure A.23: Charge Composition of the Tag and Probe lepton for the `W+jets Powheg` sample in the control-region. Additionally, the ratio over the Non Defined distribution is displayed. Electron and Muon Fakes are not distinguished.

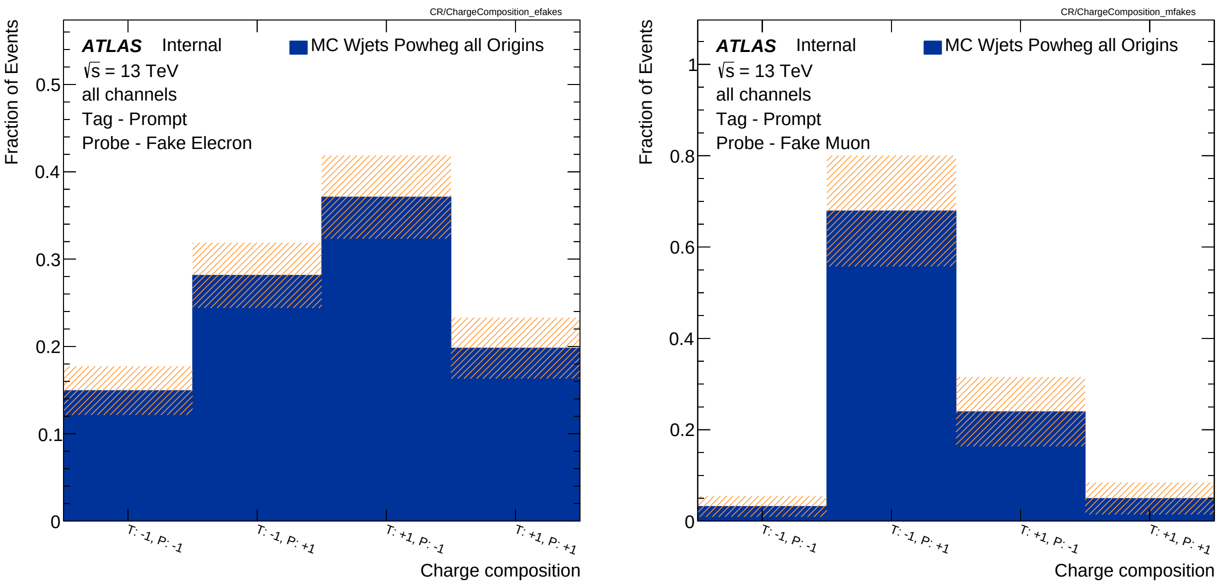


(a) Splitted fake origins, $\mathtt{t\bar{t}}$ `Powheg`



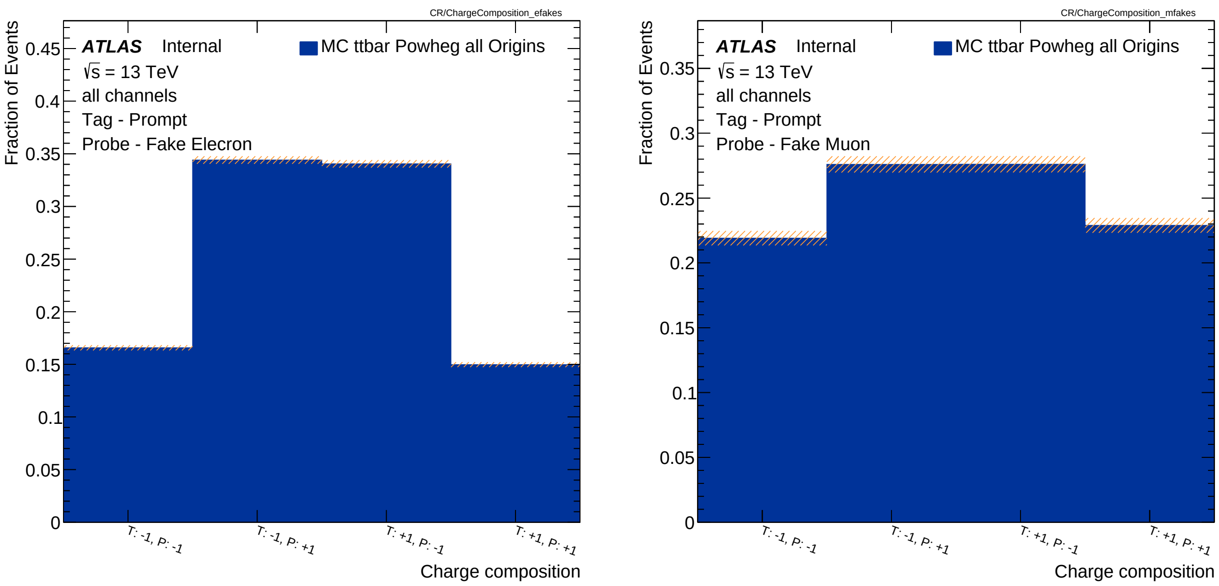
(b) All fake origins,$\mathtt{t\bar{t}}$ `Powheg`

Figure A.24: Charge Composition of the Tag and Probe lepton for the $\mathtt{t\bar{t}}$ `Powheg` sample in the control region. Additionally, the ratio over the b-jet distribution is displayed. Electron and Muon Fakes are not distinguished.



(a) Electron fakes (b) Muon fakes

Figure A.25: Charge Composition of the Tag and Probe lepton for the `W+jets Powheg` sample in the control-region. Different fake origins are not distinguished.



(a) Electron fakes (b) Muon fakes

Figure A.26: Charge Composition of the Tag and Probe lepton for the $t\bar{t}$ `Powheg` sample in the control-region. Different fake origins are not distinguished.

Table A.2: Table form of Figure 4.6 (a). The Charge Composition of the Tag and Probe lepton (fake electron) for the `W+jets Powheg` sample is displayed.

	T: -1, P: -1	T: -1, P: +1	T: +1, P: -1	T: +1, P: +1
MC Wjets Powheg NonDefined	89.2 ± 29 ± 0	216.9 ± 40 ± 0	434.5 ± 70 ± 0	63.6 ± 24 ± 0
MC Wjets Powheg PhotonConv	136.2 ± 34 ± 0	167.9 ± 40 ± 0	110.3 ± 33 ± 0	227.6 ± 50 ± 0
MC Wjets Powheg LightJets	14.3 ± 11 ± 0	41.4 ± 21 ± 0	9.0 ± 9 ± 0	19.3 ± 14 ± 0
MC Wjets Powheg CJets	9.7 ± 10 ± 0	39.3 ± 18 ± 0	54.0 ± 24 ± 0	10.1 ± 10 ± 0

Table A.3: Table form of Figure 4.6 (b). The Charge Composition of the Tag and Probe lepton (fake muon) for the `W+jets Powheg` sample is displayed.

	T: -1, P: -1	T: -1, P: +1	T: +1, P: -1	T: +1, P: +1
MC Wjets Powheg NonDefined	13.7 ± 10 ± 0	56.2 ± 22 ± 0	12.6 ± 13 ± 0	9.9 ± 10 ± 0
MC Wjets Powheg PhotonConv	0 ± 0 ± 0	0 ± 0 ± 0	0 ± 0 ± 0	0 ± 0 ± 0
MC Wjets Powheg LightJets	0 ± 0 ± 0	22.7 ± 13 ± 0	19.1 ± 14 ± 0	0 ± 0 ± 0
MC Wjets Powheg CJets	0 ± 0 ± 0	192.4 ± 40 ± 0	70.6 ± 27 ± 0	11.3 ± 11 ± 0

Table A.4: Table form of Figure A.10 (a). The Charge Composition of the Tag and Probe lepton (fake electrons and muons) for the `W+jets Powheg` sample is displayed.

	T: -1, P: -1	T: -1, P: +1	T: +1, P: -1	T: +1, P: +1
MC Wjets Powheg NonDefined	102.8 ± 30 ± 0	273.0 ± 50 ± 0	447.1 ± 70 ± 0	73.5 ± 26 ± 0
MC Wjets Powheg PhotonConv	136.2 ± 34 ± 0	167.9 ± 40 ± 0	110.3 ± 33 ± 0	227.6 ± 50 ± 0
MC Wjets Powheg LightJets	14.3 ± 11 ± 0	64.0 ± 25 ± 0	28.1 ± 16 ± 0	19.3 ± 14 ± 0
MC Wjets Powheg CJets	9.7 ± 10 ± 0	231.8 ± 50 ± 0	124.6 ± 40 ± 0	21.4 ± 15 ± 0

Table A.5: Table form of Figure 4.7 (a). The Charge Composition of the Tag and Probe lepton (fake electron) for the $t\bar{t}$ `Powheg` sample is displayed.

	T: -1, P: -1	T: -1, P: +1	T: +1, P: -1	T: +1, P: +1
MC ttbar Powheg NonDefined	5.0 ± 0.8 ± 0	25.7 ± 2 ± 0	21.9 ± 2 ± 0	5.7 ± 0.9 ± 0
MC ttbar Powheg PhotonConv	524.6 ± 9 ± 0	1167.4 ± 13 ± 0	1158.4 ± 13 ± 0	469.8 ± 8 ± 0
MC ttbar Powheg LightJets	4.1 ± 0.7 ± 0	4.1 ± 0.8 ± 0	3.4 ± 0.7 ± 0	2.8 ± 0.6 ± 0
MC ttbar Powheg BJets	89.9 ± 4 ± 0	88.8 ± 4 ± 0	92.6 ± 4 ± 0	86.8 ± 4 ± 0
MC ttbar Powheg CJets	3.8 ± 0.7 ± 0	11.5 ± 1 ± 0	7.1 ± 1 ± 0	3.1 ± 0.7 ± 0

Table A.6: Table form of Figure 4.7 (b). The Charge Composition of the Tag and Probe lepton (fake muon) for the `tt̄ Powheg` sample is displayed.

	T: -1, P: -1	T: -1, P: +1	T: +1, P: -1	T: +1, P: +1
MC ttbar Powheg NonDefined	5.0 ± 0.8 ± 0	25.7 ± 2 ± 0	28.8 ± 2 ± 0	7.9 ± 1 ± 0
MC ttbar Powheg PhotonConv	0 ± 0 ± 0	0 ± 0 ± 0	0 ± 0 ± 0	0 ± 0 ± 0
MC ttbar Powheg LightJets	0.4 ± 0.2 ± 0	4.2 ± 0.8 ± 0	4.0 ± 0.7 ± 0	0.3 ± 0.2 ± 0
MC ttbar Powheg BJets	201.3 ± 5 ± 0	199.7 ± 5 ± 0	196.7 ± 5 ± 0	205.8 ± 5 ± 0
MC ttbar Powheg CJets	10.1 ± 1 ± 0	40.7 ± 2 ± 0	41.4 ± 2 ± 0	12.2 ± 1 ± 0

Table A.7: Table form of Figure A.11 (a). The Charge Composition of the Tag and Probe lepton (fake electrons and muons) for the `tt̄ Powheg` sample is displayed.

	T: -1, P: -1	T: -1, P: +1	T: +1, P: -1	T: +1, P: +1
MC ttbar Powheg NonDefined	10.0 ± 1 ± 0	51.4 ± 3 ± 0	50.7 ± 3 ± 0	13.5 ± 1 ± 0
MC ttbar Powheg PhotonConv	524.6 ± 9 ± 0	1167.4 ± 13 ± 0	1158.4 ± 13 ± 0	469.8 ± 8 ± 0
MC ttbar Powheg LightJets	4.5 ± 0.8 ± 0	8.3 ± 1 ± 0	7.5 ± 1 ± 0	3.1 ± 0.7 ± 0
MC ttbar Powheg BJets	291.3 ± 6 ± 0	288.4 ± 6 ± 0	289.3 ± 6 ± 0	292.6 ± 6 ± 0
MC ttbar Powheg CJets	13.9 ± 1 ± 0	52.3 ± 3 ± 0	48.5 ± 3 ± 0	15.3 ± 2 ± 0

Bibliography

[1] Morad 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". In: *Phys. Rev. Lett.* 123.arXiv:1906.03203 (2019), 161801. 21 p. DOI: 10.1103/PhysRevLett.123.161801. URL: https://cds.cern.ch/record/2677868.

[2] G. Aad et al. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *JINST* 3 (2008), S08003. DOI: 10.1088/1748-0221/3/08/S08003.

[3] Giles Barr et al. *Particle physics in the LHC era*. Oxford master series in particle physics, astrophysics and cosmology. Oxford: Oxford University Press, 2016. DOI: 10.1093/acprof:oso/9780198748557.001.0001. URL: https://cds.cern.ch/record/2034442.

[4] 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. Geneva: CERN, 2018. URL: https://cds.cern.ch/record/2300058.

[5] 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. Geneva: CERN, 2018. URL: https://cds.cern.ch/record/2309552.

[6] CERN. *The Large Hardron Collider*. 2019. URL: https://home.cern/science/accelerators/large-hadron-collider (visited on 12/27/2019).

[7] The ATLAS collaboration. *Standard Model Summary Plots*. 2019. URL: https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PUBNOTES/ATL-PHYS-PUB-2019-024/fig_03.png (visited on 01/09/2020).

[8] Lyndon Evans and Philip Bryant. "LHC Machine". In: *Journal of Instrumentation* 3.08 (2008), S08001–S08001. DOI: 10.1088/1748-0221/3/08/s08001. URL: https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08001.

[9] Particle Data Group. *W^{+} Decay Modes*. 2019. URL: http://pdg.lbl.gov/2019/listings/rpp2019-list-w-boson.pdf (visited on 12/19/2019).

[10] Daniel Hay Guest. *MCTruthClassifier*. 2019. URL: https://twiki.cern.ch/twiki/bin/viewauth/AtlasProtected/MCTruthClassifier (visited on 01/10/2020).

[11] Franziska Iltzsche. “Observation of Electroweak Same-Charge $W^{\pm}W^{\pm}jj$ Production at $\sqrt{s} = 13$ TeV with the ATLAS Detector at the LHC”. 2019. URL: https://cds.cern.ch/record/2697175.

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

[13] A. M. Sirunyan et al. “Observation of Electroweak Production of Same-Sign W Boson Pairs in the Two Jet and Two Same-Sign Lepton Final State in Proton-Proton Collisions at $\sqrt{s} = 13$ TeV”. In: *Phys. Rev. Lett.* 120 (8 2018), p. 081801. DOI: 10.1103/PhysRevLett.120.081801. URL: https://link.aps.org/doi/10.1103/PhysRevLett.120.081801.

[14] Mark Thomson. *Modern particle physics.* Cambridge: Cambridge University Press, 2013. URL: https://cds.cern.ch/record/1529540.

Erklärung

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.

Moritz Poguntke
Dresden, Januar 2020