Sensitivity and optimisation studies for testing CP invariance in Higgs boson production via vector boson fusion exploiting the $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ decay mode at $\sqrt{13}$ TeV with the ATLAS detector

Albert-Ludwigs-Universität Freiburg

Master thesis

for the Degree of Master of Science (M.Sc.) in Physics

submitted by
Helena Moyano Gómez

Superviser: Prof. Dr. Markus Schumacher

Faculty of Mathematics and Physics

November 2022

Albert-Ludwigs-Universität Freiburg

Erklärung

Ich versichere hiermit, dass die vorliegende Masterarbeit selbstständig verfasst und keine weiteren als die angegebenen Hilfsmittel benutzt sowie die Stellen der Arbeit, die in anderen Werken dem Wortlaut oder dem Sinn nach entnommen sind, durch Angaben der Quellen sichtbar gemacht wurden. Außerdem versichere ich, dass die vorgelegte Masterarbeit nicht Bestandteil eines anderen Prüfungsverfahrens ist und war.

____________________________ ____________________________

(Ort, Datum) (Unterschrift)

Abstract

This thesis analyses the possible CP-violating interactions of the Higgs boson with weak gauge bosons in the vector-boson-fusion (VBF) production mode with its subsequent decay into two τ-leptons where one decays hadronically and one leptonically. The full proton-proton collision data set recorded with the ATLAS detector at the LHC between 2015 and 2018, at a center-of-mass energy of $\sqrt{s} = 13$ TeV corresponding to an integrated luminosity of 139 fb^{-1} is used. Events with jets misidentified as hadronically decaying τ-leptons represent a dominant background contribution. A data-driven method is implemented and validated to estimate this background contribution. Furthermore, a neural network is developed and applied, which improves the signal-to-background ratio compared to a cut-based approach. The CP-violating interactions are described by an effective field theory and their magnitude can be parametrized by a single parameter $\tilde{d}$. Three CP-sensitive observables are analysed using either only the mean distribution of the CP-odd observable or performing a Maximum-Likelihood fit to the full distribution: the Optimal Observable ($\mathcal{OO}$), the signed difference between the azimuthal angle of the two leading jets ($\Delta\phi_{jj}$) and an Observable constructed using machine learning methods ($\mathcal{OO}_{\text{Reg}}$). The $\tilde{d}$ is expected to be constrained to the interval $[-0.00625, 0.00625]$, $[-0.00893, 0.00893]$ and $[-0.00627, 0.00627]$ at 68% confidence level (CL) using the Gauge curve fit for the $\mathcal{OO}$, $\Delta\phi_{jj}$ and $\mathcal{OO}_{\text{Reg}}$, respectively. The Maximum Likelihood fit constraints the expected $\tilde{d}$ at a 68% CL to $[-0.0168, 0.0168]$, $[-0.0223, 0.0223]$ and $[-0.0168, 0.0168]$ for the $\mathcal{OO}$, $\Delta\phi_{jj}$ and $\mathcal{OO}_{\text{Reg}}$, respectively.

Zusammenfassung

In dieser Arbeit werden die CP-verletzenden Wechselwirkungen des Higgs-Bosons in Verbindung mit schwachen Eichbosonen im Produktionsmodus der Vektor-Bosonen-Fusion (VBF) mit seinem anschließenden Zerfall in zwei τ-Leptonen analysiert, von denen eines hadronisch und eines leptonisch zerfällt. Diese Analyse verwendet den gesamten Datensatz, der mit dem ATLAS-Detektor am LHC zwischen 2015 und 2018 bei einer Schwerpunktsenergie von $\sqrt{s} = 13$ TeV aufgenommen wurde, was einer integrierten Luminosität von 139 fb^{-1} entspricht. Jets, die fälschlicherweise als hadronisch zerfallende τ-Leptonen identifiziert wurden, stellen einen dominanten Untergrundbeitrag dar. Es wird eine datenbasierte Methode entwickelt und validiert, um diesen Untergrundbeitrag abzuschätzen. Darüber hinaus wird ein neuronales Netzwerk eingesetzt, das das Signal-zu-Untergrund-Verhältnis im Vergleich zu einem schnittbasierten Ansatz verbessert. Die CP-verletzenden Wechselwirkungen werden durch eine effektive Feldtheorie beschrieben und ihre Größe kann durch einen einzigen Parameter $\tilde{d}$ beschrieben werden. Drei CP-sensitive Observablen werden entweder mit Hilfe der Verteilung der Mittelwerte der CP-ungeraden Observablen oder einer Maximum Likelihood-Anpassung analysiert: die Optimale Observable ($\mathcal{OO}$), die vorzeichenbehaftete Differenz zwischen dem azimutalen Winkel der beiden führenden Jets ($\Delta\phi_{jj}$) und eine mit Methoden des maschinellen Lernens konstruierte Optimale Observable ($\mathcal{OO}_{\text{Reg}}$). Das erwartete $\tilde{d}$ wird mit einem Konfidenzniveau (CL) von 68 % auf das Intervall $[-0.00625, 0.00625]$, $[-0.00893, 0.00893]$ und $[-0.00627, 0.00627]$ beschränkt, indem eine Anpassung an die Eichkurve für $\mathcal{OO}$, $\Delta\phi_{jj}$ bzw. $\mathcal{OO}_{\text{Reg}}$ verwendet wird. Die Maximum Likelihood-Anpassung beschränkt das erwartete $\tilde{d}$ bei einem 68%-igen CI auf $[-0.0168, 0.0168]$, $[-0.0223, 0.0223]$ und $[-0.0168, 0.0168]$ für die $\mathcal{OO}$, $\Delta\phi_{jj}$ und $\mathcal{OO}_{\text{Reg}}$.

Contents

Chapter 1

Introduction

Elementary particle physics describes the fundamental constituents of matter, their interactions and properties. The theory that addresses this task is the Standard Model (SM) of particle physics. The SM accounts for three out of the four fundamental forces, excluding gravity: electromagnetic, weak and strong, derived from a single principle; gauge invariance. The SM has predicted several particles that have been subsequently discovered experimentally, including the massive gauge bosons Z^0 and $W^{\pm}$ [1, 2] for example. However, the theory was incomplete until, in 2012, the Higgs boson was found by the ATLAS and CMS experiments at the Large Hadron Collider (LHC) [3, 4]. Nevertheless, the SM is not able to explain some observed phenomena such as gravity, dark matter, the baryon asymmetry in the universe and neutrino masses. Therefore physics beyond the SM (BSM) is required.

The SM predicts an equal amount of baryonic and antibaryonic matter produced at the early universe [5, 6], in contrast to recent measurements of the cosmic microwave background (CMB) that found a much larger contribution from baryonic matter [7]. This is the so-called *baryon asymmetry* problem, target of multiple modern BSM physics research. This problem is addressed by introducing a process called *baryogenesis* which is in charge of explaining the different production rates of baryonic and antibaryonic matter since the beggining of the universe. This process is required to fulfill three conditions called *Sakharov conditions* introduced by Sakharov in Ref. [8]: baryon number (B) violation, violation of charge conjugation (C) and the product of charge and parity (CP), and deviations from thermal equilibrium.

CP violation is described in the SM through the Cabbibo-Kobayashi-Maskawa (CKM) matrix. However, estimations on the contribution from this source to the baryon asymmetry of the universe (BAU) are found to be insufficient to explain the observed baryon asymmetry [9]. Consequently, new sources of CP violation are being currently investigated. The discovery of the Higgs boson opened up new possibilities for the study of CP violation. Whithin the SM, the Higgs boson interactions preserve C and P symmetries at tree level. However, studies on the possibility of CP violation have been performed in the coupling of the Higgs boson to two weak gauge bosons, i.e. the HVV vertex, that can be studied in the vector boson fusion (VBF) and Higgs strahlung (VH), Higgs boson production processes.

Tests of CP violation have been previously done in the VBF production mode of the Higgs boson decaying into two τ-leptons and two vector-bosons ($H \to VV$) with the ATLAS detector in 2016 using 20.3fb^{-1} of data at a center-of-mass energy ($\sqrt{s}$) of 8 TeV [10] and in 2020 using 36.1fb^{-1} of data at $\sqrt{s} = 13$ TeV [11] showing no sign of CP violation in neither of the two analysis.

Both mentioned analysis used the *Optimal Observable* method, where a CP-odd observable is constructed from an effective field theory (EFT) in which the strenght of CP violation is parametrized by a parameter $\tilde{d}$, constrained to the interval $(-0.11, 0.05)$ [10] and $(-0.090, 0.035)$ [11] at a 68% confidence interval (CI), both consistent with the SM prediction of $\tilde{d} = 0$. Further

CP-odd observables have been suggested, such as the signed difference between the azimuthal angles of the two leading jets [12, 13] and CP-sensitive observables based on machine learning (ML) algorithms [14, 15].

This thesis tests CP invariance in the VBF production process of the Higgs boson exploiting the decay into two τ-leptons, specifically in the semileptonic decay channel $H \rightarrow \tau_{\text{lep}}\tau_{\text{had}}$, where one τ-lepton decays leptonically and one hadronically using data collected from 2015 to 2018 (full Run-2), reaching 139 fb^{-1} of integrated luminosity at $\sqrt{s} = 13$ TeV collected with the ATLAS detector.

The structure of this thesis is the following: Chapter 2 introduces the theoretical framework used, the SM is presented together with a description of the elementary particles, the fundamental forces and the Higgs mechanism, also the production processes, decay modes, properties of the Higgs boson at the LHC are described, finishing with a discussion of the CP symmetry violation (CPV). In Chapter 3, the LHC is presented, focusing on the ATLAS experiment which recorded the data used in this analysis, including an overview of the different components of the detector. Chapter 4 describes the signal and background processes considered, while Chapter 5 presents the data set, triggers and Monte Carlo simulations used for event generation. The reconstruction and identification techniques for the particles involved in this analysis are presented in Chapter 6. Chapter 7 focuses on the event selection of the di-τ semileptonic decay of the Higgs boson and its production mode (VBF), in order to create a region enriched in these events called "VBF" signal region (SR).

A data-driven method *Fake Factor method* used to estimate events where jets are misidentified as hadronically decaying τ-leptons called "fakes", representing one of the main background contributions in the $H \rightarrow \tau_{\text{lep}}\tau_{\text{had}}$ channel is presented in Chapter 8, including the results of its implementation, the discussion of the assumptions made while applying the method, its validation and some additional studies made. The analysis optimizes the signal-to-background ratio in the VBF SR through the optimisation and application of a Neural Network (NN), presented in Chapter 9. The concepts of the algorithm are introduced together with a description of the properties of the NN setup utilized in this analysis, concluding with a discussion of the obtained results and definition of the SR. This SR is used for testing the CP invariance in the VBF production mode in the $H \rightarrow \tau_{\text{lep}}\tau_{\text{had}}$ decay channel presented in Chapter 10. This Chapter displays the method used to simulate CP violating effects in the SR together with two methods to constrain the strength of CP-violation described by $\tilde{d}$. This chapter finalizes with the results of the expected uncertainties of the $\tilde{d}$ values obtained using both methods for the three CP-odd considered variables. Finally, Chapter 11 provides a discussion of the results obtained in the thesis.

Chapter 2

Theoretical background

This chapter describes the theoretical background used in this analysis. First, the Standard Model of Particle Physics is summarized in Section 2.1, including a description of the elementary particles and the fundamental interactions; quantum electrodynamics, quantum chromodynamics, and the Glashow-Weinberg-Salam theory of electroweak processes. Section 2.2 gives a general picture of the Higgs boson and its production processes.

2.1 The Standard Model of Particle Physics

Each of the four fundamental forces in nature: strong, electromagnetic, weak and gravitatinal, is described by a physical theory. The properties and interactions of all known elementary particles except gravity are described within the theoretical framework of the Standard Model (SM) of particle physics, whose mathematical framework is composed by renormalizable Quantum Field Theories (QFT) with local gauge invariance, where particles and their interactions are described as interaction terms in the Lagrangian. Each QFT is defined by a set of Feynman rules, these are a set of rules that simplify the calculation of the transition matrix element in a scattering process or decay. The Feynman rules can be easily derived from the Lagrangian.

The SM has been confirmed by every experimental test and predicted several particles, e.g. top quark [16], gauge bosons [1, 2, 17], gluons [18] and Higgs boson [3], its reliability is undeniable. However several observed phenomena can not be explained by the SM e.g. gravitational interactions, neutrino masses and the baryon asymmetry in the universe. This section gives an overview of the particles and QFTs of the SM following Ref. [19–21].

2.1.1 Elementary particles

Particles are divided depending on their spin; half integer (fermions) or integer (bosons). Fermions are further categorized into leptons and quarks where only the latter ones have colour charge i.e. can interact via strong interaction.

Leptons and quarks are organised in three generations, as shown in Figure 2.1.1.1. Each lepton (ℓ) generation is composed by a charged particle, with an electric charge value $Q_\ell = -1$, and its corresponding neutrino, $Q_\nu = 0$, ordered by mass and generation: electron and electronic neutrino (e, ν_e), muon and muonic neutrino (μ, ν_μ) and tau and tauonic neutrino (τ, ν_τ). Within the same generation, quarks are also differentiated depending on their electric charges, quarks with $Q_q = \frac{2}{3}$, ordered by mass and generation: up "u", charm "c" and top "t", and with $Q_q = -\frac{1}{3}$: down "d", strange "s" and bottom "b" quarks.

Each of the four fundamental forces is mediated by the exchange of a particle, the gravitaional mediator is the graviton (although no consistent QFT has been found yet), electromagnetic forces are mediated by photons "γ", strong forces by gluons "g" and weak forces by the gauge bosons "W" and "Z", all bosons with spin $s = 1$.

Antiparticles are defined by applying the charge conjugation operator to a particle state, the antiparticle has the same spin, mass, lifetime and strength of the particle but opposite charge and magnetic momentum, only neutral particles are eigenstates of the charge conjugation operator.

The incorporation of spontaneous symmetry breaking to the SM leads to the prediction of the spin 0 Higgs boson, which will be further described in Section 2.2.

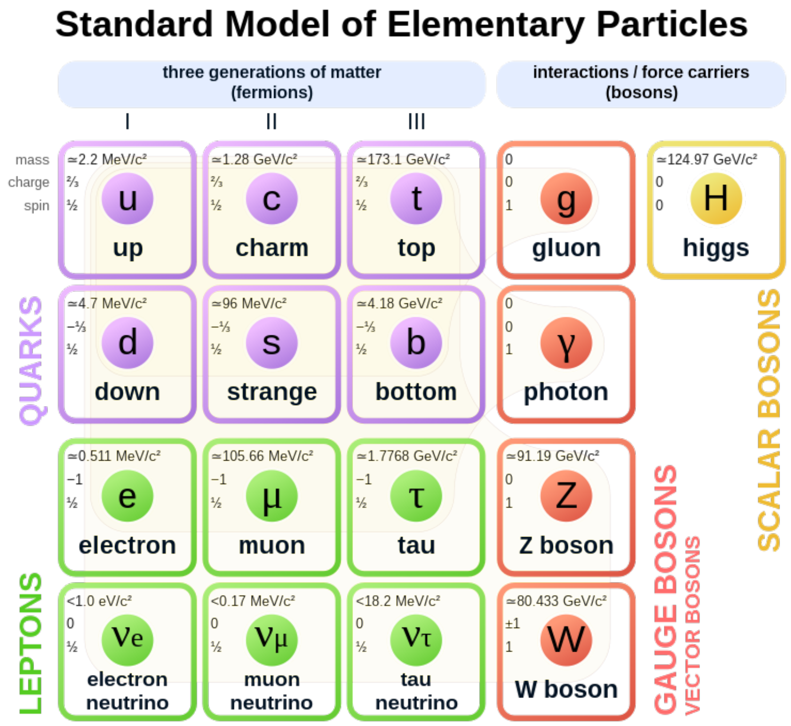


Figure 2.1.1.1: Elementary particles described in the Standard Model of Particle Physics, [23].

2.1.2 Fundamental interactions

The Feynman rules of a QFT are derived from the Lagrangian of the QFT, which is constructed depending on the particles and interactions that the QFT describes; in relativistic quantum mechanics spin $\frac{1}{2}$ particles are described with the Dirac Lagrangian (fermions), spin 1 particles with the Proca Lagrangian (gauge bosons) and spin 0 particles with the Klein-Gordon Lagrangian (Higgs boson). Furthermore, the Lagrangian of a QFT is required to be gauge invariant so that the QFT is renormalizable this is, valid for any energy scale. Gauge invariance demands two requirements; global gauge invariance and local gauge invariance. This concepts are explained using the Dirac Lagrangian, describing a spin-$\frac{1}{2}$ particle of mass m:

$$\mathcal{L}_{\text{Dirac}} = i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\bar{\psi}\psi, \tag{2.1.2.1}$$

where ψ and $\bar{\psi}$ are the four-component Dirac spinors and γ^{μ} the gamma matrices.

- **Global gauge invariance** is tested by introducing a global phase (global transformation):

$$\psi \to e^{i\theta}\psi \quad \text{and} \quad \bar{\psi} \to e^{-i\theta}\bar{\psi} \tag{2.1.2.2}$$

inserting this in the Lagrangian,

$$\mathcal{L}' = ie^{-i\theta}\bar{\psi}\gamma^{\mu}\partial_{\mu}e^{i\theta}\psi - me^{-i\theta}\bar{\psi}e^{i\theta}\psi = \mathcal{L} \tag{2.1.2.3}$$

thus, the Dirac Lagrangian is invariant under a global gauge.

- **Local gauge invariance** a local gauge transformation is applied, i.e. the transformation is not constant,

  $$\psi \to e^{i\theta(x)}\psi, \tag{2.1.2.4}$$

  introducing this transformation in the Lagrangian,

  $$\mathcal{L}' = \mathcal{L} - [\partial_{\mu}\theta(x)]\bar{\psi}\gamma^{\mu}\psi, \tag{2.1.2.5}$$

  therefore, the Lagrangian is not local invariant. However, Lagrangians are not unique, it is possible to obtain a local gauge invariant Lagrangian by introducing an interaction term into a non-local-invariant Lagrangian without changing the observables. For the Dirac Lagrangian this term is a gauge field of spin 1 "A_{μ}" (the photon in this case), multiplied by the current of the Dirac particle "$\bar{\psi}\gamma^{\mu}\psi$" and the coupling constant "g" that describes the strength of the interaction,

  $$\mathcal{L} = i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\bar{\psi}\psi - g(\bar{\psi}\gamma^{\mu}\psi)A_{\mu}. \tag{2.1.2.6}$$

  Testing local gauge invariance again setting $\theta(x) = -g\lambda$ and imposing that the field transforms as $A_{\mu} \to A_{\mu} + \partial_{\mu}\lambda$, the second term in Equation 2.1.2.5 cancels, achieving local gauge invariance.

This section provides an overview of the SM QFTs; quantum electrodynamics, the Glashow-Weinberg-Salam theory of electroweak processes and quantum chromodynamics.

Quantum Electrodynamics

Quantum Electrodynamics (QED) is a QFT that describes electromagnetic forces and whose mediators are the photons, it was first presented by Fermi in 1933 [19]. QED is a U(1) abelian unitary Lie group. The QED Lagrangian is [19]

$$\mathcal{L}_{\text{QED}} = i\bar{\psi}\gamma^{\mu}\partial_{\mu}\psi - m\bar{\psi}\psi - g_e(\bar{\psi}\gamma^{\mu}\psi)A_{\mu} - \frac{1}{4}F^{\mu\nu}F_{\mu\nu} + \frac{1}{2}m_A^2A^{\nu}A_{\nu}. \tag{2.1.2.7}$$

The coupling strength constant of QED "g_e" can be rewritten as $g_e = \sqrt{4\pi\alpha}$, related to the charge of the particle as $\alpha = Q_p^2$. This Lagrangian is the same from Equation 2.1.2.6 however, since a spin 1 field was introduced to require local gauge invariance, the Proca Lagrangian had to be inserted for consistency, where $F_{\mu\nu} = \partial_{\mu}A_{\nu} - \partial_{\nu}A_{\mu}$ is the strength-field vector of A_{μ} and $\frac{1}{4}F^{\mu\nu}F_{\mu\nu}$ the kinetic term of the field A_{μ}. However, to maintain gauge invariance, the fifth term of Equation 2.1.2.7 must vanish, therefore the spin 1 A_{μ} field must be massless this is, the photon field. Furthermore, the Noether theorem states that a continuous symmetry (gauge symmetry in this case) is related to a conservation law. In QED this is the conservation of the electric charge.

The interaction term of the photon field is proportional to the charge of the particle or antiparticle. Therefore, the photon can only couple to charged fermions or antifermions. The running coupling constant α is the result of the renormalization of QED, it solves divergences in higher order processes and represents the strength of the interaction, the behavior of α is similar to the polarization of a dielectric material in vacuum, at low momentum values (large distances) the strength of α decreases while at high momentum values (smaller distances) α is larger.

Quantum Chromodynamics

Only electrically charged particles interact via QED due to the dependence of the interaction term of the mediator particle with the charge of the particles in the Lagrangian. Therefore quarks can interact via electromagnetic forces. Additionaly, quarks carry what is called colour charge, that is the equivalent in Quantum Chromodynamics (QCD) to the electric charge in QED. Quarks can have three colours, the basis vectors are named red, green and blue, however in nature, only neutral colour hadrons exist, this is called color confinement of quarks, which states that quarks can not be isolated, so the only way to observe them is as bound states of white (colour neutral) hadrons.

QCD has SU(3) group symmetry, a unitary Lie group of 3x3 unitary matrices with determinant equal to one, therefore the Dirac spinors ψ are replaced by a dimension-3 vector of spinors that transforms as

$$\psi \to e^{i\frac{g_s}{2}\lambda^a\beta_a(x)}\psi, \tag{2.1.2.8}$$

to ensure gauge invariance. Here the index "a" ranges from 1 to 8 and represents the SU(3) colour octet states (colour singlet does not exist), "g_s" is the coupling constant of the strong force, "$\beta_a(x)$" the local gauge transformation and "λ_a" the generators of SU(3), called Gell-Mann matrices. As in QED, a gauge field is introduced to the Lagrangian to require local gauge invariance, this field transforms as

$$G^\mu_a \to G^\mu_a + \frac{1}{g_s}\partial^\mu\beta_a(x) + f_{abc}\beta^b(x)G^{\mu c}, \tag{2.1.2.9}$$

where f_{abc} are the structure constants of the SU(3) group that fulfill $[\lambda^\alpha, \lambda^\beta] = 2if^{\alpha\beta\gamma}\lambda^\gamma$ with $\gamma \in \{1, 8\}$. Then, the gauge invariant Lagrangian of QCD is

$$\mathcal{L}_{\text{QCD}} = i\bar{\psi}\gamma^\mu\partial_\mu\psi - m\bar{\psi}\psi - g_s\left(\bar{\psi}\gamma^\mu\frac{\lambda_a}{2}\psi\right)G^a_\mu - \frac{1}{4}G^a_{\mu\nu}G^{\mu\nu}_a, \tag{2.1.2.10}$$

the two first terms are the same as in Equation 2.1.2.7 since quarks are spin $\frac{1}{2}$ particles, $G^a_{\mu\nu}$ are the field-strength tensors of the gauge field, $G^{\mu\nu}_a = \partial^\mu G^\nu_a - \partial^\nu G^\mu_a - g_s f_{abc}G^{\nu b}G^{\mu c}$.

In QCD the strong force is transmitted by massless gauge bosons, gluons, whose more characteristic feature is that they can couple to each other, in contrast to photons. Referring to Noether's theorem, the conserved quantity in QCD is the colour charge, therefore if a quark radiates a gluon, the quark will change its colour.

As in QED, the renormalized coupling constant for QCD can be rewritten as $g_s = \sqrt{4\pi\alpha_s}$, where the running coupling constant of QCD "α_s" is introduced. In QED the vacuum polarization diagrams led to divergences for radiative corrections (higher order diagrams), these were "absorbed" into the running coupling constant "α" (renormalization). Since quarks also interact via QED, the behavior of α_s could be expected to be the same as α. However, in QCD virtual gluon loops provide an opposite effect, at high energies (small distances) the coupling constant is small and for lower energy values (large distances) the coupling constant is large, i.e. quarks act like free particles at high energies (small distances) while for large distances the quarks will be extremely attracted to each other due to the large α_s, this behavior is called asymptotic freedom [22] and explains the nonexistence of free quarks.

Weak interactions

Weak interactions are divided into two categories depending on the electric charge of the gauge boson that mediates the interaction; the charged current, $W^{\pm}$ and neutral current, Z. These gauge bosons are very massive and have three polarization states, in contrast with the photon or the gluon.

Similarly as the electric charge in QED and QCD and the colour charge in QCD, two charges are introduced by charged currents; the weak isospin I_w and the hypercharge Y. The weak isospin belongs to the SU(2) Lie group, therefore it shares the same mathematical properties as a spin $\frac{1}{2}$ particle. The chirality of a particle is determined by whether the particle transforms in a right- or left-handed representation of the Poincaré group, this concept is related to the helicity so that for massless particles both concepts are the same. Chirality is an important feature of weak interaction since only chiral left-hanged fermions and chiral right-handed antifermions carry the weak isospin charge. Mathematically this translates in an interaction term of the form V-A (vector−axialvector) which is maximally parity violating.

Furthermore, in the Fermi theory of weak interactions, i.e. at low energy ranges, the Fermi constant "G_F" can be derived from the muon mean lifetime, and is used to determine the magnitude of the weak coupling constant α_W, since $G_F \propto \left(\frac{g_W}{M_W}\right)^2$ and $g_W = \sqrt{4\pi\alpha_W}$.

Neutral current interactions are mediated by the electrically neutral Z boson, which has similar properties as the photons. Weak interactions are not treated as an independent interaction but together with QED as it is discussed in next Section.

Electroweak unification

The electroweak unification theory was introduced by Glashow [24], Weinberg [25] and Salam [26] in 1967 to provide a single theory that explained both electromagnetic and weak interactions. The main differences between both theories are the following: QED has only one coupling constant (g_e) while weak interaction has two, one for each current (g_W and g_Z), QED has a vector coupling and weak interaction has axial-vector coupling, provoking that weak interaction couples only with left (right) chiral particles (antiparticles), in contrast with QED. Furhtermore, the gauge boson of QED is massless, therefore it does not have longitudinal polarization, unlike weak interaction gauge bosons. Moreover, photons can only couple to charged particles, unlike weak interaction gauge bosons that can couple to all leptons and quarks. These differences are described in the unified theory.

The unified theory is based on the construction of two types of currents defined as mixtures of the weak charged and electromagnetic currents; the three weak isospin currents, corresponding to the group SU(3) and an hypercharge current corresponding to the group U(1) [27], defined as

$$Y = 2(Q - I_{w,3}), \tag{2.1.2.11}$$

where Q is the electric charge and $I_{w,3}$ the third component of the weak isospin. Table 2.1.2.1 shows the fermion quantum numbers including isospin and hypercharge. Left-handed fermions have isospin charge I_w and are grouped into doublets depending on the value of the third component of isospin $I_{w,3} = \pm 1/2$, each doublet has the same hypercharge. Right-handed fermions do not have isospin charge, only left-handed neutrinos are shown since they are considered massless in the context of this analysis.

	Generation			Quantum Number			
	1^{st}	2^{nd}	3^{rd}	I_w	$I_{w,3}$	Y	Q
Leptons	$\nu_{e,L}$	$\nu_{\mu,L}$	$\nu_{\tau,L}$	1/2	1/2	-1	0
	e_L	μ_L	τ_L	1/2	-1/2	-1	-1
	e_R	μ_R	τ_R	0	0	-2	-1
Quarks	u_L	c_L	t_R	1/2	1/2	1/3	2/3
	$d^{'}_L$	$s^{'}_L$	$b^{'}_L$	1/2	-1/2	1/3	-1/3
	u_R	c_R	t_R	0	0	4/3	2/3
	d_R	s_R	b_R	0	0	-2/3	-1/3

Table 2.1.2.1: Quantum numbers of chiral left-handed "L" and right-handed "R" fermions in the unified electroweak theory. Quarks are in the electroweak eigenstates, as further discussed in Section 2.3.2. The isospin "I_w", its third component "$I_{w,3}$", the hypercharge "Y" and the electric charge "Q" are shown.

The electroweak (EW) theory is SU(2)$_{I_w}$×U(1)$_Y$ symmetric, with the isospin charge as the generator of the SU(2) group and the hypercharge the generator of U(1). Similarly as done in QED and QCD, the theory is tested so that it is renormalizable and gauge invariant, the transformation that obtains this for the isospin states is,

$$\psi_L \rightarrow e^{i\frac{g_W}{2}\tau^{\lambda}\alpha_{\lambda}(x)}\psi_L, \tag{2.1.2.12}$$

where "g_W" is the coupling constant of the weak interaction, "τ^{γ}" the analogy to Pauli matrices (since there is SU(2) symmetry) with $\lambda \in \{1,2,3\}$ and "$\alpha(x)$" is the local perturbation. Equivalently, the gauge transformation associated to the hypercharge is

$$\psi_{L,R} \rightarrow e^{i\frac{g^{'}}{2}Y\beta(x)}\psi_{L,R}, \tag{2.1.2.13}$$

which affects both left- and right-handed fermions. The coupling constant must be different from the EW "$g^{'}$", the hypercharge is included in the local gauge transformation as "$Y\beta$(x)". The local gauge invariance requirement forces to introduce an interaction, in EW theory these are the gauge fields "W^{μ}_{λ}" for SU(2)$_{I_w}$ and "B^{μ}" for U(1)$_Y$, whose covariant derivatives

$$D^{\mu}_{L} = \partial^{\mu} + i\frac{g_W}{2}\tau^{\lambda}W^{\mu}_{\lambda} + i\frac{g^{'}}{2}YB^{\mu} \quad \text{and} \quad D^{\mu}_{R} = \partial^{\mu} + i\frac{g^{'}}{2}YB^{\mu}, \tag{2.1.2.14}$$

are used to construct the EW Lagrangian [21, 28],

$$\mathcal{L}_{EW} = \bar{\psi}_L(i\gamma_{\mu}D^{\mu}_{L})\psi_L + \bar{\psi}_R(i\gamma_{\mu}D^{\mu}_{R})\psi_R - \frac{1}{4}W^{\lambda}_{\mu\nu}W^{\mu\nu}_{\lambda} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}. \tag{2.1.2.15}$$

Where the gauge field strengths in the kinetic terms are defined as

$$W^{\mu\nu}_{\lambda} = \partial^{\mu}W^{\nu}_{\lambda} - \partial^{\nu}W^{\mu}_{\lambda} - g_W\epsilon^{bc}_{\gamma}W^{\mu}_{b}W^{\nu}_{c} \quad \text{and} \quad B^{\mu\nu} = \partial^{\mu}B^{\nu} - \partial^{\nu}B^{\mu}, \tag{2.1.2.16}$$

with the totally antisymmetric structure constants "ϵ^{bc}_{γ}" of SU(2).

The Lagrangian of EW and the gauge field strengths show that the isospin gauge fields W^{μ}_{λ} can self-couple, in contrast to the hypercharge gauge field B^{μ} which only couples with fermions. Since both fields couple to neutrinos and because a W^{μ}_{3} which interacts through a V-A coupling does not exist, none of those fields can be identified with the photon or the Z gauge boson. Instead, both the photon field A^μ and the Z boson field Z^μ are a combination of W_3^μ and B^μ,

$$\begin{pmatrix} A^\mu \\ Z^\mu \end{pmatrix} = \begin{pmatrix} \cos\theta_W & \sin\theta_W \\ -\sin\theta_W & \cos\theta_W \end{pmatrix} \begin{pmatrix} B^\mu \\ W_3^\mu . \end{pmatrix} \tag{2.1.2.17}$$

Where θ_W is the weak mixing angle that is related to the coupling constants as

$$\cos\theta_W = \frac{g_W}{\sqrt{g_W^2 + g'^2}}. \tag{2.1.2.18}$$

From a comparison of the coupling constants it can be deduced that

$$g_e = g_W \sin\theta_W = g' \cos\theta_W. \tag{2.1.2.19}$$

In QED it was shown that adding the spin 1 Proca Lagrangian that describes the photon field, the mass term associated with the photon would violate the local gauge invariance. However it was found that the photon was massless. This also happens in the EW Lagrangian; therefore, massless EW gauge fields are required to ensure gauge invariance. This does not match with observations, $m_{W^\pm} = 80.4$ GeV and $m_{Z^0} = 91.2$ GeV [29].

2.1.3 Higgs mechanism

After EW unification theory, a challenge arose in order to acommodate the observed masses of the gauge EW bosons while maintaining gauge invariance. This was solved through the Higgs mechanism, developed in 1964 [30–33]. The conceptual basis of the theory is understood when considering that a mass term of a scalar field appears when adding a harmonic potential to the Lagrangian. In the SM, a potential $V(\phi)$ of a scalar field doublet ϕ is added to the Lagrangian, this scalar field doublet is composed of the complex scalar fields ϕ^+ and ϕ^0 and defined as

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix}, \tag{2.1.3.1}$$

with the real scalar fields ϕ_i $(i = 1, 2, 3, 4)$. In order to generate the $W^\pm$ and Z^0 masses, which have $\mathrm{SU}(2)_L \times \mathrm{U}(1)_Y$ symmetry, ϕ^+ has weak isospin $I_{w,3} = \frac{1}{2}$ and ϕ^0 has $I_{w,3} = -\frac{1}{2}$, the hypercharge of the doublet is $Y = 1$. The potential of the scalar field (Higgs potential) is

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda(\phi^\dagger \phi)^2 \quad \text{with} \quad \mu^2 < 0, \lambda > 0. \tag{2.1.3.2}$$

The Higgs mechanism combines two principles; local gauge invariance and spontaneous symmetry breaking. The shape of the Higgs potential is shown in Figure 1.3.0.1, for $\mu^2 < 0$ there are infinite ground states at $\phi\phi^\dagger = \frac{v^2}{2}$, where $v^2 = -\frac{\mu^2}{\lambda}$. Spontaneous symmetry breaking happens when a specific ground state is chosen, this is

$$\phi_0 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + H(x) \end{pmatrix}, \tag{2.1.3.3}$$

where $H(x)$ is the Higgs field. Furthermore, assuming $\phi_1 = \phi_2 = \phi_4 = 0$ guarantees an electrically neutral ground state.

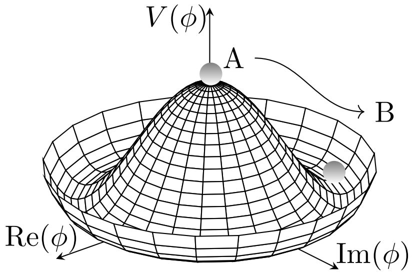


Figure 1.3.0.1: Higgs potential $V(\phi)$. The gray sphere runs from point A to a fixed ground state B spontaneously breaking the symmetry [34].

The total Lagrangian after symmetry breaking is

$$\mathcal{L} = \mathcal{L}_{\text{Higgs}} - \frac{1}{4}(\mathbf{F_w})_{\mu\nu}(\mathbf{F_w})^{\mu\nu} - \frac{1}{4}G_{\mu\nu}G^{\mu\nu}, \tag{2.1.3.4}$$

where $\mathcal{L}_{\text{Higgs}}$ corresponds to the Higgs Lagrangian,

$$\mathcal{L}_{\text{Higgs}} = (D_\mu\phi_0)^\dagger(D_\mu\phi_0) - V(\phi_0) \quad \text{with} \quad V(\phi_0) = -\lambda v^2\left(\frac{v+H(x)}{\sqrt{2}}\right)^2 + \lambda\left(\frac{v+H(x)}{\sqrt{2}}\right)^4, \tag{2.1.3.5}$$

$(\mathbf{F}_w)^{\mu\nu}$ and $G^{\mu\nu}$ are the field strengths of the SU(2) and U(1) gauge fields W^μ_λ and B^μ defined in Section 2.1.2, respectively. After spontaneous symmetry breaking, local gauge invariance is required. The transformation is done by replacing the covariant derivative in Equation 2.1.3.4 by

$$D_\mu \to \partial_\mu\mathbf{1} + \frac{1}{2}ig_W\sigma\cdot\mathbf{W}_\mu + \frac{1}{2}ig'B_\mu\mathbf{1}. \tag{2.1.3.6}$$

By introducing Equation 2.1.3.6 in $\mathcal{L}_{\text{Higgs}}$ the interaction terms and couplings between the W^μ, B^μ and $H(x)$ fields arise. From these terms the properties of the theory are extracted. The interaction term proportional to H^2 provides the mass of the Higgs scalar boson $m_H^2 = 2\lambda v^2 = -2\mu^2 =$, also found self interaction terms between three and four Higgs. Furthermore, from the Fermi constant $G_F^{-1} = \sqrt{2}v^2$, the vacuum expectation value of the Higgs field is calculated $v = 246.22$ GeV, [29]. No terms proportional to the kinetic term of the photon ($A^\mu A_\mu$) are found, thus predicting a massless photon. Comparing the self interacting term of the W^μ fields with the Proca Lagrangian, the mass term of the W boson is found to be $m_W = \frac{1}{2}g_W v$, while the mass of the Z boson is $m_Z = \frac{m_W}{\cos\theta_W}$.

Fermion masses are introduced through the Yukawa coupling presented by S.Weinberg [17] and derived from the fermion interaction with the same scalar Higgs doublet and symmetry breaking as described in the Higgs mechanism. The Yukawa coupling describes the interaction of the left-handed SU$(2)_L$ fermion doublet with the Higgs doublet ϕ and the right-handed fermion.

After symmetry breaking and gauge invariane requirement, the Lagrangian of fermion f is written as

$$\mathcal{L}^f_{\text{Yukawa}} = -\left(v\frac{\lambda_f}{\sqrt{2}}\right)\bar{\psi}^f\psi^f - \left(\frac{\lambda_f}{\sqrt{2}}\right)H\bar{\psi}^f\psi^f. \tag{2.1.3.7}$$

Where ψ^f are the Dirac spinor of fermion f and λ_f the coupling constant of the Higgs with the f quark. From the kinetic term the mass of the fermion is derived $m_f = v\frac{\lambda_f}{\sqrt{2}}$. The second term describes the fermion coupling to the Higgs boson, which is proportional to λ_f.

Therefore, the Higgs mechanism is responsible of the mass generation on the SM, resulting in a gauge invariant and renormalizable theory with the Lagrangian:

$$\mathcal{L}_{\text{SM}} = \mathcal{L}_{\text{EW}} + \mathcal{L}_{\text{QCD}} + \mathcal{L}_{\text{Higgs}} + \mathcal{L}_{\text{Yukawa}}. \tag{2.1.3.8}$$

2.2 The Higgs boson at the LHC

After its prediction, the Higgs boson was experimentally observed by the ATLAS and CMS experiments in 2012 [3, 4], with a mass of 124.97 ± 0.24 GeV measured by the ATLAS detector [35]. Among the production processes of the Higgs boson in proton-proton collisions, Figure 2.2.1 shows the example Feynman diagrams of the four dominant ones: gluon-gluon fusion (ggF), vector boson fusion (VBF), the Higgs-strahlung (VH) and the Higgs production in association with a pair of top quarks ($t\bar{t}$H).

(a) ggF

(b) VH

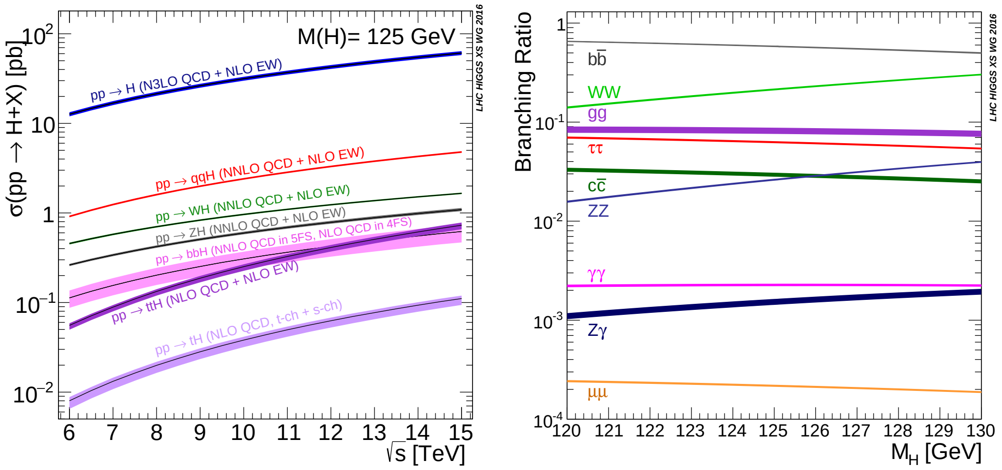
(c) $t\bar{t}$H

(d) VBF

Figure 2.2.1: Example Feynman diagrams of the main Higgs-boson production processes at the LCH: (a) gluon-gluon fusion (ggF), (b) Higgs strahlung (VH), (c) associated with a pair of top quarks and (d) Vector Boson Fusion (VBF).

Furthermore, Figure 2.2.2a shows the cross-sections for the different production modes of the Higgs boson with a mass of 125 GeV. The largest cross-section at $\sqrt{s} = 13$ TeV is 48.6 pb [36, 37] corresponding to ggF process, where the Higgs is produced from a top-quark loop. VBF process has the second largest cross-section with a value of 3.8 pb also at $\sqrt{s} = 13$ TeV [36, 37]. Here, the Higgs boson is produced from the fusion of two vector-bosons WW or ZZ resulting in two final state quarks. This process is investigated in this analysis. In the Higgs-strahlung (VH) production mode the Higgs boson is a decay product of a vector-boson W or Z produced from the annihilation of two quarks. Cross-section of this process is 2.3 pb at $\sqrt{s} = 13$ TeV [36, 37]. $t\bar{t}H$ mode of Higgs production processes have a cross section of 0.51 pb [36, 37]. Here, the Higgs boson is produced together with a top-quark pair.

Section 2.1.2 shows that weak interactions only couple to chiral left-handed states. This makes it a good source for studying CP violation, which is the objective of this analysis. Therefore, the HVV vertex is studied. This coupling exist in the VBF and VH Higgs production processes, however, the VBF production process is chosen since it has a larger cross-section and because it provides a larger momentum flow in the Higgs vertex in comparison to the VH processes [38].



(a)

(b)

Figure 2.2.2: Dominant Higgs boson production processes (a) cross sections as function of the center-of-mass-energy $\sqrt{s}$ at a Higgs boson mass of 125 GeV and (b) branching ratios as function of Higgs boson mass [36, 39].

After its production, the Higgs boson can decay leptonically or hadronically. Figure 2.2.2b illustrates the branching ratios (BR) of the decay channels of the Higgs boson depending on its mass [39]. The highest BR of 57.7% occurs when the Higgs boson decays into bottom-quarks ($H \rightarrow b\bar{b}$) followed by the decay into two W bosons ($H \rightarrow WW$) and gluons ($H \rightarrow gg$) with 27% and 9%, respectively. The Higgs decay into two τ-leptons ($H \rightarrow \tau\tau$) has a BR of 6.3% [36] which is the decay channel chosen for this analysis since it has the second highest fermionic BR. $H \rightarrow b\bar{b}$ process is not chosen due to the amount of jet activity in pp collisions.

$H \rightarrow \tau\tau$ process has three decay channels that depend on the decay modes of τ-leptons: fully leptonic ($H \rightarrow \tau_{\text{lep}}\tau_{\text{lep}}$), semileptonic ($H \rightarrow \tau_{\text{lep}}\tau_{\text{had}}$) and fully hadronic ($H \rightarrow \tau_{\text{had}}\tau_{\text{had}}$). τ_{lep} decays into a light lepton (electron or muon) and a neutrino. τ_{had} decays into a jet or quark and a neutrino. This analysis focuses on the semileptonic decay.

2.3 CP violation

Section 2.1 introduced the Standard Model (SM) of particle physics, which is the underlying theory that explains all known particles and their interactions, excluding gravity. Despite its success, the SM has not been able to explain some observed phenomena. One such example is the *baryon asymmetry* in the universe.

Recent measurements of the cosmic microwave background (CMB) found that the universe is currently composed of dark matter ($\sim 26\%$), dark energy ($\sim 69\%$) and ordinary matter ($\sim 5\%$) [7] (mostly composed of baryonic matter). Hence manifesting significant asymmetries between baryonic matter and baryonic antimatter, which is practically nil, mostly found in cosmic rays. The baryon asymmetry of the universe (BAU) can be quantified through the difference between baryon and anti-baryon densities, N_B and $N_{\bar{B}}$, respectively, to the entropy density of the universe (s) ratio, "η" defined as

$$\eta = \frac{N_B - N_{\bar{B}}}{s}, \tag{2.1}$$

Estimations of this ratio at the early universe give $\eta \sim 10^{-10}$ [5, 6], from which it follows that this asymmetry was nonexistent at the early universe. Therefore, a process called *baryogenesis* is hypothesized, which generates baryonic matter and antimatter at different rates. Baryogenesis requires three conditions to occur, known as *Sakharov conditions* [8] after Sakharov introduced them in 1967: violation of the baryon number (B) conservation, violation of C and CP conservation, and a deviation from thermal equilibrium.

The SM predicts the conservation of the baryon number and lepton number (L) individually. Nevertheless, considering non-perturbative effects, the Bell-Jackiw anomalies are found to provide the individual violation of B and L separately, while conserving the difference B-L [121–123]. The violation of the conservation of the charge conjugation (C) and charge conjugation and parity (CP) symmetries is a feature of electroweak (EW) interactions and is described in the following Sections since it is the central concept studied in this analysis. Thermal equilibrium is a time translation invariant state where expectation values of observables remain constant [5]. Interactions out of thermal equilibrium occur in the SM when including first-order phase-transitions via EW symmetry breaking [124].

The measurement of the CP symmetry in the Higgs sector is one of the widely studied topics at the LHC [38]. CP violation is present in the SM, expressed through the Cabibbo-Kobayashi-Maskawa (CKM) matrix introduced by Kobayashi and Maskawa in 1973 [125]. However, the amount of CP violation in the SM is not enough thus, new sources are required. The test of the CP symmetry violation in this analysis is done using CP-odd observables.

This section describes in detail the CP symmetry violation (CPV). First, the CP symmetry is introduced as well as the sources of CPV in the SM. Then, the theoretical framework of Effective Field theories is described in order to introduce CP-violating interactions in the SM Lagrangian, from which an Optimal Observable is defined, followed by a general description of the CP-odd observables and the introduction of the CP-odd observables used in this thesis. Three observables are used to test CP symmetry invariance.

2.3.1 CP symmetry

The CP operator combines the application of the charge "C" and parity "P" operators. The parity operator transforms the coordinate system such that it turns a right-handed coordinate system into a left one or vice versa. If the interaction Hamiltonian commutes with P, then parity is a conserved quantity of the interaction with eigenvalues ±1. The charge conjugation operator transforms a particle ψ into its anti-particle $\bar{\psi}$ e.g. $\mathrm{C}|\psi\rangle = |\bar{\psi}\rangle$. Both, parity and charge conjugation are violated separately in the SM.

A single-particle state $\psi(p, s)$ with four-momentum "p" and spin "s" is a CP eigenstate if

$$\mathrm{CP}\,|\psi(p, s)\rangle = \pm\,|\psi(-p, -s)\rangle\,. \tag{2.3.1.1}$$

When the eigenvalue is $+1$, the state is CP-even while if the eigenstate is -1, the state is CP-odd. A CP-violating process occurs if the initial and final states of the process are eigenstates of the CP operator with different eigenvalues.

2.3.2 CP violation in the Standard Model

N.Cabbibo proposed quark mixing in 1963 which explained that the weak eigenstates d' and s' are a mixture of d and s mass eigenstates of the strong interaction. This was extended to six quarks by Kobayashi and Maskawa to explain the CP-violation in the SM through the CKM matrix (V) [125]. The three quark doublets of the weak interaction are

$$\begin{pmatrix} u \\ d' \end{pmatrix}, \begin{pmatrix} c \\ s' \end{pmatrix}, \begin{pmatrix} t \\ b' \end{pmatrix} \tag{2.3.2.1}$$

so that the CKM matrix V can defined such that

$$\begin{pmatrix} d' \\ s' \\ b' \end{pmatrix}_{\mathrm{weak}} = \begin{pmatrix} V_{\mathrm{ud}} & V_{\mathrm{us}} & V_{\mathrm{ub}} \\ V_{\mathrm{cd}} & V_{\mathrm{cs}} & V_{\mathrm{cb}} \\ V_{\mathrm{td}} & V_{\mathrm{ts}} & V_{\mathrm{tb}} \end{pmatrix} \begin{pmatrix} d \\ s \\ b \end{pmatrix}_{\mathrm{mass}}, \tag{2.3.2.2}$$

where large V_{ij} values are interpreted as a preferred coupling between the i and j quarks and viceversa.

The CKM matrix is unitary, thus containing only four independent terms: three angles and a complex phase $\delta \in \mathbb{R}$. Writing V using the Euler angles θ_{12}, θ_{23} and θ_{13} leads to

$$V = \begin{pmatrix} c_{12}c_{13} & s_{12}c_{13} & s_{13}e^{-i\delta} \\ -s_{12}c_{23} - c_{12}s_{23}s_{13}e^{i\delta} & c_{12}c_{23} - s_{12}s_{23}s_{13}e^{i\delta} & s_{23}c_{13} \\ s_{12}s_{23} - c_{12}s_{23}s_{13}e^{i\delta} & -c_{12}s_{23} - s_{12}c_{23}s_{13}e^{i\delta} & c_{23}c_{13} \end{pmatrix} \tag{2.3.2.3}$$

where $s_{\mathrm{ij}} = \sin(\theta_{\mathrm{ij}})$ and $c_{\mathrm{ij}} = \cos(\theta_{\mathrm{ij}})$ [29]. The complex phase δ in a transition amplitude can cause T symmetry violation [129], since

$$T(e^{-iEt+\delta}) = e^{iEt+\delta} \tag{2.3.2.4}$$

and since CPT symmetry is conserved, CP symmetry must be violated.

The amount of CP symmetry derived from the phase in the CKM matrix is quantified by the Jarlskog invariant J [130], defined as

$$J = c_{12}c_{23}c_{13}^{2}s_{12}s_{23}s_{13}\sin\delta. \tag{2.3.2.5}$$

The most recent measurements of the Jarlskog invariant show $J = (3.18 \pm 0.15) \cdot 10^{-5}$ [29], leading to an $\eta \sim 10^{-20}$ [9], which is too small to explain the observed baryon asymmetry in the universe.

Experimentally, CP-violation was discovered in 1964 in the decays of neutral kaons [131]. Neutral spin-0 K-mesons (kaons) can be categorized into three types: flavour eigenstates produced in the strong interaction ($K^0 = d\bar{s}, \bar{K}^0 = \bar{d}s$), CP eigenstates that represent mathematical constructions (K_1, K_2), and weak interaction eigenstates observed in the weak decays of the neutral K mesons (K_S, K_L). The flavour eigenstates are not C or CP eigenstates, such eigenstates are created as linear combinations of the flavour eigenstates, that is

$$|K_1\rangle = \frac{1}{\sqrt{2}}\left(|K^0\rangle + |\bar{K}^0\rangle\right) \quad \text{and} \quad |K_2\rangle = \frac{1}{\sqrt{2}}\left(|K^0\rangle - |\bar{K}^0\rangle\right), \tag{2.3.2.6}$$

where $\mathrm{CP}|K_{1(2)}\rangle = -(+)\,|K_{1(2)}\rangle$.

The weak interaction eigenstates are found in weak interaction decays i.e. K_S decay into two pions and K_L decay into three pions. When assuming CP-invariance, K_S decays with positive CP eigenstate to two pions and K_L decays with negative CP eigenstate to three pions, thus the observed K_S and K_L were identified as the CP eigenstates K_1 and K_2, respectively.

However, Cronin and Fitch observed the decay of the K_L into two pions [131], proving CP-invariance violation. Further evidences of CP-violation processes were observed at the BaBar and Belle experiments of B mesons [132, 133].

2.3.3 Effective Field Theories

Effective field theories (EFTs) approximate physical systems by integrating out the degrees of freedom which are not relevant at the scales probed by the given experimental setup. These are replaced by a set of effective interactions between the remaining degrees of freedom [134]. The lack of new physics could indicate that new particles are heavier than the SM energy scale. The effects of this heavy particles can be described by using model-independent EFT methods. In this thesis, the SM Lagrangian is extended by a set of non-renormalizable interactions.

The Standard Model effective field theory (SMEFT)[19, 20, 134] is introduced with the same degrees of freedom, symmetry and gauge-invariant operators as the SM Lagrangian ($\mathcal{L}_{\mathrm{SM}}$) but including i interactions with an arbitrary mass dimension D. The Lagrangian of the SMEFT is [20]

$$\mathcal{L}_{\mathrm{SMEFT}} = \mathcal{L}_{\mathrm{SM}} + \frac{1}{\lambda}\sum_i c_i^{(5)}\mathcal{O}_i^{D=5} + \frac{1}{\lambda^2}\sum_i c_i^{(6)}\mathcal{O}_i^{D=6} + \frac{1}{\lambda^3}\sum_i c_i^{(7)}\mathcal{O}_i^{D=7} + \frac{1}{\lambda^4}\sum_i c_i^{(8)}\mathcal{O}_i^{D=8} + ... \tag{2.3.3.1}$$

where $\mathcal{O}_i^D$ is a gauge invariant operator of dimension D constructed from the SM fields and the parameters $c_i^{(D)}$ are the *Wilson coefficients* [20], each term is suppressed by the energy scale λ^{D-4}.

The following assumptions are made so as to reduce the number of operators in Equation 2.3.3.1: $D > 8$ operators are neglected since they are suppressed for large λ. Conservation of B-L is required, which suppresses all operators with odd mass dimensions [135, 136]. Furthermore, B conservation and a restriction to the number of fermion generations to one is required. Further operators are reduced when considering only $SU(2)_{I,L} \times U(1)_Y$ invariant CP-violating operators which are constructed using the Higgs doublet ϕ and the gauge fields $W_\mu^1, W_\mu^2, W_\mu^3$ and B_μ defined in Sections 2.1.2 and 2.1.3.

2.3.4 Effective Lagrangian

Following Refs. [10, 12, 137] and considering the assumptions from the previous section, the effective Lagrangian considered in this analysis can be written as

$$\mathcal{L}_{\mathrm{eff}} = \mathcal{L}_{\mathrm{SM}} + \frac{c_{\tilde{B}B}}{\lambda^2}\mathcal{O}_{\tilde{B}B} + \frac{c_{\tilde{W}W}}{\lambda^2}\mathcal{O}_{\tilde{W}W}, \tag{2.3.4.1}$$

where the $\mathcal{O}_{\tilde{B}}$ operator has been neglected due to existing strong constraints from measurements of CP-violation in triple gauge-boson couplings at LEP experiments [138–140]. In the HISZ basis [141, 142],

$$\mathcal{O}_{\tilde{W}W} = \phi^{\dagger}\hat{\tilde{W}}_{\mu\nu}\hat{W}^{\mu\nu}\phi \quad , \quad \mathcal{O}_{\tilde{B}B} = \phi^{\dagger}\hat{\tilde{B}}_{\mu\nu}\hat{B}^{\mu\nu}\phi \tag{2.3.4.2}$$

where $\hat{V}_{\mu\nu}$ represents the field-strength tensors of the gauge fields ($V = W^a, B$), that fulfil the relation

$$\hat{B}_{\mu\nu} + \hat{W}_{\mu\nu} = i\frac{g^{'}}{2}B_{\mu\nu} + i\frac{g}{2}\sigma^a W^a_{\mu\nu}, \tag{2.3.4.3}$$

where $B_{\mu\nu} = \partial_\mu B_\nu - \partial\nu B_\mu$ and $W^i_{\mu\nu} = \partial_\mu W^i_\nu - \partial_\nu W^i_\mu + g\epsilon^{ijk}W^j_\mu W^k_\nu$. $\tilde{V}_{\mu\nu} = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma}V^{\rho\sigma}$ are the dual field-strength tensors. The same notation from chapter 2 was used.

After the electroweak symmetry breaking, the effective Lagrangian can be written as in Ref. [143]:

$$\mathcal{L}_{\text{eff}} = \mathcal{L}_{\text{SM}} + \tilde{g}_{HAA}H\tilde{A}_{\mu\nu}A^{\mu\nu} + \tilde{g}_{HAZ}H\tilde{A}_{\mu\nu}Z^{\mu\nu} + \tilde{g}_{HZZ}H\tilde{Z}_{\mu\nu}Z^{\mu\nu} + \tilde{g}_{HWW}H\tilde{W}^{+}_{\mu\nu}W^{-\mu\nu} \tag{2.3.4.4}$$

Here, $SU(2)_{I,L} \times U(1)_Y$ invariance is required, leading to the possibility of the HVV couplings strengths in the above Lagrangian to be written in terms of two dimensionless parameters, $\tilde{d}$ and $\tilde{d_B}$ which are defined as

$$\tilde{d} = -\frac{m_W^2}{\lambda^2}c_{\tilde{W}W} \quad , \quad \tilde{d_B} = -\frac{m_W^2}{\lambda^2}c_{\tilde{B}B}\tan{\theta_W}^2, \tag{2.3.4.5}$$

where the notation from Section 2.1.2 is used. These HVV four coupling strengths are

$$\begin{aligned}
\tilde{g}_{HAA} &= \frac{g}{2m_W}(\tilde{d}\sin^2\theta_W + \tilde{d}_B\cos^2\theta_W) \\
\tilde{g}_{HAZ} &= \frac{g}{2m_W}\sin2\theta_W(\tilde{d} - \tilde{d}_B) \\
\tilde{g}_{HZZ} &= \frac{g}{2m_W}(\tilde{d}\cos^2\theta_W + \tilde{d}_B\sin^2\theta_W) \\
\tilde{g}_{HWW} &= \frac{g}{2m_W}\tilde{d}.
\end{aligned} \tag{2.3.4.6}$$

The arbitrary choice that $\tilde{d} = \tilde{d_B}$ is made since the contributions from $H\gamma\gamma$, $H\gamma Z$, HZZ and HWW to the VBF process are not experimentally distinguishable.

Under these assumptions, the parameter $\tilde{d}$ describes the amount of CP-odd contributions to the HVV coupling, resulting in a matrix element ($\mathcal{M}$) with two contributions: the CP-even part inherited from the SM and the CP-odd part that scales with $\tilde{d}$, that is

$$\mathcal{M} = \mathcal{M}_{\text{SM}} + \tilde{d} \cdot \mathcal{M}_{\text{CP-odd}}. \tag{2.3.4.7}$$

The squared matrix element is given by

$$|\mathcal{M}|^2 = |\mathcal{M}_{\text{SM}}|^2 + 2\tilde{d}\text{Re}(\mathcal{M}^*_{\text{SM}}\mathcal{M}_{\text{CP-odd}}) + \tilde{d}^2|\mathcal{M}_{\text{CP-odd}}|,$$

(2.3.4.8)

which is used in the calculation of the differential cross-section. The first and third terms of Equation 2.3.4 are CP-even, in contrast to the second term, which is CP-odd and therefore utilized to define a CP-odd observable in order to test CP-invariance in the SR.

2.3.5 CP-odd Observables

A *genuine* CP-odd observable ($\mathcal{O}$) is defined as an observable that changes its sign when applying the CP operator to a process, i.e. an observable whose expectation value vanishes if CP is conserved [126, 127]

$$< \mathcal{O} >_{\mathcal{L}=(\mathrm{CP})\mathcal{L}(\mathrm{CP})^{-1}} = 0. \tag{2.3.5.1}$$

Following the description in Ref. [126] to construct a CP-odd observable, the underlying theory must include a CP-violating phase to which the observable must depend on. This is ensured in the SM by the CKM matrix, introduced in Section 2.3.2. The objective is to find a transition between CP eigenstates with different CP eigenvalues, however, most of the processes contain states that are not CP eigenstates. Testing the CP symmetry is usually done by comparing a process and its CP conjugate via a CP-odd observable [126, 127]. Furthermore, interference terms are necessary for the observables to depend on the CP-violating phases, as shown in Section 2.3.6 [126].

The time reversal operator T transforms the state of a particle as $\mathrm{T}\,|\psi(p,s)\rangle = \langle\psi(-p,-s)|$. Further insights on CP-violation can be extracted from the properties of the “naive”-time reversal operator "T", that reverses the sign of all momentum and spin vectors i.e. $T\,|\psi(p,s)\rangle = |\psi(-p,-s)\rangle$, which in contrast to T, does not change the incoming particle into an outgoing particle. At tree-level, if the constructed genuine CP-odd observable is T-odd and has a finite, non-vanishing expectation value, the theory violates the CP symmetry. This statement requires CPT invariance, a T-symmetric phase space, an initial state that is an eigenstate or invariant under T transformation, and the absence of rescattering effects [127, 128].

2.3.6 The Optimal Observable

Following Section 2.3.5, the *Optimal Observable* ($\mathcal{OO}$) is defined using the interference term in Equation 2.3.4 [144–146] as

$$\mathcal{OO} = \frac{2\mathrm{Re}(\mathcal{M}^*_{\mathrm{SM}}\mathcal{M}_{\text{CP-odd}})}{|\mathcal{M}_{\mathrm{SM}}|^2}. \tag{2.3.6.1}$$

The $\mathcal{OO}$ incorporates the phase-space of the variables that describe the final state of the VBF production process of the Higgs boson (two reconstructed jets and the Higgs boson) and is most sensitive to small deviations from the SM prediction [145].

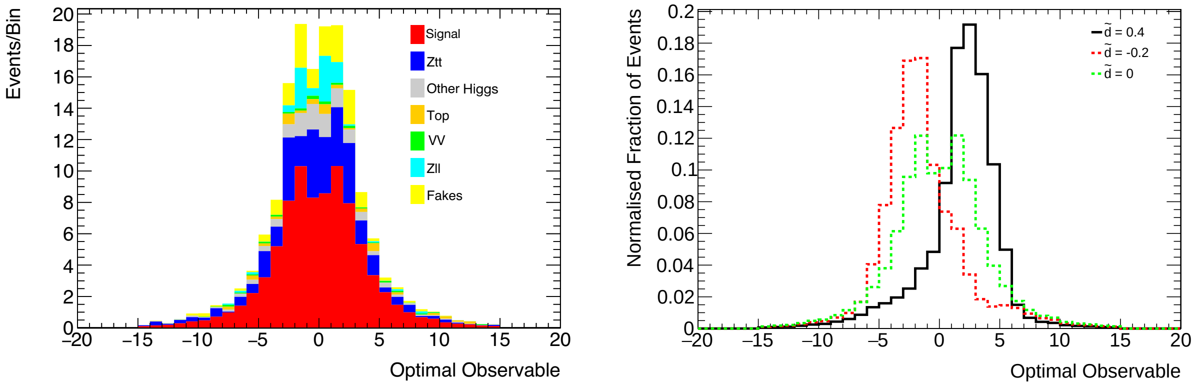
Referring to Section 2.3.5, the Optimal Observable is a CP-odd and T-odd observable. Hence, its non-vanishing expectation value would indicate CP violation or rescattering effects from new particles i.e. physics beyond the SM.

The Optimal Observable distribution in the SR is shown in Figure 2.3.6.1. Figure 2.3.6.1a illustrates the Optimal Observable for $\tilde{d} = 0$ (the SM case) with the signal and background contributions. Figure 2.3.6.1b shows the Optimal Observable for three different values of $\tilde{d}$ for the normalized signal yields.

2.3.7 Other CP-odd variables

Two other CP-odd variables are considered; signed azimuthal angle difference between the two leading-p_{T} jets ($\Delta\phi_{jj}$), and the optimal observable determined through the use of *symbolic regression* methods ($\mathcal{OO}_{\mathrm{Reg}}$).

$\Delta\phi_{jj}$ is defined as the difference between the azimuthal angle of the two VBF tagging jets, after being sorted by their pseudorapidity. This observable can be used to determine the tensor structure of the HVV vertex [12]. It is independent of the decay channel and off-shell Higgs mass



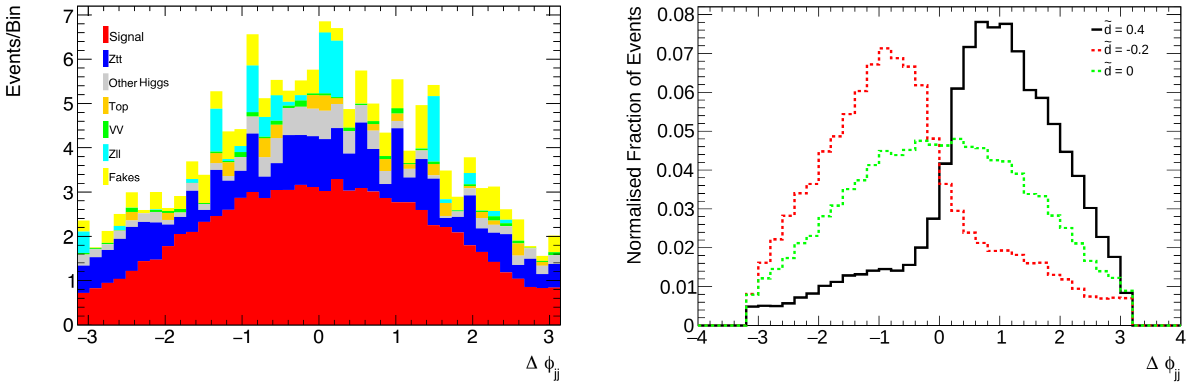
(a) Optimal Observable for $\tilde{d} = 0$

(b) Optimal Observable for three different $\tilde{d}$ values

Figure 2.3.6.1: Optimal Observable distribution for (a) SM case with signal and background contributions distinguished and (b) for different $\tilde{d}$ values with the normalized signal contribution, in the SR defined in Chapter 9. No uncertainties included.

range [13]. $\Delta\phi_{jj}$ has been previously used to test CP-invariance in the VBF [13] and ggF [95] production modes of the Higgs boson in the $H \to \gamma\gamma$ or $H \to ZZ^* \to 4\ell$ final states [147].

$\Delta\phi_{jj}$ distribution in the SR is shown in Figure 2.3.7.1. Figures 2.3.7.1a and 2.3.7.1b show $\Delta\phi_{jj}$ for $\tilde{d} = 0$ (SM case) for the signal and background processes and for three different $\tilde{d}$ values for the normalized signal, respectively.



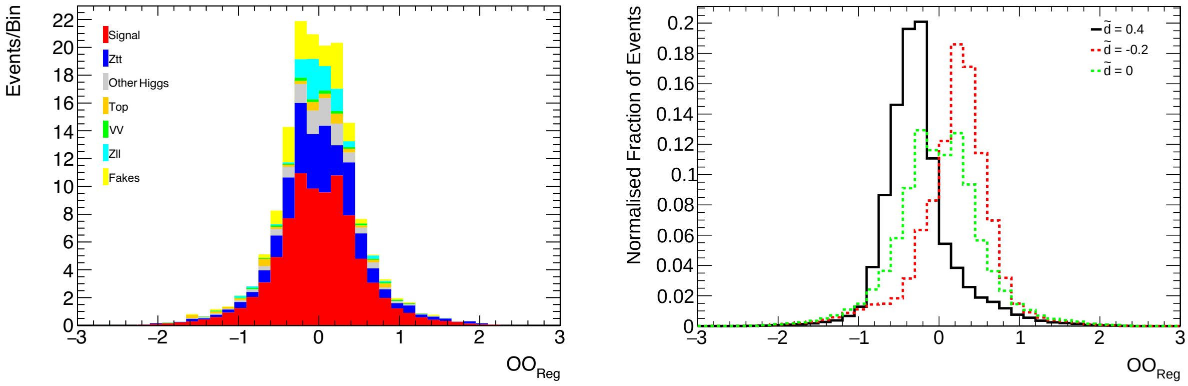
(a) $\Delta\phi_{jj}$ for $\tilde{d} = 0$

(b) $\Delta\phi_{jj}$ for different values of $\tilde{d}$

Figure 2.3.7.1: $\Delta\phi_{jj}$ distribution for (a) SM case with signal and background contributions distinguished and (b) for different $\tilde{d}$ values with the normalized signal contribution, in the SR defined in chapter 9. Overflow and underflow bins shown in (a).

An optimal observable formula for the VBF Higgs boson production process was derived in Ref. [14] from a symbolic regression method, that is, using machine learning techniques to extract symbolic formulas directly from data. In the VBF production process, the formula is found to be $\mathcal{OO}_{\text{Reg}} \sim -5.54 \cdot 10^{-5} p_{\text{T}}^{j_1} p_{\text{T}}^{j_2} \sin(\Delta\phi_{jj})$ [14] where j_1 and j_2 are the two leading jets from the VBF final state. The constant at the beginning is calculated including detector effects using an integrated LHC luminosity of 139fb^{-1}.

Figure 2.3.7.2 illustrates the $\mathcal{OO}_{\text{Reg}}$ distribution in the SR. Figure 2.3.7.2a includes all contributions from the signal and background processes for $\tilde{d} = 0$, while Figure 2.3.7.2b shows $\mathcal{OO}_{\text{Reg}}$ for three different values of $\tilde{d}$ using the normalized signal processes.



(a) $\mathcal{OO}_{\text{Reg}}$ for $\tilde{d} = 0$

(b) $\mathcal{OO}_{\text{Reg}}$ for three different $\tilde{d}$ values

Figure 2.3.7.2: $\mathcal{OO}_{\text{Reg}}$ distribution for (a) SM case with signal and background contributions distinguished and (b) for different $\tilde{d}$ values with the normalized signal contribution, in the SR defined in chapter 9. No uncertainties included.

Chapter 3

Experimental setup

This chapter gives an overview of the Large Hadron Collider (LHC) in Section 3.1, followed by a description of the ATLAS detector and its components in Section 3.2.

3.1 LHC

The Large Hadron Collider [41] at CERN (Conseil Européen pour la Recherche Nucléaire), near Geneva, Switzerland is a powerful particle accelerator [41, 42]. It can collide protons (pp) or heavy ions in a circular tunnel with a circumference of 26.7 km located about 170 m below the ground. This tunnel previously housed the LEP (Large Electron-Positron) collider. The LHC is operating since 2008. It has a design center-of-mass energy ($\sqrt{s}$) of 14 TeV for pp collisions. The injected particle beams are accelerated using 16 radiofrequency cavities. Beams are bent with superconducting magnets that can provide a field strength of up to 8.3 T. They are kept focused with 392 quadrupol magnets. The particles collide at four points in the ring where the main experiments are located: ATLAS [43], CMS [44], LHCb [45] and ALICE [46]. A simplified schematic of the LHC complex is shown in Figure 3.1.1. The ATLAS and CMS (Compact Muon Solenoid) are general purpose experiments, LHCb (LHC-beauty) focuses on b physics and ALICE (A Large Ion Collider Experiment) specializes in heavy-ion physics research.

Protons are first extracted from a hydrogen source. Their energy is increased by injecting them through a set of accelerators. In order, these accelerators are: LINAC (linear accelerators for protons or lead), the Proton Synchrotron Booster (PSB), the Proton Synchrotron (PS) and the Super Proton Synchrotron (SPS). After SPS, the proton beams are injected into the LHC ring. These beams contain up to 2808 *bunches* with 10^{11} particles per bunch for pp collisions.

The LHC performance is characterized with the instantaneous luminosity ($\mathcal{L}$) and the center-of-mass energy ($\sqrt{s}$), [48]. The luminosity is a measure of the number of collisions that can happen in a collider per cm^2 and second, defined as

$$\mathcal{L} = \frac{N^2 \gamma f_{\mathrm{rev}}}{4\pi\beta^* \epsilon_n} n_b \cdot F, \tag{3.1.1}$$

where N is the number of particles per bunch, γ the Lorentz factor, f_{rev} the revolution frequency of the protons in the accelerator, β^* the beta function at the collision point, n_b the number of bunched per beam and F the geometric luminosity reduction factor, that accounts for the beam crossing angle at the interaction point. Integrated over time leads to the integrated luminosity.

The peak design luminosity for pp collisions is 10^{34} cm^{-2} s^{-1}, whereas the design center-of-mass energy is 14 TeV.

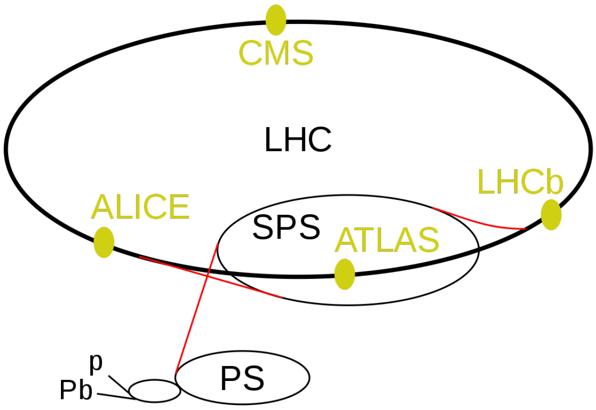


Figure 3.1.1: A simplified scheme of the LHC accelerator complex. The collision points represented by yellow dots show where the four main experiments (ATLAS, CMS, LHCb and ALICE) are located. Also, the Super Proton Synchrotron (SPS) and the Proton Synchrotron (PS) and the linear accelerators for protons (p) and lead (Pb) are shown [47].

In 2011 and 2012 the LHC had its first data-taking period called *Run-1*, with center-of-mass energies of 7 and 8 TeV, respectively. Run-2 begun in 2015 with a center-of-mass energy of 13 TeV and ended in 2018. Data collected during the full Run-2 period is used in this analysis for the fake background estimation, corresponding to an integrated luminosity of 139 fb^{-1}. Run-3 started in 2022 and is planned to exceed an integrated luminosity of 200 fb^{-1} [49].

3.2 The ATLAS experiment

The ATLAS [43] (A Toroidal LHC ApparatuS) detector is designed for multi-purpose research. It is located 100 m below ground level, has a cylindrical shape with a diameter of 25 m, length of 44 m and a weight of 7000 tonnes, [43]. The ATLAS detector is nominally forward-backward symmetric with respect to the interaction point. The different components of the ATLAS detector are shown in Figure 2.0.1 from the inside to outside, main components are: the inner detector (ID), the electromagnetic (ECAL) and hadronic calorimeters (HCAL), and the muon spectometer (MS). Figure 2.0.2 shows different types of luminosity measurements by the ATLAS detector in the course of Run-2. 139 fb^{-1} of good quality data recorded at a center-of-mass-energy of 13 TeV is used in this analysis.

3.2.1 Coordinate system

The ATLAS detector has a right-handed coordinate system, defined with its origin at the center of the detector i.e. the nominal interaction point. The cartesian coordinates are defined as shown in Figure 3.2.1.1; with the z-axis lying along the beam pipe, the x-axis pointing to the center of the main LHC ring, and the y-axis pointing upwards. The $x-y$ plane is called the *transverse plane*. Multiple variables are defined by projecting their vectors on this plane.

Two polar angles are defined. The polar angle θ, is measured with respect to the z-axis. The azimuthal angle, ϕ, is measured in the $x-y$ plane. The rapidity "y" of a particle with energy E and a momentum in the z direction of p_z is given by the formula

$$y = \frac{1}{2}\ln\left(\frac{E+p_z}{E-p_z}\right). \tag{3.2.1.1}$$

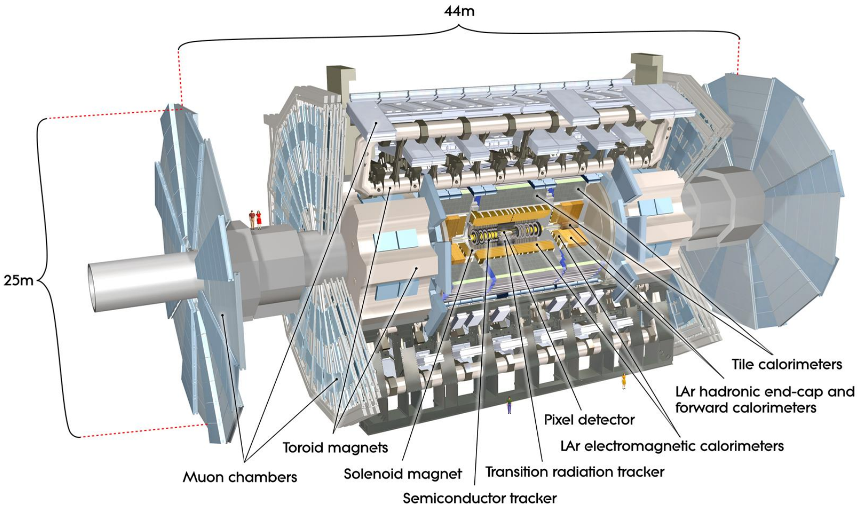


Figure 2.0.1: Overview of the ATLAS detector with its main components [51].

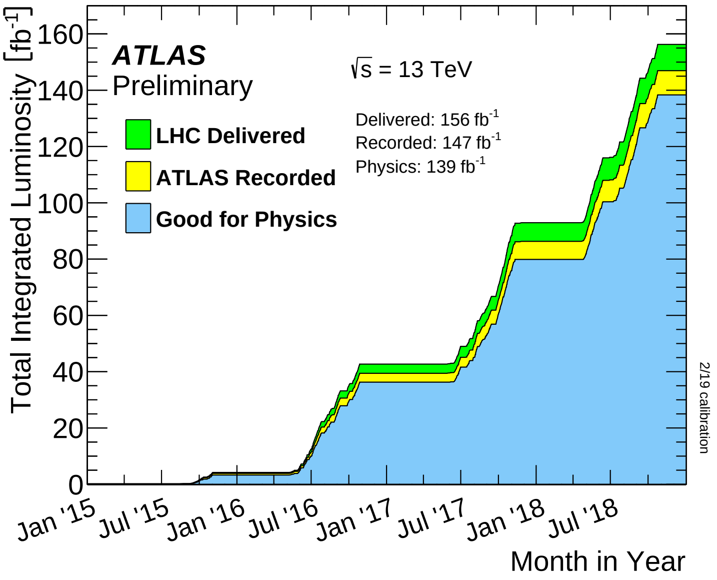


Figure 2.0.2: Total integrated luminosity at Run-2, delivered (green), recorded (yellow), and data qualified as "good for physics" that is used in analyses (blue) [50].

The rapidity difference between the two particles 1 and 2 is defined $\Delta y = y_2 - y_1$, which is a Lorentz invariant quantity under a boost in the z direction. In the limit where the momentum of the particle is much larger than its mass ($|\vec{p}| >> m$) the rapidity reduces to a quantity called *pseudorapidity*, represented by η and defined as

$$\eta = -\ln\left(\tan\left(\frac{\theta}{2}\right)\right). \tag{3.2.1.2}$$

The angular distance between two objects on the $\theta - \phi$ plane is measured with ΔR, given by

$$\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}. \tag{3.2.1.3}$$

The transverse momentum p_{T} and energy E_{T} are defined as the projections on the transverse plane $(x-y)$, this is:

$$p_{\mathrm{T}} = \sqrt{p_x^2 + p_y^2} \quad \text{and} \quad E_{\mathrm{T}} = \sqrt{E_x^2 + E_y^2}. \tag{3.2.1.4}$$

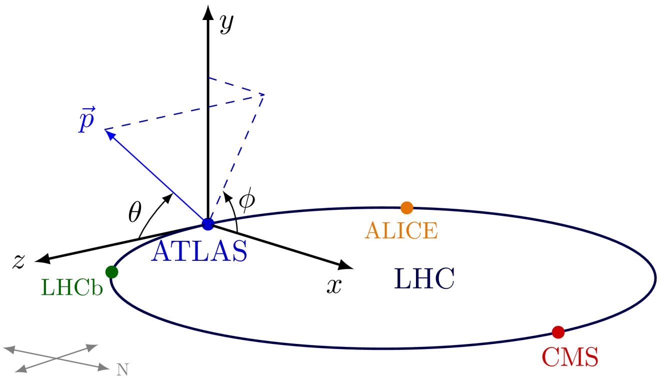


Figure 3.2.1.1: Overview of the coordinate system of the ATLAS detector . The cartesian coordinates x, y and z are defined relative to the origin at the interaction point. The polar angles θ and ϕ are shown for $\vec{p}$, which is an example representation of a momentum vector. The relative location of the three other main LHC experiments (CMS, ALICE and LHCb) are also shown [52].

3.2.2 Detector Components

Inner detector

The Inner Detector (ID) [43] is the nearest detector unit to the collision point. The magnet configuration of the ATLAS comprises a superconducting solenoid that surrounds the ID cavity, providing a magnetic field of up to 2 T. The ID has the form of a cylinder with a length of 6.2 m and a radius of 2.1 m. It consists of three systems: the pixel detector (PD), a silicon strip detector called the Semi-Conductor Tracker (SCT), and the transition radiation tracker (TRT). The ID is responsible for providing high-resolution tracking and transverse momentum measurements for charged particles.

The PD [43] is the innermost component of the ID. It consist on silicon pixel detectors that provide the highest granularity around the vertex region. Specifically, it is built with 250 μm silicon, with ≈ 80 million pixels, each with a size between $50 \times 400 \mu\mathrm{m}^2$ and $50 \times 600 \mu\mathrm{m}^2$ depending on the region. Its innermost layer, *insertable b-layer* (IBL) is designed for improving tracking performance. The PD consists of seven superconducting toroids: four concentric layers parallel to the beam pipe called *barrels* and three disk-shaped on each side, perpendicular to the beam pipe called *end-caps*, as illustrated in Figure 3.2.2.1.

The SCT [43] surrounds the PD, consisting of a silicon detector with narrow stripes of four barrel layers and nine end-cap disks on each side, extending to $|\eta| < 2.5$. It is composed of double-sided silicon modules with around 6.3 million readout channels. The SCT measures $R - \phi$ in the barrel region using stereo strips placed at small angles (40 mrad) with one set of strips in each layer parallel to the beam direction, in contrast to the end-cap region where these latter set of strips is radially located.

The TRT [43] is a combination of a straw tracker and a transition radiation detector. Straws are Xe-based filled-drift tubes with polyamide pipes. The TRT plays a crucial role in the identification of electrons since the intensity of the transition radiation produced by a particle is proportional to its γ factor. The TRT covers the $|\eta| < 2.01$ range. In the barrel region of the TRT the straws are parallel to the beam axis while arranged radially in the end-cap region.

Design resolution of the ID is shown in Table 3.2.2.1.

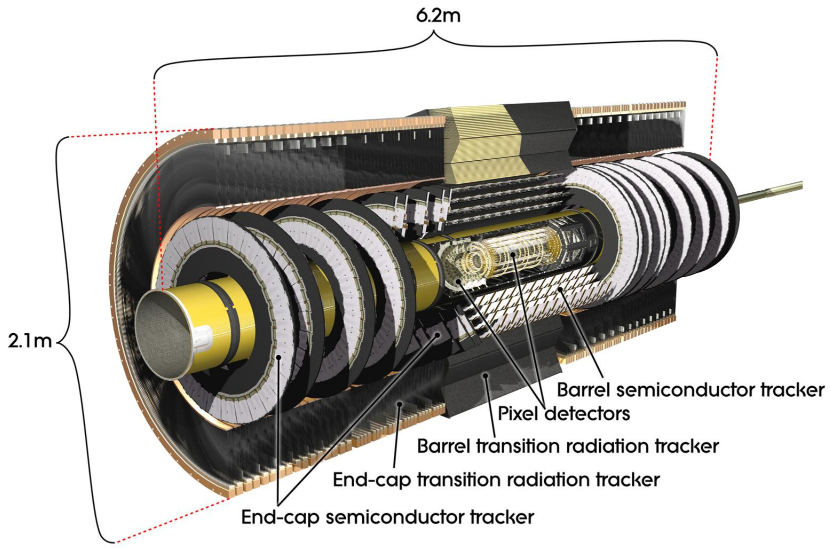


Figure 3.2.2.1: Overview of the ATLAS Inner Detector with its main components [43].

Calorimeters

The ID is in turn surrounded by the calorimeter system. First by the high granularity liquid-argon (LAr) electromagnetic calorimeter (ECAL), followed by the a scintillator-tile hadronic (HCAL) calorimeter, covering the total range of $|\eta| < 4.9$ [43]. The design resolution for the different parts of the calorimeter system is shown in Table 3.2.2.1. Figure 3.2.2.2 illustrates the main components of the ATLAS calorimetry system.

Electromagnetic calorimeter

The ECAL [43] measures the electromagnetic showers produced by photons and electrons, specifically their identification and energy measurements. It is a lead (absorber material) and liquid-argon (active medium) sampling calorimeter with "accordion-shaped" structure. It is divided into a barrel and two end-cap regions covering $|\eta| < 1.475$ and $1.375 < |\eta| < 3.2$ pseudorapidity regions, respectively. the coverage in azimuthal angle is uninterrupted. The energy loss of the electrons or photons up to the calorimeters is accounted for a presampler in the $|\eta| < 1.8$ pseudorapidity range.

Hadronic calorimeter

The HCAL [43] is dedicated to the energy measurement of hadrons. More material is used for the HCAL than for the ECAL in order to provide sufficient stopping power for the hadronic showers. It consist on three parts, as illustrated in Figure 3.2.2.2. A scintillator-tile sampling calorimeter, directly after the ECAL, it is separated into a large barrel and two smaller extended barrel cylinders, one on either side of the central barrel in the $|\eta| < 1.7$ region. The LAr hadronic end-cap calorimeter (HEC) consisting on two independent wheels per end-cap and located behind the end-cap ECAL, it extends out to $|\eta| = 3.2$ in order to reduce the drop in material density at the transition between the end-cap and the LAr forward calorimeter (FCal). The FCal improves the homogeneity in the measurements of the calorimetric coverage and reduces the radiation background levels in the muon spectrometer.

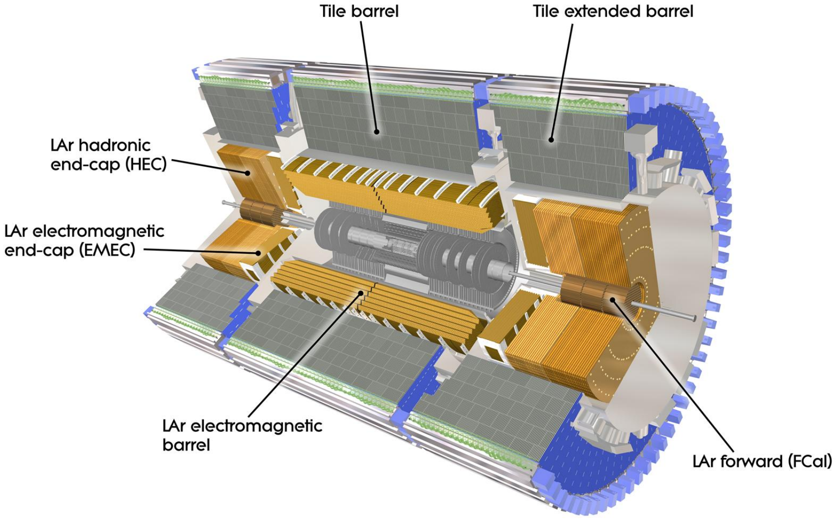


Figure 3.2.2.2: Overview of the ATLAS Calorimeter system with its main components [43].

Muon system

Muons traverse the calorimeter system. Detector signatures outside the calorimeters are assigned to muons. The Muon System (MS) [43] surrounds the calorimeter system and consist of a long barrel and two end-cap magnets that generate a strong bending power, minimising multiple-scattering effects and providing an excellent muon momentum resolution. The barrel region extends up to $|\eta| = 1.4$, a transition region to provide magnetic deflection from the barrel and end-cap fields covering the $1.4 < |\eta| < 1.6$ pseudorapidity range and the end-caps in the $1.6 < |\eta| < 2.7$. The magnetic field for the MS is provided by a toroidal magnet system.

The MS is comprised of muon chambers, including Monitored Drift Tubes (MDT's), providing precision measurements of the track coordinates in the bending direction, Cathode Strip Chambers (CSC's) to withstand the background conditions at large pseudorapidities, and Resistive Plate Chambers (RPC's) and Thin Gap Chambers (TGC's) in the barrel and end-cap regions, respectively, working as a trigger system.

The design resolution for the different parts of the muon system is shown in Table 3.2.2.1. Moreover, Figure 3.2.2.3 illustrates the main components of the ATLAS calorimetry system.

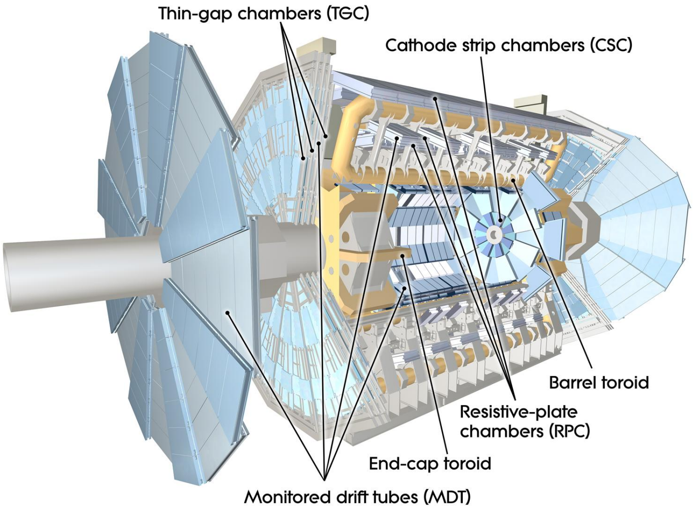


Figure 3.2.2.3: Overview of the ATLAS Muon system with its main components [43].

Trigger System

The bunch crossing rate at the LHC is 40 MHz [43]. To select relevant events during the data-taking, a two level trigger system is used at ATLAS, reducing the readouts and storage [53, 54]. The Trigger system is divided into the hardware-based Level-1 (L1) and the software-based High-level trigger (HLT). The L1 uses a subset of the total detector information to determine Regions-of-Interest (RoIs). This RoIs are processed by the HLT, which uses the high-granularity information from the calorimeters, MS and ID, reducing the event rate to approximately 1 kHz with an average decision time of 200 ms per event.

Subdetector	Design resolution
Inner Detector	$\sigma_{p_\mathrm{T}}/p_\mathrm{T} = 0.05\%\ p_\mathrm{T} \oplus 1\%$
Electromagnetic Calorimeter	$\sigma_E/E = 10\%/\sqrt{E} \oplus 0.7\%$
Hadronic Calorimeter	
- Barrel and End-cap	$\sigma_E/E = 50\%/\sqrt{E} \oplus 3\%$
- Forward	$\sigma_E/E = 100\%/\sqrt{E} \oplus 10\%$
Muon System	$\sigma_{p_\mathrm{T}}/p_\mathrm{T} = 10\%$ at $p_\mathrm{T} = 1$ TeV

Table 3.2.2.1: Design resolution of the subdetectors of the ATLAS experiment. Energy (E) and transverse momentum (p_T) values are given in units of GeV (except in the Muon System). The notation $a \oplus b = \sqrt{a^2 + b^2}$ is followed [43].

Chapter 4

Signal and background processes

This chapter describes the signal and background processes considered in this analysis. Section 4.1 presents the signal processes which is the vector-boson fusion production mode of the Higgs boson, decaying into two τ-leptons or two W bosons (VBFHWW). Section 4.2 presents the background processes that can lead to the same final state as the signal process, and processes that can be misidentified as signal processes.

4.1 Signal processes

Signal process is the vector-boson fusion (VBF) production of the Higgs boson. Other production modes of the Higgs boson are discussed in Section 2.2.

This analysis is centered on the decay of the VBF-produced Higgs boson into a pair of τ-leptons, in the semileptonic decay channel, resulting in a final state with a light lepton ($\ell = e, \mu$), visible part of hadronically decaying τ lepton (τ_{had}), and three neutrinos (ν). That corresponds to a decay chain of $H \to \tau\tau \to \ell + \tau_{\text{had}} + 3\nu$. Figure 4.1.1 illustrates the Feynman diagram of this process.

Furthermore, the analysis also includes the VBF production process of Higgs decaying into two W bosons, $H \to WW \to \ell + \tau_{\text{had}} + 3\nu$, as a signal process. The contribution from this processes to the total yield is $\sim 0.0015\%$ in the VBF SR defined in Section 7.2, as shown in Table 7.3.1.

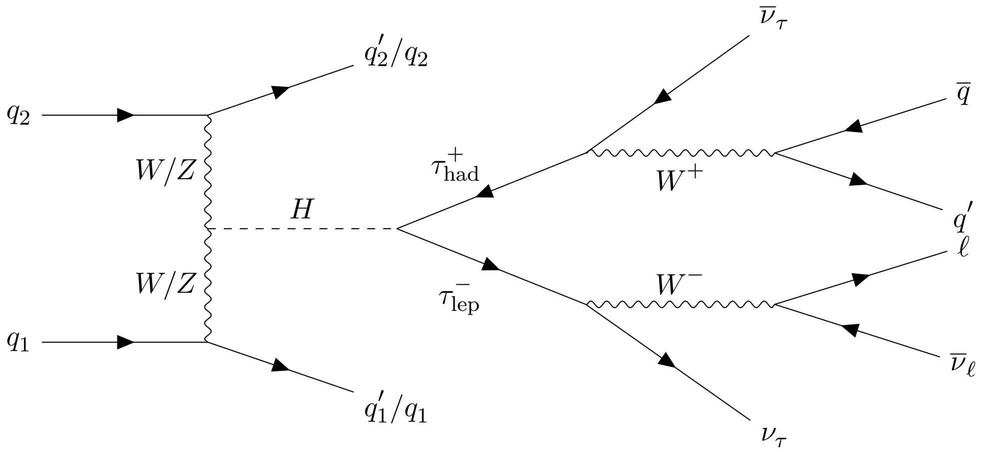


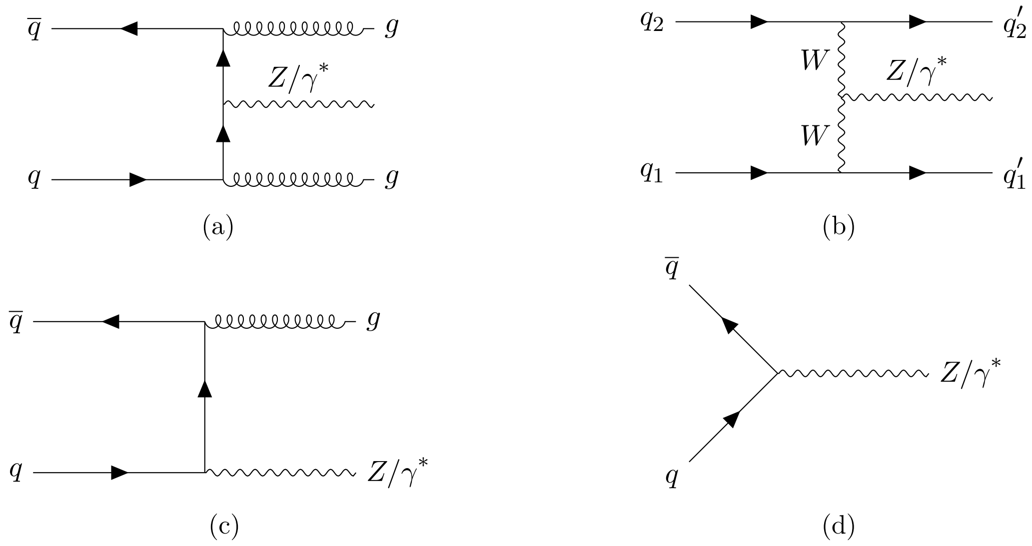
Figure 4.1.1: Example Feynman diagram of the VBF production of the Higgs boson decaying into a pair of τ leptons in the semileptonic decay channel.

4.2 Background processes

Background processes can be classified into *reducible* and *irreducible* processes. Irreducible processes provide the same final state as the signal process. In contrast, reducible processes have different final states, but due to the misidentification of objects, they can end up having the same final state composition.

Z/γ^* production process is the dominant and background contribution in this analysis, since these events can decay as $Z/\gamma^* \to \tau_{\text{lep}}\tau_{\text{had}} \to \ell + \tau_{\text{had}} + 3\nu$+jets leading to an irreducible background contribution. Figure 4.2.1 shows example Feynman diagrams of these production processes. This background contribution is suppressed by exploiting the mass difference between the Z- and Higgs-bosons by applying restrictions to the invariant mass of the di-τ system. The other two decay modes of the τ-leptons are suppressed when applying the semileptonic decay channel selection i.e. $Z/\gamma^* \to \tau_{\text{lep}}\tau_{\text{lep}}$ and $Z/\gamma^* \to \tau_{\text{had}}\tau_{\text{had}}$, (see Section 7.1).

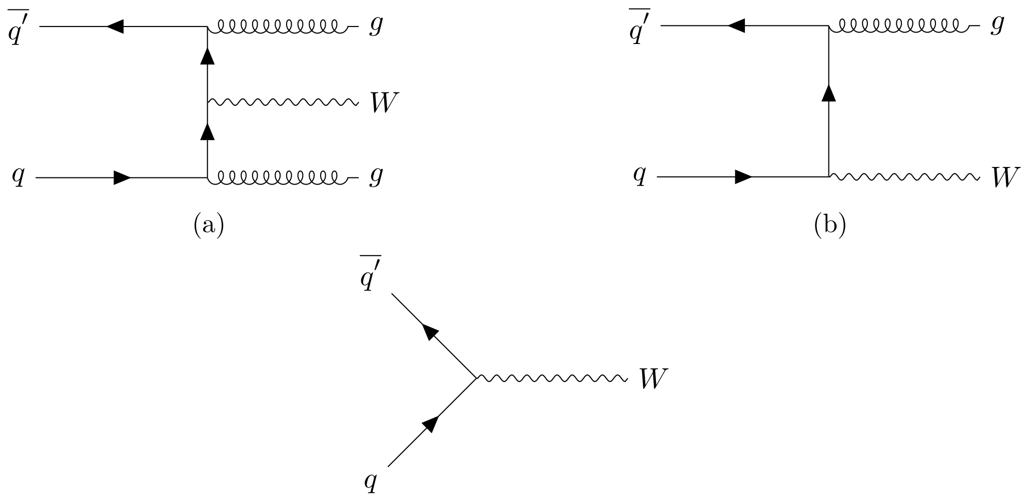
Decay of Z/γ^* into a pair of light leptons is a reducible background. Since one light lepton can be misidentified as a τ_{had} it can mimic the signal processes. This background is reduced by requiring a low amount of missing transverse energy ($E_{\text{T}}^{\text{miss}}$). In addition, Z/γ^* can be produced with additional jets that can be misidentified as τ_{had}, referred to as "fakes".



(a) (b) (c) (d)

Figure 4.2.1: Example Feynman diagrams of the production processes of the Z/γ^*.

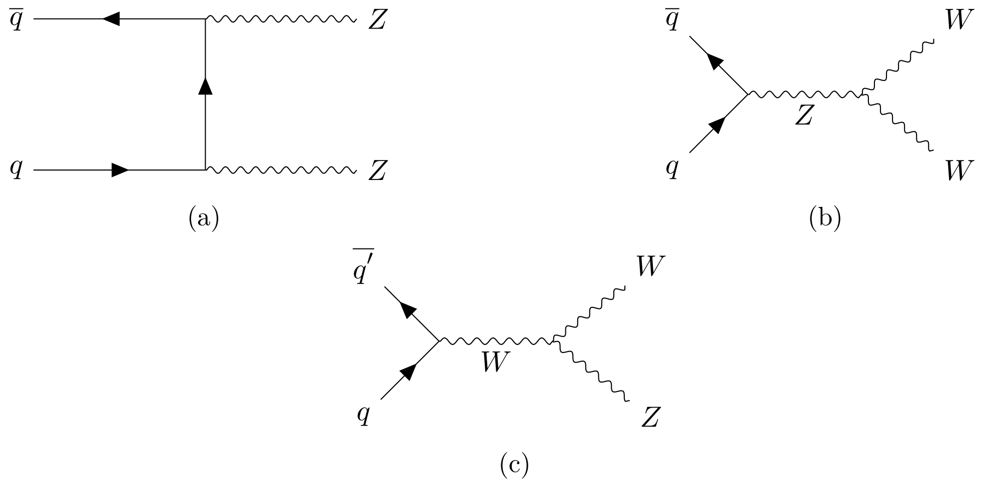
The production of W bosons with associated jets (W+jets) represent one of the main sources of jets misidentified as hadronically decaying τ-leptons (fakes). It is a reducible background. Figure 4.2.2 shows example Feynman diagrams of this process. Background from W+jets processes is suppressed by applying a selection requirement to the value of the transverse mass between the lepton and the neutrino, decay products of the W boson.



(a) (b) (c)

Figure 4.2.2: Example Feynman diagrams for (a) and (b) W+jets and (c) single W production processes.

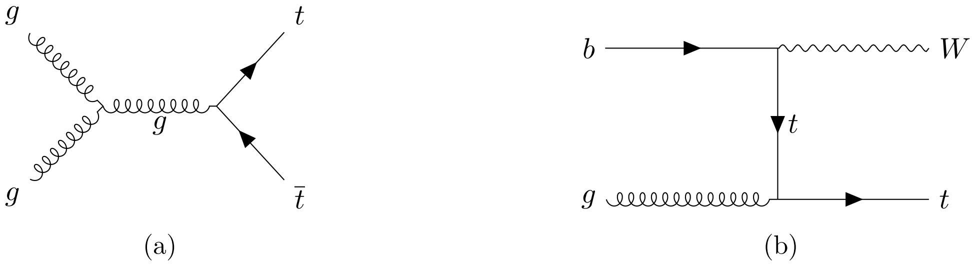
Diboson production (VV) includes $WW-$, $WZ-$ and $ZZ-$diboson production processes. The W and Z can decay leptonically or hadronically. Figure 4.2.3 shows example Feynman diagrams of this processes.



(a) (b) (c)

Figure 4.2.3: Example Feynman diagrams for diboson (VV) production.

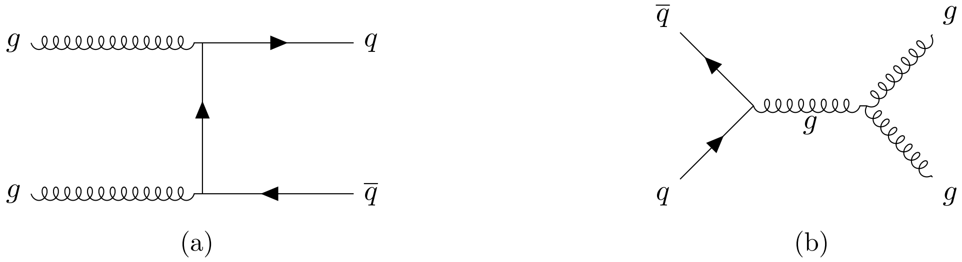
The top-quark can decay hadronically or leptonically, it can have in a final state with a real or fake τ_{had} or a lepton. Single top-quark and $t\bar{t}$ production processes are considered as background processes. Figure 4.2.4 illustrates example Feynman diagrams of these processes. Top quarks almost always decay into a W boson and a bottom quark, therefore these processes can be partially suppressed by applying a b-jet veto.



(a) (b)

Figure 4.2.4: Example Feynman diagrams for (a) $t\bar{t}$ and (b) single top quark production processes.

Multi-jet (QCD) processes involve the production of quarks and gluons. They are processes with large production cross-sections at the LHC. Furthermore, these events have a large multiplicity of jets, some of which can be misidentified as leptons or τ_{had} . Thus, they could contribute to the fake background, as explained in Chapter 8. Figure 4.2.5 illustrates two examples Feynman diagrams of the QCD background.



(a) (b)

Figure 4.2.5: Example Feynman diagrams for multi-jet (QCD) production.

The following notation is going to be used in the Figures and Tables of this thesis. The VBF process of the Higgs boson decaying both, into two τ-leptons or two W bosons is labelled "Signal". $Z \rightarrow \tau\tau$ background processes are labelled as "Ztt", single W boson production as "W", single Z boson production as "Zll", diboson as "VV", single top and $t\bar{t}$ production are combined into "Top" and all irreducible background contributions derived from jets misidentified as τ_{had} as "Fakes". The Higgs boson production processes introduced in Section 2.2, except from VBF are also considered a background contribution. These will be referred to as "Other Higgs"

Chapter 5

Data set and simulated events

This chapter describes the data set and triggers applied as part of the event selection in this analysis, described in Section 5.1 and 5.2, respectively.

In proton-proton colliders, data is compared with theoretical descriptions given by Monte Carlo (MC) simulations, which are random sampling event generators. These simulate the *hard* and *soft* processes derived from the collision, distinguished by the scale of the momentum transfer (Q). Hard processes have large Q, these can be estimated through perturbation theory. In contrast, soft processes have small Q and are obtained from non-perturbative theories. MC simulations include the parton distribution functions (PDF) of the protons, hadronization process, the radiation provoked by the hard processes, *pile-up* processes consisting of background processes originated from inelastic scattering processes (which contribute significantly to the background at a nominal luminosity of 10^{34} cm^{-2}s^{-1}); and the so-called *underlying event* derived from beam remnants and multiple parton interactions. The MC simulation tools used in this analysis for the signal and background events are described in Sections 5.3 and 5.4, respectively.

5.1 Data set

The data set used in this analysis for the fake background estimation is the full Run-2 proton-proton collision data set of the LHC, recorded between 2015 and 2018 with an integrated luminosity of 139 fb^{-1} at a center-of-mass energy ($\sqrt{s}$) of 13 TeV. Only data corresponding to fully operational subsystems of the detector are taken into account. Data satisfying this quality requirements are classified into the "Good Runs List" which is used in this analysis, as shown in Figure 2.0.2.

5.2 Triggers

In this analysis single lepton triggers, are applied to the electrons and muons [53, 55]. Different triggers are applied depending on the data-taking period. Table 5.2.1 summarizes the HLT triggers used in this analysis and reports the p_{T}-thresholds applied to the leptons and truth matching offline-reconstructed leptons for each trigger to maintain the trigger efficiency constant in p_{T} (Offline p_{T} threshold).

The name of the trigger encodes the requirements applied to the objects of the accepted events. The name "HLT Chain name" refers to the High Level Trigger described in Section 3.2.2. The first part of the name refers to the p_{T}-threshold applied to the light lepton ("e" for the electron, "mu" for muon), e.g. mu20, e24. This is followed by a "lh-" or "i-/ivar-" combined with *medium*, *loose* or *tight*, indicating the identification (lh-) or isolation (i-) criteria and the corresponding working point. The "nod0" expression means that there is no condition applied on the distance between the primary vertex and the observed track. Some trigger names contain "L1", which means that the HLT was seeded by a non-default L1 trigger decision from the ECAL "EM" or MS "MU", followed by the p_{T} threshold applied. "VH" indicates that this threshold depends on η and that a veto on energy depositions in the HCAL is applied.

Trigger	Data-taking period	HLT Chain name	HLT p_{T} threshold	Offline p_{T} threshold
Single electron	2015	`e24_lhmedium_L1EM20VH`, `e60_lhmedium` or `e120_lhloose`	$p_{\mathrm{T}}(e) > 24$ GeV	$p_{\mathrm{T}}(e) > 25$ GeV
	2016-2018	`e26_lhtight_nod0_ivarloose`, `e60_lhmedium_nod0` or `e140_lhloose_nod0`	$p_{\mathrm{T}}(e) > 26$ GeV	$p_{\mathrm{T}}(e) > 27$ GeV
Single muon	2015	`mu20_iloose_L1MU15` or `mu50`	$p_{\mathrm{T}}(\mu) > 20$ GeV	$p_{\mathrm{T}}(\mu) > 21$ GeV
	2016-2018	`mu26_ivarmedium` or `mu50`	$p_{\mathrm{T}}(\mu) > 26$ GeV	$p_{\mathrm{T}}(\mu) > 27.3$ GeV

Table 5.2.1: High level triggers (HLT) used in this analysis, divided into data-taking periods. Corresponding p_{T} thresholds applied at the trigger and offline stages are also shown.

5.3 Signal

Signal processes are simulated at NLO (Next-To-Leading Order) accuracy using POWHEG [56–59] generator and PYTHIA8 [60] for the simulation of the parton shower and non-perturbative effects. AZNLO [61] tune and the PDF4LHC15nlo [62] parton distribution function (PDF) set is used. The cross-section prediction is normalized to NNLO QCD accuracy that includes electroweak corrections [63–66].

5.4 Background

The same event generators from Ref.[66] are used. ATLAS detector simulation is done based on GEANT4 [67], that process all generated samples. Four main background production processes of the Higgs boson are ggF, VH, $t\bar{t}H$ and tH, as discussed in Section 2.2.

ggF is simulated at NNLO accuracy in QCD with POWHEG NNLOPS [56, 57, 59, 68, 69] using the PDF4LHC15 [62] NLO PDF set , and the AZNLO [61] tune of PYTHIA8 [60]. VH production samples are using the same setup as the signal processes. $t\bar{t}H$ events are generated at NLO accuracy with POWHEGBOX v2 [56, 57, 59, 70, 71] using the NNPDF3.0nlo PDF set [72] and interfaced with PYTHIA8.230 [60] e.g. using the A14 [73] tune. tH production samples are produced with MadGraph5_aMC@NLO 2.6.2 and interfaced with PYTHIA8 [60] for e.g. using the A14 tune [73] and CT10 PDF set [75].

W+jets and Z+jets production processes are simulated at NLO accuracy with Sherpa 2.2.1 [76] using the NNPDF3.0 [72] PDF set.

$t\bar{t}$ production processes are simulated at NNLO accuracy with POWHEGBOX v2 [56, 57, 59, 70, 71] and the NNPDF3.0 PDF set [72]. PYTHIA8.230 [60] is utilized for modeling the parton showering, hadronisation and underlying event activity, using the A14 [73] tune.

Single-top quark production samples are simulated at NLO accuracy with POWHEGBOX v2 [56, 57, 59, 70, 71] using the NNPDF3.0nlo [72] PDF set, and interfaced with Pythia8.230 [60] with the A14 [73] tune.

Diboson production processes (VV) are generated at NNLO accuracy with Sherpa v2.2.1 or Sherpa v2.2.2 [66, 76] using the NNPDF3.0 PDF set [72] and dedicated parton showering (PS) setup provided by the SHERPA authors.

Table 5.4.1 summarizes the generators, PDF sets and accuracy in QCD of simulated processes. The PS PDFs can be found in Ref.[66].

Process	ME Generator + PS	ME PDF set	Tune	Order in QCD
$H \to \tau\tau/WW^*$				
ggF	POWHEGBOX + PYTHIA8	PDF4LHC15	AZNLO	NNLO+NNL
VBF	POWHEGBOX + PYTHIA8	PDF4LHC15	AZNLO	NNLO
VH	POWHEGBOX + PYTHIA8	PDF4LHC15	AZNLO	NNLO
$t\bar{t}H$	POWHEGBOX + PYTHIA8	NNPDF3.0	A14	NNLO
tH	MadGraph5_aMC@NLO + PYTHIA8	CT10	A14	NLO
Background				
V+jets	Sherpa 2.2.1	NNPDF3.0	Sherpa	NNLO
$t\bar{t}$	POWHEGBOX + PYTHIA8	NNPDF3.0	A14	NNLO
Single top	POWHEGBOX + PYTHIA8	NNPDF3.0	A14	NLO
Diboson	Sherpa 2.2.1	NNPDF3.0	Sherpa	NNLO

Table 5.4.1: Overview of the signal and background event sample generators and their settings.

Chapter 6

Reconstruction and Identification of Physics Objects

This chapter defines and describes the reconstruction and identification methods of the objects used in this analysis: electrons in Section 6.1, muons in Section 6.2, hadronically decaying τ-leptons (τ_{had}) in Section 6.5, jets in Section 6.3 and neutrinos in Section 6.4. Finally, Section 6.7 describes the methods for reconstructing the invariant mass of the di-τ system.

The following concepts are used in this chapter: *Working Points* (WP) are specific requirements applied to the reconstructed objects, that define different efficiency and background rejection levels. *Baseline* objects are those which have not yet passed the overlap removal algorithm (OLR), discused in Section 6.6.

6.1 Electrons

Electron candidates in the ATLAS detector are reconstructed when tracks in the inner detector (ID) are matched to energy deposits in the electromagnetic calorimeter (ECAL). Depending on the physics process, the shower shape and track quality can vary [78]. Baseline electrons are required to have $p_{\text{T}} > 15$ and $|\eta| < 2.47$, excluding those in the $1.37 < |\eta| < 1.52$ transition region between the Barrel and the End-cap calorimeters of the ECAL.

Electron candidates passing the ORL are required to have *medium* likelihood-based identification WP [78]. The identification of the electron is done similarly as the one used for Run-1 but incorporating the improved cell clustering procedure [79]. *Tag-and-probe* [78] method is used to measure the efficiency of the identification selection. Identification efficiencies vary from 80% for the Tight WP to 93% for the Loose WP [79].

In this analysis *FCLoose* [79] isolation WP is applied to ORL candidates. This WP combines calorimeter- and track-based isolation variables. The isolation efficiency ranges from 90% and 99% for the Tight and Loose WPs, respectively. Reported identification and isolation efficiencies are based on studies conducted using 81 fb^{-1} of collision data from 2015-2017 period at $\sqrt{s} = 13$ TeV [79].

The reconstruction of the electron candidate is based on three properties of its detector signature; localized deposits of energy in the ECAL, tracking information from the ID, and a close match between the candidate that passes the Gaussian-sum filter [80] and the seeded energy cluster from the ECAL. This filter accounts for the energy losses in the material, (mostly bremsstrahlung in case of electrons),

The reconstruction efficiencies are in the range of $96 - 99\%$, based on studies using 2015-2016 data corresponding to an integrated luminosity of 36.7 fb^{-1} [78]. Similar values are found for the full Run-2 data set [79].

6.2 Muons

Muons leave the calorimeters undetected. They are reconstructed using the information from the muon spectometer (MS) and Inner Detector, as described in Section 3.2.2.3. Muon tracks are extrapolated from the MS to match with the interaction point, since external muon sources also leave signatures in the MS e.g. muons from cosmic background and cavern background radiation. Baseline muons are required to have $p_{\mathrm{T}} > 10$ GeV and $|\eta| < 2.47$. The reconstruction efficiencies of muons with $p_{\mathrm{T}} > 5$ GeV are around 99% in the $|\eta| < 2.5$ range according to the analysis of $J/\psi \to \mu\mu$ and $Z \to \mu\mu$ decays using 3.2 fb^{-1} of data [81].

Muons that pass the ORL are required to satisfy the *Loose* identification WP defined using a likelihood-based discriminant [81]. Tag-and-probe methods are used to measure the efficiency of the reconstruction and identification selections.

Muon candidates are also required to pass the *FCTightTrackOnly* [81] isolation criteria, which combines the information from the ID and ECAL to select prompt muons. The isolation efficiency varies between 93% and 100% depending on the momentum of the muon, based on measurements done using 3.2 fb^{-1} data set at $\sqrt{s} = 13$ TeV [81].

6.3 Jets

Hadrons originate from the hadronization process of partons. In a collision, there are bunches of hadrons traversing the detector, moving away from the original parton. A jet is an object that ideally captures and combines all hadrons originating from the same parton, representing a reconstruction challenge. In this analysis, jets are reconstructed using a particle-flow algorithm [83], using the anti-k$_t$ [84] algorithm with a distance parameter of $R = 0.4$, which is the standard jet reconstruction algorithm in ATLAS [84].

Jets energy calibration is done using global sequential (GS) corrections, multiplicative corrections applied to the energy measurement of the jets, derived from global jet features in the different ATLAS sub-detectors [87]. Baseline jets are required to have $p_{\mathrm{T}} > 20$ GeV, $|\eta| < 4.5$ and a LooseBad [66] quality criteria.

To reject pile-up jets and gain missing transverse energy resolution, a Jet vertex Tagger (JVT) tool [88] is utilized. The JVT *Tight* JVT [89] WP is used.

Jets that originate from b-hadrons are called b-jets and their identification is called *b-tagging.* In this analysis, b-jets are identified using the DL1r [90] b-tagging algorithm based on a deep-learning neural network that uses b-hadron features and vertex information from the inner detector. 85% b-tagging efficiency WP is used in this analysis.

6.4 Missing transverse energy

Neutrinos do not interact with the detector. Their contribution is indirectly estimated by measuring the energy imbalance in the transverse plane since the four-momentum is conserved in this plane. In practice, this is done by calculating the negative vectorial sum of the p_{T} of all recorded objects in the detector. This includes two contributions. First is the sum of transverse momenta of the reconstructed objects, called *hard term.* Second is the sum of transverse momenta of the tracks which are not associated with any of the reconstructed object but associated with the hard scatter vertex, called the *soft term* [91], that is

$$E_{\mathrm{T}}^{\mathrm{miss}} = - \sum_{i \in \{\text{objects from hard term}\}} \mathbf{p}_{\mathrm{T},i} - \sum_{k \in \{\text{objects from soft term}\}} \mathbf{p}_{\mathrm{T},k}. \tag{6.4.0}$$

The *Tight* WP defined in the official ATLAS Missing Transverse Energy Tool [92] is applied in this analysis.

6.5 Hadronic τ decays

The τ-lepton decays before reaching the innermost layer of the pixel detector due to its short decay length. Therefore, only its visible decay products can be reconstructed. These are: a light lepton (lep) (electron (e) or muon (μ)) and one or more hadrons in the $\tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ decay channel. The "τ reconstruction" here refers to the reconstruction of hadronically decaying taus using only its visible decay products (τ_{had}).

τ_{had} usually decays into one or three charged pions with up to two neutral pions, as illustrated in Figure 4.1.1. The reconstruction of the τ_{had} is done using the clustered energy deposits called *topo-clusters* [82] as inputs to the anti-k$_t$ [84] reconstruction with a radius parameter of $R = 0.4$. Baseline τ_{had} candidates are required to have one or three charged tracks, $p_{\mathrm{T}} > 20$ GeV and $|\eta| < 2.47$ excluding the $1.37 < |\eta| < 1.52$ transition region (as in the electron reconstruction). Only the highest-p_{T} τ_{had} candidate is kept and others are registered as jets.

The *medium* identification WP is applied to τ_{had} objects. Identification is based on a Recurrent Neural Network (RNN) [85] discriminant that is used to differentiate between jets originating from the τ_{had} and from partons [85]. This ID WP is exploited for estimating the background contribution from jets misidentified as τ_{had} described in chapter 8. Moreover, it was found that the RNN exceeds by more than 100% the performance on the background rejection compared to the previously Boosted Decision Tree (BDT) algorithm used for the identification of τ_{had} in the ATLAS detector [82, 86]. Furthermore, a *medium* WP derived from a BDT discriminant called eBDT [66] is used in the $\tau_e\tau_{\text{1-prong}}$ channel to suppress events where electrons are misidentified as τ_{had}. Such cases mainly originate from $Z \to ee+$ jets events [66].

6.6 Overlap Removal selection

The Overlap Removal (ORL) procedure is used to select the best assignment when two different objects are sharing most of their constituents. This can happen since the reconstructed objects used in this analysis are not constructed using disconnected calorimeter clusters or sets of tracks. The removal criteria is the same as the one used in Ref. [66], displayed in Table 6.1.

Object to remove	Object to keep	Criteria
Electron	Electron	if shared track, the electron with the highest-p_{T} is kept
τ_{had}	Electron	if $\Delta R < 0.2$ the electron is kept
τ_{had}	Muon	if $\Delta R < 0.2$ the muon is kept
Electron	Muon	if shared track, the electron is removed if the muon is associated with a signature in the MS, otherwise muon is removed
Jet	Electron	if jet within $\Delta R < 0.2$ of the electron is removed
Jet	Muon	if jet within $\Delta R < 0.2$ of the muon is removed if the tracks number< 3
Electron	Jet	electron within $\Delta R < 0.4$ of a jet is removed
Muon	Jet	muon within $\Delta R < 0.4$ of a jet is removed
Jet	τ_{had}	jet within $\Delta R < 0.2$ of a τ_{had} is removed

Table 6.1: Overlap Removal (ORL) selection based on the angular separation ΔR, identification , transverse momentum p_{T} or number of tracks. Criteria is listed in the order they are applied [66].

6.7 Invariant Mass reconstruction of the Higgs boson candidate

A precise reconstruction of the invariant mass of the di-τ system (Higgs boson candidate) is required. For this, three approaches are considered: the visible mass, the collinear approximation and the missing mass calculation.

The *visible mass* ($m_{\tau\tau}^{\mathrm{vis}}$), is reconstructed by considering only the visible decay products of the τ-leptons, that is

$$m_{\tau\tau}^{\mathrm{vis}} = \sqrt{\left(E_{\ell} + E_{\tau_{\mathrm{had}}}\right)^2 - \left(\vec{p}_{\ell} + \vec{p}_{\tau_{\mathrm{had}}}\right)^2}, \tag{6.7.1}$$

where $E_{\ell/\tau_{\mathrm{had}}}$ and $\vec{p}_{\ell/\tau_{\mathrm{had}}}$ are the energy and momentum of the light lepton (ℓ) or visible decay product of the τ_{had}, respectively.

The *collinear approximation* [93] is used to include the missing energy due to neutrinos in the reconstruction of the mass. Here, it is assumed that $E_{\mathrm{T}}^{\mathrm{miss}}$ arises exclusively from neutrinos and that τ-leptons are collinear (propagating in the same direction) to their corresponding decay products. The latter assumption leads to boosted τ-leptons, since the mass of the Higgs boson is much larger than the sum of the mass of two τ-leptons. In this approach, the *collinear mass* of the di-τ system is given by

$$m_{\mathrm{coll},\tau\tau} = \frac{m_{\mathrm{vis}}}{\sqrt{x_1 x_2}}, \tag{6.7.2}$$

where m_{vis} and $x_{1,2}$ are the mass and the momentum fractions of the visible decay products of the two τ-leptons, respectively. The momentum fractions are defined as

$$x_{1(2)} = \frac{p_{\mathrm{vis},2}^{x} p_{\mathrm{vis},1}^{y} - p_{\mathrm{vis},2}^{y} p_{\mathrm{vis},1}^{x}}{p_{\mathrm{vis},2}^{x} p_{\mathrm{vis},1}^{y} + (-) E_{\mathrm{T}}^{\mathrm{miss},x} p_{\mathrm{vis},1(2)}^{x} - p_{\mathrm{vis},2}^{y} p_{\mathrm{vis},1}^{x} - (+) E_{\mathrm{p}}^{\mathrm{miss},y} p_{\mathrm{vis},1(2)}^{x}}, \tag{6.7.3}$$

where $\vec{p}_{\mathrm{vis},1(2)} = x_{1(2)} \vec{p}_{\tau,1(2)}$. However, the collinear approximation has its limitations, since it does not consider the back-to-back production of the visible decay products of the di-τ system, situation in which Equation 6.7.3 can not be solved.

Another approach for the calculation of the invariant mass of the di-τ system is the *Missing Mass Calculator* [120] (MMC). This technique also assumes that the final state neutrinos are the unique source of missing energy, but in contrast to the collinear approach, it rejects the collinear assumption. In this method, the invariant mass of the di-τ system, called MMC *mass* ($m_{\tau\tau}^{\mathrm{MMC}}$), is estimated from the reconstruction of the unknown momenta of the neutrinos in the final state. These momenta are the solution of an underdetermined system of equations composed of the invariant masses of the τ-leptons and two components of the missing transverse energy. The system of equations is solved by performing a scan over a subset of the unknown variables, as described in Ref. [120].

Chapter 7

Event selection

This chapter describes the cut-based approach used to select VBF Higgs production processes decaying into a pair of τ-leptons, specifically, in the semileptonic channel. Section 7.1 describes the *Preselection*, which is a set of requirements applied to select the decay channel $H \to \tau_{\text{lep}}\tau_{\text{had}}$ and suppress background processes. Section 7.2 the *VBF selection*, applied to select the eponymous production mode of the Higgs boson. The region defined with both Preselection and VBF selections applied is called the VBF signal region (SR). Table 7.1 summarizes the selection criteria of the Preselection and VBF categories. The number of data and simulation yields in the Preselection and VBF selection regions can be found in Section 7.3.

Region	Requirement
Preselection	One light lepton (ℓ): electron (e) or muon (μ) and one hadronically decaying τ-lepton (τ_{had})
	$p_{\text{T}}^{\tau_{\text{had}}} > 30$ GeV
	Identification of ℓ: Medium , τ_{had}: RNN Medium
	e-veto in the $\tau_e\tau_{\text{1-prong}}$ channel (eleBDT)
	Isolation of e: FCLoose and μ: FCTightTrackOnly
	Opposite charged ℓ and τ_{had} (OS)
	$m_{\text{T}}(\ell, E_{\text{T}}^{\text{miss}}) < 70$ GeV
	b-jet veto: DL1r with 85% efficiency working point
	$E_{\text{T}}^{\text{miss}} > 20$ GeV
	Leading jet $p_{\text{T}} > 40$ GeV
	$\Delta R_{\ell,\tau_{\text{had}}} < 2.5$
	$\lvert\Delta\eta_{\ell,\tau_{\text{had}}}\rvert < 1.5$
	$0.1 < x_1 < 1.4$ and $0.1 < x_2 < 1.2$
VBF	Two jets
	Sub-leading jet $p_{\text{T}} > 30$ GeV
	$m_{jj} > 350$ GeV
	$\lvert\Delta\eta_{jj}\rvert > 3$
	$\eta_{j_1} \times \eta_{j_2} < 0$

Table 7.1: Summary of the criteria applied on the $H \to \tau_{\text{lep}}\tau_{\text{had}}$ Preselection and VBF categories. Here, p_{T} is the transverse momentum, $m_{\text{T}}(\ell, E_{\text{T}^{\text{miss}}})$ the transverse mass of the light lepton and missing transverse energy system. $\Delta R_{\ell,\tau_{\text{had}}}$ and $\lvert\Delta\eta_{\ell,\tau_{\text{had}}}\rvert$ are the angular distance and absolute value of the pseudorapidity, respectively between the two τ-candidates.

7.1 Preselection

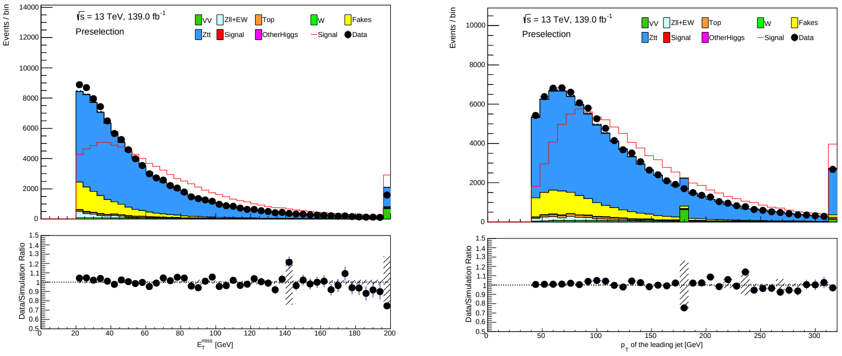
Preselection criteria targets the $H \to \tau_{\text{lep}}\tau_{\text{had}}$ decay mode. Hence, events must have one light lepton (ℓ): electron (e) or muon (μ), and one hadronically decaying τ-lepton (τ_{had}). In order to suppress background events from W+jets and top-quark production processes, the charge of the light lepton is required to be opposite to the charge of the τ_{had}. Further suppression of W+jets background processes is obtained by applying an upper limit of 70 GeV on the transverse mass of the light lepton and the missing transverse energy ($E_{\text{T}}^{\text{miss}}$) system, which is defined as

$$m_{\text{T}}(\ell, E_{\text{T}}^{\text{miss}}) = \sqrt{2p_{\text{T}}^{\ell}E_{\text{T}}^{\text{miss}}\left(1 - \cos\Delta\phi_{\ell,E_{\text{T}}^{\text{miss}}}\right)}. \tag{7.1.1}$$

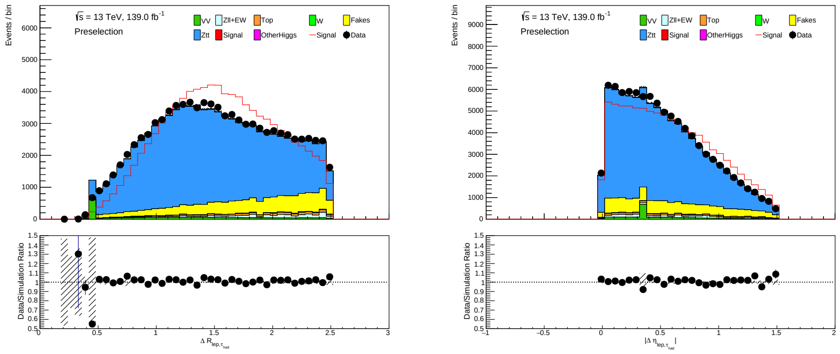
Here, $\Delta\phi_{\ell,E_{\text{T}}^{\text{miss}}}$ is the azimuthal angle between the light lepton and $E_{\text{T}}^{\text{miss}}$, and p_{T}^{ℓ} the transverse momentum of the light lepton.

To suppress top-quark production processes, a b-jet veto is applied. Also, the light lepton and the τ_{had} are required to pass their respective *medium* identification (ID) criteria (detailed in chapter 6). The isolation requirement for the light lepton is different depending on whether it is an electron (*FCLoose*) or a muon (*FCTightTrackOnly*), described in chapter 6. Additional angular selections are applied such as $\Delta R_{\ell,\tau_{\text{had}}} < 2.5$ and $|\eta_{\ell,\tau_{\text{had}}}| < 1.5$ between the light lepton and τ_{had}. Specifically in the $\tau_e\tau_{\text{1-prong}}$ channel, τ_{had} candidate is required to satisfy an electron veto, defined using a BDT-based algorithm called *eBDT* [66] to suppress background processes from Z+jets$\to$ e^+e^- events where one of the electrons could be misidentified as a hadronically decaying τ-lepton. Transverse momentum of the τ_{had} is required be larger than 30 GeV. Furthermore, in order to reduce W+jets and Z$\to$ $\tau\tau$ background events, the transverse momentum of the leading jet ($p_{\text{T}}^{j_1}$) is required to be larger than 40 GeV. Also, a selection is applied to the missing transverse energy: $E_{\text{T}}^{\text{miss}} > 20$ GeV. In order to reduce background events where the $E_{\text{T}}^{\text{miss}}$, light lepton and τ_{had} do not originate from the same source; it is also required that the momentum fractions carried by the τ_{had} are constrained by $0.1 < x_1 < 1.4$ and $0.1 < x_2 < 1.2$.

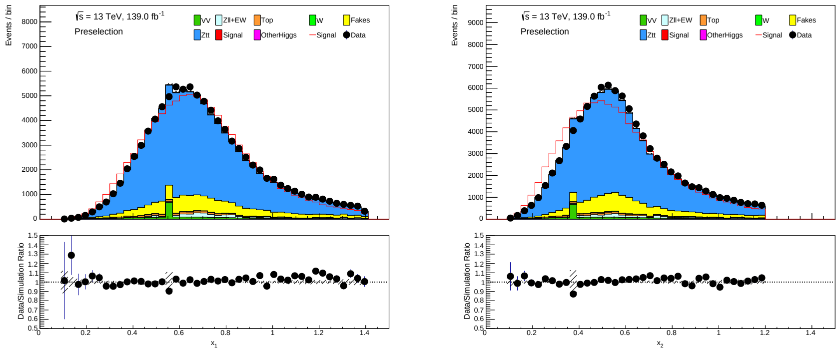
Figures 7.1.2 and 7.1.1 show the kinematic distributions after the Preselection. Fluctuations seen around some of distributions at are not present after VBF selection and not seen in other regions.



(a) E_{T}^{miss} after Preselection (b) p_{T} of leading jet after Preselection

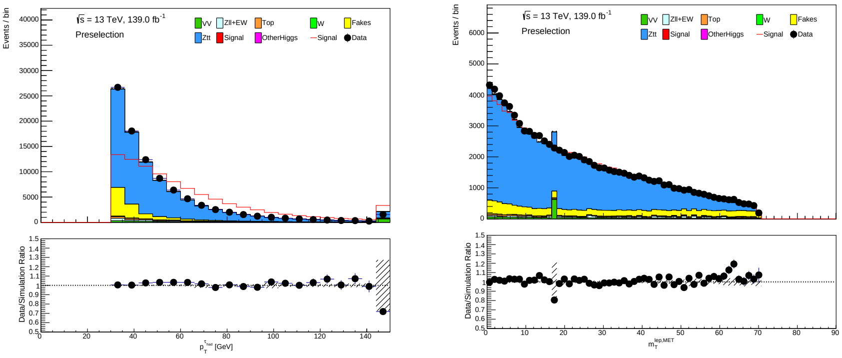


(c) $\Delta R_{\ell,\tau_{\mathrm{had}}}$ after Preselection (d) $|\Delta\eta_{\ell,\tau_{\mathrm{had}}}|$ after Preselection



(e) x_1 after Preselection (f) x_2 after Preselection

Figure 7.1.1: Various kinematic distributions in the Preselection category. In (a) the missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$), (b) the transverse momentum (p_{T}) of the leading jet, (c) the angular distance between the two τ-candidates ($\Delta R_{\ell,\tau_{\mathrm{had}}}$), (d) the absolute value difference in pseudorapidity (η) between the two τ-candidates ($|\Delta\eta_{\ell,\tau_{\mathrm{had}}}|$), and (e) and (f) the momentum fractions carried by the visible decay products (x_1 and x_2) as defined in Equation 6.7.3. Overflow bins are included in Figures (a), (b), (c) and (d). Only statistical uncertainties are shown. "Signal" line represents the signal contribution normalized to the total background. Fluctuations seen around the middle of the distributions at Figures (b), (d), (e) and (f) are not present after VBF selection and not seen in other regions.



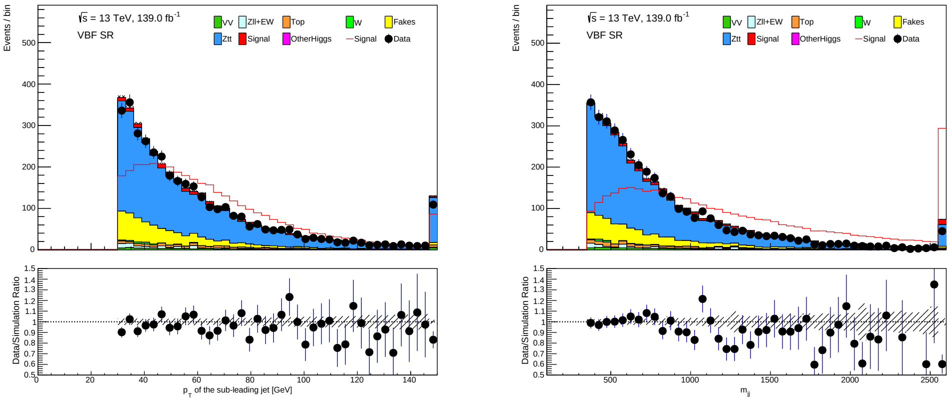
(a) p_{T} of the τ_{had} after Preselection

(b) $m_{\mathrm{T}}^{\ell,MET}$ after Preselection

Figure 7.1.2: Kinematic distributions in the Preselection category. In (a) the transverse momentum (p_{T}) of the τ_{had} ($p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$), and (b) the transverse mass between the light lepton (ℓ) and the missing transverse energy (MET) are displayed ($m_{\mathrm{T}}^{\ell,MET}$). Only statistical uncertainties are shown. "Signal" line represents the signal contribution normalized to the total background. Overflow bin are included in (a). Fluctuations seen in (b) is not present after VBF selection and not seen in other regions.

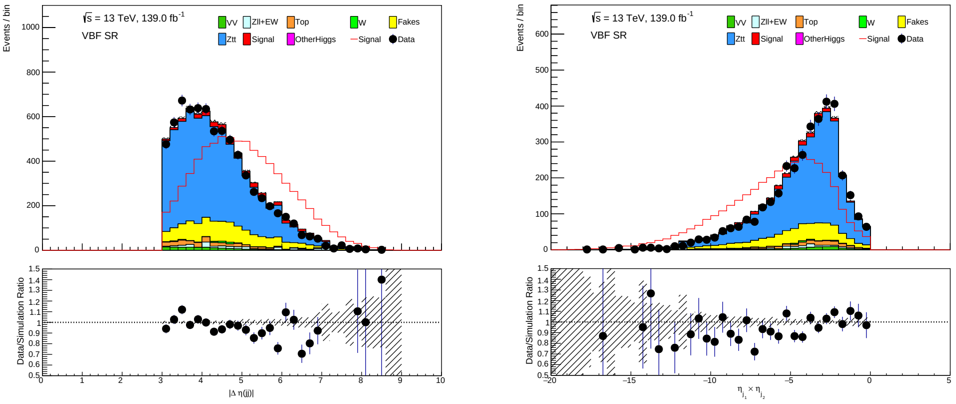
7.2 VBF region

To enhance the VBF Higgs production processes, further selections are applied in addition to Preselection. Two main background sources should be taken into account: Z→ $\tau\tau$ and ggF processes. *VBF selection* is chosen such that it exploits the kinematic differences between the two leading jets from the above-mentioned background processes. Therefore, exactly two jets are required with the transverse momentum of the sub-leading jet (j_2) being larger than 30 GeV. Moreover, a large pseudorapidity gap between the two leading jets is required, with $|\eta_{jj}| > 3$, which enhances the sensitivity to the CP-structure [95]. Furthermore, the two jets are required to be in opposite hemispheres ($\eta_{j_1} \times \eta_{j_2} < 0$). Finally, the invariant mass of the di-jet system (m_{jj}) has to be larger than 350 GeV. Figure 7.2.1 shows the kinematic distributions after the Preselection and VBF selections. The final VBF signal region is defined with the combined application of both: Preselection and VBF selections, and is referred to as "VBF SR".



(a) p_T of the sub-leading jet in the VBF SR

(b) m_{jj} in the VBF SR



(c) $|\Delta\eta_{jj}|$ in the VBF SR

(d) $\eta_{j_1} \times \eta_{j_2}$ in the VBF SR

Figure 7.2.1: Various kinematic distributions in the VBF SR. In (a) the transverse momentum, (p_T) of the sub-leading jet, (b) the mass of the di-jet system (m_{jj}), (c) the absolute value of the difference in pseudorapidity (η) between the two leading jets ($|\Delta\eta_{jj}|$) and (d) the product of the pseudorapidities of the two leading jets defined with the subindexes 1 and 2 ($\eta_{j_1} \times \eta_{j_2}$). Only statistical uncertainties are shown. "Signal" line represents the signal contribution normalized to the total background. Overflow bins are shown in Figures (a) and (b).

7.3 Event yields

Total number of data and simulation events in the Preselection and VBF SR are shown in Table 7.3.1. Signal events are divided into two contributions: VBF and VBFHWW. For backgrounds, no distinction is made and only the total yield is given. Percentage of events in each signal category relative to the total event yield is also shown. The signal yield is enhanced in the VBF SR with respect to the Preselection category.

		Signal				Proportion [%] of		
Category	Data	all	VBF	VBFHWW	Background	Signal	VBFHWW	VBF
Preselection	93532	334 ± 1	332 ± 1	1.42 ± 0.04	92594 ± 622	0.36	0.0015	0.35
VBF SR	3669	162 ± 2	161 ± 2	0.70 ± 0.03	3651 ± 23	4.42	0.019	4.39

Table 7.3.1: Number of data and simulated events in the Preselection and VBF SR, defined by applying the selections summarized in Table 7.1.

Chapter 8

Background contributions from jets misidentified as τ_{had}

This chapter shows the strategy followed to estimate the background contribution from jets misidentified as hadronically decaying τ-leptons, "fakes". Section 8.1 describes the method utilized to estimate this background processes, called the *Fake Factor* method. In Section 8.2 some of the main kinematic distributions containing the fake background for each decay channel are shown. Section 8.3 validates the method. Uncertainties are discussed in Section 8.4. Finally, Section 8.5 discusses some additional studies.

8.1 The Fake Factor method

The reconstruction method of the τ-leptons was introduced in Section 6.5. Due to its short decay length only the visible decay products of the τ-leptons can be reconstructed. As discussed in Section 4.2, one of the main sources that can mimic the signal process are the events where jets from the background processes get misidentified as the visible decay products of the hadronically decaying τ-lepton (τ_{had}). These misidentified events generally originate from W+jets or multi-jet processes, as will be discussed further in this chapter.

Generally, the τ_{had} candidates are reconstructed as narrow calorimetric showers with low track multiplicity (1 or 3 prongs), in contrast to the jets originated from background processes, which are relatively wide showers. This fact can be exploited to estimate the contribution of such misidentified events. From now on, the jets misidentified as hadronically decaying tau leptons will be called "fakes".

The estimation of fakes using Monte Carlo (MC) event generators is in general difficult [96, 97]. Furthermore, modeling appears to be poor due to statistical limitations in the number of simulated events, and the systematic uncertainties coming from misidentified objects are not known [96]. Thus, a data-driven method is preferred.

As discussed in Section 6.5, the identification (ID) of τ_{had} is constructed using a recurrent neural network (RNN) that uses information from the track and the calorimeter to differentiate between true τ_{had} and misidentified τ_{had} coming from quark and gluon-initiated jets [85]. This is the current default algorithm used in the ATLAS experiment to identify τ_{had} for the Full Run 2 data set, in contrast to the previously used Boosted Decision Tree (BDT) algorithm [85, 98].

A region enriched in fakes is needed in order to study them. Therefore a control region (CR) is defined. This CR is called "anti-τ CR" and has the same requirements as the VBF signal region (SR) defined in chapter 7 except for inverting the ID requirement applied on τ_{had}: the candidate has to fail the *medium* ID working point (see Table 7.1). Furthermore, the RNN from which the

ID working points are defined assigns low $\mathrm{NN_{score}}$ values (measurement of the probability that an event is classified as a particular process) to the τ_{had} candidates arising from gluon-initiated jets, however, for $\mathrm{NN_{score}} < 0.01$ the distribution of the RNN changes significantly [85]. Thus, a selection of $\mathrm{NN_{score}} > 0.01$ is further applied in order to obtain a similar quark-gluon ratio in the anti-τ CR as in the VBF SR.

The number of fake events in the VBF SR (sometimes simplified to SR in this chapter) ($\mathrm{N^{SR}_{fakes}}$) is calculated in this anti-τ region and then transferred using a transfer factor called the "Fake Factor" ($\mathcal{F}$), that is

$$\mathrm{N^{SR}_{fakes}} = \left(\mathrm{N^{anti\text{-}\tau}_{Data}} - \mathrm{N^{anti\text{-}\tau}_{MC,\ not\ j\rightarrow\tau}}\right) \cdot \mathcal{F}, \tag{8.1.1}$$

where $\mathrm{N^{anti-\tau}_{Data}}$ is the data in the anti$-\tau$ CR and $\mathrm{N^{anti-\tau}_{MC,\ not\mathit{j}\rightarrow\tau}}$ is the number of MC simulated events where a jet is not misidentified as a τ_{had}. However, as mentioned, the collimation of the jets is highly dependant on whether the jet was initiated by a quark or a gluon. Thus, these two contributions to fake events have to be taken into account separately, which mostly come from W+jets (W) events, multi-jet (QCD), top-quark (Top) and Z+jets (Z) production processes [11, 66, 99]. For this reason $\mathcal{F}$ is renamed the "combined Fake Factor" and is defined as:

$$\mathcal{F} = \mathrm{R_W F_W} + \mathrm{R_Z F_Z} + \mathrm{R_{Top} F_{Top}} + \mathrm{R_{QCD} F_{QCD}}. \tag{8.1.2}$$

Here, F_i represents the individual Fake Factor of the process i and R_i the fractional contribution of the process i to the anti-τ CR.

Due to the small contribution of Z + jets and top quark production to the total background, as shown in Table 8.1.3.2, they are expected to have similar individual fake factors as W + jet processes (this assumption is validated in Section 8.3.1). Considering this and based on previous analyses ([11, 66, 99]), the Fake Factor method is simplified assuming

$$\mathrm{R_W} = \mathrm{R_W} + \mathrm{R_Z} + \mathrm{R_{Top}}. \tag{8.1.3}$$

Therefore, Equation 8.1.2 can be rewritten as

$$\mathcal{F} = \mathrm{R_W F_W} + \mathrm{R_{QCD} F_{QCD}}. \tag{8.1.4}$$

This way, $\mathcal{F}$ can be calculated by obtaining $\mathrm{R_{QCD}}$ and the individual Fake Factors $\mathrm{F_W}$ and $\mathrm{F_{QCD}}$ since $\mathrm{R_W} = 1 - \mathrm{R_{QCD}}$.

8.1.1 Individual Fake Factors

The individual Fake Factor of a particular process i has to be determined in a dedicated CR. This CR must be defined so that it is enriched in events of the process i but with properties similar to those of the VBF SR, as it was assumed that fakes calculated in the anti-τ CR can be transferred to the VBF SR. Thus, the CR of the process i will be defined with the same requirements as the VBF SR, except inverting the selection that was used to suppress the process i, further information can be found in chapter 7. In this way, it is ensured that all CRs defined during the Fake Factor method application are orthogonal.

Based on the two dominant processes involved in the estimation of the fakes, two regions are defined i.e. W+jets CR (W CR) and the multi-jet CR (QCD CR). The W CR is defined by inverting the requirement on the transverse mass of the light lepton (lep) and $E_{\mathrm{T}}^{\mathrm{miss}}$ system ($\mathrm{m_T}(\mathrm{lep}, E_{\mathrm{T}}^{\mathrm{miss}})$) since this was used to remove W+jets background from the VBF SR. Similarly, the QCD CR is defined by inverting the requirement to have one isolated light lepton, since QCD events are not

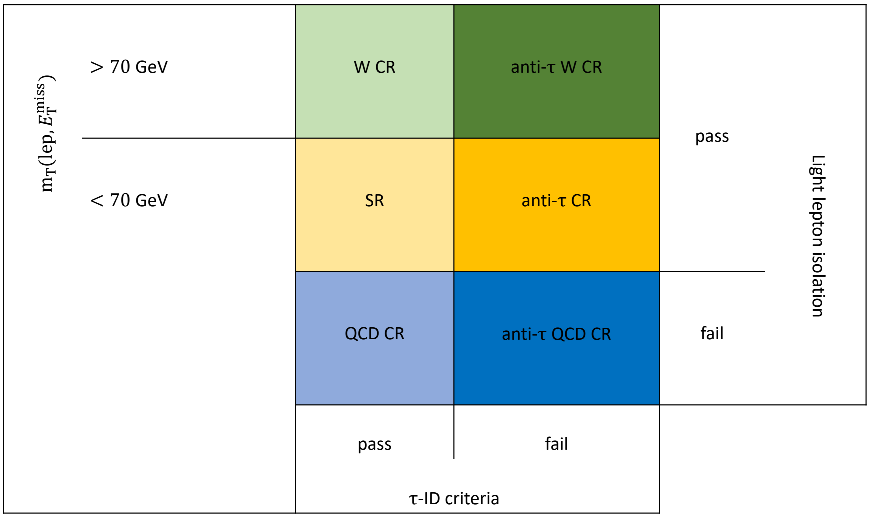


Figure 8.1.1.1: Different regions defined for the calculation of the individual Fake Factors. These are, the VBF signal region (SR), the anti-τ CR (defined in Section 8.1), the CRs and anti-τ CRs for the W+jets (W) and multi-jet (QCD) processes. $m_{\mathrm{T}}(\mathrm{lep}, E_{\mathrm{T}}^{\mathrm{miss}})$ refers to the transverse mass between the light lepton (lep) and the missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$).

expected to have isolated leptons. The number of expected events in these regions are given in Table A.1.1. An overview of the different CRs and their definitions is shown in Figure 8.1.1.1.

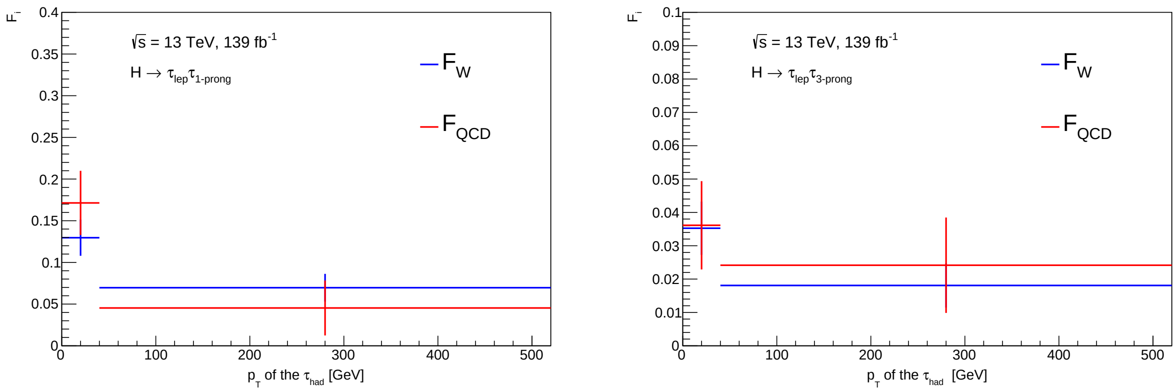
Each CR is further split into two parts depending on whether the τ_{had} candidate passes or fails (anti-τ) the ID criteria, as shown in Figure 8.1.1.1. Individual Fake Factor F_i is then calculated as the ratio of data events in these two regions:

$$\mathrm{F}_i = \frac{\mathrm{N}_{\mathrm{Data}}^{\mathrm{CR}_i} - \mathrm{N}_{\mathrm{MC,\ not\ j}\to\tau}^{\mathrm{CR}_i}}{\mathrm{N}_{\mathrm{Data}}^{\mathrm{anti}\text{-}\tau,\mathrm{CR}_i} - \mathrm{N}_{\mathrm{MC,\ not\ j}\to\tau}^{\mathrm{anti}\text{-}\tau,\ \mathrm{CR}_i}}, \tag{8.1.1.1}$$

where $\mathrm{N}_{\mathrm{Data}}^{\mathrm{CR}_i}$ and $\mathrm{N}_{\mathrm{Data}}^{\mathrm{anti,CR}_i}$ are the number of data in the region that passes or fails the τ-lepton ID criteria, respectively, and $\mathrm{N}_{\mathrm{MC,\ not}j\to\tau}^{\mathrm{CR}_i}$ and $\mathrm{N}_{\mathrm{MC,\ not}j\to\tau}^{\mathrm{anti}\text{-}\tau\mathrm{CR}_i}$ are the number of MC events in the region that passes or fails the τ-lepton ID criteria, respectively in which the τ_{had} does not originate from a jet.

The results of the individual Fake Factors as a function of the transverse momentum of the τ_{had} ($p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$), split by the number of charged tracks it has (1-prong or 3-prong), are shown in Figure 8.1.1.2 and tabulated in Table 8.1.1.1.

Due to low statistics in the high transverse momentum region (especially in the QCD CR), the individual Fake Factors are divided into two bins only. The 1-prong channel yields to larger individual Fake Factors than for the 3-prong channel, whereas no major differences are observed between the individual Fake Factors of the processes involved. However, there is an observable dependance in the $p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$: for low $p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$ values, the individual Fake Factors are higher, and for high $p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$ values, the individual Fake Factors are smaller.



(a) F_i in the 1-prong channel

(b) F_i in the 3-prong channel

Figure 8.1.1.2: Individual Fake Factors (F_i) in (a) 1-prong and (b) 3-prong channels. The individual Fake Factors corresponding to W+jets (F_W) are shown in blue and multi-jet (QCD) processes (F_{QCD}) in red for different values of the hadronic τ-lepton transverse momentum (p_T of the τ_{had}). Only statistical uncertainties are shown.

Channel	$p_T^{\tau_{had}}$ [GeV]	$F_W \times 10^2$	$F_{QCD} \times 10^2$
1-prong	< 40	13 ± 2.1	17 ± 3.8
	> 40	7.0 ± 1.7	4.5 ± 3.3
3-prong	< 40	3.5 ± 0.8	3.6 ± 1.3
	> 40	1.8 ± 0.6	2.4 ± 1.4

Table 8.1.1.1: Individual Fake Factors from W+jets (F_W) and QCD (F_{QCD}) processes in the 1-prong and 3-prong channels for different ranges of the transverse momentum of the τ_{had} ($p_T^{\tau_{had}}$).

8.1.2 Fractional contribution from multi-jet processes

The fraction of multi-jet events in the anti-τ region (R_{QCD}) is calculated as

$$R_{QCD} = \frac{N^{\text{anti-}\tau}_{\text{QCD, Data}}}{N^{\text{anti-}\tau}_{\text{Data}} - N^{\text{anti-}\tau}_{\text{MC, not j}\to\tau}}, \tag{8.1.2.1}$$

where $N^{\text{anti-}\tau}_{\text{QCD, Data}}$ is the number of multi-jet data in the anti-τ CR, and $N^{\text{anti-}\tau}_{\text{Data/MC}}$ is the total number of data/MC simulated events again in the anti-τ region. The denominator of Equation 8.1.2.1 can be determined directly from the yields in this region. However the QCD CR will be dominated by events in which a jet is misidentified as a light lepton. These "new" fakes are estimated using a data-driven method, which is applied to determine the numerator of Equation 8.1.2.1,

$$N^{\text{anti-}\tau}_{\text{QCD, Data}} = (N^{\text{QCD anti-}\tau\text{CR}}_{\text{Data}} - N^{\text{QCD anti-}\tau\text{CR}}_{\text{MC,true lep}}) \cdot \mathcal{I}. \tag{8.1.2.2}$$

Here, $\mathrm{N}_{\mathrm{Data}}^{\mathrm{QCD\ anti\text{-}\tau CR}}$ is the number of data in the QCD anti-τ CR, $\mathrm{N}_{\mathrm{MC,true\ lep}}^{\mathrm{QCD\ anti\text{-}\tau CR}}$ the yield of MC simulated events with a true leading lepton in the QCD anti-τ CR, and $\mathcal{I}$ is a transfer factor that transfers the number of QCD events in the QCD anti-τ CR to the anti-τ CR, called *Isolation Factor*. The difference between the events extracted in Equations 8.1.2.2, 8.1.1 and 8.1.1.1 is the following: in Equations 8.1.1 and 8.1.1.1, the substracted MC events were required to fulfill the truth matching condition for the τ_{had} since fake events originate from the misidentification of τ_{had}. However, in Equation 8.1.2.2 the subtracted events have to fulfill the truth matching condition for the light lepton since those fake events originate from the misidentification of the lepton and not of the τ_{had}.

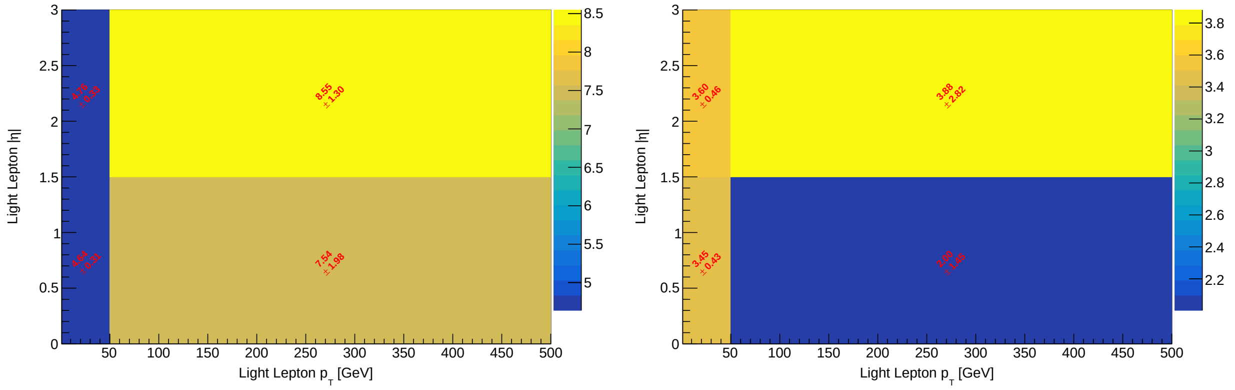
Isolation Factors

The Isolation Factors are calculated in dedicated CRs. They are defined with a similar selection to the QCD anti-τ and anti-τ CRs. In order to create orthogonal CRs, the opposite-charge condition between τ_{had} and the light lepton (lep) is inverted (OS). Since same charge events are rare in this selection, SR requirements are loosened to the Preselection level that was defined in Section 7.1. The resulting CR is named as same-sign Preselection (SSP) CR. This region is then further split into "iso" and "non-iso" CRs depending on whether the light lepton passes or fails the isolation requirement defined for the SR, respectively. The Isolation Factors are defined as the ratio between the number of data events in these two regions and calculated as

$$\mathcal{I} = \frac{\mathrm{N}_{\mathrm{Data}}^{\mathrm{iso,\ SSP}} - \mathrm{N}_{\mathrm{MC,\ true\ lepton}}^{\mathrm{iso,\ SSP}}}{\mathrm{N}_{\mathrm{Data}}^{\mathrm{non\text{-}iso,\ SSP}} - \mathrm{N}_{\mathrm{MC,\ true\ lepton}}^{\mathrm{non\text{-}iso,\ SSP}}}, \tag{8.1.2.3}$$

where $\mathrm{N}_{\mathrm{Data}}^{\mathrm{iso/non\text{-}iso,\ SSP}}$ and $\mathrm{N}_{\mathrm{MC,\ true\ lepton}}^{\mathrm{iso/non\text{-}iso,\ SSP}}$ are the number of data and MC simulated event yields with true leptons, respectively. "iso" and "non-iso" superscripts indicate the SSP CR subregion the values are calculated in. Isolation Factors are binned in the p_{T} and $|\eta|$ of the light lepton and are calculated separately for $\tau_{\mathrm{lep}}\tau_{\text{1-prong}}$ and $\tau_{\mathrm{lep}}\tau_{\text{3-prong}}$ final states. The reason for this is explained in Section 8.5.1. Furthermore, the assumption that $\mathcal{I}$ calculated in the SSP CR can be applied to the OS SR is validated in Section 8.3.2. Obtained Isolation Factors for different kinematic regions are shown in Figure 8.1.2.1 and listed in Table 8.1.2.1. Event yields from the CRs used in the calculation of the Isolation Factors are provided in Appendix A.

Similar to the individual Fake Factors, Isolation Factors are also found to be larger for the 1-prong channel than for the 3-prong channel. Isolation Factors increase with $|\eta|$ in both cases, but only in the 1-prong channel with the transverse momentum. The dependence of the Isolation Factors on p_{T} is discussed further in Section 8.5.1.



(a) $\mathcal{I}$ in the $\tau_{\text{lep}}\tau_{\text{1-prong}}$ channel

(b) $\mathcal{I}$ in the $\tau_{\text{lep}}\tau_{\text{3-prong}}$ channel

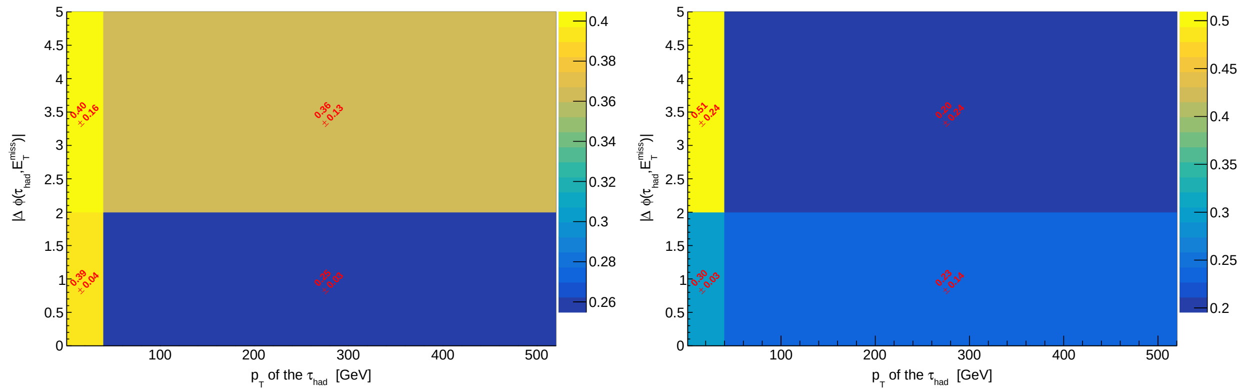
Figure 8.1.2.1: Isolation Factors ($\mathcal{I}$) for the (a) $\tau_{\text{lep}}\tau_{\text{1-prong}}$ and (b) $\tau_{\text{lep}}\tau_{\text{3-prong}}$ channels as functions of p_{T} and $|\eta|$ of the light lepton, calculated in the same-sign Preselection (SSP) anti-τ CR as detailed in Section 8.1.2. Statistical and systematic uncertainties are shown.

Channel	$p_{\text{T}}^{\text{lep}}$ [GeV]	$\lvert\eta^{\text{lep}}\rvert$	$\mathcal{I}$
1-prong	< 50	< 1.5	4.64 ± 0.31
	< 50	> 1.5	4.76 ± 0.33
	> 50	< 1.5	7.54 ± 1.98
	> 50	> 1.5	8.55 ± 1.30
3-prong	< 50	< 1.5	3.45 ± 0.43
	< 50	> 1.5	3.60 ± 0.46
	> 50	< 1.5	2.00 ± 1.45
	> 50	> 1.5	3.88 ± 2.82

Table 8.1.2.1: Isolation Factors ($\mathcal{I}$) split according to the charged track multiplicity of the τ_{had} decay (1-prong and 3-prong channels), p_{T} and $|\eta|$ of the light lepton (lep). Statistical and systematic uncertainties are shown.

Fractional Contribution of multi-jet events

Once the Isolation factors are calculated, the fractional contribution from multi-jet events to the anti-τ region ($\mathrm{R_{QCD}}$) can be estimated using Equation 8.1.2.1. $\mathrm{R_{QCD}}$ is binned in two dimensions using $p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$ and $|\Delta\phi(\tau_{\mathrm{had}}, E_{\mathrm{T}}^{\mathrm{miss}})|$, the absolute value of the azimuthal angle between the τ_{had} candidate and the missing transverse energy. Figure 8.1.2.2 shows the $\mathrm{R_{QCD}}$ values in 1- and 3-prong channels. In general, $\mathrm{R_{QCD}}$ values are found to be higher for the 1-prong channel compared to the 3-prong channel. In addition, since $\mathrm{R_W} = 1 - \mathrm{R_{QCD}}$, all categories have a higher W+jets contribution to the anti-τ CR, particularly in the 3-prong channel.



(a) $\mathrm{R_{QCD}}$ in the 1-prong channel

(b) $\mathrm{R_{QCD}}$ in the 3-prong channel

Figure 8.1.2.2: Fractional contribution of multi-jet events ($\mathrm{R_{QCD}}$) in the (a) 1-prong ($\tau_{\mathrm{lep}}\tau_{\text{1-prong}}$) and (b) 3-prong ($\tau_{\mathrm{lep}}\tau_{\text{3-prong}}$) channels. $\mathrm{R_{QCD}}$ values are binned using the transverse momentum of the τ_{had}, (p_{T} of the τ_{had}), and the absolute value of the azimuthal angle between the τ_{had} and the missing transverse energy ($|\Delta\phi(\tau_{\mathrm{had}}, E_{\mathrm{T}}^{\mathrm{miss}})|$). Statistical and systematic uncertainties are shown.

Channel	$p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$ [GeV]	$\lvert\Delta\phi(\tau_{\mathrm{had}}, E_{\mathrm{T}}^{\mathrm{miss}})\rvert$	$\mathrm{R_{QCD}} \times 10$
1-prong	< 40	< 2	3.9 ± 0.4
	< 40	> 2	4.0 ± 0.6
	> 40	< 2	2.5 ± 0.3
	> 40	> 2	3.6 ± 1.3
3-prong	< 40	< 2	3.0 ± 0.3
	< 40	> 2	5.1 ± 2.4
	> 40	< 2	2.3 ± 1.4
	> 40	> 2	2.0 ± 2.4

Table 8.1.2.2: Fractional contributions from multi-jet processes ($\mathrm{R_{QCD}}$) to the anti-τ CR calculated separately for four sub-regions that are defined by the requirements on the transverse momentum of the τ_{had} ($p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$), and the absolute value of the azimuthal angle between τ_{had} and the missing transverse energy ($\lvert\Delta\phi(\tau_{\mathrm{had}}, E_{\mathrm{T}}^{\mathrm{miss}})\rvert$). Values are evaluated in the 1-prong and 3-prong channels. Statistical and systematic uncertainties are shown.

8.1.3 Combined Fake Factors

The combined Fake Factors ($\mathcal{F}$) are binned in two dimensional $p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$–$\lvert\Delta\phi(\tau_{\mathrm{had}}, E_{\mathrm{T}}^{\mathrm{miss}})\rvert$ plane and separately calculated for 1 and 3-prong channels. The values of the combined Fake Factors are listed in Table 8.1.3.1. Despite a low dependency on the $\lvert\Delta\phi(\tau_{\mathrm{had}}, E_{\mathrm{T}}^{\mathrm{miss}})\rvert$, its addition provides more sensitivity to the analysis. In contrast, when the values of the combined Fake Factors split by the number of tracks and the $p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$ are compared, larger discrepancies are seen. These differences mainly come from the individual Fake Factors that are parametrized by these two variables. Since $\mathrm{F_W}$ and $\mathrm{F_{QCD}}$ did not show large discrepancies (see Table 8.1.1.1), their respective fractional contributions $\mathrm{R_W}$ and $\mathrm{R_{QCD}}$ does not affect the combined Fake Factors significantly. This can be seen by comparing the values of the individual Fake Factors from Table 8.1.1.1 with the combined Fake Factors in Table 8.1.3.1 for the same track multiplicity and $p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$ regions.

The combined Fake Factors show a trend of larger values in the low $p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$ region. This is inherited from the individual Fake Factors. They are also higher for the 1-prong case compared to the 3-prong case, a result that can also be traced back to the individual Fake Factors.

Once the combined Fake Factors are obtained, the final estimation of fake background events in the VBF SR is made using Equation 8.1.1. Event yields in the VBF SR, including the fake background contribution are displayed in Table 8.1.3.2 for the four different channels. In Table 8.1.3.3 the background(yields)-to-data ratio in each of the channels is shown, after the inclusion of the fake background contribution. Larger disagreement in the $\tau_{\mathrm{lep}}\tau_{\text{3-prong}}$ channel is seen as expected since this is the region with lowest statistics (leading to larger fluctuations). The results after the inclusion of the Fake background estimate provide good agreement between the data and the simulation. The validation of the full method is discussed in Section 8.3. A schematic summary of the Fake Factor method is shown in Figure 8.1.3.1.

Channel	$p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$ [GeV]	$\lvert\Delta\phi(\tau_{\mathrm{had}}, E_{\mathrm{T}}^{\mathrm{miss}})\rvert$	$\mathcal{F} \times 10^2$
1-prong	< 40	< 2	14.56 ± 2.0
	> 40	< 2	14.60 ± 2.08
	< 40	> 2	6.38 ± 1.52
	> 40	> 2	6.10 ± 1.64
3-prong	< 40	< 2	3.53 ± 2.50
	> 40	< 2	3.55 ± 1.83
	< 40	> 2	1.94 ± 1.43
	> 40	> 2	1.92 ± 1.47

Table 8.1.3.1: Combined Fake Factors ($\mathcal{F}$) calculated separately for four sub-regions that are defined by the requirements on the transverse momentum of the τ_{had} ($p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$), and the absolute value of the azimuthal angle between τ_{had} and the missing transverse energy ($\lvert\Delta\phi(\tau_{\mathrm{had}}, E_{\mathrm{T}}^{\mathrm{miss}})\rvert$). Values are evaluated in the 1-prong and 3-prong channels. Statistical and systematic uncertainties are shown.

Channel	**Signal**	**VV**	$\boldsymbol{Z \to \ell\ell}$	**Top**	**W**	**Fakes**	$\boldsymbol{Z \to \tau\tau}$	**Other Higgs**	**Data**
VBF SR inclusive	162.07 ± 0.53	68.61 ± 1.61	72.78 ± 11.01	114.35 ± 4.04	23.73 ± 5.26	580.89 ± 177.34	2729.36±20.21	60.14 ± 0.74	3669
VBF SR $\tau_e\tau_{\mathrm{had}}$	76.18 ± 0.37	32.10 ± 1.13	49.31 ± 9.65	52.18 ± 2.75	7.67 ± 2.05	320.33 ± 98.57	1248.09 ± 13.34	28.99 ± 0.52	1801
VBF SR $\tau_\mu\tau_{\mathrm{had}}$	85.88 ± 0.38	36.51 ± 1.15	23.47 ± 5.31	62.17 ± 2.96	16.06 ± 4.84	260.56 ± 78.78	1481.27 ± 15.17	31.15 ± 0.52	1868
VBF SR $\tau_{\mathrm{lep}}\tau_{\text{1-prong}}$	129.00 ± 0.48	52.83 ± 1.41	63.05 ± 9.98	86.34 ± 3.51	17.10 ± 4.57	431.65 ± 77.67	2136.99 ± 17.33	47.51 ± 0.66	2919
VBF SR $\tau_{\mathrm{lep}}\tau_{\text{3-prong}}$	33.07 ± 0.24	15.78 ± 0.77	9.73 ± 4.65	28.01 ± 1.99	6.64 ± 2.60	149.24 ± 99.91	592.37 ± 10.38	12.63 ± 0.34	750

Table 8.1.3.2: Number of data and MC simulated yields in the different VBF signal region (SR) channels including the fake background contribution. Both statistical and systematic uncertainties are shown.

VBF SR final state	b/d
Inclusive	1.04 ± 0.05
$\tau_{\mathrm{lep}}\tau_{\text{1-prong}}$	1.02 ± 0.03
$\tau_{\mathrm{lep}}\tau_{\text{3-prong}}$	1.13 ± 0.13
$\tau_e\tau_{\mathrm{had}}$	1.01 ± 0.05
$\tau_\mu\tau_{\mathrm{had}}$	1.06 ± 0.04

Table 8.1.3.3: Ratio of background yields (b) and data (d) in the VBF SR after including the fake background contribution in the inclusive region and in the four different decay channels . Statistical and systematic uncertainties are shown.

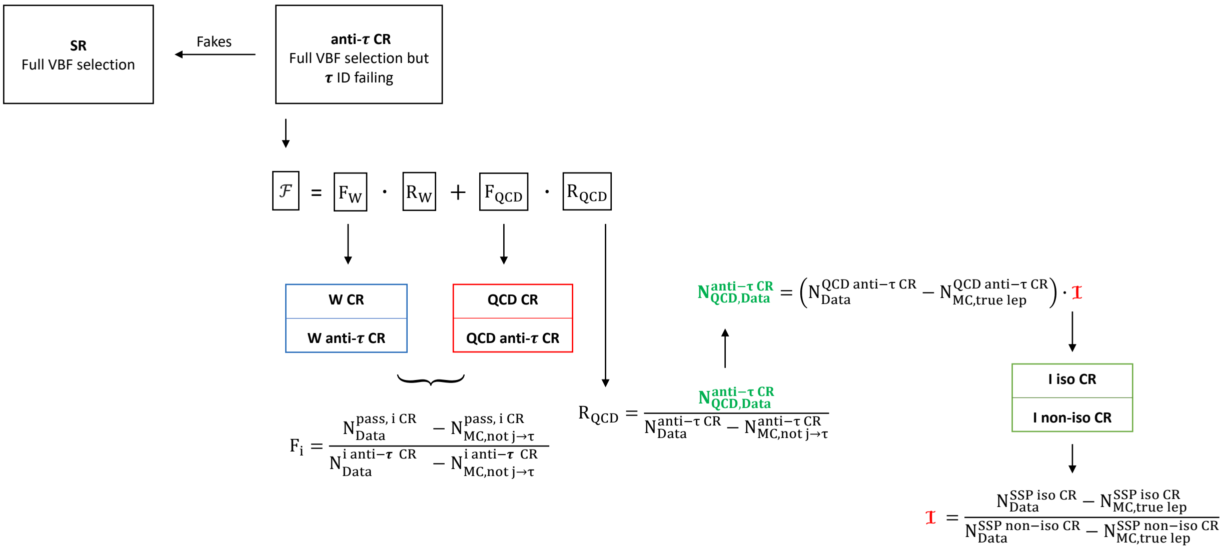
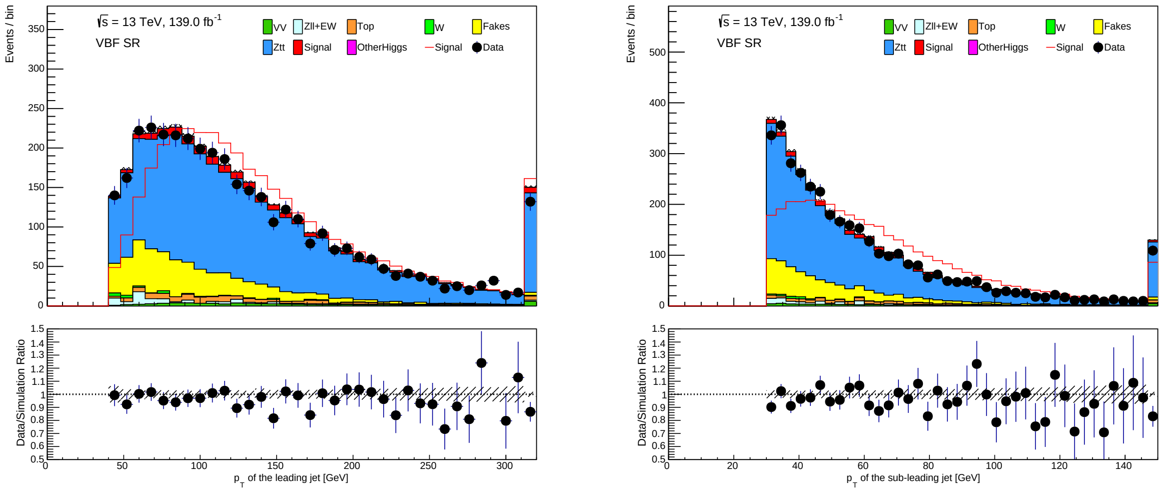


Figure 8.1.3.1: Summary of the Fake Factor method. The fake background contribution (Fakes) in the VBF Signal Region (SR) are calculated in the anti-τ control region (CR) where events have to fail the τ-lepton identification (ID) requirement. The combined Fake Factor ($\mathcal{F}$) is the sum of contributions from W+jets (W) and multi-jet (QCD) production processes, more details in Section 8.1.3. These are the Individual Fake Factors (F_i), ratio of the number of events (N) in the i CR ($i = \{\mathrm{W}, \mathrm{QCD}\}$) passing the τ-ID criteria (pass), over those that fail it (anti-τ), subtracting the MC event yields in which the τ_{had} do not originate from a jet (MC, not $j \to \tau$), as shown in Section 8.1.1 and fractional contributions R_i, displayed in Section 8.1.2. For the number of QCD events in the anti-τ CR ($\mathrm{N}^{\text{anti-}\tau\text{CR}}_{\text{QCD,Data}}$) the Isolation Factors ($\mathcal{I}$) are multiplied by the number of events in the QCD anti-τ CR (subtracting the MC event yields from true leptons $\mathrm{N}^{\text{QCD anti-}\tau\text{CR}}_{\text{MC,true lep}}$). $\mathcal{I}$ is the ratio between the number of events in the same-sign Preselection (SSP) CR that pass the isolation (iso) requirement applied on the light lepton over those that fail it (non-iso), as shown in Section 8.1.2.

8.2 Kinematic distributions in the VBF $\tau_{\text{lep}}\tau_{\text{had}}$ channel SR

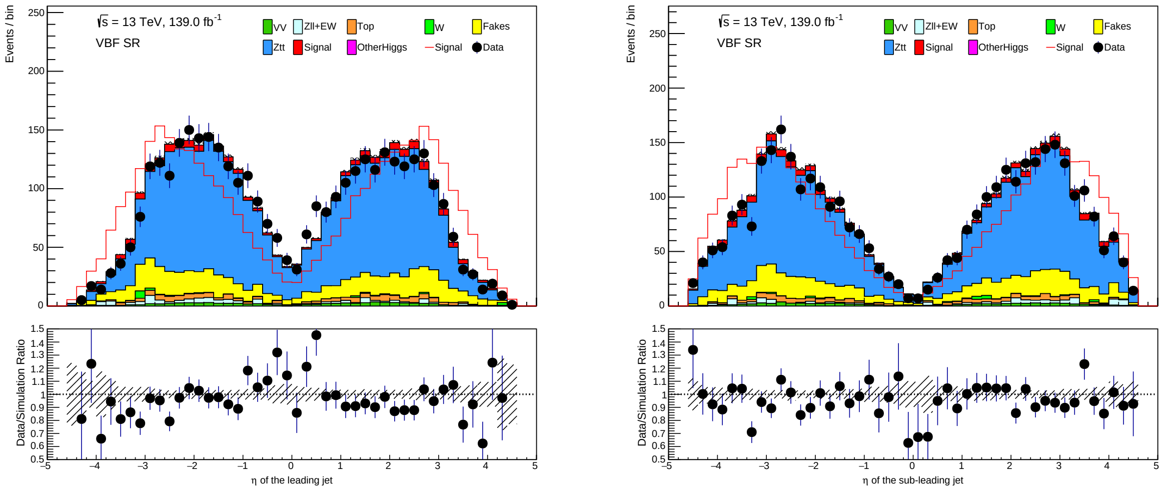
This section shows distributions of some of the main event kinematics in the VBF SR with the inclusion of the fake background contribution using the results and method described in Section 8.1. There is good agreement between the data and the prediction, as shown in Figures 8.2.1.1, 8.2.2.1, and 8.2.3.1. The p_{T} distributions display good agreement, especially for low p_{T} values compared to larger deviations in higher p_{T} ranges. Mismodeling for $p_{\text{T}}^{\tau_{\text{lep}}\tau_{\text{3-prong}}} < 50$ GeV is seen, same behavior found in Ref. [66]. Angular distributions (η), show larger differences than the p_{T} distributions. The ratio between data and simulation fluctuates around one.

8.2.1 Inclusive jet kinematics



(a) p_{T} of the leading jet

(b) p_{T} of the sub-leading jet

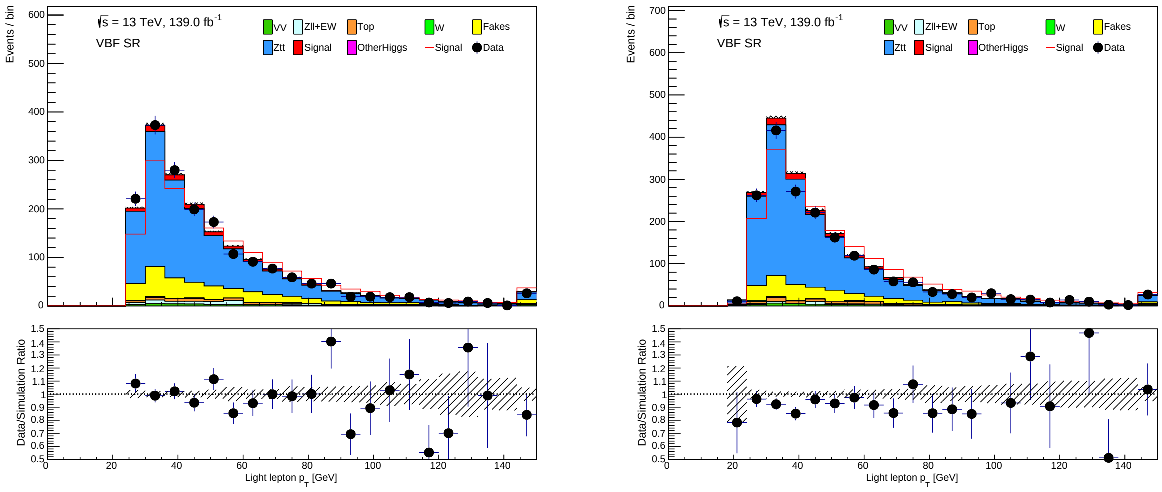


(c) η of the leading jet

(d) η of the sub-leading jet

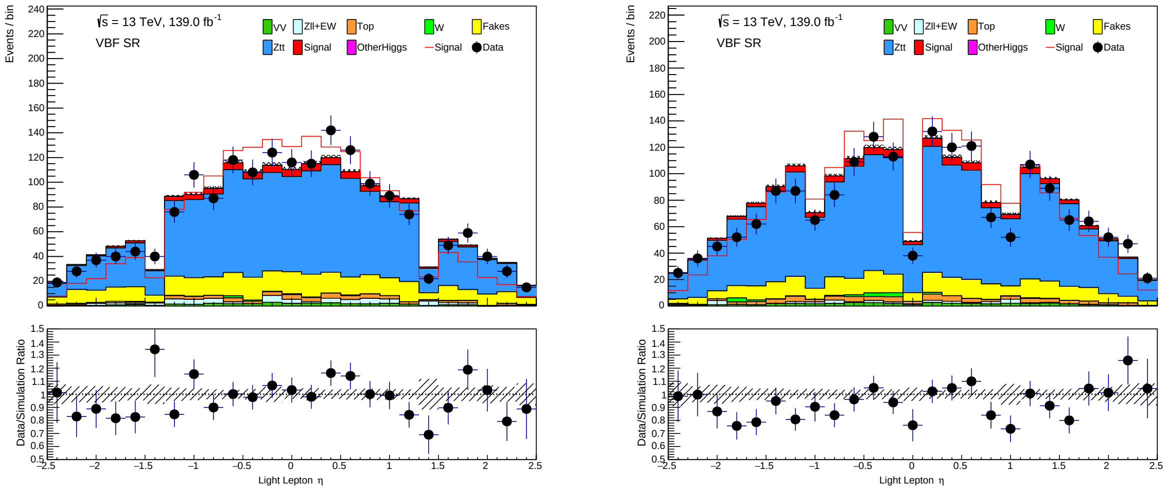
Figure 8.2.1.1: Kinematic distributions of the leading and sub-leading p_{T} jets in the VBF inclusive SR. Only statistical uncertainties are shown. Overflow bins are presented in Figures (a) and (b). "Signal" line represents the signal contribution normalized to the total background.

8.2.2 Light lepton kinematics



(a) p_{T} of the e in the $\tau_e\tau_{\mathrm{had}}$ channel

(b) p_{T} of the μ in the $\tau_\mu\tau_{\mathrm{had}}$ channel

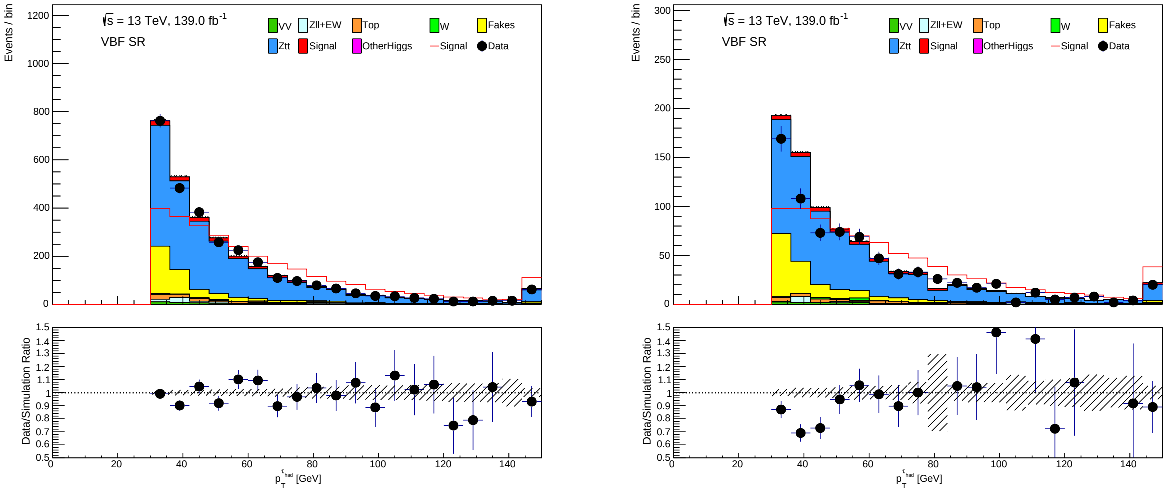


(c) η of the e in the $\tau_e\tau_{\mathrm{had}}$ channel

(d) η of the μ in the $\tau_\mu\tau_{\mathrm{had}}$ channel

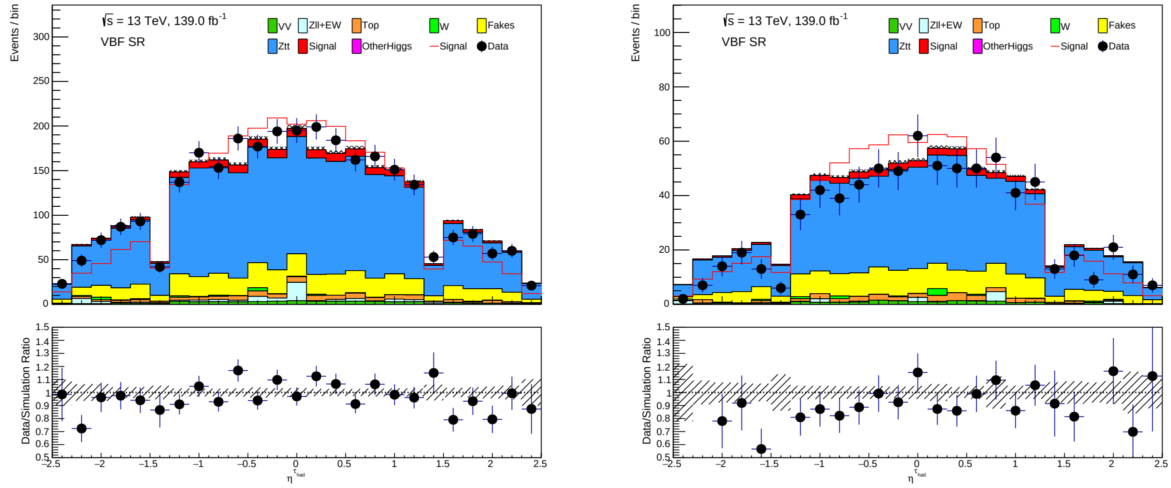
Figure 8.2.2.1: Kinematic distributions of the electron (e) and muon (μ) in the VBF SR in the $\tau_e\tau_{\mathrm{had}}$ and $\tau_\mu\tau_{\mathrm{had}}$ channels respectively. Only statistical uncertainties are shown. "Signal" line represents the signal contribution normalized to the total background.

8.2.3 τ_{had} kinematics



(a) p_{T} of the τ_{had} in the $\tau_{\text{lep}}\tau_{\text{1-prong}}$ channel

(b) p_{T} of the τ_{had} in the $\tau_{\text{lep}}\tau_{\text{3-prong}}$ channel



(c) η of the τ_{had} in the $\tau_{\text{lep}}\tau_{\text{1-prong}}$ channel

(d) η of the τ_{had} in the $\tau_{\text{lep}}\tau_{\text{3-prong}}$ channel

Figure 8.2.3.1: Kinematic distributions of the τ_{had} in the VBF signal region in the $\tau_{\text{lep}}\tau_{\text{1-prong}}$ and $\tau_{\text{lep}}\tau_{\text{3-prong}}$ final states. Only statistical uncertainties are shown. "Signal" line represents the signal contribution normalized to the total background.

8.3 Validation of the Fake Factor method

This section discusses some further studies on the impact of some of the assumptions made during the development of the Fake Factor and the validation of the method.

8.3.1 Same individual Fake Factors from all non multi-jet processes

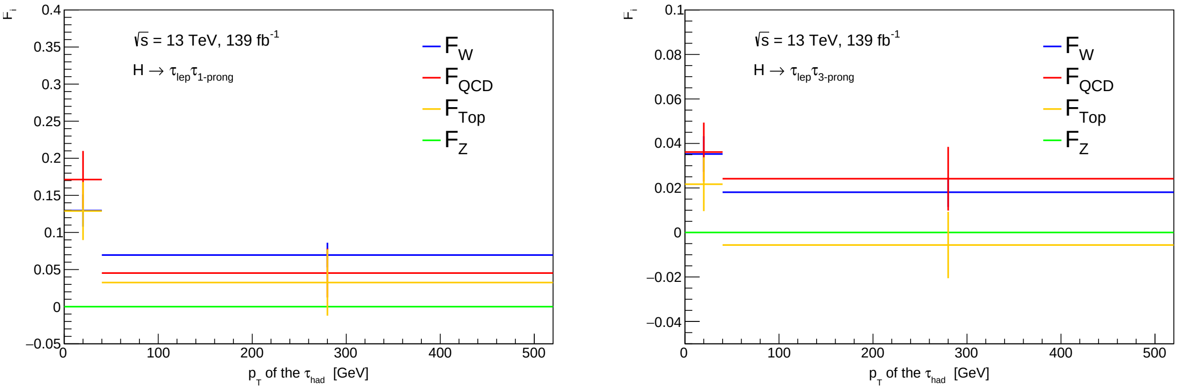
The Fake Factor method is simplified by applying Equation 8.1.3 thus, assuming that contributions from Z+jets and top-quark production processes to the background ($< 2\%$ and $\sim 3\%$ in the VBF SR, respectively) are small. Therefore, it is implicitly assumed that the individual Fake Factors from all non multi-jet contributions are approximately the same: $\mathrm{F_W} \approx \mathrm{F_{Top}} \approx \mathrm{F_Z}$.

To validate this assumption, the individual Fake Factors are calculated for all four processes using the same method as in Section 8.1.1. Two new CRs are defined for the Z+jets and top-quark production processes. The CR enriched in Z+jets events (Z CR) is defined with the same selection as the VBF SR except for requiring two leptons in the final state. The top-quark production process enhanced CR (Top CR) is defined same as the VBF SR except for requiring at least one b-tagged jet instead of a b-jet veto.

Figure 8.3.1.1 shows the F_i of all contributions as function of the transverse momentum of the τ_{had}, split by the number of charged tracks from the decay of τ_{had}. Due to the limited statistics in the Z CR, $\mathrm{F_Z} = 0$. Thus, the second term of Equation 8.1.2 (Z+jets contribution) does not affect the fake background estimation.

Individual Fake Factors $\mathrm{F_W}$ and $\mathrm{F_{Top}}$ are equal in low p_{T} range in the 1-prong channel. $\mathrm{F_{Top}}$ is the lowest factor compared to the other processes.

In Section 8.5.2 it is shown that if $\mathrm{F_W} \approx \mathrm{F_{QCD}}$ is used in the estimation of the fake background, results in a difference of 0.6% in the total simulation yield (MC simulation plus data-driven estimation) compared to the case with $\mathrm{F_W} \neq \mathrm{F_{QCD}}$. Therefore, the inclusion of $\mathrm{F_{Top}}$ into the Fake Factor method will at most lead to a difference of $\sim 0.6\%$ in the total simulation yield.



(a) p_{T} of the τ_{had} in the $\tau_{\mathrm{lep}}\tau_{\text{1-prong}}$ channel

(b) p_{T} of the τ_{had} in the $\tau_{\mathrm{lep}}\tau_{\text{3-prong}}$ channel

Figure 8.3.1.1: Individual Fake Factors for W+jets ($\mathrm{F_W}$), multi-jet ($\mathrm{F_{QCD}}$), Z+jets ($\mathrm{F_W}$) and top-quark ($\mathrm{F_{Top}}$) production processes as a function of the transverse momentum of the hadronically decaying τ-lepton ($\mathrm{p_T}$ of the τ_{had}) in the (a) 1-prong and (b) 3-prong channels. Only statistical uncertainties are shown.

8.3.2 Isolation Factor calculated in the opposite-sign region

Isolation Factors are calculated in the SSP anti-τ CR, where the Preselection is applied instead of the VBF selection. Since the VBF events of the VBF SR are included in the Preselection, the same-sign (SS) requirement (same-sign charge of the light lepton and the τ_{had}) is also applied so as to create the orthogonal SSP CR. However, Isolation Factors are then applied to the QCD anti-τ CR defined with the opposite-sign (OS) requirement. Thus, an assumption has to be made that the Isolation Factors calculated in the SSP CR could be used in the OSP CR (opposite-sign Preselection).

To validate this, Isolation Factors are also calculated using the OSP anti-τ CR. Table 8.3.2.1 lists the results of the Isolation Factors calculated in the OSP anti-τ CR and SSP anti-τ CR as functions of p_{T} and $|\eta|$ of the light lepton (lep). Isolation Factors in the OSP anti-τ CR are negative for large $p_{\text{T}}^{\text{lep}}$ values in the $\tau_{\text{lep}}\tau_{\text{3-prong}}$ channel, similar to findings reported in Section 8.5.1.

The Isolation Factors calculated with both requirements do not agree within the statistical uncertainties. The absolute difference between the central values of the Isolation Factors are considered the systematic uncertainty applied to the Isolation Factors, as described in Section 8.4.

Channel	$p_{\text{T}}^{\text{lep}}$ [GeV]	$\lvert\eta^{\text{lep}}\rvert$	anti-τ CR	$\mathcal{I}$
$\tau_{\text{lep}}\tau_{\text{1-prong}}$	< 50	< 1.5	SSP	4.64 ± 0.18
			OSP	4.89 ± 0.20
	< 50	> 1.5	SSP	4.76 ± 0.27
			OSP	4.57 ± 0.27
	> 50	< 1.5	SSP	7.54 ± 0.60
			OSP	5.65 ± 0.59
	> 50	> 1.5	SSP	8.55 ± 1.01
			OSP	7.73 ± 0.92
$\tau_{\text{lep}}\tau_{\text{3-prong}}$	< 50	< 1.5	SSP	3.45 ± 0.15
			OSP	3.05 ± 0.14
	< 50	> 1.5	SSP	3.60 ± 0.21
			OSP	3.19 ± 0.24
	> 50	< 1.5	SSP	2.00 ± 0.39
			OSP	-2.3 ± 0.44
	> 50	> 1.5	SSP	3.88 ± 0.74
			OSP	-0.64 ± 0.73

Table 8.3.2.1: Isolation Factors calculated in the same-sign Preselection (SSP) and opposite-sign Preselection (OSP) anti-τ CRs, separately for four sub-regions defined by the requirements on the p_{T} and $|\eta|$ of the light lepton (lep), split into the 1-prong ($\tau_{\text{lep}}\tau_{\text{1-prong}}$) and 3-prong ($\tau_{\text{lep}}\tau_{\text{3-prong}}$) channels. Only statistical uncertainties are shown.

8.3.3 Same-sign closure test

The validation of the Fake Factor method and the underlying assumption made therein that the transfer factors (Fake Factors) calculated in different CRs could be used in the SR is achieved through a same-sign (SS) closure test which consist on repeating the Fake Factor method using the same setup described in Section 8.1, except by inverting the opposite-sign (OS) requirement on the charge of the light lepton and the τ_{had} set in the VBF SR. So that the VBF SR will be now enriched in fakes and no VBF events. Since all CRs in the Fake Factor method are constructed with the same OS requirement as the VBF SR, this requirement is inverted in all CRs. Consequently, also inverting the same-sign condition in the SSP CR used in the calculation of the Isolation Factors, ensuring orthogonality between all regions.

Individual Fake Factors

Individual Fake Factors calculated in the SS closure test are shown in Table 8.3.3.1. Individual fake factors are calculated separately in the $\tau_{\text{lep}}\tau_{\text{1-prong}}$ and $\tau_{\text{lep}}\tau_{\text{3-prong}}$ channels. Larger individual Fake Factors are found in the $\tau_{\text{lep}}\tau_{\text{1-prong}}$ channel and no significant differences are seen between the contributions from W+jets and multi-jet processes.

Channel	$\text{F}_\text{W} \times 10^2$	$\text{F}_\text{QCD} \times 10^2$
1-prong	10.02 ± 0.21	11.39 ± 1.00
3-prong	1.43 ± 0.06	2.41 ± 0.05

Table 8.3.3.1: Individual Fake Factors from W+jets (F_W) and QCD (F_QCD) processes in the 1-prong and 3-prong channels calculated for the same-sign closure test. Only statistical uncertainties are shown.

Individual Fake Factors in the Fake Factor method are split in the 1-prong and 3-prong channels and categorized depending on the $p_\text{T}^{\tau_\text{had}}$. However in the SS closure test, due to technical issues only one categorization, either in $p_\text{T}^{\tau_\text{had}}$ or in the number of charged tracks of the τ_{had} was possible. Since the categorization of the Individual Fake Factors only on the $p_\text{T}^{\tau_\text{had}}$ caused a mismodeling of the fake background contribution when split in the $\tau_{\text{lep}}\tau_{\text{1-prong}}$ and $\tau_{\text{lep}}\tau_{\text{3-prong}}$ channels, the Individual Fake Factors in the SS closure test are not calculated for different $p_\text{T}^{\tau_\text{had}}$ values but only split into the 1-prong and 3-prong channels. Individual Fake Factors calculated as function of $p_\text{T}^{\tau_\text{had}}$ for the SS test is shown in Figure A.2.1.

Larger Individual Fake Factors are found for lower values of $p_\text{T}^{\tau_\text{had}}$, following the same trend as the F_is calculated in Section 8.1.1. Furthermore, F_W and F_QCD do not show significant differences, also it is found that $\text{F}_\text{QCD} > \text{F}_\text{W}$ for low $p_\text{T}^{\tau_\text{had}}$ values while $\text{F}_\text{W} > \text{F}_\text{QCD}$ for large $p_\text{T}^{\tau_\text{had}}$ values, as in the Fake Factor method, shown in Figure 8.1.1.2a.

Isolation Factors

The Isolation Factors calculated in the SS closure test are shown in Figure A.2.2 and Table 8.3.3.2, binned in charged track multiplicity and p_T of the τ_{had}, as well as the light lepton (lep) $|\eta|$. The Isolation Factors increase with $p_\text{T}^{\tau_\text{had}}$ in the 1-prong channel and decrease with $p_\text{T}^{\tau_\text{had}}$ in the 3-prong channel, where negative values arise. Higher dependence on $|\eta^{\text{lep}}|$ found for higher $p_\text{T}^{\tau_\text{had}}$ values.

Channel	$p_{\mathrm{T}}^{\mathrm{lep}}$ [GeV]	$\lvert\eta^{\mathrm{lep}}\rvert$	$\mathcal{I}$
1-prong	< 50	< 1.5	4.89 ± 0.20
	< 50	> 1.5	4.57 ± 0.27
	> 50	< 1.5	5.65 ± 0.59
	> 50	> 1.5	7.73 ± 0.92
3-prong	< 50	< 1.5	3.05 ± 0.14
	< 50	> 1.5	3.19 ± 0.24
	> 50	< 1.5	-2.30 ± 0.44
	> 50	> 1.5	-0.64 ± 0.73

Table 8.3.3.2: Isolation Factors ($\mathcal{I}$) split according to the number of charged tracks of the τ_{had} decay (1-prong and 3-prong channels), p_{T} and $|\eta|$ of the light lepton (lep) calculated for the same-sign closure test. Only statistical uncertainties are shown.

Fractional Contribution from QCD processes

$\mathrm{R}_{\mathrm{QCD}}$ values are calculated using Equation 8.1.2.1 and binned in two dimensional $p_{\mathrm{T}}^{\tau_{\mathrm{had}}} - |\Delta\phi(\tau_{\mathrm{had}}, E_{\mathrm{T}}^{\mathrm{miss}})|$ plane. They are further categorized according to the number of charged decay products of the τ_{had}. Figure A.2.3 and Table 8.3.3.3 show the fractional contributions from QCD processes to the anti-τ CR calculated in the SS closure test. $\mathrm{R}_{\mathrm{QCD}}$ values are larger for lower $p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$ values for both channels. Large fluctuations in the $\mathrm{R}_{\mathrm{QCD}}$ values depending on the $|\Delta\phi(\tau_{\mathrm{had}}, E_{\mathrm{T}}^{\mathrm{miss}})|$ values are found in all regions but in the $p_{\mathrm{T}}^{\tau_{\text{3-prong}}} < 40$ GeV region. Dominant contribution is found to be coming from W+jets events (since $\mathrm{R}_{\mathrm{W}} = 1 - \mathrm{R}_{\mathrm{QCD}}$), similar to the findings in Section 8.1.2.

Combined Fake Factors

The results of the combined Fake Factors ($\mathcal{F}$) calculated in the SS closure test are shown in Table 8.3.3.4. In the SS closure test the individual Fake Factors are not binned in $p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$, in contrast to the Fake Factor method. There is almost no dependence of the combined Fake Factors on the $p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$ in the SS closure test, unlike the combined Fake Factors from Table 8.1.3.1. Since the combined Fake Factors only depend on the individual Fake Factors and $\mathrm{R}_{\mathrm{QCD}}$ and the latter one does depend on the $p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$, as shown in Table 8.3.3.3, it can be concluded that the combined Fake Factors depend mostly of the individual Fake Factors, since the non-dependence on the $p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$ can only be originated by the individual Fake Factors. This fact agrees with the results from Section 8.5.2. Table A.2.3 lists the total MC and data yields for the inclusive and different decays channels of the VBF SR, result of the SS closure test.

Distributions of some of the main event kinematics in the VBF SR with the inclusion of the fake background contribution using the results of the SS closure test. There is good agreement between the data and the prediction, as shown in Figures 8.3.3.1, 8.3.3.2, and 8.3.3.3.

The p_{T} distributions display good agreement, compared to larger deviations in higher p_{T} ranges. Angular distributions (η), show larger differences than the p_{T} distributions. The ratio between data and simulation fluctuates around one. Larger fluctuations than in the Fake Factor method.

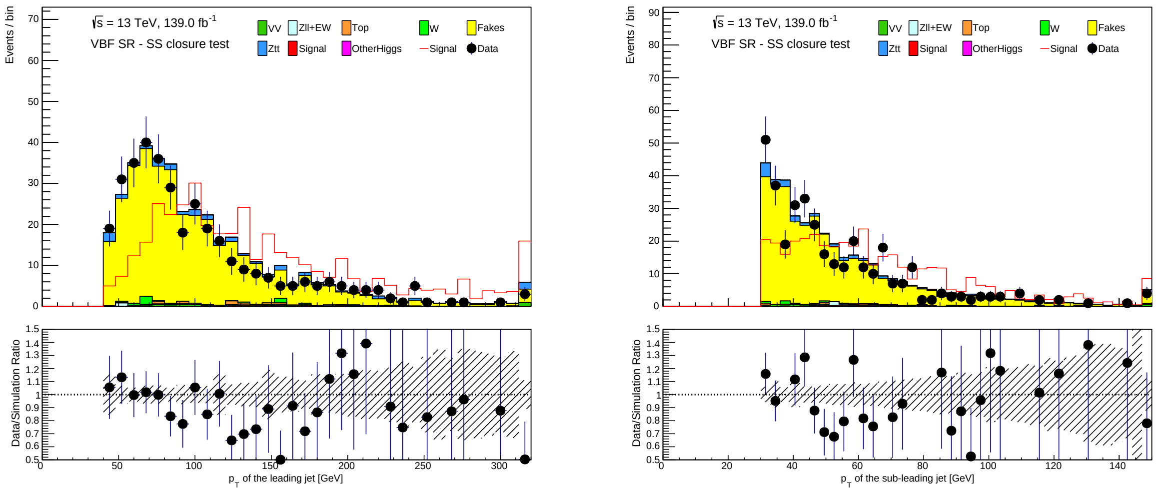
Channel	$p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$ [GeV]	$\lvert\Delta\phi(\tau_{\mathrm{had}}, E_{\mathrm{T}}^{\mathrm{miss}})\rvert$	$\mathrm{R_{QCD}} \times 10$
1-prong	< 40	< 2	4.3 ± 0.4
	< 40	> 2	9.5 ± 2.8
	> 40	< 2	3.7 ± 0.4
	> 40	> 2	8.3 ± 2.3
3-prong	< 40	< 2	2.3 ± 0.2
	< 40	> 2	2.4 ± 1.1
	> 40	< 2	1.8 ± 0.2
	> 40	> 2	0.6 ± 0.6

Table 8.3.3.3: Fractional contributions from multi-jet processes to the anti-τ CR ($\mathrm{R_{QCD}}$) calculated for the same-sign closure test. Values are obtained separately for eight subregions that are defined by the requirements on the number of charged tracks (1- or 3-prong) and the transverse momentum of τ_{had} ($p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$), and the azimuthal angle between τ_{had} and the missing transverse energy $E_{\mathrm{T}}^{\mathrm{miss}}$ ($\lvert\Delta\phi(\tau_{\mathrm{had}}, E_{\mathrm{T}}^{\mathrm{miss}})\rvert$). Only statistical uncertainties are shown.

Channel	$p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$ [GeV]	$\lvert\Delta\phi(\tau_{\mathrm{had}}, E_{\mathrm{T}}^{\mathrm{miss}})\rvert$	$\mathcal{F} \times 10^2$
1-prong	< 40	< 2	10.61 ± 0.14
	> 40	< 2	11.32 ± 0.40
	< 40	> 2	10.53 ± 0.15
	> 40	> 2	11.16 ± 0.33
3-prong	< 40	< 2	1.66 ± 0.051
	> 40	< 2	1.67 ± 0.12
	< 40	> 2	1.60 ± 0.054
	> 40	> 2	1.48 ± 0.081

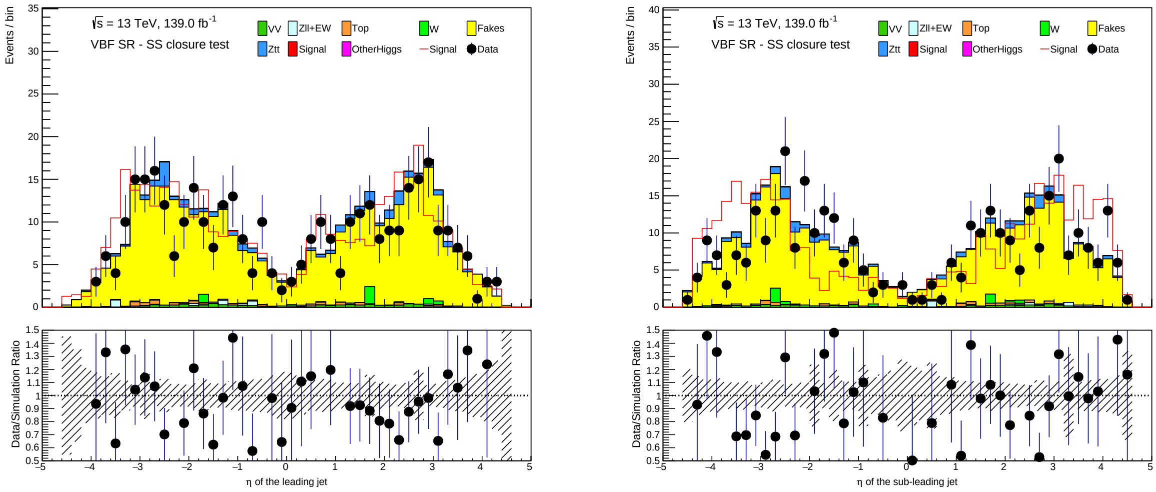
Table 8.3.3.4: Combined Fake Factors ($\mathcal{F}$) calculated for the same-sign (SS) closure test, binned in the azimuthal angle between the τ_{had} and the missing transverse energy ($\lvert\Delta\phi(\tau_{\mathrm{had}}, E_{\mathrm{T}}^{\mathrm{miss}})\rvert$), and the transverse momentum of the τ_{had} ($p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$). The $\mathcal{F}$s are further divided into the 1- and 3-prong channels. Only statistical uncertainties are shown.

- **Inclusive jet kinematics**



(a) p_{T} of the leading jet.

(b) p_{T} of the sub-leading jet

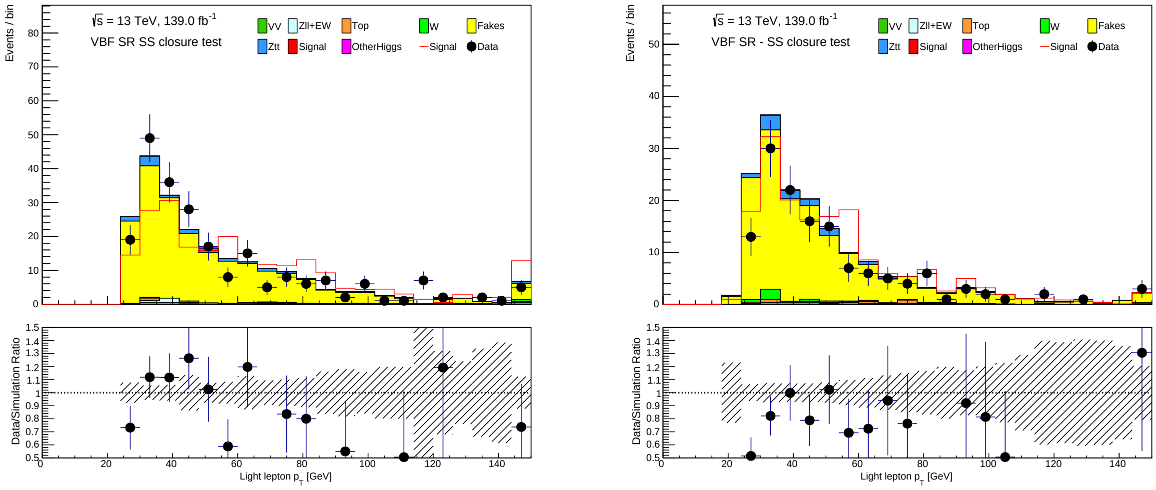


(c) η of the leading jet

(d) η of the sub-leading jet

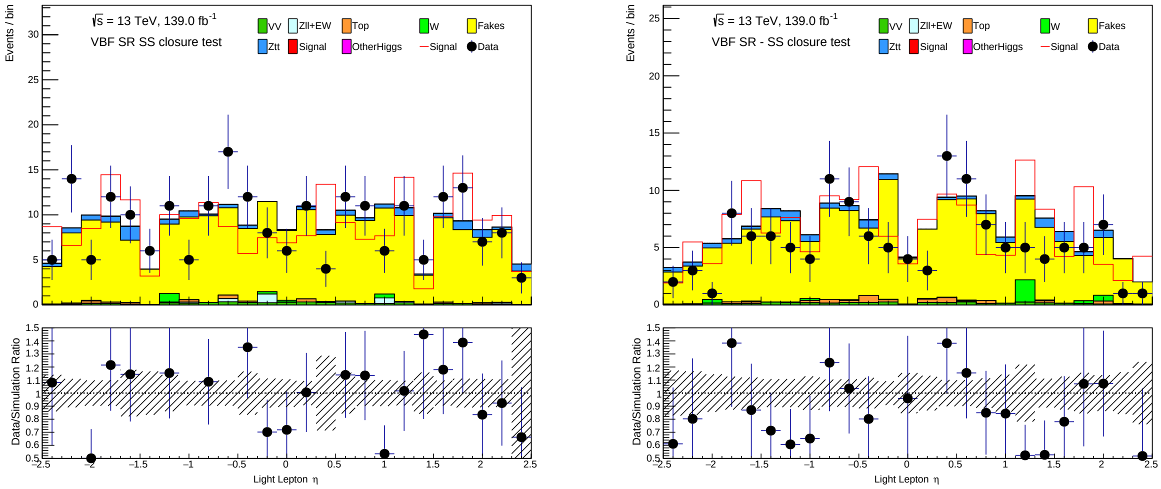
Figure 8.3.3.1: Kinematic distributions of the τ_{had} in the VBF signal region divided into the $\tau_{\mathrm{lep}}\tau_{\text{1-prong}}$ and $\tau_{\mathrm{lep}}\tau_{\text{3-prong}}$ final states calculated for the same-sign closure test. Only statistical uncertainties are shown. "Signal" line represents the signal contribution normalized to the total background.

- **Light lepton kinematics**



(a) p_{T} of the e in the $\tau_e\tau_{\mathrm{had}}$ channel

(b) p_{T} of the μ in the $\tau_\mu\tau_{\mathrm{had}}$ channel

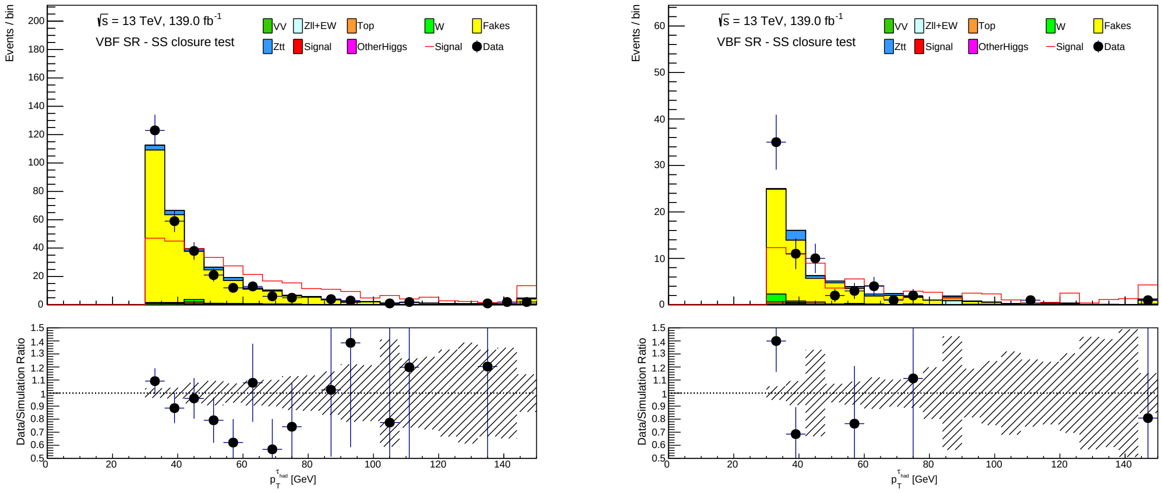


(c) η of the e in the $\tau_e\tau_{\mathrm{had}}$ channel

(d) η of the μ in the $\tau_\mu\tau_{\mathrm{had}}$ channel

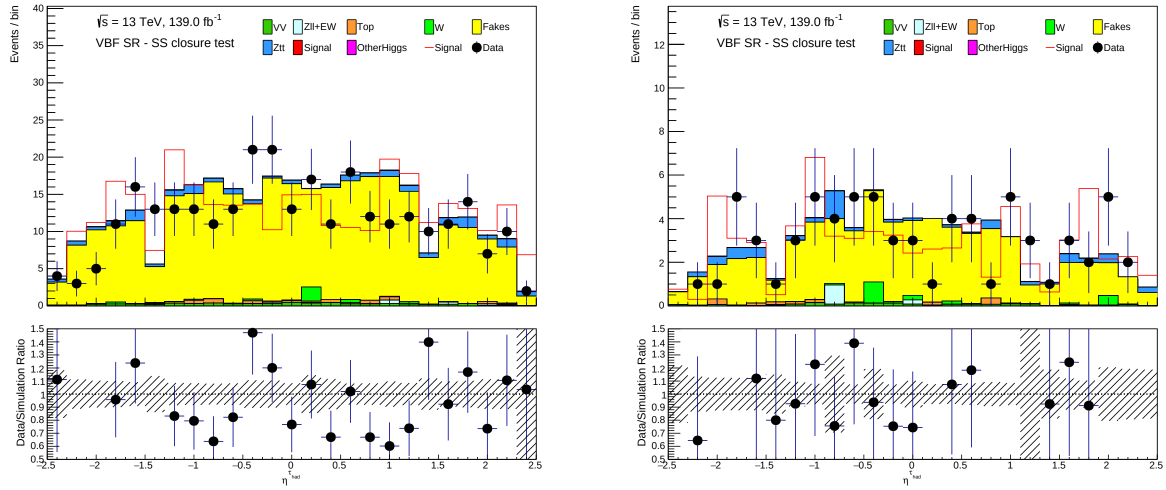
Figure 8.3.3.2: Kinematic distributions of the electron (e) and muon (μ) in the VBF SR split into the $\tau_e\tau_{\mathrm{had}}$ and $\tau_\mu\tau_{\mathrm{had}}$ channels respectively calculated for the same-sign closure test. Only statistical uncertainties are shown. "Signal" line represents the signal contribution normalized to the total background.

- τ_{had} **kinematics**



(a) p_{T} of the τ_{had} in the $\tau_{\text{lep}}\tau_{\text{1-prong}}$ channel

(b) p_{T} of the τ_{had} in the $\tau_{\text{lep}}\tau_{\text{3-prong}}$ channel



(c) η of the τ_{had} in the $\tau_{\text{lep}}\tau_{\text{1-prong}}$ channel

(d) η of the τ_{had} in the $\tau_{\text{lep}}\tau_{\text{1-prong}}$ channel

Figure 8.3.3.3: Kinematic distributions of the τ_{had} in the VBF signal region divided into the $\tau_{\text{lep}}\tau_{\text{1-prong}}$ and $\tau_{\text{lep}}\tau_{\text{3-prong}}$ final states calculated for the same-sign closure test. Only statistical uncertainties are shown. "Signal" line represents the signal contribution normalized to the total background.

8.4 Uncertainties

This section describes the uncertainties of the fake background contribution. Statistical uncertainties arise from the squared sum of the event weights, they are considered in all Fake Factor method steps. Only one systematic uncertainty is applied regarding the application of the Isolation Factors calculated in the SSP anti-τ CR to the OSP anti-τ CR, called "SS-OS" in this section. The combined uncertainty is calculated as the Pythagorean addition of the statistical and systematic uncertainties. The error propagation formulas can be found in Section A.3.

- The statistical uncertainties of the Individual Fake Factors are calculated through standard error propagation. The relative uncertainties of the Individual Fake Factors are between 16% and 73%.

- The uncertainties related to the Isolation Factors,
  - Statistical uncertainties from the CRs. The relative uncertainties vary between 4% and 20%.
  - Systematic uncertainties due to the use of the SSP CR instead of the OSP CR, as discussed in Section 8.3.2. This uncertainty is calculated as the absolute value of the subtraction between the values shown in Table 8.3.2.1 calculated for each region and channel separately, [66].

    $$\text{SS-OS} = |\mathcal{I}_{\text{SSP anti-}\tau\text{CR}} - \mathcal{I}_{\text{OSP anti-}\tau\text{CR}}| \tag{8.4.1}$$

    Subtraction values of the Isolation Factors in the high $p_{\mathrm{T}}^{\mathrm{lep}}$ regime in the $\tau_{\mathrm{lep}}\tau_{\text{3-prong}}$ channel lead to inconsistent values. This is due to the lack of statistics in this region. The systematic uncertainties are set to be 70% of the values for each of these two bins (shown in last two rows of Table 8.4.1, "original SS-OS$\rightarrow$ 70% $\mathcal{I}$"), which is the maximum relative difference between the Isolation Factors from the 1- and 3-prong channels in the $p_{\mathrm{T}}^{\mathrm{lep}} > 50$ GeV range.

- The uncertainties from the relative contribution $\mathrm{R_{QCD}}$ are the statistical uncertainties derived from the CRs, and the propagated systematic uncertainty from the Isolation Factors, (see Equation 8.1.2.2). Uncertainties of $\mathrm{R_W}$ is derived from the linear relation with $\mathrm{R_{QCD}}$, (RW $= 1 - \mathrm{R_{QCD}}$).

The uncertainties associated with the assumption that all non multi-jet contributions are approximately the same have been discarded due to their negligible influence in the final fake background estimation, as shown in Section 8.3.1.

Table 8.4.2 lists the uncertainties of the fake background contribution in four decay channels and the inclusive VBF SR. The uncertainties are divided into statistical uncertainties, around $\sim 2\%$ of the fake background contribution and systematic uncertainties of the SS-OS, which represents the main source of the uncertainties.

Moreover, the largest uncertainty in the VBF SR comes from the fake background. Including the fake background, their uncertainties represent a 4.7% of the expected yield in the inclusive VBF SR.

Channel	$p_{\mathrm{T}}^{\mathrm{lep}}$ [GeV]	$\lvert\eta^{\mathrm{lep}}\rvert$	$\mathcal{I}$	Stat. Uncertainty	SS-OS Uncertainty	Combined Uncertainty	Rel. Uncertainty [%]
1-prong	< 50	< 1.5	4.64	0.18	0.25	0.31	6.7
	< 50	> 1.5	4.76	0.27	0.19	0.33	6.8
	> 50	< 1.5	7.54	0.60	1.89	1.98	26
	> 50	> 1.5	8.55	1.01	0.82	1.30	15
3-prong	< 50	< 1.5	3.45	0.15	0.40	0.43	12.2
	< 50	> 1.5	3.60	0.21	0.41	0.46	12.8
	> 50	< 1.5	2.00	0.39	$4.30 \rightarrow 1.40$	$4.32 \rightarrow 1.45$	$216 \rightarrow 73$
	> 50	> 1.5	3.88	0.74	$4.52 \rightarrow 2.72$	$4.58 \rightarrow 2.82$	$117 \rightarrow 73$

Table 8.4.1: Uncertainties of the Isolation Factors ($\mathcal{I}$) binned in p_{T} and $|\eta|$ of the light lepton "lep". The $\mathcal{I}$, the statistical uncertainties (Stat. Uncertainty), the systematic uncertainty from the different requirements applied on the CRs used to calculate $\mathcal{I}$, same-sign, (SS) and opposite-sign, (OS), "SS-OS", the total uncertainty (Combined Uncertainty) and the relative uncertainty (Rel. Uncertainty [%]). Right arrow represents that the original SS-OS value has been changed to a 70% of $\mathcal{I}$."

VBF SR final state	Fake background yield	Stat. Uncertainty	Stat. Rel [%]	SS-OS Uncertainty	SS-OS Rel [%]	Combined Uncertainty	Combined Rel [%]
Inclusive	580.89	8.72	1.5	177.13	30.5	177.34	30.5
$\tau_{\mathrm{lep}}\tau_{\text{1-prong}}$	431.65	8.38	1.9	77.22	17.9	77.67	18.0
$\tau_{\mathrm{lep}}\tau_{\text{3-prong}}$	149.24	2.40	1.6	99.88	66.9	99.91	66.9
$\tau_e\tau_{\mathrm{had}}$	320.33	6.70	2.1	98.29	30.7	98.52	30.8
$\tau_\mu\tau_{\mathrm{had}}$	260.56	5.58	2.1	78.67	30.2	78.87	30.3

Table 8.4.2: Yields and uncertainties of the fake background in the four different decay channels and the inclusive VBF SR. Statistical uncertainty (Stat.) and systematic uncertainty same-sign (SS) and opposite-sign (OS), "SS-OS" are shown. Relative uncertainty to the total fake background yields on each channel.

8.5 Further studies

This section discusses some further approximations not used but studied for future applications.

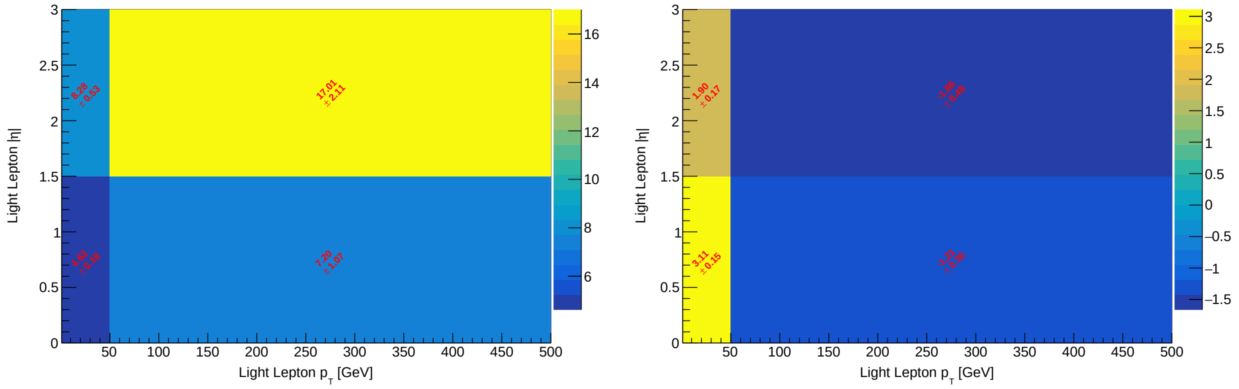
8.5.1 Isolation factors split by the lepton flavour

In the Fake Factor method, the individual Fake Factors (F_i) were evaluated on the $\tau_{\mathrm{lep}}\tau_{\text{1-prong}}$ and $\tau_{\mathrm{lep}}\tau_{\text{3-prong}}$ channels since the CRs defined for their calculation contained jets misidentified as τ_{had}. However, the CRs used for the calculation of the Isolation Factors ($\mathcal{I}$) were divided based on the isolation of the light leptons. Therefore $\mathcal{I}$ should be evaluated in the lepton flavor decay channels. First, because the working points of the light lepton isolation are different for e and μ, as shown in chapter 7. Secondly, because the fake contribution in this CRs come from jets misidentified as light leptons. This section justifies why the $\mathcal{I}$ are divided in the $\tau_{\mathrm{lep}}\tau_{\text{1-prong}}$ and $\tau_{\mathrm{lep}}\tau_{\text{3-prong}}$ channels.

The values of $\mathcal{I}$ split in the light lepton flavour channels are shown in Table 8.5.1.1, and in Figure 8.5.1.1, binned in p_{T} and $|\eta|$ of the light lepton. In the $\tau_e\tau_{\mathrm{had}}$ channel, values of $\mathcal{I}$ are larger for larger $|\eta|$ and p_{T}, in contrast to the $\tau_\mu\tau_{\mathrm{had}}$ channel. Large fluctuations between the $\mathcal{I}$ values are seen in the high $p_{\mathrm{T}}^{\mathrm{lep}}$ range.

Larger fluctuations between the Isolation Factors are shown in Table 8.5.1.1 than in Table 8.1.2.1. These arise from the yields in the distributions used as the numerator in Equation 8.1.2.3. Especially, for $p_{\mathrm{T}}^{\mu} > 50$ GeV, simulated MC events are of the same order of data, leading to mismodeling of the fake background, the distributions used for the numerator are shown in Figure 8.5.1.2.

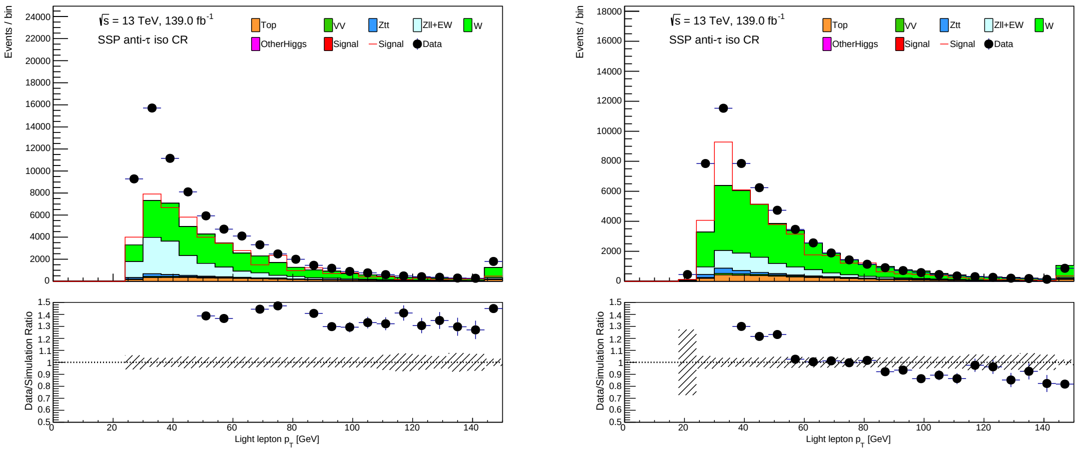
Previous analyses [11], [99] using the lepton flavour categorization for the Isolation Factors calculation addressed this issue by including an additional systematic uncertainty to $\mathcal{I}$. Following Ref.[66], in this analysis $\mathcal{I}$ values are calculated in the $\tau_{\mathrm{lep}}\tau_{\text{1-prong}}$ and $\tau_{\mathrm{lep}}\tau_{\text{3-prong}}$ channels, which provide acceptable results for the estimation of the fake background.



(a) $\mathcal{I}$ in the $\tau_e\tau_{\mathrm{had}}$ channel

(b) $\mathcal{I}$ in the $\tau_\mu\tau_{\mathrm{had}}$ channel

Figure 8.5.1.1: Isolation Factors ($\mathcal{I}$) for the (a) $\tau_e\tau_{\mathrm{had}}$ and (b) $\tau_\mu\tau_{\mathrm{had}}$ channels binned in p_{T} and $|\eta|$ of the light lepton. Values are calculated in the same-sign (SSP) anti-τ CRs. Only statistical uncertainties are shown.



(a) p_{T} of the electron in the $\tau_{\mathrm{e}}\tau_{\mathrm{had}}$ channel

(b) p_{T} of the muon in the $\tau_{\mu}\tau_{\mathrm{had}}$ channel

Figure 8.5.1.2: Transverse momentum (p_{T}) in the same-sign Preselection (SSP) anti-τ in the (a) $\tau_{\mathrm{e}}\tau_{\mathrm{had}}$ and (b) $\tau_{\mu}\tau_{\mathrm{had}}$ channels. Only statistical uncertainties are shown. The "Signal" line represents the signal contribution normalized to the total background.

Final state	$p_{\mathrm{T}}^{\mathrm{lep}}$ [GeV]	$\lvert\eta^{\mathrm{lep}}\rvert$	$\mathcal{I}$
$\tau_{\mathrm{e}}\tau_{\mathrm{had}}$	< 50	< 1.5	4.62 ± 0.18
	< 50	> 1.5	8.28 ± 0.53
	> 50	< 1.5	7.20 ± 1.07
	> 50	> 1.5	17.01 ± 2.11
$\tau_{\mu}\tau_{\mathrm{had}}$	< 50	< 1.5	3.11 ± 0.15
	< 50	> 1.5	1.90 ± 0.17
	> 50	< 1.5	-1.21 ± 0.30
	> 50	> 1.5	-1.66 ± 0.49

Table 8.5.1.1: Isolation Factors ($\mathcal{I}$) categorized by the flavour, p_{T} and $\lvert\eta\rvert$ of the light lepton (lep). Only statistical uncertainties are shown.

8.5.2 Simplified Fake Factor method

Individual Fake Factors calculated in Section 8.1.1 did not show large differences between the contribution from W+jets ($\mathrm{F_W}$) and multi-jet processes ($\mathrm{F_{QCD}}$). This section studies whether this observation can be utilized to simplify the Fake Factor method.

In the Fake Factor method, the formula for the combined Fake Factor $\mathcal{F}$ (Equation 8.1.4) can be expressed as

$$\mathcal{F} = \mathrm{F_W} - \mathrm{R_{QCD}} \cdot (\mathrm{F_W} - \mathrm{F_{QCD}}), \tag{8.5.2.1}$$

assuming that $\mathrm{F_W} \approx \mathrm{F_{QCD}}$. Then,

$$\mathcal{F} \approx \mathrm{F_W}. \tag{8.5.2.2}$$

The estimation of the fake background is repeated using the approximation given by Equation 8.5.2.2 and the values of $\mathrm{F_W}$ from Table 8.1.1.1. Results from the full and simplified Fake Factor method are compared in Table 8.5.2.1.

Larger fake background yields are found when using the full Fake Factor method. Differences between the fake background yields calculated with the two Fake Factor methods are ~3.9% in all final states. The difference between the total event yields is 0.6%. Statistical uncertainties are also found to be higher when the full Fake Factor method is applied, since the combined Fake Factor inherits the uncertainties from the rest of the terms described in Equation 8.1.4. Values listed in Table 8.5.2.1 agree in general within the range statistical uncertainties, statistical uncertainties of the full Fake Factor method can be found in Table 8.4.2.

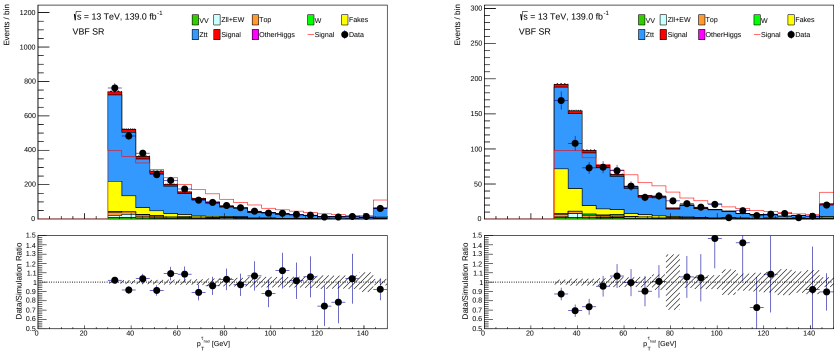
Final state	Assumption	Fake background	Total MC Simulation	s/d
Inclusive	$\mathrm{F_W} \neq \mathrm{F_{QCD}}$	580.89 ± 177.34	3811.93 ± 24.02	0.96 ± 0.02
	$\mathrm{F_W} \approx \mathrm{F_{QCD}}$	559.38 ± 8.14	3790.42 ± 24.02	0.97 ± 0.02
$\tau_e\tau_{\mathrm{had}}$	$\mathrm{F_W} \neq \mathrm{F_{QCD}}$	320.33 ± 98.57	1814.85 ± 16.87	0.99 ± 0.03
	$\mathrm{F_W} \approx \mathrm{F_{QCD}}$	307.94 ± 6.24	1802.46 ± 16.87	1.00 ± 0.03
$\tau_\mu\tau_{\mathrm{had}}$	$\mathrm{F_W} \neq \mathrm{F_{QCD}}$	260.56 ± 78.78	1997.07 ± 17.10	0.94 ± 0.02
	$\mathrm{F_W} \approx \mathrm{F_{QCD}}$	251.44 ± 5.24	1987.96 ± 17.10	0.94 ± 0.02
$\tau_{\mathrm{lep}}\tau_{\text{1-prong}}$	$\mathrm{F_W} \neq \mathrm{F_{QCD}}$	431.65 ± 77.67	2964.47 ± 20.88	0.98 ± 0.02
	$\mathrm{F_W} \approx \mathrm{F_{QCD}}$	414.86 ± 7.80	2947.68 ± 20.88	0.99 ± 0.02
$\tau_{\mathrm{lep}}\tau_{\text{3-prong}}$	$\mathrm{F_W} \neq \mathrm{F_{QCD}}$	149.24 ± 99.91	847.47 ± 11.87	0.88 ± 0.03
	$\mathrm{F_W} \approx \mathrm{F_{QCD}}$	144.51 ± 2.34	842.75 ± 11.87	0.89 ± 0.03

Table 8.5.2.1: Fake background estimates after applying the Fake Factor method where individual Fake Factors (F_i) are either determined for each process that contributes to the fake background ($\mathrm{F_W} \neq \mathrm{F_{QCD}}$) or together using the simplified Fake Factor method ($\mathrm{F_W} \approx \mathrm{F_{QCD}}$). Values are computed in the VBF SR and divided into different final states. The total MC simulation yields are also provided, followed by the simulation-to-data ratio (s/d). Statistical and systematic uncertainties are shown.

Figures 8.5.2.1 and 8.2.3.1 show the p_{T} and $|\eta|$ distributions of the τ_{had} where the fake background contribution is estimated using the simplified and full Fake Factor methods, respectively.

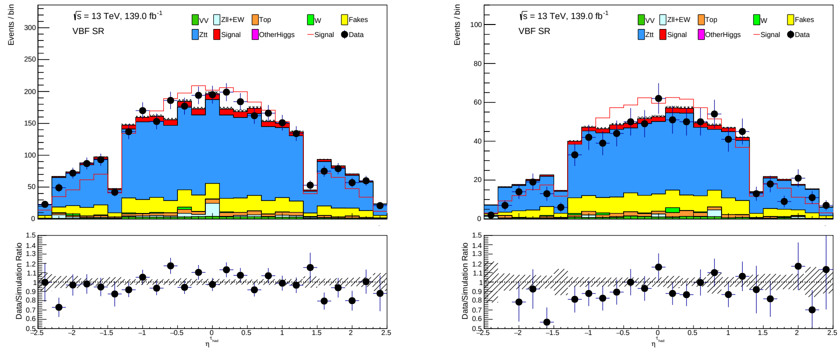
The lower contribution of the fake background is seen when using the simplified method. The differences between the results of both methods is screened by systematic uncertainties as shown in Table 8.4.2.

According to these results, the simplified Fake Factor method provides a solid estimation of the fakes. The application of the full Fake Factor or simplified method will depend on the validity of the assumption that $\mathrm{F_W} \approx \mathrm{F_{QCD}}$.



(a) p_T of the τ_had in the $\tau_\mathrm{lep}\tau_\mathrm{1\text{-}prong}$ channel

(b) p_T of the τ_had in the $\tau_\mathrm{lep}\tau_\mathrm{3\text{-}prong}$ channel



(c) η of the τ_had in the $\tau_\mathrm{lep}\tau_\mathrm{1\text{-}prong}$ channel

(d) η of the τ_had in the $\tau_\mathrm{lep}\tau_\mathrm{3\text{-}prong}$ channel

Figure 8.5.2.1: Kinematic distributions of the hadronically decaying τ-lepton (τ_had) in the VBF signal region (SR) in the 1-prong ($\tau_\mathrm{lep}\tau_\mathrm{1\text{-}prong}$) and 3-prong ($\tau_\mathrm{lep}\tau_\mathrm{3\text{-}prong}$) final states where the fake background contribution is calculated with the simplified Fake Factor method. Only statistical uncertainties are shown. The "Signal" line represents the signal contribution normalized to the total background.

Chapter 9

Neural Network for signal optimization

Chapter 7 described a cut-based approach to select signal events. However, the signal-to-background ratio reached in the VBF-enhanced signal region could be further improved for a precise measurement of the CP properties. A neural network (NN) has been employed to increase the signal sensitivity. A supervised machine learning method, which uses labeled data to train algorithms is applied in the cut-based selection region in order to classify events into background and signal.

Section 9.1 introduces the main concepts of a NN, Section 9.2 describes how a NN is defined, how does it work and how its performance can be improved. Section 9.3 displays the NN settings used in this analysis and two methods to optimize the NN performance: hyperparameter and feature optimization. Finally, Section 9.4 presents the results of the NN applied to the VBF SR.

9.1 Introduction to NN

Neural Networks are inspired from neurons in human brains, in the sense that they can recognize underlying relationships from data. These are non-linear machine learning models that have been applied to problems in numerous physics branches [100]. As most Machine Learning (ML) algorithms, NNs can help with *classification* (assigning class labels to data) or *regression* problems (make predictions of a continuous variable). This analysis addresses a classification problem.

Usually, a NN is made up of *layers*, consisting of a set of units called *neurons*, each unit i from a layer produces a scalar output $u_i(\mathbf{x})$ from an *input vector* $\mathbf{x}$ ($\dim(\mathbf{x}) = n$), which is used as the input for the next layer. Usually, the output of each neuron is defined using the same non-linear transformation $\sigma_i(z)$, called *activation function*, where z is a linear function including the weights ($\mathbf{w}_i$) that measure the relative importance of the input vector ($\dim(\mathbf{w}_i) = n$). Based on this description, the output is given by

$$u_i(\mathbf{x}) = \sigma_i(\mathbf{w}_i \cdot \mathbf{x} + b_i), \tag{9.1.1}$$

where b_i is the *bias* of neuron i.

Figure 9.1.1 A illustrates the mechanism of a neuron, while Figure 9.1.1 B shows an example architecture of a NN. Each circle represents a neuron, and the neurons of the first layer (input layer) take the input vector and produce outputs that serve as the input for the next layer of neurons (hidden layer), successively until the last layer (output layer) is reached. The output layer is generally a classifier.

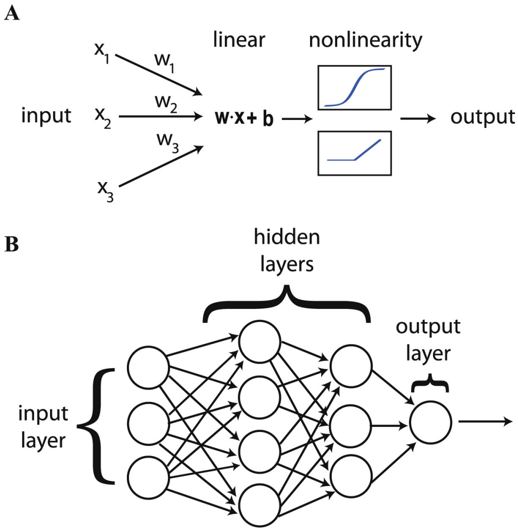


Figure 9.1.1: Schematic of a neural network (NN). **A** shows the mechanism of a neuron in a NN, starting from an input vector $\mathbf{x} = (\mathrm{x}_1, \mathrm{x}_2, \mathrm{x}_3)$, each neuron weights ($\mathbf{w} = (\mathrm{w}_1, \mathrm{w}_2, \mathrm{w}_3)$) the importance of each input using a linear function ($\mathrm{w} \cdot \mathrm{x} + \mathrm{b}$) which serves as the input for the non-linear function that generates the output of the neuron. Figure **B** shows an example architecture of a NN. Neurons are represented by circles. The first layer (input layer), intermediate layers (hidden layers) and the last layer (output layer) are shown [100].

9.2 Training of NNs

The previous section introduced the basic concepts of a NN. The *training* of a supervised machine learning model refers to the "learning" (determination) process in which the weights and biases are estimated. In general, this consists on constructing a *loss function* that calculates the error on the prediction at the output layer, then this is minimized with the application of *gradient descent* algorithms that calculate the weights and biases that minimize the loss function, finally the *backpropagation* process is applied to propagate backwards (to the first layer) the errors for the model to learn.

9.2.1 Loss function

The training of a NN starts by choosing a *loss function* (L) (also called *cost function*), which evaluates the performance of the ML model on simulation. The loss function used in this analysis is the cross-entropy.

Consider a set of M labeled classes, indexed by $m \in \{0, ..., M-1\}$ (e.g. signal, background, etc.). For all M labels, a true value y_i is assigned, where i ranges over the data points (in a binary classification problem $y_i \in \{0, 1\}$). For each data point $(\mathbf{x}_i, y_i)$ and class m, a variable y_{im} is defined as

$$y_{im} = \begin{cases} 1, & \text{if} \quad y_i = m \\ 0, & \text{otherwise.} \end{cases} \tag{9.2.1.1}$$

The likelihood that the NN assigns a data point i to the category m is given by $\hat{y}_{im}(\mathbf{w})$.For M labels and n data points the categorical cross-entropy loss-function is given by [100]

$$L(\mathbf{w}) = -\sum_{i=1}^{n}\sum_{m=0}^{M-1} y_{im}\log(\hat{y}_{im}(\mathbf{w})) + (1-y_{im})\log(1-\hat{y}_{im}(\mathbf{w})). \tag{9.2.1.2}$$

Highest resemblance between the model and the data translates into the minima of the loss function L.

9.2.2 Loss-function optimizers

The cost function optimization is done by finding the weights $\mathbf{w}$ and biases b that minimize the loss function. For this, Gradient Descent (GD) method is commonly used [100]. GD iteratively updates the weights of the model towards the minima of the loss-function. It calculates the gradient of the cost-function $\nabla_{\mathbf{w}}L(\mathbf{w})$ with respect to the weights and the biases, and moves in the direction of the negative of the gradient. The learning rate α is introduced to control how big is the step. Mathematically:

$$\begin{aligned} \mathrm{w}_i &\to \mathrm{w}_i - \alpha\frac{\partial L(\mathbf{w},b)}{\partial \mathrm{w}_i} \quad \text{for} \quad i=0,...,n,\\ b &\to b - \alpha\frac{\partial L(\mathbf{w},b)}{\partial b}, \end{aligned} \tag{9.2.2.1}$$

with $\mathbf{w} = [\mathrm{w}_1, ..., \mathrm{w}_i]$ represents the weight array of a particular neuron, length($\mathbf{w}$)=length($\mathbf{x}$)= n where $\mathbf{x}$ is the input vector. The choice of α is discussed in Section 9.3.5. In addition, to optimize the calculation of the minimum a *learning rate decay* (λ_d) can be introduced as

$$\alpha_{i+1} = \frac{\alpha}{1 + i\cdot\lambda_d}, \tag{9.2.2.2}$$

where i represents the iteration step.

However, GD has some limitations, for example when the loss curve has more than one local minima the GD has a poor performance, it also depends on the initial conditions, depending on which one or another local minima can be reached [100]. To remedy these limitations, GD can be modified.

One such modification is the Stochastic GD (SGD), which incorporates stochasticity into GD. In GD, for each epoch, the input data set is run through the NN, the model finds the predictions, calculates the loss-function (how far is the prediction from targets) and its gradient; and then backpropagates this error to update the weights. For large data sets, this is computationally expensive. SGD splits randomly the total data set into small subsets of data (*batches*), and for each batch, the whole NN procedure is repeated. This introduces stochasticity so that the probability of ending in the same local minima is lower. Furthermore, the computation time is shorter since the gradient is calculated for the batches and not for the entire data set.

SGD is usually extended by introducing a "momentum" (γ), which functions like a "memory" of the direction of the minima [100]. This can help with moving towards the global minimum in case of getting stuck in a local minimum. The momentum has a range $0 < \gamma < 1$, and leads to modification of Equation 9.2.2.1 as

$$\begin{aligned} \mathrm{w}_{i+1} &= \mathrm{w}_i - \alpha\frac{\partial L(\mathbf{w},b)}{\partial \mathrm{w}_i} - \gamma\mathrm{w}_i \quad \text{for} \quad i=0,...,n,\\ b &\to b - \alpha\frac{\partial L(\mathbf{w},b)}{\partial b} - \gamma b. \end{aligned} \tag{9.2.2.3}$$

The momentum term increases the descent in directions with small and constant gradients, reducing the path to reach the minima. To adaptively change the step size to the loss function, variations of the SGD with momentum have been introduced. One of them is the ADAM [105](Adaptive Momentum Estimator), which uses the first and second moments of the gradient to adapt the learning rate for each weight, the update rules for ADAM are [100, 105]

$$\begin{aligned}
\mathbf{g}_t &= \nabla_\theta L(\theta), \\
\mathbf{m}_t &= \beta_1 \mathbf{m}_{t-1} + (1-\beta_1)\mathbf{g}_t, \\
\mathbf{s}_t &= \beta_2 \mathbf{s}_{t-1} + (1-\beta_2)\mathbf{g}_t^2, \\
\hat{\mathbf{m}}_t &= \frac{\mathbf{m}_t}{1-\beta_1^t}, \\
\hat{\mathbf{s}_t} &= \frac{\mathbf{s}_t}{1-\beta_2^t}, \\
\theta_{t+1} &= \theta_t - \alpha_t \frac{\hat{\mathbf{m}}}{\sqrt{\hat{\mathbf{s}}_t} + \epsilon}.
\end{aligned} \tag{9.2.2.4}$$

The objective is to find the minimum of the loss function $L(\theta)$ with respect to the parameters θ (weights and bias), thus $\mathbf{g}_t$ is defined as the gradient with respect to θ of the loss function at a timestep t. Then the exponential moving averages of the gradient ($\mathbf{m}_t = \mathbb{E}[\mathbf{g}_t]$) and the squared gradient ($\mathbf{s}_t = \mathbb{E}[\mathbf{g}_t^2]$) are updated ($\mathbb{E}$ refers to the expected value). The hyperparameters $\beta_1 \in [0,1)$ and $\beta_2 \in [0,1)$ control the "memory" lifetime of the first and second moment, respectively. The moving averages are initialized such that the moment estimates are biased towards zero [105], to correct this, the bias-corrected estimates $\hat{\mathbf{m}}_t$ and $\hat{\mathbf{s}}_t$ update rules are introduced, where β_1^t, β_2^t represent β_1, β_2 to the power of t. The parameter θ update rule contain the learning rate α_t and a regularization constant ϵ to prevent divergences.

9.2.3 Backpropagation

Backpropagation is an algorithm that calculates the gradients of the weights at each step by exploiting the layered structure of the NN [100]. Consider a NN with L layers $l = [1, ..., L]$ where the weight of the connection between the neuron k in layer $l-1$ and neuron j in the layer l is denoted by w_{jk}^l. The bias of neuron j in layer l is represented by b_j^l. The activation function of neuron j in layer l, u_j^l, is defined as

$$u_j^l = \sigma\left(\sum_k \mathrm{w}_{jk}^l u_k^{l-1} + b_j^l\right) = \sigma(z_j^l), \tag{9.2.3.1}$$

which is the expanded version of Equation 9.1.1. The error of the neuron j at the output layer L is defined as the partial derivative of the loss function with respect to the weighted input z_j^l (defined in parenthesis of Equation 9.2.3.1), given by

$$\Delta_j^L = \frac{\partial L}{\partial z_j^L}, \tag{9.2.3.2}$$

therefore, the error of the neuron j in layer l can be written as [100]

$$\Delta_j^l = \frac{\partial L}{\partial z_j^l} = \frac{\partial L}{\partial u_j^l}\frac{\partial u_j^l}{\partial z_j^l} = \frac{\partial L}{\partial u_j^l}\frac{\partial \sigma(z_j^l)}{\partial z_j^l} = \frac{\partial L}{\partial u_j^l}\sigma'(z_j^l), \tag{9.2.3.3}$$

or

$$\Delta_j^l = \frac{\partial L}{\partial z_j^l} = \frac{\partial L}{\partial b_j^l}\frac{\partial b_j^l}{\partial z_j^l} = \frac{\partial L}{\partial b_j^l}, \tag{9.2.3.4}$$

where the relation $\partial b_j^l / \partial z_j^l = 1$ was used. This are the two first backpropagation equations [100]. Considering that Δ_j^l only depend on neurons in layer l through the activation of neurons in layer $l + 1$ and using the chain rule again, the third backpropagation equation is given by [100]

$$\Delta_j^l = \frac{\partial L}{\partial z_j^l} = \sum_k \frac{\partial L}{\partial z_k^{l+1}}\frac{\partial z_k^{l+1}}{\partial z_j^l} = \sum_k \Delta_k^{l+1}\frac{\partial z_k^{l+1}}{\partial z_j^l} = \sum_k \Delta_k^{l+1}\mathrm{w}_{kj}^{l+1}\sigma'(z_j^l), \tag{9.2.3.5}$$

where the relation $\mathrm{w}_{kj}^{l+1} = \frac{\partial z_k^{l+1}}{\partial u_j^l}$ is used. Finally, the fourth backpropagation equation reads

$$\frac{\partial L}{\partial \mathrm{w}_{jk}^l} = \Delta_j^l u_k^{l-1}. \tag{9.2.3.6}$$

Equations 9.2.3.3, 9.2.3.4, 9.2.3.5 and 9.2.3.6 define the four backpropagation equations [100]. They can be combined with Equation 9.2.3.1 to generate an algorithm that calculates the gradient with respect to all weights and biases [106] following a procedure:

1 Calculation of the activation functions in the first layer for all neurons u_j^1

2 Computation of u_j^l and z_j^l for all subsequent layers using Equation 9.2.3.1 (the information moves only in the forward direction from the input nodes "feed-forward")

3 Calculation of the error in the output layer L using Equation 9.2.3.3

4 Propagation of the error backwards "backpropagation" using Equation 9.2.3.5 and calculation Δ_j^l for all layers

5 Calculation of the gradient of the loss function with respect to the weights w_{jk}^l and bias b_j^l with Equations 9.2.3.6 and 9.2.3.4.

9.2.4 Regularization

When training a ML model two phenomena might occur: *overfitting* or *underfitting.* Overfitting happens when the model starts to capture the noise in data i.e., points that do not represent the overall properties of the data but of individual events [101]. Underfitting happens when the model is not capable to capture the trend of data, simplifying it.

ML models usually face with overfitting problems. To overcome overfitting and improve the interpretability of the model (exclusion of irrelevant variables not associated with the response [102]), the cross entropy is expanded with further terms, this are the "regularization terms".

Some of these regularization terms penalize large valued weights $\mathbf{w} = [w_1, ..., w_p]$ by shrinking them towards zero. Two commonly used regularization techniques are Lasso (L1) and Ridge (L2) regularization [100]. L1 and L2 regularization terms are defined as

$$\begin{aligned} \text{L1:} \quad & \lambda \sum_{i=0}^{p} |w_i|, \\ \text{L2:} \quad & \lambda \sum_{i=0}^{p} w_i^2, \end{aligned} \tag{9.2.4.1}$$

where $\lambda \geq 0$ is called the "tuning parameter". The chosen term of regularization is added to Equation 9.2.1.2. When applying L1 regularization the trend is to obtain sparse vectors, that is, small weights will tend to vanish. Therefore, L1 regularization is chosen if the objective is reducing the number of weights. This is also good for feature (input variable) selection. In contrast, L2 regularization tends not to favour sparse vectors since it tries to keep all the weights small.

Another frequently used regularization technique is the *dropout* [100, 104], which addresses overfitting by randomly removing neurons during each *epoch* (one complete pass of the training data set through the algorithm [103]) of the training . This prevents internal correlations between neurons [104]. To set the dropout rate of the neurons, a parameter is introduced to the algorithm. This parameter gives the probability that each neuron gets dropped at a particular epoch.

9.3 NN discriminant for the signal region

This Section presents the methodology used to train a NN to optimize the signal-to-background ratio in the VBF SR. The process of developing a NN discriminant for the signal is divided into the following steps:

a) Prepare the data sets

b) Define the machine learning model

c) Train the model for a fixed set of hyperparameters

d) Evaluate the model performance using the test and validation data sets

e) Optimize the hyperparameters

f) Evaluate the final performance

9.3.1 Pre-procesing data

The training region of the NN is the inclusive VBF SR defined in Chapter 7 using a cut-based approach that yields to a signal-to-background ratio of 0.044.

To test the performance of the NN model, the simulation data set is split into three subsets: *training*, *testing* and *validation*. The training set is used during the training of the model and estimation of the parameters. The performance of the model is then evaluated using the testing set, where the error prediction from the model is estimated. To avoid overfitting in the test set, the model is applied to the validation set [107, 108].

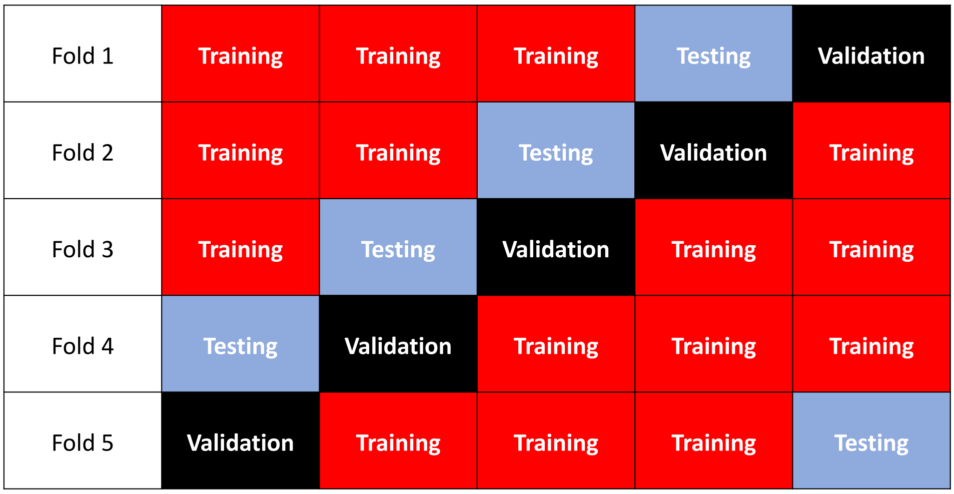
For this analysis the input variables (features) are for example the transverse momentum of the τ_{had} and the azimuthal angle between the two jets among others, and the "labels" are signal or background (considering a binary classifier). A multi-class NN is also studied. For each label, the NN estimates the probability that an event is identified with that label. The NN classifies an event into the label with the highest probability. To accomplish this, true labels provided by the MC simulations are used.

To utilize the full statistics of the data sets, a k-fold cross-validation is applied [107]. This method splits each data set into k parts (folds) with the same number of events. In this analysis $k = 5$ folds are used. The split is done using a random number generator. Each fold is further split into $k - 2$ slices for training, one slice for testing and one slice for validation i.e. 60%, 20% and 20% of events, respectively. This way, on each fold the testing and validation sets are changed, ensuring that the total available statistics are used for testing and validation, as illustrated in Figure 9.3.1.1.

This analysis applies *feature scaling* to the input variables. ML algorithms might attribute higher importance to variables with certain distributions, thus giving them a decisive role during the training of the model [109]. To bring all features to the same standing, improve the stability and the modeling of the input variables [110] a feature scaling is applied. This analysis uses the *standard* scaling, which scales each feature x such that its distribution is centered around zero with a standard deviation of one, i.e.

$$x_{\text{new}} = \frac{x - \mu}{\sigma}, \tag{9.3.1.1}$$

where μ is the mean of the distribution and σ its standard deviation.



Fold 1	Training	Training	Training	Testing	Validation
Fold 2	Training	Training	Testing	Validation	Training
Fold 3	Training	Testing	Validation	Training	Training
Fold 4	Testing	Validation	Training	Training	Training
Fold 5	Validation	Training	Training	Training	Testing

Figure 9.3.1.1: 5-fold cross-validation schematic. The data set is split into five slices on each fold, three slices for training (red), one for testing (blue) and one for validation (black).

9.3.2 NN Model

The baseline architecture of the NN is set as three hidden layers with 128 neurons each and an output layer with two neurons. The chosen activation function for each neuron of the input and hidden layers is the Rectified Linear Unit (ReLU) function, defined as [100]

$$\text{ReLU}: \quad \sigma(z_j^l) = z_j^{l+} = \max(0, z_j^l) \tag{9.3.2.1}$$

following notation from Section 9.2.3. The output layer activation function is a *softmax* function, defined as [100]

$$\text{Softmax}: \quad \sigma(z_j^L)_{im} = e^{(z_j^L)_{im}} \cdot \left(\sum_{m'=0}^{M-1} e^{(z_j^L)_{im'}} \right)^{-1}, \tag{9.3.2.2}$$

where $(z_j^L)_{im}$ is the linear function including the weights and biases of neuron j at the output layer L of the data point i with class m. The softmax function normalizes the weighted input vector $(z_j^L)_{im}$ to a probability distribution. The softmax activation function of the term i of the input vector $\sigma(z_j^l)_{im}$ represents the probability that $(z_j^l)_{im}$ is in class m, this is why the denominator of Equation 9.3.2.2 is a sum over all labels.

The loss function is set to be the categorical cross-entropy defined in Equation 9.2.1.2 to which the penalty term L2 described in Equation 9.2.4.1 was added, with a tuning parameter of $\lambda = 10^{-5}$ and a fixed learning rate decay of $\lambda_d = 0.001$. The algorithm chosen for the optimization of the NN is the ADAM optimizer described in Section 9.2.2, with a learning rate of $\alpha = 10^{-3}$ and a batch size of 512. The number of epochs is set to 50.

The number of layers, neurons per layer, the tuning parameter, the learning rate, and the batch size represent the hyperparameters of the ML model. These are the parameters that control the training process and impact the output that result from it [111].

9.3.3 Figures of merit

In Section 9.3.1 it was discussed that the total samples are divided into three subsets: training, testing and validation. This section discusses the metrics that were used to evaluate the NN performance.

After the training, the performance of the model is evaluated on the testing data set, analyzing the behavior of the loss curve for both cases. Furthermore, the *categorical accuracy* is used for the ML model evaluation. This metric measures how often the model gets the right prediction, given by counting how often the predicted label matches to the true label:

$$\text{Categorical Accuracy} = \frac{\text{Number of correct predictions}}{\text{Number of total predictions}}. \tag{9.3.3.1}$$

Evaluation of the performance of the model in the validation set is more complex. In the pre-processing step, the k-fold cross-validation was applied. In each fold, the NN score is calculated, that is, the output of the output layer that represented the probability that the data point i was from class m. For example, when using a binary classifier, if signal events are assigned to $m = 1$ and background events to $m = 0$ and a data point gets a NN score of e.g. 0.95, this point will likely get classified as signal.

The k number of NN scores are averaged and combined into the final NN score distribution [102]. Finally, the NN score distribution is scanned, calculating for each NN score (calculation is done with a step size, not infinite resolution) the significance (*Sig.*) defined as

$$Sig. = \frac{s}{\sqrt{s+b}} \tag{9.3.3.2}$$

where s and b represent the number of signal and background events after the value of the NN score, respectively. The NN score value giving the maximum significance is used to select the optimized NN model.

9.3.4 Feature selection

Selecting the features used in the NN model can optimize its performance. The larger number of features leads to a higher dimension of the weighted input vector, thus resulting in higher computation times.

However, more features can lead to a better model performance, since these are the properties that characterize the physics processes. Nevertheless, in some cases the high correlation between the variables leads to the fact that the elimination of some of these variables does not impact the final result. Two analyses were done in order to select the optimal set of features.

First the Permutation Importance test [118] which is an inspection technique of the input variables of a ML model. In this context, the importance of a variable lies in how the significance (figure-of-merit) changes after randomly shuffling its values. In this analysis, the values of each variable are randomly shuffled 10 times. Thereby the variable will lose its physical character i.e., the relation between the feature and the process disappears [118]. The importance is calculated as

$$\text{Importance} = \frac{Sig_{\text{Original}}}{Sig_{\text{Perm}}}, \tag{9.3.4.1}$$

where Sig_{Original} is the significance using the original values of the variable, and Sig_{Perm} is the average of the significance values obtained from shuffling i.e. $Sig_{\text{Perm}} = \frac{1}{i}\sum_{s=1}^{i} Sig_i$.

Secondly, for each pair of variables in the entire set of features, the Pearson correlation coefficient [119] (ρ) is calculated. ρ measures the linear correlation between two variables X and Y as the ratio of the covariance between the variables ($cov(X, Y)$), normalized to the product of their standard deviations (σ_X and σ_Y):

$$\rho_{X,Y} = \frac{cov(X, Y)}{\sigma_X \sigma_Y}. \tag{9.3.4.2}$$

If the correlation coefficient is close to 1, the variables are highly correlated. If the correlation coefficient is close to -1, variables are anti-correlated.

The set of selected features depends on the results of both studies: Permutation Importance and the Correlation test. First, the Permutation Importance test is implemented, which provides the importance for each variable. Then the Correlation within this set of variables is calculated, listing the pairs with $|\rho| \geq 0.75$. In this list, for each pair, the variable with the lowest ranking in the Permutation Importance list is discarded. The application of the method is discussed in Section 9.4.

9.3.5 Hyperparameter optimization

Hyperparameters are parameters that configure the learning process of the model. Thus, optimizing these parameters leads to a general optimization of the performance. The list of hyperparameters, their baseline values and values used during optimization are given in Table 9.3.5.1.

The *Optuna* Framework [112] is used in this analysis for hyperparameter optimization, which uses the Tree-structured Parzen Estimator (TPE) [113], a Bayesian optimization algorithm.

The Bayesian approach is characterized for using results from past evaluations to construct a probabilistic model that maps the hyperparameters (x) to a probability of a score (y) on the *objective function* which is the function to minimize (the cross-entropy). These models work by first selecting the hyperparameters that have the best performance on the *surrogate* function and then evaluating them on the objective function, the result of the evaluation is used to update the surrogate function [114, 115]. These are the so-called *surrogate* models.

The surrogate function is also called *response surface*, represented as $P(x|y)$ and is updated until a fixed number of iterations (10 iterations are used in this analysis). The criteria applied to select the best performing hyperparameters from the surrogate function is to maximize the *selection function*, commonly chosen to be the *Expected Improvement* [115], defined as

$$EI^*(x) = \int_{-\infty}^{y^*} (y - y^*) P(y|x) dy, \tag{9.3.5.1}$$

where y^* is a threshold of the objective function, x the set of hyperparameters in the iteration, y the value of the objective function using x, and $P(y|x)$ the surrogate probability model, defined as the probability of y given x, in TPE this is

$$P(y|x) = \frac{P(x|y) \cdot P(y)}{P(x)}, \tag{9.3.5.2}$$

also known as the Bayes' Rule. In turn, the probability of the hyperparameters given the score on the objective function is expressed as

$$P(x|y) = \begin{cases} l(x) & \text{if } y < y^* \\ \\ g(x) & \text{if } y \geq y^* \end{cases} \tag{9.3.5.3}$$

where $l(x)$ is the probability density function in which the value of the objective function is lower than the threshold y^* and $g(x)$ the density in which the value of the objective function is greater than this threshold. The threshold y^* is based on a fixed quantile of the observed scores y [113, 116, 117].

It is found that Equation 9.3.5.1 can be written in terms of the ratio $l(x)/g(x)$, and that by maximizing this value leads to the best performing hyperparameter set [113, 114, 116, 117], which is evaluated in the objective function. If the surrogate function is correct, this hyperparameter set should yield to a better performance than the set used in the previous iteration.

This process is usually very time-consuming therefore, a *median pruner* was added. In general, a pruner is defined as an algorithm that allows terminating the learning process prematurely. The median pruner minimizes the objective function by pruning the best intermediate result of the trial if it is lower than the median of intermediate results from previous trials in the same step [117].

Hyperparameter	Initial value	Values considered for optimization
Number of layers	3	[2,10]
Neurons per layer	128	128, 256, 512
L2 tuning parameter	10^{-5}	$[10^{-6}, 10^{-3}]$
Learning rate	10^{-3}	$[10^{-6}, 10^{-2}]$
Batch size	512	128, 512, 1024, 2048, 4096, 8192

Table 9.3.5.1: Hyperparameters, their initial values and the values or ranges ($[\cdot,\cdot]$) in which the model is evaluated.

9.4 Results

This Section presents the results of the binary NN optimization described in Section 9.3 applied to the VBF SR defined in Chapter 7.

First, the model is trained using the whole set of features and initial values of hyperparameters listed in Table 9.3.5.1. The set of features was chosen from a combination of the features studied in similar analyses, [66], [11]. The full list of variables can be found in Table 9.4.1.

The results of the Permutation Importance and the Correlation studies finalize the set of features. Figures 9.4.1 and 9.4.2 illustrate the outputs of this studies and Table 9.4.2 the optimized list of features that will be used as input for the binary NN. The significance changed from 5.74 to 5.72 when using the full list of variables compared to the setup using the optimized list of variables. Thus, using the optimized set of features does not result in an important reduction in the significance ($\sim 0.3\%$), justifying the decision to use the optimized feature list.

Variable	**Definition**
$\ell\eta$ centrality	Centrality of the light lepton (ℓ) with respect to the two leading jets
$E_{\mathrm{T}}^{\mathrm{miss}}\phi$ centrality	Centrality of the missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}}$): $E_{\mathrm{T}}^{\mathrm{miss}}\phi$ centrality $= \frac{\sqrt{2}r+s}{\sqrt{r^2+s^2}}$, $r = \frac{sin(\phi(E_{\mathrm{T}}^{\mathrm{miss}})-\phi(\tau_{\mathrm{had}}))}{sin(\phi(\ell)-\phi(\tau_{\mathrm{had}}))}$, $s = \frac{sin(\phi(\ell)-\phi(E_{\mathrm{T}}^{\mathrm{miss}}))}{sin(\phi(\ell)-\phi(\tau_{\mathrm{had}}))}$
$m_{\tau\tau}^{\mathrm{MMC}}$	Invariant di-τ mass reconstructed using the so-called missing mass calculator (MMC) [120]
$\tau\eta$ centrality	Centrality of the τ_{had} with respect to the two leading jets
$\Delta R_{\ell,\tau_{\mathrm{had}}}$	Angular distance between the two τ-candidates
$m_{\tau\tau}^{\mathrm{vis}}$	Visible mass calculated from the two τ-candidates
$p_{\mathrm{T}}^{\mathrm{total}}$	Transverse momentum of the vectorial sum of all objects in an event: ℓ, τ_{had}, leading jet (j_1), sub-leading jet (j_2) and $E_{\mathrm{T}}^{\mathrm{miss}}$
$\Delta\eta_{\ell,\tau_{\mathrm{had}}}$	Difference in pseudorapidity η between the lepton and the τ_{had}
$m^{\ell h j_1}$	Visible mass calculated from the two τ-candidates and the leading jet
$m_{\mathrm{T}}^{\ell\mathrm{MET}}$	Transverse mass calculated from the ℓ and the $E_{\mathrm{T}}^{\mathrm{miss}}$ (MET)
$p_{\mathrm{T}}^{\mathrm{H}}$	Higgs transverse momentum
$\frac{\Delta p_{\mathrm{T}}^{\ell,\tau_{\mathrm{had}}}}{\sum p_{\mathrm{T}}^{\ell,\tau_{\mathrm{had}}}}$	Ratio of the transverse momentum difference between the two τ-candidates and the scalar sum of the transverse momenta of the two τ-candidates
$\eta^{j_1} \times \eta^{j_2}$	Signed product of the pseudorapidity (η) of the leading (j_1) and sub-leading jets (j_2)
$\Delta\phi^{\ell,\tau_{\mathrm{had}}}$	Angular distance in ϕ-direction between the two τ-candidates
$\Delta\eta_{jj}$	Difference of η between the two leading jets
$\frac{E_{\mathrm{T}}^{\mathrm{miss}}}{p_{\mathrm{T}}^{\tau_{\mathrm{had}}}}$	Ratio of the missing transverse energy and the transverse momentum of the τ_{had}
$\frac{E_{\mathrm{T}}^{\mathrm{miss}}}{p_{\mathrm{T}}^{\ell}}$	Ratio of the missing transverse energy and the transverse momentum of the light lepton (ℓ)
$\frac{p_{\mathrm{T}}^{\ell}}{p_{\mathrm{T}}^{\tau_{\mathrm{had}}}}$	Ratio of the transverse momenta of the two τ-candidates
$p_{\mathrm{T}}^{\ell,\tau_{\mathrm{had}}}$	Transverse momentum of the two τ-candidates system
$m_{\mathrm{T}}^{\tau_{\mathrm{had}},\mathrm{MET}}$	Transverse mass calculated from the leading τ-candidate and the missing transverse energy ($E_{\mathrm{T}}^{\mathrm{miss}} = MET$)
p_{T}^{Hjj}	Transverse component of the vector-summed momenta of the two τ candidates, missing transverse momentum and two jets
p_{T}^{jj}	Transverse momentum of the di-jet system
$\Delta\phi_{jj}$	Difference of the angle ϕ between two jets
$p_{\mathrm{T}}^{j_2}$	Transverse momentum of the sub-leading jet

Table 9.4.1: List of features used in the neural network training analysis for signal optimization.

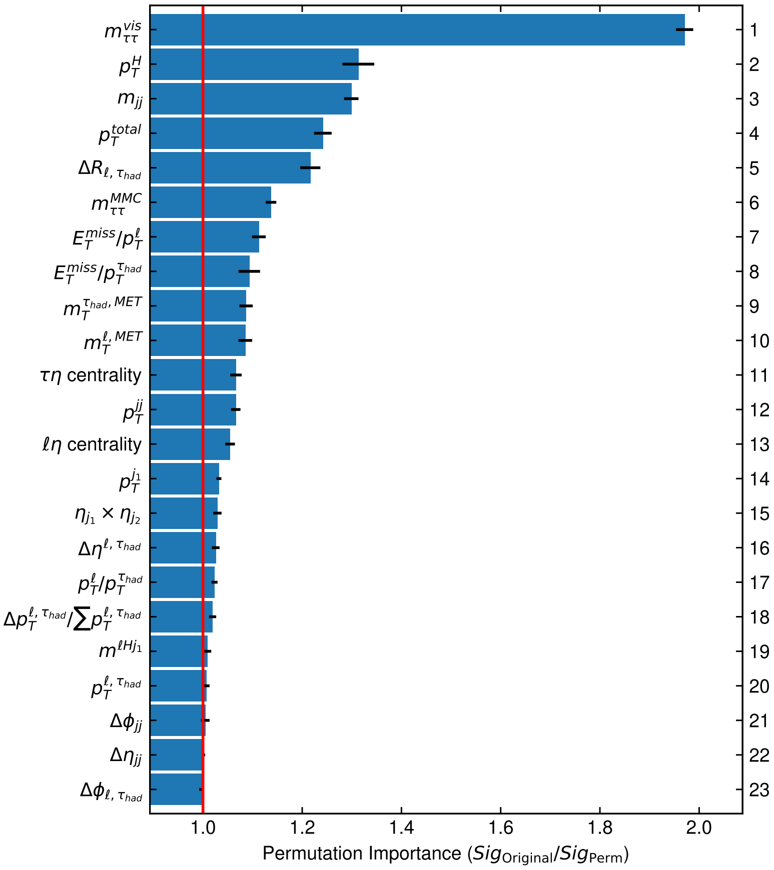


Figure 9.4.1: Ranking of variables sorted according to the Permutation Importance. Full Permutation Importance is given as the ratio of the original significance (Sig_{Original}) and the average significance after permuting the variable (Sig_{Perm}), see Section 9.3.4 for more details. The red line shows where Permutation importance is equal to one. Statistical uncertainties derived from the iterations are shown in black. Variables are defined in Table 9.4.1.

The NN hyperparameter set that provide the best performance are listed in Table 9.4.3. The maximum performance of the binary NN is found at a NN score of 0.92 resulting in a significance of $Sig. = 5.93$ ($s_{max} = 79.55$ and $b_{max} = 67.02$) and a signal-to-background ratio of $s/b = 1.19$. Figure 9.4.3a shows the NN score distribution in the VBF SR including all processes, where, the signal contribution is normalized to total background contribution. Figure 9.4.3b shows the NN score distribution for the training and testing sets separately in the VBF SR, signal contribution is normalized to total background contribution.

The accuracy and loss functions for the training and testing data sets are shown in Figure 9.4.4. In general good agreement and expected behavior is seen between the two sets. The loss curve tends to minimize over the epochs, reducing the risk of overfitting. The accuracy of the model fluctuates around 0.76 for the testing and training data sets. The testing set shows larger fluctuations than the training set.

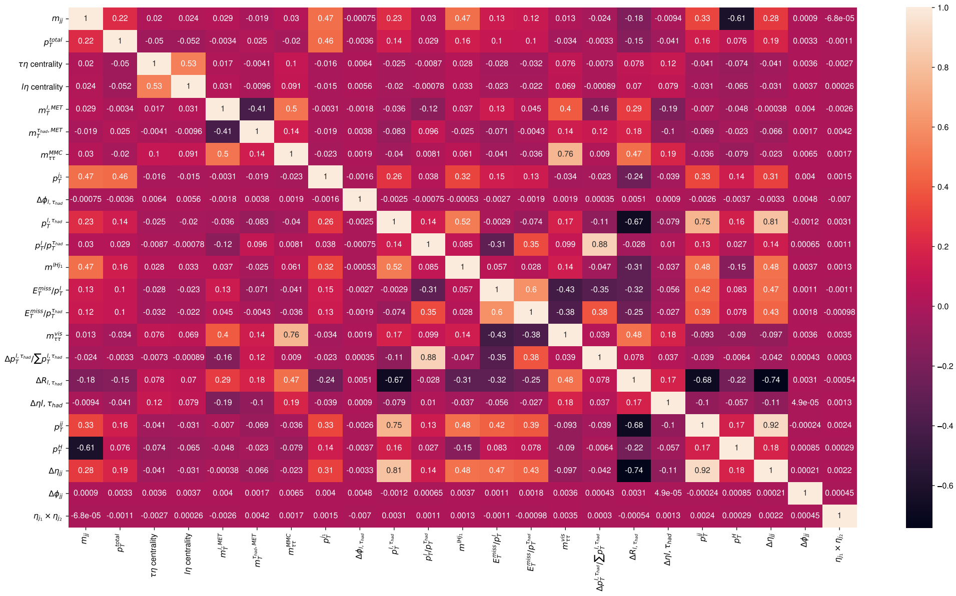


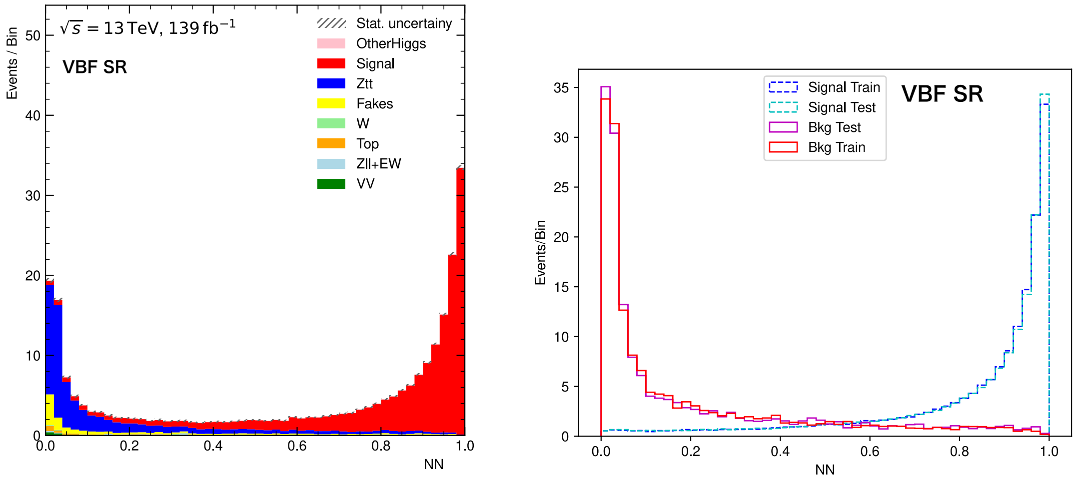
Figure 9.4.2: Pearson Correlation Coefficients (ρ) calculated for each pair of variables from Table 9.4.1 calculated in the signal region.

Variable	Importance
$m_{\tau\tau}^{\text{vis}}$	2.17
p_{T}^{H}	1.36
m_{jj}	1.34
$p_{\text{T}}^{\text{total}}$	1.26
$\Delta R_{\ell,\tau_{\text{had}}}$	1.25
$E_{\text{T}}^{\text{miss}}/p_{\text{T}}^{\ell}$	1.11
$E_{\text{T}}^{\text{miss}}/p_{\text{T}}^{\tau_{\text{had}}}$	1.10
$m_{\text{T}}^{\tau_{\text{had}},MET}$	1.08
$m_{\text{T}}^{\ell,MET}$	1.07
$\tau\eta$ centrality	1.064
p_{T}^{jj}	1.062
$\ell\eta$ centrality	1.04
$p_{\text{T}}^{j_2}$	1.03
$\eta_{j_1}\times\eta_{j_2}$	1.02
$\Delta\eta^{\ell,\tau_{\text{had}}}$	1.02
$p_{\text{T}}^{\ell}/p_{\text{T}}^{\tau_{\text{had}}}$	1.01

Table 9.4.2: Optimized set of features for the binary neural network classifier. Variables are sorted by their Importance. Upscript MET refers to the missing transverse energy $E_{\text{T}}^{\text{miss}}$.

Hyperparameter	Optimized value
Number of layers	4
Neurons per layer	512
L2 tuning parameter	4.22×10^{-6}
Learning rate	3.6×10^{-3}
Batch size	128

Table 9.4.3: Optimized set of hyperparameters for the binary neural network classifier.



(a)

(b)

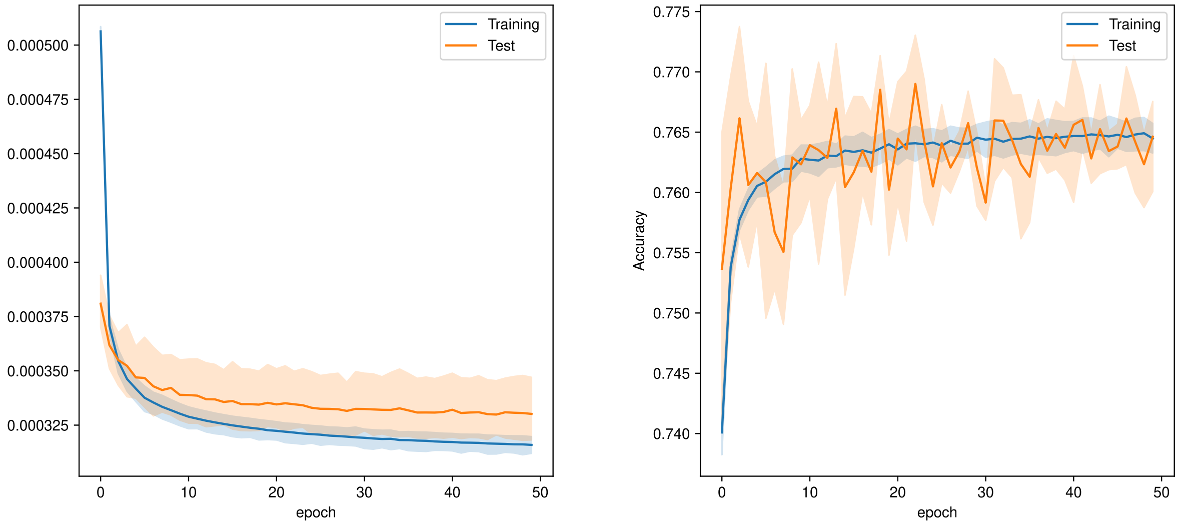
Figure 9.4.3: Distributions of the NN score of the binary (signal and background) neural network after applying feature and hyperparameter optimization in the VBF SR. (a) shows the different classified processes and (b) compares the classified signal in the training and testing sets with the background from the test data set. Signal contribution is normalized to total background contribution in both Figures.

Furthermore, the NN training and analysis is repeated using a multi-class NN, i.e. using more than two output nodes. The multi-class NN is trained with four output nodes: signal, Z$\rightarrow \tau\tau$, ggF and the rest of the background processes. After the feature and hyperparameter optimization, the obtained significance for the best performing settings was 5.92. The signal-to-background ratio provided by multi-class NN is 0.98, thus showing a diminished performance in comparison to the binary NN setup. Therefore, the binary NN setup is chosen for the signal optimization.

The relative difference of the signal-to-background ratio was optimized by a 96.3% compared to the cut-based result, (0.044) in the VBF SR using the binary NN setup.

Based on the NN score that provides the maximum significance in the best performing NN configuration (the binary NN) two regions are defined: the signal region (SR) with NN score$>$ 0.92 and the low NN CR with NN score $<$ 0.92, both including the Preselection and VBF selection described in Chapter 7. The expected SR yields of all processes are listed in Table 9.4.4 for the binary NN setup.

The modelling of the NN is validated by checking the input variables in the low NN CR, which



(a)

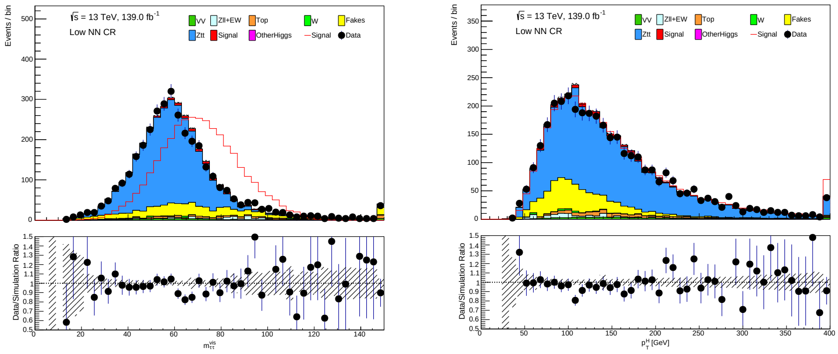
(b)

Figure 9.4.4: Loss curve (a) and accuracy (b) of the best performing NN (binary NN), after feature and hyperparameter optimization in the VBF SR. Only statistical uncertainties are shown, derived from the statistical uncertainties of the variable distributions and VBF SR yields.

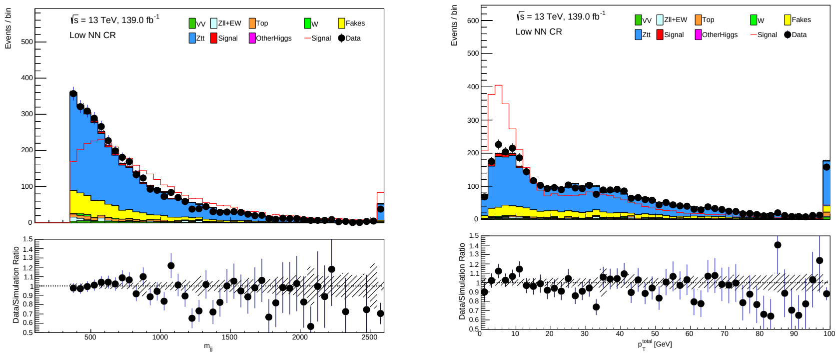
is dominated by background events. Figures 9.4.5,9.4.6 and 9.4.7 show distributions of some of the high-importance input features in the low NN CR. Good agreement between the data and the expectation is seen. Thus, this SR will be used for the test of CP invariance in the VBF Higgs production process decaying into two τ-leptons in the semileptonic decay channel in Chapter 10.

Signal	VV	Zll	Top	W	Z→ $\tau\tau$	Other Higgs	Fakes	s/b
79.55 ± 0.37	1.61 ± 0.25	7.86 ± 2.76	3.82 ± 0.78	0.27 ± 0.27	31.73 ± 2.12	9.54 ± 0.30	12.18 ± 1.07	1.19

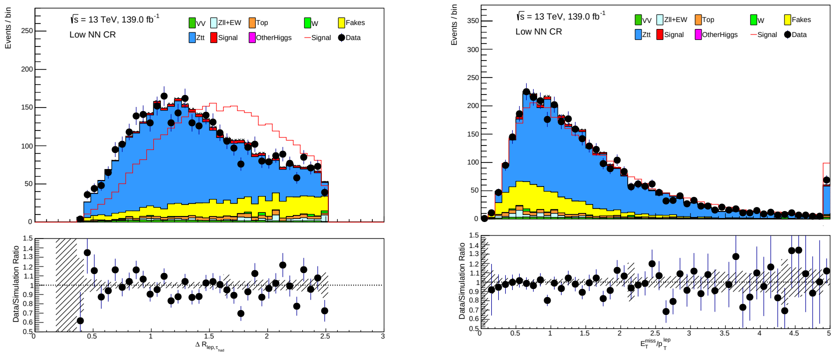
Table 9.4.4: Event yields in the SR, obtained with the binary neural network. The signal-to-background ratio (s/b) is shown. Only statistical uncertainties are included.



(a) $m_{\tau\tau}^{\text{vis}}$ (b) p_{T}^{H}

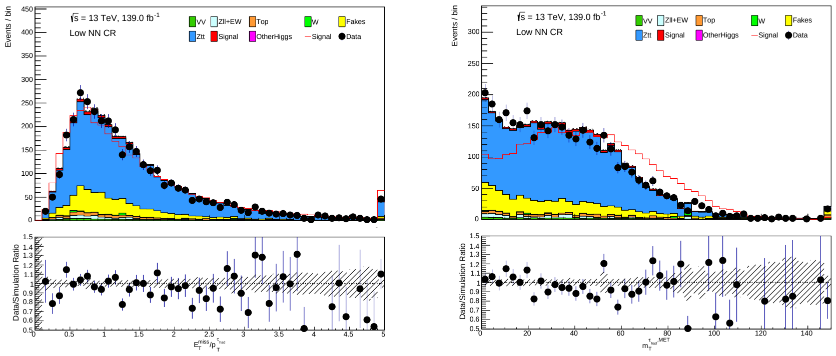


(c) m_{jj} (d) $p_{\text{T}}^{\text{total}}$

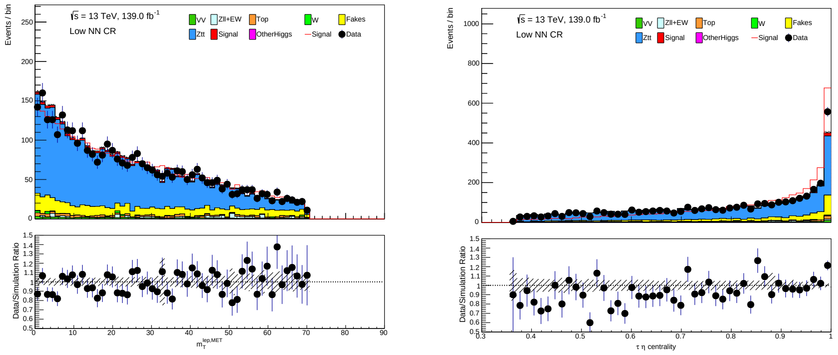


(e) $\Delta R_{\ell,\tau_{\text{had}}}$ (f) $E_{\text{T}}^{miss}/p_{\text{T}}^{\ell}$

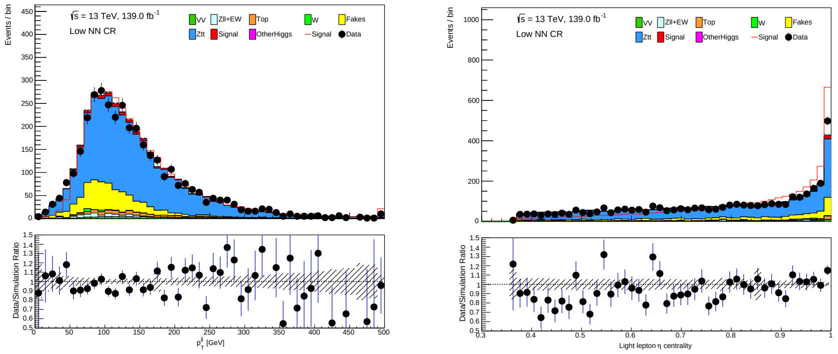
Figure 9.4.5: Kinematic distributions of some of features used for neural network analysis in the low NN control region (CR). These variables are; (a) the visible mass calculated from the two τ-candidates ($m_{\tau\tau}^{\text{vis}}$), (b) the Higgs transverse momentum (p_{T}^{H}), (c) the invariant mass of the di-jet system (m_{jj}), (d) the p_{T} of the vectorial sum of all objects in an event: light lepton (ℓ), τ_{had}, leading jet (j_1), sub-leading jet (j_2) and the missing transverse energy (E_{T}^{miss}) ($p_{\text{T}}^{\text{total}}$), (e) the angular distance between the two τ-candidates ($\Delta R_{\ell,\tau_{\text{had}}}$), and (f) the ratio of E_{T}^{miss} and the p_{T} of ℓ ($E_{\text{T}}^{\text{miss}}/p_{\text{T}}^{\ell}$). Overflow bins shown in all Figures. Figure (e) shows the cut applied at the Preselection stage (see Section 7.1). Only statistical uncertainties are shown. "Signal" line represents the signal contribution normalized to the total background.



(a) $E_{\mathrm{T}}^{miss}/p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$ (b) $m_{\mathrm{T}}^{\tau_{\mathrm{had}},MET}$

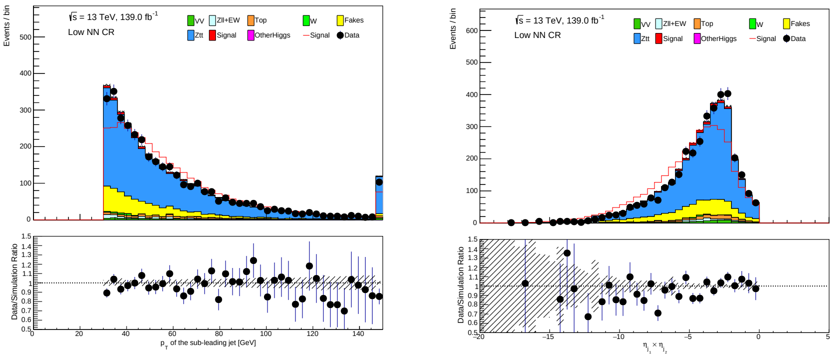


(c) $m_{\mathrm{T}}^{\ell,MET}$ (d) $\tau\eta$ centrality



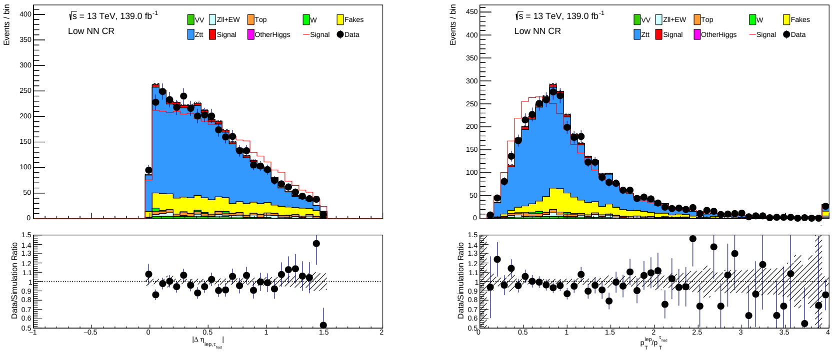
(e) p_{T}^{jj} (f) $\ell\eta$ centrality

Figure 9.4.6: Kinematic distributions of some of features used for neural network analysis in the low NN control region (CR). This variables are; (a) the ratio of the missing transverse energy (E_{T}^{miss}) and the transverse momentum (p_{T}) of the τ_{had} ($E_{\mathrm{T}}^{\mathrm{miss}}/p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$), (b) the transverse mass calculated from the leading τ-candidate and E_{T}^{miss} ($m_{\mathrm{T}}^{\tau_{\mathrm{had}},MET}$), (c) the transverse mass calculated from the light lepton (ℓ) and E_{T}^{miss} ($m_{\mathrm{T}}^{\ell,MET}$), (d) the centrality of the τ_{had} with respect to the two leading jets ($\tau\eta$ centrality), (e) the p_{T} of the di-jet system (p_{T}^{jj}), and (f) the centrality of the ℓ with respect to the two leading jets ($\ell\eta$ centrality). Overflow bins shown in Figures (a), (b) and (e). Figure (c) shows the cut applied at the Preselection stage (see Section 7.1). Only statistical uncertainties are shown. "Signal" line represents signal contribution normalized to the total background.



(a) $p_{\mathrm{T}}^{j_2}$

(b) $\eta_{j_1} \times \eta_{j_2}$



(c) $\Delta\eta^{\ell,\tau_{\mathrm{had}}}$

(d) $p_{\mathrm{T}}^{\ell}/p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$

Figure 9.4.7: Kinematic distributions of some of features used for neural network analysis in the low NN VBF control region (CR). This variables are; (a) the transverse momentum of the sub-leading jet ($p_{\mathrm{T}}^{j_2}$), (b) the signed product of the pseudorapidity of the leading and sub-leading jets ($\eta_{j_1} \times \eta_{j_2}$), (c) the difference in η between the light lepton (ℓ) and the τ_{had} ($\Delta\ell, \tau_{\mathrm{had}}$) and (d) the ratio of the p_{T} of the two τ-candidates ($p_{\mathrm{T}}^{\ell}/p_{\mathrm{T}}^{\tau_{\mathrm{had}}}$). Overflow bins shown in Figures (a) and (d). Figure (c) shows the cut applied at the Preselection stage (see Section 7.1). Figures (a) and (b) show the cut applied to select VBF events (see Section 7.2. Only statistical uncertainties are shown. "Signal" line represents signal contribution normalized to the total background.

Chapter 10

Test of CP invariance

This chapter describes in detail the performed test on CP symmetry violation (CPV) and measurements made in search of it. The concepts regardin to this chapter were discussed in Section 2.3. Three observables are used to test CP symmetry invariance. In Section 10.1, the method used to introduce CP-violating behavior to the simulated SM signal samples, is described. Sections 10.2 and 10.3 present the statistical analysis the fit models and results, respectively, that are used to test the CP invariance.

10.1 Signal reweighting

The *parton* model was introduced by Richard Feynmann in 1969 [148]. Partons are useful to interpret the showers of quarks and gluons produced in hadron colliders, since they can be described as free, point-like particles.

In proton-proton collisions, the cross-section of the production process $ij \to X$ of a particle X with initial partons i and j (σ_X) is derived from the cross-section at parton level ($\hat{\sigma}_{ij\to X}$), weighted by the parton distribution functions (PDFs) of the initial state partons: $f(x_i, Q^2)$ and $f(x_j, Q^2)$. The PDF $f(x_k, Q^2)$ represents the probability to find the parton k with a momentum fraction of the proton x_k, and depends on the squared momentum transfer Q^2 of the process. Then, σ_X is calculated by integrating over all the possible momentum fractions of these partons, that is [149]

$$\sigma_X = \sum_{i,j} \int_0^1 dx_i \int_0^1 dx_j f_i(x_i, Q^2) f_j(x_j, Q^2) \hat{\sigma}_{ij\to X}(\alpha_s, \hat{Q}^2), \tag{10.1.1}$$

where $\hat{Q}$ is the re-scaling factor of the PDFs to the Q^2 derived from QCD corrections (specifically, using the DGLAP equations [150] that describe the evolution of the PDFs with variable energy scales), α_s the running coupling constant from QCD, and the partonic cross-section: [19]

$$\hat{\sigma}_{ij\to X} = \frac{1}{F} \int \mathcal{M}(ij \to X) d\phi. \tag{10.1.2}$$

Here; F is the particle flux, $M(ij \to X)$ the matrix-element that describes the transition probability of the process and $d\phi$ the phase space factor.

Following this example, it is possible to express the matrix-elements in Equation 2.3.6.1 of the Higgs-boson production process $ij \to klH$ as

$$2\mathrm{Re}(\mathcal{M}^*_{\mathrm{SM}} \mathcal{M}_{\text{CP-odd}}) = \sum_{i,j,k,l} f_i(x_1) f_j(x_2) 2\mathrm{Re}\left(\left(\mathcal{M}^{ij\to klH}_{\mathrm{SM}} \right)^* \mathcal{M}^{ij\to klH}_{\text{CP-odd}} \right) \tag{10.1.3}$$

$$|\mathcal{M}_{\text{SM}}|^2 = \sum_{i,j,k,l} f_i(x_1) f_j(x_2) \left|\mathcal{M}_{\text{SM}}^{ij\to klH}\right|^2, \tag{10.1.4}$$

where the sum runs over the partons flavours; and x_1 and x_2 are the momentum fractions of the initial-state partons moving in the positive and negative z direction, respectively. This analysis uses the CT10 PDFs [75] at LO accuracy and the momentum fractions are calculated as

$$x_{1(2)} = \frac{m_{Hjj}}{\sqrt{s}} e^{+(-)y_{Hjj}}, \tag{10.1.5}$$

where m_{Hjj} and y_{Hjj} are the invariant mass and the rapidity of the final state, respectively.

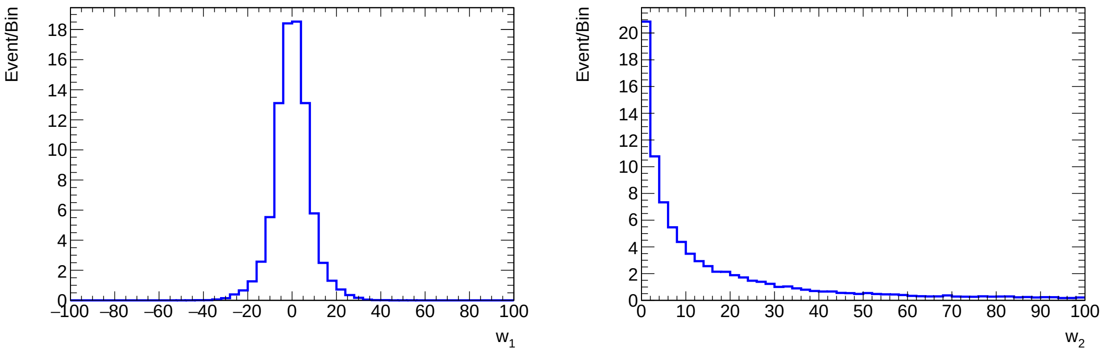
Following Ref. [66], a Matrix-Element-based reweighting procedure is applied to the SM Signal samples in order to simulate CP-odd behavior through $\tilde{d}$ dependence. These weights are defined as

$$w(\tilde{d}) = \frac{|\mathcal{M}|^2}{|\mathcal{M}_{\text{SM}}|^2} = 1 + \tilde{d} \cdot \frac{2\text{Re}(\mathcal{M}^*_{\text{SM}}\mathcal{M}_{\text{CP-odd}})}{|\mathcal{M}_{\text{SM}}|^2} + \tilde{d}^2 \cdot \frac{|\mathcal{M}_{\text{CP-odd}}|^2}{|\mathcal{M}_{\text{SM}}|^2} = 1 + \tilde{d} \cdot w_1 + \tilde{d}^2 \cdot w_2, \tag{10.1.6}$$

using the notation of Section 2.3.4. These are obtained using the four-vectors of the Higgs boson and the final state partons, the flavour of the partons involved and the *Bjorken* values $x_{1(2)}$ of outgoing partons [11]. The linear "w_1" and quadratic "w_2" weights are calculated once per signal event. Figure 10.1.1 shows their distributions in the SR.

The matrix-element calculations used for the weights' estimation (Equations 10.1.3 and 10.1.4) are done at LO accuracy using HAWK [63, 64], including the matrix-elements corresponding to three kind of events: two initial partons producing; two quarks and a Higgs boson ($qq \to qqH$), two quarks, an anti-quark and a Higgs boson ($qqg \to qq\bar{q}H$) and two initial quarks producing two quarks, a gluon and a Higgs boson ($qq \to qqgH$).

This procedure was previously used for testing the CP properties in the VBF production mode of the Higgs boson using Run-1 [151] and partial Run-2 [11] data recorded by the ATLAS detector.



(a) w_1 (b) w_2

Figure 10.1.1: Distributions of (a) the linear weight (w_1) and (b) the quadratic weight (w_2) used in the reweighting of the VBF signal samples in the SR.

10.2 Statistical Models

Estimations of sensitivity to the CP-odd couplings are obtained by implementing two types of parameter estimation methods. Section 10.2.1 describes the linear fit applied to the *Gauge curve* of the CP-odd variable distribution and Section 10.2.2 describes a Maximum Likelihood Fit.

10.2.1 Gauge curve linear fit

For each $\tilde{d}$ value within a set range; a CP-odd variable's distribution is generated, weighted following the Equation 10.1.6, and its mean value "$<$ CP-odd variable $>$" is calculated. The Gauge curve is produced as the mean value of the CP-odd variable as function of $\tilde{d}$. Figure 10.2.1.1 shows the Gauge curve for three different CP-odd variables used in this analysis, for the range $\tilde{d} \in [-0.2, 0.2]$ with a step size of 0.004.

For large positive (negative) $\tilde{d}$ values, the Gauge curve reaches a maximum (minimum) and then it merges towards zero due to the dominant CP-even contribution proportional to $\tilde{d}^2$ shown in Equation 2.3.4. In contrast, for $\tilde{d}$ values near zero, a dominant contribution from the second term in Equation 2.3.4 is expected, which is sensitive to the CP-odd coupling.

A linear relation between $<$ CP-odd variable $>$ and $\tilde{d}$ is assumed within a $\tilde{d}$ range about zero,

$$< \text{CP-odd variable} >= a \cdot \tilde{d} + b, \tag{10.2.1.1}$$

where a and b are the slope and the y-intercept of the linear fit, respectively. Solving Equation 10.2.1.1 for $\tilde{d}$ leads to

$$\tilde{d} = \frac{< \text{CP-odd variable} > -b}{a}, \tag{10.2.1.2}$$

whose error ($\Delta\tilde{d}$) can be estimated as

$$\Delta\tilde{d} = \frac{1}{|a|}\sqrt{\tilde{d}^2\Delta a^2 + \Delta b^2 + \Delta < \text{CP-odd variable} >^2}, \tag{10.2.1.3}$$

where Δa and Δb are the uncertainties of the slope and y-intercept, respectively derived from the Gauge curve linear fit and $\Delta <$ CP-odd variable $>$ is the standard error calculated from the CP-odd variable's distribution as

$$\Delta < \text{CP-odd variable} >= \frac{\sigma_{\text{CP-odd variable}}}{\sqrt{N_{\text{eff}}}} \quad \text{with} \quad N_{\text{eff}} = \frac{(\sum w)^2}{\sum w^2}, \tag{10.2.1.4}$$

where w are the event weights of the SR simulation processes. Estimations and constraints of $\tilde{d}$ ($\sigma_{\tilde{d}}$) and optimization studies for the Gauge curve linear fit are discussed in Section 10.3.

10.2.2 Maximum Likelihood fit

To evaluate the compatibility of the expected data set with a specific hypothesis ($\tilde{d} \neq 0$ in this thesis) a Maximum Likelihood method (ML) is applied.

The distribution of a set of random variables $\boldsymbol{n} = (n_1, ..., n_N)$ is given by its probability density function (p.d.f) $f(\boldsymbol{n})$, which depends on the unknown parameters $\boldsymbol{\theta}$ in an underlying model. If $\boldsymbol{n}$ correspond to an expected data set, its p.d.f. is constructed from the combination of the signal ($f_S(\boldsymbol{n};\boldsymbol{\theta})$) and background ($f_B(\boldsymbol{n};\boldsymbol{\theta})$) contributions. The *hypothesis* is then the underlying model where these p.d.fs are evaluated. The compatibility of this model with the observed data set ($\tilde{d} = 0$ in this analysis) is evaluated with the Likelihood function $\mathcal{L}(\boldsymbol{n};\boldsymbol{\theta})$.

When considering the hypothesis $f(\boldsymbol{n};\boldsymbol{\theta})$; the probability for the measurement i to be in the interval $[n_i, n_i + dn_i]$ is given by $f(n_i;\boldsymbol{\theta})dn_i$. If the measurements are independent, the probability (P) for all measurements is

$$\text{P} \ \forall \ n_i \in [n_i, n_i + dn_i] = \prod_{i=0}^{N} f(n_i;\boldsymbol{\theta})dn_i. \tag{10.2.2.1}$$

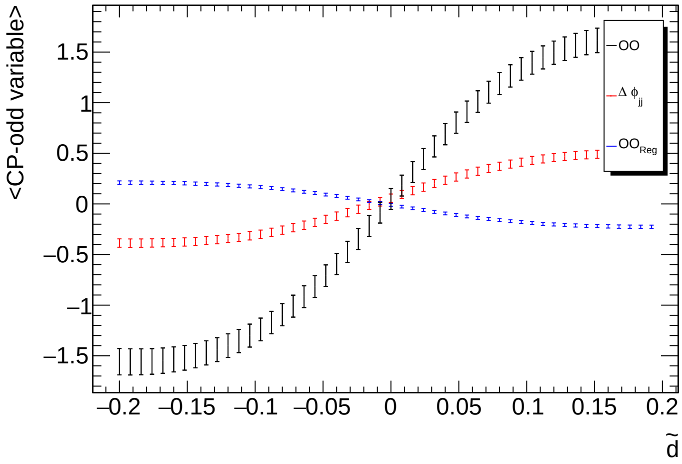


Figure 10.2.1.1: Gauge curve as a function of $\tilde{d}$ for three CP-odd variables; the Optimal Observable ($\mathcal{OO}$), the signed azimuthal angle between the two tagging jets ($\Delta\phi_{jj}$) and the Optimal Observable obtained using symbolic regression method ($\mathcal{OO}_{\text{Reg}}$) in the SR. Only statistical uncertainties are shown.

If the hypothesis for the p.d.f.s and parameters are correct, the probability will be large. The Likelihood function is defined as

$$\mathcal{L}(\mathbf{n};\boldsymbol{\theta}) = \prod_{i=0}^{N} f(n_i;\boldsymbol{\theta}). \tag{10.2.2.2}$$

Since dn_i does not depend on $\boldsymbol{\theta}$, large values of $\mathcal{L}(\boldsymbol{n};\boldsymbol{\theta})$ translates into high compatibility.

For N-binned histograms, n_i is the content of bin i in the expected histogram and n_i^{pred} the content of bin i in the underlying model histogram.

The Likelihood function for a histogram is obtained by assuming a Poisson distribution in every bin [152, 153] as

$$f(n_i;\boldsymbol{\theta}) = \text{Poisson}(n_i;n_i^{\text{pred}}) \quad \rightarrow \quad \mathcal{L}(\mathbf{n};\boldsymbol{\theta}) = \prod_{i=0}^{N} \text{Poisson}(n_i;n_i^{\text{pred}}) = \prod_{i=0}^{N} \frac{(n_i^{\text{pred}})^{n_i}}{n_i!} e^{-n_i^{\text{pred}}}. \tag{10.2.2.3}$$

To simplify calculations, the negative logarithm of $\mathcal{L}$ "Negative Log-Likelihood" (NLL) is taken, so that Equation 10.2.2.3 can be written as

$$\text{NLL}(\mathbf{n};\boldsymbol{\theta}) = -\text{log}\mathcal{L}(\mathbf{n};\boldsymbol{\theta}) = -\sum_{i=0}^{N} n_i \text{log}(n_i^{\text{pred}}) + \log(\Gamma(n_i+1)) + n_i^{\text{pred}}. \tag{10.2.2.4}$$

Thus, the maximum compatibility occurs when $\text{NLL}(\mathbf{n};\boldsymbol{\theta})$ is minimum.

Maximum Likelihood Estimators (MLE) $\hat{\boldsymbol{\theta}}$ are defined as the parameters that maximize the likelihood function $\mathcal{L}(\boldsymbol{n};\boldsymbol{\theta})$ (or minimize $\text{NLL}(\mathbf{n};\boldsymbol{\theta})$). They are solution to

$$\frac{\partial \mathcal{L}}{\partial \theta_i} = 0, \quad i = 1, ..., m. \tag{10.2.2.5}$$

The statistical error of the estimated $\hat{\boldsymbol{\theta}}$ is reported with a confidence interval. The asymptotic normality property of a MLE states that, in the large sample limit, the distribution of the MLE is approximately Gaussian (central limit theorem) [152].

The variance of a MLE is calculated by using the Rao-Cramér-Frechet (RCF) inequality, found to be

$$V[\hat{\boldsymbol{\theta}}] = \left(\mathbb{E}\left[-\frac{\partial^2 \log\mathcal{L}}{\partial\boldsymbol{\theta}^2}\right]\right)^{-1}, \tag{10.2.2.6}$$

when assuming minima variance (equality) and zero bias, which is always the case in the large sample limit [152]. The calculation of Equation 10.2.2.6 presents a challenge, since it requires the calculation of the expectation value ($\mathbb{E}$) of the second derivative of $\mathcal{L}$. In this analysis this leads to the integration of a Gamma function. However, in the large sample limit, the inverse of the variance can be estimated as

$$\hat{V} = \hat{\sigma}^2_{\hat{\boldsymbol{\theta}}} = \left(\left[-\frac{\partial^2 \log\mathcal{L}}{\partial\boldsymbol{\theta}^2}\right]\right)^{-1}\Bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}}. \tag{10.2.2.7}$$

Therefore, expanding NLL($\boldsymbol{n}$;$\boldsymbol{\theta}$) using Taylor function around the $\hat{\boldsymbol{\theta}}$ leads to

$$\mathrm{NLL}(\mathbf{n};\boldsymbol{\theta}) = -\mathrm{NLL}(\mathbf{n};\hat{\boldsymbol{\theta}}) - \left[\frac{\partial \mathrm{NLL}(\mathbf{n};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \cdot (\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}) - \frac{1}{2!}\left[\frac{\partial^2 \mathrm{NLL}(\mathbf{n};\boldsymbol{\theta})}{\partial\boldsymbol{\theta}^{\mathbf{2}}}\right]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \cdot (\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})^2 + ..., \tag{10.2.2.8}$$

whose second term vanishes by definition of $\hat{\boldsymbol{\theta}}$. Inserting Equation 10.2.2.7 in Equation 10.2.2.8, neglecting higher order terms and evaluating Equation 10.2.2.8 in [$\hat{\boldsymbol{\theta}}$±$N\hat{\sigma}^{\pm}_{\hat{\theta}}$] gives

$$\frac{N^2}{2} = \mathrm{NLL}(\hat{\boldsymbol{\theta}} \pm N\hat{\sigma}^{\pm}_{\hat{\boldsymbol{\theta}}}) - \mathrm{NLL}(\hat{\boldsymbol{\theta}}). \tag{10.2.2.9}$$

Equation 10.2.2.9 represents that a change of N standard deviation in the MLE makes the NLL($\boldsymbol{n}$;$\boldsymbol{\theta}$) increase by $N^2/2$ relative to its minimum value. The 68.3% and 95% central confidence intervals are given by $N = 1$ and $N = 1.96$, respectively. CI's indicate the $\hat{\boldsymbol{\theta}}$ value at which the NLL($\boldsymbol{n}$;$\boldsymbol{\theta}$) increases by 0.5 ($N = 1$) and 1.92 ($N = 1.96$) from its minimum.

10.3 Results

This section presents the results of the analysis performed on the test of CP invariance in the VBF Higgs boson production decaying into two τ-leptons in the $H \to \tau_{\mathrm{lep}}\tau_{\mathrm{had}}$ decay channel. The expected $\tilde{d}$ is calculated using an *Asimov* data set, constructed from the expected background and signal processes for $\tilde{d} = 0$ case, that is, the SM prediction. All studies are conducted using events in the SR defined in Section 9.4. The expected $\tilde{d}$ parameter is estimated using two methods: Gauge curve linear fit and Maximum Likelihood fit, described previously in Sections 10.2.1 and 10.2.2, respectively. Furthermore, the $\tilde{d}$ value is calculated and compared for three considered CP-odd observables. The expected mean values of these variables are listed in Table 10.3.1, which vanish as expected. All reweighted signal samples generated for $\tilde{d} \neq 0$ are normalized to their respective SM cross-section values, shown in Table 9.4.4. All uncertainties shown in this Section are statistical uncertainties of the events in the SR, displayed in Table 9.4.4. A requirement is applied to events with $|\mathcal{OO}| < 15$ to remove outlier values, results show no difference compared to the case where overflow bins are used at these accumulate the number of events with $|\mathcal{OO}| < 15$.

CP-odd variable	<CP-odd variable>
$\mathcal{OO}$	0.048 ± 0.103
$\Delta\phi_{jj}$	0.057 ± 0.041
$\mathcal{OO}_{\text{Reg}}$	-0.0085 ± 0.014

Table 10.3.1: Expected mean values of the three CP-odd variable distributions (<CP-odd variable>) for $\tilde{d} = 0$ case. Mean values for the Optimal Observable ($\mathcal{OO}$), signed azimuthal angle between the two tagging jets ($\Delta\phi_{jj}$), and the Optimal Observable obtained using symbolic regression method ($\mathcal{OO}_{\text{Reg}}$), are listed. Only statistical uncertainties are shown.

The expected $\tilde{d}$ values estimated using the Gauge curve linear fit for three CP-odd variables defined in Sections 2.3.6 and 2.3.7 are listed in Table 10.3.2. The Gauge curve is produced using 50 equidistant values of $\tilde{d}$ in the range $[-0.01, 0.01]$. Number of points and fit range are chosen based on optimization studies described in Appendix B.1 that minimized the error on $\tilde{d}$. The Gauge curve produced using these optimized parameters is shown in Figure 10.3.1. There, a linear behavior can be seen in the chosen range. Since the Asimov data set is used for the fit, the obtained values for $\tilde{d}$ are close to the assumed value $\tilde{d} = 0$ as expected.

In the Gauge curve linear fit, expected constraints on $\tilde{d}$ from the $\mathcal{OO}$ and $\mathcal{OO}_{\text{Reg}}$ variables are found to be very similar, while using $\Delta\phi_{jj}$ variable led to larger uncertainty. The main contributions to the error of $\tilde{d}$ are the values of the slope and the error on the mean value of the CP-odd variable for $\tilde{d} = 0$ ($\Delta < \text{CP-odd variable} > |_{\tilde{d}=0}$), since the first and second terms in Equation 10.2.1.3 are of the order of 10^{-13} and 10^{-4}, respectively whereas $\Delta < \text{CP-odd variable} > |_{\tilde{d}=0}$ is $\sim 10^{-2}$ for all three variables. Values for the latter parameters are listed in Table 10.3.1. The values of the slope for three CP-odd variables are provided in Table 10.3.3.

The slope value obtained from $\mathcal{OO}$ is $\sim 28\%$ larger than that of $\Delta\phi_{jj}$, leading to smaller $\tilde{d}$ constraints from $\mathcal{OO}$. In contrast, the slope derived from $\mathcal{OO}_{\text{Reg}}$ does not contribute to the constraints on $\tilde{d}$ as much as the $\Delta < \mathcal{OO}_{\text{Reg}} > |_{\tilde{d}=0}$ which has the minimum value of the three CP-odd variables, as displayed in Table 10.3.1. The values of the parameters estimated using the Gauge curve linear fit: $\tilde{d}$, slope, y-intercept, χ^2 and $\Delta < \text{CP-odd variable} > |_{\tilde{d}=0}$ can be found in Table B.2.1.

The validation of the Gauge curve linear fit is done by comparing 25 equally spaced $\tilde{d}$ values in $[-0.2, 0.2]$, range that is used to reweight the Asimov data, with the $\tilde{d}$ values from the Gauge curve linear fit of the $\mathcal{OO}$, as shown in Figure 10.3.2. As expected, for $\tilde{d}$ values close to zero, the Asimov values are recovered within uncertainties.

	Expected	
CP-odd variable	$\tilde{d} \times 10^7$	$\sigma_{\tilde{d}}$
$\mathcal{OO}$	3.71	± 0.00625
$\Delta\phi_{jj}$	1.21	± 0.00893
$\mathcal{OO}_{\text{Reg}}$	7.06	± 0.00627

Table 10.3.2: Expected $\tilde{d}$ values obtained from the Gauge curve linear fit for three CP-odd variables; Optimal Observable ($\mathcal{OO}$), signed azimuthal angle between the two tagging jets ($\Delta\phi_{jj}$) and the Optimal Observable obtained using symbolic regression methods ($\mathcal{OO}_{\text{Reg}}$), in the SR. Only statistical uncertainties are shown.

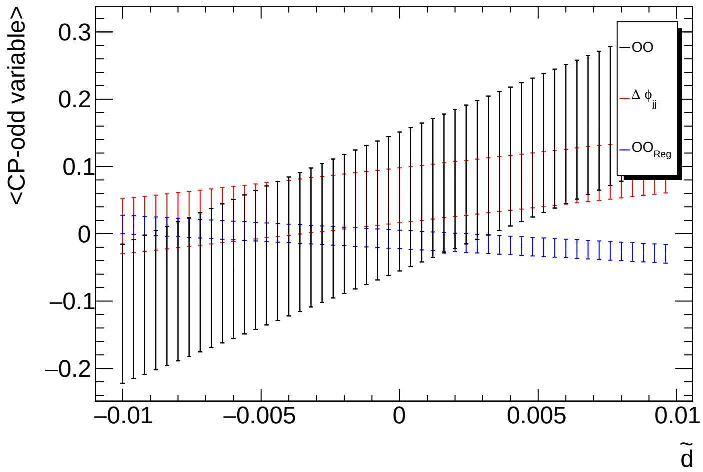


Figure 10.3.1: Gauge curve as a function of $\tilde{d}$ in the optimized fit range for three CP-odd variables: the Optimal Observable ($\mathcal{OO}$), the signed azimuthal angle between the two tagging jets ($\Delta\phi_{jj}$), and the Optimal Observable obtained using symbolic regression methods ($\mathcal{OO}_{\text{Reg}}$), in the SR. Only statistical uncertainties are shown.

CP-odd variable	Gauge curve slope
$\mathcal{OO}$	16.69 ± 2.53
$\Delta\phi_{jj}$	4.61 ± 1.00
$\mathcal{OO}_{\mathrm{Reg}}$	-2.23 ± 0.34

Table 10.3.3: Slopes of the Gauge curve linear fits shown in Figure 10.3.1 for the three CP-odd variables used; Optimal Observable ($\mathcal{OO}$), signed azimuthal angle between the two tagging jets ($\Delta\phi_{jj}$), and Optimal Observable obtained using symbolic regression methods ($\mathcal{OO}_{\mathrm{Reg}}$), in the SR. Only statistical uncertainties are included.

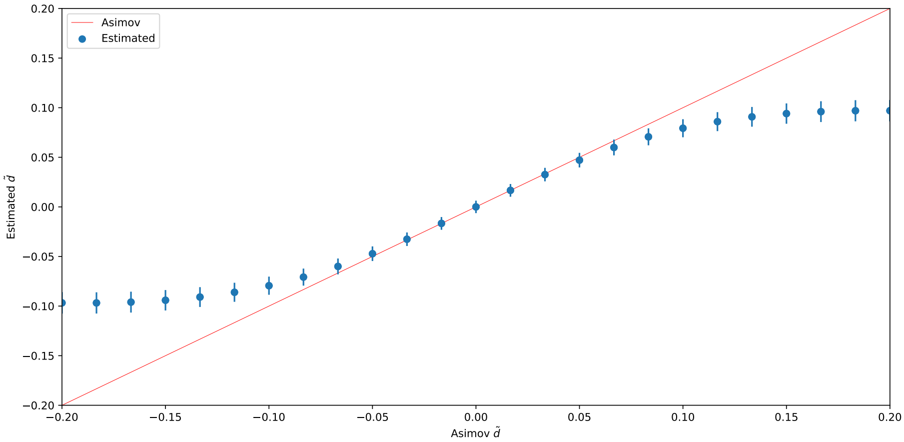


Figure 10.3.2: Validation of the Gauge curve linear fit in the SR. Red line shows $\tilde{d}$ values used in the Asimov data set. $\tilde{d}$ values estimated with the Gauge curve linear fit are shown as blue dots. Only statistical uncertainties are shown.

The Maximum Likelihood (ML) fit is done using 50 equally spaced $\tilde{d}$ values in $[-0.07, 0.07]$ range, which is different from the range used in the Gauge curve linear fit. When applying a linear fit to the Gauge curve, the fit range plays an important role since the Gauge curve is not linear over the whole $\tilde{d}$ range (see Figure 10.2.1.1). The range of the ML fit is chosen so that the ΔNLL crosses the 95% CI, as illustrated in Figure 10.3.3.

However, as described in Section 10.2.2, the ML fit is performed using histograms, from which the NLL is calculated by summing over the bins of the distribution. Therefore, the binning in the ML fit is expected to play an important role, similar to the fit range for the Gauge curve linear fit. This is studied by performing multiple ML fits using each of the three CP-odd variables using 10, 20, 60 and 100 bins. No significant impact is observed in the $\tilde{d}$ constraints from any of the three variables, as displayed in Table B.2.5. Lowest constrain on $\tilde{d}$ is found when using more than 60 bins. Thus, a common bin number of 70 is chosen for all three CP-odd variables.

The expected $\tilde{d}$ values and confidence intervals estimated using the ML fit in the SR, are shown in Table 10.3.4 for three CP-odd variables studied. Figure 10.3.3 shows the expected ΔNLL-curves of three variables. The SM case ($\tilde{d} = 0$) is contained in the expected $\tilde{d}$ constraints estimated with the ML fit. Similar to the Gauge curve linear fit results, the expected $\tilde{d}$ obtained from $\mathcal{OO}$ and $\mathcal{OO}_{\mathrm{Reg}}$ are very similar as shown in Figure 10.3.3. Competitive performance of $\mathcal{OO}_{\mathrm{Reg}}$ motivates searches for new techniques of constructing observables extracted directly from data using machine learning algorithms [14]. Furthermore, looser constraints on $\tilde{d}$ are also found for $\Delta\phi_{jj}$, represented by a wider ΔNLL curve in Figure 10.3.3. Same $\tilde{d}$ value is found for all three CP-odd variables.

	68% CI		95% CI	
	$\tilde{d} \times 10^{17}$	$\sigma_{\tilde{d}}$	$\tilde{d} \times 10^{17}$	$\sigma_{\tilde{d}}$
$\mathcal{OO}$	-4.16	± 0.0168	-4.16	± 0.0364
$\Delta\phi_{jj}$	-4.16	± 0.0223	-4.16	± 0.0448
$\mathcal{OO}_{\text{Reg}}$	-4.16	± 0.0168	-4.16	± 0.0364

Table 10.3.4: Expected $\tilde{d}$ values obtained from the Maximum Likelihood fit, for three CP-odd variables; Optimal Observable ($\mathcal{OO}$), signed azimuthal angle between the two tagging jets ($\Delta\phi_{jj}$), and the Optimal Observable obtained using symbolic regression methods ($\mathcal{OO}_{\text{Reg}}$), in the SR. Only statistical uncertainties are shown.

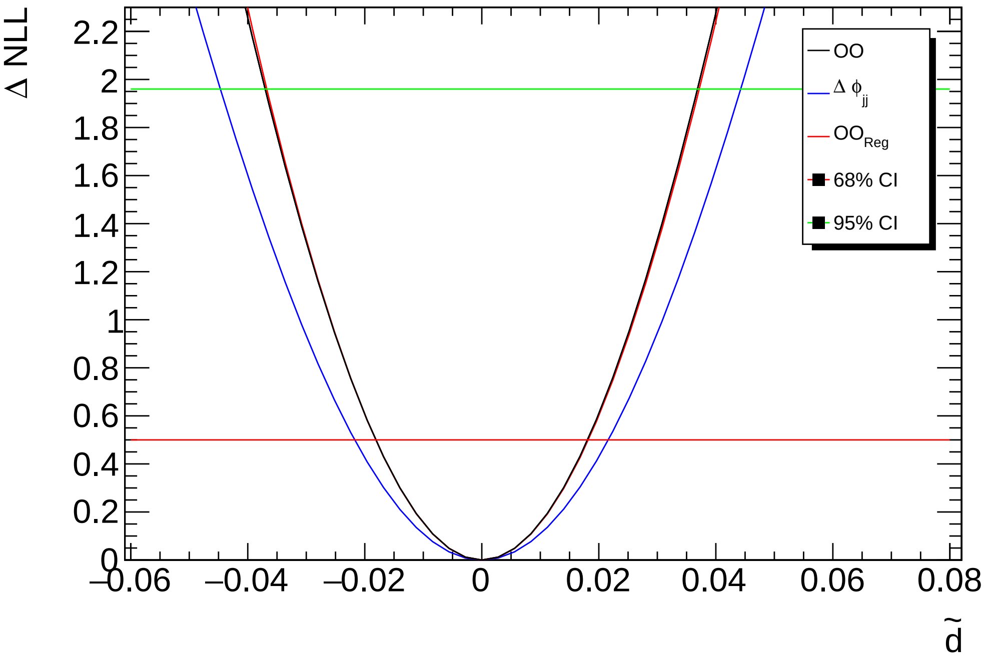


Figure 10.3.3: Negative Log-Likelihood (NLL) curves for three CP-odd variables; Optimal Observable ($\mathcal{OO}$), signed azimuthal angle between the two tagging jets ($\Delta\phi_{jj}$), and the Optimal Observable obtained using symbolic regression methods ($\mathcal{OO}_{\text{Reg}}$). 68% and 95% confidence intervals (CI) are also shown. Results are computed in the SR.

The validation method used for the Gauge curve fit is also used for the ML fit, showing agreement between the estimated $\tilde{d}$ values and the $\tilde{d}$ values used to reweight the Asimov data sets, as shown in Figure 10.3.4. In contrast to the Gauge curve linear fit, $\tilde{d}$ values are recovered over the whole $\tilde{d}$ range, verifying that the ML fit results do not depend on the fit range, as it does in the Gauge curve linear fit.

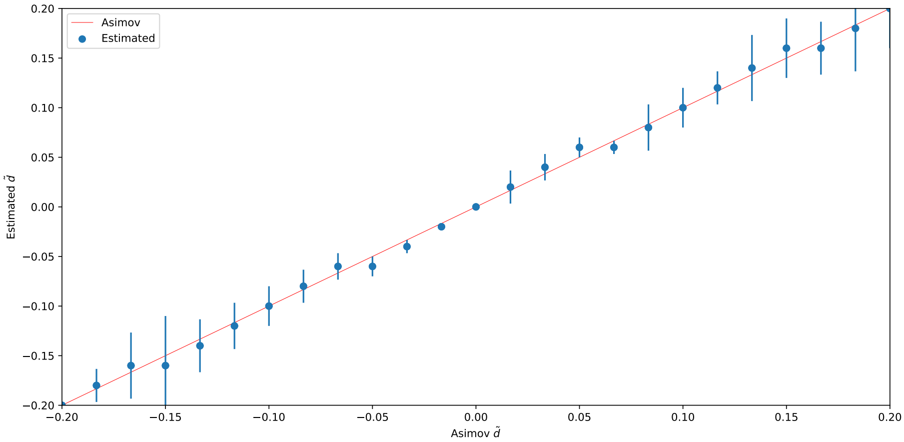


Figure 10.3.4: Validation of the Maximum Likelihood fit in the SR. Red line shows $\tilde{d}$ values used in the Asimov data set. $\tilde{d}$ values estimated with the Gauge curve linear fit are shown as blue dots. Only statistical uncertainties are shown. Only the 68% CI is shown.

The behavior of the expected $\tilde{d}$ observed from the three CP-odd variables is the same for the two fits. Similar values and constraints are found for $\mathcal{OO}$ and $\mathcal{OO}_{\text{Reg}}$ variables whereas weaker constraints are found when using the $\Delta\phi_{jj}$ variable. This behavior was expected, since the $\mathcal{OO}$ is constructed using the full phase space of the signal while $\Delta\phi_{jj}$ represents a CP-odd sensitive variable.

In contrast, it was not expected to find weaker constraints on $\tilde{d}$ from the Gauge curve linear fit than from that of the ML fit. This expectation derives from the premise that the ML fit has no restrictions, it uses all the information of the variables, while the Gauge curve linear fit is only valid in a range where the Gauge curve behaves linearly and estimates $\tilde{d}$ from the $<$ CP-odd variable $>$, which might "screen" the actual shape of the CP-odd distributions. Some studies have been made in this regard, none pointed to a possible error in the implementation of the Statistical analyses:

Both fits were repeated using only signal processes. Since the background processes are CP-even, they should not impact the expected $\tilde{d}$ values. Constraints on $\tilde{d}$ are reduced by a factor of 10 when using the Gauge curve linear fit, as can be seen in Table B.2.2. For the ML fit such significant reduction is not found, as displayed in Table B.2.8. The use of only signal samples in the Gauge curve linear fit has not significant impacts on the expected $<$ CP-odd variable $>|_{\tilde{d}=0}$ values, as expected.

Furthermore, $\tilde{d}$ values are estimated using different step sizes of the $\tilde{d}$ range using both fit setups, in order to verify that there are no anomalies in the fitted distributions. Appendix B.2 shows the results of the estimated $\tilde{d}$ values using step sizes of 0.0004, 0.0002 and 0.001 for both fits. No significant changes are seen between the results.

In addition, no correlation between the NN score and the $\tilde{d}$ constraints are found. The minimum $\tilde{d}$ constraint corresponds to the applied NN score requirement of 0.92.

Chapter 11

Conclusion and outlook

Limitations on the Standard Model of particle physics have led to analyses focused on the search of deviations from the theory. The Higgs boson was theoretically predicted and introduced in the Standard Model through the Higgs mechanism and its discovery in 2012 provided a new field of study on its properties and possible disparities from the Standard Model. However, differences between the measurements and the predictions are not found so far. Furthermore, improvements on particle detectors have enabled the increase of measured data, that can be used for precision measurements on the effects of unknown phenomena.

This thesis focuses on testing CP invariance of VBF production process of the Higgs boson exploiting the decay $H \to \tau_{\text{lep}}\tau_{\text{had}}$. The utilized data in this analysis corresponds to proton-proton collisions recorded in the full Run-2 data-taking period (between 2015 and 2018) by the ATLAS detector at the LHC at a center-of-mass energy of $\sqrt{s} = 13$ TeV corresponding to an integrated luminosity of 139 fb^{-1}.

First, a signal region (VBF SR) is defined using a cut-based approach to select signal events. A data-driven method is introduced, the "Fake Factor" method, to estimate the background contribution from jets that are misidentified as the hadronically decaying τ-lepton (τ_{had}), called "fakes". This method has been previously used in other analyses [11, 66, 99]. The method uses the identification variable of the τ_{had}, which exploits the differences in shapes of the calorimetric showers between the jets originated from background processes and the τ_{had} [85]. A control region (anti-τ CR) is defined inverting the identification criteria applied in the VBF SR. The fakes are transferred from this region to the VBF SR using a transfer factor called "combined Fake Factor", calculated by summing the individual contributions from the background processes that participate in the fakes production: W+jets, multi-jet, Z+jets and top-quark production processes. The method is simplified by including the background contributions from Z+jets and top-quark into the contribution from W+jets, this simplification is validated in Section 8.3.1.

An expectation-to-data ratio of 1.04 ± 0.05 is obtained in the VBF SR after the inclusion of the estimated fake background contribution. Furthermore, a difference of 0.8% is found between the fake background yield in the VBF SR from this analysis and Ref. [66], which uses the same data set and signal events as this analysis. The method is validated and a simplified method is proposed for future applications.

A neural network (NN) algorithm is developed and applied to the VBF SR to enhance the signal-to-background ratio. Studies were performed for the optimization of the NN, including the optimization of the set of input variables, hyperparameters and output nodes. The best performing NN is found to be a binary classifier. The initial set of input features was optimized, 7 variables were removed with no significant reduction ($\sim 0.3\%$) of the significance (figure of merit) of the NN. The set of hyperparameters of the NN was optimized using the *Optuna* framework [112]. The best performing NN is found to have 4 hidden layers, 512 neurons per layer, a tuning parameter of $4.22 \cdot 10^{-6}$, a learning rate of $3.6 \cdot 10^{-3}$ and a batch size of 128. The binary classifier provides a signal-to-background ratio of 1.19 and categorical accuracy of $\sim 76\%$.

The SR defined from the optimised binary NN is used to test CP invariance. First, the Standard Model signal samples are reweighted with CP-mixing effects that scale with the parameter $\tilde{d}$, which controls the strength of the CP symmetry violation. Three CP-sensitive observables are analysed: the Optimal Observable ($\mathcal{OO}$) constructed from the matrix element of the EFT, the signed difference in the azimuthal angle between the two leading jets ($\Delta\phi_{jj}$) and the Optimal Observable derived from data using machine learning methods ($\mathcal{OO}_{\text{Reg}}$) [14]. An Asimov data set was created with the Standard Model prediction ($\tilde{d} = 0$). Two statistical methods are performed to determine expected constraints on $\tilde{d}$ using either the mean value of the CP-odd observable or a Maximum Likelihood to the full distribution. The first one estimates $\tilde{d}$ by applying a linear fit on the Gauge curve, defined as the mean of the CP-odd distribution weighted with different $\tilde{d}$ values within a fixed range. In the second method the negative-log-likelihood (NLL) function of the CP-odd distributions weighted with different $\tilde{d}$ values is calculated, the minimum of the NLL function provides the estimation of $\tilde{d}$. Two confidence intervals (CI) are defined for the NLL curve, the 68% and 95% CIs. Only expected results are discussed. The expected constraints on $\tilde{d}$ estimated with the Gauge curve linear fit are $[-0.00625, 0.00625]$, $[-0.00893, 0.00893]$ and $[-0.00627, 0.00627]$ for the $\mathcal{OO}$, $\Delta\phi_{jj}$ and $\mathcal{OO}_{\text{Reg}}$, respectively at a 68% CI. The expected constraints on $\tilde{d}$ found using the ML fit at the 68% CI are $[-0.0168, 0.0168]$, $[-0.0223, 0.0223]$ and $[-0.0168, 0.0168]$ and at a 95% CI $[-0.0364, 0.0364]$, $[-0.0448, 0.0448]$ and $[-0.0364, 0.0364]$ for the $\mathcal{OO}$, $\Delta\phi_{jj}$ and $\mathcal{OO}_{\text{Reg}}$, respectively. The expected $\tilde{d}$ constraints are successfully found to be consistent with the Standard Model prediction of zero. Furthermore, as obtained better constraints were expected from the $\mathcal{OO}$ variable compared to that from $\Delta\phi_{jj}$. High resemblances found between the $\tilde{d}$ constraints from $\mathcal{OO}$ and $\mathcal{OO}_{\text{Reg}}$, which motivates the use and research of the methods utilized for defining the $\mathcal{OO}_{\text{Reg}}$ variable. Nevertheless, better constraints on $\tilde{d}$ were expected from the ML fit, in contrast with the obtained results, this has to be addressed in future analysis.

Improvements in the sensitivity to constrain $\tilde{d}$ can be expected from future analyses that use larger data sets, as future analysis using data from the Run-3 of the LHC, currently being collected. Increasing the statistics would lead to improvements in the Fake Factor method, since CRs with low statistics were found in this analysis, larger statistics could reduce the uncertainties and provide more accurate values of the combined Fake Factors. In addition, larger data sets used in the training of the NN can lead to less risk of overfitting.

Future analysis must include observed results, which were not used in this thesis. The expected constraints on $\tilde{d}$ obtained from the ML fit in this thesis can be compared to two previous analyses Ref.[11] and [154] that used the same Maximum-Likelihood method. The first one, tests the CP properties in the VBF process of the $H \to \tau\tau$ final state using 36 fb^{-1} of data at $\sqrt{s} = 13$ TeV with the ATLAS detector, an expected $\tilde{d} \in [-0.035, 0.033]$ was found at 68% and $\tilde{d} \in [-0.21, 0.15]$ at 95% for the $\mathcal{OO}$. The second analysis focuses on the VBF process in the $H \to \gamma\gamma$ decay, using 139 fb^{-1} of data at a $\sqrt{s} = 13$ TeV with the ATLAS detector, an expected $\tilde{d} \in [-0.027, 0.027]$ was found at 68% and $\tilde{d} \in [-0.055, 0.055]$ at 95% for the $\mathcal{OO}$. The constraints on $\tilde{d}$ from both analyses ([11] and [154]) were combined leading to an expected $\tilde{d}$ constraint of $\tilde{d} \in [-0.022, 0.021]$ at 68% and $\tilde{d} \in [-0.046, 0.045]$ at 95% for the $\mathcal{OO}$ [154]. These combined expected constraints are $\sim 23.6\%$ larger than the results from this thesis. Also, better $\tilde{d}$ constraints ($\sim 52\%$) is found compared to the analysis that uses the same decay channel ([11]). Since this thesis does not include systematic uncertainties, in contrast to the mentioned analyses, it is not possible to affirm that the sensitivity was improved, therefore a rigorous analysis of the systematic uncertainties must be addressed.

In addition, the ML fit could be improved by introducing further unknown parameters, such as the signal-strength parameter, defined as the ratio between the measured and predicted signal cross-sections; and other nuisance parameters such as the statistical and systematic uncertainties and normalization factors from $Z \to \tau\tau$, $Z \to \ell\ell$ and top-quark background contributions, previously included in Ref. [11]. The analysis can be expanded, including the other two decay channels of the $H \to \tau\tau$ (this analysis is focused in the semileptonic decay channel), providing a wider scope for finding CP violation effects. Furthermore, the performed analysis can be adapted to further decay processes of the Higgs boson, meaning that the challenge followed by this thesis remains open for future investigations.

Appendix A

Supplementary information of the fake background estimation

This section displays some tables and distributions used in the Fake Factor method for estimating the fake background contribution.

A.1 Event yield Tables of the CRs

Table A.1.1 lists the MC simulation and data yields of the CRs used for the calculation of the Individual Fake Factors shown in Section 8.1.1.

Region	**Signal**	**VV**	**Zll**	**Top**	**W**	**Ztt**	**Other Higgs**	**Data**
SR	162.07±0.53	68.61±1.61	72.78±11.01	114.35±4.04	23.73±5.26	2729.36±20.21	60.14 ± 0.74	3669
anti-τ SR	48.92±0.29	32.58±1.37	74.44±33.25	197.69±5.58	190.12±21.36	993.07±13.16	22.13 ±0.45	11383
W CR 1-prong	5.80±0.10	28.52±0.88	23.14±5.17	70.33±3.17	8.65±3.06	60.60±3.47	3.25 ± 0.17	348
W CR 3-prong	0.93±0.04	6.43±0.38	-2.23±2.78	15.79±1.53	4.43±2.05	15.40±1.40	0.59 ± 0.07	94
W anti-τ CR 1-prong	1.17±0.04	9.57±0.90	8.76±7.52	55.71±2.94	43.34±6.84	18.72±1.52	0.95 ± 0.09	1690
W anti-τ CR 3-prong	0.37±0.02	7.69±0.86	13.50±3.85	76.67±3.54	118.04±14.77	10.58±1.09	0.33 ± 0.06	2302
QCD CR 1-prong	0.56±0.03	0.41±0.11	-0.08±0.20	0.90±0.33	0.00±0.00	12.41±1.36	0.19±0.04	47
QCD CR 3-prong	0.14±0.01	0.00±0.00	0.00±0.00	0.42±0.27	0.00±0.00	4.34 ± 0.71	0.05 ± 0.02	19
QCD anti-τ CR 1-prong	0.18±0.02	0.10±0.08	0.10±0.10	0.86 ± 0.33	0.21±0.20	4.96 ± 0.83	0.08 ± 0.03	290
QCD anti-τ CR 3-prong	0.09±0.01	0.02±0.01	0.49 ± 0.49	0.88 ± 0.43	0.57 ± 0.37	2.36 ± 0.55	0.10 ± 0.04	465

Table A.1.1: MC simulation and data yields in the control regions (CR) used for the calculation of the Individual Fake Factors. W+jets (W) and multi-jet (QCD) CR and anti-τ CRs are split into the 1-prong and 3-prong channels. Only statistical uncertainties are shown.

Table A.1.2 lists the MC simulation and data yields of the regions used for the calculation of the Isolation Factors in the Fake Factor method shown in Section 8.1.2.

	Region	Signal	VV	Zll	Top	W	Ztt	Other Higgs	Data
$\tau_{\text{lep}}\tau_{\text{1-prong}}$	SSP anti-τ iso CR	3.37 ± 0.07	619.99±9.86	8158.16 ± 299.00	2524.25 ± 20.70	19380.07 ± 391.89	1144.95 ± 15.90	12.98 ± 0.32	55375
	SSP anti-τ non-iso CR	0.02 ± 0.01	3.25 ± 0.68	16.85 ± 7.67	39.77 ± 2.98	118.26 ± 42.12	8.91 ± 1.11	0.10 ± 0.03	4714
$\tau_{\text{lep}}\tau_{\text{3-prong}}$	SSP anti-τ iso CR	2.08 ± 0.06	1045.17 ± 26.01	12300.74 ± 351.97	4474.21 ± 25.69	34359.56 ± 491.42	1135.20 ± 16.54	8.91 ± 0.25	74087
	SSP anti-τ non-iso CR	0.01 ± 0.00	3.66 ± 0.77	42.16 ± 18.20	55.04 ± 3.41	147.53 ± 34.42	6.61 ± 1.28	0.08 ± 0.02	6465

Table A.1.2: MC simulated events and data yields in the same-sign Preselection (SSP) anti-τ inverting (non-iso) or without inverting (iso) the isolation requirement on the light lepton control regions (CR) for the $\tau_{\text{lep}}\tau_{\text{1-prong}}$ and $\tau_{\text{lep}}\tau_{\text{3-prong}}$ channels. Only statistical uncertainties are shown.

Table A.1.3 lists the total MC simulation and data-driven expected yields and data in the four different VBF decay channels and inclusive SR where the fake background contribution has been estimated using the simplified Fake Factor method shown in Section 8.5.2.

Region	Signal	VV	Zll	Top	W	Fakes	Ztt	Other Higgs	Sum bkg	Data
VBF SR inclusive	162.07±0.53	68.61±1.61	72.78±11.01	114.35±4.04	23.73±5.26	559.38±8.14	2729.36±20.21	60.15 ± 0.74	3790.42±24.02	3669
VBF SR $\tau_e\tau_{\text{had}}$	76.18± 0.37	32.10±1.13	49.31±9.65	52.18±2.75	7.67±2.05	307.94±6.24	1248.09±13.34	28.99 ± 0.52	1802.46±16.87	1801
VBF SR $\tau_\mu\tau_{\text{had}}$	85.88±0.38	36.51±1.15	23.47±5.31	62.17±2.96	16.06±4.84	251.44.48±5.24	1481.27±15.17	31.15 ± 0.52	1987.96±17.10	1868
VBF SR $\tau_{\text{lep}}\tau_{\text{1-prong}}$	129.00±0.48	52.83±1.41	63.05±9.98	86.34±3.51	17.10±4.57	414.86±7.80	2136.99±17.33	47.51 ± 0.66	2947.68±20.88	2919
VBF SR $\tau_{\text{lep}}\tau_{\text{3-prong}}$	33.07±0.24	15.78±0.77	9.73±4.65	28.01±1.99	6.64±2.60	144.51±2.34	592.37±10.38	12.63±0.34	842.75 ± 11.87	750

Table A.1.3: MC simulated and data yields in the VBF signal region (SR) and its four decay channels including the estimation of the fake background contribution calculated with the simplified Fake Factor method. Only statistical uncertainties are shown.

A.2 Same-sign closure test

Table A.2.1 lists the MC expected yields and data in the CRs used for the calculation of the Individual Fake Factors in the same-sign (SS) closure test, i.e. inverting in all CRs the same-sign charge between the light lepton and the τ_{had} requirement.

Region	Signal	VV	Zll	Top	W	Ztt	Other Higgs	Data
W CR 1-prong	0.06 ± 0.01	3.97 ± 0.88	-0.27 ± 0.62	1.71 ± 0.46	3.27 ± 0.96	0.46 ± 0.18	0.10 ± 0.02	86
W anti-τ CR 1-prong	0.07 ± 0.01	2.56 ± 0.35	4.63 ± 3.35	19.12 ± 1.69	21.82 ± 2.86	1.01 ± 0.30	0.11 ±0.03	815
W CR 3-prong	0.23 ± 0.02	1.51 ± 0.14	1.20 ± 0.83	1.22 ± 0.39	0.06 ± 2.29	5.55 ± 1.50	0.15 ± 0.03	70
W anti-τ CR 3-prong	0.02 ± 0.01	2.43 ± 0.55	13.08 ± 2.93	43.69 ± 2.61	61.03 ± 8.67	0.54 ± 0.18	0.03 ± 0.02	1193
QCD CR 1-prong	0.01 ± 0.00	0.07 ± 0.04	0.00 ± 0.00	0.20 ± 0.20	0.05 ± 0.05	0.10 ± 0.08	0.01 ± 0.01	27
QCD anti-τ CR 1-prong	0.01 ± 0.00	0.01 ± 0.00	0.00 ± 0.00	0.35 ± 0.22	0.23 ± 0.22	0.19 ± 0.15	0.00 ± 0.00	234
QCD CR 3-prong	0.00 ± 0.00	0.00 ± 0.00	0.00 ± 0.00	0.27 ± 0.19	0.00 ± 0.00	0.04 ± 0.04	0.00 ± 0.00	9
QCD anti-τ CR 3-prong	0.00 ± 0.00	0.07 ± 0.06	0.14 ± 0.13	0.51 ± 0.26	0.96 ± 0.55	0.16 ± 0.10	0.00 ± 00	362

Table A.2.1: MC yields and data in the control regions (CR) used in the Fake Factor method validation, same-sign closure test. W+jets (W) and multi-jet (QCD) CRs are split into the 1-prong and 3-prong channels. Only statistical uncertainties are shown.

Figure A.2.1 illustrates the Individual Fake Factors calculated in the SS closure test depending on the p_{T} of the τ_{had}.

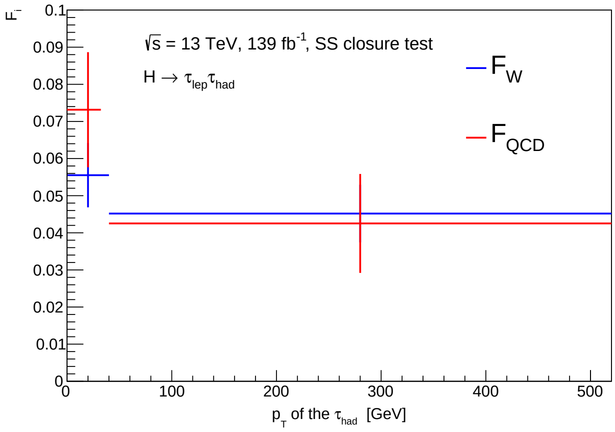
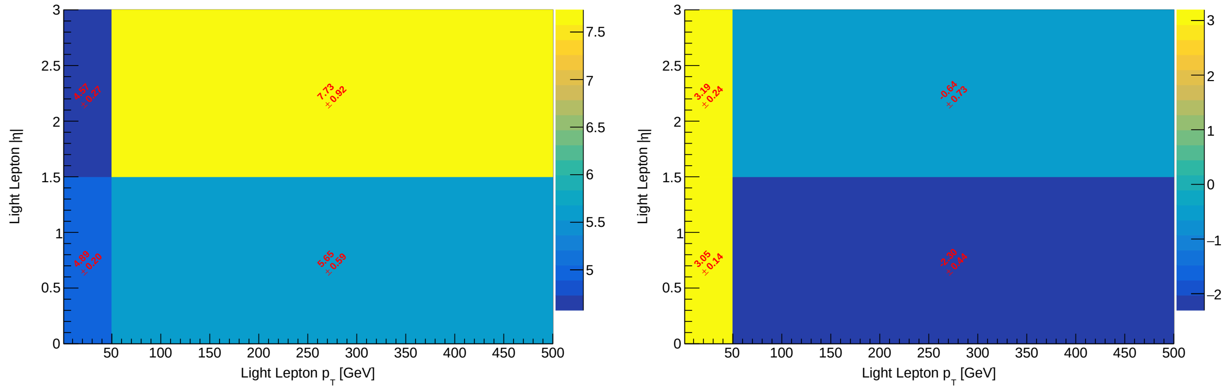


Figure A.2.1: Individual Fake Factors calculated for the same-sign (SS) closure test, binned in the transverse momentum (p_{T}) of the τ_{had}. Individual Fake Factors corresponding to W+jets (F_{W}) in blue and multi-jet processes (F_{QCD}) in red. Only statistical uncertainties are shown.

Figure A.2.2 illustrates the results of the Isolation Factors calculated in the OSP anti-τ CR for the SS closure test described in Section 8.3.3.



(a) $\mathcal{I}$ in the $\tau_{\text{lep}}\tau_{\text{1-prong}}$ channel (b) $\mathcal{I}$ in the $\tau_{\text{lep}}\tau_{\text{3-prong}}$ channel

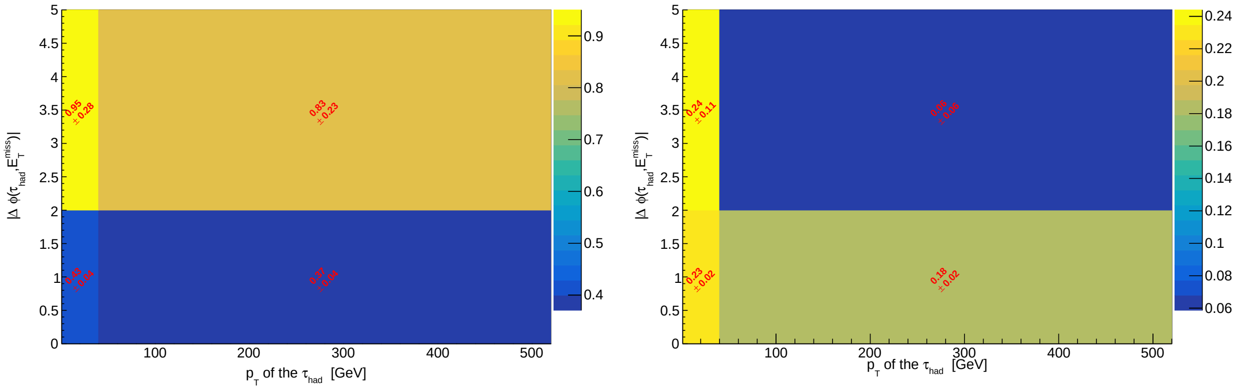
Figure A.2.2: Isolation Factors ($\mathcal{I}$) for the same-sign (SS) closure test binned in the light lepton p_{T} and $|\eta|$ split into the (a) 1- ($\tau_{\text{lep}}\tau_{\text{3-prong}}$) and (b) 3-prong ($\tau_{\text{lep}}\tau_{\text{3-prong}}$) channels, calculated in the opposite-sign Preselection (OSP) anti-τ CR as detailed in Section 8.3.3. Only statistical uncertainties are shown.

Table A.2.2 lists the expected MC yields and data on each CR used in the calculation of the Isolation Factors for the SS closure test described in section 8.3.3.

	Region	**Signal**	**VV**	**Zll**	**Top**	**W**	**Ztt**	**Data**
1-prong	OSP anti-τ iso CR	326.85±1.58	1530.63±15.82	8830.15±299.80	4631.57±26.15	48640.61±551.15	16282.11±60.89	107582
	OSP anti-τ non-iso CR	2.03±0.12	6.50±0.90	58.80±17.53	64.51±3.57	297.73±47.55	111.96±5.09	5970
3-prong	OSP anti-τ iso CR	176.59±1.14	2068.68±19.18	12862.60±357.38	6845.86±31.87	75581.91±709.86	10437.55±45.49	125858
	OSP anti-τ non-iso CR	1.13± 0.10	9.80±1.22	55.85±21.71	78.55±4.15	266.88±75.96	73.66±4.58	8258

Table A.2.2: MC simulated and data yields in the opposite-sign Preselection (OSP) anti-τ CRs, inverting (iso) or not (non-iso) the isolation requirement on the light lepton, split in the (a) $\tau_{\text{lep}}\tau_{\text{1-prong}}$ and (b) $\tau_{\text{lep}}\tau_{\text{3-prong}}$ channels for the same-sign (SS) closure test.

Figure A.2.3 illustrates the fractional contribution from QCD processes ($\mathrm{R_{QCD}}$) calculated for the fake background estimation in the SS closure test.



(a) $\mathrm{R_{QCD}}$ in the $\tau_e\tau_{\mathrm{had}}$ channel

(b) $\mathrm{R_{QCD}}$ in the $\tau_\mu\tau_{\mathrm{had}}$ channel

Figure A.2.3: Fractional contribution from multi-jet events $\mathrm{R_{QCD}}$ calculated for the SS closure test split into the (a) 1-prong ($\tau_{\mathrm{lep}}\tau_{\text{1-prong}}$) and (b) 3-prong($\tau_{\mathrm{lep}}\tau_{\text{3-prong}}$) channels, divided in four regions depending on the transverse momentum (p_{T}) of the τ_{had} and the azimuthal angle between the τ_{had} and the missing transverse energy ($|\Delta\phi(\tau_{\mathrm{had}}, E_{\mathrm{T}}^{\mathrm{miss}})|$). Only statistical uncertainties are shown.

Table A.2.3 lists the total number of events and MC simulated yields in the VBF inclusive SR and its decay channels result of the SS closure test, the signal region must be enriched in fake events.

Channel	**Signal**	**VV**	$Z \to \ell\ell$	**Top**	**W**	**Fakes**	$Z \to \tau\tau$	**Other Higgs**	**Data**
VBF SR inclusive	1.23 ± 0.05	7.68 ± 0.34	0.32 ± 2.29	5.48 ± 0.94	2.64 ± 3.25	350.87 ± 6.15	21.96 ± 2.06	0.75 ± 0.08	362
VBF SR $\tau_e\tau_{\mathrm{had}}$	0.70 ± 0.04	4.06 ± 0.26	0.26 ± 2.29	2.22 ± 0.56	-1.14 ± 2.56	203.35 ± 4.79	12.01 ± 1.73	0.39 ± 0.06	225
VBF SR $\tau_\mu\tau_{\mathrm{had}}$	0.54 ± 0.03	3.62 ± 0.21	0.06 ± 0.03	3.26 ± 0.75	3.78 ± 2.00	147.51 ± 3.84	9.95 ± 1.12	0.36 ± 0.05	137
VBF SR $\tau_{\mathrm{lep}}\tau_{\text{1-prong}}$	1.00 ± 0.04	6.18 ± 0.31	-0.88 ± 2.13	4.27 ± 0.85	2.58 ± 2.30	290.58 ± 6.05	16.41 ± 1.41	0.60 ± 0.07	292
VBF SR $\tau_{\mathrm{lep}}\tau_{\text{3-prong}}$	0.23 ± 0.02	1.51 ± 0.14	1.20 ± 0.83	1.22 ± 0.39	0.06 ± 2.29	60.29 ± 1.07	5.55 ± 1.50	0.15 ± 0.03	70

Table A.2.3: Number of data and MC simulated yields in the different VBF signal region (SR) channels including the fake background contribution estimated in the same-sign closure test. Only statistical uncertainties are shown.

A.3 Uncertainties formulas

Uncertainties are calculated using standard error propagation, for $z = h(x, y)$:

$$\Delta z = \sqrt{\left(\frac{\partial h}{\partial x}\right)^2 \Delta x^2 + \left(\frac{\partial h}{\partial y}\right)^2 \Delta y^2} \tag{A.3.1}$$

Individual Fake Factors

Statistical uncertainties are calculated as

$$\Delta \mathrm{F}_i = \sqrt{\left(\frac{-\Delta \mathrm{N}^{\text{pass, CR}_i}_{\text{MC, not j}\to\tau}}{\mathrm{N}^{\text{anti,CR}_i}_{\text{Data}} - \mathrm{N}^{\text{anti, CR}_i}_{\text{MC, not j}\to\tau}}\right)^2 + \left(\frac{(\mathrm{N}^{\text{pass,CR}_i}_{\text{Data}} - \mathrm{N}^{\text{pass, CR}_i}_{\text{MC, not j}\to\tau}) \cdot \Delta \mathrm{N}^{\text{anti, CR}_i}_{\text{MC, not j}\to\tau}}{(\mathrm{N}^{\text{anti,CR}_i}_{\text{Data}} - \mathrm{N}^{\text{anti, CR}_i}_{\text{MC, not j}\to\tau})^2}\right)^2}. \tag{A.3.2}$$

Isolation Factors

Statistical uncertainties are calculated using standard error propagation:

$$\Delta \mathcal{I}_{\text{Stat.}} = \sqrt{\left(\frac{-\Delta \mathrm{N}^{\text{iso, SSP}}_{\text{MC, true lepton}}}{\mathrm{N}^{\text{non-iso, SSP}}_{\text{data}} - \mathrm{N}^{\text{non-iso, SSP}}_{\text{MC, true lepton}}}\right)^2 + \left(\frac{(\mathrm{N}^{\text{iso, SSP}}_{\text{data}} - \mathrm{N}^{\text{iso, SSP}}_{\text{MC, true lepton}}) \cdot \Delta \mathrm{N}^{\text{non-iso, SSP}}_{\text{MC, true lepton}}}{(\mathrm{N}^{\text{non-iso, SSP}}_{\text{data}} - \mathrm{N}^{\text{non-iso, SSP}}_{\text{MC, true lepton}})^2}\right)^2}. \tag{A.3.3}$$

Systematic uncertainty is added in quadrature to the statistical uncertainty:

$$\Delta \mathcal{I} = \sqrt{\Delta \mathcal{I}^2_{\text{Stat.}} + \Delta \mathcal{I}^2_{\text{SS-OS}}}. \tag{A.3.4}$$

Fractional Contribution $\mathbf{R_{QCD}}$

Total uncertainty:

$$\Delta \mathrm{R}_{\text{QCD}} = \sqrt{\left(\frac{-\mathcal{I}\cdot\Delta \mathrm{N}^{\text{QCD CR,anti-}\tau}_{\text{MC,true lep}}}{\mathrm{N}^{\text{anti-}\tau}_{\text{Data}} - \mathrm{N}^{\text{anti-}\tau}_{\text{MC, not j}\to\tau}}\right)^2 + \left(\frac{\mathcal{I}\cdot\Delta \mathrm{N}^{\text{anti-}\tau}_{\text{MC, not j}\to\tau}\cdot(\mathrm{N}^{\text{QCD CR,anti-}\tau}_{\text{Data}} - \mathrm{N}^{\text{QCD CR,anti-}\tau}_{\text{MC,true lep}})}{(\mathrm{N}^{\text{anti-}\tau}_{\text{Data}} - \mathrm{N}^{\text{anti-}\tau}_{\text{MC, not j}\to\tau})^2}\right)^2 + \left(\frac{\Delta\mathcal{I}\cdot \mathrm{R}_{\text{QCD}}}{\mathcal{I}}\right)^2}. \tag{A.3.5}$$

Decomposition into the two contributions of the uncertainty:

$$\Delta \mathrm{R}^2_{\text{QCD}} = A + B + \left(\frac{\mathrm{R}_{\text{QCD}}}{\mathcal{I}}\right)^2 \cdot \Delta \mathcal{I}_{\text{Stat.}^2} + \left(\frac{\mathrm{R}_{\text{QCD}}}{\mathcal{I}}\right)^2 \cdot \Delta \mathcal{I}_{\text{SS-OS}^2} = \Delta \mathrm{R}_{\text{QCD, Stat.}} + \Delta \mathrm{R}_{\text{QCD, SS-OS}} \tag{A.3.6}$$

with

$$A = \left(\frac{-\mathcal{I} \cdot \Delta \mathrm{N}^{\text{QCD CR,anti-}\tau}_{\text{MC,true lep}}}{\mathrm{N}^{\text{anti-}\tau}_{\text{Data}} - \mathrm{N}^{\text{anti-}\tau}_{\text{MC, not j}\to\tau}}\right)^2 \quad ; \quad B = \left(\frac{\mathcal{I} \cdot \Delta \mathrm{N}^{\text{anti-}\tau}_{\text{MC, not j}\to\tau} \cdot (\mathrm{N}^{\text{QCD CR,anti-}\tau}_{\text{Data}} - \mathrm{N}^{\text{QCD CR,anti-}\tau}_{\text{MC,true lep}})}{(\mathrm{N}^{\text{anti-}\tau}_{\text{Data}} - \mathrm{N}^{\text{anti-}\tau}_{\text{MC, not j}\to\tau})^2}\right)^2 \tag{A.3.7}$$

and

$$\mathrm{R}_{\mathrm{QCD,\,Stat.}} = A + B + \left(\frac{\mathrm{R}_{\mathrm{QCD}}}{\mathcal{I}}\right)^2 \cdot \Delta\mathcal{I}_{\mathrm{Stat.}^2} \ \ , \ \ \mathrm{R}_{\mathrm{QCD,\,SS\text{-}OS}} = \left(\frac{\mathrm{R}_{\mathrm{QCD}}}{\mathcal{I}}\right)^2 \cdot \Delta\mathcal{I}^2_{\mathrm{SS\text{-}OS}}, \tag{A.3.8}$$

so that the systematic uncertainty SS-OS from the Isolation Factor can be propagated to $\mathrm{R}_{\mathrm{QCD}}$.

Fractional Contribution $\mathbf{R_W}$

$$\Delta\mathrm{R}_{\mathrm{W}} = \Delta\mathrm{R}_{\mathrm{QCD}} \tag{A.3.9}$$

Combined Fake Factor F

Including both statistical and systematic uncertainties, systematic uncertainties are inherited from $\mathrm{R}_{\mathrm{QCD}}$,

$$\Delta\mathcal{F} = \sqrt{((1 - \mathrm{R}_{\mathrm{QCD}}) \cdot \Delta\mathrm{F}_{\mathrm{W}})^2 + (-(\mathrm{F}_{\mathrm{W}} - \mathrm{F}_{\mathrm{QCD}}) \cdot \Delta\mathrm{R}_{\mathrm{QCD}})^2 + (\mathrm{R}_{\mathrm{QCD}} \cdot \Delta\mathrm{F}_{\mathrm{QCD}})^2}. \tag{A.3.10}$$

Appendix B

Supplementary information used for the Test of CP invariance

B.1 Optimization of the fitting parameters

The selection of the fit range and step size for the Gauge curve fit is based in the following studies.

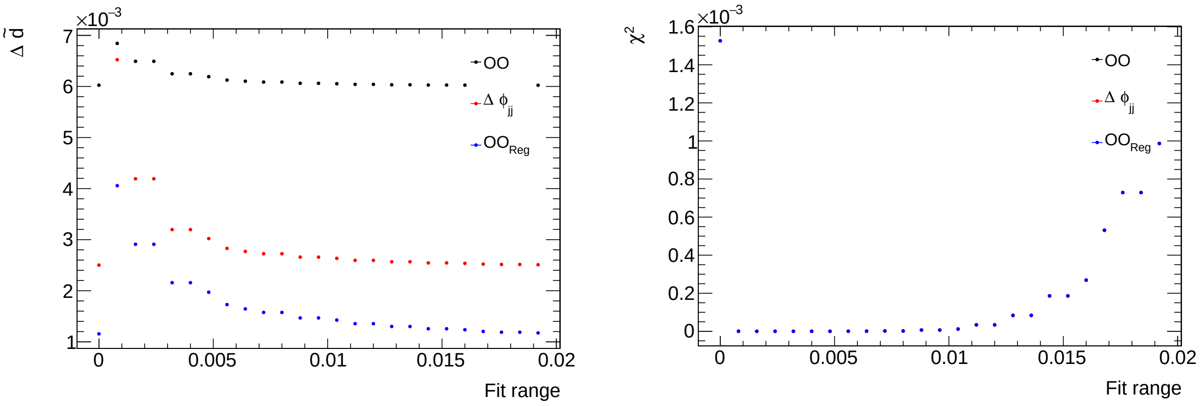
The fit range is chosen such that the fit is done using the maximum number of $<$ CP-odd variable $>$ values while maintaining the linear behavior of the Gauge curve, that is, for small $\tilde{d}$ values. The error on $\tilde{d}$ is plotted with respect to the fit range and the fit range with respect to χ^2, as shown in Figure B.1.1.

The expected sensitivity on $\tilde{d}$ is approximately constant from fit ranges around 0.01. Around this value value, Figure B.1.1b shows that the χ^2 value starts to rise. This is observed for the three CP-odd variables. Considering this, the chosen value for the range of the Gauge curve fit is [-0.01,0.01].

Furthermore, the value and error of $\tilde{d}$ was estimated for different step sizes. The step size is defined as

$$\text{Step size} = \frac{x_{\text{max}} - x_{\text{min}}}{\text{N}_{\text{bin}}} \tag{B.1.1}$$

where x_{max} and x_{min} are the extremes of the fit range and N_{bin} the total number of $\tilde{d}$ values used for the fit. $\tilde{d}$ was estimated for $\text{N}_{\text{bin}} = 20$, 50 and 100. $\text{N}_{\text{bin}} = 50$ was chosen considering the computation time and the amount of sensitivity on $\tilde{d}$ lost ($\sim$ 0.5% between 50 and 100 and $\sim$ 1.9% between 20 and 100 for the three CP-odd variables), as can be seen in Tables B.2.4, B.2.1 and B.2.3.



(a) Fit range (b) χ^2

Figure B.1.1: Distributions of (a) the fit range and (b) the χ^2 values for the Gauge curve fit for the three different CP-odd variables; Optimal Observable ($\mathcal{OO}$), azimuthal angle between the two leading jets ($\Delta\phi_{jj}$) and the Optimal Observable defined using symbolic regression methods ($\mathcal{OO}_{\text{Reg}}$).

B.2 Further results

This Section shows the complete results (extended version) of the fits done for estimating the $\tilde{d}$ values described in Section 10.3.

Table B.2.1 lists the results of the Gauge curve fit done using the optimized parameters that is, the final results, using 50 $\tilde{d}$ values in $\tilde{d} \in [-0.01, 0.01]$, i.e. a step size of 0.0004.

CP-odd variable	$\tilde{d} \times 10^{-7}$	±	Slope	y-intercept	$\chi^2 \times 10^{-5}$	$<$ CP-odd variable $>|_{\tilde{d}=0}$
$\mathcal{OO}$	3.708	0.00625	16.69 ± 2.53	0.0480 ± 0.0146	2.12	0.0480 ± 0.103
$\Delta\phi_{jj}$	1.211	0.00893	4.61 ± 1.00	0.0572 ± 0.0058	1.04	0.0572 ± 0.0408
$\mathcal{OO}_{\text{Reg}}$	7.056	0.00627	-2.23 ± 0.34	-0.00846 ± 0.0020	2.00	-0.00846 ± 0.0138

Table B.2.1: Gauge curve fit parameter results using 50 $\tilde{d}$ values in $\tilde{d} \in [-0.01, 0.01]$.

Table B.2.2 lists the results of the Gauge curve fit done using 50 $\tilde{d}$ values in $\tilde{d} \in [-0.01, 0.01]$, i.e. a step size of 0.0004, with only signal samples, no background contribution included.

CP-odd variable	$\tilde{d} \times 10^{-7}$	±	Slope	y-intercept	χ^2	$<$ CP-odd variable $>|_{\tilde{d}=0}$
$\mathcal{OO}$	5.797	0.000606	30.24 ± 0.45	0.0192 ± 0.0026	0.0022	0.0193 ± 0.0182
$\Delta\phi_{jj}$	2.843	0.000828	8.36 ± 0.17	0.0108 ± 0.0010	0.0012	0.0012 ± 0.0069
$\mathcal{OO}_{\text{Reg}}$	8.781	0.000626	-4.03 ± 0.06	-0.0020 ± 0.0004	0.0020	-0.00196 ± 0.00250

Table B.2.2: Gauge curve fit parameter results using 50 $\tilde{d}$ values in $\tilde{d} \in [-0.01, 0.01]$ with only signal samples.

Table B.2.3 lists the results of the Gauge curve fit done using 100 $\tilde{d}$ values in $\tilde{d} \in [-0.01, 0.01]$, i.e. a step size of 0.0002.

CP-odd variable	$\tilde{d} \times 10^{-7}$	$\pm$	Slope	y-intercept	$\chi^2 \times 10^{-5}$	$<$ CP-odd variable $>|_{\tilde{d}=0}$
$\mathcal{OO}$	4.779	0.00622	16.69 ± 1.79	0.0480 ± 0.0103	4.27	0.0480 ± 0.103
$\Delta\phi_{jj}$	2.280	0.00890	4.61 ± 0.71	0.0572 ± 0.0041	2.08	0.0572 ± 0.0408
$\mathcal{OO}_{\text{Reg}}$	8.098	0.00624	-2.23 ± 0.24	-0.00846 ± 0.0014	4.06	-0.00846 ± 0.0138

Table B.2.3: Gauge curve fit parameter results using 100 $\tilde{d}$ values in $\tilde{d} \in [-0.01, 0.01]$.

Table B.2.4 lists the results of the Gauge curve fit done using 20 $\tilde{d}$ values in $\tilde{d} \in [-0.01, 0.01]$, i.e. a step size of 0.001.

CP-odd variable	$\tilde{d} \times 10^{-7}$	$\pm$	Slope	y-intercept	$\chi^2 \times 10^{-6}$	$<$ CP-odd variable $>|_{\tilde{d}=0}$
$\mathcal{OO}$	5.073	0.00634	16.69 ± 4.00	0.0479 ± 0.0232	8.45	0.0480 ± 0.103
$\Delta\phi_{jj}$	-1.974	0.00907	4.61 ± 1.58	0.0572 ± 0.0092	4.23	0.0572 ± 0.0408
$\mathcal{OO}_{\text{Reg}}$	3.925	0.00634	-2.23 ± 0.54	-0.00846 ± 0.0031	7.73	-0.00846 ± 0.0138

Table B.2.4: Gauge curve fit parameter results using 20 $\tilde{d}$ values in $\tilde{d} \in [-0.01, 0.01]$.

Table B.2.5 lists the results of the Maximum Likelihood (ML) fits done using 50 $\tilde{d}$ values in $\tilde{d} \in [-0.07, 0.07]$ estimated from the CP-odd variables fitted using 10,20,60 and 100 bins.

	10	20	60	100
$\mathcal{OO}$	± 0.0172	± 0.0168	± 0.0168	± 0.0168
$\Delta\phi_{jj}$	± 0.0227	± 0.0227	± 0.0227	± 0.0223
$\mathcal{OO}_{\text{Reg}}$	± 0.0172	± 0.0172	± 0.0168	± 0.0168

Table B.2.5: Expected $\tilde{d}$ errors obtained from the Maximum Likelihood fit, described in Section 10.2.2 for the three CP-odd variables; Optimal Observable ($\mathcal{OO}$), signed azimuthal angle between the two tagging jets ($\Delta\phi_{jj}$), and the Optimal Observable obtained using symbolic regression methods ($\mathcal{OO}_{\text{Reg}}$). Distributions were fit using 10, 20, 60 and 100 bins. Only statistical uncertainties are shown.

Table B.2.6 lists the results of the Maximum Likelihood (ML) fit done using 100 $\tilde{d}$ values in $\tilde{d} \in [-0.07, 0.07]$.

	68% CI		95% CI	
	$\tilde{d} \times 10^{-17}$	$\pm$	$\tilde{d} \times 10^{-17}$	$\pm$
$\mathcal{OO}$	-6.24	-0.0182	-4.16	-0.0378
$\Delta\phi_{jj}$	-6.24	-0.0224	-4.16	-0.0448
$\mathcal{OO}_{\text{Reg}}$	-6.24	-0.0182	-6.24	-0.0378

Table B.2.6: Expected $\tilde{d}$ values obtained from the Maximum Likelihood fit, described in Section 10.2.2 for the three CP-odd variables; Optimal Observable ($\mathcal{OO}$), signed azimuthal angle between the two tagging jets ($\Delta\phi_{jj}$), and the Optimal Observable obtained using symbolic regression methods ($\mathcal{OO}_{\text{Reg}}$), using 100 $\tilde{d}$ values. Only statistical uncertainties are shown.

Table B.2.7 lists the results of the Maximum Likelihood (ML) fit done using 20 $\tilde{d}$ values in $\tilde{d} \in [-0.07, 0.07]$.

	68% CI		95% CI	
	$\tilde{d} \times 10^{-18}$	$\pm$	$\tilde{d} \times 10^{-18}$	$\pm$
$\mathcal{OO}$	-1.73	-0.0140	-1.73	-0.0350
$\Delta\phi_{jj}$	-1.73	-0.0201	-1.73	-0.0412
$\mathcal{OO}_{\text{Reg}}$	-1.73	-0.0140	-1.73	-0.0350

Table B.2.7: Expected $\tilde{d}$ values obtained from the Maximum Likelihood fit, described in Section 10.2.2 for the three CP-odd variables; Optimal Observable ($\mathcal{OO}$), signed azimuthal angle between the two tagging jets ($\Delta\phi_{jj}$), and the Optimal Observable obtained using symbolic regression methods ($\mathcal{OO}_{\text{Reg}}$), using 20 $\tilde{d}$ values. Only statistical uncertainties are shown.

Table B.2.8 lists the results of the Maximum Likelihood (ML) fit done using 50 $\tilde{d}$ values in $\tilde{d} \in [-0.07, 0.07]$ using only signal processes, no background processes are included.

	68% CI		95% CI	
	$\tilde{d} \times 10^{-17}$	$\pm$	$\tilde{d} \times 10^{-17}$	$\pm$
$\mathcal{OO}$	-4.16	-0.0140	-4.16	-0.0280
$\Delta\phi_{jj}$	-4.16	-0.0168	-4.16	-0.0336
$\mathcal{OO}_{\text{Reg}}$	-4.16	-0.0140	-4.16	-0.0280

Table B.2.8: Expected $\tilde{d}$ values obtained from the Maximum Likelihood fit, described in Section 10.2.2 for the three CP-odd variables; Optimal Observable ($\mathcal{OO}$), signed azimuthal angle between the two tagging jets ($\Delta\phi_{jj}$), and the Optimal Observable obtained using symbolic regression methods ($\mathcal{OO}_{\text{Reg}}$) using only signal processes. Only statistical uncertainties are shown.

Bibliography

[1] UA2 Collaboration, *Evidence for* $Z^0 \rightarrow e^+e^-$ *at the CERN pp collider*, Phys. Lett. B **129** (1983) 130, `https://www.sciencedirect.com/science/article/abs/pii/037026938390744X?via%3Dihub`.

[2] UA1 Collaboration, *Experimental observation of isolated large transverse energy electrons with associated missing energy at* $\sqrt{s} = 540$ *GeV*, Phys. Lett. B 122:103, (1983), `https://www.sciencedirect.com/science/article/abs/pii/0370269383911772?via%3Dihub`.

[3] ATLAS Collaboration, *Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC*, Phys. Lett. B **716**, 1, (2012), 1-29, CERN-PH-EP-2012-218, `https://arxiv.org/abs/1207.7214`.

[4] CMS Collaboration, *Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC*, Phys. Lett. B **716** (2012) 30-61, `http://dx.doi.org/10.1016/j.physletb.2012.08.021`.

[5] L. Canetti, M. Drewes and M. Shaposhnikov, *Matter and Antimatter in the Universe*, New J. Phys. 1**4** (2012) 095012, `https://arxiv.org/abs/1204.4186`.

[6] D.N. Spergel, L. Verde, H.V. Peiris, E. Komatsu, M.R. Nolta, C.L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, A. Kogut, M. Limon, S.S. Meyer, L. Page, G.S. Tucker, J.L. Weiland, E. Wollack and E. L. Wright, *First Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Determination of Cosmological Parameters*, ApJS **148** 175 (2003), `https://arxiv.org/abs/astro-ph/0302209`.

[7] Planck Collaboration, *Planck 2018 results. VI. Cosmological parameters*, A&A 641, A6 (2020), `https://arxiv.org/abs/1807.06209`.

[8] A.D. Sakharov, *Violation of CP Invariance, C asymmetry, and baryon asymmetry of the universe*, Pisma Zh. Eksp. Teor. Fiz. **5** (1967) 32-35, `https://inspirehep.net/literature/51345`.

[9] J.M. Cline, *Baryogenesis*, Les Houches Summer School - Session 86: Particle Physics and Cosmology: The Fabric of Spacetime, 9, (2006) `https://arxiv.org/abs/hep-ph/0609145`.

[10] ATLAS Collaboration, *Test of CP invariance in vector-boson fusion production of the Higgs boson using the Optimal Observable method in the ditau decay channel with the ATLAS detector*, Eur. Phys. J. C **76** (2016) 658, `https://arxiv.org/abs/1602.04516`.

[11] ATLAS Collaboration, *Test of CP invariance in vector-boson fusion production of the Higgs boson in the* $H \rightarrow \tau\tau$ *channel in proton–proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Physics Letters B Volume 805, (2020), `https://doi.org/10.48550/arXiv.2002.05315`.

[12] T. Figy, V. Hankele, G. Klamke and D. Zeppenfeld, *Anomalous Higgs boson couplings in vector boson fusion at the CERN LHC*, Phys. Rev. D. **74**, (2006), 095001, `https://arxiv.org/abs/hep-ph/0609075`.

[13] T. Plehn, D. Rainwater and D. Zeppenfeld, *Determining the Structure of Higgs Couplings at the LHC*, Phys. Rev. Lett. 88:051801, (2002), `https://doi.org/10.48550/arXiv.hep-ph/0105325`.

[14] A. Butter, T. Plehn, N. Soybelman and J. Brehmer, *Back to the Formula – LHC Edition*, `https://doi.org/10.48550/arXiv.2109.10414`.

[15] N.C. Hall, I. Criddle, A. Crossland, C. Englert, P. Forbes, R. Hankache and A.D. Pilkington, *Machine-enhanced CP-asymmetries in the electroweak sector*, Phys. Lett. B **832** (2022) 137246, `https://arxiv.org/pdf/2209.05143.pdf`.

[16] D0 Collaboration, *Search for High Mass Top Quark Production in pp* $^{-}$ *Collisions at* $\sqrt{s} = 1.8$ *TeV*, Phys. Rev. Lett. **74** (1995) 2422, `https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.74.2422`.

[17] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967) 1264, `https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.19.1264`.

[18] PLUTO Collaboration, *Jet analysis of the* γ *(9.46) decay into charged hadrons*, Phys. Lett. B **82** (1979) 449, `https://www.sciencedirect.com/science/article/abs/pii/037026937990265X?via%3Dihub`.

[19] D. Griffith, *Introduction to Elementary Particle Physics*, Wiley (1984).

[20] F. Halzen and A. D. Martin, *Quarks Leptons: An Introductory Course in Modern Particle Physics*, John Wiley & Sons (1984), ISBN: 978-0-471-8874-16.

[21] M. Böhm, A. Denner and H. Joos, *Gauge Theories of the Strong and Electroweak Interaction*, B.G. Teubner Stuttgart-Leipzig-Wiesbaden (2001), ISBN: 978-3-322-80160-9.

[22] H.D. Politzer, *Reliable Perturbative Results for Strong Interactions?*, Phys. Rev. Lett. **30** (1973), 1346, `http://dx.doi.org/10.1103/PhysRevLett.30.1346`.

[23] C. MissMJ, Own work by uploader, Fermilab, Office of Science, United States Department of Energy, Particle Data Group (2019), `https://commons.wikimedia.org/w/index.php?curid=4286964`.

[24] S.L. Glashow, *Partial-symmetries of weak interactions*, Nuclear Phys. Vol. 22, 4 (1961) 579-588, `https://www.sciencedirect.com/science/article/abs/pii/0029558261904692`.

[25] S. Weinberg, *A Model of Leptons*, Phys. Rev. Lett. **19** (1967), 1264, `https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.19.1264`.

[26] A. Salam, *Weak and Electromagnetic Interactions*, Conf. Proc. C 680519 (1968), 367-377, `https://inspirehep.net/literature/53083`.

[27] K. Nishijima, *Charge Independence Theory of V Particles*, Progress of Theoretical Physics, Vol. 13, 3 (1955), 285–304, `https://academic.oup.com/ptp/article/13/3/285/1924498`.

[28] A. Pich, The Standard Model of Electroweak Interactions, Proceedings, High-energy Physics, ESHEP (2010), 1-50, CERN-2012-001, `https://arxiv.org/abs/1201.0537`.

[29] Particle Data Group, *Review of Particle Physics*, Phys. Rev. D **98**, 030001, (2018), `https://journals.aps.org/prd/abstract/10.1103/PhysRevD.98.030001`.

[30] P.W. Higgs, *Broken symmetries, massless particles and gauge fields*, Phys. Lett. **12**, Vol. 12, 2, (1964), 132-133, `https://www.sciencedirect.com/science/article/abs/pii/0031916364911369?via%3Dihub`.

[31] P.W. Higgs, *Spontaneous Symmetry Breakdown without Massless Bosons*, Phys. Rev. **145** (1966), 1156, `https://journals.aps.org/pr/abstract/10.1103/PhysRev.145.1156`.

[32] F. Englert and R. Brout, *Broken Symmetry and the Mass of Gauge Vector Mesons*, Phys. Rev. Lett. **13** (1964), 321, `https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.13.321`.

[33] P.W. Higgs, *Broken Symmetries and the Masses of Gauge Bosons*, Phys. Rev. Lett. **13** (1964), 508, `https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.13.508`.

[34] J. Riebesell, *Higgs Potential*, `https://tikz.net/higgs-potential/`.

[35] ATLAS Collaboration, *Measurement of the Higgs boson mass in the* $H \to ZZ^* \to 4\ell$ *and* $H \to \gamma\gamma$ *channels with* $\sqrt{s} = 13$ *TeV pp collisions using the ATLAS detector*, Phys. Lett. B **784** (2018), 345, CERN-EP-2018-085, `https://arxiv.org/abs/1806.00242`.

[36] C. Grojean, *Higgs Physics*, Proceedings of the 2015 CERN–Latin-American School of High-Energy Physics, Vol. 5, (2016), 143, `https://e-publishing.cern.ch/index.php/CYR/article/view/435`.

[37] LHC Higgs Cross Section Working Group, *Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector*, CERN Yellow Reports: Monographs 2 (2017), CERN–2017–002-M, `https://arxiv.org/abs/1610.07922`.

[38] J. Brehmer, F. Kling, T. Plehn and T.M.P. Tait, *Better Higgs-CP tests through information geometry*, Phys. Rev. D **97**, (2018), 095017, `https://journals.aps.org/prd/pdf/10.1103/PhysRevD.97.095017`.

[39] ATLAS Open Data, *The Higgs boson*, `http://opendata.atlas.cern/books/current/get-started/_book/the-higgs-boson.html`.

[40] Particle Data Group. τ *BRANCHING FRACTIONS*, Particle Data Group (2016), `https://pdg.lbl.gov/2016/reviews/rpp2016-rev-tau-branching-fractions.pdf`.

[41] L. Evans and P. Bryant, *LHC Machine*, JINST **3**, (2008), S08001, `https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08001`.

[42] CERN, *The Large Hadron Collider*, `https://home.cern/science/accelerators/large-hadron-collider`.

[43] ATLAS Collaboration, *The ATLAS Experiment at the CERN Large Hadron Collider*, JINST **3**, (2008), S08003 `https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08003`.

[44] CMS Collaboration, *Detector*, `https://cms.cern/`.

[45] LHCb Collaboration, *The LHCb Detector at the LHC*, JINST **3**, (2008), S08005, `https://iopscience.iop.org/article/10.1088/1748-0221/3/08/S08005/meta`.

[46] ALICE Collaboration, *Performance of the ALICE Experiment at the CERN LHC*, CERN-PH-EP-2014-031, `https://arxiv.org/abs/1402.4476`.

[47] A. Horvath, vector image created with Inkscape, `https://en.wikipedia.org/wiki/Large_Hadron_Collider`.

[48] M. Hostettler and G. Papotti, *Luminosity lifetime at the LHC in 2012 proton physics operation*, 4th International Particle Accelerator Conference, Shanghai, China, 12-17, (2013), 1403, `https://s3.cern.ch/inspire-prod-files-b/bf8be5e2dcdbbc56bafd5bb6e8278fd6`.

[49] S. Fartoukh, S. Kostoglou, M.C. Solfaroli,G. Arduini, H. Bartosik, *et al.*, *LHC Configuration and Operational Scenario for Run 3*, CERN-ACC-2021-0007, `https://cds.cern.ch/record/2790409?`.

[50] ATLAS Collaboration, ATLAS public twiki, `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2`.

[51] ATLAS Collaboration Open Data, *ATLAS detector at the LHC*, `http://opendata.atlas.cern/release/2020/documentation/atlas/experiment.html`.

[52] I. Neutelings, *CMS coordinate system*, TikZ.net, Graphics with TikZ in LaTeX, `https://tikz.net/axis3d_cms/`.

[53] A.R. Martinez on behalf of the ATLAS Collaboration, *The Run-2 ATLAS Trigger System*, J. Phys. Conf. Ser. **762** 012003, (2016), `https://iopscience.iop.org/article/10.1088/1742-6596/762/1/012003`.

[54] The ATLAS Collaboration, *Trigger menu in 2018*, ATL-DAQ-PUB-2019-001, (2019), `https://cds.cern.ch/record/2693402`.

[55] ATLAS Collaboration, *Performance of the ATLAS Trigger System in 2015*, Eur. Phys. J. C 77 (2017) 317, `https://arxiv.org/abs/1611.09661`.

[56] P. Nason, *A new method for combining NLO QCD with shower Monte Carlo algorithms*, JHEP **11**, (2004), 040, `https://iopscience.iop.org/article/10.1088/1126-6708/2004/11/040`.

[57] S. Frixione, P. Nason, and C. Oleari, *Matching NLO QCD computations with parton shower simulations: The POWHEG method*, JHEP **11**, (2007), 070, `https://arxiv.org/abs/0709.2092`.

[58] P. Nason and C. Oleari, *NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG*, JHEP **02**, (2010), 037, `https://link.springer.com/article/10.1007/JHEP02(2010)037`.

[59] S. Alioli, P. Nason, C. Oleari, and E. Re, *A general framework for implementing NLO calculations in shower Monte Carlo programs: The POWHEG BOX*, JHEP **06**, (2010), 043, `https://arxiv.org/abs/1002.2581`.

[60] T. Sjöstrand, S. Askb, J.R. Christiansena, R. Corkea, N. Desaic, P. Iltend, S. Mrennae, S. Prestelf, C.O. Rasmussena, P.Z. Skandsh, *An introduction to PYTHIA 8.2*, Comput. Phys. Commun. **191**, (2015) 159, 2815, CERN-PH-TH-2014-190, `https://arxiv.org/abs/1410.3012`.

[61] ATLAS Collaboration, *Measurement of the Z/γ^* boson transverse momentum distribution in pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, JHEP **09**, (2014), 145, CERN-PH-EP-2014-075, `https://arxiv.org/abs/1406.3660`.

[62] J. Butterworth, S. Carrazza, A. Cooper-Sarkar, A.D. Roeck, J. el Feltesse, S. Forte, J. Gao, S. Glazov, J. Huston, Z. Kassabov, R. McNulty, A. Morsch, P. Nadolsky, V. Radescu, J. Rojo and R. Thorne, *PDF4LHC recommendations for LHC Run II*, J. Phys. G **43**, (2016), 023001, `https://arxiv.org/abs/1510.03865`.

[63] M. Ciccolini, A. Denner, and S. Dittmaier, *Strong and electroweak corrections to the production of a Higgs Boson + 2 Jets via weak interactions at the Large Hadron Collider*, Phys. Rev. Lett. **99**, (2007), 161803, `https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.99.161803`.

[64] M. Ciccolini, A. Denner, and S. Dittmaier, *Electroweak and QCD corrections to Higgs production via vector-boson fusion at the CERN LHC*, Phys. Rev. D **77**, (2008), 013002, `https://arxiv.org/abs/0710.4749`.

[65] P. Bolzoni, F. Maltoni, S.O. Moch and M. Zaro, *Higgs Boson Production via Vector-Boson Fusion at Next-to-Next-to-Leading Order in QCD*, Phys. Rev. Lett. **105**, (2010), 011801, `https://arxiv.org/abs/1003.4451`.

[66] ATLAS Collaboration, *Measurements of Higgs boson production cross-sections in the $H \to \tau_+\tau_-$ decay channel in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector*, JHEP **08**, (2022), 175, CERN-EP-2021-217, 2022, `https://doi.org/10.48550/arXiv.2201.08269`.

[67] S. Agostinelli et al., *Geant4—a simulation toolkit*, Nucl. Instrum. Meth. A **506**, (2003), 250.

[68] K. Hamilton, P. Nason, E. Re and G. Zanderighi, *NNLOPS simulation of Higgs boson production*, JHEP **10**, (2013), 222, `https://arxiv.org/abs/1309.0017`.

[69] K. Hamilton, P. Nason and G. Zanderighi, *Finite quark-mass effects in the NNLOPS POWHEG+MiNLO Higgs generator*, JHEP **05**, (2015), 140, `https://arxiv.org/abs/1501.04637`.

[70] S. Frixione, P. Nason and G. Ridolfi, *A positive-weight next-to-leading-order Monte Carlo for heavy flavour hadroproduction*, 2907 JHEP 09 (2007) 126, `https://iopscience.iop.org/article/10.1088/1126-6708/2007/09/126`.

[71] H. B. Hartanto, B. Jäger, L. Reina and D. Wackeroth, *Higgs boson production in association with top quarks in the POWHEG BOX*, Phys. Rev. D **91**, (2015), 094003, `https://journals.aps.org/prd/abstract/10.1103/PhysRevD.91.094003`.

[72] NNPDF Collaboration, *Parton distributions for the LHC run II*, JHEP **04**, (2015), 040, `https://arxiv.org/abs/1410.8849`.

[73] ATLAS Collaboration, *ATLAS Pythia 8 tunes to 7 TeV data*, ATL-PHYS-PUB-2014-021, (2014), `https://cds.cern.ch/record/1966419`.

[74] D. J. Lange, *The EvtGen particle decay simulation package*, Nucl. Instrum. Meth. A **462**, (2001), 152, `https://www.sciencedirect.com/science/article/abs/pii/S0168900201000894?via%3Dihub`.

[75] H.-L. Lai, M. Guzzi, J. Huston, Z. Li, P.M. Nadolsky, J. Pumplin, and C.-P. Yuan, *New parton distributions for collider physics*, Phys. Rev. D **82**, (2010), 074024, `https://arxiv.org/abs/1007.2241`.

[76] E. Bothmann, G.S. Chahal, S. H oche, J. Krause, F. Krauss, S. Kuttimalai, S. Liebschner, D. Napoletano, M. Shönherr, H. Schulz, S. Schumann and F. Siegert, *Event generation with Sherpa 2.2*, SciPost Phys. **7**, (2019), 034, `https://arxiv.org/abs/1905.09127`.

[77] C. Anastasiou, L.J. Dixon, K. Melnikov and F. Petriello, *High precision QCD at hadron colliders: Electroweak gauge boson rapidity distributions at next-to-next-to leading order*, Phys. Rev. D **69** (2004) 094008, `https://arxiv.org/abs/hep-ph/0312266`.

[78] ATLAS Collaboration, *Electron reconstruction and identification in the ATLAS experiment using the 2015 and 2016 LHC proton-proton collision data at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **79**, (2019), 639, `https://arxiv.org/abs/1902.04655`.

[79] ATLAS Collaboration, *Electron and photon performance measurements with the ATLAS detector using the 2015-2017 LHC proton-proton collision data*, Journal of Instrumentation, Vol. **14**, (2019), `https://doi.org/10.48550/arXiv.1908.00005`.

[80] R. Frühwirth, *A Gaussian-mixture approximation of the Bethe–Heitler model of electron energy loss by bremsstrahlung*, Comp. Phys. Comm. **154**, (2003), 131, `https://www.sciencedirect.com/science/article/abs/pii/S0010465503002923?via%3Dihub`.

[81] ATLAS Collaboration, *Muon reconstruction performance of the ATLAS detector in proton-proton collision data at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **76**, (2016), 292, `https://doi.org/10.48550/arXiv.1603.05598`.

[82] ATLAS Collaboration, *Reconstruction, Energy Calibration, and Identification of Hadronically Decaying Tau Leptons in the ATLAS Experiment*, ATLAS-CONF-2011-077, (2011), `https://cds.cern.ch/record/1353226`.

[83] ATLAS Collaboration, *Jet reconstruction and performance using particle flow with the ATLAS Detector*, Eur. Phys. J. C **77**, (2017), 466, `https://link.springer.com/article/10.1140/epjc/s10052-017-5031-2`.

[84] M. Cacciari, G. P. Salam and G. Soyez, *The anti-kt jet clustering algorithm*, JHEP **04**, (2008), 063, `https://iopscience.iop.org/article/10.1088/1126-6708/2008/04/063`.

[85] ATLAS Collaboration, *Identification of hadronic tau lepton decays using neural networks in the ATLAS experiment*, ATL-PHYS-PUB-2019-033, (2019), `https://cds.cern.ch/record/2688062`.

[86] ATLAS Collaboration, *Measurement of the tau lepton reconstruction and identification performance in the ATLAS experiment using pp collisions at* $\sqrt{s} = 13$ *TeV*, ATLAS-CONF-2017-029, (2017), `https://cds.cern.ch/record/2261772`.

[87] ATLAS Collaboration, *Jet global sequential corrections with the ATLAS detector in proton–proton collisions at* $\sqrt{s} = 8$ *TeV*, ATLAS-CONF-2015-002, (2015), `https://cds.cern.ch/record/2625233`.

[88] ATLAS Collaboration, *Forward Jet Vertex Tagging: A new technique for the identification and rejection of forward pileup jets*, ATL-PHYS-PUB-2015-034, (2015), `https://cds.cern.ch/record/2042098`.

[89] A. Hrynevich on behalf of the ATLAS collaboration, *ATLAS jet and missing energy reconstruction, calibration and performance in LHC Run-2*, JINST **12**, (2017), C06038, `https://cds.cern.ch/record/2263777?`.

[90] The ATLAS Collaboration, *Search for new resonances in mass distributions of jet pairs using 139 fb*^{1} *of pp collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, JHEP **03**, (2020), 145, `https://arxiv.org/pdf/1910.08447.pdf`.

[91] The ATLAS Collaboration, E_T^{miss} *performance in the ATLAS detector using 2015–2016 LHC pp collisions*, ATLAS-CONF-2018-023, (2018), `https://cds.cern.ch/record/2625233`.

[92] The ATLAS Collaboration, *Performance of missing transverse momentum reconstruction with the ATLAS detector using proton–proton collisions at* $\sqrt{s} = 13$ *TeV*, Eur. Phys. J. C **78**, (2018), 903, `https://arxiv.org/pdf/1802.08168.pdf`.

[93] R. K. Ellis, I. Hinchliffe, M. Soldate, and J. J. Van Der Bij, *Higgs decay to* $\tau^{+}\tau^{-}$ *A possible signature of intermediate mass Higgs bosons at high energy hadron colliders*, Nucl. Phys. B **297**, (1988), 221, `https://doi.org/10.1016/0550-3213(88)90019-3`.

[94] A.Elagin, P.Murat, A.Pranko, A.Safonov, *A New Mass Reconstruction Technique for Resonances Decaying to di-tau*, Nucl. Instrum. Methods Phys. Res. A **654**, (2011), 481, `https://arxiv.org/abs/1012.4686`.

[95] G. Klämke and D. Zeppenfeld, *Higgs plus two jet production via gluon fusion as a signal at the CERN LHC*, JHEP **04**, (2007), 052, `https://doi.org/10.48550/arXiv.hep-ph/0703202`.

[96] M. Bahmani, on behalf of the ATLAS Collaboration. *Data-driven estimation of fake* τ *background in Higgs searches in ATLAS*, 7th workshop on prospects for charged Higgs discovery at colliders (PoS CHARGED2018), (2019), 018, ATL-PHYS-PROC-2019-00612, `https://cds.cern.ch/record/2710046`.

[97] ATLAS Collaboration, *Measurement of the* $H \to \tau^{+}\tau^{-}$ *cross-section in 13TeV Collisions with the ATLAS Detector*, ATL-COM-PHYS-2017-446, (2018), `https://cds.cern.ch/record/2261605`, internal.

[98] ATLAS Collaboration. *Identification and energy calibration of hadronically decaying tau leptons with the ATLAS experiment in pp collisions at* $\sqrt{s} = 8\,TeV$, Eur. Phys. J. C **75**, (2015), 303, `https://doi.org/10.48550/arXiv.1412.7086`.

[99] ATLAS Collaboration, *Cross-section measurements of the Higgs boson decaying into a pair of* τ*-leptons in proton-proton collisions at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Phys. Rev. D **99**, (2019), 072001, `https://doi.org/10.48550/arXiv.1811.08856`.

[100] P. Mehta, C. Wang, Al.G.R. Day and C. Richardson, *A high-bias, low-variance introduction to Machine Learning for physicists*, Phyics Reports **810**, (2019), 1-124, `https://doi.org/10.48550/arXiv.1803.08823`

[101] P. Gupta, *Regularization in Machine Learning*, Towards Data Science, (2017), `https://towardsdatascience.com/regularization-in-machine-learning-76441ddcf99a`.

[102] G. James, D. Witten, T. Hastie and R. Tibshirani, *An Introduction to Statistical Learning with applications in R*, Springer, DOI 10.1007/978-1-4614-7138-7.

[103] Ajay Sarangam, *Epoch in Machine Learning: A Simple Introduction*, Jigsaw Academy, (2021), `https://www.jigsawacademy.com/blogs/ai-ml/epoch-in-machine-learning`.

[104] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, *Dropout: a simple way to prevent neural networks from overfitting*, Journal of machine learning research **15**, (2014), 1929–1958, `https://jmlr.org/papers/v15/srivastava14a.html`.

[105] D.P. Kingma and J. Ba, *Adam: A method for stochastic optimization*, (2014), `https://arxiv.org/abs/1412.6980`.

[106] E. van Nieuwenburg, E. Bairey, G. Refael, *Learning phase transitions from dynamics*, Phys. Rev. B **98**, (2018), 060301, `https://arxiv.org/abs/1712.00450`.

[107] T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning: Data mining, Inference and Prediction*, Springer, (2009), ISBN: 978-0-387-84858-7.

[108] S. Haykin, *Neural Networks and Learning Machines*, McMaster University, (2009), ISBN: 0131471392.

[109] B. Roy, *All about feature scaling*, Towards Data Science, (2020), `https://towardsdatascience.com/all-about-feature-scaling-bcc0ad75cb35`.

[110] J. Brownlee, *How to use Data Scaling Improve Deep Learning Model Stability and Performance*, Machine Learning Mastery, (2019), `https://machinelearningmastery.com/how-to-improve-neural-network-stability-and-modeling-performance-with-data-scaling/`.

[111] K. Nyuytiymbiy, *Parameters and Hyperparameters in Machine Learning and Deep Learning*, Towards Data Science, (2020), `https://towardsdatascience.com/parameters-and-hyperparameters-aa609601a9ac`.

[112] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, *Optuna: A Next-generation Hyperparameter Optimization Framework*, Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, (2019), `https://arxiv.org/abs/1907.10902`.

[113] J. Bergstra, R. Bardenet, Y. Bengio and B. Kégl, *Algorithms for hyper-parameter optimization*, Advances in Neural Information Processing Systems **24**, (2011), `https://proceedings.neurips.cc/paper/2011/file/86e8f7ab32cfd12577bc2619bc635690-Paper.pdf`.

[114] W. Koehrsen, *A Conceptual Explanation of Bayesian Hyperparameter Optimization for Machine Learning*, Towards Data Science, (2018), `https://towardsdatascience.com/a-conceptual-explanation-of-bayesian-model-based-hyperparameter-optimization-for-machin`

[115] J. Schmidhuber, *Deep learning in neural networks: An overview*, Neural Networks, Vol. **61**, (2015), 85-117, `https://arxiv.org/abs/1404.7828`.

[116] M. Kastner, N. Nellen and C. Jahn, *Model-Based Optimisation with Tree-structured Parzen Estimation for Discrete Event Simulation at Container Terminals*, TU Hamburg, Hamburg (Germany), (2019), `http://www.asim-fachtagung-spl.de/asim2019/papers/47_Proof_164.pdf`.

[117] M. Shibata, *An Introduction to the Implementation of Optuna, a Hyperparameter Optimization Framework*, Optuna, (2021), `https://medium.com/optuna/an-introduction-to-the-implementation-of-optuna-a-hyperparameter-optimization-framework`

[118] L. Breiman, *Random Forests*, Machine Learning **45**, (2001), 5–32, `https://doi.org/10.1023/A:1010933404324`.

[119] C. Jebarathinam, D. Home, U. Sinha, *Pearson Correlation Coefficient as a measure for Certifying and Quantifying High Dimensional Entanglement*, Phys. Rev. A **101**, (2020), 022112, `https://arxiv.org/abs/1909.01372`.

[120] A. Elagin, P. Murat, A. Pranko, and A. Safonov, *A new mass reconstruction technique for resonances decaying to di-tau*, Nucl. Instrum. Methods Phys. Res.A **654**, (2011), 481, `https://arxiv.org/abs/1012.4686`.

[121] S.L. Adler, *Axial-Vector Vertex in Spinor Electrodynamics*, Phys. Rev. **177**, (1969), 2426, `https://journals.aps.org/pr/abstract/10.1103/PhysRev.177.2426`.

[122] J.S. Bell and R. Jackiw, *A PCAC puzzle: $\pi^0 \to \gamma\gamma$ in the σ-model*, Il Nuovo Cimento A **60**, (1965-1970), 47–61, `https://link.springer.com/article/10.1007/BF02823296`.

[123] G.'t Hooft, *Symmetry Breaking through Bell-Jackiw Anomalies*, Phys. Rev. Lett. **37**, (1976), 8, `https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.37.8`.

[124] V.A.KuzminV, A.Rubakov and M.E.Shaposhnikov, *On anomalous electroweak baryon-number non-conservation in the early universe*, Physics Letters B, Vol. **155**, (1985), 1–2, `https://www.sciencedirect.com/science/article/abs/pii/0370269385910287`.

[125] M. Kobayashi and T. Maskawa, *CP-Violation in Renormalizable Theory of Weak Interaction*, Progress of Theoretical Physics, **49**, (1973), 652-657, `https://doi.org/10.1143/PTP.49.652`.

[126] G. Valencia, *Constructing CP-Odd Observables*, AMES-HET-94-12, (1994), `https://inspirehep.net/literature/380768`.

[127] J. Brehmer, F. Kling, T. Plehn, and T.M.P. Tait, *Better Higgs-CP tests through information geometry*, Phys. Rev. D **97**, (2018), `https://journals.aps.org/prd/pdf/10.1103/PhysRevD.97.095017`.

[128] D. Atwood, S. Bar-Shalom, G. Eilam, A. Soni,*CP Violation in Top Physics*, Phys. Rept. **47**, (2001), 1-222, `https://doi.org/10.48550/arXiv.hep-ph/0006032`.

[129] G. Barr, R. Devenish, R. Walczak and T. Weidberg, *Particle Physics in the LHC Era*, Oxford University Press, ISBN: 9780191811203, (2016).

[130] C. Jarlskog, *Commutator of the Quark Mass Matrices in the Standard Electroweak Model and a Measure of Maximal CP Nonconservation*, Phys. Rev. Lett. **55**, (1985), 1039, `https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.55.1039`.

[131] J. H. Christenson, J. W. Cronin, V. L. Fitch and R. Turlay, *Evidence for the 2π Decay of the K_2^0 Meson*, Phys. Rev. Lett. **13**, (1964), 138, `https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.13.138`.

[132] BaBar Collaboration, *Observation of CP Violation in the B^0 Meson System*, Phys. Rev. Lett. **87**, (2001), 091801, `https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.87.091801`.

[133] Belle Collaboration, *Observation of Large CP Violation in the Neutral B Meson System*, Phys. Rev. Lett. **87**, (2001) 091802, `https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.87.091802`.

[134] A. Falkowski, *Effective field theories in particle physics*, IPhT Courses, `https://courses.ipht.fr/?q=en/node/167`.

[135] A. de Gouvêa, J. Herrero-García and A. Kobach, *Neutrino masses, grand unification, and baryon number violation*, Phys. Rev. D **90**, (2014), 016011, `https://journals.aps.org/prd/abstract/10.1103/PhysRevD.90.016011`.

[136] A. Kobach, *Baryon number, lepton number, and operator dimension in the Standard Model*, Phys. Rev. D **758**, (2016), 455, `https://www.sciencedirect.com/science/article/pii/S0370269316301976?via%3Dihub`.

[137] W. Buchmuller and D. Wyler, *Effective Lagrangian Analysis of new interactions and flavor conservation.* Nucl. Phys. B. **268**, (1986), 621–653, `https://www.sciencedirect.com/science/article/abs/pii/0550321386902622`.

[138] OPAL Collaboration, *Measurement of W boson polarizations and CP violating triple gauge couplings from* W^+W^- *production at LEP*, Eur. Phys. J. C. **19**, (2001), 229–240, `https://arxiv.org/abs/hep-ex/0009021`.

[139] DELPHI Collaboration, *Study of W boson polarisations and Triple Gauge boson Couplings in the reaction* $e^+e^- \to W^+W^-$ *at LEP 2*, Eur. Phys. J. C **54**, (2008), 345–364, `https://arxiv.org/abs/0801.1235`.

[140] ALEPH Collaboration, *Improved measurement of the triple gauge-boson couplings* γWW *and* ZWW *in* e^+e *collisions*, Phys. Lett. B **614**, (2005), 7–26, `https://www.sciencedirect.com/science/article/abs/pii/S0370269305004120?via%3Dihub`.

[141] K. Hagiwara, S. Ishihara, R. Szalapski and D. Zeppenfeld, *Low energy effects of new interactions in the electroweak boson sector*, Phys. Rev. D **48**, (1993), 2182, `https://journals.aps.org/prd/abstract/10.1103/PhysRevD.48.2182`.

[142] K. Hagiwara, T. Hatsukano, S. Ishihara and R. Szalapski, *Probing non-standard bosonic interactions via W-boson pair production at lepton colliders*, Nucl. Phys. B **496**, (1997), 66, `https://www.sciencedirect.com/science/article/abs/pii/S0550321397002083?via%3Dihub`.

[143] L3 Collaboration, *Search for anomalous couplings in the Higgs sector at LEP.* Phys. Lett. B **589**, (2004), 89–102, `https://arxiv.org/abs/hep-ex/0403037`.

[144] D. Atwood and A. Soni, *Analysis for magnetic moment and electric dipole moment form factors of the top quark via* $e^+e^- \to tt^-$, Phys. Rev. D **45**, (1992), 2405, `https://journals.aps.org/prd/abstract/10.1103/PhysRevD.45.2405`.

[145] M. Davier, L. Duflot, F. Le Diberder and A. Rougé, *The optimal method for the measurement of tau polarization*, Phys. Lett. B **306**, (1993), 411, `https://www.sciencedirect.com/science/article/abs/pii/037026939390101M?via%3Dihub`.

[146] M. Diehl and O. Nachtmann, *Optimal observables for the measurement of three gauge boson couplings in* $e^+e^- \to W^+W^-$, Z. Phys. C **62**, (1994), 397, `https://link.springer.com/article/10.1007/BF01555899`.

[147] F.U. Bernlochner, C. Englert, C. Hays, K. Lohwasser, H. Mildner, A. Pilkington, D. D. Price and M. Spannowsky, *Angles on CP-violation in Higgs boson interactions*, Physics Letters B **790**, (2019), `https://arxiv.org/pdf/1808.06577.pdf`.

[148] R.P. Feynman, *The behavior of hadron collisions at extreme energies*, Conf.Proc.C **690905**, (1969), 237-258, `https://inspirehep.net/literature/57407`.

[149] S.D. Drell and T.-M. Yan, *Partons and their applications at high energies*, Annals of Physics, Vol. **66**, 2, (1971), 578-623, `https://www.sciencedirect.com/science/article/pii/0003491671900716?via%3Dihub`.

[150] A.D. Martin, *Proton structure, Partons, QCD, DGLAP and beyond*, Acta Phys. Polon. B **39**, (2018), 2025-2062, `https://arxiv.org/abs/0802.0161`.

[151] ATLAS Collaboration, *Test of CP Invariance in vector-boson fusion production of the Higgs boson using the Optimal Observable method in the ditau decay channel with the ATLAS detector*, Eur. Phys. J. C **76**, (2016), 658, `https://arxiv.org/abs/1602.04516`.

[152] G. Cowan, *Statistical data analysis*, Oxford University Press, (1998), ISBN: 9780198501558.

[153] G. Ganis, L. Moneta, *Fitting and Parameter Estimation in ROOT*, ROOT Training at La Plata, (2013), `https://twiki.cern.ch/twiki/pub/Main/ROOTLaPlataTutorial/ROOT_LaPlata2013_Fitting.pdf`

[154] ATLAS Collaboration, *Study of the CP property of the Higgs boson to electroweak boson coupling in the VBF $H \to \gamma\gamma$ channel with the ATLAS detector*, submitted to Phys. Rev. Lett., CERN-EP-2022-134, (2022), `https://arxiv.org/abs/2208.02338`.

Acknowledgements

I would like to take this opportunity to thank all the people who supported and accompanied me during the course of my Master's degree. In the following I highlight some of them.

First of all I would like to thank Prof. Dr. Markus Schumacher for introducing me to a part of physics that was unknown to me before the master's degree, especially his support during the course of this project and, in particular, for always facilitating the working conditions and taking my personal conditions into account.

I am very grateful for the work, time and enthusiasm that my supervisor, Ö. Oğul Öncel, put into my work. His patience, support and dedicated work have made a significant contribution to this thesis and to my future as a physicist.

Thanks to the research group, without whom I would not have enjoyed and learned so much this year, in addition to the assistance and discussions provided. In particular, thank you Ye Joon Kim for always solving my questions and helping me through my coding-based difficulties.

Esta parte va en español para mis padres, sin los que nada de esto hubiera sido posible, gracias por la confianza y el apoyo, sin ellos no sería quien soy. Mamá, por esto quizás no me dan el Nobel, pero ya llegará. Papá también te menciono que sino te pones celoso. Un abrazo muy fuerte y nos vemos en España.

Quiero agradecer especialmente a Lucas por el apoyo incondicional, el tragarse mis penas y basicamente por exisitir y soportarme. Te quiero mucho.

No pueden faltar en esta sección mis amigos José y Gonzalo sin los que seguiría en la carrera aunque digan lo contrario. Gracias por apoyarme y siempre estar para mí.

Finalmente quería agradecer al resto de mi familia incluyendo a las spooky hijas de Diana, que no voy a mencionar por si alguna se va del grupo, gracias por las risas y el apoyo.