UNIVERSITÀ DEGLI STUDI FIRENZE

Scuola di Scienze Matematiche Fisiche e Naturali
Corso di Laurea Magistrale in Scienze Fisiche e Astrofisiche

Differential measurement of the Higgs boson production in the WW decay channel with model independent neural networks

Misura differenziale della produzione del bosone di Higgs nel canale di decadimento WW con reti neurali indipendenti dal modello

Relatore:
Dott. Lorenzo Viliani

Correlatore:
Prof. Piergiulio Lenzi

Candidato:
Benedetta Camaiani

Anno Accademico: 2020/2021

Contents

1 Phenomenology of the Higgs sector of the Standard Model 3
- 1.1 The Standard Model of particle physics . . . 3
- 1.2 The electroweak interaction . . . 5
  - 1.2.1 The Higgs mechanism . . . 6
- 1.3 The strong interaction . . . 8
- 1.4 Higgs boson phenomenology . . . 9
- 1.5 Beyond the Standard Model . . . 12

2 The LHC and the CMS experiment 17
- 2.1 The Large Hadron Collider . . . 17
  - 2.1.1 Proton-proton collisions . . . 20
- 2.2 The Compact Muon Solenoid . . . 23
- 2.3 Event reconstruction . . . 28

3 Measurement of the Higgs boson properties in the HWW channel 31
- 3.1 The *VBF* mechanism in the $H \to W^+W^-$ decay channel . . . 31
- 3.2 Background processes . . . 32
- 3.3 Data and simulated samples . . . 37
- 3.4 Monte Carlo corrections . . . 38
- 3.5 Systematic uncertainties . . . 39
- 3.6 Differential and fiducial analysis . . . 40
  - 3.6.1 The unfolding problem . . . 43
- 3.7 Model dependence . . . 46

4 Analysis strategy 47
- 4.1 Preselection and categorization . . . 47
- 4.2 Neural Networks for event classification . . . 49
- 4.3 Adversarial Neural Network . . . 52
  - 4.3.1 Training procedure . . . 55
  - 4.3.2 Optimization . . . 57

4.3.3 Performances 59
4.4 Fiducial phase space and generator level binning 63
4.5 Signal regions 64
4.6 Control regions 70
4.7 Signal extraction 72
4.7.1 Nuisance parameters 73
4.7.2 Statistical treatment of background processes 73
4.7.3 Unfolding procedure 74

5 Results **77**
5.1 Expected results 77
5.2 Model dependence studies 79
5.2.1 Bias estimation 80
5.2.2 Results 83

A ANN performance **93**

Introduction

The observation of a scalar resonance compatible with the long sought Higgs boson was announced in 2012 by the CMS and ATLAS collaborations at the CERN Large Hadron Collider. This was a clear experimental confirmation for the Standard Model (SM) of elementary particles, which set off several precision measurements aiming to measure the properties of the newly discovered particle. Presently, the Higgs boson has been observed in all accessible production mechanisms via combination of several decay channels, with no significant deviations from the SM expectations.
However, the Higgs sector could be more complex with respect to what observed so far and Beyond the Standard Model (BSM) effects could arise from accurate measurements of the Higgs boson couplings with fermions and bosons. In fact, several BSM theories predict couplings between the Higgs boson and vector bosons which are different from the SM predictions. Since these *HVV anomalous couplings* have not only different coupling strength but show also alternative tensor structures with respect to the SM case, they can lead to different characteristics of the Higgs boson production processes, which in turn may give rise to measurable changes in the kinematic distributions of the final state particles. One of the most sensitive observables to BSM physics is the signed azimuthal angle difference ($\Delta\Phi_{jj}$) between the two leading jets produced in association with the Higgs boson during a proton-proton collision.

The work presented in this thesis is a measurement of the differential production cross section of the Higgs boson as a function of $\Delta\Phi_{jj}$. The $H \to WW \to 2\ell 2\nu$ decay channel is selected as the final state, due to its relatively high branching ratio and low background contamination, and the Vector Boson Fusion (VBF) production mechanism is considered as the signal process. The analysis is performed on proton-proton collision data recorded by the CMS detector in 2018 at a center of mass energy of 13 TeV. The signal has been extracted through a Maximum Likelihood template fit, where the Monte Carlo simulation of an observable with good signal to background discrimination power is fit to data. In addition, the measurement is carried out whitin a restricted phase space in order to avoid the extrapolation to the full phase space using an underlying theoretical assumption. Since an unfolding procedure is also used, the result is independent on the detector response and can thus directly be compared to any physics model. Obviously, this remains true if no sources of model depencence are present in the analysis procedure, that would otherwise spoil the re-interpretability of the measurement.

The main source of model dependence affecting this analysis lies in the signal extraction procedure since the shape of the distribution of the fit variable depends on the Higgs boson properties themselves such as, for instance, the HVV coupling properties.

In this thesis, I propose the usage of the *Domain Adaptation* (DA) technique to reduce this degree of dependence. DA is a sub-field of Machine Learning that aims to define an algorithm with the ability to perform well also in target domains that are different from the one it was trained on. Considering the BSM theories that predict the aforementioned HVV anomalous couplings as target domains, this technique allows to define a discriminating variable which is agnostic to different signal models.

To achieve this I implemented a new discriminator based on an Adversarial Neural Network approach. The use of this new discriminator as a fit variable ensures that the additional bias in the analysis result introduced by the signal extraction procedure is minimal.

Chapter 1

Phenomenology of the Higgs sector of the Standard Model

1.1 The Standard Model of particle physics

The Standard Model (SM) is a quantum field theory [1], based on the local simmetry group $SU(3)_C \otimes SU(2)_L \otimes U(1)_Y$ and formulated in the 1970s. It describes *elementary particles*, the smallest known building blocks of the universe without discernible structure, and three of the four known fundamental forces: electromagnetism, weak interaction and strong nuclear force. The elementary particles are characterized by two Lorentz invariants, the *mass* m and the *spin* s, and are classified into *fermions* or *bosons*, according to the statistics they obey.

Fermions

Fermions are spin-$\frac{1}{2}$ particles that constitute ordinary matter and according to the spin–statistics theorem, they respect the Pauli exclusion principle. They are subdivided into two families based on the force that defines their interaction: *quarks* and *leptons*. Both classes consist of six particles, organized into three doublets, called *generations*, and ordered according to an increasing mass hierarchy.

- Leptons can interact only via the electromagnetic and weak force and they can either carry one unit of electric charge or be neutral. The charged leptons are the *electron* (e), *muon* (μ), and *tau* (τ). Each of these particles has a negative charge (Q $= -1$) and a distinct mass. The corresponding neutral partners are named *neutrinos* (ν_e, ν_μ, ν_τ)

Figure 1.1: *Standard Model of elementary particles, excluding the Higgs boson.*

and have no electric charge and very small mass.

$$\begin{pmatrix} \nu_e \\ e \end{pmatrix}, \begin{pmatrix} \nu_\mu \\ \mu \end{pmatrix}, \begin{pmatrix} \nu_\tau \\ \tau \end{pmatrix}$$

- Quarks can interact via electromagnetic, weak and strong interactions. The six existing types or *flavours* differ from one another in their mass and charge characteristics: *up* (u), *charm* (c) and *top* (t) have Q = + 2/3, while *down* (d), *strange* (s) and *bottom* (b) have Q = − 1/3. In addition, quarks carry a *colour* charge but they always occur in combination with other ones to form colourless bound states called *hadrons*.

$$\begin{pmatrix} u \\ d \end{pmatrix}, \begin{pmatrix} c \\ s \end{pmatrix}, \begin{pmatrix} t \\ b \end{pmatrix}$$

Each one of these particles has an associated *antiparticle*, having the same mass but opposite electric charge and magnetic moment.

Gauge Bosons

Bosons are integer spin particles, defined as force carriers that mediate fundamental forces. In fact, within the SM, the interaction between fermions occurs through the exchange of spin-1 bosons: the *photon* (γ) is the massless mediator of the electromagnetic force between electrically charged particles and it is well-described by the theory of *Quantum Electrodynamics* (QED); the massive $W^{\pm}$ and Z bosons mediate the weak interactions, while 8 massless *gluons* (g) are responsible for the strong force between coloured quarks.

As far as the theory group structure is concerned, $SU(3)_C$ is assumed to be an exact symmetry of the model and it describes strong interaction phenomena. On the contrary, the $SU(2)_L \otimes U(1)$ simmetry describes the electromagnetic and weak part of the theory and it needs to be broken. If not, all particles would be massless and this would be in deep constrast with the experimental evidences. In fact, it is shown that all particles in the SM, except for the photon and the gluon, carry a mass as illustrated in Figure 1.1.
This problem is solved by introducing an additional scalar (or spin-0) field in the theory known as *Higgs boson*, as detailed in the following sections.

1.2 The electroweak interaction

In the 1970s, experiments showed that the weak and electromagnetic interactions were actually the manifestations of the same fundamental interaction, the electroweak force. The unified decription, which embeds QED, had already been proposed in the late 1960s by S. L. Glashow, A. Salam and S. Weinberg [2], [3], [4] as an $SU(2)_L \otimes U(1)_Y$ local gauge theory.
By Noether's Theorem there are four conserved charges: three weak isospin charges T_i related to the $SU(2)_L$ group and one weak hypercharge Y associated to the $U(1)_Y$. It can be shown that the electric charge Q, which is the generator of the QED gauge group $U(1)_{em}$, is given by the following equation:

$$Q = T_3 + \frac{Y}{2} \tag{1.1}$$

where T_3 is the component of the weak isospin along the quantization axis.
The particle vector state ψ can be decomposed into *chirality components*: ψ_L and ψ_R with *left* and *right* chirality respectively. This decomposition is crucial because the ψ_L and ψ_R are treated differently by the gauge interactions. In particular, under weak isospin $SU(2)$ transformations the left (right)-handed particles (antiparticles) are $T = 1/2$ weak-isospin doublets, whereas the right(left)-handed are singlets, i.e. the weak isospin of ψ_R is

Table 1.1: *First family lepton and quark quantum numbers. The fermions of the second and third families have the same quantum numbers as the corresponding fermions of the first one.*

Lepton	T	T_3	Y	Q
ν_L	$\frac{1}{2}$	$\frac{1}{2}$	-1	0
e_L	$\frac{1}{2}$	-$\frac{1}{2}$	-1	-1
e_R	0	0	-2	-1

Quark	T	T_3	Y	Q
u_L	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{2}{3}$
d_L	$\frac{1}{2}$	-$\frac{1}{2}$	$\frac{1}{3}$	-$\frac{1}{3}$
u_R	0	0	$\frac{4}{3}$	$\frac{2}{3}$
d_R	0	0	-$\frac{1}{3}$	-$\frac{2}{3}$

zero. Furthermore, $U(1)_Y$ acts differently on ψ_L and ψ_R because they have different weak hypercharges. These distinctions is due to the parity symmetry violation in the electroweak interaction. Fermions quantum numbers are summarized in Table 1.1.

As mentioned previously, this description allows neither fermions nor bosons to have mass. For fermions, the problem arises from the impossibility to construct a Dirac mass term using only ψ_L. This compels to couple the left- and right-handed part of the particle vector state, thus breaking $SU(2)_L$ and $U(1)_Y$ simmetries. Similarly, the gauge symmetry of non-abelian gauge theories prohibits massive gauge bosons. If a mass term is artificially added by hand, the gauge symmetry is violated and the theory becomes non-renormalizable due to the mass-induced ultraviolet behavior for the gauge particle propagator. Therefore, none of the fermions or bosons of the Standard Model can be introduced with mass into the theory, but must acquire it by some other mechanism, as explained below.

1.2.1 The Higgs mechanism

Based on the Standard Model description that has been presented so far, all particles need to have zero mass and the addition of mass terms to the Lagrangian density spoils the gauge invariance and the renormalizability of the theory.

The solution to this inconsistency between theory and experiments has been devised in 1964 by Peter Higgs, François Englert and Robert Brout and it is nowdays known as *the Higgs mechanism*. It is based on the *spontaneous symmetry breaking* (SSB) mechanism, whose main idea is the following [5]. Consider a system whose Lagrangian has a particular symmetry, i.e. is invariant under the corresponding symmetry transformations. If its lowest energy level is non-degenerate, the ground state of the system is unique and possesses the same symmetry properties of the Lagrangian. On the contrary, in

case of degeneracy, there are more eigenstates to represent the ground state and, if one of the degenerate states is arbitrarily chosen, then the ground state no longer shares the symmetries of the Lagrangian. Once such a choice is performed, the original symmetry group is said to be spontaneously broken.
In the SM case, therefore, the gauge symmetry is intrinsically maintened by the Lagrangian density of the theory but it is not revealed in its vacuum state, which is degenerate under $SU(2)_L \otimes U(1)_Y$ transformations.
The Higgs mechanism postulates the exitence of an additional complex scalar field which is an isospin doublet:

$$\Phi = \begin{pmatrix} \phi^+ \\ \phi_0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ \phi_3 + i\phi_4 \end{pmatrix} \tag{1.2}$$

where $\phi_1, \phi_2, \phi_3, \phi_4$ are real fields. The most general interaction term for this field, that preserves the renormalizability and gauge invariance of the theory, has the form:

$$V_H(\Phi) = \mu^2 \Phi^\dagger \Phi + \lambda (\Phi^\dagger \Phi)^2 \tag{1.3}$$

The shape of the potential $V_H(\Phi)$ depends on the λ and μ parameters. The first one is taken greater than zero ($\lambda > 0$) in order to have vacuum stability, while different situations are possible when choosing the value of μ parameter: if it is chosen so that $\mu^2 < 0$, infinite values are possible for the minimum of the potential so the simmetry of $V_H(\Phi)$ may be broken. More precisely, the degenerate states are arranged in a 4-dimensional hypersphere defined by the condition

$$\Phi^\dagger \Phi = -\frac{\mu^2}{2\lambda} = \frac{v^2}{2} \tag{1.4}$$

where the v factor correspond to the *vacuum expectation value* (VEV). Since the exact electromagnetic symmetry has to be preserved in order to have electric charge conservation, the original $SU(2)_L \otimes U(1)_Y$ symmetry group must be broken as

$$SU(2)_L \otimes U(1)_Y \longrightarrow U(1)_{em}$$

i.e. after SSB the $U(1)_{em}$ must remain as a symmetry of the vacuum. This requirement guarantees the existance of the neutral massless photon. For this reason the ground state, around which to expand Φ, has to be electrically neutral and can be chosen as

$$\hat{\Phi} = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \tag{1.5}$$

The Φ field is now redifined as a fluctuation Φ' around $\hat{\Phi}$:

$$\Phi = \hat{\Phi} + \Phi' = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} + \begin{pmatrix} \phi^+ \\ \frac{h+i\eta}{\sqrt{2}} \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \phi_1 + i\phi_2 \\ v + h + i\eta \end{pmatrix} \tag{1.6}$$

In this construction, ϕ_1, ϕ_2 and η are massless scalar fields called *Goldstone bosons*, while h is the massive field associated to the Higgs bosons. It can be shown that, with a particular gauge transformation, Φ can be written as

$$\Phi = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v + h \end{pmatrix} \tag{1.7}$$

Introducing this field in the kinetic term of the scalar field, the bosons acquire mass given by:

$$m_W = \frac{v}{2}g, \qquad m_Z = \frac{v}{2}\sqrt{g^2 + g'^2}, \qquad m_H = v\sqrt{2\lambda}$$

with g, g', v and λ free parameters of the theory.
The origin of the boson masses are now explained but unfortunately the Higgs mechanism does not provide mass to the fermions. This problem can be solved introducing in the lagrangian density a coupling term, known as *Yukawa coupling*, between the fermion doublets and the Higgs field. After symmetry breaking, it results in a gauge invariant mass term for the fermionic field and in an interaction term with the Higgs boson, which is proportional to the fermion mass itself.
The prediction of the electroweak theory have been experimentally confirmed over the years: the $W^{\pm}$ and the Z gauge bosons were discovered in the 1983 by the UA1 and UA2 experiments while in 2012 the discovery of a new particle consistent with the predicted Higgs boson was announced by the CMS and ATLAS collaborations.

1.3 The strong interaction

The theory describing the strong interaction is an unbroken gauge non-abelian theory based on the colour group $SU(3)_C$, known as *Quantum Chromodynamics* (QCD). The fact that the theory is unbroken ensures the conservation of a physics quantity called *colour*. The colour charge of a quark has three different values, say red, blu and green, while antiquarks carry anticolour, whose possible values are anti-red, anti-blu and anti-green. As mentioned above, eight massless gluons mediate the interaction, each of these carring a colour and an anticolour. Although nine gluons colored combinations are expected, one is a colour singlet and, since it carries no net charge,

has to be excluded. A peculiar feature that distinguishes gluons from photons, is their strong self-interaction, due to their coloured nature. This is a direct effect of the non-abelian nature of the theory, which makes the gauge fields carring the charge of the symmetry group.
The QCD coupling constant α_s, which gives a measurement of the force intensity, is a function of the scale of the interaction Q. This leads to two other important consequences:

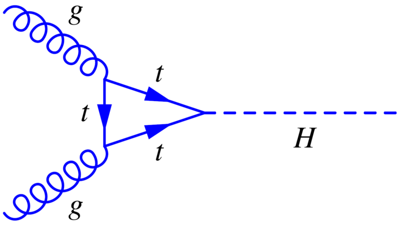
- *asymptotic freedom*: the coupling constant decreases at large scales Q and quarks can be considered as quasi-free interacting particles. This is the reason why a perturbative approach to QCD (pQCD) can be used to describe hard (i.e. high energy) process.

- *colour confinement*: is the phenomenon that colour charged particles can not be isolated, and therefore can not be directly observed. At low energy scale, the value of α_s is so large that the perturbative approach can no longer be applied. In this regime the intensity of the force increases with the distance so, as two colour charges are separated, at some point becomes energetically advantageous for a new quark-antiquark pair to appear from the vacuum than increasing further the interaction strength. This is the reason why free quarks have never been observed.

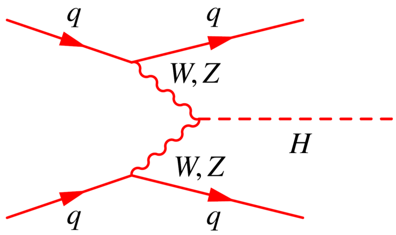
In order to obey colour confinement, quarks and gluons produced in strong interaction processes create other colored particles to form hadrons clustered in *jets*, i.e. collimated groups of colorless objects. This process is known as *hadronization*.

1.4 Higgs boson phenomenology

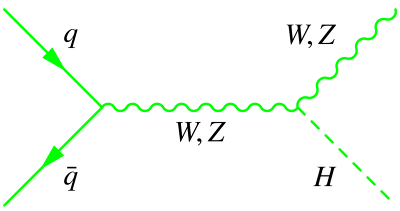
The Higgs boson H is a scalar and electrically neutral field whose mass is a free parameter of the electroweak theory and therefore needs to be measured. It has been discovered in 2012 by the ATLAS and CMS experiments at the CERN Large Hadron Collider (LHC). Its mass has been measured with increasing precision over the years and currently is reported to be of 125.25 ± 0.17 GeV [6]. Since the Higgs boson can interact with any other massive field, it is very unstable and decays into other particles almost immediately. Both production mechanisms and decay modes of this particle are discussed below.



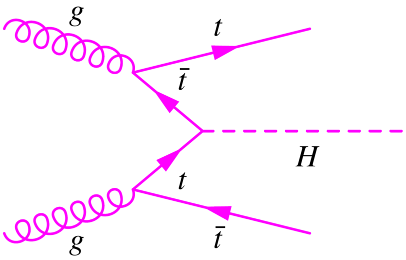
(a) *Gluon fusion*



(b) *Vector Boson fusion*



(c) *Higgsstrahlung*



(d) *Top quark associated production*

Figure 1.2: *Feynman diagrams for the main Higgs boson production modes.*

Production mechanisms

The main production processes of the Higgs boson at hadron colliders are represented by the Feynman diagrams in Figure 1.2. In order of decreasing cross section, the main Higgs boson production modes are:

- **Gluon fusion** (ggH): two incoming gluons interact, giving rise to a Higgs boson via a quark loop, whose main contribution comes from heavy quarks, like the top quark (Figure 1.2.a). This is due to the fact that the Higgs boson coupling strength increases linearly with the fermion masses. The ggH process is the main Higgs boson production mode, as shown in Figure 1.3 with a cross section of 48.5 pb at a center of mass energy equal to 13 TeV [7];

- **Vector Boson fusion** (VBF): two interacting quarks radiate two vector bosons (ZZ or W^+W^-), which then combine to generate a Higgs boson, as shown in Figure 1.2.b. The VBF process has the second largest cross section among Higgs boson production mechanisms, which is equal to 3.78 pb at a center of mass energy of 13 TeV;

- **Higgsstrahlung** (VH): this process is also known as *vector boson associated production* because a vector boson, produced by two incoming

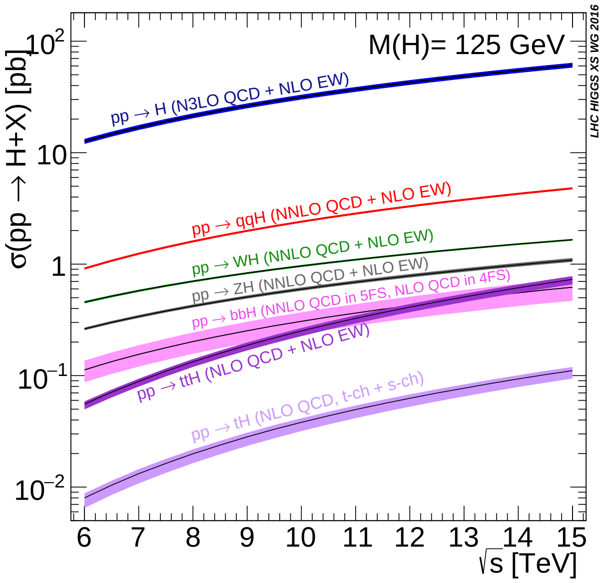


Figure 1.3: *Cross sections for the main Higgs boson production mechanism in proton-proton collisions, as a function of center of mass energy.*

quaks, radiates a Higgs boson (Figure 1.2.c). At a center of mass energy equal to 13 TeV, the W- and Z-associated production mechanisms have a cross sections of 1.37 and 0.88 pb, respectively;

- **Top quark associated production** ($t\bar{t}H$): a $t\bar{t}$ quarks pair, origininated from the splitting of two gluons, combine into a Higgs boson, as illustrated in Figure 1.2.d. This process has a cross section of 0.51 pb at a center of mass energy of 13 TeV.

Decay channels

The possible decay modes of the Higgs boson are essentially determined by the value of its mass. In Figure 1.4 the Higgs boson branching ratios for m_H = 125 GeV are presented. The most probable decay channel of the Higgs boson is $H \to b\bar{b}$. Although being the channel with the largest branching ratio, its identification is very complicated because of the very large background from QCD processes. A similar problem affects the $H \to \gamma\gamma$ channel, since jets can be misidentified as photons. On the other hand, the $H \to ZZ$ decay mode, with the subsequent decay of the Z bosons into four leptons, has a very clear signature that would allow a unequivocal detection. Unfortunately, it has a relatively low branching ratio. Other leptonic Higgs

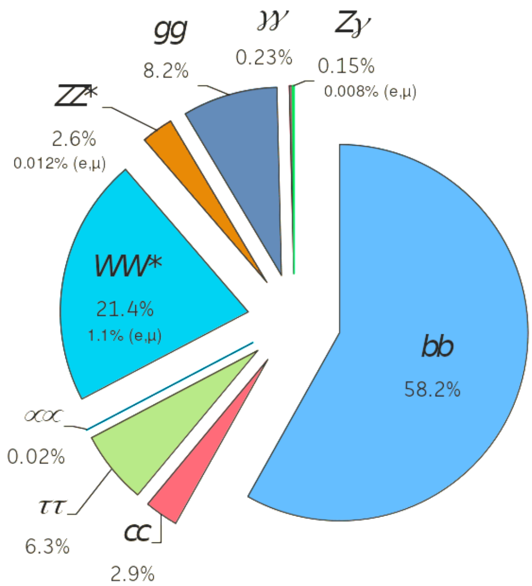


Figure 1.4: *Branching ratios for the Higgs boson decay modes.*

boson decays, like $H \rightarrow \mu^+\mu^-$ or $H \rightarrow e^+e^-$, have even smaller probabilities to occur.
The channel taken into consideration in the present work is $H \rightarrow W^+W^-$, which is the second most probable. As will be explained in the next chapters, it has been selected because of its clear signature, relatively low background contamination and its relatively high branching ratio, which gives it sensitivity to most Higgs boson production modes.

1.5 Beyond the Standard Model

Following the Higgs boson discovery, its properties and interactions with other SM particles have been widely studied. Indeed, the Higgs boson has been observed in all accessible production modes (ggH, VBF, VH and $t\bar{t}H$) and all measurements are consistent with the SM prediction within the uncertainties. The Higgs boson decay channels into gauge bosons (ZZ, WW, $\gamma\gamma$), the third generation quarks (t, b) and τ leptons have been observed as well and no significant deviations from SM expectations have been reported so far.
Nevertheless, there are some shortcomings and problems with the Standard

Model. For instance, it does not account for gravity, the fourth fondamental force, and does not provide an explanation for the hierarchy problem [8] or the nature of dark matter.

In order to provide an effective description of such long-standing and unsolved questions, several theoretical models have been proposed, generically referred to as *Beyond the Standard Model* (BSM) theories. For example, the quantum numbers of the SM Higgs boson are expected to be $J^{CP} = 0^+$ but many BSM models introduce the so-called *Anomalous Couplings* (AC) between vector bosons and the Higgs particle which do not necessarily conserve C-parity or CP-parity.

The most general scattering amplitude describing the interaction between a spin-0 H boson and two spin-1 gauge bosons V_1V_2, such as WW, ZZ, $Z\gamma$, $\gamma\gamma$ or gg, has the form:

$$\begin{aligned}\mathcal{A}(HVV) \sim \left[a_1^{VV} + \frac{k_1^{VV} q_{V_1}^2 + k_2^{VV} q_{V_2}^2}{(\Lambda_1^{VV})^2}\right] m_{V_1}^2 \epsilon_{V_1}^* \epsilon_{V_2}^* \; + \; a_2^{VV} f_{\mu\nu}^{*(1)} f^{*(2),\mu\nu} + \\ + \; a_3^{VV} f_{\mu\nu}^{*(1)} \tilde{f}^{*(2),\mu\nu}\end{aligned} \tag{1.8}$$

where $f^{(i)\mu\nu} = \epsilon_{V_i}^{\mu} q_{V_i}^{\nu} - \epsilon_{V_i}^{\nu} q_{V_i}^{\mu}$ and $\tilde{f}_{\mu\nu}^{(i)} = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma} f^{(i),\rho\sigma}$ are the field strength tensor and the dual field strength tensor of a gauge boson with momentum q_{V_i}, polarization ϵ_{V_i} and pole mass m_{V_i} [9]. Λ_1^{VV} is the energy scale of the BSM physics and is a free parameter of the model. For the sake of clarity, let L_1 be

$$L_1 \stackrel{def}{=} \frac{k_1^{VV} q_{V_1}^2 + k_2^{VV} q_{V_2}^2}{(\Lambda_1^{VV})^2} \tag{1.9}$$

Equation (1.8) takes four couplings into account, as explained below.

The leading order (LO) SM-like contribution corresponds to $a_1^{VV} = 1$ with $VV = ZZ, WW$, and $L_1, a_2, a_3 = 0$. As already mentioned, there can not be LO couplings to massless gauge bosons, so there are no contributions from photons and gluons at this order. HZZ and HWW are CP-even interactions, whose vertex is represented in Figure 1.5.a.

The rest of VV couplings are ascribed to AC, which can be either small SM corrections due to loop effects or new BSM phenomena. In particular, $a_2^{VV} = 1$ describes loop-induced CP-even couplings (like $HZ\gamma$, $H\gamma\gamma$ and Hgg), which are parametrically suppressed by the coupling constants α and α_s. Possible Feynman diagrams for these processes are shown in Figure 1.5.b and 1.5.c. These phenomena are described by the model which will be referred to as the H0PH physics model in the following.

On the other hand, the L_1 term is associated to CP-even $HVf\bar{f}$ and $Hf\bar{f}f\bar{f}$

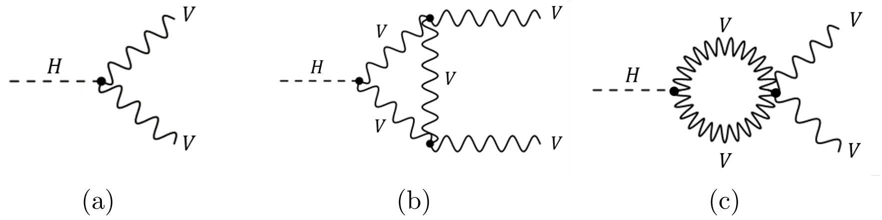


Figure 1.5: *Vertices for the HVV interaction.*

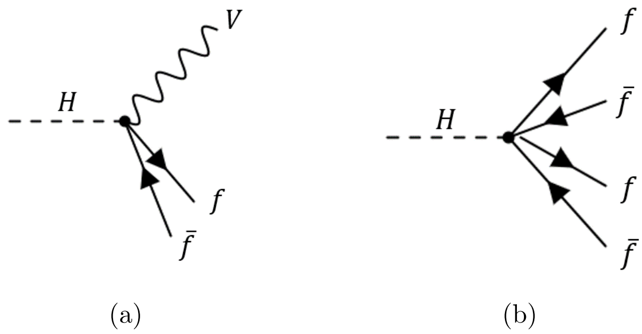


Figure 1.6: *Vertices for the $HVf\bar{f}$ and $Hf\bar{f}f\bar{f}$ interactions.*

interactions (Figure 1.6.a and 1.6.b respectively). These couplings are predicted by the physics hypothesis which will be denominated H0L1 model in this work.

Finally, $a_3^{VV} = 1$ accounts for three loop-induced CP-odd coupling in the SM. This kind of processes is identified by the so-called H0M model.

The a_1^{VV} coupling is assumed to be zero by the H0PH, H0L1 and H0M models.

In addition, three physics theories are defined as mixtures between the SM and one of the previous BSM hypothesis. Such models correspond to have a_1^{VV} and one of the BSM operators both equal to 0.5 and will be labeled by adding the “f05” tag to the model names introduced above. A brief summary of the notations adopted in this work is reported in Table 1.2.

The CP-odd couplings together with any of the other CP-even couplings can lead to CP violation in a given process and hence any deviation from the SM prediction may be outlined.

Table 1.2: *Anomalous coupling values of the HVV scattering amplitude and associated physics models.*

Anomalous coupling	Physics model
$a_1^{VV} = 1$	SM
$a_2^{VV} = 1$, $a_1^{VV} = 0$	H0PH
$L_1 = 1$, $a_1^{VV} = 0$	H0L1
$a_3^{VV} = 1$, $a_1^{VV} = 0$	H0M
$a_1^{VV} = 0.5$, $a_2^{VV} = 0.5$	H0PHf05
$a_1^{VV} = 0.5$, $L_1 = 0.5$	H0L1f05
$a_1^{VV} = 0.5$, $a_3^{VV} = 0.5$	H0Mf05

Chapter 2

The LHC and the CMS experiment

2.1 The Large Hadron Collider

The Large Hadron Collider (LHC) is the largest particle accelerator in the world [10]. It is located at CERN, the European Organization for Nuclear Research, on the Franco-Swiss border near Geneva. It is installed in the underground tunnel at an average depth of 100 m, which hosted the Large Electron Positron Collider (LEP) until 2001. The LHC has a circular shape whith a length of 27 km in circumference and it is designed to accelerate and collide two beams of protons up to an energy of 7 TeV, resulting in a center of mass energy of 14 TeV.
The LHC is the last element of the CERN accelerator complex, in which the proton beams reach their highest energies. A schematic representation of the accelerator chain is shown in Figure 2.1: the protons are produced through the ionization of Hydrogen and are injected into the linear accelerator LINAC2, where they are accelerated up to an energy of 50 MeV. The beam is sent to the Proton Synchrotron Booster (PSB) and subsequently to the Proton Synchrotron (PS), which increase the protons energy up to 1.4 GeV and 25 GeV, respectively. The combined action of these two synchrotrons groups the protons in bunches interspersed by 25 ns, which are then accelerated up to 450 GeV by the Super Proton Synchrotron (SPS). Finally, the protons are extracted and injected into the LHC ring through two different trasmission lines, where they flow in opposite directions and are further accelerated by radio frequency cavities.
The beams are made to collide in four interaction points where the four main experiments are placed:

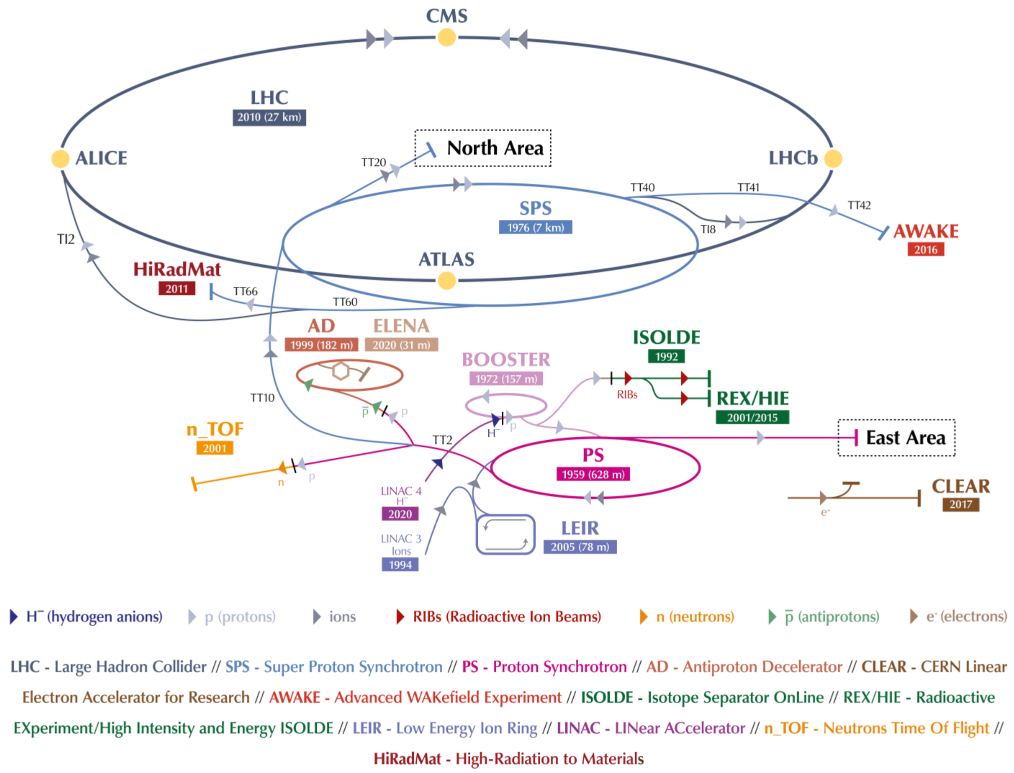


Figure 2.1: *Schematic representation of the accelerator complex installed at CERN.*

- **CMS** (Compact Muon Solenoid) [11] and **ATLAS** (A Toroidal LHC ApparatuS) [12] are two general-purpose experiments, designed to investigate a wide spectrum of physics. The main objectives are the observation and study of the Higgs boson and searches for any possible evidence of new physics beyond the SM. The two independent detectors pursuing the same goals are a fundamental tool for a cross-check of any possible new discovery;

- **LHCb** (LHC beauty) [13] is designed to study CP violation in electroweak processes and the matter-antimatter asymmetry through the analysis of rare decays of hadrons containing b quarks;

- **ALICE** (A Large Ion Collider Experiment) [14] focuses on heavy ions collisions, with the aim of studying a new state of matter, known as quark-gluon plasma.

In order to keep protons on the circular trajectory of the accelerator ring plane, a set of about 120 magnetic dipoles, having magnetic field oriented orthogonally with respect to the ring, is employed. For a particle with unitary charge and momentum p, it can be shown that the radius of its curvature in a magnetic field B is given by:

$$\rho\,[\mathrm{m}] = \frac{p\,[\mathrm{GeV}]}{0.3 \cdot B\,[\mathrm{T}]} \tag{2.1}$$

Therefore, to keep beams within the LHC ring, a magnetic field of about 5.5 T is required. However, since the LHC is built by alternating curved and linear sectors, the magnet system have to produce a 8 T magnetic field along the beam pipe. This is possible only by using Niobium-Titanium dipole magnets, which need to be cooled down to ~ 2 K to reach the superconducting phase and be able to sustain the strong currents needed to generate these magnetic fields. Moreover, over 8000 higher order magnets, such as quadrupoles and octupoles, are employed as focusing systems.

One of the most important parameter of a particle accelerator is the *instantaneous luminosity* $\mathcal{L}$, which provides a measure of the expected rate of events given the process cross section σ:

$$R = \sigma\mathcal{L} \tag{2.2}$$

The LHC design luminosity is $\mathcal{L} = 10^{34}$ cm^{-2}s^{-1}, which corresponds to around 10^9 proton interactions per second, since the total inelastic cross section is about 100 mb. The instantaneous luminosity depends on the construction characteristics of the accelerator according to the following formula:

$$\mathcal{L} = f\,\frac{n_1 n_2}{4\pi\sigma_x\sigma_y} \tag{2.3}$$

where n_1 and n_2 are the numbers of particles contained in the two bunches colliding at a frequency f, while σ_x and σ_y are the beam sizes in the transverse plane at the interaction point. The temporal displacement between two successive bunch crossing of 25 ns corresponds to a collision frequency of $f = 40$ MHz and the transverse size of the beams can be reduced down to 15 μm.
Integrating the instantaneous luminosity over the period of data taking gives the *integrated luminosity*:

$$\mathcal{L}_{int} = \int \mathcal{L}(t)dt \tag{2.4}$$

which is directly related to the total number of events of a given process. The Run 2 (2016-2018) data taking period corresponds to a integrated luminosity

of 138 fb^{-1}.
Since the protons are grouped in bunches, multiple proton-proton (pp) interactions occur when the beams are made to collide and more than one interaction vertex per bunch crossing is expected. This additional collisions are known as *pile up*.

2.1.1 Proton-proton collisions

The LHC is currently able to reach energies of 6.5 TeV per beam, corresponding to a center of mass energy of 13 TeV. At this energy scale the internal structure of the protons is completely resolved and the collision can be described as occurring between its constituents.
Protons are hadrons made up by three valence quarks (uud) which are surrounded by virtual quark-antiquark pairs, known as *sea* quarks, continuously created and annihilated through the absorption and the emission of gluons. All these particles are collectively called *partons* and carry a fraction x of the proton four-momentum, also known as Bjorken's scaling variable:

$$p^{\mu}_{parton} = xp^{\mu}_{proton}. \tag{2.5}$$

Therefore, the center of mass energy available for the creation of new particles in hard scattering processes is:

$$\sqrt{s_{parton}} = \sqrt{x_1 x_2 s} \tag{2.6}$$

where $\sqrt{s}$ is the center of mass energy of the incoming protons, and x_1, x_2 are the four-momentum fractions carried by the two interacting partons. Since generally $x_1 \neq x_2$, the center of mass frame of the two partons is boosted along the beam axis.
The probability density function $f_i(x)$ associated to the Bjorken's scaling variable x is known as *parton distribution function* (PDF) and it is normalized by requiring that :

$$\sum_i \int_0^1 x f_i(x) dx = 1 \tag{2.7}$$

where the index i runs over all the partons making up the proton. PDFs depend on the energy scale μ_F at which they are measured, which is called factorization scale, and their evolution with scale is governed by the DGLAP equation [15]. Figure 2.2 shows examples of parton distribution functions at two different factorization scales.
When two protons collide, several partons can interact. Tipically, interactions with high transferred momentum are the relevant ones for high energy

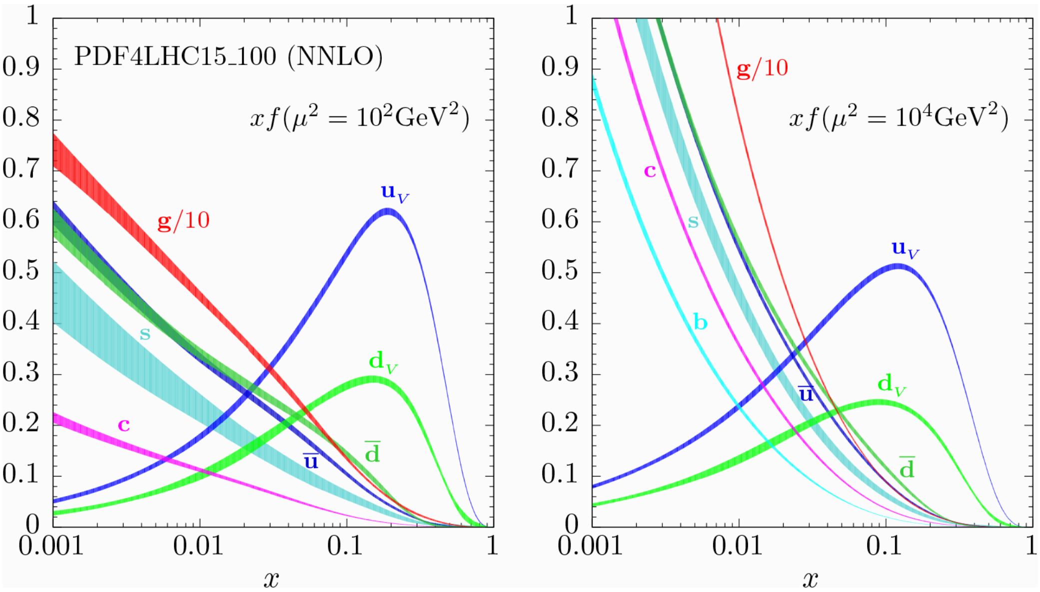


Figure 2.2: *Parton distribution functions for factorization scale* $\mu_F^2 = 10^2$ *GeV* ^{2} *(left) and* $\mu_F^2 = 10^4$ *GeV* ^{2} *(right).*

physics experiments, since they allow for the production of a variety of particles with high transverse momentum. These processes are generally called *hard scattering*. The remnants of both the colliding protons give rise to the *underlying event* (UE), i.e. softer scattering processes.
According to the QCD factorization theorem, the hadronic interaction can be factorized into partonic hard scattering convoluted with the parton distribution functions. Consider a process $p + p \to F + X$, where F is the final state of interest and X refers to any other product of the scattering interaction. The total cross section can be written as:

$$\sigma_{pp\to F+X} = \sum_{a,b} \int dx_1 dx_2 f_{p_1}(x_1, \mu_F^2) f_{p_2}(x_2, \mu_F^2) \hat{\sigma}_{p_1 p_2 \to F}(\mu_F^2, \mu_R^2) \tag{2.8}$$

In this expression the sum runs over all the initial state partons a and b that generate F. The μ_R parameter, known as renomarlization scale, is introduced in pQCD to treat ultraviolet divergences while $f_{p_1}(x_1, \mu_F^2)$ and $f_{p_2}(x_2, \mu_F^2)$ are the probability densities for parton a and b to be found respectively within the protons p_1 and p_2. Finally, $\hat{\sigma}_{p_1 p_2 \to F}(\mu_F^2, \mu_R^2)$ is the partonic hard scattering cross section. At LHC, the total pp cross section is ~ 100 mb.

As mentioned before, at the LHC the center of mass frame of the parton hard scattering is generally boosted along the beam direction. It is therefore useful to describe the final state in terms of variables that are invariant under Lorentz transformations along that direction. By choosing the beam direction as the z axis of the coordinate system, a convenient set of kinematic variables is composed of the *transverse momentum* p_T and the rapidity interval between two objects Δy. The former is defined as the projection of the total momentum on the plane orthogonal to z:

$$p_T = \sqrt{p_x^2 + p_y^2} \tag{2.9}$$

where p_x and p_y are the cartesian components of the particle spatial momentum $\vec{p}$.
The *rapidity* of a particle with energy E is defined as:

$$y = \frac{1}{2} \ln \left(\frac{E + p_z}{E - p_z} \right) \tag{2.10}$$

where p_z is the projection of particle momentum along the beam direction. The rapidity is not invariant under boosts along the beam direction. In fact, given the boost velocity β, it transforms according to the law:

$$y \longrightarrow y + \frac{1}{2} \ln \left(\frac{1 + \beta}{1 - \beta} \right) \ . \tag{2.11}$$

Therefore, rapidity differences Δy are Lorentz invariant.
Experimentally, it is more convenient to use the *pseudorapidity* η, whose definition is:

$$\eta = - \ln \tan \frac{\theta}{2} \tag{2.12}$$

where θ is the polar angle between the particle momentum and the z axis. For ultra-relativistic particles the pseudorapidity coincides with the rapidity. In terms of these variables, the four-momentum of a particle of mass m can be written as:

$$p^\mu = (E, p_x, p_y, p_z) = (m_T \cosh y, p_T \sin \phi, p_T \cos \phi, m_T \sinh y) \tag{2.13}$$

where ϕ is the azimuthal angle on the x-y plane and $E_T = \sqrt{m^2 + p_T^2}$.

Another important variable is the *missing transverse energy* E_T^{miss}, defined as:

$$E_T^{miss} = |\vec{E}_T^{miss}| = | - \sum_i \vec{p}_{T,i}| \tag{2.14}$$

where the i index runs over all the reconstructed particles of the final state. Since the transverse momentum of the initial state is null, E_T^{miss} is expected to vanish if all products of a collision were detected. However, neutrinos or exotic particles eventually produced are too weakly interactive to be directly detected, resulting in a non vanishing missing transverse energy value.

2.2 The Compact Muon Solenoid

The Compact Muon Solenoid (CMS) apparatus [11] is a general purpose detector installed in one of the four LHC interaction point. It has a cylindrical structure coaxial to the beam pipe (*barrel* region), which measures approximately 22 m in length and 15 m in diameter and it is closed at both ends with detecting disks (*endcap* region), ensuring the hermetic closure of the apparatus.
The coordinate system adopted by the CMS collaboration is shown in Figure 2.3: it is a right-handed Cartesian system with the origin fixed in the nominal collision point inside the experiment; the x-axis is oriented radially toward the center of LHC, the y-axis points vertically upward and the z-axis coincides with the beam direction. The azimuthal angle ϕ is measured starting from the x-axis in the x-y plane in which the radial coordinate is denoted by r, while the polar angle θ is evaluated from the z-axis.

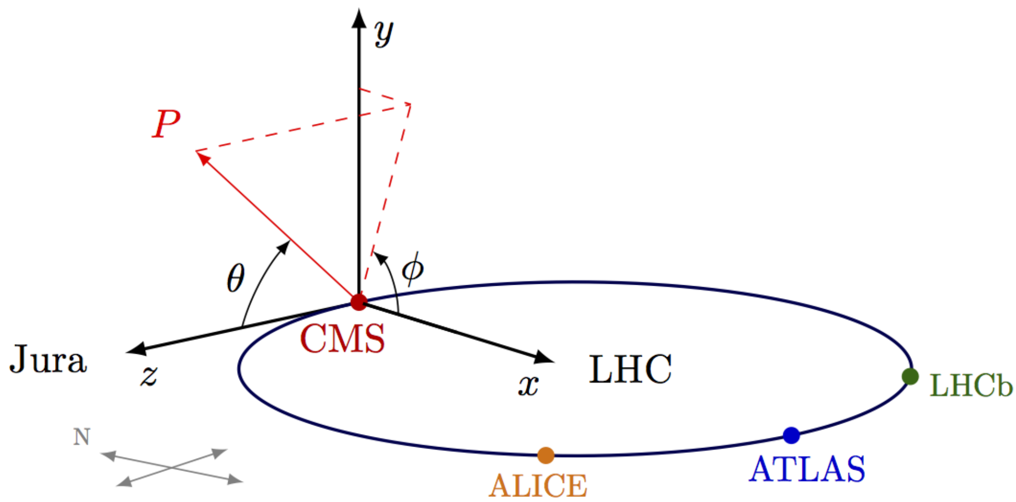


Figure 2.3: *CMS coordinate system.*

The CMS apparatus (Figure 2.4) is composed of a complex system of subdetectors, designed to detect electrons, photons, muons and hadronic jets produced in the collisions.
The CMS magnet produces a uniform magnetic field of 3.8 T within the solenoid, which has a diameter of 6 m and is 12.8 m long. It is composed of four Niobium-Titanium layers, cooled down to 4 K in order to show superconductor properties. The resulting energy stored in the magnet is about 2.7 GJ at full corrent ($\sim$ 19 kA). Outside the solenoid is the return yoke, made up of three sections along the z-axis holding the muon chambers in the gaps.

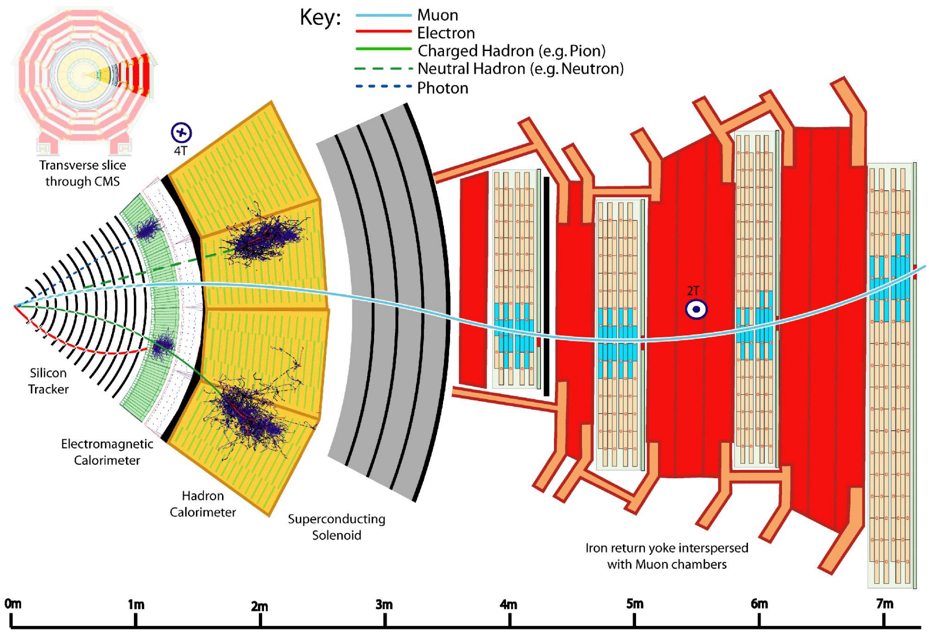


Figure 2.4: *Schematic view of a section of CMS.*

The magnetic field intensity in this region is about 1.8 T.
In the following a brief description of the main subdetectors in given, from the inner region to the outer one.
The closest detector to the beams collision point is the silicon tracker, which covers the pseudorapidity range $|\eta| < 2.5$. It is designed to reconstruct the trajectories (or tracks) of charged particles originating from the primary interaction point and to identify the position of secondary vertices produced by particles with a short mean life time, such as hadrons containing b quarks. The tracker is immersed in the magnetic field generated by the superconducting solenoid and this allows to determine the momentum of charged particles by measuring the curvature of their trajectories according to the equation (2.1). The inner part of the tracker ($r = 3$ cm) is a pixel detector. Each pixel has a pitch of $100 \times 150\ \mu\text{m}^2$, for a total of 124 million pixels. Spatial resolution on a single hit varies between 10 μm in the barrel regions and 15 μm in the endcaps. Around the pixel detector, a microstrip tracker extends up to $r = 110$ cm. It consists of 15148 silicon modules, covering an area of about 198 m^2 with a total of 9.3 million strips. Single hit reconstruction resolution varies between 10 μm and 50 μm depending on the η of the track. A schematic representation of the silicon tracker is shown in Figure 2.5.

The electromagnetic calorimeter (ECAL) surrounds the tracker system. It is a hermetic and homogeneous detector, whose main function is to identify

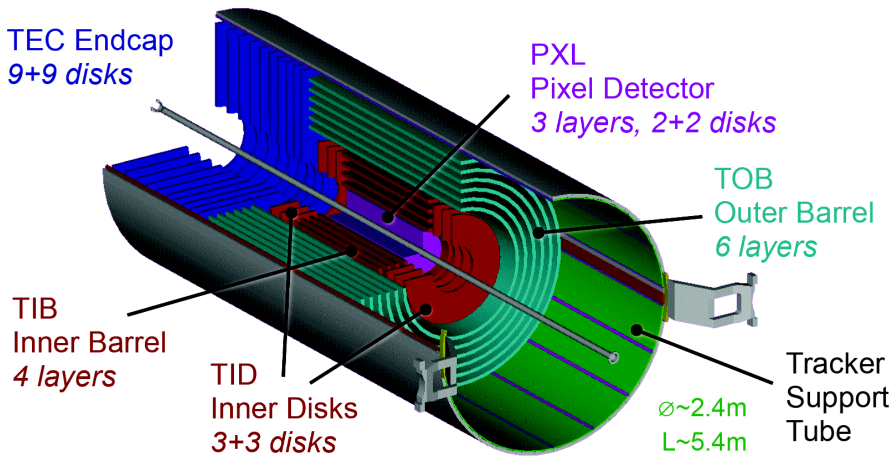


Figure 2.5: *Three-dimensional schematic view of the CMS silicon tracker.*

electrons and photons and to accurately measure their energy. This goal is achieved using scintillating crystals of lead tungstate ($PbWO_4$) with a truncated pyramidal shape. $PbWO_4$ is a suitable material because with its high density, short radiation length and small Molière radius allows to design a compact calorimeter with high granularity. In addition, this material has a fast scintillation decay time ($\sim$ 10 ns), which permits to collect about 80% of the light produced by the particle shower within the 25 ns interval between two consecutive bunch crossings.
The ECAL is divided into a cylindrical central part (EB) closed by two endcaps (EE), as shown in Figure 2.6. The former extends in the region $|\eta| < 1.479$, while the latter cover the region $1.479 < |\eta| < 3$ each. A preshower system (ES) is used to separate the shower produced by photons coming from the primary vertex from those produced from π^0 decay.
The energy resolution of a homogeneous calorimeter is given by:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{a}{\sqrt{E}}\right)^2 + \left(\frac{b}{E}\right)^2 + c^2 \tag{2.15}$$

where the first term is the stochastic contribution which accounts for fluctuations in the number of generated and collected photoelectrons; the second term is due to electric noise in the readout system and finally, the third term is dominant at high energies and and is related to calorimeter imperfections and other systematic errors. The energy resolution for electrons in the barrel

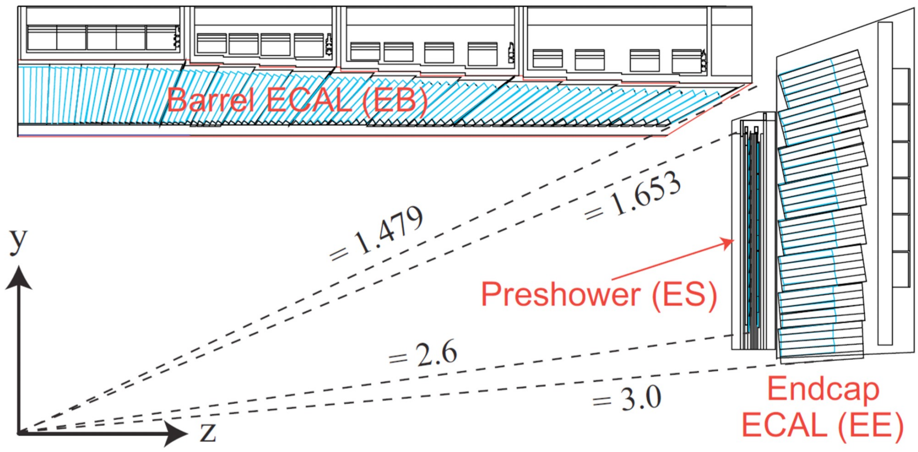


Figure 2.6: *Longitudinal view of an ECAL sector.*

region EB was measured to be:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{2.8\% \sqrt{\text{GeV}}}{\sqrt{E}}\right)^2 + \left(\frac{12\% \text{ GeV}}{E}\right)^2 + (0.3\%)^2 \qquad (2.16)$$

In order to make a complete calorimetric system for the jets and hadrons energy, the hadronic calorimeter (HCAL) is used togheter with ECAL. HCAL is a hermetic sampling calorimeter with brass absorbers and plastic scintillators as active elements, which covers the region within $|\eta| < 5$. It is divided into four distinct regions: the barrel (HB), the endcap (HE) and the outer (OH) and forward (HF) regions. A schematic view of these subdetectors is shown in Figure 2.7.
Because of its hermetc structure, it is also useful to indirectly measure weakly interacting particles, such as neutrinos, by means the energy imbalance in the transverse plane.
The scintillation light is collected by wavelength shifting (WLS) fibres and read out by hybrid photodiodes (HPD) and silicon photomultipliers. The energy resolution in the different regions of HCAL is expressed by the following formulas:

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{90\%\sqrt{GeV}}{\sqrt{E}}\right)^2 + (4.5\%)^2 \quad \text{in the barrel/endcap} \qquad (2.17)$$

$$\left(\frac{\sigma_E}{E}\right)^2 = \left(\frac{172\%\sqrt{GeV}}{\sqrt{E}}\right)^2 + (9\%)^2 \quad \text{in the HF} \qquad (2.18)$$

where E is expressed in GeV.

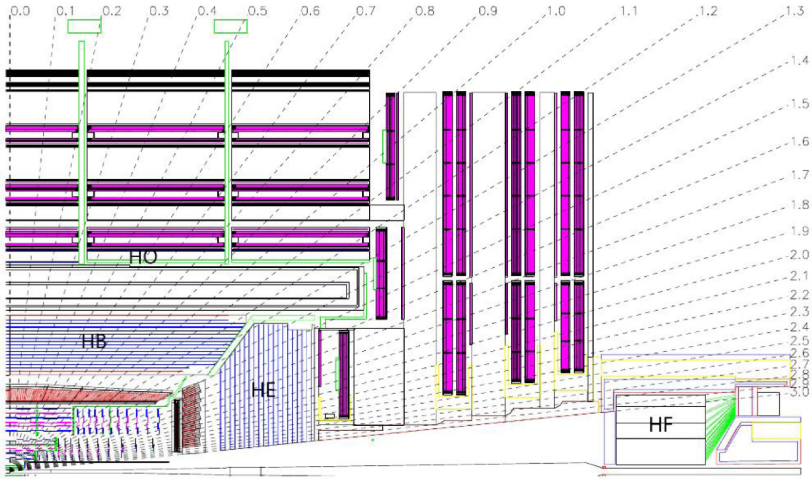


Figure 2.7: *Longitudinal view of an HCAL sector.*

Finally, the muon system is placed outside the solenoid magnet, embedded in the return yoke. The choice of its placement is dictated by the fact that muons lose little energy when passing through matter. Togheter with the silicon tracker, this detector complements the measurement of charged particles momentum. As shown in Figure 2.8, the muon system consists of several types of indipendent gaseous particle detectors: drift tubes are placed in the barrel regions, while cathode strip chambers in the endcaps. In addition, resistive plate chambers are placed in both regions.

Since the temporal displacement between two consecutive bunch crossings is 25 ns and multiple interactions occur for each bunch crossing, it is not feasible to collect and process all the data associated with the resulting number of events. In fact, the speed at which data can be written to mass storage is limited and moreover, only a fraction of the events involves high transverse momentum interactions. Therefore, a system trigger is required to select on the fly collision events which are interesting for physics analysis. The CMS trigger is organized in two stages, as explaind below. The first one is the Level 1 Trigger (L1) which is based on programmable electronics and aims to reduce the event rate from 40 MHz to 100 kHz. It is designed to make decisions within 3.2 µs, by collecting information coming from the calorimeters and the muon systems.
The second stage is the High Level Trigger (HLT), which is software based and reduce the events rate down to 1 kHz. The event selection is based on all the information coming from the subdetectors of the CMS apparatus.
Finally, the selected events are permanently stored in the CMS computing center and undergo the offline reconstruction.

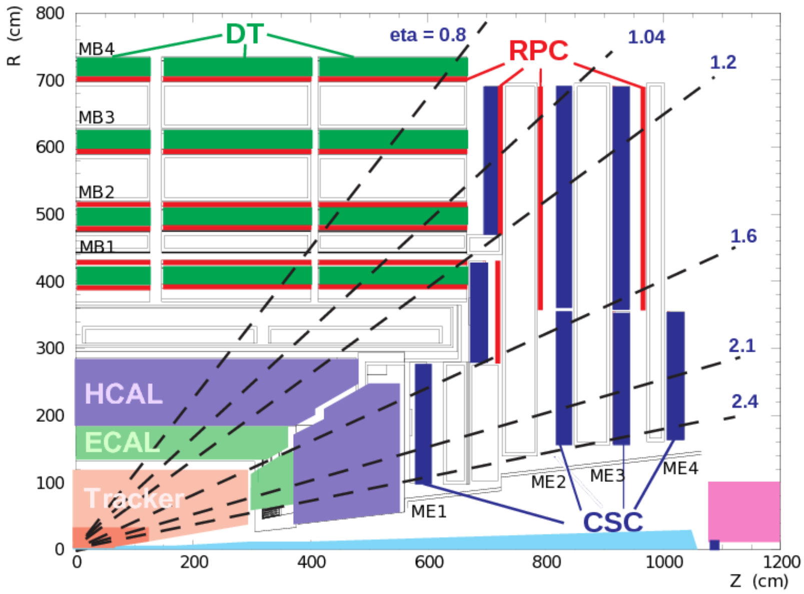


Figure 2.8: *Longitudinal view of the muon system.*

2.3 Event reconstruction

The reconstruction and identification of all the physics objects are based on the *Particle Flow* (PF) algorithm [16].
This technique uses the information from all CMS subdetectors, in order to determine the energy, direction and type of all the stable particles in the event: first, the silicon tracker allows to identify charged particle tracks; then, the energy deposits of the ECAL are merged into clusters and finally, the muon system provides the reconstruction of muon tracks.
The next step is to combine these elements to reconstruct the particles.
The reconstruction happens in the following order:

- muons are reconstructed by requiring a match between charged tracks in the silicon tracker and in the muon chambers. If the p_T of the candidate muon is compatible within 3 standard deviation with the one measured by the sole tracker, then a PF muon is generated and the corresponding track is removed from the calculation.

- Electrons reconstruction is based on the information of both the tracker and ECAL: since electrons can lose energy due to bremmstrahlung radiation, a PF electron is identified if a charged particle track matches one or more ECAL energy clusters. After the identification, the corresponding track and clusters are removed from the list.

- The remaining tracks give rise to PF charged hadrons. These tracks can be linked to ECAL and HCAL clusters and the energy is measured combining information from the calorimeters.

- Energy deposits of ECAL and HCAL which have no associated tracks in the silicon tracker or muon chamber are classified as photons and neutral hadrons respectively.

After the reconstruction of all PF particles in the events, PF hadron candidates are used to define jets that may be produced in the collision. The jet identification is based on the *anti-k_t* algorithm [17].
For each particle i with measured momentum $p_{T,i}$, two distances are calculated:

$$d_{i,B} \equiv p_{T,i}^{-2} \tag{2.19}$$

and

$$d_{i,j} \equiv \min(d_{i,B}, d_{j,B}) \frac{\Delta_{ij}^2}{R^2} \tag{2.20}$$

Equation (2.19) gives a measure of the distance from the beam, while equation (2.20) measures the distance from another j-th particle.
The Δ_{ij}^2 variable is defined as:

$$\Delta_{ij}^2 = (y_i - y_j)^2 + (\phi_i - \phi_j)^2$$

where y_i and ϕ_i are respectively the rapidity and azimuth angle of particle i. Finally, R is an adimensional radial parameter, which recalls the jet radius. The algorithm then calculates $d_{i,j}$ for all possible particle pairs in order to find the smallest one, which is then compared with $d_{i,B}$:

- if $d_{i,j}$ is smaller than $d_{i,B}$, particles i and j are combined into a new single object, summing their four-momenta, and the algorithm restarts from the first step;

- if $d_{i,B}$ is smaller than $d_{i,j}$, i is declared as a jet and removed from the list of input particles.

This procedure is repeated until no particles are left.

From a theoretical point of view, a jet clustering algorithm must satisy the infrared safety and the collinear safety conditions. These requirements are fulfilled if no infrared and collinear singularities appear in the perturative calculation, respectively. Therefore, a jet algorithm is infrared safe if it yields the same set of jets after modifying an event by adding a soft radiation. Similarly, a jet algorithm is collinear safe if the final set of jets is not changed

after introducing a collinear splitting of one of the inputs.
The *anti-k_t* algorithm is protected by these QCD divergences since the $d_{i,B}$ and $d_{i,j}$ are collinear safe by definition and the infrared safety is ensured by requiring a minimum threshold on the jet momentum.

Chapter 3

Measurement of the Higgs boson properties in the HWW channel

3.1 The *VBF* mechanism in the $H \to W^+W^-$ decay channel

This analysis focuses on the study of the *VBF* production mode, already described in section 1.4, which is considered as the signal process. The *VBF* mechanism is interesting because a precision measurement of its cross section allows to put tight constraints on the Higgs boson couplings to vector bosons and thus is an important test of the SM predictions. The measurement is performed by selecting events in which the Higgs boson decays in two opposite sign W bosons. Since $m_H < 2m_W$, the decay to two real W bosons is energetically forbidden and therefore at least one is required to be *virtual*, or equivalently *off-shell*. This means that its four-momentum does not satisfy the dispersion relation $p^\mu p_\mu = m^2$.
Because of its short mean lifetime ($\simeq 10^{-25}s$), both W bosons are only observed through their decay products. The W boson can decay to hadrons, with a branching ratio of 67%, or to leptons ($W^+ \to \ell^+\nu$ or $W^- \to \ell^-\bar{\nu}$) with a probability of 10.9% per lepton family. Although the fully hadronic decay mode is the most probable one, it is highly contaminated by multi-jet backgrounds, which make its identification hard. Even in the case of a semileptonic final state, the background contribution remains very large.
For this reason, the fully leptonic final state

$$H \to W^+W^- \to 2\ell 2\nu \tag{3.1}$$

is selected in this work. Despite its lower branching ratio with respect to the other decay modes, it shows a clear experimental signature and relatively low background contamination. The decay into leptons of the third family is not taken into consideration, since the identification of the tau lepton is experimentally difficult. Moreover, the measurement is affected by uncertainties higher than those of the $e\mu$ final state and thus, it does not bring significant improvements to the analysis result.
The experimental signature of the considered final states is characterized by two oppositely-charged leptons and missing transverse energy due to the escaping neutrinos. Although both leptons have high p_T values, the one coming from the off-shell W boson has on average a smaller p_T with respect to the one arising from the on-shell W boson. In addition, two jets emerge from the outgoing partons.
However, not only the decay of the Higgs boson in two W bosons gives rise to this final state. In fact, it can be mimicked by other background processes, which have to be accounted for in order to define a pure phase space enriched in signal events. These processes are detailed in the next section.
The main difficulty of this channel is that, due to the presence of the neutrinos, the Higgs boson invariant mass peak can not be reconstructed from the kinematics of the final state and other methods must be used to discriminate between signal and backgrounds.

3.2 Background processes

Non-resonant WW

One of the most important background processes that contributes to the considered final state is the production of W^+W^- pairs without involving a Higgs boson. This process is known as non-resonant WW production and the corresponding Feynman diagram is illustrated in Figure 3.1. When the process only involves electroweak vertices, it is called WW electroweak production.
If the two W bosons decay leptonically and two jets are originated from quarks or gluon emission, this kind of events has the same experimental signature of the signal. Such backgrounds are named *irreducible*, because experimentally they share the same final state of the signal process and can be distinguished only by means of kinematic properties.
According to the angular momentum conservation and due to the scalar nature of the Higgs boson, in the resonant case the spins of the W bosons must be antiparallel and fermions ones must sum up to 0. Moreover, since

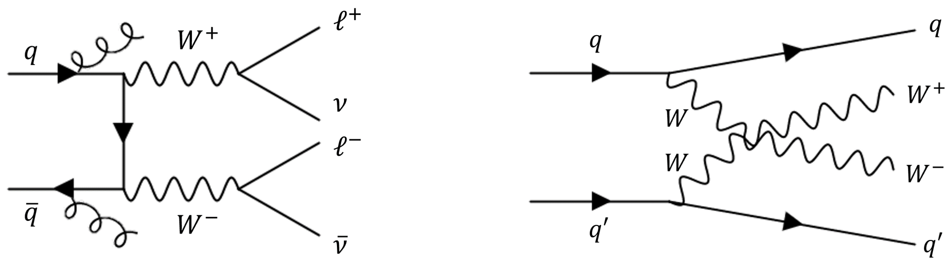


Figure 3.1: *Example of a Feynman diagram for non-resonant WW production with two jets coming from initial state radiation (left) and for WW electroweak process (right).*

the W boson is a spin-1 particle, the lepton and neutrino originating from its decay must have parallel spins. Considering also that weak interactions select left-handed (right-handed) components for particles (antiparticles), the spin correlation of the overall decay chain, shown in Figure 3.2, forces the directions of the two charged leptons to be nearly collimated. In contrast, in the non-resonant WW production case there is no preferential spin direction. This means that the signal process shows an azimuthal angle difference $\Delta\phi$ between the two leptons smaller than the background one.
Therefore, the dilepton invariant mass, defined as

$$m_{\ell\ell} = \sqrt{2p_T^{\ell_1}p_T^{\ell_2}(1-\cos(\Delta\phi))}$$

where the $p_T^{\ell_1}$ and $p_T^{\ell_2}$ are the transverse momentum of the candidate leptons coming from the W bosons decay, results smaller in the resonant case with respect to the non-resonant events. Hence, this observable is highly discriminating between this background and the signal.

Top quark

Another background that must be taken into account is due to events involving top quarks. The top quark weakly decays as

$$t \to bW^+$$

with a branching ratio of about 96% [6]: b quarks in turn give rise to an hadronic jet while the W boson may decay leptonically. Processes of single top quark, where a top and a W boson are produced, or $t\bar{t}$ production can therefore lead to final states similar to the signal, as illustrated in Figure

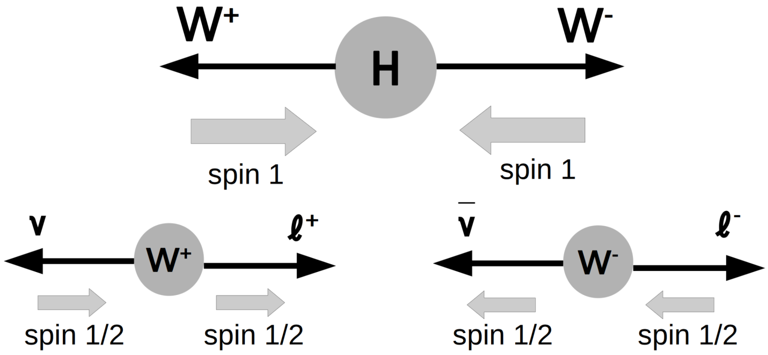


Figure 3.2: *Spin correlation of the $H \to W^+W^- \to 2\ell 2\nu$ decay chain.*

3.3. However, jets originating from b quarks, generically referred to as *b-jets*, show unique properties that allow them to be distinguished from light quark- or gluon-induced ones: b-jets typically contain bound states of b quarks, named B-hadrons, having high mass ($\sim$ 5 GeV) and relatively long lifetime ($\sim 10^{-12}s$) which allows them to travel for sizable distances before decaying. This leads to the presence of a secondary vertex due to the B-hadron decay within the b-jets, which is absent in the light-jets.
The CMS collaboration has developed specific algorithms to identify b-jets, called b-tagging algorithms, and top quarks events are therefore reduced by imposing the absence of b-tagged jets in the final state. However, this background can not be completely suppressed because the tagging efficiency on real b-jets is about 80%, and because of the relatively high cross section for top quarks production at the LHC energy scale.

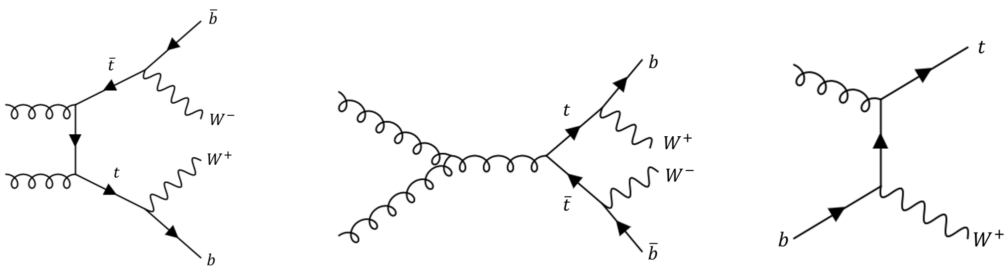


Figure 3.3: *Example of Feynman diagrams for $t\bar{t}$ pair production (left, middle) and tW production (right).*

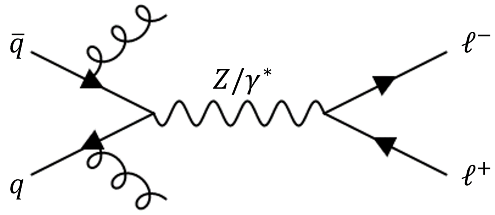


Figure 3.4: *Feynman diagram for leptonic Drell-Yann with two jets coming from initial state radiaiton.*

Drell-Yann

Drell-Yann (DY) processes occur when a $q\bar{q}$ pair annihilates to a photon or Z boson, which subsequently decays into two charged leptons with the same flavour. These events may be reconstructed with non-zero missing transverse energy due to detector inefficiencies and finite resolution, simulating the presence of neutrinos. If two jets come from initial state radiation, as pictured in Figure 3.4, this background may mimic the final state of interest. The DY contamination is reduced by selecting only events with two different flavour leptons, i.e. one electron and one muon. This requirement effectively suppresses the ee and $\mu\mu$ final states, but does not exclude the $DY \to \tau\tau$ contribution. In fact, τ leptons may decay leptonically in a e or μ lepton and two neutrinos, resulting in a final state with the same experimental signature as the signal. To distinguish such background, it is useful to introduce the dilepton momentum and the transverse mass of the system, defined as

$$p_T^{\ell\ell} = \left|\vec{p}_T^{\,\ell\ell}\right| = \left|\vec{p}_T^{\,\ell_1} + \vec{p}_T^{\,\ell_2}\right| \tag{3.2}$$

$$m_T^H = \sqrt{2p_T^{\ell\ell}E_T^{miss}[1 - \cos\Delta\phi(\vec{p}_T^{\,\ell\ell}, \vec{E}_T^{miss})]} \tag{3.3}$$

respectively. Here $\Delta\phi(\vec{p}_T^{\,\ell\ell}, \vec{E}_T^{miss})$ is the azimuthal angle between $\vec{p}_T^{\,\ell\ell}$ and $\vec{E}_T^{miss}$.
It can be shown that DY background gives rise to low values of $p_T^{\ell\ell}$ and m_T^H, given that the leptons coming from tau decays tend to be less energetic then the ones originating from a W boson. This is because the Z boson mass is smaller than the Higgs boson one and there are more neutrinos in this final state. Moreover, the spin correlation effect, that affects leptons coming from the Higgs boson decay, is absent in DY processes.
This kind of events is therefore reduced by means of a cut on m_T^H and $p_T^{\ell\ell}$ and requiring a threshold on the missing transverse energy variable.

Non-prompt leptons

Additional background contributions are due to events in which only a single W boson is present but it is produced in association with jets, and from semi-leptonic $t\bar{t}$ decays, in which a W boson decays into a $q\bar{q}$ pair. These processes constitute a background for the analysis when one of the jets is misidentified as a lepton, which then plays the role of the second lepton in the considered final state. The misidentification can be owed to a pion erroneously reconstructed as an electron, as well as a semileptonic B-hadron decay. In the latter case, the lepton actually exists but it is not coming from the primary interaction point. Such leptons are said to be *non-prompt* and are usually produced together with additional nearby particles.
Since prompt leptons originating from the hard scattering processes are expected to be isolated, this background is suppressed by requiring that signal electrons and muons meet specific isolation criteria.

For muons, an isolation variable is defined as:

$$I_{\mu}^{rel} = \left[\sum_{ChH} p_T + max \left(0, \sum_{NH} p_T + \sum_{pH} p_T - 0.5 \sum_{chHPU} p_T \right) \right] / p_T^{muon} \ , \quad (3.4)$$

where the sums are carried out in a cone of radius $\Delta R < 0.4$ around the muon direction. This variable is called *relative isolation* and the subscripts ChH, NH, Ph, $ChHPU$ refer to charged hadrons from the primary vertex, neutral hadrons, photons and charged hadrons from pileup vertices, respectively. The substractive term is a correction for the neutral hadrons coming from pile up. It accounts for the fact that a zero-charged particles does not leave any track in the tracker and therefore can not be assigned to a vertex with sufficient precision. It is estimated that the contribution due to neutral hadrons not arising from primary vertex is about half of that of charged hadrons. In this analysis, it is required that $I_{\mu}^{rel} < 0.15$.

A quite similar isolation variable is defined for the electrons:

$$I_{e}^{rel} = \left[\sum_{ChH} p_T + max \left(0, \sum_{NH} p_T + \sum_{pH} p_T - \rho A \right) \right] / p_T^{electron} \ , \quad (3.5)$$

The pile up effect is taken into account by the *effective area* correction technique, as described in [18]. The ρ parameter is the energy density associated to pile up events while A is the electron effective catching area. The cut applied on this isolation variable is $I_{e}^{rel} < 0.04$.

Other backgrounds

Other minor background processes are the following:

- WZ: this process mimics the signal final state when one of the three charged leptons, produced from the leptonic decays of the W and Z bosons, is outside of the detector acceptance, or is not reconstructed due to inefficiencies. The contribution of such events is estimated from Monte Carlo simulation;

- $W\gamma$: due to an interaction with the tracker material, the photon can be converted to an e^+e^- pair. If one of the two electrons goes out of the detector acceptance and if the W boson decays leptonically, this process simulate the signal signature. It is also estimated from simulation;

- $W\gamma^*$: since the photon is produced as an off-shell particle, it quickly decays into an e^+e^- or $\mu^+\mu^-$ pair. If one lepton is not revealed due to inefficiencies or acceptance effects, this background can enter the signal selection;

- Other VZ and VVV with $(V = \gamma, W, Z)$: diboson or triboson events, without the production of a Higgs boson, can enter the selection.

3.3 Data and simulated samples

Data sets and triggers

The data set used for this analysis was recorded by the CMS detector during the 2018 data taking and corresponds to 59.74 fb^{-1} of proton-proton collisions at a center of mass energy of $\sqrt{s} = 13$ TeV.

Events are demanded to pass either single-lepton or double-lepton trigger, which require the presence of one or two high transverse momentum electrons or muons, respectively. The single lepton trigger selects electrons with $p_T >$ 25 GeV in the central region ($|\eta| < 2.1$) and with $p_T > 27$ GeV for $2.1 < |\eta| < 2.5$, while muons must have $p_T > 24$ GeV for $|\eta| < 2.4$. The dilepton $e\mu$ trigger requires different p_T threshold depending on which is the leading lepton: 23 GeV for the leading electron and 8 GeV for the subleading muon, while 23 GeV for the leading muon and 12 GeV for the electron.

Monte Carlo simulations

Monte Carlo (MC) samples are employed in this analysis to simulate background and signal events and estimate systematic uncertainties. For background processes, these are useful to predict the yields of the residual contributions in the signal region, while for the signal these are used to extract the shape for the fit template. They are simulated at various perturbative orders in pQCD, using different event generators.

Gluon-gluon fusion and vector boson fusion production mechanisms are generated with POWHEG v2 [19] at next-to-leading order (NLO) accuracy. The VH production mode is instead simulated with MINLO HVJ. The subsequent decay of the Higgs boson in two W boson is permorfed using JHUGEN v7.1.4, which executes also the leptonic decays of the vector bosons. The BSM samples which describe HVV anomalous couplings, previously introduced in section 1.5, are generated with POWHEG v2, considering both the ggH and VBF production modes.

For what background processes are concerned, non-resonant WW event samples are simulated using two different generators, according on the production processes: $q\bar{q} \to WW$ is generated with POWHEG v2 at NLO accuracy while the $gg \to WW$ events simulation is perfomed by MCFM v7.0.1 at leading order (LO) accuracy. Single top and top-antitop pair production are generated at NLO. Finally, DY contributions are provided both at LO and NLO accuracy by MADGRAPH5_AMC@NLO event generator.

In order to provide simulation of initial and final radiations, hadronization and underlying events, all the event generators are interfaced to PYTHIA 8.1 [20]. Finally, the detector response is simulated using the GEANT4 software package [21], which is based on a detailed description of the CMS detector.

3.4 Monte Carlo corrections

Despite best efforts, MC samples do not always faithfully reproduce the data. For this reason *scale factors* (SF) are applied to MC samples as multiplicative event weights, with the aim to correct any disagreement with data. Such corrections are directly derived from data and calculated in phase spaces where the signal is not present. A short description of the most important SFs is reported below:

- Luminosity SF: it is used to normalize the number of events in the MC samples to the integrated luminosity in data;

- Pile up SF: it is needed to equalize the PU profile between simulation and data. In fact, MC samples are typically generated before the data taking and thus details of the PU distributions are not known beforehand but can only be predicted. The MC events are simulated using a PU profile which is a guess of what will occur in data. This SF takes into account also possible variations of the LHC instantaneous luminosity during tha data taking period. It is calculated as the ratio between the measured number of pile up interactions in data and the estimated one in the simulation;

- Trigger efficiency SF: differently from data, trigger selections are not applied to simulated samples for reasons similar to those explained at the previous bullet. Once data become available and the trigger menu is specified, MC samples have to be corrected. This is done with the *tag and probe* method [22] by exploiting the leptonic decay of the Z boson. These events present two leptons in the final state: if one passes tight isolation and identification criteria, then is used as a "tag" and paired with a "probe" in order to select a lepton pair whose invariant mass is close to that of the Z boson. The trigger efficiency is estimated as the fraction of the events in which the probe lepton passes the trigger thresholds and SFs are measured as functions of p_T and η of the leptons;

- b-tagging SF: it is necessary to correct for different b-tagging efficiencies in MC simulation and data;

- Lepton identification and isolation SF: it takes into account data/simulations differences in the lepton identification and isolation efficiencies and it is calculated with a tag and probe method quite similar to the one described above.

3.5 Systematic uncertainties

The MC distributions of physics observables are affected by several systematic uncertainties, which need to be taken into account. These are mainly divided into two categories, depending on the source from which they derive: *experimental* and *theoretical uncertainties*. In this section a brief description of both types of systematic uncertainties is given, while their parameterization and implementation in the analysis procedure is postponed to the subsection 4.7.1.

Experimental uncertainties

The following experimental uncertainties have been taken into account:

- the uncertainty on the measured integrated luminosity, estimated to be 2.5% for the 2018 data;

- the uncertainty on the b-tagging efficiency;

- the uncertainties on the trigger efficiencies;
- uncertainties on the lepton reconstruction and identification efficiencies; they depend on the η and p_T of the particle and result to be 2-5% for electrons and 1-2% for muons;
- the uncertainty on the measured jet energy scale, due to detector effects;
- the uncertainty on the modelling of E_T^{miss}, which is derived from lepton and jet energy scale uncertainties;
- the uncertainty on the pile up SF.

Theoretical uncertainties

Unlike the experimental ones, the theoretical uncertainties are independent of the detector response and affect the MC predictions of both signal and background processes. The source of uncertainties considered in this analysis are listed below:

- the uncertainty on the PDF;
- uncertainties on the theoretical cross section due to the missing higher order calculations;
- uncertainties on the modeling of the initial and final state radiation and the underlying event.

3.6 Differential and fiducial analysis

The increasing amount of data provided by the LHC in recent years has made it possible to measure the cross section of the process of interest as a function of one or more relevant kinematic variables, performing the so-called fiducial and differential measurements. The cross section is measured within a restricted phase space, called *fiducial volume*, that mimics the experimental selection in order to avoid an extrapolation to the full phase space using an underlying theoretical assumption. In addition, as will be described in subsection 3.6.1, experimental effects due to the finite resolution of the detector are removed from the observed distributions of the kinematic variables, in order to recover the original physical spectra.
This type of measurements therefore provides results that can be directly

compared to any physics model and thus, can be easily reinterpreted in the future, representing the legacy of today's experiments.

In this thesis work I present a measurement of the differential production cross section of the Higgs boson, as a function of the signed azimuthal angle difference $\Delta\Phi_{jj}$ between the two leading [1] jets in the H + 2 jets candidate events. The $\Delta\Phi_{jj}$ observable is defined as

$$\Delta\Phi_{jj} = \phi_{j_k} - \phi_{j_l} \tag{3.6}$$

with

$$\eta_{j_k} > \eta_{j_l}$$

where η_{j_k} and η_{j_l} are the pseudorapidity of the two jets with higher transverse momenta in the event, j_k and j_l.

The sign of this observable flips when a CP transformation acts on the system, as illustrated in Figure 3.5. Consider the two leading jets j_1 and j_2, originated from q and $\bar{q}$ respectively, and assume $\eta_1 > \eta_2$. The signed azimuthal angle difference between the two jets is given by $\Delta\Phi_{jj} = \phi_{j_1} - \phi_{j_2}$ (Figure 3.5.a). If a P transformation is applied to the system, the sign of the spatial coordinates of both jets is inverted and the $\Delta\Phi_{jj}$ observable becomes $\Delta\Phi_{jj} = \phi_{j_2} - \phi_{j_1}$ (Figure 3.5.b). Finally, the C operator transforms q into $\bar{q}$ and vice versa, so the jets flavour changes but there is no exchange between j_1 and j_2 (Figure 3.5.c) and the corresponding azimuthal angle difference reimains $\Delta\Phi_{jj} = \phi_{j_2} - \phi_{j_1}$.
Note that the $\Delta\Phi_{jj}$ definition (3.6) is P-odd but it is invariant under the exchange of the two protons. In fact, since $\Delta\Phi_{jj}$ is defined as the azimuthal angle of the "forward" jet (i.e. with larger pseudorapidity) minus the azimuthal angle of the "backward" jet (i.e. with smaller pseudorapidity), when the two protons directions are exchanged with each other, the "backward" and the "forward" jets also switch place. This leaves the sign of the $\Delta\Phi_{jj}$ observable intact [23].
The main aspect that makes $\Delta\Phi_{jj}$ a very interesting observable is its sensitivity to the tensor structure of the Higgs boson coupling to vector bosons, and therefore to possible anomalous Higgs boson couplings. Figure 3.6 shows the different shapes of the $\Delta\Phi_{jj}$ distribution on VBF signal events obtained by assuming the SM HVV coupling or the possible BSM HVV interactions introduced in section 1.5. Moreover, the SM predicts a symmetric distribution between $-\pi$ and π while the alternative model labeled H0Mf05 shows an asymmetric shape. The asymmetric behavior is due to the presence of

[1] All physics objects are listed in decreasing p_T order; the first and the second one are named *leading* and *subleading*, respectively.

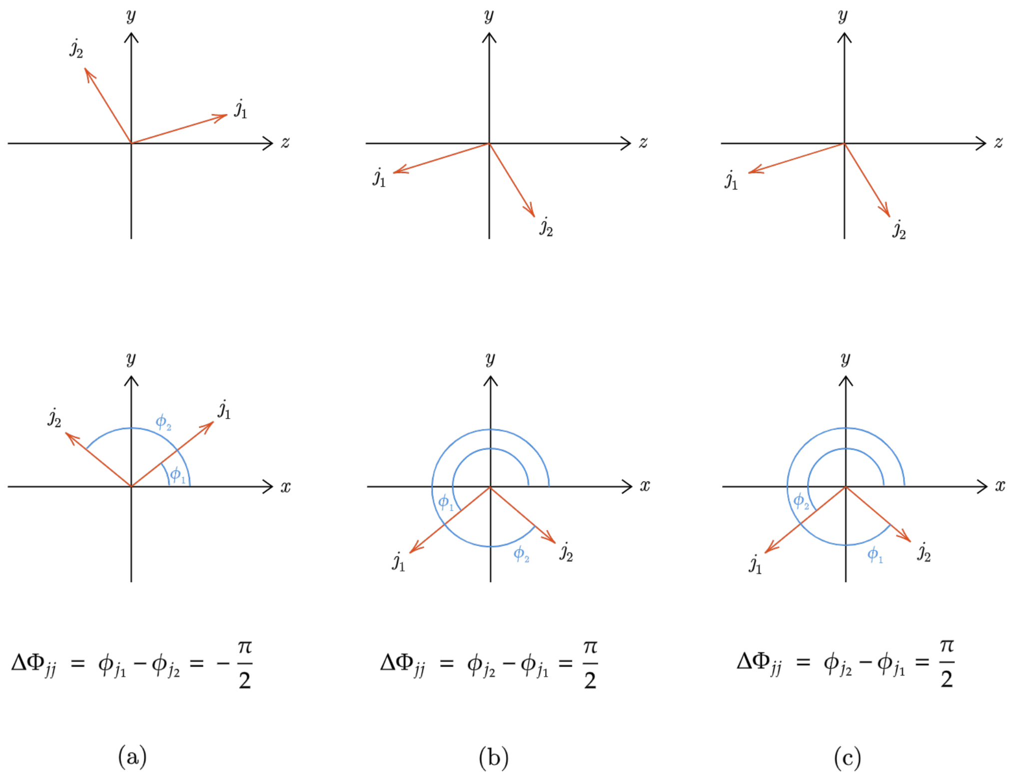


Figure 3.5: *The effect of a CP transformation on a two jets system:* (a) *shows the initial state of the system in the zy (top) and xy (bottom) planes;* (b) *shows the system after applying a P transformation; finally,* (c) *shows the effect of a C transformation on the system* (b).

interaction terms associated with CP-odd quantum numbers in the scattering amplitude (1.8). Such terms may interfere with the SM ones and reveal themselves as deviations from the SM expectation, as explained below. The full amplitude (1.8) can be rewritten according to the H0Mf05 model as:

$$\mathcal{A}(HVV) \sim a_1 M_{SM} + a_3 M_{CP\text{-}odd} \tag{3.7}$$

which results in the matrix element squared:

$$\begin{aligned} |\mathcal{A}(HVV)|^2 \sim a_1^2 |M_{SM}|^2 + a_3^2 |M_{CP\text{-}odd}|^2 + \\ + a_1 a_3 M_{SM} M^*_{CP\text{-}odd} + a_1 a_3 M^*_{SM} M_{CP\text{-}odd} \end{aligned} \tag{3.8}$$

For a small anomalous coupling a_3 the last two terms of equation (3.8), which are the interference ones, are sensible to the sign of the anomalous coupling.

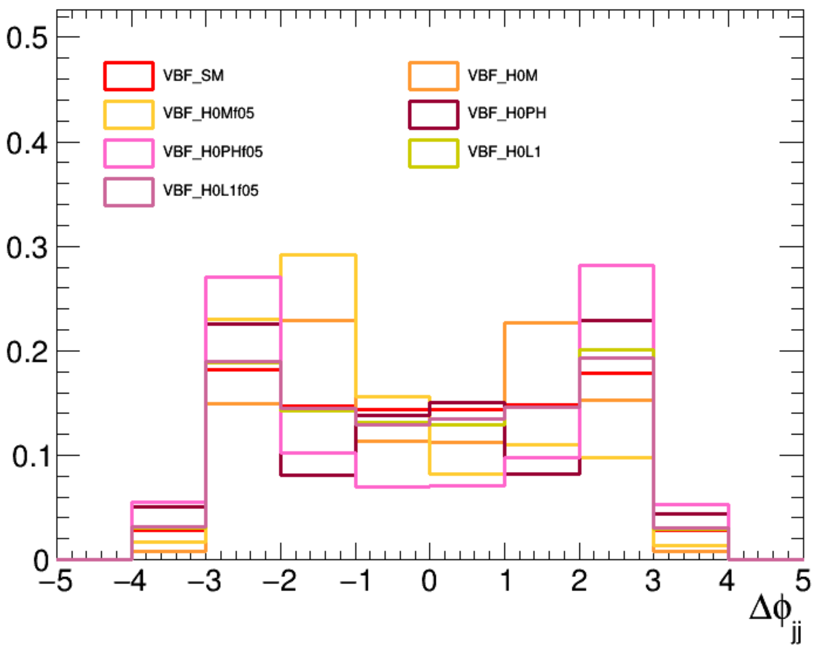


Figure 3.6: *$\Delta\Phi_{jj}$ distributions in $[-\pi, \pi]$ according to SM and several BSM theories.*

A measurement of the $\Delta\Phi_{jj}$ observable, and in particular of its asymmetry, can therefore provide hints of new physics.

3.6.1 The unfolding problem

As outlined before, the distinctive feature of differential measurements is their model independence. In order to make the comparison with theoretical predictions of different models or other experimental results easier, it is also necessary that the measured differential cross section would not be affected by apparatus effects. In fact, the kinematic properties of each event, such as four-momenta of particles and derived quantities, are measured only with limited precision due to the finite resolution of the detector. This implies that the observed spectra of the kinematic variables are blurred versions of the true ones: estimating these true spectra starting from observations distorted by the limited resolution of a particle detector constitutes the *Unfolding* problem.
In the following a brief formalization of the unfolding problem is reported [24],[25]. Consider a given observable, denote by x its true value and y the corresponding measured quantity. Let $f(x)$ be the true distribution of x and

$r(y|x)$ the probability density function to observe y given the true value x. The p.d.f. $g(y)$, that describes the distribution of the measured value y, is given by the convolution of $f(x)$ and $r(y|x)$ defined as:

$$g(y) = \int r(y|x)f(x)dx \tag{3.9}$$

One says that the true distribution is *folded* with $r(y|x)$ and hence the goal of estimating $f(x)$ is called unfolding. Generally speaking, the true phase space is named *generator* or *particle* level, while the smeared one is the *reconstructed* level; $r(y|x)$ is instead named *response function* or *kernel function*. In this analysis both phase spaces are subdivided in bins: let $\{T_j\}$, $j = 1, ...n$ and $\{S_i\}$, $i = 1, ...m$ be the partitions of the true and smeared space, respectively. The measured quantity is then the number of events in each of the bins at reconstructed level

$$\boldsymbol{\lambda} \equiv \left(\int_{S_1} g(y)dy, \; ... \; , \int_{S_m} g(y)dy \right), \tag{3.10}$$

while the true values are represented by

$$\boldsymbol{\mu} \equiv \left(\int_{T_1} f(x)dx, \; ... \; , \int_{T_n} f(x)dx \right). \tag{3.11}$$

These two quantities are connected by the response matrix (RM) of the detector $\boldsymbol{R}$, according to the expression

$$\boldsymbol{\lambda} = \boldsymbol{R}\boldsymbol{\mu}. \tag{3.12}$$

The RM element R_{ij} is defined as

$$R_{ij} \equiv \frac{\int_{S_i} \int_{T_j} r(y|x)f(x)dxdy}{\int_{T_j} f(x)dx} \tag{3.13}$$

and corresponds to the conditional probability that an event with true value x in bin T_j will be observed with measured value y in bin S_i. This formalization reduces the smearing carried out by the detector to a migration effect from a truth level original bin to a non corresponding smeared level bin. Therefore, the effect of the response matrix is to smudge any fine structure possibly present in the true spectrum. An example of the bin migration is illustrated in Figure 3.7, where, for simplicity of representation, $m = n$ has been chosen. Moreover, the unfolding is reformulated as the estimation of $\boldsymbol{\mu}$ starting from

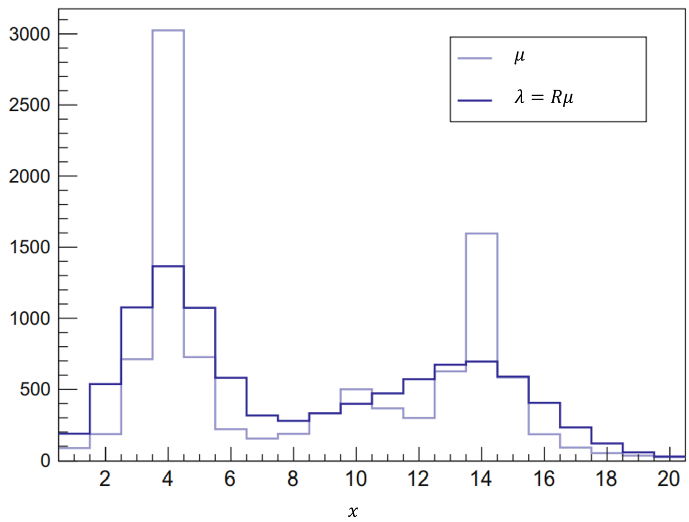


Figure 3.7: *Example of migration effect: the true distribution* $\boldsymbol{\mu}$ *(light blue line) with superimposed the expected smeared distribution* $\boldsymbol{\lambda}$ *(dark blue line)* [26].

a single observation of the quantity of interest, said $\boldsymbol{y}$, which fluctuates with a Poisson distribution around the mean value $\boldsymbol{\lambda}$:

$$\boldsymbol{y} \sim \mathcal{P}oisson(\boldsymbol{R\mu}). \tag{3.14}$$

The limiting factor with any unfolding construction is that the statistical fluctuations in the obserations $\boldsymbol{y}$ are amplified when propagated to the estimate of $\boldsymbol{\mu}$. In fact, when $\boldsymbol{\mu}$ is computed according to $\boldsymbol{\mu} \sim \boldsymbol{R}^{-1}\boldsymbol{y}$, if $\boldsymbol{R}$ is not sufficiently well behaved, a strongly oscillating spectrum can be found, with negative correlations between adjacent bins. Therefore, a regularization prescription is applied to the unfolding procedure, in order to reduce this oscillating behavior.
Up to here, backgrounds are not accounted for in this description: all types of events which are possibly observed in one of the reconstructed level bins but do not originate from any of the truth level bins have to be included. Suppose that $\boldsymbol{b}$ is the vector of expectation values of the entries observed in each reconstructed level bin coming from background processes. The relation 3.14 has to be modified to:

$$\boldsymbol{y} - \boldsymbol{b} \sim \mathcal{P}oisson(\boldsymbol{R\mu}). \tag{3.15}$$

The estimation of $\boldsymbol{\mu}$ can be achieved in several ways. The method adopted in this analysis is based on an implicit inversion of the response matrix within the likelihood maximization, which will be detailed in section 4.7.3.

3.7 Model dependence

As highlighted before, differential and fiducial analysis represents a valuable tool for probing alternative hypotheses with respect to the SM. For this reason, it is desiderable to limit the model depencence of the analysis, providing a result with no bias towards the SM expectation. However, some sources of model dependency may arise from the analysis strategy.
The most important one lies in the signal extraction procedure. In fact, as mentioned earlier, the presence of neutrinos in the $H \rightarrow WW$ final state makes the full kinematic reconstruction, and thus the measurement of the invariant mass of the system, impossibile. If the invariant mass was reconstructed, it would have allowed a straightforward discrimination of signal and background processes, because only signal events would gather at invariant mass values around the Higgs boson mass, forming a relatively narrow peak. Moreover, the selection of signal events would have been completely model independent, since the shape of the invariant mass peak does not depend on the considered physics model. Therefore, the signal extraction relies on an alternative procedure, known as template fit, which will be detailed in section 4.7. The general idea is to fit the MC simulation of an observable with good signal to background discriminating power to data.
However, the shape of the observable generally depends on the Higgs boson properties such as, for instance, the HVV coupling properties. Since these properties are predicted to be different from the SM ones by several BSM theories, the differential measurement keeps a degree of dependence on the signal hypotesis used to generate the template. Such behavior spoils the reinterpretability of the result and must be taken into account.
Another possible source of model dependency can arise from the unfolding procedure. The response matrix $\boldsymbol{R}$ may indeed depend on the considered signal model and only if it is quite diagonal it introduces a small dependence which can be neglected.
Finally, the event selection efficiencies may be different according to the signal models.
The purpose of this thesis is to reduce the impact of these sources of model depencence on the analysis result, in particular the one due to the shape effect of the fit variable.

Chapter 4

Analysis strategy

4.1 Preselection and categorization

The analysis strategy adopted in this thesis to measure the signal yield is inspired to the ones developed by the CMS collaboration for previous works [27]. The phase space is split into different regions, constructed by cutting on kinematic variables that show some discriminating capability between signal and backgrounds. Two orthogonal macro-categories are defined: Signal Region (SR) and Control Region (CR). The former is defined to maximize the sensitivity with respect to the process of interest, while the latter is expected to contain a low fraction of signal events in favor of backgrounds, and is used to constrain the main background contributions.
The CRs are also used to check that the data are well modeled by the MC simulations of the main background processes. In the first stages of the analysis data are shown only in the CRs, following the "blinding" policy of the analyses in the CMS collaboration. This approach aims to prevent the experimenter bias, i.e. the unintentional tendency to choose specific cuts in an attempt to interpret statistical fluctuations of the data as a signal. Only once the complete procedure will be reviewed by an independent committe of members of the CMS collaboration, the data will be revealed in the SRs.

The event categorization is summarized in Table 4.1. In addition to the SR, two control regions are defined: the top CR and the $DY \to \tau\tau$ CR, enriched in top quark and $DY \to \tau\tau$ events respectively.
Preselection criteria are applied to all phase space regions. Two jets having $p_T > 30$ GeV each and an invariant mass higher than 120 GeV are required in the event. The leading and subleading leptons (ℓ_1 and ℓ_2 respectively) must pass p_T thresholds dictated by the trigger requirements, while the p_T of the third lepton, if present, is demanded to be below 10 GeV in order to suppress

Table 4.1: *Events preselection and definition of signal and control regions.*

Region	Requirements
Preselection	oppositely-charged $e\mu$ final state, at least two jets with $p_T > 30$ GeV, $p_T^{\ell_1} > 25$ GeV, $p_T^{\ell_2} > 13$ GeV, $p_T^{\ell_3} < 10$ GeV, $p_T^{\ell\ell} > 30$ GeV, $m_{\ell\ell} > 12$ GeV, $m_{jj} > 120$ GeV, $E_T^{miss} > 20$ GeV
SR	$m_T^H > 60$ GeV, $m_T^{\ell_2} > 30$ GeV, no b-tagged jets with $p_T > 20$ GeV
Top CR	$m_{\ell\ell} > 50$ GeV, $m_T^{\ell_2} > 30$ GeV at least one b-tagged jet
$DY \to \tau\tau$ CR	40 GeV $< m_{\ell\ell} <$ 80 GeV, $m_T^H < 60$ GeV, no b-tagged jets with $p_T > 20$ GeV

minor backgrounds, such as WZ and triboson production. Moreover, the dilepton invariant mass is required to be higher than 12 GeV to reduce the contribution from QCD and low mass resonances. The requirement of having different flavour lepton pairs in the event suppresses $DY \to ee$ and $DY \to \mu\mu$ events, while the $p_T^{\ell\ell}$ and m_T^H requirements are necessary to reduce the $DY \to \tau\tau$ process. A moderate amount of E_T^{miss} value is requested because of the presence of the neutrinos in the event. The $m_T^{\ell_2}$ variable is analogous to m_T^H considering only the subleading lepton. It is defined as

$$m_T^{\ell_2} = \sqrt{2p_T^{\ell_2} E_T^{miss}[1 - \cos\Delta\phi(\vec{p}_T^{\,\ell_2}, \vec{E}_T^{\,miss})]} \tag{4.1}$$

and it helps to reduce the non-prompt lepton background. Finally, no b-tagged jets with $p_T > 20$ GeV must be present in the SR and $DY \to \tau\tau$ CR in order to reduce background contributions involving the top quark. This requirement is not applied to the Top CR, where at least a b-tagged jet is requested to enhance the sensitivity to top quarks events.
Figure 4.1 shows the distribution of the $\Delta\Phi_{jj}$ observable in the SR.

In order to perform a differential cross section measurement, all regions are further divided according to the value of the azimuthal angle difference between the two leading jets in the final state. Four categories or bins are defined at reconstructed level, as detailed in Table 4.2.

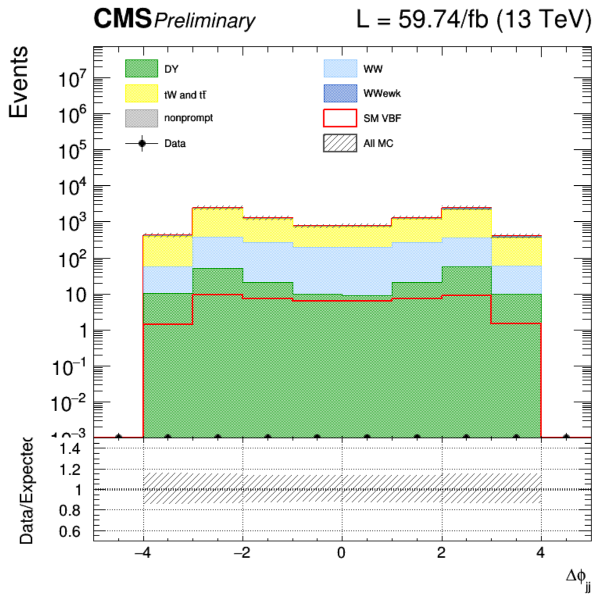


Figure 4.1: *Distribution of the $\Delta\Phi_{jj}$ observable in the SR. The signal contribution (red line) is shown both stacked on top of the backgrounds and superimposed on it.*

Table 4.2: *Definition of the $\Delta\Phi_{jj}$ bins.*

Bin 0	Bin 1	Bin 2	Bin 3
$-\pi < \Delta\Phi_{jj} < -\frac{\pi}{2}$	$-\frac{\pi}{2} < \Delta\Phi_{jj} < 0$	$0 < \Delta\Phi_{jj} < \frac{\pi}{2}$	$\frac{\pi}{2} < \Delta\Phi_{jj} < \pi$

The selection criteria outlined in this section and the intrinsic acceptance of the detector define the subset of the phase space at reconstructed level which is experimentally accessible.

4.2 Neural Networks for event classification

The challenge of measuring rare signals dominated by the towering LHC backgrounds has lead to the introduction of Machine Learning (ML) techniques in the analysis procedure. One of the most common uses is to employ Neural Networks (NNs) for classification purposes [28].
A NN is a parametric approximation of an unknown function f, which maps a set of inputs x_i to the corresponding outputs y_i so that, ideally, $f(x_i) = y_i$. The NN architecture is based on a collection of connected units called *neu-*

rons or *nodes*, typically organized into multiple *layers*. Each node has a real number associated to it, called *activation*. In the simplest architecture, called a feed forward NN, neurons of one layer are linked only to neurons of the immediately preceding and immediately following layers. Each connection has an associated weight, which determines the strength of one node influence on another.

Let A_i^k be the activation of the i-th node of the k-th layer. A_i^k is given by a linear combination of the activations of all nodes of the previous layer, subsequently passed through a non linear function called *activation function*:

$$A_i^k = \sigma\left(\sum_j w_{ij}^k A_j^{k-1} + b_i^k\right) \tag{4.2}$$

where σ is the activation function of the k-th layer, w_{ij}^k and b_i^k are the weights and the biases associated to the connections between the layers $k-1$ and k and A_j^{k-1} are the activations of the nodes belonging to the $(k-1)$-th layer which are linked to A_i^k.

The layer that receives external data is the *input layer*. Intermediate layers are commonly called *hidden layers*, while the last layer is called the *output layer* whose activations form the output of the NN. Between the two outermost layers, multiple connection patterns are possible. A layer is said to be fully connected if each of its neurons is connected to every neuron in the previous one.

Once its structure is defined, the parameters of the NN must be chosen to optimally reproduce the desired function. This process is called *training* (or *learning*) and aims to adjust all the weights and biases to improve the accuracy of the results. To achieve this, a set of example inputs-outputs pairs is used (*supervised learning*). For the sake of simplicity, let us call $\boldsymbol{W}$ the set of weights and bias parameters.

The training proceeds by minimizing a *loss function* ($\mathcal{L}$), which depends on all the parameters $\boldsymbol{W}$ and provides a measure of the distance between the value predicted by the algorithm and the actual target value.

The minimization is performed using an iterative method, known as *gradient descent*. The main idea is to compute the gradient of the loss function with respect to $\boldsymbol{W}$

$$\vec{\nabla}_{\boldsymbol{W}}\mathcal{L}(\boldsymbol{W})$$

and then update the parameters by adjusting their values in the direction opposite to the gradient vector:

$$\boldsymbol{W'} \leftarrow \boldsymbol{W} - \eta\, \vec{\nabla}_{\boldsymbol{W}}\mathcal{L}(\boldsymbol{W})$$

This procedure is repeated iteratively until a minimum of $\mathcal{L}$ is reached. The η parameter is called *learning rate* and controls the size of the steps taken to reach the minimum.
Toghether with the number of layers, nodes and any other parameter that controls the overall structure of the NN and the training procedure, they form the set of so called *hyperparameters*.

Classification algorithms are a type of ML algorithms where the outputs are restricted to a limited set of values, or classes like signals or backgrounds. Suppose to have a set of input values $\boldsymbol{x}$, corresponding to the kinematic variables of an event, and want to classify the events into N classes. One way to achieve this is to build the NN such that for each class $n = \{1, ..., N\}$ it generates a positive number $p_n(\boldsymbol{x})$ which can be interpreted as the probability of the input event to belong to the class n. This is achieved by using the *softmax function* as activation function of the output layer. Given the N input values $z_n(\boldsymbol{x})$ of the output layer, it is defined by the formula:

$$p_n(\boldsymbol{x}) = \frac{e^{z_n(\boldsymbol{x})}}{\sum_{m=1}^{N} e^{z_m(\boldsymbol{x})}} \tag{4.3}$$

This function has the desirable property of converting its inputs into N positive numbers whose sum is bound to 1:

$$\sum_{n=1}^{N} p_n(\boldsymbol{x}) = 1 \tag{4.4}$$

Each $p_n(\boldsymbol{x})$ represents the best estimation of the true target value y_n, defined as:

$$y_n = \begin{cases} 1 & \text{if class} = n \\ 0 & \text{if class} \neq n \end{cases} \tag{4.5}$$

In order to describe the performance of a classification model, the so-called *confusion matrix* (M) is often computed. By definition, M_{ij} is equal to the fraction of observations known to be in class i and predicted to be in class j. If the algorithm has good classification capabilities, the confusion matrix is expected to be close to diagonal.
The loss function usually used in binary classification, where the number of classes equals 2, follows from the Binomial distribution and is called Binary Cross-Entropy, defined as:

$$BCE = -y \log p(\boldsymbol{x}) - (1 - y) \log(1 - p(\boldsymbol{x})) \tag{4.6}$$

where p is the binary indicator (0 or 1) associated to the first class predicted by the neural network and y is the corresponding target value. Being the

outputs forced to sum up to 1, $(1-p)$ and $(1-y)$ are the predicted probability and the target value associated to the second class, respectively.
In extention to multi-class classification tasks, the Categorical Cross-Entropy (CCE) is used as loss function. In the case of a NN with N classes and for a single event, it corresponds to:

$$CCE(N) = -\sum_{n=1}^{N} y_n \cdot \log p_n(\boldsymbol{x}) \tag{4.7}$$

where p_n is the probability associated to the n-th class as predicted by the algorithm and y_n is the corresponding target value. In order to optimize the training procedure, the loss function is not evaluated event by event but rather is updated after an entire batch has been run. The *batch size* is the hyperparameter that defines the number of events that will be processed by the network before updating the internal model parameters.
Finally, the last hyperparameter that needs to be specified is the number of *epochs*, which corresponds to the number of cycles that the algorithm will work through the entire data set during the training procedure. An accurate optimization of this hyperparameter allows reducing the *overtraining*, which is the tendency of the algorithm to learn the statistical fluctuations of the training sample and interpret them as real features.

4.3 Adversarial Neural Network

As detailed in section 3.7, in order to preserve the model independence of the $\Delta\Phi_{jj}$ differential analysis, it is necessary to define a learning algorithm that classifies events correctly and at the same time does not implicitly distinguish different physics models with which signal events are generated.
I achieved this goal using a technique known as *Domain Adaptation* (DA) [29]. DA is a branch of ML concerned with studying how an algorithm performs when evaluated on a data set different from the one it was trained on. A set of techniques has been developped that allows the definition of NNs whose performance is robust when applied on a range of different data sets, called *Domains*. In case the different data sets correspond to events generated assuming different signal models, DA consists in building a NN that is agnostic with respect to the considered signal hypothesis, meaning that the shape of the output discriminant is the same regardless of the model assumption.
The implementation of the DA I developed in this work is based on an *Adversarial Neural Network* (ANN). The ANN is a system of two NNs, consisting

of a classifier (C) and an adversary (A), which are trained in a competitive way. Both C and A have the usual structure of a NN for classification purposes introduced in the previous section but are trained to perform different tasks: the role of the classifier is to determine if the event is signal- or background-like and the network is trained on a data sample including events arising from different domains, i.e. different signal models. On the other hand, the goal of the adversary is to guess the physics model of signal events, regressing the domain from the second-to-last layer of the classifier.
The two networks are trained so that C learns to discriminate between signal and background, but is penalized if the output contains too much information on the domain of origin of signal events. To achieve this, I have implemented a two-step training procedure.
First the adversary is pre-trained with a CCE loss $\mathcal{L}(A)$; next the classifier is trained with a combined loss function, while keeping the weights of A frozen. The combined loss is given by:

$$\mathcal{L}(C + A) = \mathcal{L}(C) - \alpha \cdot \mathcal{L}(A) \tag{4.8}$$

where $\mathcal{L}(C)$ is the CCE loss of the classifier and α is a configurable hyperparameter that regulates how much the network is penalized for learning the signal model.
All hyperparameters defining the ANN have to be optimized so that C is able to discriminate events while simultaneously preventing A from regressing the domain. When an equilibrium between the performance of each of the networks is reached, the output of the classifier is independent of the signal model.
The structure of the ANN that I implemented is represented in Figure 4.2 and is described in the following.

The Classifier

For what the classifier structure is concerned, I defined three classes as target functions, one for *VBF* and two for the backgrounds that I want to separate from the signal: *ggH* and *BKG*. The label *ggH* refers to gluon-gluon fusion processes while the *BKG* embeds both top quark events and non-resonant WW production, the two main background processes in this analysis.
The input variables $x_1, ..., x_m$ of this network are the measurable kinematic variables of an event, which are detailed in the following subsection.
The activation function of the hidden layers is the *rectified linear activation function* (ReLU), which is defined as

$$\text{ReLu}(x) = \max(0, x) .$$

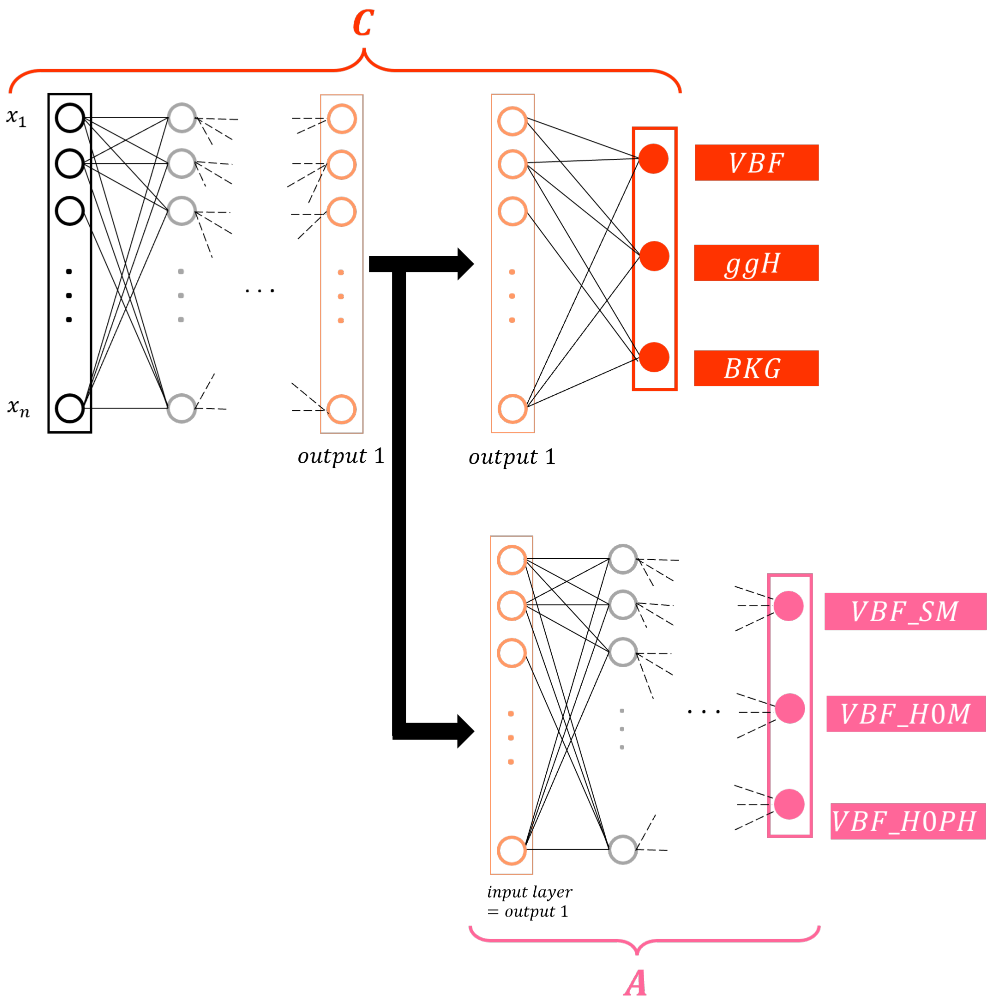


Figure 4.2: *Schematic view of the Adversarial Neural Network.*

The output layer uses the softmax function.
The hyperparameters n_l^C, η^C and n_{nodes}, which correspond to the number of hidden layers, the learning rate and the number of nodes in the hidden layers respectively, need to be optimized through a procedure detailed in the following.

The Adversary

Since two possible alternative physics model with respect to the SM have been considered, the adversary also has three classes, labeled by *VBF_SM*, *VBF_H0M* and *VBF_H0PH*. The first one refers to the *VBF* signal events generated as-

suming the SM hypothesis; the second and the third ones label the signal events obtained according to H0M and H0PH models, respectively. Referring to equation (1.8), these models correspond to $a_1^{VV} = 0,\ a_3^{VV} = 1$ and $a_1^{VV} = 0,\ a_2^{VV} = 1$ respectively. These models are chosen because they show physical properties that are very different with respect to the SM ones.
The input layer of this network is the second-to-last layer of the classifier. The ReLU activation function is used for the hidden layers, while softmax is used for the output layer.
As for the classifier, the number of hidden layers n_l^A, the learning rate η^A and the number of nodes in the hidden layers n_{nodes} have to be optimized. Note that n_{nodes} is chosen to be the same for both C and A.

4.3.1 Training procedure

The training has been performed on MC simulations and the following samples have been chosen: SM ggH production, SM non-resonant WW and top quark events, which constitute the background event samples, and finally the VBF signal. The latter process has been generated according to the three considered physical models: SM, H0M and H0PH.
Before starting the learning, events are selected through the kinematic requirements listed in Table 4.3, which correspond to the preselection and SR criteria discussed in detail in section 4.1:

Table 4.3: *Selection criteria applied on the training sample.*

Selection criteria
oppositely-charged $e\mu$ final state, at least two jets with $p_T > 30$ GeV,
$p_T^{\ell_1} > 25$ GeV, $p_T^{\ell_2} > 13$ GeV, $p_T^{\ell_3} < 10$ GeV,
$m_{\ell\ell} > 12$ GeV, $p_T^{\ell\ell} > 30$ GeV,
$E_T^{miss} > 20$ GeV, $m_{jj} > 120$ GeV,
$m_T^H > 60$ GeV, $m_T^{\ell_2} > 30$ GeV,
no b-tagged jets with $p_T > 20$ GeV,

The training sample was randomly split in two parts, one used for the training and the other for the *validation*, i.e. the ANN is tested on a small fraction of the training samples to check the overtraining of the algorithm.
The learning procedure has been performed using Keras [30] and TensorFlow [31] libraries and exploiting Google GPUs[1]. GPUs allow to reach optimal performances in terms of the time needed for completing the training procedure. For instance, if compared to the CPUs usage, the time employed to find the

[1]Google GPUs can be exploited through the free cloud service Google Colab.

loss minimum is reduced by about a factor 20 and it takes only five minutes instead of an hour and an half to finish the training procedure.
For the construction of the ANN, a set of input variables that could highlight the signal characteristics with respect to the other processes has been defined. The input variables chosen for the training procedure are the following:

- $p_{T_{j_1}}$, $p_{T_{j_2}}$: the magnitudes of the transverse momenta of the leading and subleading jet respectively;

- η_{j_1}: the pseudorapidity of the leading jet j_1. Since it is typically generated by the hadronization of a q quark with a high fraction of the proton four-momentum, j_1 coming from the VBF process tends to be emitted at large $|\eta|$. Instead, background events are characterized by the presence of a leading jet in the central region, i.e. at low $|\eta|$;

- η_{j_2}: the pseudorapidity of the subleading jet. Its distribution is similar to η_{j_1} for the VBF process, while background events show homogeneous shapes throughout the range of possible values of this variable;

- $|\Delta\eta_{jj}|$: the separation in η between the two jets in the final state. The signal shows a larger pseudorapidity gap with respect to the backgrounds;

- m_{jj}: the invariant mass of the dijet system. It has a good discrimination power since it is typically larger for the signal than for the background;

- $p_T^{\ell\ell}$, $p_T^{\ell_1}$, $p_T^{\ell_2}$: the magnitudes of the transverse momenta of the dilepton system, the leading lepton and the subleading lepton respectively;

- η_{ℓ_1}, η_{ℓ_2}: the pseudorapidity of the leading and subleading lepton respectively. Similarly to $p_T^{\ell\ell}$, $p_T^{\ell_1}$ and $p_T^{\ell_2}$, these variables are not expected to have a high discrimination capability individually, but are included to exploit their correlation with other observables;

- $m_{\ell\ell}$: the invariant mass of the lepton pair. As explained in section 3.2, due to the spin correlation effect this variable is peaked at low values for VBF and ggH mechanisms and it shows a broadened shape for non-resonant events;

- $\Delta\phi_{\ell\ell}$: the angular separation in ϕ between the two leptons in the final state. For signal events, it has a smaller value than in the non-resonant case, since the spin correlation effect forces the directions of the two charged leptons to be nearly collimated;

- $\Delta R_{\ell\ell}$: the radial separation between the two leptons in the final state. Similarly to $\Delta\phi_{\ell\ell}$, this variable is almost flat for events without a resonance;

- $m_{\ell j}$: the invariant mass of the system composed by the ℓ-th lepton and j-th jet, where $\ell = \{\ell_1,\ \ell_2\}$ and $j = \{j_1,\ j_2\}$. There are four possible combinations which have limited, but not completely negligible, discrimination power;

- $C_{tot} = \log\Big(\sum_{\ell} |(2\eta_\ell - \sum_{j} \eta_j)/|\Delta\eta_{jj}|\Big)$, where $\ell = \{\ell_1,\ \ell_2\}$, $j = \{j_1,\ j_2\}$: it represents a measure of how much the charged leptons are emitted centrally with respect to the two-jets system;

- qgl_j: the quark-gluon likelihood discriminant [32] evaluated for the leading and subleading jet ($j = \{j_1,\ j_2\}$). This variable is constructed with a multivariate technique and it gives a probability that the j-th jet has been produced by the hadronization of a quark or a gluon: the discriminator tends to assume values close to 1 for quark-like jets and close to 0 for gluon-like jets. Background events present jets usually emitted by gluons, since they are coming from initial or final state radiation. On the other hand, the signal always has quark-like jets;

- m_T^H: the transverse mass of the system. As already shown in section 3.2, background events are more likely to present a small value of this varible;

- $m_T^I = \sqrt{(p_T^{\ell\ell} + E_T^{miss})^2 - (\vec{p}_T^{\,\ell\ell} + \vec{E}_T^{\,miss})^2}$: the improved transverse mass. It tends to have a different shape for events containing the production of a Higgs boson;

- E_T^{miss}: the missing transverse energy;

- $\Delta\phi(\vec{p}_T^{\,\ell\ell}, \vec{E}_T^{\,miss})$: the azimuthal opening angle between $\vec{p}_T^{\,\ell\ell}$ and $\vec{E}_T^{\,miss}$. It has a good discriminating power and it is included in the definition of m_T^H;

- H_T: the scalar sum of the transverse momenta of all jets in the event. It gives a measure of the hadronic activity of the event.

4.3.2 Optimization

In order to get the best training of the ANN, it was necessary to optimize its hyperparameters. The optimization has been carried out using the Optuna

software [33], which allows the simultaneous maximization or minimization of multiple objective functions.
The classifier has been optimized by maximizing its accuracy, i.e. the fraction of events correctly classified.
Moreover, in order to reduce the model dependence of the overall network the Kolmogorov-Smirnov (K-S) test statistic has been employed, to have a measure of compatibility between the VBF-output shapes of signal events simulated under the SM and H0M hypotheses: a small value corresponds to a good compatibility between the two shapes. If no residual model dependence is left, the two distributions are expected to be compatible within the statistical accuracy. The same test statistic has been repeated for the SM and H0PH signal hypotheses.
Finally, the optimization has been performed by minimizing the objective function defined as the mean between the two K-S test statistics and by maximizing the classifier accuracy.
In order to find the best estimation of the hyperparameters, 100 training trials of 500 epochs each have been executed, and during every training the values of the hyperparameters have been changed according to a bayesian optimization approach [34] within defined ranges.
The intervals associated with possible outcomes of the hyperparameters are reported in the Table 4.4.

Table 4.4: *Ranges within the optimization procedure can vary the hyperparameters.*

Hyperparameter	Interval
α	$[0, 100]$
n_{nodes}	$[10, 100]$
n_l^C	$[1, 10]$
n_l^A	$[1, 10]$
η^C	$[10^{-5}, 10^{-3}]$
η^A	$[10^{-4}, 10^{-2}]$

The accuracy and the K-S test values returned by the optimization procedure are summarized in Figure 4.3: each training trial is represented by a blue dot; the red dots correspond to the best attemps that provide extreme values of both the objective functions. The pink dot highlights the best trial which has been chosen as working point: it simultaneously provides a high accuracy value (0.75) and a low K-S test (0.03). The numerical values of the hyperparameters corresponding to the best trial are listed in Table 4.5.

Table 4.5: *Hyperparameters numerical values associated to the chosen working point.*

Hyperparameter	Value
α	46.2
n_{nodes}	94
n_l^C	5
n_l^A	10
η^C	0.0003
η^A	0.0016

4.3.3 Performances

As a first step after the optimization, the ANN has been implemented and trained using the optimized values of the hyperparameters. Given that the kinematic properties of signal and background events depend on the $\Delta\Phi_{jj}$ value, in order to add more flexibility to our model and improve the bias reduction, the optimal approach was found to be the implementation of two ANNs. Each ANN has been trained on events with different $\Delta\Phi_{jj}$ ranges as explained below.
The first network is trained on events with

$$|\Delta\Phi_{jj}| < \pi/2 \tag{4.9}$$

while the second one learns on events with

$$\pi/2 < |\Delta\Phi_{jj}| < \pi \ . \tag{4.10}$$

Requirements (4.9) and (4.10) are added to selection criteria listed in Table 4.3 of the first and second ANN respectively. With reference to the event categorization Tables 4.1 and 4.2, this means training the first ANN on events belonging to the $\Delta\Phi_{jj}$ bins 1 and 2, while the second one on events coming from bins 0 and 3 of the SR.
The training samples of both the ANNs are composed by about 85000 events, equally divided between the three considered processes, i.e. VBF, ggH and background events. The VBF signal sample is in turn made up of events coming from the three possible domains in equal proportions. As already mentioned, the domains correspond to the physical models according to which the signal events have been generated, i.e. the SM, H0M and H0PH hypothesis. Instead, the sample of background events is made up half by non-resonant WW and half of top quark events.
The learnings processes have been carried out with 2000 epochs each and

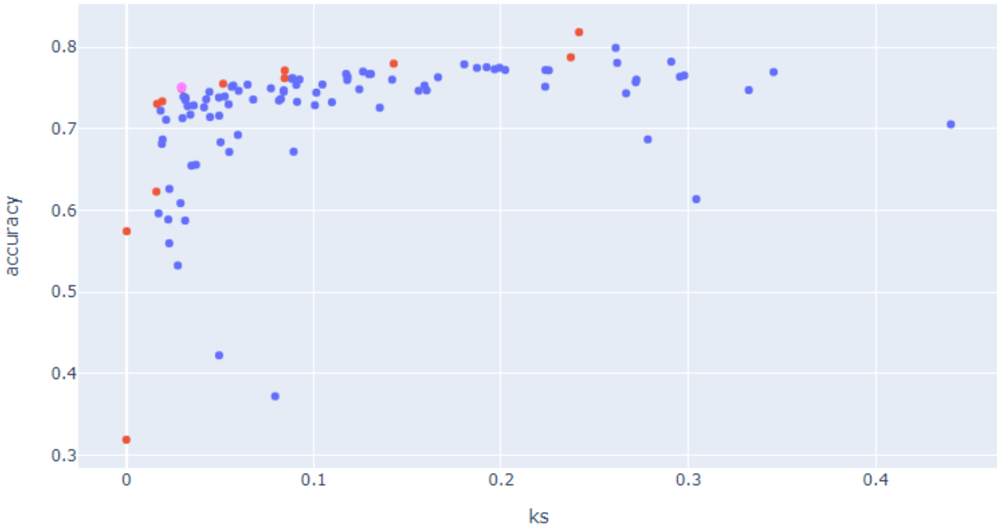


Figure 4.3: *Scatter plot of the accuracy (y axis) and K-S test (x axis) values obtained from 100 training trials: each blue dot represents a trial while the red ones correspond to the best attemps. The pink dot highlights the chosen working point.*

with a batch size equal to the entire training sample.
Figure 4.4 shows the loss function of the adversary and the combined loss function as a function of the number of epochs of the ANN trained in $\Delta\Phi_{jj}$ bins 1 and 2.

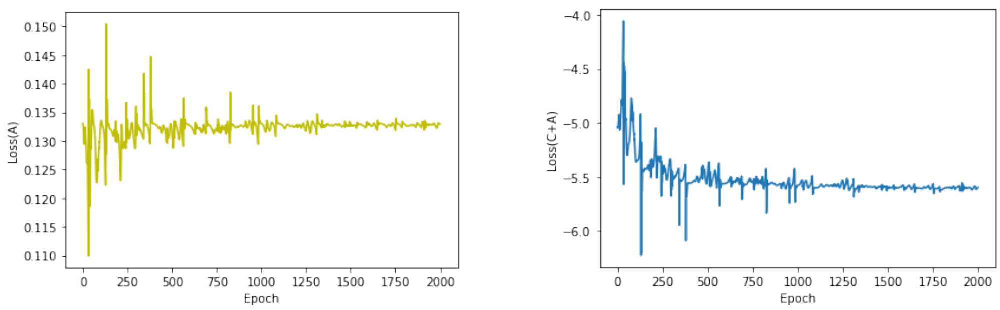


Figure 4.4: *Loss function of the adversary (left) and combined loss function (right) of the ANN trained on events with* $|\Delta\Phi_{jj}| < \pi/2$.

In order to check the effectiveness of the procedure, the distributions of the VBF-output for SM, H0M and H0PH events have been compared separately for each of the two ANNs. As shown in Figure 4.5, the distributions are consistent within the statistical uncertainties, meaning that the classifier is not

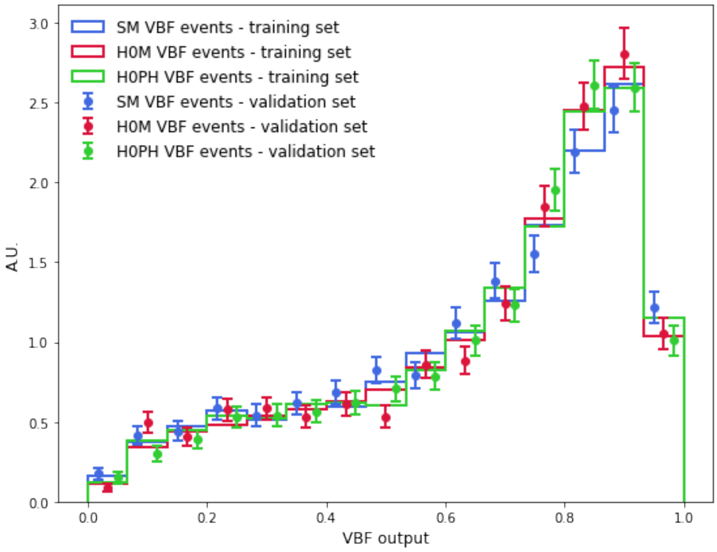


Figure 4.5: *Performance of the ANN trained on events with* $|\Delta\Phi_{jj}| < \pi/2$*: Normalized distributions of the* VBF*-output on* VBF *events simulated under the SM (blue), H0M (red) and H0PH (green) hypotheses. The solid lines correspond to the distributions obtained on training sample, while the dots with statistical error bar stand for validation sample.*

able to recognize the domain of origin, i.e. the signal model, of the VBF events.

In order to check the overtraining of the algorithm, the three distributions obtained from the validation sample have been compared to the ones obtained in the training. The results are shown in Figure 4.5: the normalized distributions evaluated on the training samples are drawn with solid colored lines, while the corresponding distributions of the validation sample are represented using colored dots with statistical error bars. Since the shapes of the training distributions are compatible with the validation ones within the statistical acccuracy, the algorithm does not overtrain the training sample.

The signal to background discriminating power of the ANN is highlighted in Figure 4.6. The VBF-output exhibits the expected behavior: its distribution has a maximum near to zero on background and ggH events, while it peaks at values close to one on signal events. The performances on the validation sample are also taken into consideration and are depicted superimposed to the training ones using dots with statistical error bars.

The classification accuracy is quantified by the associated confusion matrix, pictured in Figure 4.7. The matrix shows that 80%, 70% and 83% of the VBF, ggH and background events respectively, are correctly categorized

into the corresponding class.
The same set of plots describing the performances of the ANN trained on events with $\pi/2 < |\Delta\Phi_{jj}| < \pi$ can be found in appendix A.

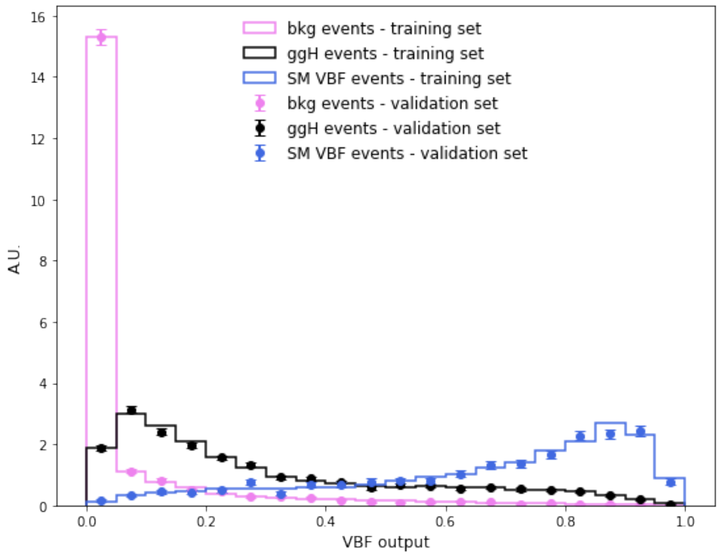


Figure 4.6: *Performance of the ANN trained on events with $|\Delta\Phi_{jj}| < \pi/2$: Normalized distributions of the VBF-output on SM signal (blue), ggH (black) and background (pink) events. The solid lines correspond to the distributions obtained on training sample, while the dots with statistical error bar stand for validation sample.*

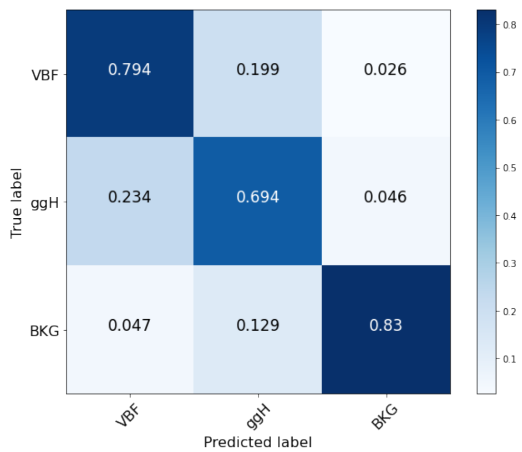


Figure 4.7: *Confusion matrix of the ANN trained on events with $|\Delta\Phi_{jj}| < \pi/2$.*

4.4 Fiducial phase space and generator level binning

The Higgs boson cross section is measured within the so-called fiducial volume, i.e. a phase space at particle level whose definition is chosen in order to minimize the dependence of the results on the underlying signal model of the Higgs boson production. Moreover, limiting the measurement to a restricted fiducial region allows to avoid large theoretical uncertainties arising from the extrapolation of the result from the experimentally accessible kinematic regions to the full phase space.
The fiducial volume is constructed through the kinematic requirements summarized in Table 4.6, defined using only particle level quantities obtained from the MC generator before the simulation of the CMS detector.

Table 4.6: *Definition of the fiducial volume.*

Fiducial selection
oppositely-charged $e\mu$ final state, at least two jets with $p_T > 30$ GeV, $p_T^{\ell_1} > 25$ GeV, $p_T^{\ell_2} > 13$ GeV, $p_T^{\ell_3} < 10$ GeV, $p_T^{\ell\ell} > 30$ GeV, $m_{\ell\ell} > 12$ GeV, $m_{jj} > 120$ GeV, $m_T^H > 60$ GeV, $\lvert\eta_{j_1}\rvert < 4.7$, $\lvert\eta_{j_2}\rvert < 4.7$, $\lvert\eta_{\ell_1}\rvert < 2.5$, $\lvert\eta_{\ell_2}\rvert < 2.5$

The selection mimics the experimental one described in previous sections, except for the absence of the E_T^{miss} and $m_T^{\ell_2}$ cuts. Moreover, the pseudorapidity absolute value of the charged leptons is requested to be smaller than 2.5, while the leading and subleading jet must have $|\eta_j| < 4.7$. The entire set of requirements is chosen in order to maximize two correlated quantities: the purity of the selected sample and the reconstruction efficiency. The former corresponds to the number of signal events coming from the fiducial volume and passing the analysis event selection after the detector simulation and reconstruction (called 'reco' events for brevity) over the total number of signal events selected by the analysis requirements:

$$p = \frac{N^{reco\ \&fid}}{N^{reco}} \ , \tag{4.11}$$

while the latter is equal to the fraction of signal events entering the fiducial region that pass the analysis selection:

$$\epsilon = \frac{N^{reco\ \&fid}}{N^{fid}} \ . \tag{4.12}$$

The kinematic requirements listed in Table 4.6 correspond to $p = 74\%$ and $\epsilon = 64\%$. Generally speaking, an event satisfying the fiducial requirements

is said to be fiducial; otherwise it is an out-of-fiducial event.
In order to obtain an unfolded result, the VBF signal samples are subdivided, based on generator level quantities, into 4 bins of $\Delta\Phi_{jj}$, using the same bin boundaries as for the definition of the analysis categories described in section 4.1. In this way the unfolded (gen level) and folded (reco level) distributions have the same binning. In addition, I decided to apply the same binning also to out-of-fiducial signal events that fall into the reco level categories. Therefore, the signal sample has four fiducial and four out-of-fiducial $\Delta\Phi_{jj}$ bins at generator level. Same division is applied also to the ggH sample.
The response matrix for both the fiducial and out-of-fiducial parts of the SM VBF signal is reported in Figure 4.8. Each column is normalized to 1.

4.5 Signal regions

Once the networks have been trained and the model parameters fixed to their best values, the ANNs have been tested on the events selected by the analysis requirements. Figure 4.9 shows the distributions of the outputs of the ANN trained on events with $|\Delta\Phi_{jj}| < \pi/2$, evaluated in the SR. To maximize the analysis sensitivity, instead of a single SR, a finer event categorization has been designed. By looking at Figure 4.9, it was natural to use the ANNs scores in order to divide the inclusive SR into three signal sub-region: VBF-, ggH- and BKG-like SRs. Each of these is defined by requiring the corresponding ouput to be higher than the other ones. For instance, the VBF-like SR selects events for which the VBF-output is greater than ggH- and BKG-scores; likewise, other categories are defined as illustrated in Table 4.7.

Table 4.7: *Definition of signal regions using the outputs of the classifier.*

Signal Regions		
VBF-like	**ggH-like**	**BKG-like**
$\text{output}_{VBF} > \text{output}_i$ $i = \{ggH, BKG\}$	$\text{output}_{ggH} > \text{output}_i$ $i = \{BKG, VBF\}$	$\text{output}_{BKG} > \text{output}_i$ $i = \{VBF, ggH\}$

All SRs are in turn split into four $\Delta\Phi_{jj}$ bins, similarly to what is prescribed in Table 4.2. For the definitions of bins 1 and 2 of all SRs, the ANN trained on events with $|\Delta\Phi_{jj}| < \pi/2$ has been used, while in bins 1 and 3 the ANN trained on events with $\pi/2 < |\Delta\Phi_{jj}| < \pi$ has been employed.
In the following, the distributions of the VBF-, ggH- and BKG-outputs in the $\Delta\Phi_{jj}$ bins of the corresponding SR are reported in logarithmic scale. According to the blinding policy of the CMS collaboration, data are not

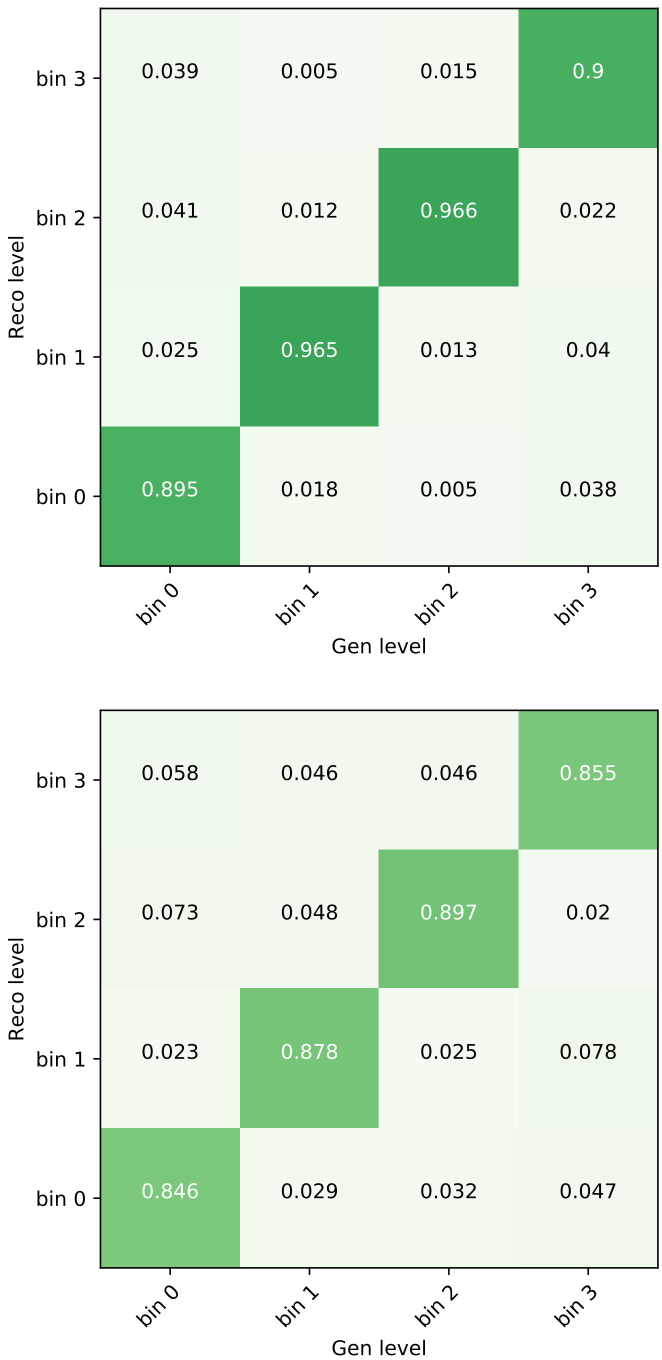
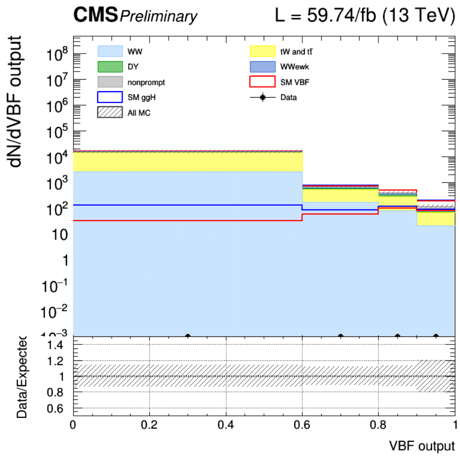
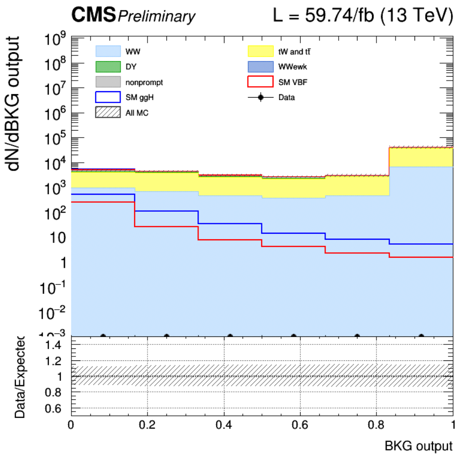


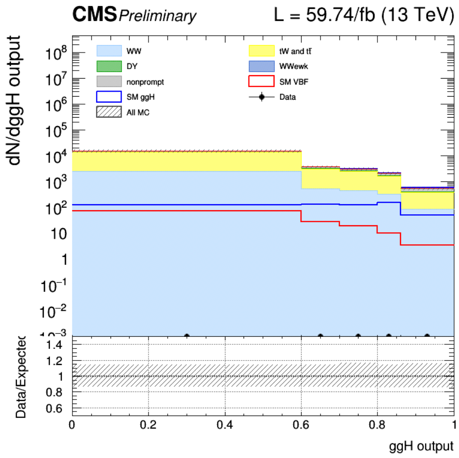
Figure 4.8: *Response matrix of the fiducial (above) and out-of-fiducial (below) components of the VBF signal sample.*



(a) *VBF-output.*



(b) *BKG-output.*

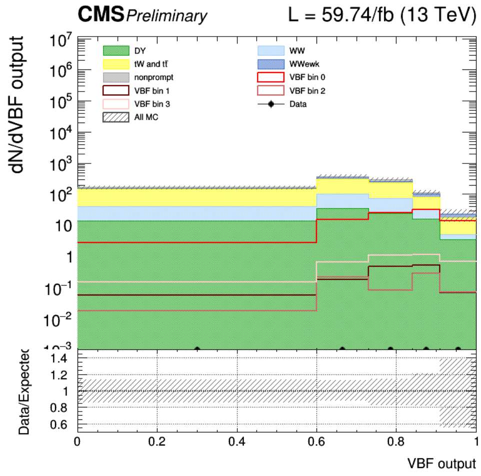


(c) *ggH-output.*

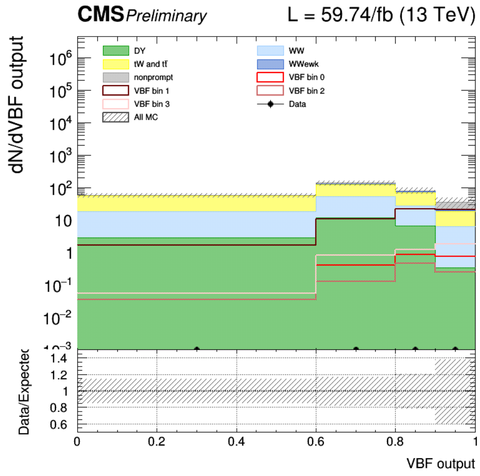
Figure 4.9: *Distributions of the VBF-, ggH- and BKG-outputs of the ANN trained on events with* $|\Delta\Phi_{jj}| < \pi/2$ *evaluated in the SR. The VBF (red line) and ggH (blue line) signal contributions are shown both stacked on top of the background and superimposed on it. A logarithmic scale is used. The dashed lines represent both MC statistical and systematic uncertainties.*

shown. For the VBF-like and ggH-like regions, the binning of the histogram is chosen in order to have at least two expected signal events and at least five expected signal and background events in each bin, while keeping the statistical uncertainty on MC simulations smaller than 30%. On the other hand, for the BKG-like SR a uniform splitting into six bins is adopted.

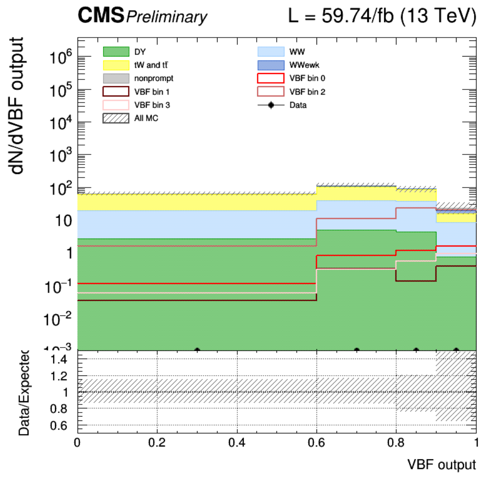
VBF-like SR



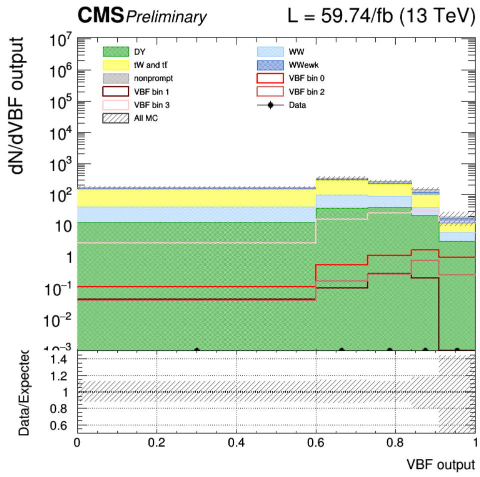
(a) $-\pi < \Delta\Phi_{jj} < -\frac{\pi}{2}$



(b) $-\frac{\pi}{2} < \Delta\Phi_{jj} < 0$.



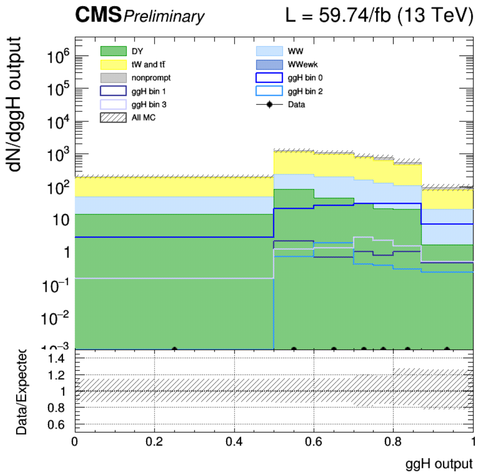
(c) $0 < \Delta\Phi_{jj} < \frac{\pi}{2}$.



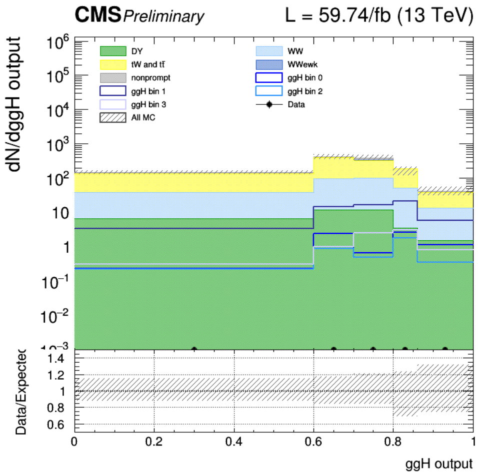
(d) $\frac{\pi}{2} < \Delta\Phi_{jj} < \pi$

Figure 4.10: *Distributions of the VBF-output on VBF signal events evaluated in all $\Delta\Phi_{jj}$ bins of the VBF-like SR. The fiducial and out-of-fiducial components of the signal belonging to the same $\Delta\Phi_{jj}$ bin are drawn together, superimposed on the backgrounds. Dashed lines represent both MC statistical and systematic uncertainties.*

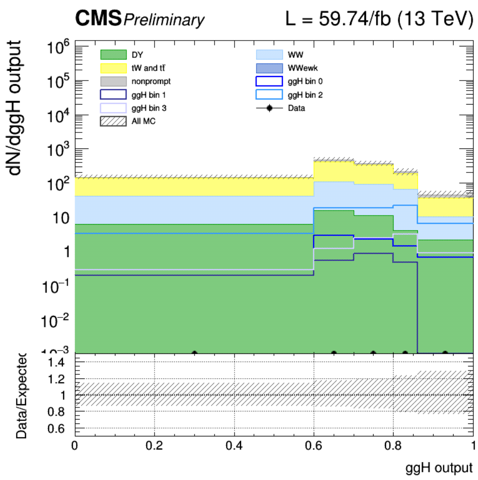
ggH-like SR



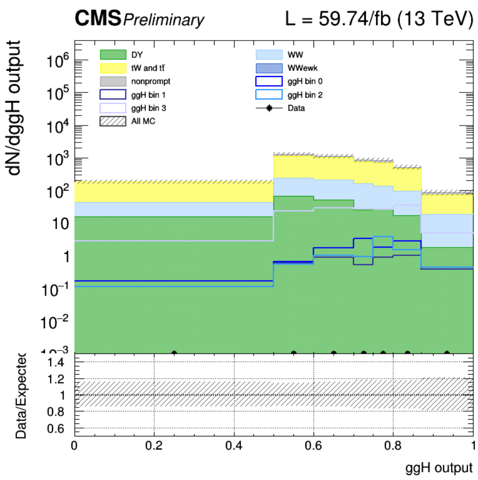
(a) $-\pi < \Delta\Phi_{jj} < -\frac{\pi}{2}$



(b) $-\frac{\pi}{2} < \Delta\Phi_{jj} < 0$.



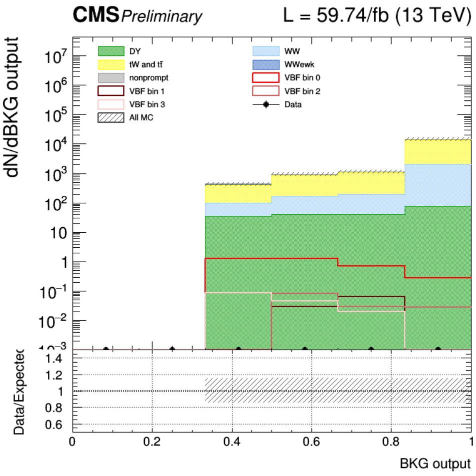
(c) $0 < \Delta\Phi_{jj} < \frac{\pi}{2}$.



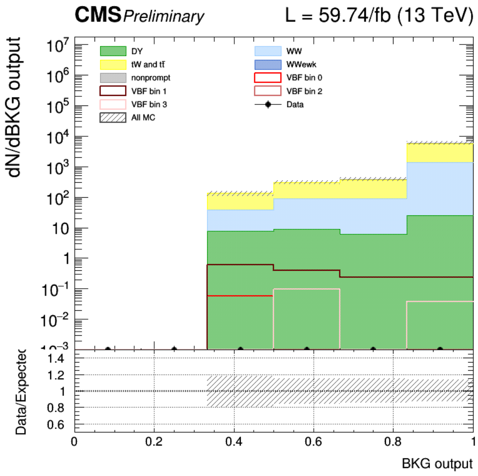
(d) $\frac{\pi}{2} < \Delta\Phi_{jj} < \pi$

Figure 4.11: *Distributions of the ggH-output on ggH events evaluated in all $\Delta\Phi_{jj}$ bins of the ggH-like SR. The fiducial and out-of-fiducial components of the ggH sample belonging to the same $\Delta\Phi_{jj}$ bin are drawn together, superimposed on the backgrounds. Dashed lines represent both MC statistical and systematic uncertainties.*

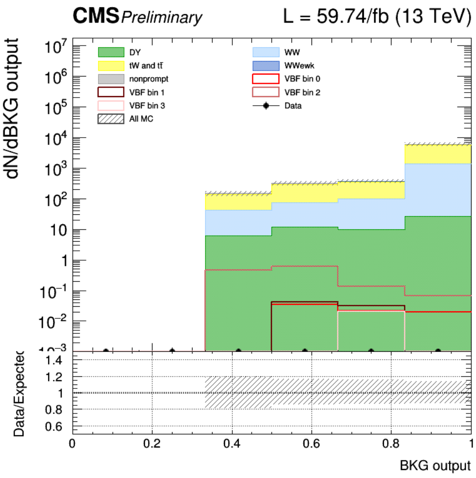
BKG-like SR



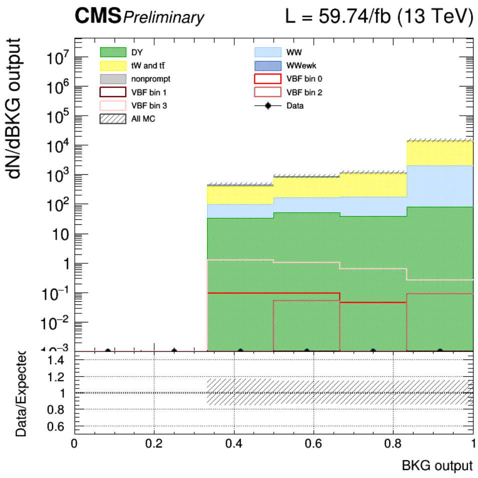
(a) $-\pi < \Delta\Phi_{jj} < -\frac{\pi}{2}$



(b) $-\frac{\pi}{2} < \Delta\Phi_{jj} < 0$.



(c) $0 < \Delta\Phi_{jj} < \frac{\pi}{2}$.



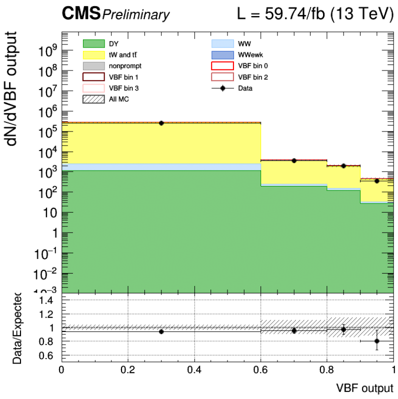
(d) $\frac{\pi}{2} < \Delta\Phi_{jj} < \pi$

Figure 4.12: *Distributions of the BKG-output on VBF events evaluated in all $\Delta\Phi_{jj}$ bins of the BKG-like SR. The fiducial and out-of-fiducial components of the signal belonging to the same $\Delta\Phi_{jj}$ bin are drawn together, superimposed on the backgrounds. Dashed lines represent both MC statistical and systematic uncertainties.*

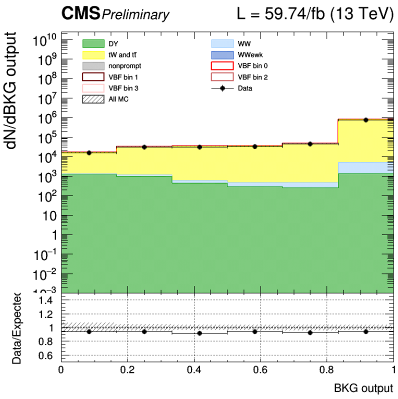
4.6 Control regions

Top CR

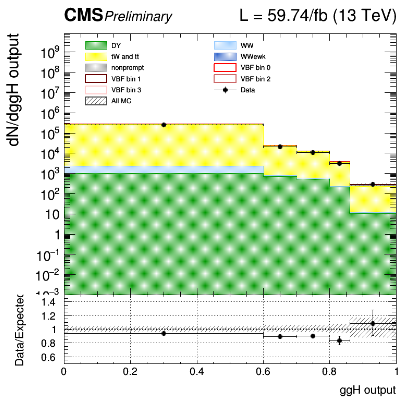
The Top CR is enriched with top quark production. In addition to the preselection criteria, it is defined by requiring at least one of the two jets in the final state tagged as a b-jet. Moreover, only events with $m_{\ell\ell}$ and $m_T^{\ell_2}$ higher than 50 and 30 GeV respectively are considered. Figure 4.13 shows a good data/MC agreement for the VBF-, ggH- and BKG-outputs of the ANN trained on events with $|\Delta\Phi_{jj}| < \pi/2$, meaning that the behaviour of these variables is well predicted by simulations.



(a) VBF-output.



(b) BKG-output.

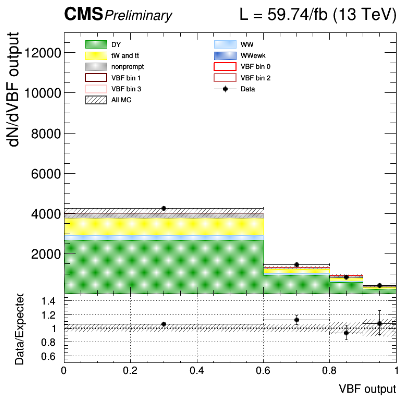


(c) ggH-output.

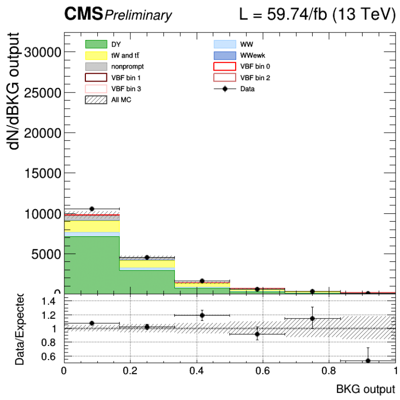
Figure 4.13: *Distributions of the VBF-, ggH- and BKG-outputs evaluated in the Top CR, shown in logarithmic scale. Error bars are associated to statistical uncertainties on data, while dashed lines represent both MC statistical and systematic uncertainties.*

$DY \to \tau\tau$ CR

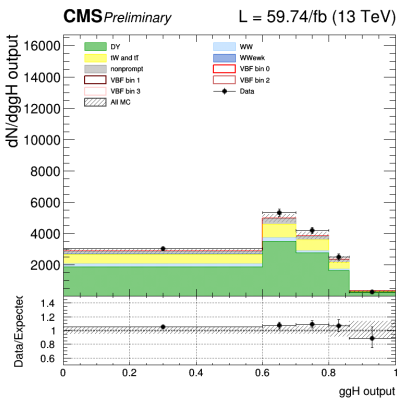
The $DY \to \tau\tau$ CR is expected to maximize the sensitivity to $DY \to \tau\tau$ processes. It is constructed through the preselection criteria and by further requiring the invariant mass of the dilepton system to be between 40 and 80 GeV, m_T^H lower than 60 GeV and no b-tagged jet with $p_T > 20$ GeV. A good agreement between data and MC is achieved even in this region for all the outputs, as shown in Figure 4.14. More plots of the ANN trained on events with $\pi/2 < |\Delta\Phi_{jj}| < \pi$ are available in appendix A.



(a) *VBF-output.*



(b) *BKG-output.*



(c) *ggH-output.*

Figure 4.14: *Distributions of the VBF-, ggH- and BKG-outputs evaluated in the $DY \to \tau\tau$ CR. Error bars are associated to statistical uncertainties on data, while dashed lines represent both MC statistical and systematic uncertainties.*

4.7 Signal extraction

The signal extraction is perfomed by employing the Maximum Likelihood (ML) method [25], which is a statistical tool for estimating the parameters of a model given a finite sample of data.
Suppose to have n indipendent measurements $(x_1, ..., x_n)$ of a random variable x, which is distributed according to a probability density function (p.d.f.) $f(x, \boldsymbol{\theta})$ depending on a set of parameters $\boldsymbol{\theta}$. The likelihood function is constructed as

$$L(\boldsymbol{\theta}) = \prod_{i=1}^{n} f(x_i; \boldsymbol{\theta}) \tag{4.13}$$

and it is a function only of the $\boldsymbol{\theta}$ parameters.
The ML method consists in selecting the values of the model parameters that maximize the likelihood function. This intuitively corresponds to maximize the agreement of the considered model with the observed data. Therefore, the ML estimators $\hat{\boldsymbol{\theta}}$ of the true parameters $\boldsymbol{\theta}$ are defined by the following condition:

$$\max_{\theta} L(\boldsymbol{\theta}) \equiv L(\hat{\boldsymbol{\theta}}) \ . \tag{4.14}$$

However, when the functional form of the probability density function is not analytically known, a binned ML fit is used: the unknown p.d.f. is replaced with histograms of the random variable x representing each signal and background contribution obtained from MC simulations. These histograms are generally called *templates*.

Consider a histogram containing N bins and let s_i and b_i be the expected signal and background yields in the i-th bin, respectively. The number of observed data $N_{i,obs}$ in bin i is expected to have a Poisson distribution with mean value $s_i + b_i$. To check the agreement between the measured number of signal events and the prediction of the considered physics model, the *signal strength* parameter μ is usually introduced. It corresponds to the ratio between the measured signal cross section and the expected one, i.e.

$$\sigma = \mu\ \sigma^{exp} \tag{4.15}$$

Thus, $N_{i,obs}$ follow a Poisson distribution with mean value $(\mu s_i + b_i)$ and the likelihood function is:

$$L(\mu) = \prod_{i=i}^{N} \frac{(\mu s_i + b_i)^{N_{i,obs}}}{N_{i,obs}!} e^{-(\mu s_i + b_i)} \equiv \prod_{i=i}^{N} \mathcal{P}\left(N_{i,obs}\ ; \mu s_i + b_i\right) \tag{4.16}$$

Determing the value of the μ parameter that maximizes $L(\mu)$ corresponds to finding the ML estimator $\hat{\mu}$ for the true value of μ. This procedure is called *ML template fit*.

4.7.1 Nuisance parameters

To take any systematic error that affects the MC simulation into account, additional parameters are included in equation (4.16): the so-called *nuisance* parameters $\boldsymbol{\nu}$.
Each source of uncertainty is modeled as an individual parameter and the likelihood function is redefined including a constraint $\mathcal{N}(\boldsymbol{\nu})$, determined by the set of nuisance parameters:

$$L(\mu, \boldsymbol{\nu}) = \prod_{i=i}^{N} \mathcal{P}\left(N_{i,obs}\ ; \mu s_i(\boldsymbol{\nu}) + b_i(\boldsymbol{\nu})\right) \cdot \mathcal{N}(\boldsymbol{\nu}) \tag{4.17}$$

All ν parameters are determined through the fit procedure described above but they are not relevant for the analysis purpose, in contrary of the μ parameter which is referred to as *parameter of interest* (POI).
The systematic uncertainties are divided into two categories, according to the effect they have on the templates:

- *normalization uncertainties*: they vary the overall normalization of the template such as, for instance, the uncertainties on the normalization of background processes. Each of these is modeled by introducing a multiplicative factor common to all bins of a given sample and constraining it with a log-normal prior distribution centered around the expected value and with a standard deviation corresponding to the given normalization untertainty. The log-normal functional form is chosen instead of normal distribution because it is suitable to describe parameters that can not fluctuate below zero.

- *shape uncertainties*: they modify the shape of signal and background templates; they can not be estimated using a single flat distribution since they alter the expected number of events bin by bin. For each uncertainty, three histograms with same normalization have to be considered: one with the nominal value and other two with $+1\sigma$ and -1σ variations of it.

All the MC corrections listed in section 3.4 are both of shape- and normalization-type except for the systematic uncertainty related to the luminosity measurement which has only a normalization effect.

4.7.2 Statistical treatment of background processes

Data are not only used to extract the POI but also to determine the normalization of backgrounds which have dedicated CRs. In fact, the normalization

of the major backgrounds is not fixed to the theoretical prediction but it is directly evaluated from data.
More precisely, each $\Delta\Phi_{jj}$ bin of the CRs is included in the likelihood function as an extra bin in order to quantify the number of events of the corresponding process. A additional parameter, called *rate parameter*, with a flat prior distribution is added to the yield of the bin as multiplicative factor. Thus, the term b_i in equation (4.17) can be expanded as:

$$b_i = k_{top} b_i^{top} + k_{DY} b_i^{DY} + k_{WW} b_i^{WW} + {b_i}^{other} \tag{4.18}$$

where ${b_i}^{other}$ stands for minor backgrounds processes, while k_{top} and k_{DY} are rate parameters associated to top and $DY \to \tau\tau$ events, respectively.
For a given $\Delta\Phi_{jj}$ bin, k_{top} and k_{DY} do not depend on the SR or CR to which the bin belongs, meaning that backgrounds in different categories float together. For instance, suppose that one rate parameter is associated to all contributions from top events in $\Delta\Phi_{jj}$ bin 1. This parameter will be determined primarily by data in the top CR, since this region is where most of the top events are present. However, its value will adjust the normalization of the top background in $\Delta\Phi_{jj}$ bin 1 of all categories, including the signal regions.
The normalization of the non-resonant WW process, which has no dedicated CR, is measured with the same procedure but it is primarily determined by the signal region itselft.

4.7.3 Unfolding procedure

Equation (4.17) has to be generalized to the case of study of this work, i.e. it must include all the generator level signal contributions defined in section 4.4, as well as the unfolding procedure.
The method I adopted to unfold the result avoids the explicit construction and inversion of the response matrix but rather consists in embedding it into the likelihood function. The key point of this procedure is to express the reconstructed level signal contributions as a function of generetor level quantities.
Recall that the generator level signal sample is binned into four fiducial and four out-of-fiducial $\Delta\Phi_{jj}$ bins.
Let s_j^{fid} and $s_j^{non\text{-}fid}$ with $j = 0, ..., 3$ be the contributions in the fiducial and out-of-fiducial j-th bin, respectively. The reconstructed level signal components s_i are therefore defined as:

$$s_i = R_{ij}^{fid}\ s_j^{fid} + R_{ij}^{non\text{-}fid}\ s_j^{non\text{-}fid} \tag{4.19}$$

where repeated indices are implicitly summed over, and R_{ij}^{fid} and $R_{ij}^{non\text{-}fid}$ are the response matrices of the fiducial and out-of-fiducial components, respectively. They are thus built implicitly by calculating contributions from each generator level sample in every reconstructed level bin.
A signal strength parameter μ_j is introduced for each bin j: the fiducial and out-of-fiducial components coming from the same generator level bin are scaled together within the maximum likelihood fit.
Each μ_j is equal to the ratio between the measured fiducial signal cross section and the expected one in the corresponding $\Delta\Phi_{jj}$ bin:

$$\mu_j = \frac{\sigma_j}{\sigma_j^{exp}} \tag{4.20}$$

Equation (4.19) is thus rewritten as:

$$s_i = R_{ij}^{fid}\ \mu_j\ s_j^{fid} + R_{ij}^{non\text{-}fid}\ \mu_j\ s_j^{non\text{-}fid} \tag{4.21}$$

Finally, substituting in equation (4.17), the likelihood used in this analysis is:

$$L(\boldsymbol{\mu}, \boldsymbol{\nu}) = \prod_{i=0}^{N_{bins}^{reco}} \mathcal{P}\left(N_{i,obs}\ ;\ \sum_{j=0}^{N_{bins}^{gen}} (R_{ij}^{fid}\mu_j s_j^{fid} + R_{ij}^{non\text{-}fid}\mu_j s_j^{non\text{-}fid}) + b_i\right) \cdot \mathcal{N}(\boldsymbol{\nu}) \tag{4.22}$$

The fit provide the complete set of $\boldsymbol{\mu} = \{\mu_0, \mu_1, \mu_2, \mu_3\}$, which corresponds to unfolded ratios between observed and expected yields per generator level bin.

The maximization of the likelihood function can in principle be very complex. Therefore, usually it is convenient to introduce the *negative log likelihood* NLL, defined as follows:

$$NLL \equiv -\log(L(\boldsymbol{\mu}, \boldsymbol{\nu})) \tag{4.23}$$

This function is thus minimized in order to find the estimators $\hat{\boldsymbol{\mu}}$ of the true values of $\boldsymbol{\mu}$.

Chapter 5

Results

5.1 Expected results

In this section the results of the analysis are presented. The work presented in this thesis is currently in progress and still covered by the blinding policy of the CMS collaboration at the time of this writing, therefore only expected results will be reported.
The procedure described in section 4.7 is applied to the so-called *Asimov data set*, which allows to estimate the signal sensitivity before looking at the real data. This is a pseudo-dataset constructed by generating the observed numbers of events in each bin such that they are equal to the sum of the expected signal and background events ($N_i^{asimov} = s_i + b_i$) and with a Poisson uncertainty ($\Delta N_i^{asimov} = \sqrt{s_i + b_i}$). This leads to deal with a pseudo data set matching exactly to the sum of all MC histrograms, in which the values of $\boldsymbol{\mu}$ returned by the template fit are equal to 1 by construction. Performing the fit on such data set is useful to estimate the uncertainties which will affect the measurement on true data.

As a first step, the entire fit procedure is carried out on the Asimov data set, corresponding to the assumption of SM cross sections. The likelihood function shown in equation (4.22) is built using the template distributions of the outputs of the classifier: each output has been evaluated in its respective SR. For instance, the VBF-output has been fitted in each $\Delta\Phi_{jj}$ bin of the VBF-like signal region; likewise, for the ggH- and BKG-outputs. As usual, the ANN trained on events with $|\Delta\Phi_{jj}| < \pi/2$ has been used in $\Delta\Phi_{jj}$ bins 1 and 2 while the ANN trained on events with $\pi/2 < |\Delta\Phi_{jj}| < \pi$ in bins 0 and 3. The VBF process is considered as the signal process.
The Table 5.1 summarizes the total errors of signal strength parameters, corresponding to 68% Confidence Level (CL) intervals.

Table 5.1: *Expected results for the total (statistical and systematic) and statistical errors of the signal strength parameters, evaluated on the Asimov data set corresponding to the SM hypothesis.*

Estimator	Total errors (68% CL)	Statistical errors (68% CL)
$\hat{\mu}_0$	$-1.0/+1.1$	$-0.9/+1.0$
$\hat{\mu}_1$	$-1.0/+1.1$	$-0.8/+1.0$
$\hat{\mu}_2$	$-0.9/+1.0$	$-0.8/+0.9$
$\hat{\mu}_3$	$-1.1/+1.2$	$-0.9/+1.0$

The fiducial cross section of the j-th bin is than measured by multiplying the expected value (σ_j^{fid}) with the estimator $\hat{\mu}_j$: $\hat{\sigma}_j^{fid} = \sigma_j^{fid} \cdot \hat{\mu}_j$
The expected fiducial cross section σ_j^{fid} has been estimated as the product of the inclusive cross section with the acceptance related to the j-th bin of the fiducial region. The acceptance of the j-th bin is defined as the fraction of total events that enters the j-th bin of the fiducial volume. The measured fiducial cross sections are summarized in Table 5.2.

Table 5.2: *Fiducial cross section measured on the Asimov data set.*

j	$\hat{\sigma}_j^{fid}$ [fb]
0	$0.8^{+0.9}_{-0.8}$
1	$0.4^{+0.4}_{-0.4}$
2	$0.4^{+0.4}_{-0.4}$
3	$0.8^{+1.0}_{-0.8}$

The main contribution to the total error on the signal strength parameters is the statistical one. This can be checked by repeating the fit removing all the nuisance parameters that account for the systematic uncertainties. The results obtained with this procedure are reported in Table 5.1: they correspond only to the statistical contribution to the total error on $\hat{\mu}_j$ which is evidently dominant with respect to the stystematic one.
The fact that the statistical error is the dominant component is expected since the VBF is an extremely rare process and the measurement is performed using only the 2018 data set (59.74 fb^{-1}). Future plans include to extend the differential analysis to all the data collected during Run 2 (138 fb^{-1}), therefore reducing the expected uncertainty rouglhy by a factor $\sqrt{2.3} \sim 1.5$.
Finally, the template fit has been repeated considering both the gluon-gluon fusion and the vector boson fusion production mechanism as the signal pro-

cess of the analysis. This is technically implemented in the likelihood function by scaling both ggH and VBF signal yields with the same signal strength parameters. Table 5.3 shows the expected errors of the estimator $\hat{\mu}_j$ for each $\Delta\Phi_{jj}$ bin, evaluated on the Asimov data set. As expected, the errors are lower than the ones shown in Table 5.1, since now the amount of expected signal events is larger than in the previous case.

Table 5.3: *Expected results for the total errors of the signal strength parameters, evaluated on the Asimov data set corresponding to the SM hypothesis and considering both the gluon-gluon fusion and the vector boson fusion as the signal process.*

Estimator	Total errors (68% CL)
$\hat{\mu}_0$	$-0.8/+0.9$
$\hat{\mu}_1$	$-0.7/+0.8$
$\hat{\mu}_2$	$-0.7/+0.8$
$\hat{\mu}_3$	$-0.8/+0.9$

5.2 Model dependence studies

The second part of this work focuses on quantifying the model dependence of the analysis results. As already described in section 3.7, the main source lies on the signal extraction procedure because the shape of the discriminating variable used for the template fit generally depends on the underlying physics hypothesis assumed to generate the histograms. Therefore, this inevitably leads to a bias on the fit results, whose quantification is not trivial since it depends on both the unknown true model and the one assumed to simulate the templates.

The implementation of the Adversarial Neural Network described in previous chapter allows to take this behavior under control, providing a fit variable whose shape is the same whatever the true physics model describing the data is.

The interest is now focalized on the bias estimation, in order to quantify the reduction made possible by the exploitation of the ANN. This goal is achieved by comparing the bias values obtained in this work with those estimated by replacing the ANN with a simple Neural Network: the entire analysis procedure has been repeated employing for the event categorization and for the template fit a Deep Neural Network (DNN) developed by the CMS collaboration for another measurement which is still ongoing. This network has

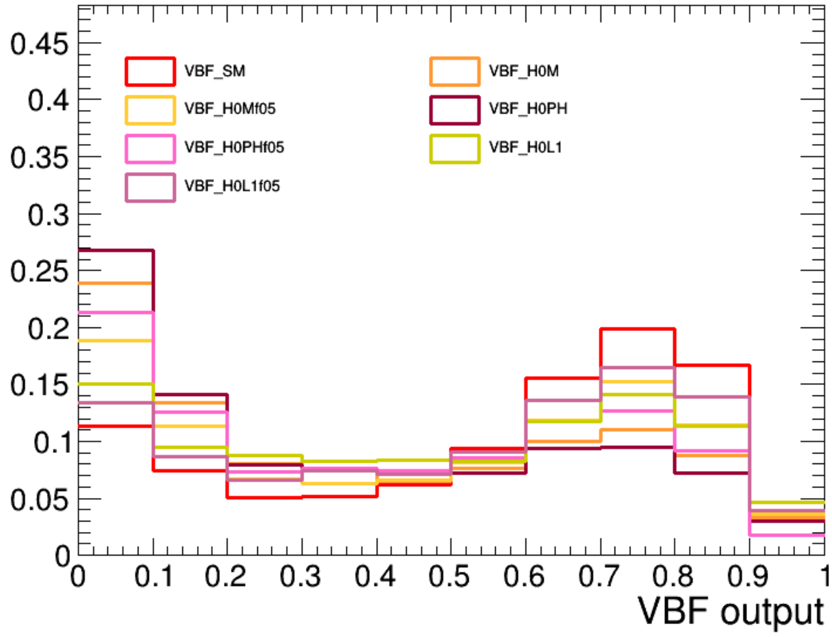


Figure 5.1: *Normalized distribution of the VBF-output of the DNN on VBF signal events generated assuming the SM and several BSM hypothesis.*

the classical structure of a NN for events classification illustrated in section 4.2, but without the adversarial term. The classification is performed taking the same kinematic distributions of the ANN as input variables. Since the adversarial term is missing, the DNN performance strongly depends on the physics model of signal events. This behavior is shown in Figure 5.1, where the normalized distribution of the VBF-output on SM signal events is pictured togheter with the ones obtained on signal events generated according to the BSM theories introduced in section 1.5.

5.2.1 Bias estimation

The bias estimation has been perfomed through the procedure described below.

An Asimov pseudo-dataset is generated assuming one of the considered signal BSM hypothesis and all background contributions. This dataset is then fitted using the SM signal and background MC templates. For the j-th $\Delta\Phi_{jj}$ bin, the fit provides the estimator $\hat{\mu}_j$ which is equal to the number of BSM signal events fitted in the j-th bin over the number of SM signal events expected in the same bin:

$$\hat{\mu}_j = \frac{s^{fit}_{BSM,j}}{s^{exp}_{SM,j}} \tag{5.1}$$

In this particular case of study, $\hat{\mu}_j$ does not correspond to the ratio between the measured BSM fiducial cross section and the expected SM one, since the two models have different selection efficiency and acceptance factors. Let's then define $\tilde{\mu}_j$ as the ratio of the BSM and SM fiducial cross sections in bin j:

$$\tilde{\mu}_j = \frac{\sigma^{fid}_{BSM,j}}{\sigma^{fid}_{SM,j}} \tag{5.2}$$

Therefore, $\tilde{\mu}_j$ would be the signal strength that one would expect to measure if there was no bias in the procedure.
The *total bias* has been introduced as:

$$b_j = \hat{\mu}_j - \tilde{\mu}_j \tag{5.3}$$

This quantity corresponds to the most general bias that can be defined since it embeds all possible sources of model dependence that enter the analysis procedure. In fact, it includes both the contribution coming from the shape effects of the discriminating fit variable and the one due to the different acceptance and selection efficiencies.
However, the latter contribution can be taken into account and the total bias corrected for it.
For the sake of simplifying the notation, in the expressions below the j index, representing the j-th generator level bin, is omitted. The numerator and denominator of equation (5.1) can be rewritten as follows:

$$s^{fit}_{BSM} = \mathcal{L}\left(\epsilon_{BSM}\hat{\sigma}^{fid}_{BSM} + \bar{\epsilon}_{BSM}\hat{\sigma}^{non\text{-}fid}_{BSM}\right) \tag{5.4}$$

$$s^{exp}_{SM} = \mathcal{L}\left(\epsilon_{SM}\sigma^{fid}_{SM} + \bar{\epsilon}_{SM}\sigma^{non\text{-}fid}_{SM}\right) \tag{5.5}$$

where

- $\mathcal{L} = 59.74$ fb^{-1} is the integrated luminosity;
- $\epsilon_{(B)SM}$ is the reconstruction efficiency of the fiducial bin. It is defined as the efficiency of reconstructing fiducial (B)SM events belonging to one bin at generator level in any of the $\Delta\Phi_{jj}$ bins at reconstructed level;

- $\bar{\epsilon}_{(B)SM}$ is the reconstruction efficiency of the out-of-fiducial bin. It is defined as the efficiency of reconstructing out-of-fiducial (B)SM events belonging to one bin at generator level in any of the $\Delta\Phi_{jj}$ bins at reconstructed level;

- σ_{SM}^{fid} and $\sigma_{SM}^{non\text{-}fid}$ are the fiducial and out-of-fiducial production cross section respectively, predicted by the SM;

- $\hat{\sigma}_{BSM}^{fid}$ and $\hat{\sigma}_{BSM}^{non\text{-}fid}$ are the fiducial and out-of-fiducial production cross section under the BSM hyphothesis, which have been measured through the template fit.

By introducing the total cross section, defined as the sum of the fiducial and out-of-fiducial components of the (B)SM signals, equations (5.4) and (5.5) can be rewritten as:

$$s_{BSM}^{fit} = \mathcal{L} \cdot \left[\epsilon_{BSM} + \bar{\epsilon}_{BSM} \left(\frac{1 - A_{BSM}}{A_{BSM}} \right) \right] \cdot \hat{\sigma}_{BSM}^{fid} \tag{5.6}$$

$$s_{SM}^{exp} = \mathcal{L} \cdot \left[\epsilon_{SM} + \bar{\epsilon}_{SM} \left(\frac{1 - A_{SM}}{A_{SM}} \right) \right] \cdot \sigma_{SM}^{fid} \tag{5.7}$$

where:

- $\sigma_{(B)SM} = \sigma_{(B)SM}^{fid} + \sigma_{(B)SM}^{non\text{-}fid} +$ is total cross section predicted by the (B)SM hyphothesis;

- $A_{(B)SM} = \frac{\sigma_{(B)SM}^{fid}}{\sigma_{(B)SM}}$ is the ratio between the fiducial and total cross section predicted by the (B)SM hyphothesis.

For the sake of clarity, let θ_{BSM} and θ_{SM} be

$$\theta_{BSM} = \left[\epsilon_{BSM} + \bar{\epsilon}_{BSM} \left(\frac{1 - A_{BSM}}{A_{BSM}} \right) \right] \tag{5.8}$$

$$\theta_{SM} = \left[\epsilon_{SM} + \bar{\epsilon}_{SM} \left(\frac{1 - A_{SM}}{A_{SM}} \right) \right] \tag{5.9}$$

respectively. Equation (5.6) and (5.7) are therefore given by:

$$s_{BSM}^{fit} = \mathcal{L} \cdot \theta_{BSM} \cdot \hat{\sigma}_{BSM}^{fid} \tag{5.10}$$

$$s_{SM}^{exp} = \mathcal{L} \cdot \theta_{SM} \cdot \sigma_{SM}^{fid} \tag{5.11}$$

and equation (5.1) becomes:

$$\hat{\mu} = \frac{\theta_{BSM}}{\theta_{SM}} \cdot \frac{\hat{\sigma}^{fid}_{BSM}}{\sigma^{fid}_{SM}} \tag{5.12}$$

Therefore, the *residual bias* has been defined as:

$$b^{res} = \hat{\mu} \cdot \frac{\theta_{SM}}{\theta_{BSM}} - \tilde{\mu} = \frac{\hat{\sigma}^{fid}_{BSM} - \sigma^{fid}_{BSM}}{\sigma^{fid}_{SM}} \tag{5.13}$$

Since the θ_{SM} and θ_{BSM} factors take the acceptance effects into account, b^{res} is only due to the shape effects. Finally, the bias due to acceptance effects is quantified by introducing:

$$b^{acc} = b - b^{res} = \hat{\mu}\left(1 - \frac{\theta_{SM}}{\theta_{BSM}}\right) \tag{5.14}$$

The fiducial (B)SM cross section is given by the number of fiducial (B)SM events that enter in the considered bin ($N^{fid}_{(B)SM}$) times the integrated luminosity:

$$\sigma^{fid}_{(B)SM} = N^{fid}_{(B)SM} \cdot \mathcal{L} \tag{5.15}$$

Similarly, the out-of-fiducial (B)SM cross section is defined as:

$$\sigma^{non\text{-}fid}_{(B)SM} = N^{non\text{-}fid}_{(B)SM} \cdot \mathcal{L} \tag{5.16}$$

where $N^{non\text{-}fid}_{(B)SM}$ is the number of events that do not pass the fiducial selection but enter in considered $\Delta\Phi_{jj}$ bin. The signal events which have a number of jets with $p_T > 30$ GeV at generator level less than two, are not considered in the analysis.

5.2.2 Results

The total bias b, residual bias b^{res} and b^{acc} have been estimated for both the BSM samples used for the training of the ANN, i.e. H0M and H0PH models. In addition, the H0L1f05 model has been studied in order to test the ANN performance on events not employed for the learning procedure.
Table 5.4 summarizes the reconstruction efficiencies values, the number of fiducial and out-of-fiducial events and the ratio between the fiducial and total cross section for the SM, H0M and H0PH signal samples. The errors on the efficiencies have been calculated according to the Clopper-Pearson method [25].
The bias quantities estimated with the DNN and the ANN for both the H0M

Table 5.4: *Expected reconstruction efficiency, number of fiducial and out-of-fiducial events and ratio between the fiducial and total cross section of the SM, H0M and H0PH signal samples.*

SM					
$\Delta\Phi_{jj}$ bin	ϵ_{SM}	$\bar{\epsilon}_{SM}$	N_{SM}^{fid}	$N_{SM}^{non\text{-}fid}$	A_{SM}
0	0.696 ± 0.008	0.061 ± 0.002	48	765	0.058
1	0.730 ± 0.010	0.055 ± 0.002	25	442	0.054
2	0.716 ± 0.010	0.054 ± 0.002	26	442	0.055
3	0.696 ± 0.008	0.060 ± 0.002	47	767	0.058

H0M					
$\Delta\Phi_{jj}$ bin	ϵ_{BSM}	$\bar{\epsilon}_{BSM}$	N_{BSM}^{fid}	$N_{BSM}^{non\text{-}fid}$	A_{BSM}
0	0.749 ± 0.007	0.082 ± 0.002	240	4344	0.052
1	0.789 ± 0.008	0.065 ± 0.002	161	2838	0.054
2	0.780 ± 0.008	0.065 ± 0.002	160	2822	0.054
3	0.752 ± 0.007	0.084 ± 0.002	242	4339	0.053

H0PH					
$\Delta\Phi_{jj}$ bin	ϵ_{BSM}	$\bar{\epsilon}_{BSM}$	N_{BSM}^{fid}	$N_{BSM}^{non\text{-}fid}$	A_{BSM}
0	0.785 ± 0.006	0.082 ± 0.002	1070	14223	0.070
1	0.803 ± 0.008	0.065 ± 0.002	440	5998	0.068
2	0.811 ± 0.008	0.067 ± 0.002	446	6035	0.069
3	0.773 ± 0.006	0.086 ± 0.002	1068	14205	0.070

H0L1f05					
$\Delta\Phi_{jj}$ bin	ϵ_{BSM}	$\bar{\epsilon}_{BSM}$	N_{BSM}^{fid}	$N_{BSM}^{non\text{-}fid}$	A_{BSM}
0	0.742 ± 0.007	0.065 ± 0.002	145	2267	0.060
1	0.755 ± 0.010	0.060 ± 0.002	77	1165	0.062
2	0.756 ± 0.009	0.058 ± 0.002	77	1163	0.062
3	0.750 ± 0.003	0.070 ± 0.002	144	2259	0.062

and H0PH models are instead shown in Tables 5.5. As already said, these BSM models describe HVV couplings that are extremely different from the SM case and this is also outlined by the fact that the fit provides values of the $\hat{\mu}$ that are much higher than one.

The most important bias term to reduce is b^{res}, since it is the contribution to the total bias that accounts only for the shape effects of the fit variable by definition. In fact, b^{acc} is due to the fact that the SM and the BSM samples have different acceptance factors, as well as different events selection efficiencies. However, the former can be estimated a posteriori using only generator level quantities and thus, they can be accounted for. On the contrary, the latter depend on the event reconstruction and can not be quantified using only generator level quantities. Nevertheless, they do not strongly depend on the considered physics model, as shown in Table 5.5. Therefore, the acceptance factors are expected to be the dominant contribution to b^{acc}.
Performing the analysis with the DNN, the main contribution to the total bias b comes from the b^{res}. This behavior is expected since, as illustrated in Figure (5.1), distributions of the VBF output of the DNN on both H0M and H0PH signal events differ greatly from the one on SM signal events.
Moreover, this contribution goes in the opposite direction with respect to the one due to the acceptance effects.
On the other hand, using the ANN for the analysis procedure allows to strongly reduce the absolute value of the residual bias in each $\Delta\Phi_{jj}$ bin. This makes the main contribution to the total bias now be the one coming from acceptance effects which, as explained above, does not constitute a limiting factor for the analysis.
The situation is quite different for the H0L1f05 model, which has not been employed for the training procedure of the ANN. The reconstruction efficiencies and the number of fiducial and out-of-fiducial events are reported in Table 5.4, while the bias quantities are summarized in Table 5.5. In this case, the ANN does not provide the improvements seen for the other BSM samples. In fact, the absolute values of b^{res} remain high in all $\Delta\Phi_{jj}$ bins with respect to the ones obtained with the DNN and it is around the 20% of $\tilde{\mu}$.
In order to check if this behavior is due to the fact that the H0L1f05 sample is not used in the training procedure, I tried to re-train the ANN using all the BSM samples described in section 1.5. In this case the absolute value of b^{res} gets reduced to about 10% of $\tilde{\mu}$, confirming the expected improvement when the H0L1f05 model is also employed in the training procedure.

The residual bias can be used to quote an uncertainty on the measured signal cross section, which takes the model dependence of the result into account. This goal is pursued by computing the impact of the residual bias b^{res} on the expected signal strength $\tilde{\mu}$ as the percentage on the final result $\hat{\sigma}^{fid}$:

$$\Delta\sigma^{bias} = \frac{|b^{res}|}{\tilde{\mu}} \cdot \hat{\sigma}^{fid} \tag{5.17}$$

Table 5.5: *Resulting bias quantities evaluated on the Asimov data set corresponding to the H0M and H0PH hypothesis with both the DNN and the ANN.*

H0M

		DNN				ANN			
$\Delta\Phi_{jj}$ bin	$\tilde{\mu}$	$\hat{\mu}$	b	b^{res}	b^{acc}	$\hat{\mu}$	b	b^{res}	b^{acc}
0	5.0	3.6	$-1,5$	-2.3	0.9	7.1	2.1	0.4	1.7
1	6.4	4.1	-2.4	-2.9	0.5	6.9	0.4	-0.4	0.8
2	6.2	4.2	-2.1	-2.6	0.6	6.8	0.5	-0.4	0.9
3	5.1	3.7	-1.5	-2.3	0.9	7.3	2.1	0.4	1.7

H0PH

		DNN				ANN			
$\Delta\Phi_{jj}$ bin	$\tilde{\mu}$	$\hat{\mu}$	b	b^{res}	b^{acc}	$\hat{\mu}$	b	b^{res}	b^{acc}
0	22.4	10.5	-11.9	-13.0	1.1	27.3	4.9	2.01	2.9
1	17.5	11.6	-5.9	-5.9	0.0	16.8	-0.7	-0.8	0.1
2	17.4	12.2	-5.1	-6.4	1.2	18.1	0.7	-1.1	1.8
3	22.6	11.5	-11.1	-12.8	1.7	23.6	5.1	1.0	4.1

H0L1f05

		DNN				ANN			
$\Delta\Phi_{jj}$ bin	$\tilde{\mu}$	$\hat{\mu}$	b	b^{res}	b^{acc}	$\hat{\mu}$	b	b^{res}	b^{acc}
0	3.0	2.7	-0.3	-0.6	0.3	4.3	1.3	0.9	0.4
1	3.1	2.8	-0.3	-0.4	0.1	3.5	0.4	0.3	0.1
2	3.0	2.8	-0.2	-0.3	0.2	3.6	0.6	0.4	0.2
3	3.1	3.1	0.0	-0.3	0.3	4.6	1.5	1.1	0.4

The $\hat{\sigma}^{fid}$ measured values of each $\Delta\Phi_{jj}$ bin are listed in Table 5.2.
The $\Delta\sigma^{bias}$ uncertainty has been first calculated using the residual bias and the expected signal strength values obtained with the H0M pseduo data set; then with the ones corresponding to the H0PH data set. The results are reported in Table 5.6.

Table 5.6: *Model dependence uncertainty evaluated with both the ANN and the DNN.*

	ANN		*DNN*	
	H0M	**H0PH**	**H0M**	**H0PH**
$\Delta\Phi_{jj}$ bin	$\Delta\sigma^{bias}$[fb]	$\Delta\sigma^{bias}$[fb]	$\Delta\sigma^{bias}$[fb]	$\Delta\sigma^{bias}$[fb]
0	0.06	0.07	0.4	0.4
1	0.03	0.02	0.2	0.1
2	0.03	0.03	0.2	0.2
3	0.07	0.03	0.4	0.5

For each $\Delta\Phi_{jj}$ bin, the largest $\Delta\sigma^{bias}$ contribution among those obtained with the H0M and H0PH sample is chosen and quoted as model dependence uncertainty on the corresponding $\hat{\sigma}^{fid}$ value.
The same procedure has been repeated using the outputs of the DNN as fit variables. The measured fiducial cross sections are reported in Table 5.7 using both the ANN and the DNN, while the model dependence uncertainty estimated from both the H0M and H0PH samples are summarized in Table 5.6.

Table 5.7: *Fiducial cross section measured on the Asimov data set using the outputs of the ANN and the DNN as fit variables.*

	ANN	*DNN*
$\Delta\Phi_{jj}$ bin	$\hat{\sigma}^{fid}$ [fb]	$\hat{\sigma}^{fid}$ [fb]
0	$0.8^{+0.9}_{-0.8}$	$0.8^{+1.0}_{-0.9}$
1	$0.4^{+0.4}_{-0.4}$	$0.4^{+0.5}_{-0.5}$
2	$0.4^{+0.4}_{-0.4}$	$0.4^{+0.5}_{-0.5}$
3	$0.8^{+1.0}_{-0.8}$	$0.8^{+1.0}_{-0.9}$

Results are presented in the following. Figure 5.2 shows the fiducial cross section value measured in each $\Delta\Phi_{jj}$ bin with both the ANN and DNN: the solid black error lines stand for the total uncertainty (statistical and systematic) evaluated with the ANN, while the dashed ones are estimated with the DNN. Instead, the red and blue bars quantify the uncertainties due to the model dependence of the results related to the ANN and DNN usage, respectively.

The measured fiducial cross sections are the same in both cases. This is because the fit has been performed using the Asimov data set and thus the measured values depend only on the number of fiducial signal events expected in each $\Delta\Phi_{jj}$ bin. On the contrary, statistical and systematic errors are slightly different in the two cases since the variables used for the template fit change. In particular, the ANN provides smaller errors than those obtained with the DNN.

For what model dependence uncertainty is concerned, a sizeable improvement is observed when performing the analysis with the ANN. In fact, $\Delta\sigma^{bias}$ is an order of magnitude smaller than the value obtained with the DNN, for all the $\Delta\Phi_{jj}$ bins.

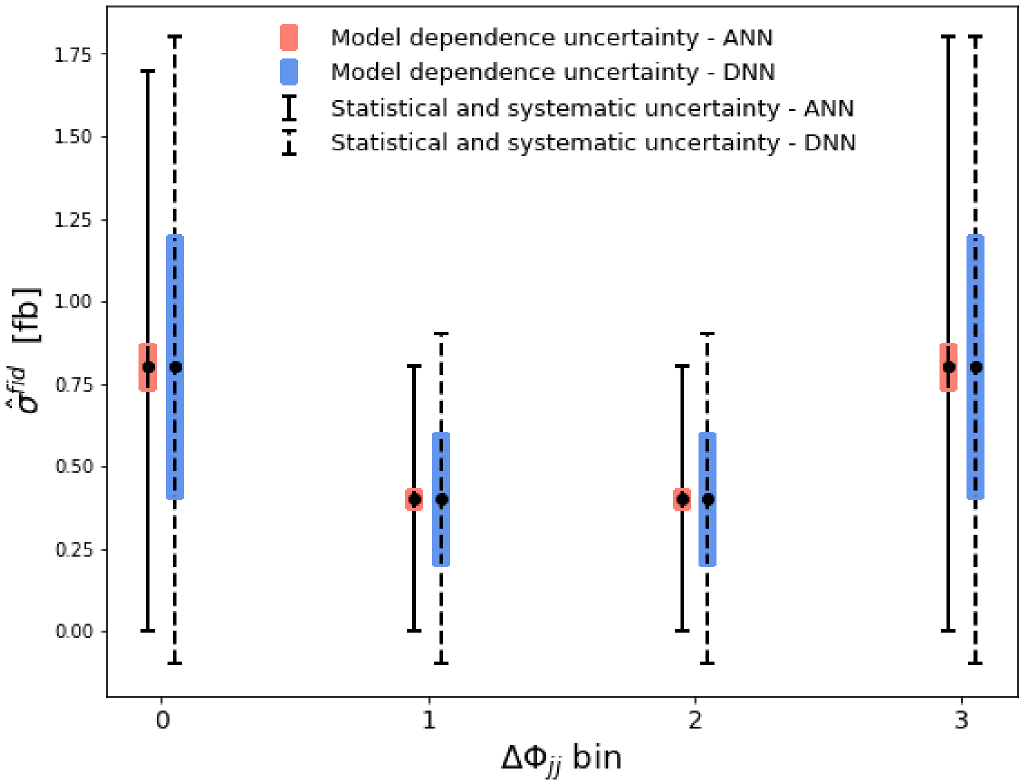


Figure 5.2: *Fiducial cross section measured in each $\Delta\Phi_{jj}$ bin using both the ANN and DNN. Solid black lines and red bars stand for the statistical and systematic errors and for the model dependence uncertainty estimated with the ANN; while dashed black lines and blue bars stand for the statistical and systematic errors and for the model dependence uncertainty estimated with the DNN.*

Conclusions

In this work, I presented a differential measurement of the production cross section of the Higgs boson as a function of the signed azimuthal angle difference between the two leading jets in the final state ($\Delta\Phi_{jj}$). The analysis has been performed on proton-proton collision data at $\sqrt{s} = 13$ TeV, collected by the CMS experiment during 2018. The VBF production mechanism has been chosen as the signal process, and the $H \to WW \to 2\ell 2\nu$ decay channel has been selected as the final state. The result is provided within a fiducial volume defined at particle level, constructed to minimize the extrapolation from the experimentally accessible region to the full phase space, that would otherwise introduce a model dependence uncertainty due to the underlying theoretical assumption. The signal has been extracted through a Maximum Likelihood template fit, which encodes the unfolding procedure, allowing to unfold the result to the generator level phase space, correcting it for the effects introduced by the presence of the detector.

In order to reduce as much as possible the introduction of a bias towards a certain physics model in the analysis result, I adopted the Domain Adaptation (DA) technique to build a discriminating variable that could distinghuish the signal from the main backgrounds and was simultaneously agnostic to different signal hypoteses.
Several BSM signal models have been considered. These models describe anomalous couplings between the Higgs boson and the vector bosons, reflecting on different tensor structures of the HVV vertex present in both the Higgs boson production mechanism and decay mode.
The implementation of DA I developed is based on an Adversarial Neural Network (ANN), which is a neural network system consisting of a classifier and an adversary trained in a competitive way. The former performs the actual discrimination between signal and background events, while the latter is designed to identify the physics model corresponding to a signal event. The performances of the two neural networks have been finely balanced in order to let the classifier being able to discriminate between signal and background events and prevent the adversary from guessing the domain. By doing so, I

obtained an ANN output that has good signal to background discriminating power and at the same time is signal model independent. This new discriminator has been used to build the templates for signal and background processes.

In order to check that the discriminator distribution is well modeled by the data, according to the blinding policy of the CMS collaboration, I have defined control regions orthogonal to the signal ones and I found a good agreement between the Monte Carlo simulations and data.

The expected uncertainty on the final result has been calculated on the Asimov data set. Since the analysis is limited by the limited number of data events, improvements in the uncertainty are expected when the analysis will be extended to the full data set collected during the Run 2 data taking period. In addition, a detailed study regarding the model dependence of the result has been carried out. In particular, the bias introduced on the final result by the fit procedure has been estimated and compared to the one obtained using a Deep Neural Network (DNN), previously developed by the CMS collaboration. The DNN is composed only by the classifier and, since the adversarial term is missing, its performance has shown a strong depencence on the considered signal model.
The ANN has allowed a sizeable reduction of the bias on the final result with respect to the case where the DNN is used for the analysis procedure.
Moreover, the bias has been used to quote an uncertainty on the measured signal cross section, which takes the model dependence of the result into account. The uncertainty provided with the ANN is an order of magnitude smaller than the value obtained with the DNN.

Appendix A

ANN performance

In this appendix, the performances of the ANN trained on events with $\pi/2 < |\Delta\phi_{jj}| < \pi$ are shown. In addition, the distributions of VBF-, ggH- and BKG-outputs in CRs are reported: a good data/Monte Carlo agreement is achieved for all of them.

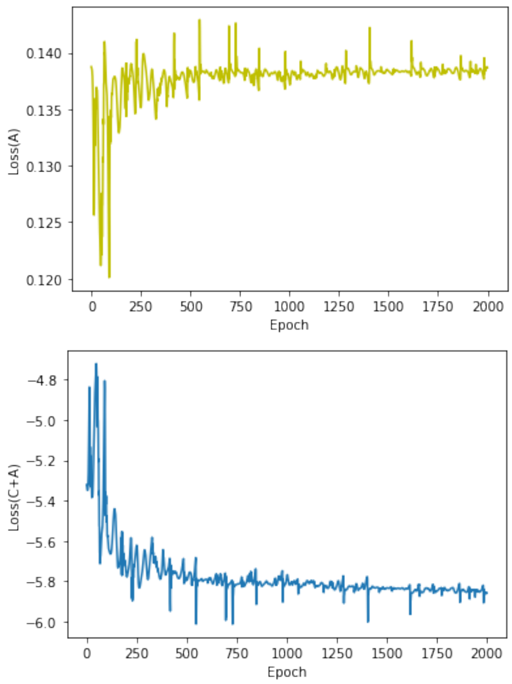


Figure A.1: *Loss function of the adversary (above) and combined loss function (below).*

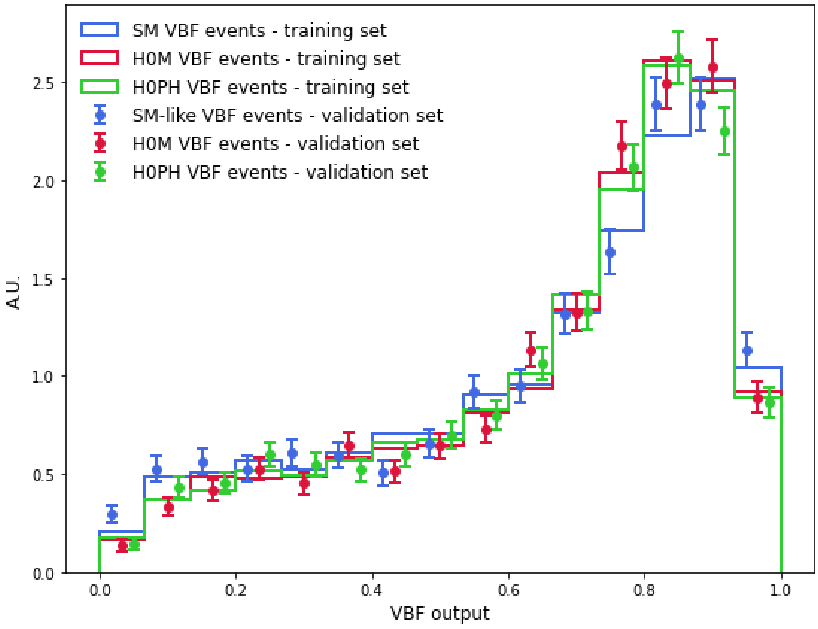


Figure A.2: *Normalized distributions of the VBF-output on VBF events simulated under the SM (blue), H0M (red) and H0PH (green) hypotheses. The solid lines correspond to the distributions obtained on training sample, while the dots with statistical error bar stand for validation sample.*

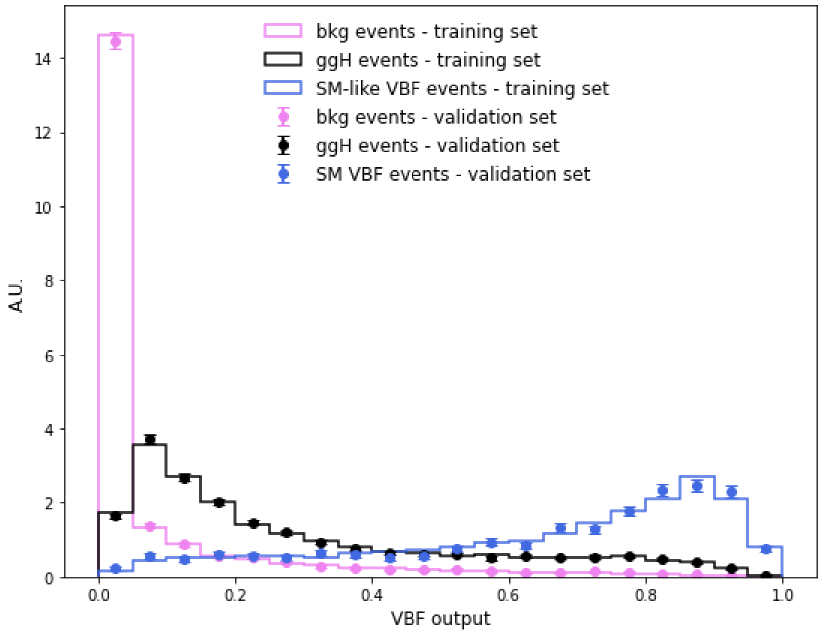


Figure A.3: *Normalized distributions of the VBF-output on SM signal (blue), ggH (black) and background (pink) events. The solid lines correspond to the distributions obtained on training sample, while the dots with statistical error bar stand for validation sample.*

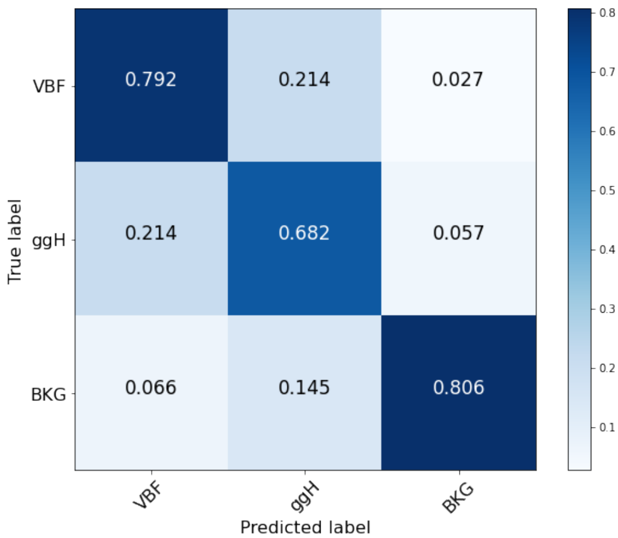
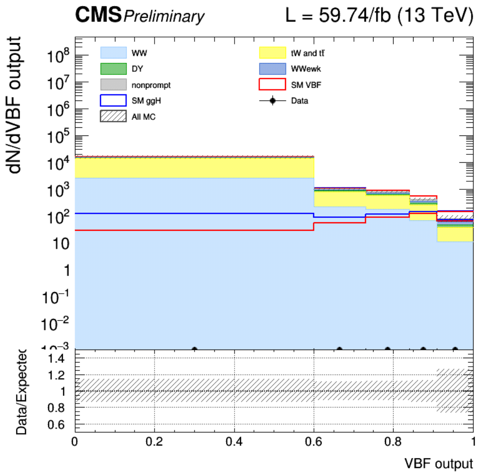
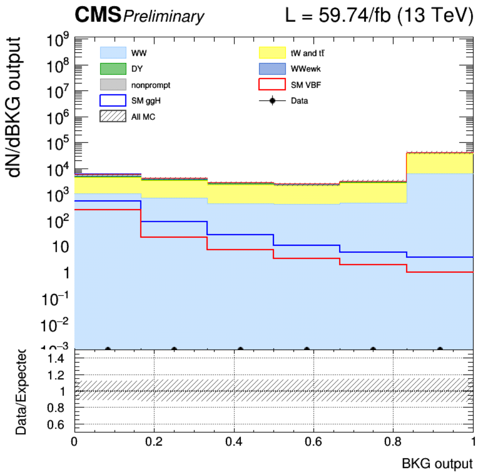


Figure A.4: *Confusion matrix.*

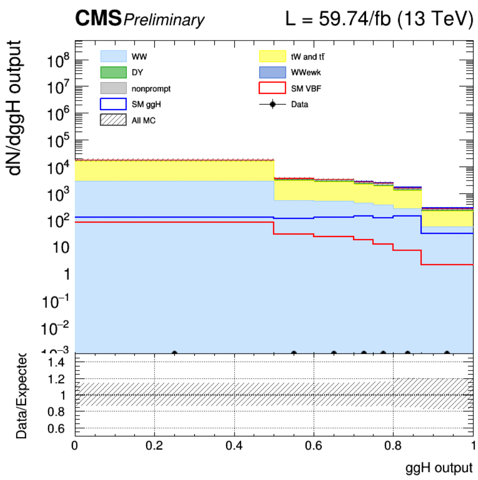
Signal Region



(a) *VBF-output.*



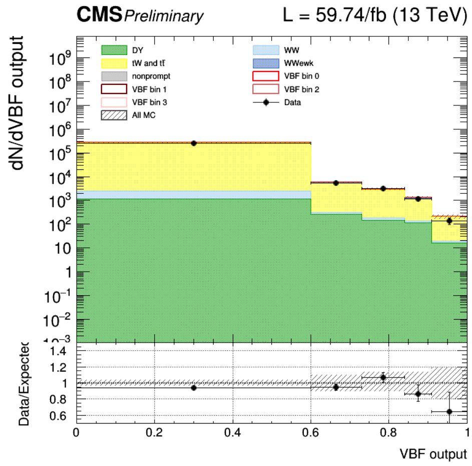
(b) *BKG-output.*



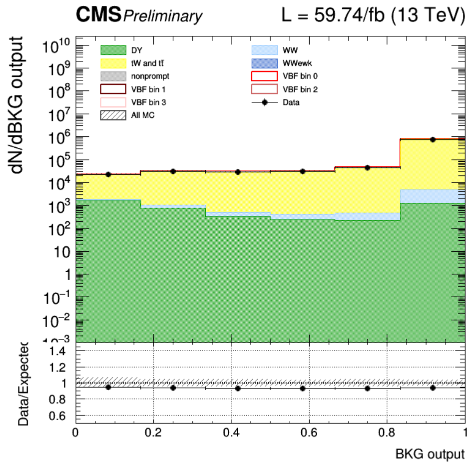
(c) *ggH-output.*

Figure A.5: *Distributions of the VBF-, ggH- and BKG-outputs evaluated in the SR. The VBF (red line) and ggH (blue line) signal contributions are shown both stacked on top of the background and superimposed on it. A logarithmic scale is used. The dashed lines represent both MC statistical and systematic uncertainties.*

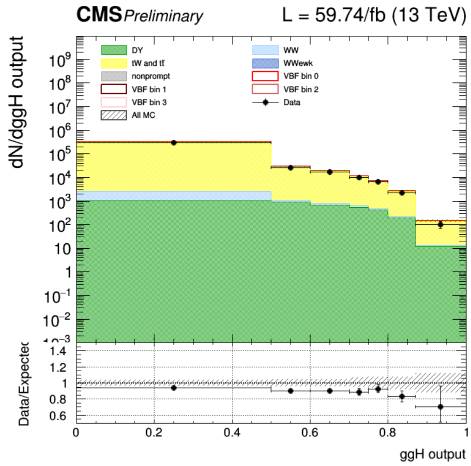
Top CR



(a) *VBF-output.*



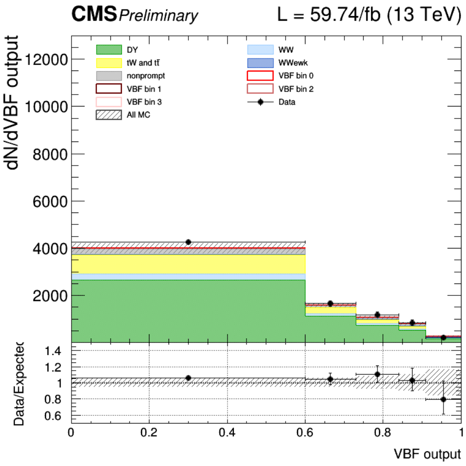
(b) *BKG-output.*



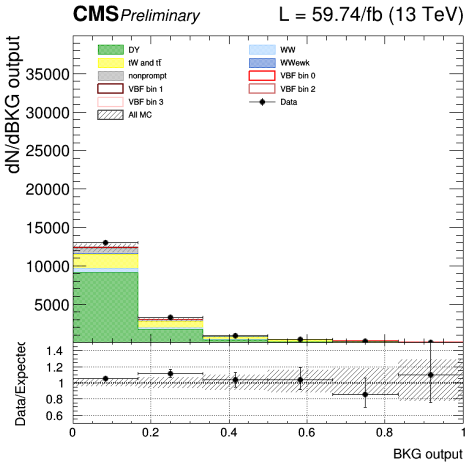
(c) *ggH-output.*

Figure A.6: *Distributions of the VBF-, ggH- and BKG-outputs evaluated in the Top CR, shown in logarithmic scale. Error bars are associated to statistical uncertainties on data, while dashed lines represent both MC statistical and systematic uncertainties.*

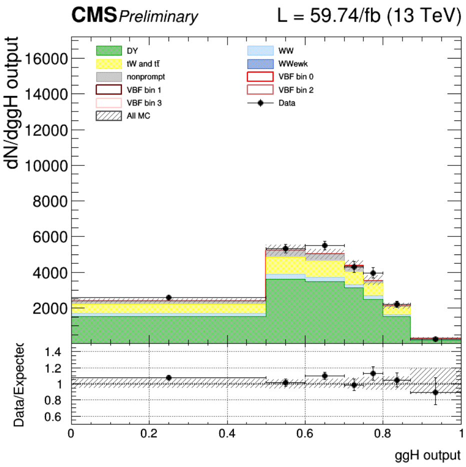
$DY \rightarrow \tau\tau$ CR



(a) *VBF-output.*



(b) *BKG-output.*



(c) *ggH-output.*

Figure A.7: *Distributions of the VBF-, ggH- and BKG-outputs evaluated in the $DY \rightarrow \tau\tau$ CR. Error bars are associated to statistical uncertainties on data, while dashed lines represent both MC statistical and systematic uncertainties.*

Bibliography

[1] M. E. Peskin and D. V. Schroeder, *An Introduction to quantum field theory.* Addison-Wesley, 1995.

[2] S. L. Glashow, "Partial-symmetries of weak interactions," *Nuclear Physics*, vol. 22, no. 4, pp. 579 – 588, 1961.

[3] A. Salam, "Weak and Electromagnetic Interactions," *Conf. Proc.*, vol. C680519, pp. 367–377, 1968.

[4] S. Weinberg, "A model of leptons," *Phys. Rev. Lett.*, vol. 19, pp. 1264–1266, Nov 1967.

[5] M. F. and S. G., *Quantum Field Theory. 2nd Edition.* John Wiley and Sons, 2010.

[6] M. T. et al., "Review of particle physics," *Phys. Rev. D*, vol. 98, p. 030001, Aug 2018.

[7] D. de Florian, C. Grojean, F. Maltoni, C. Mariotti, A. Nikitenko, M. Pieri, P. Savard, M. Schumacher, and R. Tanaka, *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector.* CERN Yellow Reports: Monographs, Geneva: CERN, Oct 2016. 869 pages, 295 figures, 248 tables and 1645 citations. Working Group web page: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHXSWG.

[8] G. Bhattacharyya, "Hierarchy problem and BSM physics," *Pramana*, vol. 89, p. 53, Oct. 2017.

[9] A. V. Gritsan, J. Roskes, U. Sarica, M. Schulze, M. Xiao, and Y. Zhou, "New features in the jhu generator framework: Constraining higgs boson properties from on-shell and off-shell production," *Physical Review D*, vol. 102, Sep 2020.

[10] L. Evans and P. Bryant, "LHC Machine," *JINST*, vol. 3, p. S08001, 2008.

[11] The CMS Collaboration, "The CMS experiment at the CERN LHC," *Journal of Instrumentation*, vol. 3, pp. S08004–S08004, aug 2008.

[12] G. Aad *et al.*, "The ATLAS Experiment at the CERN Large Hadron Collider," *JINST*, vol. 3, p. S08003, 2008.

[13] A. A. Alves, Jr. *et al.*, "The LHCb Detector at the LHC," *JINST*, vol. 3, p. S08005, 2008.

[14] K. Aamodt *et al.*, "The ALICE experiment at the CERN LHC," *JINST*, vol. 3, p. S08002, 2008.

[15] G. Altarelli and G. Parisi, "Asymptotic Freedom in Parton Language," *Nucl. Phys.*, vol. B126, pp. 298–318, 1977.

[16] F. Beaudette, "The CMS Particle Flow Algorithm," in *Proceedings, International Conference on Calorimetry for the High Energy Frontier (CHEF 2013): Paris, France, April 22-25, 2013*, pp. 295–304, 2013.

[17] M. Cacciari, G. P. Salam, and G. Soyez, "The anti-ktjet clustering algorithm," *Journal of High Energy Physics*, vol. 2008, p. 063–063, Apr 2008.

[18] M. Cacciari and G. P. Salam, "Pileup subtraction using jet areas," *Physics Letters B*, vol. 659, no. 1, pp. 119 – 126, 2008.

[19] P. Nason, "A New method for combining NLO QCD with shower Monte Carlo algorithms," *JHEP*, vol. 11, p. 040, 2004.

[20] T. Sjostrand, S. Mrenna, and P. Z. Skands, "A Brief Introduction to PYTHIA 8.1," *Comput. Phys. Commun.*, vol. 178, pp. 852–867, 2008.

[21] S. Agostinelli *et al.*, "Geant4—a simulation toolkit," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, vol. 506, no. 3, pp. 250 – 303, 2003.

[22] The CMS collaboration, "Measuring Electron Efficiencies at CMS with Early Data," 2008.

[23] V. Hankele, G. Klämke, D. Zeppenfeld, and T. Figy, "Anomalous higgs boson couplings in vector boson fusion at the cern lhc," *Physical Review D*, vol. 74, Nov 2006.

[24] M. Kuusela and V. M. Panaretos, "Statistical unfolding of elementary particle spectra: Empirical bayes estimation and bias-corrected uncertainty quantification," *The Annals of Applied Statistics*, vol. 9, Sep 2015.

[25] G. Cowan, *Statistical Data Analysis.* Oxford science publications, Clarendon Press, 1998.

[26] L. Lista, "Statistical Methods for Data Analysis in Particle Physics," *Lect. Notes Phys.*, vol. 941, pp. 1–257, 2017.

[27] A. Sirunyan *et al.*, "Measurements of properties of the Higgs boson decaying to a W boson pair in pp collisions at $\sqrt{s}$=13 TeV," *Physics Letters B*, vol. 791, pp. 96 – 129, 2019.

[28] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining, Fourth Edition: Practical Machine Learning Tools and Techniques.* San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., 4th ed., 2016.

[29] S. Ben-David, J. Blitzer, K. Crammer, A. Kulesza, F. Pereira, and J. Vaughan, "A theory of learning from different domains," *Machine Learning*, vol. 79, pp. 151–175, 05 2010.

[30] F. Chollet *et al.*, "Keras," 2015.

[31] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, and X. Zheng, "TensorFlow: Large-scale machine learning on heterogeneous systems," 2015. Software available from tensorflow.org.

[32] The CMS Collaboration, "Performance of quark/gluon discrimination in 13 TeV data," Nov 2016.

[33] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proceedings of the 25rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.

[34] A. Klein, S. Falkner, S. Bartels, P. Hennig, and F. Hutter, "Fast Bayesian Optimization of Machine Learning Hyperparameters on Large

Datasets," in *Proceedings of the 20th International Conference on Artificial Intelligence and Statistics* (A. Singh and J. Zhu, eds.), vol. 54 of *Proceedings of Machine Learning Research*, pp. 528–536, PMLR, 20–22 Apr 2017.