SAPIENZA
UNIVERSITÀ DI ROMA

Higgs boson anomalous couplings through the W,Z associated production in the diphoton decay channel with CMS experiment

Facoltà di Fisica
Particle and Astroparticle Physics

Riccardo Ursini
ID number 1756071

Advisor
E. Di Marco

Academic Year 2022/2023

Thesis not yet defended

Higgs boson anomalous couplings through the W,Z associated production in the diphoton decay channel with CMS experiment
Specialization thesis. Sapienza University of Rome

This thesis has been typeset by LaTeX and the Sapthesis class.

Author's email: riccardo.ursini@cern.ch

Abstract

The study of CP violation and anomalous couplings of the Higgs boson to massive gauge bosons is presented. The thesis focuses on the two-photons decay channel of the Higgs boson, produced in association to a vector boson V (=W,Z) either decaying into a hadronic - with two jets - or leptonic - with electrons/muons - final state. The data used were acquired by the CMS experiment during the LHC Run-2, corresponding to an integrated luminosity of 137 fb^{-1} at a proton-proton collision energy of 13 TeV. The kinematic effects of Beyond Standard Model contributions, modeled with Monte-Carlo simulations, are considered to train a multivariate algorithm to identify the different production mechanisms and to discriminate among the backgrounds, both the overwhelming non resonant two photons production and the Standard Model production of Higgs bosons. The limits of possible anomalies in the coupling of the Higgs boson to massive gauge bosons - extracted by a fit on the diphoton invariant mass distribution - are reported as direct constraints on additional terms to the Standard Model Lagrangian within the framework of Effective Field Theories.

Contents

Chapter 1

Introduction

Discovered at the CERN's Large Hadron Collider (LHC), the Higgs boson has so far been found to be compatible with Standard Model (SM) predictions. Its spin-parity quantum numbers are consistent with $\mathrm{J}^{PC} = 0^{++}$, according to measurements performed by CERN's CMS [1] and ATLAS [2] experiments, nonetheless leaving room for the possibility of small anomalous HVV couplings ($\mathrm{V}=\mathrm{W}^{\pm}$, Z^0) which could alter its $\mathbb{CP}$ properties. The Higgs boson couplings, once the Higgs boson mass is defined, are precisely predicted by the Standard Model and finding a deviation on these values would lead to evidence of new physics. The present thesis aims at searching possible effects beyond the Standard Model (bSM) through the measurement of the Higgs boson couplings with the electroweak massive vector bosons. Because non-zero spin hypotheses for the Higgs boson have been ruled out, the analysis will focus on investigating anomalous couplings of a scalar Higgs. Previous studies of anomalous HVV couplings were performed by both the CMS and ATLAS experiments, also including off-shell Higgs boson production [3]. In this thesis an extensive study of HVV couplings for a Higgs boson in associated production to a weak vector boson (VH) is performed, looking at the $H \to \gamma\gamma$ decay channel of the Higgs, which has a low branching fraction but a clean signature in the CMS detector. The data used were recorded by the CMS experiment during LHC Run 2 in 2016, 2017 and 2018 at proton-proton collision center-of-mass energy of 13 TeV, and correspond to an integrated luminosity of $137\mathrm{fb}^{-1}$.

The Z/W weak boson produced alongside the Higgs boson, can either decay leptonically into a $l^+l^-/l\nu$ pair or hadronically into a $q\bar{q}$ pair. The analysis will focus on the hadronic decay channel of the vector boson due to several reasons. Firstly, the higher branching ratio [4] resulting in higher statistics:

$$\mathcal{BR}(Z \to q\bar{q}) = 69.91\% \qquad \mathcal{BR}(W \to q\bar{q}^{'}) = 67.41\% \tag{1.1}$$

$$\begin{aligned} \mathcal{BR}(Z \to e^+e^-) &= 3.36\% & \mathcal{BR}(W \to e\,\nu_e) &= 10.71\% \\ \mathcal{BR}(Z \to \mu^+\mu^-) &= 3.36\% & \mathcal{BR}(W \to \mu\,\nu_\mu) &= 10.63\% \\ \mathcal{BR}(Z \to \tau^+\tau^-) &= 3.36\% & \mathcal{BR}(W \to \tau\,\nu_\tau) &= 11.38\% \end{aligned}$$

In addition, the leptonic channel for VH production is also currently being investigated by another CMS analysis. Nonetheless, the leptonic VH categories will be included in the final fit, despite not being explicitly optimized in this work. It should be mentioned that the leptons that are included in the analysis are only electrons and muons, as tauons decay into low p_T objects - either jets or further charged leptons - that are difficult to reconstruct and neutrinos that are only detected in terms of missing transverse energy (MET).

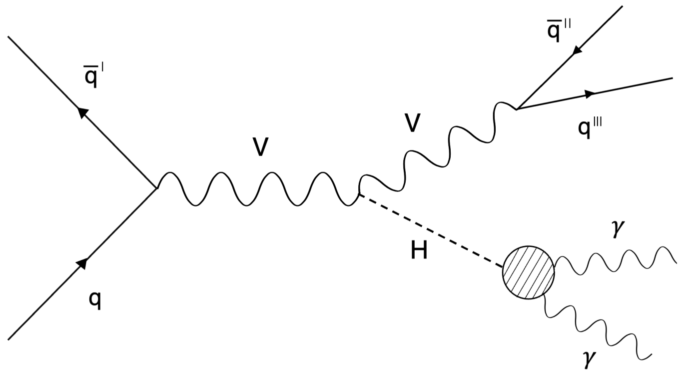


Figure 1.1. Feynmann diagram of the hereby studied case of interest: diphoton decaying Higgs boson in associated production with a weak vector boson V (=W,Z) decaying in the hadronic channel. The HVV vertex is present at production stage.

The V(→jj)H→ $\gamma\gamma$ signal process, illustrated in Figure [1.1], exhibits a clear final state: a diphoton state resonant to the Higgs mass, plus a dijet state resonant to the Z/W mass, thus providing a clear signature for the event.

The number of VH produced Higgs events over the whole Run-2 is expected to be:

$$N_{VH} = \sigma_{VH} \cdot \mathcal{BR}(H \to \gamma\gamma) \cdot \mathcal{L}_{int} \cdot \epsilon_\gamma^2 = \\ (\underbrace{1.37}_{\text{WH}} + \underbrace{0.88}_{\text{ZH}}) pb \cdot 2.27 \times 10^{-3} \cdot 137 fb^{-1} \cdot (0.5)^2 \simeq 170 \text{ events} \tag{1.2}$$

where σ is the cross section for VH production (at M$_H$ = 125 GeV, $\sqrt{s} = 13$ TeV), $\mathcal{BR}$ the branching ratio of the $H \to \gamma\gamma$ decay channel, and ϵ_γ the approximate reconstruction efficiency of a photon in the CMS Electromagnetic Calorimeter (ECAL). The identification of the Higgs boson signal, which is at the basis of the measurement of the HVV couplings, relies on a fit of the diphoton invariant mass distribution, where the Higgs boson appears a narrow peak over a decreasing background distribution, which will be directly fitted from Run-2 data. In order to be sensitive to possible anomalous couplings, a multivariate analysis technique consisting of an implementation of a Boosted Decision Tree (BDT) fed with reconstructed kinematic variables of photons and jets - whose distributions varies depending on the CP structure of the coupling - will be performed. The thesis is structured as follows: Chapter 2 provides a theoretical overview of The Standard Model theory and the Higgs mechanism, as well as a Higgs boson characterization in terms of production mechanisms, decay channels and quantum numbers. Additionally, a focus on Effective Field Theories (EFT), in which possible bSM effects can be framed, is presented.

Chapter 3 describes the experimental setup, including the Large Hadron Collider and the CMS detector along with their characterizing components. The fourth chapter describes the selection criteria applied on the final state to isolate the event of interest from background sources. A general overview on kinematic distributions and their shape response to different couplings and production methods is provided. Chapter 5 exploits the sensibility of kinematic distributions to different production mechanism and SM/bSM scenarios to train the BDT for performing multivariate analysis. Subsequently, a classification of the events according to a categorization based on the probability output of the BDT is performed. The categorization scheme is then optimized to maximize the overall statistical significance. Chapter 6 finally performs the likelihood fit over several different categories of different signal purity, with the purpose of extracting the yield deviation, by production mechanism, from Standard Model scenario and the Higgs anomalous couplings to massive vector bosons.

Chapter 2

The Standard Model

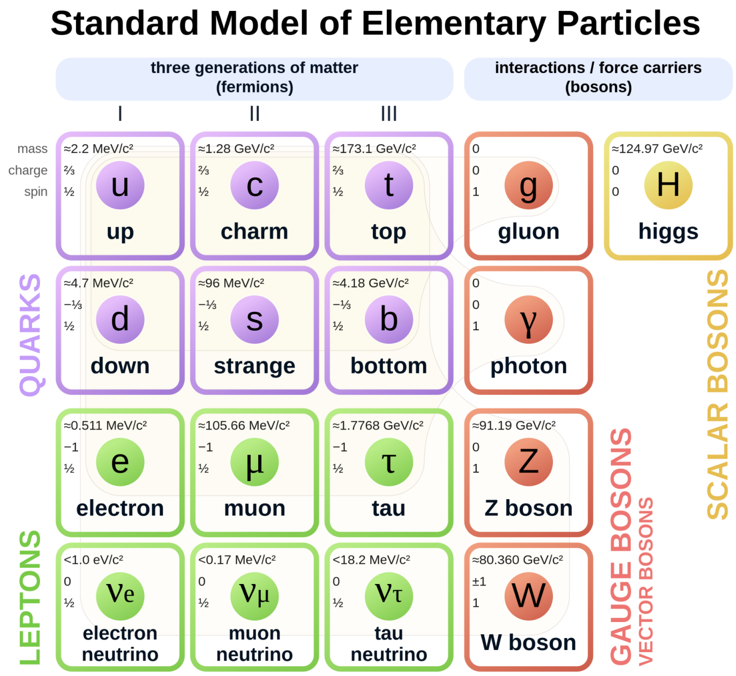


Figure 2.1. Fundamental constituents of nature, divided into matter fermions and gauge bosons, which mediate the fundamental interactions. The Higgs boson is the last addition to the model, discovered in 2012, despite being theorized several decades earlier

The Standard Model of particle physics is a comprehensive theory that describes the behaviour and the fundamental structure of matter and interactions, gravity excluded, at the smallest currently achievable scale. The theory was formulated by gathering results and theoretical predictions from the 1960s onwards, after the pioneering work of Weinberg, Salam, and Glashow [5, 6], who proposed a gauge theory based on a $SU(2) \otimes U(1)$ symmetry group that unified weak and electromagnetic interactions. The Standard Model is founded on the gauge paradigm, which postulates that the fundamental forces are described by gauge theories. This is due to the various properties of the three fundamental forces and their low-energy manifestations, or infrared (IR) phases:

- **Electromagnetic** (Coulomb phase): a long range interaction corresponding to a massless vector field;
- **Weak** (Higgs phase): a short range interaction, requiring a spontaneous symmetry breaking (SSB) in the theory to produce massive spin 1 particles mediating the force via the Higgs mechanism;
- **Strong** (Confined phase): an even shorter-range interactions that exhibits confinement at low energy, accountable for binding quarks together into color-singlet asymptotic states known as hadrons.

The use of gauge theories is advantageous due to the broad phenomenology that can be spanned by a single gauge Lagrangian, reproducing the IR phases of the fundamental interactions. This results from the renormalization group equation for the beta function, which characterizes the behavior of the coupling in a specific theory as a function of the energy scale μ. The perturbative expansion at 1 loop corrections in the coupling constant yields a comprehensive understanding of the interactions at low-energy scales:

$$\beta(g) = \mu \frac{d}{d\mu} g(\mu) = \frac{\beta_0}{16\pi^2} g^3 + o(g^5)^1 \tag{2.1}$$

In the case of QED, which is an abelian gauge theory:

$$\beta_0 > 0 \tag{2.2}$$

so, at least for small values of the coupling (i.e. the fine structure constant) where the perturbative expansion holds, the coupling increases as the energy scale rises. This trend keeps on holding even out of the perturbative regime, with the coupling eventually blowing up at a finite energy scale, where the QED shows a Landau pole and the theory ceases to be valid.
In the case of a SU(N) non-abelian gauge theory, the lowest order coefficient writes:

$$\beta_0 = -\frac{11}{3} C_2(adj) + \frac{4}{3} T(r_f) n_f + \frac{1}{3} T(r_s) n_s \tag{2.3}$$

where $C_2(adj)$(=N_C, the number of colors of the theory) is the quadratic Casimir of the adjoint representation of the gauge group in which the generators, hence the force mediators, transform. n_f (n_S) is the number of coloured fermion (boson) flavours and $T(r_f)$ ($T(n_S)$) the Dynkin index of the representation in which the fermions (bosons) transform. In the case of QCD, based on a SU(3) gauge group, $N_c = 3$ and the Dynking index for the Dirac fermions, which transform in the fundamental representation of the group, is equal to $\frac{1}{2}$. Since $nf \ll \frac{11}{2} N_C$ in the SM:

$$\beta_0 = \frac{11}{2} N_C + \frac{2}{3} n_f < 0 \tag{2.4}$$

[1]This expansion comes from a MS regularization scheme. It can however be taken without lossof generality

the strong running coupling decreases as the the energy scale increases, making QCD a strongly coupled theory in confined phase in IR regime.

The gauge group of the Standard model is:

$$SU(3) \otimes SU(2) \otimes U(1) \tag{2.5}$$

The first factor arises from QCD, while the remaining part of the gauge group represents the electroweak (EW) sector, which spontaneously breaks into the abelian invariant subgroup of the QED (more on SSB in section 2.1.1):

$$SU_L(2) \otimes U(1)_Y \longrightarrow U(1)_Q \tag{2.6}$$

where, being pedantic, $SU(2)_L$ is generated by weak isospin (I) and represents the weak force, which explicitly couples only to the left-handed component of the matter spinors, making the SM a chiral theory. $U(1)_Y$ group is generated by weak hypercharge (Y), which is related to weak isospin and to the electric charge (Q) generating the invariant subgroup $U(1)_Q$ of electromagnetism, by the Gell-Mann-Nishijima relation:

$$Q = I_3 + \frac{Y}{2} \tag{2.7}$$

Matter particles in the SM are divided into quarks and leptons, depending on whether they interact through strong force or not. Leptons do not interact thorugh strong force and therefore they are singlets of $SU(3)$. Depending on their chirality however, they can either transform as doublets of $SU(2)$, if they are left-handed:

$$I = \frac{1}{2}: \qquad l^i_L = \begin{pmatrix} \nu^i_L \\ e^i_L \end{pmatrix} \tag{2.8}$$

or as singlets if they have right chirality:

$$I = 0: \qquad e^i_R \tag{2.9}$$

where i= 1,2,3 (e^i = e$^-$, μ^-, τ^- ; $\nu^i = \nu_e, \nu_\mu, \nu_\tau$) is the family index.
Quarks on the other hand, interact strongly through gluon exchange, thus always transforming as triplets of SU(3). As for the leptons, with respect to the weak interaction, they either transform as doublets:

$$I = \frac{1}{2}: \qquad q^i_L = \begin{pmatrix} u^i_L \\ d^i_L \end{pmatrix} \tag{2.10}$$

or as singlets:

$$I = 0: \qquad u^i_R, d^i_R \tag{2.11}$$

for any of the 3 families ($u^i = u, c, t$; $d^i = d, s, b$).

The chiral nature of the SM directly prevents explicit Dirac mass terms from appearing in the Lagrangian:

$$\mathcal{L}_{SM} \not\supset m(\overline{\psi}_L \psi_R + h.c.) \tag{2.12}$$

as left-handed components of the spinors transform as doublets of SU(2), while right-handed components do transform as singlets of the group, thus spoiling gauge invariance. Majorana mass terms do not suit either, since they always break gauge invariance. Therefore in the SM masses are dynamically generated, either via a Yukawa coupling with the Higgs field in the case of fermions, or through the Higgs mechanism for the weak gauge bosons.

The abelian factor in SM gauge group causes the SM itself to have a Landau pole (at an energy scale of $\sim 10^{42}$ GeV) [2], thus implying that the SM is not a theory that can be interpolated at arbitrary high energy, but that can only be regarded as an effective field theory (EFT) (more about EFT in Section 2.2).

The SM Lagrangian density is hence constituted by a renormalizable Lagrangian term plus effective terms including operators of mass dimension D>4:

$$\mathcal{L}_{SM} = \mathcal{L}^{(4)} + h.o. \tag{2.13}$$

the renormalizable term can be written as:

$$\mathcal{L}^{(4)} = \mathcal{L}_{kin} + \mathcal{L}_{Yuk} + \mathcal{L}_{\theta} - V(\phi) \tag{2.14}$$

where:

1. **Kinetic term**

$$\begin{aligned}\mathcal{L}_{kin} = &\underbrace{-\frac{1}{4}G^a_{\mu\nu}G^{a,\mu\nu} - \frac{1}{4}W^i_{\mu\nu}W^{i,\mu\nu} - \frac{1}{4}B_{\mu\nu}B^{\mu\nu}}_{\text{Gauge bosons}} + \\ &\underbrace{+(D_\mu\phi)^\dagger(D^\mu\phi)}_{\text{Higgs}} + \underbrace{i\sum_{\{\psi\}}\sum_{j}\overline{\psi}_j \not{D}\psi_j}_{\text{Matter fermions}}\end{aligned} \tag{2.15}$$

The fermionic field ψ in the Dirac term varies in $\{q_L, u_R, d_R, l_R, e_R\}$ and j $(= 1, 2, 3)$ is the family index. The index a (= 1,...,8) stands for the 8 gauge fields (gluons), corresponding to the SU(3) generators, i (=1,2,3) for the gauge fields corresponding to the SU(2) generators and there is only one gauge field that corresponds to the single generator of U(1). The field strengths are defined as:

$$F^a_{\mu\nu} = \partial_\mu\mathcal{A}^a_\nu - \partial_\nu\mathcal{A}^a_\mu - igf^{abc}\mathcal{A}^b_\mu\mathcal{A}^c_\nu \tag{2.16}$$

where $\mathcal{A}$ is the gauge field and f^{abc} is the structure constant characterizing the specific Lie group, with $f^{abc} = 0$ in the abelian case. The covariant derivative in the Higgs term couples the Higgs field to the gauge fields in the EW sector:

$$D_\mu = \partial_\mu - ig\frac{\sigma^i}{2}W^i_\mu - ig'YB_\mu \tag{2.17}$$

[2]The Landau pole is still several orders of magnitudes above the Planck scale (10^{19} GeV) at which the quantum nature of gravity should be addressed

σ^i are the Pauli matrices, the generators of SU(2), whereas Y$\cdot\mathbb{1}$ generates U(1). The covariant derivative in the Dirac term has a further term for the quark spinors, coupling them with gluon field:

$$D_\mu \supset -ig_3\lambda^a G^a_\mu \tag{2.18}$$

where g_3 is the strong coupling and λ^a are the Gell-Mann matrices, generating SU(3).

By working out the terms, one could re-write all in terms of the physical fields of the EW sector, which are defined as combinations of the W/B gauge fields:

$$W^\pm_\mu = \frac{W^{(1)}_\mu \mp W^{(2)}_\mu}{\sqrt{2}}$$

$$Z_\mu = -sin\theta_W B_\mu + cos\theta_W W^{(3)}_\mu \qquad A_\mu = cos\theta_W B_\mu + sin\theta_W W^{(3)}_\mu \tag{2.19}$$

with θ_W being the Weinberg angle. The $W^\pm$ bosons couple only to left-handed components of the spinors, the Z and the photon couple to both left and right-handed components.

2. **θ-term** [3]
There is another possible gauge invariant, Lorentz invariant and renormalizable way to write kinetic terms for gauge fields, which is the so-called θ-term. It has a physical meaning only in the case of non-abelian gauge theories, and here is specifically present only for QCD, since for the SU(2)-related gauge fields the term can be absorbed by a field re-definition:

$$\mathcal{L}_\theta = \frac{\theta_0}{32\pi^2} G^a_{\mu\nu}\tilde{G}^{a,\mu\nu} \tag{2.20}$$

3. **Yukawa term**
This term couples matter fermions to the Higgs, generating the fermion masses, which is evident the unitary gauge.

$$\mathcal{L}_{Yuk} = -\overline{q}^i_L \phi^c y^{ij}_a u^j_R - \overline{q}^i_L \phi y^{ij}_d d^j_R + \overline{l}^i_L \phi y^{ij}_a e^j_R + h.c. \tag{2.21}$$

ϕ is the Higgs doublet, and $\phi^C = i\sigma^{(2)}\phi^*$ its charge conjugated, introduced to guarantee the first term to be a hypercharge singlet. The y are the Yukawa matrices which can be diagonalized, by properly rotating the fields, into matrices with fermions/leptons masses on the diagonal. This comes at the expenses of introducing an additional SU(3) matrix in the quark kinetic term, the CKM matrix, accountable for flavour mixing. By doing so the mass of the fermions are explicitly written in terms of the matrices eigenvalues:

$$m_{u^i} = \frac{v\cdot y^{ii}_u}{\sqrt{2}} \qquad m_{d^i} = \frac{v\cdot y^{ii}_d}{\sqrt{2}} \qquad m_{l^i} = \frac{v\cdot y^{ii}_a}{\sqrt{2}} \tag{2.22}$$

[3] The θ-term is accountable for a possible CP violation in the strong sector, which is yet to be observed. Therefore, at current knowledge, θ_0 is considered to be really small and the term could be neglected

4. **Scalar potential**
   Apart from matter fermions and vector gauge bosons, the SM posits the existence of a scalar field ϕ (focus in the next section). Consequently, the following scalar potential is introduced:

$$V(\phi) = \mu^2\phi^\dagger\phi + \lambda(\phi^\dagger\phi)^2 \tag{2.23}$$

   that is the most general way to write a renormalizable SU(2)-invariant scalar potential.

2.1 The Higgs boson

The need for introducing the Higgs boson emerged from the otherwise inexplicable observation that, in contrast to the photon, the weak gauge bosons must possess mass. In addition, as previously reported, explicit mass terms, either Dirac's or Majorana's, could not be included in the SM Lagrangian. Therefore, also the generation of mass for matter fermions appeared to lack a clear explanation. These were the main reasons that led to theorize the presence of a scalar field, of which the Higgs boson is the quantum mainfestation, by Englert and Higgs himself [8]. However, this is not the only issue solved by the Higgs boson theorization. In a spontaneously broken, non-abelian gauge theory mediated by some gauge bosons, $\mathcal{A}^i$, a 2→2 self-interaction of such vector bosons, longitudinally polarized, could be considered:

$$\mathcal{A}^a_L\mathcal{A}^b_L \longrightarrow \mathcal{A}^c_L\mathcal{A}^d_L \tag{2.24}$$

Evaluating the corresponding scattering amplitude's high-energy limit, which can be computed in terms of the Nambu-Goldstone bosons of the theory in the ξ-gauge (according to the Equivalence Theorem), it diverges as the square of the energy. This leads to a spoiling of the unitarity of the theory in the UV regime. The introduction of an additional degree of freedom (dof), namely the Higgs field, endows the matter fields with the correct number of dof_s for them to transform as a complete representation of the gauge group, hence preventing the scattering amplitude from diverging at high energy and restoring unitarity.

2.1.1 Higgs mechanism

The most direct way to spontaneously break the $SU(2) \otimes U(1)$ EW symmetry is to posit the existence of a field ϕ, a $SU(2)_L$ complex doublet of weak hypercharge Y=1, namely the Higgs doublet :

$$\phi(x) = \begin{pmatrix} \phi^+ \\ \phi^0 \end{pmatrix} = \frac{1}{\sqrt{2}} \begin{pmatrix} \sqrt{2}\phi^+ \\ \phi^0 + ia^0 \end{pmatrix}$$

(2.25)

where ϕ^0/a_0 are the CP-even/odd neutral components of the Higgs doublet, and ϕ^+ is the complex charged component. The field is associated to a Lagrangian density given by the sum of the previously cited scalar potential and kinetic term for the Higgs:

$$\mathcal{L}_{Higgs} = (D_\mu \phi)^\dagger (D^\mu \phi) - \mu^2(\phi^\dagger \phi) - \lambda(\phi^\dagger \phi)^2 \tag{2.26}$$

which is manifestly symmetric with respect to the EW gauge group. The condition $\lambda > 0$ [4], ensures that the energy is bounded from below and therefore that there exists a stable ground state. If $\mu^2 < 0$ the potential gets the peculiar shape of Figure 2.2, causing the neutral component of the Higgs doublet to acquire a non-vanishing vacuum expectation value (VEV):

$$\langle \phi_0 \rangle \equiv v = (G_F\sqrt{2})^{-1/2} \simeq 246 \text{ GeV} \tag{2.27}$$

Consequently, on the manifold of degenerate vacua (3-sphere), the doublet takes the form:

$$\langle \phi \rangle = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ v \end{pmatrix} \tag{2.28}$$

(2.27) implies that the vacuum is not invariant under the action of the whole gauge group, but only under the action of a subgroup of it, thus triggering the SSB (2.6) that characterizes the EW sector in the SM.

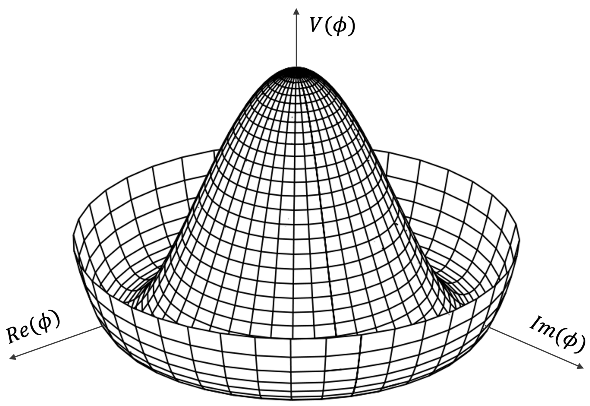


Figure 2.2. Typical shape of the Higgs scalar potential for $\mu^2 < 0$ in the $Re(\phi) - Im(\phi)$ space

From the 4 generators of $SU(2) \otimes U(1)$ only 1 for the abelian invariant subgroup U(1) is left unbroken. There are consequently 3 broken generators to which, according to Goldstone's theorem, correspond 3 Nambu-Goldstone bosons (NGB$_s$).

Pre-SSB, the Higgs doublet is a complex 2-component vector, thus 4 dof$_s$, parametrized as for (2.1.1). After SSB, ϕ can be parametrized as:

[4] This is not a condition to be taken for granted at all energy ranges. According to the current measurements of Higgs and top quark mass (which provides the dominant loop contribution), the EW vacuum is most likely metastable, implying that, at an energy scale below the Planck scale ($\mathcal{O}(10^{11})$ GeV), the quartic self interaction parameter becomes negative and the Higgs potential stability is spoiled [11],[12]

$$\phi(x) = e^{i\chi^{\hat{a}}(x)T^{\hat{a}}/v} \begin{pmatrix} 0 \\ \frac{v+h(x)}{2} \end{pmatrix} \tag{2.29}$$

where $T^{\hat{a}}$ ($\hat{a} = 1, 2, 3$) are the broken generators in the SSB, $\chi^{\hat{a}}$ are the just cited NGB$_s$, corresponding to massless angular dof$_s$ rotating the fields on the manifold of vacua, and $h(x)$ is a radial excitation along the walls of the potential which corresponds to the massive dof that we call Higgs boson.

In the unitary gauge[5], the vector gauge bosons absorb the NGB$_s$ and the Higgs mechanism becomes evident. The only real scalar field left is $h(x)$ and the doublet takes the form:

$$\phi(x) = \begin{pmatrix} 0 \\ \frac{v+h(x)}{2} \end{pmatrix} \tag{2.30}$$

The additional 3 scalar dof$_s$ arising from NGB$_s$, are spent to turn the 3 massless weak gauge fields (2 dof$_s$ each) into 3 massive gauge fields (3 dof$_s$ each) of mass:

$$m_W^2 = \frac{g^2v^2}{4} \qquad m_Z^2 = \frac{(g^2 + g'^2)v^2}{4} \tag{2.31}$$

Therefore the Higgs mechanism is just a reshuffling of dof$_s$ in a gauge theory with SSB, thus providing a way for the $W^{\pm}, Z$ bosons to acquire mass.

The mass of the posited Higgs boson can be expressed in terms of the quartic self-coupling and VEV:

$$m_H = \sqrt{2\lambda}v \tag{2.32}$$

which makes the Higgs mass a parameter of the Standard Model.

2.1.2 Discovery at LHC

The Higgs boson was jointly discovered by CMS and ATLAS collaborations in 2012 [14, 15], gathering data from LHC Run1, at an integrated luminosity of 5.1fb^{-1} at 7 TeV in 2011, and 5.3fb^{-1} at 8 TeV in 2012 for CMS (and similar luminosity values for ATLAS). The discovery was made by looking at the bump in the invariant mass of the final state for the Higgs decay channels with the higher mass resolution: $H \to \gamma\gamma$ and $H \to ZZ$ (more on Higgs decay channels in section 2.1.4), at an energy of:

$$m_H = 125.3 \pm 0.4(stat) \pm 0.4(syst) \text{ GeV} \tag{2.33}$$

that is the first estimation (from CMS experiment) of the mass of the Higgs resonance. As for the CMS collaboration, the Higgs boson discovery has been declared to be proven at a statistical significance of 5.8σ, above the 5σ threshold for claiming a discovery, which corresponds to a rejection of the zero hypothesis at a (1 - 1.7$\cdot$10^{-9})CL.

[5] In the unitary gauge the number of unpaired scalar dof$_s$ is minimal, i.e. the NGB$_s$ are suppressed

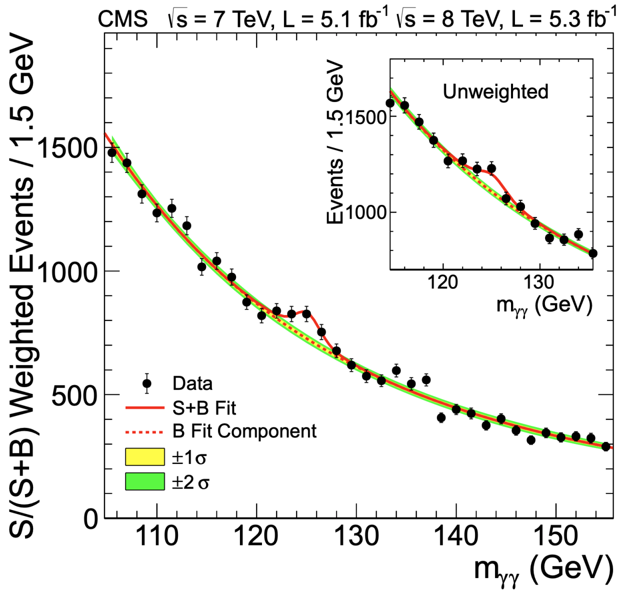


Figure 2.3. Original 2012 plot of the invariant mass of the final state, in the H→ $\gamma\gamma$ channel from the CMS collaboration [14]

Further data from LHC Run2 (2015-2018), with a higher center of mass energy, of 13 TeV, corroborated the hypothesis of the SM Higgs boson with a mass around 125 GeV. Then, several other precision measurements Higgs couplings, cross sections, branching ratios have been performed.

At the Large Hadron Collider (LHC) Higgs bosons are produced according to several modes from initial proton-proton collision. With a measured decay width $\Gamma_H = 3.2^{+2.4}_{-1.7}$ MeV [16], compatible with the SM prediction (4.1 MeV), the Higgs has such a short lifetime that it is only observed as a resonance, i.e. a short-living intermediate state between an initial (hadronic) state and the final state composed by the products of Higgs decay. It is thus worth exploring which are the possible production modes of the Higgs boson and which final states the Higgs boson can decay into.

2.1.3 Production at LHC

Being the LHC a proton-proton machine, Higgs bosons can only be produced from a primary quark-quark or gluon-gluon interaction. At the centre-of-mass (CoM) energy of 13 TeV indeed, rather than coherent point-like particles, protons appear as a composition of valence quarks (uud) and a plethora of sea-quarks and gluons, which carry a small fraction of the proton momentum. These constituent partons, interacting at a fundamental level, might produce a Higgs boson in the final state according to several distinct production mechanisms, depicted in Figure 2.5:

- **Gluon fusion (ggH)**: At a $\sqrt{s}$ of 13 TeV and for the Higgs mass $\sim$ 125 GeV, gluon fusion is the dominant production mechanism. From of pair of

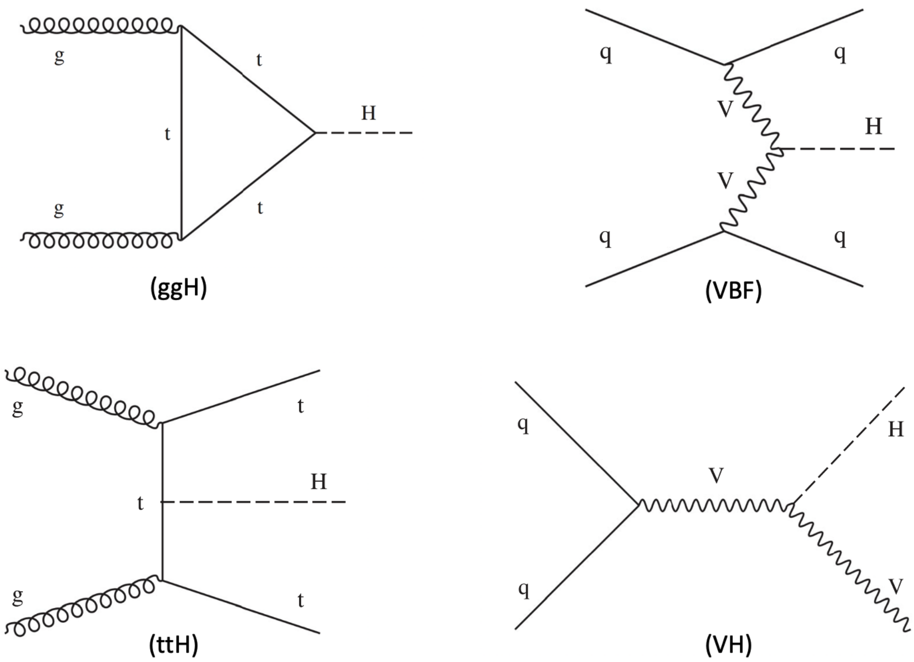


Figure 2.4. Main leading order Feynman diagrams of the Higgs production modes at LHC

gluons, the Higgs boson is produced at loop level through a mediation of a virtual top quark. The contribution from the other quarks are suppressed as $\sim m_q^2$, and the top quark contribution is the only relevant one. This kind of production mechanism is largely affected by radiative corrections which are nowadays computed at next-to-next-to-next-to leading order (N3LO) [17]. However, since no explicit HVV vertex (with V = W/Z) is present, this mode can not be considered for estimating Higgs-weak vector bosons couplings;

- **Vector boson fusion (VBF)**: it is the second most common Higgs production mode at LHC, and arises from a $q\overline{q}'$ initial state radiating a couple of weak vector bosons that scatter and produce the Higgs boson. This production mode is characterized by a pair of hard forward, almost back-to-back jets in the final state -originating from the quarks in the initial state- thus allowing to distinguish this kind of process both from overwhelming QCD background and ggH (+2 jets)/V($\rightarrow$ jj)H productions;

- **Associated production with a vector boson (VH)**: also called *Higgs-strahlung* due to the fact that the Higgs is radiated from a weak vector boson, it arises from a $q\overline{q}'$ initial state. In the final state the Higgs boson is produced alongside a weak vector boson V (=$W^{\pm}, Z^0$), which either decays hadronically in a quark/anti-quark pair, or leptonically in a lepton/anti-lepton pair. The following analysis will focus on such production method, specifically looking for the vector boson hadronic decay channel;

- **Associated production with a top quark pair (t$\bar{t}$H)**: the last process to be considered is the Higgs production in association with a pair of top quarks.

Despite being far from dominant, this process is useful for investigating Higgs properties due to the strong coupling with the top quark. It is indeed the main channel where to look for Higgs-fermions couplings.

This list however is not exhaustive and only comprises the production modes with higher cross section at $\sqrt{s} = 13$ TeV. Those constitute the main sources of resonant background to the VH production, hereby under study, and they are going to be included in the following analysis. In the dominant ggH mode, for Higgs decay channel into two weak vector bosons [6], the Higgs off-shell production is quite sizeable and the off-shell/on-shell ratio is estimated to be around 8% in the SM [10]. However, this scenario is still small compared to the on-shell case of interest, particularly for the other production methods, and it is going to be neglected.
Higgs production cross section depends both on CoM energy of the system and on the mass of the Higgs.

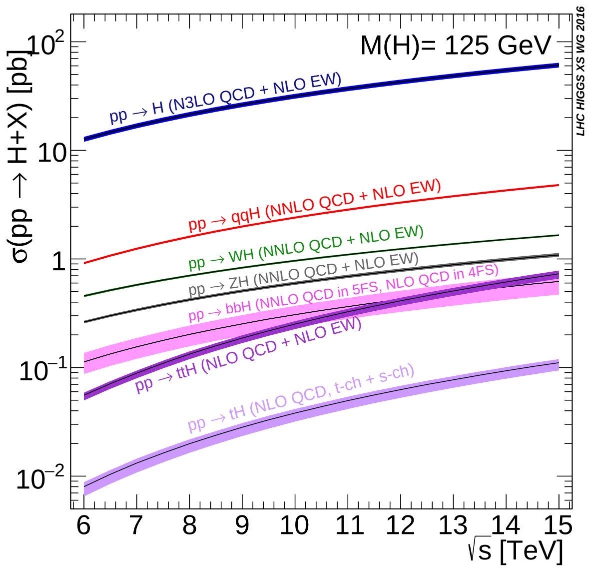


Figure 2.5. Cross section of Higgs production modes as a function of proton-proton the center of mass energy (ggH is indicated as $pp \to H$ and VBF as $pp \to qqH$). The gain in statistics with the Run-1/Run-2 transition (from 8 to 13 TeV) is evident, as the cross section approximately doubled with the CoM energy increase.[20]

In Figure 2.5 the mild dependence of the cross sections on the energy in the 6-15 TeV window is reported, for a Higgs boson mass of 125 GeV. Among all the Higgs production events during LHC Run 2, around 88% are attributed to ggH, 7% to VBF, 4% to Higgs-strahlung and 1% to ttH production [18]. Each of these production mechanisms exhibit a different kinematic topology, reflecting the different underlying physics of the process. For instance, unlike gluon fusion and VH production, VBF is a t-channel process, implying that, unlike the aforementioned production modes, the jets in the final state will have a large angular separation and a smaller angular deflection in terms of the polar angle with respect to the beam axis (pseudo-rapidity)

[6] due to the strong Higgs coupling to the longitudinal polarization of vector bosons around the electroweak scale

and a larger invariant mass. On the other hand, due to their s-channel nature, both ggH (+2j) - whose jets are produced at NLO in QCD by the initial state radiation as ggH does not exhibits jets in the final state at LO - and V($\rightarrow jj$)H, show two less separated jets in the phase space. Altogether, VBF and V($\rightarrow jj$)H constitute a EW Hjj production method, i.e. an electroweak-based Higgs production process with 2 jets in the final state. Since they are both $q\overline{q}'$ originated and can yield the same final state, they could interfere with each other. However, the interference is suppressed as 1/N$_C$, where N$_C$ is the number of colors in the theory. Moreover, the t-channel VBF propagator and the s-channel VH propagator can not be simultaneously on-shell, so the interference is further suppressed and hence negligible. Thus, EW Hjj production is just the incoherent sum of VBF + V($\rightarrow jj$)H production modes. The two contributions are comparable up to a certain dijet invariant mass range ($\sim$ 500 GeV), above which the VBF largely dominates the EW Hjj production [19]. By looking at the kinematic distributions, some difference in shape among different production modes can be observed, suggesting proper kinematic cuts for discriminating Higgs bosons produced according to different production modes. A more extensive comparison of the kinematic distributions of production modes is provided in Section 4.4.

2.1.4 Decay

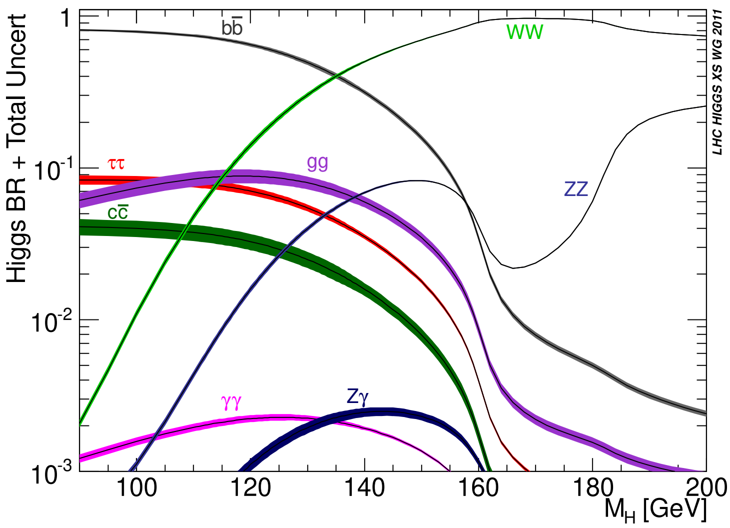


Figure 2.6. Higgs decay branching ratios as a function of Higgs mass value [20], at $\sqrt{s}$ = 13 T

Within the framework of the Standard Model, the Higgs boson can decay into various final states. The sensitivity and the branching ratio (BR) of each decay channel dependent on Higgs boson mass, as well as CoM energy. This relationship is clearly illustrated in Figure 2.6. For a Higgs boson with a mass of approximately 125 GeV, the dominant decay channel is $H \rightarrow b\bar{b}$, followed by $H \rightarrow WW$, $H \rightarrow \tau\tau$, $H \rightarrow WW$, $H \rightarrow c\bar{c}$, $H \rightarrow \gamma\gamma$, and $H \rightarrow Z\gamma$. A quantitative estimation of the BRs is provided in Table 1.1. However, the BR is not the sole figure of merit for a decay channel, despite a higher BR guarantee more statistics. For instance, despite possessing some of the highest BRs, $H \to b\bar{b}$ and $H \to \tau\tau$ are challenging channels for Higgs detection at a hadron collider. Higgs hadronic decay channels are plagued by a substantial QCD background - consisting in a mulit-jet production from QCD processes - constantly generated by proton-proton collisions LHC. Instead, the τ decay channel is problematic due to the difficulty in reconstructing the decay chain of its hadronic + MET or leptonic + MET channels. The $H \to WW$ decay channel can lead to either a final state with hadrons ($W \to q\bar{q}$), which also suffers from QCD background, or a final state with neutrinos ($W \to l\nu$), which are not reconstructed in the final state and can only be detected in terms of missing transverse energy.

Decay channel	Branching ratio	Rel. uncertainty
$H \to \gamma\gamma$	2.27×10^{-3}	2.1%
$H \to ZZ$	2.62×10^{-2}	$\pm$1.5%
$H \to W^+W^-$	2.14×10^{-1}	$\pm$1.5%
$H \to \tau^+\tau^-$	6.27×10^{-2}	$\pm$1.6%
$H \to b\bar{b}$	5.82×10^{-1}	$^{+1.2\%}_{-1.3\%}$
$H \to c\bar{c}$	2.89×10^{-2}	$^{+5.5\%}_{-2.0\%}$
$H \to Z\gamma$	1.53×10^{-3}	$\pm$5.8%
$H \to \mu^+\mu^-$	2.18×10^{-4}	$\pm$1.7%

Table 1.1 Higgs branching ratios values and relative uncertainty at $M_H = 125$ GeV, $\sqrt{s} = 13$ TeV [20] . In the case of unstable particles, these BR_s must be multiplied by the BR for the specific final state observed.

The two most commonly employed channels in this kind of Higgs analysis are $H \to ZZ$, where each Z boson decays into two charged leptons ($Z \to l\bar{l}$), resulting in a four-lepton final state, or $H \to \gamma\gamma$. The latter channel is the focus of investigation in this analysis. This choice is based on several reasons. Firstly, anomalous couplings of the Higgs boson to weak vector bosons have been extensively studied in the $H \to ZZ$ channel, as it exhibits a HVV vertex both at production and decay stage [21]. Secondly, despite its small branching ratio, the $H \to \gamma\gamma$ channel offers a clean environment for analysis, as it guarantees high reconstruction efficiency and mass resolution. This is mainly due to the excellent performance of both the electromagnetic calorimeter (ECAL) and the reconstruction algorithms of the CMS detector. Nonetheless, also the $H \to \tau\tau$, $H \to WW$ and $H \to b\bar{b}$ are employed for the study of Higgs boson couplings.

Due to the abovementioned reasons, the final state invariant mass resolution varies depending on the decay channel. While it stands around 10% for $H \to b\bar{b}$, 15% for $H \to \tau\tau$ and 20% for $H \to WW \to 2l2\nu$, it significantly improves to (1-2)% for $H \to ZZ \to 4l$ and $H \to \gamma\gamma$, highlighting the advantage of considering those decay channels [10].

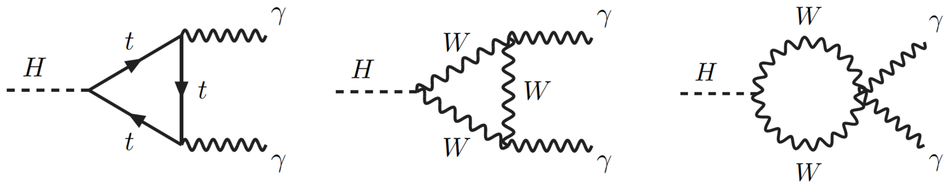


Figure 2.7. Examples of $H \to \gamma\gamma$ Feynman diagrams at 1-loop order [22]

As the photon, being a massless particle, does not directly couple to the Higgs boson, the $H \to \gamma\gamma$ decay process occurs at loop order, with either a massive vector boson or a massive fermion mediating the loop, as shown in Figure 2.7 .

2.1.5 Quantum numbers

In nature there are 3 fundamental discrete symmetries: charge conjugation ($\mathbb{C}$), parity $\mathbb{P}$ and time-reversal ($\mathbb{T}$). The parity operation performs an inversion of the spatial coordinates of a particle:

$$\mathbb{P}\,\psi(\vec{r}) = \psi(-\vec{r}) \tag{2.34}$$

All the polar vector observables characterizing a particle change sign under parity, as the momentum $\vec{p}$, whereas axial vector quantities are fixed point of the transformation, as the particle's angular momentum $\vec{J}$. The charge conjugation operation flips the sign of the internal, additive quantum numbers of a particle (electric charge, baryon number, lepton number, colour charge, hypercharge, strangeness,...), without affecting mass, momentum or spin and mapping a particle to its own antiparticle. A third discrete transformation is time reversal, which acts like parity but on time component of the particle's wave function

$$\mathbb{T}\,\psi(t) = \psi(-t) \tag{2.35}$$

All these operators are unitary ($\mathbb{O}\mathbb{O}^\dagger = \mathbb{O}^\dagger\mathbb{O} = \mathbb{I}$) and satisfy the condition $\mathbb{O}^2 = \mathbb{I}$. As a consequence, they exhibit a discrete spectrum of possible eigenvalues equal to ± 1. Fundamental particles may be eigenstates of these symmetries, thus being endowed with intrinsic multiplicative quantum numbers corresponding to the eigenvalues of those operators. $\mathbb{P}$ and $\mathbb{C}$ symmetry are neither conserved separately, nor together. Parity violation in weak interaction has been firstly discovered in 1956 by Madame Wu's experiment on cobalt nuclei decay, whereas the joint violation of $\mathbb{C}$ and $\mathbb{P}$ symmetries, can be traced back to the 1964 experiment of Cronin and Fitch, who observed $\mathbb{CP}$ violation in the decay of K mesons. The only discrete symmetry that, at present, still persists as a fundamental symmetry of nature is $\mathbb{CPT}$, the composition of the three single symmetries.

After its discovery, the main concern about the Higgs boson has been to probe its quantum numbers. As for the spin, the SM clearly predicts spin-0 particle: as reported in the previous section, the Higgs field introduction arose from the urge of providing the theory with the single additional missing dof that would have healed the UV scattering amplitudes divergence and provide a framework for SSB in order for the weak bosons to get mass.

The spin-parity quantum numbers can be mainly investigated by analyzing the angular distribution of helicity amplitudes both in production and decay processes. The J^{PC} properties of Higgs boson have been primarily studied in ggH, VBF and VH production processes, looking at the output of $H \to ZZ \to 4l$, $H \to WW \to l\nu l\nu$ and $H \to \gamma\gamma$ decay channels, which constitute the processes that present kinematic variables that are sensible to spin-parity properties of the Higgs boson [10, 23]:

- The $H \to \gamma\gamma$ channel is particularly sensitive to gluon-initiated, $J^P = 2^+$, case. The analysis is performed by taking into account the diphoton transverse momentum and the production angle $cos\theta^*$. While the $cos\theta^*$ distribution is expected to peak around 1 for 2^+ scenario, in the case of SM 0^+, it should be uniform, apart from a cut off due to selection criteria introduced to identify diphoton pairs.

- For the $H \to WW \to l\nu l\nu$ channel the production angles can not be easily reconstructed, as neutrinos are present in the final state. The $\gamma^\mu(1 - \gamma_5)$ nature of weak charged currents however, leads to having two close charged leptons in the transverse plane in the final state, making some variables, as their azimuthal angular separation $\Delta\Phi_{ll}$ and invariant mass m_{ll}, sensible to several spin-parity scenarios.

- In the $H \to ZZ \to 4l$ channel the discriminating variables are the several production and decaying angles than can be defined. This channel is mostly sensible to the pure 0^- scenario.

In particular, the observation of isotropic angular distributions in the ZZ or $\gamma\gamma$ channels, corroborate the idea of a scalar Higgs of positive parity.

Combining the results together, a pure pseudoscalar hypothesis has been excluded at 98% confidence level (CL), vector or pseudo-vector have been excluded at a 99.95%, and several spin-two boson hypotheses were excluded at a 98% C.L. or higher (2^+ model alone has been excluded at 99.99% CL) [24]. To further rule out the hypothesis of a vector Higgs is Lang's theorem, according to which an on-shell vector boson can not decay into a pair of on-shell photons. About charge conjugation, the Higgs boson is an eigentsate with eigenvalue C=+1, as it is observed decay into a $\gamma\gamma$ state [7] in a C-invariant theory. All of these reasons led to identify the Higgs as a $J^{PC} = 0^{++}$ particle, in agreement with Standard Model prediction. However, this does not definitely preclude the Higgs from being a CP-mixed state, endowed with a (currently unobserved) CP-odd component, arising from some bSM effect that could prevail at higher energy ranges. Probing such possibility is the purpose of the present thesis.

[7]Photons are eigenstates of the $\mathbb{C}$ operator with eigenvalue $C_\gamma = -1$

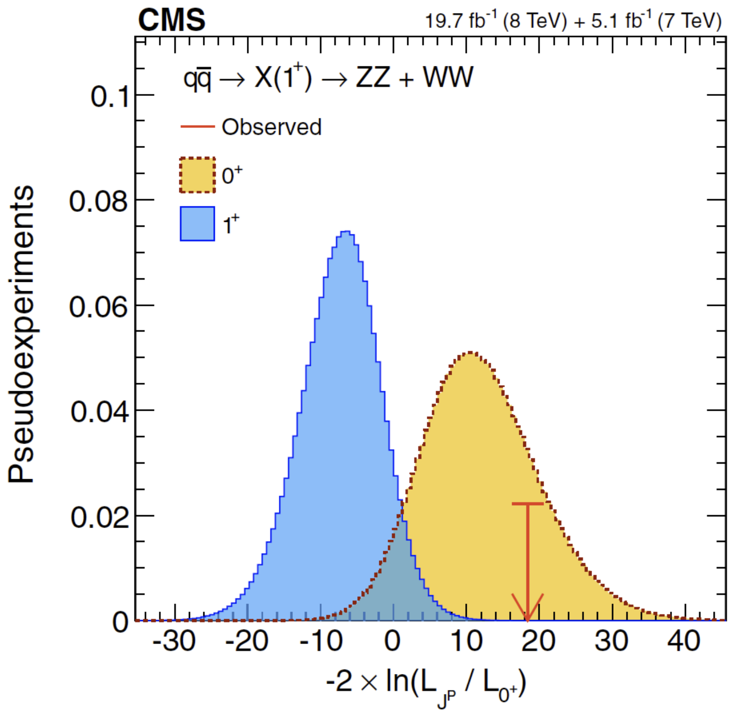


Figure 2.8. Distribution of test statistics $q = -2 \cdot ln(L_{J^P}/L_{0^+})$ for probing SM Higgs against pure vector Higgs hypotheses. The observed value falls on the tail of the 1^+ pdf, ensuring the rejection of vector Higgs hypothesis at 99.95% CL [24].

2.2 EFT framework

Effective Field Theories are widely used in physics to provide a simplified, yet comprehensive, description of a theory, for the parameters on which such a theory depends taking values on a subset of the parameter domain. EFT$_s$ serve as a crucial framework for parametrizing unknown high-energy beyond Standard Model effects in a systematic manner and constitute the so called SMEFT. This class of EFT$_s$ are known as bottom-up EFT$_s$ as they parametrize deviation from a known low-energy theory due to an uknwon UV theory, conversely to top-down EFT$_s$ which exploit a known high-energy theory to make predictions on the low-energy limit. It is worth noting, as highlighted in Section 2, that the SM itself constitutes an EFT, as it cannot be extrapolated to arbitrary high energies and must therefore be regarded as a low-energy approximation, at the characteristic EW scale $\Lambda = v = 246$ GeV, for a broader theory. This approach enables the integration of heavy dof$_s$ out of the theory, by adding further effective terms to the low energy Lagrangian, of which it constitutes an UV completion. A typical EFT Lagrangian can be expressed as:

$$\mathcal{L}_{EFT} = \mathcal{L}^{(4)} + \sum_i c_i^{(d)} \mathcal{O}_i^{(d)}(x) \tag{2.36}$$

where $\mathcal{L}^{(4)}$ is the low-energy, renormalizable Lagrangian. The additional,

effective term is an expansion in local operators of mass dimension $d > 4$, the *Wilson operators*. The acceptable local operators, which are some combinations of the fields, are all those preserving the symmetries of the low-energy theory. These operators are multiplied by the *Wilson coefficients*, $c_i^{(d)}$, which encode the virtual effects of heavy new physics in low-energy observables, scaling as:

$$c_i^{(d)} \propto \frac{\alpha^{n_i-2}}{\Lambda^{d-4}} \tag{2.37}$$

for tree level-generated Wilson operators, with Λ being the characteristic energy scale of the effective term and α the coupling of the UV theory constituted by n_i fields. For an EFT of known UV theory, the explicit form of Wilson coefficients is extracted by matching the EFT prediction of scattering amplitudes with the prediction of the theory, while for an EFT of unknown UV theory it generally occurs by making a power counting [26]. The contribution of each local operator $\mathcal{O}_i^{(d)}$ to amplitudes of physical processes at an energy scale of order E, scales as $(\mathrm{E}/\Lambda)^{d-4}$. Since $(\mathrm{E}/\Lambda) < 1$ by construction, the SMEFT in its validity domain describes small deviations for SM predictions. Therefore, a bottom-up EFT is just a power series expansion in (E/Λ) on top of the known low-energy Lagrangian.
The hereby studied bSM CP deviations from the 0^{++} SM-like Higgs, are described in terms of EFT$_s$. In the most general case, an effective Lagrangian could be written for any of the possible spin scenarios. However, as previously mentioned, the spin-0 hypothesis for the Higgs boson is highly privileged and it is the only one that is going to be considered. The construction of an effective Lagrangian for the spin-0 Higgs is obtained by requiring that the parametrization:

- Allows to easily recover SM case;
- Includes all the local operators above the EW scale compatible with gauge invariance;
- Accounts for CP mixing between $0^+$and 0^- states.

The effective term comprises only 6-dimensional operators, as d=5 operators violate lepton number L, and all the higher order odd-dimensional operators violate $B-L$ (with B baryon number) [27]. Moreover, d=8 operators contribution scales as $(\mathrm{E}/\Lambda)^4$, unlike d=6 operators which scale as $(\mathrm{E}/\Lambda)^2$, and they are therefore neglected. In the current scenario, Eq. 2.36 takes the simplified form:

$$\mathcal{L}_{EFT} = \mathcal{L}_{SM} + \frac{1}{\Lambda^2}\sum_i f_i O_i^{(6)} \tag{2.38}$$

The 6-dimensional Wilson operators taken into consideration are going to be manifestly $SU(3) \otimes SU(2) \otimes U(1)$ invariant operators, written in terms of the SM fields, conserving both lepton and baryon number. No constraint on CP parity is imposed, as parametrizing possible CP-odd components of the HVV vertex is right the purpose that is being pursued.

These $\mathcal{O}_i^{(6)}$ operators form a complete, minimal set of operators (a basis) which poses an arbitrary framework where to express the EFT.

This reasoning is going to be restricted to the EFT accounting for HVV anomalous couplings, which is the object of the present dissertation, however neglecting possible

$Hgg/H\gamma\gamma$ anomalous couplings as well as neglecting effective corrections to the other terms of SM Lagrangian. By writing the Effective Lagrangian for HVV couplings in terms of mass eigenstates, the small deviation of bSM Higgs scenarios from the the baseline SM case appears evident:

$$\begin{aligned}\mathcal{L}_{hvv} = \frac{h}{v}\Bigg[(1+\delta c_w)\frac{g^2v^2}{2}W^+_\mu W^-_\mu + (1+\delta c_z)\frac{(g^2+g^{'2})v^2}{4}Z_\mu Z_\mu \\ +c_{ww}\frac{g^2}{2}W^+_{\mu\nu}W^-_{\mu\nu} + \tilde{c}_{ww}\frac{g^2}{2}W^+_{\mu\nu}\tilde{W}^-_{\mu\nu} + c_{w\Box}\, g^2(W^+_\mu\partial_\nu W^-_{\mu\nu} + h.c.) \\ +c_{zz}\frac{g^2+g^{'2}}{4}Z_{\mu\nu}Z_{\mu\nu} + \tilde{c}_{zz}\frac{g^2+g^{'2}}{4}Z_{\mu\nu}\tilde{Z}_{\mu\nu} + c_{z\Box}\, g^2 Z_\mu\partial_\nu Z_{\mu\nu} \\ +c_{z\gamma}\frac{e\sqrt{g^2+g^{'2}}}{2}Z_{\mu\nu}A_{\mu\nu} + \tilde{c}_{z\gamma}\frac{e\sqrt{g^2+g^{'2}}}{2}Z_{\mu\nu}\tilde{A}_{\mu\nu} + c_{\gamma\Box}\, gg^{'}Z_\mu\partial_\nu A_{\mu\nu}\Bigg]\end{aligned} \tag{2.39}$$

where the coefficients hold the $\mathcal{O}(\Lambda^{-2})$ dependence [28]. The $\delta c_w, \delta c_w$ are the deviations to SM Higgs couplings to W/Z bosons, while the further terms implement the effective terms that introduce a tensorial structure for the HVV vertex which is not present in the SM. The SM Lagrangian for the HVV vertex (in which at least one V = W,Z) is thus recovered by imposing $\delta c_W = \delta c_Z = 0$ and all the other $c_i = 0$. The tilde terms are contracted with the - totally anti symmetric - Levi-Civita tensor and parametrize CP-odd terms. The $c_{Z\gamma}$ ($c_{\tilde{Z\gamma}}$) parametrize a $HZ\gamma$ vertex, realized at loop order, accounting for a possible variation of the $W^{(3)}$-B mixing with respect to the SM. $c_{(z/w/\gamma)\Box}$ account for 2-derivative interactions. Despite exhibiting a clear structure, manifestly showing Lagrangian terms parametrizing small variations on top of the already knwon SM terms, the 2.39 Lagrangian is not ideal for having coefficients to probe experimentally. This leads to the introduction of another basis, the *Higgs basis*, in which the d=6 EFT expansion is spanned by a subset of the mass eigenstate coefficients. Indeed, it can be shown that not all the coefficients reported in 2.39 are independent, and the Higgs basis provides a minimal independent subset to span the effective operators that can be more directly connected to observable quantities. On the top of this basis, the so-called *κ-formalism* is introduced [29]. The *κ-formalism* characterizes the Higgs anomalous couplings in terms of the κ coefficients, to be meant as coupling strength modifier parameters defined as the ratios of the HVV couplings to their SM value. The κ framework assumes a single narrow resonance so that the zero-width approximation can be used to decompose the cross section as product two factors characterising the production and the decay of the Higgs boson. The κ parameters are introduced by expressing each of the these factors as their SM expectation multiplied by the square of a coupling strength modifier for the corresponding process at leading order [10]:

$$(\sigma\cdot BR)(i\to H\to f) = \frac{\sigma_i^{SM}\kappa_i^2\cdot\Gamma_f^{SM}\kappa_f^2}{\Gamma_H^{SM}\kappa_H^2} \longrightarrow \mu_i^f \equiv \frac{\sigma\cdot BR}{(\sigma\cdot BR)_{SM}} = \frac{\kappa_i^2\cdot\kappa_f^2}{k_H^2} \tag{2.40}$$

where κ_H is a parameter accounting for the Higgs boson width adjustment due to anomalous couplings, and μ is the *signal strength*, i.e. the ratio of the cross section by branching ratio product, modified by anomalous couplings, to the SM expectation for the analogous quantity. These κ coefficient are directly related to the experimental observable quantities corresponding to Higgs boson production and decay modes, and can be explicitly expressed in terms of the EFT coefficients of $\mathcal{L}_{hvv}$ Lagrangian 2.39.

2.2.1 HVV scattering amplitude

In the κ-formalism the most general scattering amplitude for a HVV vertex can be computed in a clean way [21] [8]:

$$\mathcal{A}(HV_1V_2) = \frac{1}{v}\left[a_1^{VV} + \frac{\kappa_1^{VV} q_{V1}^2 + \kappa_2^{VV} q_{V2}^2}{(\Lambda_1^{VV})^2} + \frac{k_3^{VV}(q_{V1}+q_{V2})^2}{(\Lambda_Q^{VV})^2}\right] m_{V1}^2 \epsilon_{V1}^* \epsilon_{V2}^* \\ + \frac{1}{v} a_2^{VV} f_{\mu\nu}^{*(1)} f_{\mu\nu}^{*(2)} + \frac{1}{v} a_3^{VV} f_{\mu\nu}^{*(1)} \tilde{f}_{\mu\nu}^{*(2)} \tag{2.41}$$

where m_V, q_V, ϵ_V are, respectively, the vector boson mass, its four-momentum and its polarization vector. $f^{(i)\mu\nu} = \epsilon_{Vi}^{\mu} q_{Vi}^{\nu} - \epsilon_{Vi}^{\nu} q_{Vi}^{\mu}$, $\tilde{f}_{\mu\nu}^{(i)} = \frac{1}{2}\epsilon_{\mu\nu\rho\sigma} f^{(i),\rho\sigma}$ provide the tensorial structure of the last two terms. Λ_1 and Λ_Q are the new energy physics scales. Apart from a HVV (V=W,Z) vertex, Eq. 2.41 holds also for a Hgg, which however is not going to be considered. The coefficients appearing in the amplitude formula are real and acquire the following physical meaning:

- a_1^{VV} is the CP-even coefficient paramerizing the SM vertex, according to which $a_1^{ZZ} = a_1^{WW} = 2$;
- $\kappa_1^{VV}, \kappa_2^{VV}$ parametrize a CP-even interaction dependent by the energy scale Λ_1;
- κ_3^{VV} is the coefficient of a CP-even HVV contact interaction;
- a_2^{VV} parametrize a tensorial CP-even vertex;
- a_3^{VV} parametrize a tensorial CP-odd vertex;

By imposing symmetry and gauge invariance arguments, some constraints on the parameters are imposed:

$$\kappa_3^{VV} = a_1^{Z\gamma} = \kappa_1^{Z\gamma} = 0, \;\; \kappa_1^{ZZ} = \kappa_2^{ZZ}, \;\; \kappa_1^{WW} = \kappa_2^{WW} \tag{2.42}$$

For V=W,Z there are therefore 11 independent parameters for accounting Higgs anomalous couplings to vector bosons:

$$\underbrace{a_1^{ZZ}, a_1^{WW}}_{0PM}, \underbrace{\kappa_1^{WW}, \kappa_1^{ZZ}}_{L1}, \underbrace{\kappa_2^{Z\gamma}}_{L1Zg}, \underbrace{a_2^{ZZ}, a_2^{WW}, a_2^{Z\gamma}}_{0PH}, \underbrace{a_3^{ZZ}, a_3^{WW}, a_3^{Z\gamma}}_{0M} \tag{2.43}$$

[8] This writing of the amplitude is done by using the anomalous amplitude decomposition approach. It is however fully equivalent to the parametrization provided in Eq. 2.39. As a matter of convention, some coefficients are written as a_i coefficients instead of κ_i coefficients.

Eq. 1.42 The underbraced text highlights the naming convention adopted by the Monte Carlo samples that are employed in the analysis, accounting for different Higgs scenarios for the anomalous coupling: SM (0PM), Λ_1-dependent (L1), Λ_1-dependent $Z\gamma$ vertex (L1Zg), scalar higher-order corrections (0PH) and pseudoscalar (0M).

In the narrow width approximation, the sole cross section for a $(i \to H \to f)$ process can be written in terms of a factorization of initial/final state cross sections, which can themselves be written as a sum of contributions arising from different Higgs scenarios, parametrized by the anomalous couplings in 2.41:

$$\sigma(i \to H \to f) \propto \frac{\left(\sum \alpha_{jk}^{(i)} a_j a_k\right)\left(\sum \alpha_{lm}^{(f)} a_l a_m\right)}{\Gamma_{tot}} \tag{2.44}$$

where a_i are a general way to denote both a_i and κ_i coefficients of 2.41 and $\alpha_{ij}^{(i/f)}$ are energy-dependent coefficient which could be computed from simulation [29]. From 2.44, any single coupling a_n can be parametrized in terms of an effective cross section ratio f_{a_n}:

$$f_{a_n}^{(i,f)} = \frac{\alpha_{nn}^{(i,f)} a_n^2}{\sum_m \alpha_{mm}^{(i,f)} a_m^2} \times sign\left(\frac{a_n}{a_1}\right) \tag{2.45}$$

corresponding to the fractional contribution of the the coupling a_n to the total cross section of the $(i \to H \to f)$ process.

Together with the previously introduced signal strength $\mu_i = (\sigma{\cdot}BR)_i/(\sigma{\cdot}BR)_i^{SM}$, the f_{a_i} fractional cross sections provide a direct formulation to be probed experimentally. Being expressed in terms of ratios of cross sections, they are indeed invariant with respect to the coupling convention. Moreover, several statistical uncertainties cancel out in the ratio.

Chapter 3

LHC and the CMS experiment

3.1 LHC

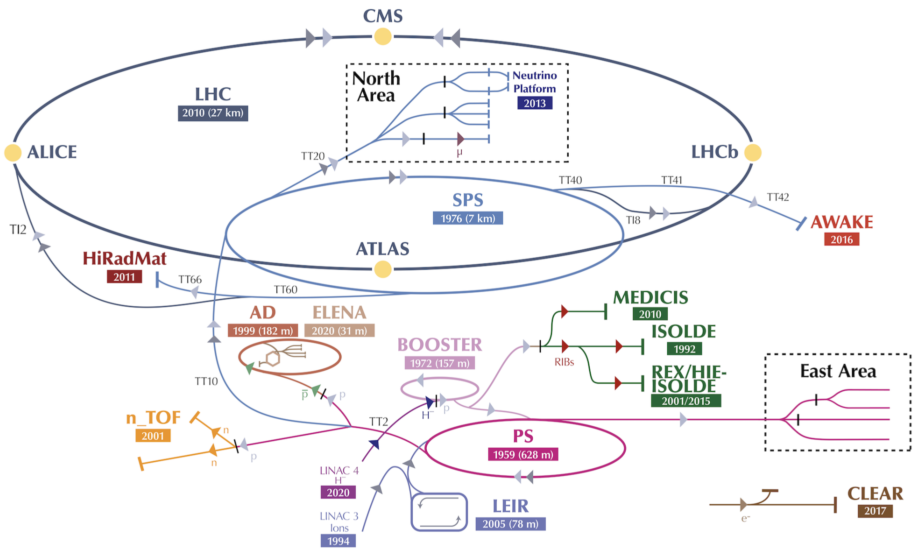


Figure 3.1. CERN's accelerators injection complex, whose final stadium is the LHC [33]

The Large Hadron Collider (LHC) is a 27-kilometer long proton-proton synchrotron collider that accelerates proton beams up to an energy of 6.8 TeV, resulting in a back-to-back collision at a center-of-mass (CoM) energy of up to 13.6 TeV [1], making it the most powerful accelerator in the world. The LHC had been primarily designed to discover the Higgs boson, which was successfully achieved in 2012. However, the LHC now hosts several experiments that aim to both make precision measurements of Standard Model predictions and search for beyond-the-Standard-Model (bSM) physics. The beams in the LHC undergo a multi-step acceleration process where they transit through a sequential chain of accelerators, gradually increasing their energy at each step, until reaching the aforementioned capped energy value.

[1] This is the recorded energy limit measured during the Run-3 of LHC, which is not the focus of the present analysis

The first step in the chain of accelerators is the Linear Accelerator *Linac4*, which accelerates negatively charged hydrogen ions (H^-), endowed with an additional electron, to a momentum of 16 MeV, before sending them to the *Proton Synchrotron Booster* (PSB). The hydrogen ions are then stripped of their two electrons, leaving only protons that are accelerated to 2 GeV, after which they enter the *Proton Synchrotron* (PS), where the beam's momentum is further enhanced to 26 GeV. Subsequently, the protons are directed to the *Super Proton Synchrotron* (SPS), which accelerates them up to 450 GeV. At this stage, the proton beams are injected into the two beam pipes of the LHC, which are kept in an ultra-vacuum state, where they circulate in opposite directions and are accelerated to (almost) the nominal energy. It takes approximately 4 minutes to inject the proton beams, which are produced in packs, also called *bunches*, into the LHC ring and around 20 minutes to make them reach the asymptotic energy value.

Proton bunches are characterized both by a longitudinal length (*bunch length*), and by a lateral spread (*bunch radius*), which should be thought in terms of the area of the beam section in the the transverse plane with respect to the ideal flight axis, also called *focal plane*. Once the proton bunches are inside the ring, there is a complex system of superconducting magnets - operating slightly above the absolute zero temperature - that allow to modify the beam trajectory both on the ring plane and the transverse plane with respect to their flight direction:

- *Dipole magnets*: in order for a charged particle to bend - by Lorentz law - and be guided into an overall circular trajectory, dipole magnets must be employed. These magnets generate a perpendicular magnetic field with respect to the ring plane, deflecting the proton bunches trajectory within the beam pipe;

- *Quadrupole magnets*: during the travelling in the ring, proton bunches could deviate from the ideal spot in the focal plane. The correction of these possible deviations in the transverse plane are achieved through magnetic quadrupoles, which work similarly to a spring: the further the bunch gets from the ideal focal point, the greater the force the quadrupole exerts to bring it back to the optimal position. Quadrupole magnets ensure that the beam is well focused, minimizing its bunch radius and thus maximizing the number of interactions per bunch, leading to a greater luminosity.

- *Sextupole magnets* and higher order multipoles serve to make fine adjustments of the beam and to correct further imperfections. Sextupole magnets, for instance, correct possible momentum deviations, thereby ensuring the monochromaticity of the beam.

  As the proton bunches are bent through the ring, they lose energy due to synchrotron radiation, requiring a proper supply of energy to keep on traveling at the nominal energy. This is achieved by utilizing *radiofrequency (RF) cavities*, consisting in metallic chambers which contain an electric field. The field in an RF cavity oscillates at a fixed frequency of 400 MHz, so timing the arrival of particles is important. Depending on the phase ϕ at which each proton reaches the RF cavity, it receives an electric impulse proportional to

$|sin\phi|$, which accelerates it. As a result, protons totally in phase are not accelerated, while protons with a different energy, thus reaching the RF cavity off-phase, receive some energy transfer that tend to bring them back on phase.

Circumference	26659 m
Nominal energy, proton	7 TeV
Number of dipoles	1232
Number of quadrupoles	858
Peak magnetic dipole field	8.33T
Dipole operating temperature	1.9K
RF	400.8 MHz
Bunch spacing	25 ns
Bunch length σ_z	75 mm
Bunch radius $\sigma_x = \sigma_y$	16 μm
Design luminosity	$10^{-34} cm^{-2} s^{-1}$

Table 3.1. LHC design parameters

The istanteneous luminosity produced is equal to:

$$L = \frac{n_b N^2 f}{4\pi\beta^* \epsilon} \gamma R \tag{3.1}$$

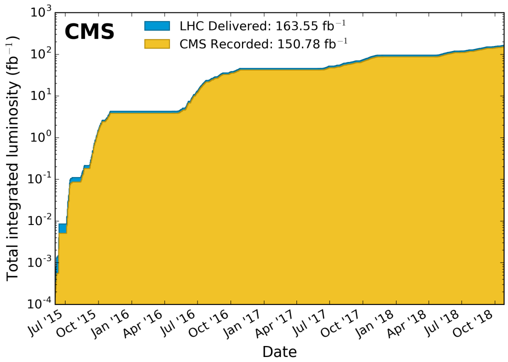
where γ is the relativistic factor, n_b is the number of bunches at the interaction point, N is the number of protons per bunch, f is the revolution frequency in the ring, β^* is the beam focal length, ϵ is the beam transverse normalized emittance and R is a luminosity geometrical reduction factor. Luminosity only depends on parameters of the detector, hence being an intrinsic parameter of the detector itself. Another way of writing luminosity is in terms of *integrated luminosity*, i.e. the luminosity integrated over a fixed time interval of data-taking:

$$\mathcal{L} = \int L dt \tag{3.2}$$

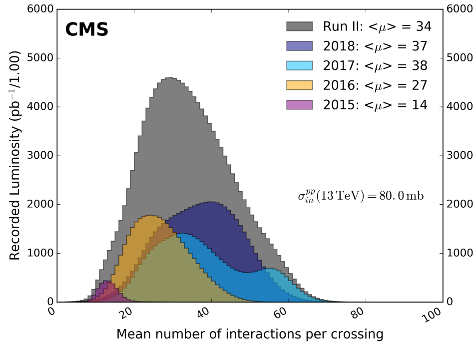
Let σ_i be the cross-section of some process occurring in the LHC, the number of expected of expected events for the process, given the integrated luminosity $\mathcal{L}$ and the efficiency ϵ for the process final state detection in the detector, is:

$$N_i = \sigma_i \cdot \mathcal{L} \cdot \epsilon \tag{3.3}$$

At the present day, LHC has completed 2 whole Runs of data-taking, with the third one currently underway:



((a)) Cumulative LHC delivered/CMS acquired luminosity



((b)) Pile-up distribution in the CMS experiment, divided by year

Figure 3.2. Run-2 overview of the event generated by LHC and acquired by CMS [34]

- Run-1: in 2011 (at 7 TeV) and 2012 (8 TeV);
- Run-2: from 2015 to 2018 (at 13 TeV);
- Run-3 from 2022 to 2025 (at 13.6 TeV).

After Run-3 is completed, the LHC will undergo the Long Shutdown 3, focused on detector upgrades in view of High Luminosity-LHC (HL-LHC) that will start operating in 2029 at an instant luminosity at least a order of magnitude greater than the current. This will require CMS, as well as the other CERN experiments, to comply to the HL-LHC by updating the detectors to make them faster, more granular and resistant, in order to cope with the higher pile up and integrated radiation dose.

3.2 The CMS experiment

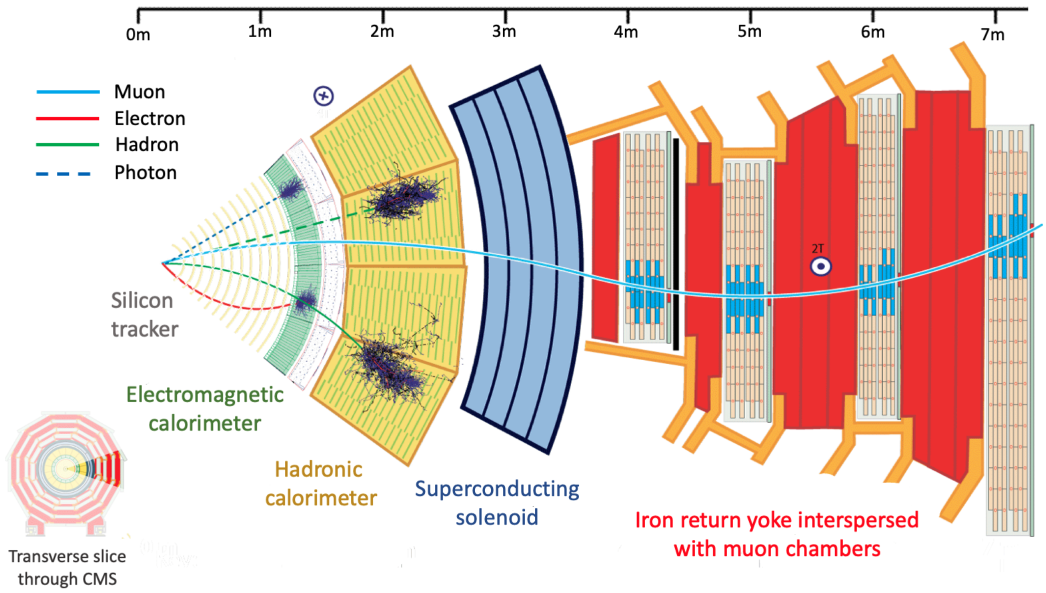
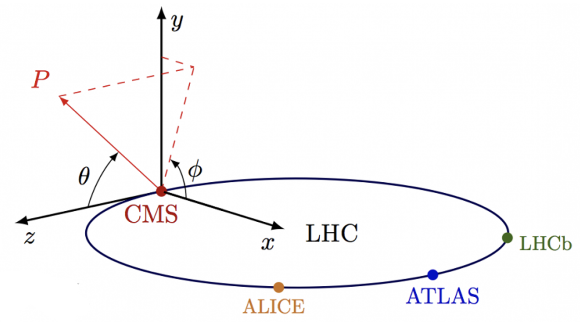


Figure 3.3. Transverse section of the CMS detector, highlighted in its components, and the flow of different particle types through it.

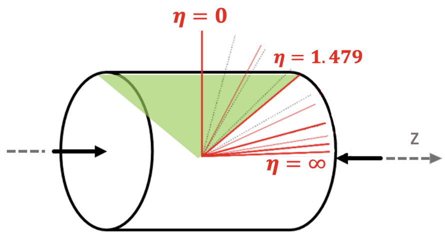
The Compact Muon Solenoid (CMS) is a general purpose detector, designed to investigate a wide range of physics, proving the existence of a Higgs boson together with ATLAS experiment. The detector consists in a 21.6m long cylinder barrel, of 14.6m of diameter, closed at its ends by two endcaps. Figure 3.6, shows the concentric layer structure of the detector, with the inner layer being the silicon tracker, followed by an electromagnetic calorimeter (ECAL), an hadronic calorimeter (HCAL) and a superconducting solenoid producing a 4T magnetic field for deflecting charged particles and measure their momentum according to the track deflection observed in the tracker. Everything is embedded in a iron return yoke interspersed with muon chambers. The beam pipe passes through the detector, perpendicularly to the transverse plane shown in Figure 3.6, defining what is conventionally identified as the z-axis. The x-axis is taken as the radial coordinate in the LHC ring plane and the y-axis as the normal unit vector to such a plane (Figure 3.4 (a)). Due to the cylindrical geometry of the detector, it is useful to introduce a cylindrical coordinate system (r, η, ϕ), where ϕ is the azimuthal angle spanning the transverse plane of Figure 3.6, and η the *pseudorapidity* defined as a function of the polar angle θ with respect to the beam direction x:

$$\eta = -ln\left(tan\frac{\theta}{2}\right) \tag{3.4}$$

As for Figure 3.4 (b), for the ECAL, the barrel region is identified by the condition $|\eta| < 1.479$, whereas the endcaps are comprised in the $|\eta| > 1.479$ region.



((a)) Conventional Cartesian reference frame on the LHC ring plane



((b)) Pseudorapidity mapping in the cylindrical ECAL shape

Figure 3.4

The angular distance in the η-ϕ space is defined as:

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \tag{3.5}$$

3.2.1 Trigger

At the peak nominal luminosity of LHC, the event rate reaches $\mathcal{O}(10^9)$Hz. As the typical storage size of an event is about 1 MB, storing all the events in real time is not a viable solution, neither is it useful, as a large fraction of the events is due to QCD background, without having a particular relevance in the search for new physics. In CMS, a two-level trigger is implemented to properly select events of potential physics interest [35]:

- The **Level-1 Trigger (L1)** is implemented in hardware, and selects events containing detector signals consistent with an electron, photon, muon, τ lepton, jet, or missing transverse energy. The trigger thresholds are adjusted to the LHC instantaneous luminosity during data taking in order to restrict the output rate to 100 kHz, the upper limit imposed by the CMS readout electronics. This trigger has a fixed latency of $\sim 4\mu s$, in which it has to decide which data are worth to be collected. Since this time window is not large enough to process the information of the whole detector, the L1 trigger only processes information from the calorimeters and the muon chambers;

- The **High-Level Trigger (HLT)** is a software-implemented trigger system that runs high-level physics algorithms to further refine the purity of the output stream, selecting an average rate of 400 Hz for offline event storage. It is split into three sublevels: the first only access data from the calorimeters and the muon chambers, the second adds the tracker information and the third processes the whole information of the event.

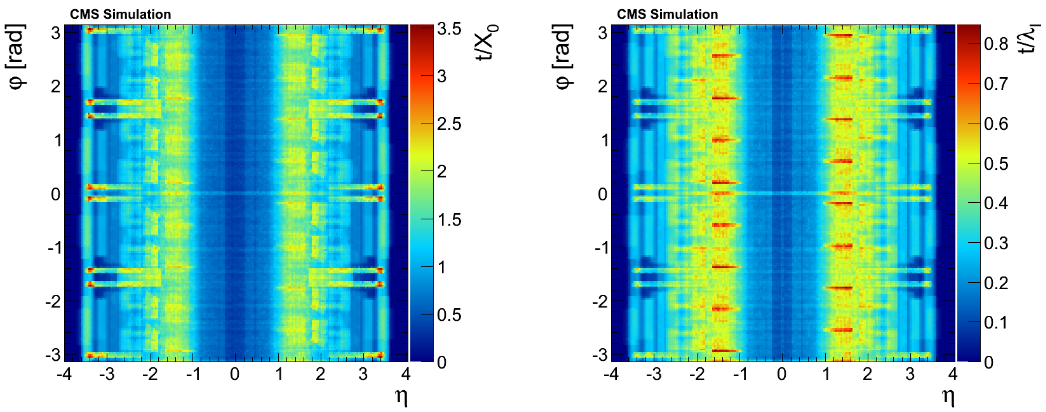
3.2.2 Silicon tracker

The tracker [36] is the innermost part of the detector, the closest to the beam pipe and the Interaction Point (IP). It consists of silicon-based sensors for detecting the track of charged particles with $|\eta| < 2.4$ - which deposit energy in it by ionization - from whose deflection the momentum can be extracted, as they experience a track deflection according to a radius $R = p_T/0.3B$. The great challenge of the tracker is being enough light as not to contaminate the kinematics of the particles, yet being accurate in tracking particles and hard resisting to the huge amount of radiation it is exposed to, which however decreases as the radial distance from the IP increases. The CMS tracker is composed of different substructures:

- **Pixel sensors**: the innermost part of the tracker, is constituted by about 66 million $100 \times 150 \mu m^2$ pixels arranged in a radial distance range from 3 to 16 cm with respect to the beam pipe;

- **Strip sensors**: after passing through the pixel sensors, particles encounter the several layers of $80 - 100 \mu m$ wide silicon strips, which extend up to a radius of 130 cm, for a total of 10 million channels.

The particles traveling through these several steps of highly segmented silicon tracker detectors, produce small electric signals which are amplified, stored in michrochip memory for several microseconds, then processed by readout electronics, converted into infrared pulses, transported by an optic fiber system in a radiation safe environment and analyzed.



((a)) Tracker path length in the η-ϕ space in units of radiation length

((b)) Tracker path length in the η-ϕ space in units of hadronic interaction length

Figure 3.5

The tracker material thickness, either in terms of radiation length (X_0) or hadronic interaction length (λ_I) is mostly uniform in the azimuth angle ϕ, but strongly dependent on the pseudorapidity η, as clearly shown in Figure 3.5, with the barrell-endcap transition being the most delicate, photon conversion-prone, zone of the detector. The material thickness is indeed very important, as it is proportional to the probability for a photon to be converted in an electron-positron pair inside it, thus triggering an electromagnetic shower before entering the ECAL and reducing its energy reconstruction precision.
The high segmentation of the silicon tracker allows the detector to get position

measurements with a precision as high as 10 μm, which leads to making momentum measurements with high precision, as the relative uncertainty on transverse momentum measurement is given by:

$$\frac{\sigma(p_T)}{p_T} = \frac{\sigma(x)p_T}{0.3BL^2}\sqrt{\frac{720}{N+4}} \tag{3.6}$$

where B is the magnetic field intensity, L the distance between the first and the last sampled points on the track, N the number of sampled points on the track, and $\sigma(x)$ right the uncertainty on position measurements.

3.2.3 ECAL

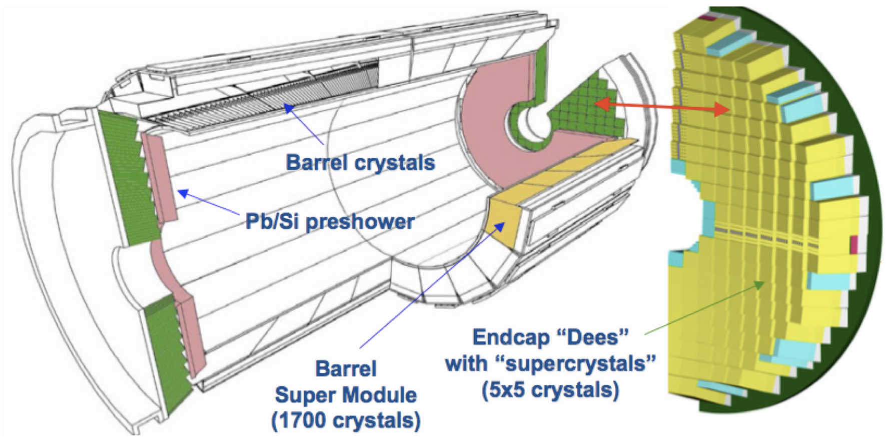


Figure 3.6. Section of the CMS ECAL

The Electromagnetic Calorimeter [37] of the CMS experiment is a hermetic, fine grained, homogeneous calorimeter, containing 75,848 lead tungstate ($PbWO_4$) scintillating crystals. The ECAL purpose is to measure the energy of particles triggering EM showers inside it, i.e $e^{\pm}$ and γ and it does so with high precision. The ECAL design in CMS had been guided from the start by the idea of optimizing the detection of the $H \to \gamma\gamma$ channel in the search for the Higgs boson, implying that, also nowadays, the accuracy with which the diphoton final state is reconstructed is excellent. The main ECAL design requirements were:

- Excellent energy and position/angle resolution up to $|\eta| < 2.5$, to match the tracker coverage;
- Hermeticity, compactness and high granularity;
- Fast response;
- Large dynamic range (5 GeV to 5 TeV) and excellent linearity;
- Radiation resistence.

As for the whole CMS detector, the ECAL has a cylindrical structure. It is constituted by a barrel (EB), consistsing of 61200 crystals of lead tungstate, also called *stolzite*, closed by 7324 crystals in each endcap (EE). The choice for stolzite scintillating crystals arises from the specific properties of the material: high density (8.26 g/cm^3), small radiation legnth (X_0 = 0.89 cm) and Muliere radius (R_M = 2.2 cm) allow for shower containment in a relatively compact volume. $PbW0_4$ crystals exhibit an emission peak at 440 nm, in the blue region, and a small time constant ($\tau = 15ns$) that produce a fast response, with around 80 % of the scintillation light emitted within 25 ns. Moreover, stolzite crystals are relatively easy to grow and have good resistance to radiation. On the other hand, they exhibit reduced light yield, only 100γ /MeV for a 23 cm crystal - requiring the coupling to photodetectors with high internal gain - and a strong dependence of the light yielded from temperature, $\Delta(LY)/\Delta T = -2\%/°C$ at $T > 18°C$, which imposes strict requirements on temperature stability.
The EB covers the region $|\eta| < 1.479$ and is composed by 36 *supermodules*, of 1700 crystals each. The crystals have trapezoidal shape of about 23 cm (25.8 X_0) in length with a transverse section of 22 × 22 mm^2 in the frontal extremity, and of 26 × 26 mm^2 in the opposite extremity (1-1.2 R_M). The crystals are arranged in a quasi-projective geometry, and employ avalanche photodiodes (APD) for photo-detection. The EEs consist of two detectors, a preshower (ES) device followed by a the $PbWO_4$ scintillator calorimeter, coupled to vacuum phototriodes (VPTs) for photo-detection. The preshower detector is used for particle identification in the EE regions of CMS. Each ES is made of two orthogonal layers of silicon sensors, interspersed with lead layers that serve to generate electromagnetic showers. The main aim of the preshower detector is to distinguish photons from pions (π^0) that decay in two close photons at high energy, being thus difficult to be discriminated. Each endcap calorimiter is made by 7324 rectangular and quasi-projective crystals of 1.3 R_M lateral and 24.7 X_0 longitudinal size. Notice that in the region $|\eta| > 2.5$, both the radiation level and high particle multiplicity forbid precision measurements.
ECAL energy resolution depends on both the particle energy and some intrinsic parameter of the calorimeter:

$$\frac{\sigma}{E} = \frac{S}{\sqrt{E}} \oplus \frac{N}{E} \oplus C \tag{3.7}$$

where S is the stochastic term, N is the noise term and C is the constant term characterizing the calormeter. The symbol denotes the sum in quadrature. The stochastic term depends on the fluctuations in the number of scintillating photons detected and in the number of processes through which the particles lose their energy in the crystals. The noise term comes from the electronic noise and the pileup. The constant term has different causes: losses due to failures in the longitudinal containment, non uniformity in the light collection, intercalibration between crystals and geometrical imperfections. The parameters appearing in Eq.3.7 have been extrapolated in test beams, and are estimated as:

$$S = 2.8\% \sqrt{\text{GeV}} \qquad N = 124 \text{ GeV} \qquad C = 0.3\% \tag{3.8}$$

3.2.4 HCAL

The hadronic calorimeter employed in CMS is a crucial component for measuring the energy and direction of particles within hadronic jets, and for estimating the missing energy of events in conjunction with the electromagnetic calorimeter. To fulfill these objectives, the hadronic calorimeter must possess good hermeticity and transverse granularity. Additionally, the resolution of the energy measurement must be of high quality, and there must be sufficient longitudinal containment of the hadronic showers.
The HCAL is comprised of a central calorimeter ($|\eta| < 3$) and two calorimeters situated at high pseudorapidities ($3 < |\eta| < 5$). It was chosen to use a sampling calorimeter consisting of copper absorber layers and plastic scintillators as active material. The calorimeter possesses a tile structure with tiles parallel to the beam axis. It is segmented into a central cylindrical structure ($|\eta| < 1.3$) and two endcaps ($1.3 < |\eta| < 3$), consisting of a total of 2593 trigger towers without longitudinal segmentation. The calorimeter possesses a granularity of $|\Delta\eta| \times |\Delta\phi| = 0.087 \times 0.087$, which corresponds to the granularity of the trigger towers of the electromagnetic calorimeter (ECAL).
The central calorimeter has a depth of about 7 interaction lengths λ_I and provides an energy resolution given by:

$$\frac{\sigma}{E} = \frac{S}{\sqrt{E}} \oplus C \tag{3.9}$$

with S = 125% stochastic term and C= 5% constant term.
However, a 7 λ_I depth is insufficient to fully contain hadronic showers longitudinally. Thus, an additional layer was included behind the solenoid to provide 3 additional λ_I, resulting in a 10% improvement in the energy resolution for pions of 300 GeV.
The calorimeters at high pseudorapidity values, situated in a radiation-rich and highly-multiplicative environment, are sampling calorimeters composed of iron and quartz chambers. The chambers come in two distinct lengths, with the longer chambers starting from the frontal face of the calorimeter, while the shorter ones start 22 cm from the longer ones. This allows for subtraction of the electromagnetic component of the shower, which is deposited in the initial part of the calorimeter.
The calorimeters consist of a total of 1728 trigger towers and possess a granularity of $|\Delta\eta| \times |\Delta\phi| = 0.175 \times 0.175$.

3.2.5 Muon chamber

The muon detection system is responsible for identifying and measuring muons, particularly penetrating particles capable of passing through the calorimeters without being absorbed. The presence of muons in the final state is a hallmark of numerous physical processes, rendering the performance of the muon detector critical. Specifically, the $H \to ZZ \to 4\mu$ decay channel imposes the most rigorous demands on the muon detector functionality. The muon detector is situated beyond the magnet and covers the pseudorapidity domain $|\eta| < 2.4$. It comprises a barrel and two endcaps, each composed of four measurement stations interspersed with the iron return yoke of the magnet. In the barrel region, planes of drift tubes form the system, with each station consisting of a chamber constructed from 12 planes of tubes, amounting to a total of 195000 tubes. The endcaps incorporate cathodic strip chambers (CSCs) to provide precise measurements despite strong magnetic fields and high particle multiplicities. CSCs are multiwire proportional chambers with segmented cathode planes in strips, arranged in modules of six layers. In both the barrel and endcaps, resistive plate chambers (RPCs) serve as triggers. RPCs are gas chambers with parallel planes offering discrete spatial resolution and excellent time resolution (3 ns), similar to that of a scintillator. They form a swift trigger system, proficient in identifying candidate muons with high efficiency. The barrel contains six stations of RPCs, while each endcap contains four stations, resulting in a total of 612 chambers.

The *Particle flow* algorithm [38] aims to reconstruct each particle inside an event, by correlating the basic elements from all detector layers (tracks and clusters) to identify each final-state particle, and by combining the corresponding measurements to reconstruct the particle properties on the basis of this identification.

Chapter 4

Signal selection

The current thesis aims at investigating Higgs anomalous couplings in VH production mode, using at the $H \to \gamma\gamma$ decay channel. As discussed in Section 2.1.4 this decay channel, despite having quite a small branching ratio, yields a clear signature in CMS: the high reconstruction efficiency and great mass resolution - of the order of 1% - are mainly due to the performance of the ECAL, illustrated in Section 3.2.3. In fact, the diphoton invariant mass is calculated as:

$$m_{\gamma\gamma} = \sqrt{2E_{\gamma_1}E_{\gamma_2}(1 - cos\theta_{\gamma\gamma})} \tag{4.1}$$

where $E_{\gamma_{1,2}}$ denotes the energy of the photons and $\theta_{\gamma\gamma}$ the angle between the two. The resolution can be computed as:

$$\frac{\sigma(m_{\gamma\gamma})}{m_{\gamma\gamma}} = \frac{1}{2}\left[\frac{\sigma(E_{\gamma_1})}{E_{\gamma_1}} \oplus \frac{\sigma(E_{\gamma_2})}{E_{\gamma_2}} \oplus \frac{\sigma(\theta_{\gamma\gamma})}{tan(\theta_{\gamma\gamma})/2}\right] \simeq 1\% \tag{4.2}$$

so, the excellent ECAL energy resolution of about $\sigma(E)/E \sim 1\%$ translates to $\sim 1\%$ mass resolution, provided that the direction of the photons can be measured with good precision. Regarding the final state associated to the vector boson produced jointly to the Higgs, the focus will be on its hadronic decay channel, for the reason exposed in Section 1. The overall final state characterizing the V($\to$jj)H$\to \gamma\gamma$ process under study therefore comprises one diphoton and one dijet. The (2γ+2j) final state is the same as other H production mechanisms, as it is also common to VBF/ttH production modes as well as to gluon fusion with two jets initiated by QCD gluon radiation. The VH production can be distinguished by the VBF production, which results in the same final state, but with a cross section $\sim$ 10 times larger, because the former is characterized by two almost central jets, with an invariant mass peaked at the mass of the decaying vector boson. The first aim of this work is characterizing the kinematics of the final state, which is fundamental to make a discrimination with respect to other production mechanisms. Maximizing this separation helps to increase the sensitivity to this production mode, and thus the precision on the VH coupling. This categorization is hereby achieved by implementing a multivariate analysis algorithm, which will be described in the next chapter.

4.1 Background sources

The fit process for extracting the signal may be afflicted by several sources of background, either resonant or not. By *resonant background* is meant all the background processes whose $m_{\gamma\gamma}$ distribution is peaked at 125 GeV. In other words, resonant background is constituted by all the Higgs production mechanisms but VH (Figure 4.1). Disentangling the contribution of those different production mechanism to the VH signal will require a multivariate analysis algorithm as well as a dedicated fit procedure.

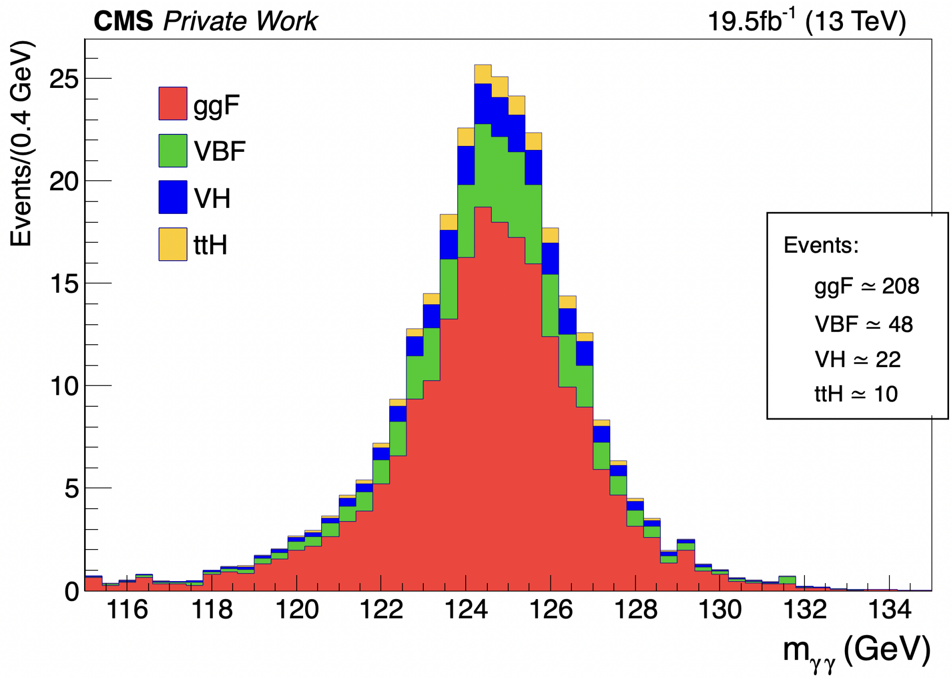


Figure 4.1. Stack histogram for the diphoton invariant mass, where the contributions to the peak from VH signal and resonant backgrounds are highlighted. The distribution is plotted from Monte-carlo samples referring to a limited data-taking period of Run-2, corresponding to an integrated luminosity of 19.5 fb^{-1}. The events are subject to the preselection described in the following sections 4.2, 4.3, in addition to the requirement of two jets in the event.

Another kind of background is the *non-resonant background*, constituted by those processes that exhibit a diphoton final state, whose invariant mass distribution however, is not peaked on Higgs mass but defined on a broad energy range and smoothly decreasing. The three sources of non resonant background arise from the following processes, listed below in decreasing order of cross section:

- $pp \to jj$ where both of the jets are mis-identified at reconstruction stage as photons: this is the most probable process at LHC. However, as the probability of mis-identifying a jet as a photon is $\mathscr{P}(j \to \gamma) \sim \mathcal{O}(10^{-5})$, the chance of mis-identifying both the jets is $\mathscr{P}(jj \to \gamma\gamma) \sim \mathcal{O}(10^{-10})$, making this kind of background negligible;

- $pp \to \gamma + j$ with the jet mis-identified as a photon: in this case the reduction factor is just $\mathcal{O}(10^{-5})$ which makes this scenario not totally negligible. Nevertheless, as the cross section of this process is not predicted theoretically with high precision and the modelling of the interaction of the hadronic shower inside the calorimeters is not trivial to simulate in a Monte-Carlo, this background is not considered.

- $pp \to \gamma\gamma$: despite being the process with the smaller cross section, as it only occurs at loop order, this is the only source of non-resonant background that will be considered for BDT training because it produces real photons in the final state and thus it is selected with large efficiency by the photon ID.

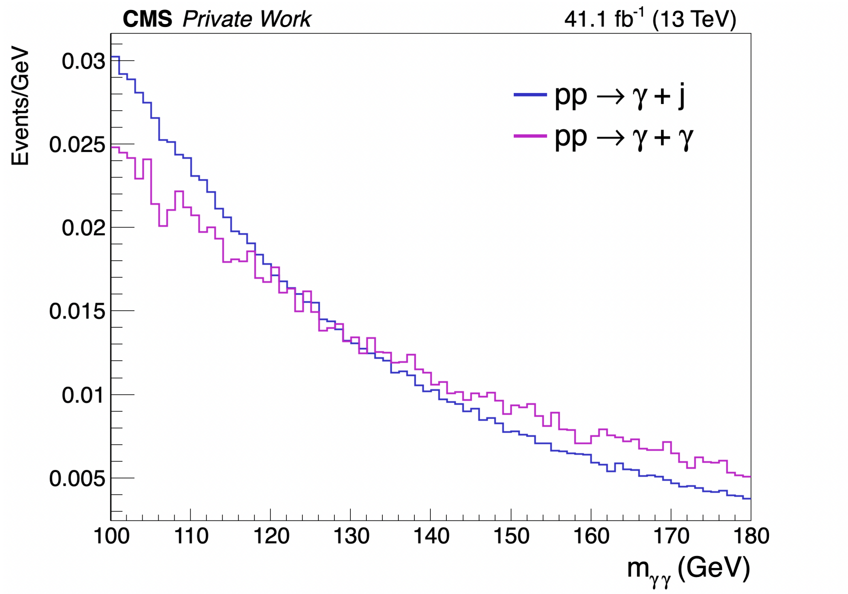


Figure 4.2. Diphoton invariant mass distribution shape-comparison between $pp \to \gamma\gamma$ and $pp \to \gamma + j$. The histograms are normalized to unity. The Monte-Carlo samples employed are from the 2017 era of Run-2.

Neglecting the $pp \to \gamma + j$ background in the optimization of the analysis can be justified as follows. The total contribution of this kind of background to the inclusive non-resonant processes is expected to be small in terms of event rate (15-20% at most to the overall non-resonant background, to give a ballpark number). The crucial aspect is that $pp \to j + \gamma$ has a $m_{\gamma\gamma}$ shape which does not substantially modify the non resonant contribution, which in the following will be modelled by the sole $pp \to \gamma\gamma$ process. This assumption seems quite reasonable by looking at Figure 4.2, as the $pp \to \gamma + j$ does not exhibit any peak that would deviate from the typical decreasing behaviour observed for $pp \to \gamma\gamma$. A normalization comparison between the two processes had not been possible since the since the $pp \to \gamma + j$ sample corresponds to an equivalent luminosity of few fb^{-1}, thus suffering from sizeable statistical fluctuations . Once the aforementioned hypotheses are assumed, neglecting $pp \to \gamma + j$ and modelling non resonant background only with $pp \to \gamma\gamma$, will not influence the analysis, as the only possibly appreciable effect might be a slight underestimation of the non resonant background. This will not affect the signal extraction in the fit procedure since the background shape will be entirely extracted by data.

4.2 Photon selection

Since the final fit will be performed on the diphoton invariant mass, correctly identifying and reconstructing the prompt photons is crucial for the analysis.
In CMS, photons are reconstructed from crystal clusters in the ECAL [13]. Some photons however, may trigger a shower before entering the ECAL (*converted* photons), as they may produce electrons by pair production in the tracker that could trigger an EM shower before the ECAL, leading to energy dispersion. On the other hand, the *unconverted* photons yield an EM shower in the ECAL, consisting in a plethora of photons/electrons depositing their energy inside the scintillating crystals of ECAL, until the shower is extincted. A cluster is identified in the ECAL starting from a central *seed* - corresponding to the crystal with the highest energy deposit above a given threshold - which is the first element of the cluster. Around the seed the cluster is grown by topological clustering, i.e. by taking into consideration all those crystals sharing an edge with the cluster itself, and experiencing an energy deposit above another given, lower threshold [1].

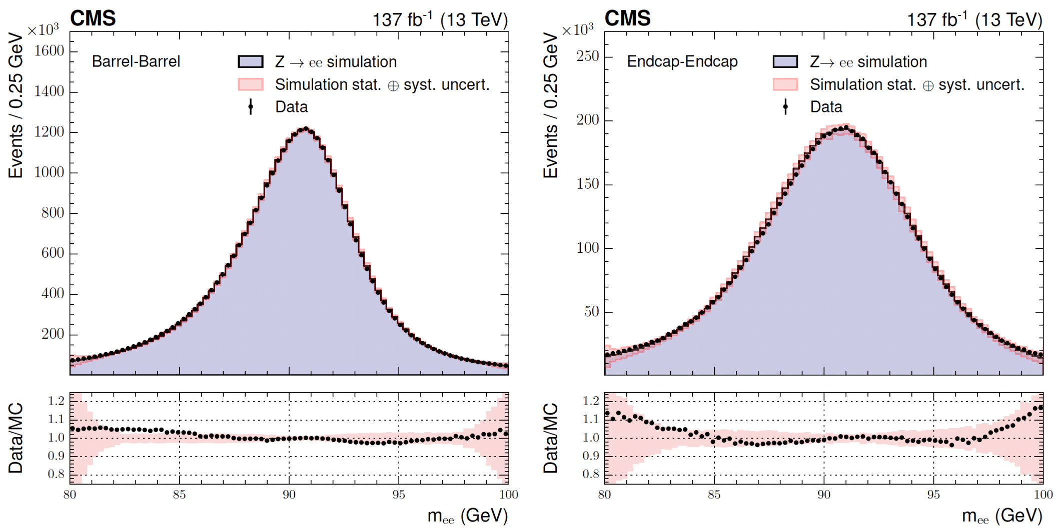


Figure 4.3. Data-MC comparison for $Z \rightarrow ee$ from which correction factors are extracted via a multivariate regression [13]

The clusters detected during an event are then merged into *superclusters*, to ensure good containment of the shower, optimising robustness of the energy resolution against pileup and accounting for geometrical variation, particularly for converted photons, whose electrons are bent by the magnetic field producing a broader spread along ϕ. The photon energy is then obtained by summing the energy deposits registered into the superclusters, and in the preshower detectors covering the $1.65 <$

[1]Typically 80 MeV in the barrel and 80-300 MeV in the encapds depending on η

$|\eta| < 2.6$. This energy estimation is then corrected both in scale and resolution [39]. The first correction stage aims at correcting dispersive effects arsing from imperfect shower containment and converted photons, through a multivariate regression trained on Monte-Carlo samples. After applying these corrections however, some discrepancies between data and simulation remain, and this is why a second, data-driven, step of corrections is performed. By looking at a high statistics process with clear signature, namely $Z \to ee$ with electrons reconstructed as photons (i.e. neglecting the information in the tracker), the correction is performed by matching simulation to the data.

In order to perform a photon efficiency able to identify prompt photons with good efficiency some selection is perfored on the most significant cluster shape variables :

- R_9: defined as the ratio of the energy deposited in a $3\times$ 3 crystal matrix built around the seed over the total photon energy, it is particularly useful for discriminating converted from unconverted photons. Whereas in the first case the shower is initiated in the tracker and has a broader lateral diffusion, thus a smaller R_9 value, for unconverted photons, as stolzite crystals are $\sim 1R_M$ broad, the shower should be mostly contained in the 3×3 matrix, giving a R_9 value close to 1;

- $\sigma_{\eta\eta}$: measures the lateral extension of the shower in terms of the energy-weighted standard deviation of single crystal within a 5×5 array of crystals centered on the seed. This variable is not only useful for discriminating converted from unconverted photons (the latter produce shower with smaller lateral extension, thus $\sigma_{\eta\eta}$) but also for discriminating against jets. A jet indeed can mimick the signal of a photon by producing a shower in ECAL, mainly due to $\pi^0 \to \gamma\gamma$. The jet shower, as for converted photons, has a higher lateral development, and a greater $\sigma_{\eta\eta}$;

- H/E: the hadronic over electromagnetic energy ratio is defined as the ratio between the energy deposited in the HCAL tower behind the supercluster's seed (i.e. in a cone of radius $R = 0.15$ around the supercluster direction) and the energy of the photon or electron candidate.

- I_{ph}: photon isolation, the p_T sum of particles identified as photons inside a cone of size R=0.3 around the photon candidate direction;

- I_{tr}: track isolation, the p_T sum of of all tracks in a cone of size $R = 0.3$ around the photon candidate direction, excluding tracks in an inner cone of size $R = 0.04$ to avoid counting tracks arising from photon conversion into electron-positron pairs;

The main preselection photon criteria on these variables, which are already embedded in the Monte Carlo samples that will be employed, are reported in Table 3.1.

Further constraints are imposed on the kinematic variables of the reconstructed diphoton, in order to optimize the selection of photons coming from the $H \to \gamma\gamma$ decay :

	R_9	H/E	$\sigma_{\eta\eta}$	$\mathcal{I}_{\text{ph}}$ (GeV)	$\mathcal{I}_{\text{tk}}$ (GeV)
Barrel	[0.50, 0.85]	<0.08	<0.015	<4.0	<6.0
	>0.85	<0.08	—	—	—
Endcaps	[0.80, 0.90]	<0.08	<0.035	<4.0	<6.0
	>0.90	<0.08	—	—	—

Table 3.1 Preselection thresholds applied on photon candidate for the $H \to \gamma\gamma$ decay channel. Selections on R_9 and $\sigma_{\eta\eta}$ are imposed to reject ECAL energy deposit incompatible with a single, isolated EM shower, such as those from neutral mesons. Preselection on H/E is performed in order to reject hadronic showers. Further conditions are imposed on photon and track isolation.

- The minimum transverse momentum p_T of leading and subleading photons must be greater than 35 GeV and 25 GeV, respectively;
- Pseudorapidity of photons must be $|\eta| < 2.5$ and not in the barrel-endcap transition of $1.44 < |\eta| < 1.57$;
- electron veto: candidate photon rejection if its supercluster in the ECAL is near to the extrapolated path of a track compatible with an electron;
- loose selection of charged hadron isolation, i.e. the p_T sum of charged hadrons inside a cone of size R=0.3;
- The candidate photon must satisfy at least one of the following conditions: $R_9 > 0.8$, $I_{ch}/p_T^{\gamma} < 0.3$, and $I_{ch} < 20$ GeV.

Photons overcoming the preselection constraints are subject to a multivariate analysis via a Boosted Decision Tree (BDT) to distinguish prompt photons from jets mimicking a photon signals. An ID BDT is trained on simulated samples of $\gamma + j$ where the photon is taken as signal and the jet as background, receiving cluster shape variabes as input. As a result a BDT score (IDMVA score), comprised in [-1,1], is produced. An additional selection constraint is for both of the photons in the final state to have an IDMVA score of at least -0.9, which occurs in 99 % of cases [13].

Another crucial aspect for the sensibility of the analysis is correctly identifying the diphoton production vertex, as it has a direct impact on $m_{\gamma\gamma}$ resolution. If the vertex position on the z-axis is known to better than 1cm the invariant mass uncertainty is dominated by the energy reconstruction. If not, the vertex identification dominates the uncertainty on $m_{\gamma\gamma}$. The RMS of the distribution of the reconstructed vertices z coordinate in data in 2016–2018 varies in the range 3.4–3.6 cm [13]. A BDT (*vertex identification* BDT) is dedicated to reconstruct the vertex position and, as photons do not leave sign in the tracker, the input variables are related to the charge particles produced by particles such as gluons (or the vector boson in this case), recoiling against the diphton system. A second BDT (*vertex probability* BDT) estimates the probability that the vertex, chosen by the vertex identification BDT, is within 1 cm of the vertex from which the diphoton originated.

A final selection on the two leading photons, γ_1 and γ_2, requires them to satisfy the condition: $p_T^{\gamma_1} > m_{\gamma\gamma}/3$, $p_T^{\gamma_2} > m_{\gamma\gamma}/4$ and $100 < m_{\gamma\gamma} < 180$ GeV.

4.3 Jet selection

Hadrons are identified by matching the tracker information (in the case of charged hadrons) with ECAL and HCAL energy deposits. A hadron, both charged or neutral, triggers a shower in the HCAL. If the hadron is charged, it leaves a measurable track in the tracker and possibly small energy deposits in ECAL due to the ionization of the material. The energy of a hadron is thus reconstructed by summing the energy deposits on the HCAL tower built upon the correspondent tracker/ECAL section. Jets are collimated sprays of hadrons that arise from the fragmentation, and subsequent hadronization, of primary partons, which can not be directly observed as they are not color singlet. The jets are reconstructed by employing the infrared and collinear safe *anti-k_T* algorithm, with a size parameter R=0.4 [40]. This is a sequential clustering algorithm that iteratively groups pairs of the closest particles in the transverse plane, with the distance between particles defined as:

$$d_{ij} = min({k_T}_i^{2p}, {k_T}_j^{2p})\frac{\Delta_{ij}^2}{R^2} \tag{4.3}$$

where k_T are the transverse momenta, and Δ the the distance in the $\eta - \phi$ space. The parameter R sets radius threshold, while the parameter p defines a whole class of algorithms. For the anti-k$_T$ algorithm, $p = -1$. The anti-k$_T$ algorithm favors the clustering of particles around the particle with the highest momentum, leading to the reconstruction of well-behaved and stable jets. Indeed, this makes the algorithm collinear safe, i.e. not prone to collinear emission of particles in a narrow cone around the jet axis. The infrared safeness arises from the distance expression (4.3) that is employed to cluster particles, as is proportional to the inverse of transverse momenta, thus making infrared, soft emissions irrelevant. Jet energy/momentum is estimated as the sum of the energy/momentum of the particles composing it. In the central region of the detector, the resolution on the reconstructed jet energy is 15–20% at 30 GeV, about 10% at 100 GeV, and 5% at 1 TeV [41]. In the following analysis, jets are implicitly required to have transverse momentum $p_T > 25$ GeV, pseudorapidity $|\eta| < 4.7$ and an angular separation from photons $\Delta R(j, \gamma) > 0.4$.

On top of this baseline photon/jet selection, common to each $H \to \gamma\gamma$ analysis in CMS, some more specific, production mode-related, cuts are imposed for selecting VH hadronic events. The events are required to have at least $n_j = 2$ reconstructed leading jets in the event, whose invariant mass is consistent with the decay of a weak vector boson: $60 < m_{jj} < 120$ GeV. Moreover, as the hadronic decay products are expected to be two close central jets, the constraint on pseudorapidity is tightened to $|\eta| < 2.4$, and the transverse momentum threshold is raised to $p_T > 30$ GeV. A stricter bound is also applied to the photon ID BDT output, which is required to be greater than -0.2 for both of the final state photons.

4.4 Kinematic distributions

In this section, the study of the main kinematic distribution is described. This will be achieved by making use of MC samples that account for a detailed simulation of both the kinematics of the particles originated from the processes of interest and their full interaction in the CMS detector. This study will be carried out by comparing the kinematic variables of the VH signal with other production methods and non resonant background, as well as with beyond standard model simulated samples.

The *JHUGen* generator [29] , allows to simulate in detail the kinematics of bSM Higgs scenarios, evaluating the matrix element at LO: pseudoscalar (0M), scalar with higher order corrections (0PH), Λ_1 term (L1), Λ_1 term for $Z\gamma$ vertex (L1Zg) in addition to the SM Higgs (0PM), for each production mode. These possible scenarios directly correspond to the various Wilson effective coefficients that parametrize possible bSM corrections to the HVV vertex, as for Eq. 2.43. The kinematics of SM Higgs production modes however, is simulated through the *powheg* generator, which allows computation at NLO. The PYTHIA 8 generator is employed for simulating the parton shower. In order to better match data accounting for year-dependant variations in the detector condition, MC samples are divided by year of Run-2 operation (2016,2017, 2018), with the year 2016 further divided into two subsamples to account for sizeably different conditions in the tracker. The plots reported in this section are generated by employing one of these 2016 subsamples (denoted as 2016preVFP), the only one available for the present study, corresponding to an integrated luminosity of $19.5fb^{-1}$.
However, the JHUgen-provided cross section for the bSM samples is not considered reliable for counting bSM events from MC samples. Furthermore, as the primary aim of this section is to compare the shapes of kinematic distributions, all the distributions presented are then normalized to unity by default. The samples are endowed with the general $H \to \gamma\gamma$ preselection, as described in the two previous sections, with no VH hadronic selection imposed on top of that, if not for the necessary constraint of requiring at least two jets in the event.

A comparison of the transverse momentum probability density functions (pdf$_s$) for different production mechanisms is presented in Figure 4.4. It can be observed that none of the mechanisms yields a significantly more energetic Higgs boson, with only the gluon fusion exhibiting a *pdf* shifted towards lower P_T values.
Instead, based on the simulated bSM Higgs boson samples, it is possible to identify peculiar features , in terms of modifications to the shape of the kinematics pdf$_s$. Figure 4.5 shows that all the considered fully bSM models result in Higgs bosons with greater transverse momentum, where this (in magnitude) corresponds to the magnitude of the sum of the transverse momenta of the diphoton system. It should be clarified that the visible peaks in Figure 4.5, arise from random fluctuations

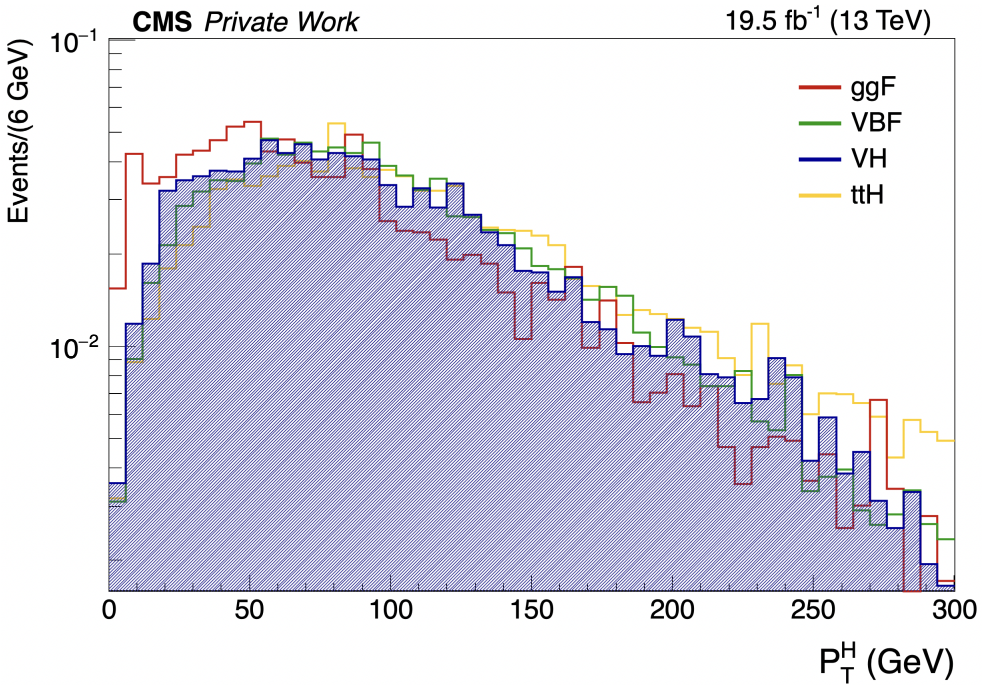


Figure 4.4. Higgs transverse momentum distribution, normalized to unity, for different SM production mechanisms

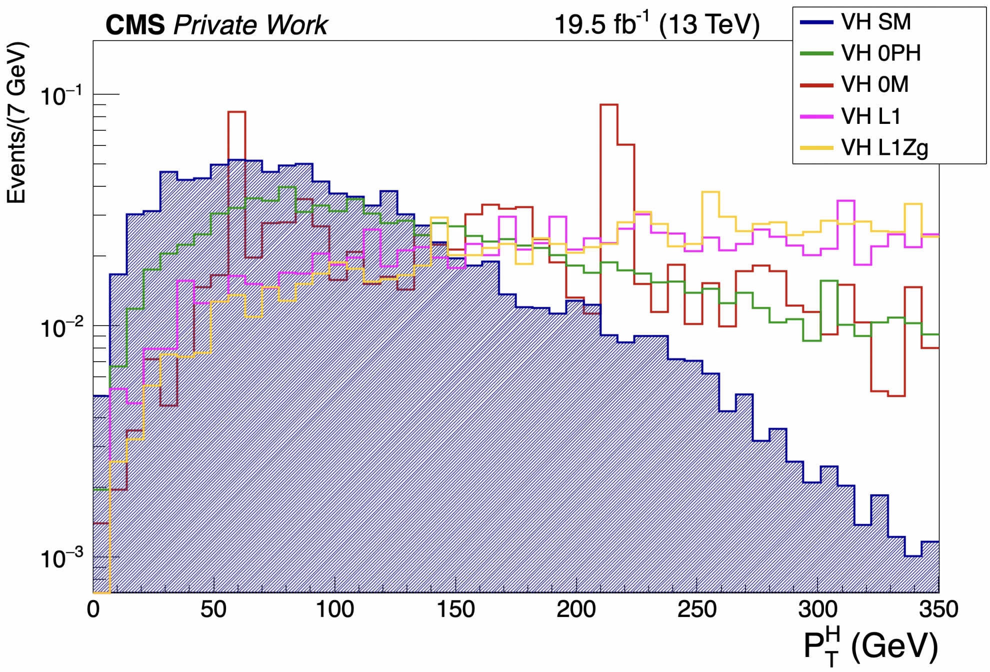


Figure 4.5. Comparison on the VH-produced Higgs boson's transverse momentum distribution, normalized to unity, between SM case and several bSM scenarios

affecting the pseudoscalar (0M) sample due low statistics. In fact the statistics of the bSM samples is about 2% of the SM samples.

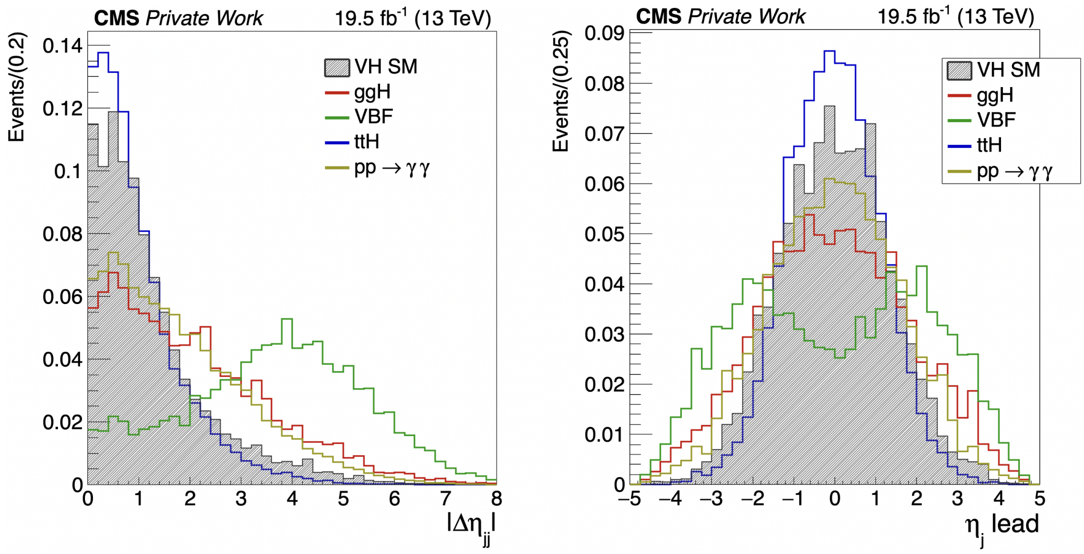


Figure 4.6. Characterization of the dijet system in terms of pseudorapidity distributions, normalized to unity, for the SM production mechanisms and non resonant background. On the left, the module of the η separation of the two leading jets. On the right, the pseudorapidity of the leading jet.

Figure 4.6 corroborates what has been previously discussed: the jets in the hadronic channel of the VH associated production are central, η-closed jets, contrary to VBF production that yields two forward, almost back-to-back jets.
Figure 4.7 shows the effect of modifying the VH coupling on the dijet distributions. It is important to note that, when dealing with the dijet kinematic distributions, all the anomalous Higgs scenarios are merged into a unique bSM histogram. This is intended to provide an average overview of the anomalous couplings (AC$_s$) effects on pdf$_s$, which is also the strategy that will be adopted when performing the multivariate analysis.
In the case of VH production, the dijet invariant mass displays the characteristic peak between the masses of the W and Z bosons, which can not be distinguished through the preselection as the resolution on the jet energy does not allow to. The dijet invariant mass peak is shown to be shifted upwards in energy by the presence of bSM contributions. This observation suggests that the AC$_s$ taken into account enhance the Higgs coupling to the Z boson more than they do with the W bosons, however lacking a clear explanation for this. The dijet leading transverse momentum for bSM samples is shifted up in energy, implying that the jets produced in this case are in average more energetic, thus enhancing the resolution on the dijet invariant mass M_{jj} as in a calorimeter the higher the energy, the better the resolution. An analogous effect that is also observable in Figure 4.5 , with the AC$_s$ producing a more energetic Higgs boson with respect to the SM case. The distribution of the opening in the azimuthal direction, $\Delta\Phi_{jj}$ between the two leading jets is mostly uniform for SM production mechanisms, but are peaked on lower values for bSM samples, since the

anomalous contributions to HVV vertex might alter the polarization of the diphoton and dijet system and thus the angular correlation among the final states particles. The discussed low η dijet separation characterizing the VH associated production (as well as the ttH) appears to be further enhanced by the bSM contributions, as the jets produced are on average more energetic.

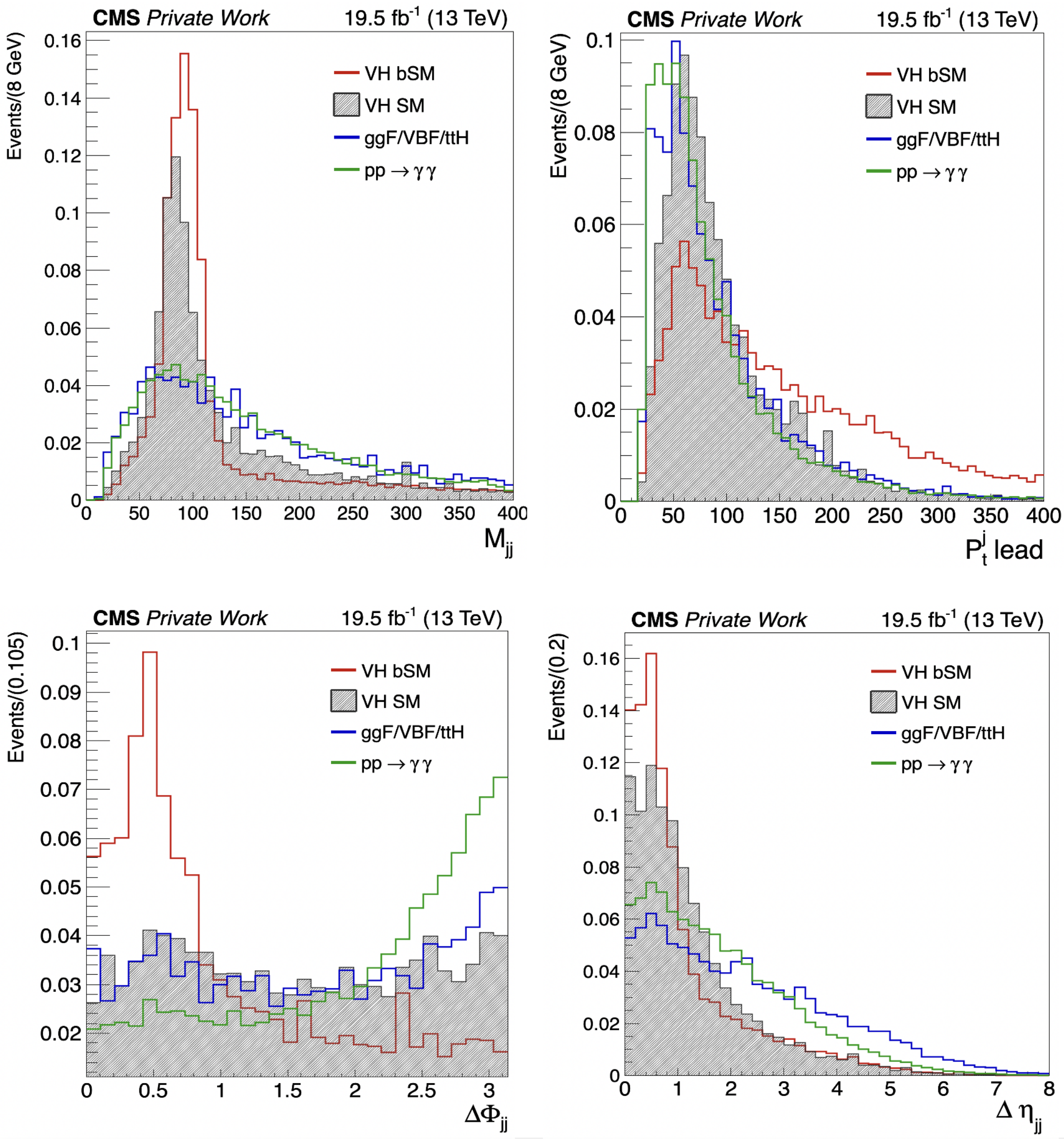


Figure 4.7. Some of the discriminating leading dijet kinematic pdf_s, normalized to unity: invariant mass, transverse momentum of the leading jet, difference in both of the azimuth angle and pseudorapidity between the two jets. The contribution of SM VH production vs bSM VH, resonant and non-resonant background is highlighted

Chapter 5

Multivariate analysis

As already mentioned in the introduction in Section 1, the present analysis implements a machine learning algorithm to optimize a multivariate analysis on the Monte-Carlo samples. This targets possible bSM effects on top of the observation of a SM-like Higgs boson from the other decay and production channels. The first step in the process is hence to distinguish bSM/SM events based on the possible difference in the shape of kinematic variables, and this is the purpose of the algorithm described in this chapter. In this context, the study on kinematic pdf$_s$ performed in the previous chapter results crucial for finding the most discriminating ones to be used in this context. In Figure 4.7 the most useful variables for such separation, not only between VH SM/bSM events but also with respect to the other sources of background (both resonant and non resonant), are reported. For this purpose the kinematic distributions taken into account are mostly those of the dijet, as the diphoton variables provide information about the Higgs decay, which could be sufficient to distinguish SM/bSM events, but not so much for the production mechanisms (an example is provided in Figure 4.4).
The variables that are going to be considered are the number of jets in the event, the dijet invariant mass, the transverse momentum and the pseudorapidity of the leading and the subleading jets, plus the magnitude in difference for pseudorapidity and the difference in the azimuth angle. Some more dependent variables are provided, as the rapidity of leading/subleading jets, the $\Delta\Phi$ truncated variable and the *Zeppenfeld variable*.
Higgs kinematics is considered only through its transverse momentum, additionally to the total transverse momentum of the diophoton+dijet state.

As the choice of the algorithm is somehow arbitrary, some criteria must guide the selection. Eventually, it boils down to two popular and general-purpose classes of algorithms, which well fit this kind of multi-label classification task: a Deep Neural Network (DNN) [42] or a Boosted Decision Tree (BDT) [43]. The working dataset is made of 1.7 million entries from the MC samples for Run-2 2016preVFP era, the only available samples with the proper preselection containing all the kinematic variables that were required for the training. It should be noticed that, as the number of entries is not uniform in each MC sample, the proportion of the several types of labelled events in the dataset is imbalanced:

	Process	Fraction
Class 1	VH bSM	8.3 %
Class 2	VH SM	3.1 %
Class 3	ggH/VBF/ttH	15.8 %
	$pp \to \gamma\gamma$	72.7 %

Table 5.1. Fraction of events divided by class in the total dataset of MC events, used for training the algorithms

This leads to train the algorithms in quite a highly imbalanced classification task, which will require the adoption of some measures to compensate the skewness in the dataset.

5.1 DNN vs BDT

As a reference, a feed-forward DNN with 4 hidden dense layers is implemented, whose architecture is shown in Table 5.2. The present discussion on the DNN is carried out by implicitly assuming the optimal classification criteria, consisting in splitting the events into 3 classes (VH bSM/VH SM/Background), choice which will be discussed in detail only in Section 5.2.2.
The dataset is split into three subsets: a training set for the learning phase of the algorithm in which the trainable parameters are tuned through the back-propagation process until reaching a minimum of the loss function, employed to quantify the algorithm performance in term of output distance to the true labels of the events. A validation set is used to test the performance on a previously unseen dataset, with the purpose of evaluating the over/under-fitting of the algorithm and tune its hyperparameters. Finally, a test set is defined for testing the performance and generalization power of the network on a totally independent subset. The proportion adopted for the train/validation/test split is 60/20/20 % of the original dataset.
The 16-components input feature vector, made of the aforementioned kinematic variables, is passed to a series of 4 dense hidden layers with a number of neurons equal to (64, 128, 128, 64). The final layer is constituted by 3 neurons, as the classification task will be performed on 3 classes. The activation function applied on each layer is a rectified linear unit (ReLU), with a batch normalization layer added to properly renormalize the output at each stage into a gaussian of null mean and unit standard deviation, to avoid the phenomenon of vanishing gradient which typically afflicts deep networks. The selected loss function is a categorical cross-entropy, a typical choice for a multi-classification task, employing a stochastic gradient descent (SGD) technique to perform the minimization procedure. The learning rate is defined through a learning rate scheduler function: starting from a constant and large learning rate (10^{-2}) to globally probe the hyperfsurface of the loss function, the value is kept constant for 10 epochs, after which it is decreased by an order of magnitude. This strategy is repeated until the 50th epoch (corresponding to a value

Layer (type)	Output Shape	Param #
Input (InputLayer)	[(None, 16)]	0
Dense_1 (Dense)	(None, 64)	1088
Dropout_1 (Dropout)	(None, 64)	0
Batch_norm_1 (BatchNormalization)	(None, 64)	256
Dense_2 (Dense)	(None, 128)	8320
Dropout_2 (Dropout)	(None, 128)	0
Batch_norm_2 (BatchNormalization)	(None, 128)	512
Dense_3 (Dense)	(None, 128)	16512
Dropout_3 (Dropout)	(None, 128)	0
Batch_norm_3 (BatchNormalization)	(None, 128)	512
Dense_4 (Dense)	(None, 64)	8256
Dropout_4 (Dropout)	(None, 64)	0
Batch_norm_4 (BatchNormalization)	(None, 64)	256
dense_1 (Dense)	(None, 3)	195

Total params: 35,907
Trainable params: 35,139
Non-trainable params: 768

Table 5.2 Architecture of the DNN employed for the classification task

of 10^{-5}), when the learning rate starts to be decreased exponentially, (reaching a value of $6 \cdot 10^{-7}$ at 80th epoch) in order to perform a more and more fine grained investigation of the loss function nearby the minimum, as the epochs increase. With a total number of 35 thousands trainable parameters, corresponding to the weights parametrizing the connections among neurons in the layers, such a DNN is typically prone to over-fitting. This is why in each hidden stage a dropout layer is added, randomly turning off a certain percentage of the neurons in the layer (in this case the value is set to 30 %), in order to prevent the network to systematically learn noise in the data or ad hoc pattern correlations in the mapping among neutrons. To further improve the generalization power of the algorithm, a L2 regularization in each layer is inserted, which introduces a penalty on the loss function given by the square of the weights of the model (with a multiplicative penalty factor which is set to 0.02), thus inducing the network not to set the weights on a large range and preventing it from learning peculiar fluctuations of the training set. Still for the purpose of regularization, an early stopping procedure is utilized, stopping the training if no improvement is observed in a set number of epochs (here 20 epochs are chosen). The hyperparameters of the network, such as the number of layers, the number of neurons per layer, the dropout percentage etc. have been fine tuned to regularize the output of the network history, both with regard to finding the right working point of the DNN in terms of bias/variance trade-off and to optimize the absolute performance. As previously mentioned, this is an imbalanced classification task, so if the network was trained on the plain dataset, it would exhibit a bias towards the

most represented class, which is, as for Table 5.1, the resonant background class. The *sklearn* package in Python [44] provides a useful tool to make up for the problem, the *sample weights*, defined for each class as a parameter inversely proportional to the representation fraction of that class in the dataset. In this way, to each event in the sample is assigned a weight leveling the imbalance in the class representation: the events of most represented class, $pp \to \gamma\gamma$, are associated to the smallest weigth, while the events of the less represented class, namely VH SM, are assigned to the largest weight. This leads to a loss penalty term for the most represented events, which will be less weighted by the algorithm, thus compensating for the skewness of the dataset.
The performance of the algorithms is tested with regard to receiving operating characteristic (ROC) curves, and subsequent area under the curve (AUC) score. A ROC curve is a profiling of the discriminator performance in terms of true positive rate (tpr) vs false positive rate (fpr) relation. The further the curve is from the bisecting line, which represents a random classifier, and the more it tends to the upper-left corner of the plane, which guarantees the ideal working point with high signal recognition and low background mis-identification, the better the discriminator is.

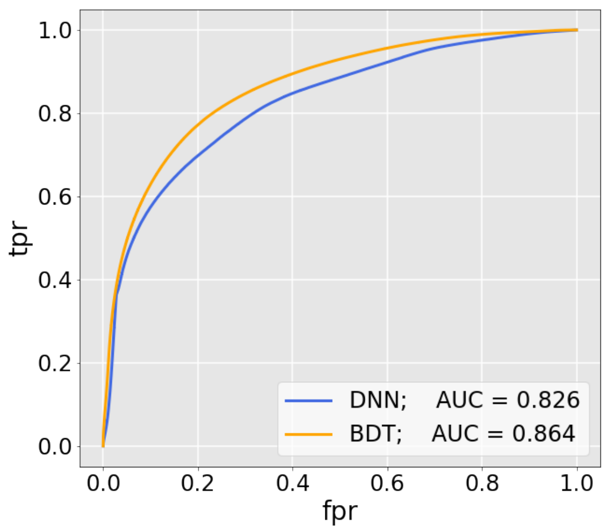


Figure 5.1. 3 classes DNN-BDT ROC curve comparison

Figure 5.1 shows the comparison in terms of ROC curve performance between the just discussed DNN and the optimized BDT, whose implementation will be described in detail in Section 5.2.1. Being a multiclass problem, the ROC curve is obtained through a micro average over the several classes. Differently to a macro average, which simply averages the performance over the 3 classes, a micro average computes the fpr/tpr by ravelling the true one-hot encoded label vector and the probability vector predicted by the algorithm, allowing in this way to take into account the relative representation ratios of the classes in the dataset. This fits the imbalanced simulated sample classification problem under investigation and thus results in a weighted average over the non-uniformly represented classes. Although it could be questioned whether the algorithms are fully optimized and are really operating at their ideal working point, they show a similar performance on the given dataset, with the BDT apparently working slightly better. Therefore the choice will fall on the BDT, which, at similar performance, appears to be less of a black-box with respect to a DNN, allowing to better predict the behaviour of the algorithm and facilitating its optimization. The next section will therefore be focused on the criteria that led to the optimization of the algorithm.

5.2 BDT

A Boosted Decision Tree is a powerful non-parametric algorithm for supervised learning, consisting in an ensemble of weaker learners (decision trees). It features several desirable characteristics:

- *high performance* in both classification and regression tasks, as it is able to capture complex relations in data;
- *robustness* with respect to noisy samples, as it is constituted by an ensemble of decision trees, resulting less prone to over-fitting with respect to a DNN;
- *interpretabilty* as it exhibits a clear graph-based decision structure;
- *scalability* as it is fast, can handle huge datasets and is easily parallelizable.

The ensemble is implemented by the means of a boosting technique, which concatenates the outputs of several decision trees in order to adjust the training weights to prioritize the mis-classified samples in each iteration, thus allowing subsequent trees to focus on correcting the errors of the previous ones.

A binary decision tree is an acyclic graph starting from an initial root node, which gets split into two branches, each of whose nodes are further split into two branches and so on, until all of the leaves of the tree reach some stopping condition (Figure 5.2). A global purity score is defined for the tree (Gini index, cross-entropy, ...) which guides the tree growth as the branches get further and further split. The feature variables on which to impose the separation condition, and the splitting threshold itself, are indeed imposed by trying to maximize the gain of purity, i.e. the variation of purity after and before the splitting point, at each node. The growth of a branch is terminated when some stopping condition is met: it could be the minimal number of events per leaf, the maximum number of leaves, the maximum depth of the tree or the purity gain below a certain threshold. A binary decision tree can be regarded as a d-dimensional histogram - whose bins are iteratively defined by the sequence of binary splits - which is able to approximate any generic separation surface in the feature space of data, corresponding in this case to the space spanned by the aforementioned kinematic variables.

A popular process for testing and optimizing a decision tree is the so-called *k-fold cross validation*, consisting in a re-sampling technique for an iterative split of

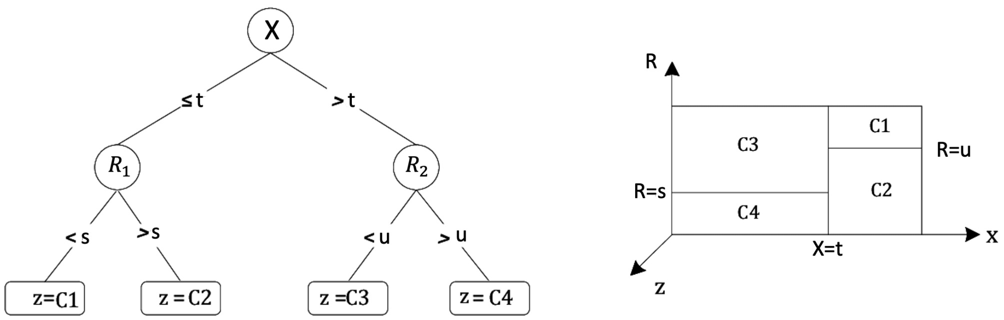


Figure 5.2. Trivial example of the graph structure of a binary decision tree, and its interpretation as a d-dimensional histogram in the space of features

the training set in cyclic (training + validation) subsamples, for the purpose of optimizing the hyperparameters and estimating the performance of the algorithm. The advantage in doing so relies in the minimization of statistical fluctuations for samples of finite statistics.

A decision tree works as a recursive partitioning of the sample in regions of gradually higher purity. The pros of such an algorithm rely in its transparency, its intuitive graph-based representation, and its reduced sensibility to weak variables. On the other hand, this type of algorithm suffers from overtraining (if not properly controlled in its growth), might be sensible to small variations in the training set and has reduced performance by its own. The first issue is addressed by implementing *pruning* techniques which remove some of the branches, according to some selection criteria, after letting them grow without constraints. The second issue is mitigated by the already cited cross validation technique. The last one is solved by resorting to ensembles of decision trees through *bagging* or *boosting* techniques.

Decision trees can therefore be utilized as a building block for a more complex algorithm such as a boosted decision tree, which combines weak models to get a more predictive one. The basic idea of the gradient boosting (GB) is adding decision trees to pre-existing ones in such a way to iteratively minimize the gap of the prediction of the ensemble with respect to the loss function value obtained by the true label of the event. In this way the learning process of the algorithm is traced back to a minimization problem and GB results as a gradient descent algorithm in a space of functions, given by the weak classifiers added at each step on top of the ensemble, in order to match the true output of the loss function.
The actual loss function for a GB algorithm is written as a sum of two terms: one term for measuring the prediction distance from the actual value, and one term, independent from data, that regularizes the trees by associating to each leaf of the estimators a weight. In such a way, the loss function depends by a set of weights - as well as by a set of hyperparameters, such as the number of leaves per tree - which must be tuned in order to minimize the loss function and converge.

The GB technique is further extended to the *extreme gradient boosting* (XGB) algorithm, which adds to the standard GB some features that improve the performance, such as a smaller learning rate to improve generalization power, stochastic gradient descent and regularizing penalty functions (L1, L2, L1+L2, ...).
The choice for implementing the multivariate analysis technique, will fall on a XGB algorithm for the several reasons exposed, receiving the kinematic variables previously listed in the introduction to this chapter as input. Next sections are thus going to focus on the XGB-based BDT that has been instructed, dealing with the optimization of its hyeperparameters and comparing two different kind of classification schemes.

5.2.1 Hyperparameters optmization

The BDT utilized for discriminating MC events on their kinematic variables has been implemented via the *xgboost* package available in python [45]. The algorithm has been written, trained and tested on the Kaggle online platform by making use of the available GPU acceleration for speeding up the training.

A BDT depends on several *hyperparameters* that constrain both the topology of each decision tree, as well as the global structure of the ensemble. Some of the most relevant parameters are here reported:

- *learning rate*: defines the step with which the SGD is performed. As previously mentioned, a learning rate $\eta \ll 1$ guarantees a higher generalization power for the BDT, making the boosting process more conservative;
- *gamma*: minimum loss reduction required to make a further partition from a leaf node;
- *max depth*: puts a cap on the growth of the branches in the trees. A high value makes the BDT more powerful, but also more prone to overfitting;
- *minimum_child_weight*: the minimum threshold value for the sums of the weights in a node to be further partitioned. The higher the value the more conservative the algorithm;
- *subsample*: fraction of the training data that is subsampled by the BDT, helping to prevent overfitting;
- *lambda*: magnitude of the L2 regularization term on the weigths;
- *alpha*: magnitude of the L1 regularization term on the weigths;

During the training phase of the algorithm a k-fold cross validation with k=5 is implemented.
The objective function considered is a *multiprobability softmax* function which is the standard loss function for multilabel classification problems in *xgboost* and a *negative log-likelihood* is taken as a metric to probe the validation set. The learning rate and the number of estimators have been tuned trying different possible values for a parameter, while constraining the other to be fixed, finding the best value

Hyperparameter	**Range** *(min,max,step)*	**Optimal value**
max depth	(4, 8, 1)	6
min child weight	(2, 5, 1)	4
subsample	(0.5, 1, 0.1)	0.7
gamma	(0, 0.5, 0.1)	0
alpha	(10^{-5}, 10^{-2}, 0.1, 1, 100)	0.1

Table 5.2. Hyperparameters grid optimization. The triples represent the minimum of the varying range, the maximum of the varying range and the step in the scan. In the case of alpha the explicit vector of possible values is reported

as the one optimizing the AUC performance score. The learning rate is set to 0.1 and the number of estimators in the ensemble is fixed to 1000. Furthermore, the value of *lambda* parametrizing the L2 regularization is kept at the default value of 1. The optimization of the other hyperparameters has been performed by exploiting the *GridSearchCV* module available in python's *SciKitLearn* package. This tool automatizes the processes of making a scan of the parameters to find the configuration maximizing the performance score, by fixing a discrete sequence of values for each parameter within a fixed range of variability. By doing so for each element in a chosen subset of the hyperparameters, it is defined a grid over which the algorithm is tested, thus evaluating the best possible configuration for the hyperparameter values. The optimization is carried out in turn as only a single, or at most a pair, of hyperparamters is chosen to be grid-optimized together, keeping the other fixed. Table 5.2 shows the hyperparameters considered for optimization, their range of variability and the optimal value maximizing the AUC score.
As in the case of the DNN, being the sample imbalanced, each event in the training sample is assigned to a sample weight compensating for the under/over-representation of its class in the dataset. The several constraints imposed in terms of a moderate learning rate and number of estimators, limited branching growth, as well as the adoption of (L1+L2) regularization technique, makes the BDT more conservative, enhancing its generalization power and reducing its proneness to overfitting and overtraining.

5.2.2 3 classes vs 4 classes model

A fundamental matter in the development of the algorithms is deciding how to group the events. Since the purpose is optimizing the VH discrimination, either SM or bSM, these two output classes are fixed. There is still freedom to chose whether to merge the remaining resonant (ggH/VBF/ttH) and non-resonant ($pp \to \gamma\gamma$) background or not. The optimized BDT has been therefore tested both on a 3-classes sample and on a 4-classes sample, as reported in Figure 5.3, assuming that the optimization of the BDT itself does not drastically depend on the arbitrary choice of the event grouping. Looking at the ROC curves, which are still computed by micro averaging, the 3-classes (VH bSM/VH SM/bgd) BDT exhibit a better performance, and it will

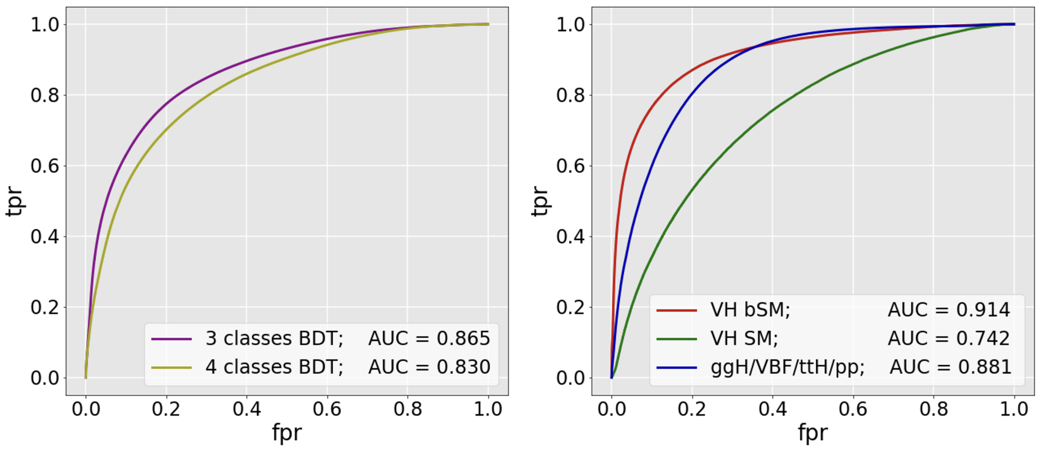


Figure 5.3. On the left, the ROC curve comparison between a 3-classes and a 4-classes model. On the right the ROC curve performance differentiated by class, in a one-vs-rest scenario.

be taken as the default configuration choice in the following analysis.
As this is a multi-classification task, the BDT sensibility to different classes can be tested. The ROC curves, which are intrinsically a dichotomous signal/background comparison, must be defined according to a convention which bring back this case to a binary scenario. Two are the options: the *one-vs-one* hypothesis separately tests the class taken as signal with respect to the other two classes and then makes an average. The *one-vs-rest* scenario, which is the one that will be adopted, tests the signal detection/background mis-identification rate for one of the three classes by considering the other two altogether as background. From the right plot in

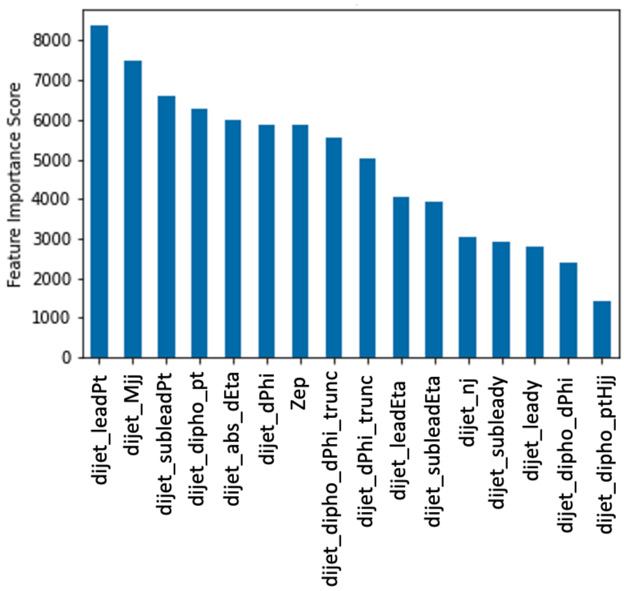


Figure 5.4. Feature importance ranking made by the BDT

Figure 5.3, which shows the BDT discrimination power on the 3 different classes, it clearly emerges a poorer performance on the VH SM class. Most of the VH events are quite easy to identify, as they show the peculiar peak of the dijet mass due to the W/Z bosons hadronic decay. However, both SM and bSM events exhibit the characterizing peak on the dijet invariant mass, making them less easily separable. Due to the already discussed skewness of the sample, the BDT could be likely biased towards predicting the the bSM class, as the VH SM is less known to the algorithm being the less represented class in the sample. This partly biased behaviour of the BDT appears to be an ineliminable factor, despite the regularization techniques adopted and the introduction of the sample weights for compensating the skewness of the sample. Nonetheless, it should be noticed that the BDT manages to reach an overall good performance, guaranteeing a good event discrimination.

Figure 5.4 reports the input features ranking, according to the variables the the BDT considers to be the most significant for discrimination. The dijet transverse momentum, its invariant mass, as well as the angular separation between the two jets, result to be the most discriminating kinematic observables for labelling the events in the classes defined.

5.2.3 Probability output

Consequently to its optimization, training and validation, the BDT is finally employed for the purpose it has been implemented for. Before proceeding to predict the probabilities of each event belonging to one of the three defined classes, the legitimacy of testing the BDT on the events belonging to the training set should be questioned.

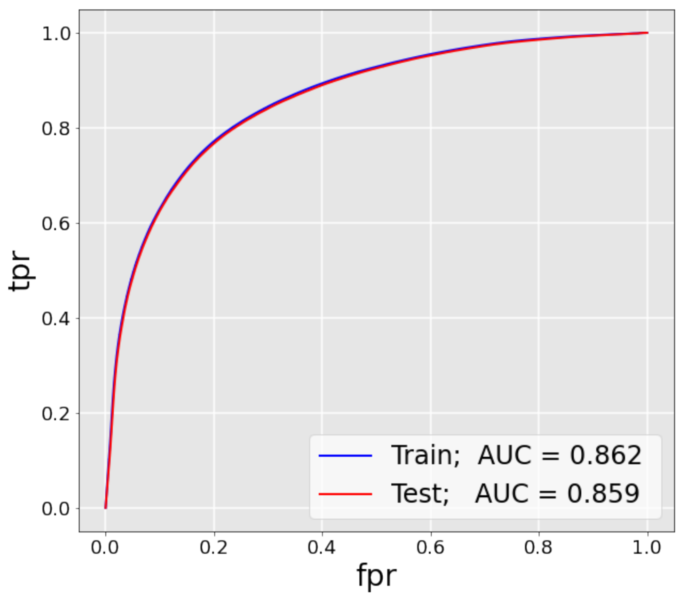


Figure 5.5. BDT performance on training and test set

The training set events indeed, are part of the sample over which the algorithm is trained and validated to find the best weight values for the purpose of optimizing some objective function. Since they do not constitute a statistically independent sample, the algorithm might be strongly biased toward those events. The several regularization techniques employed however, as the k-fold CV training, the (L1+L2) regularization term applied, as well as constraints on hyperparameters such as *max_depth* and a moderate number of estimators in the ensemble, prevent the BDT from learning the noise of the training sample and make it less prone to overfitting. Figure 5.5 reports the BDT performance on both the training set, which constitutes 80 % of the overall dataset, and on the statistically independent test set, constituting the remaining 20 %. The datasets are randomly shuffled each time before being split or being tested over, to avoid any bias. The train/test ROC curves are almost perfectly coinciding, implying that there is a negligible overtraining, thus guaranteeing an algorithm with good generalization properties and low bias towards the data it has been trained over. This legitimates the use training data for testing the BDT probability output.

In Figure 5.6 the distribution of the probability outputs for the 3 classes is reported, as the softmax objective function allows for a probabilistic classification of the events. Out of the 3 probability components in the output, only 2 are independent. This forces to choose the 2 most discriminating ones for optimization. As already noted in the previous section, the BDT appears to be weak in predicting the label of VH SM events, mainly due to the low statistics in the sample. This translates into a $\mathscr{P}$(SM) probability output that does not exhibit VH SM event distribution peaking towards one - but rather mostly covering some intermediate probability values - leading this component to poorly separate the events of different classes. On the other hand, $\mathscr{P}$(BSM)/$\mathscr{P}$(BGD) appear to well discriminate VH bSM/background events, so they are going to be considered for optimization.

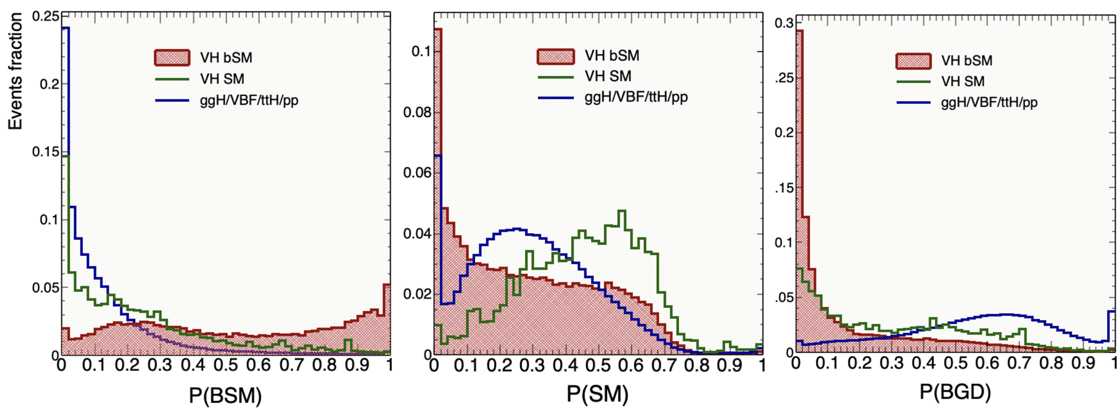


Figure 5.6. Probability output from the 3-classes BDT model

5.3 Categories optimization

A categorization on the events is performed, through a proper partition of the optimal probability outputs identified. The purpose of this is to isolate events in regions with high VH signal purity from others largely background-dominated. Apart from a separation by production mechanism, the categorization also splits with respect to the SM/bSM hypothesis, thus allowing to define categories more sensible to bSM couplings.

Multiple categorization schemes are tested:

a) Single 1D cut on the $\mathscr{P}$(BSM) output;

b) Single 1D cut on the $\mathscr{P}$(BGD) output;

c) Uniform 1D multi-binning on the $\mathscr{P}$(BSM) output;

d) Non uniform 1D multi-binning on the $\mathscr{P}$(BSM) output;

e) 2D partition with a single cut per axis on the joint $\mathscr{P}$(BSM)-$\mathscr{P}$(BSM) output;

The first two scenarios consist in identifying a single threshold value on the variable - thus defining two sole bins - then varying the value in order to find the optimal value. Scenarios (c) and (d) and are only performed on $\mathscr{P}$(BSM) since, after observing the results of the 1D cut on $\mathscr{P}$(BGD), the option of solely partitioning on this output is discarded, as it leads to significantly worse results, as for Table 5.3. $\mathscr{P}$(BSM) clearly results as the most discriminating variable. The (c) scenario is a uniform binning on the domain [0,1] in steps of 0.2, for (d) the binning is non uniform with the purpose of making broader bins where less events are expected, i.e. in the central region. The defined domain partition is therefore (0, 0.1, 0.3, 0.7, 0.9, 1). The last scheme considered is the 2D binning with a single cut both on $\mathscr{P}$(BSM) and $\mathscr{P}$(BGD).

The figure of merit guiding the choice of the best categorization scheme is the maximization of the statistical significance, defined for each i category of the scheme as:

$$\sigma_i = \frac{s_i}{\sqrt{s_i + b_i}} \quad \longrightarrow \quad \sigma_{tot} = \oplus_i \, \sigma_i \tag{5.1}$$

It should be noted that this is just an approximated formula for Poissonian processes, holding in the limit $s \ll b$, which however is reasonably assumed throughout the whole analysis. The statistical significance is evaluated for each subregion of the partitioned space defined by the specific scenario, and then summed in quadrature to get the total statistical significance.
In Eq. 5.1 the signal s_i is the number of bSM events falling in the i-th category of the specific scheme. The number of bSM events in each category are evaluated by taking into account the upper limit of the coefficient f_{a_3}, parametrizing the fractional cross section of a pseudoscalar Higgs, from [21]:
The number of signal events is thus evaluated as:

$$s_i = UL(f_{a_3}) \cdot N^{tot}_{\text{VH SM}} \cdot \epsilon_i \tag{5.2}$$

where $\epsilon_i = N^i_{bSM}/N^{tot}_{bSM}$ is the efficiency of the i-th category, meant as the ratio of bSM MC events falling in the category over the total number of bSM events.
The background events are a sum of contributions arising both from the events in the background class - comprising ggH/VBF/ttH events as well as non-resonant $pp \to \gamma\gamma$ events - and the expected VH SM events execept for the fraction attributed to bSM processes:

$$b_i = N^i_{\text{VH SM}} \cdot (1 - UL(f_{a_3})) + N^i_{bgd} \tag{5.3}$$

where $N^i_{bgd} = N^i_{ggH} + N^i_{VBF} + N^i_{ttH} + N^i_{pp\to\gamma\gamma}$.

The event yield is thus estimated from the MC samples after a normalization to the total Run-2 integrated luminosity (137 fb^{-1}). Since the signal yield will be extracted by a maximum likelihood fit on the $m_{\gamma\gamma}$ distribution, and the signal ranges over a narrow window over the total fit interval, the background and signal yields for this optimization are calculated over a $\sim 5\sigma$ window around the Higgs mass, corresponding to [120,130] GeV, to mimick the fit behaviour.
As previously anticipated, in this context neglecting the $pp \to \gamma + j$ events might lead to a small underestimation in the number of background events. However, this will likely not influence the optimization procedure, as it should equally affect all the categories as long as the $pp \to \gamma + j$ kinematic distributions do not drastically deviate from the ones of $pp \to \gamma\gamma$ events, which has been shown in Figure 4.2 for $m_{\gamma\gamma}$.

In Table 5.3, the computed statistical significance for the different categorization scenarios is reported. While for (c) and (d) the binning is fixed a priori, for the other schemes the results reported in table come from an optimization. Scenarios (a) and (b) have been optimized by scanning the probability threshold in steps of 0.1 in the [0,1] domain of the probability variables. The best values result as 0.6 for $\mathscr{P}$(BSM) and 0.8 for $\mathscr{P}$(BGD). The 2D scheme (e) has been optimized by grid-scanning all the possible (p_{BSM}, p_{BGD}) pairs in steps of 0.1 per axis, with the optimal result found to be (0.5, 0.1). While it is very low for the 1D partition on $\mathscr{P}$(BGD), the remaining schemes reach similar values of statistical significance. In particular, (c),(d) and (e) all exhibit almost the same value. Scenario (e) however proves that, despite employing a weakly discriminating variable as $\mathscr{P}$(BGD), this kind of 2D cutting scheme enhances the discrimination power.

Cat. scheme	σ
(a)	0.155
(b)	0.054
(c)	0.178
(d)	0.176
(e)	0.178

Table 5.3. Statistical significance for the different optimized categorization schemes

For this reason, as well as for a greater simplicity in performing a finer optimization on it, this is the categorization scheme that is going to be employed in the following.

Once the 2D categorization scheme is chosen, a fine-grained optimization procedure is performed, consisting in just decreasing the scanning step from 0.1 to 0.01. In Figure 5.7 the result of the 2D grid-scan is shown, where the statistical significance as a function of the threshold cut on $\mathscr{P}$(BSM)/$\mathscr{P}$(BGD) is reported for the $\mathscr{P}$(BGD)/$\mathscr{P}$(BSM) probability fixed at its optimal value.

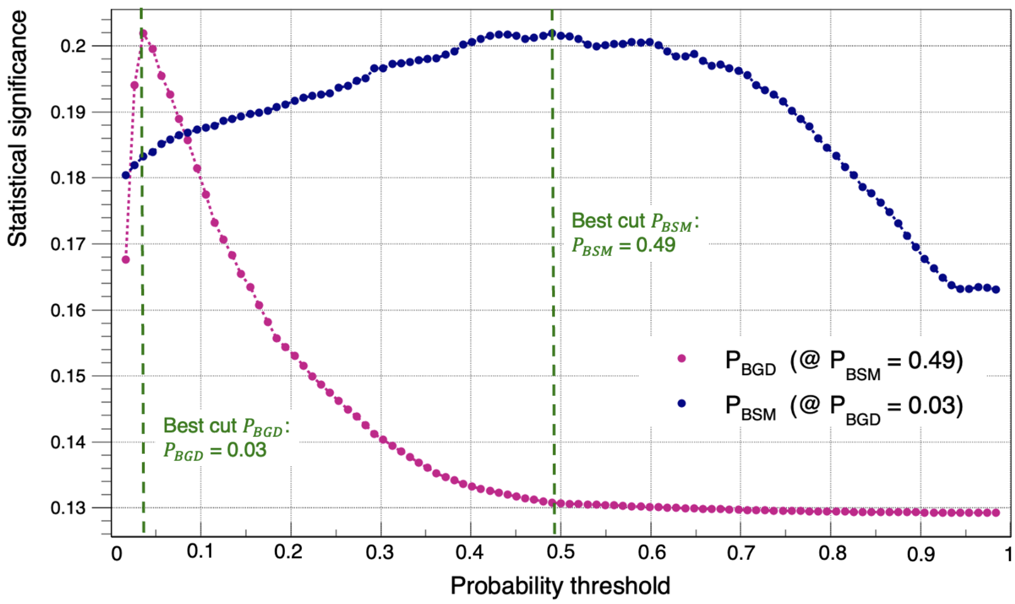


Figure 5.7. Scan of $\mathscr{P}$(BSM)/$\mathscr{P}$(BGD) outputs keeping the other fixed at its optimal value

The best threshold value are therefore found to be:

$$\mathscr{P}(BSM) = 0.49 \qquad\qquad \mathscr{P}(BGD) = 0.03 \tag{5.4}$$

The $\mathscr{P}$(BSM)-$\mathscr{P}$(BGD) space is thus split into 4 regions, corresponding to 4 categories of different signal purity.

From Table 5.4 few things can be noticed. Firstly, by estimating the number of expected VH hadronic event in LHC Run-2, i.e. by multiplying Eq. 1.2 presented in the Introduction chapter by the average $\mathscr{BR}(V \rightarrow hadr.) \simeq 68\%$ branching ratio:

$$N_{VH}^{hadr} = \sigma_{VH} \cdot \mathscr{BR}(H \rightarrow \gamma\gamma) \cdot \mathscr{BR}(V \rightarrow q\bar{q}) \cdot \mathscr{L}_{int} \cdot \epsilon_{\gamma}^{2} \simeq 115 \text{ events} \tag{5.5}$$

the expected number of events perfectly fits with the result reported in table, allowing to certifying the overall goodness of the preselection on data and validating the steps and assumptions made. The number of background events being much greater than the estimated number of signal events, strengthens the legitimacy of the approximation in Eq. 5.1. Moreover, it clearly emerges that most of the significance of the categorization scheme is attributable to only one category, namely *Cat 2*,

	Cat 1	**Cat 2**	**Cat 3**	**Cat 4**	
	$\begin{cases} P_{bsm} \geq 0.49 \\ P_{bgd} \geq 0.03 \end{cases}$	$\begin{cases} P_{bsm} \geq 0.49 \\ P_{bgd} < 0.03 \end{cases}$	$\begin{cases} P_{bsm} < 0.49 \\ P_{bgd} \geq 0.03 \end{cases}$	$\begin{cases} P_{bsm} < 0.49 \\ P_{bgd} < 0.03 \end{cases}$	**Tot**
VH BSM	**1.60**	**2.80**	**3.44**	**0.22**	**8.06**
VH SM	**8.39**	**9.11**	**93.99**	**3.51**	**115.00**
W^+H	2.26	2.92	21.67	1.12	27.95
W^-H	2.39	2.58	32.06	1.23	38.26
ZH	3.75	3.61	40.27	1.17	48.80
Bgd	**969.40**	**196.68**	**30373.50**	**139.74**	**31679.32**
ggH	60.74	12.73	648.03	3.32	724.83
VBF	9.85	1.82	240.99	1.71	253.83
ttH	1.25	0.63	70.80	0.39	73.07
pp$\rightarrow \gamma\gamma$	898.21	182.14	29424.30	135.10	30639.76
s/$\sqrt{s+b}$	**0.0511**	**0.1943**	**0.0197**	**0.0182**	**0.2027**

Table 5.4. Run-2 expected H$\rightarrow \gamma\gamma$ yields divided by category and production mode, in the 120 GeV $\leq m_{\gamma\gamma} \leq$ 130 GeV $m_{\gamma\gamma}$window. The several components contributing to VH SM/Bgd classes are highlighted

corroborating the goodness of the classifier implementation. The purpose of the multivariate analysis implemented to identify the bSM by their kinematics, does not rely in merely counting the events, but rather to look at how they distribute in the several categories defined on the most bSM-sensitive output of the BDT. This is summarized by the category efficiency ϵ_i that evaluates the sensitivity of the the $i-th$ category to the anomalous coupling-induced kinematics. It should be noted that this is an average bSM efficiency parameter, as it does not distinguish among the several bSM scenarios. This is justified as it is implicitly assumed that the difference in the kinematics introduced by the bSM effects altogether, with respect to the SM HVV vertex, is much more evident than difference among the several bSM hypotheses, as shown in Figure 4.6 for the Higgs transverse momentum. A finer analysis would differentiate the effects of such different anomalous scenarios, at the expenses of a much heavier computational complexity, as in the following fit procedure further categorization splits are going to be performed, and the number of such categories would sizeably increase.

5.4 Data/MC comparison

This last section of the Chapter deals with the comparison between MC samples and LHC Run-2 data. The MC samples employed are the MC samples previously utilized, not yet split into categories, nor including the *scale factors* accounting for MC samples correction to match data. The following plots are based on the available 2016preVFP and 2017 MC $pp \to \gamma\gamma$ samples, and on the Run-2 data referring to the same period of data taking, in order to minimize the impact of year-dependant systematic uncertainties. The preselection applied on both data and MC is the general preselection for $H \to \gamma\gamma$ analysis exposed in Sections 4.2 and 4.3, without imposing the VH hadronic selection - which would dramatically reduce the statistics of the samples - except for the requirement of at least two reconstructed jets in the events. The MC distributions are normalized to data, as the present purpose is only to compare the distributions shape-wise, thus avoiding to include non accurate estimation of the contributions of neglected background sources to the normalization of the distributions for simulated events. It should be premised that, rather than for strictly validating the MC samples with respect to recorded data, the following plots should be interpreted as an attempt to provide a qualitative comparison for general corroboration. While the data constitute a totally inclusive set of the largely dominant non resonant background, in addition to the negligible resonant events, the MC samples employed only include one source of background, $pp \to \gamma\gamma$. For the reasoning exposed in Section 4.1, this kind of process has been considered sufficient for modeling non resonant background, also considering the restrictions in terms of availability of reliable MC samples accounting for $pp \to \gamma + j$ or $pp \to jj$. This leads to an unavoidable incomplete modelling of the background by the MC samples, which might result in a partial mismatch in the shape of the distributions. Nonetheless, the following comparison can be regarded as an estimation by subtraction of the supposed $pp \to \gamma + j$ contribution to the shape of kinematic distribution or, alternatively, as an evaluation of the goodness of the $pp \to \gamma\gamma$ background-only assumption made in Section 4.1.

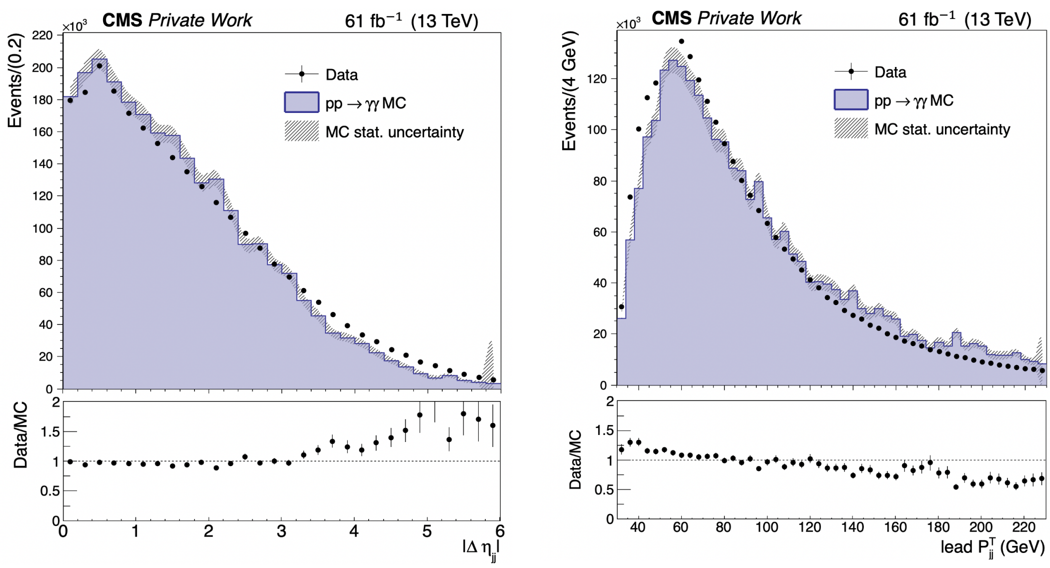


Figure 5.8. Data/MC comparison for the dijet pseudrapidity difference in magnitude (left) and the transverse momentum of the leading dijet (right). MC is normalized to data.

Figure 5.8 exhibits the data/MC comparison for two of the most discriminating BDT inputs. Despite data reproduce quite well the overall shape of the distribution, some local mismatch is evident, particularly on the tails. The mismatch is considered to be attributable to the just exposed causes, with particular regard to the missing $p \to \gamma + j$ component in the MC samples.

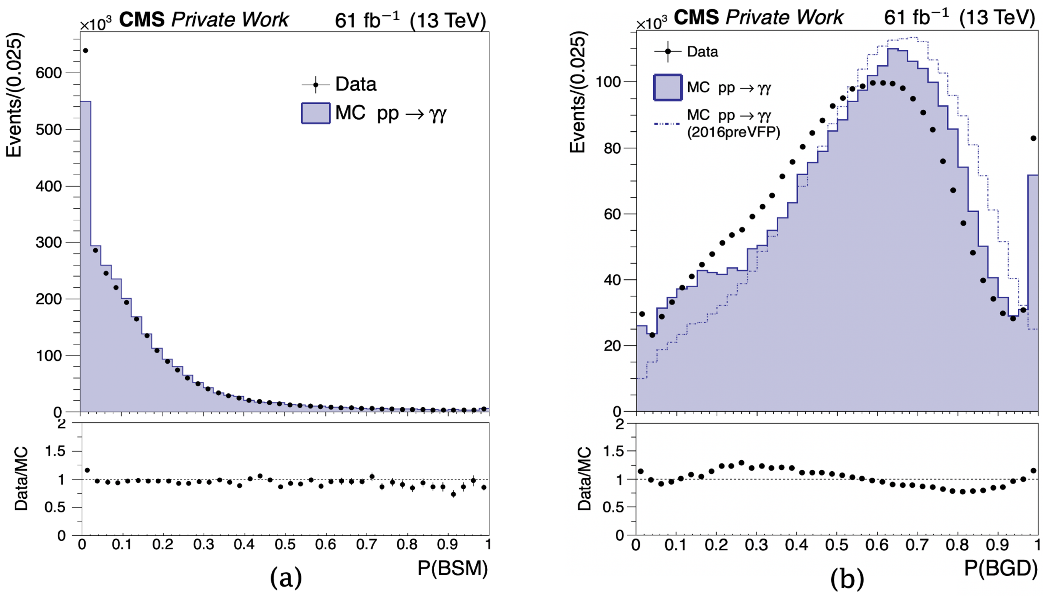


Figure 5.9. Data/MC comparison of the two most discriminating BDT outputs. MC is normalized to data.

The optimized BDT has been tested on both the data and MC samples, in order to compare the coherence of its outputs in the two cases. The result is reported in Figure 5.9 for the two most discriminating BDT outputs utilized for defining categories, namely $\mathscr{P}(BSM)$ and $\mathscr{P}(BGD)$. Due to the previously exposed reasoning about the uncompleteness of MC samples with respect to totally inclusive data, a certain fraction of events to match MC to data are missing, and those could make a contribution to the shape difference in the distribution of the $\mathscr{P}(BGD)$ output. Moreover, the analysis performed in Section 5.2.2, proved the 3 classes model as the best classification scenario in terms of algorithm performance. The BDT's *bgd* class was trained to learn the discriminating feature of both resonant and non resonant background together, while the hereby tested samples are either sole $pp \to \gamma\gamma$ simulated events or totally inclusive data. This could justify a biased, sub optimal performance of the algorithm in terms of predicting the $\mathscr{P}(BGD)$ output for the data sample, contributing to the mismatch observed in Figure 5.9 (b). The dashed line represents the BDT output on the sole 2016preVFP sample, which is part of the original training sample of the algorithm and constitutes one component of the background distributions reported in Figure 5.6. The 2016preVFP sample, already known to the algorithm, presents a peak compatible with the 2016preVFP+2017 MC sample reported, implying that the unseen 2017 events are not accountable for shifting the peak, but only for raising the tails of the distribution. The deviation of data distribution should then be traced back to the additional, unmodeled, components this sample comprises. These same missing components however, do not equally affect the distribution of $\mathscr{P}(BSM)$ shown in Figure (a), as they constitute, just as the MC events, an orthogonal sample to the *bSM* class of the algorithm, thus leading to a better agreement between data and MonteCarlo.

The effect in the mismatch of the data/MC shape for the BDT output probability will only partially influence the final fit procedure. While the non resonant background estimation will not be affected, as it will be performed on data, on the signal peak (VH+resonant background) this will likely lead to a variation in the efficiency of the categories defined. This should be included as a systematic uncertainty in the model, estimated through a proper comparison with a $Z \to ee$ control sample, which however has not been performed in the present work and could be considered as a finer optimization of the analysis.

Chapter 6

Maximum Likelihood Fit

The procedure for extracting the *parameters of interest* (POIs) relies on a binned maximum likelihood fit in the $100<m_{\gamma\gamma}<180$ GeV region, as exposed in the present section. The Higgs boson signature in the diphoton decay channel is a peak around the Higgs mass, $m_H \simeq 125$ GeV, over a smoothly decreasing background distribution, with the peak width driven by the experimental resolution on diphoton invariant mass ($\mathcal{O}$(GeV)) which is largely predominant with respect to Higgs boson's decay width ($\mathcal{O}$(MeV)).
Due to the possible low statistics of $H \rightarrow \gamma\gamma$ events in some categories - over the total number of $\gamma\gamma$ events overcoming the already exposed selection criteria in Section 4.2 - the events are considered to be Poisson-distributed. Denoting as s and b the expected Higgs and background yields respectively, the *signal strength modifier* $\mu = (\sigma \cdot BR)/(\sigma \cdot BR)_{SM}$ can be defined, as already introduced in Section 2.2. Both signal and background yields are subject to the several systematic uncertainties, presented in Section 6.3, that affect the experimental procedure and that are going to be taken into consideration as *nuisance parameters* θ, such that the yields are function of these parameters $s = s(\theta)$, $b = b(\theta)$. The nuisance parameters are considered to be distributed according to some *pdf* ρ, typically a log-normal or a gaussian distribution, centered in zero and whose width is given by some a priori knowledge on the magnitude of the systematic effect ($\hat{\theta}$).
Denoting the observed data as x, the overall likelihood for either *signal+background* or *background-only* hypothesis writes as a product of likelihoods :

$$\mathcal{L}_{s+b}(x|\mu,\theta) = \prod_k \mathcal{L}^{(k)}(x|\mu s + b) \cdot \rho(\hat{\theta}|\theta) \tag{6.1}$$

$$\mathcal{L}_{b}(x|\mu,\theta) = \prod_k \mathcal{L}^{(k)}(x|b) \cdot \rho(\hat{\theta}|\theta) \tag{6.2}$$

where k is the bin index. In each bin, assuming a diphoton invariant mass distribution $\mathcal{P}_{sig}(m_{\gamma\gamma})$ for the signal and $\mathcal{P}_{bgd}(m_{\gamma\gamma})$ for the background, the likelihood in the $s+b$ hypothesis writes as a weighted sum of signal and background pdf_s, normalized to the expected Poisson-distributed yields:

$$\mathcal{L}^{(k)}(x|\mu s+b) = \left(\prod_j \sum_i \mathcal{L}(x|\mu_i s_{i,j}+b_{i,j}) \cdot \mathcal{P}^{(s+b)}_{i,j}(m_{\gamma\gamma})\right)_k =$$
$$= \left(\prod_j \sum_i \frac{(\mu_i s_{i,j}+b_{i,j})^{n_{i,j}}}{n_{i,j}!} e^{-(\mu_i s_{i,j}+b_{i,j})} \cdot \frac{\mu_i s_{j,j}\, \mathcal{P}^{i,j}_{sig}(m_{\gamma\gamma}) + b_{i,j}\, \mathcal{P}^{i,j}_{bgd}(m_{\gamma\gamma})}{\mu_i s_{i,j}+b_{i,j}}\right)_k \tag{6.3}$$

similarly, for the *b-only* hypothesis:

$$\begin{aligned}\mathcal{L}^{(k)}(x|b) &= \left(\prod_j \sum_i \mathcal{L}(x|b_{i,j}) \cdot \mathcal{P}^{i,j}_{bgd}(m_{\gamma\gamma})\right)_k = \\ &= \left(\prod_j \sum_i \frac{(b_{i,j})^{n_{i,j}}}{n_{i,j}!} e^{-b_{i,j}} \cdot \mathcal{P}^{i,j}_{bgd}(m_{\gamma\gamma})\right)_k\end{aligned} \tag{6.4}$$

where i is the process index, either referring to ggH,VBF,VH,ttH production modes or to non-resonant background, and j is the category index labelling the *reconstructed categories* over which the fit will be performed. The signal and the background yields depend indeed on both the category and the production mode that is taken into consideration. The signal strength modifier $\mu = \mu_i$ is assumed to be varying with the process i, contrarily to the fully SM-approach which implies the same cross section scaling regardless of the production mechanism. By letting the signal strengths float for the various production mechanisms, one makes a less model-dependent assumption, indeed generalizing the picture of the SM which, given the Higgs boson mass, precisely predicts the cross sections of the different production mechanisms. The overall $\mathcal{L}(x|\mu,\theta)$ likelihood does therefore depend on a set of free parameters, namely the set of signal strengths for each production mode: $\vec{\mu} = (\mu_{ggH}, \mu_{VBF}, \mu_{VH}, \mu_{ttH})$, which is going to be extracted through a maximum likelihood procedure.
To probe data, either actual or MC toy-generated, against both *s+b* and *b-only* hypotheses, the test-statistics $\tilde{q}_\mu$ is constructed by the means of a profile likelihood ratio:

$$\tilde{q}_\mu = -2ln\left(\frac{\mathcal{L}(x|\mu,\hat{\theta}_\mu)}{\mathcal{L}(x|\hat{\mu},\hat{\theta}_\mu)}\right) \tag{6.5}$$

where $\hat{\mu}$ are the best-fit values i.e the values corresponding to the global minimum of the likelihood and $\hat{\theta}$ are the nuisance parameters obtained by minimizing the likelihood at fixed μ. The maximum likelihood problem is thus traced back to minimizing *-2 ln(*$\mathcal{L}(\mu)$*)*.

The maximum likelihood fit is carried out by fitting the $m_{\gamma\gamma}$ distribution over several reconstructed categories, including the 4 VH hadronic-optimized categories previously defined in terms of the output of the multiclass BDT defined in the previous chapter. For the purpose of extracting the VH production parameters, one might be tempted to consider only the high VH signal purity regions, which would result in a sub optimal approach. As the signal peak is an inclusive compendium of several production mechanism and, possibly, bSM scenarios, it is required to carefully estimate the contribution of resonant background sources in high VH-purity categories by simultaneously measuring the other processes in all the other categories. This procedure, based on a likelihood depending on a common set of parameters, includes all the categories regardless of their signal purity, thus exploiting all the available information in a global minimization.

Generally speaking, considering a process i and a reconstructed category j, the normalization of the signal peak in the (i,j) category is set to:

$$N_{i,j}^{sig} = (\sigma \cdot \mathscr{BR})_i \cdot (\epsilon \times \mathscr{A})_{ij} \times \mathscr{L}_{int} \tag{6.6}$$

where σ_i is the cross section of the physical process, and the efficiency ϵ_j is a multiplicative factor arising from several sources of selection (trigger, kinematic cuts, fraction of events falling in the category j, ...) and $\mathscr{A}$ is the geometrical acceptance.

Analysis categories	SM Higgs boson expected signal			
	Total	Fraction vH	σ_{eff} (GeV)	S/S+B
RECO_VH_MET_Tag2	4.7	82.6%	1.96	0.18
RECO_VBFTOPO_VHHAD_Tag0	13.4	66.3%	1.57	0.26
RECO_VBFTOPO_VHHAD_Tag1	40.9	35.6%	1.58	0.11
RECO_VH_MET_Tag1	2.7	96.6%	2.03	0.33

Table 6.1. Expected yields and VH signal purity in the optimized VH-hadronic categories. From top to bottom, the reported categories correspond to the *Cat i (i=1,...,4)* categories reported in Table 5.4

.

Table 6.1 shows the expected yields and VH signal purity, for each of the 4 categories previously optimized. The significance and the effective sigma, corresponding to the FWHM of the signal peak, are also reported. It should be noticed that the table cannot be directly compared to Table 5.4 in Section 5.3 as the reconstructed categories included in the fit are subject to a stricter VH-hadronic selection on top of the BDT-based categorization, thus reducing the yields of resonant background in the categories and increasing VH signal purity. Furthermore, these categories are not orthogonal and a hierarchy for assigning events passing the selection of multiple categories is defined, in order to privilege the production mechanisms with lower cross section. It should be also considered that the expected yields in Table 5.4 are estimated from the sole 2016preVFP MC sample, then normalized to Run-2 integrated luminosity, thus neglecting year-dependant systematic uncertainties and calibrations that could lead to event migration in the categories.

In order to perform the fit to extract the signal yields, the signal peak and the background $m_{\gamma\gamma}$ distributions, in each considered category, must be modeled. While the signal modelling is performed on Monte-Carlo samples, for background it is directly made directly on data, on properly defined control regions.

6.1 Signal modeling

A parametric signal model is defined for fitting the mass peak on the MC samples in a number categories equal to $N_{\text{year}} \times N_{\text{process}} \times N_{\text{cat}} \times N_{\text{vtx}} = 4 \times 6 \times 4 \times 2 = 192$, if considering the 4 VH hadronic-optimized categories defined. By taking into account all the 12 VH reconstructed categories, including the leptonic ones, the number of overall categories reaches 576. By further extending the approach to the whole set of 78 reconstructed categories, the number rises to 3744 different categories. Even if it complicates the model, considering all the categories is necessary to constrain the large amount of ggH and VBF contribution to VH categories, so the last scenario is the one that is going to be considered. The other splitting criteria are the Run-2 year of operation, which might lead to differences in mass resolution and to year-dependent systematic uncertainties, the production process (with VH production split in W^{+}H, W^{-}H, ZH) and 2 different vertex-scenarios. Being a traceless neutral particle, the photon vertex reconstruction is not trivial and might lead to a vertex identification suffering from quite a sizeable uncertainty. If the vertex is known with an uncertainty greater than 1cm, its contribution to the smearing of the diphoton invariant mass distribution is dominant with respect to ECAL's. For this purpose, a cut on the output of the *vertex* BDT cited in Section 4.2 is performed, thus identifying the right-vertex (RV) and wrong-vertex (WV) scenarios, characterized by different mass resolutions.

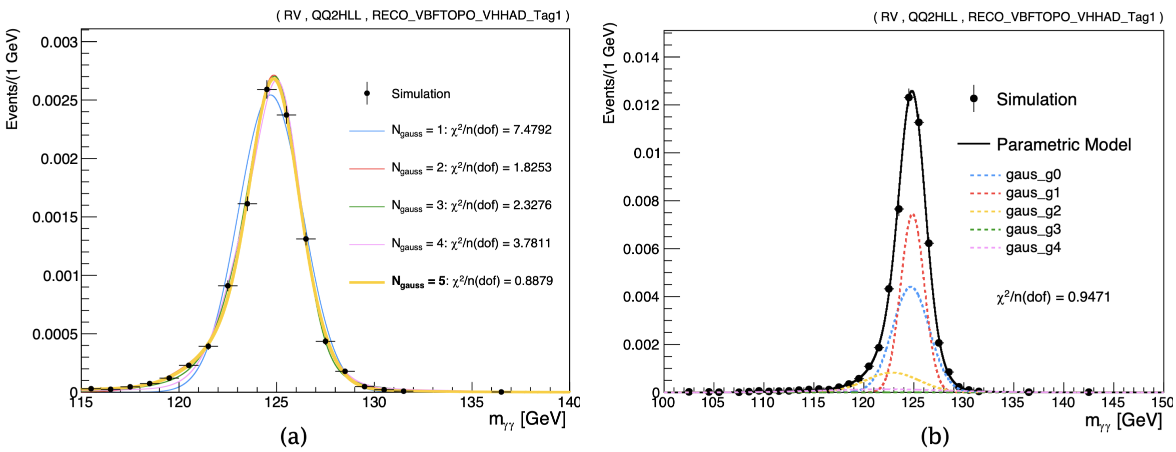


Figure 6.1. Signal fTest for 2017 ZH events, RV scenario

Despite a Crystal-Ball function - consisting in a Gaussian core with power law tails - could suit for the purpose as the peak has non a Gaussian left tail due to ECAL

leakage, the signal is going to be fitted assuming a sum of N gaussians as fit function. This is because the N-Gaussians fit is more robust and more easily implementable on the large amount of categories defined. An *fTest* is performed to extract, for each category, the optimal number of Gaussian functions in the sum, testing the hypothesis from N=1 to N=5, in order to minimize the χ^2. Figure 6.1 (a) shows the N-Gaussians models tested, while Figure 6.1 (b) reports the best fitting function and the N (=5) Gaussian components making up the sum, for one specific VH-hadronic optimized category (*RECO_VBFTOPO_VHHAD_Tag1*), which corresponds to the *Cat 3* ($P_{BSM} < 0.49, P_{BGD} \geq 0.03$) category of Table 5.4 and is taken as a reference for the plots in the following.
In the fitting parameters of the Gaussian terms making up the signal, the mean value of the distribution is a free fit parameter. Shifting the maximum of the *pdf* however, means shifting the Higgs mass and, consequently, all the parameters in Eq. 6.6 depending by it, affecting the expected yields in the category. Thus, the MC samples for three different m_H points (120,125,130 GeV) are considered, in order to interpolate each signal model parameter as a function of m_H, as shown in Figure 6.2 (a).
The cross section and the $H \to \gamma\gamma$ branching ratio are analytically computed as function of m_H, while the energy dependent efficiency ϵ at different mass points is directly extracted from samples. The interpolated functions out of these 3 mass points are used when fitting, either for the signal strengths or the anomalous couplings, leaving the Higgs mass free to float around the m_H=125 GeV starting point. By varying Higgs mass, the peak signal varies both in shape and normalization, and could also lead to a variation in the number of expected yields.
Once the fTest is done and the m_H dependence of the parameters is extracted, the proper fit of the N gaussians is carried out in each category. The outcomes are then packaged, combining the results from different years altogether. An example is given by Figure 6.3, where the peak of the associated production with the Z boson is reported.

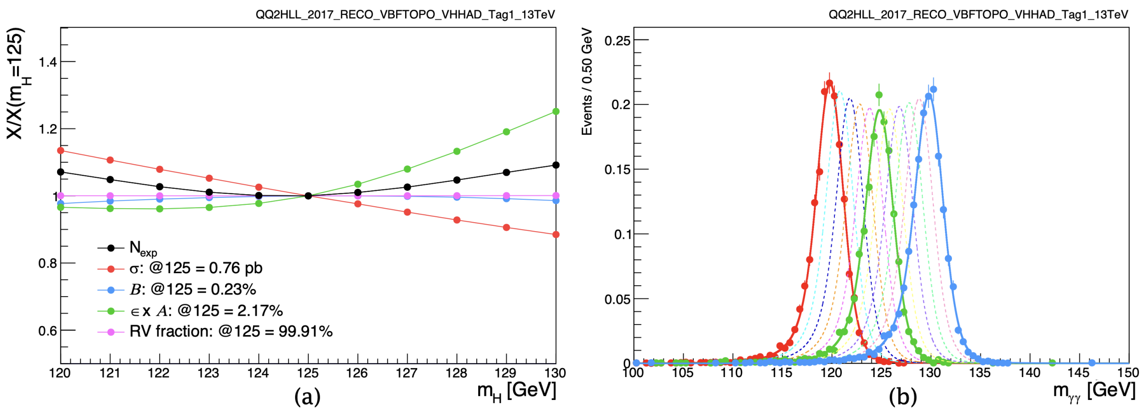


Figure 6.2. On the left, extrapolated trend for m_H-depending parameters. On the right, signal peak at the 3 different mass points from the MC samples, with the extrapolated intermediate shapes reported with dashed lines.

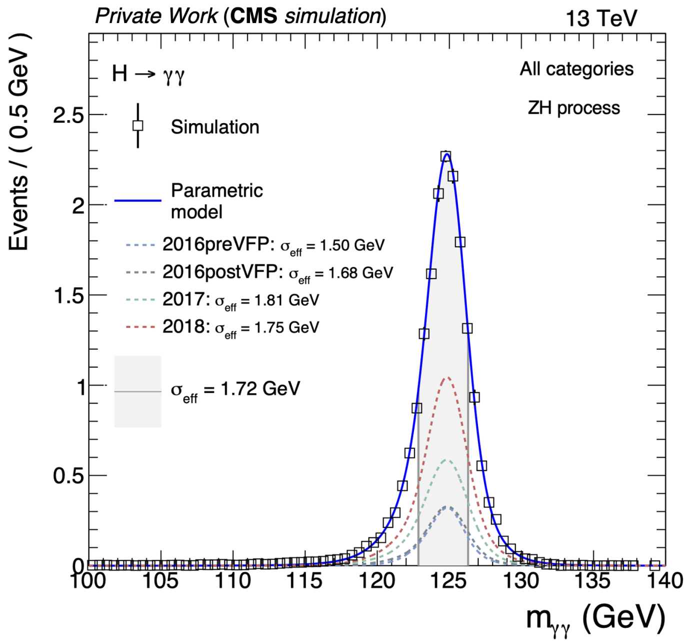


Figure 6.3. Inclusive plot for ZH associated production, all categories and years merged. Contribution of different years to the peak are highlighted

6.2 Background modeling

Under the reasonable assumptions about both its dominant contribution to non resonant background and similar kinematic pdf$_s$ shape with respect to $pp \to \gamma + j$ events, the $pp \to \gamma\gamma$ has been so far considered the sole source of non resonant background. This appears legitimate for the purpose of training a multivariate analysis algorithm for a kinematic-based discrimination of events. When it comes to fitting the yields and evaluating the contribution of such background source to the categories however, this kind of non rigorous discussion, together with the absence of reliable MC samples for modeling every source of non resonant background, is no longer suitable and is replaced by a data-driven approach. The Run-2 data available, when restricted to proper control regions, provides a totally inclusive source of non resonant background that can be considered for estimating the yields also in the region where the Higgs peak is present. Blinding the signal region comprised in 115$\leq m_{\gamma\gamma} \leq$135 GeV, the data can be fitted on the sidebands of the $m_{\gamma\gamma}$ distribution. Several fitting functions are tried out, including power law and exponential functions, Laurent series and Bernstein polynomials. Utilizing the best modeling function, the number of non resonant background events under the resonant peak can be computed by integrating the function in the signal region. The peculiar choice of a function for modeling the background introduces a bias in the analysis, as the yields and eventually the POIs extracted also depend by the specific function picked, so an uncertainty should be assigned to this choice. A *discrete profiling* approach to the problem is adopted, labelling the possible fitting function selections with a discrete index and randomly sampling them in order to generate toy MC. The POIs extrapolated in this scenario, account for systematic uncertainties introduced by the background modeling selection.

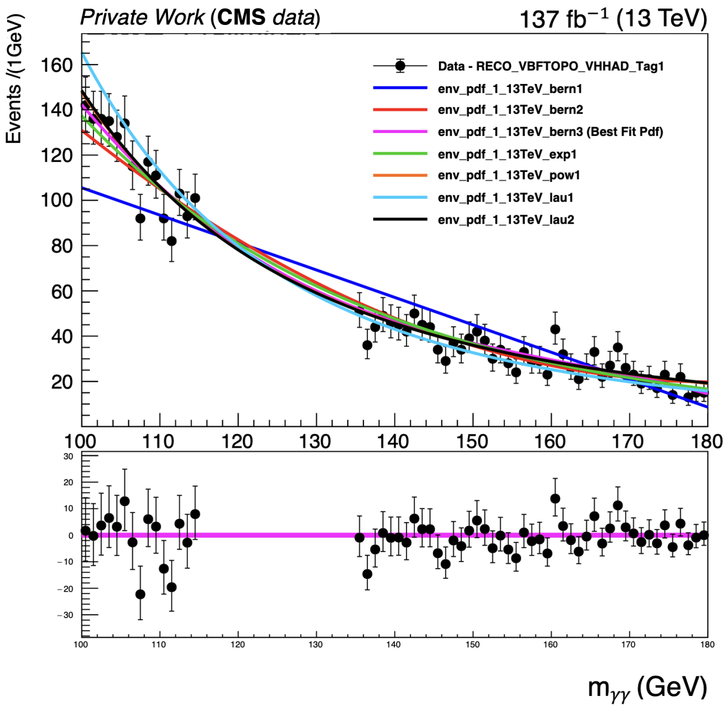


Figure 6.4. Background fit on the sidebands of the $m_{\gamma\gamma}$ distribution, blinding the signal region. The third order Bernstein polynomial is the best-fitting function for the VH-hadronic optimized reference category

6.3 Systematic uncertainties

The several sources of systematic uncertainties can be split into two large subgroups: experimental uncertainties and theoretical uncertainties. Systematic uncertainties, as already mentioned in the introduction to this chapter, are taken into account as nuisance parameters. Among the possible sources of uncertainties, one could identify the ones affecting the $m_{\gamma\gamma}$ signal shape, such as the ones regarding the energy reconstruction of the photon, or those which do only modify the expected event yields, including the theoretical ones and the experimental uncertainties on the BDT used for categorization. The MC samples employed throughout the analysis include a non complete list of systematic uncertainties, either experimental or theoretical, possibly leading to an underestimation of the error estimated on the POI_s.

6.3.1 Experimental uncertainties

A subgroup of the experimental uncertainties contribute to modify the shape of the $m_{\gamma\gamma}$ distribution:

- *Photon energy scale and resolution*: the estimation of the energy of the reconstructed photons in the ECAL may suffer from several systematic effects

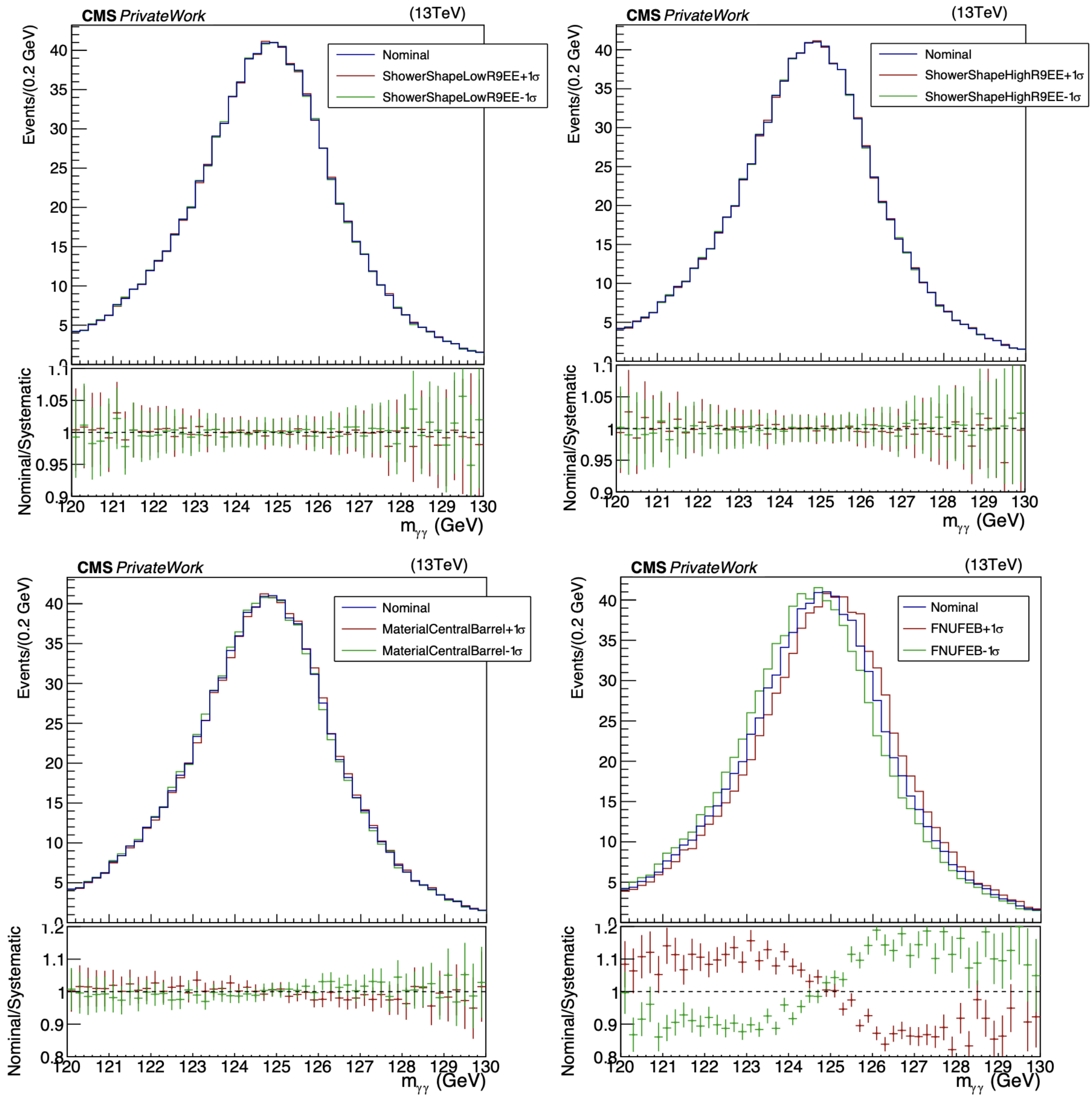


Figure 6.5. Some of the experimental systematic uncertainties affecting the diphoton mass measurement, the whole Run-2 MC sample is reported. For each case, the nominal $m_{\gamma\gamma}$ distribution is compared to the equivalent distribution corrected by $\pm 1\sigma$ the uncertainty produced by the systematic in consideration

such as: incomplete containment of the shower, loss of the transparency of the crystals, conversion in the tracker and pileup. To correct these deviations some correction factors are applied on the reconstructed energy in order to match data with MC through a common tag and probe technique with a $Z \to ee$ sample with electrons reconstructed as photons. Nonetheless, some mismatch can still persist and it is considered as a nuisance parameter in the fit. Since the photons converted in the tracker tend to produce more laterally diffused shower, this kind of systematic uncertainty is computed separately not only for endcap/barrel regions but also according to two different (Low/High) R_9 intervals. Most of the photons are affected by an energy scale uncertainty of 0.05-0.15%, with the most energetic ones reaching 0.5-3.0%.

- *Non linearity of photon energy scale* possible non linearity effects in the photon energy reconstruction may arise from non uniformity in the longitudinal crystal response which could lead to a coordinate-dependent light yield, from saturation effects which tend to affect scintillating crystals, as well as from readout electronics noise. A 0.2 % uncertainty on the whole p_T range is assigned.
- *Shower shape corrections* this kind of uncertainty accounts for imperfect modeling of shower shape in simulation, leading to an energy scale uncertainty ranging in 0.01-0.15%.
- *Longitudinal non uniformity of light collection* the modeling of light yield and collection might be dependent by the longitudinal starting point of the shower in the ECAL. The shower maximum for photons occurs typically deeper than for electrons, introducing a mismatch in the calibration of this effect through simulation. The uncertainty assigned is 0.16-0.25% for photons with R_9>0.96, whilst for low R_9 photons amounts to less than 0.07%.
- *Modelling of the material in front of the ECAL* the simulation of the interactions (photon conversion and multiple scattering mainly) in the amount of material crossed by photons before reaching the ECAL might be not perfectly reproduced in simulation. The associated uncertainty ranges from 0.02-0.05% for the most central photons, increasing to as much as 0.24% for the photons in the endcap.

These kind of uncertainties are typically considered to be gaussian-distributed. Figure 6.5 reports some of the shape-affecting experimental uncertainties. The *FNUFEB* uncertainty, among those reported, appears to be the most significant one. It accounts for the mismatch in the prediction of the shower longitudinal starting point, leading to a possible imperfect containment of the shower, in the η region covered by the barrel. *MaterialCentralBarrel* describe the $m_{\gamma\gamma}$ shape modification due to the uncertainty in the simulation of the material before ECAL in the central region of the detector. *ShowerShape(Low/High)R9EE* accounts for imperfect simulation of the supercluster shape in ECAL's endcap for converted/unconverted photons.

Another type of experimental systematic uncertainties are those who do not affect the distribution shape, but only the number of expected yields. These are generally treated as log-normally distributed corrections to pdf_s normalization. Some of the most important systematic uncertainties of this kind are:

- *Integrated luminosity*: the uncertainty on the total luminosity of the Run-2 period is 1.8 %, with a single-year uncertainty of 2.5, 2.3, 2.5 % for 2016, 2017, 2018 eras respectively. The single-year uncertainties are partially correlated as they are measured according to a common scheme.

- *Photon ID BDT score*: also the BDT score for the discrimination of prompt photons from mis-reconstructed photons is subject to a systematic uncertainty, which is estimated by varying the sample over which the algorithm is trained. The impact on most sensible categories is estimated to be around 3 %.

- *Trigger efficiency*: estimated through a tag and probe technique with a $Z \to ee$ sample, the uncertainty is $\lesssim 1$ %.

- *Photon preselection*: the preselection imposed to constrain a specific region of phase space introduces a further source of uncertainty as a different fraction of events can pass the cut in data and MC. The uncertainty is computed as the ratio of preselection efficiency in data and MC, which results as less than 1 %.

- *Jet energy scale*: some p_T and η-dependent scale correction to the reconstructed energy of the jets constitute a source of systematic error, which is estimated in few percent of the total energy. This source of uncertainty has a direct effect on the yields in the categories, with those targeting VH hadronic events being particularly sensible to it. The effect of these uncertainties on the yields is evaluated by varying the input jet energies within their uncertainties and then estimating the yields variation in each category. The discrepancy in term of yields can be as high as 20 % for the most sensible categories.

6.3.2 Theoretical uncertainties

Theoretical uncertainties arise from imperfect knowledge about the underlying theoretical models to the physical processes that are being simulated. In the MC samples that are utilized, the sole source of theoretical uncertainties considered are :

- *ggH associated scales and migrations* due to QCD scale corrections for the dominant ggH production mechanism. This leads to both an uncertainty on the total production cross section and to a migration of event among categories.

Some additional theoretical uncertainties could regard the imperfect knowledge about the strong coupling constant due to effects of higher order terms, which could account for inaccurate prediction of the kinematics involving QCD processes occurring at parton level, the uncertainty on the $H \to \gamma\gamma$ branching ratio, the uncertainty on the momentum *pdf* of the partons and on the parton shower physics.

6.4 Fit results

The results extracted from signal and background modelling are stored into a text file including all the necessary information about pdf shapes, extracted yields and systematic uncertainties corrections to the expected yields and/or pdf momenta, for each category and each year. This constitutes the input to the simultaneous fit model, whose purpose is minimizing the negative delta log likelihood 6.5.

6.4.1 Signal strength

The fit is either performed by a single global minimization of the likelihood to extract the best estimation for the vector $\vec{\mu}$ of POIs, or with a one-dimensional scan of the likelihood. For a specific process, the scan is implemented by fixing its signal strength μ to a certain value within its range of definition, then performing the minimization with respect to all the other free parameters. By default, μ_{ggH} and μ_{VBF} are allowed to vary in the [0,2] interval, while μ_{VH} and μ_{top} vary in [0,3] and [0,4], respectively. This technique allows to reconstruct to shape of the likelihood nearby the minimum, thus providing a graphical way to estimate both the uncertainty on the POI and the statistical significance of the result obtained.

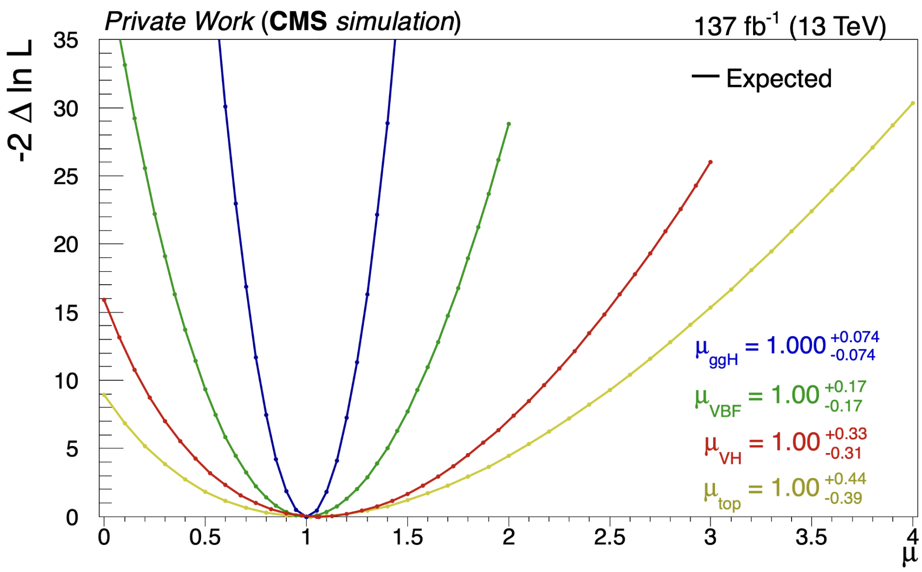


Figure 6.6. Profile 1D likelihood scan for the signal strength of different production mechanisms on the Asimov dataset. The bestfit values, corresponding to the likelihood minima, are reported with their uncertainty, both statistical and systematical.

At first stage, data is kept blind, and the extraction of the *expected* parameters of interest is performed over an *Asimov dataset*, accounting for an idealized scenario where background and signal contributions distribute according to the mean values of the fitted pdf_s. This provides a useful benchmark to assess the sensitivity of the analysis. The result of the profile 1D scan on such a dataset is reported in Figure 6.6 for the whole set of production modes. The expected signal strengths identify with the SM hypothesis $\mu = 1$, by construction. It can be noticed that the higher the cross section of the mechanism, the smaller the uncertainties affecting the estimation and the higher the statistical significance. The analysis with respect to the VH production mechanism appears to suffer from lower sensitivity with respect to ggH/VBF due to the limited number of events. The expected signal strength for the VH associated production:

$$\mu_{VH}^{exp.} = 1.000 \ (Syst)^{+0.013}_{-0.016} \ (Stat)^{+0.331}_{-0.310} \tag{6.7}$$

is estimated with a statistical significance of:

$$\sqrt{q(\mu = 0)} \simeq \sqrt{16} = 4\sigma \tag{6.8}$$

corresponding to the square root of the value assumed by the test statistics in $\mu = 0$ which accounts for the no signal hypothesis.

Data is then unblinded to probe the results obtained on the Asimov dataset for the MC-modeled signal and the data-modeled background. Figure 6.7 shows the likelihood profile for the fit performed on data. Unlike the *expected* scenario, the *observed* yields in the categories provided by data suffer from fluctuations due to limited statistics, particularly for the production modes with lower cross section. This is the case of the VH production, whose best value for the fitted signal strength:

$$\mu_{VH}^{obs} = 1.56^{+0.36}_{-0.34} \tag{6.9}$$

deviates 1.6σ from the expected value. This value is observed with a $\simeq 6\sigma$ statistical significance with respect to the background-only hypothesis, furthermore showing an error in line with the expected prediction. The best estimation of μ for ggH production on the other hand, is considerably closer to the SM hypothesis and, being by far the most dominant production mechanism, benefits of a larger statistics, nonetheless exhibiting a smaller uncertainty. The ttH production case, apart from

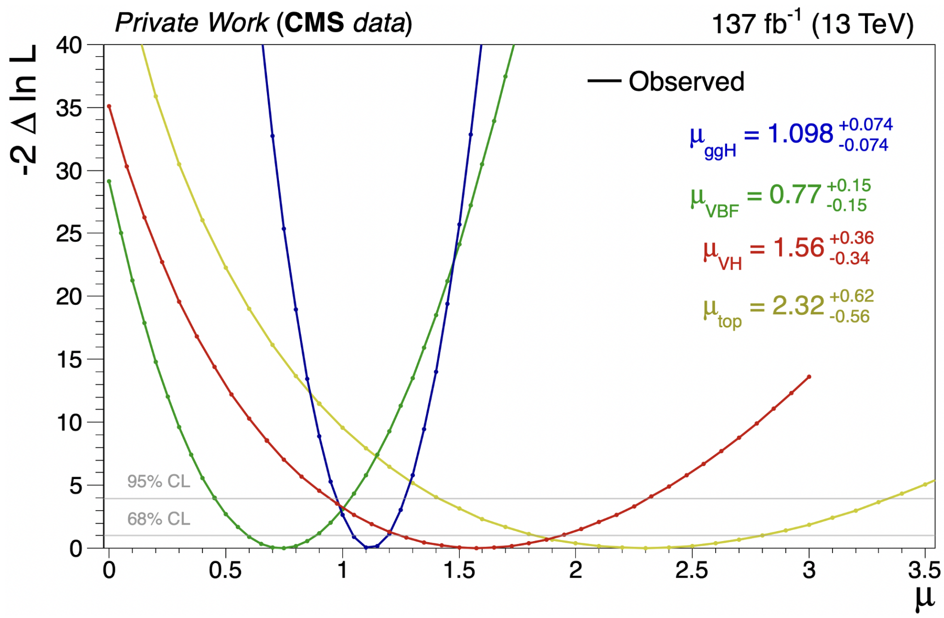


Figure 6.7. Profile 1D likelihood scan for the signal strength of different production mechanisms on the observed data. The bestfit values corresponding to the likelihood minima, are reported with their uncertainty, both statistical and systematical.

being the production mode with the lowest cross section, exhibits a greater deviation from SM. It must be considered that during the analysis it has been necessary to remove some ttH categories due to technical issues, thus spoiling the sensitivity of the analysis to this mechanism and leading to a poor estimation of its POI.
It should be recalled that the set of systematic uncertainties considered is incomplete, leading to an underestimation of the error on the POI_s and to a greater distance, in terms of number of standard deviations, from the expected values.

After performing the fit over the whole set of categories, in Figure 6.8 are reported the fitted pdf_s imposed on data, once weighted by the statistical significance of each category. The Higgs peak clearly emerges at an energy value of 125.38 GeV, with data perfectly fitting the background + signal models within the uncertainty bands.

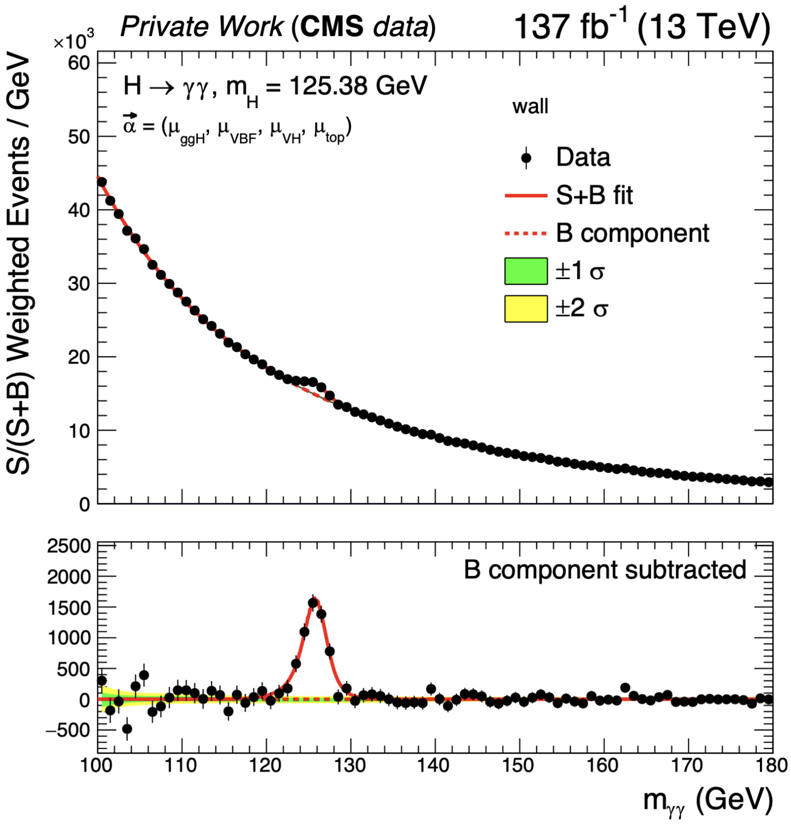


Figure 6.8. Diphoton invariant mass distribution with each event weighted with the (S/S+B) value of its category, with S being the inclusive number of Higgs production events under the peak. All categories are included. The 1σ and 2σ uncertainty bands are produced through toy MC_s.

The same result is reported in Figure 6.9 for the sole VH categories, accounting for the VH signal under study in the present work.
Table 6.2 reports the signal strengths extracted, both expected and observed

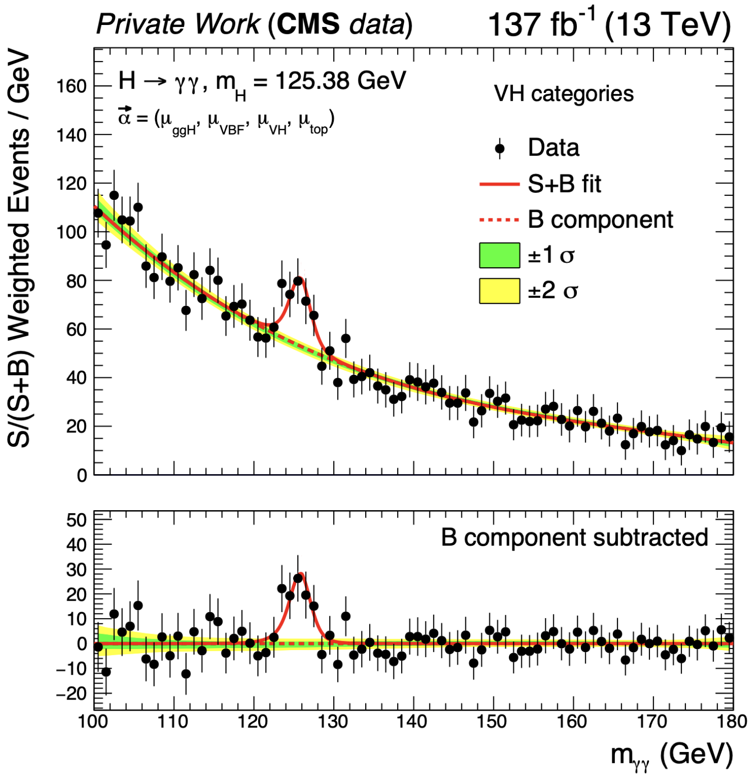


Figure 6.9. Diphoton invariant mass distribution with each event weighted with the (S/S+B) value for the 12 VH categories, accounting for a VH signal observed with a statistical significance of 6σ.

Parameter	Expected	Observed
μ_{ggH}	1.000 $\pm$0.074	1.098 $\pm$0.074
μ_{VBF}	1.00$\pm$0.17	0.77 $\pm$0.15
μ_{VH}	$1.00^{+0.33}_{-0.31}$	$1.56^{+0.36}_{-0.34}$
μ_{top}	$1.00^{+0.44}_{-0.39}$	$2.32^{+0.62}_{-0.56}$

Table 6.2. Fitted signal strengths

6.4.2 Anomalous couplings

The estimation of possible bSM contributions to the HVV vertex in the (V)H$\rightarrow \gamma\gamma$ process is given in terms of the fractional cross sections f_{a_i} that have been introduced in Section 2.2.1. Being complex numbers, whose magnitude is comprised in [0,1], they possess a phase $\phi_{a_i} = arg(a_i/a_1)$, where a_i are the proper anomalous couplings accounting for anomalous contribution to the HVV vertex appearing in the scattering amplitude in Eq. 2.41. The results reported in this section come from the fit extraction of the $f_{a_i} cos(\phi_i)$ quantity, where the physical constraint $cos(\phi_{a_i}) = \pm 1$ is imposed, in order for the couplings to be real. From Eq. 2.45, the magnitude of the fractional cross sections are explicitly written as the ratio of the specific bSM scenario contribution over the totally inclusive cross section, weighted by the square magnitude of the relative anomalous coupling:

$$
\begin{aligned}
f_{a_3} &= \frac{|a_3|^2\sigma_3}{|a_1|^2\sigma_1 + |a_2|^2\sigma_2 + |a_3|^2\sigma_3 + |\kappa_1|^2\sigma_{\Lambda_1} + |\kappa_2^{Z\gamma}|^2\sigma_{\Lambda_1}^{Z\gamma}} \\
f_{a_2} &= \frac{|a_2|^2\sigma_2}{|a_1|^2\sigma_1 + |a_2|^2\sigma_2 + |a_3|^2\sigma_3 + |\kappa_1|^2\sigma_{\Lambda_1} + |\kappa_2^{Z\gamma}|^2\sigma_{\Lambda_1}^{Z\gamma}} \\
f_{\Lambda_1} &= \frac{|\kappa_1|^2\sigma_{\Lambda_1}}{|a_1|^2\sigma_1 + |a_2|^2\sigma_2 + |a_3|^2\sigma_3 + |\kappa_1|^2\sigma_{\Lambda_1} + |\kappa_2^{Z\gamma}|^2\sigma_{\Lambda_1}^{Z\gamma}} \\
f_{\Lambda_1^{Z\gamma}} &= \frac{|\kappa_2^{Z\gamma}|^2\sigma_{\Lambda_1}^{Z\gamma}}{|a_1|^2\sigma_1 + |a_2|^2\sigma_2 + |a_3|^2\sigma_3 + |\kappa_1|^2\sigma_{\Lambda_1} + |\kappa_2^{Z\gamma}|^2\sigma_{\Lambda_1}^{Z\gamma}}
\end{aligned}
\tag{6.10}
$$

where σ_i is the cross section for the process corresponding to $a_i = 1$ with all other couplings set to zero and σ_1 refers to the SM case.

The new single-bin likelihood function writes as:

$$
\mathcal{L}(x|\mu, \theta, f_{a_n}) = \prod_j \sum_i e^{-(\mu s_{i,j} + b_{i,j})} \left[\mu\, s_{i,j}\, \mathcal{P}_{sig}^{i,j}(x; \mu, f_{a_n}) + b_{i,j}\, \mathcal{P}_{bgd}^{i,j}(x)\right] \tag{6.11}
$$

where $\mathcal{P}$ are the distributions defined for each signal and background process i in each category j. Two signal strength parameters are considered: μ_V for the VBF and VH production which are related by the same HVV coupling, and μ_f for ggH/ttH production which are characterized by Higgs-top interaction. These parameters are left as free parameters in the fit.
The anomalous couplings are included in the picture by generalizing the signal model as:

$$
\mathcal{P}_{sig}^{i,j}(x; \mu, (f_{a_n}, \phi_{a_n})) = (1 - f_{a_n})\mathcal{P}_{a_1}^{i,j}(x) + f_{a_n}\mathcal{P}_{a_n}^{i,j}(x) + \sqrt{f_{a_n}(1 - f_{a_n})}\mathcal{P}_{a_1 a_n}^{i,j}(x; \phi_{a_n}) \tag{6.12}
$$

where $\mathcal{P}_{a_1}$ is the pure SM *pdf*, $\mathcal{P}_{a_n}$ is the pure bSM *pdf* and $\mathcal{P}_{a_1 a_n}$ models the interference between the two cases, all properly normalized to the number of yields for the category (i,j). When evaluating the process amplitude indeed, an interference between SM and bSM scenarios can occur since they are characterized by the same initial and final state. The resulting interference term in the amplitude is evaluated as a linear combination of pure SM hypothesis (f_{a_i}=0), pure bSM (f_{a_i}=1), or an hybrid case where events are equally generated SM/bSM (f_{a_i}=0.5), each of which is

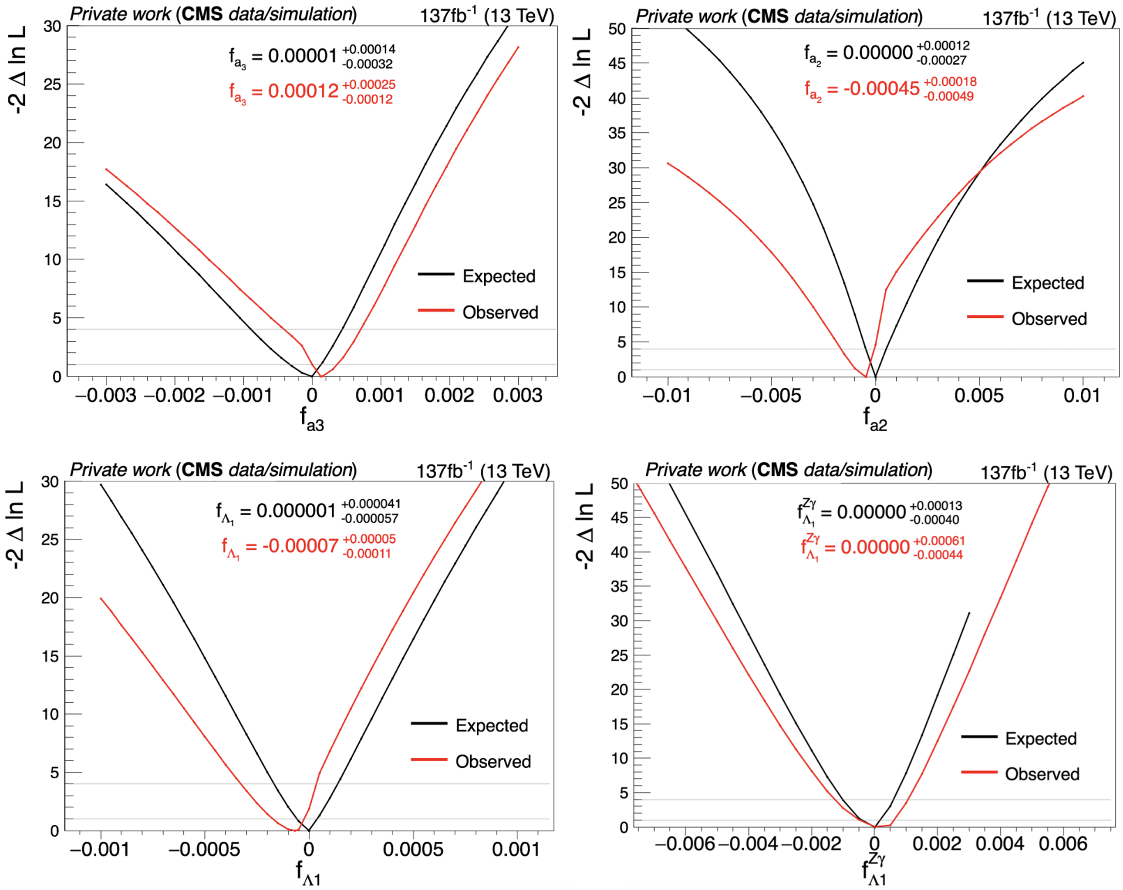


Figure 6.10. Profile likelihood scan for $\vec{f} = (f_{a2}, f_{a3}, f_{\Lambda_1}, f^{Z\gamma}_{\Lambda_1})$ POIs, both on the Asimov dataset and on data. The approach used scans the f_{a_i} under study while keeping the others to zero.

modeled by a MC sample.

The $\vec{f} = (f_{a2}, f_{a3}, f_{\Lambda_1}, f^{Z\gamma}_{\Lambda_1})$ POI_s are then extracted by letting one of the f_{a_n} components float while fixing the others to zero. For reasons of convention, and in order to compare anomalous coefficients extracted from different production modes and decay channels, the fractional cross sections f_{a_n} are defined with respect to the cross section of the $H \to ZZ \to 2e2\mu$ process, which exhibits a HVV vertex both in production and decay in the case of VBF/VH production .
The likelihood in Eq. 6.11 is maximized by the means of the same profile likelihood ratio as for Eq. 6.5, with the likelihood now depending also on the f_{a_i} parameters. The likelihood is optimized with respect to the POIs f_{a_i}, the yield parameters μ and with respect to the nuisance parameters which include the constrained parameters describing the systematic uncertainties. The confidence intervals are determined from profile likelihood scans of the respective parameters. The allowed 68% and 95.4% CL intervals are defined using the profile likelihood function, and correspond to negative delta log-likelihood values of -2 $\Delta\mathcal{L}$ = 1 and -2 $\Delta\mathcal{L}$ = 3.99 respectively.

Figure 6.10 reports the results obtained for the parameters of interest, which do not appear to be significantly different from zero.

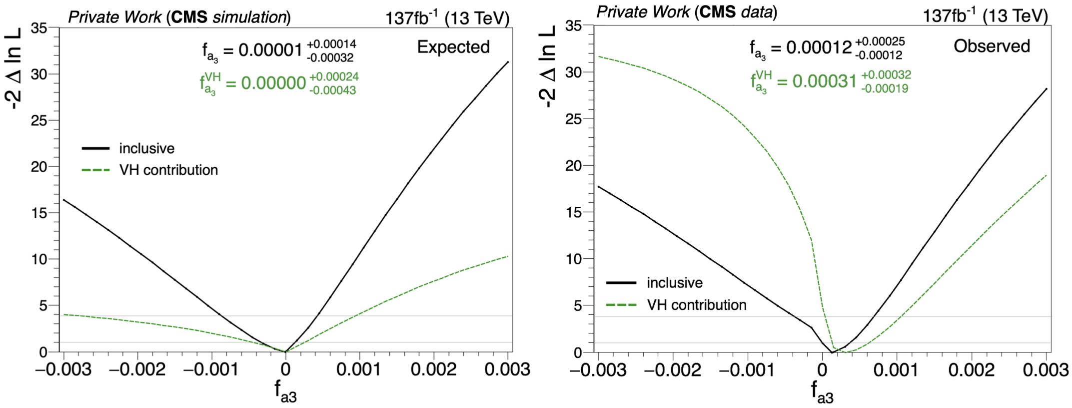


Figure 6.11. VH-only f_{a_3} scan compared to the inclusive case, both expected and observed on data.

As the VH associated production is the main mechanism under study in the present thesis, it is meaningful to evaluate its contribution to at least one of the anomalous fractional cross sections. Since in the fit procedure for the anomalous couplings the VH production is tied to the VBF production, by a common scaling parameter μ_V, immediately disentangling the sole VH contribution appears not feasible. An alternative, approximated way to do so is to freeze during the fit the scaling parameters μ_V and μ_f to their SM values - from which they do not appear to significantly deviate as for the previous Section - and to perform the fit only over the reconstructed VH categories. Once the sole VH categories are taken into account, the ggH and VBF contribution in such categories can not be accurately constrained and this is why their magnitude is fixed a priori to the expected SM value. In this way, the contribution to the fit arises only from categories with high VH signal purity, thus taking into account other production mechanisms only to a small extent. This however leads to evaluating the likelihood scan in different conditions with respect to the ones in Figure 6.10, as the signal strengths are not free to float. The VH contribution reported in Figure 6.11 thus appears to be partly optimistic, or at least affected by an underestimated statistical uncertainty, which is the result of constraining the μ parameters. The results reported in Figure 6.11, in the attempt of quantifying the VH contribution to the pseudoscalar Higgs hypothesis, show a result that is totally compatible with zero.

Table 6.3 summarizes the results obtained for the extraction of the Higgs anomalous couplings to the massive vector bosons in terms of the fractional cross sections. The parameters do not appear to deviate significantly from zero. The sole f_{a_2} parameter, accounting for a tensor structure for the HVV vertex that could lead to a higher

Parameter	Expected/(10^{-3})		Observed/(10^{-3})	
	68% CL	95%CL	68% CL	95% CL
f_{a_3}	$0.00^{+0.14}_{-0.32}$	[-0.86,0.41]	$0.12^{+0.21}_{-0.12}$	[-0.37,0.68]
f_{a_2}	$0.00^{+0.12}_{-0.77}$	[-0.46,0.49]	$-0.45^{+0.18}_{-0.49}$	[-1.62,-0.07]
f_{Λ_1}	$0.00^{+0.04}_{-0.06}$	[-0.16,0.13]	$-0.07^{+0.05}_{-0.11}$	[-0.32,0.03]
$f^{Z\gamma}_{\Lambda_1}$	$0.00^{+0.13}_{-0.48}$	[-0.97,0.58]	$0.00^{+0.61}_{-0.44}$	[-1.20, 1.04]

Table 6.3. Fitted anomalous fractional cross sections reported at 68% and 95% CL. Each parameter is extracted while fixing the others to zero.

order CP-even correction to the Higgs-massive vector boson coupling, results to be compatible with zero slightly off 2σ. This does not result sufficient to prove any bSM hypothesis, also recalling the error underestimation affecting the parameters extracted in the analysis, and requires a further corroboration on new data.

Conclusions

In the present thesis several new physics hypotheses for the Higgs boson have been probed on Run-2 data, with the purpose of giving some constraints on the possibility of bSM effects emerging from the Higgs boson interaction with massive gauge bosons. With particular regard to the VH associated production, in the vector boson hadronic decay channel, a blind optimization procedure has been performed to enhance the sensitivity of the analysis with respect to this kind of process. The final fit procedure however, took into account an inclusive set of categories, in order to properly evaluate resonant background contribution in VH categories. The signal strength for the VH production extracted from data: $\mu_{VH}^{obs} = 1.56^{+0.36}_{-0.34}$, observed with a statistical significance of $\simeq 6\sigma$, appears to deviate 1.6σ from SM hypothesis. An underestimation of the error on the value is considered to be attributable to an underestimation of the systematic uncertainties, as some of them were not included in the analysis, with a specific reference to the theoretical ones.
The EFT framework provided an expression for the most general scattering amplitude for an HVV vertex (V=W,Z), including possible bSM scenarios. These alternative hypotheses account for a pseudo-scalar CP violating coupling, a CP-even higher order correction and an energy dependent coupling, defined also for the HZγ vertex. Each of these scenarios are parameterized by additional Wilson operators, and their relative coefficients, from which the fractional cross sections f_{a_i} are defined. The fit procedure has been extended in a way to directly extract f_{a_i} parameters from data in a unique minimization with the Higgs boson signal extraction. The results, reported in Table 6.3, show no significant deviation from the $f_{a_i} = 0$ SM hypothesis, thus rejecting the possibility of a bSM Higgs boson as regards the current analysis performed on LHC Run-2 data.

Bibliography

[1] The CMS collaboration, "Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV," Physical Review D - vol. 92, 7 2015.

[2] The ATLAS collaboration, "Study of the spin and parity of the Higgs boson in diboson decays with the ATLAS detector", arXiv:1506.05669 [hep-ex]

[3] T. C. collaboration, "Measurements of the Higgs boson width and anomalous HVV couplings from on-shell and off-shell production in the four-lepton final state," Physical Review D - vol. 99, 6 2019.

[4] M. Tanabashi et al. (Particle Data Group), Phys. Rev. D 98, 030001 (2018)

[5] Sheldon L. Glashow, "Partial-symmetries of weak interactions", Nucl. Phys. 20, 579 (1961);

[6] Steven Weinberg, "A model of leptons", Phys. Rev. Lett., 19:12641266, Nov 1967.

[7] Southampton University notes, "The renormalization group", https://www.southampton.ac.uk/ doug/qft/aqft4.pdf

[8] Peter W. Higgs, "Broken symmetries and the masses of gauge bosons", Phys. Rev. Lett., 13:508509, Oct 1964.

[9] John M. Cornwall, David N. Levin, and George Tiktopoulo "Derivation of gauge invariance from high-energy unitarity bounds on the S matrix", Phys. Rev. D 10, 1145 (1974)

[10] M. Tanabashi et al. (Particle Data Group), Phys. Rev. D 98, 030001 (2019), "Status of Higgs boson physics"

[11] G.Degrassi, S.Di Vita, J. Elias-Miró et al., "Higgs mass and vacuum stability in the Standard Model at NNLO", arXiv:1205.6497v2 [hep-ph]

[12] V. De Luca, A. Kehagias, A. Riotto, "On the Cosmological Stability of the Higgs Instability", arXiv:2205.10240v2 [hep-ph]

[13] The CMS collaboration, "Measurements of Higgs boson production cross sections and couplings in the diphoton decay channel at $\sqrt{s} = 13$ TeV" (CMS-HIG-19-015), arXiv:2103.06956v2 [hep-ex]

[14] The CMS collaboration, "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC" (CMS-HIG-12-028), arXiv:1207.7235v2 [hep-ex]

[15] The ATLAS collaboration "Observation of a New Particle in the Search for the Standard Model Higgs Boson with the ATLAS Detector at the LHC", arXiv:1207.7214v2 [hep-ex]

[16] The CMS collaboration "Measurement of the Higgs boson width and evidence of its off-shell contributions to ZZ production" (CMS-HIG-21-013), Nature Physics volume 18, pages 1329–1334 (2022)

[17] M.Bonvini, S.Forte, G.Ridolfi, "The threshold region for Higgs production in gluon fusion", arXiv:1204.5473 [hep-ph]

[18] V.Cavaliere, "Associated production of Higgs VH, HH and other states with Higgs", presentation at Brookhaven National Laboratory, https://cds.cern.ch/record/2780345/files/ATL-PHYS-SLIDE-2021-496.pdf

[19] A. Buckley, X. Chen et al., "A comparative study of Higgs boson production from vector-boson fusion", arXiv:2105.11399v2 [hep-ph]

[20] LHC Higgs Working Group, CERN twiki web page, https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHWG

[21] A. M. Sirunyan et al. (CMS collaboration), "Constraints on anomalous Higgs boson couplings to vector bosons and fermions in its production and decay using the four-lepton final state", (HIG-019-009), Phys. Rev. D 104, 052004 (2021)

[22] S. Actis, G. Passarino, C. Sturm, S. Uccirati, "NNLO Computational Techniques: the Cases $H \to \gamma\gamma$ and $H \to gg$", arXiv:0809.3667v1 [hep-ph]

[23] M. M. Muhlleitner, "Determination of Higgs Spin and Parity at the LHC", Recontre de Moriond workshop (2013)

[24] V. Khachatryan et al., "Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV", Physical Review D 92, 012004 (2015)

[25] The CMS collaboration, "A portrait of the Higgs boson by the CMS experiment ten years after the discovery" (CMS-HIG-22-001), arXiv:2207.00043v3 [hep-ex]

[26] R.Penco "An Introduction to Effective Field Theories", arXiv:2006.16285v1 [hep-th]

[27] D. de Florian, C.Grojean, F.Maltoni et al., "Handbook of LHC Higgs cross sections:4. Deciphering the nature of the Higgs sector" (Report of the LHC Higgs Cross SectionWorking Group, CERN–2017–002-M), arXiv:1610.07922v2 [hep-ph]

[28] P.Artoisenet, P.de Aquino, F.Demartin et al., "A framework for Higgs characterisation" (CERN-PH-TH/2013-148), arXiv:1306.6464v3 [hep-ph]

[29] V.Gritsan et al., "New features in the JHU generator framework: Constraining Higgs boson properties from on-shell and off-shell production", arXiv:2002.09888v4 [hep-ph]

[30] F..F Freitas, C.K.Khosa, V.Sanz, "Exploring the standard model EFT in VH production with machine learning", Physical Review D 100, 035040 (2019)

[31] E.Massò, V.Sanz, "Limits on Anomalous Couplings of the Higgs to Electroweak Gauge Bosons from LEP and LHC", arXiv:1211.1320v2 [hep-ph]

[32] A.V.Gritsan, J.Roskes,U.Sarica et al., "New features in the JHU generator framework: Constraining Higgs boson properties from on-shell and off-shell production", arXiv:2002.09888v4 [hep-ph]

[33] CERN web page:
https://home.cern/science/accelerators/accelerator-complex

[34] CMS luminosity -public results, Twiki CERN web page:
https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResultsOverview

[35] The CMS collaboration, "The CMS trigger system" (CMS-TRG-12-001), arXiv:1609.02366v2 [physics.ins-det]

[36] V. Veszpremi (on behalf of the CMS Collaboration), "Operation and performance of the CMS tracker", arXiv:1402.0675 [physics.ins-det]

[37] The CMS collaboration, "Electron and photon reconstruction and identification with the CMS experiment at the CERN LHC" (CMS-EGM-17-001), arXiv:2012.06888v2 [hep-ex]

[38] The CMS collaboration, "Particle-flow reconstruction and global event description with the CMS detector", arXiv:1706.04965

[39] The CMS collaboration, "A measurement of the Higgs boson mass in the diphoton decay channel" (CMS-HIG-19-004), arXiv:2002.06398v2 [hep-ex]

[40] M. Cacciari, G. P. Salam, and G. Soyez, “The anti-kT jet clustering algorithm”, arXiv:0802.1189

[41] V. Khachatryan et al., " Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV", JINST 12 P02014 (2017)

[42] D. A. Roberts, S.Yaida, "The Principles of Deep Learning Theory", arXiv:2106.10165v2 [cs.LG]

[43] Y. Coadou, "Boosted decision trees", arXiv:2206.09645v1 [physics.data-an]

[44] SciKit-learn web page: https://pypi.org/project/scikit-learn/

[45] Xgoost documentation web page
https://xgboost.readthedocs.io/en/stable/python/index.html