SCUOLA NORMALE SUPERIORE

CLASSE DI SCIENZE
Corso di Perfezionamento in Fisica

Tesi di Perfezionamento

Search for the Higgs boson decaying into two muons with the CMS experiment

Candidato:
Giulio Mandorli

Relatori:
Prof. Luigi Rolandi
Prof. Andrea Rizzi
Dr. Paolo Azzurri

Anno Accademico 2019/2020

Introduction

The Higgs boson was discovered at the CERN LHC by both the ATLAS and CMS collaborations with a mass near 125 GeV in 2012. The measurement of the Higgs boson properties has been one of the main targets of LHC experiments since the discovery.

In order to probe the mechanism giving mass to fermions and bosons, the Higgs boson couplings have to be measured. The Standard Model Higgs couplings can be extracted by measuring Higgs boson production and decay rates, Therefore a lot of effort has been done to observe every single experimentally accessible decay in various production modes in the last few years. The bosonic decay modes of the Higgs in $\gamma\gamma$, ZZ, and WW have been established using data collected until 2012 (LHC Run 1) while the data collected in 2016 and 2017 allowed to measure the decays to the third generation fermions. The decay of the Higgs boson into $\tau\tau$ has been observed using 2016 data and it confirmed the coupling of the Higgs with leptons. The decay of the Higgs boson into $\mathrm{b\bar{b}}$ has been observed in 2018, hence establishing the coupling with down-type quark. The $\mathrm{t\bar{t}H}$ production mode has also been observed, thus probing directly the coupling to up-type quarks.

So far, the Higgs boson couplings have been observed only with the third generation of fermions, thus, the next step would be the search for decays into the second generation. Among those, the Higgs decay into c-quark pairs is the one with the higher event rate. Probing the Standard Model expected event rate is notwithstanding impossible at the LHC because of the large amount of irreducible background and because of the low mass resolution in the dijet final state. Instead, the Higgs boson decay into muons has smaller rates but it also has a clean final state that can be reconstructed with high efficiency and excellent mass resolution, thanks to the CMS tracking and muon systems. However, since the branching fraction $\mathcal{B}(\mathrm{H}\rightarrow \mu^{+}\mu^{-})$ is very small $(2 \cdot 10^{-4})$ and the LHC produces abundant backgrounds, an observation of this decay has not been reached yet and upper limits only have been sets.

My work took part in the most recent search for the Higgs boson decaying to two muons performed by the CMS collaboration. The analysis uses the full proton-proton collision data collected by the CMS experiment at $\sqrt{s}$ = 13 TeV during Run 2, corresponding to an integrated luminosity of 137.4 fb^{-1}. The search of the Higgs boson to muon decay targets each of the four main Higgs production channels at the LHC. The four searches use orthogonal data and their results are combined in a single measurement.

My work has been dedicated to the improvement of the search in the vector boson fusion (VBF) production mode. In this work, we demonstrate that the VBF channel is the most sensitive among the four $\mathrm{H}\rightarrow \mu\mu$ searches. The VBF dedicated analysis is based on the fit of a multivariate discriminator, in contrast with the other three dedicated Higgs searches where the fit is performed on the dimuon mass distribution. Moreover, the background estimation is completely based on Monte-Carlo simulations. These two aspects make the strategy to be very similar to the one adopted by two important CMS measurements: the measurement of electroweak production of two jets in association with a Z boson, and the measurement of the Higgs boson decaying into b-quark pairs.

Machine learning is a fundamental tool to improve the sensitivity of this search. Two different machine learning techniques are tested for the H$\rightarrow$ $\mu\mu$ search: boosted decision tree (BDT) and deep neural network (DNN). The performance of the two algorithms are compared and the results are extracted from the algorithm that separates better the signal from the background on simulated data.

The Higgs boson signal strength is extracted among several regions with a simultaneous maximum likelihood fit and the results are computed for a Standard Model Higgs boson with a mass of 125.38 GeV. In the VBF channel, an excess of events is observed with a strength of $1.36^{+0.69}_{-0.61}$, defined as the ratio between the expected and the observed event rate σ_{obs}(pp$\rightarrow$H$\rightarrow$ $\mu\mu$)/σ_{SM}(pp$\rightarrow$H$\rightarrow$ $\mu\mu$). The observed excess of events corresponds to a signal significance of 2.4 standard deviations above the background-only hypothesis. The excess of events observed from the combination of the four dedicated searches corresponds to a signal significance of 3.0 standard deviations and it constitutes the evidence of the Higgs boson decay into muons. The results obtained with the data collected at $\sqrt{s}$ = 13 TeV are combined with the ones obtained during Run 1 at the center of mass energies of $\sqrt{s}$ = 7 and 8 TeV. The final combination improves the results of about 1%, resulting in a signal strenght of $1.19^{+0.40}_{-0.30}$(stat)$^{+0.15}_{-0.14}$(syst).

The thesis is structured as follows. The Standard Model and the Higgs mechanism are introduced in chapter 1. Moreover, the most important measurements of the Higgs boson properties are summarized in that chapter. In chapter 2 the experimental apparatus is described. Firstly, the LHC machine is introduced, then the CMS experiment and the various subdetectors are presented. The algorithms used by CMS to reconstruct physics objects are introduced in chapter 3. In particular, lepton reconstruction and identification are described in the first half of the chapter, while the second half is dedicated to jets. Chapter 4 introduces the search for H$\rightarrow$ $\mu\mu$ with CMS Run 2 data. An overview of the entire analysis is given and the dedicated searches in the ggH, VH, and ttH channels are summarized. The VBF analysis that I have carried out is described in detail in chapter 5. Finally, chapter 6 presents the results of the VBF dedicated search and the results obtained from the combination with the other three dedicated searches and with the data collected during Run 1.

Contents

Chapter 1

The Standard Model and the Higgs Boson

A brief overview of the Standard Model and of the spontaneous symmetry breaking is given in this chapter. The theoretical introduction aims to build the frame in which the analysis is set and to motivate the search performed in this work. Then, an overview of the state of the art research on the Higgs boson properties and couplings is given.

1.1 The Standard Model

The Standard Model (SM) is a renormalizable non-abelian gauge theory that describes all known fundamental particles and their interactions. It has been built in the latter half of the last century and, since then, a large number of experiments confirmed its predictions with great precision. Three of the four fundamental interactions are described by the Standard Model: the electromagnetic, the weak, and the strong interaction. The gravitational interaction is not described by the model, however, gravity effects at the electroweak scale are much weaker than the other interactions and their contribution is negligible in the processes currently investigated in high energy physics.

1.1.1 The fermions

In quantum field theory, particles are described by fields and they are divided according to their spin: particles with integer spin are called bosons, and particles with half-integer spin are called fermions. In the Standard Model, all fundamental fermions have spin 1/2 and they are divided into two groups: quarks and leptons. The quarks interact with the strong and the electroweak forces, while leptons interact only with the electroweak force.

Quarks and leptons are further divided as a function of their electrical charge. Up-type quarks have charge +2/3 while down-type quarks have charge -1/3. Leptons are divided in neutrinos, which are electrically neutral, and leptons with charge -1. Matter fermions are classified in three generations or families. Each generation is composed of one charged lepton, one neutrino, one up-type quark, and one down-type quark. Since members of different generations have identical quantum numbers, the generations are three copies of the same structure. The only difference between particles of different generation is the mass: first generation fermions are the lighter and third generation fermions are the heaviest. The exception are the neutrinos, whose mass hierarchy has not been measured.

Each fermion has a corresponding antiparticle that has the same mass and opposite quantum numbers. Moreover, fermion fields of quarks and charged leptons can be divided in a left-handed component and a right-handed one $\psi = \psi_L + \psi_R$[1], the two components interacts differently with the electroweak force.

1.1.2 The Higgs boson

The only scalar boson in the Standard Model is the Higgs boson. The Higgs boson is a massive electrically neutral boson, it interacts via electroweak force, and it is predicted to interact with all the Standard Model massive particles. Figure 1.1 shows all the Standard Model particle with their quantum numbers.

1.1.3 The interactions

In gauge theories, the lagrangian is invariant under certain groups of transformations called gauge groups. The action of the groups on the fields fixes the quantum number and the interaction properties of each particle. Because of the Lorentz invariance, those transformations must be local and an additional field has to be added in the theory for each generator of the group. The new fields transform with the adjoint representation of the Lie group and they correspond to spin-1 particles that are the mediators of the forces.

The three forces described by the Standard Model are carried by specific particles called gauge boson. The electromagnetic interaction is carried by photons (γ). The weak interaction is carried by W^+ and W^- that are electrically charged, and by Z that is neutral. The strong interaction is carried by 8 gluons (g). Photons and gluons are massless mediators, while the weak bosons have a mass of $m_W \simeq 80$ GeV and $m_Z \simeq 91$ GeV. The gauge bosons are shown in Figure 1.1 together with the matter particles.

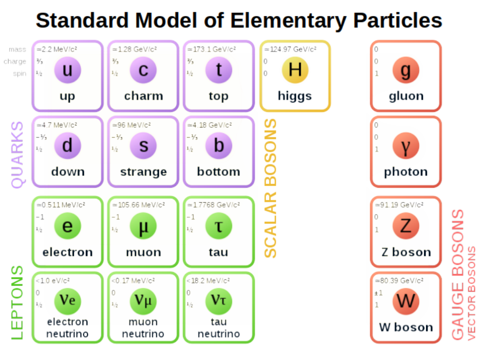


FIGURE 1.1: Table of Standard Model particles [2].

The symmetry group that describe the Standard Model interactions is

$$SU(3)_c \otimes SU(2)_L \otimes U(1)_Y \tag{1.1}$$

The dynamic that derives from the $SU(3)_c$ symmetry is called quantum chromodynamic (QCD). The QCD defines the strong interaction and the corresponding force

[1] The experiment of Goldhaber [1] shown that there are no light right-handed neutrinos, therefore they are not included in the Standard Model description

mediators are called gluons. The quarks transform as a triplet under the action of the $SU(3)_c$ group, therefore three equivalent copies of each quark exist. Each copy is characterized by a different state called color charge. This peculiar interaction has some specific characteristic that forces the quarks to bond in colorless states. These bound states are composed of two or three quarks and they are called hadrons. Free quarks and gluons can not exist in nature. Leptons and the other bosons do not interact via strong force and they do not have any color charge.

The electroweak interactions [3, 4, 5] are described by the action of $SU(2)_L \otimes U(1)_Y$ on the matter fields. The four generators correspond to the boson fields W^i_μ and B_μ. The action of both $SU(2)_L$ and $U(1)_Y$ are different on the two chiral components of each fermion. The right-handed chirality components are singlet under $SU(2)_L$ while left-handed ones are arranged in doublets and are mixed by the $SU(2)_L$ transformations. Specifically, the left-handed components of up-type and down-type quarks of the same generation form a doublet $q_L = \begin{pmatrix} u_L \\ d_L \end{pmatrix}$ as well as the left-handed components of the leptons $\ell_L = \begin{pmatrix} e_L \\ \nu_L \end{pmatrix}$. Since $U(1)$ is an abelian group, its representations are identified by a constant called hypercharge. The representations of the three gauge groups are summarized in table 1.1.

	q_L	u_R	d_R	ℓ_L	e_R
$SU(3)_c$	triplet	triplet	triplet	singlet	singlet
$SU(2)_L$	doublet	singlet	singlet	doublet	singlet
$U(1)_Y$	1/6	2/3	-1/3	-1/2	-1

Table 1.1: Representations of the gauge groups on fermions.

In order to have local invariance under the gauge transformations, the gauge fields have to be added in the theory. Moreover, all the space-time derivatives ∂_μ must be promoted to covariat derivative

$$D_\mu = \partial_\mu + ig_S t^a G^a_\mu(x) + igT^i W^i_\mu(x) + ig' YB_\mu(x) \tag{1.2}$$

where g_S, g, and g' are the coupling constant of the strong, weak, and hypercharge fields, $G^a_\mu(x)$ are the fields of the gluons, and t^a, T^i, and Y are the generators of the Lie group in the rappresentation required by the field. Such a substitution generates coupling terms between the matter particles and the gauge bosons.

1.1.4 Gauge boson kinetic term

The local invariance requirement introduces the gauge bosons in the theory. The kinetic term for the gauge fields can be written as

$$\mathcal{L}_g = -\frac{1}{4} G^a_{\mu\nu} G^{a\,\mu\nu} - \frac{1}{4} W^i_{\mu\nu} W^{i\,\mu\nu} - \frac{1}{4} B_{\mu\nu} B^{\mu\nu} \tag{1.3}$$

where $G^a_{\mu\nu}$, $W^i_{\mu\nu}$, and $B_{\mu\nu}$ are called field strength tensors and they are defined to be the commutator of the covariant derivative D_μ associated to the corresponding field. For any generic field A^a_μ with coupling g, the field strength tensor $F^a_{\mu\nu}$ can be expanded as

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + igf^{abc} A^b_\mu A^c_\nu \tag{1.4}$$

The constants f^{abc} are called structure coefficients, they are totally antisymmetric and they are related to the algebra of the generators. The hypercharge group is abelian, therefore its structure coefficient is zero and its signal strength tensor does not contain the third term of equation 1.4. Instead, $SU(3)$ and $SU(2)$ have several non zero structure coefficients, and the third term generates both third and quartic vertices between strong and weak boson fields.

A mass term for the gauge boson is not allowed because it would not be gauge invariant. In the Standard Model, the weak boson masses, as well as the masses of the fermions, arise from the Higgs mechanism.

1.2 The Higgs sector

Spontaneous symmetry breaking is a phenomenon in which a physical system is described by a Lagrangian that is invariant under certain transformations, but the lowest energy state does not exhibit the same symmetry. The actual implementation of the symmetry breaking in the Standard Model is realized through a complex scalar field whose states are not invariant under $SU(2)_L \otimes U(1)_Y$. This process is called the Higgs mechanism [6, 7, 8, 9, 10, 11].

A new field $H(x)$ that is a doublet under $SU(2)$ and with hypercharge 1/2 is introduced in the theory. A potential of the form

$$V(H) = -\mu^2 H^\dagger H + \lambda (H^\dagger H)^2 \tag{1.5}$$

and the kinetic term associated to the $H(x)$ field

$$K(H) = D_\mu H^\dagger D^\mu H \tag{1.6}$$

are added in the lagrangian. The potential produces a new ground state therefore the lagrangian has to be expanded around the new minimum

$$H(x) = \frac{1}{\sqrt{2}} e^{-i\theta^a(x)\sigma^a/2} \begin{pmatrix} 0 \\ v + h(x) \end{pmatrix} \tag{1.7}$$

where $v = \sqrt{\mu^2/\lambda}$ is called vacum expectation value. The covariant derivative $D_\mu H(x)$ contains also the gauge boson fields as seen in equation 1.2, therefore such an expansion produces in the lagrangian the following quadratic terms

$$\frac{1}{8} g^2 \left(W^{(1)2}_\mu + W^{(2)2}_\mu \right) (v + h(x))^2 + \frac{1}{8} \left(g W^{(3)}_\mu - g' B_\mu \right)^2 (v + h(x))^2 \tag{1.8}$$

The mass terms for the gauge bosons arise from this procedure. The first term of the equation contains the fields $W^{(1)}_\mu$ and $W^{(2)}_\mu$ which are combined in W^-_μ and W^+_μ

and produce the mass of the W boson $m_W = g^2 v^2/4$. The fields $W_\mu^{(3)}$ and B_μ are rotated to obtain the fields of the Z boson and the photon. The last quadratic term in equation 1.8 provides a mass for the Z boson $m_W = (g^2 + g'^2)v^2/4$ while the photon remains massless. The massless photon corresponds to an unbroken generator that is by definition the electrical charge.

The minimum produced by the Higgs potential is not invariant under the initial symmetry, however, it is invariant under the transformation produced by Q. Therefore, the pattern associated with this symmetry breaking is $SU(2)_L \otimes U(1)_Y \rightarrow U(1)_{em}$. The entire procedure does not affect the color symmetry $SU(3)_c$.

This mechanism introduces in the lagrangian a scalar field $h(x)$ that derived directly from $H(x)$ and it is the field associated with the Higgs boson. The Higgs mass can be obtained from the potential $V(H)$ and it is a free parameter in the Standard Model.

1.2.1 Couplings with gauge bosons

The equation 1.8 can be rearranged in the following

$$\left(m_W W_\mu^+ W^{\mu -} + \frac{1}{2} m_Z Z^2 \right) \left(1 + 2\frac{h(x)}{v} + \frac{h(x)^2}{v^2} \right) \tag{1.9}$$

The left factor is the mass term for the two weak boson fields while the right factor contains two terms in $h(x)$ that form the couplings between the Higgs and the gauge bosons. Therefore, the Higgs mechanism predicts the existence of triple and quartic vertices of the Higgs and the weak bosons fields. Moreover, it predicts also the values of the corresponding coupling constant:

$$\begin{aligned} g_{HVV} &= \frac{2m_V^2}{v} \\ g_{HHVV} &= \frac{m_V^2}{v^2} \end{aligned} \tag{1.10}$$

In particular, the triple couplings are very important for the Higgs boson studies because they contribute significantly to the Higgs production at hadron colliders.

1.2.2 Couplings with fermions

Mass terms for fermions $m(\bar{\psi}_L \psi_R + \bar{\psi}_R \psi_L)$ are not allowed in the lagrangian because they are not invariant under $SU(2)$ transformations. However, such terms must be present because fermions are observed to be massive. Such an issue is repaired by exploiting the non-zero vacuum expectation of the field $H(x)$. In fact, the coupling between fermions and $H(x)$ is gauge invariant and can be used to construct mass terms for the fermions [12, 13]

$$y_\phi\,(\bar{\psi}_L^{(u)}, \bar{\psi}_L^{(d)}) H\, \psi_R^{(d)} + h.c. = \frac{y_\phi v}{\sqrt{2}}\, \bar{\psi}_L^{(d)} \psi_R^{(d)} + \frac{y_\phi}{\sqrt{2}}\, h \bar{\psi}_L^{(d)} \psi_R^{(d)} + h.c. \tag{1.11}$$

The interaction term is called Yukawa coupling. Equation 1.11 produces the mass term for down-type quarks and for charged leptons. For up-type quarks, the coniugate field $\tilde{H}(x) = i\sigma^2 H(x)$ is used:

$$y_\phi\,(\bar{\psi}_L^{(u)}, \bar{\psi}_L^{(d)})\widetilde{H}\,\psi_R^{(u)} + h.c. = \frac{y_\phi v}{\sqrt{2}}\,\bar{\psi}_L^{(u)}\psi_R^{(u)} + \frac{y_\phi}{\sqrt{2}}\,h\bar{\psi}_L^{(u)}\psi_R^{(u)} + h.c. \tag{1.12}$$

The difference between the original mass terms and the Yukawa couplings is the interaction between the fermion and the Higgs field $h(x)$. This interaction comes directly from the requirement of gauge invariance.

Therefore, the Standard Model explains the fermion masses through the Yukawa term and it predicts the existence of a coupling between the Higgs boson and each fermion. The value of the coupling is expected to be proportional to its mass for all fermions:

$$\lambda_f = \frac{m_f}{v} \tag{1.13}$$

1.3 The Higgs boson at the LHC

One of the main driving motivations for the Large Hadron Collider (LHC) construction was to find proves about the electroweak spontaneous symmetry breaking. In 2012, the ATLAS [14] and CMS [15] Collaborations descoverd a new particle with mass around 125 GeV [16, 17, 18]. Subsequently, both the experiments established that the properties of the new particle, including its spin and CP properties, are compatible with those predicted by the SM for the Higgs boson.

1.3.1 Higgs production channels

The Higgs boson at the Large Hadron Collider is produced through four main mechanisms, called modes or channels. The exact values of the cross-sections of the four channel productions depend on the center of mass energy of the proton-proton collisions, however, the relative contribution to the total Higgs production is about constant at the energies explored by the LHC. Figure 1.2 shows the cross-sections of the main Higgs production channels for energies ranging from 7 to 14 TeV. The accuracy of the theoretical calculations of the Higgs cross-section is reported in Figure 1.2 for each process. The uncertainty due to strong interactions is dominant in the calculation. Table 1.2 shows the predicted cross-sections of the main Higgs production at $\sqrt{s} = 13$ TeV collisions for a Higgs boson mass of 125 GeV. The cross-sections of the Higgs production at the Large Hadron Collider are smaller of several order of magnitude than the cross-sections of strong processes. Therefore the Higgs searches are performed in conditions where the signal over background ratio is very small.

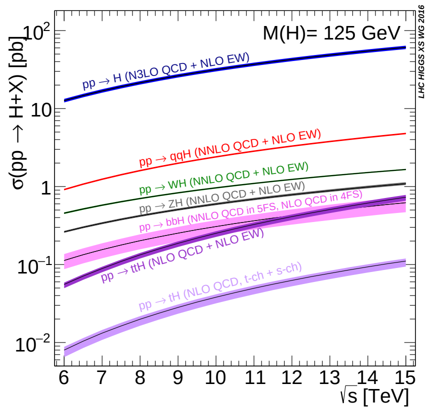


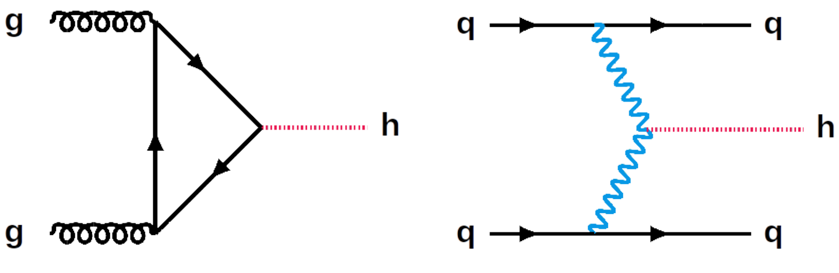
FIGURE 1.2: Cross-sections of the main Higgs production channels as a function of the center of mass energy of the collisions. The solid line indicates the central values while the band shows the uncertainties [19]

Channel	$\sigma(pp \to H)$[pb]
ggH	$48.58\ ^{+2.93}_{-3.77}$
VBF	$3.782\ ^{+0.081}_{-0.080}$
ZH	$0.884\ ^{+0.036}_{-0.031}$
$W^{+}H$	$0.840\ ^{+0.025}_{-0.026}$
$W^{-}H$	$0.533\ ^{+0.011}_{-0.011}$
ttH	$0.507\ ^{+0.035}_{-0.050}$

TABLE 1.2: Cross-sections of the main Higgs production channels at $\sqrt{s}$ = 13 TeV proton-proton collisions for a Higgs boson mass of 125 GeV [20].

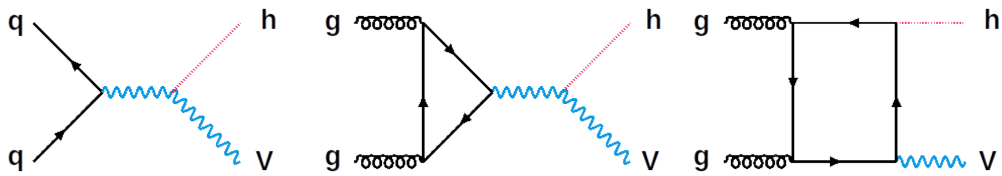
The gluon-gluon fusion (ggH) is the Higgs production mechanism with the largest cross-section at the LHC. In this process, the Higgs boson is produced by the fusion of two gluons through a heavy quark loop. The quark that contributes mostly to this process is the t-quark because of the proportionality between the mass and the coupling with the Higgs. The ggH production has not any characteristic features that can be used to identify these events, therefore it is not ideal for hadronic Higgs decays searches. However, it contributes significantly to Higgs searches because of the high cross-section. The main diagram of the gluon-fusion Higgs production is shown in Figure 1.3.

The vector boson fusion (VBF) is defined by the virtual collision of two gauge bosons, which may be both Z or W bosons. It has a cross-section about ten times smaller than the ggH one and it is the second most frequent Higgs production mechanism at the LHC. In the VBF channel, the initial partons exchange a transverse momentum that is typically lower than the center of mass energy of the quarks, therefore they hadronize in two separated high rapidity jets. Moreover, since the



(a) (b)

FIGURE 1.3: Main Feynman diagrams for the ggH (a) and VBF (b) Higgs productions.



(a) (b) (c)

FIGURE 1.4: Quark initiated (a) and gluon initiated (b) Feynman diagrams for the Higgs-strahlung production.

VBF is a pure electroweak process, low hadron activity is expected in the central part of the detector where the produced Higgs boson is typically directed at. These particular features are extremely important to tag this process and to reduce the huge amount of QCD background that arises from proton-proton collisions. Figure 1.3 shows the main Feynmann diagrams for the VBF production at the LHC.

The Higgs-strahlung (VH) process consist in the emission of a Higgs boson from a gauge boson. The final state consists of a Higgs boson and a gauge boson. The presence of the gauge boson is an important signature that is exploited in Higgs searches. In particular, the presence of charged leptons or neutrinos coming from the gauge boson decay is extremely efficient to reduce the QCD background. The Higgs-strahlung process is divided into $W^{+}H$, $W^{-}H$, and ZH productions that have about the same cross-section. The main diagrams of the three VH processes are quark initiated, however, a small contribution of ZH production arises from gluon initiated process. Figure 1.4 shows the main Feynmann diagrams for the VH production, the quark initiated process contributes to about 90% of the total Higgs-strahlung production at the LHC.

The top pair associate production (ttH) is the production of the Higgs boson together with two t-quarks. This channel allows the direct measurement of the Higgs boson coupling with the t-quark. The signature of this process consists of having two t-quarks that can be identified from the presence of W bosons and b-quarks in the events. Moreover, it is also characterized by a high jet multiplicity. This signature makes the ttH channel to be worth studying even if it has a very small cross-section. The main Feynman diagram of the ttH production is shown in Figure 1.5.

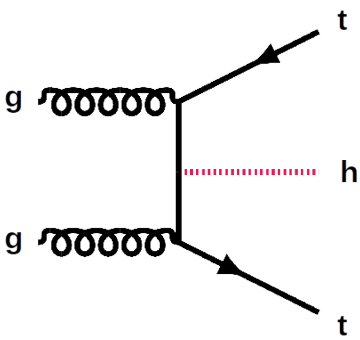


FIGURE 1.5: Main diagram of the top associated Higgs production.

There are several other Higgs productions like bbH and tH and their contribution to the total Higgs production has to be taken into account in Higgs measurement. However, they have not been target of direct searches so far because of their small cross-sections and weak signatures.

1.3.2 Higgs decay channels

The Standard Model predicts the decay amplitude of the Higgs boson and its branching fraction in each possible final state. At 125 GeV, the total decay amplitude is about 4 MeV, corresponding to a lifetime of the order of 10^{-22} seconds. In such a short lifetime, the Higgs boson cannot travel significantly in the detector, therefore it is not directly measured but it has to be identified from its decay products.

The Higgs boson can decay into each quark or charged lepton pair because of the Yukawa interaction, the only exception is H→ $\mathrm{t\bar{t}}$ which is kinematically forbidden. The branching fractions $\mathcal{B}(\mathrm{H} \rightarrow f\bar{f})$ strongly depend on the fermion mass. The dominant decay is H→ $\mathrm{b\bar{b}}$ with a branching fraction of about 58%. Other significant decays are H→ $\tau\tau$ and H→ $\mathrm{c\bar{c}}$. The remaining Higgs decays into fermions are very small because of the fermion masses.

The Higgs boson can decay also in gauge boson pairs. For the cases of W or Z, one of the bosons has to be virtual since $2m_V > m_H$. For the cases of gluon and photon, the final state is reached via a fermionic loop because the Higgs boson does not couple directly with them. In the case of H→ $\gamma\gamma$, a loop of W is possible too. Decays into different neutral bosons, like Zγ, are also possible via a fermionic loop. Table 1.3 shows the predicted branching fractions to fermions and to bosons for a Higgs boson with a mass of 125 GeV.

Decay channel	Branching fraction [%]
H→ $\mathrm{b\bar{b}}$	$58.24\,^{+0.72}_{-0.74}$
H→ $\tau\tau$	$6.27\,^{+0.10}_{-0.10}$
H→ $\mathrm{c\bar{c}}$	$2.89\,^{+0.16}_{-0.06}$
H→ $\mu\mu$	$\left(2.18^{+0.04}_{-0.04}\right) \cdot 10^{-2}$
H→ WW	$21.37\,^{+0.33}_{-0.33}$
H→ gg	$8.19\,^{+0.42}_{-0.42}$
H→ ZZ	$2.62\,^{+0.04}_{-0.04}$
H→ $\gamma\gamma$	$0.227\,^{+0.005}_{-0.005}$
H→ Zγ	$0.153\,^{+0.009}_{-0.009}$

TABLE 1.3: Main predicted branching fractions to fermions and to bosons for a Higgs boson with mass 125 GeV [20].

1.3.3 Previous Higgs measurements

The first observation of the Higgs boson has been reached from both the ATLAS and CMS experiments via the H→ 4ℓ and H→ $\gamma\gamma$ decay channels. Immediately after, both the decays have been observed independently by CMS and ATLAS [21, 22, 23, 24] as well as the H→ WW $\rightarrow 2\ell 2\nu$ decay channel [25, 26]. Moreover, the first estimate of its mass [27], which was the missing parameter in the Standard Model, has been given.

Spin and parity quantum numbers of the Higgs boson have been extracted from the angular distributions of the Higgs decay products in the H$\rightarrow 4\ell$ and H$\rightarrow \gamma\gamma$, and H$\rightarrow$ WW $\rightarrow 2\ell 2\nu$ channels. Data are compatible with the Standard Model $J^P = 0^+$ quantum numbers, while they exclude alternative hypotheses [28, 29, 30].

Several studies have been performed from the CMS and ATLAS collaboration with the data collected in 2011 and 2012, called Run 1. After Run 1, the data collected have been used to perform tests about the overall consistency between the predicted and measured properties of the newly discovered particle[31, 32]. Moreover, the two experiments published two common papers to combine the measurements of the mass and of the couplings [27, 33].

The large significances obtained from the measurements of the decays to vector bosons verify the spontaneous symmetry breaking in the electroweak sector predicted by the Standard Model. These observations close a big chapter in high energy physics since they successfully confirmed the electroweak unification. However, the Run 1 data was not sufficient to probe also the Yukawa couplings with the same precision. In fact, none of the fermionic decays had reached the 5σ threshold in Run 1, leaving the entire fermionic sector still to be confirmed.

The LHC started a new period of data-taking in 2015, called Run 2. The large amount of data collected in Run 2 allowed to observe some decays and production modes that were missing after Run 1. In particular, observing the Higgs to fermion couplings was crucially important to end verifying Higgs sector in the Standard Model.

The 5σ significance of the H$\rightarrow \tau\tau$ was reached by the combination of ATLAS and CMS results obtained with Run 1 data [33]. However, the two collaborations observed independently the decay in Run 2 using the 2016 data [34, 35]. The dedicated studies of the VBF Higgs production channel contributed significantly in obtaining the observation of the $\tau\tau$ decay. Such a result confirmed the coupling of the Higgs with leptons.

The coupling of the Higgs boson with the t-quarks has been measured by observing the ttH production mode. The Higgs to t-quark coupling could have been indirectly extracted from the gluon-fusion Higgs production, instead, the observation of the ttH production constitutes the first direct measurement of that coupling and it certificated the interaction between the Higgs and the up-type quarks. Both ATLAS and CMS obtained the 5σ significance over the background-only expectation during Run 2[36, 37].

The observation of H$\rightarrow$ b$\bar{\text{b}}$ has been reached in 2018 by CMS and ATLAS using 2016 and 2017 data [38, 39]. The most sensitive production channel for this search is VH. These measurements probed the coupling of the Higgs boson with the down-type quarks, thus demonstrated that the equations 1.11, and 1.12 hold for all flavors of the third generation fermion.

The huge amount of data collected in Run 2 have been used to perform also several other important measurements about the Higgs. In particular, its mass has been fitted using H$\rightarrow \gamma\gamma$ and H$\rightarrow ZZ$ decay channels [40, 41]. The combination obtained with the Run 1 data is

$$m_H = 125.38 \pm 0.11(\text{stat}) \pm 0.08(\text{syst}) \quad \text{GeV} \tag{1.14}$$

and it is the most precise measurement of the Higgs boson mass up to date.

The total width of the Higgs mass peak has been measured in a model-dependent way by comparing the on-shell and off-shell production rates in the ZZ decay mode [42]. The comparison between measured and predicted Higgs width is very important to spot physic beyond the Standard Model. In particular, the total width is sensible to possible Higgs decays into dark matter particles. The Higgs width has been measured to be $3.2^{+2.8}_{-2.2}$ MeV, in agreement with the expectation. The observed width is shown with the 68% and 95% confidence level in Figure 1.6.

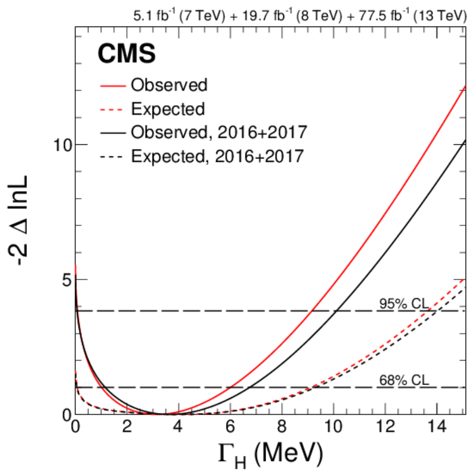


FIGURE 1.6: Observed (solid) and expected (dashed) likelihood scans of the Higgs total width. Results are shown using only the data collected in 2016 and 2017 (black) or from the combination with the Run 1 (red) [42].

1.3.4 Previous H$\rightarrow \mu^+\mu^-$ measurements

The Yukawa couplings of the Higgs boson with the third generation fermions have been observed consistently with the Standard Model predictions. In order to prove that the Higgs mechanism works with all fermions, the Higgs interaction with the first and second generations has to be established. Measuring the Higgs decay to first generation fermions is impossible at the LHC since only a few events are expected to be produced during Run 2. Therefore, particular attention is focused on measuring the Higgs coupling to the second generation.

The coupling of the Higgs with the c-quark is the highest one among the second generation fermions because of the higher mass. However, the Higgs decay to c-quarks is difficult to identify at hadron colliders because of the large amount of QCD background. For the same reason, the Higgs decay to strange quarks is even more difficult.

On the other hand, the Higgs decay into muons has the advantage of better signal-to-background ratio, but the disadvantage of a low branching fraction. Moreover, the dimuon invariant mass can be measured with great precision and it can be exploited in the H→ $\mu\mu$ searches. In addition, the measurement of the dimuon invariant mass is much more precise than the one of two jets. Therefore, the Higgs decay into muons is the best decay channel to probe the coupling with the second generation.

Higgs to muon decay searches have been performed by ATLAS and CMS collaborations since Run 1. The observed upper limits on the production rate were found to be about 7 times the Standard Model prediction [43, 44]. The data collected in Run 1 proved the non flavor-universality of the leptonic couplings with the Higgs, however, they were not sufficient to probe the expected coupling value.

During Run 2, both ATLAS and CMS collected enough data to obtain more significant results. The ATLAS collaboration published a first result obtained with the data collected in 2016 and 2017 [45]. Subsequently, a reanalysis of the entire Run 2 dataset has been performed, resulting in an observed (expected) signal significance of 2.0σ (1.7σ). The obtained upper limit exclusion at 95% con dence level is 2.2 times the Standard Model expectation [46]. The CMS collaboration first published a result using only the 2016 dataset, setting an upper limit of 2.9 the standard model cross-section [47].

Subsequently, a great effort has been devoted to the analysis of the full CMS Run 2 dataset, which is the main topic of this thesis. Several improvements were added to the search, allowing to reach a sensitivity previously not thought possible with only the Run 2 data. Thanks to the excellent detector features of CMS and to the strong analysis effort, the CMS result is significantly better than all the previous results for the channel.

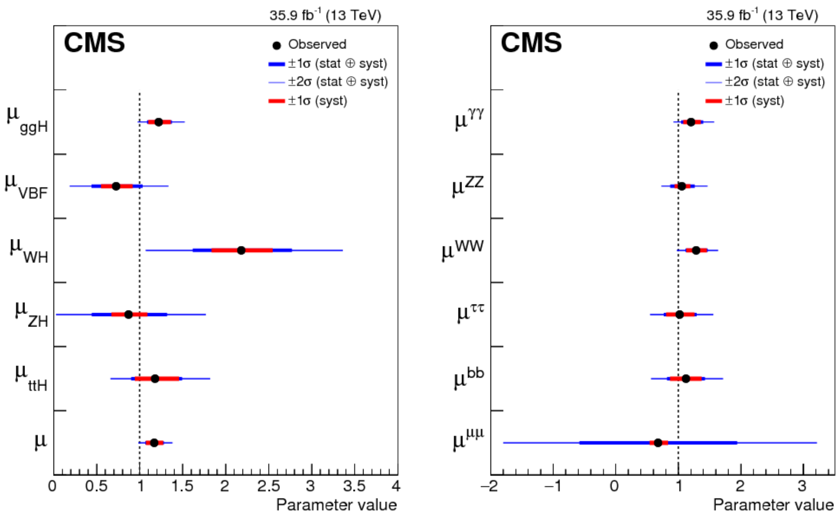
1.3.5 Combined production and decay measurements

The observed rates of the Higgs productions and decays that are measured by CMS and ATLAS collaborations have to be compared to the Standard Model predictions in order to test the agreement between experiment and theory. Such a comparison is described by the signal strength modifier μ, defined as the ratio of the measured Higgs boson rate to its SM prediction. For each production and decay channel the signal strengths for productions μ_i and for decays μ^f are

$$\mu_i = \frac{\sigma(i \to H)}{\sigma_{SM}(i \to H)} \quad \text{and} \quad \mu^f = \frac{\mathcal{B}(H \to f)}{\mathcal{B}_{SM}(H \to f)} \tag{1.15}$$

where i is the Higgs production process and f is the final state. In the Standard Model, all the signal strengths are equal to unity. Production and decay strengths cannot be individually measured without additional assumptions, therefore experimental searches are sensitive to the product $\mu_i^f = \mu_i \cdot \mu^f$.

All the results obtained from the various Higgs searches can be combined with a simultaneous fit in order to better constrain the signal strength parameters. The overall fit can be performed with several assumptions. All the signal strengths μ_i and μ^f can be considered independent and they can be extracted simultaneously in the overall fit. Otherwise, a single overall signal strength can be extracted varying simultaneously all the signal strength modifiers. Figure 1.7 shows the post-fit signal strength modifiers for productions and for decays. The production signal strengths are extracted assuming the Standard Model branching fractions, while the decay signal strengths are extracted assuming the Standard Model cross-sections. The two simultaneous fits have been performed using the results obtained at the beginning of Run 2 [48]. All the obtained results are consistent with the Standard Model expectations.



(a) (b)

FIGURE 1.7: Signal strength modifiers for each production (a) and decay (b) channel. The horizontal bars show the 1σ and 2σ uncertainties while the red bars indicates the 1σ systematic uncertainties. The overall signal strength is shown in the last point in (a). The signal strengths shown in (a) and (b) do not includes the measurement of the Higgs boson decay into muons described in this thesis [48].

The signal yields can also be expressed in terms of the coupling strength modifiers within the κ-framework [49]. A set of coupling modifiers $\vec{k}$ is introduced to parametrize deviations from the Standard Model couplings. For any given production or decay process j, the coupling modifier k_j is

$$k_j^2 = \frac{\sigma_j}{\sigma_j^{SM}} \qquad \text{or} \qquad k_j^2 = \frac{\Gamma_j}{\Gamma_j^{SM}} \tag{1.16}$$

The difference between signal strength and coupling strength parametrizations is that the latter describe single processes either in production and in decay channels. In fact, production and decay anomalies are described by the same set of parameters in the κ-framework, while they are described by different parameters in the signal strength framework.

At tree-level, the coupling modifier can be interpreted as multiplicating factors of the Standard Model couplings:

$$y_j = k_j y_j^{SM} \tag{1.17}$$

therefore they probe directly the consistency of the data with the Standard Model vertices. Additional effective coupling modifiers can be defined for gluons and photons k_g and k_γ, which describe the loop process for ggH production and for $\gamma\gamma$ decay.

Similarly to the signal strength modifiers extraction, a simultaneous fit among all the Higgs results obtained by CMS is performed to estimate the values of the coupling modifiers. Six free couplings parameters are introduced in the fit k_W, k_Z, k_t, k_τ, k_b, k_μ while the effective coupling k_g is expressed in as a function of k_t and k_b. Figure 1.8 shows the Higgs couplings best fit with the data collected at the beginning of Run 2. The overall fit is in agreement with the Standard Model prediction. The fit shown in Figure 1.8 takes into account the Higgs to muon decay search of Ref. [47] while it does not consider the results presented in this thesis. In the Standard Model, the Higgs coupling with the fermions is proportional to the ratio m/v (equation 1.13) while the couplings with the weak bosons are proportional to m^2/v^2 (equation 1.10). For this reason, anomalies in the coupling modifier with fermions k_f and in the square of the coupling modifiers with bosons $\sqrt{k_V}$ are proportional to m/v.

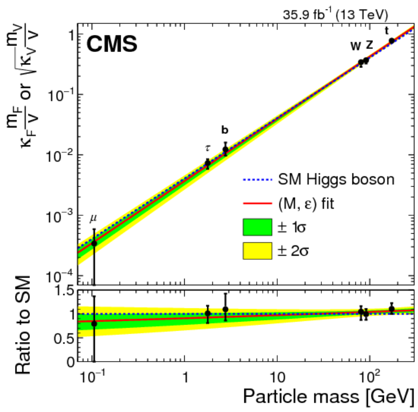


FIGURE 1.8: Coupling modifiers and best fit values as a function of particle mass. The dashed blue line indicates the predicted dependence on the particle mass in the case of the SM Higgs boson. The couplings and the fit shown in this plot do not includes the measurement of the Higgs boson decay into muons described in this thesis [48].

Chapter 2

The CMS experiment at the LHC

The following chapter describes the Large Hadron Collider (LHC) and the CMS experiment whose data have been used in the analysis presented in this thesis. The main CMS subdetectors and their performance will be summarized.

2.1 The Large Hadron Collider

The large Hadron Collider (LHC) is a circular proton-proton accelerator, located at CERN, on the Swiss-French border. It is currently the largest and most powerful particles accelerator ever built. The LHC is built inside the tunnel previously used for the Large Electron-Positron Collider (LEP). The tunnel is about 27 km in length, it is located about 100 m [50] underground and the circle rests on a plane inclined at 1.4% respect to gravity.

The LHC tunnel has not exactly a circular shape but is formed by 8 rectilinear sections and 8 circular portions. Arcs has an internal radius of 3.7 km and contains magnetic dipoles that are used to curve the beams reaching a maximum magnetic field of 8.33 T. The rectilinear sections in the tunnel are approximately 528 m long. They are named from 1 to 8 and they host experiments or utilities. They also contain magnetic quadrupole in order to compensate the transverse dispersion of the protons in the bunches, caused by their mutual repulsion and by synchrotron radiation.

Experimental insertions are located in the rectilinear section. ATLAS (A Toroidal LHC Apparatus) and CMS (Compact Muon Solenoid) experiments are located at Point 1 and 5 respectively. They are both multi-purpose experiments, designed to investigate a wide range of scenarios. Their focus includes Higgs boson properties study and new physics properties searches at TeV scale. They have similar structures and similar subdetectors in order to have comparable results and to have the possibility to cross check each other's studies. ALICE (A Large Ion Collider Experiment) studies a phase of matter called *quark-gluon plasma* that is formed in high energy nuclear collisions. Those studies are performed collecting heavy-ion collisions. LHCb is an experiment designed to perform precision measurement of processes related to b-quark and CP-symmetry violation.

Point 2 and Point 8 contain ALICE and LHCb experiments and the injection systems for Beam 1 and Beam 2. Points 3 and 7 contain two collimation systems. Radio frequency systems are located at point 4 and point 6 contains the beam dump insertion. Points 3 and 7, as well as 4 and 6, do not contain beam cross interaction points. Figure2.1 shows LHC Layout and the content of rectilinear sections.

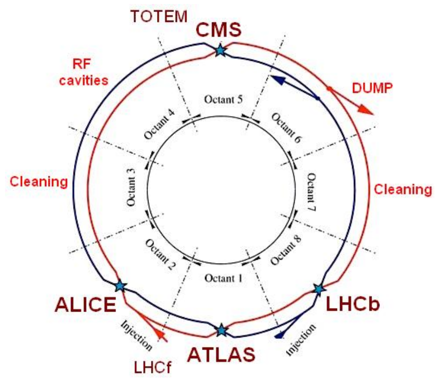


FIGURE 2.1: LHC Layout[50].

Protons are accelerated and injected in the LHC by a chain of four accelerators. The first one is a linear accelerator that injects particles in a chain of three circular accelerators, and finally, they are injected in the LHC with opposite directions in two separate beam pipes with an energy of 450 GeV. The particles are accelerated inside LHC by radio-frequency cavities in Point 4 from 450 GeV to 7 TeV, therefore the center of mass energy of particles collision is 14 TeV. The particles circulate grouped in bunches; each beam pipe contains about 2800 bunches in LHC and each proton bunch 10^{11} protons. The bunches are spaced 25 ns in time in the nominal design therefore collisions happen every 25 ns.

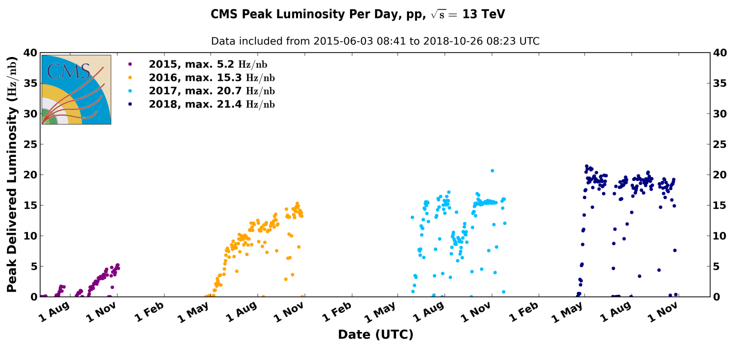


FIGURE 2.2: Peaks of istantaneous luminosity versus day delivered by CMS per day during stable beams and for pp collisions[51].

The collision rate of an accelerator is measured by the luminosity. The luminosity is the ratio of the rate of produced events to the interaction cross section

$$R = \mathcal{L} \cdot \sigma$$

Searching rare events requires a large number of collisions; for this reason, the beams are "compressed" in the transverse plane just before entering in the experiments, in order to increase the luminosity of each bunch crossing. The CMS designed luminosity of 10^{34} cm^{-1}s^{-1} has been achieved and exceeded in the last three years of data taking. Figure2.2 shows the maximum istantaneous luminosity reached by CMS during the Run 2. Integrated luminosity recorded by CMS and usable for physics analyses are 35.9 fb^{-1} for 2016, 41.53 fb^{-1} for 2017, and 60.0 fb^{-1} for 2018.

Rare events searches need high luminosity, but at high luminosity the rate of minimum bias collisions exceeds the bunch crossing rate and there is more than one minimum bias interaction per bunch crossing. This effect is known as "pile-up". Pile-up interactions produce noise and tracks that makes object reconstruction more difficult, therefore a compromise has to be found: the higher the luminosity is, the rarest events can be searched, but simultaneously the higher luminosity is the worse the event reconstruction is. Figure2.3 shows the number of interactions per bunch crossing per year. The distribution is different year by year, average number of interactions is related to luminosity and it is shown in the legend for every year.

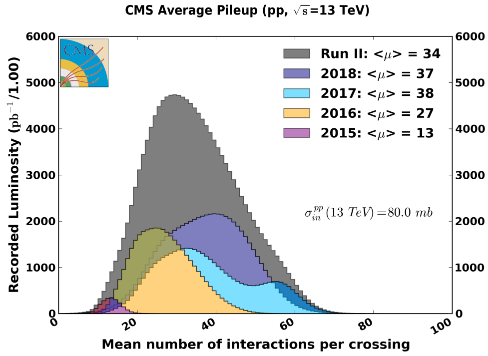


FIGURE 2.3: Distribution of the average number of interactions per bunch crossing (pileup) for pp collisions in Run 2, the overall mean values are shown year by year[51].

2.2 The CMS Experiment

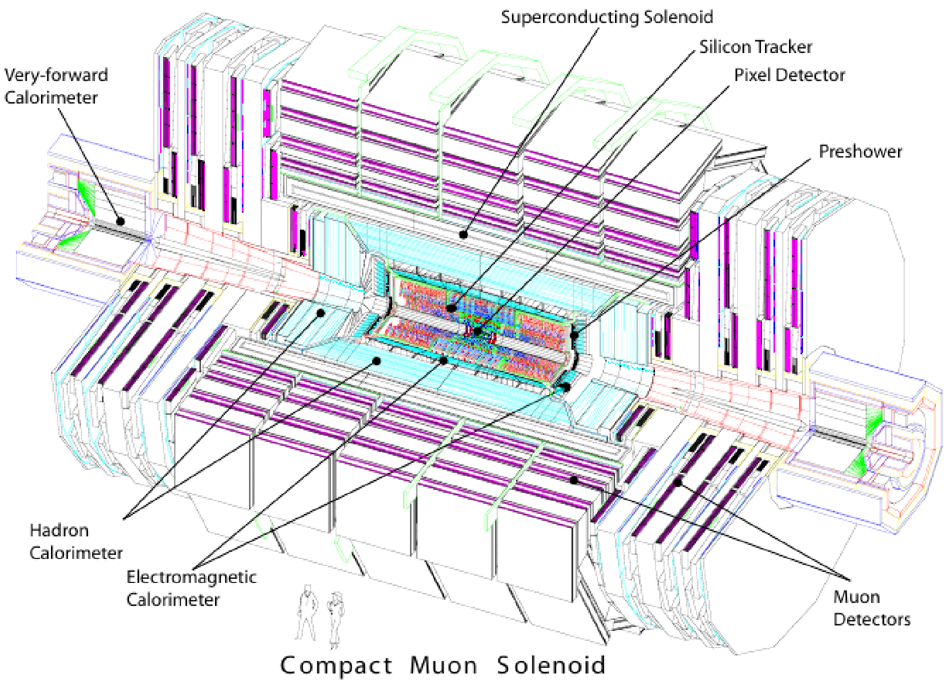


FIGURE 2.4: CMS and all its sub-detectors [52]

The Compact Muon Solenoid (CMS) is a barrel shaped detector, centered at the nominal point where the beams collide. It consists of a central part called barrel and two external parts called endcaps, placed at the ends of the cylindrical barrel.

CMS weighs 12500 t, has a diameter of 14.6 m and a length of 21.4 m, Figure 2.4 shows the layout of the CMS experiment and all its subdetectors.

The heart of CMS is its superconducting solenoid that provides a magnetic field of 3.8 T with the axis aligned along the beams direction. The bore is 13 m long and it has a radius of 3 m. The magnetic flux is returned by an iron yoke composed by 3 wheel in the barrel and three disk in each endcaps. The flux saturates the 1.5 m of iron and ensures enough bending power to measure the muon momentum that crosses the detector.

Starting from the interaction point a particle crosses the silicon tracker, the electromagnetic calorimeter (ECAL) and the hadron calorimeter (HCAL) before reaching the coil. Outside the coil the muon detection system is imbibed in the iron of the yoke of the magnet. The endcaps are shaped in layers perpendicular to the beam line and they have ECAL and HCAL sections and the muon system imbibed in the iron of the yoke like in the barrel.

The right-handed coordinate system of CMS has the center at the collision point and is defined with the $\hat{x}$ axis pointing towards LHC center, the $\hat{y}$ axis pointing upwards and the $\hat{z}$ axis along the beam direction. The azimuthal angle ϕ is defined in the x-y plane, it is measured from the positive direction of the x-axis and it varies in ($\varphi \in [-\pi, \pi)$). The polar angle θ is measured from the positive direction of the z-axis and r is the radial distance defined by $r = \sqrt{x^2 + y^2}$.

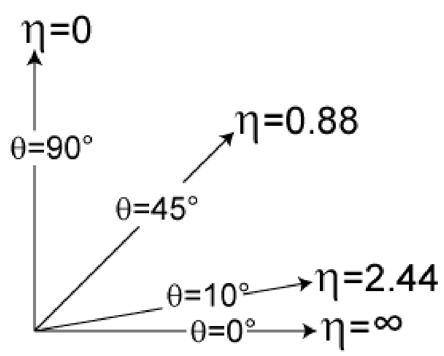


Figure 2.5: Some pseudorapidity values [53].

In hadron-hadron collision the center of mass of the interaction is boosted along the z-axis with respect to the laboratory frame. The rapidity

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

is additive under boost along the z-axis and therefore rapidity differences are constant in those reference frames. Pseudorapidity, defined by

$$\eta \equiv -ln \left[tan \left(\frac{\theta}{2} \right) \right]$$

is an approximation of rapidity. The two match in the relativistic limit and pseudorapidity is used as a coordinate in place of the polar angle.

Figure 2.5 shows some pseudorapidity values with the corresponding polar angle θ.

A brief description of CMS sub-detectors is given in the following sections, a more detailed description can be found in reference [54, 52].

2.2.1 The Silicon Tracker

The silicon tracker [55, 56] is the closest detector to the interaction point and it provides precise and efficient measurements of the trajectories of the charged particles. Proton-proton interactions produce order of 10^3 particle every 25 ns, therefore the tracker has been designed to have high granularity and fast response in order to identify reliably the particles and associate them with the correct bunch crossing. These features have been achieved using as low material budget as possible in order to minimize multiple scattering, nuclear interaction, photon conversion and bremsstrahlung.

Charged particles crossing the detector have bended courses because of the intense magnetic field generated by the superconducting solenoid. The tracker measures the trajectory of each particle and its curvature is used to derive the momentum. Moreover, trajectories are backward extrapolated in order to reconstruct the position of the primary and secondary interaction vertices.

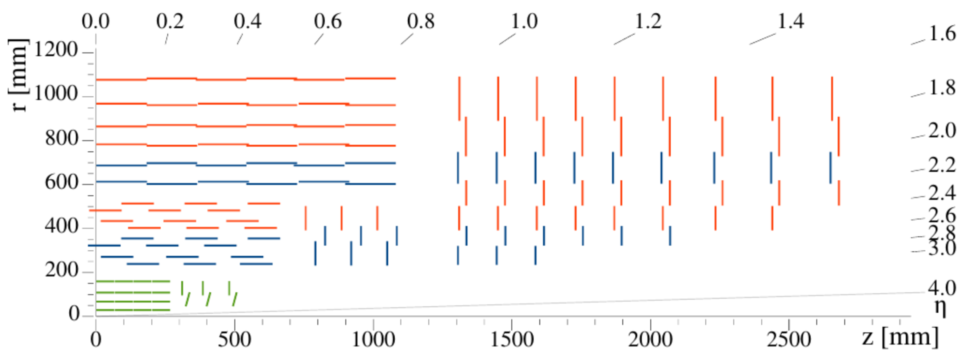


FIGURE 2.6: Longitudinal section of one quarter of the CMS tracker. The pixel detector is shown in green. The single side strip modules are shown in red while the double side strip modules are blue [57].

The silicon tracker is divided in two main sections: the closest to the beamline, made of pixel modules (PIXEL), and the outer part, made of silicon strips. The Silicon Strip detector, in turn, is divided in two: the inner one, called tracker inner barrel (TIB) and tracker inner disks (TID), and the outermost one, calle tracker outer barrel (TOB) and tracker endcaps (TEC).The total length of the tracker is 5.8 m and its diameter 2.5 m, and it can detect particles up to $|\eta| = 2.5$. Figure 2.6 shows an $r - z$ section of the tracker.

The tracker is composed by several modules and the modules are arranged in cylindrical shape and in disks placed at the two endcaps in order to cover as most solid angle as possible. Modules are made of pixel and strip shaped diodes. The charged particles that cross the detector excite the electrons of the semiconductor liberating electron-hole pairs. The charge carriers would normally follow the electric field but, in the barrel, the magnetic fields make the electrons and holes drift along a direction that is tilted with respect to the electric field. The angle between the direction of the electric field and the electron drift is called Lorentz angle. This angle depends on the module position and on the intensity of the electric and magnetic field, thus it is not the same in all the modules of the traker.

In the pixel detector, the Lorentz angle is about 25° (see Figure 2.7) therefore the collected charges are spreaded among more than one pixel. In the modules of the strip detector, the Lorentz angle is very high and this effect is compensated by inclining properly the modules in order to obtain the optimal charge sharing between adjacent strips.. The particle position is measured by interpolating the charge read by adjacent pixels (or strips), reaching a precision much greater than the pixel (strip) width.

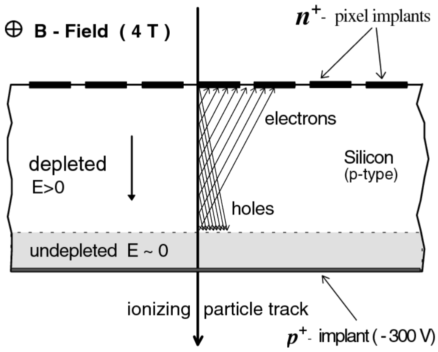


FIGURE 2.7: Drawing of the path of the electrons and holes inside the pixel detector [58].

Pixel detector

The pixel system [59] is the part of the tracker that is closest to the interaction region. The modules have silicon pixels with a surface of $100 \times 150\ \mu m^2$ and 285 μm thick. Their spatial resolution is 15-20 μm in the transverse coordinate ($r\phi$ for the barrel) and 20-40 μm in the longitudinal one.

The pixel detector is exposed to large radiation and it suffers most the radiation damage due to proton-proton collision. CMS has substituted the original pixel detector with a new one during the 2016-2017 shutdown, therefore the data analyzed in this work have been taken with two different pixel detectors. The original one has been used for 2016 data taking, while the new one for 2017 and 2018 data. The pixels that compose the detectors are identical, the differences lie in the electronics and in the layer positions.

The pixel detector is divided in a barrel region (BPIX) and a forward region (FPIX). The original BPIX had three cylindrical layers at radii 4.4, 7.2, and 10.2 cm while the new BPIX has four cylindrical layers at radii 3.0, 6.8, 10.9, and 16.0 cm and 54.9 cm long. Adjacent modules overlap 0.5-1.0 mm in the ϕ direction and there is no overlap in the z-direction.

The original FPIX consisted of four disks, two at each endcaps, located at $z = \pm 34.5$ and $z = \pm 46.5$ from 6 to 15 cm radius. The new FPIX is composed of three disks in each encap and each disk is divided into inner and outer rings. The disks are set at 29, 40, and 52 cm from the interaction point, they radially cover from 4.5 to 16.1 cm, and the modules are rotated of 20° and placed in a turbine-like geometry. The modules of the inner rings are also tilted of 12° respect to the radial direction to enhance the charge sharing above mentioned. New BPIX and FPIX together ensure a four-hit measurement for all the tracks with $|\eta| < 2.5$. Figure 2.6 shows transverse and longitudinal sections of the original and the upgraded detectors.

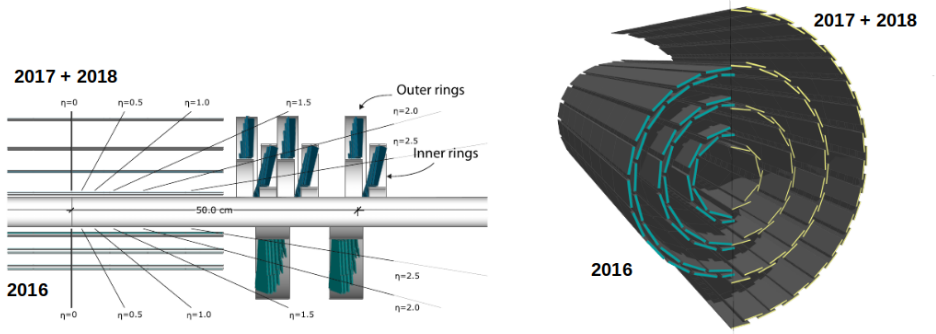


FIGURE 2.8: (left) Transverse sections of the pixel detectors, the region under the beam line shows the original detector while the region above shows the current one. (right) Longitudinal sections of the pixel detectors, the left side of the drawing shows the original detector while the right side shows the current one. [59].

Strip detector

The four subdetectors TIB, TID, TOB and TEC are made of microstrip modules and surround the pixel detector. The strip shaped diodes have different dimensions, their width vary from 80 to 205 μm and their length from 8.5 to 20 cm. The strips length runs along the beam in order to have a precise measurement of $r\phi$ at the expense of the z coordinate.

The strip subdetectors are equipped with both single-side and double-side modules. The latters are made of two modules mounted back-to-back with the strip tilted by an angle of 100 mrad. This setting allows to measure the longitudinal coordinate and improves the measurement of $r\phi$. Moreover, when more than one particle crosses the same module, the solution of the ambiguities in pairing the coordinates is simplified as shown in Figure 2.9. Figure 2.6 shows the single-side and the double-side modules of the strip detectors.

The spatial resolution of the silicon microstrips is very different from one module to the other, depending on the strip dimensions and the location. The resolution of the TIB and TID modules varies from 15 to 40 μm, while in TOB and TEC from 20 to 50 μm. The double-side modules can measure the longitudinal coordinate with a resolution that is larger by a factor of 10.

2.2.2 The Calorimeters

The calorimeters are located outside the tracker and inside the solenoid. They are used to measure the energy of the electromagnetic and hadronic showers and therefore they can detect photons and neutral hadrons. Therefore the tracker and the calorimeters are able to provide an energy measurement of all known long living particles that proton-proton interactions may produce.

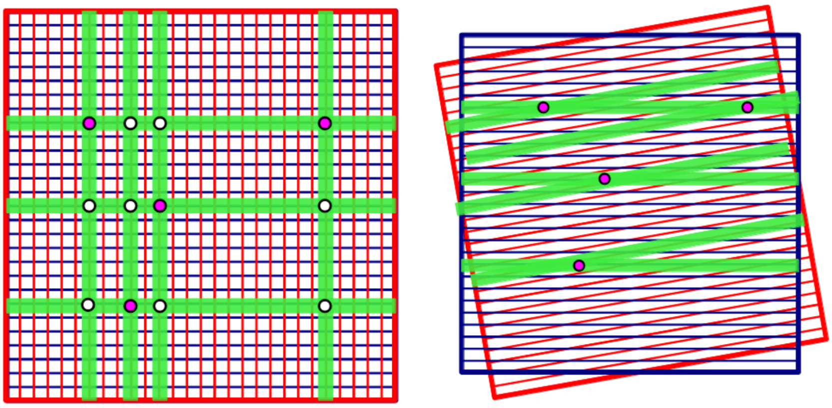


FIGURE 2.9: Two different strips disposition possibilities. On the left two strips are tilted by an angle of 90°, on the right they are tilted by 100 mrad. Magenta filled circles represent the positions in which particles cross the detector. The left disposition produces ambiguous results [60].

The electromagnetic calorimeter

The electromagnetic calorimeter (ECAL) [61] is the calorimeter closer to the interaction point and it is divided into two parts: the barrel (EB) and the endcaps (EE). The barrel covers up to $|\eta| < 1.479$, while the encaps cover the regions $1.479 < |\eta| < 3.0$, moreover a preshower is placed in front of EE covering the regions $1.653 < |\eta| < 2.6$. Figure 2.10 shows the longitudinal section of the Electromagnetic Calorimeter.

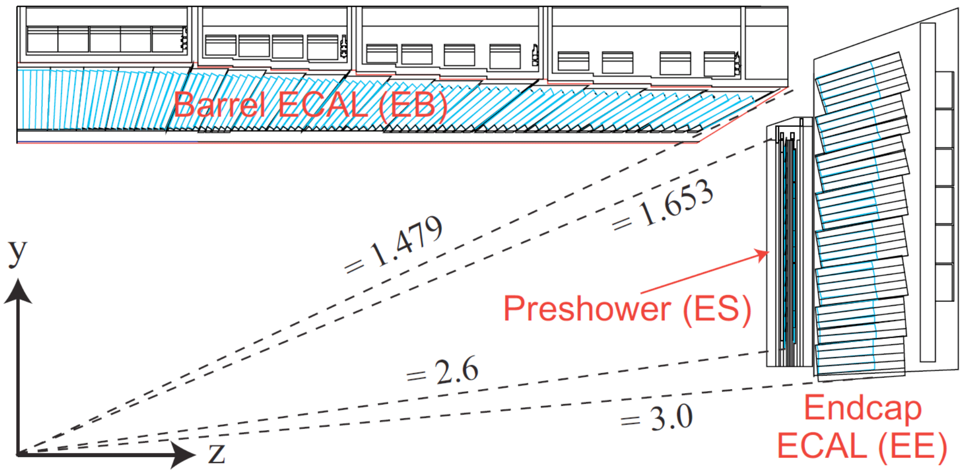


FIGURE 2.10: Section and coverage in η of the Electromagnetic Calorimeter [52]

The ECAL is made of lead tungstate ($PbWO_4$) crystals, 61200 are in the EB while 14684 in the EE. The $PbWO_4$ has been chosen because it has a high density $\rho = 8.28$ g/cm^3, a short radiation length $X_0 = 0.89$ cm, a sall Moliere radius[1] $R_M = 2.2$ cm, and a very fast response. This properties allows the calorimeter to have a compact

[1]Moliere radius is the radius of the cylinder containing on average 90% of the shower energy.

shape, fine granularity and is radiation resistant.
The crystals are active materials, therefore they both induce electromagnetic shower, and generate scintillation light to measure the energy. The scintillation light produced is in the visible range with a peak at 420-430 nm and it is detected by avalanche photodiodes placed at one end of the crystals. Crystals reaction has to be very fast in order to distinguish subsequent bunch crossings, about 80% of the light is emitted in 25 ns.

The crystals have a truncated square pyramid shape with the smaller base positioned towards the interaction points. In order to avoid cracks to be aligned with particle trajectories, the pyramid axis does not pass exactly through the interaction point but it is tilted by a few degrees in both ϕ and η. The barrel crystals are 23 cm long corresponding to 26 X_0 while endcaps crystals are 22 cm which is 25 X_0. The angle coverage is approximately 0.0174×0.0174 in $\eta - \phi$ (about 1°) therefore also the cross sections are different among barrel and endcaps. The bases are 22×22 cm^2 and 26×26 cm^2 in the barrel while 28.6×28.6 cm^2 and 30×30 cm^2 in the encaps. Figure 2.11 shows a picture of a crystal with the photodetector at the rear end.

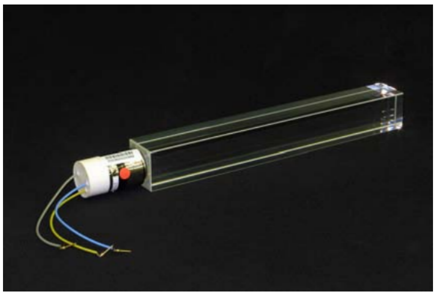


FIGURE 2.11: Endcap crystal with photodetector at the end. The crystal emits and it is transparent to visible light [52].

The preshower is a sampling calorimeter made by two layers. Each layers is composed by a lead radiator that initiate the shower and a silicon strip detector placed just after the radiator. In total the preshower is 20 cm thick and about $3X_0$. The preshower is used to help distinguish high energy photons from π_0 that appear as two very close photons.
The energy resolution of the electromagnetic calorimeter has been measured with electron beams with momenta between 20 and 250 GeV. The resolution can be parametrized as the sum of three source

$$\frac{\sigma_E}{E} = \frac{a}{E} \oplus \frac{b}{\sqrt{E}} \oplus c$$

The first term is the noise and it is due to electronics and pile-up, independently of the energy. The second term is the stochastic term and it models the stochastic processes of the electromagnetic shower and of the electronics readout. The last term is the constant term and it is mainly due to energy scale calibration.
A typical energy resolution is shown in Figure 2.12 and it is

$$\frac{\sigma_E}{E} = \frac{0.28\%}{E} \oplus \frac{12\%}{\sqrt{E}} \oplus 0.3\%$$

with E in GeV.

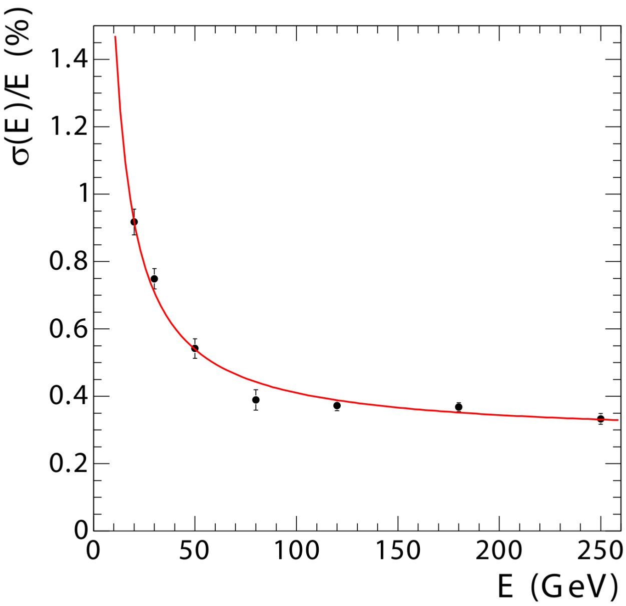


FIGURE 2.12: ECAL energy resolution as a function of the electron energy measured during a test-beam [52].

Hadronic calorimeter

The CMS hadron calorimeter (HCAL) [62, 63] is a sampling calorimeter designed for hadron measurements and it is the only detector available to measure the energy of the neutral hadrons. It provides a wide angular coverage to allow a good measurement of the missing transverse energy (see sec.3.2). The CMS hadron calorimeter has four major sections: the HCAL barrel (HB), the HCAL endcaps (HE), the HCAL outer (HO), and the HCAL forward (HF). Figure 2.13 shows the four main regions of the hadronic calorimeter.

HB and HE are placed in between ECAL and the superconducting solenoid therefore it is essential to use non ferromagnetic materials. The absorber is made of brass because brass has a short interaction length and it is easy to shape. Plastic scintillators are used as active material and the light is converted by wavelength-shift fibers to the silicon photomultipliers. HB covers the region $|\eta| < 1.4$ while HE is in $1.3 < |\eta| < 3.0$. The plastic scintillator tiles have different dimensions in barrel and encaps: the granularity is 0.087×0.087 in HB and it is 0.17×0.17 in HE in the $\eta - \phi$ plane. The thickness of the calorimeter varies from 5.8 λ_I at $\eta = 0$ to 10.6 λ_I at $|\eta| = 1.3$; the electromagnetic calorimeter adds about 1.1 λ_I of material.

The HO is arranged outside the magnetic coil, it covers $|\eta| < 1.26$ and it is made of 10 mm thick scintillator with the HB angular segmentation. It is divided in 5 rings, each of them has one layer of scintillators, except the central ring which has two of them. The HO uses the solenoid as an absorber therefore a dedicated absorber is not added. It has been installed in order to complement the central part of the calorimeter and to ensure that hadronic showers are sampled with nearly 11 hadronic interaction lengths.

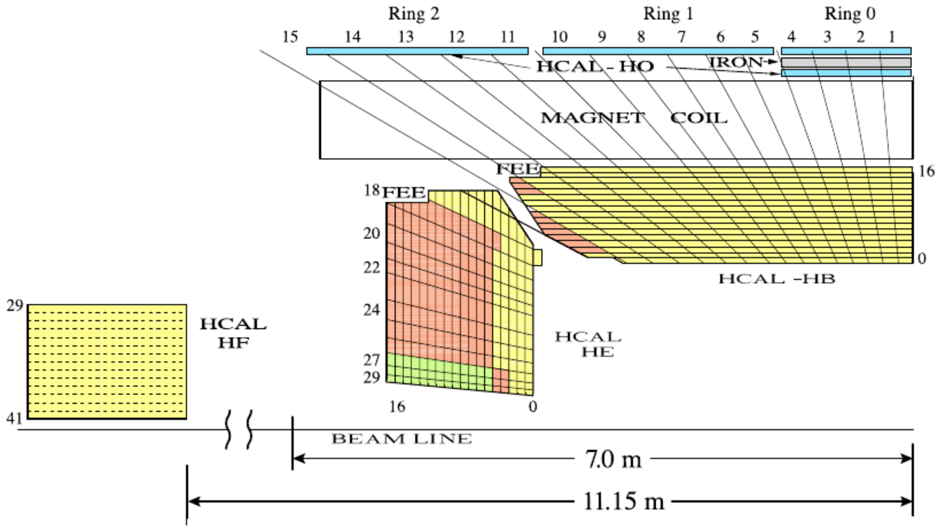


FIGURE 2.13: Longitudinal section of HCAL showing the location of HB, HE, HO and HF [63].

The HF is a Cherenkov calorimeter made of a steel absorber and quartz fibers, it is located at 11.2 m from the interaction region and it covers up to $|\eta| = 5.2$. The fibers are arranged longitudinally through the absorber with a spacing of 5 mm and they are used to collect Cherenkov light generated from the shower. The fibers in the same $\eta \times \phi$ channel bring the light to multi-anode tubes that convert the optical signal into an electrical one.

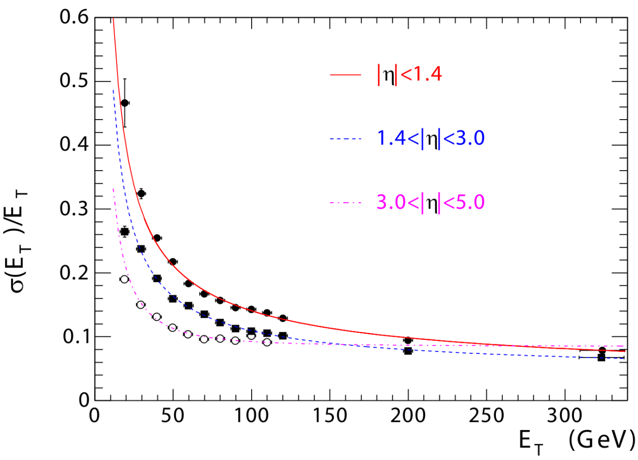


FIGURE 2.14: HCAL transverse energy resolution as a function of the measured jet divided in pseudorapidity sections [62].

The resolution of the calorimeter is parametrized as

$$\frac{\sigma(E)}{E} = \frac{A}{\sqrt{E}(GeV)} \oplus B$$

The resolution measurements have been fitted for different range in pseudorapidity, corresponding to different parts of the HCAL and it is shown in Figure 2.14. The coefficients are about $A \simeq 1$ and $B \simeq 0.1$.

2.2.3 The Muon System

Electrons and photons produced in proton-proton interactions are absorbed by the ECAL. Hadrons are almost entirely absorbed by the HCAL because of the high number of interaction lengths. Muons do not interact strongly and their energy loss due to ionization is about 3-4 GeV when crossing the detectors. Therefore muons with momentum larger than 3-4 GeV can be detected by dedicated chambers mounted in the external part of the detector, those chambers are called muon chambers [64, 65, 66, 67].

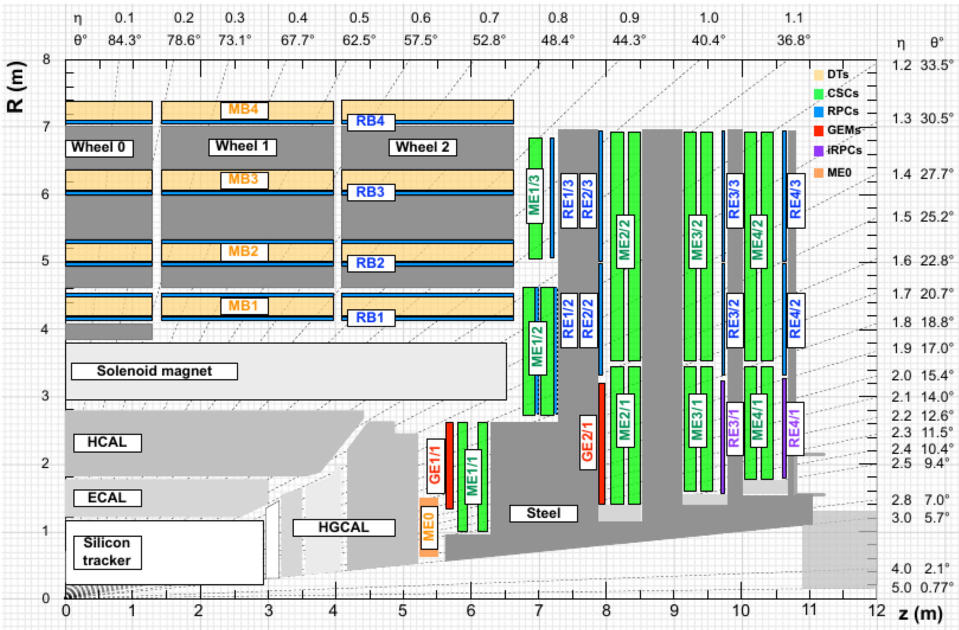
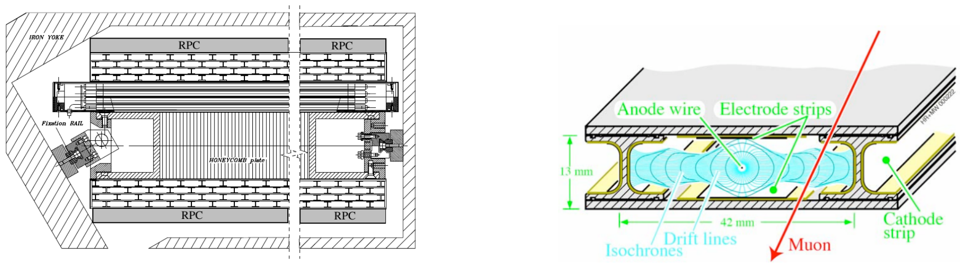


FIGURE 2.15: An r-z cross section of the CMS detector, the interaction point is at the low left corner. The various muon stations are shown in colors while the iron yoke is represented in dark grey. M stands for Muon, B for Barrel and E for Endcap [67].

Four different types of gas detector are used for muons: the drift tube chambers (DT), the cathode strip chambers (CSC), the resistive plate chambers (RPC), and the gas electron multipliers (GEM). The chambers are placed outside the solenoid and they are interleaved with the iron of the return yoke to make full use of the magnetic return flux that is about 1.8 T. This arrangement allows a standalone measurement of the muon momentum which is useful for the trigger and for the off-line matching between the muon chamber track and the tracker one. The presence of electromagnetic showers due to the iron and the unavoidable cracks due to supports and cabling complicates the track reconstruction. Therefore a highly redundant design has been chosen for the muon chambers setting. Figure 2.15 shows the layout of the muon system.

In the barrel, muons are measured by drift tube chambers composed by 2.4 m long, 13 mm high and 42 mm wide drift tubes (DT). The DT are gas detectors made by a mixture of 85% Ar + 15% CO_2 and they are organized in 4 stations. A DT chamber is composed in two or three superlayers (SL), each one split in four layers of drift tubes. The 3 innermost stations have three SLs: two SL have the wires parallel to the beam line in order to measure the ϕ coordinate, one has the wire perpendicular to the beam line in order to measure the z coordinate. The outermost station has only the two SLs that measure ϕ. The SLs are the smallest independent unit of the design. The resolution of a single tube is 250 μm on $r\phi$ coordinate and the resolution of one station is 100 μm. Figure 2.16 shows the structures of one chamber and the sketch of a cell.



(a) (b)

FIGURE 2.16: (a) DT chamber inside the return yoke in the $r - \phi$ plane. The three SLs are visible: the ones close to the RPC measure the ϕ coordinate while the third measure the z coordinate.
(b) Drawing of a cell of the drift chambers [52]

The cathods strip chambers (CSC) are multiwire proportional chambers with trapezoidal shape covering the region $0.9 < |\eta| < 2.4$. Each chamber consists in 6 layers of anode wires interleaved by cathods strips as shown in Figure 2.17. The strips run radially and they have widths from 3.2 mm and 16 mm. They measure the ϕ coordinate by interpolating the induced charges, similarly to what is done for the tracker measurements. The obtained resolution is in the range between 80 μm and 450 μm for one layer and between 75 μm and 150 μm for the chambers. The wires are perpendicular to the strips in order to measure the pseudorapidity of the muons. Their diameter is 50 μm and they are spaced by 2.5 or 3.175 mm. The pseudorapidity resolution is 0.01-0.04.

The resistive plate chambers (RPC) and the gas electron multipliers (GEM) are gas detectors that use an electric field to amplify the ionization. In both detectors the signal is read by silicon strips with a coarse resolution and both RPCs and GEMs measure the crossing time of a muon faster than the 25 ns between two consecutive bunch crossings, therefore they are used as fast triggers.

Figure 2.18 shows the muon resolution as a function of the transverse momentum in barrel and endcaps. The resolution is shown for muon system only, for tracker

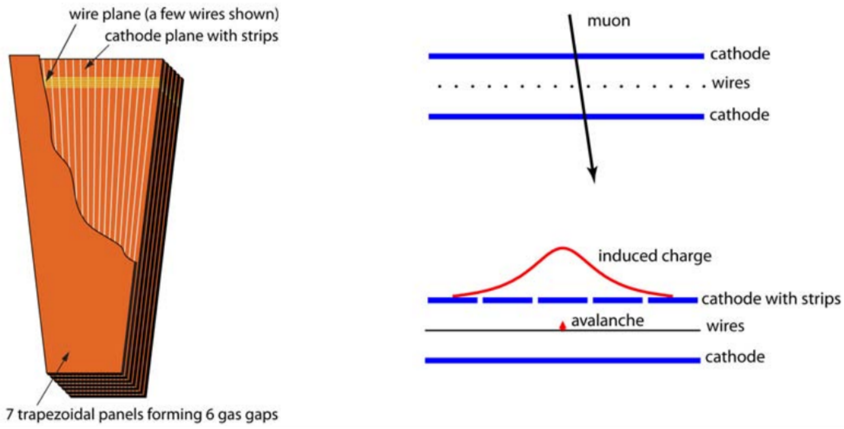


FIGURE 2.17: (left) Layout of a CSC. The cut-out reveals the anode wires and the cathods strips, only few wires are shown. (right) Drawing of a cell of the drift chambers [52].

only, and for the combined measurement. Muon system became relevant for high energy muons.

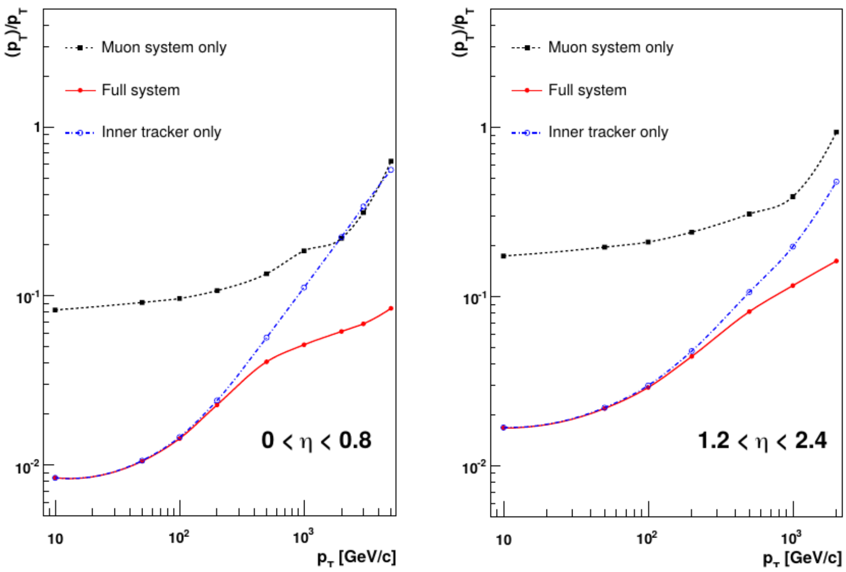


FIGURE 2.18: Momentum resolution for simulated muon tracks measured by the muon chambers only, the tracker only and the full system [52].

2.2.4 The Trigger System

The high frequency collisions produce an amount of data impossible to store by CMS therefore a drastic reduction has to be achieved. A trigger system performs the event selection rejecting the events considered less important for physics analysis. The selection is applied in two steps: the first one is called Level-1 trigger (L1)[68] and it is made by programmable electronics, the second one is called High Level Trigger (HLT)[69] and it is a software program.

Level-1 Trigger

The task of the Level-1 trigger is to perform the first event selection. It reduces the number of events from the bunch crossing rate 40 MHz to less than 100 kHz. During the decision time, the information of the events is recorded in local buffers with 128 memory cells. Since the time between two bunch crossing is 25 ns, a decision has to be taken in 3.2 μs. Taking into account signal propagation delays, L1 calculations have to be made in less than 1 μs. L1 performs the skim mainly based on single object events, if the event is accepted the information are read from every pipeline memory and they will be used by the HLT.

The L1 trigger uses only coarsely segmented data from calorimeters and muon detectors and it is organized in a tree level architecture. At the bottom end there are the calorimeter trigger towers, which measure the energy deposit in the calorimeters, and the muon signal from DT, CSC and RPC, they are also called local triggers. The second level is composed by the global muon trigger and the global calorimeter trigger which combines the infomation from the local triggers. At the top, the L1 global trigger puts together all the information and performs the final decision.

Each muon system has its own trigger logic. The trigger front-end of the Drift Tubes provides a rough measurement of η and ϕ. Muon segments are found separately in anode and cathode electronics in Cathode Strip Chambers. Cathode signal is used to measure the ϕ coordinate while anode signal is used to determine the exact bunch crossing. Either DT and CSC are used to fit straight lines. The muons are assumed to originate from proton-proton collisions and their curvature is computed using the position of the beam interaction point and the line measured by the muon station. This measurement of the muon curvature provides an estimate of the transverse momentum. The multiple scattering strongly affects the p_T measurement resulting in an uncertainty of 10%. Figure 2.18 shows the transverse momentum resolution of the muon measured by the muon system only. The uncertainties at low momentum is dominated by multiple scatering effects. The RPC logic is based on spatial and temporal coincidence of hits in four RPC muon stations. The RPC also helps in resolving spatial and temporal ambiguities in multi muon events.

Tracks measured by DT, CSC and RPC are gathered by the global muon trigger, which tries to correlate the measurements among each other, if two candidates match their parameters they are combined to give optimum precision. It also accesses the calorimeter tower to compute the isolation of the muons and sort the found tracks by p_T and quality. The top 4 muons are sent to the global trigger. The global trigger collects all the objects sent by the regional triggers. Every object contains its coordinate (η, ϕ) so that different thresholds can be applied as a function of the location. It also handles triggers requiring objects to be close or opposite. The final decision is passed to the Time and Trigger control system which transmits the result to each subdetector and lets the events be read.

High Level Trigger

The HLT is a software trigger executed on a multiprocessor farm placed aside the experiment. It has access to the whole detector readout and therefore it can perform complex calculations. The HLT is made by chains of filters and producers called trigger paths. The producers compute physics objects (track, cluster, energy, mass...) from the information of the event, the filters decide if the event has to be rejected or not based on the available information. HLT producer modules are programs similar to the ones used for the offline reconstruction, many of those are simplified in order to reduce execution time. If an event passes at least one trigger path, it is recorded on disk. Thanks to the HLT, the recorded event rate is about 1 kHz.

Some trigger paths may select events with a rate which is too large for CMS. When this happens only a fraction of selected events are recorded. This cut is random and it is called prescale. Prescaled trigger paths are present in both L1 and HLT. Those triggers are used to measure the efficiency of triggers with tighter requirements with respect to the prescaled one.

Chapter 3

Physics Objects reconstruction

The interesting events for the H→ $\mu\mu$ search are identified by the presence of two muons in the final state. Moreover, the Higgs production via VBF channel is characterized by the presence of two high energy quark-jets and low hadronic activity in the rapidity gap among the two jets. In this chapter, it is described how the physical objects of interest for the analysis are identified and reconstructed in CMS.

3.1 Track Reconstruction

Each detector described in section 2.2 measures signals generated from specific particles crossings. Signals from all CMS detectors are collected and saved whenever the event is triggered as explained in 2.2.4. Once data are stored, offline software runs all over the events in order to reconstruct and identify all the detected particles. This section describes the algorithm that reconstructs the trajectories of the charged particles from the tracker hits.

The first step of the reconstruction consists of processing the tracker readout to build hits. The tracker hits are the best estimates of the points where the charged particles have crossed the tracker modules. Each charged particle produces a sequence of tracking hits that identify the particle trajectory. The software used to reconstruct particle trajectories at CMS is called Combinatorial Track Finder (CTF) [70] and it is based on the Kalman filter [71]. The final collection of tracks is produced by a multiple steps procedure called iterative tracking.

The iterative tracking consists of repeating several times the CTF track reconstruction sequence. The first iterations search for the easier tracks to find, tracks with high transverse momentum and several pixel hits. The algorithm moves progressively to searching for more complex tracks, loosening track quality criteria. At each iteration, the hits associated with reconstructed tracks are not considered in order to reduce the combinatorial complexity and simplify the next iterations.

Figure 3.1 shows the efficiency of track reconstruction of CTF as a function of the particle p_T. All the iteration are visible in different colors. Specialized tracking

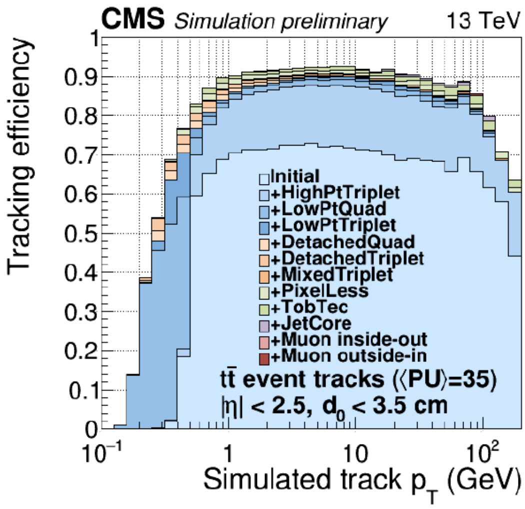


FIGURE 3.1: Tracking efficiency as a function of the track p_T in simulated events with the new pixel detector. The iterations are shown in different colors [72].

iterations are included in the plot as final steps but they are not included in the CTF algorithm.

3.1.1 Tracking iteration

Each iteration of CTF is composed of four steps, the main differences between the iterations lie in the configuration of the seed generation (first step) and in the final track selection (last step). The steps are the following:

- **Seeding** - The initial parameters of the tracks are defined by the seeding. The helical paths are described by 5 parameters, therefore at least three hits are needed to identify a candidate track (alternatively, two hits with additional constraints on the origin of the track can be used). The seeding algorithm generates all the possible combinations of initial hits that identify all the possible tracks. The initial hits used to generate the candidate tracks are called seeds. To limit the number of hit combinations, the seeds are required to meet certain conditions. In particular, in the first iterations of the CTF algorithm, the seeds must have small curvature and they must be compatible with a track originating from the proton-proton interaction region. The seed conditions are gradually loosening at each iteration of the CTF algorithm. Pixel seeding is preferred because of the lower occupancy and the better resolution. Moreover, nuclear interactions rise the number of tracks in the outer part of the tracker and it would raise the number of seeding and the computational cost in the outer tracker. Strips with double-side modules are also used in the last iterations.

  The current pixel detector has four layers and the best seeds have four hits now. Seeding configuration is shown in table 3.1 for all iterations.

Iteration	Seeding	Target
1	4 pixel	prompt, high p_{T} (initial)
2	3 pixel	prompt, high p_{T}
3	4 pixel	prompt, low p_{T}
4	3 pixel	prompt, low p_{T}
5	4 pixel	displaced
6	3 pixel	displaced
7	3 pixel+strip	displaced
8	3 inner strip	displaced
9	3 outer strip	displaced

TABLE 3.1: Seeding and target tracks of each tracking iteration.

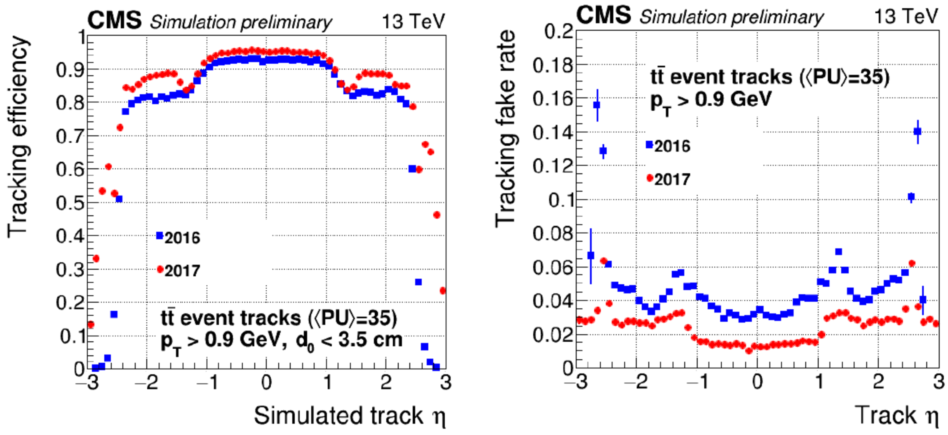
- **Track finding** - The track finding procedure is based on the Kalman filter method. The tracks found by seeding are extrapolated to the external layers one by one. A new hit is searched in each layer, compatible with the estimated crossing point, and if it is found the parameters of the track are updated with the new measurement. After the parametrization update, the procedure keeps extrapolating with the next layer. Because of tracker inefficiencies, some tracker hits may be missing. Therefore, when no hits are found in the next layer the track finding algorithm proceeds searching for compatible hits in the layer after.

  More than one compatible hit cluster can be found during the procedure. In this case, all the compatible hits are considered and a different candidate track is built for each one. The best candidate is chosen on the normalized χ^2 and the number of valid hits and it is promoted to track, while the other candidates are discarded. When the outward extrapolation is completed, an inward searched is performed from the outer hits to improve the hit collection efficiency. Lastly, a trajectory cleaner is applied at each pair of tracks in order to identify particles reconstructed more than once.

- **Track fitting** - After the track finding stage, the complete set of hits is available for each track. A fit procedure is then performed in order to obtain an optimal measurement of the track parameters. Hit positions and uncertainties are also improved by the track fitting.

- **Track selection** - A final track quality selection is applied as the final step in order to reduce fake rate. The selection is a function of p_T, η, the number of layers with valid hits, and the iteration number.

The tracking efficiency and fake rate are shown in Figure 3.2 as a function of track pseudorapidity. The 2016 performance is compared with 2017 one, the two are different mainly because of the pixel detector upgrade. The detector efficiency is almost flat in the barrel region and it significantly improves in the region $1.6 <|\eta|< 2.0$ where the new disks provide one hit more than than the previous ones. Either efficiency and fake rate improved with the new pixel detector. Efficiency and fake

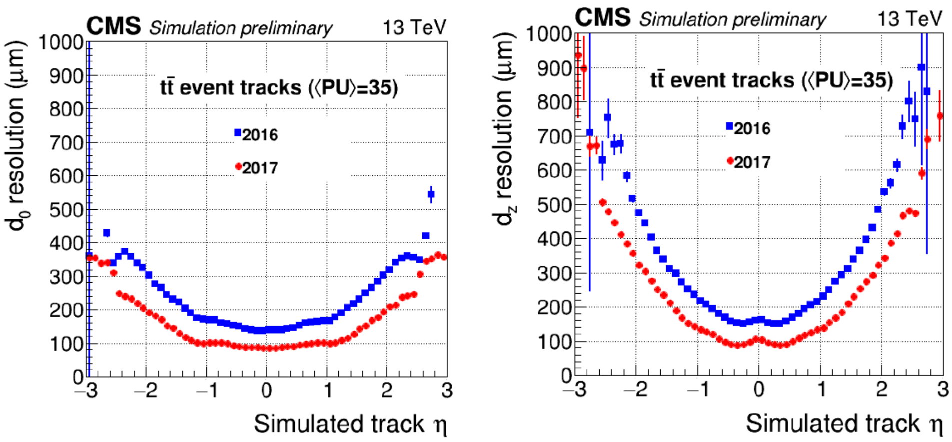


(a) (b)

FIGURE 3.2: Vertex transverse resolution as a function fo the number of associated tracks (a) and the number of reconstructed vertices as a function of the generated ones (b) [72].

rate have been also measured in data with similar results.

Charged particle tracks are identified by five parameters: the momentum, polar and azimuthal angles, and the transverse and longitudinal impact parameters. Figure 3.3 shows the resolutions of the transverse and longitudinal impact parameters in 2016 and 2017 in $t\bar{t}$ simulated events. The impact parameters resolutions are 100÷200 μm in the transverse plane and 100÷500 μm in the z-direction.



(a) (b)

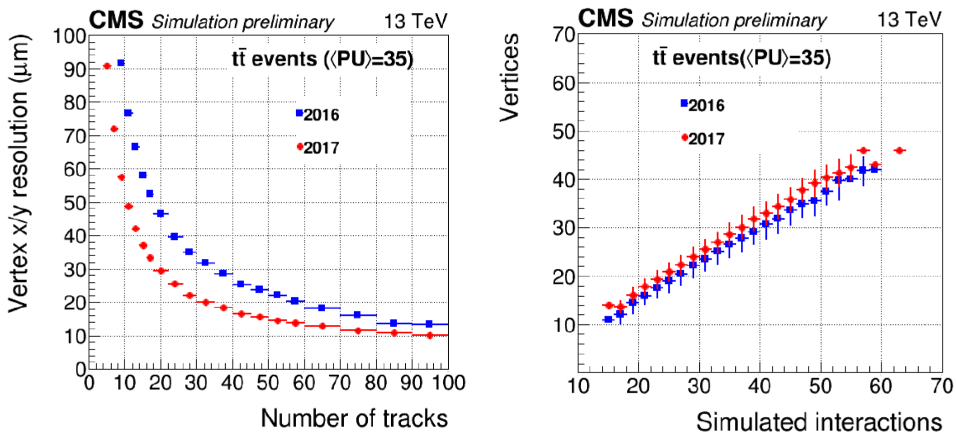
FIGURE 3.3: Transverse (a) and longitudinal (b) impact parameter resolutions as a function of the track pseudorapidity [72].

3.1.2 Primary vertex reconstruction

After the described reconstruction, all the tracks are extrapolated backward to find the position of proton-proton interaction they originate from. In a bunch crossing, there are several proton-proton collisions, therefore not all tracks point back to the same point. The interaction points, called primary vertices, are found using only well reconstructed tracks with no displacement.

At first, a track clustering is performed in z using the annealing algorithm [73]. Then the adaptive vertex fitter (AVF) [74, 75] is used to find the primary vertex from a collection of tracks. The AVF is an algorithm that fits the vertex position iteratively. At each iteration, each track is weighted with a function of the χ^2 obtained from the previous iteration. The AVF consists of performing an initial fit to the primary vertex using all the tracks in the collection. Then, at each iteration tracks with high χ^2 obtain smaller weights thus their contribution gradually vanishes. Instead, the weights of the tracks that have good compatibility with the vertex go asymptotically to one. Therefore only the tracks with low χ^2 are used to find the vertex position.

The more tracks are used in the fit, the more precise is the vertex. Figure 3.4a shows the vertex resolution in the transverse plane, longitudinal resolution is slightly worse mainly because of the size of the pixels. Figure 3.4b shows the number of reconstructed vertices as a function of the number of generated proton-proton collisions.



(a) (b)

FIGURE 3.4: Vertex transverse resolution as a function of the number of associated tracks (a) and number of reconstructed vertices as a function of the generated ones (b) [72].

Not all proton-proton collisions are studied in generic searches. One vertex in each bunch crossing is chosen and studied, it is called primary vertex. All the others are treated as noise and their products are not considered, those vertices are called pile-up vertices. The primary vertex is chosen as the one with the largest $\sum p_T^2$ where the sum is performed over all the objects reconstructed with the particle flow algorithm described in the next section.

3.2 Particle Flow

CMS, like many other high energy physics experiments, is composed of several subdetectors. The subdetectors are made by different technologies and they aim to measure different kinds of particles. As many other experiments do, also CMS could merely use the only subdetectors information to particle measurement. Nevertheless, in order to improve particles identification and momenta resolution, an algorithm that combine all the single subdetector measurements has been developed and used since Run 1. This algorithm is the particle flow algorithm [76].

Particle flow provides a global event reconstruction and it reconstructs a comprehensive list of final-state particles. This approach significantly improves the mitigation of pile-up effects, moreover, it allows a much better reconstruction of jets (see section 3.5), missing energy, and tau leptons, which are not covered in this thesis. The final list of objects returned from the particle flow algorithm is used to derive composite physics objects. They are named PF candidates.

The first step of the object reconstruction is performed independently within each CMS subdetector: silicon tracker reconstructs charged particles as described in 3.1, ECAL and HCAL cluster energy deposits separately, and muon chambers detect muons outside the solenoid. Then, signals in different detectors that may be caused by the same physic object are linked. The links are made if there is geometrical compatibility between the concerned objects and a link distance is defined to quantify the quality of the link. In order to reduce the computational cost, only the nearest neighbors are considered for the link building.

The particle flow algorithm proceeds in steps. The steps are described below and they run in the same order as follows.

Muon reconstruction
Muon reconstruction is the first step in the PF algorithm. In order to identify muons, it associates segments in muon chambers with inner tracker hits and it checks the compatibility of the two track momenta. Different selection criteria are applied to identify isolated muon and muons in jets, in order to deal with hadron misidentification. Whenever a muon is identified, the associated tracks in the tracker and in the muon system are removed from the list of unassociated objects.

Electron and photon reconstruction
Tracks are extrapolated to the calorimeters in order to link them with energy deposits. The extrapolation is computed up to one radiation length in ECAL and one interaction length in HCAL. If the extrapolation falls within the cluster area enlarged by one cell in each direction the link is built and the relative link distance is defined as the distance between the two objects in the η-ϕ plane. If one track is linked with more than one calorimeter cluster, or one cluster is linked with more than one track, the PF algorithm keeps only the link with the shortest distance.

Electrons and photons identification is performed using tracks from the tracker and

ECAL energy deposit. Photon candidates are identified by a cluster in ECAL without any linked track. Moreover, photons often convert into e^+e^- pair inside the tracker material. In order to take care of photon conversion, a dedicated conversion finder has been developed. If the sum of the two tracks momenta are compatible with the original photon direction, the tracks are removed from the PF element collections and their sum is promoted to photon.

Electrons are identified by a track linked to an electromagnetic cluster. Electrons have non-negligible probability to emit bremsstrahlung photons inside the tracker material, therefore tangents to the electron trajectory are extrapolated to the ECAL from each intersection point between the track and the tracker layers. An ECAL energy deposit is linked to the track as a potential bremsstrahlung photon if it is compatible with any tangent to the electron trajectory. Bremsstrahlung photon clusters have enlarged ϕ windows in order to take into account photon conversion at the edge of the tracker material.

The energy of the electrons is measured by combining the information of ECAL and the tracker, while the direction is taken from the track direction. Photon energy is measured by ECAL and its direction is taken as the direction from the primary vertex.

Hadron reconstruction
After muon, electron and isolated photon are identified and removed from the PF element lists, PF proceeds with hadron identification. Energy deposits in HCAL with no linked tracks are identified with neutral hadrons while energy deposits in HCAL with associated tracks are identified with charged hadrons. In the latter case, the energy cluster is compared with the track momentum. If the two energies are compatible, a charged hadron is identified and its momentum is determined by a weighted average of the HCAL energy and the track momentum. If the two energies are not compatible, a neutral hadron is identified out of the cluster energy excess. Similarly, if an energy excess is found in ECAL that is not compatible with charged hadron identified deposit, a new non-isolated photon is identified. This is very common in jets, where about 25% of the energy is carried by photons. Outside the tracker acceptance, there is no way to recognize charged particles, therefore ECAL+HCAL are simply tagged as hadronic shower.

Secondary vertices reconstruction
Prompt hadrons can have nuclear interactions with the tracker material. These interactions are reconstructed and the tracks coming from the secondary vertex are linked by the PF algorithm. Then, the secondary charged particles are replaced by a single primary charged hadron.

Event post processing
The particle identifications described above are the result of an optimized combination of the information of all subdetectors, however particle misreconstructions and misidentifications are inevitable. In general, these errors in reconstructions are difficult to identify, however in rare cases, they are detectable by looking at global

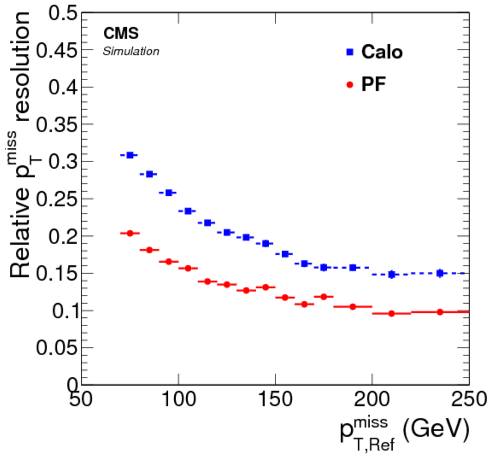


FIGURE 3.5: Resolution of the missing transverse momentum in $t\bar{t}$ events [76].

event variables, like the missing transverse momentum p_T^{miss}. The missing transverse momentum is defined by the negative sum of all the PF object momenta and it should be compatible with zero if neutrinos are not present in the event. Sometimes very high p_T^{miss} events are due to cosmic muon matching with a tracker particle, sometimes they are due to misreconstruction or misidentification of a genuine muon. Figure 3.5 shows the p_T^{miss} resolution when PF algorithm is run and when calorimeters information only are used.

3.3 Muons

Muon reconstruction [77] is almost independent of the particle flow algorithm. Muon tracks are reconstructed independently in the tracker as explained in section 3.1 and in the muon chambers. Therefore three different kinds of muon objects are built.

Standalone muon tracks are build using muon chambers information only. Their reconstruction is performed from CSC and DT segment used as seeds and using Kalman filter similarly to what is done for the tracker.
Tracker muon tracks are built "inside-out" from the tracker. Tracker tracks with $p_T > 0.5$ GeV and $p > 2.5$ GeV are propagated from the tracker outside the solenoid. Tracks are matched to the segments in the muon chambers if the relative distance is less than 3 cm in the rϕ or its relative uncertainty is smaller than 25%.
Global muon tracks are built "outside-in", matching standalone muon tracks with the associated tracker track. A final fit is performed using the Kalman filter in order to provide a better measurement of the muon parameters.

Muon momentum is computed from a simultaneous fit of the standalone muon track and the tracker track by the Tune-P algorithm [78]. The Tume-P algorithm performs several different fits and chooses the p_T measurement from the one with the best quality of the fit. The algorithm has been validated with cosmic rays, in proton-proton collisions, and in simulated events with different misalignment scenarios. Moreover, momentum and invariant mass distributions have been studied in order to ensure than the algorithm does not bias the momentum measurement.

Several muon variables have been studied to define different selection criteria to identify muons. Those criteria divide muons in several identification types with different efficiency and purity so that any analysis can choose the more convenient muon collection. The main indentification types are "loose", "medium", and "tight". Loose muons are PF muons that are either tracker or global muon. Medium muons are loose muons that have valid hits in at least 80% of the tracker layer they traverse. Moreover, medium muons must have $\chi^2/dof < 3$ and a position compatibility between the tracker track and the standalone muon trajectory of $\chi^2 < 12$. Tight muons are loose muons with at least 6 associated hits in the inner tracker. The muon must be compatible with the primary vertex and the global fit must have $\chi^2 < 10$.

An important criterion to distinguish prompt muons from muons coming from hadron decays in jets is the isolation. The isolation is the estime of the total p_T of the PF particles emitted in a cone of R=0.4 around the muon:

$$I_{Iso} = \sum_{charged} p_T + max(0, \sum_{neutral} p_T + \sum_{\gamma} p_T - p_T^{PU}) \tag{3.1}$$

where the sum over the charged particles considers the ones coming from the primary vertex only, and the p_T^{PU} is an estime of the pile-up neutral particles. The contribution of neutral particles coming from pile-up is computed by summing all charged hadrons originating from pile-up vertices and scaling it by a factor of 0.5. The factor 0.5 is an estimate fo the neutral to charged hadron production ratio in inelastic proton-proton collisions. The energy contribution from neutral particles in isolation is set to be positive.

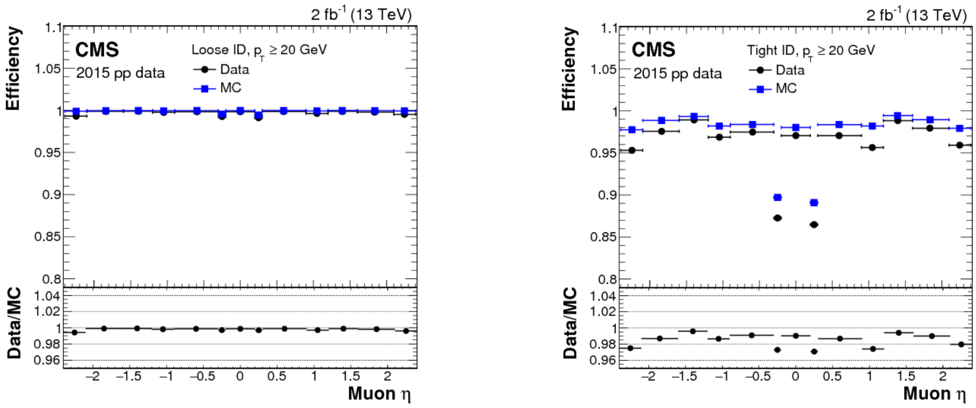
Two working points are generally used to select isolated muons. The loose isolation is defined to have 98% efficiency and it selects muons with $I_{Iso}/p_T < 0.25$. The tight isolation selects muons with $I_{Iso}/p_T < 0.15$ and it have efficiency of 95%. These values have been tuned and tested in simulated $Z \to \mu^- \mu^+$ and multijet events.

The efficiencies of muon identification and isolation have been studied with the tag-and-probe method. The total muon efficiency is factorized as:

$$\epsilon_\mu = \epsilon_{track} \times \epsilon_{reco+ID} \times \epsilon_{Iso} \times \epsilon_{trig}$$

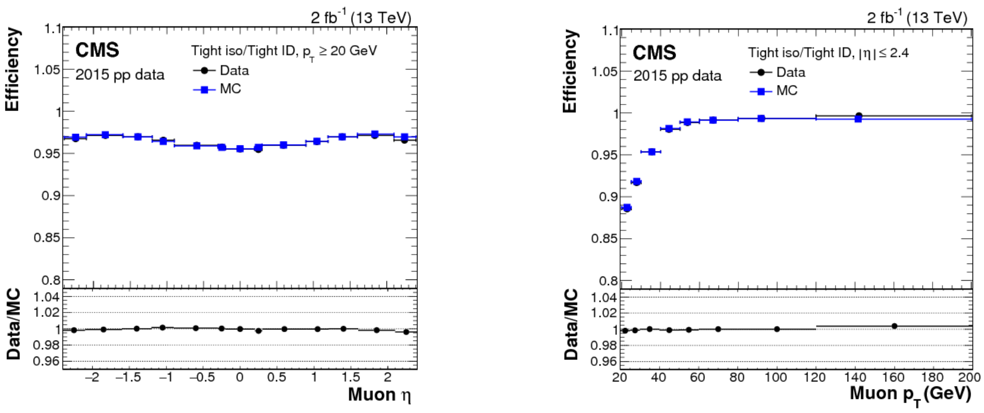
and the four factors have been studied separately. The efficiency of track reconstruction is ϵ_{track}. The term $\epsilon_{reco+ID}$ contains reconstruction and identification efficiencies. The isolation efficiency is ϵ_{Iso}, and ϵ_{trig} is the trigger efficiency. Scale factors for muon efficiency are derived from data/MC comparison and applied for the analysis.

Figure 3.6 shows the reconstruction and identification efficiency $\epsilon_{reco+ID}$ measured in data in 2015. The loose identification efficiency exceeds 99% over the entire η range, and agreement between data and MC is within 1%. Figure 3.7 shows isolation efficiency for muons that pass the tight identification criteria as a function of the pseudorapidity and of the momentum.



(a) (b)

FIGURE 3.6: Tag-and-probe efficiency of muon identification and reconstruction for loose (a) and tight (b) identifications. The statistical uncertainties are smaller than the symbols used to display the measurements [77].



(a) (b)

FIGURE 3.7: Tag-and-probe efficiency of muon isolation as a function of the pseudorapidity (a) and of the momentum (b) of the muon. The statistical uncertainties are smaller than the symbols used to display the measurements [77].

3.3.1 Transverse momentum resolution

The sensitivity of the $H \to \mu\mu$ search strongly depends on the width of the dimuon mass peak. The contribution of the expected theoretical Higgs boson decay width to the mass peak width is negligible, therefore the width of the dimuon mass peak depends uniquely on the experimental resolution. The direction of the muons is measured with high precision (about 0.1 ÷ 1 mrad for muons with p_T >10 GeV), hence the resolution on the invariant mass is dominated by the muon momentum uncertainty.

There are two main sources of uncertainty that determine the transverse momentum resolution: the first one is the multiple scattering of the muon with the detector, the second one is the non-infinite precision of the hit positions.

Multiple scattering effects depend on the material budget of the detector, thus the associated uncertainty depends on the pseudorapidity. The introduced angle of deflection is proportional to $1/p_T$, therefore the corresponding uncertainty is proportional to p_T. The error on the hit position leads to an uncertainty that is proportional to p_T^2. The multiple scattering effects are the major source of uncertainty in the transverse momentum measurement for muon with $p_T \lesssim$ 30 GeV, while the second effect is dominant for high energy muons. The transverse momentum uncertainty is estimated in simulated events by computing the residuals, which are defined as the differences between the generated and the reconstructed track p_T. Figure 3.8 shows the resolution of the transverse momentum uncertainty as a function of the muon p_T. For each bin, the half-widths of the residual distribution containing 68% and 90% of the events are computed and they are shown in the plot.

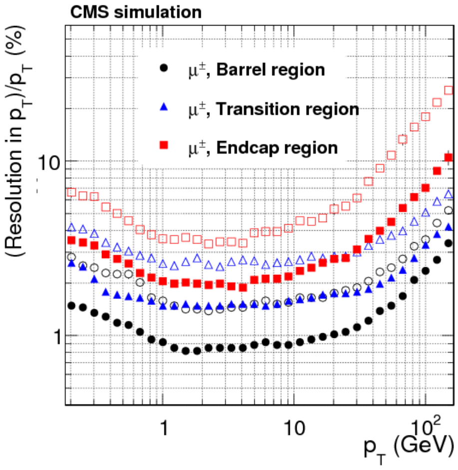


FIGURE 3.8: Resolution of the transverse momentum as a function of the pseudorapidity or single, isolated muons with pT = 1, 10, and 100 GeV. For each bin, the solid (open) symbols correspond to the half-width for 68% (90%) intervals of the residuals distribution [70].

Figure 3.9 shows the transverse momentum resolution for the muons with momentum 70 GeV, that is the typical muon momentum in the H$\rightarrow$ $\mu\mu$ search. The resolution is 1 $\div$ 2% in the barrel region where differences in resolutions are caused by the multiple scattering. Instead, in the endcaps, the resolution gets significantly worse as a function of the pseudorapidity. The reason is that the level arm used to perform the measurement is reduced for muons with $|\eta| \gtrsim 1.5$.

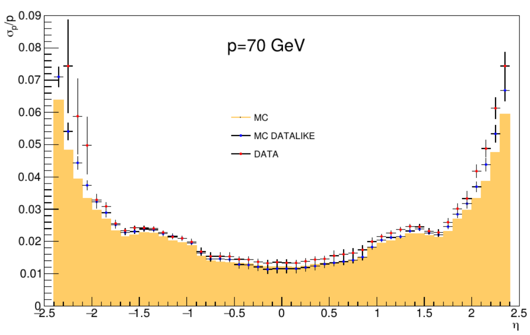
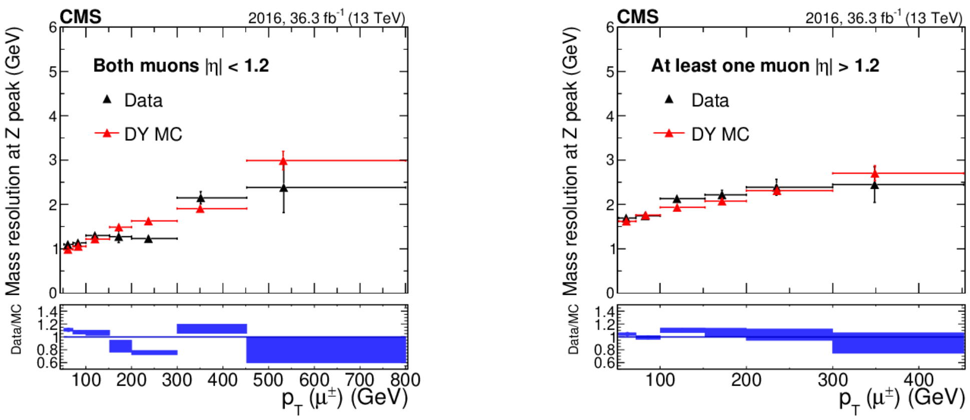


FIGURE 3.9: Transverse momentum resolution in data and in simulation as a function of the pseudorapidity for muons with momentum of 70 GeV [79].

The resolution of the muon momentum strongly depends on the muon pseudorapidity. Consequently, the resolution of the dimuon invariant mass depends on the pseudorapidities of the two muons. Whenever the two muons are central, the dimuon mass is measured precisely. In the case of one muon is central and one is forward, the mass is measured with less precision. When both the muons have high pseudorapidity the invariant mass has low resolution. Figure 3.10 shows the average dimuon mass resolution under the Z peak. In the left plot, both the muons in the events are central while in the right plot at least one muon is not in the barrel region. The resolution in the second case is worst in the entire p_T spectrum and the reason is that one of the muons is measured with smaller precision.



(a) (b)

FIGURE 3.10: Dimuon mass resolution at the Z peak as a function of muon transverse momentum. Each dimuon event is counted twice since each muon in the event is filling the histograms. The values in each bin are obtained from the average of the distribution within the bin for data and simulation [80].

An additional source of error in the momentum measurement is the uncertainty due to the muon momentum scale. In CMS, the muon momentum scale is measured as a function of η using Z and J/Ψ events. Its contribution is less than 0.2% [81] therefore it does not contribute significantly to the dimuon mass uncertainty.

3.3.2 Final state radiation recovery

In a small fraction of events, muons from Higgs decay radiate a photon, this process is called final state radiation (FSR). In those cases, the dimuon invariant mass is reduced because the photon momentum is not taken into account. This effect causes a long tail in the invariant mass distribution, degrading the resolution of the signal peak. Moreover, if the photon falls in the isolation cone of the muon it can increase the relative isolation of the muon and reducing the selection efficiency. In order to avoid these effects, a procedure has been developed to recover FSR of selected muons, similarly to the one used in [82].

The procedure is applied to all muons with $p_T > 20$ GeV and $|\eta| < 2.4$. Firstly, all the photons within a cone of $\Delta R = 0.5$ around the muon tracks are considered. Photons that lie in the window $1.44 < |\eta| < 1.57$ or that are already associated with the bremsstrahlung of electrons are ignored. Remaining FSR photon candidates are discarded if the sum of the p_T of all particles in a cone of 0.3 around the photon, muon excluded, is greater than 180% of the p_T of the photon or if $\Delta R(\mu, \gamma)/p_T^2(\gamma) > 0.012$. Furthermore, in order to avoid $H \to Z\gamma \to \mu\mu\gamma$ decays the photons are required to have $p_T(\gamma)/p_T(\mu) < 0.4$. In case of multiple choice, the photon with the smallest value of $\Delta R(\mu, \gamma)/p_T^2(\gamma)$ is chosen.

Whenever this procedure associates a photon with the muon radiation, the momenta of the two particles are added and the sum is identified with the muon momentum. Moreover, the photon is ignored from the isolation sum defined in equation 3.1. The described FSR recovery procedure improves the signal efficiency of about 2% and the mass peak resolution of about 3%.

3.4 Electrons

Electrons are reconstructed by a mixture of standalone approach [83, 84] and PF algorithm. Electrons lose energy via bremsstrahlung effect in tracker material, therefore part of the energy of the electrons is spread over several ECAL crystals. The energy radiated before ECAL is on average 33% of the original electron energy at $\eta = 0$, while it is about 80% at $\eta = 0$. Figure 3.11 shows the fraction of energy lost by electrons in data and MC in $Z \to ee$ events.

The energy is estimated by a multivariate technique that takes into account either the track measurement either the cluster energy. The resolution is about 5÷10% in the endcaps and it reaches 2% in the barrel for high energy electrons.

Bremsstrahlung photons alter electron trajectory in the tracker, therefore charge can be mismeasured, especially at $|\eta| > 2$. In order to measure the electron charge, three different methods are used and the charge sign is chosen as the one shared

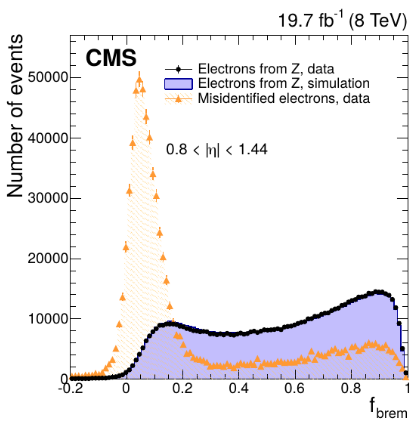


FIGURE 3.11: Fraction of the momentum lost due to bremsstrahlung in $Z \to ee$ data (dots) and simulated (solid histogram) events, and from background enriched events (triangles). The distributions are normalized to the area of $Z \to ee$ data distribution [84].

by at least two methods. Two measurements are provided by the curvatures of the track fitted with KF and with the gaussian sum filter (GSF) method [85]. The third one is given by the sign of the difference in ϕ of the ECAL energy deposit and the first tracker hit. The combined three methods provide a misidentification probability of 1.5% for $Z \to ee$ events without any further selection.

CMS developed many algorithms to identify prompt electrons and efficiently reject other background sources like photon conversions or hadron misidentifications. Some of them are simply cut based selections, while others combine several variables in multivariate techniques. The electron identification algorithm currently used by CMS is multivariate discriminant that exploit a great number of discriminating variables. Figure 3.11 shows one of them in the range $0.8 < |\eta| < 1.44$.

The discriminator efficiency and misidentification have been measured in data with the tag-and-probe method in $Z \to ee$ events. Two working point are commonly used, one has 90% efficiency and the other one has 84% efficiency on electrons with from Z decay with $p_T > 20$. Figure 3.12 shows the efficiency of the 90% working point discriminator as a function of the electron momentum for differen η regions.

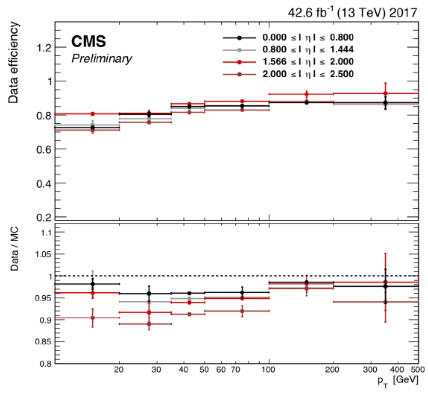


FIGURE 3.12: Electron efficiency as a function of the electron p_T measured for the multivariate identification [86].

A significant fraction of reconstructed electrons are misidentified hadrons or are genuine electrons originated from a hadron decay. In both cases, several particles are close to the fake electron in the $\eta - \phi$ plane. In order to identify those non-prompt electrons an isolation selection can be applied. Electron isolation is equivalent to muon one and it is defined by eq. 3.1, but in case of electrons, the sums are performed with PF particles that are in a cone of R=0.3 within the electron.

In the $H \to \mu\mu$ search, only electrons with $|\eta| < 2.5$ and $p_T > 20$ GeV have been considered: they are used to tag vector bosons. Moreover, because of nonoptimal reconstruction performance, electrons in the region between the ECAL barrel and endcaps $1.44 < |\eta| < 1.57$ have not been considered.

3.5 Jets

Almost every collision at LHC produces scattered high momentum quarks and gluons. Quarks and gluons cannot be observed free due to QCD confinement, they instead fragment, producing a narrow cone of hadrons that carry most of the momentum of the original parton. Those sets of collimated particles are the evidence of quarks and gluons and they are detected as jets. High energy jets usually originate from hard scattering processes but many more jets are present in a generic event.

These additional jets come from the underlying event and from pile-up collisions, they are also called "hadron activity".

Jet clustering algorithms are defined to select the particles originating from one high momentum quark or gluon. Their task is to identify collimated tracks and gather them into a single object called jet. The momentum of the jet is defined by the sum of the momenta of all the particles that compose it and it is considered the momentum of the original parton. The jet momentum has to be a good proxy of the original quark or gluon momentum. This feature is very important when jets are used to mass resonance searches.

A good jet clustering algorithm has to be infrared and collinear safe. It is infrared safe if the final selection of particles does not change because of a very low energy gluon emission from one of the particles in the event. It is collinear safe if the final selection of particles does not change when a particle is replaced by two almost colinear particles with a total momentum which equals the one of the initial particle. Collinear and infrared safeness is very crucial for a clustering algorithm because it ensures the stability of the jet collection under low Q^2 QCD processes.

Several jet clustering algorithms exist, the most used ones are the k_T [87], Cambridge/Aachen [88] and anti-k_T algorithm[89]. They belong to a class of sequential recombination algorithms and their use depends on the particular study. They all are infrared and collinear safe. In CMS, jets are clustered from particle-flow objects using the anti-k_T algorithm, it is implemented by FASTJET [90, 91] with small computational cost [92].

The sequential procedure to reconstruct the jets depends on two parameters, n and R, and it is as follows.

Make a list with all the candidate particles for the algorithm (usually thay are all the particles from the primary vertex): they are called protojet. Then execute the following steps:

1. *Compute the distance d_{ij} for each protojet pair and the distance d_i for each protojet:*

$$d_{ij} = \min(k_{Ti}^n, k_{Tj}^n)\frac{\Delta_{ij}}{R}$$
$$d_i = k_{Ti}^n$$

   where $\Delta_{ij} = \sqrt{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2}$

2. *Let d_{min} be the smallest among all the d_{ij} and the d_i. If d_{min} is one of the d_{ij}, merge the protojets i and j in a single protojet with four-momentum equal to the sum of the four-momenta of i and j. If d_{min} is one of the d_i, promote the protojet i to jet and remove it from the protojet list.*

3. *It there are some remaining protojets, go back to point 1.*

This procedure merges protojets only if their angular distance is smaller than R. In CMS, the standard clustering distance used is $R = 0.4$, but jet with $R = 0.8$ are also used in heavy particle searches.

The parameter n characterizes the algorithm. The k_T algorithm corresponds to the choice of $n = 1$, so it associates first particles with small transverse momenta. The Cambridge/Aachen is identified by $n = 0$, therefore it associates the closest particle in the $\eta - \phi$ plane without considering the momenta. The anti-k_T algorithm has $n = -1$, thus particles are clustered around the most energetic ones.

3.5.1 Calibration

Like all experimentally-reconstructed objects, jets need to be calibrated in order to have the correct energy scale. The jet energy correction (JEC) is computed using the same approach of the 8 TeV calibration [93] and it is factorized in three main steps: the pile-up offset corrections, the simulated response corrections, and the residual corrections.

Pile-up offset refers to the additional contribution of jet energy and momentum because of pile-up interactions. Proton-proton collisions in previous and subsequent bunch crossing also contribute to energy deposit due to the finite signal decay time in the calorimeter therefore two different effects are considered: the in-time pile-up (IT PU) and out-of-time pile-up (OOT PU). The OOT PU is mitigated by calorimeter signal processing. The IT PU is corrected by removing the particles coming from pile-up interaction. The charged ones are simply identified and removed, the neutral ones are subtracted by estimating the median energy density ρ per-event. All the corrections are parametrized as a function of η, p_T, and the jet area.

The simulated response corrections are derived and applied after the pile-up offset corrections. They are determined in simulation by comparing the jet energy with respect to the generator level jet one.

The residual corrections are determined by comparing data/MC. Firstly, the response of jet over a wide range in p_T is corrected relative to the one of barrel jets ($|\eta| < 1.3$). This is called the relative jet correction ad it is determined with a sample of dijet events with low statistical uncertainty. Then, jets are calibrated up to $\eta < 5.2$ exploiting the transverse momentum balance at hard-scattering level with a reference object. This second step is performed in Z/γ+jet events and it is called absolute correction.

Corrections for the jet energy scale (JES) are derived for jets with $p_T > 10$ GeV and $|\eta| < 5.2$ with uncertainties less than 3%. Figure 3.13 shows the pile-up offset and the simulated response corrections for barrel jet.

The jet energy resolution (JER) is measured after applying JES in data and MC events. The method used for determining the JER is the same as the one used for JES, but the width of the distributions is studied instead of the means. The JER is

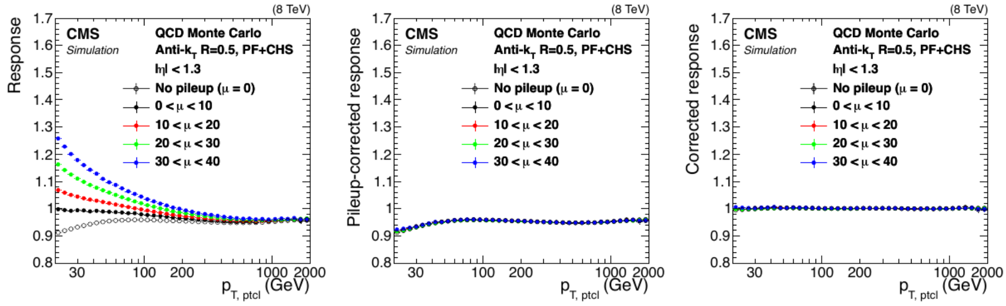
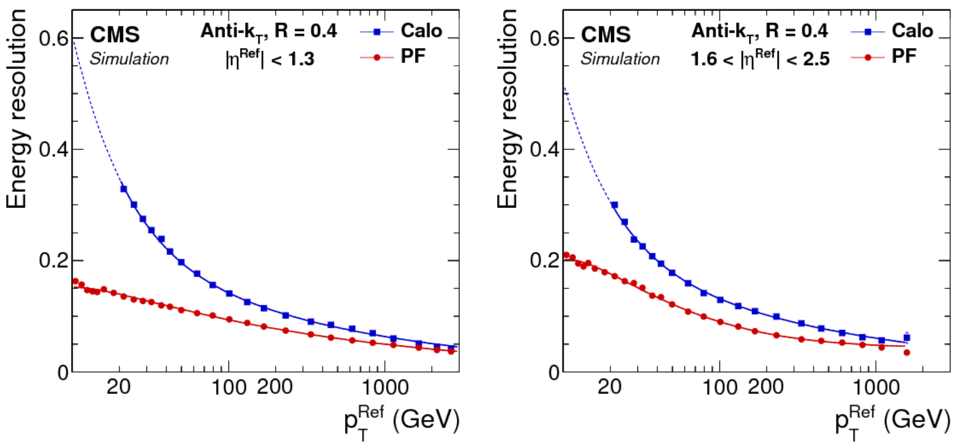


FIGURE 3.13: Average values of the ratii of measured jet p_T to particle-level jet p_T as a function of particle-level jet $p_{T,ptcl}$ at various stages of JES: before any corrections (left), after pile-up corrections (middle), and after all JESs (right). Distributions corresponding to different average numbers of pileup interactions per bunch crossing (μ) are shown separately [93].

extracted as a function of p_T, η, jet size R, and the pile-up interactions. The typical resolution is 15–20% at 30 GeV, about 10% at 100 GeV, and 5% at 1 TeV in the central rapidity region. As mentioned in 3.2 jet energy measurements are accurate due to PF algorithm. Figure 3.14 compares the energy resolution of PF jets momentum with jets measured from calorimeters only.



(a) (b)

FIGURE 3.14: Jet energy resolution for barrel (a) and for endcaps (b). The comparison between PF jets and calorimeter jets is shown [76].

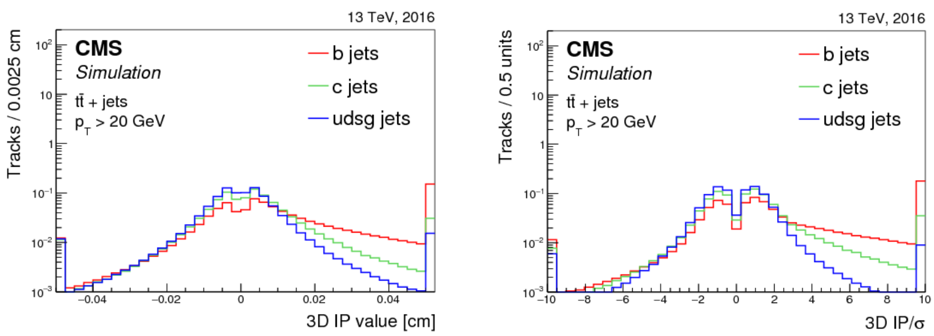
Jets are usually ordered by transverse momentum. In this work, jets referred as leading are the one with the highest transverse momentum in the event, while the subleading jets are the one with the second highest transverse momentum.

3.5.2 Identification of b-jets

The identification of b quarks is very important for either new physics searches either standard model measurements. Several algorithms have been developed by

CMS [94, 95] in order to identify jets originating from b quarks, called b-jets. Jets coming from light quarks and b-jets look similar, in fact, there is no difference in their reconstruction. However, B hadrons are always produced in b hadronization and the B hadrons can be exploited to identify the presence of b-quark.

B hadrons are long living particles with $c\tau \simeq 500$ μm, thus a B with p_T in the range 10÷100 GeV decay after 0.1÷1 cm on average. The first layer of the CMS pixel barrel is placed at 3 cm from the interaction point therefore B hadrons cannot be directly measured but they are identified by measuring their decay products.
Particles coming from the B decay show a sizable impact parameter (IP) with respect to the primary vertex because of the B long lifetime. In addition, B hadron semileptonic branching fraction is significantly higher than for other hadrons, and the leptons show a high transverse momentum relative to the jet direction due to the large B hadron mass ($\sim$ 5 GeV). Moreover, the B decay point can be measured using the tracks of the B decay inside the b-jet and the invariant mass of the particles compatible with the secondary vertex can be computed and compared with the B hadron mass. B hadrons decay to charmed hadrons (D) with a large branching fraction. D hadrons have non-negligible lifetime ($c\tau \simeq 300$ μm) therefore more than one secondary vertex can be reconstructed within b-jets.
Many observables can be used to discriminate between b-jets and non-b-jets. Some b-tagging algorithms use just one of those observables, while others combine more than one for better discrimination.



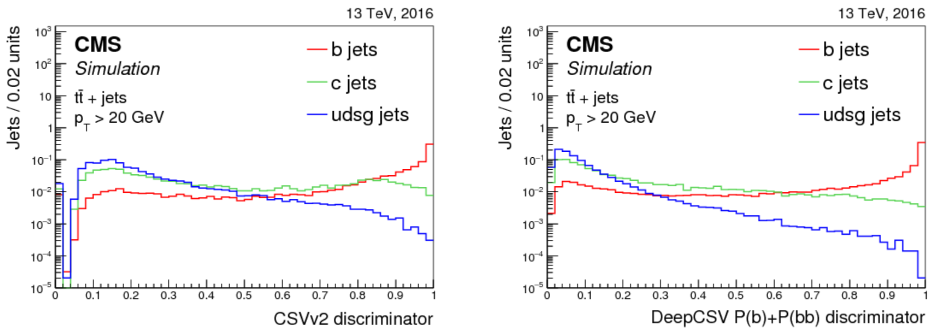
(a) (b)

FIGURE 3.15: Distributions of the IP (a) and S_{IP} (b) for tracks in jets originating from different quarks hadronizations. The first and the last bins include overflow and underflow entries [95].

The track counting (TC) is a b-tagging algorithm that relies on one observable only: the impact parameter significance (S_{IP}). Particles coming from B decay are characterized by having an IP of 100÷500 μm and they are measured with a resolution of about 30 μm. Also, the sign of the IP has discriminating power, it is defined as the sign of the scalar product of the jet direction with the IP direction. Tracks from B decay have mostly positive IP. Prompt tracks can have non zero IP because of resolution and multiple scattering but the distribution of their IP is symmetric around zero. TC algorithm computes the impact parameter significance for each track

$$S_{IP} = \frac{IP}{\sigma_{IP}}$$

where is the IP uncertainty, and it sorts the tracks by this parameter. Two versions of the algorithm are used: the track counting high efficiency (TCHE) where the output of the algorithm is the S_{IP} of the second track in the ordered list and the track counting high purity (TCHP) where the output is the S_{IP} of the third track in the ordered list. Figure 3.15 shows the distribution of the S_{IP} for tracks in b-jets and in non-b-jets.



(a) (b)

FIGURE 3.16: Distribution of the CSVv2 (a) and DeepCSV (b) discriminators for jets of different flavours in tt events [95].

Several algorithms exploit more than one variable to discriminate b-jets from non-b-jets. The jet probability (JP) combines the information of IPs of several tracks therefore it can be considered a natural extension of the TC algorithm. The combined secondary vertex (CSV) is another algorithm that uses even more information about tracks within the jet, moreover, whenever secondary vertices are found, CSV exploit their four-momentum. An improved version of the combined secondary vertex CSVv2 has been developed during Run 2, Figure 3.16 shows its distribution for jets originating from quarks of different flavours.

Deep learning techniques have been largely used in the last years. In particular, deep learning has been employed for improving the identification of the b-jets. The DeepCSV algorithm has been developed using a feed neural network with four fully connected hidden layers, each with 100 nodes. The input variables were the same as the CSVv2 ones, but they have been linearly transformed in order to have mean zero and a root-mean-square value of unity. Figure 3.16 shows the distribution of the DeepCSV tagger.

Three working points are defined for each b-tagging algorithm: loose (L), medium (M) and tight (T). The working points have been extracted in multijet events with p_T =30 GeV on average and correspond to a misidentification rate of 10%, 1%, and

0.1% for loose, medium, and tight respectively. Figure 3.17 shows the misidentification probability of several algorithms as a function of the b-quark identification efficiency. This analysis makes use of the DeepCSV discriminator at the loose and medium working point.

Efficiency and misidentification of b-tagging algorithms have been measured in data and scale factors have been extracted in bins of η and p_T and the discriminator itself. Efficiency has been calibrated in b-enriched samples obtained selecting leptons or in $t\bar{t}$ events. Misidentification scale factors have been derived in Z+jets samples where one jet is used as tag. Scale factors are close to one but all algorithms perform slighly better in MC than in data.

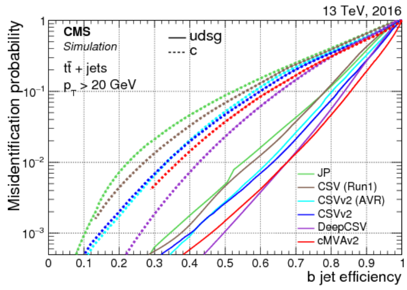


FIGURE 3.17: Misidentification probability versus b-jet identification efficiency in simulated applied to jets in $t\bar{t}$ events for several b-tagging algorithms [95].

3.5.3 Gluon/Quark jet discrimination

As known both from first principles and a large collection of experimental measurements, hadronic jets produced by gluon fragmentation show differences with respect to jets initiated by quarks. These differences have been extensively studied and validated [96, 97], and they have been exploited by many analysis to discriminate between gluon-jets and quark-jet in Higgs and electroweak searces [98, 99, 100]. The Gluon/Quark likelihood (QGL)[101] is a probability tagger that discriminates gluon-jets from quark-jets.

The QGL is built from three variables based on the jet constituent's information provided by the PF algorithm. Only particles that are well associated with the jet, or with $p_T > 1$ in case of neutral, are taken into account. The three discriminating variables are described below:

- **The jet constituents multiplicity**
  The multiplicity is a simple and well studied property. The quark-jets tend to have a smaller number of constituents respect to gluon-jets. This variable is particularly discriminating for high p_T jets where the average quark-to-gluon multiplicity ratio would converge to 4/9 [102].

- **The jet constituents minor RMS in the $\eta\phi$ plane**
  The jet p_T-weighted shape in the $\eta\phi$ plane is approximated by an ellipse. The ellipse is identified by the minor and a major axis length and orientation. The minor axis length is used for building the Gluon/Quark likelihood and it is particularly useful for low p_T jets. In order to find the minor axis length, the

p_T-weighted axis of the jet is computed and the 2×2 matrix M is build

$$M_{11} = \sum_i p_{T,i}^2 \Delta\eta_i^2$$
$$M_{22} = \sum_i p_{T,i}^2 \Delta\phi_i^2$$
$$M_{12} = M_{21} = \sum_i p_{T,i}^2 \Delta\phi_i \Delta\eta_i$$

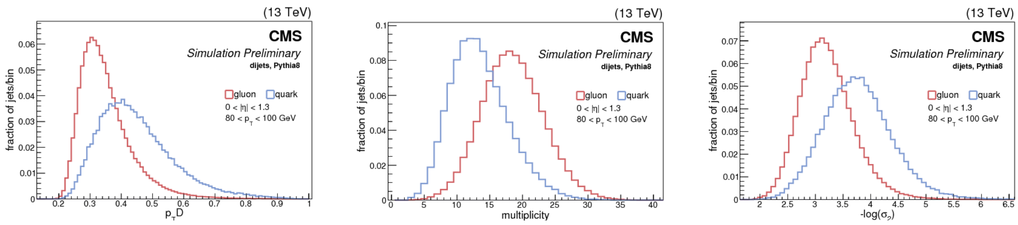
where the sum is performed all over the constituents, $p_{T,i}$ is the transverse momentum and $\Delta\eta_i$ and $\Delta\phi_i$ are the pseudorapidity and azimuthal distance between the constituent and the jet axis.
Let λ_1 and λ_2 be the major and the minor eigenvalues of the matrix M, the p_T-weighted minor axis length is

$$\sigma_2 = \sqrt{\frac{\lambda_2}{\sum_i p_{T,i}^2}}$$

- **The jet internal p_T distribution**
  Quarks fragmentation produces hard constuents that carry a significant fraction of the jet energy. The jet fragmentation distribution is defined by

$$p_T D = \frac{\sqrt{\sum_i p_T^2}}{\sum_i p_T^2}$$

  where the sum runs over the consituents. This variables is generally smaller for gluon-jets.



(a) (b) (c)

FIGURE 3.18: QGL input variables distribution for quark and gluon simulated jets [101].

Figure 3.18 shows the distribution of the three described variables.

The QGL has been validated using Z+jet events as quark enriched sample and dijet events, which are gluon enriched. Moreover, data-driven scale factors have been computed and fitted with polynomial functions to be applied separately for quarks and gluon jets in MC events. Figure 3.19 shows the QGL output distribution in data and MC for the two enriched samples.
In this analysis, all the MC events include those polynomial corrections while uncorrected events have been used to estimate the systematic uncertainty.

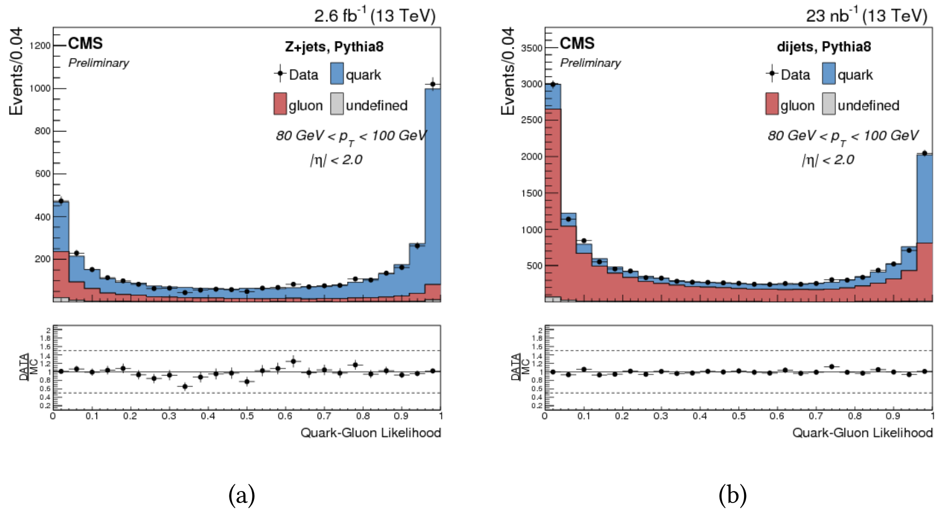


FIGURE 3.19: Data-MC comparison for the QGL output distribution in (a) Z+jet events and (b) dijet events. MC events include the polynomial corrections and the contribution of quark and qluon jets are shown separately [101].

3.5.4 Additional "soft" hadron activity

Hadronic jets are reconstructed taking into account neutral particles measured in calorimeter towers. Such a method provides a good measurement of the original parton energy but it is very sensible to pile-up effects. An alternative jet collection is build using charged particles only, called “track jets”. Track jets have an excellent angular resolution and they contain tracks unambiguously associated with the primary vertex. Track jets [103, 104] are a clean and well studied object used to reconstruct the hadronization of partons, they can efficiently measure low energy jets down to few GeV and they are mainly used to avoid pile-up noise. Therefore track jets are often used to estimate the additional hadron activity.

In order to build the track jets, tracks satisfying the following requirements have been considered

- have a *high purity* quality flag,
- have p_T > 300 MeV,
- are not associated to the Higgs decay leptons, nor to the selected two p_T-leading jets in the event,
- make minimum $|d_z(\mathrm{PV})|$ when associated to the event hardest primary vertex (PV),
- satisfy $|d_z(\mathrm{PV})|$ <2 mm with respect to the hardest PV.

Track jets have been build by clustering the selected tracks with the anti-k_T clustering algorithm with distance parameter R = 0.4. In this analysis, track jets kinematic variables have been considered in order to separate signal from backgrounds. In paricular, the following variables have been taken into account

- the transverse momentum of the leading track jets.
- H_T^{soft}, $H_{(2)T}^{\text{soft}}$, $H_{(5)T}^{\text{soft}}$, $H_{(10)T}^{\text{soft}}$: the scalar p_T sum of the soft track-jets with transverse momentum p_T >1, 2, 5, and 10 GeV respectively;
- N_2^{soft}, N_5^{soft}, N_{10}^{soft}: the multiplicity of the track jets with transverse momentum p_T >2, 5, and 10 GeV respectively;

Chapter 4

Search for $H \rightarrow \mu\mu$ with CMS Run 2 data

We implemented four dedicated searches of $H \rightarrow \mu\mu$ decay for the four main Higgs production channels. The search in the channels ggH, ttH, and VH adopt the same data-driven strategy and they are presented in this chapter while the VBF dedicated search that I developed is described more in detail in the next chapter. The results are combined among the four searches and they are shown in chapter 6.

4.1 Introduction

The analysis presented in this work aims to provide a measurement of the Higgs-to-muon couplings. In order to maximize the sensitivity of the Higgs to muons decay, a dedicated search has been performed for each production channel: gluon-fusion (ggH), vector boson fusion (VBF), vector boson associated production (VH), and top quark-antiquark associated production (ttH). Each search exploits the signature of the channel under study, and the final results have been combined and published in a common paper [105]. In VBF, VH, and ttH channels, the amount of signal is much smaller than the ggH channel. Nevertheless, the channels with less Higgs events have lower amounts of background events, therefore all the channels contribute significantly to the combination.

This thesis dedicates focuses in particular on the VBF channel search, to which I contributed directly. The VBF channel is the second one by cross-section and has the best overall sensitivity. Chapter 5 is fully dedicated to the VBF channel. The searches in the channels ggH, VH, and ttH, which were carried out by other groups inside the CMS collaboration, and the common preselection and categorization (that is also relevant for VBF search) are presented in the following sections, in order to provide a complete description of the Higgs to muons decay measurement. The combination of the results from all the channels is shown in chapter 6.

The study has been performed using the proton-proton collision data collected by the CMS detector in the years 2016-2018, corresponding to an integrated luminosity of 137.4 fb^{-1}. The result from this work has been combined with the one from data recorded at $\sqrt{s}$ =7 and 8 TeV corresponding to 5.1 fb^{-1} and 19.1 fb^{-1}. The final combination improves the results of about 1% and it is shown at the end of chapter 6.

4.2 Common preselection and categorization

A loose preselection is common to all the abovementioned searches. The preselected events are divided among the four searches with complementary cuts. The cuts exploit the signatures of the VBF, VH, and ttH production modes and define the phace space under study. The priority has been given to those three channels because they have larger signal-to-background ratio. All the remaining events are classified as ggH candidates and they are considered in the ggH dedicated search.

The triggers perform the first step of the selection. The signal events contain two prompt muons coming from the Higgs decay. The p_T distribution for the leading muon (see Figure 4.2) has a turn on at about 40 GeV and peaks around 60 GeV, therefore signal events can be selected with high efficiency using single muon triggers. The L1 trigger requires at least one standalone muon in the muon chambers with a p_T threshold of 22 GeV. After that, the HLT Single Muon trigger selects events with at least one reconstructed muon with $p_T >$ 27 GeV in 2017 and with $p_T >$ 24 GeV in 2016 and 2018 in order to ensure fully trigger efficiency.

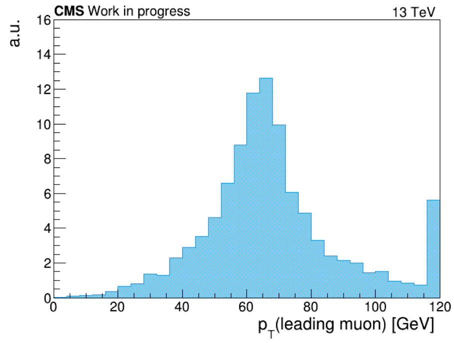


FIGURE 4.1: Distribution of the transverse momentum of the leading muon in simulated Higgs events. The last bin includes overflow and underflow entries.

After the trigger selection, the events are reconstructed as described in chapter 3. The offline selection requires the presence of two reconstructed muons with medium identification criteria. Moreover, the muons must have opposite charge, $p_T >$ 20 GeV, $|\eta| <$ 2.4, and they have to pass the loose isolation. Lastly, at least one muon must have $p_T >$ 29 GeV for 2017 events and $p_T >$ 26 GeV for 2016 and 2018 events.

Higgs events produced via ttH channel contain two t-quarks in the final state. The t-quarks decay to b-quarks [106], therefore ttH events contain also two b-jets. Those events are extracted from the preselected sample thanks to the DeepCSV algorithm described in section 3.5.2. Events containing two loose b-tagged jets or one medium tagged jet with $p_T >$ 25 GeV and $|\eta| <$ 2.5 are classified as ttH events. These events are further divided into leptonic and hadronic categories: leptonic ttH are the events containing any additional leptons while hadronic ttH are the events with no additional leptons. Events that do not fit the ttH category are subsequently divided among the other three channels.

Events without any b-tagged jet but containing additional leptons fall into the VH

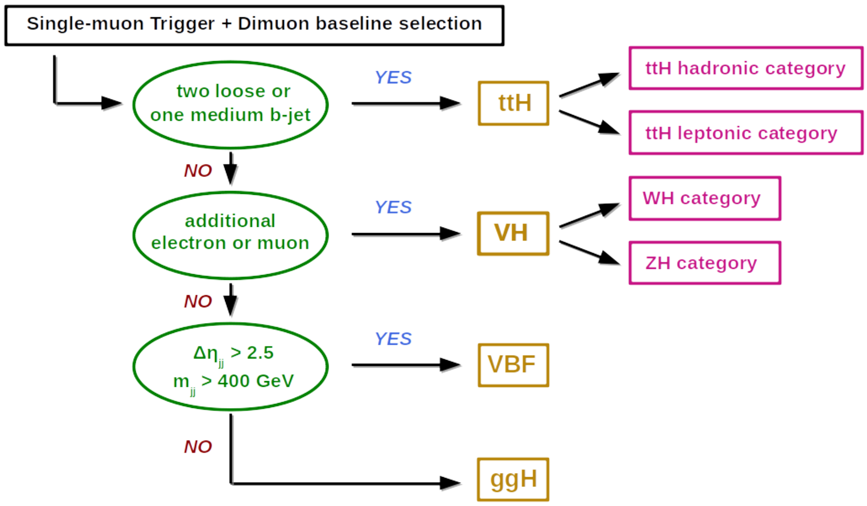


FIGURE 4.2: Summary of the logical definition of each category from the baseline selection.

category. VH category is split into two subsets: ZH and WH. VH events containing two additional muons or two additional electrons are labeled as ZH, while events with one additional lepton only are in the WH category.

Events with two hadronic jets with momenta $p_T > 35$ GeV for the leading jet and $p_T > 25$ GeV for the subleading one, with a pseudorapidity distance $\Delta\eta_{jj} > 2.5$ and an invariant dijet mass $m_{jj} > 400$ GeV are classified into the VBF category. The residual events are about 96% of the initial sample and they fall into the ggH category which is mainly composed of either events with less than two jets or events with no additional leptons. Figure 4.2 shows the tree defining the event categorization.

The VBF, ttH, and VH searches apply additional cuts in order to further reduce the background and obtain a better signal-to-background ratio. Each analysis aims to select and study Higgs bosons produced by a specific process. However, it may also select events coming from the other channels. Table 4.1 shows the number of signal expected events divided by production channels in the four $\mathrm{H} \to \mu\mu$ searches.

Specific multivariate techniques are used in each channel to enhance the signal-to-background separation. Moreover, the events are eventually split into further subcategories in order to improve the final results. The ggH, VH, and ttH searches follow the same analysis strategy consisting of a fully data-driven background estimation from fits to the dimuon mass spectrum. The three analyses are described in the following sections and the final results will be shown in chapter 6.

Higgs production	ggH search	VBF search	ttH search	VH search
ggH	825	20.7	2.8	0.0
VBF	38	26.4	0.3	0.0
ttH	0	0.0	6.7	0.3
VH	27	0.6	1.3	3.7

TABLE 4.1: Expected signal events in the four searches divided by production channel.

4.3 Gluon-fusion channel

In this section, the search in the ggH channel is described. Signal events in this category have no aditional signature beyond the dimuon narrow resonance, therefore they are difficult to identify and distinguish from the background ones. The VH and ttH dedcated searches can rely on much specific signal signatures that help in signal-to-background separation. However, the gluon-fusion cross-section is much larger than the other production channels ones therefore the search in the ggH channel relies on the signal statistical power.

The most important background in the ggH category consists of the Drell-Yan (DY) production of off-shell Z/γ bosons decaying into a muon pair. DY process represents about 90% of the total amount of background. Other sources of background are present even if less relevant. The most significant ones are diboson production (WW, WZ, and ZZ), $t\bar{t}$ events, and single top events where the t-quarks decay to muons.

The most discriminating feature between signal and background is the dimuon mass ($m_{\mu\mu}$), which has a sharp peak around 125 GeV for the signal and it decreases monotonically for the background. However, many other variables can help in signal-to-background separation. The most important among those is the p_T of the dimuon system, which is slightly harder in case of Higgs boson events with respect to the DY events.

The analysis strategy can be summarized in the following steps. A $m_{\mu\mu}$ independent multivariate discriminator is built in order to separate as best as possible signal-like events from background-like ones. Then, the discriminator is used to divide the events in different subcategories. The dimuon mass distributions are used for each subcategory: at first, they are fitted in a mass blinded window in order to describe analytically of signal and background, and lastly, the signal strength is extracted from a simultaneous fit among all the subcategories.

In order to adopt the described strategy, the multivariate discriminant output and the dimuon mass have to be highly uncorrelated. Therefore the variables used in the discriminator have to be uncorrelated with the mass, moreover, the mass has to be uncorrelated with any combination of the input variables.

4.3.1 Multivariate discriminator

The Multivariate discriminator is built using the TMVA package [107]. A Boosted decision tree (BDT) is trained on MC samples separately on each of the three years in order to take into account differences in detector performance and reconstruction. The BDTs are trained using the Gradient Boost method with the same setting. The settings used for the training are shown in table 4.2 Many other settings or methods have been tested but no significant gain has been observed. A complete description of Boosted Decision Trees is available in appendix B.

Category	Trees	Shrinkage	Bagged fraction	node size	Depth	Separation
ggH	1000	0.1	0.5	>3%	4	CrossEntropy

TABLE 4.2: Settings used in training the BDTs of ggH search.

The trainings are performed using half of the available simulated samples in the mass range $115 < m_{\mu\mu} < 135$ GeV. The background samples used in the training consist of DY, $t\bar{t}$, single top, and diboson events, while signal samples include events from ggH, VBF, VH, and ttH productions that pass the ggH category selection.

The kinematic variables used as inputs in building the BDT are several. Four variables related to the two muons are used. Two of these variables are the dimuon transverse momentum $p_T^{\mu\mu}$ and pseudorapidity and $\eta_{\mu\mu}$. In addition, the Collins-Soper variables ϕ_{CS} and $\cos\theta_{CS}$ described in appendix A are used. This set of variables does not contain information about the dimuon mass. Single muon variables are also included: the pseudorapidity of the muons η_{μ_1} and η_{μ_2} and the the transverse momenta relative to the dimuon mass $p_T^{\mu_1}/m_{\mu\mu}$ and $p_T^{\mu_2}/m_{\mu\mu}$ are added as BDT input variables.

Further, in order to enhance signal-to-background separation, variables carrying jet information are added in case jets are present. First of all, the total number of jets in the events is used. Then, whenever the events contain one jet only with $p_T > 25$ GeV, the jet momentum, jet pseudorapidity, and the azimuthal and pseudorapidity distances between the dimuon system and the jet $\Delta\eta(\mu\mu, j)$ and $\Delta\phi(\mu\mu, j)$ are included in the input list.
Whenever the events contain two or more jets with $p_T > 25$ GeV, additional variables that can help to target the presence of VBF and VH signal are considered. The variables related to the dijet system m_{jj}, $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$ of the two leading jets are used as well as the angular separation of the dimuon system with the subleading jet $\Delta\eta(\mu\mu, j_2)$ and $\Delta\phi(\mu\mu, j_2)$. Lastly, the BDT takes as input the Zeppenfeld variable defined by

$$z^\star(H) = \frac{\eta_{\mu\mu} - (\eta_{j_1} + \eta_{j_2})/2}{|\eta_{j_1} - \eta_{j_2}|} \tag{4.1}$$

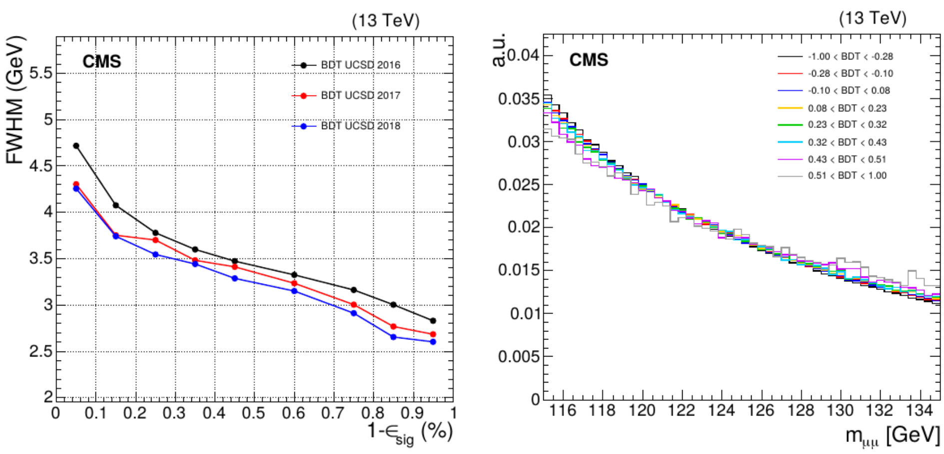
This variable is a linear transformation of the dimuon pseudorapidity, and its value is $|z^\star| < 0.5$ when the dimuon pseudorapidity falls in between the jet pseudorapidities, and it is $|z^\star| > 0.5$ otherwise.

The analysis is very sensitive to the dimuon mass resolution, therefore each event is weighted by $1/\sigma_{\mu\mu}$ in the training. The mass resolution is not used as input in the training thus it is not used in evaluating the BDT and the final score is not an explicit function of $\sigma_{\mu\mu}$. Several tests have been done using also the mass resolution as input variable but no significant changes have been observed: the reason is that the BDT could learn the mass resolution from single muon variables such as $p_T^\mu/m_{\mu\mu}$ and η_μ.

In order to understand the dependence of the mass resolution from the BDT score, the full width half maximum (FWHM) of ggH signal mass distribution is extracted for different cuts in the BDT output from a fit with a double-side crystal ball function (DCB)

$$\mathrm{DCB}(m_{\mu\mu}) = \begin{cases} e^{-(m_{\mu\mu}-\hat{m})^2/(2\sigma^2)} & -\alpha_{\mathrm{L}} < (m_{\mu\mu} - \hat{m})/\sigma < \alpha_{\mathrm{R}} \\ (\frac{n_{\mathrm{L}}}{|\alpha_{\mathrm{L}}|})^{n_{\mathrm{L}}} \times e^{-\alpha_{\mathrm{L}}^2/2} \times (\frac{n_{\mathrm{L}}}{|\alpha_{\mathrm{L}}|} - |\alpha_{\mathrm{L}}| - (m_{\mu\mu} - \hat{m})/\sigma)^{-n_{\mathrm{L}}} & (m_{\mu\mu} - \hat{m})/\sigma \leq -\alpha_{\mathrm{L}} \\ (\frac{n_{\mathrm{R}}}{|\alpha_{\mathrm{R}}|})^{n_{\mathrm{R}}} \times e^{-\alpha_{\mathrm{R}}^2/2} \times (\frac{n_{\mathrm{R}}}{|\alpha_{\mathrm{R}}|} - |\alpha_{\mathrm{R}}| - (m_{\mu\mu} - \hat{m})/\sigma)^{-n_{\mathrm{R}}} & (m_{\mu\mu} - \hat{m})/\sigma \geq \alpha_{\mathrm{R}} \end{cases} \tag{4.2}$$

The DCB consists of a gaussian core defined by the mean $\hat{m}$ and the standard deviation σ. The tails are described by power-low functions with parameters n_{L} and α_{L} for the low mass tail and n_{R} and α_{R} for the high mass tail. The FWHM is shown in Figure 4.3a for each cut in the BDT score. The monotone trend demonstrates that the BDT has learned the importance of the mass resolution.



(a) (b)

FIGURE 4.3: (a) FWHM of the dimuon mass distributions as a function of the BDT cut. (b) Dimuon mass distribution for background events in specific BDT score ranges.

Two tests are performed in order to confirm that the BDT does not contain any information about the signal peak. One is comparing the BDT output for gluon-fusion simulated events at different Higgs mass hypotheses. The distributions are compatible for masses within the range $120 < m_H < 130$ GeV. The other one is comparing the background dimuon mass shapes as a function of the BDT score, as shown in Figure 4.3b. The BDT bias does not sculpt the mass distribution either at low score or high score, thereby demonstrating that the BDT output is not sensitive to the dimuon mass.

Given such a BDT, a cut in its output score would have two main effects. Firstly, the cut would increase the signal-to-background ratio. Then, the selected signal sample would have a better mass resolution than the initial sample. The improved resolution would lead to a better signal extraction.

4.3.2 Categorization

At this stage, a further categorization of the events belonging to the ggH phase space is performed using the BDT output. The adopted procedure is iterative and it is applied to the sum of the signals and backgrounds from the three years.

An initial fit of the dimuon mass distribution is performed using all the events in the range $110 < m_{\mu\mu} < 150$ GeV, and the expected significance is extracted from the fit. The signal is fitted with the DCB, while the background is modeled with a modified Breit-Wigner function (mBW) [47]

$$\mathrm{mBW}(m_{\mu\mu}; m_Z, \Gamma_Z, a_1, a_2, a_3) = \frac{\Gamma_Z\, e^{a_2 m_{\mu\mu} + a_3 m_{\mu\mu}^2}}{(m_{\mu\mu} - m_Z)^{a_1} + (\Gamma_Z/2)^{a_1}} \tag{4.3}$$

where the parameters a_1, a_2, and a_3, are free to float and, $m_Z = 91.19$ GeV and $\Gamma_Z = 2.49$ GeV are the Z boson mass and width [106].

Then, the BDT output is scanned in quantiles of 1% of signal efficiency from low to high score. For each quantile, the events are divided in two sets, therefore 99 pairs of sets are considered (0% and 100% are excluded). For each possible pair of sets, a simultaneous fit in the two sets is performed and the significance is extracted. The optimal position of the subcategory boundary is chosen as the one that provided the best significance. This procedure is repeated from two subcategories to three and so on, in order to define additional subcategory boundaries. The iteration is stopped when a further split would not increase the significance of at least 1%.

Five subcategories are defined by this iterative procedure and they are labelled "ggH-cat1", "ggH-cat2", "ggH-cat3", "ggH-cat4", "ggH-cat5", corresponding to 0-30, 30-60, 60-80, 80-95, 95-100 signal efficiency quantiles. The "ggH-cat1" subcategory contains the events with the lower BDT scores while the "ggH-cat5" subcategory contains the events with the higher. In order to verify the robustness of the subcategory boundaries and of the method used, the same procedure is repeated with a signal sample produced by a different generator. The same number of boundaries is found and the ranges are in agreement within 2%, suggesting that they are not related to a particular generator.

The boundaries and the BDT output are shown in Figure 4.4. The distributions of all the backgrounds are stacked and the signals distributions are shown separately.

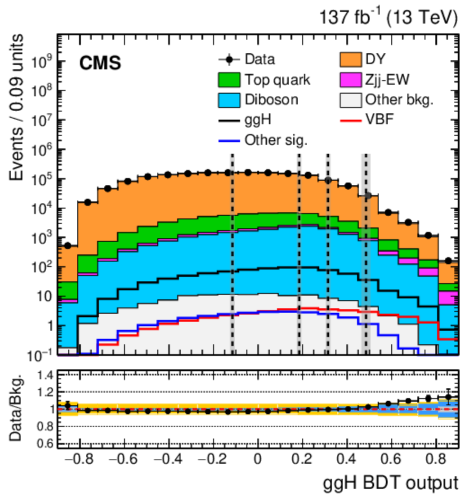


FIGURE 4.4: BDT output distribution for signal and background for the events passing the ggH category selection with $110 < m_{\mu\mu} < 150$ GeV. The grey vertical bands indicate the range between the minimum and maximum BDT score considered to define the boundaries. The lower panel shows the ratio between data and the expected background. The grey band indicates MC statistical uncertainties, while the azure and orange bands represent the experimental and theoretical uncertainties.

The lower panel shows the ratio between data and backgrounds with the associated uncertainties.

4.3.3 Signal extraction

The signal is extracted by a simultaneous maximum-likelihood fit to the dimuon mass spectrum in the five defined subcategories. Signal and background distributions are modeled via analytical functions and the overall normalizations are extracted from the fit.

The mass distribution of the gluon-fusion Higgs is modeled with a DCB in each subcategory. Few residual Higgs events from the other three production modes are present in every subcategory. The distributions of VBF, VH, and ttH productions are also modeled with DCBs in each subcategory. The DCBs parameters are determined from simulation and they are not allowed to vary in the final fit. Instead, the mass $\hat{m}$ and the resolution σ parameters are constrained to be the same for all the DCBs are they are allowed to float in the fit.

There is not any theoretical principle that states the best parametric function for the background description, therefore several functions have been taken into consideration. The adopted description tries to maximize the expected sensitivity while introducing a negligible bias and it is referred to as the *core-pdf method*.

The core-pdf method has been chosen because of the following two features of the background. The background composition is constant across the BDT output, therefore the background composition is the same for all the subcategories. On the other hand, the kinematic properties of the events changes along with the BDT score, and the background shape of the dimuon mass is not exactly the same in all the subcategories. The core-pdf method consist of modeling the background with a core function common to all the subcategories multiplied by a different function for each subcategory. The parameters of the core function are extracted from a simultaneous fit across all the categories. The core function is multiplied by a third-order polynomial in ggH-cat1 and ggH-cat2, while it is multiplied by a second-order polynomial in ggH-cat3, ggH-cat4, and ggH-cat5. The parameters of the polynomials are fully uncorrelated among the subcategories. The core function is obtained by an envelope of three functions: the modified Breit-Wigner defined in Eq. 4.3, a sum of two exponentials, and the product of a third-order Bernstein polynomial and a non-analytical shape. The non-analytical shape is obtained by DY simulation events generated at NNLO precision in QCD and NLO precision in EW theory.

The parameters of the functions obtained with the core-pdf method are directly constrained in the S+B fit, therefore by the observed data. The background contribution under the peak is estimated from the signal-free sidebands and its uncertainty is mainly affected by the statistical power of the mass sidebands.

In principle, any parametric function used to model the background can bias the result. The uncertainty due to the choice of the parametric function is estimated by the following strategy. A background-only fit is performed in each subcategory using several different functions. The post-fit shapes are used to generate pseudodata sets for all event subcategories. Then, the pdf-core method is used to fit the signal-plus-background pseudodata and extract the injected signal. The median difference between the extracted and the injected signal gives a measure of the bias due to the choice of the background model. The bias estimated from this method would change the total uncertainty of less than 1%, and therefore it is neglected.

Experimental and theoretical uncertainties are common among all the channel searches. In order to avoid repetition, a complete description of error treatment is provided in chapter 5. Figure 4.5 shows the data dimuon mass distribution and the post-fit function used to describe signal and background in each subcategory.

4.4 ttH channel

The event selection of ttH search starts with the dimuon selection and with the b-tagging requirement explained in section 4.2. The presence of tagged b-jets ensures mutual exclusivity with the other searches and reduces the background arising from DY and diboson events. Moreover, the oppositely charged muons must have an invariant mass between 110 and 150 GeV. The background at this point is mainly composed by dileptonic $t\bar{t}$ events.

Events are divided in two categories: ttH hadronic and ttH leptonic. The leptonic category contains events with one or two additional muons or electrons. In case two additional leptons are present, they must have opposite charge, they must have $p_T > 20$ and they must have $|\eta| < 2.4$ if muons or $|\eta| < 2.5$ if electrons. In addition,

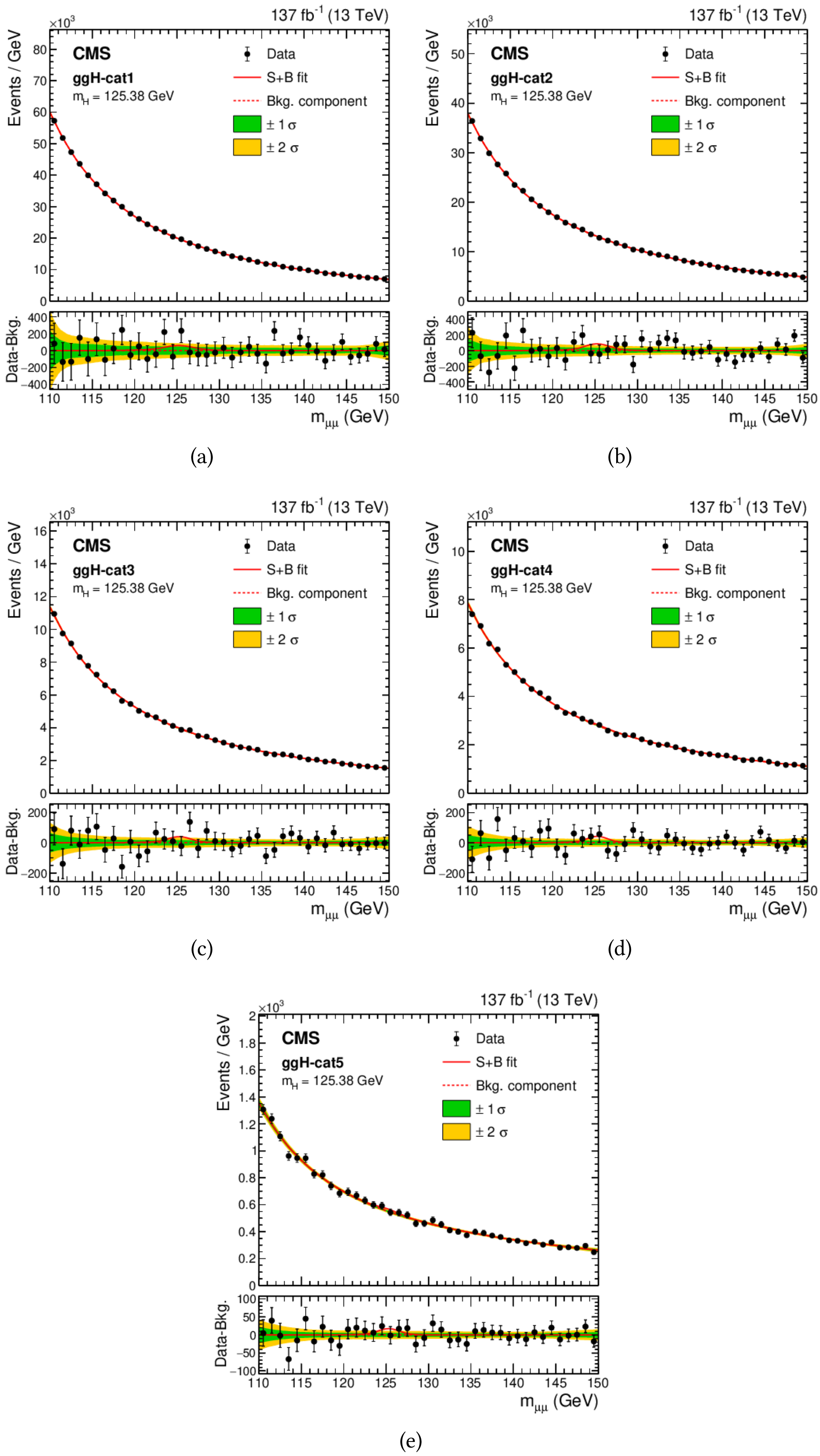


FIGURE 4.5: Dimuon mass spectrum in all the ggH subcategories, the background has been extracted from the fit. The lower panel shows the residual background subtraction and the red line represents the post-fit signal.

their invariant mass has to be greater than 12 GeV and outside the Z mass window $|m_{\ell\ell} - m_Z| > 10$ GeV. Whenever more than one oppositely charged muon pair have invariant mass within 110 and 150 GeV, the pair with the highest transverse momentum is chosen as the Higgs candidate. Moreover, ttH leptonic events must have at least two jets with $p_T > 25$ GeV and $|\eta| < 4.7$, and the leading one must have $p_T > 35$ GeV.

Events with no additional leptons fall in the ttH hadronic category. Events are further reduced by requiring at least three jets with $p_T > 25$ GeV and $|\eta| < 4.7$ and the leading one must have $p_T > 50$ GeV. Moreover, at least one combination of three jets must have an invariant mass within 100 and 300 GeV. Table 4.3 shows the selection requirements of hadronic and leptonic categories.

Observable	ttH hadronic	ttH leptonic
DeepCSV score	> 0 medium or > 1 loose	
Number of leptons	2	3 or 4
Jet multiplicity	≥ 3	≥ 2
Leading jet p_T	≥ 50 GeV	≥ 35 GeV
Jet triplet mass	100 GeV $< m_{jjj} <$ 300 GeV	-
Z mass veto	-	$\|m_{\ell\ell} - m_Z\| > 10$ GeV
Low mass resonance veto	-	$m_{\ell\ell} > 12$ GeV

TABLE 4.3: Summary of the event selection criteria for the ttH hadronic and leptonic categories.

The same strategy of ggH channel is applied in this search. Firstly, a BDT is built in order to separate signal from background in each category, then, the BDT score is used to further divide the events in subcategories. Lastly, the dimuon mass spectrum is fitted simultaneously among all the categories.

4.4.1 Multivariate discriminants

Distinctive kinematic properties of ttH process help to distinguish signal from background. Two different BDTs are trained: one for ttH hadronic and one for ttH leptonic. The main backgrounds in the two categories are used in the trainings and they are $t\bar{t}$ and DY+Jets for the hadronic category, while $t\bar{t}$, $t\bar{t}Z$, $t\bar{t}W$, and WZ for the leptonic one.

In order to avoid bias in the final fit, all the chosen variables are highly uncorrelated with the dimuon mass. The input variables can be divided in three families: the ones containing information about the dimuon system, the one describing the hadronic activity, and the one related to the t-quarks decay products. The majority of the observables used as input are common between the two BDTs, then, dedicated variables are added to exploit the particular features of the specific categories.

Common inputs comprehend muons kinematic variables. The candidate Higgs is described by the dimuon transverse momentum $p_T^{\mu\mu}$, the pseudorapidity $\eta_{\mu\mu}$, and

the Collins-Soper variables ϕ_{CS} and $\cos\theta_{CS}$. Additionally, the pseudorapidities η and the transverse momenta relative to the dimuon mass $p_T^{\mu}/m_{\mu\mu}$ of the two muons are considered.
The hadronic activity is described by the p_T and η of the three leading jets in the event, the number of selected jets, and the maximum score of DeepCSV of jets not overlapping with leptons ($\Delta R(\ell, j) > 0.4$). Moreover, several global variables are included. The scalar (H_T) and vectorial ($|\vec{H}_T^{miss}|$) sum of the transverse momenta of all leptons and jets are added as well as the missing transverse momentum p_T^{miss}. Also the projection of p_T^{miss} on the bisector of the dimuon system, called $\Delta\zeta$ [108], is added.

The variables describing the kinematic of the t-quarks are different between the two categories. The leptonic category considers the azimuthal angle between the dimuon system and the p_T-leading additional lepton. The invariant mass between the p_T-leading additional lepton and the jets with the highest DeepCSV score, and the transverse invariant mass between the p_T-leading additional lepton and p_T^{miss} are included too. Moreover, also the number of additional electrons is used as input.

In ttH hadronic events, identifying the jets produced from t-quark decay is not straightforward. A dedicated algorithm has been developed in order to identify t-quark decay, called resolved hadronic top tagger (RHTT). For each event, among all combination of three jets with $100 < m_{jjj} < 300$ GeV, the one with the highest score of RHTT is selected and the following variables are built and used in the BDT: the p_T of the t-quark candidate, the p_T balance of the t-quark candidate and the dimuon pair, and the RHTT score of the jet triplet.

The trainings are performed in the mass window $115 < m_{\mu\mu} < 135$ GeV and 2016, 2017, and 2018 samples are combined in order to improve the statistical power. Moreover, the events are weighted as the inverse of the dimuon mass resolution as done in the ggH search. The settings used for the trainings are shown in table 4.4 for leptonic and hadronic categories. Both the BDTs show strong separation between signal and backgrounds.

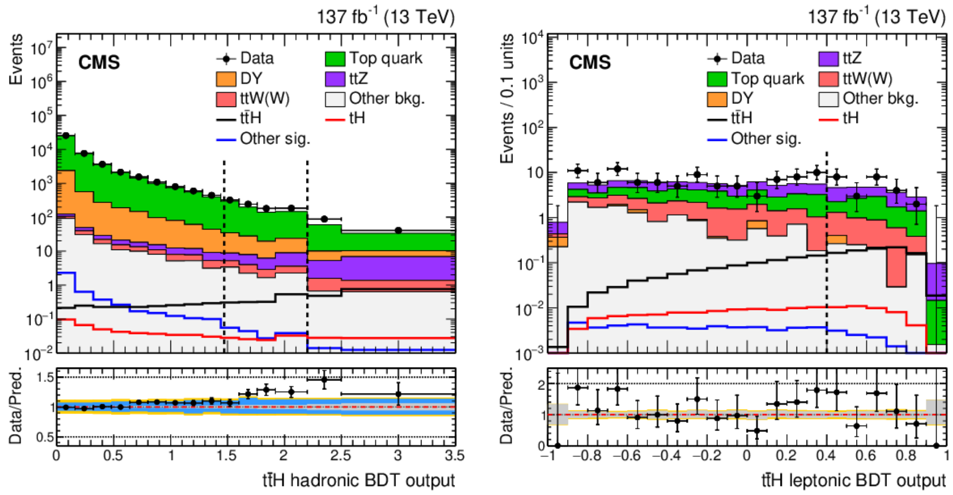
Category	Trees	Shrinkage	Bagged fraction	node size	Depth	Separation
ttH leptonic	200	0.1	0.5	>5%	3	CrossEntropy
ttH hadronic	600	0.1	0.5	>5%	3	CrossEntropy

TABLE 4.4: BDT settings of the ttH leptonic and ttH hadronic BDTs.

4.4.2 Categorization and signal extraction

The events are divided in subcategories applying the iterative procedure used in the ggH search. The procedure is applied with simulated events in the range $110 < m_{\mu\mu} < 150$. During this procedure, the signals dimuon distributions are fitted separately with DCBs while the backgrounds ones are modeled with exponential functions in each category.

The hadronic category is divided in three subcategories called "ttHhad-cat1", "ttHhad-cat2", and "ttHhad-cat3", corresponding to 0-70, 70-86, 86-100 signal efficiency quantiles respectively. The leptonic category is divided in two subcategories at 52% signal efficiency and they are labeled "ttHlep-cat1", and "ttHlep-cat2". Figure 4.6 shows the BDT output distributions, the various backgrounds are stacked and the signals are shown separately. The figure also shows the category boundaries in ttH leptonic and ttH hadronic category.



(a) (b)

FIGURE 4.6: BDT output distribution for data, signal, and background for the events with $110 < m_{\mu\mu} < 150$ GeV in ttH hadronic (a) and ttH leptonic (b). The dashed vertical lines indicate the range between the minimum and maximum BDT score considered to define the boundaries. The lower panel shows the ratio between data and the expected background and its description is the same as in Figure 4.4.

Signal extraction is performed with a data-driven approach as done in ggH search. The signal shapes are fitted with DCBs for each production mode in each subcategory. The background is modelled with a second order Bernstein polinomial in "ttHhad-cat1" and "ttHhad-cat2", a sum of two exponentials in "ttHhad-cat3", and a single exponential in "ttHlep-cat1" and "ttHlep-cat2".

Since the underlying model of the background spectrum is not known, the choice of the functions used to describe it may lead to a bias in the signal strength extraction. In order to estimate the bias, the same strategy of ggH search is adopted. Pseudodata sets are generated using different background models. Then, the signal is extracted with a signal plus background fit and the comparison between the injected and extracted signal provides a measurement of the bias. The chosen background functions maximize sensitivity while keeping a neglectable bias.

Figure 4.7 shows the dimuon mass distributions in each ttH subcategory. The post-fit functions used to model signal and background are shown.

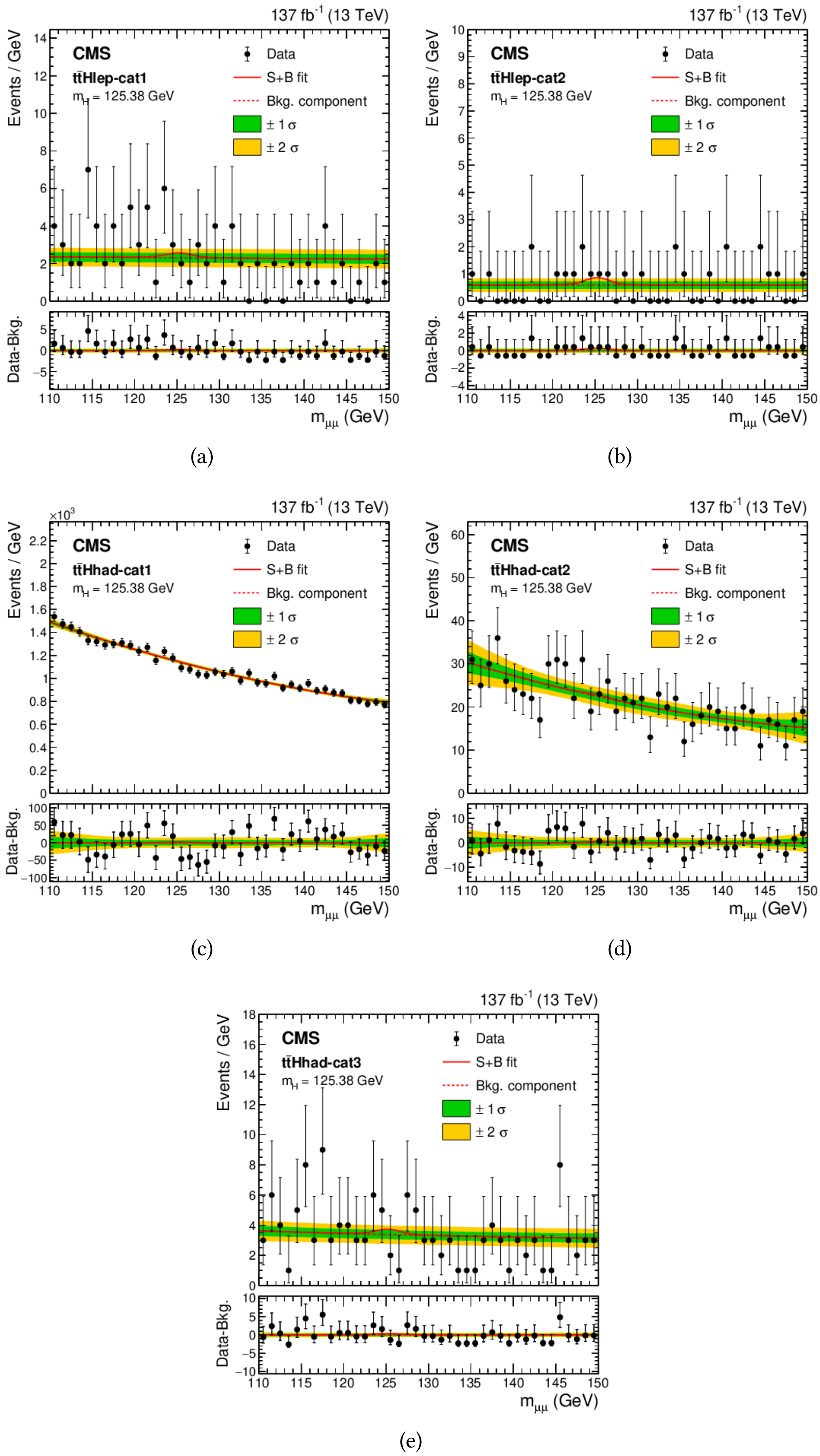


FIGURE 4.7: Dimuon mass spectrum in all the ttH leptonic (a,b) and hadronic (c,d,e) categories, the background has been extracted from the fit. The lower panel shows the residual background subtraction and the red line represents the post-fit signal.

4.5 VH channel

VH search starts with the selection described in section 4.2. The b-jets veto excludes events that enter in ttH channel, moreover, it reduces the $t\bar{t}$ background. Also, the additional leptons requirement have the same purpose: it reduces the background and it ensures mutual exclusivity with the other searches. Moreover, at least one pair of oppositely charged muon must have $110 < m_{\mu\mu} < 150$. The events are divided as a function of the number of additional leptons: WH category contains events with one additional lepton while ZH category contains events with two same flavor oppositely charged additional leptons. In each category, several other dedicated cuts are performed in order to further reject backgrounds while not significantly reducing the signals.

In WH category, if the additional lepton is a muon, none of the two oppositely charged muon pairs must have $m_{\mu\mu} < 12$ GeV or $81 < m_{\mu\mu} < 101$ GeV in order to suppress quarkonium and WZ backgrounds. Moreover, if both the two muon pairs have the invariant mass in the range 110-150 GeV, the one with higher transverse momentum is chosen as Higgs candidate.

In the ZH category, equivalently to the WH category, any same flavor oppositely charged lepton pair must have $m_{\mu\mu} < 12$ GeV. In addition, the events are rejected if they contain two distinct muon pairs with $m_{\mu\mu}$ in 81-101 GeV in order to reduce ZZ background. If the additional leptons are electrons, their invariant mass is required to be in 70-110 GeV. If the additional leptons are muons, their invariant mass must be in 81-101 GeV. If, after the described selection, more than one muon pair has invariant mass compatible with the Z boson, the combination with the mass closer to 91 GeV is taken as Z candidate.

The main backgrounds of WH category are WZ production, ZZ process where one lepton is not reconstructed and DY with associated lepton production. In particular, in WZ events the muon pair comes from an off-shell Z boson decay and its kinematic is very much similar to the signal one. The main backgrounds of ZH category are the ZZ and ggZZ to 4 leptons. All the other contributions are negligible in this channel.

The VH channel follows the same analysis strategy described in the previous analyses. Within this baseline selection, the events are further divided in subcategories by multivariate discriminators. Finally, the dimuon mass spectrum is fitted simultaneously using analytical functions to describe signals and backgrounds.

4.5.1 Multivariate discriminants

One BDT is trained in each category. The input variables are selected in order to build BDTs not significantly correlated with the dimuon mass. The kinematic variables that discriminate most signal from backgrounds in the VH phase space are the angles between the leptons in the events. The majority of inputs used are not in common between the two BDTs and they are mentioned below. The only exceptions are the transverse momentum and the pseudorapidities of the dimuons system. Moreover, in both categories, a flag is used to take into account the flavor of the additional leptons.

In the WH category, BDT includes $\Delta\eta$ and $\Delta\phi$ between the additional lepton ℓ and the candidate Higgs, and between the additional lepton and both muons. The transverse momentum of the additional lepton $p_T(\ell)$ is also considered. In addition, the transverse mass and the azimuthal separation between the additional lepton and the $\vec{H}_T^{miss}$ are added as inputs.

In the ZH category, the azimuthal difference of the muons $\Delta\phi(\mu_1, \mu_2)$ is added. Kinematic variables related to the Z boson are also considered. In particular, the mass $m_{\ell\ell}$, the transverse momentum $p_T^{\ell\ell}$, the pseudorapidity $\eta_{\ell\ell}$ and the angular distance of the two leptons $\Delta R(\ell_1, \ell_2)$ are included. Moreover, the pseudorapidity difference and the cosine of the Collins-Soper angle between Z and Higgs boson candidates $\Delta\eta(\mu\mu, \ell\ell)$ and $\cos\theta_{CS}(\mu\mu, \ell\ell)$ are added as inputs.

Differently from the previous analysis, the BDTs are trained in the mass window $110 < m_{\mu\mu} < 150$ GeV, and signal samples with different mass assumptions m_H = 120, 125, 130 GeV are all used to improve the signal statistical power. In both the categories, all years samples are combined and used in the same training. Moreover, the events are weighted with the inverse of the dimuon mass resolution as described for ggH and ttH searches. The settings used for the trainings are shown in table 4.5 for ZH and WH categories.

Category	Trees	Shrinkage	Bagged fraction	node size	Depth	Separation
ZH	400	0.1	0.5	>5%	3	CrossEntropy
WH	500	0.1	0.5	>5%	4	CrossEntropy

TABLE 4.5: BDT settings of the ZH and WH BDTs.

4.5.2 Categorization and signal extraction

The subcategories optimization strategy adopted by ggH and ttH searches is applied to ZH and WH too. The signals are fitted with DCBs where the parameters n_L and n_R are kept fixed at value 2.0 since they float in a large range without changing the quality of the fit. In the procedure, the background estimation is performed using the exponential modified Breit-Wigner function (BWZ) for each category. The BWZ function is defined by:

$$\mathrm{BWZ}(m_{\mu\mu}; m_Z, \Gamma_Z, a) = \frac{\Gamma_Z \, e^{a m_{\mu\mu}}}{(m_{\mu\mu} - m_Z)^2 + (\Gamma_Z/2)^2} \tag{4.4}$$

The procedure divides WH in "WH-cat1", "WH-cat2", and "WH-cat3" while ZH in "ZH-cat1", and "ZH-cat2". The boundaries correspond to 0-22, 22-70, and 70-100 efficiency quantiles for WH and to 0-52, 52-100 efficiency quantiles for ZH. Figure 4.8

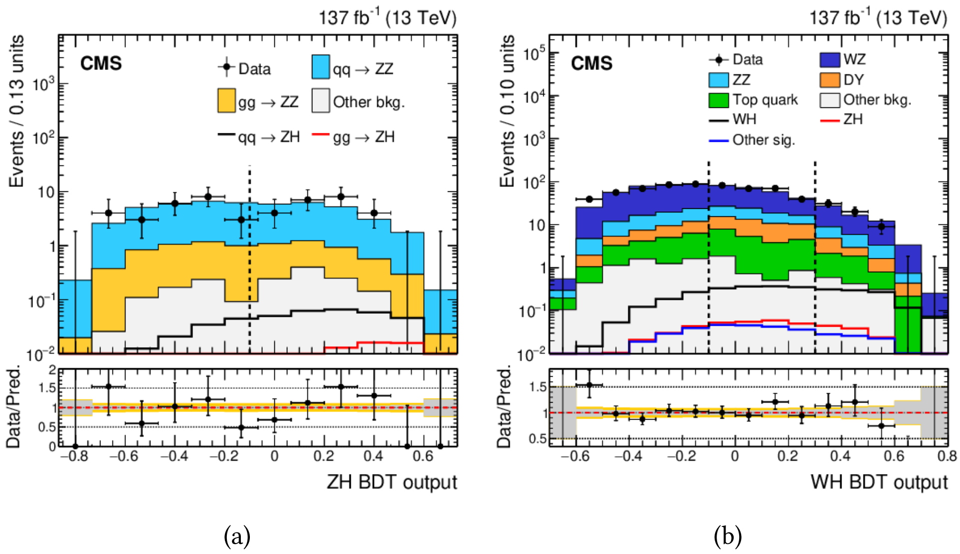


FIGURE 4.8: BDT output distribution for data, signal, and background for the events with $110 < m_{\mu\mu} < 150$ GeV in ZH hadronic (a) and WH leptonic (b). The dashed vertical lines indicate the range between the minimum and maximum BDT score considered to define the boundaries. The lower panel shows the ratio between data and the expected background and its description is the same as in Figure 4.4.

shows the BDT output distributions for data and MC. The vertical dashed lines indicate the positions of the boundaries.

The signal is extracted performing a simultaneous binned maximum-likelihood among all the categories. The signals are modeled with DCBs while the backgrounds are estimated by BWZ except for WH-cat1 where BWZγ defined by

$$\mathrm{BWZ}\gamma(m_{\mu\mu}; m_Z, \Gamma_Z, a, f) = f \frac{\Gamma_Z \, e^{a m_{\mu\mu}}}{(m_{\mu\mu} - m_Z)^2 + (\Gamma_Z/2)^2} + (1-f)\frac{e^{a m_{\mu\mu}}}{m_{\mu\mu}} \tag{4.5}$$

is used.

The potential bias arising from the choice of the parametric functions are estimated with the same method used in ggH and ttH analysis. Data are fitted with several background models. The analytical functions describing the background that are obtained from the fit are used to generate pseudodata sets. These pseudodata sets are fitted with different background descriptions and they are used to extract the signal. The median difference between the measured and injected signal yields estimates the bias due to the used parametric model. The chosen functions maximize the sensitivity while keeping the bias negligible.

Figure 4.9 shows data distributions of the dimuon mass in each category along with the post-fit functions describing signals and backgrounds.

Uncertainties and results of the three analyses will be presented in the next chapters.

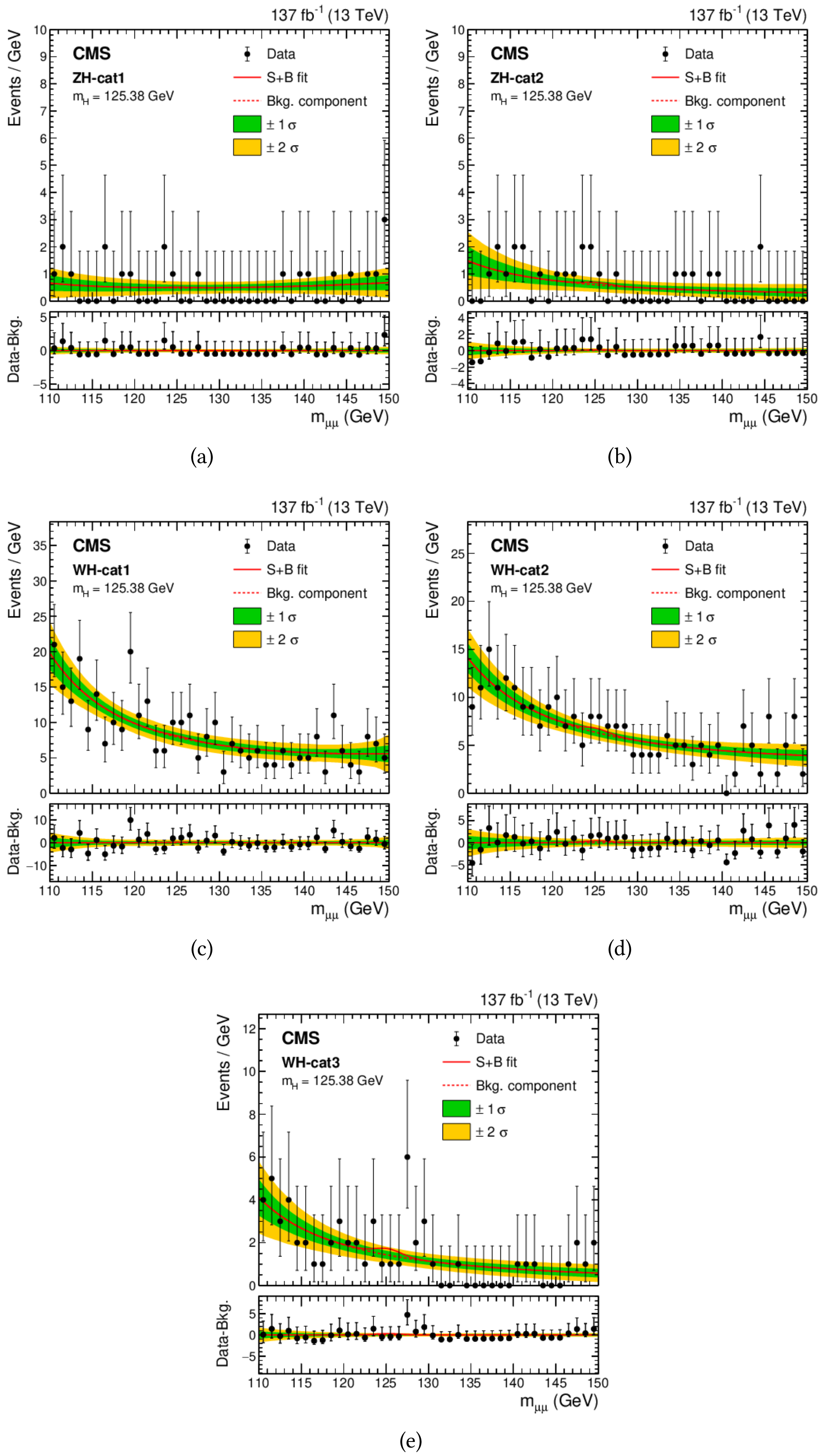


FIGURE 4.9: Dimuon mass spectrum in ZH (a,b) and WH (c,d,e) categories, the background has been extracted from the fit. The lower panel shows the residual background subtraction and the red line represents the post-fit signal.

Chapter 5

The H→ $\mu\mu$ search in the VBF channel

In this chapter, the Higgs search in the VBF production channel is described in details. The analysis starts with a loose selection, orthogonal to the searches described in the previous chapter. Then, two multivariate techniques are tried to separate signal from background: the Boosted Decision Tree and the Deep Neural Network. The discriminator building is described and an overview of the expected performance is provided. Finally, the systematic uncertainties are discussed.

5.1 Introduction

The $H \rightarrow \mu\mu$ dedicated search in the VBF production channel completes the description of the Higgs-to-muon coupling measurement. The VBF channel provides the most important contribution in the combined result of the four analyses and it is the one where my personal effort focused.

The VBF cross-section is much smaller than the ggH one, in fact, only the 7% of Higgs events are coming from VBF production and about 215 events of VBF Higgs are expected by CMS collisions in the years 2016, 2017, and 2018 in the H→ $\mu\mu$ channel. On the other hand, the VBF channel has a signature that can be exploited in signal-to-background separation. In the VBF production, two partons emit a weak vector boson each thus they recoil with a transverse momentum of about half of the boson mass. On average, the initial momentum of the partons is much larger than the recoil therefore they are detected as jets at high rapidity in the two opposite endcaps. Moreover, since there is no exchange of color between the two partons, additional hadron activity is suppressed in VBF events. Finally, The two weak bosons interact and produce a Higgs boson that is usually at central pseudorapidities, in between the two jets.

The Higgs boson decays promptly, therefore in Higgs events, the trajectories of the muons are compatible with the proton-proton collision vertex. The the invariant mass of the Higgs decay products have to be compatible with the Higgs boson mass, therefore the main feature of the H→ $\mu\mu$ events is the presence of two prompt muons with an invariant mass close to 125 GeV. The signal mass peak in the dimuon mass distribution can be exploited to perform a fully data-driven search of

VBF Higgs similarly to the ggH, VH, and ttH searches. Nevertheless, a different approach is chosen for the VBF analysis.

The search for H→ $\mu\mu$ via VBF is performed following a fully MC based strategy. Monte-Carlo simulations can model the signal and the relevant backgrounds with great accuracy, therefore the background estimation is entirely derived from MC. As opposed to the fully data-driven analysis, the MC based approach has the advantage that the background estimation is not affected by the data statistical uncertainty. The Higgs search via VBF production with the data-driven approach has been performed and it is described in D, however it provides an expected significance about 20% less than the MC based approach.

The fully MC based signal extraction follows the strategy adopted to study and observe the vector boson fusion production of the Z boson decaying into leptons with 7 TeV [100], 8 TeV [109], and with a greater precision with 13 TeV [110]. Similar measurement of single leptons plus two jets have been carried out to confirm the actual existence of the production of W boson via VBF production (VBF W) with 8 TeV [111] and 13 TeV [112]. A fully MC based strategy have been adopted also in the search of H→ $b\bar{b}$ via VH production channel [39].

The strategy is simple and straightforward. A first loose selection is applied in order to reduce the background while keeping high the signal efficiency. The selected events are used to train a multivariate discriminator that separates best signal from background. Then, the distribution of the discriminator output is fitted to extract the signal strength. In this search, two different multivariate techniques are tested: the boosted decision tree and the deep neural network. The final fit is performed with the discriminator that provides better expected results.

Both the selection and the multivariate discriminants exploit the characteristic of the VBF process to improve the signal-to-background separation. Precisely for this reason, the VBF production of Z boson (VBF Z) is the background that is most difficult to distinguish from the signal. The VBF Z background consists in the production of an off-shell Z boson via the triple gauge boson vertex and the topology of the diagram is exactly the same as the one of the VBF Higgs production. Figure 5.1c shows the Feynmann diagram of VBF Z production.

Together with VBF Z, also ditop and Drell-Yan (DY) events are irreducible backgrounds for the VBF H→ $\mu\mu$ search. The Drell-Yan production of muon pairs is the most abundant background in this search. The two jets that take the place of the VBF jets may arise either from the radiation of a parton or from pile-up events. The $t\bar{t}$ events are a less relevant background because it is suppressed by the cut $m_{jj} > 400$ GeV and by the low branching fraction of W into muons. Figure 5.1 shows an example of the Feynmann diagrams of DY and $t\bar{t}$ events.

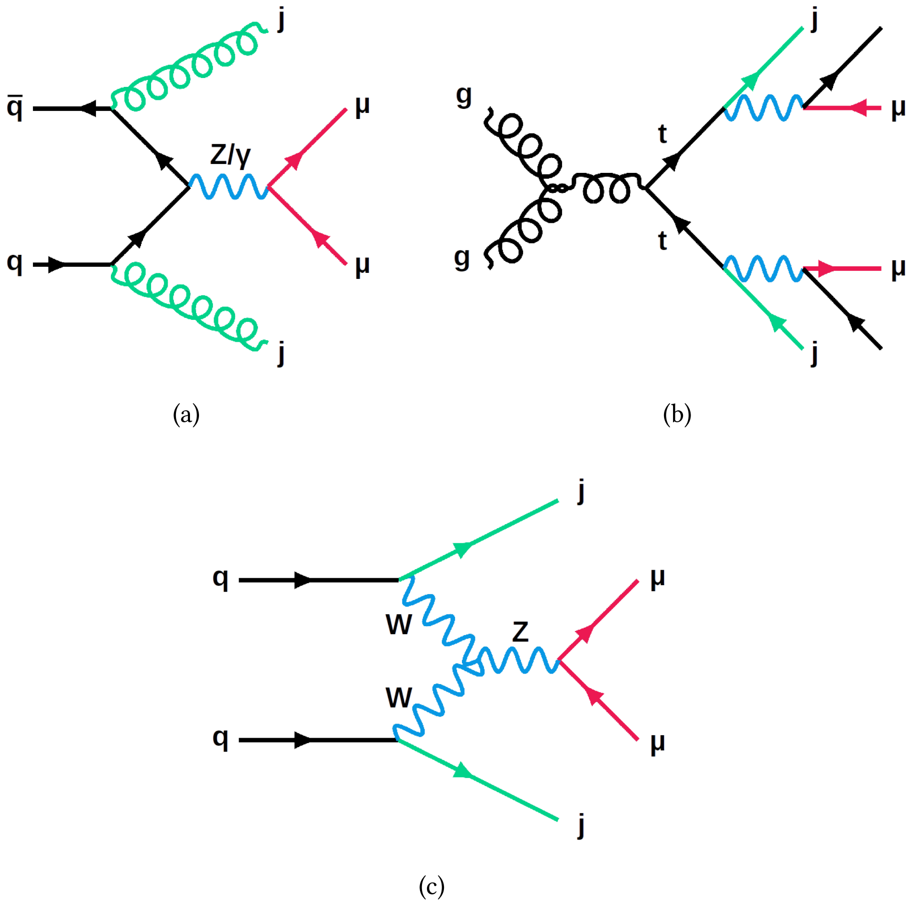


FIGURE 5.1: Feynmann diagrams of the three main backgrounds: Drell-Yan (a), ditop (b), and Z production via vector boson fusion (c). The green fermions represent the two quarks detected as jets, the red fermions are the two muons, and the internal blue lines indicates the electroweak gauge bosons.

5.2 Event simulation

Simulated events are extremely important to study the kinematic features of signal and background process, as well as to optimize the analysis strategy or understand detector effects. Moreover, simulated samples are used to build the multivariate discriminants that help to isolate the Higgs boson from the backgrounds. The matrix element generators used to produce MC events are POWEG v2 [113, 114, 115] and MADGRAPH5_AMC@NLO [116]. Instead, the simulation of the hadronization and of the underlying event is performed by PYTHIA (v8.2 or grater) [117] or HERWIG [118]. Whenever MADGRAPH5_AMC@NLO is used, jets from the matrix element calculations are matched to the parton shower following the FxFx (MLM) prescription [119, 120]. The PYTHIA generator is used with CUETP8M1 tune [121] for the 2016 data-taking period and CP5 tune [122] for 2017 and 2018, while HERWIG is used with UE5C tune [121] for 2016 and CH3 tune [123] for 2017 and 2018. The simulations use NNPDF 3.0 (3.1) parton distribution functions (PDF) for 2016 (2017/2018) data-taking period [124, 125].

In all simulated events, the interactions of all final state particles with the CMS detector are simulated with GEANT4 [126] and the events are reconstructed with the same algorithms that are used for data. The pile-up interactions are produced by overlaying elastic proton-proton collisions on the generated event. The MC events are weighted in order to match the distribution of the number of pile-up interactions per bunch crossing observed in data.

5.2.1 Signal simulation

The MC samples for the ggH process are simulated at next-to-leading order (NLO) accuracy in QCD with two generators: POWEG and MADGRAPH5_AMC@NLO. In both cases, up to two additional partons are included in the final state in the matrix element calculation. The parton shower is simulated by PYTHIA for both the ggH samples. Generated events are reweighted to match the distribution of the p_T of the Higgs boson with the POWEG NNLOPS predictions [127, 128]. The reweighting is performed in bins of jet multiplicity, where jets are required to have $p_T > 30$ GeV at generator level.

The VBF process has peculiar features: two jets with large separation and suppressed hadron activity in between them. The parton shower generators PYTHIA and HERWIG show a significant difference in modeling these features. In fact, the standard p_T-ordered shower implemented by PYTHIA may lead to unphysical additional hadron activity [129]. In order to have a correct description of the kinematics of VBF signal events, two VBF H samples are produced at NLO precision in QCD using POWEG [130] for the matrix element generation and HERWIG or with PYTHIA in dipole shower mode for the parton shower. Those two samples are used to extract the final results, while an additional sample generated by POWEG and hadronized by PYTHIA is used to train the multivariate discriminators.

VH and ttH samples are produced with POWEG [131, 132] at NLO accuracy in QCD interfaced with PYTHIA for parton showering. For all processes, table 5.1 shows the generators used, the cross-section, and the equivalent luminosity that indicates the statistical power of the sample. The signal samples of the four channels are generated with a Higgs mass m_H = 125 GeV and in addition, simulated samples are generated with MadGraph5_amc@nlo for ggH and VBF modes with two additional mass hypoteses: m_H = 120 GeV and m_H = 130 GeV.

The cross-section and $\mathcal{B}(\mathrm{H} \to \mu\mu)$ values are taken from the recommendations of Ref. [20]. The cross-section of ggH process is computed with next-to-next-to-NLO ($\mathrm{N^3LO}$) precision in QCD and NLO precision in electroweak (EW) theory [133]. The VBF [134] and the qq→VH [135] cross-sections are calculated at next-to-NLO (NNLO) in QCD and NLO in EW theory while ttH cross-section is computed at NLO accuracy in QCD and EW theory [136, 137]. The H→ $\mu\mu$ branching fraction is computed with hdecay [138] at NLO accuracy in QCD and EW theory.

sample	generator + PS	$\int \mathcal{L}$ / 137 fb^{-1}	$\sigma \times \mathcal{B}$ (fb)
ggH($\mu\mu$)	MadGraph5_amc@nlo + Pythia	980.0	10.57
	Powheg + Pythia	3714	10.57
VBF H($\mu\mu$)	Powheg + Pythia dipole	23991	0.8211
	Powheg + Pythia	37209	0.8211
	Powheg + Herwig	26203	0.8211
$\mathrm{W^+H}(\mu\mu)$	Powheg + Pythia	24889	0.1828
$\mathrm{W^-H}(\mu\mu)$	Powheg + Pythia	52543	0.1159
ZH($\mu\mu$)	Powheg + Pythia	46474	0.192
ttH($\mu\mu$)	Powheg + Pythia	110221	0.1103
ggH($\mu\mu$)(m_H=120)	MadGraph5_amc@nlo + Pythia	927.0	12.652
ggH($\mu\mu$)(m_H=130)	MadGraph5_amc@nlo + Pythia	1410	8.5047
VBF H($\mu\mu$)(m_H=120)	MadGraph5_amc@nlo + Pythia	2247	0.9535
VBF H($\mu\mu$)(m_H=130)	MadGraph5_amc@nlo + Pythia	3111	0.6827

Table 5.1: Summary of the Monte Carlo signal samples used in the analysis by process. The generator and PS simulator used are listed in the second column of the table while third column shows the equivalent luminosity of each sample. The cross sections used in the analysis are reported in the fourth column.

5.2.2 Background simulation

The largest contribution to the background of this search comes from the Drell-Yan process. Several samples of DY events are generated with specific cuts on the dilepton mass at generator level. All the DY samples are generated with MadGraph5_amc@nlo with up to two additional partons in the final state and Pythia is used to simulate the hadronization. Two samples of dimuon production via Drell-Yan process are generated within the mass window 105-160 GeV, one is produced at

NLO precision in QCD and the other at LO in QCD. In order to gain statistics in VBF-like phase space, two samples of dimuon production via Drell-Yan process are generated requiring also that the two leading jets at generator level have $m(qq) > 350$ GeV and distance $\Delta R > 0.4$ with the muons. The additional samples have different precisions too: one has NLO accuracy, the other one has LO accuracy. Dimuon events under the Z peak are also needed to study the kinematic properties of the DY background and to assess the reliability of the simulation. Therefore, dilepton events via DY production are generated with 0, 1 or more jets at generator level and with $m(\ell\ell) > 50$ GeV.

The electroweak production of Z boson in association with two jets is extremely important in this search. As done for the DY, two alternative sets of samples are produced: one with the invariant mass of the two leptons larger than 50 GeV, and one with two muons with $m_{\mu\mu}$ in between 105 and 160 GeV. The samples are produced with MadGraph5_amc@nlo with LO accuracy in QCD with different parton shower generators. The Herwig parton shower is known to better model the VBF Z model without color connection [139] than Pythia therefore it is used as nominal sample in the analysis. The sample showered by Pythia is used to estimate the parton shower uncertainty.

The DY production in association with two jets has the same final state as the VBF Z process, thus the interference between the two processes has to be taken into account in order to correctly model the data. The interference is generated directly with MadGraph5_amc@nlo at LO accuracy and it is showered by Pythia. Table 5.2 shows the DY and VBF Z simulated samples with the generators, the cross-section, and the equivalent luminosity. A column is added to show the cuts at generator level that identify the samples.

The background with the most significant contribution after DY and VBF Z is the $t\bar{t}$ production. Ditop events are generated in the three decay channels separately: leptonic, semileptonic, and hadronic. The most important for the analysis is the fully leptonic channel but little contributions comes also from the semileptonic one. Single t-quark processes are less relevant but not neglectable. Both $t\bar{t}$ and single t-quark processes have been generated with NLO precision in QCD with Powheg and Pythia generators.

Small contributions arise from single W and diboson production. The generators used for these samples are either MadGraph5_amc@nlo or Powheg for matrix element, while Pythia is used in shower simulation. All these minor backgrounds are computed at NLO precision in QCD. Several other background processes have been considered and found to be negligible, therefore no other samples are used in background modeling in this analysis. table 5.3 shows the generators, the cross-sections, and the equivalent luminosity for the described processes.

When possible, cross-sections are computed with higher precision than the event generations. DY process cross-sections are calculated at NNLO precision in QCD and NLO in EW theory with fewz v3.1b2 [140]. The cross-section of $t\bar{t}$ processes

sample	generator selection	generator + PS	$\int \mathcal{L}$ / 137 fb^{-1}	$\sigma \times \mathcal{B}$ (fb)
DY($\mu\mu$)+jets	$105 < m(\mu\mu) < 160$ GeV	MadGraph5_aMC@NLO (NLO) + Pythia	10.56	47170
		MadGraph5_aMC@NLO (LO) + Pythia	18.93	47170.
DY($\mu\mu$)+jets	$105 < m(\mu\mu) < 160$ GeV, $m(qq) > 350$ GeV	MadGraph5_aMC@NLO (NLO) + Pythia	57.80	1770.0
		MadGraph5_aMC@NLO (LO) + Pythia	111.0	1770.0
DY($\ell\ell$)+jets	$m(\ell\ell) > 50$ GeV, 0 jet	MadGraph5_aMC@NLO + Pythia	0.291	462052
DY($\ell\ell$)+jets	$m(\ell\ell) > 50$ GeV, 1 jet	MadGraph5_aMC@NLO + Pythia	0.513	859590
DY($\ell\ell$)+jets	$m(\ell\ell) > 50$ GeV, ≥ 2 jet	MadGraph5_aMC@NLO + Pythia	0.414	338260
VBF Z($\ell\ell$)	$m(\ell\ell) > 50$ GeV	MadGraph5_aMC@NLO + Herwig	12.02	1664.0
		MadGraph5_aMC@NLO + Pythia	24.14	1664.0
VBF Z($\mu\mu$)	$105 < m(\mu\mu) < 160$ GeV	MadGraph5_aMC@NLO + Pythia dipole	580.0	74.86
		MadGraph5_aMC@NLO + Pythia	419.0	50.889
		MadGraph5_aMC@NLO + Herwig	4160	50.889
VBF Z($\mu\mu$)	$105 < m(\mu\mu) < 160$ GeV	MadGraph5_aMC@NLO + Herwig	871.0	74.86
interference	$m(\ell\ell) > 50$ GeV	MadGraph5_aMC@NLO + Pythia	5.360	128.0

Table 5.2: Summary of the Monte Carlo DY and VBF Z samples used in the analysis by process. The second column shows the cuts that identify the samples. The generator and PS simulator used are listed in the third column of the table while the fourth column shows the equivalent luminosity of each sample. The cross-sections used in the analysis are reported in the last column. The phase space selections at the generator level are specified in the third column.

are computed with top++ v2.0 [141] that included NNLO and NNLL corrections. Diboson cross-sections are computed NLO and NNLO/NLO k-factors from Ref. [142, 143, 144] are applied.

5.3 Selection

The selection starts with the trigger and the muon requirements described in section 4.2. The events are selected if two jets with $p_T > 35$ GeV and $p_T > 25$ GeV whitin $-4.7 < \eta < 4.7$ are present. In order to avoid double-counting with VH and ttH analyses, the events must not have any additional leptons, any medium b-tagged jets, or more than one loose b-tagged jet. The additional lepton requirement does not change significantly the event rate while the b-tag vetoes help in reducing the ditop background. About the 70% of $t\bar{t}$ events are rejected by the b-tag conditions.

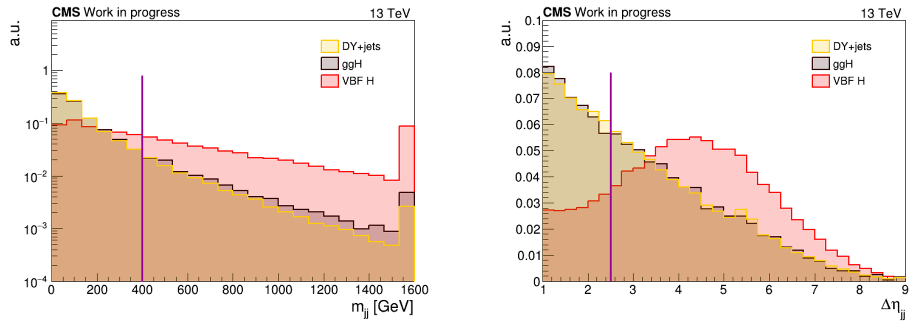
The events must have a pseudorapidity distance between the two leading jets $\Delta\eta_{jj} > 2.5$ and their invariant mass have to be $m_{jj} > 400$ GeV. Figure 5.2 shows the distribution of m_{jj} and $\Delta\eta_{jj}$ for the VBF H, ggH, and DY processes before applying the

sample	generator + PS	$\int \mathcal{L}$ / 137 fb^{-1}	$\sigma \times \mathcal{B}$ (fb)
$t\bar{t}$ leptonic decay	Powheg + Pythia	16.96	85650
$t\bar{t}$ semileptonic decay	Powheg + Pythia	9.62	356186
single top, t-channel top	Powheg + Pythia	16.97	136020
single top, t-channel anitop	Powheg + Pythia	16.86	80950
single top, tW top	Powheg + Pythia	7.61	35850
single top, tW antitop	Powheg + Pythia	7.20	35850
single top, s-channel leptonic decay	MadGraph5_amc@nlo + Pythia	49.4	3360.0
W($\ell\nu$) + 0 jet	MadGraph5_amc@nlo + Pythia	0.041	501319
W($\ell\nu$) + 1 jet	MadGraph5_amc@nlo + Pythia	0.072	842609
W($\ell\nu$) + 2 jet	MadGraph5_amc@nlo + Pythia	0.096	317296
Z($\ell\ell$)Z(qq)	Powheg + Pythia	115.3	1554.5
W($\ell\nu$)Z($\ell\ell$)	MadGraph5_amc@nlo + Pythia	80.0	1015.7
W(qq)Z($\ell\ell$)	MadGraph5_amc@nlo + Pythia	30.46	6321.0
W($\ell\nu$)Z($\nu\nu$)	MadGraph5_amc@nlo + Pythia	9.10	2011.5
W($\ell\nu$)W($\ell\nu$)	Powheg + Pythia	8.18	12178
W($\ell\nu$)W(qq)	Powheg + Pythia	7.98	34150

Table 5.3: Summary of minor bakground samples used in the analysis by process. The generator and PS simulator used are listed in the second column of the table while third column shows the equivalent luminosity of each sample. The cross sections used in the analysis are reported in the fourth column.

VBF selection. The distributions are normalized to unity to highlight the different shapes and the purple vertical lines are added to show the applied cuts. The unselected events are considered in the data-driven ggH analysis therefore the values of the cuts are a trade-off between ggH and VBF searches. In fact, the numbers of signal events in ggH analysis and in VBF analysis are very sensitive to this cut. Those values are chosen in order to maximize the signal events in ggH search without reducing significantly the sensitivity of the VBF search. After applying the VBF selections, the remaining signal is about half VBF H and half ggH.

At this point, the events of interest are divided in three regions. The regions are defined by cuts on the invariant mass of the two muons $m_{\mu\mu}$. The *Signal Region* contains the events in which $m_{\mu\mu}$ is within 115-135 GeV. The *SideBand* is composed by events with $110 < m_{\mu\mu} < 115$ GeV and by events with $135 < m_{\mu\mu} < 150$ GeV. The events within 15 GeV of the Z mass belongs to the *Z Region*. The event selection



(a) (b)

FIGURE 5.2: Dijet invariant mass (a) and pseudorapidity distance (b) for the VBF H, ggH, and DY + jets processes before applying the VBF selection. The distributions are normalized to unity and the lines represent the cuts defining the VBF category.

and the region definitions are summarized in table 5.4.

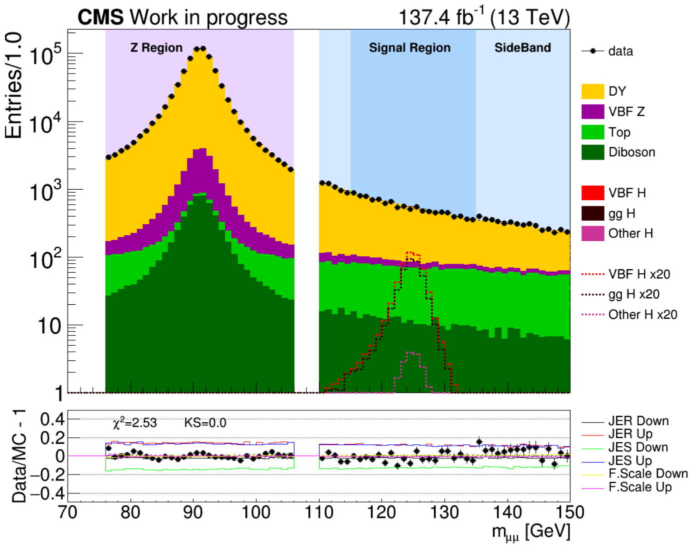


FIGURE 5.3: Dimuon invariant mass distribution after the event selection. Events in different regions are shown in the same plot. The lower panel shows the ratio data over Monte-Carlo. The dashed area represents the MC statistical uncertainties while the lines in the lower panel show the uncertainties related to the energy of the jets and to the factorization scale.

Figure 5.3 shows the distribution of the dimuon invariant mass with the region definition. The figure is composed of two panels. The upper panel shows the distribution of all the processes. The backgrounds are divided into four categories. The DY and the VBF Z are the most signal-like backgrounds and they placed on the top. The $t\bar{t}$ and the single top events are merged and they are referred to as "Top". All the other backgrounds are on the bottom of the distribution and they are indicated as "Diboson". The signal is shown by the three lines and it is scaled in order to be more evident. Moreover, the signal yields are stacked on the top of the background distribution, and that their contribution is not visible on the chart.

The lower panel of Figure 5.3 contains the data-MC ratio and it shows the main systematic uncertainties: the jet energy scale (JES) and resolution (JER), and the accuracy of renormalization calculation due to QCD processes. The jet energy scale and resolution are introduced in section 3.5.1 while the uncertainties due to QCD effects are discussed in section 5.6.1. All the single variable distributions that are shown in the following have a structure identical to the one described for Figure 5.3.

Table 5.5 shows the yields of signals and backgrounds in the signal region for the three years. Tables 5.6 and 5.7 show the yields in the sideband and in the Z region. The signal yields are negligible in sideband and Z region. The classification of the backgrounds shown in the tables is the one adopted also in the plots.

Important effects in the signal and background yields arise from the jet energy scale and resolution uncertainties. The main reason is the cut on the transverse momentum of the subleading jet. In fact, the JER and JES uncertainties of low energy jets are high and they produce important variations of the subleading jet p_T when applied. These effects are particularly enhanced in the region $|\eta| > 2.5$ where the energy measurement is provided by the calorimeters only.

Significant detector noise has been recorded in the calorimeters during the 2017 data-taking, especially in the pseudorapidity window 2.6÷3.0. The noise adds up to pile-up contribution and it worsens the energy resolution and scale. Therefore JES and JER uncertainties have very large effects in signal and background normalizations during the 2017 data-taking.

object	selection cut
leptons	$p_T(\mu) > 30,\ 20$ GeV $\lvert\eta(\mu)\rvert < 2.4$ no additional leptons
jets	$p_T(j) > 35,\ 25$ GeV $\lvert\eta(j)\rvert < 4.7$ no DeepCSV medium
global	$m_{jj} > 400$ GeV $\Delta\eta_{jj} > 2.5$ GeV $<$ 2 loose DeepCSV
Signal region	$125 < m_{\mu\mu} < 135$ GeV
Sideband	$110 < m_{\mu\mu} < 115$ GeV or $135 < m_{\mu\mu} < 150$ GeV
Z region	$76 < m_{\mu\mu} < 106$ GeV

TABLE 5.4: Event selection criteria and definition of the regions.

sample	yield 2016	yield 2017	yield 2018
DY	2061(328)	2510(505)	3688(608)
Top	275(35)	291(45)	423(59)
VBF Z	74(3)	79(32)	127(10)
Diboson	50(6)	57(9)	77(10)
VBF H	6.977(0.682)	7.384(0.677)	11.774(1.203)
ggH	4.689(1.165)	6.438(1.828)	9.608(2.530)
VH	0.165(0.020)	0.182(0.031)	0.269(0.039)
ttH	0.011(0.003)	0.011(0.004)	0.017(0.006)
Total MC	2472(363)	2951(555)	4337(669)
Data	2302	3113	3950

TABLE 5.5: Event yields in data and for different MC sources in signal region for 2016, 2017, and 2018. The statistical uncertainties are shown in parantheses.

sample	yield 2016	yield 2017	yield 2018
DY	1776(286)	2169(439)	3163(517)
Top	253(32)	255(40)	386(56)
VBF Z	65(3)	68(28)	109(9)
Diboson	44(6)	42(8)	65(9)
Total MC	2137(318)	2535(483)	3723(572)
Data	2118	2753	3383

TABLE 5.6: Event yields in data and for different MC sources in sideband for 2016, 2017, and 2018. The statistical uncertainties are shown in parantheses.

sample	yield 2016	yield 2017	yield 2018
DY	154334(31726)	182051(50422)	279452(51998)
Top	538(69)	536(81)	804(103)
VBF Z	4090(59)	4052(96)	6608(137)
Diboson	974(121)	964(159)	1282(161)
Total MC	159936(31944)	187604(50722)	288147(52348)
Data	157187	219345	274617

TABLE 5.7: Event yields in data and for different MC sources in Z region for 2016, 2017, and 2018. The statistical uncertainties are shown in parantheses.

5.4 Search for discriminating variables

A multivariate discriminator needs several variables as input. The more those variables separate signal from background, the more powerful the discriminator is, therefore a search of the variables with high separation power is performed. The parameter d is defined in order to have a tangible value of the separation power each variable can provide. The parameter d is defined by

$$d = \frac{1}{2}\sum_i |S_i - B_i| \tag{5.1}$$

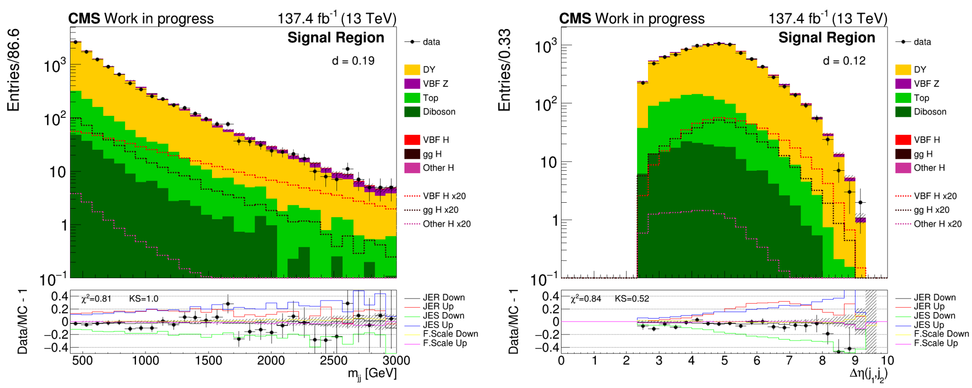
where the sum is performed all over the bins, and S_i and B_i are bin yields of the normalized signal and background histograms. The value of d ranges from 0 to 1. A value of 0 means that signal and background shapes are identical, while a value of 1 indicates a perfect separation between signal and background. The d value is shown in all the plots.

As mentioned above, this analysis relies on the great accuracy of the MC prediction, therefore, checking the agreement between data and MC is extremely important. In the following, several variables are shown. Each plot has a lower panel containing the data/MC ratio with its uncertainty bands. All the variables that are important for this analysis are shown year by year in appendix G.

Dijet invariant mass and pseudorapidity separation

The invariant mass m_{jj} and the pseudorapidity difference $\Delta\eta_{jj}$ of the two selected jets are the variables that discriminate most between VBF processes and other processes, so much so that they have been used in the analysis selection. The selected events can still be classified as a function of those two variables since the VBF processes have different distributions with respect to the other processes.

The two jets produced in VBF processes are directed in the opposite part of the detector, therefore they easily have high invariant mass and great distance in pseudorapidity. Therefore, both m_{jj} and $\Delta\eta_{jj}$ distributions have longer tails in VBF processes than in other processes. The exact shapes and peaks of m_{jj} and $\Delta\eta_{jj}$ distributions depend on the selection cuts for both signal and background. Figure 5.4 shows the distributions of the two variables for signal region events.



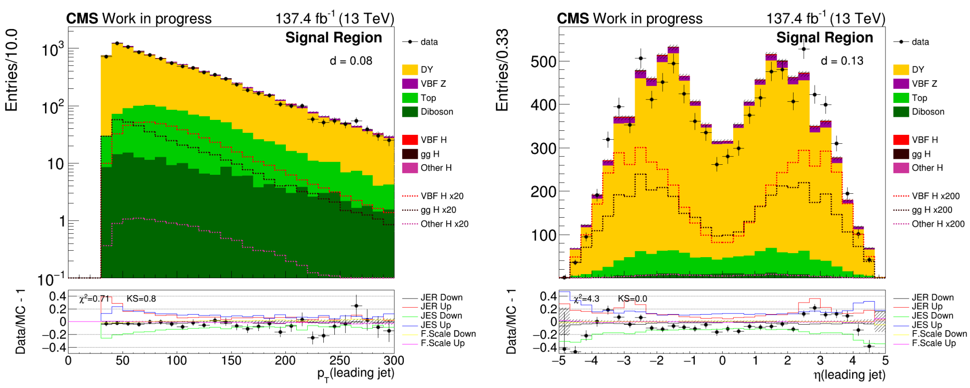
(a) (b)

FIGURE 5.4: (a) Invariant mass and (b) pseudorapidity distance of the two leading jets after the event selection in the signal region in all the years. Captions are the same of Figure 5.3.

Jets kinematic

As already mentioned, one of the main characteristic of the VBF production is having two jets with high pseudorapidity, moreover, also the momentum distributions are not identical in signal and background. Therefore, jets momenta and pseudorapidities can be used to separate signal from background. Figure 5.5 shows the distribution of the transverse momentum and pseudorapidity of the leading jet, the distributions of the same variables for the subleading jet are very similar.

Jets in DY events tend to be central and less energetic than the ones of VBF events. The bumps in the η distribution are caused by the selection and they are at lower pseudorapidities than the ones of VBF signal.



(a) (b)

FIGURE 5.5: Transverse momentum (a) and pseudorapidity (b) of the leading jet after the event selection in the signal region in all the years. Captions are the same of Figure 5.3.

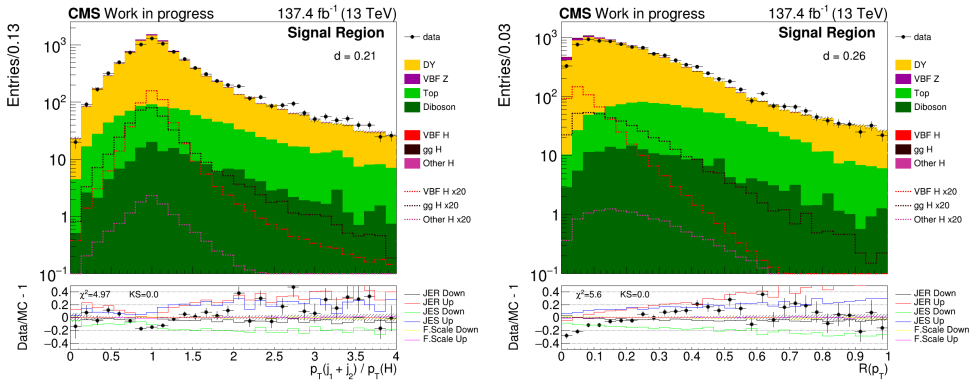
Transverse momentum balance

The main objects in VBF events are the two jets and the muons. They are the most energetic objects and their momenta should be balanced in the transverse plane, however, additional particle production or pile-up interactions may result in an unbalance of the vectorial sum of the main objects. In background events instead, the two jets and muons may not be the only protagonists of the hard scattering therefore they are, on average, less balanced than in VBF events. For this reason, balancing variables can have great discriminating power.

The ratio between the transverse momentum of the leading jets and the one of the muons is usually close to 1 but its distribution much more peaked in signal than in background events. Similarly, the transverse momentum balance of the main objects $\mathrm{R}(p_T)$ defined by

$$\mathrm{R}(p_T) = \frac{|\vec{p_T}(jj) + \vec{p_T}(\mu\mu)|}{|\vec{p_T}(j_1)| + |\vec{p_T}(j_2)| + |\vec{p_T}(\mu\mu)|} \tag{5.2}$$

has different shape for signal and backgrounds: the backgrounds have longer tails than VBF events, expecially the top processes that contain additional hadronic activity in great abundance. Figure 5.6 shows the distribution of $\mathrm{R}(p_T)$ and of the transverse momenta ratio in the signal region for all the years.



(a) (b)

FIGURE 5.6: Transverse momentum balance of the two jets and the Higgs candidate after the event selection in the signal region in all the years. Captions are the same of Figure 5.3.

Soft jet multiplicity and $\mathrm{H_T}$

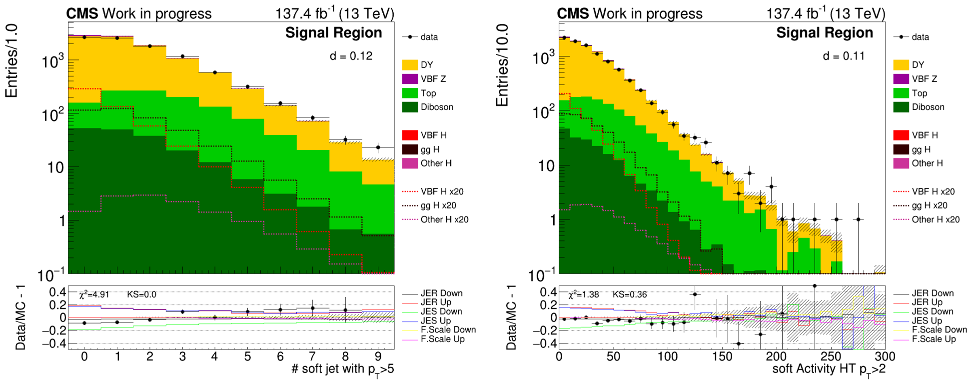
The two leading jets and the two muons are the four objects coming from the hard scattering and other hadron activities are suppressed in VBF process. Variable related to additional hadron activity are therefore well discriminating between signal and background. Track jets explained in section 3.5.4 are a powerful tool that can be used to separate signal from background. Both the transverse momentum balance R(p_T) and the soft activity variables monitor the additional hadron activity. The first is sensible to pile-up but also it is also sensible to neutral particles coming from the primary vertex, while the soft jets are independent of both.

Among many, the most discriminating soft activity variables are the multiplicity of track jets with $p_T > 5$ GeV, called N_5^{soft}, and the scalar sum of the transverse momenta of all the track jets with $p_T > 2$ GeV, called $H_{(2)T}^{\text{soft}}$. Figure 5.7 shows the distribution of N_5^{soft} and $H_{(2)T}^{\text{soft}}$ in the signal region for all the years.

Angular variables

One of the main VBF characteristics is having two opposite very forward jets. The produced boson is typically directed in the central part of the detector in between the two jets. Therefore, the pseudorapidity of the boson relative to the pseudorapidities of the jets has a peculiar shape in VBF process.

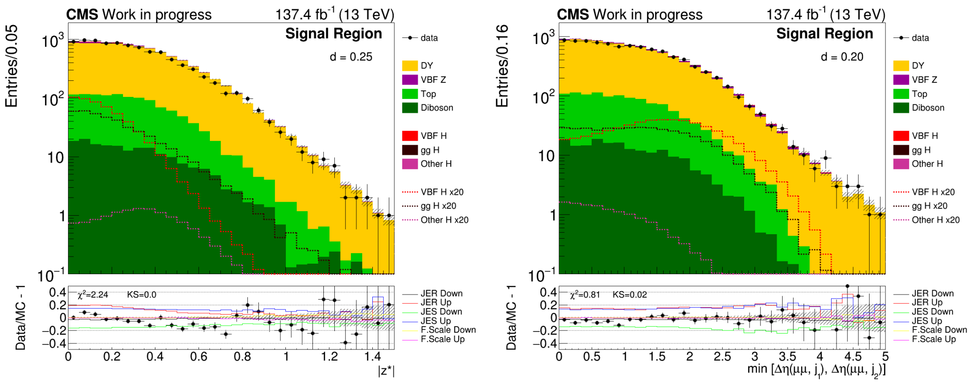
The Zeppenfeld variable $z^\star$ [145] is a linear transformations of the dimuon pseudorapidities and it is defined in 4.1. The Zeppendeld variable show if the Higgs candidate is directed in between the jets or not. Similarly to the Zeppenfeld variable, the angular distances $\Delta\eta(\mu\mu, j_2)$ and $\Delta\eta(\mu\mu, j_1)$. In particular, the minimum of the two distances is the most discriminating angular variable. Figure 5.8 shows the distribution of $z^\star$ and min[$\Delta\eta(\mu\mu, j_2), \Delta\eta(\mu\mu, j_1)$] in the signal region for all the years. The background distribution of $z^\star$ has a longer tail while the VBF shape is more peaked in 0. Instead, min[$\Delta\eta(\mu\mu, j_2), \Delta\eta(\mu\mu, j_1)$] show that background shapes are close to



(a) (b)

FIGURE 5.7: (a) Multiplicity of track jets with $p_T > 5$ GeV and (b) scalar sum of the transverse momenta of the track jets with $p_T > 2$ GeV after the event selection in the signal region in all the years. Captions are the same of Figure 5.3.

zero while VBF process peak around 1.7. Both the variables show that the dimuon pseudorapidity is mainly central and well distanced from the jets.



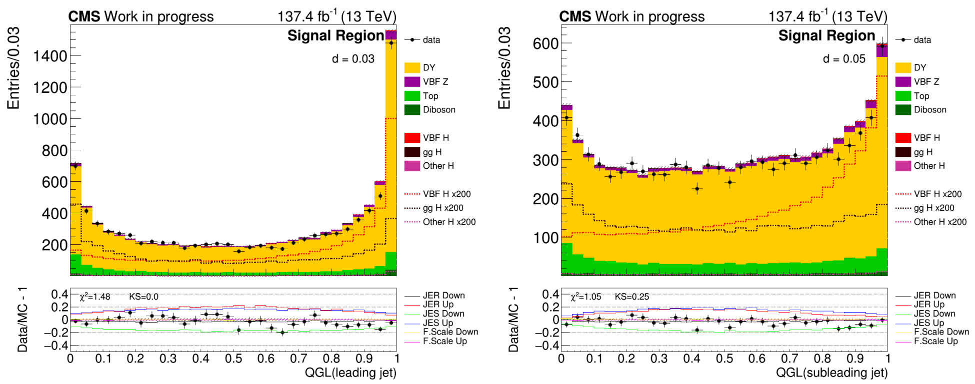
(a) (b)

FIGURE 5.8: Distributions of the Zeppenfeld variable $z^\star$ (a) and of the minimum pseudorapidity distance $\min[\Delta\eta(\mu\mu, j_2), \Delta\eta(\mu\mu, j_1)]$ (b) after the event selection in the signal region in all the years. Captions are the same of Figure 5.3.

Quark-Gluon likelihood

Jets in VBF events are expected to originate from quarks whereas events produced by other sources can have jets originated from both gluons or quarks. This difference is exploited by the QGL tag described in section 3.5.3. The QGL discriminator is evaluated on the two leading jets and the output value has discriminating power between signal and background. The crucial point is that QGL tagger does not use the kinematic of the jet as single objects, but it uses only internal variables. Therefore the QGL output is completely independent from the from four-momenta of the jets or of other physic objects, thus the QGL provides a transversal separation between signal and background.

Figure 5.9 shows the QGL outputs for the leading and subleading jets in signal region events for all the years. VBF events have more likely high QGL values and the QGL distributions peak at 1 for both leading and subleading jets. The shapes of the backgrounds peak in both 0 and 1, meaning that leading jets are often gluon-jets. Any of the leading jets can come from the radiation of a gluon from initial or final states. In that case, the gluon-jets have a softer p_T-spectrum and therefore subleading jets come from gluon hadronization more frequently than the leading ones. This is the reason why the QGL output of the subleading jet is more discriminating than the leading one.



(a) (b)

FIGURE 5.9: Quark-Gluon likelihood of the leading (a) and the subleading (b) jets after the event selection in the signal region in all the years. Captions are the same of Figure 5.3.

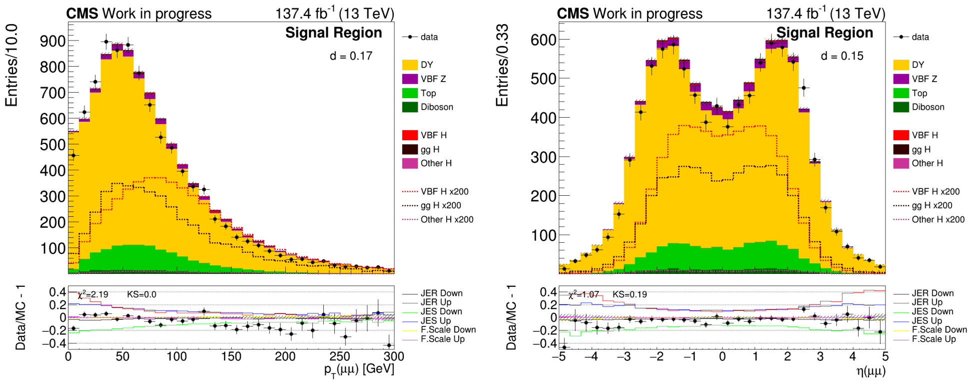
b-tagging

The b-tagging algorithm DeepCSV is used in the event selection. The cut on the DeepCSV score reduces substantially the amount of background arising from t-quarks. In principle, the DeepCSV output could still help in signal-to-background separation even after the selection, similarly to m_{jj} and $\Delta\eta_{jj}$. However, this variable would help in separating the signal from the backgrounds coming from t-quarks only and it would not help in separating signal from DY or VBF Z events. The backgrounds related to t-quarks have much fewer events than DY, moreover, the t-quark backgrounds are very dissimilar from the signal in terms of kinematic distributions: many variables have shapes that are very different from signal ones. Therefore, the separation provided by this variable is not relevant in signal-to-background discrimination.

Higgs candidate properties

The kinematic of the Higgs candidate can be different in signal and backgrounds since the two muons have different origins. In particular, the dimuon system has a harder p_T-spectrum and it is directed more centrally in the detector. The azimuthal angle instead does not contains relevant information because of the symmetry of the collision.

The momentum and the pseudorapidity can help in signal-to-background separation, moreover, they do not contain any information about the dimuon invariant mass therefore their separation is independent from the dimuon mass shape. Figure 5.10 shows the transverse momentum and pseudorapidity of the dimuon system in the signal region for all years.



(a) (b)

FIGURE 5.10: (a) Transverse momentum and (b) pseudorapidity of the dimuon system after the event selection in the signal region in all the years. Captions are the same of Figure 5.3.

Collins-Soper angles

In order to describe the directions of the two muons, the Collins-Soper angles are used. The Collins-Soper angles correspond to the angles of decay of the Higgs candidate into muons in the dimuon system and they are described in appendix A.
Their discriminating power derives from the differences between scalar boson and vector boson differential cross-sections. However, the shape of the angle distributions are modified by acceptance and reconstruction efficiency, therefore their discriminating power is much reduced with respect to ideal conditions.

The Collins-Soper angles are called θ_{CS} and ϕ_{CS} and they are completely independent of the invariant mass of the muons. Figure 5.11 shows θ_{CS} for all the selected events.

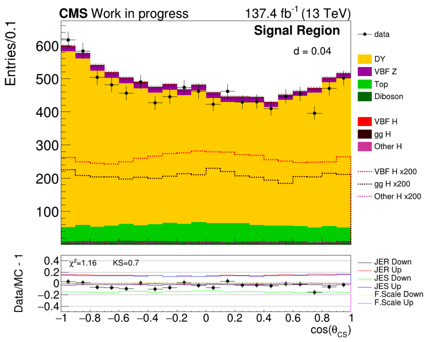


FIGURE 5.11: Collins-Soper angles θ_{CS} (a) and ϕ_{CS} (b) after the event selection in the signal region in all the years. Captions are the same of Figure 5.3.

Dimuon mass resolution

Theoretical Higgs width is 4 MeV while the muon momentum uncertainty is around 2 GeV, therefore experimental uncertainties are dominant in the dimuon mass resolution. The sensitivity of this search strongly depends on the dimuon invariant mass uncertainty: better resolution would mean a smaller window of interest, consequently reducing the background. For this reason, the final state radiation recovery described in section 3.3.2 is applied so to improve the mass resolution and the final result.

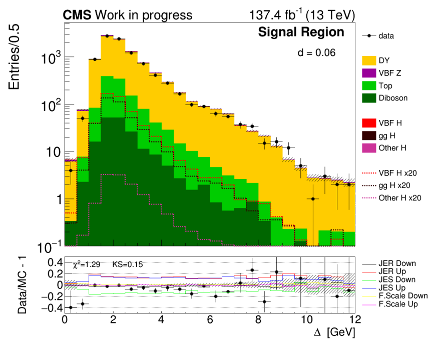


FIGURE 5.12: Dimuon invariant mass resolution after the event selection in the signal region in all the years. Captions are the same of Figure 5.3.

Figure 5.12 shows the distribution of the dimuon mass uncertainty. The variable does not show a high d parameter therefore it can not separate signal from background very well on its own. However, it can be exploited in signal-to-background separation because it affects the invariant mass distribution. Figure 5.13 shows the dimuon mass distribution in signal region for the events with mass uncertainty smaller and greater than 2 GeV.

Signal events with smaller uncertainty are closer to 125 GeV and show a more pronounced peak. In the same way, DY and VBF Z events are closer to 91 GeV thus they have mainly low values of $m_{\mu\mu}$. The background is steeply falling and the signal-to-background ratio is higher in the bins around 125 GeV. Events with higher mass uncertainties instead, show more flat distributions including a broader signal peak that is more difficult to exploit.

The mass resolution is a useful variable even if it does not have a high d parameter or have a significant correlation with the dimuon mass. Instead, it can improve the separation provided by the mass therefore it is a powerful tool that can be used in a multivariate analysis.

5.5 Multivariate discriminators

The Higgs signal can be extracted from data performing a single variable fit. The dimuon invariant mass seems to be the perfect choice for that task since the signal events are bunched in a narrow peak while the background events show a continuous monotonically descending trend. However, such a signal extraction would not exploit as best the VBF signature that has been used just in the event selection. Therefore, a multivariate discriminator is used in signal extraction in place of the dimuon mass. The discriminator is built through machine learning techniques that can optimally combine the dimuon mass with the VBF topology information provided by other kinematic variables.

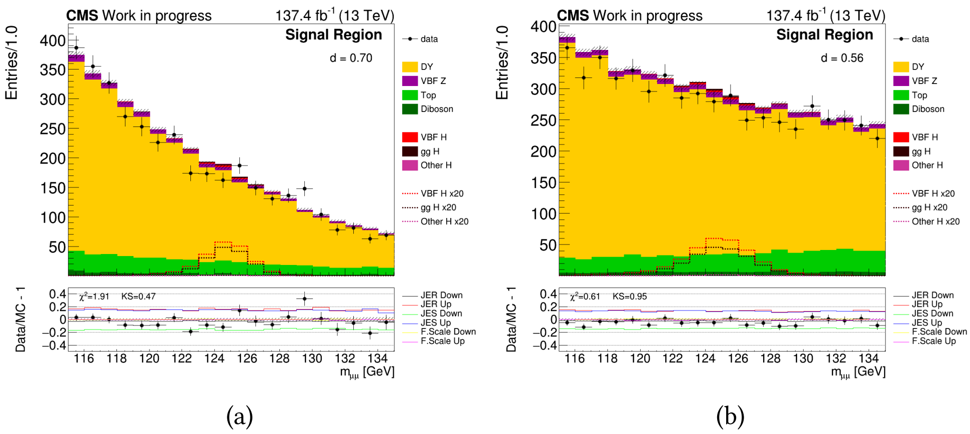


Figure 5.13: Dimuon invariant mass for signal region events with (a) mass uncertainty smaller than 2 GeV and (b) mass uncertainty greater than 2 GeV. Captions are the same of Figure 5.3.

Two machine learning techniques are explored for building the discriminator used in signal extraction: boosted decision trees (BDTs) and deep neural networks (DNNs). Both methods have been widely used in high energy physic for classification problems. Neural networks have been used since the 90s but they were replaced at LHC by BDTs for Run 1 analyses. Recently, due to the increase of computing power, DNNs made a comeback, and their performance often equal or exceed the ones of BDTs. In this thesis, the two techniques are adopted in order to build two discriminators. The one that provides the best expected sensitivity will be used to extract the final results.

Machine learning introduction

Some important concepts about machine learning have to be defined before proceeding with the discriminators descriptions.

The two machine learning techniques used in building the discriminator algorithm belong to the family of supervised learning. They make use of events in which the desired output is known in order to build a model able to classify the events in classes. The events used for building the models are called training events and they are taken from Monte-Carlo samples. The process that builds the discriminator model from the simulated events is called training. The model is built through the minimization of a function called loss function, or cost function. The loss function measures the distance of the output value of the discriminator from the truth and it is minimized via gradient descending.

The main goal of all the machine learning algorithms is to perform well on new, previously unseen data, called test dataset. In the specific case of the H$\rightarrow$ $\mu\mu$ search, the test dataset is the dataset used to extract the final results. Since any algorithm always performs better on the training dataset than on the test dataset, the task of performing well on new data can be divided in two different sub-tasks: the algorithm has to perform well on the training dataset, and the performance gap between training and test dataset has to be small. Those two sub-tasks are the two main challenges in machine learning. When an algorithm does not fulfill the first one, it is said to underfit the training dataset, while an algorithm that does not fulfill the second one is said to overfit the training dataset. Underfitting typically occurs when the model is not able to well describe the data, while overfitting occurs when the model learns the statistical fluctuation of the training dataset.

Several settings can control the behavior of the training procedure and define the model used by the discriminator algorithm. These settings are called hyperparameters and they are not set during the training process but they are fixed a priori. For example, an hyperparameter for the model build with neural network is the number of layers. For the BDT, an hyperparameter is the number of trees used in the boosting procedure. The values of the hyperparameters determine the ability to describe the data and they can be modified to correct underfitting and overfitting. So, for both the discriminators, several available hyperparameters are tested in order to tune the optimal configuration for signal-to-background separation.

Dataset used in discriminator building

The signal and background events that are used to extract the final results constitute the test dataset. The training dataset is the dataset used to train the discriminator and it is completely orthogonal to the test dataset. Whenever a discriminator is trained, an additional dataset is needed in order to verify its expected performance, this step is called validation. The validation dataset has to be independent of both the training dataset and the test dataset. The validation step is mandatory to verify the absence of overtraining, to compare discriminators built with different hyperparameters, and consequently to optimize the results.

The three datasets have to contain either signal and background events, therefore, signals and backgrounds have to be divided into three parts each. Two VBF Higgs samples are available: one showered with Pythia without dipole recoil and the other with Pythia with the dipole recoil option. The first sample is used for testing, while the second one is divided into two parts: one for training and one for validation. Adding the powheg ggH sample in the training has been tried but no significant improvement has been found. The reason is that gluon-fusion events have kinematic properties similar to the backgrounds. Consequently, ggH events are not used in training the discriminators.

The background samples used in training and validation are composed of a mixture of DY, VBF Z, and $t\bar{t}$ events. The DY samples are generated either with NLO precision and with LO precision. The samples generated with NLO precision are used for testing in order to provide a better background prediction when extracting the final result. The samples generated with NLO precision are used for training and validating the discriminators. Only one sample is available for the VBF Z and $t\bar{t}$ processes, therefore those samples are used either in training, validation, and testing. Tables 5.8 shows the samples used in training and validation and it shows the generators and the total number of events that pass the signal region requirements for each sample. The equivalent luminosities of the samples in the table are shown in section 5.2.

sample	generator + PS	total number of events
VBF H($\mu\mu$)	Powheg + Pythia	1000059
DY 105 VBF	MadGraph5_amc@nlo (LO) + Pythia	327737
DY 105	MadGraph5_amc@nlo (LO) + Pythia	69374
VBF Z	MadGraph5_amc@nlo + Herwig	211939
$t\bar{t}$ leptonic decay	Powheg + Pythia	15885

Table 5.8: Samples used for training and validating the discriminators. The last column shows the number of simulated events that pass the signal region selection.

In order to build a multivariate discriminator, several event variables are needed. They are the bases of the discriminator building and the final separation between signal and background is proportional to the ones provided by each variable. Therefore having a large number of variables with different distribution for signal and background is crucial. Moreover, many variables may be correlated with each other, so having variables containing orthogonal information is extremely important too. In the following paragraphs, the construction of the two discriminators is described.

5.5.1 Boosted Decision Tree

The first multivariate discriminant built to separate signal from background is a boosted decision tree. In this work, the BDT is implemented with the TMVA package [107] in the ROOT framework [146]. An exhaustive description of BDTs is available in appendix B.

The events used in the training are the simulated events of the samples listed in table 5.8 that pass the signal region selection. Signal sample and LO DY samples are not used in the signal extraction, therefore they are split in two subsets: one for the training and one for the validation. Instead, the VBF Z sample and the $t\bar{t}$ sample are used in the fit to measure the signal strength, therefore they have to be divided in three parts: one for the training, one for the validation, and one for the signal extraction.

Having a limited number of available MC events has repercussions also on the settings and on the input variables used to build the discriminant. In fact, the optimal hyperparameters and the optimal number of input variables are sensitive to the dataset size. The limiting factor is the number of DY simulated events. The background events are almost entirely produced by the DY process, therefore using a low equivalent luminosity sample for DY influences significantly the discriminator performance.

The DY samples generated at LO precision are used in BDT building in place of the NLO ones even if the latters are more accurate. This approach is chosen because BDT building process cannot handle properly large fractions of negative weights. The DY samples generated at NLO precision contain a large fraction of negative weight that would produce unstable discriminators. This choice leads to a worst description of the DY process but also increases the statistical power of either the events used in signal extraction and the ones used in BDT building.

Parameters optimization

Several options provided by the TMVA package are available to build the discriminator. There is not a general prescription to tune the BDT hyperparameters for a given training dataset, in this work the metric used to compare different BDTs is the expected signal significance. The signal significance is chosen because it is a raw estimate of the distance of the signal strength produced by the final fit from zero in standard deviations. The BDT distributions for signal and background are used to compute the signal expected significance bin by bin, then the bin significances are summed to obtain the total significance. The total significance is thus defined by

$$Significance = \sqrt{\sum_{i=1}^{N} \frac{S_i^2}{B_i}} \tag{5.3}$$

where S_i and B_i are the expected events of signal and background in the i-th bin and N is the number of bins of the histogram.

The described significance contains a strong dependence on the binning of the BDT histogram. This dependence has to be removed to obtain reliable comparisons between BDTs. The dependence on the binning has several causes. A too fine binning leads to high fluctuations of the bin yields of the background, the signal sample does not suffer of this fluctuation because the available equivalent luminosity is much higher than the background one. A too row binning would merge too many background events and it would reduce the significance computed by $S_i/\sqrt{B_i}$ in the most signal-like bins. Any binning with a fixed size could have both the problems, in particular, BDTs with different signal-to-background separations may have the same significance and BDTs with equal signal-to-background separation may have different significances. In order to solve these problems, the binning is build directly from the specific BDT distribution, therefore the performance of different BDTs are extracted from different binnings.

Given a certain BDT, the binning construction starts from the higher values of the BDT score and it proceeds to the bottom ones. Bin limits are defined one by one with the following criteria:

- the bin has to contain at least 0.5 expected signal events,
- the bin yield of the background must have a relative error smaller than 30%,
- the bin width has to be the smallest that verifies the two previous condition.

Such a procedure produces binnings that are independent of BDT rescalings and the BDT distributions are spread over a fixed number of bins.

The described binning has the desired features that allow it to be used in sensitivity calculation. However, the last bins of the BDT distributions still suffer from DY low statistics. Therefore, in order to optimize the BDT options and parameters, an ensemble of 20 different BDTs are built for each specific configuration. Half of the events of table 5.8 that are not used in the final signal extraction are picked randomly so to gather 20 different training datasets. At this stage, 20 BDTs are trained with the same configuration, and with the 20 training datasets. The 20 unused half datasets are used to validate the BDTs.

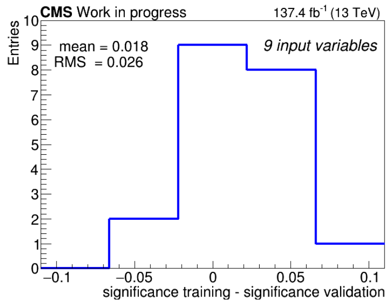


FIGURE 5.14: Difference of the sensitivities computed with the training sample and the validation sample for 20 BDTs. The BDTs are trained with 9 input variables and with the settings shown in table 5.9.

This method allows to check performance and overtraining. The performance is defined to be the average of the 20 sensitivities computed with the 20 validation datasets. The overtraining can be verified in the following way. For each BDT, the difference between the sensitivity computed with the validation datasets and the sensitivity computed with the training datasets is found. If the distribution of those 20 differences is compatible with zero there is no overtraining. Figure 5.14 shows the distribution of the 20 sensitivity differences for one of the tested configurations.

The averages of total sensitivities computed with the described method are used to optimize the BDT settings. The BDT hyperparameters described in appendix B are probed and compared to obtain the optimal configuration. They are tuned to achieve a model with the highest possible performance that does not overfit the training samples. Table 5.9 shows the optimal configuration founded with this procedure.

Boosting type	Trees	Shrinkage	Bagged fraction	node size	Depth	Separation
Adaptive boost	300	0.1	0.6	>5%	3	Gini index

TABLE 5.9: Optimized hyperparameters of the BDT.

N+1 test

The limited statistic sets a limit on the number of input variables used in the BDT, therefore not all the variables shown in the previous section can be exploited. At this point, a procedure that finds the set of variables that maximize the BDT performance is needed, the adopted procedure is called N+1 test. An N+1 test builds the optimal set of input variables by selecting the variables one by one from the ones that are availables.

Before performing any training, the input variables undergo two transformation processes. Firstly, each variable is gaussianized. It means that each variable is transformed so that its distribution become a gaussian with mean 0 and standard deviation 1. Subsequently, the variables are decorrelated by a rotation, therefore the inputs used to build the BDT model are not exactly the gaussianized variables but a set of their independent linear combinations.

After the two transformations, the N+1 test is performed starting from training BDTs with two input variables: $m_{\mu\mu}$ and m_{jj}. The invariant masses $m_{\mu\mu}$ and m_{jj} are used as a starting point because they are the most discriminating variables. The performance of the BDT trained with the two input variables is compared with the performance of several BDTs trained with three variables. The sets of three variables used are composed of $m_{\mu\mu}$, m_{jj} and an additional variable that is different from set to set.

As previously done, 20 different BDTs are trained with the same settings and inputs but with different datasets therefore for each set of inputs the performance of the 20 BDTs are averaged. The averaged significances are computed with both validation samples and training samples separately and they are compared in order to ensure the absence of overtraining. Moreover, for each variable, the 20 BDTs are used to compute the significance distribution and to extracts its standard deviation that is used as sensitivity uncertainty. Figure 5.15 compares the performance of the BDTs with two inputs and the BDTs with three inputs computed with training and validation samples.

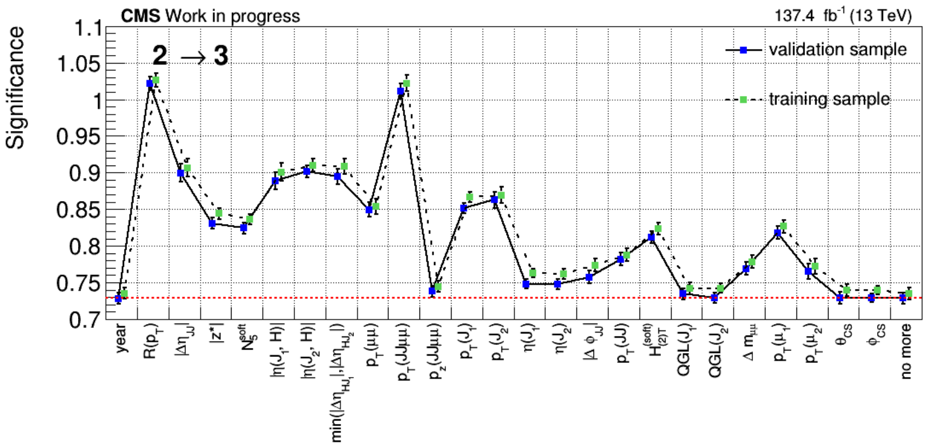


FIGURE 5.15: Comparison between three input BDTs and two input BDT. Each bin shows the sensitivity of the BDTs with three variables, except for the rightmost one that shows the sensitivity of the BDT with two inputs. The labels on the x-axis show the variables added to the initial set of two variables. The performance and their uncertainties are the mean and the standard deviation of 20 discriminators trained with different datasets and they are computed with the training dataset (green) and with the validation dataset (blu) separately.

The plot shows that the transverse momentum balance R(p_T) is the variable that performs better therefore it is added to the final set of input variables. Then, the described procedure is repeated to find the fourth variable: the performance of the BDTs trained with $m_{\mu\mu}$, m_{jj}, and $\mathrm{R}(p_T)$ is compared with the performance of BDTs trained with four input variables. After the choice of the fourth variable, the procedure is repeated until no significant gain is obtained by adding any variables. The N+1 test leads to an optimizing set of 9 variables:

- $m_{\mu\mu}$ - the dimuon invariant mass,
- m_{jj} - the dijet invariant mass,
- $\mathrm{R}(p_T)$ - the transverse momentum balance,
- $\min(|\eta(\mu\mu, j_1)|, |\eta(\mu\mu, j_2)|)$ - the minimum distance in pseudorapidity of the Higgs candidate with the two leading jets,
- N_5^{soft} - the number of soft jets with $p_T > 5$ GeV,
- $p_T(j_2)$ - the transerse momentum of the subleading jet,
- $\Delta\phi(jj)$ - the azimuthal distance between the two jets,
- $\Delta\eta(jj)$ - the pseudorapidity distance between the two jets,
- $|z^\star|$ - the absolute value of the Zeppenfeld variable.

The plots showing all the steps of the test are shown in appendix E, Figure 5.16 shows the last step, where adding the tenth variable does not improve the performance of the BDT.

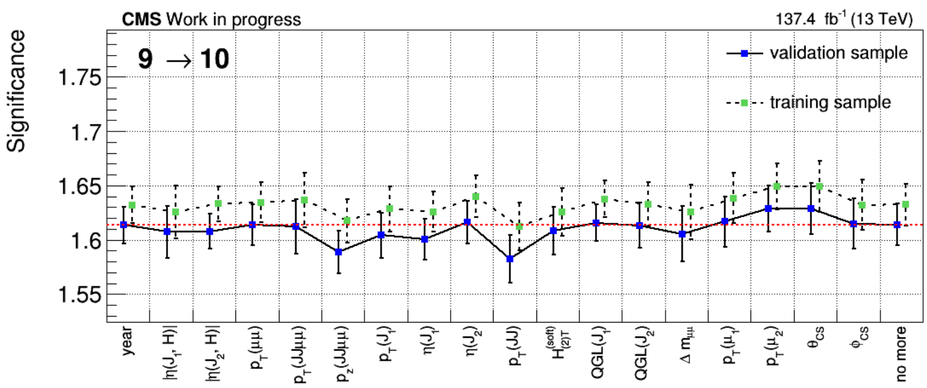


FIGURE 5.16: Comparison between 10 input BDTs and 9 input BDT. Captions are the same as Figure 5.15.

Several remarks arise from the result of the N+1 test. Firstly, the variable "year" has been used in the procedure so that the BDT could take into account any differences between the data taking periods. However, "year" does not show any particular improvement, meaning that those differences are small. Then, the QGL of the leading jets has not been selected even if their separation is completely independent of the other variables. This means that the procedure did not stop because it depleted all the available information, but it stopped because it was not able to exploit the remaining variables. In fact, several other variables contained transversal information and they could not be used. These unexploited variables lead to try an alternative method to separate signal from background, a method capable of combining more inputs than the BDT.

5.5.2 Neural Netwok

In recent years neural networks have been successfully used in several fields, including high energy physics, thanks to the development of deep learning. Therefore, alongside the BDT, a feed-forward neural network is developed in order to better separate signal from background. The DNN is trained with the KERAS [147] package and Tensorflow backend [148]. An introduction to deep learning concepts that are used is available in appendix C.

Network architecture

The DNN is trained using simulated samples listed in 5.8. The VBF Z sample and $t\bar{t}$ sample are used in the final signal extraction and they have to be divided in training, validation, and test datasets. The VBF Z process is the most signal-like background and its limited statistic in the training and in the signal extraction have a non-negligible impact on the final result. In order not to lose statistical power of that sample a 4-fold procedure [149] is used at training time. The 4-fold procedure consists of dividing each sample into four subsets of the same size. Four different DNN are trained and each DNN uses two subsets for training, one for validation and one for testing in rotation as sketched in Figure 5.17. Such a scheme allows to train, validate, and test with all the available events.

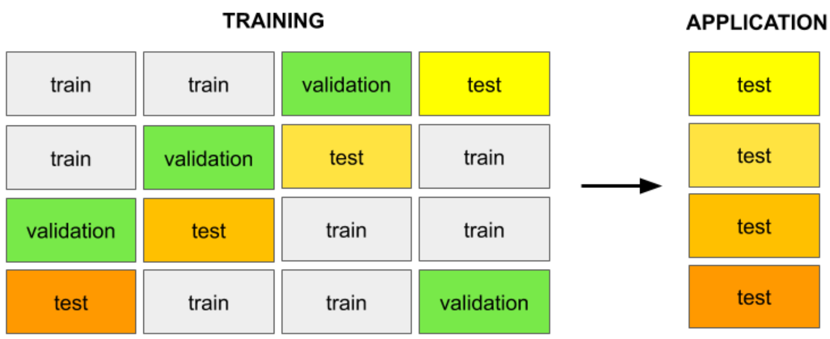


FIGURE 5.17: Schematic representation of the 4-fold training, validation and evaluation procedure.

The 4-fold procedure is performed for all the samples used in the training, including the ones that are not used as test samples. The evaluation of the DNN on the simulated samples that are used in the training is performed evaluating the four DNN in rotation. The result is a single DNN histogram that is composed of the DNN score of the four DNNs. The evaluation of the DNN on the simulated samples that are not used in the training is performed with a different method: the four DNNs are evaluated on the entire sample, then the DNN histograms are averaged. In data instead, the events, are randomly divided into four subsets of equal size, and the four DNNs are evaluated on one subset each. In the end, for either data and all the samples, the final distribution of the DNN is formed by the four DNNs and exploits the four trainings. This allows the partial regularization of the final result because the fluctuations of single subsets are reduced by a factor of 2.

Since the 4-fold implementation allows not to lose statistics in training and testing, the DY samples with NLO accuracy can be used in place of the LO ones, and at the same time, negative event weights can be used in the training. In principle, the NLO samples describe the kinematic of the DY process better than the LO one, therefore several DNNs are built with different DY samples in order to find the training dataset that optimizes the DNN performance. A DNN is trained with the NLO DY samples and a DNN is trained with the LO samples. Moreover, several DNNs are trained with the NLO and the LO DY samples combined with different mixtures. Those DNNs are validated with the 4-fold technique using only NLO DY samples because it is the most accurate and it is the one used at test level. All the DNNs have similar performance and they show better results when fewer NLO DY events are used. Therefore, the LO DY samples are chosen to be used in the training despite the possibility to use the NLO ones.

Neural networks, as opposed to BDTs,do not suffer limitations when adding extra input features to the training. In fact, BDTs break the input space into as many regions as there are leaves therefore they use a fraction of the total events to parametrize the model in each region. On the contrary, DNNs are smoothness-based learners and use all the events to parametrize a single model that covers the entire space. For this reason, in DNNs all the variables that can be useful can be added without loss in performance.

The DNN input variables are chosen based on the BDT optimizations and N+1 test. However DNN allows the use of many more variables, therefore performance improvements are sought by adding additional inputs based on previous analysis experience. The used input variables are 25, three of those are related to the dimuon invariant mass while the remaining 22 have little correlation with it. Before being fed to the DNN all the inputs are standardized. The sample mean is subtracted from each value and the result is divided by the sample standard deviation: As a result, the new input distributions have a mean 0 and a standard deviation of 1.

The input variables are:

- $m_{\mu\mu}$, $\Delta m_{\mu\mu}$, $\Delta m_{\mu\mu}/m_{\mu\mu}$ - the dimuon mass and the absolute and relative mass resolutions,
- m_{jj}, $\log m_{jj}$ - the dijet invariant mass and its logarithm,
- $R(p_T)$ - the transverse momentum balance,

- $|z^{\star}|$ - the absolute value of the Zeppenfeld variable,
- $\Delta\eta(jj)$ - the pseudorapidity distance between the two jets,
- $p_T^{\mu\mu}$, $\log p_T^{\mu\mu}$, $\eta_{\mu\mu}$ - the dimuon transverse momentum and pseudorapidity,
- $\cos\theta_{CS}$, ϕ_{CS} - the *phi* angle and the cosine of the θ angle in the Collins-Soper reference frame,
- $\min(|\eta(\mu\mu, j_1)|, |\eta(\mu\mu, j_2)|)$ - the minimum distance in pseudorapidity of the Higgs candidate with the two leading jets,
- N_5^{soft} - the number of soft jets with p_T > 5 GeV,
- $H_{(2)T}^{\text{soft}}$ - the scalar sum of the transverse momentum of soft jets with p_T > 2 GeV,
- $p_T(j_1)$, $p_T(j_2)$, $\eta(j_1)$ $\eta(j_2)$, $\phi(j_1)$, $\phi(j_2)$ - the vector components of the leading jets,
- $QGL(j_1)$, $QGL(j_2)$ - the quark-gluon likelihood discriminators for the two leading jets,
- *year* - the data-taking period.

The DNN exploits several variables that BDT is not able to use. Primarily, the quark-gluon likelihoods contain information that is orthogonal to all the other variables so QGLs are used in the DNN. Also, the separation provided by the soft activity information is independent of the other variables, therefore soft activity is further exploited with respect to the BDT by adding $H_{(2)T}^{\text{soft}}$. Then, the dimuon mass resolution is used to improve the signal-to-background separation provided by the dimuon mass. Moreover, the entire kinematic of the two leading jets is exploited in the DNN, while the BDT could just use $p_T(j_2)$. Finally, the variable "year" is added to the training so that possible discrepancies due to the different data-taking conditions can be taken into account.

Some of the above listed variables are redundant. Redundant variables help in emphasizing important information and do not impact the performance because DNN is capable of handling a large number of inputs.

In order to optimize the training setup, several architectures have been tested. The optimal performance is obtained when the training is performed in multiple steps. Firstly, four networks are optimized with different inputs and for different tasks. The four outputs of the last hidden layer nodes are merged and combined to solve the actual classification problem. The five networks are the following:

(1) **signal -vs- VBF Z**
The VBF Z proces is the most signal-like background. Distinguishing as best VBF H from VBF Z is crucial to isolate the signal, therefore a dedicated network with 3 hidden layers is trained in order to separate signal from this background.

(2) **signal -vs- DY**
A network with 3 hidden layers is dedicated to separate signal from DY events.

(3) **mass independent signal -vs- background**
A 3 hidden layers network is trained with the 22 input variables uncorrelated with the dimuon mass. This network is trained against the three backgrounds.

(4) **mass + mass resolution signal -vs- background**
A network with two hidden layers is trained using only three input variables: the dimuon mass and its absolute and relative resolutions. In this way, the network can try to exploit as best the dimuon mass discriminating power. All the backgrounds are used in the training.

(5) **merge**
A final network with 2 hidden layers is used to combine the outputs of the 4 preliminary networks. The last fine-tuning of the model is performed in two steps. In the first step (5a), the node weights of the networks (1,2,3) are fixed as obtained by the previous training while the network (4) weights are free to float. In the second step (5b), also the node weights of the network (3) are unfrozen.

A schematic rapresentation of the DNN architecture is shown in Figure 5.18. The input variables are shown in orange and they are divided in two blocks: the three related to the mass and the others. The DNNs optimized for the single tasks are shown in grey boxes and their output is blue. The last layers of the 4 networks are merged in a single vector and used as input for the combination. The final output is shown in red.

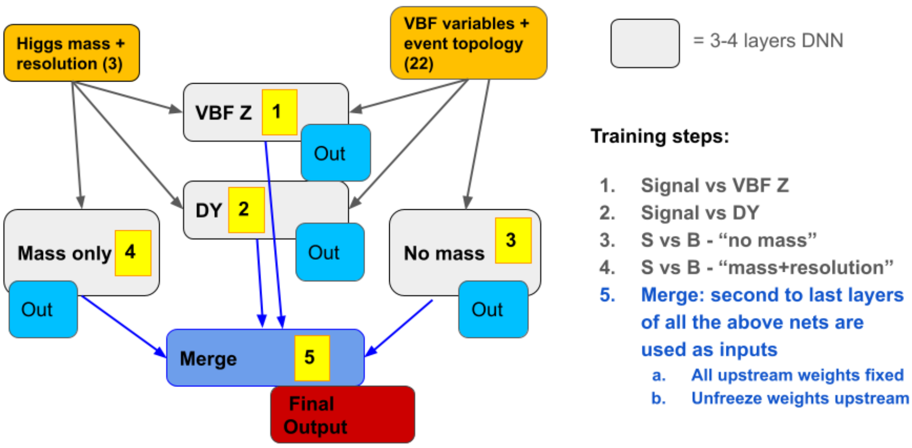


FIGURE 5.18: Schematic representation of the DNN architecture. The training procedure consists in optimizing the first four networks for single tasks, then combining the outputs and fine-tuning the model by unfreezing upstream weights.

Each network has 2 or 3 hidden layers with few tenths of nodes with a descending pyramidal architecture. A dropout of 20% is used in all the training steps in order to regularize the model and a batch size of 1024 events is used, except for step (5b) where 10240 events are used. The adopted loss function is the binary cross-entropy, which results in the maximum likelihood estimator for binary classification problems. A starting learning rate of 0.05 is used and it is gradually decreased at training time based on validation loss, thus preventing overfitting.

The adopted architecture divides the signal-to-background separation provided by the mass from the one provided by the remaining variables. In fact, network (3) focuses entirely on the VBF and event topology while it lets the network (4) to make the most from the dimuon mass information. Instead, the networks (1) and (2) use all the input variables simultaneously but they aims at single background rejection. The four networks contain different informations and they can be combined. Moreover, having several independent pretraining helps the final signal-to-background discrimination.

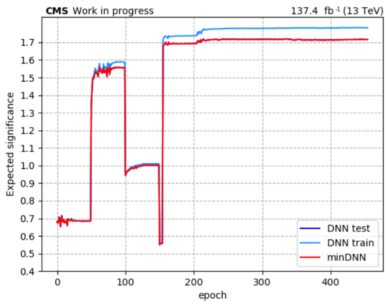
In order to prove that no other variables would increase the discriminator performance, a DNN with the same architecture is trained using as inputs all the event variables at disposal, about one hundred. The performance of the two DNNs do not show significant differences, meaning that additional inputs would not improve the signal-to-background separation.

All the described networks are trained for several epochs. The network (4) is trained for 5 epochs and (5b) for 200 epochs. The remaining models are trained individually for 50 epochs each. The best trained model is chosen among the 250 training epochs of (5a) and (5b) using the estimated Asimov significance [150]. The significance is computed for the training and for the validation dataset for each fold, and the minimum significance is used to pick the best model. The Asimov significance is computed by summing in quadrature the bin by bin significances defined by

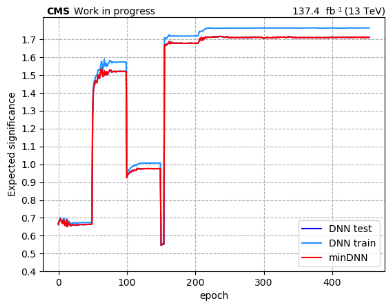
$$Asimov\ \ significance = \sqrt{\sum_{i=1}^{N} 2\left[(S_i + B_i)\log\left(1 - \frac{S_i}{B_i}\right) - S_i\right]} \tag{5.4}$$

where S_i and B_i are the expected of signal and background events in the i-th bin. The binning used to evaluate the DNN performance is computed similarly to the one used in BDT evaluation: each bin has to contain 0.5 expected signal events and the uncertainty of the background yields has to be smaller than 30%.

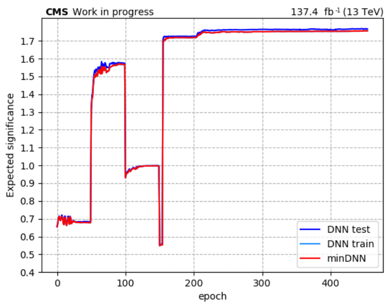
Figure 5.19 shows the training (light blue) and validation (blue) significances for each epoch and for each fold of the networks in order from (1) to (5b). The minimum significance (red) is used to choose the best performing model. The sudden changes in the performance trend are due to the different training steps. The performance of the network (1) is shown in the first 50 epochs. The 50 training epochs of the network (2) are the ones from 51 to 100. The performance of (2) is much higher than (1) because the VBF Z process is much more signal-like than the DY, therefore (1) obtain a looser separation than (2). The epochs from 101 to 150 show the performance of (3) while the epochs of (4) are the ones from 151 to 155. Networks (5a) and (5b) are shown in epochs 156-205 and 206-406 respectively. The performance of the



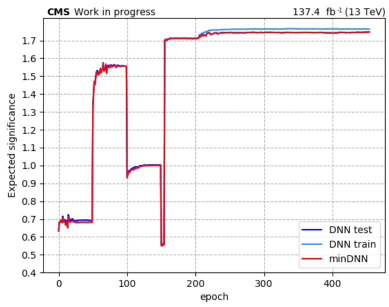
(a)



(b)



(c)



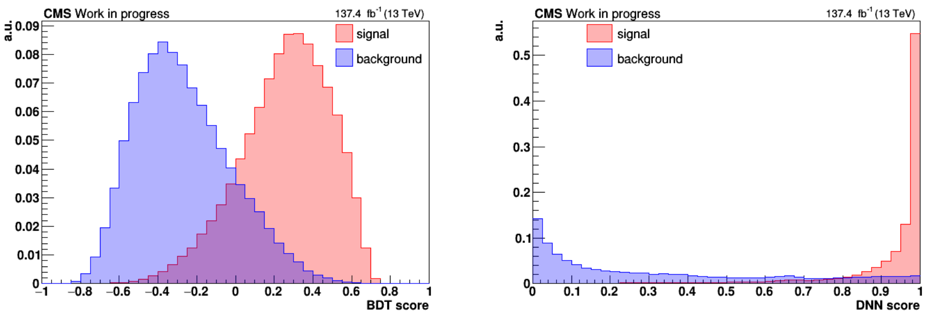
(d)

FIGURE 5.19: Training and validation significances for each epoch for each fold of the networks. In order, there are 50 epochs of (1), 50 epochs of (2), 50 epochs of (3), 5 epochs of (4), 50 epochs of (5a), and 200 epochs of (5b).

various steps are characterized by the same typical features: an initial increasing trend and initial instabilities. Moreover, the increases at the 206-th epoch are due to the unfreezing of the event weights of network (3).

5.5.3 Performance of the discriminators

The BDT and the DNN have to be compared in order to choose the best discriminator for the analysis. Figure 5.20 shows the distribution of the optimized BDT and DNN computed with the validation dataset. The BDT score range from -1 to 1, the signal-like events have high BDT values while the background-like events have low BDT values. Instead, the DNN score is always positive: signal events are generally close to 1 and background events are peaked at 0. The plots do not show any visible outperformance of one of the two discriminators with respect to the other one.



(a) (b)

FIGURE 5.20: Signal and background distributions of the BDT (a) and DNN (b) computed with the validation dataset.

The present analysis is very sensitive to the signal efficiency at about 99-99.9% background rejection. Therefore an efficient method to compare the discriminators is through the Receiver Operating Characteristic (ROC) curve. The ROC curve is the background efficiency as a function of the signal efficiency and it is computed with the validation dataset. Figure 5.21 shows the ROC curves of the optimized BDT and DNN. The curves show that the DNN has higher signal efficiency for each background rejection. In particular, the DNN has higher signal efficiency at the desired working point.

Another estimator of the discriminator performance is the expected significance. The significance can be computed either with equation 5.3 or with equation 5.4. Both sensitivities show an improvement of the DNN performance of about 10% compared to the BDT. Therefore, both ROC and significance show that the DNN performs better than the BDT on the validation set.

The signal strength extraction is performed with the test dataset. The performance of the discriminators with the test dataset are not exactly the ones found with the validation dataset: the test dataset contains a better description of the background and it includes also other signal channel productions. However, the comparisons

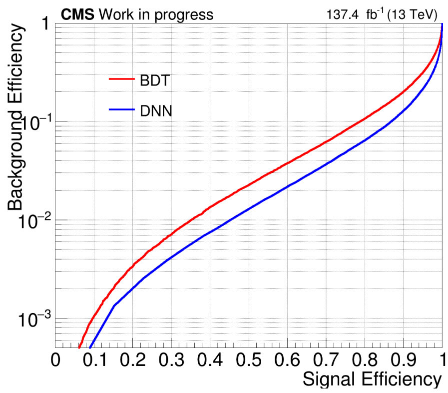


FIGURE 5.21: ROC curves of the optimized BDT and DNN evaluated with the validation dataset. The LO DY sample is used in the evaluation.

obtained with the validation dataset are accurate enough to confirm that the DNN can separate signal from background better than the BDT. Therefore the final fit is performed with the DNN discriminator.

5.6 Systematic uncertainties

In this section, the main sources of errors are described. The systematic uncertainties are propagated consistently to the distributions used in the fit and they affect the significance and the signal strength uncertainty. The following uncertainties are coherently considered in all the channels of the $H \to \mu\mu$ search and they are correctly treated to obtain the combined result.

The systematic uncertainties are divided in theoretical and experimental uncertainties. Theoretical uncertainties are the sources of error that arise from mismodeling in the MC sample generation. Experimental uncertainties are the sources of error related to differences between measurements and the true values of the observables. Either the experimental uncertainties and the theoretical ones can affect the shape of the distributions, their normalizations, or both. The shape uncertainties are either applied to the relevant objects and propagated through the event selection and the discriminator evaluation, or treated as an event weight when building the variations about the nominal distribution. Systematic uncertainties are modeled in the fit as nuisance parameters and they are constrained by log-normal or gaussian priors.

5.6.1 Theoretical uncertainties

Signal cross-sections

The uncertainties of cross-sections and branching fractions for ggH, VBF, VH, and ttH processes are taken from Ref. [20]. These include uncertainties in the PDF distributions and in the choice of QCD normalization and factorization scales.

ggH and VBF Simplified Template Cross Sections

An additional set of uncertainties is considered in order to take into account theoretical mismodelings of some kinematic distributions of the Higgs processes. These uncertainties are applied following the recommendations of the LHC Higgs cross-section working group [151]. They account for variations in the signal acceptance as a function of $p_T(H)$, $p_T(Hjj)$, jet multiplicity, and m_{jj} for the VBF process and as a function of $p_T(H)$ and jet multiplicity for the ggH process. The uncertainty due to this systematic is about $2 \div 4\%$ for VBF process and $15 \div 25\%$ for ggH process in the VBF category.

Perturbative QCD scale variation

The perturbative QCD renormalization and factorization uncertainties are estimated by changing the scales μ_R and μ_F up and down by a factor of 2 from the default values used in the matrix element calculation. These uncertainties are considered by applying the modified event weights and they affect both shape and normalization. In the fit, they are treated correlated between regions and years but uncorrelated between processes.

Parton shower uncertainty for VBF H

The parton shower uncertainty is estimated for the VBF signal from the differences observed in the predictions obtained with different generators. These differences affect either shape and acceptance and they are about $5 \div 10\%$. The nominal prediction is taken from POWHEG + PYTHIA with dipole recoil, while the full difference with POWHEG + HERWIG is used as variation. The uncertainty is symmetrized by inverting the ratio between the two predictions in order to have a two side nuisance.

Parton shower uncertainty for VBF Z

Similarly to VBF H, the parton shower uncertainty for the VBF Z process is estimated by the difference between the predictions obtained with different generators. However, the PYTHIA shower generator without dipole recoil is known to mismodel the additional hadron activity in VBF events, and a VBF Z sample generated by POWHEG + PYTHIA with dipole recoil is not available. Therefore the sample generated with POWHEG + HERWIG is used as central prediction and the 20% difference between POWHEG + PYTHIA and POWHEG + HERWIG is taken as parton shower uncertainty. This fraction is chosen because it corresponds to twice the parton shower uncertainty provided by the PYTHIA generator with the PS-weight mechanism. Here too, the uncertainty is symmetrized by inverting the ratio between nominal and variation.

PDF variations

A specific recipe is designed to take into account PDF uncertainties. This method uses the 100 replicas of the PDFs provided by NNPDF3.0 in 2016 samples and by NNPDF3.1 in 2017 and 2018 samples. For each replica and process, the expected discriminator distribution is computed. The ratio between the 100 distributions and the nominal one is computed and fitted with two different functions: $y = mx + b$ and $y = ax^2 + b$, where x is the value of the discriminator. Let a', m', and b' be the RMS of the ensamble of fitted parameters a, m and b. Three variations of the nominal histogram are built multiplying the yields by either ax^2, mx, or b. Normalization effects are removed from the two non-constant variations. These uncertainties are symmetrized by flipping the sign of the a', m', or b'. The three PDF variations are treated as correlated across regions and eras and uncorrelated between processes.

5.6.2 Experimental uncertainties

Luminosity

The uncertainty on the integrated luminosity affects directly the expected yields of signals and backgrounds. The luminosity of 2016, 2017, and 2018 data-taking periods are known with uncertainties of 2.5%, 2.3%, and 2.5% respectively [152, 153, 154].

Prefiring correction

A shift in the timing of the ECAL L1 trigger in the forward region ($2.4 < |\eta| < 3.0$) led to a specific trigger inefficiency in the 2016 and 2017 data-taking periods. This inefficiency has been measured and a correction has been provided using minimum bias events. The correction is applied by reweighting the MC events as a function of the η and p_T of the jets and it is about 2÷3% at m_{jj} = 400 GeV and 6÷9% at m_{jj} > 2 TeV. The corresponding error is obtained by propagating the uncertainty associated with the correction that is 20%. This systematic affects either shapes and normalizations and it is correlated across signal and control regions.

Single-muon trigger

Scale factors and the corresponding uncertainties are applied to take into account inefficiencies from the single-muon trigger. Their contribution is less than 0.5% and they are included in the overall identification efficiency described below.

Muon energy scale and resolution

The muon scale and momentum resolution are calibrated as a function of p_T and η of the muons using dilepton resonances [155]. The relative momentum resolution varies from 1.5 to 2% when the muon passes through the barrel while it ranges from 2 to 4% for endcap muons.

Lepton identification and isolation efficiency

The muon and electron identification and isolation efficiencies are determined centrally for the CMS analyses using the tag-and-probe method in Z events. The obtained scale factors are applied to correct the prediction from the Monte Carlo samples. The corresponding uncertainties are propagated through all the steps of the analysis. The electron identification and isolation uncertainties affect only VH and ttH categories.

Quark-gluon likelihood discriminator

The polynomial corrections described in chapter 3 are considered in the event weights. The variations to the nominal distributions are computed by applying those corrections twice for up variations and not applying the corrections for the down variations. QGL systematic variations do not change the signal and background yields, therefore only the shape variations are considered [101].

Pile-up
A shape uncertainty on the number of pile-up events is obtained by varying the minimum bias cross-section used in the pile-up reweighting applied to simulated events.

b-tagging
The b-tagging scale factors are needed in order to take into account differences in selection efficiency between data and simulated events. They are computed centrally for the CMS collaboration for each working point and they are implemented via reweighting the events. The sources of b-tagging uncertainty are varied independently and correctly propagated through the analysis steps. The corresponding uncertainties influence both shapes and normalizations of all the distributions.

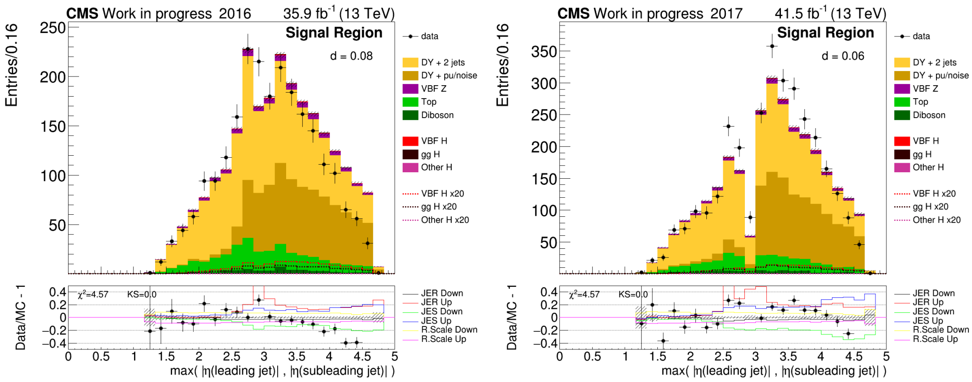
Monte Carlo simulation size
The uncertainty due to the limited size of the simulated events is taken into account by varying independently each bin yield for signal and background with the corresponding uncertainty with the Barlow-Beeston method [156, 157]. These uncertainties are uncorrelated across the bins and they affect mostly the high score region of the discriminators.

Jet energy scale (JES)
The transverse momentum of each jet is varied up and down within one standard deviation for each source of uncertainty as recommended centrally for CMS data analysis [93]. The event selection and discriminator evaluation are re-applied accounting for the shifted energies and the corresponding variations in each bin yield for each process are used to estimate the uncertainty. Therefore the jet energy scale affects the selection acceptance as well as the shape of all the distributions.

Jet energy resolution (JER)
The jet energy resolution is taken into account by smearing each jet by the smearing factors provided centrally for the CMS collaboration [93]. Instead of having a single nuisance parameter to define the JER uncertainty, jets are divided in six exclusive categories and the resolution parameters in the categories are set to be uncorrelated in the final fit. Two categories divide the central jets: the first is composed of jets with $|\eta| < 1.93$ and the second is composed of jets with $1.93 < |\eta| < 2.5$. Forward jets are divided in four categories as a function of p_T and η: jets are divided in high energy jets with p_T >50 GeV, and low energy jets with p_T <50 GeV, and either high and low energy jets are divided in pseudorapidity at $|\eta| = 3.139$.

Drell-Yan contribution from pile-up and noise
About 35÷40% of the DY background is composed of events where either the leading or the subleading jet are not matched with a jet at generator level. These jets come from soft emissions produced mainly by the parton shower or by the pile-up interactions. They usually fall in the forward regions of the detector and they pass the p_T threshold used in the analysis because their energy is wrongly measured

by unprecise detector response. In fact, the forward region is sensitive to pile-up, noise, and other sources of background that are difficult to predict. In particular, the rate of this component of the DY background is not well predicted.



(a) (b)

FIGURE 5.22: Maximum pseudorapidity between the two leading jets of signal region events in 2016 (a) and 2017 (b). Captions are the same of Figure 5.3.

This effect is particularly enhanced during the 2017 data-taking because of the noise in the calorimeters. Figure 5.22 shows the pseudorapidity of the most forward jet between leading and subleading in the signal region in 2016 and 2017. The DY contribution is split into two components: the first have leading and subleading jets matched with a jet at generator level, the second contains the remaining DY events. The ratios data/MC have similar trends in the two years, however, the 2017 data-taking shows an excess of data in the range within 3 and 5. The difference between the two years is compatible with a wrong normalization of the component where one jet is not matched at generator level.

The adopted solution is treating the DY process as two different backgrounds, called DY+2jets and DY+pu/noise. The normalization of the DY+pu/noise is left floating in the fit and it is directly constrained by the observed data. Because of significant variation of the detector response over the data taking period, this normalization is treated uncorrelated across years. The normalization of the DY+2jets is taken from the simulation and, as well as all the other backgrounds, constrained by the data within the described uncertainties.

5.7 Signal extraction method

The variable that separates most signal from background is the DNN score, therefore it is used also in the signal extraction. In order to check the data-MC agreement of the DNN distribution, the DNN score is evaluated in the sideband and in the Z region. However, such an evaluation does not probe the agreement in the region of higher value because events that have $m_{\mu\mu}$ far from 125 GeV have also the DNN output is close to 0. The reason is that the dimuon mass is very important in the

DNN evaluation and events with $m_{\mu\mu}$ far from the Higgs peak are classified as background. A solution is found by replacing the DNN input value of the dimuon mass with the default value of 125 GeV during the evaluation in both data and simulation. Such a method allows having a wider distribution of the DNN score that reproduces the main features of the DNN distribution in the signal region.

The signal samples used for extracting the signal strength are generated with a Higgs mass of $m_H = 125.00$ GeV, however, the most precise measurement of the Higgs boson mass is 125.38 GeV. In order to obtain results with the correct mass, the dimuon mass is biased of 0.38 GeV before evaluating the DNN in data and backgrounds. This is possible because the other DNN inputs are not significantly correlated to $m_{\mu\mu}$ therefore there is no way that the DNN obtains information about the true dimuon mass from other sources. This bias is equivalent to evaluate a DNN that is trained with signal samples generated with $m_H = 125.38$ GeV.

The signal strength is extracted with a maximum likelihood fit performed over the DNN distribution in the signal region. In order to better constrain the nuisances, the fit is performed simultaneously in signal region and in sideband. The variables used in the fit can be several for the sideband. The optimal distribution for the sideband is the distribution of the DNN with the fixed value of the dimuon mass input. Figure 5.23 shows the distribution of the DNN score in the signal region and the one of the DNN at fixed mass in the sideband.

The fit can be performed in two different ways. One method is to compute the shapes of the discriminator distributions using all the events of the data taking period and then fit one distribution for the signal region and one for the sideband. The other one is to compute the shapes of the discriminator distributions separately for the three years and then perform a simultaneous fit over the three years. The method that provides the best result has to be chosen.

In case the same binning is used for the two methods, fitting simultaneously the three years would perform better. The reason is that the significance in each bin, that is about S^2/B, is higher when combiend from different distributions

$$\frac{S_{16}^2}{B_{16}} + \frac{S_{17}^2}{B_{17}} + \frac{S_{18}^2}{B_{18}} \geq \frac{(S_{16} + S_{17} + S_{18})^2}{B_{16} + B_{17} + B_{18}} \tag{5.5}$$

where S and B are the signal and background yields in 2016, 2017 and 2018[1].

On the other hand, a single histogram in the signal region allows a finer binning therefore improving the result. However, fitting the single distribution do not exploit as best the year by year differences. In fact, several CMS operating conditions are different between the various year, and they consequently affect the shapes of the single year discriminator distributions. Moreover, the DNN distribution is a function of the year since the year is one of the input variables. Merging different

[1] The inequality of equation 5.5 derives directly from the Cauchy-Schwarz inequality.

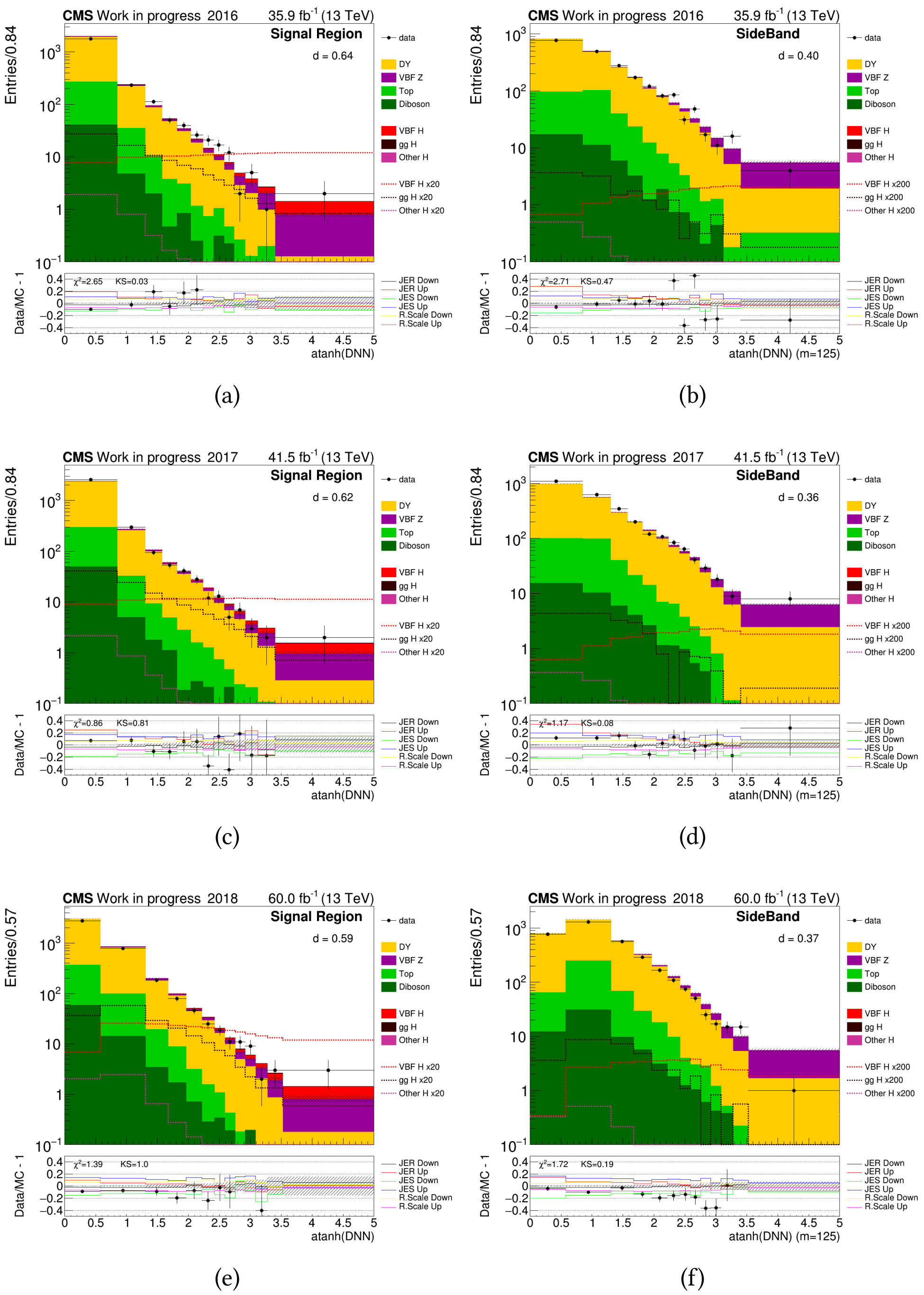


FIGURE 5.23: DNN distributions after the event selection for 2016, 2017, and 2018 in signal region and sideband. Captions are the same of Figure 5.3.

distributions reduce the signal-to-background separation, therefore fitting simultaneously the distributions of the three years produces the best result.

A crucial point of the signal extraction is the binning choice. In fact, as explained in section 5.5.1, the final result depends on the binning. The appropriate bin boundaries are chosen in order to achieve optimal sensitivity while minimizing the total number of bins. In each data-taking period, the DNN output distribution of the VBF H sample is used to define thirteen bins. The bins are defined starting from the rightmost one which is set to have a fixed signal yield of 0.6 expected events. All the other bins are defined such that the signal yields change linearly down to the leftmost bin. This procedure defines different bin boundaries in each data-taking period. The explained binning construction is applied to the distributions of the DNN in the signal region separately for the three years. The binnings that are found by this procedure are applied also in the two control regions. The DNN scores of the simulated signal events usually are close to 1, therefore the described binning construction produces several bins in the high score region and few in the low score region. In order to zoom the distribution in the signal-like part of the distribution and obtain a readable view of the DNN output, the transformation DNN$'$ = atanh(DNN) is applied to the DNN output. The same transformation is applied to the bin boundaries in order to mantain the binnings of the DNN distributions.

The expected results are extracted from the distributions shown in Figure 5.23. The presence of differences among the distributions of different years is highlighted by the d parameter shown in each plot. The DNN and BDT output distributions of the three regions are shown in appendix F.

5.7.1 Statistical treatement

The signal extraction is performed by maximizing a likelihood function that takes into account all the systematic uncertainties. The procedure that computes the exclusion limits is called $\mathrm{CL_s}$ method and it is described in Ref. [158].

Each uncertainty is parametrized by a nuisance parameter θ_i and it is described by a probability density function (pdf) $p(\theta_i|\tilde{\theta}_i)$, where $\tilde{\theta}_i$ is the initial estimate of the parameter. The systematic uncertainties are usually modeled by Gaussian or log-normal pdfs that are defined by the up and down variation histograms. The results are expressed in terms of the signal strength modifier μ, defined as the ratio between the measured cross-section and the Standard Model cross-section. In all the steps of signal extraction, only positive signal strengths are considered.

The likelihood is a function of the signal strenght and the nuisance parameters, and it is defined by

$$L(\,\mathrm{data}\,|\,\mu,\theta\,) = \prod_k \mathrm{Poisson}\,(\,n_k\,|\,\mu S_k(\theta) + B_k(\theta)\,) \;\times\; \prod_i p(\theta_i|\tilde{\theta}_i) \qquad (5.6)$$

where k represents the bin, and S_k and B_k are the signal and background expected yields in the k-th bin. The parameters that correspond to the global maximum of the likelihood are called $\hat{\mu}$ and $\hat{\theta}$. For fixed factors $\mu > \hat{\mu}$ the profile likelihood ratio is defined by:

$$\tilde{q}_{\mu} = -2\ln\left(\frac{L(\text{ data} \mid \mu, \hat{\theta}_{\mu})}{L(\text{ data} \mid \hat{\mu}, \hat{\theta})}\right) \tag{5.7}$$

where $\hat{\theta}_{\mu}$ is the parameter that maximize the likelihood at fixed μ.

A test statistic is constructed to compare the compatibility of data with the background+signal and background-only hypothesis. Firstly, the parameters $\hat{\theta}_{\mu}$ that best describes the observed data for the chosen μ and $\hat{\theta}_0$ that best describes the observed data for the background-only hypothesis ($\mu = 0$) are used to generate two sets of toy Monte-Carlo pseudodata. The profile likelihood ratio is computed for each toy and the probability distributions of the likelihood ratios $f(\tilde{q}_{\mu}|0, \hat{\theta}_0)$ and $f(\tilde{q}_{\mu}|\mu, \hat{\theta}_{\mu})$ are found.

The constructed probability distributions are used to compute the probability associated to the two hypotesis:

$$p_{\mu} = P(\tilde{q}_{\mu} \geq \tilde{q}_{\mu}^{obs} \mid \text{background+signal}) = \int_{\tilde{q}_{\mu}^{obs}}^{\infty} f(q|\mu, \hat{\theta}_{\mu})dq$$

$$p_b = P(\tilde{q}_{\mu} < \tilde{q}_{\mu}^{obs} \mid \text{background-only}) = \int_{0}^{\tilde{q}_{\mu}^{obs}} f(q|\mu, \hat{\theta}_{\mu})dq$$

where $\tilde{q}_{\mu}^{obs}$ is the observed profile likelihood for the desired signal strength μ. The two values are the probability of measuring a signal strength greater than the observed one in the background+signal hypothesis and the probability of measuring a signal strength lower than the observed one in the background+only hypothesis. The ratio of the probabilities

$$CL_s = \frac{p_{\mu}}{1 - p_b} \tag{5.8}$$

is the limit exclusion of the signal strength modifier μ. The upper limit on the signal strength at 95% confidence level $\mu^{95\%\text{CL}}$ is found by finding the parameter μ that provide CL_s= 0.05. The p-value is defined for $\mu = 1$ as

$$\text{p-value} = \frac{p_1}{1 - p_b} \tag{5.9}$$

and it is the probability that a fluctuation larger than the observed data is reproduced by the background-only model. The p-value is often converted into the equivalent significance Z, defined by the area under one tail of the Gaussian

$$\text{p-value} = \frac{1}{\sqrt{2\pi}} \int_{Z}^{\infty} e^{-x^2/2} dx \tag{5.10}$$

The 3σ (Z = 3) and 5σ significances (Z = 5) correspond to a probability of $1.3\,10^{-3}$ and $2.9\,10^{-7}$ respectively.

In order to find the expected 95% confidence level upper limit, a large amount of toy Monte-Carlo pseudodata that model the simulated samples of the analysis is generated. The upper limit $\mu^{95\%\mathrm{CL}}$ is computed for each toy and its distribution is built. The median, 68% band, and 95% band define the nominal, $\pm 1\sigma$, and $\pm 2\sigma$ expected upper limit.

Chapter 6

Results

In this chapter, the final results of the H→ $\mu\mu$ searches are presented. The signal significance and upper limit obtained from the VBF dedicated search are shown. Then, the results of the other three dedicated searches are reported and they are combined with the ones of the VBF search. Finally, the analysis is combined with the data recorded by CMS at 7 TeV and 8 TeV and the measurement of the Higgs boson coupling to muon is given. The results are extracted for the SM Higgs boson with m_H=125.38 GeV.

6.1 VBF Results

The signal strength is extracted by performing a simultaneous maximum likelihood fit over the signal region and the sideband of the three data-taking periods. The DNN score is fitted in the three signal regions and the DNN evaluated with fixed dimuon mass is fitted in the three sidebands. The histograms obtained from the Monte Carlo samples are used in the fit to predict the signal and background shapes. The results for the signal yields are expressed in terms of the signal strength modifier $\mu = (\sigma\mathcal{B}(H \to \mu\mu))_{\text{obs}}/(\sigma\mathcal{B}(H \to \mu\mu))_{\text{SM}}$ in order to have a direct comparison with the Standard Model.

The same fit is used to compute also the upper limit and the significance in order to quantify the compatibility of data with the background-only and the signal+background model. Upper limit, significance, and signal strength are extracted either from each year independently and simultaneously from the three years.

The theoretical uncertainties are treated as fully correlated among regions and periods. The experimental uncertainties are instead considered correlated among regions and uncorrelated among periods.

The expected and observed significances are shown in table 6.1 for each fit. The best fit signal strength for the Higgs boson with mass 125.38 GeV is $\hat{\mu} = 1.36^{+0.69}_{-0.61}$, corresponding to an observed significance of 2.40 σ. Data and simulated DNN distributions in signal region and sideband are shown in Figure 6.1. The lower panels show the pre-fit and post-fit ratios between data and simulation. The grey area indicates the MC uncertainty and the blue histogram shows the best fit signal contribution.

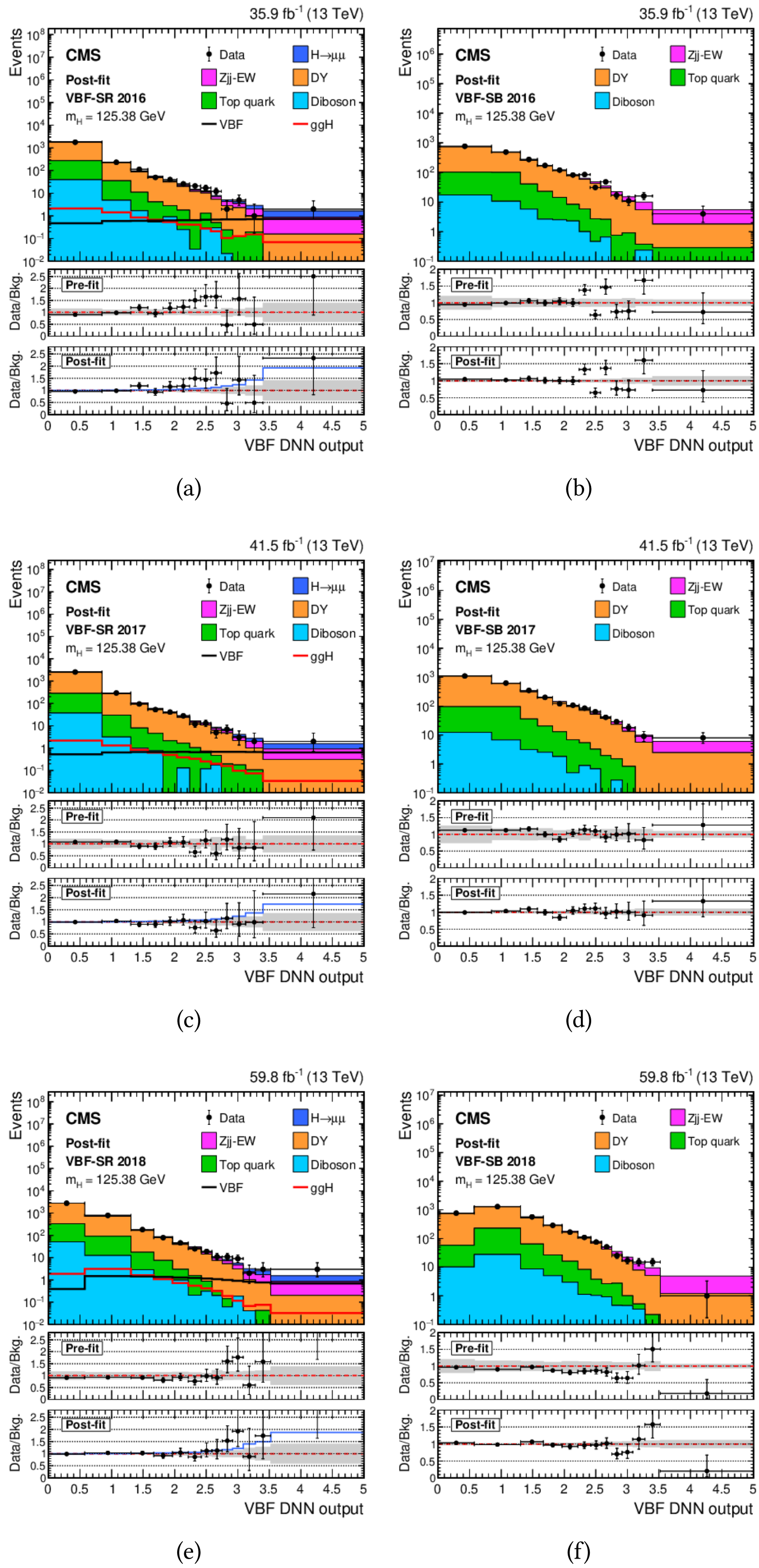


FIGURE 6.1: DNN distributions for 2016 (first row), 2017 (second row), and 2018 (third row) in signal region (first column) and sideband (second column). Signal and background predictions are obtained from the fit performed across the entire data-taking period. The lower panels show the residuals after the pre-fit and post-fit background subtraction. The grey bands show the background uncertainty while the blue lines indicate the best fit signal contribution.

	2016	2017	2018	combined
significance	2.01(1.04)	0.00(0.88)	2.21(1.21)	2.40(1.77)

TABLE 6.1: Expected and observed signal significances for the exclusion of the background-only hypothesis for m_H=125.38 GeV for each year and for the entire data-taking period.

Figure 6.2 shows the expected and observed upper limits obtained from the fit for each year and for the entire data-taking period. Overall the uncertainties are dominated by statistical effects. The leading systematic uncertainties are the ones originating from the parton shower modeling of VBF H and VBF Z processes. The leading experimental uncertainties are those arising from the normalization of the DY with noise and pile-up jets.

The distributions of the DNN score for the entire data-taking period in the signal region and in the sideband are computed by summing the distributions of each year. Since the bin boundaries adopted to extract the results are optimized separately per data-taking period, the combined DNN distributions are computed by summing the observed and predicted yields bin by bin. Figure 6.3 shows the obtained DNN distributions.

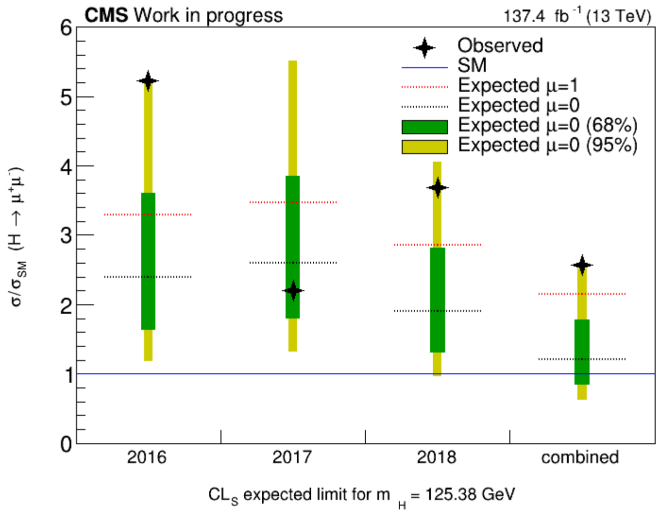
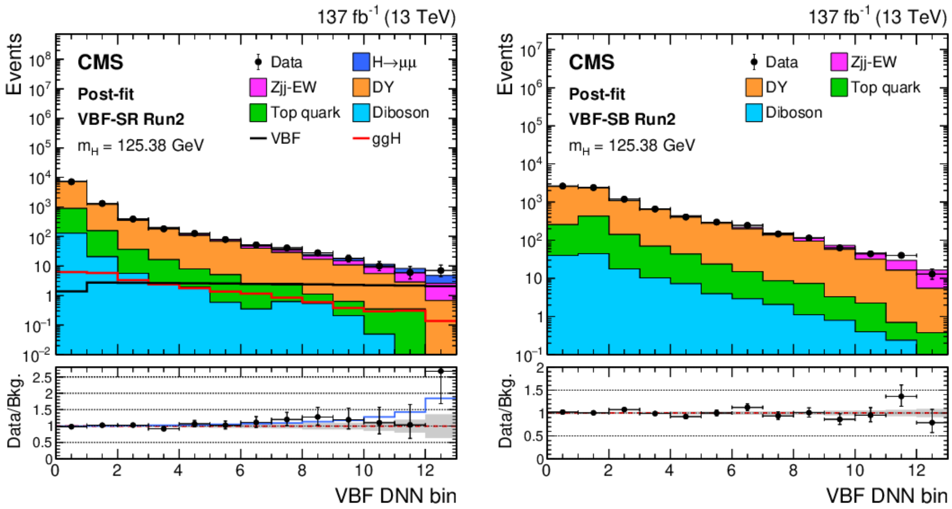


FIGURE 6.2: Expected and observed upper limit exclusions for the background-only hypothesis for m_H=125.38 GeV for each year and for the entire data-taking period.

6.2 Combined Results

The result of the search in VBF channel is combined with the ones coming from ggH, VH, and ttH searches to maximize the sensitivity. A simultaneous fit is performed across all events subcategories of the four analyses: the subcategories described in chapter 4 for ggH, VH, and ttH searches, and signal region and sideband for the VBF search. A single value for the signal strength is extracted from the fit while the relative contributions from different Higgs production modes are fixed to the SM



(a) (b)

FIGURE 6.3: DNN distributions in signal region (a) and sideband (b) for the combination of 2016, 2017, and 2018 data. Signal and backround predictions are obtained from the fit preformed across the entire data-taking period. The lower panel shows the residuals after post-fit background subtraction. The grey band shows the background uncertainty while the blue line indicates the best fit signal contribution.

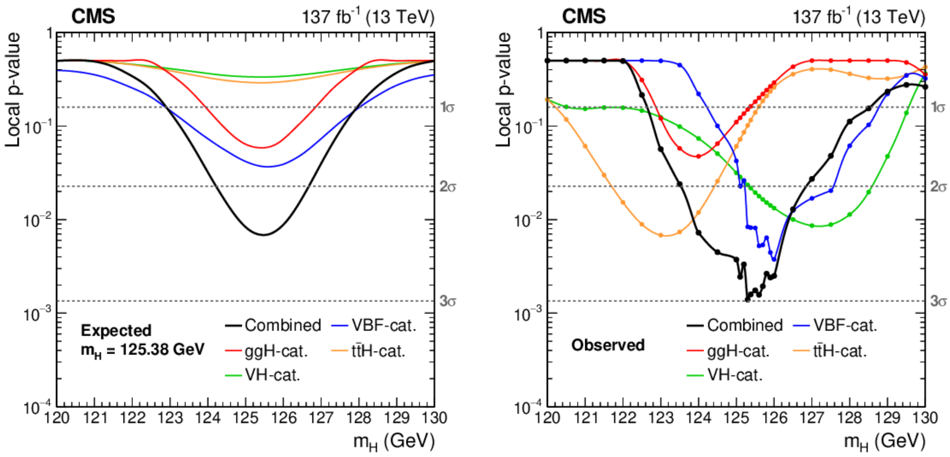
prediction.

The systematic uncertainties taken into consideration are those described in section 5.6. All the theoretical uncertainties are assumed to be fully correlated among all the categories. The acceptance uncertainties from the muon energy scale and resolution are treated as correlated across the ggH, VH, and ttH categories while they are uncorrelated with the same uncertainties in the VBF category. Moreover, their effects on the peak position and width are considered uncorrelated among the categories. The pile-up and noise contribution of DY acceptance is considered only in the VBF category. The remaining experimental uncertainties are treated as fully correlated among all the categories included in the fit.

The results on the combined search are quoted as p-value as a function of the Higgs mass. The expected and observed local p-values are computed for several mass hypotheses in the range 120-130 GeV for each individual search and for the combined fit. Figure 6.4 shows the p-values for 11 mass hypotheses within 125 and 126 GeV and for 18 mass hypotheses in the remaining windows. The points show the masses for which the observed p-values are computed while the expected p-values are computed with an injected signal with m_H = 125.38 GeV.

The four colored curves show the results obtained from the four searches independently. The black line indicates the results obtained from the combined fit. The VBF category provides the best expected result among the four searches in the entire mass range.

In the ggH, VH, and ttH categories, signal samples generated with m_H=120, 125,



(a) (b)

FIGURE 6.4: Expected (a) and observed (b) p-values as a function of m_H for each category and for the combined fit. The points show the masses for wich the the observed p-values are computed.

and 130 GeV are interpolated in order to provide signal models for any mass in the range 120-130 GeV. The parameters describing the signal shape are derived from those models, then a new fit is performed. In order to compute the correct signal yields at each mass hypothesis, the signal cross-section and branching fraction are modified as predicted from the Standard Model calculation [20].

Instead, in the VBF search, the p-values at different masses are computed by shifting the dimuon mass both for data and MC, similarly to the shift used to compute the results for a Higgs boson of 125.38 GeV. This is possible because of the little correlation between the DNN input variables and the dimuon mass. The mass shift is applied only to data and background, therefore the bin boundaries of the DNN histograms are not changed since they depend only on the signal distribution. The p-values are derived from the simultaneous fit of the mass biased DNN distributions in signal region and sideband. The discontinuities observed in the p-value trend of the VBF category are due to event migrations in data between different bins when evaluating the DNN for different mass hypotheses.

The obtained best fit signal strength for the Higgs boson with mass 125.38 GeV is $\hat{\mu} = 1.19^{+0.41}_{-0.40}(\text{stat})^{+0.17}_{-0.16}(\text{syst})$, corresponding to an excess of events over the background expectation with a significance of 3.0 σ. The signal strength measurement refers to the product of the cross-sections with the branching fraction of the Higgs boson to muon. However, assuming the cross-section to be the Standard Model predicted ones, the branching fraction is constrained to be $0.8 \cdot 10^{-4} < \mathcal{B}(H \to \mu\mu) < 4.5 \cdot 10^{-4}$ at 95% confidence level.

The best fit signal strength is shown in Figure 6.5 with its 1σ and 2σ CL intervals. The black markers show the signal strength obtained from a profile likelihood

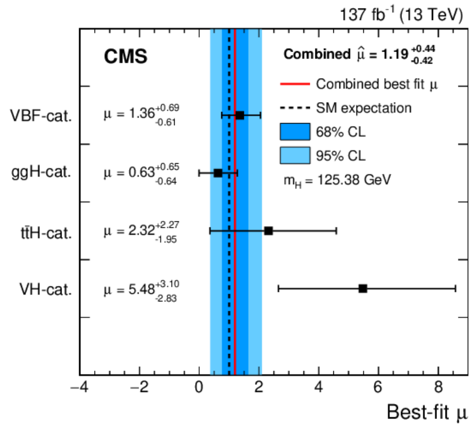


FIGURE 6.5: Observed signal strength for each category and for the combination with the associated uncertainty.

scan in each category, and the associated uncertainties indicate the 68% CL interval. Table 6.2 shows the statistical and systematic contribution to the post-fit signal strength uncertainty. The table shows also the individual contribution of the limited size of MC events, the experimental and theoretical systematics. The statistical component is computed by performing a likelihood scan in which the systematic nuisance parameters are fixed to their post-fit values. The systematic component is taken as the difference in quadrature between the total uncertainty and the statistical one. The other contributions are evaluated with a similar method. The contribution of the limited size of MC events only affects the VBF category.

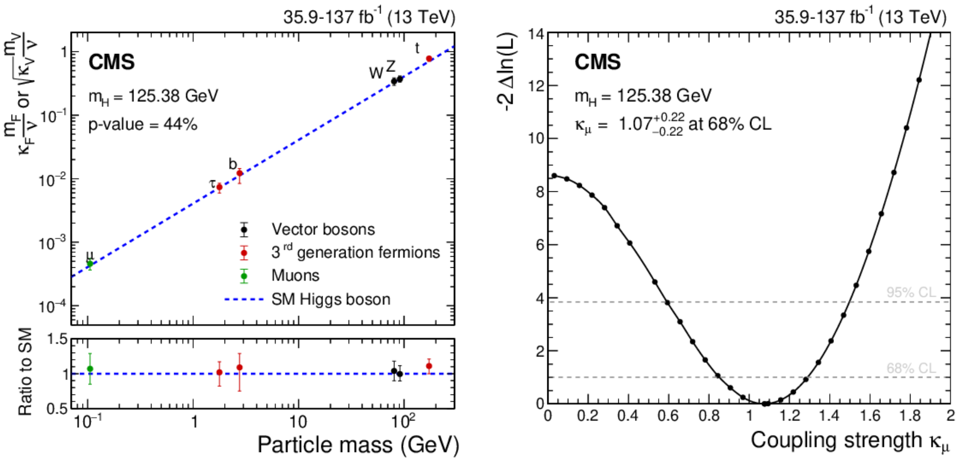
Source of uncertainty	$\Delta\mu$	
total post-fit uncertainty	+0.44	-0.42
Statistical uncertainty	+0.41	-0.40
Systematic uncertainty	+0.17	-0.16
Experimental uncertainty	+0.12	-0.11
Theoretical uncertainty	+0.10	-0.11
MC limited size uncertainty	+0.07	-0.06

TABLE 6.2: Impact of the main source of uncertainties in the measurement of the signal strength.

Higgs to muon coupling

The Higgs coupling to muon cannot be directly extracted from the presented signal strength because the Higgs production cross-sections depend on several other couplings of the Higgs with other Standard Model particles, mainly t-quark and vector bosons. Therefore a unified fit is performed in order to constrain all the available couplings as well as the Higgs to muon one. The Higgs couplings are constrained by combining the presented results with the ones of Ref. [159]. No additional Beyond Standard Model particle is considered. The k-framework introduced in chapter 1 is used: six Higgs coupling modifiers κ_μ, κ_τ, κ_b, κ_t, κ_Z, and κ_W are considered to parametrize the signal strength and they are left free left to float independently in a simultaneous likelihood fit across all the event categories.

The best fit values of the coupling parameters are in agreement with the Standard Model prediction and they are shown as a function of the particle mass in Figure 6.6a. Figure 6.6b shows the best fit and the profile likelihood of κ_μ. The corresponding 68% and 95% CL intervals are shown in the likelihood scan and they are $0.85 < \kappa_\mu < 1.29$ and $0.59 < \kappa_\mu < 1.50$ respectively. Figure 6.6a can be directly compared with Figure 1.8.



(a) (b)

FIGURE 6.6: (a) Observed profile likelihood ratio as a function of κ_μ with the associated 68% CL interval. (b) Best fit for the coupling modifiers and comparison with the Standard Model prediction. The error bars correspond to the corresponding 68% CL interval.

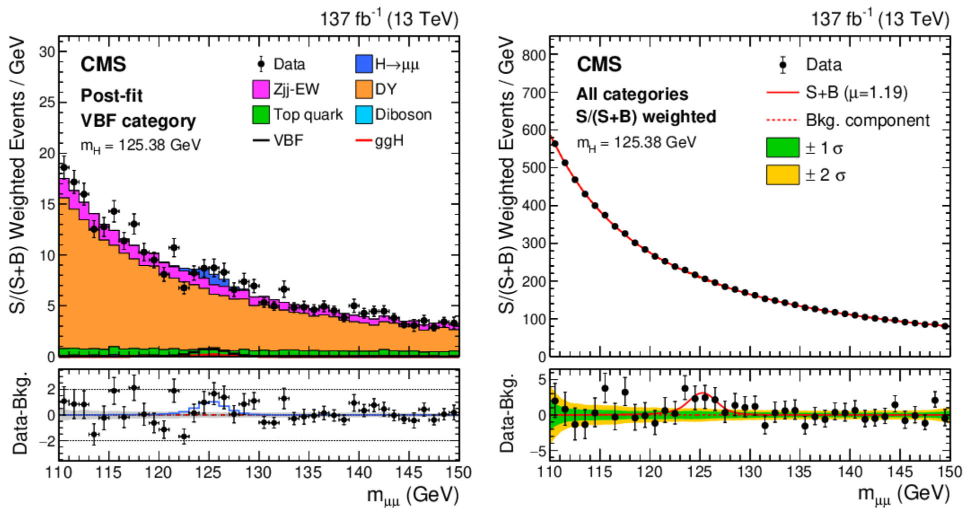
Weighted mass distribution

The major signal signature of H$\rightarrow$ $\mu\mu$ signal in all the four channels of the analysis is the dimuon invariant mass peak around 125 GeV. However, the expected signal peak is not significantly visible in any of the analysis categories because of the huge amount of background. Cumulative mass distribution is computed enhancing the signal-like events in order to visualize the signal as a peak in the dimuon mass over the monotone background trend. A mass unbias weight is applied to all events of all the categories, then a single mass distribution is computed.

The parametric mass distributions of the subcategories of the ggH, VH, and ttH dedicated searches shown in chapter 4 are summed to obtain a single distribution. Each data and MC event is weighted by the signal purity S/(S+B) under the signal peak within $\pm$HWHM of the corresponding subcategory.

In the VBF category, the distribution of $m_{\mu\mu}$ is produced by weighting properly the events using a mass independent DNN. A mass independent DNN distribution is computed by replacing the input value of the $m_{\mu\mu}$ with the fixed value of 125 GeV when evaluating the mass dependent DNN. The distribution of the mass independent DNN is computed using the same bin boundaries of Figure 6.1 and the purity S/(S+B) is computed for each bin. The dimuon mass shown in Figure 6.7a is obtained by weighting each event of both data and MC by the corresponding bin purity of the mass independent DNN. The distribution contains the events of the entire data-taking period from signal region and sideband.

The bin yields of the purity weighted mass distribution histogram of the VBF category are interpolated with a continuous and derivable function. The continuous function obtained with this method is summed to the purity weighted mass distributions of ggH, VH, and ttH categories and it is shown in Figure 6.7b. For the ggH, VH, and ttH categories, the signal and background are the sums of the post-fit functions obtained from the simultaneous fit. For the VBF category, they are obtained interpolating the signal and background post-fit histograms with a spline. The weighted data events in the distribution come mainly from ggH category, where the purity is small but there are many events. The observed data are compatible with an excess of events observed around 125 GeV.



(a) (b)

FIGURE 6.7: Weighted dimuon mass distributions for 2016, 2017, and 2018 data in the range 110-150 GeV in the VBF search (b) and for the combinations of all the categories (b). The lower panel shows the residuals after the background subtraction. The lower panel shows the residuals after post-fit background subtraction with the associated uncertainty and the best fit signal contribution.

6.3 Combination with Run 1

The presented results are combined with the proton-proton collision data at $\sqrt{s}$=7 and 8 TeV recorded by CMS during Run 1, following the strategy adopted in Ref. [160]. The 7+8 TeV analysis has been performed with an integrated luminosity of 5.1 fb^{-1} at $\sqrt{s}$=7 TeV and 19.7 fb^{-1} at $\sqrt{s}$=8 TeV and it is described in Ref. [44]. The search is updated with the improvement in the cross-sections and branching fractions predictions [20] and a simultaneous maximum likelihood fit is performed over the 7, 8, and 13 TeV data. The experimental uncertainties are conservatively taken uncorrelated between the 13 TeV and the 7+8 TeV analyses while the theoretical uncertainties are treated as fully correlated.

The combination produces an improvement of about 1% on both the expected and observed significance of an excess of events at m_H=125.38 GeV over the background-only expectation. The corresponding signal strength and 68% CL interval is $\hat{\mu} = 1.19^{+0.40}_{-0.39}(\text{stat})^{+0.15}_{-0.14}(\text{syst})$. Table 6.3 contains the observed and expected significances and upper limits derived from the analysis combination and from each category production. The expected upper limits are computed assuming the Standard Model without the Higgs Boson. Figure 6.8 shows the observed and expected p-values as a function of the Higgs boson mass hypoteses computed from the 7, 8, and 13 TeV combination similarly to Figure 6.4.

Production category	significance	upper limit
VBF	2.40(1.77)	2.57(1.22)
ggH	0.99(1.56)	1.77(1.28)
VH	2.02(0.42)	10.8(5.13)
ttH	1.20(0.54)	6.48(4.20)
Combined $\sqrt{s}$ = 13 TeV	2.95(2.46)	1.94(0.82)
Combined $\sqrt{s}$ = 7, 8, 13 TeV	2.98(2.48)	1.93(0.81)

TABLE 6.3: Expected and observed signal significances and upper limits for the exclusion of the background-only hypothesis for m_H=125.38 GeV for each category. The expected upper limits are computed witht the background-only hypoteses $\mu = 0$.

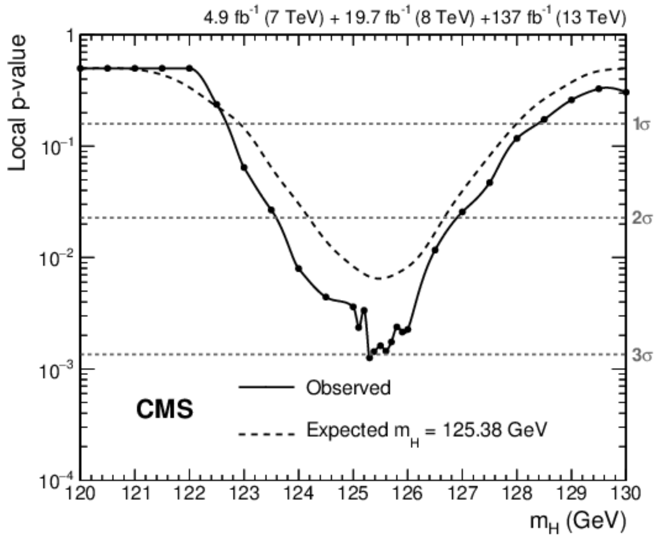


FIGURE 6.8: Expected and observed p-values as a function of m_H for the combination of 7, 8, and 13 TeV data. The points show the masses for wich the the observed p-values are computed.

6.4 Conclusion

The analysis described in this thesis presents the first evidence for the Higgs boson decay to second generation fermions. The search is performed using the proton-proton collision data at $\sqrt{s}$=13 TeV recorded by the CMS experiment in the period 2016-2018, corresponding to an integrated luminosity of 137 fb^{-1}. The results are in agreement with the Standard Model prediction and they provide the best constraint on the coupling between the Higgs boson and the muon to date.

This analysis largely improves on the previous measurement published by CMS and obtained with the 2016 data. The improvements are due both to the larger amount of data analyzed and an improved the analysis strategy. The previous measurement targeted mainly the ggH production mode and it did not exploit as best as possible the other production channels, especially the VBF one which turned out to be the most sensitive in this search. The largest improvements with respect to the previous result are due to the dedicated search for the VBF production and the analysis strategy adopted in this channel.

The analysis strategy adopted in the VBF channel is different from the one adopted in the other H$\rightarrow \mu\mu$ searches. The signal extraction is performed by fitting the output of a multivariate discriminator, in contrast with other approachs where the fit is performed on the dimuon mass distribution. Two different multivariate discriminators have been tested, in contrast with the other H$\rightarrow \mu\mu$ searches where only the BDT algorithm has been tried. A large improvement derives from the choice of the deep neural network, which shows to perform better than the BDT discriminator. In addition, further gains come from the effort devoted to the optimization of the the multivariate discriminator, in order to separate as best as possible signal from background. These elements provide an improvement in sensitivity of about 20% with respect to an alternative analysis where a data driven approach is adopted.

The results of the H$\rightarrow \mu\mu$ analysis are combined with the data collected at center of mass energies of 7 and 8 TeV, corresponding to integrated luminosities of 5.1 and 19.7 fb^{-1} respectively. The combination increases both the observed and expected results by 1%. The measured signal strength, relative to the Standard Model Higgs with mass 125.38 GeV, is $\hat{\mu} = 1.19^{+0.40}_{-0.39}(\text{stat})^{+0.15}_{-0.14}(\text{syst})$. The excess of events observed over the background expectation corresponds to a significance of 3.0 standard deviation, thus it constitutes the first evidence of the Higgs boson decay to muons.

6.5 Outlook

The LHC is preparing for a new period of data-taking that will take place between Spring 2022 and Autumn 2024, called Run 3. Thanks to the data collected during Run 3 the LHC experiments will have an integrated luminosity of about 300 fb^{-1}. After this period the LHC will prepare for the High-Luminosity LHC (HL-LHC) project. The HL-LHC is planned to start in 2027 and aims to collect 3000 fb^{-1} of proton-proton collisions. The new data that the LHC experiments will collect will be sufficient to significantly improve most of the Higgs boson measurement available today.

The data collected in the next years allow improving the Higgs to muon decay measurement. The dedicated VBF search is limited by the statistical uncertainties therefore collecting new data is the only way to improve significantly the result. In particular, the VBF search is expected to obtain the 3σ evidence with the Run 3 data. The entire H$\rightarrow\mu\mu$ analysis will be performed with the new data by both ATLAS and CMS collaborations in order to reach the 5σ significance similarly to the other Higgs couplings. The 5σ significance will be achieved by the CMS collaborations at the early stage of the HL-LHC data-taking. At the end of HL-LHC, the expected uncertainties on the Higgs-to-muon coupling will be about 10% [161].

Several projects for studying new phenomena at future accelerators are under discussion. The high energy LHC (HE-LHC) is a program designed to reach the center of mass energy of 27 TeV in the LHC tunnel [162]. Building a new accelerator for electrons (FCC-ee) or hadrons (FCC-hh) with a radius of about 100 Km is also under study [163, 164]. Each of these projects aims to collect new data and opens new possibilities for Higgs studies and for probing effective field theory models. In particular, the HE-LHC program would improve the H$\rightarrow\mu\mu$ precision of about 3 times while the FCC-ee is not expected to improve significantly the results obtained after HL-LHC. Instead, the FCC-hh is expected to measure the Higgs decay to muons with a precision of about 1%.

Several details of the H$\rightarrow\mu\mu$ search may change in the future. In the VBF channel, the MC based approach has shown better performance than the data driven approach. A comparison between the two approaches may be performed also in the other channels and, eventually, all the four channels may adopt the MC based strategy.

With the increase of data, differential measurements of the Higgs decay products are accessible in the muon decay channel. Several properties can be probed by measuring the Higgs differential cross-sections; for example, the Higgs spin and parity can be studied and confirmed in this channel, where the Higgs decay products are directly measured. Moreover, differential cross-section measurements are sensitive to beyond the Standard Model physic.

The evidence of the Higgs to muon decay is the first step in the validation of the Higgs coupling to muons. However, also the couplings with the first generation and the remaining second generation fermions have to be confirmed. The couplings of the Higgs boson with the light quarks and the electron are very small, and there is not any defined plan that would reach to probe of the Standard Model expected coupling values. Instead, the coupling of the Higgs boson with the c-quark could make significant progress in the next decades.

The Higgs decay to c-quark pairs has a higher event rate than the Higgs decay to muons. However, the measurement is challenged by the large multijet final state events produced by the LHC. Only upper limits have been set on the Higgs to c-quarks in the last years. The CMS and ATLAS collaborations will keep searching for this decay during the HL-LHC and they are expected to exclude about 6 times the Standard Model expected production [165].

Appendix A

Collin-Soper frame

The Drell-Yan (DY) process is the production of a lepton pair through Z or γ boson. The Higgs decaying in two muons have exactly the same final state particles, therefore DY is, in principle, an irreducible background for the $H \to \mu\mu$ searches. However, since Z and γ are spin 1 bosons and the Higgs boson have spin 0, the lepton pairs angular distributions are different in the two processes.

The angular distributions of the leptons are typically expressed in the rest frame of the dilepton system, i.e. in the frame where the original boson was at rest. Several choices of coordinate systems have been considered in the literature, the one used in this work is Collin-Soper's coordinate system[166]. In the Collin-Soper frame, the hadron plane is identified with the plane spanned by the two protons (or hadrons in general) momenta. The y-axis is defined as the normal vector to the hadron plane, while the z-axis is defined as the axis bisecting the momentum of one of the proton and the inverse of the momentum of the second one. The sign of the z-axis is defined as the sign of the z-component of the lepton pair momentum in the laboratory frame. The x-axis is chosen in order to have a right-handed cartesian coordinate system.

The polar and azimuthal angles θ and ϕ are the angles identifying the negatively charged lepton. The two leptons have opposite directions and equal momentum in this frame, therefore also the direction of the positively charged lepton is identifiable by θ and ϕ. Figure A.1 shows the Collin-Soper frame and the θ and ϕ angles.

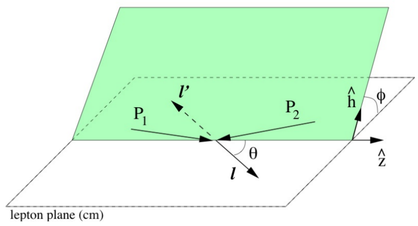
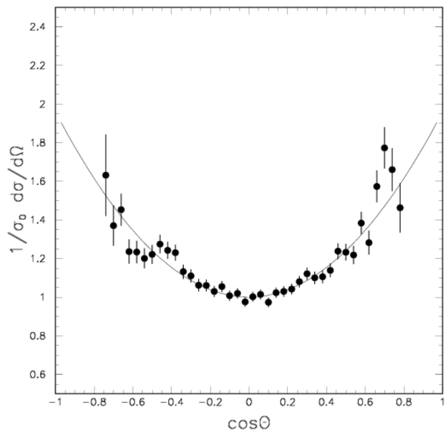


FIGURE A.1: Illustration of the Collin-Soper frame. The angles θ and ϕ are shown [167].

A spin 0 decay has not any preferential direction, thus the lepton direction has a constant angular distribution. The DY leptons distributions instead are affected by the spin of the Z/γ and it is not constant. The angular distribution in the Collin-Soper frame is given by

$$\frac{d^2\sigma}{d\Omega} \propto (1 + \cos^2\theta) + A_0\frac{1}{2}(1 - 3\cos^2\theta) + A_1 \sin(2\theta)\cos\phi + A_2 \sin^2\theta\cos(2\phi) + \\ + A_3 \sin\theta\cos\phi + A_4\cos\theta + A_5\sin^2\theta\sin(2\phi) + A_6\sin(2\theta)\sin\phi + A_7\sin\theta\sin\phi$$

where the A_i are second order coefficients coming from the scattering calculation and, except for A_4, they vanish as the boson transverse momentum approach zero. Figure A.2 shows the DY distribution of the θ angle measured in $p + Cu$ collision at 800 GeV. The events have $11 < m_{\mu\mu} < 17$ GeV, the trend as been fitted with the shape $1 + \lambda \cos^2\theta$ and it shown in the figure.



(a)

FIGURE A.2: Drell-Yan angular distribution from Fermilab $p + Cu$ collisions at 800 GeV. The solid line is a ft to the data with $1 + \lambda \cos^2\theta$ where λ is $0.96 \pm .04 \pm .06$ [168].

Unfortunately, multivariate analyses cannot easily exploit such complex distribution differences. Moreover, the A_i are functions of p_T and rapidity of the boson, in particular A_0 reverses the $\cos^2\theta$ trend for $p_T \sim 80$ GeV [169]. In addition, due to the acceptance of the CMS detector, the distribution of θ is sculpted and it reduces the differences between Higgs and DY process. Nevertheless, the angular distributions of θ and ϕ can help in discriminating between Higgs and DY, and therefore they have been used in $H \to \mu\mu$ searches.

Multivariate analyses used in $H \to \mu\mu$ searches can take dimuon system kinematic variables as input. Since each muon momentum has three degrees of freedom, the kinematic of the dimuon system is described by 6 variables. The ϕ-symmetry of the detector makes one degree of freedom redundant. Three variables are the invariant mass $m_{\mu\mu}$, the transverse momentum $p_T^{\mu\mu}$, and the pseudorapidity $\eta_{\mu\mu}$ of the two muons. The remaining two variables correspond to the decay of the boson in the rest frame of the dimuon system and, coherently in the four searches, Collin-Soper frame angles θ and ϕ have been used and called θ_{CS} and ϕ_{CS}.

Appendix B

Boosted Decision Tree

B.1 Introduction

Decision trees are tree-like structures of decision commonly used in decision analyses. A decision tree checks a sequence of conditions about event attributes and it classifies events belonging to a single sample into different classes accordingly to the conditions results.

Figure B.1 shows an example of a decision tree where the events are classified in signal or background. The classification starts from a root node that tests an attribute of the event. The result of the first condition leads to a second node that tests a different condition. The second result leads to the third node and so on. At the end, the events are divided into several subsets, called leaves. The events in the leaves can be associated to classes or to real numbers.

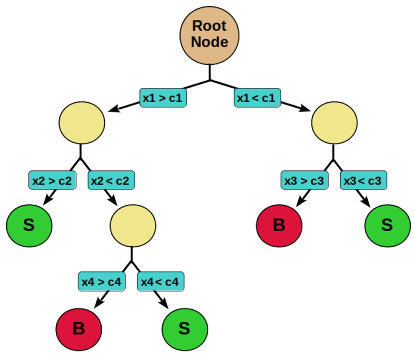


FIGURE B.1: Sketch of a dicision tree.

Since the computer age began and the computing power went across rapid improvements, a huge amount of machine learning algorithms have been developed. Machine learning algorithms are programs able to make predictions or decisions without being explicitly programmed but learning automatically through the experience. Several programs among those can build decision trees. Those programs usually use boosting methods that produce more robust decision trees. The decision trees built with the boosting method are called boosted decision trees (BDTs).

BDTs have been widely used in high energy physics in recent years since they are fast and reliable even when trained on small datasets. In particular, they have been employed in several Run 1 analyses at LHC. BDTs are also used in $H \to \mu\mu$ search since they are able to discriminate between signal and background exploiting simultaneously several features of the events. The BDTs built for $H \to \mu\mu$ search have been implemented with the TMVA package [107] in the ROOT framework [146].

In order to build a BDT, the machine learning program has to learn from a set of examples called training dataset. The process that uses the training dataset and defines the discriminator is called training. The training algorithms used in building the various BDTs used in $H \to \mu\mu$ search have little differences from each other in methods and in settings. In the following, the boosted decision tree training algorithms are described.

B.2 Training

In order to build a BDT, two main ingredients are needed: the training dataset and the input variables list. The performance of the BDT depends on the amount of signal and background examples provided, therefore having a large number of training events is very important for the final result. Moreover, in order to define the splitting criteria for each node, a list of event attributes to use is needed. The splitting criteria will be a cut on one of the variables used as inputs.

The algorithm proceeds iteratively starting from the root node. An initial splitting criterion is defined using the entire training dataset. The split is determined by finding the variable and the corresponding cut which perform the best separation between signal and background. The two resulting subsets go under the same algorithm thus they are further divided into two by different splitting criteria. The algorithm is repeated until the tree is built.
In principle, the iterations could continue until every leaf contain signal only or background only events. In that case, the decision tree would be strongly overtrained and therefore unable to make predictions. In order to avoid such an issue, a stop condition is set. Among the various possibilities, the BDTs built for $H \to \mu\mu$ search have constraints on the depth of the tree, i.e. the maximum number of nodes between the root node and the leaves.

The splitting criteria are found by optimizing a gain function G that depends on a certain function of the purity $f(p)$. The purity is defined as the ratio of the signal events over the total events

$$p = \frac{S}{S + B}$$

where the events can be weighted. For each input variable and for each cut, the gain function

$$G(p, p_1, p_2) = f(p) - f(p_1) - f(p_2)$$

is computed, where p_1 and p_2 are the purity of the two subsets divided by the cut. The variable and the cut that maximize G is chosen as the node splitting criteria. Several functions $f(p)$ are commonly used in BDT building. The ones used in $H \to \mu\mu$ search are the Gini index and the cross-entropy. The Gini index is $f(p) = p(1 - p)$ and it correspond to the probability of misidentifying signal as background or vice versa in the split sample. The cross entropy method use $f(p) = -p\ln(p) - (1 - p)\ln(1 - p)$ and it represents the average of the logarithm of the accuracy.
In order to avoid the node to provide an unbalanced splitting, an additional condition is required. The two subsamples produced by the splitting have to be greater

than a certain fraction of the original sample. Usually the chosen fraction float around a few percent.

The final tree divides the original phase space in many regions. The output of the tree is defined for every region and it can be either a real variable or just a classification in signal or background, depending on the majority of training events that end up in the final leaf. In the latter case, the output is set to be +1 for signal and -1 for background.

BDTs built with the described procedure are unstable to small fluctuations in the input dataset. In fact, small changes can give rise to great differences in the tree structure. The adopted solution is using an ensemble of trees and combine their output into a single result. The ensemble method used is called boosting and it consists of training several different trees by reweighting the training events. Boosting methods stabilize the response of the decision tree and it improves the performance with respect to single tree construction.

B.2.1 Adaptive boost

The adaptive boost, or AdaBoost [170], has been developed in the mid-90s and it is the first boosting algorithm. The adaptive boost grows an ensemble of trees and it combines them into a single function. The trees are trained sequentially and the event weights are modified at each iteration in order to give more importance to the previous misclassified events. This procedure lets adaptive boost be sensitive to noisy data and outliers and less sensitive to overfitting.

Let $\{x_i, y_i\}_1^N$ be the dataset, where the item x_i are event attributes and their associated class is $y_i \in \{-1, +1\}$. Let w_1^i be the original weights of the events. The algorithm starts training a tree using the original weights and comparing the output $h_1(x_i)$ with the true class y_i. The misclassification weight ϵ_1 defined by

$$\epsilon_1 = \frac{\sum_{y_i \neq h_1(x_i)} w_1(x_i)}{\sum_{y_i = h_1(x_i)} w_1(x_i)} \tag{B.1}$$

is used to compute the boost weight

$$b_1 = \frac{1 - \epsilon_1}{\epsilon_1} \tag{B.2}$$

Then, the misclassified events ($h_1(x_i) \neq y_i$) are weighted by b_1 and all the events weights are renormalized such that the sum of the weights remains constant. A second tree is trained with the new weights and the procedure is repeated several times.

After that the desired number of tress have been trained, they are combined in a single function $F(x)$

$$F(x) = \frac{1}{M < \ln(b) >} \sum_{1}^{M} \ln(b_m) h_m(x) \tag{B.3}$$

A boosted decision tree is the combination of several trees and its output is the real number $F(x)$ in $[-1, 1]$. The output is close to +1 for signal-like events and close to -1 for background-like events.

B.2.2 Adaptive boost as loss function minimization

The adaptive boost procedure described above can be reformulated as the minimization of a function called loss function [171, 172]. The loss function $L(F(x), y)$ is sensitive to the deviations of the predicted results $F(x)$ with the true values y, therefore it is a measurement of the accuracy of a decision tree. The loss function used in adaptive boost is

$$L(F(x), y) = \sum_{i=1}^{N} e^{-y_i F(x_i)} \tag{B.4}$$

The loss function is minimized by constructing iteratively the boosted decision tree output $F(x)$. After m-1 iteration, the function $F(x)$ will bee the combination of m-1 trees:

$$F_{m-1}(x) = \alpha_1 h_1(x) + ... + \alpha_{m-1} h_{m-1}(x) \tag{B.5}$$

At the m-step another tree is added in the combination, therefore

$$F_m(x) = F_{m-1}(x) + \alpha_m h_m(x) \tag{B.6}$$

The new tree output $h_m(x)$ has to minimize the loss function which can be managed as

$$L = \sum_{i=1}^{N} e^{-y_i F_m(x_i)}$$

$$L = \sum_{i=1}^{N} e^{-y_i F_{m-1}(x_i)} e^{-y_i \alpha_m h_m(x_i)}$$

$$L = \sum_{i=1}^{N} w_i^{(m)} e^{-\alpha_m y_i h_m(x_i)}$$

where $w_i^{(m)} = e^{-y_i F_{m-1}(x_i)}$. The sum can be split in the events that are correctly classified, thus $y_i h_m(x_i) = 1$, and the events that are wrongly classified where

$y_i h_m(x_i) = -1$.

$$L = \sum_{y_i = h_m(x_i)} w_i^{(m)} e^{-\alpha_m y_i h_m(x_i)} + \sum_{y_i \neq h_m(x_i)} w_i^{(m)} e^{-\alpha_m y_i h_m(x_i)}$$
$$L = \sum_{y_i = h_m(x_i)} w_i^{(m)} e^{-\alpha_m} + \sum_{y_i \neq h_m(x_i)} w_i^{(m)} e^{\alpha_m}$$
$$L = \sum_{i=1}^{N} w_i^{(m)} e^{-\alpha_m} + \sum_{y_i \neq h_m(x_i)} w_i^{(m)} \left(e^{\alpha_m} - e^{-\alpha_m}\right)$$
$$L = e^{-\alpha_m} \left(\sum_{i=1}^{N} w_i^{(m)}\right) + \left(e^{\alpha_m} - e^{-\alpha_m}\right) \left(\sum_{y_i \neq h_m(x_i)} w_i^{(m)}\right)$$

Since the $w_i^{(m)}$ are fixed at the m-th iteration, the loss function is minimized by choosing the appropriate α_m and $h_m(x)$. In order to minimize the second right-hand side sum, $h_m(x)$ is chosen as the output of the tree trained with the event weight $w_i^{(m)}$. The α_m parameter is found by

$$\frac{\partial L}{\partial \alpha_m} = 0 \tag{B.7}$$

and it is

$$\alpha_m = \frac{1}{2} \ln \left(\frac{\sum_{y_i = h_m(x_i)} w_i^{(m)}}{\sum_{y_i \neq h_m(x_i)} w_i^{(m)}} \right) = \frac{1}{2} \ln \left(\frac{1 - \epsilon_m}{\epsilon_m} \right) \tag{B.8}$$

where ϵ_m is defined for the m-th step similarly to equation B.1. This procedure defines the trees $h_m(x)$ and the parameters α_m[1] used in the combination:

$$F(x) = \frac{1}{M < \alpha >} \sum_{1}^{M} \alpha_m h_m(x) \tag{B.9}$$

In order to demonstrate the equivalence of the two methods, the weights $w_i^{(m)}$ have to be shown they are the ones used in section B.2.1. Such validation is done by verifying that the misclassified events gain the boost defined in equation B.2. The weights definition lead to

$$w_i^{(m+1)} = w_i^{(m)} e^{-\alpha_m y_i h_m(x_i)} \tag{B.10}$$

Let i and j be a misclassified and a correctly classified event respectively. Then:

$$\frac{w_j^{(m+1)}}{w_i^{(m+1)}} = \frac{w_j^{(m)} e^{-\alpha_m y_j h_m(x_j)}}{w_i^{(m)} e^{-\alpha_m y_i h_m(x_i)}} = \frac{w_j^{(m)}}{w_i^{(m)}} e^{2\alpha_m} = \frac{w_j^{(m)}}{w_i^{(m)}} \left(\frac{1 - \epsilon_m}{\epsilon_m} \right) \tag{B.11}$$

[1] The misclassification weight ϵ_m is smaller than 0.5 by definition, therefore the parameter α_m is always positive.

Therefore the misclassified events gain a boost of $b_m = \frac{1-\epsilon_m}{\epsilon_m}$ at the m-th iteration.

B.2.3 Gradient boost

Gradient boosting algorithms are a generalization of the adaptive boost described above and they derive from the iterative minimization of a loss function[173]. The procedure is similar to a function expansion where every added term points to the negative gradient direction. Many loss functions can be chosen, and they fully determine the boosting procedure. Among all the possible loss functions, the one used in $H \to \mu\mu$ search is the binomial log-likelihood

$$L(F, y) = \sum_{i=1}^{N} \ln(1 + e^{-y_i F(x_i)}) \tag{B.12}$$

In general, not every the loss function leads to a straightforward procedure like the Adaboost algorithm, therefore a steepest-descent approach has to be used to perform the minimization.

The induction step finds $F_m(x)$ from $F_{m-1}(x)$. The procedure starts with the residual calculation:

$$r_i^{(m)} = \left[\frac{\partial L(F, y)}{\partial F}\right]_{F=F_{m-1}} \tag{B.13}$$

Then, a decision tree $h_m(x)$ is trained using the residuals $\{x_i, r_i^{(m)}\}_1^N$ and it is added to $F_{m-1}(x)$

$$F_m(x) = F_{m-1}(x) + \gamma_m h_m(x) \tag{B.14}$$

where the constant γ_m is found by minimizing the loss function

$$\frac{\partial L(F_m, y)}{\partial \gamma_m} = 0 \tag{B.15}$$

A very popular regularization method used in boosting algorithm is the regularization by shrinkage. It consists of reducing the weight of the individual trees in the final combination by multiplying them by a factor ν called learning rate.

$$F_m(x) = F_{m-1}(x) + \nu\gamma_m h_m(x) \qquad 0 < \nu \leq 1 \tag{B.16}$$

Boosting with a small learning rate requires more iteration and therefore it increases the computational time. However, small learning rates significantly improve the accuracy of the predictions.

The bagging is another technique very often used in gradient boost. It consists of using a different subsample of the original dataset at each iteration. The subsamples are composed of events randomly picked from the training dataset and their sizes are a constant fraction f of the training dataset. Bagging fractions f within 0.5 and 0.8 does not reduce significantly the statistical power of the training dataset while they introduce randomness in the algorithm and therefore they help prevent overfitting.

Appendix C

Deep Neural Networks

This appendix introduces machine learning and deep neural network concepts. The deep neural networks are a fundamental tool in the VBF H→ $\mu\mu$ analysis, where a feedforward neural network is trained on simulated data with dropout regularization. Therefore, in this appendix the neural network architectures, the used functions and the regularization methods are described.

C.1 Introduction

Machine learning is a field of computer science studying algorithms that automatically learn from experience. This kind of machine learning algorithms are very useful to face a multitude of tasks and they find application in several fields. The reason is that for simple tasks, programmers can specify all the steps needed to be solved. Instead, for more advanced tasks, manually writing the code needed to solve the problem may be challenging and it is more convenient to let the computer write its own algorithm to solve the task.

The most common tasks for which machine learning is employed are classification and regression problems. Classification algorithms are used when the desired output is restricted to a limited set of values. For example, they are used to separate the studied objects into different categories. In the case of two categories, the learning algorithm solves the tasks by producing a function $f : \mathbb{R}^n \to [0, 1]$. Regression algorithms aims to estimate the relationship between known variables and dependent unknown variables called outputs. In the case of one output variable, the algorithm solves the tasks by producing a function $f : \mathbb{R}^n \to \mathbb{R}$.

Machine learning algorithms have been used in high energy phisics since the 90s. In recent years, neural network algorithms have been used more and more in high energy physics thanks to the success in other fields. Nowadays, deep neural networks (DNNs) are slowly replacing BDTs, which where the most popular tool during Run 1 at the LHC, in either physics analysis and reconstruction tasks.

Thi appendix introduces the main concepts of the neural networks that are used in the H→ $\mu\mu$ search via VBF channel. The Neural network have been implemented in PYTHON with the KERAS [147] package and Tensorflow backend [148].

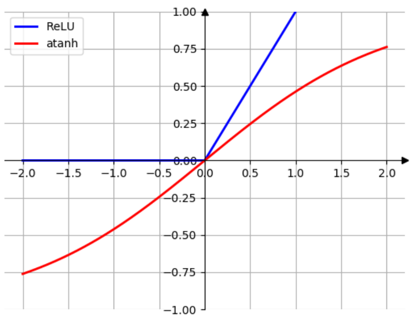
C.2 Deep neural networks

Deep feedforward neural networks are a particular type of machine learning algorithm. They are organized similarly to the biological neural networks and they are composed of basic units called artificial neurons. Artificial neurons are mathematical functions that take one or more inputs and return a single output. The functions applied by the neurons are a composition of two transformations: the first one is an affine transformation, and the second one is a non-linear activation function. Let $\vec{x}$ be the input vector of a certain neuron, the output can be written as

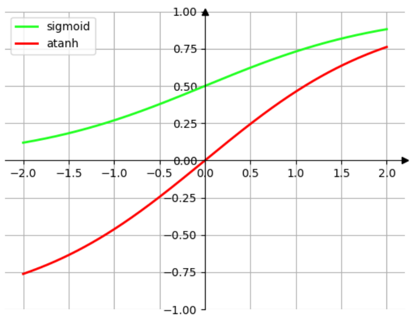
$$y = h(\vec{w} \cdot \vec{x} + \vec{b}) \tag{C.1}$$

where $\vec{w}$ and $\vec{b}$ are the weight and bias vectors and they are the tunable parameters of the neuron, while the function h is the activation function.

The design of activation functions is an active area of research and there is not any theoretical principle stating that one function is better than one other. In general, it is impossible to predict in advance which one works best for a given problem. Nowadays, the most used activation function is the rectified linear unit $\mathrm{ReLU}(x) = \max(0, x)$ which is linear for positive numbers and zero for negative ones. A similar activation function is the leaky ReLU, defined to be $\mathrm{LeakyReLU}(x) = \max(0.01x, x)$. The main difference between the two activation functions is the flat output in the negative region, therefore the ReLU does not distinguish between different negative inputs. Shallow activation functions have been very popular in the past years. The most used was the sigmoid $\sigma(x) = 1/(1+e^{x})$ or equivalently the arctangent function. The ReLU, sigmoid and arctangent activation functions are shown in Figure C.1.



(a)



(b)

FIGURE C.1: Examples of activation functions. The ReLU and is shown in (a) and the sigmoid $\sigma(x)$ in (b). The arctangent function is shown in both plots.

In neural networks the artificial neurons are organized in multiple layers and each neuron takes as input the output of all the neurons in the previous layer. Such an organized structure lets the information flow in one single direction, resulting in a function $f : \mathbb{R}^n \to \mathbb{R}^m$. Typically the input vector has large dimensionality, while the output vector has one or few components. Figure C.2 shows an example of neural network.

The neuron layers that compose the DNNs are categorized as input, hidden, and output layers. The input layer is the vector of variables that are fed to the network. The final layer is called output layer and it contains values that the network wants to reproduce. The layers in between are called hidden layers because their output values are not explicitly seen by the user. The overall length of the chain of layers is called depth of the model. The dimensionalities of the input and output layers are fixed by the problem to deal with. Instead, the number of artificial neurons of the hidden layers, called width of the layer, can be freely chosen.

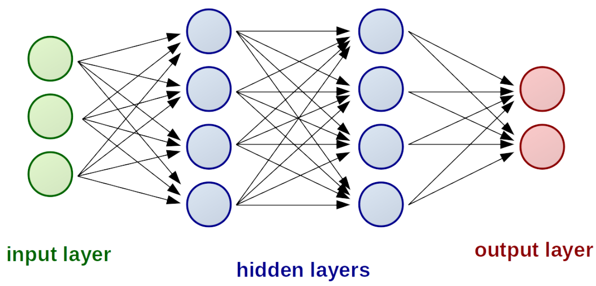


FIGURE C.2: Sketch of a 3 layer neural network with three inputs. The two hidden layers have of 4 neurons each.

The described computing system has a very particular architecture. Let $f^{(i)}$ be the vector function of the i-th layer, then the entire DNN is the composition $f = f^{(depth)} \circ ... \circ f^{(2)} \circ f^{(1)}$. Therefore DNNs belong to a very peculiar class of functions. It has been demonstrated that, given appropriate depth and layer widths, the functions of this class can approximate any given function [174, 175].

C.2.1 Loss function minimization

The deep neural networks are functions that have a several parameters. The parameters of the networks are the weights and biases of the affine transformations of the artificial neurons. These parameters can be optimized in order to solve specific tasks. The optimization of a neural network is performed by providing a set of examples to the network. The set of examples is called training dataset and the optimization process is called training. The training of the network is performed by tuning the parameters of the model via the minimization of a function called loss function.

The loss function is an estimate of the distance between the target values and the network predictions. Therefore minimizing the loss function means tuning the network to have an output as close as possible to the desired target. In classification problems, the adopted loss function is the cross-entropy between the training data and the model's prediction. In regression problems, the adopted loss function is the mean squared error defined to be the sum of the squared distance between the outputs and the target values. In either cases, the loss function is defined as the expectation value of the negative log-likelihood, therefore its minimization is equivalent to maximizing the the likelihood of the output distribution.

The minimization of the loss function is performed using a gradient descending procedure called stochastic gradient descent (SGD). This procedure is iterative and it consists of moving the model parameters in the direction of the negative gradient of the loss function. The gradient of the loss function is computed using small sets of data called minibatches, or simply batches. Therefore the descending has a stochastic component due to only a fraction of the dataset is used to compute the gradient. The loss function is, in general, a non-convex function and there is no guarantee that the global minimum exists.

The gradient of the loss function is computed using the back-propagation algorithm [176]. This algorithm allows computing numerically the derivative of the loss function with respect to the parameters of each neuron. Normally, the numerical evaluation of the gradient is computationally expensive, however, the algorithm reduce substancially the computation cost using the chain rule of calculus:

$$\frac{dz}{dx} = \frac{dz}{dy}\frac{dy}{dx} \tag{C.2}$$

The equation holds also for vector-value functions and the transformation is described by the Jacobian matrix. The chain rule allows computing the gradient of the loss function using Jacobian transformations between different layers. For example, the derivative of the network output with respect to a parameter w of the first layer is computed by

$$\frac{df}{dw} = \left(\frac{\delta f}{\delta f^{(depth)}}\right) \cdots \left(\frac{\delta f^{(2)}}{\delta f^{(1)}}\right) \frac{df^{(1)}}{dw} \tag{C.3}$$

where $\left(\frac{\delta f^{(i)}}{\delta f^{(j)}}\right)$ are the Jacobian transformations between adjacent layers.

The analytical form of the gradient may contain several repeated sub-expressions. The sub-expressions that are repeated are storaged in order to reduces the computational cost. However, storing the sub-expressions increases exponentially the memory consumption, therefore several strategies can be adopted in order to reduce also the memory usage, for example by trying to simplify the network graph or repeating some calculations of some simple sub-expressions.

The back-propagation is applied to compute the gradient in the SGD algorithm. The model parameters are modified in the direction of the gradient, then the new value of the loss function is computed using the events in a single minibatch. The gradient calculation and the parameters update are repeated iteratively. This procedure relies on the assumption that the gradient can be correctly computed using a small fraction of the available events. In order to find the optimized model parameter, the SGD algorithm passes through the training dataset several times. An entire passage through the training dataset is called epoch and the number of times that the algorithm passes completely through the training dataset is called number of epochs.

In the SGD algorithm, the DNN parameters $\vec{\theta}$ are modified by the negative gradient multiplicated by the learning rate ϵ

$$\vec{\theta} \rightarrow \vec{\theta} - \epsilon\vec{\nabla}(\text{loss}) \tag{C.4}$$

The learning rate is a crucial parameter that determines the step size while moving toward a minimum. Choosing the appropriate learning rate is very important to reach low values of the loss function. Figure C.3 shows the behavior of the loss function for different learning rates. Learning rates too high can cause the loss function to diverge. Instead, when learning rates too low are adopted, the convergence became very slow. In most cases, the learning rate is automatically changed at training time. An initial learning rate is set and it is reduced as a function of the loss in order to obtain a better convergence.

Several improvements of the stochastic gradient descent algorithm have been proposed to speed up and improve the convergence. An example is the momentum based methods. In the most basic definition of the momenum method the parameters modification is defined by a velocity $\vec{\theta} \rightarrow \vec{\theta} + \vec{v}$. The velocity is defined as the weighted sum of the gradient and of the velocity at the previous iteration $\vec{v} \rightarrow \alpha\vec{v} - \epsilon\vec{\nabla}(\text{loss})$. The value of the parameter α is tunable. This method is especially useful for small and noisy gradients.

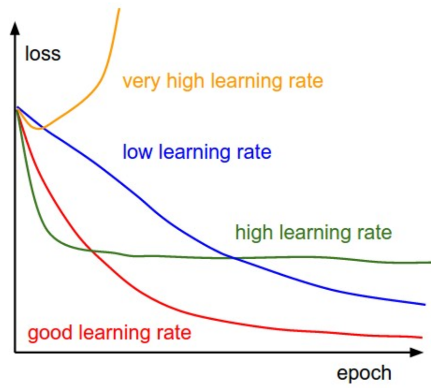


FIGURE C.3: Typical behavior of the loss as a function of the training step for different learning rates [177].

C.2.2 Regularization

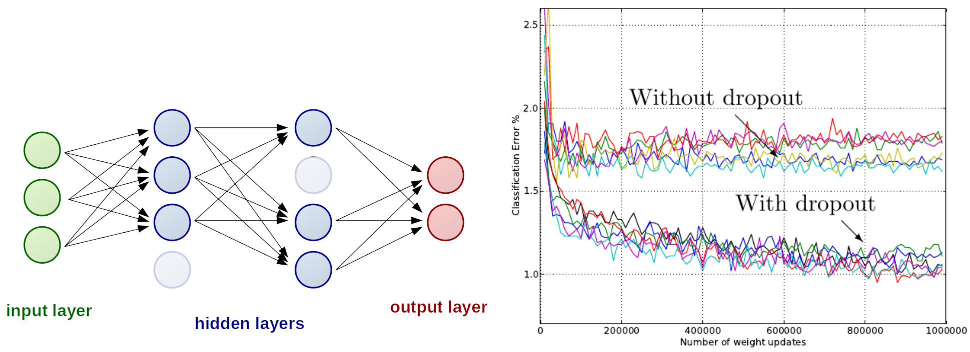
A central problem in machine learning is how to make algorithms that perform well not just on the training data, but also on new inputs. Regularization is any modification performed on the machine learning algorithm that aims to improve the performance on previously unseen data without affecting the performance on the training data. This class of modifications helps also in reducing an eventual overfit of the training data. Several techniques are commonly adopted. Adding a penalization for high weights in the loss function helps in keeping the weights small and improves the gradient descending convergence. Another method is sharing parameters between different neurons. This strategy is adopted under the assumption that some inputs have symmetry properties or should be treated similarly.

Ensemble methods are also frequently used. These techniques consist of training several different models using subsets of the training dataset. The final score of the algorithm is computed by averaging the output of the trained models. The reason why this strategy works well is that different models usually do not make all the same errors on the test set, therefore the averaged score performs substantially better than its members.

Another important method of regularization, called dropout [178], consists of masking fractions of the hidden nodes of the DNN. When the SGD algorithm is executed with the dropout, a random fraction of the nodes of the network are selected at each iteration and their outputs are set to zero. This is equivalent to apply the gradient descending method to a DNN without the selected nodes.

The dropout method is very popular because it is computationally cheap and it improves significantly the performance of the trained DNN. Training a network with dropout is equivalent to train a very large ensemble of networks simultaneously. Therefore dropout is similar in this respect to ensamble methods. In fact, each masking produces a different DNN model therefore an exponentially large ensemble of sub-networks are generated. Training simultaneously all the sub-networks is possible thanks to the sharing of their parameters. This method forces the DNN to perform well regardless of which hidden units are considered. However, in the final evaluation of the test dataset, all the nodes are used.

Figure C.4a shows a sketch of a neural network. The drawing highlights the nodes that are removed by the dropout method. Figure C.4b show the loss function evolution with and without the dropout evaluated on the test dataset, showing the benefits of regularization.



(a) (b)

FIGURE C.4: Example of a deep feed-forward neural network where a fraction of the nodes are removed by the dropout (a). Loss function evolution for different architectures with and without dropout .evaluated on the test dataset (b) [178].

Appendix D

VBF H$\to$ $\mu\mu$: a data-driven approach

An alternative search to the Higgs boson decay to muons is described in this appendix. The presented analysis adopts a data-driven approach similar to the one used in the ggH, VH, and ttH searches. The expected results are shown and a comparison with the MC-based is discussed.

D.1 Introduction

The H$\to$ $\mu\mu$ analysis consists of four different searches that aims to extract the signal from the four main production channels. Among all the Higgs production channels, the VBF is one of the most important in the H$\to$ $\mu\mu$ search. The reason is that it has a very clear signature that helps in reducing the large amount of background and it has the second higher cross-section among the Higgs production channels. These features make the VBF production to be crucial in rare decays searches such as H$\to$ $\mu\mu$. Therefore the VBF channel has to be exploited as best as possible.

Two different searches for Higgs decay to muons via VBF production are performed at the same time. One is the MC-based analysis described in chapter 5 and the other one is a fully data-driven search. The data-driven analysis consists of a fully data-driven estimation of the background through an analytical fit to the dimuon mass distribution. Its analysis strategy is similar to the one adopted by the ggH, VH, and ttH searches. Firstly, a $m_{\mu\mu}$ independent discriminator is built in order to separate signal-like from background-like events. Then, the events are divided into subcategories as a function of the discriminator score. Finally, the signal is extracted with a simultaneous fit to the dimuon mass distributions among the defined subcategories.

The data-driven approach is much more aligned to the searches of the other channels, however, this method is not a priori chosen for the VBF category. Instead, either the MC-based and the data-driven analyses are performed. The one that will be combined with the other three searches is chosen by comparing the expected results.

D.2 Selection

The analysis selection is identical to the one described in section 4.2. It starts with the trigger and muon requirement explained in section 4.2. The two oppositely charged muons must have an invariant mass within 110 and 150 GeV. Events containing additional leptons, medium b-tagged jets, or more than one loose b-tagged jet are not selected because they are considered in the VH and ttH analyses. The events must have two jets with $|\eta| < 4.7$ and transverse momentum greater than 35 and 25 GeV. Moreover, the two leading jets must have a pseudorapidity distance $\Delta\eta_{jj} > 2.5$ and an invariant mass $m_{jj} > 400$ GeV.

D.3 Multivariate discriminator

A Boosted decision tree (BDT) is build with the TMVA package [107] in the ROOT framework [146]. The training is performed separately for 2016, 2017, and 2018 in order to take into account differences in detector performance among the three data-taking periods. The setting used to build the BDT is shown in table D.1

Boosting type	Trees	Shrinkage	Bagged fraction	node size	Depth	Separation
Gradient boost	1000	0.1	0.5	>3%	4	CrossEntropy

TABLE D.1: Settings used in training the BDTs of the VBF data-driven search.

The BDT is trained using half of the available simulated events with an invariant mass in the range 115-135 GeV. The samples used as signals in the training are simulated Higgs events via ggH, VBF, VH, and ttH with a mass m_H=125 GeV. Background events are composed by DY, $t\bar{t}$, single top, and diboson processes.

In order to avoid biases generated by the categorization, the BDT has not to be correlated with $m_{\mu\mu}$. Therefore all the chosen input variables are highly uncorrelated with the dimuon mass. Moreover, the events are weighted by $1/\sigma_{\mu\mu}$ in the training. In this way, besides separating signal-like events from background-like ones, the BDT will also separate events with small and large dimuon mass uncertainty. Consequently, events with high BDT scores will show a more pronounced signal peak. The dimuon mass uncertainty is not correlated with $m_{\mu\mu}$, thus the weighting is not biasing the BDT output.

The input variables related to the dimuon system are the transverse momentum $p_T^{\mu\mu}$, the pseudorapidity $\eta_{\mu\mu}$, and the Collins-Soper variables ϕ_{CS} and $\cos\theta_{CS}$ described in appendix A. The pseudorapidities η_{μ_1} and η_{μ_2} and the transverse momentum of each muon divided by the dimuon mass $p_T^{\mu_1}/m_{\mu\mu}$ and $p_T^{\mu_2}/m_{\mu\mu}$ are included in the input variables. The transverse momentum and pseudorapidity of the two leading jets are used as input as well as their invariant mass m_{jj}, and their azimuthal and pseudorapidity differences $\Delta\eta_{jj}$ and $\Delta\phi_{jj}$. Additional input variables are the number and the scalar p_T sum of the track jets with $p_T > 2$ GeV, called N_2^{soft} and $H_{(2)T}^{\text{soft}}$. Moreover, the total number of jets is used.

Several variables that consider the kinematic of both muons and jets are included as inputs. The pseudorapidity and azimuthal distance of the dimuon system with the leading jet $\Delta\eta(\mu\mu, j_1)$ and $\Delta\phi(\mu\mu, j_1)$ and the minimum pseudorapidity and azimuthal distance of the dimuon system with the two leading jets $\min_i -\Delta\phi(\mu\mu, j_i)$ and $\min_i -\Delta\eta(\mu\mu, j_i)$ are added. Then, the Zeppenfeld variable $z^{\star}$ (see equation 4.1), the transverse momentum balance $\mathrm{R}(p_T)$ (see equation 5.2) and the $\mathrm{p_T}$-centrality are also used. The $\mathrm{p_T}$-centrality is defined by

$$\mathrm{p_T} - \text{centrality} = \frac{p_T^{\mu\mu} - |\vec{p}_T^{\,j_1} + \vec{p}_T^{\,j_2}|/2}{|\vec{p}_T^{\,j_1} - \vec{p}_T^{\,j_2}|} \tag{D.1}$$

and it provides an comparison between the transverse momentum of the dimuon system and of the dijet system.

D.4 Categorization

All the selected events with dimuon mass within 110 and 150 GeV are divided into subcategories as a function of the BDT output. The subcategories boundaries are defined with an iterative procedure in order to maximize the signal sensitivity. The procedure identical to the one applied in the ggH, VH, and ttH analyses.

An initial fit is performed to extract the signal peak in the dimuon mass distribution. The signal is modelled by a double-side crystal ball function (defined by equation 4.2) while the monotone background is described by a BWZγ function (defined by equation 4.5). The signal significance is computed using the result of the fit.

Then, the BDT output is scanned in signal quantiles and, for each quantile, the events are divided in two sets: 99 pairs of sets are considered in total. For each pair of subsets, the signal significance is extracted from a simultaneous fit across the two sets of events. The quantile that provides the best significance is defined to be the subcategory boundary. The described procedure is repeated from two event subsets to three and so on until further divisions do not provide additional gains in significance larger than 1%. The boundaries obtained with this method are the ones that define the subcategories. Four subcategories are found with this approach and they are defined by the quantile boundaries of 10%, 22%, and 45% signal efficiency.

The procedure is repeated using signal samples with different MC generators. The same number of subcategories are found, and the corresponding boundaries are in agreement within 2%, proving that they are robust and not dependent on the generators. The four subcategories are called "VBF-cat0", "VBF-cat1", "VBF-cat2", and "VBF-cat3" respectively from the one with lower BDT scores to the one with higher scores. Figure D.1 shows the distribution of the BDT score after the transformation $BDT' = \mathrm{atanh}((BDT + 1)/2)$ for data and simulated events. The vertical lines indicate the subcategories boundaries obtained with the described iterative procedure.

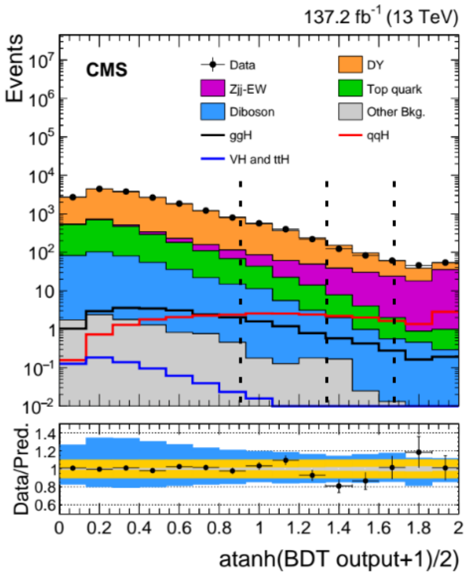


FIGURE D.1: BDT output distribution for signal and background for the events passing the VBF category selection with 110 < $m_{\mu\mu}$ < 150 GeV. The rightmost bin contains the overflow events. The dotted lines indicate the range between the minimum and maximum BDT score considered to define the boundaries. The lower panel shows the ratio between data and the expected background. The grey band indicates MC statistical uncertainties, while the azure and orange bands represent the experimental and theoretical uncertainties.

D.5 Siganl extraction

The signal is extracted by fitting simultaneously the dimuon mass spectrum distributions via analytical functions across the four subcategories. The signal events are mainly produced via VBF and ggH, however, small contributions arise from VH and ttH production. The signal processes are modeled independently for ggH, VBF, VH, and ttH productions in the four subcategories with DCB functions. The DBC parameters are extracted from the simulation and they are fixed in the fit to the data.

In principle, any parametric distribution for modeling the background can inject a potential bias in the signal extraction. The reason is that the background yields under the signal peak are estimated from the S+B fit in the sideband. Therefore, several analytical forms are taken into consideration for modeling the background distributions. Either physics-inspired functions and agnostic ones are taken into account.

All the considered functions for the background description are studied in order to estimate the bias introduced in the signal extraction. The bias introduced by the background description is studied with the following strategy. Different background models are used to perform a background-only fit and the postfit parameters are used to generate pseudodatasets in each subcategory. The function under study is used to fit the pseudodatasets and extract the injected signal strengths. The difference between the extracted and injected signal is an estimate of the bias introduced by the studied function.

The analytical functions chosen to estimate the total background are modified Breit-Wigner functions (defined in equation 4.3) in "VBF-cat0", and "VBF-cat1" and BWZγ functions in "VBF-cat2", and "VBF-cat3". The bias introduced by the background models is less than 1% and therefore it is neglected in the uncertainty treatment.

The expected signal strength modifier with the associated uncertainty is extracted with a simultaneous binning fit among the four categories. The fit is performed using the expected yields obtained from the simulation in the mass window 115-135 GeV and using the data in the sideband. The data and the functions extracted from the fit to the dimuon distributions are shown in Figure D.1. The background-only component is shown with its one and two standard deviation uncertainties obtained from the fit.

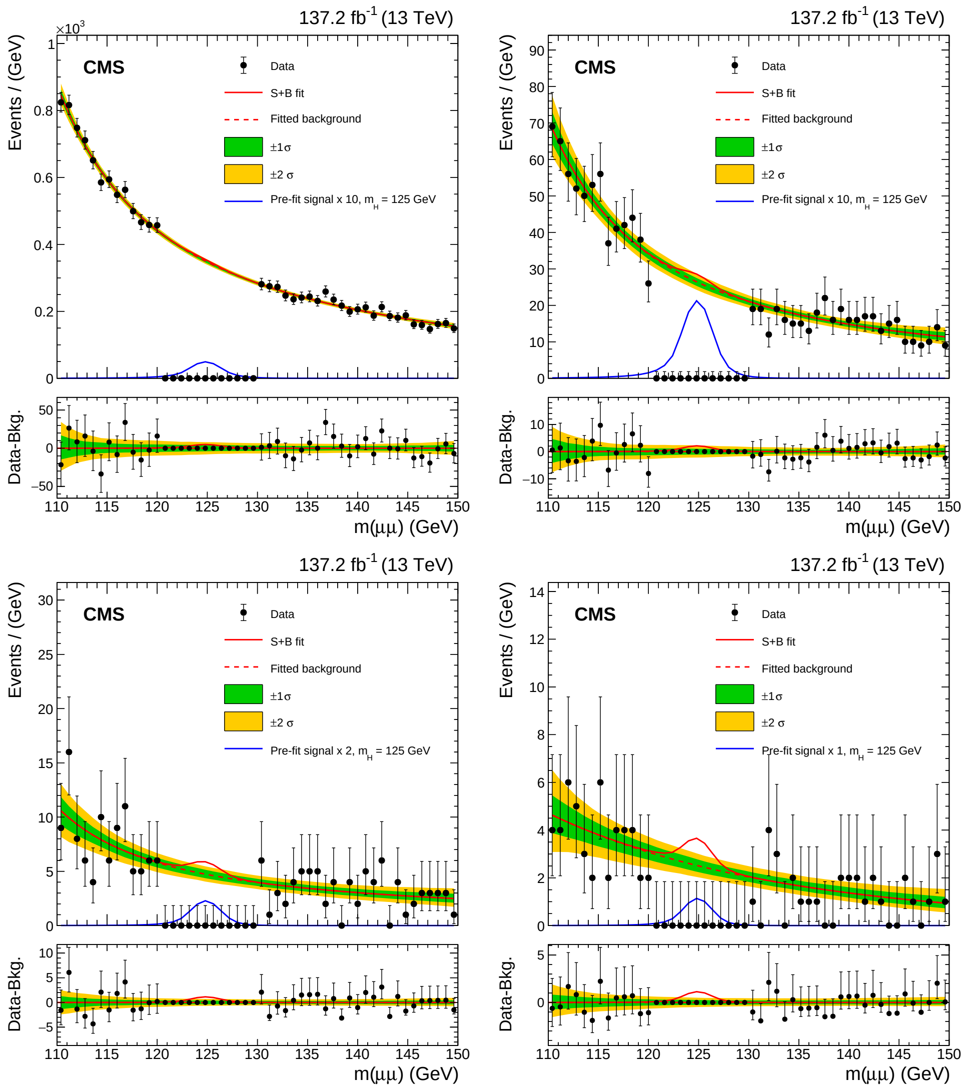


FIGURE D.2: Dimuon mass spectrum in all the VBF subcategories, the background functions are extracted from the fit. The fit is performed using simulated prediction in the mass window 115-135 GeV. The lower panel shows the residual background subtraction and the red line represents the signal.

Expected upper limits on the signal strength modifiers are computed at the 95% CL alongside the signal significance for each category independently and for the simultaneous fit. The expected results of the H→ $\mu\mu$ search via data-driven approach are shown in table D.2.

VBF subategory	expected significance	expected upper limit
VBF-cat0	0.38σ	5.36
VBF-cat1	0.61σ	3.42
VBF-cat2	0.74σ	2.79
VBF-cat3	1.12σ	2.20
Combination	1.52σ	1.42

TABLE D.2: Expected upper limits and signal significances for each BDT subcategory and for their combination.

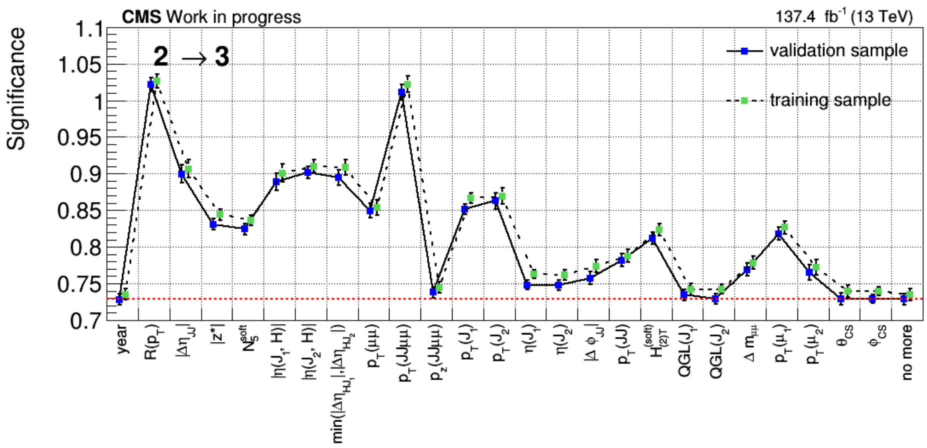
D.6 Conclusion

Two searches of the Higgs decay to muons are performed with two different methods. The search via data-based approach is presented and the expected results are shown in table D.2. The MC-based search is described in chapter 5 and the associated expected results are shown in chapter 6.

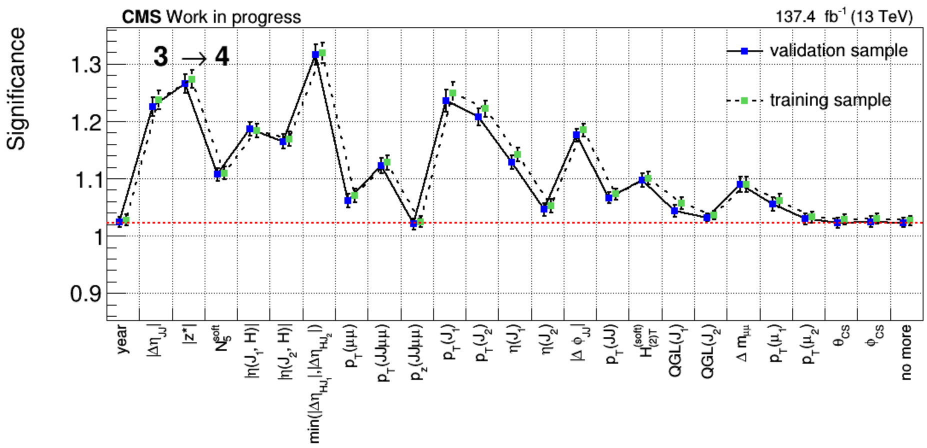
The expected results obtained from the two analyses are compared. The MC-based approach shows an improvement in the expected results of more than 15% with respect to the data-driven approach. Therefore the MC-based analysis is the one that is chosen to be combined with the ggH, VH, and ttH searches. The results obtained from the combination are published in a common paper [105] and they are shown in chapter 6.

Appendix E

N+1 test

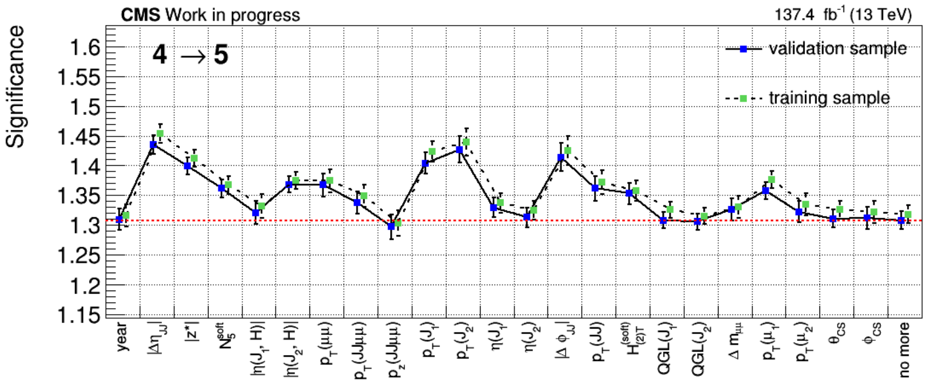


(a)

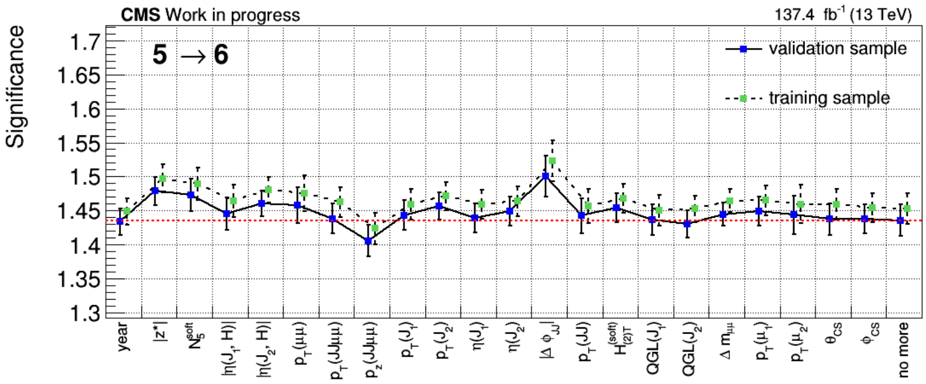


(b)

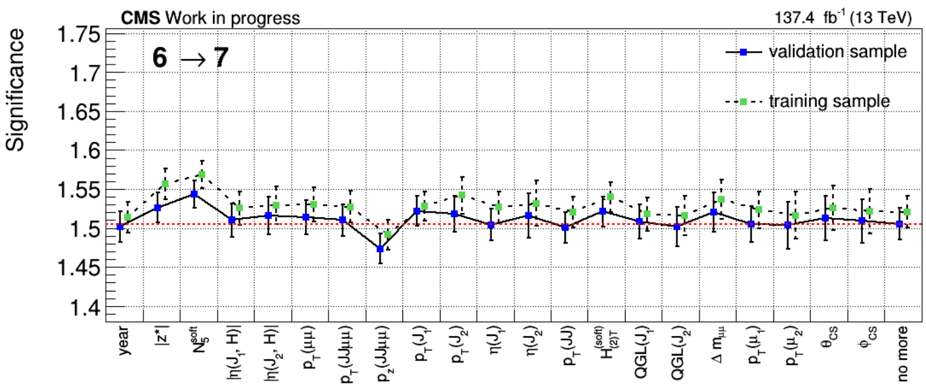
FIGURE E.1: (a) First and (b) second step of the N+1 test. The performance and their uncertainties shown are the mean and the standard deviation of 20 discriminators trained with different dataset and they are computed with the training dataset (green) and with the validation dataset (blu) separately.



(a)

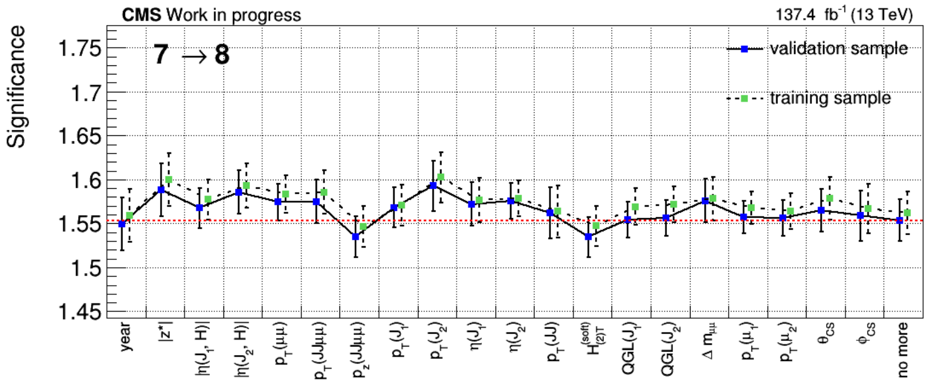


(b)

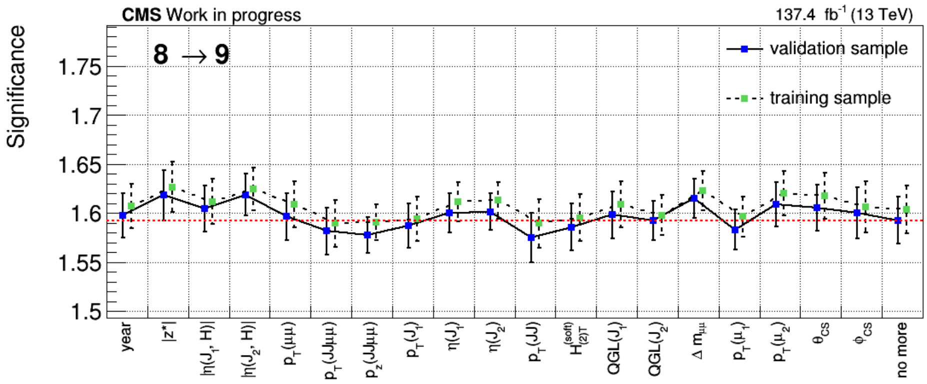


(c)

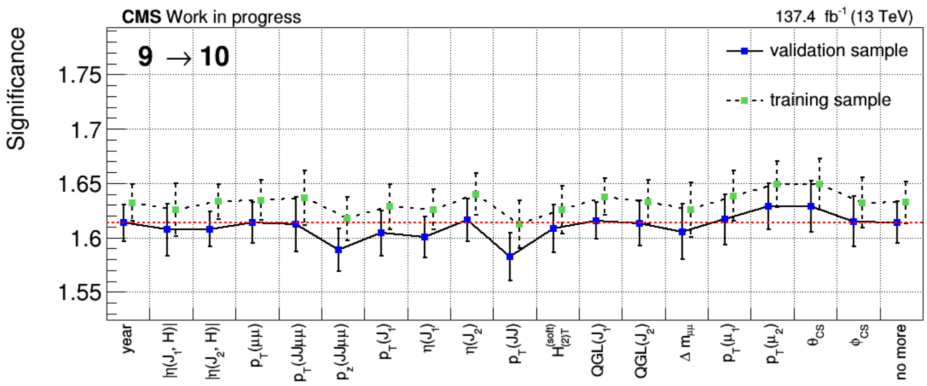
FIGURE E.2: (a) Third, (b) fourth and (c) fifth step of the N+1 test. Caption is the same as Figure E.1.



(a)



(b)



(c)

FIGURE E.3: (a) Sixth, (b) seventh and (c) eighth step of the N+1 test. Caption is the same as Figure E.1.

Appendix F

BDT and DNN distributions

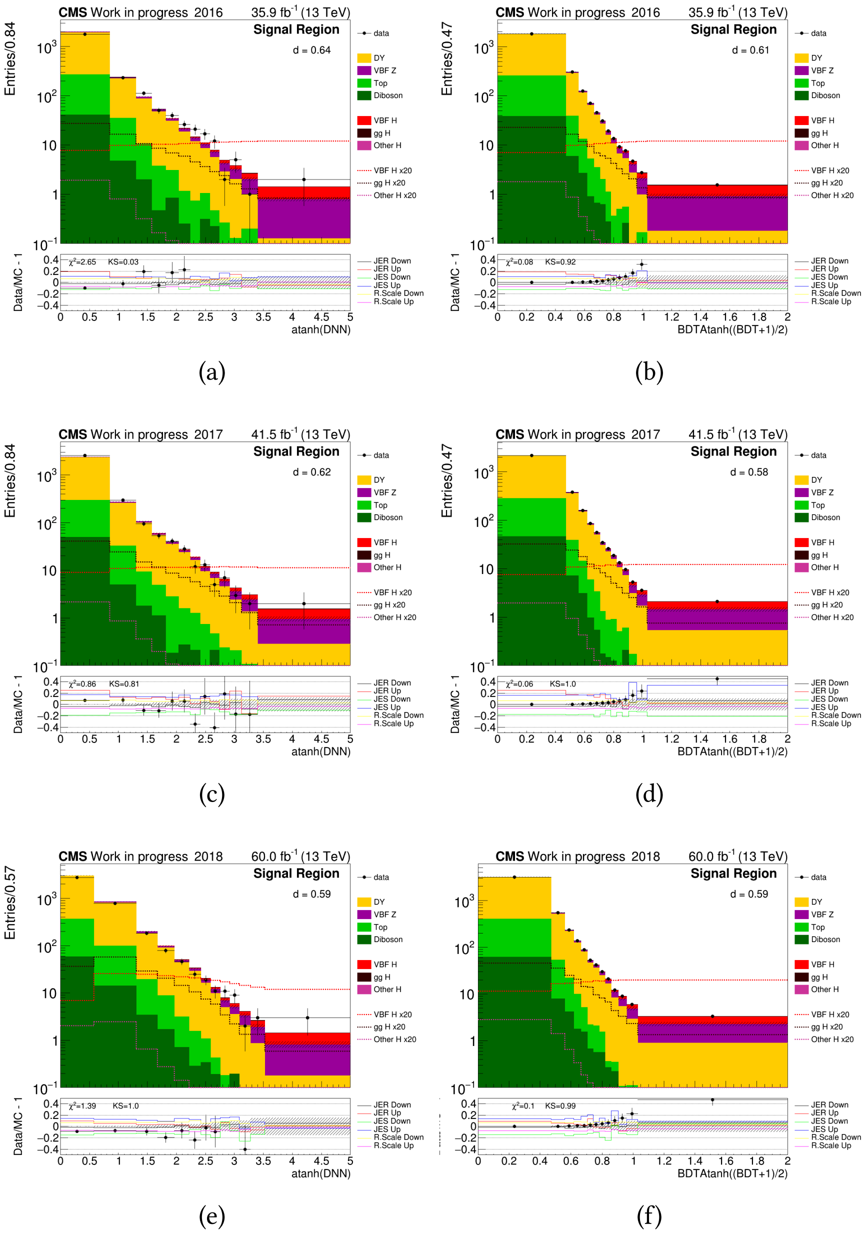
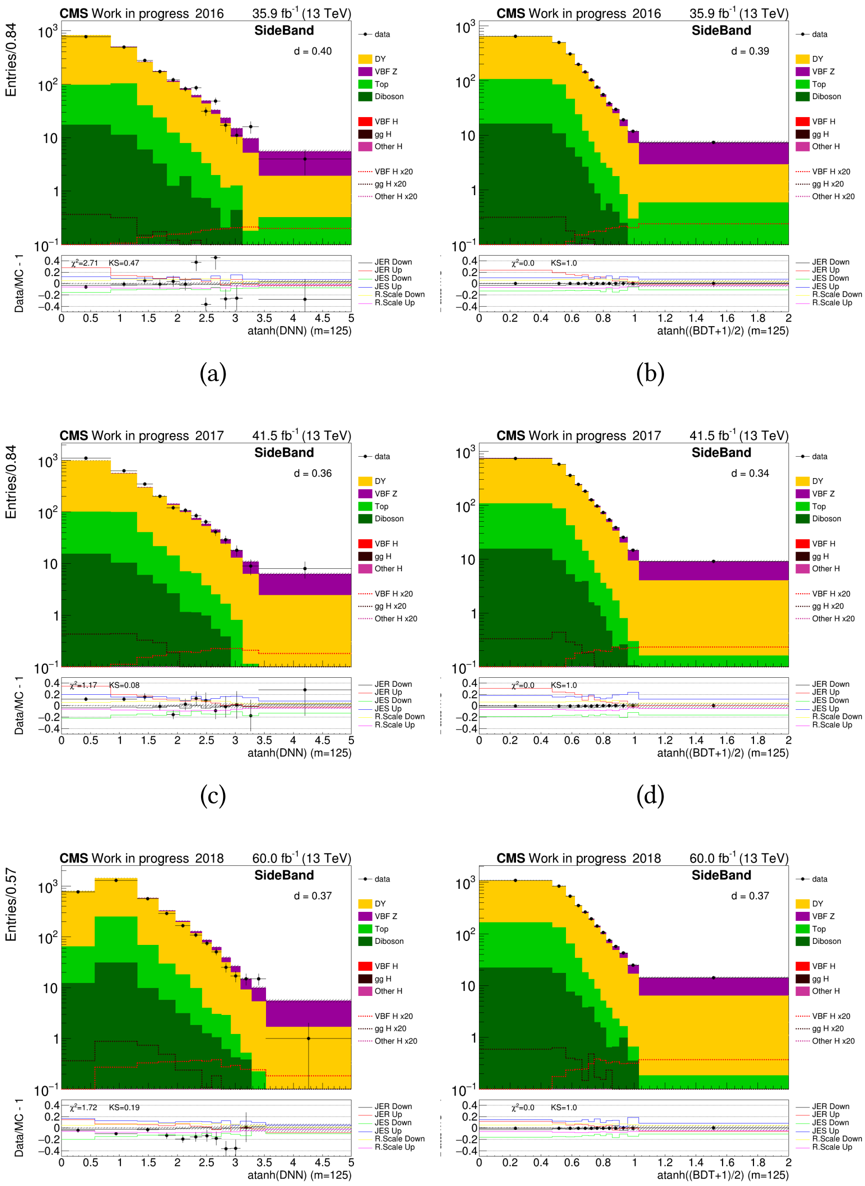


Figure F.1: BDT and DNN distributions after the event selection for 2016 (a) 2017 (b) and 2018 (c) in signal region. The lower panel shows the ratio data over Monte-Carlo. The dashed area represents the MC statistical uncertainties while the lines in the lower panel show the uncertainties related to the energy of the jets and to the factorization scale.



(a) (b)

(c) (d)

(e) (f)

FIGURE F.2: BDT and DNN distributions after the event selection for 2016 (a) 2017 (b) and 2018 (c) in sideband. Captions are the same of Figure F.1.

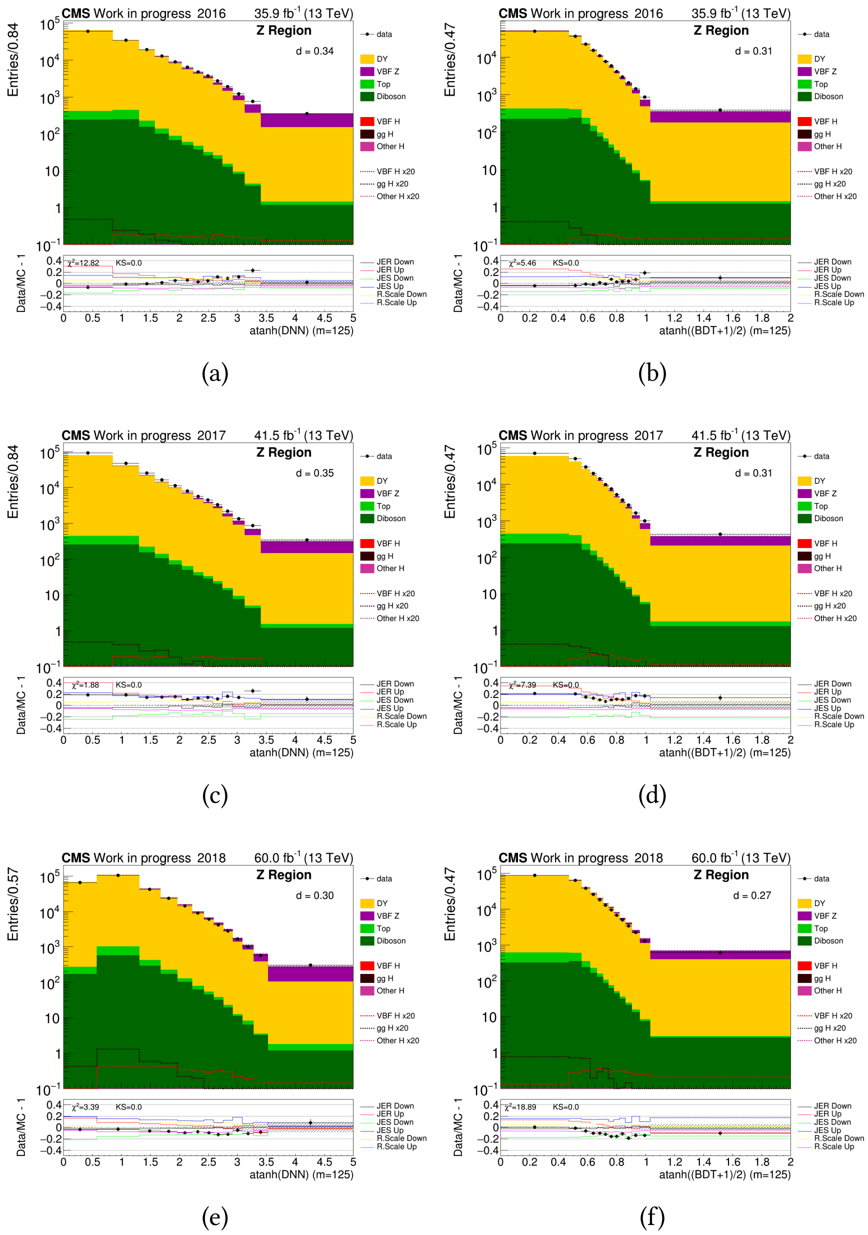
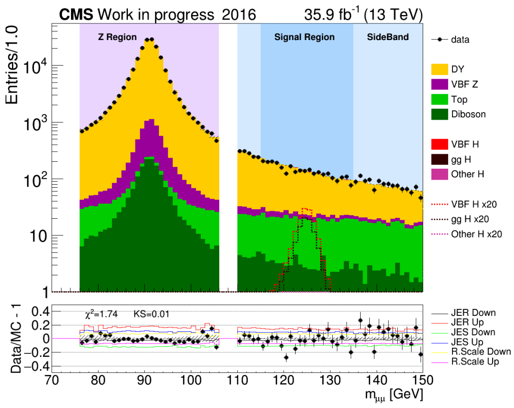


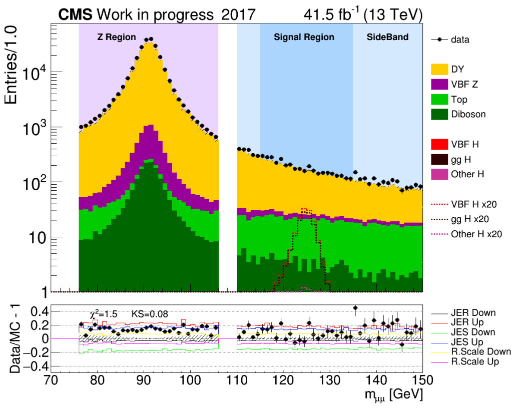
Figure F.3: BDT and DNN distributions after the event selection for 2016 (a) 2017 (b) and 2018 (c) in Z region. Captions are the same of Figure F.1.

Appendix G

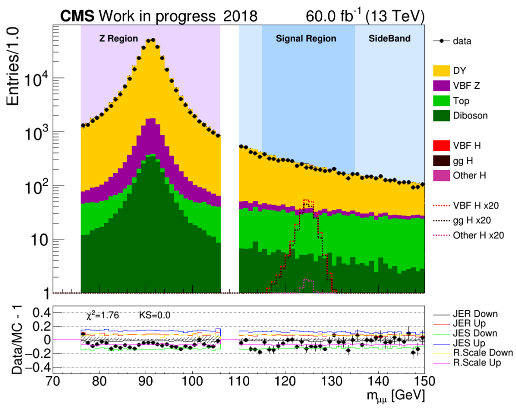
Data vs MC comparison



(a)

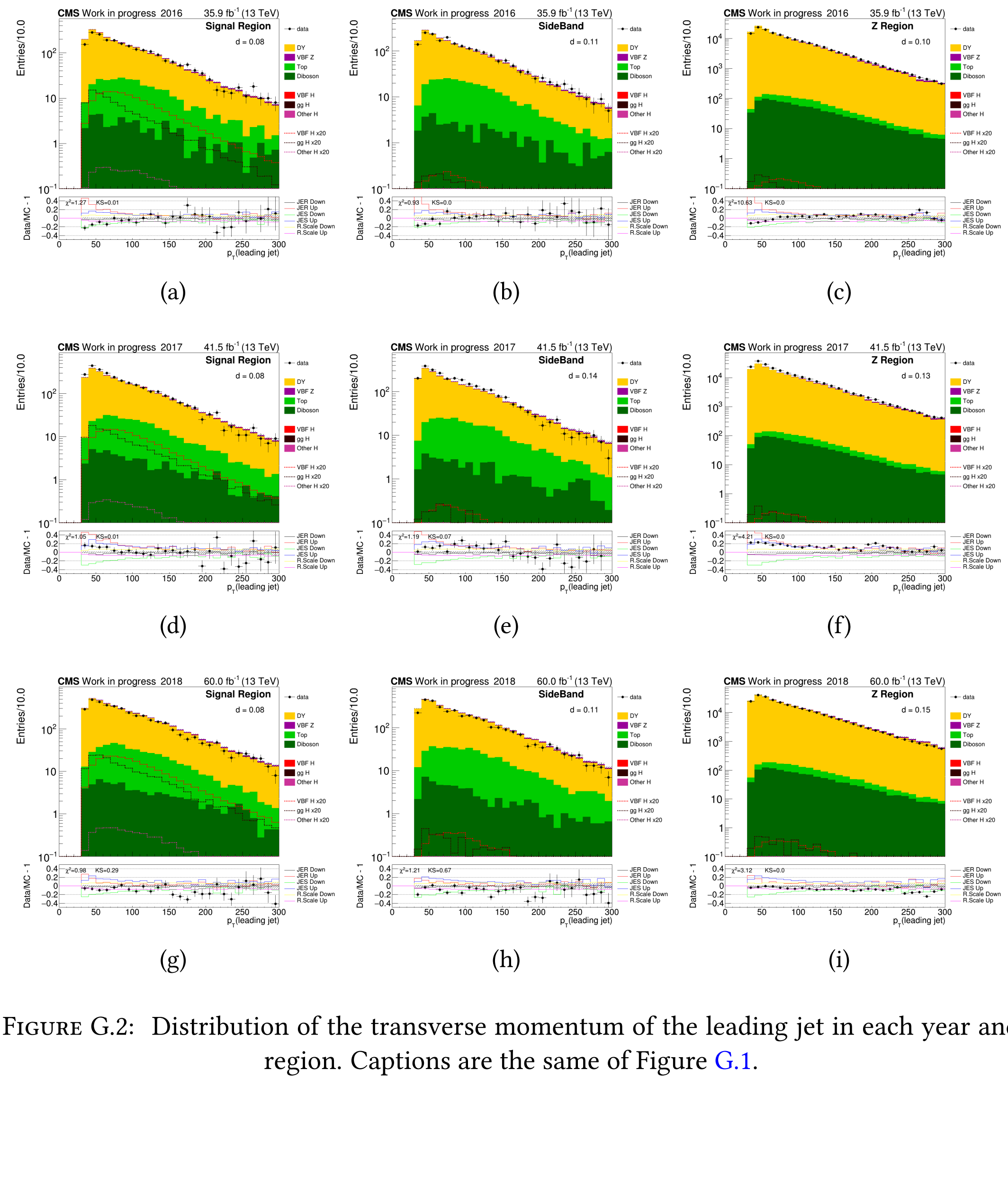


(b)



(c)

FIGURE G.1: Dimuon invariant mass distribution after the event selection for 2016 (a) 2017 (b) and 2018 (c). Events in different regions are shown in the same plot. The lower panel shows the ratio data over Monte-Carlo. The dashed area represents the MC statistical uncertainties while the lines in the lower panel show the uncertainties related to the energy of the jets and to the factorization scale.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

FIGURE G.2: Distribution of the transverse momentum of the leading jet in each year and region. Captions are the same of Figure G.1.

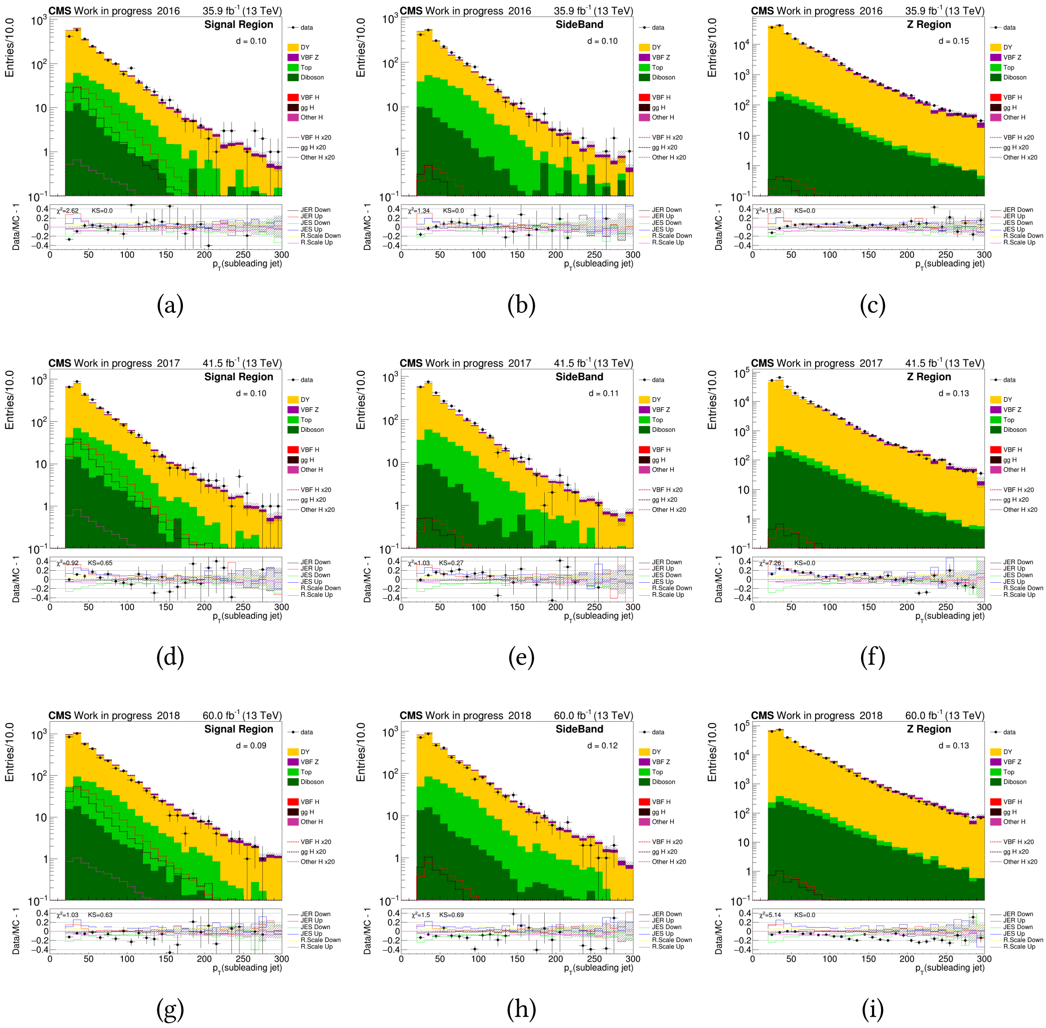
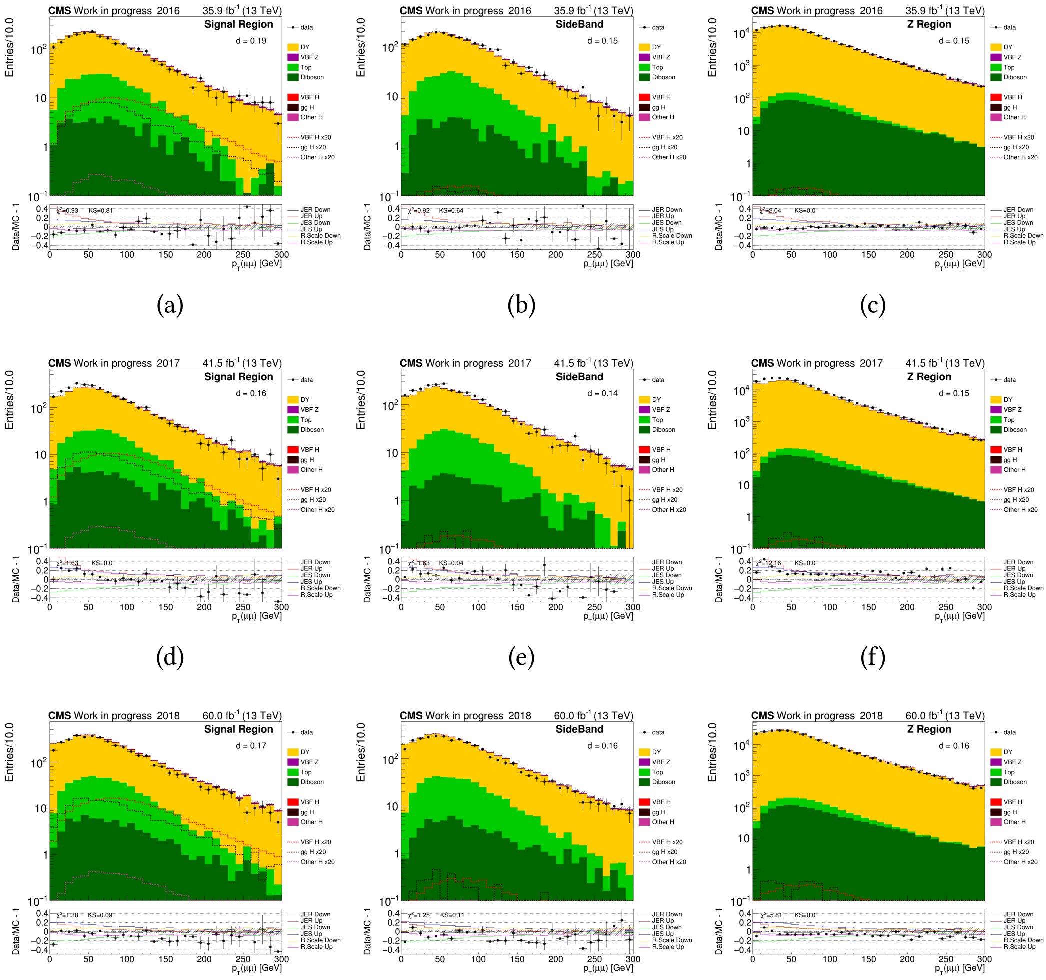


FIGURE G.3: Distribution of the transverse momentum of the subleading jet in each year and region. Captions are the same of Figure G.1.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure G.4: Distribution of the transverse momentum of the dimuon system in each year and region. Captions are the same of Figure G.1.

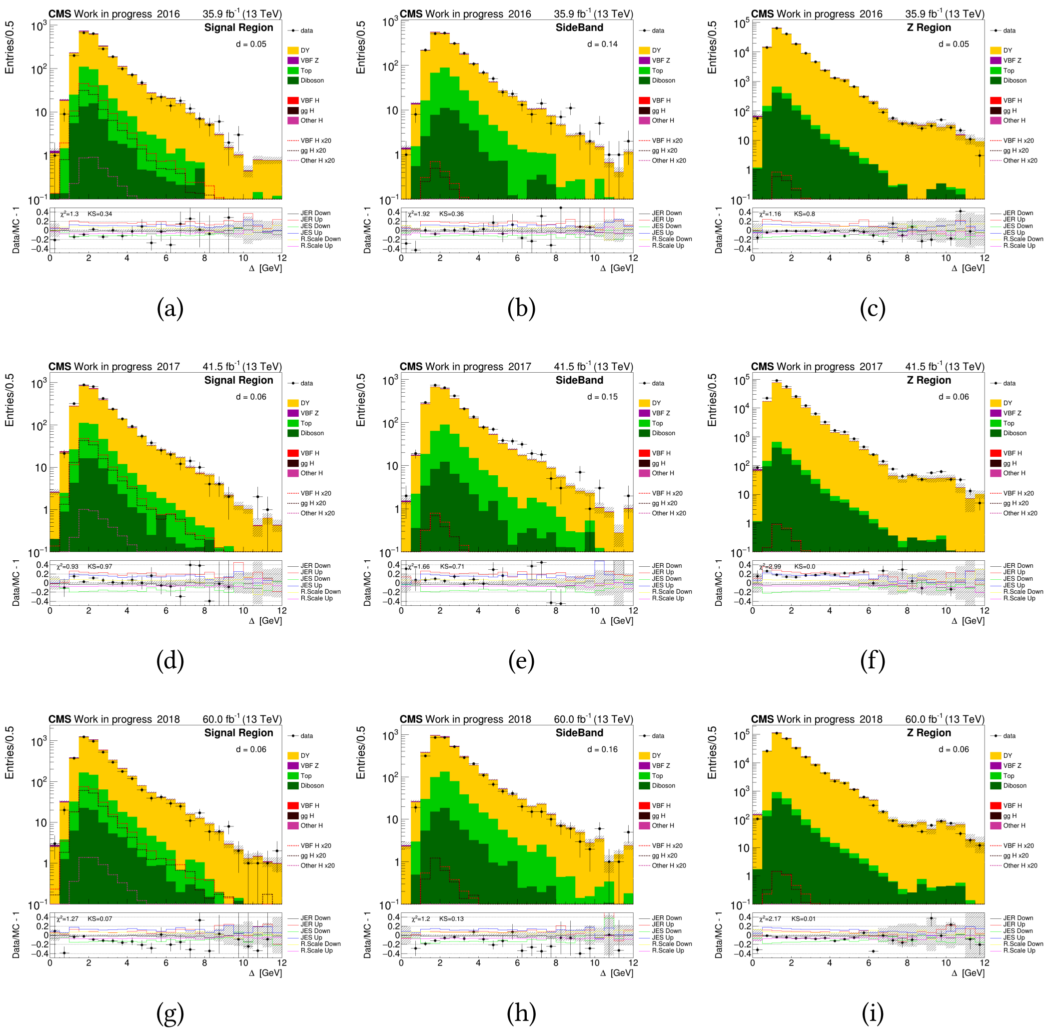
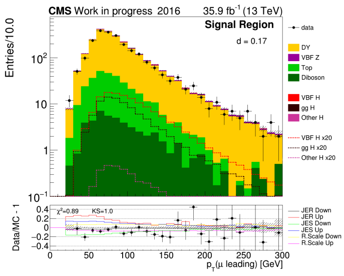
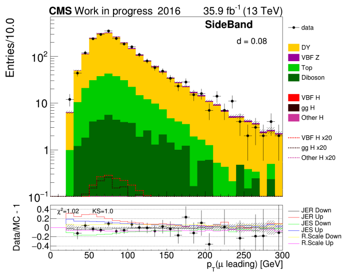


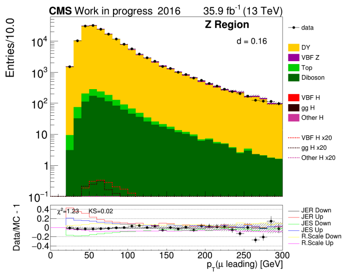
FIGURE G.5: Distribution of the mass resolution of the dimuon invariant mass in each year and region. Captions are the same of Figure G.1.



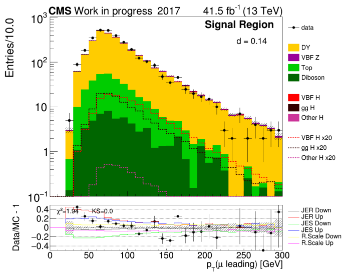
(a)



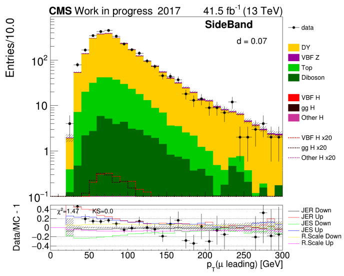
(b)



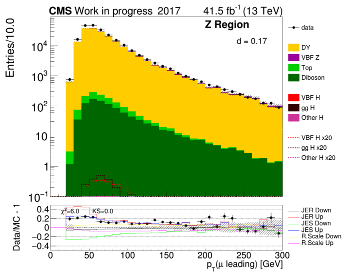
(c)



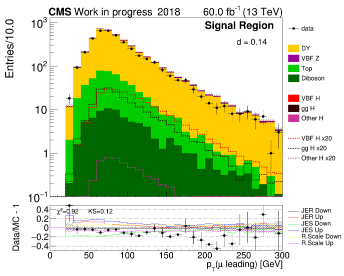
(d)



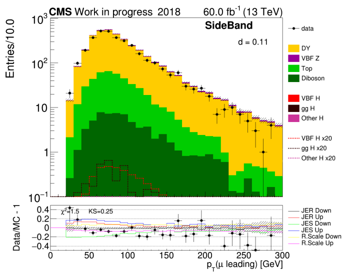
(e)



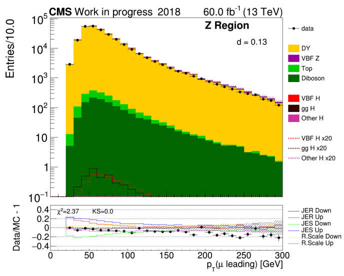
(f)



(g)



(h)



(i)

FIGURE G.6: Distribution of the transverse momentum of the leading muon in each year and region. Captions are the same of Figure G.1.

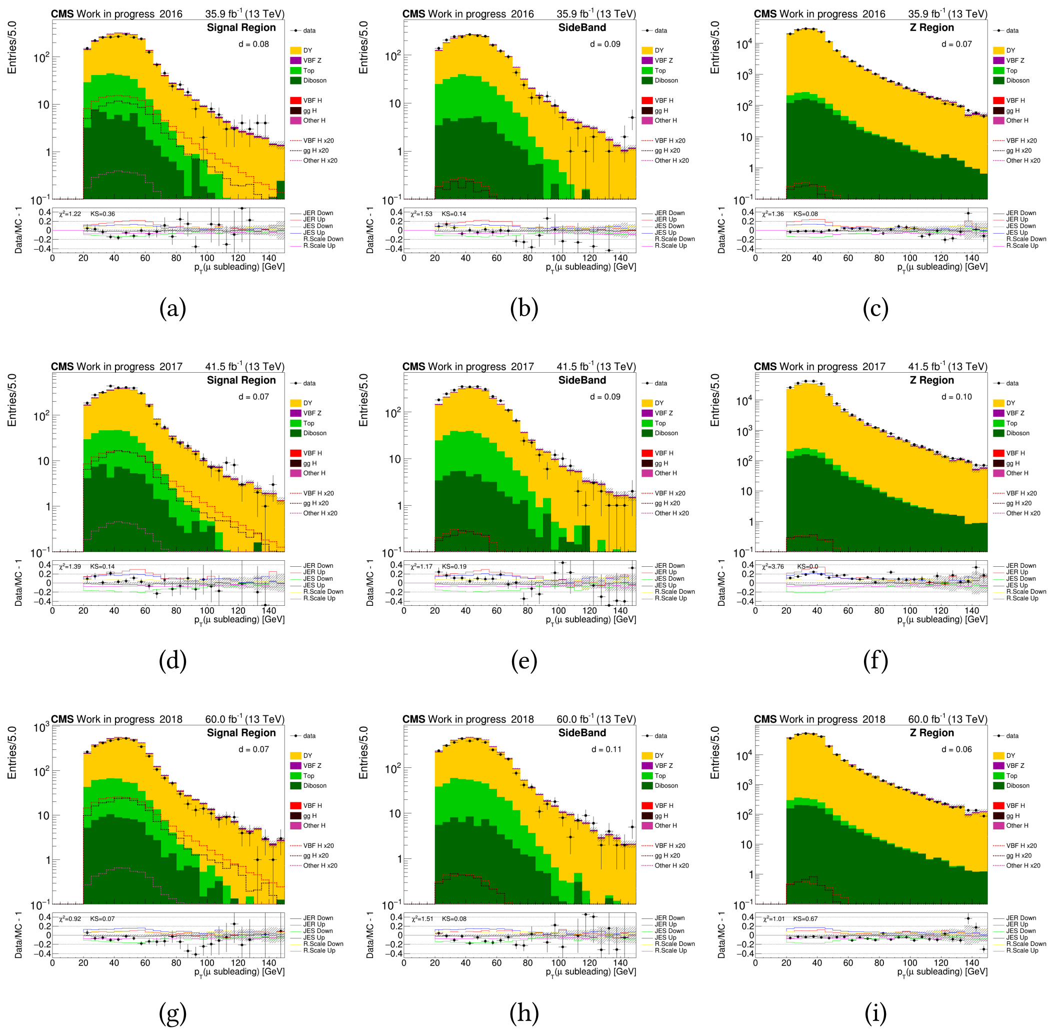


FIGURE G.7: Distribution of the transverse momentum of the subleading muon in each year and region. Captions are the same of Figure G.1.

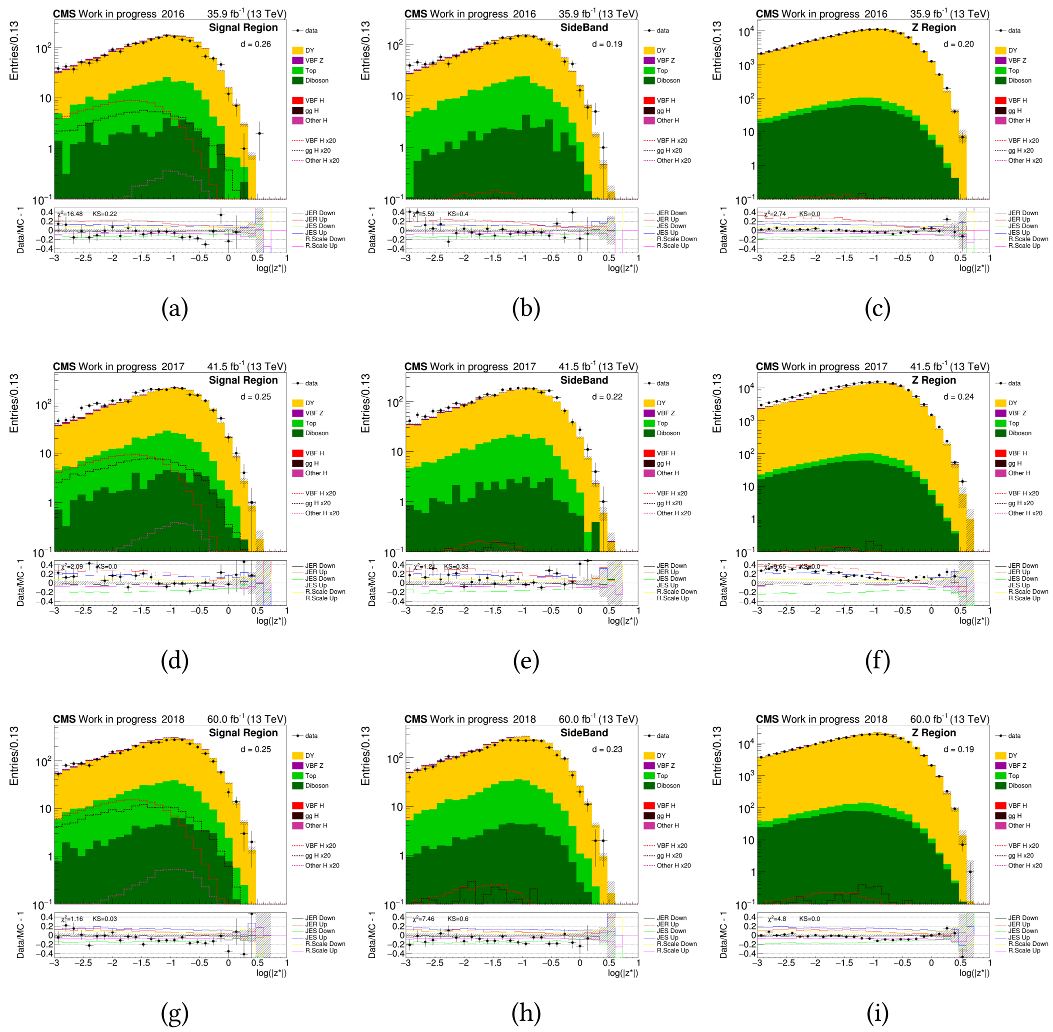
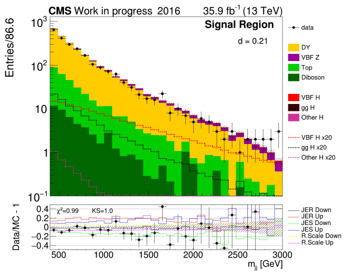
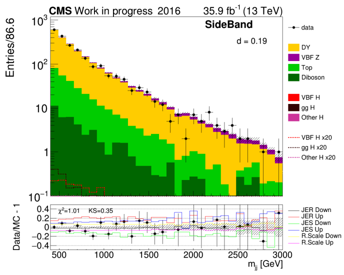


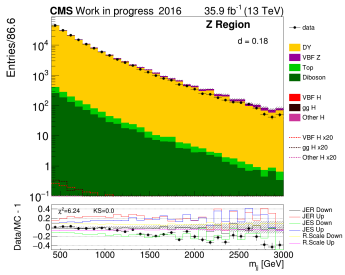
Figure G.8: Distribution of the zeppenfeld variable $z^{\star}$ in each year and region. Captions are the same of Figure G.1.



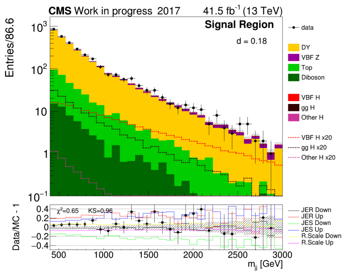
(a)



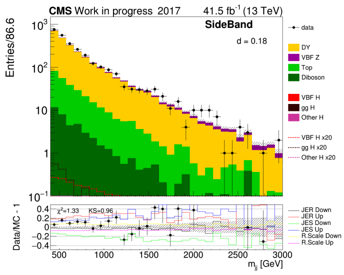
(b)



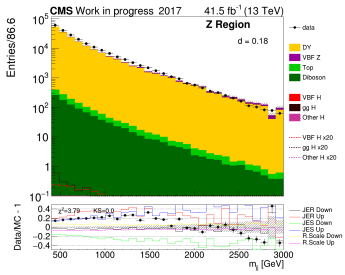
(c)



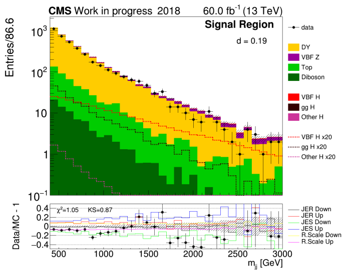
(d)



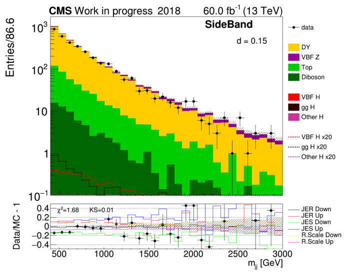
(e)



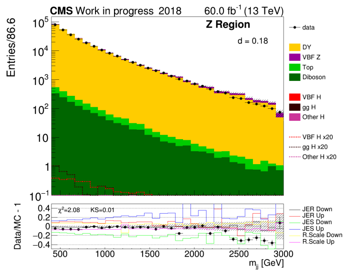
(f)



(g)

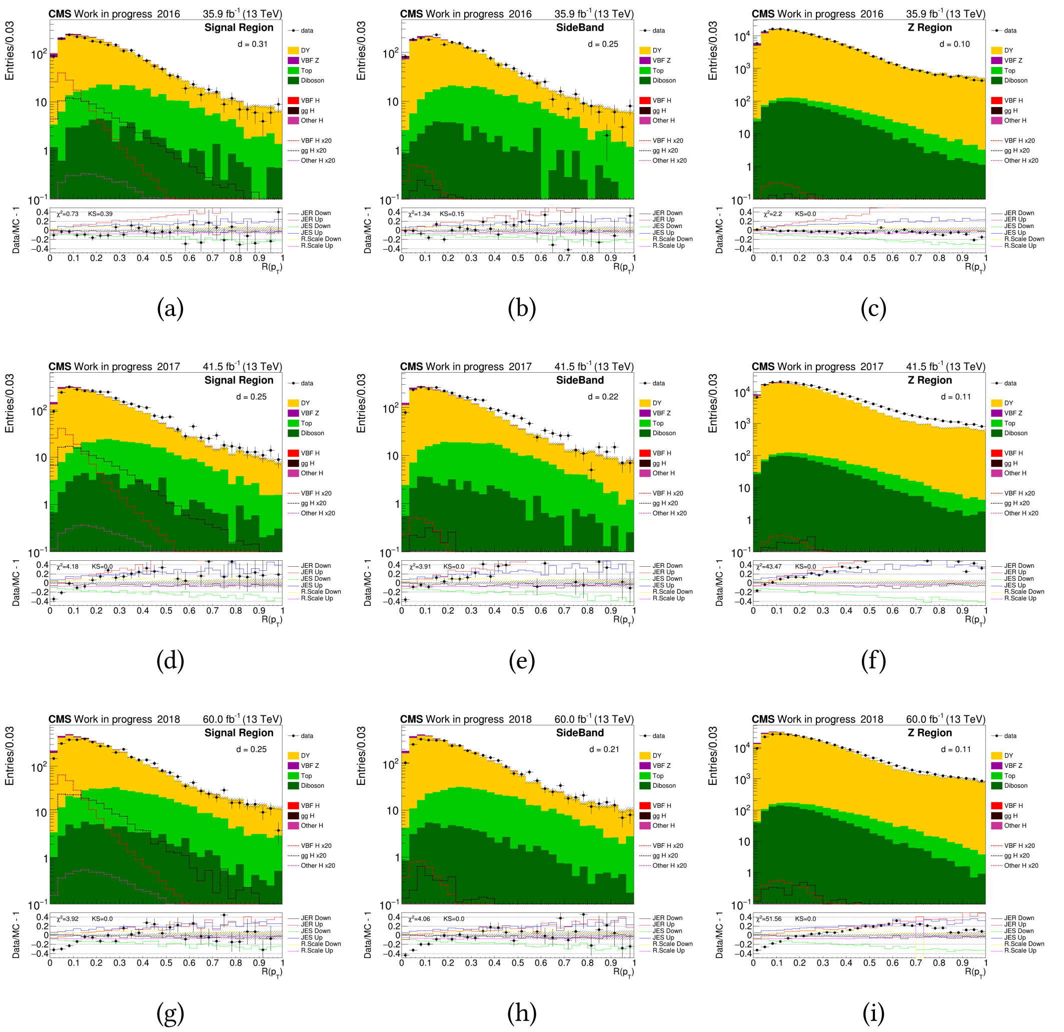


(h)



(i)

Figure G.9: Distribution of the dijet invariant mass in each year and region. Captions are the same of Figure G.1.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

FIGURE G.10: Distribution of the transverse momentum balance R(p_T) in each year and region. Captions are the same of Figure G.1.

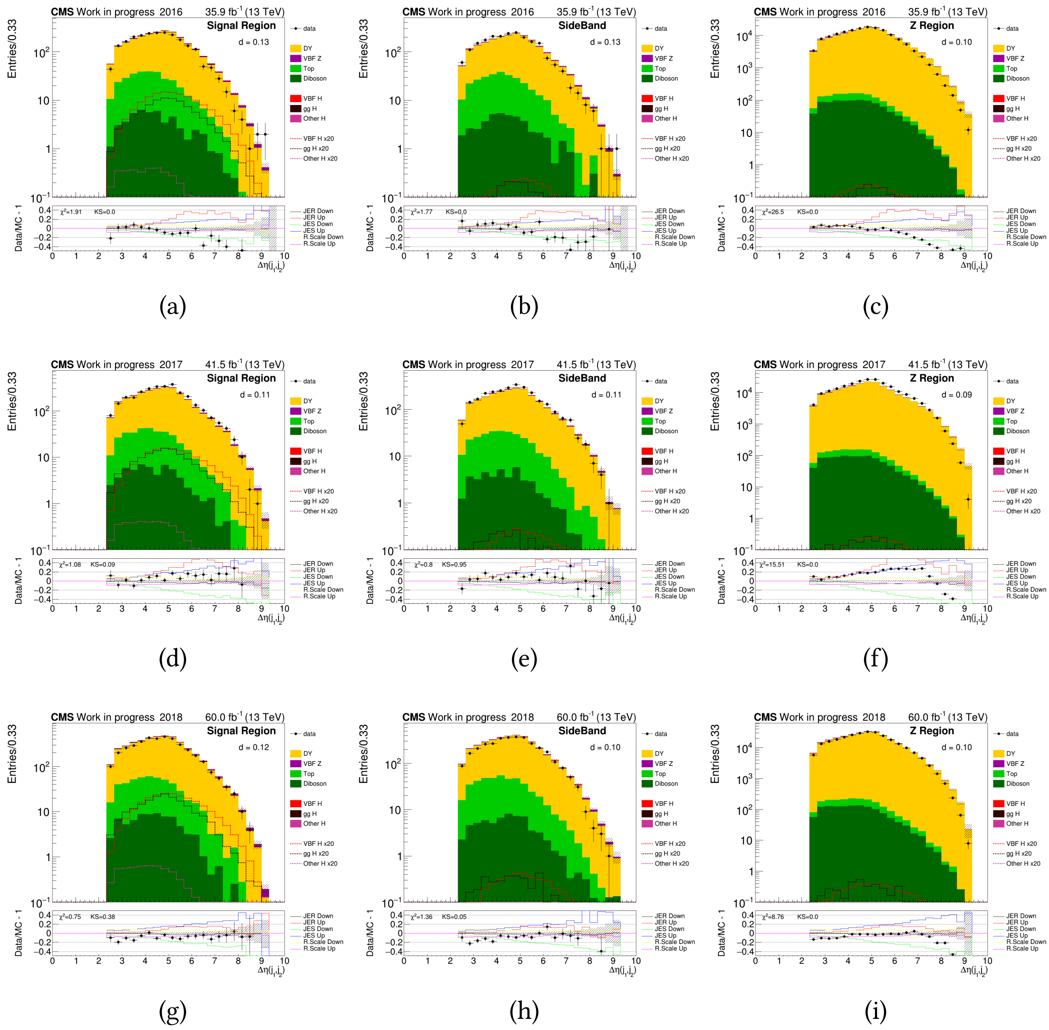


FIGURE G.11: Distribution of the pseudorapidity distance between the two leading jets in each year and region. Captions are the same of Figure G.1.

Bibliography

[1] M. Goldhaber, L. Grodzins, and A. W. Sunyar. "Helicity of Neutrinos". In: *Phys. Rev.* 109 (1958), pp. 1015–1017. DOI: 10.1103/PhysRev.109.1015.

[2] *Standard Model (Wikipedia, the free encyclopedia)*. URL: https://en.wikipedia.org/wiki/Standard_Model#/media/File:Standard_Model_of_Elementary_Particles.svg.

[3] S. L. Glashow. "Partial Symmetries of Weak Interactions". In: *Nucl. Phys.* 22 (1961), pp. 579–588. DOI: 10.1016/0029-5582(61)90469-2.

[4] Steven Weinberg. "A Model of Leptons". In: *Phys. Rev. Lett.* 19 (1967), pp. 1264–1266. DOI: 10.1103/PhysRevLett.19.1264.

[5] Abdus Salam. "Weak and Electromagnetic Interactions". In: *Conf. Proc.* C680519 (1968), pp. 367–377.

[6] Peter W. Higgs. "Broken symmetries, massless particles and gauge fields". In: *Phys. Lett.* 12 (1964), pp. 132–133. DOI: 10.1016/0031-9163(64)91136-9.

[7] F. Englert and R. Brout. "Broken Symmetry and the Mass of Gauge Vector Mesons". In: *Phys. Rev. Lett.* 13 (9 1964), pp. 321–323. DOI: 10.1103/PhysRevLett.13.321. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.321.

[8] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (16 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.508.

[9] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble. "Global Conservation Laws and Massless Particles". In: *Phys. Rev. Lett.* 13 (20 1964), pp. 585–587. DOI: 10.1103/PhysRevLett.13.585. URL: https://link.aps.org/doi/10.1103/PhysRevLett.13.585.

[10] Peter W. Higgs. "Spontaneous Symmetry Breakdown without Massless Bosons". In: *Phys. Rev.* 145 (4 1966), pp. 1156–1163. DOI: 10.1103/PhysRev.145.1156. URL: https://link.aps.org/doi/10.1103/PhysRev.145.1156.

[11] T. W. B. Kibble. "Symmetry Breaking in Non-Abelian Gauge Theories". In: *Phys. Rev.* 155 (5 1967), pp. 1554–1561. DOI: 10.1103/PhysRev.155.1554. URL: https://link.aps.org/doi/10.1103/PhysRev.155.1554.

[12] Y. Nambu and G. Jona-Lasinio. "Dynamical Model of Elementary Particles Based on an Analogy with Superconductivity. I". In: *Phys. Rev.* 122 (1 1961), pp. 345–358. DOI: 10.1103/PhysRev.122.345. URL: https://link.aps.org/doi/10.1103/PhysRev.122.345.

[13] M. Gell-Mann and M. Lévy. "The axial vector current in beta decay". In: *Il Nuovo Cimento (1955-1965)* 16.4 (1960), pp. 705–726. ISSN: 1827-6121. DOI: 10.1007/BF02859738. URL: https://doi.org/10.1007/BF02859738.

[14] The ATLAS Collaboration. "The ATLAS Experiment at the CERN Large Hadron Collider". In: *Journal of Instrumentation* 3.08 (2008), S08003–S08003. DOI: 10.1088/1748-0221/3/08/s08003. URL: https://doi.org/10.1088/1748-0221/3/08/s08003.

[15] The CMS Collaboration. "The CMS experiment at the CERN LHC". In: *Journal of Instrumentation* 3.08 (2008), S08004–S08004. DOI: 10.1088/1748-0221/3/08/s08004. URL: https://doi.org/10.1088/1748-0221/3/08/s08004.

[16] The CMS Collaboration. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 30–61. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2012.08.021. URL: https://www.sciencedirect.com/science/article/pii/S0370269312008581.

[17] The CMS Collaboration. "Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV". In: *Journal of High Energy Physics* 2013.6 (2013), p. 81. ISSN: 1029-8479. DOI: 10.1007/JHEP06(2013)081. URL: https://doi.org/10.1007/JHEP06(2013)081.

[18] The ATLAS Collaboration. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Physics Letters B* 716.1 (2012), pp. 1–29. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2012.08.020. URL: https://www.sciencedirect.com/science/article/pii/S037026931200857X.

[19] *CMS Luminosity public plots*. URL: https://twiki.cern.ch/twiki/bin/view/LHCPhysics/LHCHWG?redirectedfrom=LHCPhysics.LHCHXSWG.

[20] LHC Higgs Cross Section Working Group. *Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector*. CERN Yellow Reports: Monographs. 2016. DOI: 10.23731/CYRM-2017-002. URL: https://cds.cern.ch/record/2227475.

[21] The CMS collaboration. "Observation of the diphoton decay of the Higgs boson and measurement of its properties". In: *Eur. Phys. J.* C74.10 (2014), p. 3076. DOI: 10.1140/epjc/s10052-014-3076-z. arXiv: 1407.0558 [hep-ex].

[22] The ATLAS collaboration. "Measurement of Higgs boson production in the diphoton decay channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector". In: *Phys. Rev.* D90.11 (2014), p. 112015. DOI: 10.1103/PhysRevD.90.112015. arXiv: 1408.7084 [hep-ex].

[23] The ATLAS collaboration. "Measurements of Higgs boson production and couplings in the four-lepton channel in pp collisions at center-of-mass energies of 7 and 8 TeV with the ATLAS detector". In: *Phys. Rev.* D91.1 (2015), p. 012006. DOI: 10.1103/PhysRevD.91.012006. arXiv: 1408.5191 [hep-ex].

[24] The CMS collaboration. "Measurement of the properties of a Higgs boson in the four-lepton final state". In: *Phys. Rev.* D89.9 (2014), p. 092007. DOI: 10.1103/PhysRevD.89.092007. arXiv: 1312.5353 [hep-ex].

[25] The ATLAS collaboration. "Observation and measurement of Higgs boson decays to WW^* with the ATLAS detector". In: *Phys. Rev.* D92.1 (2015), p. 012006. DOI: 10.1103/PhysRevD.92.012006. arXiv: 1412.2641 [hep-ex].

[26] The CMS collaboration. "Measurement of Higgs boson production and properties in the WW decay channel with leptonic final states". In: *JHEP* 01 (2014), p. 096. DOI: 10.1007/JHEP01(2014)096. arXiv: 1312.1129 [hep-ex].

[27] The ATLAS collaboration and the CMS collaboration. "Combined Measurement of the Higgs Boson Mass in *pp* Collisions at $\sqrt{s} = 7$ and 8 TeV with the ATLAS and CMS Experiments". In: *Phys. Rev. Lett.* 114 (2015), p. 191803. DOI: 10.1103/PhysRevLett.114.191803. arXiv: 1503.07589 [hep-ex].

[28] The CMS collaboration. "Study of the Mass and Spin-Parity of the Higgs Boson Candidate Via Its Decays to Z Boson Pairs". In: *Phys. Rev. Lett.* 110.8 (2013), p. 081803. DOI: 10.1103/PhysRevLett.110.081803. arXiv: 1212.6639 [hep-ex].

[29] The ATLAS collaboration. "Evidence for the spin-0 nature of the Higgs boson using ATLAS data". In: *Phys. Lett. B* 726 (2013), pp. 120–144. DOI: 10.1016/j.physletb.2013.08.026. arXiv: 1307.1432 [hep-ex].

[30] The CMS collaboration. "Constraints on the spin-parity and anomalous HVV couplings of the Higgs boson in proton collisions at 7 and 8 TeV". In: *Phys. Rev. D* 92.1 (2015), p. 012004. DOI: 10.1103/PhysRevD.92.012004. arXiv: 1411.3441 [hep-ex].

[31] The ATLAS collaboration. "Measurements of the Higgs boson production and decay rates and coupling strengths using pp collision data at $\sqrt{s} = 7$ and 8 TeV in the ATLAS experiment". In: *Eur. Phys. J. C* 76.1 (2016), p. 6. DOI: 10.1140/epjc/s10052-015-3769-y. arXiv: 1507.04548 [hep-ex].

[32] The CMS collaboration. "Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV". In: *Eur. Phys. J. C* 75.5 (2015), p. 212. DOI: 10.1140/epjc/s10052-015-3351-7. arXiv: 1412.8662 [hep-ex].

[33] The ATLAS collaboration and the CMS collaboration. "Measurements of the Higgs boson production and decay rates and constraints on its couplings from a combined ATLAS and CMS analysis of the LHC pp collision data at $\sqrt{s}$ = 7 and 8 TeV". In: *JHEP* 08 (2016), p. 045. DOI: 10.1007/JHEP08(2016)045. arXiv: 1606.02266 [hep-ex].

[34] The CMS collaboration. "Observation of the Higgs boson decay to a pair of τ leptons with the CMS detector". In: *Phys. Lett.* B779 (2018), pp. 283–316. DOI: 10.1016/j.physletb.2018.02.004. arXiv: 1708.00373 [hep-ex].

[35] The ATLAS collaboration. "Cross-section measurements of the Higgs boson decaying into a pair of τ-leptons in proton-proton collisions at $\sqrt{s}$ = 13 TeV with the ATLAS detector". In: *Phys. Rev.* D99 (2019), p. 072001. DOI: 10.1103/PhysRevD.99.072001. arXiv: 1811.08856 [hep-ex].

[36] The CMS collaboration. "Observation of $\mathrm{t\bar{t}H}$ production". In: *Phys. Rev. Lett.* 120.23 (2018), p. 231801. DOI: 10.1103/PhysRevLett.120.231801. arXiv: 1804.02610 [hep-ex].

[37] The ATLAS collaboration. "Observation of Higgs boson production in association with a top quark pair at the LHC with the ATLAS detector". In: *Phys. Lett.* B784 (2018), pp. 173–191. DOI: 10.1016/j.physletb.2018.07.035. arXiv: 1806.00425 [hep-ex].

[38] The ATLAS collaboration. "Observation of $H \to b\bar{b}$ decays and VH production with the ATLAS detector". In: *Phys. Lett.* B786 (2018), pp. 59–86. DOI: 10.1016/j.physletb.2018.09.013. arXiv: 1808.08238 [hep-ex].

[39] The CMS collaboration. "Observation of Higgs boson decay to bottom quarks". In: *Phys. Rev. Lett.* 121.12 (2018), p. 121801. DOI: 10.1103/PhysRevLett.121.121801. arXiv: 1808.08242 [hep-ex].

[40] The CMS collaboration. "A measurement of the Higgs boson mass in the diphoton decay channel". In: *Physics Letters B* 805 (2020), p. 135425. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2020.135425. URL: http://www.sciencedirect.com/science/article/pii/S037026932030229X.

[41] The CMS Collaboration. "Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at $\sqrt{s}$ = 13 TeV". In: *Journal of High Energy Physics* 2017.11 (2017), p. 47. ISSN: 1029-8479. DOI: 10.1007/JHEP11(2017)047. URL: https://doi.org/10.1007/JHEP11(2017)047.

[42] The CMS collaboration. "Measurements of the Higgs boson width and anomalous HVV couplings from on-shell and off-shell production in the four-lepton final state". In: *Phys. Rev. D* 99.11 (2019), p. 112003. DOI: 10.1103/PhysRevD.99.112003. arXiv: 1901.00174 [hep-ex].

[43] The ATLAS collaboration. "Search for the Standard Model Higgs boson decay to $\mu^+\mu^-$ with the ATLAS detector". In: *Physics Letters B* 738 (2014), pp. 68–86. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2014.09.008. URL: https://www.sciencedirect.com/science/article/pii/S0370269314006583.

[44] The CMS collaboration. "Search for a standard model-like Higgs boson in the $\mu^+\mu$ and e^+e decay channels at the LHC". In: *Physics Letters B* 744 (2015), pp. 184 –207. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2015.03.048. URL: http://www.sciencedirect.com/science/article/pii/S0370269315002117.

[45] The ATLAS collaboration. *A search for the rare decay of the Standard Model Higgs boson to dimuons in pp collisions at $\sqrt{s} = 13$ TeV with the ATLAS Detector*. Tech. rep. ATLAS-CONF-2018-026. Geneva: CERN, 2018. URL: https://cds.cern.ch/record/2628763.

[46] The ATLAS collaboration. "A search for the dimuon decay of the Standard Model Higgs boson with the ATLAS detector". In: *Phys. Lett. B* 812 (2021), p. 135980. DOI: 10.1016/j.physletb.2020.135980. arXiv: 2007.07830 [hep-ex].

[47] The CMS collaboration. "Search for the Higgs boson decaying to two muons in proton-proton collisions at $\sqrt{s} = 13$ TeV". In: *Phys. Rev. Lett.* 122.2 (2019), p. 021801. DOI: 10.1103/PhysRevLett.122.021801. arXiv: 1807.06325 [hep-ex].

[48] The CMS collaboration. "Combined measurements of Higgs boson couplings in proton–proton collisions at $\sqrt{s} = 13$ TeV". In: *Eur. Phys. J. C* 79.5 (2019), p. 421. DOI: 10.1140/epjc/s10052-019-6909-y. arXiv: 1809.10733 [hep-ex].

[49] LHC Higgs Cross Section Working Group. *Handbook of LHC Higgs Cross Sections: 3. Higgs Properties: Report of the LHC Higgs Cross Section Working Group*. Ed. by S Heinemeyer. CERN Yellow Reports: Monographs. 2013. DOI: 10.5170/CERN-2013-004. URL: https://cds.cern.ch/record/1559921.

[50] Lyndon Evans and Philip Bryant. "LHC Machine". In: *JINST* 3 (2008), S08001. DOI: 10.1088/1748-0221/3/08/S08001.

[51] *CMS Luminosity public plots*. URL: https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[52] The CMS collaboration. "The CMS Experiment at the CERN LHC". In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[53] *Pseudorapidity*. URL: https://en.wikipedia.org/wiki/Pseudorapidity#/media/File:Pseudorapidity2.png.

[54] The CMS collaboration. *CMS Physics: Technical Design Report Volume 1: Detector Performance and Software*. Technical design report. CMS. Geneva: CERN, 2006. URL: https://cds.cern.ch/record/922757.

[55] The CMS collaboration. *The CMS tracker system project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/368412.

[56] The CMS collaboration. *The CMS tracker: addendum to the Technical Design Report*. Technical design report. CMS. Geneva: CERN, 2000. URL: https://cds.cern.ch/record/490194.

[57] The CMS collaboration. *The Phase-2 Upgrade of the CMS Tracker*. Tech. rep. CERN-LHCC-2017-009. CMS-TDR-014. Geneva: CERN, 2017. URL: http://cds.cern.ch/record/2272264.

[58] J. D'Hondt. "The CMS tracker system". In: *IEEE Nuclear Science Symposium Conference Record, 2005* 2 (2005), pp. 1084–1087.

[59] The CMS collaboration. *CMS Technical Design Report for the Pixel Detector Upgrade*. Tech. rep. CERN-LHCC-2012-016. CMS-TDR-11. 2012. URL: https://cds.cern.ch/record/1481838.

[60] Silvio Donato. "Sviluppo di un nuovo trigger con identificazione di quark-b per il canale Z($\nu\nu$)H(bb) nell'esperimento CMS ad LHC". 2013. URL: http://cds.cern.ch/record/1563462.

[61] The CMS collaboration. *The CMS electromagnetic calorimeter project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/349375.

[62] The CMS collaboration. *The CMS hadron calorimeter project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/357153.

[63] The CMS collaboration. *CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter*. Tech. rep. CERN-LHCC-2012-015. CMS-TDR-10. 2012. URL: https://cds.cern.ch/record/1481837.

[64] The CMS collaboration. *The CMS muon project: Technical Design Report*. Technical design report. CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/343814.

[65] The CMS collaboration. *The CMS muon project: Technical Design Report*. Technical Design Report CMS. Geneva: CERN, 1997. URL: https://cds.cern.ch/record/343814.

[66] The CMS collaboration. *A GEM Detector System for an Upgrade of the High-eta Muon Endcap Stations GE1/1 + ME1/1 in CMS*. Tech. rep. 2012. arXiv: 1211.1494. URL: https://cds.cern.ch/record/1492937.

[67] The CMS collaboration. *The Phase-2 Upgrade of the CMS Muon Detectors*. Tech. rep. CERN-LHCC-2017-012. CMS-TDR-016. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2283189.

[68] The CMS collaboration. *CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems*. Technical design report. CMS. URL: https://cds.cern.ch/record/706847.

[69] The CMS collaboration. "The CMS high level trigger". In: *Eur. Phys. J.* C46 (2006), pp. 605–667. DOI: 10.1140/epjc/s2006-02495-8. arXiv: hep-ex/0512077 [hep-ex].

[70] The CMS collaboration. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *JINST* 9.10 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569 [physics.ins-det].

[71] R. Fruhwirth. “Application of Kalman filtering to track and vertex fitting”. In: *Nucl. Instrum. Meth. A* 262 (1987), pp. 444–450. DOI: 10.1016/0168-9002(87)90887-4.

[72] The CMS collaboration. “2017 tracking performance plots”. In: (2017). URL: https://cds.cern.ch/record/2290524.

[73] K. Rose. “Deterministic annealing for clustering, compression, classification, regression, and related optimization problems”. In: *IEEE Proc.* 86.11 (1998), pp. 2210–2239. DOI: 10.1109/5.726788.

[74] The CMS collaboration. *Vertex Fitting in the CMS Tracker*. Tech. rep. CMS-NOTE-2006-032. Geneva: CERN, 2006. URL: https://cds.cern.ch/record/927395.

[75] The CMS collaboration. “Adaptive vertex fitting”. In: *J. Phys. G* 34 (2007), N343. DOI: 10.1088/0954-3899/34/12/N01.

[76] The CMS collaboration. “Particle-flow reconstruction and global event description with the CMS detector”. In: *JINST* 12.10 (2017), P10003. DOI: 10.1088/1748-0221/12/10/P10003. arXiv: 1706.04965 [physics.ins-det].

[77] The CMS collaboration. “Performance of the CMS muon detector and muon reconstruction with proton-proton collisions at $\sqrt{s}$ = 13 TeV”. In: *JINST* 13.06 (2018), P06015. DOI: 10.1088/1748-0221/13/06/P06015. arXiv: 1804.04528 [physics.ins-det].

[78] The CMS collaboration. “Performance of CMS Muon Reconstruction in *pp* Collision Events at $\sqrt{s}$ = 7 TeV”. In: *JINST* 7 (2012), P10002. DOI: 10.1088/1748-0221/7/10/P10002. arXiv: 1206.4071 [physics.ins-det].

[79] Elisabetta Manca. “Validation of the muon momentum resolution in view of the W mass measurement with the CMS experiment”. 2016. URL: http://cds.cern.ch/record/2233647.

[80] The CMS collaboration. “Performance of the reconstruction and identification of high-momentum muons in proton-proton collisions at $\sqrt{s}$ = 13 TeV”. In: *JINST* 15.02 (2020), P02027. DOI: 10.1088/1748-0221/15/02/P02027. arXiv: 1912.03516 [physics.ins-det].

[81] The CMS collaboration. “Performance of CMS muon reconstruction in pp collision events at$\sqrt{}$s = 7TeV”. In: *Journal of Instrumentation* 7.10 (2012), P10002–P10002. DOI: 10.1088/1748-0221/7/10/p10002. URL: https://doi.org/10.1088/1748-0221/7/10/p10002.

[82] The CMS collaboration. “Measurements of properties of the Higgs boson decaying into the four-lepton final state in pp collisions at $\sqrt{s}$ = 13 TeV”. In: *JHEP* 11 (2017), p. 047. DOI: 10.1007/JHEP11(2017)047. arXiv: 1706.09936 [hep-ex].

[83] The CMS collaboration. “Electron reconstruction in CMS at the LHC”. In: *Nucl. Phys. B Proc. Suppl.* 172 (2007). Ed. by P.S. Marrocchesi et al., pp. 257–259. DOI: 10.1016/j.nuclphysbps.2007.08.149.

[84] The CMS collaboration. "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s}$ = 8 TeV". In: *JINST* 10.06 (2015), P06005. DOI: 10.1088/1748-0221/10/06/P06005. arXiv: 1502.02701 [physics.ins-det].

[85] The CMS collaboration. *Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC*. Tech. rep. CMS-NOTE-2005-001. Geneva: CERN, 2005. URL: https://cds.cern.ch/record/815410.

[86] The CMS collaboration. "Electron and Photon performance in CMS with the full 2017 data sample and additional 2016 highlights for the CALOR 2018 Conference". In: (2018). URL: https://cds.cern.ch/record/2320638.

[87] S. Catani et al. "Longitudinally invariant K_t clustering algorithms for hadron hadron collisions". In: *Nucl. Phys.* B406 (1993), pp. 187–224. DOI: 10.1016/0550-3213(93)90166-M.

[88] M. Wobisch and T. Wengler. "Hadronization corrections to jet cross-sections in deep inelastic scattering". In: *Monte Carlo generators for HERA physics. Proceedings, Workshop, Hamburg, Germany, 1998-1999*. 1998, pp. 270–279. arXiv: hep-ph/9907280 [hep-ph].

[89] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti-k_t jet clustering algorithm". In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[90] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "FastJet User Manual". In: *Eur. Phys. J.* C72 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].

[91] Stephen D. Ellis and Davison E. Soper. "Successive combination jet algorithm for hadron collisions". In: *Phys. Rev.* D48 (1993), pp. 3160–3166. DOI: 10.1103/PhysRevD.48.3160. arXiv: hep-ph/9305266 [hep-ph].

[92] Matteo Cacciari and Gavin P. Salam. "Dispelling the N^3 myth for the k_t jet-finder". In: *Phys. Lett.* B641 (2006), pp. 57–61. DOI: 10.1016/j.physletb.2006.08.037. arXiv: hep-ph/0512210 [hep-ph].

[93] The CMS collaboration. "Jet energy scale and resolution in the CMS experiment in pp collisions at 8 TeV". In: *JINST* 12.02 (2017), P02014. DOI: 10.1088/1748-0221/12/02/P02014. arXiv: 1607.03663 [hep-ex].

[94] The CMS collaboration. "Identification of b-Quark Jets with the CMS Experiment". In: *JINST* 8 (2013), P04013. DOI: 10.1088/1748-0221/8/04/P04013. arXiv: 1211.4462 [hep-ex].

[95] The CMS collaboration. "Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV". In: *JINST* 13.05 (2018), P05011. DOI: 10.1088/1748-0221/13/05/P05011. arXiv: 1712.07158 [physics.ins-det].

[96] R. Barate et al. "Studies of quantum chromodynamics with the ALEPH detector". In: *Phys.Rept.* 294 (1998), pp. 1–165. DOI: 10.1016/S0370-1573(97)00045-8.

[97] The CMS collaboration. *Performance of quark/gluon discrimination using pp collision data at $\sqrt{s}$=8 TeV*. CMS Physics Analysis Summary CMS-PAS-JME-13-002. CERN, 2013. URL: http://cdsweb.cern.ch/record/1599732.

[98] The CMS collaboration. "Search for a Higgs boson in the decay channel H to $ZZ^{(*)}$ to $q\bar{q}\ell^{-}\ell^{+}$ in pp collisions at $\sqrt{s}$ = 7 TeV". In: *JHEP* 04 (2012), p. 036. DOI: 10.1007/JHEP04(2012)036. arXiv: 1202.1416 [hep-ex].

[99] The CMS collaboration. "Search for the standard model Higgs boson produced through vector boson fusion and decaying to $b\bar{b}$". In: *Phys. Rev. D* 92 (3 Aug. 2015), p. 032008. DOI: 10.1103/PhysRevD.92.032008. URL: https://link.aps.org/doi/10.1103/PhysRevD.92.032008.

[100] The CMS collaboration. "Measurement of the Hadronic Activity in Events with a Z and Two Jets and Extraction of the Cross Section for the Electroweak Production of a Z with Two Jets in pp Collisions at $\sqrt{s}$ = 7 TeV". In: *JHEP* 10 (2013), p. 062. DOI: 10.1007/JHEP10(2013)062. arXiv: 1305.7389 [hep-ex].

[101] The CMS collaboration. *Jet algorithms performance in 13 TeV data*. Tech. rep. CMS-PAS-JME-16-003. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2256875.

[102] E.D. Malaza. "Multiplicity Distributions in Quark and Gluon Jets to $O(\alpha_S)$". In: *Z. Phys. C* 31 (1986), p. 143. DOI: 10.1007/BF01559605.

[103] The CMS collaboration. *Commissioning of TrackJets in pp Collisions at $\sqrt{s}$=7 TeV*. CMS Physics Analysis Summary CMS-PAS-JME-10-006. CERN, 2010. URL: http://cdsweb.cern.ch/record/1275133.

[104] The CMS collaboration. *Performance of Jet Reconstruction with Charged Tracks only*. CMS Physics Analysis Summary CMS-PAS-JME-08-001. CERN, 2009. URL: http://cdsweb.cern.ch/record/1198681.

[105] The CMS collaboration. "Evidence for Higgs boson decay to a pair of muons". In: *JHEP* 01 (2021), p. 148. DOI: 10.1007/JHEP01(2021)148. arXiv: 2009.04363 [hep-ex].

[106] P.A. Zyla et al. "Review of Particle Physics". In: *PTEP* 2020.8 (2020), p. 083C01. DOI: 10.1093/ptep/ptaa104.

[107] H. Voss et al. "TMVA, the Toolkit for Multivariate Data Analysis with ROOT". In: *PoS* ACAT (2007), p. 040. DOI: 10.22323/1.050.0040.

[108] The ATLAS collaboration. "Search for the neutral Higgs bosons of the Minimal Supersymmetric Standard Model in pp collisions at $\sqrt{s}$ = 7 TeV with the ATLAS detector". In: *JHEP* 02 (2013), p. 095. DOI: 10.1007/JHEP02(2013)095. arXiv: 1211.6956 [hep-ex].

[109] The CMS collaboration. "Measurement of electroweak production of two jets in association with a Z boson in proton-proton collisions at $\sqrt{s}$ = 8 TeV". In: *Eur. Phys. J. C* 75.2 (2015), p. 66. DOI: 10.1140/epjc/s10052-014-3232-5. arXiv: 1410.3153 [hep-ex].

[110] The CMS collaboration. "Electroweak production of two jets in association with a Z boson in proton–proton collisions at $\sqrt{s} = 13$". In: *Eur. Phys. J.* C78.7 (2018), p. 589. DOI: 10.1140/epjc/s10052-018-6049-9. arXiv: 1712.09814 [hep-ex].

[111] The CMS collaboration. "Measurement of electroweak production of a W boson and two forward jets in proton-proton collisions at $\sqrt{s} = 8$ TeV". In: *JHEP* 11 (2016), p. 147. DOI: 10.1007/JHEP11(2016)147. arXiv: 1607.06975 [hep-ex].

[112] The CMS collaboration. "Measurement of electroweak production of a W boson in association with two jets in proton–proton collisions at $\sqrt{s} = 13$ TeV". In: *Eur. Phys. J. C* 80.1 (2020), p. 43. DOI: 10.1140/epjc/s10052-019-7585-7. arXiv: 1903.04040 [hep-ex].

[113] Paolo Nason. "A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms". In: *Journal of High Energy Physics* 2004.11 (2004), pp. 040–040. DOI: 10.1088/1126-6708/2004/11/040. URL: https://doi.org/10.1088%2F1126-6708%2F2004%2F11%2F040.

[114] Stefano Frixione, Paolo Nason, and Carlo Oleari. "Matching NLO QCD computations with parton shower simulations: the POWHEG method". In: *Journal of High Energy Physics* 2007.11 (2007), pp. 070–070. DOI: 10.1088/1126-6708/2007/11/070. URL: https://doi.org/10.1088%2F1126-6708%2F2007%2F11%2F070.

[115] Simone Alioli et al. "A general framework for implementing NLO calculations in shower Monte Carlo programs: the POWHEG BOX". In: *Journal of High Energy Physics* 2010.6 (2010), p. 43. ISSN: 1029-8479. DOI: 10.1007/JHEP06(2010)043. URL: https://doi.org/10.1007/JHEP06(2010)043.

[116] J. Alwall et al. "The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations". In: *JHEP* 07 (2014), p. 079. DOI: 10.1007/JHEP07(2014)079. arXiv: 1405.0301 [hep-ph].

[117] Torbjörn Sjöstrand, Stephen Mrenna, and Peter Skands. "A brief introduction to PYTHIA 8.1". In: *Computer Physics Communications* 178.11 (2008), pp. 852 –867. ISSN: 0010-4655. DOI: https://doi.org/10.1016/j.cpc.2008.01.036. URL: http://www.sciencedirect.com/science/article/pii/S0010465508000441.

[118] Johannes Bellm et al. "Herwig 7.0/Herwig++ 3.0 release note". In: *The European Physical Journal C* 76.4 (2016), p. 196. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-016-4018-8. URL: https://doi.org/10.1140/epjc/s10052-016-4018-8.

[119] J. Alwall et al. "Comparative study of various algorithms for the merging of parton showers and matrix elements in hadronic collisions". In: *The European Physical Journal C* 53.3 (2008), pp. 473–500. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-007-0490-5. URL: https://doi.org/10.1140/epjc/s10052-007-0490-5.

[120] Rikkert Frederix and Stefano Frixione. "Merging meets matching in MC@NLO". In: *Journal of High Energy Physics* 2012.12 (2012), p. 61. ISSN: 1029-8479. DOI: 10.1007/JHEP12(2012)061. URL: https://doi.org/10.1007/JHEP12(2012)061.

[121] The CMS collaboration. "Event generator tunes obtained from underlying event and multiparton scattering measurements". In: *The European Physical Journal C* 76.3 (2016), p. 155. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-016-3988-x. URL: https://doi.org/10.1140/epjc/s10052-016-3988-x.

[122] The CMS collaboration. "Extraction and validation of a new set of CMS pythia8 tunes from underlying-event measurements". In: *The European Physical Journal C* 80.1 (2020), p. 4. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-019-7499-4. URL: https://doi.org/10.1140/epjc/s10052-019-7499-4.

[123] The CMS collaboration. *Extraction and validation of a set of HERWIG 7 tunes from CMS underlying-event measurements*. Tech. rep. CMS-PAS-GEN-19-001. Geneva: CERN, 2020. URL: https://cds.cern.ch/record/2719158.

[124] Richard D. Ball et al. "Parton distributions for the LHC run II". In: *Journal of High Energy Physics* 2015.4 (2015), p. 40. ISSN: 1029-8479. DOI: 10.1007/JHEP04(2015)040. URL: https://doi.org/10.1007/JHEP04(2015)040.

[125] Richard D. Ball et al. "Parton distributions from high-precision collider data". In: *The European Physical Journal C* 77.10 (2017), p. 663. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-017-5199-5. URL: https://doi.org/10.1140/epjc/s10052-017-5199-5.

[126] S. Agostinelli et al. "GEANT4–a simulation toolkit". In: *Nucl. Instrum. Meth. A* 506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.

[127] Keith Hamilton et al. "NNLOPS simulation of Higgs boson production". In: *Journal of High Energy Physics* 2013.10 (2013), p. 222. ISSN: 1029-8479. DOI: 10.1007/JHEP10(2013)222. URL: https://doi.org/10.1007/JHEP10(2013)222.

[128] Keith Hamilton, Paolo Nason, and Giulia Zanderighi. "Finite quark-mass effects in the NNLOPS POWHEG+MiNLO Higgs generator". In: *Journal of High Energy Physics* 2015.5 (2015), p. 140. ISSN: 1029-8479. DOI: 10.1007/JHEP05(2015)140. URL: https://doi.org/10.1007/JHEP05(2015)140.

[129] Barbara Jäger et al. "Parton-shower effects in Higgs production via Vector-Boson Fusion". In: *Eur. Phys. J. C* 80.8 (2020), p. 756. DOI: 10.1140/epjc/s10052-020-8326-7. arXiv: 2003.12435 [hep-ph].

[130] Paolo Nason and Carlo Oleari. "NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG". In: *Journal of High Energy Physics* 2010.2 (2010), p. 37. ISSN: 1029-8479. DOI: 10.1007/JHEP02(2010)037. URL: https://doi.org/10.1007/JHEP02(2010)037.

[131] Gionata Luisoni et al. "$HW^{\pm}/HZ$ + 0 and 1 jet at NLO with the POWHEG BOX interfaced to GoSam and their merging within MiNLO". In: *Journal of High Energy Physics* 2013.10 (2013), p. 83. ISSN: 1029-8479. DOI: 10.1007/JHEP10(2013)083. URL: https://doi.org/10.1007/JHEP10(2013)083.

[132] H. B. Hartanto et al. "Higgs boson production in association with top quarks in the POWHEG BOX". In: *Phys. Rev. D* 91 (9 2015), p. 094003. DOI: 10.1103/PhysRevD.91.094003. URL: https://link.aps.org/doi/10.1103/PhysRevD.91.094003.

[133] Charalampos Anastasiou et al. "High precision determination of the gluon fusion Higgs boson cross-section at the LHC". In: *Journal of High Energy Physics* 2016.5 (2016), p. 58. ISSN: 1029-8479. DOI: 10.1007/JHEP05(2016)058. URL: https://doi.org/10.1007/JHEP05(2016)058.

[134] Matteo Cacciari et al. "Fully Differential Vector-Boson-Fusion Higgs Production at Next-to-Next-to-Leading Order". In: *Phys. Rev. Lett.* 115 (8 2015), p. 082002. DOI: 10.1103/PhysRevLett.115.082002. URL: https://link.aps.org/doi/10.1103/PhysRevLett.115.082002.

[135] Oliver Brein, Abdelhak Djouadi, and Robert Harlander. "NNLO QCD corrections to the Higgs-strahlung processes at hadron colliders". In: *Physics Letters B* 579.1 (2004), pp. 149 –156. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2003.10.112. URL: http://www.sciencedirect.com/science/article/pii/S0370269303017234.

[136] S. Dawson et al. "Associated Higgs boson production with top quarks at the CERN Large Hadron Collider: NLO QCD corrections". In: *Phys. Rev. D* 68 (3 2003), p. 034022. DOI: 10.1103/PhysRevD.68.034022. URL: https://link.aps.org/doi/10.1103/PhysRevD.68.034022.

[137] S. Frixione et al. "Weak corrections to Higgs hadroproduction in association with a top-quark pair". In: *Journal of High Energy Physics* 2014.9 (2014), p. 65. ISSN: 1029-8479. DOI: 10.1007/JHEP09(2014)065. URL: https://doi.org/10.1007/JHEP09(2014)065.

[138] A. Djouadi, J. Kalinowski, and M. Spira. "HDECAY: a program for Higgs boson decays in the Standard Model and its supersymmetric extension". In: *Computer Physics Communications* 108.1 (1998), pp. 56 –74. ISSN: 0010-4655. DOI: https://doi.org/10.1016/S0010-4655(97)00123-9. URL: http://www.sciencedirect.com/science/article/pii/S0010465597001239.

[139] The CMS collaboration. *Measurement of electroweak production of two jets in association with a Z boson in proton-proton collisions at $\sqrt{s}$ = 13 TeV*. Tech. rep. CMS-PAS-SMP-16-018. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2261499.

[140] Ye Li and Frank Petriello. "Combining QCD and electroweak corrections to dilepton production in the framework of the FEWZ simulation code". In: *Phys. Rev. D* 86 (9 2012), p. 094034. DOI: 10.1103/PhysRevD.86.094034. URL: https://link.aps.org/doi/10.1103/PhysRevD.86.094034.

[141] Michał Czakon and Alexander Mitov. "Top++: A program for the calculation of the top-pair cross-section at hadron colliders". In: *Computer Physics Communications* 185.11 (2014), pp. 2930 –2938. ISSN: 0010-4655. DOI: https://doi.org/10.1016/j.cpc.2014.06.021. URL: http://www.sciencedirect.com/science/article/pii/S0010465514002264.

[142] T. Gehrmann et al. "W^+W^- Production at Hadron Colliders in Next to Next to Leading Order QCD". In: *Phys. Rev. Lett.* 113 (21 2014), p. 212001. DOI: 10.1103/PhysRevLett.113.212001. URL: https://link.aps.org/doi/10.1103/PhysRevLett.113.212001.

[143] Massimiliano Grazzini, Stefan Kallweit, and Dirk Rathlev. "ZZ production at the LHC: Fiducial cross sections and distributions in NNLO QCD". In: *Physics Letters B* 750 (2015), pp. 407 –410. ISSN: 0370-2693. DOI: https://doi.org/10.1016/j.physletb.2015.09.055. URL: http://www.sciencedirect.com/science/article/pii/S0370269315007303.

[144] Massimiliano Grazzini et al. "$\mathrm{W}^{\pm}\mathrm{Z}$ production at the LHC: fiducial cross sections and distributions in NNLO QCD". In: *Journal of High Energy Physics* 2017.5 (2017), p. 139. ISSN: 1029-8479. DOI: 10.1007/JHEP05(2017)139. URL: https://doi.org/10.1007/JHEP05(2017)139.

[145] David L. Rainwater, R. Szalapski, and D. Zeppenfeld. "Probing color singlet exchange in Z + two jet events at the CERN LHC". In: *Phys. Rev. D* 54 (1996), pp. 6680–6689. DOI: 10.1103/PhysRevD.54.6680. arXiv: hep-ph/9605444.

[146] R. Brun and F. Rademakers. "ROOT: An object oriented data analysis framework". In: *Nucl. Instrum. Meth. A* 389 (1997). Ed. by M. Werlen and D. Perret-Gallix, pp. 81–86. DOI: 10.1016/S0168-9002(97)00048-X.

[147] François Chollet et al. *Keras*. https://keras.io. 2015.

[148] Martín Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. Software available from tensorflow.org. 2015. URL: http://tensorflow.org/.

[149] Payam Refaeilzadeh, Lei Tang, and Huan Liu. "Cross-Validation". In: *Encyclopedia of Database Systems*. Ed. by LING LIU and M. TAMER ÖZSU. Boston, MA: Springer US, 2009, pp. 532–538. ISBN: 978-0-387-39940-9. DOI: 10.1007/978-0-387-39940-9_565. URL: https://doi.org/10.1007/978-0-387-39940-9_565.

[150] Glen Cowan et al. "Asymptotic formulae for likelihood-based tests of new physics". In: *The European Physical Journal C* 71.2 (2011), p. 1554. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-011-1554-0. URL: https://doi.org/10.1140/epjc/s10052-011-1554-0.

[151] Marco Delmastro et al. *Simplified Template Cross Sections – Stage 1.1*. Tech. rep. Geneva: CERN, 2019. arXiv: 1906.02754. URL: https://cds.cern.ch/record/2669925.

[152] The CMS collaboration. *CMS Luminosity Measurements for the 2016 Data Taking Period*. Tech. rep. CMS-PAS-LUM-17-001. Geneva: CERN, 2017. URL: https://cds.cern.ch/record/2257069.

[153] The CMS collaboration. *CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s}$ = 13 TeV*. Tech. rep. CMS-PAS-LUM-17-004. Geneva: CERN, 2018. URL: https://cds.cern.ch/record/2621960.

[154] The CMS collaboration. *CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s}$ = 13 TeV*. Tech. rep. CMS-PAS-LUM-18-002. Geneva: CERN, 2019. URL: https://cds.cern.ch/record/2676164.

[155] A. Bodek et al. "Extracting muon momentum scale corrections for hadron collider experiments". In: *The European Physical Journal C* 72.10 (2012), p. 2194. ISSN: 1434-6052. DOI: 10.1140/epjc/s10052-012-2194-8. URL: https://doi.org/10.1140/epjc/s10052-012-2194-8.

[156] J.S. Conway. "Incorporating Nuisance Parameters in Likelihoods for Multi-source Spectra". In: (2011), 115–120. 6 p. DOI: 10.5170/CERN-2011-006.115. arXiv: 1103.0354. URL: http://cds.cern.ch/record/1333496.

[157] Roger Barlow and Christine Beeston. "Fitting using finite Monte Carlo samples". In: *Computer Physics Communications* 77.2 (1993), pp. 219 –228. ISSN: 0010-4655. DOI: https://doi.org/10.1016/0010-4655(93)90005-W. URL: http://www.sciencedirect.com/science/article/pii/001046559390005W.

[158] The ATLAS collaboration the CMS collaboration and the LHC Higgs Combination Group. *Procedure for the LHC Higgs boson search combination in Summer 2011*. Tech. rep. CMS-NOTE-2011-005. ATL-PHYS-PUB-2011-11. Geneva: CERN, 2011. URL: https://cds.cern.ch/record/1379837.

[159] The CMS collaboration. *Combined Higgs boson production and decay measurements with up to 137 fb^{-1} of proton-proton collision data at $\sqrt{s}$ = 13 TeV*. Tech. rep. CMS-PAS-HIG-19-005. Geneva: CERN, 2020. URL: https://cds.cern.ch/record/2706103.

[160] The CMS collaboration. "Measurement of inclusive and differential Higgs boson production cross sections in the diphoton decay channel in proton-proton collisions at $\sqrt{s}$ = 13 TeV". In: *JHEP* 01 (2019), p. 183. DOI: 10.1007/JHEP01(2019)183. arXiv: 1807.03825 [hep-ex].

[161] M. Cepeda et al. "Report from Working Group 2: Higgs Physics at the HL-LHC and HE-LHC". In: *CERN Yellow Rep. Monogr.* 7 (2019). Ed. by Andrea Dainese et al., pp. 221–584. DOI: 10.23731/CYRM-2019-007.221. arXiv: 1902.00134 [hep-ph].

[162] A. Abada et al. "HE-LHC: The High-Energy Large Hadron Collider: Future Circular Collider Conceptual Design Report Volume 4". In: *Eur. Phys. J. ST* 228.5 (2019), pp. 1109–1382. DOI: 10.1140/epjst/e2019-900088-6.

[163] A. Abada et al. "FCC-ee: The Lepton Collider: Future Circular Collider Conceptual Design Report Volume 2". In: *Eur. Phys. J. ST* 228.2 (2019), pp. 261–623. DOI: 10.1140/epjst/e2019-900045-4.

[164] A. Abada et al. "FCC-hh: The Hadron Collider: Future Circular Collider Conceptual Design Report Volume 3". In: *Eur. Phys. J. ST* 228.4 (2019), pp. 755–1107. DOI: 10.1140/epjst/e2019-900087-0.

[165] The ATLAS collaboration. *Prospects for $H \to c\bar{c}$ using Charm Tagging with the ATLAS Experiment at the HL-LHC*. Tech. rep. ATL-PHYS-PUB-2018-016. Geneva: CERN, 2018. URL: http://cds.cern.ch/record/2633635.

[166] John C. Collins and Davison E. Soper. "Angular Distribution of Dileptons in High-Energy Hadron Collisions". In: *Phys. Rev. D* 16 (1977), p. 2219. DOI: 10.1103/PhysRevD.16.2219.

[167] L. D. Isenhower et al. *Polarized Drell-Yan measurements with the Fermilab Main Injector*. Tech. rep. DOI: 10.2172/1127901. URL: https://www.osti.gov/biblio/1127901.

[168] P.L. McGaughey, J.M. Moss, and J.C. Peng. "High-energy hadron induced dilepton production from nucleons and nuclei". In: *Ann. Rev. Nucl. Part. Sci.* 49 (1999), pp. 217–253. DOI: 10.1146/annurev.nucl.49.1.217. arXiv: hep-ph/9905409.

[169] The CMS collaboration. "Angular coefficients of Z bosons produced in pp collisions at $\sqrt{s}$ = 8 TeV and decaying to $\mu^+\mu^-$ as a function of transverse momentum and rapidity". In: *Phys. Lett. B* 750 (2015), pp. 154–175. DOI: 10.1016/j.physletb.2015.08.061. arXiv: 1504.03512 [hep-ex].

[170] Yoav Freund and Robert E Schapire. "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting". In: *Journal of Computer and System Sciences* 55.1 (1997), pp. 119 –139. ISSN: 0022-0000. DOI: https://doi.org/10.1006/jcss.1997.1504. URL: http://www.sciencedirect.com/science/article/pii/S002200009791504X.

[171] Leo Breiman. "Arcing classifier (with discussion and a rejoinder by the author)". In: *Ann. Statist.* 26.3 (June 1998), pp. 801–849. DOI: 10.1214/aos/1024691079. URL: https://doi.org/10.1214/aos/1024691079.

[172] L. Breiman. "Prediction Games and Arcing Algorithms". In: *Neural Computation* 11.7 (1999), pp. 1493–1517. DOI: 10.1162/089976699300016106.

[173] Jerome H. Friedman. "Greedy function approximation: A gradient boosting machine." In: *Ann. Statist.* 29.5 (Oct. 2001), pp. 1189–1232. DOI: 10.1214/aos/1013203451. URL: https://doi.org/10.1214/aos/1013203451.

[174] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. "Multilayer feedforward networks are universal approximators". In: *Neural Networks* 2.5 (1989), pp. 359–366. ISSN: 0893-6080. DOI: https://doi.org/10.1016/0893-6080(89)90020-8. URL: https://www.sciencedirect.com/science/article/pii/0893608089900208.

[175] Moshe Leshno et al. "Multilayer feedforward networks with a nonpolynomial activation function can approximate any function". In: *Neural Networks* 6.6 (1993), pp. 861–867. DOI: 10.1016/S0893-6080(05)80131-5. URL: https://doi.org/10.1016/S0893-6080(05)80131-5.

[176] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. "Learning representations by back-propagating errors". In: *Nature* 323.6088 (1986), pp. 533–536. ISSN: 1476-4687. DOI: 10.1038/323533a0. URL: https://doi.org/10.1038/323533a0.

[177] *CS231n: Convolutional Neural Networks for Visual Recognition.* URL: https://cs231n.github.io/neural-networks-3/.

[178] Nitish Srivastava et al. "Dropout: A Simple Way to Prevent Neural Networks from Overfitting". In: *J. Mach. Learn. Res.* 15.1 (Jan. 2014), 1929–1958. ISSN: 1532-4435.