SEARCH FOR DECAYS OF THE HIGGS BOSON INTO MUONS IN P-P COLLISIONS WITH THE CMS DETECTOR

LUISA FERNANDA CHAPARRO SIERRA

ADVISOR:

Juan Carlos Sanabria, PhD

Physics Department

Faculty of Science

Universidad de Los Andes

2017

CERN-THESIS-2017-323
18/01/2018

ACKNOWLEDGMENTS

Thanks to God, my parents and my partner Gaston for their support specially in the difficult times.
Thanks to my advisor, for all the long days giving me his knowledge and advice.
Special thanks to Doctor Adrian Perieanu and his team who allowed me to be part of an amazing working group.
Finally, but not less important, thanks to Universidad de Los Andes and Colciencias who gave me the financial support for this adventure.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

On July of 2012 the spokespersons of CMS and ATLAS experiments announced at CERN the discovery a neutral boson with a mass of 125 GeV and decay channels consistent with those of the SM Higgs boson. On March of 2013 at the *Rencontres du Moriond* Conference new evidence gathered by both experiments allowed the identification of this new boson as the Higgs boson, putting an end to a search that had started 30 years before, right after the discovery of the W and Z^0 vector bosons by the UA1 and UA2 experiments of the Sp$\bar{\text{p}}$S collider at CERN (announced in 1983). At that time the experimental confirmation of the existence of massive electroweak vector bosons represented the consolidation of the Standard Model of Particles and Fields. This model relied upon the spontaneous symmetry breaking of the $SU(2) \times U(1)$ gauge symmetry of the electromagnetic and weak interactions, mediated by a neutral scalar field, via a mechanism that has come to be known as the "Higgs mechanism" [1]. The presence of this scalar field and its coupling to the matter fields allowed the inclusion of mass terms for bosons and fermions, while preserving the gauge invariance, and therefore the renormalizability of the model. The great phenomenological success of this theory, that eventually became known as *The Standard Model*, is exemplified by the discovery of the W and Z^0 just as predicted. The very important role played by the scalar boson (Higgs boson) in providing mass for the other fields and preserving the necessary gauge symmetry made its search the most important and urgent problem in physics since 1983.

The mass of the Higgs boson in the SM is a free parameter, therefore its value was unconstrained by the theory, making its experimental search difficult, since a wide range of energies had to be scanned. The initial searches were performed at the S$p\bar{p}$S accelerator at CERN with negative results. Next came the Tevatron at Fermilab, that started in 1988 producing collisions between protons and antiprotons with the center-of-mass energy of 2 TeV. During the 1990's the electron-positron collider LEP of CERN joined the search without any positive outcome. Finally, in 2010 the Large Hadron Collider (LHC) at CERN started its operations colliding protons at a center-of-mass energy of 7 TeV, and by 2012 at 8 TeV. The first hints of a scalar boson were announced on december of 2011 and were confirmed by July of 2012. By 2013 strong evidence had convinced the particle physics community that the discovered particle was the Higgs boson.

The two experiments that performed the initial search for the Higgs boson at the LHC were CMS and ATLAS. In the scientific goals of those experiments, and of the whole LHC project, the main priority was the search of the Higgs boson. The CMS experiment concentrated on five channels: H$\to\gamma\gamma$, H$\to$ ZZ, H$\to$W$^+$W$^-$, H$\to$ $\tau^+\tau^-$ and H$\to$ b$\bar{\text{b}}$.

The initial announcements of the Collaboration were based on the first three channels, that have the highest signal yields and/or the lowest background contamination. By the time of the Moriond conference (Feb 2013) the H$\to$ $\tau^+\tau^-$ analysis also provided a signal. Currently, the effort of the experiment focused in channels with lesser yields and/or higher backgrounds like H$\to$ b$\bar{\text{b}}$, H$\to$ Zγ, H$\to\gamma\gamma^*$ and H$\to$ $\mu^+\mu^-$.

The High Energy Physics Group at Universidad de los Andes (Uniandes) is a member of the CMS Collaboration since 2006. As a PhD student in physics working at this university and with this group, I am also part of the CMS Collaboration, since 2012. I have worked directly at CERN between January of 2012 and July of 2014. During summer of 2015 and the first semester of 2016, as a part of the LHC Physics Centre (LPC) program at Fermilab and in collaboration with Northwestern University (Evanston, IL, USA), I continued working on CMS data analysis for the H$\rightarrow\gamma\gamma^*$ and H$\rightarrow$ $\mu^+\mu^-$ channels.

As a member of the CMS collaboration I acquired responsibilities at three levels: hardware, software and physics. Our group works in the operation of the Resistive-Plate-Chamber muon detectors (RPC) and in the prototyping of Gas Electron Multipliers (GEM) at hardware level, in muon reconstruction at the software level and in several physics analyses. For my dissertation, the Uniandes group focused its interest in Higgs physics, specially in the Higgs muonic decay channel. To perform this kind of analysis, I joined the CMS H2Mu Group.

The H$\rightarrow$ $\mu^+\mu^-$ branching fraction is very small (2.4×10^{-4}) however it is very important because it involves fermionic coupling (Yukawa coupling), and the mass of the final state fermions is small, which allows to study the dependence of the Higgs couplings on the mass of the different states. Therefore in the experimental studies of the Higgs properties being able to observe H$\rightarrow$ $\mu^+\mu^-$ and measure the branching fraction is a necessary step. For the second run of the LHC, the CMS collaboration has made an emphasis on this an other similar channels. For my dissertation I have focused on the search of H$\rightarrow$ $\mu^+\mu^-$ events, and specially in the optimization of the search through the recovery of events lost due to the emission of photons by the muons (final state radiation, FSR). In this thesis I present results of the methods that I have implemented to recover Higgs events (FSR Recovery) and the latest results of the CMS search of the evidences of the H$\rightarrow$ $\mu^+\mu^-$ decay.

During my time as a CMS collaboration member I also worked with the $HZ\gamma$ Analysis Group, focusing on the H$\rightarrow$ $\gamma\gamma^* \rightarrow \gamma\mu^+\mu^-$ analysis (Higgs Dalitz decay). During this time, I developed a High Level Trigger for the 13 TeV analysis and a complementary multivariate analysis for the identification of photons in order to improve the signal.

This document is organized as follows: Chapter 1 presents the physics of the Higgs boson including the state of the art of the experimental studies performed by CMS and ATLAS. In Chapter 2, the main features of the Large Hadron Collider (LHC) and the CMS Detector are described. Chapter 3 shows the details of the reconstruction techniques used by the CMS experiment making an emphasis on the physical objects used in this analysis (muons, photons, jets and electrons). Afterwards, a detailed description of the H$\rightarrow$ $\mu^+\mu^-$ channel, and the analysis performed by CMS H2Mu group with the 13 TeV data, is presented in Chapter 4. In Chapter 5 the recovery of the final state radiation events using a cut-based method is discussed. An alternative method to do the recovery using a multivariate analysis approach, is discussed in Chapter 6. Finally, the results of the two methods of recovery and the results of the analysis of the dimuon channel are presented in Chapter 7. The other contributions that I have made to CMS are presented in Appendixes A and B.

1

THE HIGGS BOSON

The main topic of this dissertation is the study of the Higgs boson decay channel into muons, therefore in this chapter a summary of the physics of the Higgs boson will be presented. The chapter will start with a discussion of the spontaneous symmetry breaking mechanism in the Standard Model (SM), the different mechanisms of production and the subsequent decay channels; emphasizing on the so called "discovery channels". Next, we will discuss in more detail the rare decay channels in which the dissertation is based. Finally, we will present briefly some of the physics scenarios beyond SM that are related with the Higgs sector.

1.1 THE ELECTROWEAK SYMMETRY BREAKING IN THE STANDARD MODEL

The Standard Model of Particles is the best theoretical framework for the description of the dynamics of particles and interactions at the most fundamental level that we have up to now. The current formulation was developed in the early 1970's. Over time, and through many experiments, the SM has been established as a well-tested physics theory. The main ingredients are:

- It is a quantum-mechanical theory
- It is a relativistic theory
- It gives a quantum-mechanical treatment to the interactions (it is a quantum field theory)
- It is a gauge field theory
- It describes the electromagnetic, weak and strong interactions
- It is a renormalizable theory

The electromagnetic and strong sectors of the SM can be described in terms of "pure" gauge field theories, Quantum Electrodynamics (QED) in the electromagnetic case and Quantum Chromodynamics (QCD) in the strong interaction case. QED is a gauge field theory based on the symmetry group $U(1)_{EM}$, which describes the electromagnetic interactions between quarks and leptons. QCD is a gauge field theory based on the symmetry group $SU(3)_C$ and describes the strong interactions between quarks [1]. As a result of this character, the interaction fields are massless: the photon field for QED and eight gluon fields for QCD. The gauge invariance of these theories guarantees their renormalizability. Massive interaction fields would break the gauge invariance and make the theories non-renormalizable.

The experimental evidence collected up to the 1950's for the weak interaction suggested that it was mediated by massive vector fields, both charged and neutral; therefore, this interaction could not be described by a pure gauge field theory. Even more, the discovery of parity violation by the weak interaction demanded a separation of the quiralities of the matter fields (fermion fields). The mass terms associated with the matter fields mix quiralities, and therefore also break any gauge invariance associated with the weak interaction. In conclusion, the situation by the end of the 1950's was that, due to the weak interaction, no quantum fields could have mass and preserve gauge invariance, and therefore the renormalizability of the theory was also broken.

The solution to this problem came in the form of the inclusion of a neutral scalar field with a degenerated vacuum state, the Higgs field. The field coupled to the matter fields through Yukawa terms. The spontaneous symmetry breaking of the Higgs vacuum state would generate the coupling of the matter fields to the non-zero vacuum expectation value, generating effective mass terms for these fields. The implementation of the symmetry breaking also resulted in massive and non-massive vector fields (interaction fields). The resulting massive fields were $W^{\pm}$ and Z^0, and the massless field was γ. Through this mechanism the theory recovered its renormalizability, solving the problem of the mass terms, and unifying the electromagnetic and weak interactions. The mass terms of all massive fields are the result of the interaction with the vacuum of the Higgs field. The Higgs field should also be massive, and should interact with any massive field with a strength dependent on the mass of the field.

The discovery of neutral-current interactions in the 1970's, represented the first direct evidence for the existence of the Z^0. At the beginning of the 1980's, the direct observation of $W^{\pm}$ and Z^0 consolidated the SM as the model for the electro-weak and strong interactions. The theory was renormalizable and very successful phenomenologically, but depended on the existence of a massive and neutral scalar field with positive parity, the Higgs field. The Higgs should interact with any other massive field, and the coupling should depend on the mass of the field. Unfortunately, in the SM the mass of the Higgs was not fixed, therefore the experimental search had to be driven by m_H, an unknown parameter, making it much more difficult [1].

The Higgs of the SM has several couplings which, in experimental terms, means that there are several production and decay channels. The very precise measurements of the electroweak sector of the SM, specially those performed at LEP in the 1990's, where such that, if m_H became known, cross sections and branching ratios for the Higgs production and decay would be pre-determined with high accuracy. The discovery of a scalar state with a mass of $\sim$125 GeV, announced in July of 2012 by the experiments ATLAS and CMS [2, 3], and the latter updates that indicate, to a high level of confidence, that it has spin 0 and positive parity, as well as couplings that depend on the mass of the state to which it is coupling, suggest very strongly that this state has all the features of the SM Higgs [4]. Therefore, one could fix $m_H = 125$ GeV, and have predictions for cross sections and branching ratios for the different Higgs channels. Any significant deviations from these values would represent important indications of physics beyond the SM. The detailed study of each and every one of the experimental Higgs channels becomes very important to understand the electroweak symmetry breaking mechanism of the SM, and what lies beyond.

1.2 HIGGS BOSON PRODUCTION AT THE LHC

The SM predicts that the Higgs boson can be produced in a number of ways, although the probability of any of these production channels in proton-proton collisions, in the energy range of the TeVs is very small; for example, only 1 Higgs boson per 10 billion collisions in the case of the Large Hadron Collider [5].

The main production mechanisms are associated to the fact that the Higgs boson couples preferentially to heavy particles, like the $W^{\pm}$ and Z^0 vector bosons, top quarks and bottom quarks. These mechanisms are: gluon fusion, vector-boson fusion, associated production with a gauge boson and associated production with heavy quarks. The Feynman diagrams of the main Higgs boson production mechanisms are presented in Figure 1 [6].

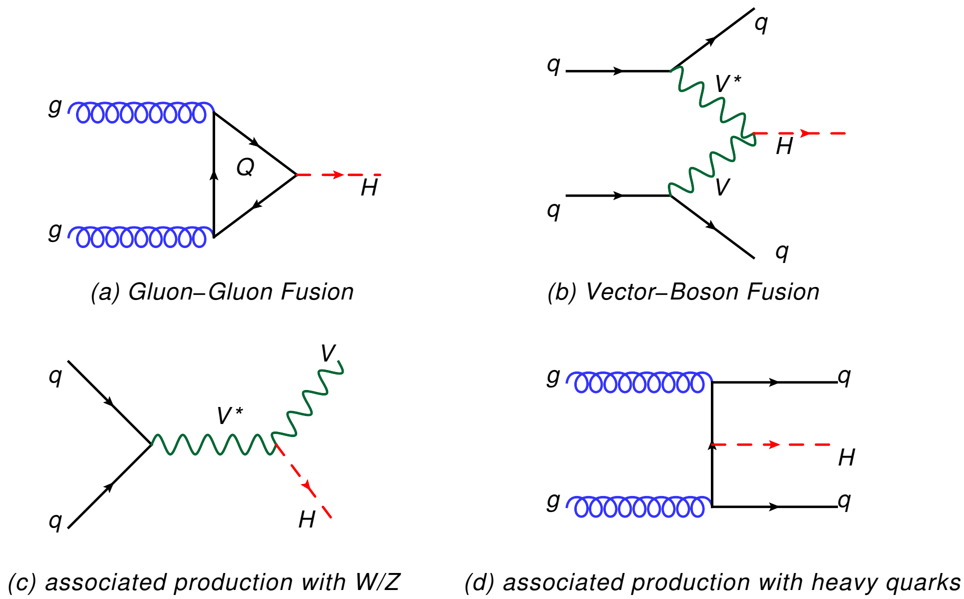

Figure 1: Generic Feynman diagrams for the dominant SM Higgs boson production mechanism at the LHC [6].

The main LHC Higgs production cross sections at a center-of-mass energy ($\sqrt{s}$) of 7, 8 and 14 TeV are summarized in Figure 2. For a 125 GeV Higgs boson the production cross sections, as a function of center of mass energy, are shown in Figure 3. Table 1 also summarizes these values with the relative uncertainties.

At LHC energies the main production mechanism is gluon fusion, but the other production channels have significant contributions. Looking at the topology of the whole event one can achieve a level of identification of these channels.

$\sqrt{s}$ (TeV)	Production cross section (in pb) for $m_H = 125$ GeV					
	ggF	VBF	WH	ZH	$t\bar{t}$H	total
7	$15.1^{+15\%}_{-15\%}$	$1.22^{+3\%}_{-2\%}$	$0.58^{+4\%}_{-4\%}$	$0.33^{+6\%}_{-6\%}$	$0.09^{+12\%}_{-18\%}$	17.4
8	$19.3^{+15\%}_{-15\%}$	$1.58^{+3\%}_{-2\%}$	$0.70^{+4\%}_{-5\%}$	$0.41^{+6\%}_{-6\%}$	$0.13^{+12\%}_{-18\%}$	22.1
14	$49.8^{+20\%}_{-15\%}$	$4.18^{+3\%}_{-3\%}$	$1.50^{+4\%}_{-4\%}$	$0.88^{+6\%}_{-5\%}$	$0.61^{+15\%}_{-28\%}$	57.0

Table 1: The SM Higgs boson production cross sections in pp collisions with $m_H = 125$ GeV.

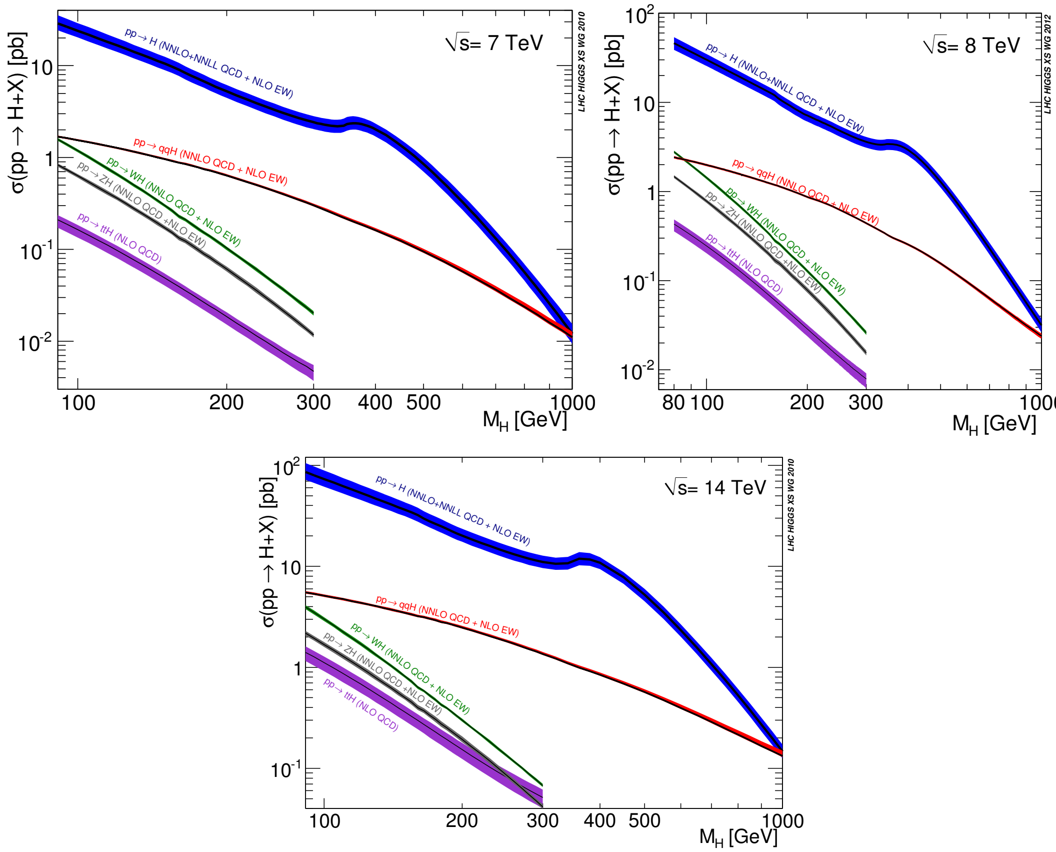

Figure 2: LHC Higgs boson production cross sections at a center of mass energy of 7 TeV (Upper left), 8 TeV (Upper right) and 14 TeV (Below) [3, 6].

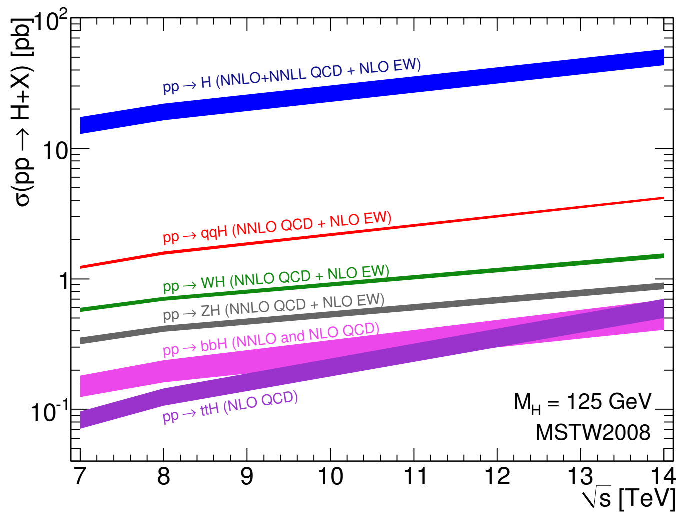

Figure 3: LHC Higgs boson production cross sections as a function of center of mass energy, $\sqrt{s}$, for pp collisions. The theoretical uncertainties are indicated as a band [6].

1.2.1 GLUON FUSION PRODUCTION MECHANISM

Gluon Fusion is the dominant Higgs boson production mechanism at high-energy hadron colliders (see Figure 1a). The Higgs boson is produced from the coupling of two gluons mediated by a loop of heavy fermions. Since the coupling of particles to the Higgs boson is proportional to their mass, the main contributions to the fermionic loop comes from the heaviest quarks. In practice, the top and bottom quark contributions dominate.

The production of the Higgs boson through gluon fusion would be sensitive to a fourth generation of quarks. Because the Higgs coupling is proportional to the fermion mass, the contribution of a heavier generation of quarks would not be suppressed as one would expected for any massive loops. In consequence, the Higgs cross section is sensitive to a fourth generation of quarks, even if they are very heavy for a direct discovery at the LHC. The mass range is limited by the scale of new physics where the standard model breaks down. In conclusion, the production of the Higgs boson throw a fermion loop provides a way to explore the physics beyond the SM.

1.2.2 VECTOR BOSON FUSION PRODUCTION MECHANISM

The second-largest SM Higgs boson production mechanism at the LHC is vector boson fusion (VBF). Higgs production via VBF proceeds by the scattering of two quarks or anti-quarks, mediated by $W^{\pm}$or Z^0 bosons that fuse into a Higgs (see Figure 1b). The scattered quarks give rise to two hard jets in the forward and backward regions of the detector. Because of the color-singlet nature of the boson exchange, the gluon radiation that comes from the central rapidity regions is suppressed. This characteristic behavior of VBF can be used to distinguish these production channels from the QCD backgrounds. The VBF channel provides a particularly clean environment, not only for Higgs searches, but also for the determination of Higgs couplings [7].

1.2.3 WH AND ZH ASSOCIATED PRODUCTION MECHANISMS

Higgs bosons can be produced also in association with Z^0 or $W^{\pm}$ via a process called "Higgs-strahlung". When quarks collide and merge to form a virtual $W^{\pm}$or Z^0 bosons with enough energy, they can radiate a Higgs boson (see Figure 1c). At the LHC this process is the third largest [7].

1.2.4 HIGGS PRODUCTION IN ASSOCIATION WITH $t\bar{t}$

The final process that is commonly considered is the production of a Higgs boson in the fusion of top and anti-top quarks (see Figure 1d). This channel provides important information about the top-Higgs Yukawa coupling. The production mechanism involves two colliding gluons, each of them producing a $t\bar{t}$ pair. The t from one gluon and the $\bar{t}$ from the other one combine to form a Higgs boson [5].

1.3 HIGGS BOSON DECAY CHANNELS

A Higgs mass of around 125 GeV provides an excellent opportunity to explore the Higgs couplings to many SM particles. The dominant decay modes for this mass are: $H\to b\bar{b}$ and $H\to W^-W^+$, followed by $H\to\gamma\gamma$, $H\to \tau^+\tau^-$, $H\to c\bar{c}$ and $H\to ZZ^*$. The $H\to\gamma\gamma$ channel has a smaller rate but very low background, making it a very important channel, while $H\to Z\gamma$ and $H\to \mu^+\mu^-$ have even smaller rates such that in the first run of the LHC no events were detected.

Since the decays into gluons, diphotons and $Z\gamma$ are loop induced, they provide information on the Higgs couplings to WW, ZZ and $t\bar{t}$ in different combinations [7]. Some of these channels are excellent candidates to show possible deviations from the SM predictions and might open the door to the observation of physics beyond the SM.

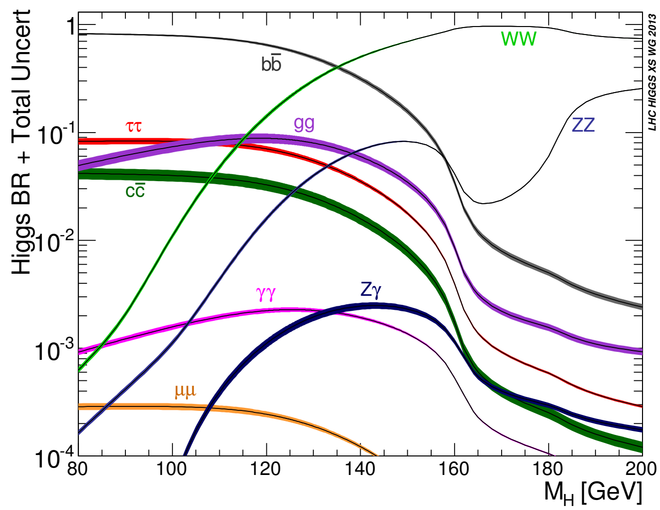

Figure 4: Branching fractions as a function of the Higgs boson mass. The width of the lines represents the total theoretical uncertainty [2].

For a given value of m_H, the search sensitivity depends on the production cross section, the decay branching ratio into the selected final state, the signal selection efficiency, the mass resolution and the level of background of similar or identical final state topologies.

Decay channel	Branching ratio	Rel.uncertainty
$H\rightarrow \gamma\gamma$	2.28×10^{-3}	$^{+5.0\%}_{-4.9\%}$
$H\rightarrow ZZ$	2.64×10^{-2}	$^{+4.3\%}_{-4.1\%}$
$H\rightarrow W^+W^-$	2.15×10^{-1}	$^{+4.3\%}_{-4.2\%}$
$H\rightarrow \tau^+\tau^-$	6.32×10^{-2}	$^{+5.7\%}_{-5.7\%}$
$H\rightarrow b\bar{b}$	5.77×10^{-1}	$^{+3.2\%}_{-3.3\%}$
$H\rightarrow Z\gamma$	1.54×10^{-3}	$^{+9.0\%}_{-8.9\%}$
$H\rightarrow \mu^+\mu^-$	2.19×10^{-4}	$^{+6.0\%}_{-5.9\%}$

Table 2: The branching ratios and the relative uncertainties for a SM Higgs boson with $m_H = 125$ GeV.

Combining the Higgs boson search results requires a simultaneous analysis of the data selected by the individual decay modes, accounting for their correlations and for all the statistical and systematic uncertainties. The statistical methodology used for this combination was developed by the ATLAS and CMS Collaborations in the context of the LHC Higgs Combination Group [3, 8–10].

The magnitude of the Higgs boson decay signals are characterized by the branching fractions and can be presented relative to the standard model expectations, and referred to as signal strengths μ. The predicted branching fractions for a Higgs boson as a function of Higgs-boson mass are shown in Figure 4.

The branching ratios for the most relevant decay modes of the SM Higgs boson, including the most recent theoretical uncertainties, are listed in Table 2 for $m_H = 125$ GeV.

1.3.1 THE DISCOVERY CHANNELS

For a low mass Higgs boson ($110 < m_H < 150$ GeV), where the natural width of the state is only a few MeV, the five decay channels that play an important role (the discovery channels) at the LHC are those listed in Table 3. In order to distinguish between different production modes, CMS and ATLAS usually split the Higgs boson candidates into several mutually exclusive tags based on the topological and/or kinematic features present in the event. In the next subsections we will describe the main features of these channels.

Decay Mode	Production tagging	# of subchannels	m_H range (GeV)	Mass Resolution
$\gamma\gamma$	untagged	4	110-150	1-2%
	dijet (VBF)	1or 2		
ZZ	untagged	3	110-600	1-2%
W^+W^-	untagged	4	110-600	20%
	dijet (VBF)	4		
$\tau^+\tau^-$	untagged	16	110-145	15%
	dijet (VBF)	4		
$b\bar{b}$	lepton, E_T^{miss} (VH)	10	110-135	10%

Table 3: The discovery channels. SM Higgs boson searches at the LHC. The numbers reported are for $m_H = 125$ GeV.

1.3.1.1 $H \to \gamma\gamma$

The $H\to\gamma\gamma$ channel is a rare decay mode, which is only detectable in a limited Higgs boson mass region between 80 and 150 GeV/c^2. In this region, the production cross section and the decay branching ratio are relatively large. The $H\to\gamma\gamma$ channel was the most important one in the initial search for the Higgs boson, because it provides a clean final state, which can be reconstructed with high precision [11].

The search is performed by looking for a narrow peak over a smoothly falling background, in the invariant mass distribution of two high p_T photons. In order to observe this peak above the irreducible $\gamma\gamma$ background an excellent energy and angular resolutions are required.

The background in this channel is high and come from prompt $\gamma\gamma$, $\gamma+$jet and dijet processes. Isolation is a powerful tool to reject the non-prompt background due to electromagnetic showers originating in jets. In addition to isolation criteria the following variables were also taken into account by ATLAS and CMS in the analyses for this channel: the ratio between hadronic energy associated to the reconstructed photon and its energy, the transverse width of the electromagnetic shower and an electron track veto. In the case of CMS, the photon candidates detected in the Barrel have less background than those detected in the Endcaps. For this reason it is useful to divide the events with photon candidates into categories and apply different selection criteria for each of them.

In order to optimize search sensitivity and also to separate the various Higgs production modes, ATLAS and CMS experiments split events into several mutually exclusive categories:

- The VH production category: diphoton events containing a high p_{T} muon, electron, dijets or missing energy ($\not{E}_{\mathrm{T}}$) consistent with the decay of a W or Z boson.
- The VBF production category: the events contain energetic dijets with a large mass and pseudo-rapidity (η) difference.
- The gluon fusion production category: the remaining events are considered.

While the VH category is relatively pure, the VBF category has significant contamination from the gluon fusion process [7].

Another layer of the categorization comes from the diphoton invariant mass distribution. The events are categorized in exclusive classes: five classes (four untagged and one VBF-tagged classes) for the 7 TeV data and six classes (four untagged and two VBF-tagged classes) for the 8 TeV data [3]. Categories with good resolution and larger signal-to-background ratio contribute the most to the sensitivity of the search.

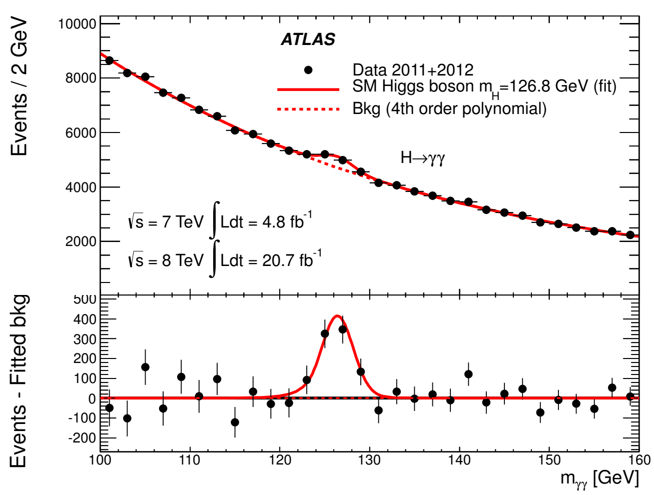

Figure 5: The combined invariant mass distribution of diphoton candidates observed by ATLAS [12].

The functional form of the background is determined by a fit to the full $m_{\gamma\gamma}$ distribution in each category. All categories are fitted simultaneously to determine the signal yield at a particular mass. With the full dataset, after combining all categories, the $m_{\gamma\gamma}$ distribution for the ATLAS experiment is shown in Figure 5 and for the CMS experiment is shown in Figure 6. In both experiments a significant excess of events with diphoton masses around 125 GeV was observed. All diphoton analyses give observed exclusion limits near $m_H = 125$ GeV that are much weaker than expected for the background. In conclusion, the H→$\gamma\gamma$ channel shows evidence of a new state that decays into two photons with an invariant mass of 125 GeV, with a CL well beyond 95% (confirming previous announcements).

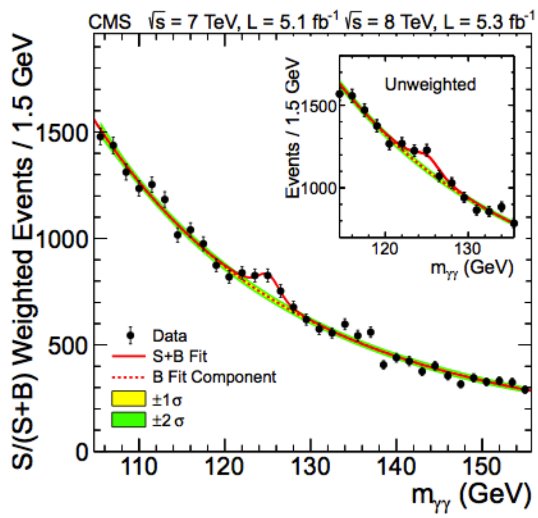

Figure 6: The combined CMS $m_{\gamma\gamma}$ distribution with each event weighted by the ratio of signal-to-background in each event category [3].

1.3.1.2 H→ZZ

The studies of the decay channel H→ ZZ → 4ℓ, are crucial for the determination of some properties like the spin and the parity of the discovered boson. The SM predicts that the Higgs boson should have zero spin and positive parity (represented as 0+). The spin zero of the state can be checked through the lack of polar (and azimuthal) angle correlations between the initial and final state particles in the combined process of production and decay $gg \to$ H $\to$ ZZ [13].

In the H $\to$ ZZ $\to 4\ell$ channel, in which $\ell = e,\mu$, a search is performed for a narrow mass peak over a small continuos background dominated by non-resonant ZZ* production from q$\bar{\text{q}}$ annihilation and gluon-gluon fusion processes. To enhance the sensitivity to VBF and VH production processes, the ATLAS and CMS experiments divide the 4ℓ final state into mutually exclusive categories. The SM Higgs boson signal was searched for in a simultaneous statistical analysis of six 2D distributions of the four-lepton invariant mass $m_{4\ell}$. The six distributions correspond to the three lepton final states: 4e, 4μ, 2e2μ for the 7 and 8 TeV datasets. Each subchannel is analyzed separately and the results are then combined.

The background sources include an irreducible four-lepton contribution from the direct ZZ production via q$\bar{\text{q}}$ and gluon-gluon process. Reducible contributions come from Z + b$\bar{\text{b}}$ and t$\bar{\text{t}}$ production, where the final states contain two isolated leptons and two b-quark jets producing secondary leptons. Additional backgrounds arise from Z+jets and WZ+jets events where the jets are misidentified as leptons.

The combined ATLAS $m_{4\ell}$ distribution is shown in Figure 7. The largest deviation from the SM background-only expectation is observed at m_H = 124.3 GeV, where the significance of the observed peak is 6.7 σ for the full 7 and 8 TeV datasets. The expected significance for the SM Higgs boson at that mass is 4.4 σ. The CMS experiment observes its largest excess at $m_H = 125.8$ GeV with an observed significance of 6.7 σ, as shown in Figure 8.

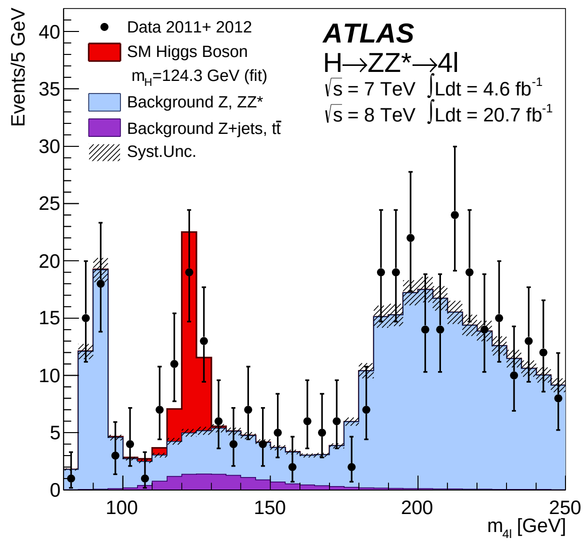

Figure 7: The distribution of the four-lepton invariant mass, $m_{4\ell}$, for the selected candidates for the data recovered by ATLAS during Run 1. The estimated background, as well as the expected SM Higgs boson signal are shown [12].

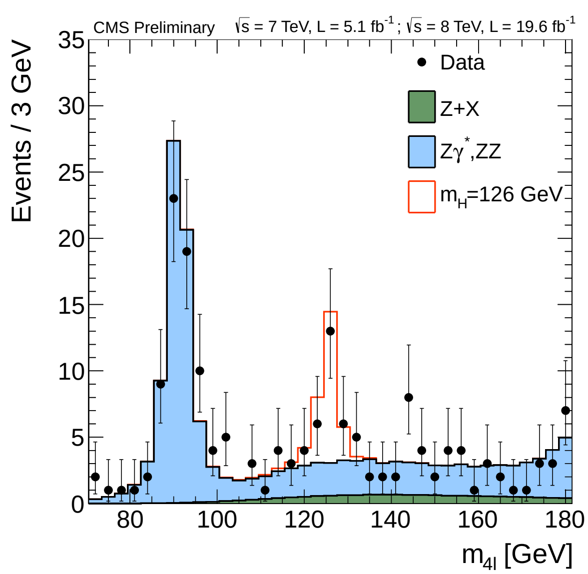

Figure 8: Distribution of the four-lepton reconstructed mass in the full mass range for the sum of 4e, 4μ and 2e2μ channels [14].

1.3.1.3 H→ W$^+$W$^-$

Another important decay channel is the H→W$^+$W$^-$. The final state for this decay is $2\ell 2\nu$, where ℓ is a charged lepton: electron or muon, and ν the corresponding neutrino. While the rate for the channel is large, the m_H resolution is quite poor ($\approx$ 20% m_H) due to the presence of neutrinos. For this channel it is not possible to reconstruct a Higgs boson mass peak. Instead, an excess of events above the expected backgrounds can be observed and used to establish the presence of a Higgs boson signal. The search is reduced to a counting experiment of the event yield in broad bins in m_H [11].

In the case of CMS this channel was analyzed by selecting events in which both W bosons decay leptonically, resulting in a signature with two isolated, oppositely charged, high p_T leptons (electrons or muons) and a large missing transverse energy ($\not{E}_T$) due to the undetected neutrinos. To improve the sensitivity of the signal, the events were separated according to the lepton flavor and jet multiplicity, into the samples: e^+e^-, $\mu^+\mu^-$, $e^\pm\mu^\mp$, 0−jet and 1−jet. Events with more than one jet were rejected [9].

The backgrounds contributing to this channel are numerous and depend on the kind of sample of selected events. For events with opposite flavor leptons and without the presence of high p_T jets, the dominant background comes from non-resonant WW production. Events with same flavor leptons have large Drell-Yan contamination. Additionally, $t\bar{t}$, Wt and W + jets are backgrounds for all the categories. Non-resonant WZ, ZZ and Wγ processes also contribute to the background at a sub-leading level.

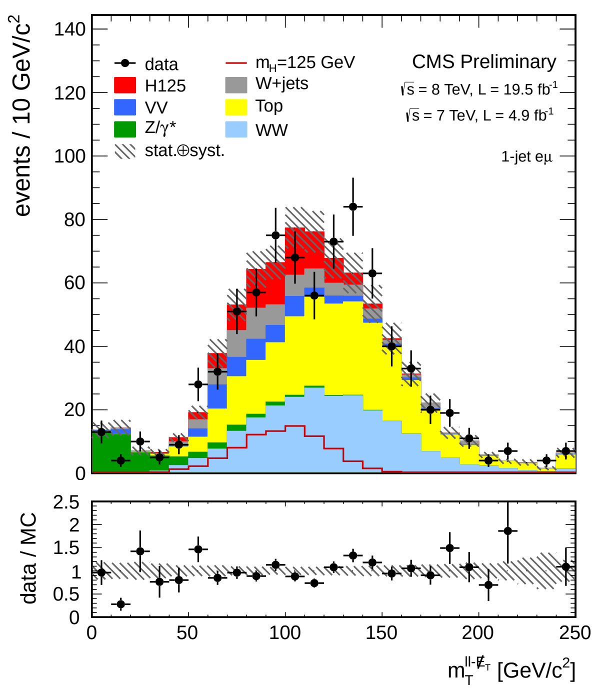

Figure 9: The m_T distribution for events, selected with a cut-based analysis, summed over all lepton flavors and with 1 reconstructed jets. The contributions of all SM background sources and a SM Higgs with $m_H = 125$ GeV are stacked together [15].

To suppress the top quark background, CMS use top-tagging technique. A requirement of large missing transverse energy ($\not{E}_T$) was used to reduce the Drell-Yan and multi-jet backgrounds. In the e^+e^-

and $\mu^+\mu^-$ categories events with $m_{\ell\ell}$ consistent with the Z mass are vetoed. All residual background rates, except for the small contributions from non-resonant WZ, ZZ and Wγ are evaluated from control samples obtained from the data.

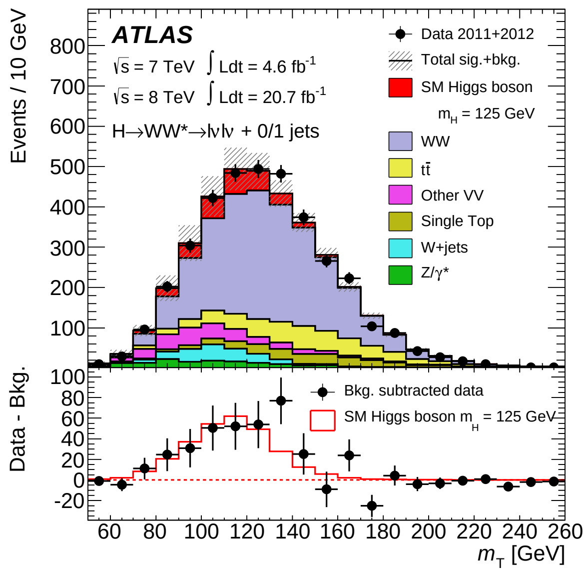

Figure 10: The transverse mass distributions for events passing the full selection of the H→ WW* $\rightarrow \ell\nu\ell\nu$ analysis summed over all lepton flavors for final states with number of jets less or equal than 1. The signal is stacked on top of the background [12].

The m_T distributions of selected events is shown in Figure 9 and Figure 10 for the CMS and ATLAS experiments respectively. CMS analysis using all lepton flavor combinations shows an excess with an observed significance of 4 σ for a 125 GeV Higgs boson. ATLAS fitted the m_T distributions and observed an excess with a significance of 3.8 σ for $m_H = 125.5$ GeV.

The CMS and ATLAS experiments have also performed dedicated searches for the associated Higgs boson production (VH) in this channel. The signal consists of three high p_T isolated leptons with missing transverse energy and low hadronic activity for the WH case or four leptons for the ZH case [7, 12, 15].

1.3.1.4 H→ $\tau^+\tau^-$

According to the spontaneous symmetry breaking of the SM, the Higgs not only couples to bosons but also to fermions (quarks or leptons). An important test of whether the discovered boson is indeed the SM Higgs is to observe its decay into fermions. For the H→ $\tau^+\tau^-$ channel, τ leptons decaying to electrons (τ_e), muons (τ_μ) and hadrons (τ_{had}) are considered.

In the reconstruction of the ditau system four final states were used: $e\tau_{had}$, $\mu\tau_{had}$, $e\mu$, $\mu\mu$, where the charge leptons come from $\tau \rightarrow e\nu\nu$ or $\tau \rightarrow \mu\nu\nu$ decays. Due to the presence of missing neutrinos, the invariant mass resolution is poor ($\approx$ 15%). As a result, the search looks for a broad excess over the expected background in the ditau invariant mass distribution. The major sources of background come from Drell-Yan, Z→ $\tau^+\tau^-$ and Z→ e^+e^-, W+jets, tt̄ and multijet production.

Events in all sub-channels are divided into categories based on the number and kinematic properties of additional energetic jets. The search sensitivity is generally higher for those categories that contain one or more additional jets. The best signal-to-background and search sensitivity category is the VBF one, consisting of a τ pair plus two energetic jets separated by a large pseudorapidity.

The ATLAS experiment observed a deviation from the background-only hypothesis with a local significance of 4.1 σ [16]. These results are shown in Figure 11.

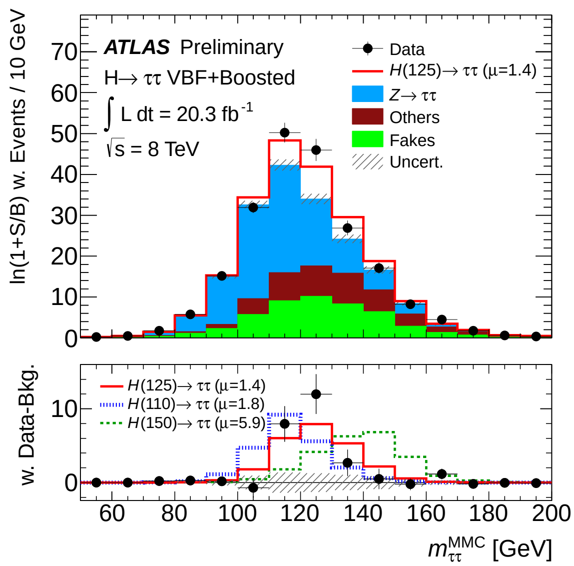

Figure 11: Distributions for $m_{\tau\tau}$ where events are weighted by $ln(1+S/B)$ for all channels [16].

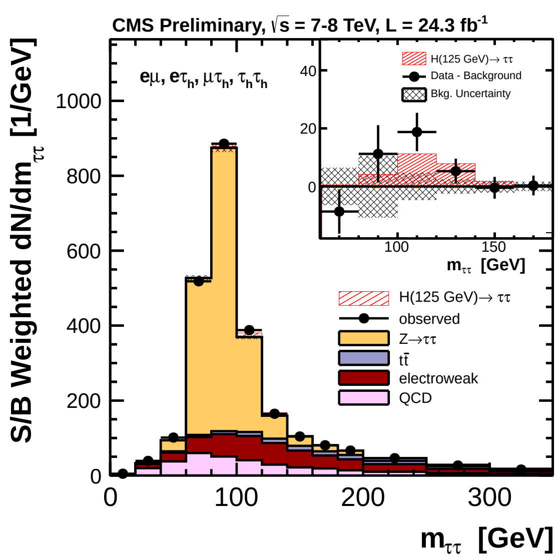

Figure 12: The combined observed and expected $m_{\tau\tau}$ distributions for all sub-channels combined [17].

Figure 12 shows the CMS $m_{\tau\tau}$ distributions combining all non-VH categories. The inset plot shows the difference between the observed data and the expected background distributions. In a first publication, the significance of the observed excess at $m_H = 125$ GeV was 2.85 σ. The updated result, shows an observed excess of 3.4 σ. Both ATLAS and CMS measurements provide substantial evidence of the coupling of the Higgs boson to leptons.

1.3.1.5 H $\rightarrow$ b$\bar{\text{b}}$

For a Higgs boson with $m_H \approx 125$ GeV, the branching fraction to b$\bar{\text{b}}$ is about 57%, it is the highest one. However, the presence of very large backgrounds makes the isolation of a Higgs boson signal quite difficult.

For this channel, the studies were focused on Higgs bosons produced in association with at least one b quark. The Higgs boson candidate mass is reconstructed from two b-tagged jets in the event. The crucial elements in this search are: b-jet tagging with high efficiency and low fake rate, accurate estimate of b-jet momentum, and good estimation of backgrounds from various signal control samples.The dominant background comes from QCD multijet production with two or more jets containing b hadrons and can not be fully reduced by kinematic selection. For this reason a method based on control data samples was applied by CMS [18].

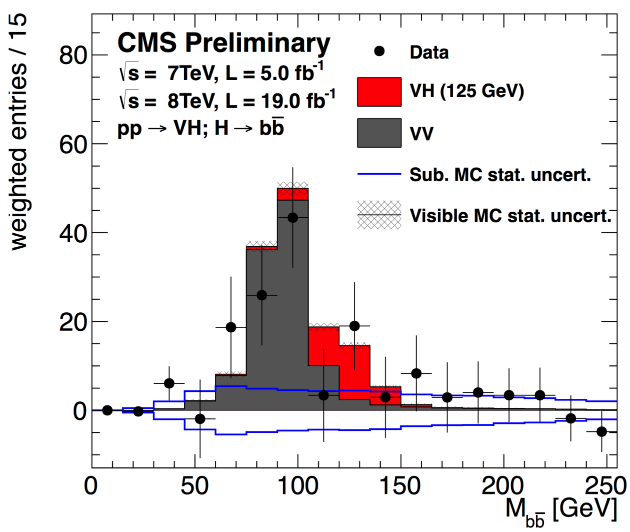

Figure 13: The $m_{b\bar{b}}$ distribution with all backgrounds, except dibosons, subtracted [18].

CMS uses multivariate analysis (MVA) classifiers based on kinematic, topological and quality properties of the tag b-jets. The algorithms were trained for different values of m_H to separate the Higgs boson signal in each category from backgrounds. The excess of events observed in bins with the largest signal-to-background ratios is consistent with the production of a 125 GeV SM Higgs boson, with a significance of 2.1 σ [18]. Figure 13 shows the $m_{b\bar{b}}$ distribution for all categories combined, weighted by the signal-to-background ratio for each of them.

ATLAS performed a cut-based analysis with the events selection divided into several p_T categories. The discriminating variable used is the invariant mass of the dijet. An excess of events with a significance of 1.4 σ was observed [19]. Figure 14 shows the $m_{b\bar{b}}$ distribution for 7 and 8 TeV after subtraction of all backgrounds, except for the diboson processes.

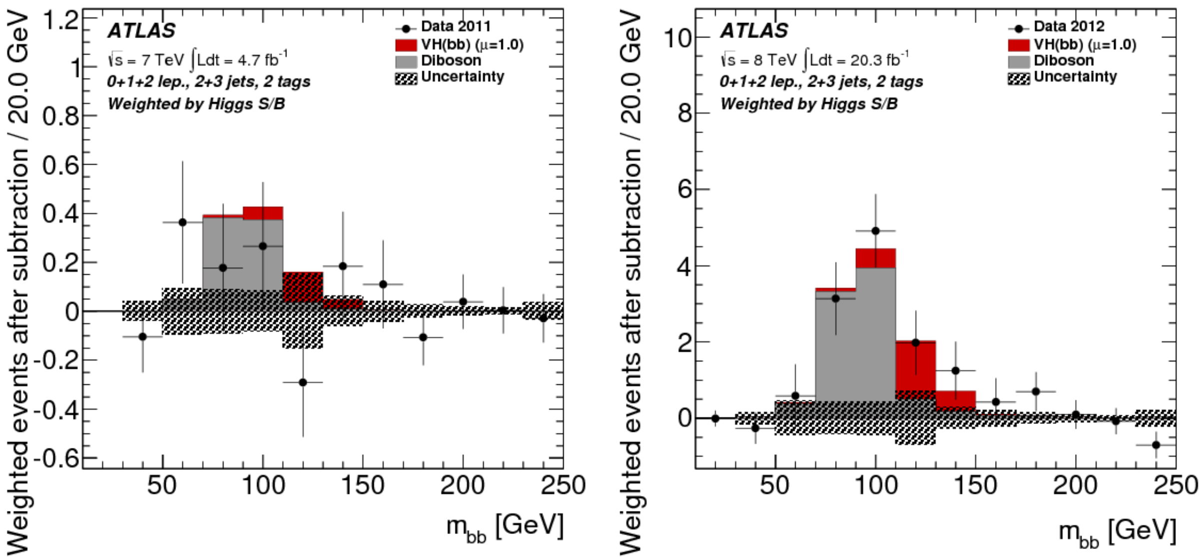

Figure 14: The distribution of $m_{b\bar{b}}$ after subtraction of all backgrounds except for the diboson processes, as obtained with the dijet-mass analysis for the (Left) 7 TeV and (Right) 8 TeV data [19].

1.3.2 OTHER DECAY CHANNELS

Additional to the discovery channels, the SM Higgs can decay into the denominated *rare* decay channels. These channels have small branching ratios, or they do not couple directly to the Higgs boson. They are: H→ $Z\gamma$, H→$\gamma\gamma^*$ and H→ $\mu^+\mu^-$.

1.3.2.1 H→ $Z\gamma$

Among the Higgs couplings, the H→ $Z\gamma$ is interesting in many ways, for instance, the interaction of the Higgs with the photon and the Z does not occur at tree-level because the photon is massless. The H→ $Z\gamma$ coupling is induced by loops of charged particles, dominated by the heaviest ones, like the W boson and the top quark as can be seen in Figure 15. In this regard this channel is similar to H→ $\gamma\gamma$.

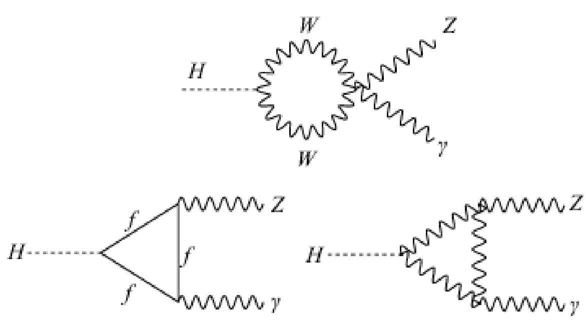

Figure 15: Feynman diagrams for the Higgs decay into a Z boson and a photon [20].

Due to the presence of the loop of massive charged particles, the channel is sensitive to physics beyond the standard model; if new heavy charged particles exist, it would result in a different branching fraction from the one predicted by the SM. Despite the small value of this branching fraction, which in the SM varies between 0.111% and 0.246%, the CMS and ATLAS experiments will be sensitive to this channel in the near future. Additionally, since there is no missing energy this channel has a good mass resolution, and the only relevant background is due to the proton-proton to diboson process.

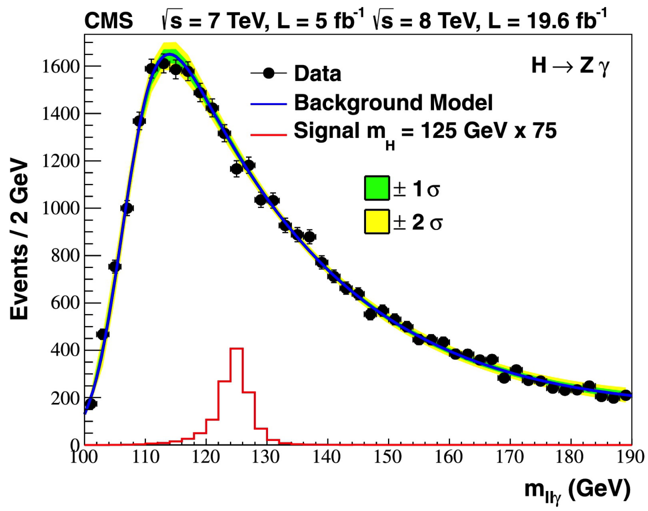

Figure 16: CMS results for the $m_{\ell\ell\gamma}$ spectrum in the electron and the muon channel for the 7 and 8 TeV data combined, without weighting by the expected signal-to-background ratio of the individual data samples. The histogram shows the expected signal due to a 125 GeV standard model Higgs boson, scaled by 75 [20].

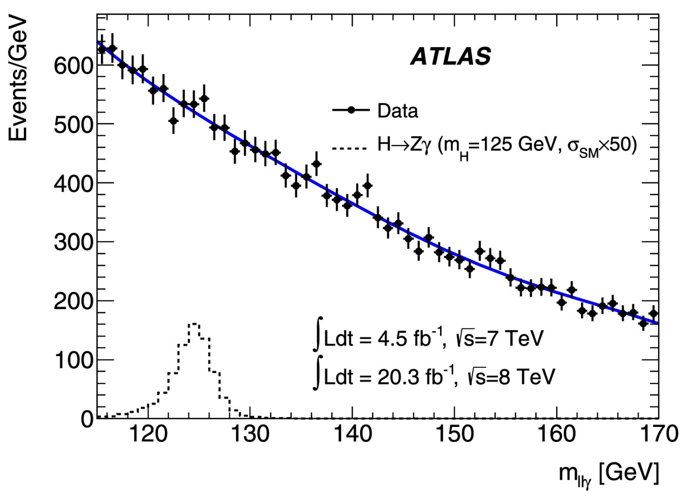

Figure 17: ATLAS results for the distribution of the reconstructed $m_{\ell\ell\gamma}$ invariant mass in data, after combining all the event categories used. The dashed histogram corresponds to the signal expectation for a Higgs boson mass of 125 GeV decaying to $Z\gamma$ at 50 times the SM-predicted rate [21].

The H→ $Z\gamma$ decay has been searched for by the ATLAS and CMS experiments using the full data sets collected at $\sqrt{s} = 7$ and 8 TeV [20, 21]. No significant excess over the expected background was observed at these energies and luminosities.

CMS performed an event classification according to the expected mass resolution and the signal-to-background ratio in order to improve the sensitivity of the search. Figure 16 shows the $m_{\ell\ell\gamma}$ invariant mass for the combination of the electron and the muon channels for the 7 and 8 TeV data. An observed limit of 9.5 times the SM prediction for a 125 GeV Higgs Boson was reported.

ATLAS instead performed an event selection into four categories, based on the pp centre-of-mass energy and the lepton flavor. Figure 17 shows the reconstructed invariant mass after the combination of all the categories. An observed limit of 11 times the SM prediction was reported.

1.3.2.2 H→$\gamma\gamma^*$

This decay channel could be seen as a QED radiative correction of the H→ $\gamma\gamma$ decay in which one of the γ converts to a dilepton. Due to the different nature of the detection of a photon as compared to the detection of charged leptons, the H→$\gamma\gamma^*$ or Dalitz decay, represents a different challenge from the experimental point of view and, is actually closer to the H→ γZ→ $\gamma\ell^+\ell^-$. The name "Dalitz decay" has historical origins and comes from one of the decay channels of the neutral pion: $\pi^0 \to \gamma e^- e^+$. This decay is actually the result of the dominant decay $\pi^0 \to \gamma\gamma$, in which one of the γ is virtual (γ^*) and converts into an electron-positron pair due to a QED radiative process. A similar situation can happen to the H→ $\gamma\gamma$ decay channel, where one of the γ is virtual and then converts to a dilepton.

After the searches for H→Zγ were performed by the ATLAS and CMS experiments, CMS searched for a Higgs boson Dalitz decay as an extension of the previous channel [22]. A search was performed at 7 and 8 TeV in the mass range between 120 and 150 GeV. In order to differentiate the Dalitz decay contribution from H→Zγ, CMS required a dilepton mass smaller than 20 GeV. The complex nature of the tau decay makes the τ-dilepton case a completely different experimental problem, therefore the work was focused on the electron and muon dileptons.

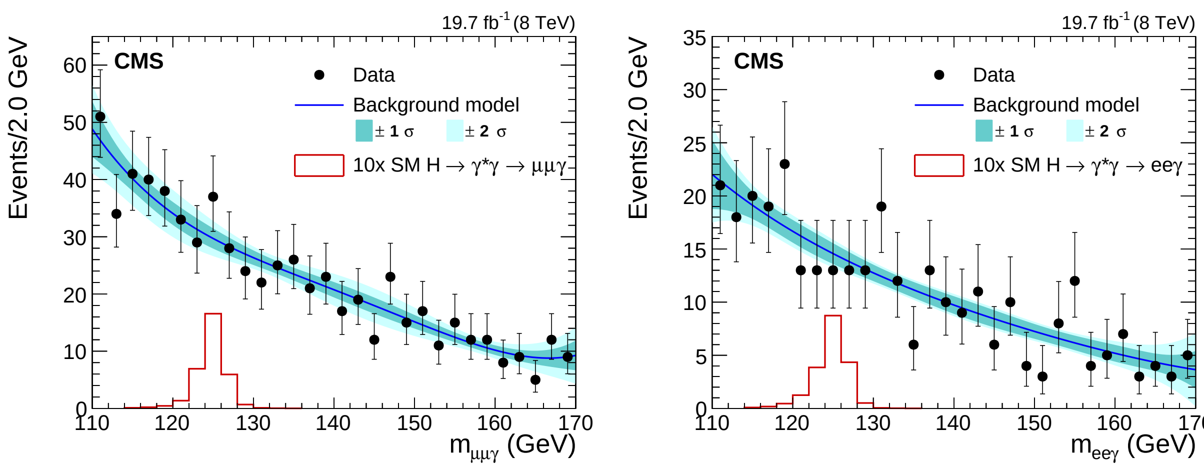

Figure 18: The $m_{\mu\mu\gamma}$ (Left) and $m_{ee\gamma}$ (Right) spectra for 8 TeV data, together with the result of a background-only fit to the data. The expected contribution from the SM Higgs boson signal with $m_H = 125$ GeV, scaled up by a factor of 10, is shown as a histogram [22].

Figure 18 shows the $m_{\mu\mu\gamma}$ (Left) and $m_{ee\gamma}$ (Right) spectra for the 8 TeV data. No significant excess above background was reported in the full mass range. The observed limit for m_H =125 GeV is 6.7 times the SM prediction.

1.4 H→ $\mu^+\mu^-$ DECAY CHANNEL

The H→ $\mu^+\mu^-$ is at this stage the only decay channel that can be measured at LHC, where the Higgs boson couples to second-generation fermions. The branching fraction for this channel in the SM has a value of 2.2×10^{-4} for a 125 GeV Higgs boson, and is among the smallest accessible at the LHC, being about ten times smaller than the one for H→$\gamma\gamma$ [23, 24].

The H→ $\mu^+\mu^-$ cross section measurement is important because it can provide three unique probes of the SM Higgs boson properties:

- It is a way to measure the Higgs-muon Yukawa coupling strength[25].
- The ratio of the H→ $\mu^+\mu^-$ to H→ $\tau\tau$ provides restrictions on the Higgs couplings to fermions for distinct generations, and on the couplings of leptons with different masses.
- It gives some hints about possible lepton flavor violations in the H→ $\mu\tau$ channel.

From the experimental point of view this decay has a clean final-state signature but a very small yield. CMS and ATLAS subdivide the events using the pseudorapidity of the muons, based on their good muon momentum resolution. This is a very challenging measurement due to the small branching ratio and high SM backgrounds. The dominant and irreducible background comes from the $Z/\gamma^* \to \mu^+\mu^-$ whose rate is several orders of magnitude larger than the SM Higgs signal. Nevertheless, the advantage of a narrow mass peak improves the search sensitivity.

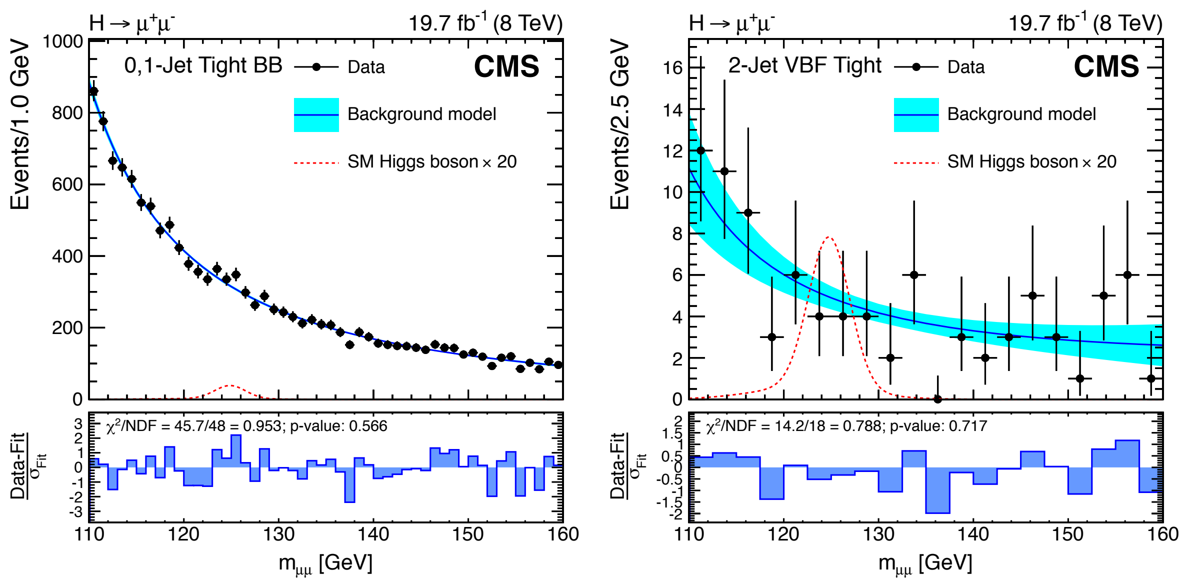

Figure 19: The dimuon invariant mass at 8 TeV and the background model are shown for the 0/1 Jet Tight category when both muons are reconstructed in the Barrel (Left) and the 2-Jet VBF Tight category (Right). The dotted line illustrates the expected SM Higgs boson signal scaled up a factor of 20. The lower histograms show the residual for each bin (Data-Fit) normalized by the Poisson statistical uncertainty of the background model [23].

The SM H→ $\mu^+\mu^-$ search is performed on events with two oppositely charged muons, in which at least one of them has a p_T larger than 25 GeV. In order to suppress the Drell-Yan background, ATLAS requires a dimuon p_T larger than 15 GeV, while CMS separates the dimuon into p_T categories. CMS also categorizes events using information from the number and the topology of additional jets in the event. For both searches no significant excesses have been found yet.

CMS has set an observed limit on the production rate at 7.4 times the SM value. Figure 19 shows the dimuon invariant mass at 8 TeV for the 0/1 Jet Tight category (Left) and the 2-Jet VBF Tight category (Right). The dotted line illustrates the expected SM Higgs boson signal, scaled by a factor of 20, for m_H =125 GeV.

In the case of ATLAS an observed limit on the signal strength was set to 7.0 times the SM value. Figure 20 shows the dimuon invariant mass distribution for the ATLAS selection.

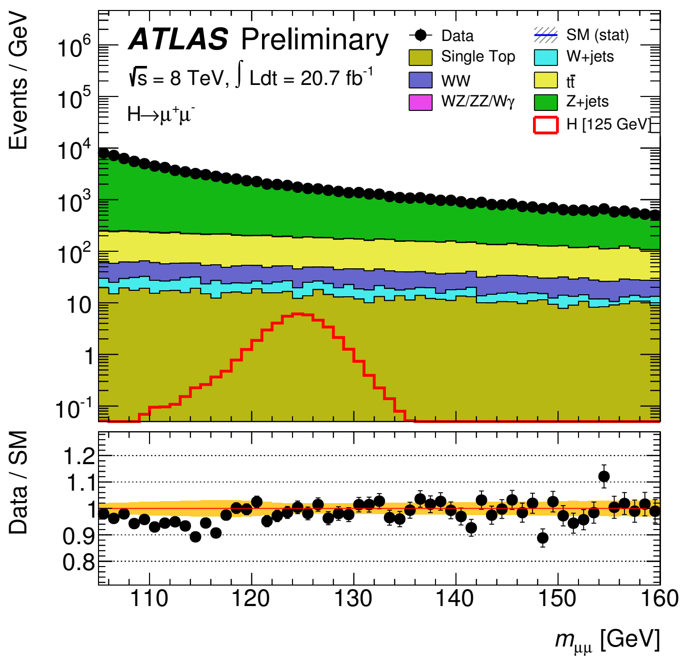

Figure 20: Dimuon invariant mass distribution for the selected events using logarithmic scale [24].

The subject of my PhD dissertation is this channel. It will be studied using the data collected in proton-proton collisions at 13 TeV by the CMS detector. A detailed discussion about the search performed with this channel is presented from chapter 4 on.

1.5 HIGGS BOSONS IN MODELS BEYOND THE STANDARD MODEL

The data available on the Higgs boson decay has already imposed constrains, not only on the properties of the SM version of the state, but on any possible composed nature of this particle resulting from other models. However, the results presented by CMS and ATLAS at the 2017 Moriond Conference do not exclude several alternatives of physics BSM, like many SUSY-inspired models. In minimal SUSY models the simplest version of the Higgs sector is composed by five states: three neutral ones and two charged ones. In some parametrizations of the Minimal Standard SUSY Model (MSSM) the lightest neutral Higgs will have practically the same properties of the SM Higgs, making it very difficult to distinguish them.

Since the couplings of the Higgs to all the other SM particles are fixed once its mass is known, the branching ratios for the different decay modes are also known. Any significant departure from the expected values could give us a hint of the emergence of physics BSM. The coupling of the Higgs to other particles is proportional to the mass of the states (and data has already shown that [4]), therefore the contribution of BSM massive states to the intermediate loops in the H→ $\gamma\gamma$, γZ, $\gamma\gamma^*$ channels should have a significant contribution. The statistics available for these processes at 13 TeV may show significant deviations from SM predictions, and therefore evidence of physics BSM.

In summary, the experimental results for the 125 GeV Higgs boson discovered by the CMS and ATLAS experiments have not shown significant deviations from SM. However, the new statistics available at 13 TeV for some of the channels mentioned previously may show some hints of new physics, including the $H \to \mu^+\mu^-$ channel.

2

THE CMS EXPERIMENT

The data used to make the studies for this dissertation have been collected by the Compact Muon Solenoid (CMS) Experiment. The main features of the Large Hadron Collider (LHC) and CMS Detector will be presented in this chapter. The chapter starts with the description of the LHC and the CMS Experiment. The different subdetectors and their functionality are discussed. Finally a brief description of the Data Acquisition and the Trigger systems is presented.

2.1 THE LARGE HADRON COLLIDER

The CMS Experiment is one of the four experiments at the LHC, the particle accelerator located at CERN in the border between France and Switzerland. The LHC is a proton-proton collider with energies that could go up to 14 TeV. It is installed inside a tunnel, original built for the LEP accelerator, with a circunference of 26.7 km. The LHC can also collide heavy-ions (Lead-Lead).

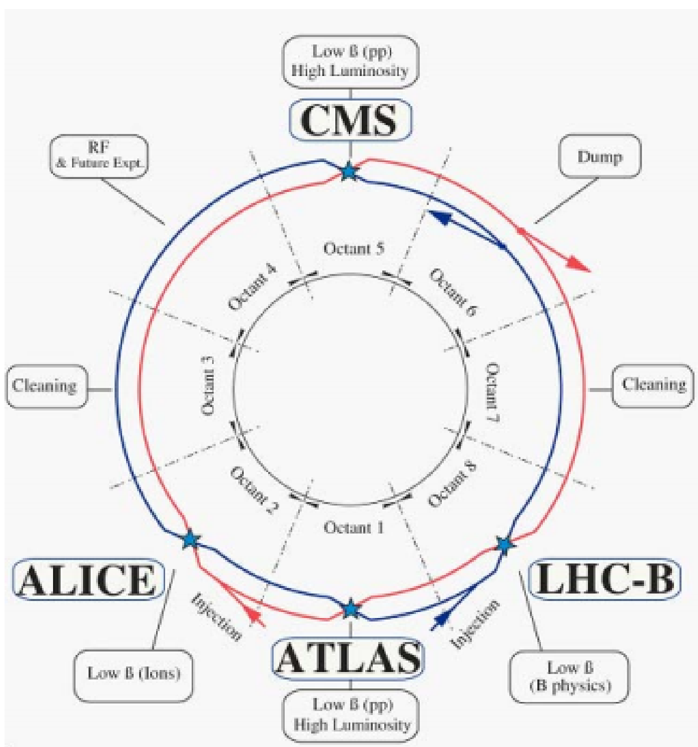

Figure 21: Schematic layout of the LHC and the four experiments at the interaction points [26].

Each proton beam of the LHC was designed to reach an energy up to 7 TeV, equivalent to a center-of-mass energy of 14 TeV for every collision. Currently it collides beams of protons a center-of-mass energy of 13 TeV. A set of older CERN accelerators, pre-accelerate the proton bunches up to 450 GeV per proton. Two beams, each with 2808 bunches, are injected to twin synchrotrons in opposite directions and are accelerated up to 7 TeV per proton. The time separation between bunches is 25 ns. Each bunch is made out of approximately 1.15×10^{11} particles.

Bunches traveling in opposite directions cross each other in four points along the circumference of the LHC every 25 ns, defining what is called "Bunch-Crossing" (BX). A BX is the fundamental time unit at the LHC and its experiments. In each of these interaction points there is a large particle detector. In Point 1 ATLAS, in Point 2 ALICE, in Point 5 CMS and in Point 8 LHCb (See Figure 21). In addition to these four experiments, there are two small detectors designed to measure proton-proton scattering at very small angles, in order to measure the total proton-proton cross section, among other things. These are TOTEM, which is installed around Point 5 and LHCf which is installed around Point 1.

ATLAS and CMS are both multi-purpose experiments whose main physics goals are the search for the Higgs boson, and exploration of any possible new physics in the energy ranges provided by the LHC. The aim of the LHCb detector is to study the charge-parity symmetry violation (CP violation), and to explore the nature of the matter-antimatter asymmetry in the universe. ALICE detector is specialized in studying the physics of the quark-gluon plasma [27].

The instantaneous luminosity given by the accelerator is different at each of these experiments, it has a maximum at CMS and ATLAS with a designed optimal value of $\mathcal{L} = 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$. This value has not been reached yet. The total integrated luminosities delivered to CMS during the different periods or LHC Run 1 (2010, 2011 and 2012) are presented in Figure 22.

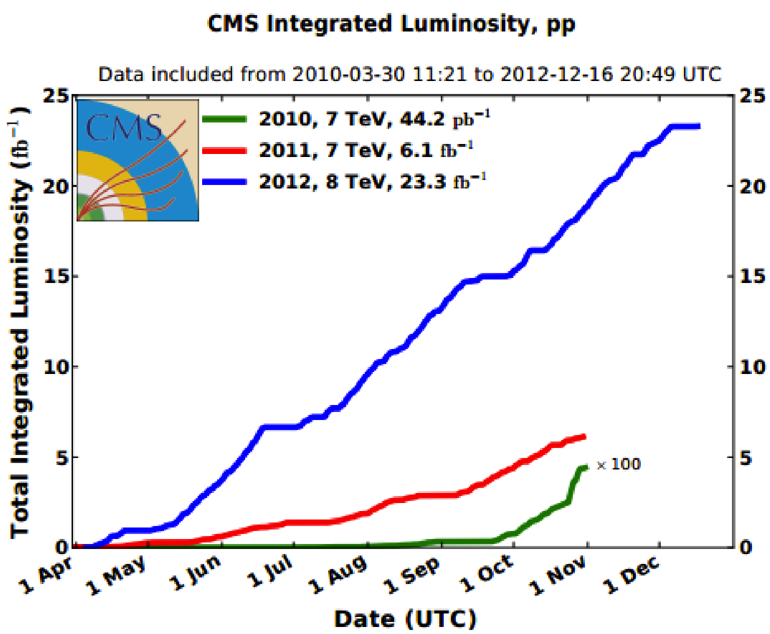

Figure 22: Integrated Luminosity delivered by the LHC to the CMS Experiment during the first Run (2010, 2011, 2012) [28].

To reach the energy and luminosity levels needed for the physics program, the LHC uses superconductive magnets. These magnets work at 1.9 K, which is the lowest temperature of any superconducting accelerator to date. The main dipoles of the LHC can produce a magnetic field of 8.4 T at a current of 11700 A approximately. The magnets have two side-by-side apertures, one for each of the counter rotating proton beams. Each one of these magnets has a length of 14.3 m. The entire accelerator has 1232 of these dipoles, 1104 in the arc and 128 in the Dispersion Suppression (DS) regions.

In both rings, each sector is equipped with two pairs of quadrupole magnets. Each short straight section not equipped with quadrupoles, contains an octupole magnet. Each ring contains 168 octuple magnets. These magnets are subdivided into focusing and defocusing magnets.

In addition to the magnets, the LHC has 16 Radio Frequency (RF) cavities. These cavities are located in cylindrical refrigerators, called cryomodules, two modules per beam. The cryomodules keep the RF cavities working in a superconducting state.

2.2 THE CMS DETECTOR

The CMS detector is a multi-purpose machine. It is installed inside a cavern about 100 meters underground in Point 5 at the LHC. This Point is located at Cessy, France on the north side of the LHC ring, as can be seen in Figure 21.

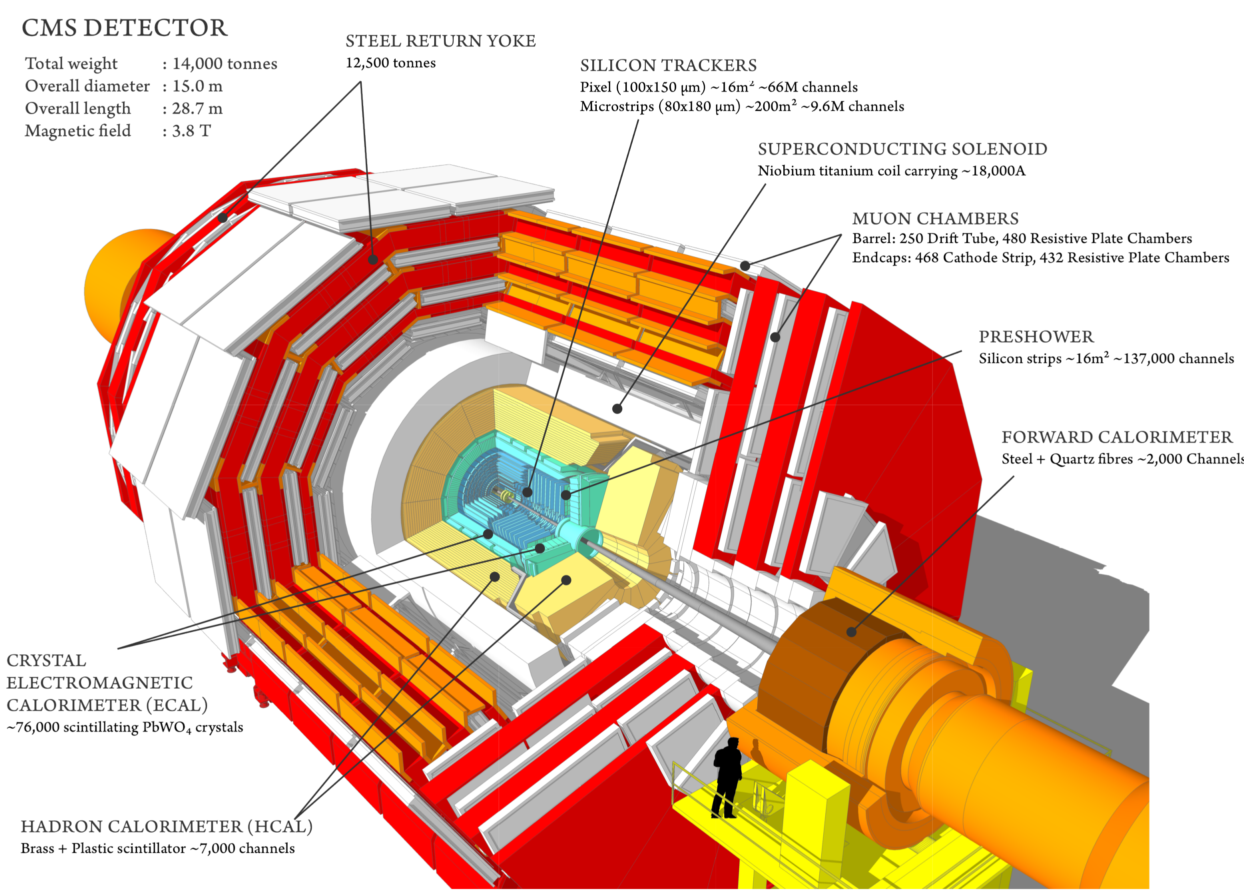

Figure 23: Sectional view of the CMS detector [29].

The detector requirements for CMS to meet its scientific goals are:

- Good muon identification and momentum resolution.
- Good charged-particle momentum resolution and reconstruction efficiency.
- Good electromagnetic energy resolution.
- Good missing-transverse-energy and dijet mass resolution.

The main distinguishing features of CMS are: a large high-field solenoid, a full-silicon-based inner tracking system, and a homogeneous scintillating-crystals-based electromagnetic calorimeter[30].

The CMS has a cylindrical shape with a length of 21.6 m and a diameter of 14.6 m. Its mechanical structure is defined by a large iron yoke that is used to collect the magnetic field that comes out of the central solenoid. This yoke gives CMS its cylindrical shape. The yoke is segmented into 5 wheels that define the central Barrel and three rings in each side of the Barrel that define the two end-caps. Inside the yoke there is a central region delimited by a large superconducting solenoid. Inside the solenoid there is a silicon tracker (TRK), an electromagnetic calorimeter (ECAL) and a hadron calorimeter (HCAL). Outside of the magnet, and inserted in the yoke structure, are the muon detectors composed of layers of Drift Tubes (DT) and Resistive Plate Chambers (RPC) in the Barrel and of Cathode Strip Chambers (CSC) and RPCs in the Endcaps. The geometrical structure of the CMS Detector systems can be seen in Figure 23.

Due to its large iron yoke the CMS Detector is very heavy (14 kt). It is also compact in comparison to the ATLAS detector since it does not require an external magnetic field for the muon system. This is because the iron of the massive yoke compactifies the magnetic field coming out of the solenoid, making it intense enough to deflect the muons. It is optimized to detect and trigger on muons. The central magnet is a very large superconductive solenoid, 13 m long and 6 m in diameter, that provides an inner magnetic field up to 3.8 T, and an outer magnetic field of the order of 2 T (for the muon tracking system).

2.2.1 THE TRACKER SYSTEM

A high resolution silicon tracker is located in the inner most part of the CMS Detector, surrounding the interaction point. This tracker has also a cylindrical shape with the length of 5.8 m and a diameter of 2.5 m. The tracker is composed of a pixel detector with three layers and ten silicon strip detector layers in the Barrel. In the Endcaps there are two pixel and twelve strip layers [30].

Due to its proximity to the interaction point the tracker is very important for detecting short-lived particles that would not leave a direct signal in the detectors, but would be reconstructed from their decay products. The resolution with which the tracks of the decay products are measured will determine if the original state can or cannot be reconstructed and the accuracy with which it can be done. This is the crucial role of the silicon tracker, because most of the interesting states are very short-lived and their reconstruction is very important. (For example the identification of the b-quarks out of b-jets).

During the Long Shutdown 1 (LS1) [1], a new central beam pipe with a smaller outer diameter was installed in Point 5, leaving an additional radial space for an inner layer of the Pixel Tracker. This layer

1 The LHC shut down for two year of planned maintenance and upgrades. The work prepared the experiments and the accelerator to take data at centre-of-mass energy of 13 TeV. The Long Shutdown 1 started in 2013

was installed during the technical stop in 2016-2017 [31]. In the Long Shutdown 2 (LS2)[2], the whole system will have four layers in the Barrel and three disks in the End-caps. With these modifications the tracker will improve significantly vertex reconstruction (primary and secondary), even for the high pile-up values expected for Run 3. The tracker cooling system was upgraded for Run 2 and now is able to reach a temperature of 253 K, necessary for the optimal operation of the silicon detectors.

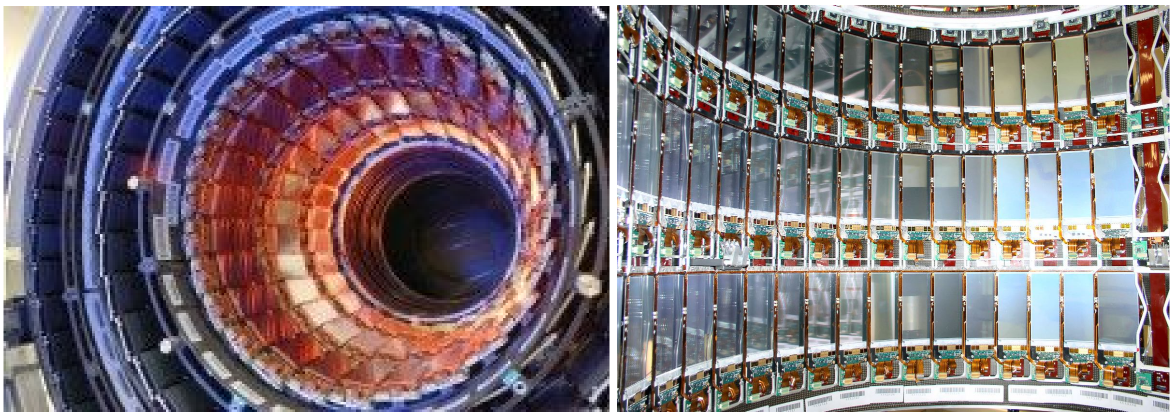

Figure 24: Left: Picture of the CMS Tracker showing silicon strips detectors in the Barrel module. Right: Picture of the inside of the Tracker Inner Barrel Layer 3. Each panel in both pictures is compose by several silicon strips [32].

2.2.2 ELECTROMAGNETIC CALORIMETER (ECAL)

The second sub-detector encountered by particles coming out of the interaction point is the Electromagnetic Calorimeter (ECAL). The ECAL absorbs and measures the energies of electrons, photons and electromagnetic showers.

In the design of the calorimeter, emphasis was made on good resolution for photon and electron energy, position and direction measurements, and wide geometric coverage. The measurements made by the ECAL play a crucial role in the functioning of the CMS trigger system.

The ECAL is made of 61200 lead tungstate crystals ($PbWO_4$) (see Figure 25) mounted in the Barrel, and 7324 crystals in each of the two End-caps [30]. The use of high density crystals allows for a fast response calorimeter, with a fine granularity, and resistance to the high radiation dose. The lead tungstanate crystals have a very good energy resolution and the time resolution is of the same order of the bunch crossing time of the LHC.

In the Barrel, avalanche photodiodes (APDs) are used as photodetectors. A pair of those are mounted on the back of each crystal. They can operate in a 4 T environment. They have a high quantum efficiency over the wavelength of emission, stable gain, small enough capacitance, small leakage current and they are radiation hard [27]. In the Endcaps, the photodetectors are vacuum phototriodes (VPTs). The VPTs are photomultipliers having a single gain stage. One VPT is attached to the back of each crystal.

2 The Long Shutdown 2 is scheduled to start in July 2018. As the LS1 will be a year and a half of planned maintenance and implement important modifications in order to move on with the High Luminosity LHC (HL-LHC)

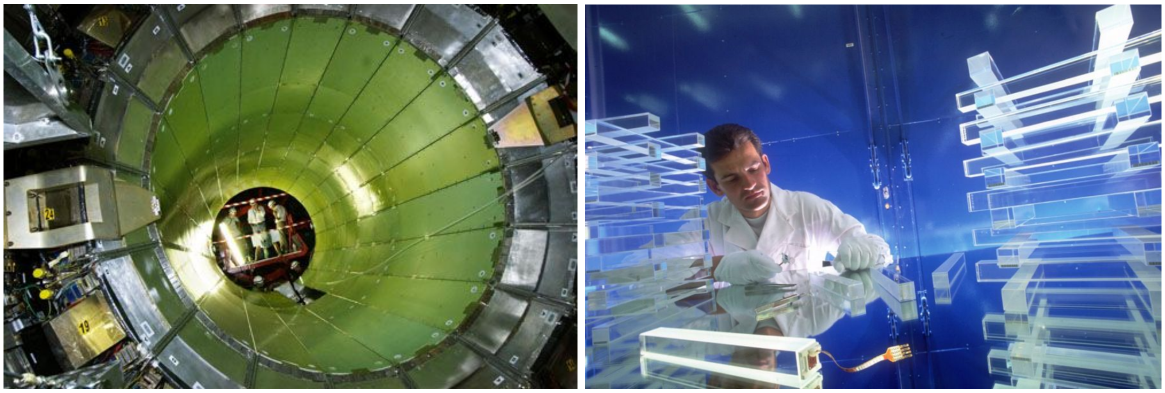

Figure 25: Left: CMS ECAL during construction. Right: Lead tungstate crystals for the CMS ECAL [33].

Additional detectors, called preshower detectors, are placed in front of the two Endcaps. They are used to increase the resolution in the forward regions with the purpose of identifying neutral pions decaying into two photons in that angular region. Each preshower detector has two layers: a layer of lead radiators initiate electromagnetic showers from incoming photons and electrons, while silicon strip sensors measure the deposited energy and the transverse shower profiles. The total thickness of a preshower detector is 20 cm.

The ECAL read-out system has to acquire the small signals produce by the photodetectors, with high speed and precision. In terms of data acquisition and triggering, the ECAL presents its data in the form of "towers" . A trigger tower is a 0.087×0.087 $\eta - \phi$ region. For the ECAL this corresponds to a 5×5 crystal grid.

2.2.3 HADRONIC CALORIMETER (HCAL)

The third subdetector coming out the interaction point is the Hadronic Calorimeter (HCAL), designed to measure the energy of hadrons produced as a result of the collisions [27]. The HCAL surrounds the ECAL and acts in conjunction with it to measure the energy and other kinematical information of jets. It also provides hermetic coverage for missing transverse energy, an indirect measurement of the presence of non-interacting neutral particles, such a neutrinos or other unknown particles. The HCAL plays an important role in the CMS Trigger, in which the missing energy is one of the main variables when looking for SUSY and other kinds of physics signatures beyond SM.

The HCAL is composed of plastic scintillator tiles with embedded wavelength-shifting fiber readouts (see Figure 26). The tiles are interleaved with stainless steel and copper layers [30]. The metal layers absorb hadrons, creating showers and the plastic scintillator regions detect the particles coming from the particles shower that cross them.

The HCAL Barrel (HB) is divided into two half-Barrel sections and is surrounded by the superconducting solenoid. This design constraints the total amount of material used to develop and absorb the hadronic showers. The HCAL Endcaps (HE) cover a big portion of the rapidity range. At least, 34 % of the particles produced in the final state cross that region. Since the HE is inserted into the ends of the solenoid, it is made out of non-magnetic material (brass, that was recycled from Russian navy shells). The HCAL extends outside the solenoid, into what is called the Outer Calorimeter (HO). The HO uses the solenoid coil to identify late starting showers and to measure the energy deposited beyond HB.

The HO design was constrained by the muon system geometry.

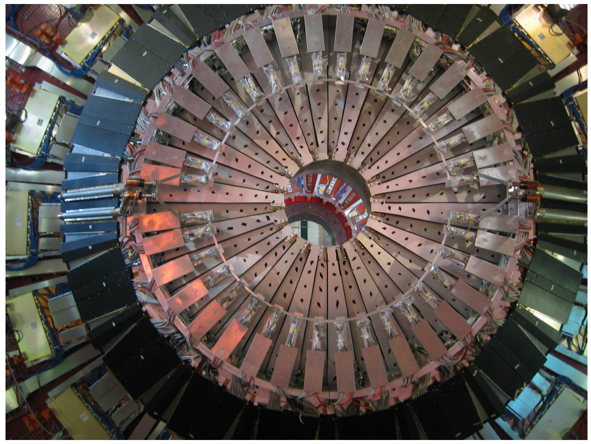

Figure 26: View of a CMS HCAL Endcap.

CMS has two separate Forward Calorimeters (HF) located at 10.9 m from the interaction point in each side. These calorimeters use quartz fibers as active material, embedded in copper. The HF distinguishes showers generated by electrons and photons from those generated by hadrons.

In order to improve the gain and the granularity of the calorimeter, and to increase particle identification and background rejection capabilities, CMS is planning to replace the photodetector technology in the HCAL with multi-anode photo multipliers (PMs) and Silicon Photomultipliers (SiPMs). During LS1, the photo-multipliers were replaced in the HF and HO. The installation in the HB and HE will continue in LS2 [34].

2.2.4 MUON SYSTEM

The Muon detector is a powerful tool to recognize signatures of interesting processes, specially for analyses searching for topologies involving high p_T particles. As the middle name of the experiment indicates, the muon detection plays an important role at CMS. The central concept of the CMS is the ability to trigger and to reconstruct muons at the highest luminosities [30].

The muon system is located outside the solenoid magnet. Due to the shape of the solenoid, the system has a cylindrical Barrel section and two planar Endcap regions. It consists of about 25000 m^2 of detection planes.

The CMS muon system has a redundant and complementary trigger capability over nearly the entire rapidity range. The muon system was designed to respond very fast in order to provide the trigger and to reconstruct the tracks with high momentum resolution. The three principal functions of the muon system are: muon identification, momentum measurement, and triggering. Good momentum resolution and trigger capability are enabled by the high-field solenoidal magnet and its flux-return yoke. The yoke serves as a hadron absorber cleaning up the muon signal and providing the necessary magnetic field to curve the track outside the solenoid [36].

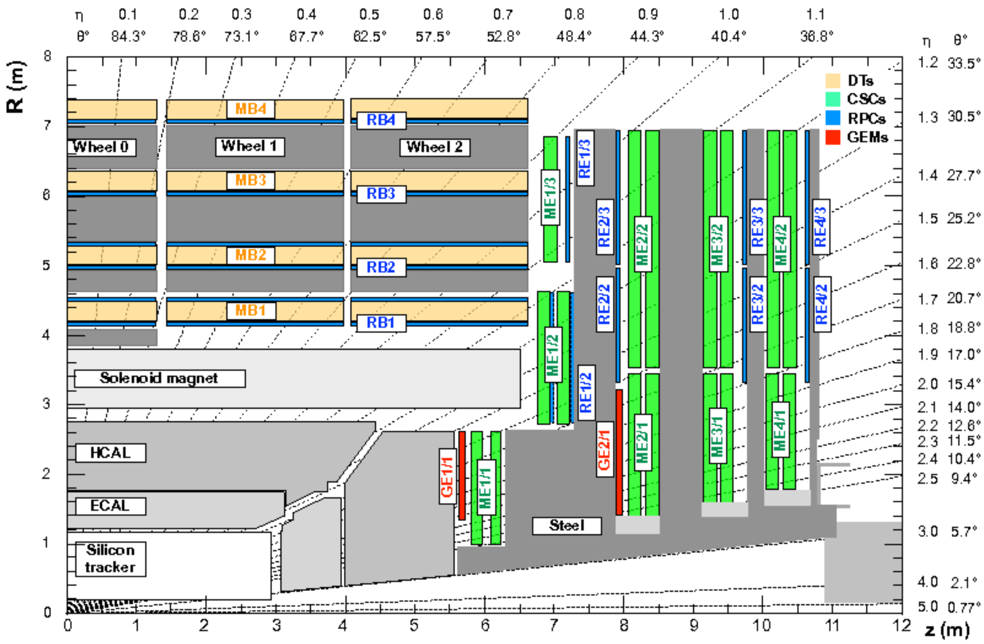

Figure 27: Barrel and Endcap muon chamber in CMS [35].

The muon system uses three different technologies to detect and measure the muons: Drift Tubes (DT), Resistive Plate Chambers (RPC) and Cathode Strip Chambers (CSC). The DTs are located in the Barrel region, the CSCs are located in the Endcaps, and the RPCs are located in both, Barrel and Endcaps. These technologies will be presented in the next subsections.

The Muon System will be improved during the LS2. Gas Electron Multiplier detectors(GEMs) will be installed in the Endcaps in order to improve the muon identification efficiency and momentum measurement resolution in these regions. In addition to their high granularity, the advantage that the GEMs have over the other technologies is their high time resolution, for these reasons GEMs will be used for triggering.[35].

2.2.4.1 DRIFT TUBE SYSTEM (DT)

In the Barrel, the muon detector consists of four stations that form concentric cylindrical shells around the beam line. The three inner shells have 60 DTs each while the outer shell has 70.

The DTs are aluminum cells with a thickness of a few centimeters. These cells are filled with a gas mixture of $Ar - CO_2$ and have an anode wire in the center (50 μm diameter gold-plated stainless-steel wire). The anode collects electrons resulting from charged particles passing through the gas. The drift cells operate at a gas gain of 10^5, allowing them to work within an efficiency plateau [36].

These cells are assembled in units called chambers, with four layers of cells each. These layers have different orientation in order to provide 2D information, as well as timing information. From this signal plus drift time information, two coordinates can be measured for every detected particle: the position in the wire where the electrons hit, and the distance that those electrons traveled before hitting the wire.

A chamber is assembled by gluing three super layers (SL). Each SL is made of five aluminum sheets. An SL has an independent gas and electronics enclosure.

There are four levels of DT chambers moving radially out of the interaction point. These DT chambers are complemented by RPC detectors, in what is a called "muon station", therefore there are four levels of muon station in the Barrel. These stations are segmented in the azimuthal angle into twelve sectors (forming a dodecagonal structure). Along the z-axis there are five wheels. Therefore, it has four levels of muon stations radially, twelve sectors azimuthally and five wheels along the z-axis.

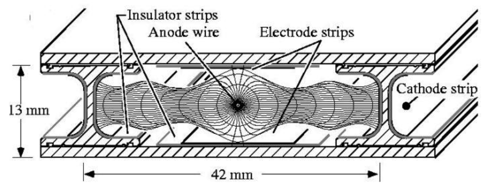

Figure 28: The drift tube cell of CMS [37].

2.2.4.2 CATHODE STRIP CHAMBERS (CSC)

The CSCs are designed to trigger on and to reconstruct the tracks of muons in the Endcaps. They can operate at high rates and in large and non uniform magnetic fields. The CSCs do not require precise gas, temperature or pressure control.

This type of detector consists of positively charged anode wires and negatively charged cathode strips made out of copper, orientated 90 degrees with respect to the wires. This grid is located inside a chamber filled with ionizing gas. When a muon passes through a CSC chamber, it knocks out electrons from the atoms of the gas. These electrons drift until they reach the anode wires, leaving behind positive ions that move to the copper strips. This process creates two currents that could be registered, giving the CSC a very good spatial resolution in two dimensions [30].

Figure 29: Cathode Strip Chambers mounted on a CMS Endcap [37].

Each Endcap region is composed of four disks. The inner most disk has three rings of chambers, while the other three disks are composed by two rings of chambers. Each ring is segmented in 18 azimuthal sections, with the exception of the outer ring which has 36 chambers. Each section is composed by a CSC. This arrangement allows for accurate identification of muon tracks, and matching with tracks reported by the inner tracker at large η [36].

2.2.4.3 RESISTIVE PLATE CHAMBERS (RPC)

CMS has added planes of RPCs in both, the Barrel and Endcaps, to provide complementary trigger information. These are gaseous parallel-plate cameras that combine adequate spatial resolution with excellent time resolution ($\sim 1ns$). The temporal resolution and fast response of RPCs make them very important in the trigger system, because they identify the correct bunch crossing associated with the detected track.

An RPC is constructed of two parallel plates, these plates are made of phenolic resin, with good surface flatness and high bulk resistivity. The plate separation is of a few millimeters, creating a gap in which the gas is located. This gas is ionized by the particles when they pass through it; the ionization generates an electron avalanche that will induce a signal in metal strips located on the outside of the plates [38].

A total of six layers (coaxial cylindric shells, all around the beam axis) of RPCs are embedded in the Barrel iron yoke. Each layer of RPCs forms a dodecagon with full 2π azimuth coverage. There are four levels of muon stations conforming concentric dodecagons. In the first two levels of muon stations there are two layers of RPCs detectors, and in the last two levels of the muon stations there is one. (See figure 30 (right)).

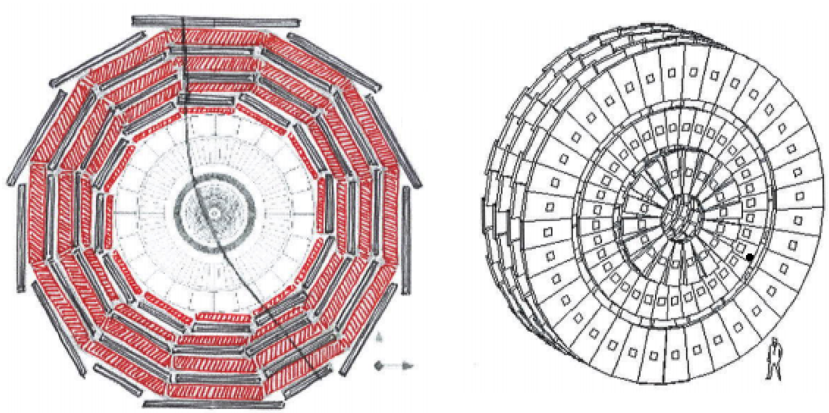

Figure 30: Left: Perspective view of the CMS Endcap muon chamber layout. Right: Barrel design of a muon system in CMS [30].

In the forward and backward regions, three iron disks constitute the Endcap yoke. As in the Barrel, two complementary muon detector systems are used. The Endcaps are instrumented with four layers of RPCs corresponding to the four disks (muon stations). The fourth disk was installed during LS1, following the same geometric distribution of the CSCs and complementing their detection capabilities. In every station the RPCs are arranged in three concentrical rings, as shown in Figure 30 (Left).

2.2.5 CMS TRIGGER SYSTEM

The LHC provides proton-proton collisions at high interaction rates. The CMS sub-detectors sum up to about 100 millions electronic readout channels, which provide information about the signal produced as the result of each BX (at the highest design luminosity). This amount of data cannot be recorded. Also, most of these interactions are not interesting for physics studies, since they involve low p_T collisions; therefore the data must be filtered quickly by a system that selects automatically only the interesting events: the CMS Trigger System.

The trigger system reduces the rate of collisions and selects only the events that have high p_T signatures. This is done in two levels: a Level-1 trigger (L1) that quickly selects events based on the presence of muon-like tracks and high energy depositions in the calorimeters; and a High Level Trigger (HLT) that uses more elaborated software algorithms to search for signatures of different kinds of interesting physics. The rate reduction capability is at least a factor of 10^6.

2.2.5.1 L1-TRIGGER

The Level-1 Trigger is custom-designed largely programmable electronics. The design output rate of the L1 is 100 kHz. The L1 uses segmented data from the calorimeters and the muon system.

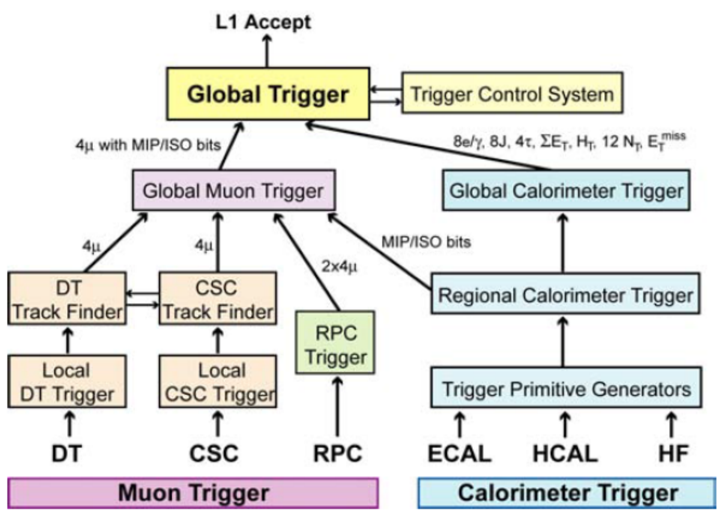

Figure 31: Architecture of the Level-1 Trigger [30].

At the first level all the information about the event is saved. The L1 system must readout the signal, make a decision, and send the accepted events to a buffer for latter use by the HLT in approximately 3.2 μs. Every operation in the L1 should be repeated every 25 ns [39].

The L1 Trigger System is organized into three major subsystems: the L1 calorimeter trigger, the L1 muon trigger, and the L1 global trigger. Also, the L1 has local, regional and global components. The local triggers are called Trigger Primitive Generators (TPG), and are related with the energy deposits in the calorimeters or hit patterns in the muon chambers. The regional trigger combines the information of the local triggers and determines what kind of objects are contained in the event. Finally, the global trigger determines the highest-rank object over the entire detector.

CALORIMETER TRIGGER

For the trigger, the calorimeters are subdivided in towers. The TPGs sum the transversal energies of the ECAL and of the HCAL to get the trigger towers as well as the correct bunch crossing. The Regional Calorimeter trigger determines electron/photon candidates, sums of the transverse energy, isolation bits, etc.

The Global Calorimeter Trigger defines jets, the total missing transverse energy ($\not{E}_T$) and the total transverse energy for the BX. It also gets the highest-rank, isolated and non isolated, electron/photon candidates. In addition, it monitors the rates for some trigger algorithms in order to get information about the LHC luminosity.

MUON TRIGGER

The information coming from DTs, the CSCs and the RPCs is used to issue the L1 muon trigger. Within the Barrel the DTs are equipped with Bunch and Track Identifiers (BTI), and they provide local trigger information in the form of tracker segments. In the Endcaps the CSCs give 3D tracker segment information. The Regional Muon Trigger joins all these track segments and assigns physical parameters to the tracks. The RPCs reconstruct and provide their own trigger candidates.

The Global Muon Trigger combines the information that comes from the three subdetectors. It also improves the trigger efficiency and reduces trigger rates. Using the complementarity and redundancy of the three sources, it suppresses background. In each BX it receives up to four Barrel and four Endcap muon candidates.

GLOBAL TRIGGER

The Global trigger has five stages: input, logic, decision, distribution and readout. The Global Trigger takes the decision to accept or reject an event based on the objects selected by the L1 Muon Trigger and the L1 Calorimeter Trigger. These objects are up to four muon candidates plus all the calorimeter information. Also, it synchronizes all the information related with the BX, since it comes at different times. The L1 accept decision is sent to the timing, trigger and control systems for distribution to the sub-systems to initiate the readout.

2.2.5.2 HIGH LEVEL TRIGGER

The goal of the High Level Trigger is to select interesting physics candidates, doing a more sophisticated reconstruction and selection on the events that passed the L1 trigger, reducing the trigger rate from about one hundred thousand to only about one hundred events per second. The HLT trigger is implemented in software and runs in a farm of more than a thousand very fast CPUs, where simplified reconstruction algorithms are executed. The criteria of the HLT algorithms are more sophisticated than the ones used by the L1 trigger [40].

The reconstruction of the objects in the HLT should be as close as possible to the online reconstruction. At the HLT level several algorithms run in parallel to search for signatures of different kinds of interesting physics, putting a ‘tag’ on the events that pass pre-established criteria. Once the HLT has accepted events, it splits them into datasets depending on the trigger path they followed, and sends them to hard disks.

2.2.6 DATA ACQUISITION SYSTEM

The CMS Data Acquisition (DAQ) system plays an important role in the experiment because the collisions and the data rates are much higher than the rate at which it is possible to store the data. At the LHC design luminosity ($\mathcal{L} = 10^{34}\,\mathrm{cm}^{-2}\,\mathrm{s}^{-1}$) it is expected that the rate will be around 20 collisions per bunch crossing every 25 ns. This implies that one MByte of data should be readout every 25 ns by CMS. This amount of data is many orders of magnitude larger than the current storage capability [30]. The DAQ system was designed to read the information coming from collisions at a rate of 10^2 Hz in order to store them for later analysis. The computational power needed to do this process in a single step is very high, for this reason the selection task is done in two steps, the first step is done by the L1 Trigger, which decreases the rate of accepted events to less than 100 kHz, an a further reduction that is done by the HLT, which reduces the L1 accept rate from 100 kHz to around 100 Hz.

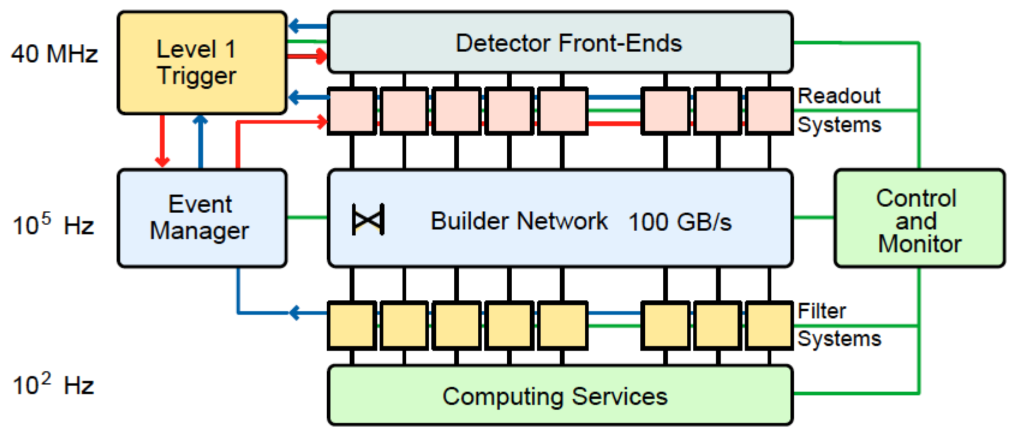

Figure 32: The architecture of DAQ system [30].

The DAQ system performs the following functions: the readout of the front-end system after the L1 trigger step, the execution of algorithms in order to accept the most interesting physics events (HLT), and the storage of these events. Also, the DAQ system operates the Detector Control System (DCS) in order to supervise the infrastructure of the subdetectors and to provide high-quality data for physics purposes [40]. The DAQ system is composed of several subsystems: detector Front-Ends, readout systems, the builder network, filter systems, the event manager, computing services and controls. A diagram of the whole system is shown in Figure 32.

The detector Front-Ends temporarily store the data that comes from the L1 Trigger. Afterwards, the data is read by the readout system, which feeds it to the filter system. The readout and filter systems are interconnected through a 100 GBps network named Builder Network. The filter system is responsible for the execution of the HLT algorithms and for event selection for offline studies. The data flow is controlled by the event manager (EVM) which also assigns event numbers, used for identifying events during processing.

The architecture of the DAQ system allows the L1 Trigger to spend 3 μs in the processing and storing the data in the front-end electronics. For the HLT case, the DAQ allows the processing time be up to 1 s, for the storing of the data in random-access memory.

The DAQ system was redesigned during the Long shutdown 1. The system has about twice the event building capacity of the original DAQ system. The new DAQ architecture takes advantage of the latest developments in the computing industry to be able to manage the amount of data that has been collected during Run 2 by the CMS detector.

3

EVENT RECONSTRUCTION AT CMS

The CMS detector stores the data that come from the pp collisions in a raw format (RAW). A RAW file is a collection of unprocessed data that contains only detector information. In order to obtain data that can be used in physics analysis, it is necessary to convert the RAW files into reconstructed files (RECO). The RECO files contain information of the physics quantities after running a series of algorithms over the RAW data that produce physical quantities (this processing of the RAW data is called reconstruction). CMS has two ways to reconstruct the ”physical objects”: the standard reconstruction and the particle-flow reconstruction. In this chapter we will present the main features of both techniques. The chapter starts with a description of the "standard" reconstruction, and after which the Particle Flow method will be discussed. Finally, some details about the reconstruction of each of the physical objects defined by CMS will be presented, focusing in the objects that will be used for the H→ $\mu^+\mu^-$ channel and for this dissertation.

3.1 STANDARD RECONSTRUCTION

In CMS the standard reconstruction process uses information coming from different subdetectors as input. Each subdetector gives a set of locally reconstructed objects. These objects are used to perform the global reconstruction of the physics objects. Therefore the "standard" reconstruction is a software process that is done in three steps: local reconstruction, global reconstruction and combination of these reconstructed objects [41].

The local reconstruction occurs in an individual subdetector module. The module uses real data coming from the DAQ system or from simulated data. The data format at this level is called "digis". After the reconstruction algorithms are applied, one goes from "digis" to reconstructed hits (RecHits). RecHits contain information about positions and energy clusters of the particles signatures in each subdetector. In the tracker the RecHits have information coming from pixels or strips in which the signal threshold is exceeded. In the calorimeters the RecHits store position, time and localized energy deposit information. In the Muon system, the RecHits store the effective drift velocity for the DTs, the distribution of charge induced for the CSCs and the position for the muon hits in the RPCs. All that information is added to the event and is used in the global reconstruction.

The global reconstruction uses the RecHits coming from the different subdetector systems to obtain the best version of this kind of objects. In the tracker system, the combination gives tracks with low or high p_T, displaced vertices, etc. In the calorimeter system, the different clusters are linked to created projective towers, useful in the jet reconstruction. Meanwhile in the Muon system, the tracks segments coming from all the sub detectors are combined. The objects obtained in this step are added to the event.

The last reconstruction step combines local and global reconstruction to obtain higher-level objects based on the information of the whole detector. Some examples are a combination of the ECAL and HCAL clusters to produce jet candidates; tracks produced by the Muon system are extrapolated into the Tracker to match with tracks reconstructed by this subdetector; a discrimination between electrons and photons is performed combining the ECAL clusters and Tracker information. The resulting objects are stored like Analysis Object Data (AOD) and are used for direct physics analyses.

3.2 PARTICLE FLOW RECONSTRUCTION

The Particle Flow (PF) algorithm aims to improve the reconstruction of physics objects using the information in a global way. It uses an optimal combination of the information gathered from all the subdetectors to better reconstruct all physics objects. In order to implement the PF algorithm, it is necessary to maximize the separation between charged and neutral hadrons, to achieve a large field integral and to have a good calorimeter granularity. The CMS detector is able to implement the PF algorithm due to the excellent resolution and granularity of its calorimeters [42].

The CMS PF algorithm is based on an efficient reconstruction of the tracks, the ability to disentangle overlapping showers and the efficient correlation between the information coming from each subdetector. The algorithm begins by extrapolating the tracks from the tracker into the calorimeters. If the extrapolated track falls into or near one or several clusters, those clusters will be associated to the track. If the clusters show the behavior of a charged hadron, the reconstruction ends in this step and the clusters are not considered any more in the rest of the algorithm. In the case of electron identification, due to the existence of photons coming from radiation it is important to carefully attach the photon clusters to the electron cluster to avoid double counting in the energy. To this end an special track reconstruction mechanism is used [43]. Muon reconstruction is based on the standard algorithm with improved selection criteria that will be discussed further in subsection 3.3.1. After reconstruction of the tracks, the remaining energy deposits are identified as photons in the case of the ECAL, and neutral hadrons in the case of the HCAL. The missing energy $\not{E}_T$ and tau reconstruction is done using the resulting charged hadrons, photons, neutral hadrons, electrons and muons.

3.3 RECONSTRUCTED PHYSICAL OBJECTS

As shown previously, the output from the CMS detector electronics is converted into physics objects using either the standard reconstruction or the PF reconstruction algorithms. The resulting objects depend on the signature left on the detector by the passing particles. Figure 33 shows a diagram with the different signatures produced by the particles when crossing the subdetectors. Starting from the collision point, the particles enter into the Tracker, in which all the charged particles leave a track. Due to the magnetic field in which the system is immersed, the curvature of the tracks provides information to measure the momentum and charge of the particles. Electrons and photons leave energy depositions in the ECAL. For the hadrons the energy is deposited mainly in the HCAL. The neutral hadrons do not have an associated track, while the charged ones do. Muons pass by the calorimeters leaving very small energy depositions and continue into the Muon system, where they leave additional tracks [44]. The neutrinos traverse all the detector system without leaving a signal but generating missing energy. All of the reconstructed physics objects used in this dissertation (muons, electrons, jets, b-jets and photons) will be described in the following subsections.

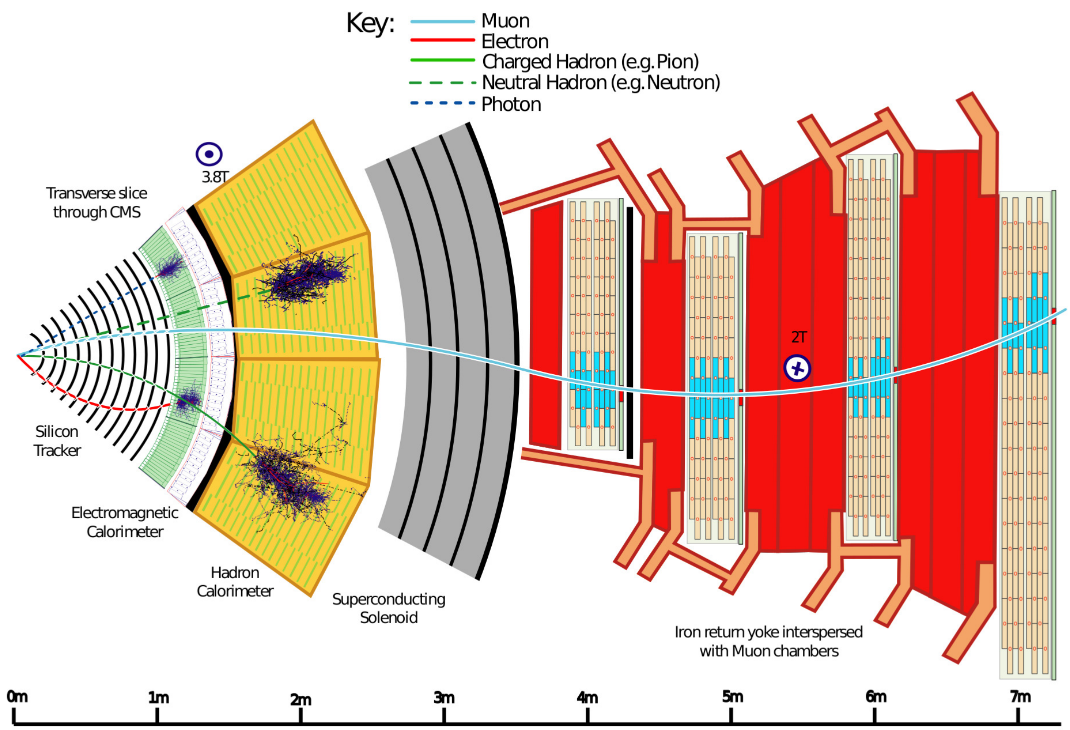

Figure 33: A diagram of the specific particle interactions in a transverse slice of the CMS detector, from the beam interaction region to the muon detector [44].

3.3.1 MUONS

Muons are charged particles that are about 200 times heavier than electrons (105.65 MeV/c^2) and whose lifetime at rest is 2.2×10^{-6} seconds, which means that in the lab frame they will travel long enough distances such that they will decay outside a detector the size of CMS. A muon can be detected with high probability because it deposits a small amount of energy in the calorimeters, reaching the Muon system where they leave very clean tracks .

In the standard reconstruction, the tracks are reconstructed in the inner tracker (*tracker track*) and in the Muon system (*standalone-muon track*). Using these tracks, CMS can provide two kinds of reconstructed muons: Global muons and Tracker muons [41]. Global muon reconstruction is performed by doing a matching between a standalone-muon track and a tracker track. The muon trajectory is extrapolated from the inner station at the Muon system back to the outer tracker surface. Because of this it is called *outside-in* reconstruction. The global muon fit can improve the momentum resolution as compared to the value obtained using only the tracker fit. Tracker muon reconstruction consists of taking all the tracks with at least 0.5 GeV in p_T and 2.5 GeV in total momentum, and extrapolating them into the Muon system. If one of the segments matches with the extrapolated track, the track is considered a Tracker muon. Tracker muon reconstruction is more efficient for the low momentum muons, because it only needs a match with one muon track segment. This is called an *inside-out* reconstruction.

In order to have a better and more efficient reconstruction, CMS uses a combination of different algorithms. It also applies a selection based on some identification variables. CMS has three basic muon identification algorithms: soft, tight and Particle-Flow [45]. The soft muon selection requires a tracker muon candidate, in which the muon segment that is matched must be matched in x and y coordinates. The tight muon selection instead requires that the candidate be reconstructed as a global and tracker muon. The PF muons are identified by three different selections: isolated, pf-tight and pf-loose. For the isolated muons the PF algorithm requires a global muon candidate. After all the isolated muons in the event have been selected, the pf-tight and pf-loose muon selections are applied to the remaining muons. These two selections optimize the identification of muons within jets and tag them as fake muons. The PF muons are used by CMS except in high p_{T} muon analyses. A detailed description of the muon identification criteria can be found at the CMS MuonPOG page [46].

3.3.2 ELECTRONS

Electrons are charged particles that can lose energy by ionizing atoms when passing through the detector. They can leave a trace in the Tracker system and deposit a large amount of energy in the ECAL, mainly by bremsstrahlung and pair production mechanism. This is called an electromagnetic shower. CMS reconstructs the electrons by associating a track in the silicon detectors with an energy cluster in the ECAL.

The reconstruction of the electrons starts by the identification of clusters in the ECAL. Usually the electron energy is collected in a small number of crystals. Approximately 97% of the incident energy of a single electron is contained in a cluster of 5×5 crystals. In the case of electrons that radiate photons before entering the ECAL, the average energy radiated could go from 33% up to 86%. In order to measure the total energy of the electron it is necessary to take into account the energy of these radiated photons.

CMS uses two complementary algorithms to reconstruct the clusters: the *hybrid-algorithm* in the Barrel and the *multi 5×5* algorithm in the Endcap. The hybrid algorithm collects the energy in a small window in η and an extended window in ϕ. The multi 5×5 algorithm builds clusters connecting rows of crystals that have energies that decrease monotonically when moving away from the seed crystal (More details in references [47, 48]).

For the electron tracks, since the bremsstrahlung emission causes a deviation in the electron trajectory, a dedicated electron track reconstruction is used, modeling the bremsstrahlung energy loss distribution by a Gaussian mixture in an algorithm called the Gaussian-Sum Filter (GSF) [43]. The electron candidates are reconstructed from the association of a GSF track with a cluster in the ECAL, evaluating the distances between the track extrapolation and the cluster position. The clustering can be created also by using the PF algorithm. In this case, the clustering is driven by the GSF tracks. The cluster from the electron at the entrance of the ECAL is matched to the track at the exit of the tracker [49].

In order to identify electrons and to differentiate them from photon conversions (electrons produced by photons), misidentified jets, or electrons from semileptonic decays, the discrimination variables are divided into three groups: variables that compare ECAL and tracker measurements, purely calorimetric variables, and tracking observables. A detailed description of the electron identification criteria can be found at the CMS EGammaPOG page [50, 51].

3.3.3 PHOTONS

In CMS, photons are primarily reconstructed from energy deposits in the ECAL. Approximately 95% of the energy of unconverted photons is contained in a cluster of 5×5 crystals. However, due to the interaction of the photons with the tracker material, some of them are converted into e^+e^- pairs. The photons conversion probability goes from 27% in the center of the barrel up to 62% in the forward region [52]. The e^+ and e^- tracks bend in the magnetic field and then the energy is deposited in the ECAL, producing some spreading in ϕ. This results in a degradation of the energy resolution for the reconstruction of photons.

Using the standard reconstruction, a very convenient measure of the lateral spread of energy deposition is provided by the *R9* variable, defined as the ratio of energy contained in a 3×3 array of crystals to the total supercluster energy. Values approaching unity are obtained for unconverted photons, or for photons that have converted very close to the ECAL [52].

In order to reduce the background for *prompt photons* or unconverted photons, it is necessary to apply isolation requirements. Fake photon signals due to jets can be rejected by looking for additional energetic particles in a cone around the reconstructed ECAL cluster. In the case of converted photons, CMS requires two tracks each with at least three hits and matching in (η, ϕ) with a supercluster with transverse energy bigger than 25 GeV. Each of these photons can have more than one track or track pairs. If at least one reconstructed track or track pair is found, the converted photon is reconstructed [53].

As electrons emit photons and photons are converted into electron pairs, it is necessary to save all the information about tracks, energy recovery, and clusters in order to differentiate one from the other. Using PF reconstruction, the photon candidates are retained if they are isolated from other tracks and calorimeter clusters in the event. This selection is not so strict as the one applied at the analysis level to select isolated photons. After isolated photons are selected, the non-isolated ones can be identified. All the ECAL clusters with η values less than 2.5 are reconstructed as photons.

The PF algorithm has an advantage over the standard reconstruction: it is able to reconstruct photons with p_T larger than 2 GeV as compared to the standard reconstructions that has a 10 GeV minimum. Photons with small p_T values are used for final-state radiation recovery. This kind of photons are important in the H→ $\mu^+\mu^-$ channel and will be used to recover dimuon events in which the muons radiated photons. The recovery methods that we studied are described in Chapters 5 and 6.

After reconstruction, CMS has two photon identification algorithms: an approach that uses selection requirements applied to a set of individual variables, and a multivariate technique. Both algorithms include an "electron-veto" to reject electrons misidentified as photons. More details of those methods can be found at the CMS EGammaPOG page [54, 55].

3.3.4 JETS

A jet can be defined as a collimated spray of particles arising from the fragmentation and hadronization of high p_T quarks and gluons coming out of the collision. In CMS, jets leave energy deposits in the ECAL and the HCAL, together with tracks pointing in the same direction. Three main jet reconstruction algorithms have been used in CMS: the iterative cone, the midpoint cone and the inclusive k_T jet algorithms [56].

In the iterative cone algorithm a list of particles or calorimeter towers ordered by transverse energy is used as an input. From this list one object is taken as a seed for a *proto-jet*. A cone of size R in η, ϕ space around the seed is used to find the largest transverse energy. The objects inside the cone allow the calculation of a *proto-jet* direction and energy. The procedure is repeated until the energy of the proto-jet changes by less than 1%. When a stable proto-jet is found, the particles in the pro-jet are removed from the list, and the proto-jet is added to the jet list [56].

The midpoint-cone algorithm also uses an iterative search of stable cones. In contrast to the iterative cone algorithm, the midpoint-cone algorithm does not remove the input objects from the list, resulting in overlapping proto-jets. Once a proto-jets list is created, for each pair of jets laying in a cone radius closer than R, the midpoint is used as an additional seed and therefore some proto-jets are merged [56].

The inclusive k_{T} jet algorithm is a cluster-based jet algorithm. For each proto-jet a distance is assigned ($d_i = E_{\mathrm{T}i}^2$). Each proto-jet is compared with all other proto-jets present in the event. The distance between each pair of proto-jets is defined by $d_{ij} = min(E_{\mathrm{T}i}^2, E_{\mathrm{T}j}^2)[(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2]/R^2$. The smallest d_i and d_{ij} is found and they are compared; if the smallest value is d_i, this jet cannot be merged and is taken out of the list of available proto-jets. If, on the contrary, d_{ij} is the smallest value then both jets are merged [56, 57].

Jets are classified according to the information used in their reconstruction. In CMS there are three kinds of jets: calorimeter jets, track jets, and calorimeter plus track jets. The calorimeter jets or *Calo-jets* are constructed using only information coming from the ECAL and the HCAL. Once the *Calo-jets* are reconstructed, the tracks are associated with a cluster based on a spatial separation in the η-ϕ plane. Jets with a p_{T} lower than 20 GeV are rejected. Track jets are reconstructed from tracks of charged particles, independent of calorimetric information. In the case of calorimeter plus track jets or JPT jets, information from the calorimeters and the tracker is combined.

Jets are also reconstructed using the PF algorithm. All particles reconstructed with the PF algorithm, without distinction of type and without any energy threshold, are clustered into *particle-flow jets*. PF jets are studied down to a p_{T} of 15 GeV. Since PF jets use the totality of the available information from the subdetectors, their reconstruction quality is much better [58].

B-JETS

The algorithm which is implemented at the detector level in order to discriminate *b-jets* from the other jets is known as "b-tagging". The b-tagging algorithm is based on the properties of the b-quarks and B-hadrons. When a b-quark is produced due to hard scattering or is emitted through hard radiation, it immediately hadronizes, producing a B-hadron that travels a long-mean path before decaying. An event in which a B-hadron is produced is detected as a collection of tracks. Most of these point to the interaction point, but a subset of them come from the decay vertex of the B-hadron; i.e. a displaced vertex. A detailed description of the b-tagging methods can be found at the CMS BTagPOG page [59].

In CMS, jets and b-jets have different identification algorithms. For jets, CMS has developed the PF identification algorithm (*PFJetID*) in order to reject fake, badly reconstructed, and noisy jets. About 1-2% of the jets are rejected due to this. Four levels of jet tagging are also defined according to the jets η value: "loose" and "tight" for $|\eta| < 3$, and "loose" and "tight" for $3 < |\eta| < 5$. A detailed description of these identification criteria can be found in Jet Identification page [60]. For the b-jets, CMS uses a set of discriminators listed in the CMS BTagPOG page [61].

All of the physics objects previously described are employed in the searches of the H→ $\mu^+\mu^-$ decay channel. There are other physics objects such as taus and missing transverse energy (MET) that are also used. As they are not necessary for the main subject of this dissertation, we have not discussed them in this chapter. Detailed information about them can be found in CMS TauPOG page [62] and CMS JetMETPOG page [63].

4

THE H→ $\mu^+\mu^-$ DECAY CHANNEL

In the SM, once the mass of the Higgs boson is known, all its couplings are fixed, including the coupling to muons. Any significant deviations in the experimental results from the values expected from the SM predictions could indicate the onset of new physics. Therefore, it is very important to measure all possible decay channel parameters. A search for events in which the recently discovered Higgs boson decays into a pair of muons is presented in this chapter. The search was performed using the data collected by the CMS experiment during Run 2 of the LHC. The data comes from p-p collisions at $\sqrt{s} = 13$ TeV, corresponding to an integrated luminosity of 35.9 fb^{-1}. The chapter starts with a description of the decay channel and of the data and Monte Carlo simulations involved on the analysis. Then, the event selection and categorization are discussed. Afterwards, the signal and background modeling are presented. A detailed description of this analysis has been presented by the CMS H2Mu Analysis Group in the Analysis Note CMS AN-17-098 [64] and a paper submitted to Physical Review Letters [65].

4.1 THE DECAY CHANNEL

In the standard model, the fermion mass depends on its coupling to the Higgs boson field. In the LHC there are four standard model fermions that couple to the Higgs, in which their couplings can be measured: the top quark, the bottom quark, the tau lepton and the muon. Except for the muon coupling, the previous analyses, done by CMS and ATLAS experiments, have extracted values for these couplings [16–19].

In the case of muons, the H→ $\mu^+\mu^-$ decay rate, for a SM Higgs mass of 125 GeV, is approximately 0.02%. This decay rate is much smaller than that of the $b\bar{b}$ and $\tau^+\tau^-$ channels. The branching ratio for the channel is 2.4×10^{-4} for a 125 GeV Higgs, which makes it the smallest branching ratio accessible at the LHC during the Run 2 [23, 24]. However, thanks to the good momentum reconstruction for muons in CMS, most of the H→ $\mu^+\mu^-$ events can be measured with good dimuon mass resolution which allows for an enhanced discrimination between signal and background events.

As we mentioned in section 1.4, information of interest can be gathered from the cross-section of the Higgs decay to muons, including information about the Yukawa coupling, possible violations of lepton flavor in the H→ $\mu\tau$ channel, and restrictions to the couplings to fermions of different generations [23, 25].

After the searches for the Higgs boson in this channel at 7 and 8 TeV by CMS and ATLAS, the CMS H2Mu Analysis Group is performing the first search at 13 TeV. In this analysis we take into account the contribution of the most relevant Higgs production channels: gluon-gluon fusion (ggH), vector boson fusion (VBF), Higgs-strahlung (VH), and association with a top quark ($t\bar{t}H$). We have developed an event categorization and selection criteria that maximizes the search sensitivity. We will discuss the event selection in section 4.3 and the categorization in section 4.4.

4.2 MONTE CARLO AND DATA SAMPLES

Background and signal simulated data was used during the development of the analysis. For the signal four of the main production methods were considered: gluon fusion (ggH), vector boson fusion (VBF), Higgs-strahlung (VH) and production via association with top quarks ($t\bar{t}H$), as described in section 1.2. Monte Carlo (MC) samples from the central production at CMS, generated using POWHEG [66], for a 125 GeV Higgs boson, were used. The samples used for signal during the studies are listed in Table 4.

Higgs signal MC samples	Events	XSec [pb]	XSec×BR [fb]
/GluGlu_HToMuMu_M125_13TeV_powheg_pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	250000	48.58	10.571
/VBF_HToMuMu_M125_13TeV_powheg_pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	249200	3.7817	0.8229
/WPlusH_HToMuMu_M125_13TeV_powheg_pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	124547	0.09426	0.02051
/WMinusH_HToMuMu_M125_13TeV_powheg_pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	125000	0.05983	0.013019
/ZH_HToMuMu_M125_13TeV_powheg_pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	249748	0.17762	0.03865

Table 4: The Higgs signal MC samples used in the H→ $\mu^+\mu^-$ analysis [64, 67].

The main sources of background are due to the Drell–Yan (DY), $t\bar{t}$, single top, single top + X and production of diboson and triboson. In the case of Drell–Yan, scenarios with 1, 2 and no jets were considered. For the diboson production channels, WW, ZZ and WZ samples with several decay variants were used. A similar set of samples and decay variants were used in the triboson case. The majority of samples were produced using MadGraph [68] or MC@NLO, except for the single top samples, which were produced using POWHEG. The samples used are listed in Tables 5 and 6.

The data used in the H→ $\mu^+\mu^-$ search come from proton-proton collisions at a center-of-mass energy of 13 TeV. This data was collected by CMS during Run 2 of the LHC. The total data sample used for this analysis corresponds to an integrated luminosity of 35.9 fb^{-1}. The datasets used during the analysis are summarized in Table 7.

Drell–Yan MC samples	Events	XSec [pb]
/DYJetsToLL_M-50_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext2-v1/MINIAODSIM	122055388	5765
/DYToLL_0J_13TeV-amcatnloFXFX-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM	49579613	4754
/DYToLL_1J_13TeV-amcatnloFXFX-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM	49902571	888.9
/DYToLL_2J_13TeV-amcatnloFXFX-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v2/MINIAODSIM	42324802	348.8
/DYToLL_2J_13TeV-amcatnloFXFX-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM	47974554	348.8
Single Top MC samples	Events	XSec [pb]
/ST_tW_top_5f_NoFullyHadronicDecays_13TeV-powheg_TuneCUETP8M1/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	5372991	35.85
/ST_tW_top_5f_NoFullyHadronicDecays_13TeV-powheg_TuneCUETP8M1/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM	3256650	35.85
/ST_tW_antitop_5f_NoFullyHadronicDecays_13TeV-powheg_TuneCUETP8M1/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	5425134	35.85
/ST_tW_antitop_5f_NoFullyHadronicDecays_13TeV-powheg_TuneCUETP8M1/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM	3256407	35.85
Single Top + X MC sample	Events	XSec [pb]
/tZq_ll_4f_13TeV-amcatnlo-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_-mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM	14509520	0.0758
Top Pair MC samples	Events	XSec [pb]
/TTJets_DiLept_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	6094476	85.656
/TTJets_DiLept_TuneCUETP8M1_13TeV-madgraphMLM-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM	24350202	85.656
/TTJets_Dilept_TuneCUETP8M2T4_13TeV-amcatnloFXFX-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	14529280	85.656

Table 5: The DY, Single Top and Top Pair MC samples used as background sources in H→ $\mu^+\mu^-$ analysis [64, 67].

Diboson MC samples	Events	XSec [pb]
/WWTo2L2Nu_13TeV-powheg/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	1999000	12.46
/WZTo3LNu_TuneCUETP8M1_13TeV-amcatnloFXFX-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	11887464	2.113
/WZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	26517272	4.409
/ZZTo2L2Nu_13TeV_powheg_pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	8842475	0.564
/ZZTo2L2Q_13TeV_amcatnloFXFX_madspin_pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	15345572	3.22
/ZZTo4L_13TeV-amcatnloFXFX-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM	10709784	1.212
Triboson MC samples	**Events**	**XSec [pb]**
/WWW_4F_TuneCUETP8M1_13TeV-amcatnlo-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	240000	0.2086
/WWZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	250000	0.1651
/WZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	246800	0.05565
/ZZZ_TuneCUETP8M1_13TeV-amcatnlo-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6-v1/MINIAODSIM	249237	0.01398
Top Pairs MC samples	**Events**	**XSec [pb]**
/TTWJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v3/MINIAODSIM	2160168	0.2043
/TTWJetsToLNu_TuneCUETP8M1_13TeV-amcatnloFXFX-madspin-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext2-v1/MINIAODSIM	3120397	0.2043
/TTZToLLNuNu_M-10_TuneCUETP8M1_13TeV-amcatnlo-pythia8/RunIISummer16MiniAODv2-PUMoriond17_80X_mcRun2_asymptotic_2016_TrancheIV_v6_ext1-v1/MINIAODSIM	1992438	0.2529

Table 6: Diboson, Triboson and Top pairs MC samples used as background sources in H→ $\mu^+\mu^-$ analysis [64, 67].

In order to have a better match between simulated and real data, it is necessary to re-weight the pile-up distribution for MC samples. MC samples are generated for the expected distributions of number of pile-up interactions for the data-taking period. However, because the primary vertex reconstruction depends on the offline selection criteria and trigger, differences between MC and real data arise. In order to correct for these differences, a re-weighting procedure is employed using the luminosity distribution of the data sample, calculated from the pile-up JSON file, and the distribution of the number of pile-up events in the MC samples. CMS has developed algorithms to calculate the re-weight, whose details can be found in the CMS Pileup Re-weighting utilities page [69]. In this analysis, the re-weighting was done using a minimum cross-section bias of 69.2 mb in data, and a 69.2 mb, 65 mb, and 72 mb minimum cross-section bias for MC samples. Figure 34 shows the distribution of measured data and MC truth pile-up, and the re-weighting factors.

Dataset	Run Range	Integrated Lum. [fb^{-1}]
`/SingleMuon/Run2016B-03Feb2017_ver2-v2/MINIAOD`	272007-275376	5.788
`/SingleMuon/Run2016C-03Feb2017-v1/MINIAOD`	275657-276283	2.573
`/SingleMuon/Run2016D-03Feb2017-v1/MINIAOD`	276315-276811	4.248
`/SingleMuon/Run2016E-03Feb2017-v1/MINIAOD`	276831-277420	4.009
`/SingleMuon/Run2016F-03Feb2017-v1/MINIAOD`	277772-278808	3.102
`/SingleMuon/Run2016G-03Feb2017-v1/MINIAOD`	278820-280385	7.540
`/SingleMuon/Run2016H-03Feb2017_ver2-v1/MINIAOD` `/SingleMuon/Run2016H-03Feb2017_ver3-v1/MINIAOD`	280919-284044	8.606
Luminosity mask: `Cert_271036-284044_13TeV_23Sep2016ReReco_Collisions16_JSON.txt`		

Table 7: Single muon data collected during the proton-proton collisions at $\sqrt{s} = 13$ TeV by CMS at LHC in 2016 [64, 67].

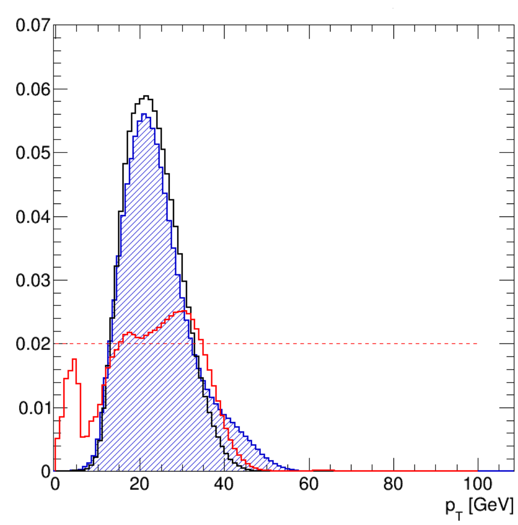

Figure 34: Measured data (black line) and MC (blue hashed line) truth pile-up distributions, as well as the pile-up reweighting factors (red line) multiplied by 0.02 [64].

4.3 EVENT SELECTION

All the physics objects used in the event selection were reconstructed using the PF algorithm. These objects were reconstructed as described in chapter 3. The candidate identification was done as recommended by each of the CMS POGs.

The event selection requires the lowest un-prescaled trigger available for the analyzed runs, and at least one isolated muon with a p_T greater than 24 GeV. The triggers: `HLT_IsoMu24` and `HLT_IsoTkMu24`, meet the trigger requirements. Another requirement for the selected muons is that at least one of their reconstructed primary-vertex (PV) is valid. This vertex must also have at least four degrees of freedom (NdfPV).

The muon candidates must be global and tracker muons. They must be in a range of pseudo-rapidity lower than 2.4 and have a p_T greater than 20 GeV. At least one of the muons must match one of the triggers and have a p_T larger than 26 GeV. Additionally, the muons must be isolated in a cone with radius ($\Delta R \equiv \sqrt{\Delta\eta^2 + \Delta\phi^2}$) less than 0.4. We have employed the relative isolation criterion of $I_{rel}^{PF} < 0.25$, which corresponds to a "loose" PF-based combined relative isolation. The selected muons in the pair must have complementary charges. Only events with a dimuon invariant mass between 110 and 150 GeV were considered for signal extraction [70].

Since the width of the Higgs-boson-reconstructed mass distribution is dominated by the detector resolution, it is important to improve the dimuon mass resolution as much as possible. CMS provides a software package to allow for the correction and calibration of the muons momenta named *Rochester Muon Momentum Corrections* [71, 72]. These corrections were validated on the Z^0 peak mass distribution. When applied to data, the corrections reduced the variation on ϕ from 0.1 GeV down to 0.0025 GeV; an improvement of 98%. The resolution of the Z^0 peak was improved in 1.6%, corresponding to about 20 MeV. The Rochester corrections allow us to align the scale and resolution of real data, as well as simulated data [64].

The PF jets considered in the event are those that do not overlap the selected muons. To select them, all jets with $\Delta R < 0.4$ between them and the muons are removed. The jets must have a p_T greater than 30 GeV and be located in a pseudo-rapidity range $|\eta| < 4.7$. The PF jets "loose" identification, mentioned in section 3.3.4, is used to reject pile-up or misidentified jets.

Events that have b-jets with $p_T >$ 20 GeV, located in a pseudo-rapidity range of $|\eta| < 2.4$, and in which the value for the CSVv2 tag is less than 0.8484 are considered. They will allow for the discrimination of events that come from $t\bar{t}$ (See Section 3.3.4). This CSVv2 tag value corresponds to a 60-65% b-tagging efficiency. In the case of FSR studies, this selection is used as a b-jet veto [64, 67].

Initially an electron veto was considered; events in which there were electrons identified as "MediumID", with a $p_T >$ 10 GeV. Also, events with an isolation cone to the selected muons $\Delta R(\mu, e) < 0.4$ were rejected. However, this criterion was not included in any of the reference categories in the final analysis. For the FSR studies the electron-veto was used in order to better select the FSR photons [67]. Table 8 summarizes the selection criteria over the 13 TeV samples.

Muon selection	
Trigger	`HLT_IsoMu24 OR HLT_IsoTkMu24`
Pseudo-rapidity	$\|\eta\| < 2.4$
Isolation	$I_{rel}^{PF} < 0.25$
p_T	> 26 GeV (leading muon – trigger match) > 20 GeV (sub-leading muon)
PF-Jet selection	
PF-Jet ID	Loose ID
p_T	> 30 GeV
Pseudo-rapidity	$\|\eta\| < 4.7$
"Isolation"	$\Delta R < 0.4$
B-Jet selection	
p_T	> 20 GeV
Pseudo-rapidity	$\|\eta\| < 2.4$
CSVv2 tag	less than 0.8484
Electron Veto	
Electron ID	MediumID
p_T	> 10 GeV
Isolation	$\Delta R(\mu, e) < 0.4$

Table 8: Summary of the selection criteria.

4.4 CATEGORIZATION

After selecting the events, they are categorized in order to improve the analysis sensitivity. A multivariate analysis method was implemented using a Boosted Decision Tree (BDT) in order to categorize the events. The BDT was trained to discriminate between signal and background events.

The variables used in the training of the BDT, and the categorization of events are:

- p_T and η of the dimuon.
- $|\Delta\eta|$ and $|\Delta\phi|$ between the selected muons.
- η for the highest p_T jet among those selected.
- Mass of the jet pair with the highest mass.
- $|\Delta\eta|$ for the previous jet pair.
- Number of jets with $|\eta| < 2.4$ and $|\eta| > 2.4$.
- Number of b-jets.
- Missing transverse energy (MET).

Both, signal and background samples, are normalized according to their SM cross-sections.

The resulting categories divide the event sample according to the pseudo-rapidity of the muons. The BDT score ranges between -1.0 for background-like events to 1.0 for signal-like events. The categorization algorithm chooses between events using the maximum $|\eta|$ value for the muon or the BDT score, creating two event categories. After this, the algorithm iterates within the two categories and splits each into two, so that the maximum gain is achieved. The process is repeated until the desired number of categories is reached.

The maximum value of $|\eta|$ for the muons allows us to isolate the events with the best mass resolution, which is achieved for $|\eta| < 0.9$, in the central section of the detector, and degrades as the muons become forward, especially in sectors with $|\eta| > 1.9$. The final categories were obtained through a simplification of the ones obtained through the BDT. Categories with the same behavior were merged together. At the end a total of 15 categories were used, and they are shown in Figure 35 [64, 70].

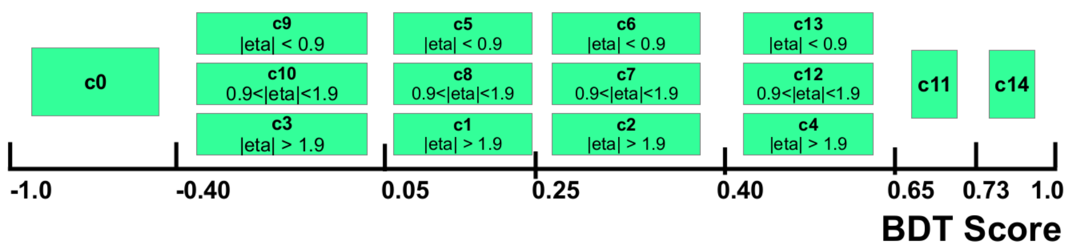

Figure 35: The final categorization of the events. The η value refers to the maximum pseudo-rapidity value of the two muons in the dimuon pair [70].

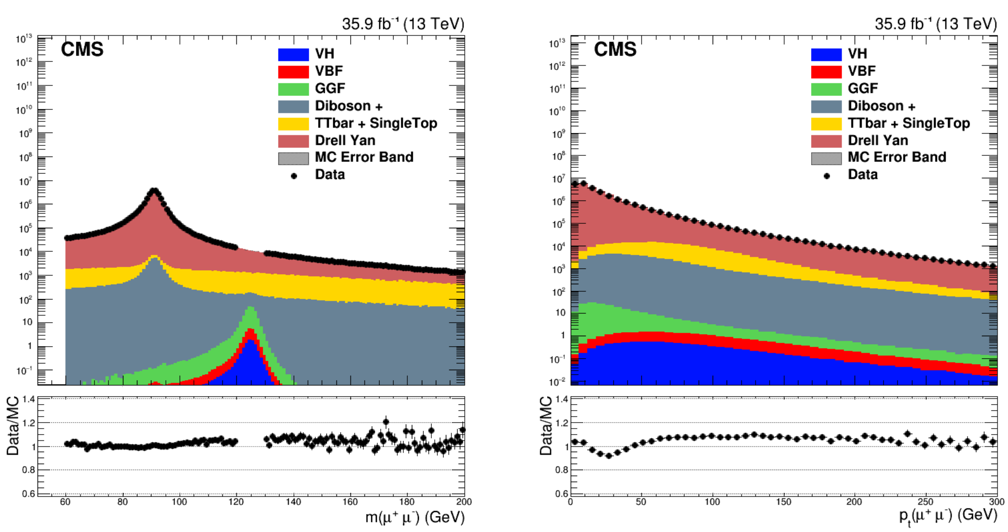

Figure 36: Data and MC comparison. The left panel shows the dimuon invariant mass distribution and the right panel shows the dimuon p_{T} distribution for the inclusive events [70].

To validate the categories and to ensure that the signal was modeled correctly, it is important that real data and simulated data are in agreement. The results show a good agreement between them, as can be seen in Figure 36.

4.5 SIGNAL MODELING

The signal is modeled as a composition of gaussian functions. Each category and production process is modeled separately. The modeling depends only on the Higgs boson mass and the nuisance parameters.

The signal model includes both, shape and normalization, and is defined as:

$$s(x, m_H, \theta) = \sum_{n=1}^{N} \tilde{f}_i \mathcal{N}(x_i, \mu_i, \sigma_i) \qquad , \tag{1}$$

where $\mathcal{N}(x_i, \mu_i, \sigma_i)$ is the normal distribution and $\mu_i(m_H, \theta)$, $\sigma_i(m_H, \theta)$ the mean and sigma values for each gaussian. x is the reconstructed dimuon mass. Finally, $\tilde{f}_i$ are recursive fractions used to model the relative importance of each pdf.

The final model uses three gaussian functions ($N = 3$), which allows for a good representation of the modeled distribution, even at its tails. The second and third gaussians are used to model the low invariant mass events, which are produced as an effect of FSR processes. For each category and production process the invariant mass is adjusted for three different Higgs boson masses: $m_H = 120, 125, 150$ GeV. This is done in order to test various hypotheses about the Higgs mass, and to test any mass in the range 120-130 GeV [64, 70].

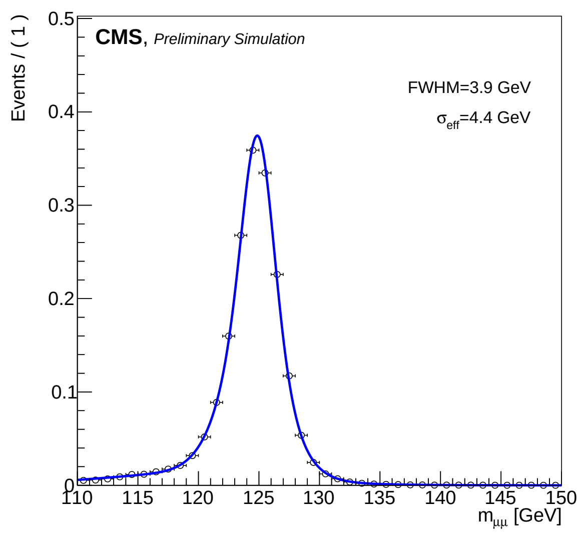

Figure 37: Signal model compared to MC predictions, summing up the contribution from all process and all categories [64].

It is necessary to calculate the yield for each category and production mode in order to normalize the signal model. The expected yield is calculated as:

$$\text{Yield} = \mathcal{L}\sigma(pp \to H)B(H \to \mu\mu)\epsilon A \qquad , \tag{2}$$

where ϵ is the efficiency and A the acceptance. The values for the production cross-section for each process, and the decay branching ratio were taken from the Handbook of LHC Higgs Cross Sections: 4 (Yellow Report) [73].

Figure 37 shows the signal model prediction compared to the MC sample for all the categories, summing up the contribution from all the processes.

4.6 BACKGROUND MODELING

Due to the dominant nature of the background processes, it is important to model each of the contributions optimally. The dominant background sources are: Drell-Yan and top pair production ($t\bar{t}$). The background can be modeled as a slowly and softly decaying exponential plus a polynomial contribution. Two kinds of functions were considered for the modeling: the first model is motivated by the physics, in which the functions are modifications of the Breit-Wigner distribution coming from the Z peak; the second includes Bernstein polynomials and sums of exponentials [70].

Figure 38 shows examples of some of the background fits to the data for the aforementioned functional forms.

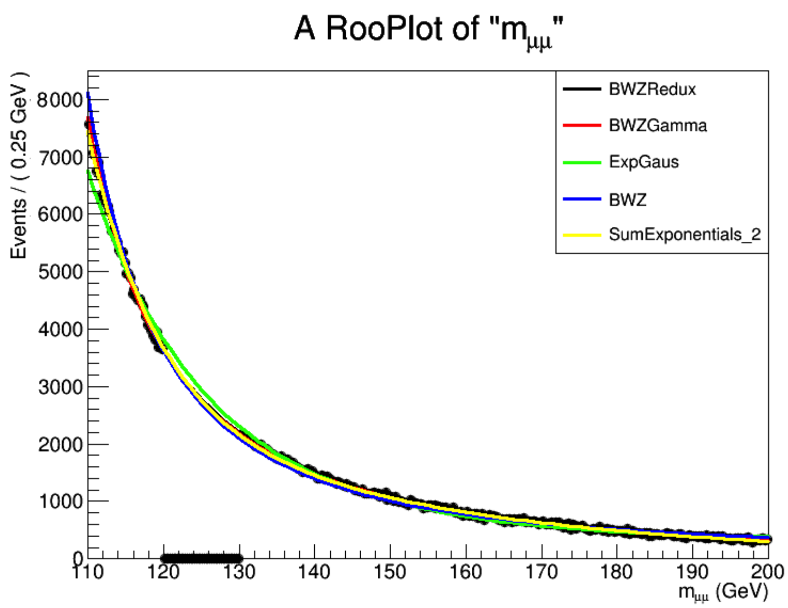

Figure 38: Background-only fits to the dimuon mass spectrum in data. Any category is applied for this plot [64].

The functions from the first group that were used are:

$$\text{BWZ:}B(x) = \frac{e^{ax}\sigma_z}{(x-\mu_z)^2 + (\frac{\sigma_z}{2})^2}, \tag{3}$$

$$\text{BWZRedux:}B(x) = \frac{e^{a_2 x + a_3 x^2}}{(x-\mu_z)^{a_1} + (\frac{2.5}{2})^{a_1}}, \tag{4}$$

$$\text{BWZGamma:}B(x) = f\frac{e^{ax}\sigma_z}{(x-\mu_z)^2 + (\frac{\sigma_z}{2})^2} + (1-f)\frac{e^{ax}}{x^2}. \tag{5}$$

And for the second group:

$$\text{Bernsteins:}B(x) = \sum_{i=0}^{n} \alpha_i [\binom{n}{i} x^i (1-x)^{n-i}], \quad (6)$$

$$\text{SumExponentials:}B(x) = \sum_{i=1}^{n} \beta_i e^{\alpha_i x}. \quad (7)$$

For each of the 15 categories a functional form was chosen so that any possible bias would be minimized. Each category has different functional forms due to differences in the invariant mass resolution of the dimuons, specially for those events coming from DY processes.

In Chapters 5 and 6 ,corrections to this analysis due to FSR processes, are presented. These chapters correspond to my main contribution for this analysis. The final results of the H→ $\mu^+\mu^-$ searches at 13 TeV will be discussed in Chapter 7.

5

FINAL-STATE RADIATION RECOVERY

Due to the small branching fraction for the H$\rightarrow \mu^+\mu^-$ channel, predicted by the SM, the signal-to-background ratio is expected to be small as well [73]. Thus, it is important to recover every possible dimuon event that can increase the signal sensitivity. Studies of the recovery of events that have final-state radiation are an important topic of this dissertation, therefore in this Chapter I will present them in detail. The Chapter starts with a description of the importance of the recovery of events that have photons coming from final state radiation (FSR) off muons, next the selection of FSR candidates using a cut-based procedure and the FSR recovery algorithm that I have implemented will be discussed. Finally, validation studies of this algorithm, using real and simulated data for Z^0, will be presented. In Chapter 6 similar studies are presented using a Multivariate Analysis (MVA) technique for the selection of the FSR candidates.

5.1 FINAL STATE RADIATION

In order to maximize the amount of recovered dimuon Higgs events, increasing the signal sensitivity, we study events in the low-mass tail of the Higgs dimuon distribution that could be the result of emission of photons by the muons. These photons are named final-state radiation photons (FSR-photons).

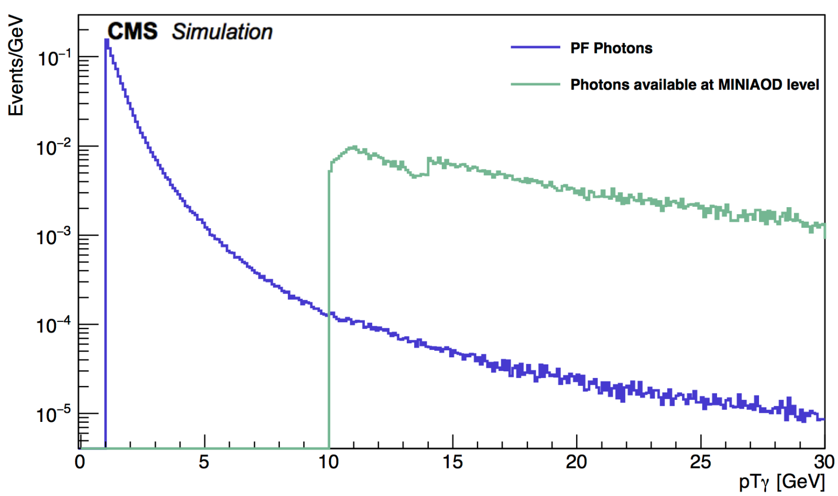

Figure 39: The transverse momentum (p_T) distribution of photons in the *Particle Flow Collection (PF photons)*, and from the *Photon Collection, available at the MINIAOD level* [74, 75].

Even though the number of events that have FSR photons is small compared to the total expected number of events ($\sim 5\%$ in a typical mass window), these events should be identified and their effect should be included in the analysis because:

- It will improve the invariant mass resolution.
- It will improve the signal- to- background ratio.

FSR photons have a soft p_T spectrum that makes their reconstruction, selection and identification non-trivial. The standard photon reconstruction algorithms are not applicable in this context because they aim for high energy isolated objects that have p_T larger than 10 GeV. The Particle Flow (PF) algorithm allows for the reconstruction of photons with p_T as low as 2 GeV, as it can be seen from Figure 39. Therefore, the Particle-Flow photons will be used for FSR recovery of the dimuon final state.

For our $\mu^+\mu^-$ sample we selected photons from the PF collection with a small $\Delta R(\gamma,\mu)/E_{T\gamma}^2$ as presented in Section 5.2. Once the FSR photons candidates have been selected, we study their effect in the reconstruction of the invariant mass, in order to accept or reject them in the analysis (see Section 5.3). In order to validate our procedure we used the FSR algorithm on Monte Carlo and data samples.

5.2 CUT-BASED FSR PHOTON CANDIDATE SELECTION

Particles reconstructed using the PF algorithm are stored in the `particleFlow` collection of the mini AODs. The FSR photon candidates come from this collection. For a photon candidate we required a p_T>2 GeV, as well as $|\eta| < 2.4$ [76]. We also required a photon PF relative isolation [77], calculated using a cone of radius $R = 0.3$. We also used a threshold of 0.2 GeV on charged hadrons, with a veto cone of 0.0001, and a threshold of 0.5 GeV on neutral hadrons and photons, with a veto cone of 0.01. We also took into account the contribution from Pile-up (PU) vertices. For these studies we used a photon PF relative isolation less than 1.8 [78]. The FSR photon candidates must also satisfy the criterion: $\Delta R(\gamma,\mu)/E_{T\gamma}^2 < 0.012$ to assure that the closest photons to a muon with the highest p_T, are selected. There may be more than one candidate associated to each muon; only one of them is selected, the one with the smallest $\Delta R(\gamma,\mu)/E_{T\gamma}^2$ value.

Our FSR recovery procedure is similar to the one used in the H→ZZ channel [78]. The two selection criteria used in that channel are: $\Delta R(\gamma,\mu)/E_{T\gamma}^2 < 0.012$ and $\Delta R(\gamma,\mu)< 0.5$. In our analysis we studied the effect of the second criterion by setting its value to: less than 0.5, greater than 0.5, or excluding the criterion. The results of this study is summarized in section 7.1. The main conclusion is that, in our case, the $\Delta R(\gamma,\mu)$ criterion does not have a significant contribution, and therefore we do not use it.

5.3 FSR RECOVERY

In order to include the effect of the FSR photons in the invariant mass reconstruction of dimuon events (FSR Recovery), the muon relative isolation has to be updated when an FSR photon can be associated to the respective muon. In the initial selection we chose muons whose PF-based combined relative isolation – with $\Delta\beta$ correction– passes the loose cut criterion (smaller than 0.25) [46]. When a muon does not pass the isolation criterion, it could be due to a close photon or photons. In this case there is a certain probability that the isolation value is larger due to the FSR process. The four-momenta of the photons should be included in the invariant mass reconstruction of the dimuon to see if it improves the final selection. Therefore, in order to do the FSR recovery, we also look at non-isolated muons, in order to determine if this criterion was not met due to a photon, or photons. If that is the case we also include those events in our sample and correct the invariant mass using the photon information.

For muons that do pass the isolation criterion, there is still the possibility of having FSR photons, that did not fall into the isolation cone. By enlarging that cone one can find events with photons that, when included in the invariant mass reconstruction, improve the resolution. Therefore, in our analysis we do FSR recovery for isolated and non-isolated muons using information from the PF photon collection.

In order to include the FSR effect , each photon that satisfies the FSR selection criteria is associated to the closest muon. If it complies with the additional requirements: $m_{\mu\mu\gamma} < 130$ GeV and $|m_{\mu\mu\gamma} - m_H| < |m_{\mu\mu} - m_H|$ for the Higgs case, and $|m_{\mu\mu\gamma} - m_Z| < |m_{\mu\mu} - m_Z|$ for the Z^0 case, the invariant mass of the system is calculated for the muon pair plus the selected FSR photon or photons (one associated to each muon). In all other cases the invariant mass is calculated using only the muon pair. These criteria are used in order to select only the photons that could improve the dimuon invariant mass.

5.4 FSR RECOVERY VALIDATION STUDIES

In order to validate the FSR Recovery algorithm, we first tried it with simulated samples for H$\rightarrow \mu^+\mu^-$ and $Z^0 \rightarrow \mu^+\mu^-$, then we tested it with proton-proton collision data collected in 2016, for the case of Z^0 reconstruction. For the case of H$\rightarrow \mu^+\mu^-$ we used the 125 GeV gluon-gluon fusion Higgs sample. In the case of $Z^0 \rightarrow \mu^+\mu^-$, when we used simulated data we took the `ZToMuMu _NNPDF30_-13TeV-powheg_M_50_120` sample.

For the Higgs case we selected the events that pass the b-jet veto and the minimum p_T unprescaled trigger: `HLT_IsoMu24` or `HLT_IsoTkrMu24`. The selection criteria also required a pair of opposite-charged muons, at least one of them having a p_T greater than 26 GeV; the other muon must have a p_T greater than 10 GeV. For the Z^0 case we require a pair of opposite- charged muons with p_T>24 GeV, that pass the minimum p_T unprescaled trigger: `HLT_IsoMu22 or HLT_IsoTkrMu22`.

In this study we measured the percentage of events correctly reconstructed, the efficiency of the FSR recovery method, its purity, and the overall gain.

5.4.1 GEOMETRICAL CATEGORIZATION AND MASS WINDOWS

In order to check the improvement in the mass reconstruction, the events were split in three geometrical categories. The categories are defined by muon pseudorapidity, as shown in Tab 9. These categories are implemented for both, Z^0 and Higgs studies.

Geometrical Category	η value
BB (Barrel)	$\lvert\eta\rvert < 0.8$
OO (Overlap)	$0.8 < \lvert\eta\rvert < 1.6$
EE (Endcap)	$1.6 < \lvert\eta\rvert < 2.4$

Table 9: Geometrical categories divided by η value.

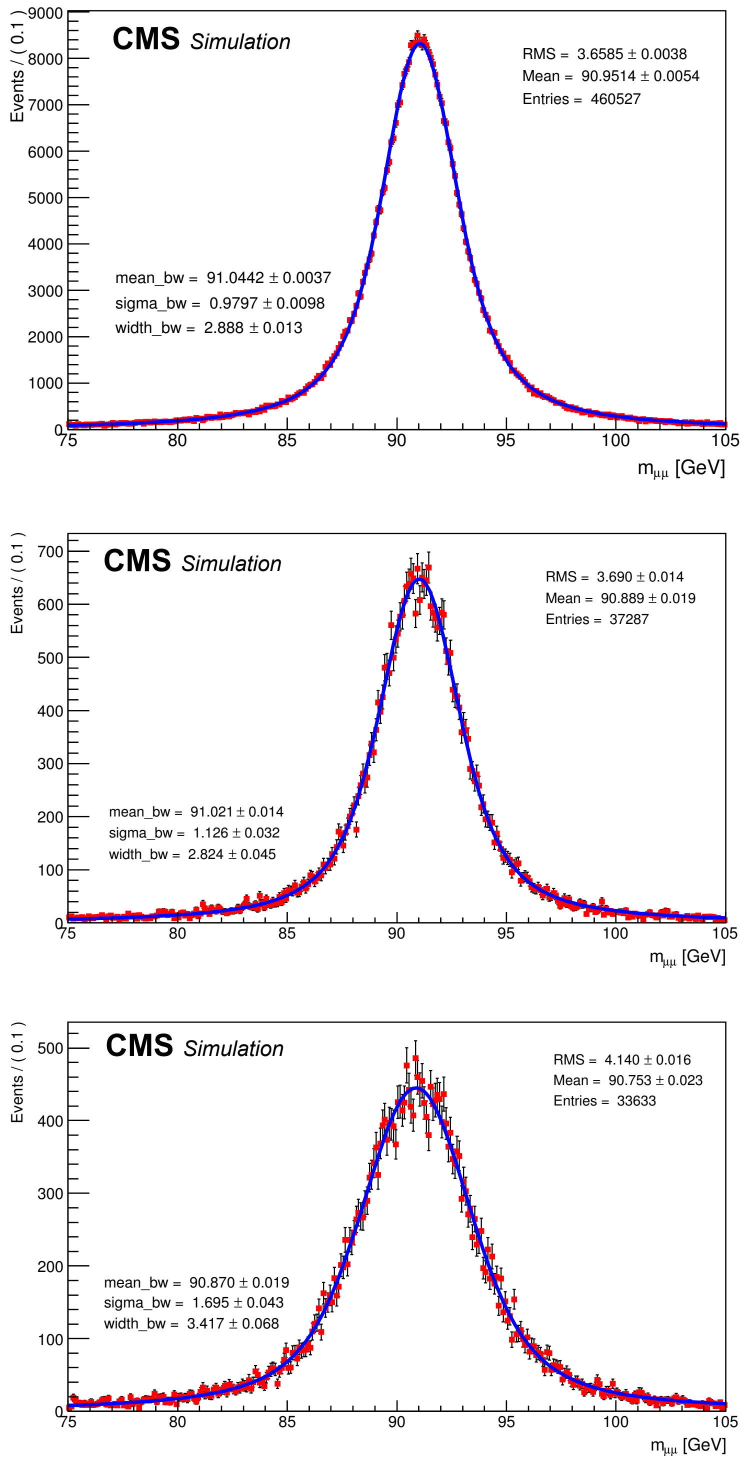

Figure 40: The reconstructed Z^0 mass peak and its fits for: a.) Barrel b.) Overlap c.) Endcap geometrical categories when no FSR photon is emitted. Due to detector resolution, the Breit-Wigner is convolved with a gaussian function in order to obtain a better fit.

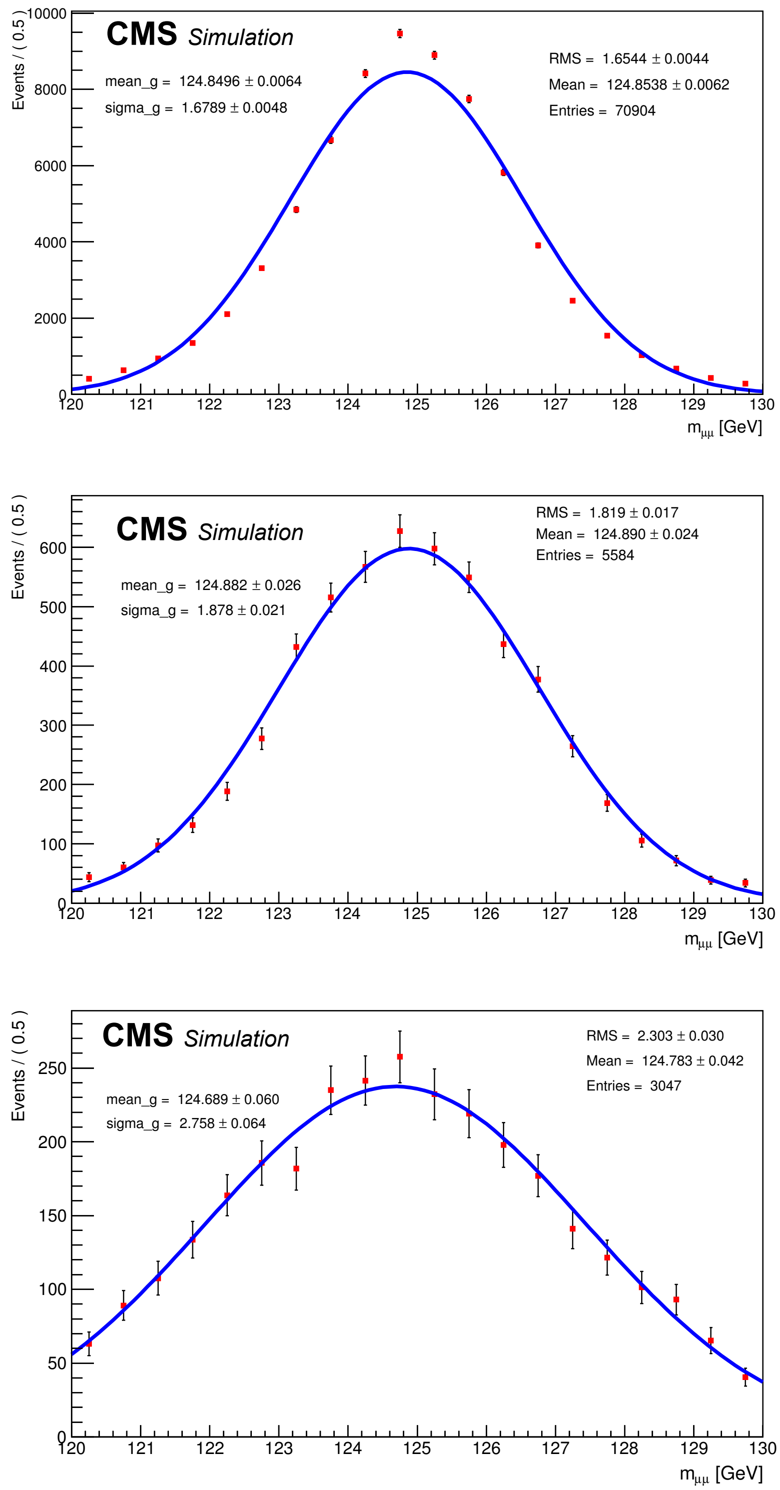

Figure 41: The reconstructed Higgs mass and its fits for: a.) Barrel b.) Overlap c.) Endcap geometrical categories when no FSR photon is emitted.

For each of the previous categories, we look at events that do not have FSR photons at the generation level and plot the mass distribution for the muon pair at the reconstruction level, which gives us a reference to compare with the mass distribution obtained when the algorithm is applied. The mass distribution is fitted with a Breit-Wigner function for Z^0 case, and with a gaussian function for the Higgs case.

For the fit we used the `RooFit` toolkit [79, 80]. In both cases we defined a mass window using the mean value from the fit plus/minus 3 standard deviations. Figure 40 shows the results of the fits for the Z^0 case. From the fits we obtained the mass window for each geometrical category. These mass windows are shown in Table 10. For the Higgs, the fit can be seen in Figure 41. The value of the mass window for each geometrical category is presented in Table 11.

Geometrical Category	mass window (GeV)
BB	$80.5 < m_Z < 101.5$
OO	$80.2 < m_Z < 101.8$
EE	$76 < m_Z < 105$

Table 10: Mass windows for the Z^0 case.

Geometrical Category	mass window (GeV)
BB	$120.0 < m_H < 130.0$
OO	$119.3 < m_H < 130.5$
EE	$116.0 < m_H < 133$

Table 11: Mass windows for Higgs case.

5.4.2 RECONSTRUCTION EFFICIENCY

In order to determine the gain of the FSR recovery algorithm, we need to know how many events are well-reconstructed. Since in Monte Carlo samples we know which particles were generated, we can identify which events have photons that are coming from a muon, so we select and quantify those events. Next, we count the number of dimuons (at reconstruction level) whose invariant mass are within the defined mass windows. Then, we compare these two numbers and obtain the reconstruction efficiency, for events with FSR photons, resulting from the standard CMS reconstruction algorithms.

In other words, the efficiency of the reconstruction can be calculated as:

$$\text{Efficiency} = \frac{\text{\#Events at RecoLevel with a FSR photon at GenLevel}}{\text{\#Events at GenLevel with a FSR photon}} \times 100 \tag{8}$$

Table 12 shows the percentage of events with correct mass reconstruction for the Z^0 and Higgs channels. As we can see from the table, due to the CMS reconstruction algorithm [81], we lose around

3% of the generated events. For the Z^0 case, in all the geometrical categories, the percentage of correctly reconstructed events is around 97%, which is slightly higher than that for the Higgs case.

Geometrical category	Reconstructed events with FSR photon at generation level	Events with a FSR photon at generation level	% correctly reconstructed events
	Z^0 Boson		
BB	24017	24670	97.35 ± 0.10
OO	2229	2276	97.93 ± 0.30
EE	2245	2300	97.60 ± 0.32
	H Boson		
BB	2078	2189	97.25 ± 0.35
OO	2229	269	92.56 ± 1.60
EE	2245	211	91.46 ± 1.92

Table 12: Reconstruction efficiency for Z^0 and Higgs dimuon channels in events with FSR photons.

5.4.3 FSR RECONSTRUCTION EFFICIENCY

From the study presented in the previous section we conclude that we lose around 3% of the FSR events in the reconstruction process. Therefore we need to measure the efficiency of the FSR selection using the other 97% of the events. We count the number of events that have a FSR photon(s) at the reconstruction level, in the defined mass windows, to obtain the efficiency of our FSR selection.

Geometrical category	Reconstructed events with FSR photon identified by the FSR algorithm	Events with an FSR photon at generation level	FSR reconstruction efficiency (%)
	Z^0 Boson		
BB	10085	24670	40.87 ± 0.31
OO	943	2276	41.43 ± 1.03
EE	706	2300	30.69 ± 0.96
	H Boson		
BB	566	2189	25.85 ± 0.94
OO	64	269	23.79 ± 2.60
EE	46	211	21.80 ± 2.84

Table 13: FSR event selection efficiency.

The FSR algorithm efficiency is defined as:

$$\text{FSR Algorithm Efficiency} = \frac{\text{\#Events at RecoLevel identified by the FSR algorithm}}{\text{\#Events at GenLevel with a FSR photon}} \times 100 \quad (9)$$

Table 13 shows this efficiency for Z^0 and Higgs cases. For the Z^0 case the efficiency is around 40%, while for the Higgs case is around 23%. It should be noted that the size of the mass window for the Z^0 is larger than for the Higgs. However, when using a window of similar size, the efficiency for the Z^0 decreases to around 30%.

5.4.4 FSR RECONSTRUCTION PURITY

After using our FSR-candidate selection, it is important to know if these events correspond to photons emitted by muons or by other sources, or if they are the result of misidentification of the photon candidates. In order of determine the FSR purity of the sample, we select the events that have an FSR photon at the reconstruction level, that matches a FSR photon at the generation level, with a $\Delta R < 0.2$. The events at reconstruction level are counted in each of the geometrical categories. The fraction of these events in the sample define the purity of our FSR selection.

This means that the purity can be obtained as:

$$\text{Purity} = \frac{\text{\#Events at RecoLevel with } \Delta R(\gamma_{Reco}, \gamma_{Gen}) < 0.2}{\text{\#Events at RecoLevel identified by the FSR algorithm}} \times 100 \, , \quad (10)$$

Table 14 shows the purity for the Z^0 and Higgs cases. In both cases the average purity is around 85%.

Geometrical category	Reconstructed events with FSR photon (the photon is matched at generation level)	Reconstructed events with FSR photon	FSR reconstruction purity (%)
Z^0 Boson			
BB	8788	10085	87.13 ± 0.33
OO	799	943	84.72 ± 1.17
EE	626	706	88.60 ± 1.20
H Boson			
BB	477	566	84.27 ± 1.53
OO	54	64	84.37 ± 4.54
EE	37	46	80.43 ± 5.85

Table 14: FSR algorithm reconstruction purity. The purity for Z^0 and Higgs cases have an average of 85%.

5.4.5 FSR RECOVERY ALGORITHM GAIN

Many of the events selected by the isolation criteria, vetoes and trigger, among others, are rejected because the dimuon invariant mass is outside the mass window of interest. When we perform the FSR recovery described in Section 5.3, and we include photon information in the process of reconstructing the mass of the dimuon pair, several events that had been rejected are now inside the mass window. These will increase the number of signal events and will also improve the resolution of the mass peak. We called the percentage of extra events *gain* .

If one takes the sample of events for which one knows that the dimuon has FSR photons and compares it to the number of the events in the chosen mass window, after FSR recovery, with the number of events without FSR recovery, one gets what we call *partial gain*. If one compares the number of recovered events with the total number of the events in the sample, that are in the mass window, one gets what we called *overall gain*.

In other words, the *partial gain* can be obtained using the relation:

$$\text{Partial Gain} = \frac{\text{\#Events Reconstructed with FSR photon } (m_{\mu^+\mu^-\gamma}) - \text{\#Events Reconstructed with FSR photon } (m_{\mu^+\mu^-})}{\text{\#Events Reconstructed with FSR photon } (m_{\mu^+\mu^-})} \times 100 \,, \quad (11)$$

while the *overall gain*, is calculated it as:

$$\text{Overall Gain} = \frac{\text{\#Events Reconstructed with FSR photon } (m_{\mu^+\mu^-\gamma})}{\text{\#Events Reconstructed without FSR photon } (m_{\mu^+\mu^-})} \times 100 \,. \quad (12)$$

Geometrical category	Reconstructed events with FSR photon $(\mu^+\mu^-\gamma)$	Reconstructed events with FSR photon $(\mu^+\mu^-)$	FSR partial gain (%)
	Z^0 Boson		
BB	15993	10085	58.58 ± 0.49
OO	1544	943	63.73 ± 1.57
EE	964	706	36.54 ± 1.81
	H Boson		
BB	3874	566	584 ± 22.3
OO	317	64	395 ± 42.7
EE	153	46	232 ± 25.8

Table 15: FSR reconstruction partial gain.

The results for the partial gain for Z^0 are shown in Table 15. Due to the width of the mass windows, for the Higgs case the gain exceeds 100%. Nevertheless, if we use a similar width for the Z^0 case, the gain could reach around 200% in average. However, although the partial gain is large, we are interested in knowing what extra percentage of the events can be recovered. Table 16 shows the results of the overall gain for our Z^0 and Higgs samples: around 3% for the Z^0 and 4% for Higgs.

Geometrical category	Reconstructed events with FSR photon ($\mu^+\mu^-\gamma$)	Reconstructed events ($\mu^+\mu^-$)	FSR overall gain (%)
Z^0 Boson			
BB	15993	532694	3.00 ± 0.02
OO	1544	43635	3.53 ± 0.09
EE	964	40077	2.40 ± 0.08
H Boson			
BB	3874	81764	4.73 ± 0.07
OO	317	6624	4.78 ± 0.26
EE	153	4151	3.68 ± 0.29

Table 16: FSR reconstruction overall gain.

5.4.6 FSR RECOVERY ALGORITHM APPLIED TO DATA FOR THE Z^0 CASE

In order to validate our FSR recovery algorithm with pp collision data we focused on the Z^0 peak, as we mentioned before. In this case, in order to obtain the overall gain it is necessary to take into account all the relevant background processes for the channel $Z^0 \rightarrow \mu^+\mu^-$. We have identified five possible sources of background:

- DY+Jets to ll,
- TT+Jets,
- WW to 2l 2ν,
- ZZ to 4l,
- ZZ to 2l 2ν.

Monte Carlo samples	Number of events	Cross section (pb)	(%) Event yield
`DYJetsToLL_M-50_TuneCUETP8_M1_13TeV-madgraphMLM-pythia8`	48833528	5767	73.575
`TTJets_DiLept_TuneCUETP8M1_13TeV-madgraphMLM-pythia8`	6049044	85.656	1.091
`WWTo2L2Nu_13TeVpowheg`	1986222	12.46	0.159
`ZZTo4L_13TeVamcatnloFXFX-pythia8`	10644072	1.212	0.007
`ZZTo2L2Nu_13TeV_powheg_pythia8`	8777448	0.564	0.015

Table 17: Main background processes for Z^0 dimuon decay channel

In order to model signal and background in the dimuon invariant mass distribution, we used the Monte Carlo simulation samples listed in Table 17.

Due to the small event yield of the diboson background channels, as can be seen in Table 17, we modeled the background in the dimuon mass distribution using only the DYJetsToLL_M-50 and TTJets_DiLept Monte Carlo samples.

In order to model the data we used the following functions, which we will discuss later:

$$\mathrm{BW}(x; x_0, \gamma) = \frac{1}{\pi\gamma\left[1 + \left(\frac{x - x_0}{\gamma}\right)^2\right]} \tag{13}$$

$$\mathrm{Gauss}(\mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{14}$$

$$\mathrm{Landau}(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} e^{s\log s + xs}\, ds \tag{15}$$

The DY Jets sample has Z^0 as an intermediary state, contributing to the signal and the background. Therefore, in order to describe this channel we used a function that is a sum of a Breit-Wigner for the signal and a Gaussian for the background:

$$\mathrm{DYJets}_{\mathrm{fit}} = f * \mathrm{BW}(x; x_0, \gamma) + (1 - f)(\mathrm{Gauss}_{\mathrm{DYJets}}(\mu_{\mathrm{DYJets}}, \sigma_{\mathrm{DYJets}})). \tag{16}$$

The results of the fits to the corresponding Monte Carlo sample are presented in Figure 42.

For the TTJets case, this background source was modeled using a Landau function convolved with a Gaussian function:

$$\mathrm{TTJets}_{\mathrm{fit}} = \mathrm{Landau}_{\mathrm{TTJets}}(x) * (\mathrm{Gauss}_{\mathrm{TTJets}}(\mu_{\mathrm{TTJets}}, \sigma_{\mathrm{TTJets}})). \tag{17}$$

The results of the fits to the corresponding Monte Carlo sample are presented in Figure 43.

In order to model the pp collision data, we ran our FSR algorithm over the `PromptReco-2016H` dataset. This sample corresponds to a small fraction of the pp collision data collected by the CMS experiment at 13 TeV. After modelling the background sources (as shown in Figures 42 and 43), we performed a fit over the dimuon mass distribution obtained with this data. The signal was modeled using a Breit-Wigner distribution, while the background was modeled using a sum of a Gaussian (DYJets) plus a Landau convolved with a Gaussian (TTJets):

$$\mathrm{Signal} = \mathrm{BW}(x; x_0, \gamma) \tag{18}$$

$$\mathrm{Background} = \mathrm{TTJets}_{\mathrm{fit}} + \mathrm{Gauss}_{\mathrm{DYJets}}(\mu_{\mathrm{DYJets}}, \sigma_{\mathrm{DYJets}}) \tag{19}$$

$$\mathrm{Data}_{\mathrm{fit}} = f * \mathrm{Signal} + (1 - f)(\mathrm{Background}) \tag{20}$$

The procedure was applied to all the geometrical categories presented in Table 9. We considered the dimuon mass distribution when no FSR photons are emitted (the results are presented in Figure 44). We also considered all the events in the dimuon-plus-FSR-photon mass distribution (the results are shown in Figure 45). For both cases we can obtain the fraction of signal events from the fit functions; these results are shown in Table 18.

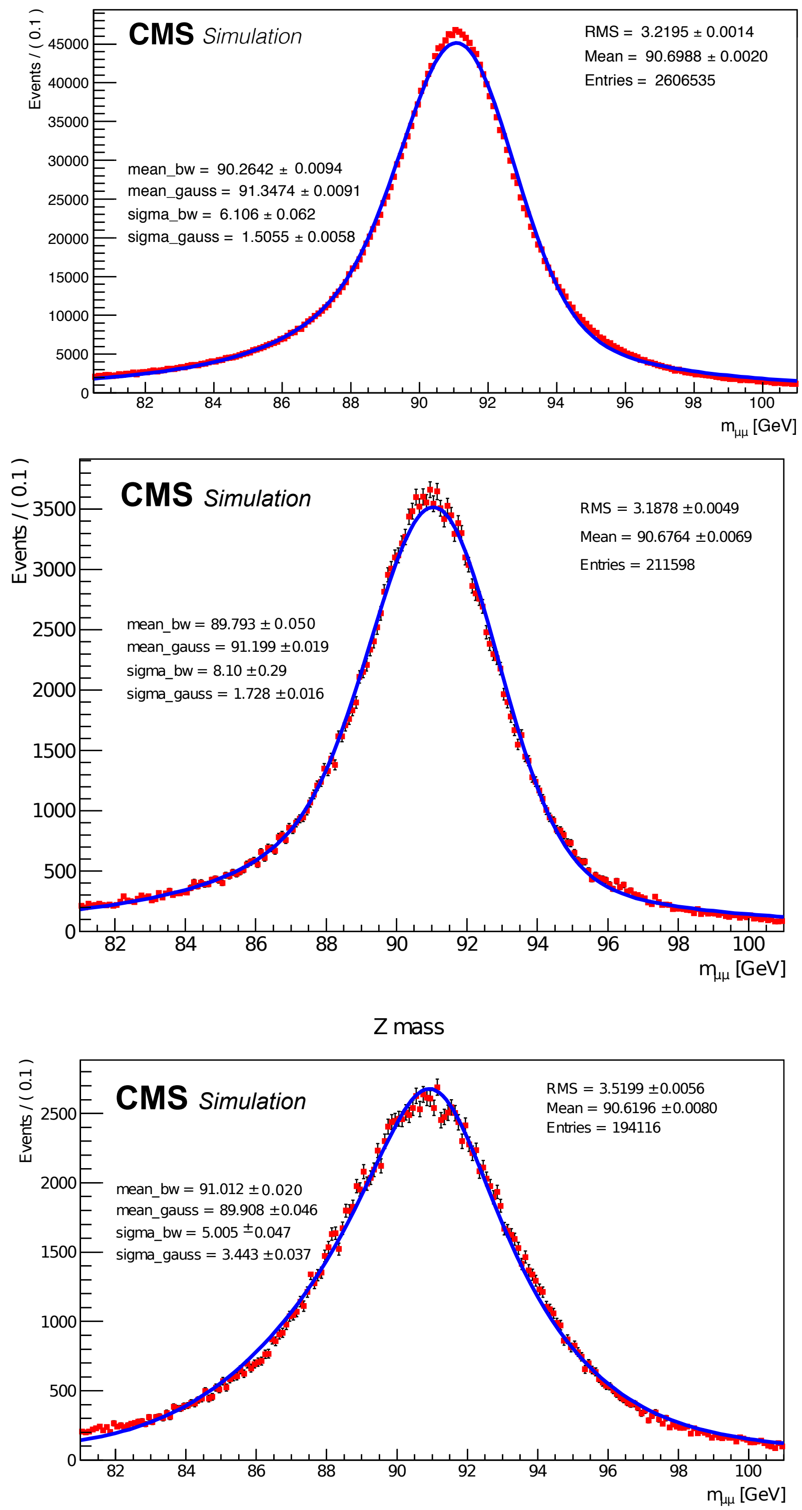

Figure 42: The reconstructed dimuon mass peak and its fits for the DY+Jets sample: a.) Barrel b.) Overlap c.) Endcap geometrical categories when no FSR photon is emitted.

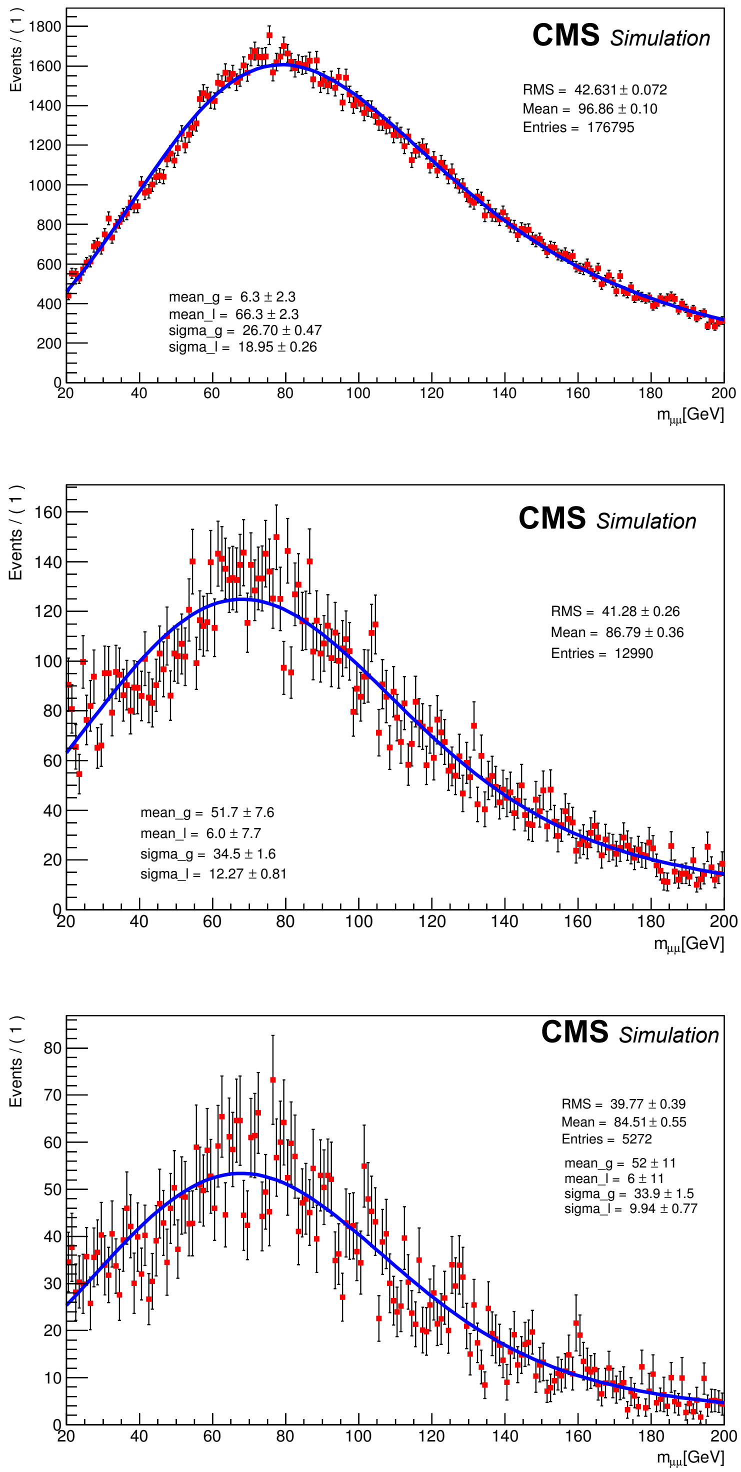

Figure 43: The reconstructed dimuon mass peak and its fits for the TT+Jets sample: a.) Barrel b.) Overlap c.) Endcap geometrical categories when no FSR photon is emitted. The dimuon mass shape is given by the two muons from the W-boson decays in $t\bar{t}$ events.

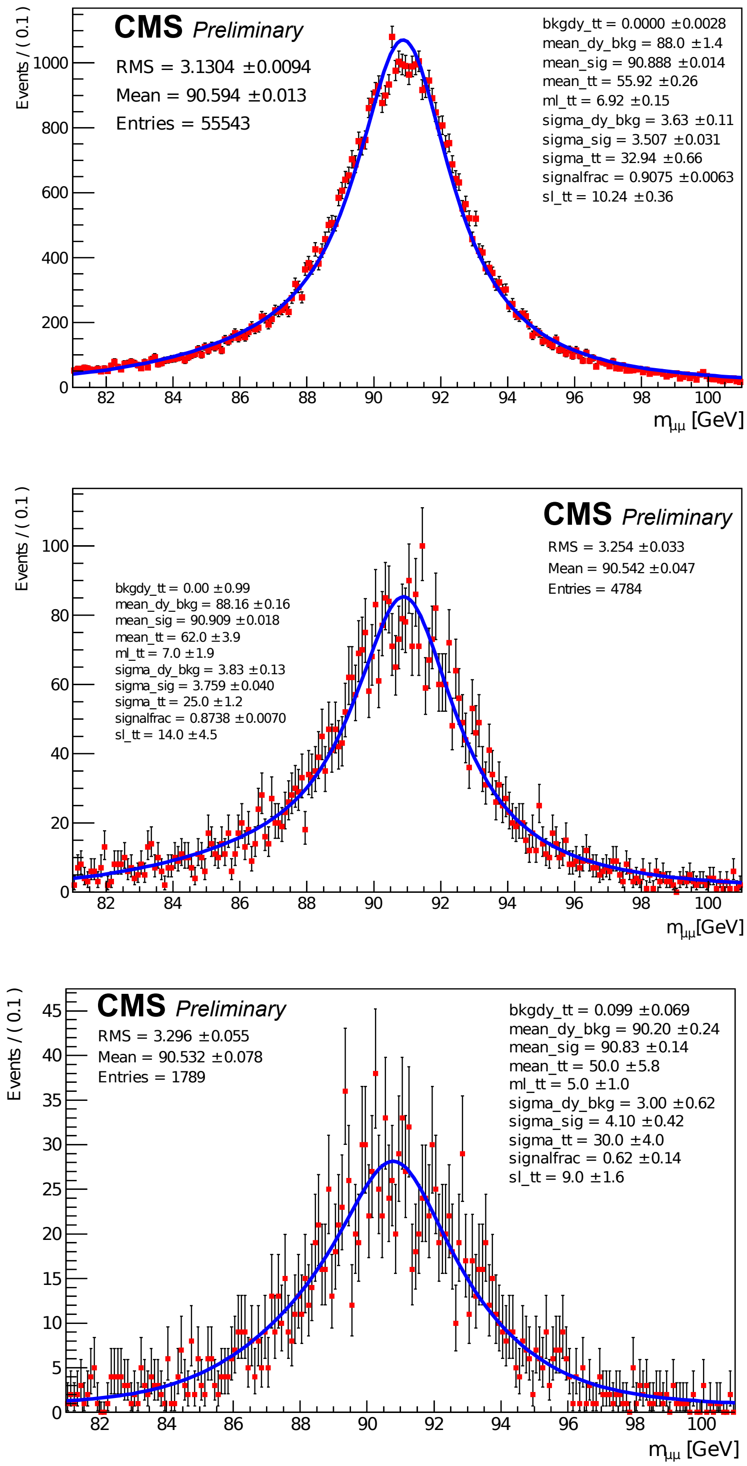

Figure 44: The reconstructed dimuon mass peak and the fit using the equation 20 for pp collisions data: a.) Barrel, b.) Overlap, c.) Endcap, geometrical categories when no FSR photon is emitted.

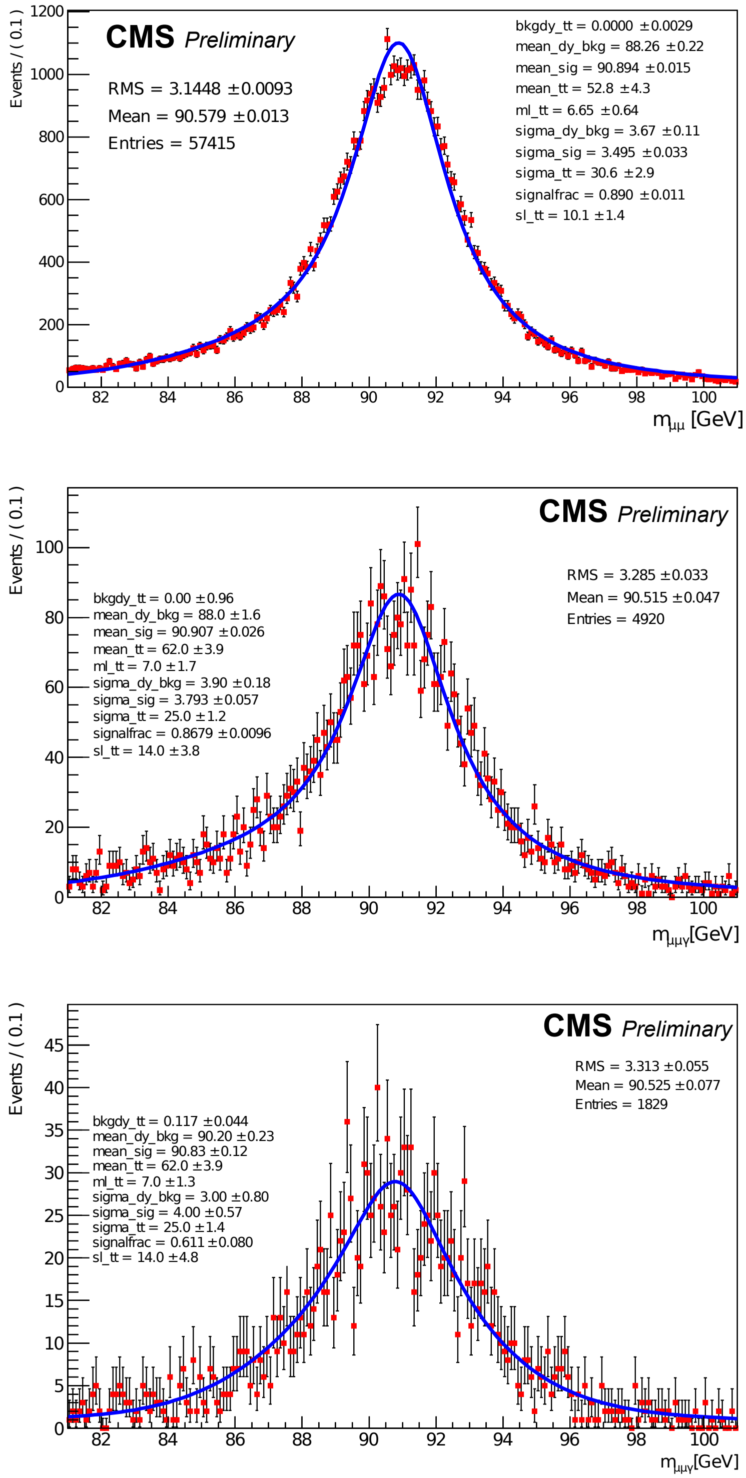

Figure 45: The reconstructed dimuon mass plus photon peak and the fit using the equation 20 for pp collisions data: a.) Barrel, b.) Overlap, c.) Endcap, geometrical categories with and without emitted FSR photons.

Geometrical Category	Number of events in mass distribution	Signal fraction	Number of signal events
Events when no FSR photon is emitted			
BB	55543	0.9	49988
OO	4784	0.8738	4180
EE	1789	0.62	1109
Events with FSR photon			
BB	57415	0.89	51099
OO	4920	0.87	4280
EEs	1829	0.62	1133

Table 18: Z^0 Signal events over pp collision data collected at 13 TeV

Geometrical category	Overall Gain (%)
BB	2.1 ± 0.06
OO	2.34 ± 0.23
EE	2.18 ± 0.44

Table 19: Overall gain for a Z^0 signal events in pp collision data collected at 13 TeV

After obtaining the number of signal events, we can get the *overall gain*. The results are presented in Table 19. The comparison with a gain using simulated events, presented in Table 16, shows that the results using collision data are slightly smaller. In two of the three categories (BB and EE) the implementation of the FSR algorithm also improves the resolution of the dimuon mass distribution.

After these validation studies, the FSR algorithm gave us a gain of around 2% for the Z^0 case. In most of the cases considered, the recovery, not only increased the number of signal events, but also improved the resolution of the dimuon mass distribution.

The studies show that for the H→ $\mu^+\mu^-$ channel we expected a gain of the order of 4%, that would improve the chance to observe the signal. The results of the FSR Recovery procedure applied over the Higgs dedicated categories will be shown in Chapter 7.

6

MULTIVARIATE ANALYSIS METHOD FOR FSR RECOVERY

After studying the FSR recovery using the cut-based approach, we explored the recovery of the events using a multivariate technique. The main goal of using a multivariate approach is to improve the percentage of recovered events as compared to the cut-based technique. In this chapter the procedures for training and testing of a multivariate analysis (MVA) classifier used in FSR recovery will be presented. The chapter starts with a description about decision trees and boosting. Afterwards, the toolkit for multivariate analysis (TMVA) will be described. Next, the choice of the relevant variables used in the training of the MVA will be discussed. Finally, working points obtained for the implementation of the FSR dimuon recovery will be presented. The results of the implementation of the MVA in the FSR dimuon recovery for the Higgs reconstruction will be discussed in Chapter 7.

6.1 DECISION TREES

Decision trees are classifiers with a binary structure. They sequentially cut on discrimination variables until a predefined criterion is achieved. An event is processed through the tree yielding a response for signal or for background. A scheme of how a decision tree works can be seen in the Figure 46 [82].

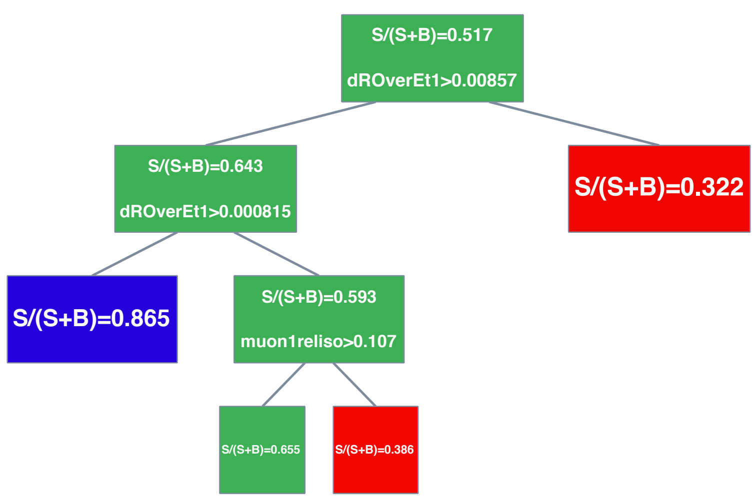

Figure 46: Example of a decision tree showing intermediate nodes (green boxes), signal (blue box) and background (red boxes) leafs. The variables used stem from the H→ $\mu^+\mu^-$ analysis.

Decision trees have a straightforward interpretation, they can be visualized as a simple two-dimensional tree structure. They are in this respect similar to rectangular cuts. Via this method the space of discrimination variables is split in many rectangular regions, which are labeled as signal-like or background-like. The trees grow as they are trained in a sequential manner, i.e. starting with the root node (*mother node*) a new node is added to the tree in each iteration. The cut value and discrimination variable at each node are chosen such that they provide the best separation between signal and background, i.e. the next node will be chosen so that the selected variable splits the training dataset into two groups: signal events and background events. Each node on the tree sets a threshold on the evaluated variable. The leafs, i.e. the final nodes (*daughter nodes*), determine wether the event is signal or background, and they are labeled according to their purity. The purity is calculated as: $p = S/(S+B)$. When $p > 0.5$ the event is considered signal, otherwise it is considered background. The quality of separation is defined by the *impurity function*. In our case, this function is called *Gini Index*: $p \cdot (1-p)$.

Since a single tree has a limited depth, it is not enough to discriminate signal from background for all the possible events in the training set. If a tree is too deep, it will adjust too much to the training set, limiting its usability for other inputs. In order to avoid this, a set of shallower trees (*a forest*) is trained. Each tree in the forest is trained with a subset of the training set and the final classification is given by a majority vote over all these trees.

6.2 BOOSTING

Decision trees are sensitive to statistical fluctuations of the training sample. In order to make decision trees more robust, we use boosting. When training the forest, instead of training each tree independently we use the output of each tree to train the following one. Figure 47 shows a schematics diagram of a boosting process. The main purpose of using boosting is to combine a collection of weak classifiers into one that is strong. TMVA provides three different boosting types: adaptive boosting, gradient boosting and bagging.

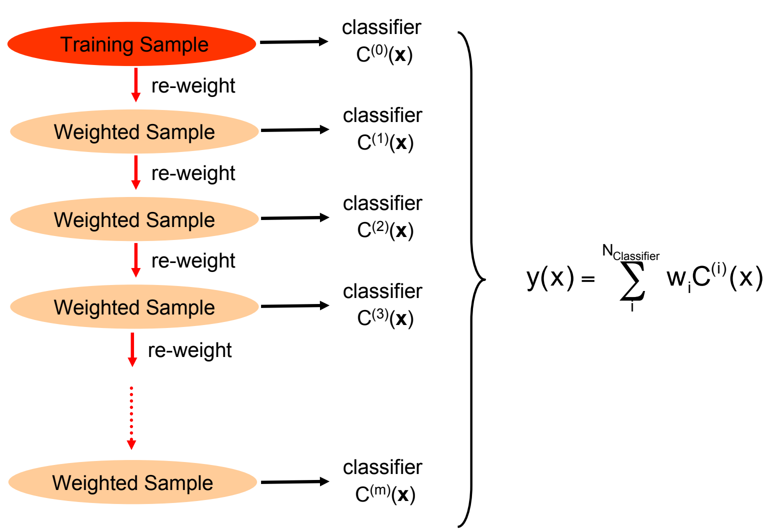

Figure 47: Schematics of a boosting method [83].

6.2.1 ADAPTIVE BOOST

Adaptive boosting, called *AdaBoost*, in TMVA, is the most popular boosting algorithm. This method rates a strong learner by training weak learners sequentially in a way that they "learn" from the errors of the previously grown trees. The events that were misclassified by the current tree will have a higher weight when training the next one, ensuring that this new tree learns how to better classify these events. It starts with the original event weights when training the first decision tree. The subsequent tree is trained using a modified event sample, where the events are weighted by a common *boost weight* α. The boost weight is derived from the misclassified events fraction f_1, of the previous tree: $\alpha = \frac{1-f_1}{f_1}$. The weights of the entire sample are reweighted and renormalized so that the sum of the weights remain constant. The deviation between the true value and the value predicted by the set of currently trained trees is measured by the *loss-function*. The boosting should minimize this deviation during training.

AdaBoost performs well on weak classifiers, meaning small individual decision trees with a tree depth as small as 2 or 3, that have very little discrimination power by themselves. When we use small trees as an ensemble, they are much less prone to overtraining compared to simple decision trees. The performance usually improves when the training is forced to a "slow-learning" and a large number of boost steps are used in the method.

6.2.2 GRADIENT BOOST

A Gradient Boost produces a prediction model in the form of an ensemble of weak prediction models; typically decision trees. The boosting procedure is now employed to adjust the parameters such that the deviation between the model response and the true value obtained from the training sample is minimized.

AdaBoost is based on an exponential loss function. One shortcoming of using this loss function is that it degrades in the presence of outliers or mislabelled data points. The GradientBoost algorithm allows using other, potentially more robust, loss functions. However, it requires much more tuning during the development of the classifier.

Gradient Boost works well on weak classifiers. This means small individual decision trees with a depth of usually 2 to 4. It can be robust if the learning rate is reduced through the weight of individual decision trees. Small weight implies more trees to be grown.

6.2.3 BAGGING

Bootstrap aggregating, also called *bagging*, is a technique where a classifier is repeatedly trained. To do this it is necessary to use training events previously sampled, such that the combined classifier represents an average of the individual classifiers. The bagging technique is not a boosting method in the strict sense. Bagging does not improve a weak classifier in the way that adaptive or gradient boosting does.

For bagging there exists the possibility of replacement in the resampling. Meaning that the same event could be picked several times from the initial sample. Training several classifiers with different resampled training events, and combining them, results in an averaged classifier that is more stable with respect to statistical fluctuations in the training sample.

6.3 OVERTRAINING AND PRUNING

When classifiers are too adapted to a specific training sample, similar testing samples can have a completely different response from the trained classifier. This also happens if the training was performed using too few degrees of freedom, because too many parameters of the model were adjusted to too few data points. Moreover, the same behavior appears when the training is done with inappropriate statistics. In all these cases we are talking about overtraining. Overtraining is problematic, as real data will be classified differently from the training sample and the BDT output distribution can not be used to determine the optimal cut value. A convenient way to detect overtraining, and to measure its impact, is by comparing the performance results between training and test samples.

An efficient solution to combat overtraining is pruning. Pruning is the process of cutting back a tree from the bottom up, after it has been built to its maximum size. The purpose of this method is to remove statistically insignificant nodes, thus reducing the overtraining. For decision trees the pruning starts by first growing the tree to its maximum size and then cutting it back, rather than interrupting the node splitting at an earlier stage. This is very beneficial, as compared to limiting the tree depth, because apparently insignificant splits can nevertheless lead to good splits further down the tree. For Boosted Decision trees, since the boosting performs best on weak classifiers, pruning is unnecessary.

6.4 TOOLKIT FOR MULTIVARIATE ANALYSIS (TMVA)

The toolkit for Multivariate analysis (TMVA) gives an environment for the processing, evaluation and application of multivariate classification and regression techniques for ROOT users. The TMVA was designed to supply the need of high-energy physics applications, but is open to other fields. The TMVA includes many algorithms for training, testing and classifications, like linear and nonlinear discriminants, artificial neural networks, and boosted/bagged decision tress, among others [82].

The TMVA software is based on an object-oriented implementations in C++ and ROOT, for each multivariate analysis method. Training and testing is performed using ROOT trees or text files, each event can have its own weight. The linear correlation of the input variables is not only used in the calculations, it is also shown using visualization scripts (the majority of the plots presented in this chapter are made using these scripts). A preliminary ranking is derived for the input variables used in the MVA. Variables overlap and classification problems may emerged, in order to reduce this, the variables can be linearly transformed to obtain a set of non-correlated ones.

In order to develop a subsequent analysis, TMVA provides the training results in a readable XML format. These files are know as *weight files*.

6.4.1 THE TMVA FACTORY AND READER

In order to start the training phase, for each event in the sample, the set of input variables is saved. TMVA divides the process in training and testing stages that are coordinated by the TMVA Factory, and an analysis stage in which we use the TMVA Reader to get the MVA value. Figure 48 shows a flow of a typical TMVA training and analysis application.

The *Factory* organizes the user interaction with the TMVA modules. First, it registers all the discriminating variables. Then, it selects the MVA method, e.g. BDT. Afterwards, it calls the TMVA training, testing and performance evaluation methods from the *Factory*. The input events are internally copied and split into one training and one test ROOT tree. This guarantees a statistically independent evaluation of the MVA algorithms based on the test sample. Finally, the training results are stored in XML files and the relevant histograms are saved in a ROOT output file.

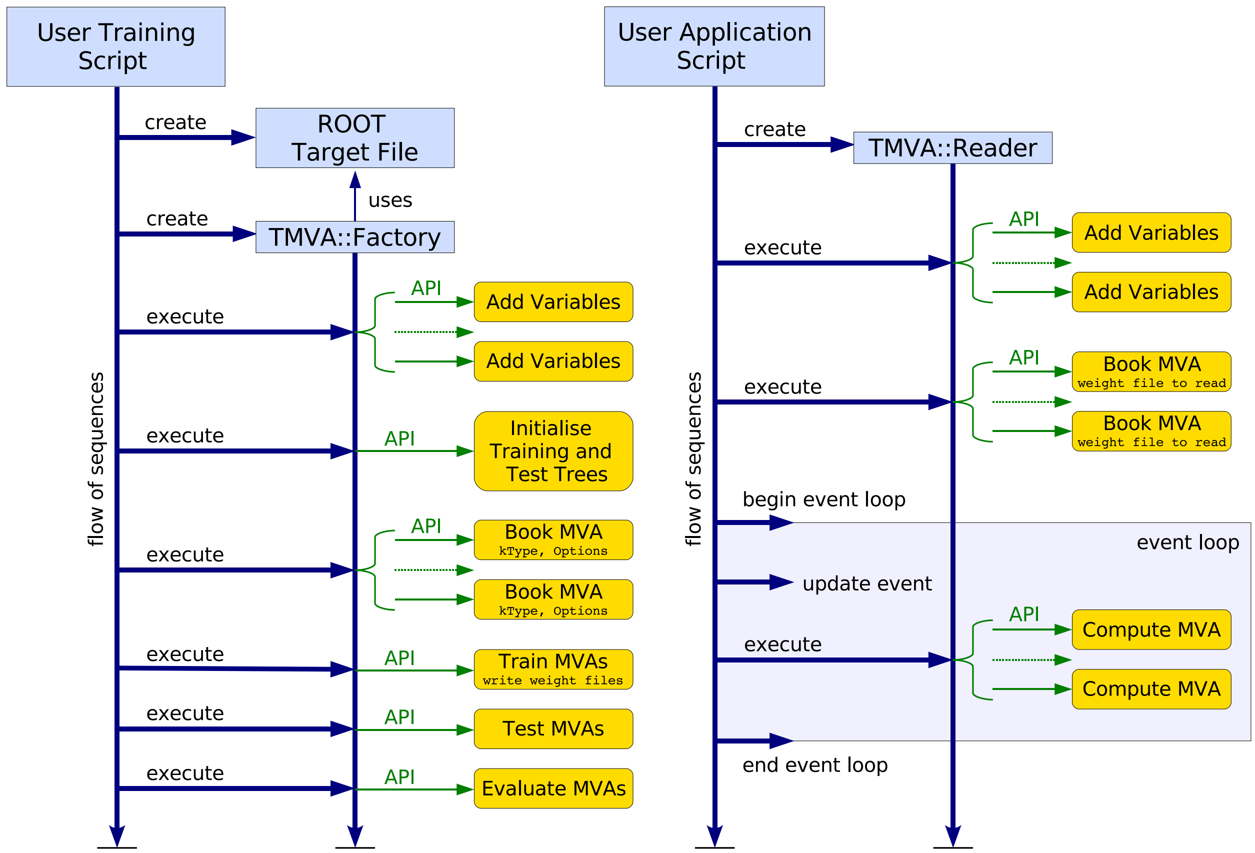

Figure 48: Flow of a typical TMVA training application (Left) and flow of a typical TMVA analysis application (Right) [82].

After training and testing, the linear correlation coefficients among the classifier outputs are printed. These values aim to discriminate between the relevant input variables and those that are not. Also, overlap matrices are calculated for signal and background. These matrices help to determine the fractions of signal and background events that are equally tagged by the classifier.

The MVA methods are chosen and used to classify events in data samples with unknown signal and background composition. Similar to the *Factory*, the communication between the user application and the MVA methods is done using the TMVA *Reader* class, which is instanced by the user. The *Reader* evaluates and gives an MVA value for each event. The same variables used in training must be input when using the Reader class to calculate the MVA value.

6.5 TRAINING AND TESTING PROCEDURES

In order to start training over the chosen multivariate analysis method, it is necessary to select a set of relevant variables. These variables must be present for each of the events in the sample. The training sample has a set of events labeled as signal or background. In order to recover dimuon events for the reconstruction of Higgs using FSR photons, we implemented a BDT and trained it to discriminate the real FSR photons from misidentified photos (mostly a product of the reconstruction process) using the cut-based results as training data and testing. The training was performed using gluon fusion H→ $\mu^+\mu^-$ Monte Carlo samples for different Higgs masses. The kinematical variables used are:

- Pseudorapidity for leading muon (η_{μ_1}) and sub-leading muon (η_{μ_2}).
- Maximum value for pseudorapidity of muons, Max($|\eta_{\mu_1}|, |\eta_{\mu_2}|$).
- Transverse momentum for each muon, $p_{\mathrm{T}\mu_1}$ and $p_{\mathrm{T}\mu_2}$.
- Relative isolation for each muon.
- Relative isolation for each muon when the photon is excluded.
- $\Delta\eta(\mu_1, \mu_2)$.
- Transverse momentum, pseudorapidity and PF Relative isolation for photon.
- $\Delta R(\gamma, \mu)$ and $\Delta R(\gamma, \mu)/E_{\mathrm{T}}{}^2$.
- Number of jets.
- Minimum $\Delta R(jet, \gamma)$.
- Number of electrons.
- Minimum $\Delta R(e, \gamma)$.

To discriminate between true and false FSR events among those that were classified by the cut-based criteria as true events, we focused on the FSR photons at generation level. To decide whether the events are label as signal or background, we matched the FSR photons in the reconstruction stage to the generated FSR photons in the Monte Carlo sample, using a maximum ΔR criterion. If the photon at reconstruction level matched to the generated FSR one, the events were considered signal. On the contrary, all the events in which the reconstructed FSR photon did not match a generated photon, or in which only photons at reconstruction level existed, were considered background.

After some training iterations and testing, using different parameters in the algorithm, we selected the number of trees to train in the forest, and the depth of each tree to obtain the best significance on the "Receiver Operation Characteristics" (ROC) curves, and therefore best classification performance on the testing dataset. We also selected the number of signal and background events for the training, in order not to not have either overtraining nor undertraining. To assure a correctly trained model, it is necessary to compare the performance results when the trained classifier is applied to different sets of training and test samples, for signal and background events. If the algorithm performance is good on training samples but degrades significantly on test samples, the forest is too adjusted to the values of the variables in the training sample, and thus overtrained. If the forest performance is not much better than random guessing on both samples, the forest has not inferred sufficient knowledge from the training set and is undertrained.

Taking into account previous considerations, we used 350 trees for training with a depth of 3 for each tree. Figure 49 shows the BDT response for our training and testing samples. The agreement between the training and testing samples with the set of input variables is quite good.

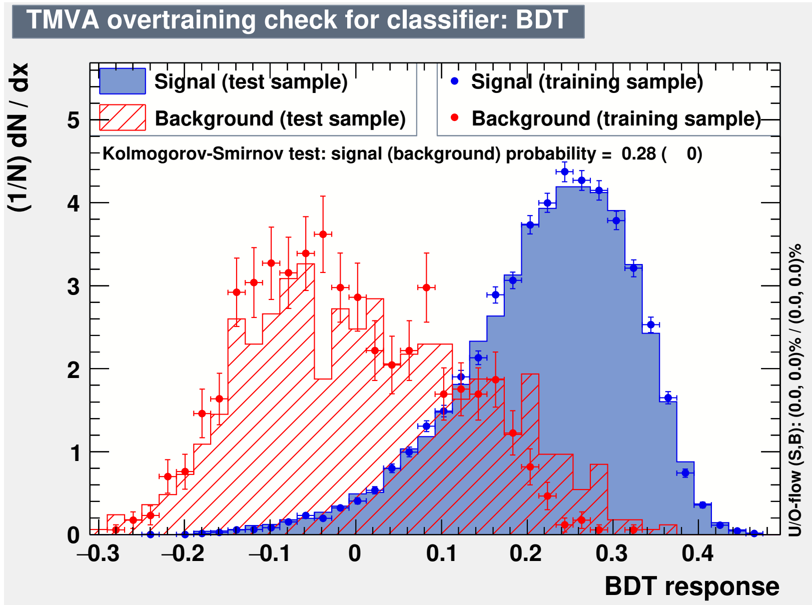

Figure 49: BDT Response for our training and testing procedures. This plot is used to check the overtraining for the classifier. If the training (dots) and testing (histogram) samples match or are very close between them, the BDT is not overtrained.

To avoid overtraining, a parameter called `nCuts`, whose standard value is 20, is used. This parameter limits the number of times the tree is grown and provides a good compromise between time of calculation and size of steps [82].

After fixing all the parameters, we checked the linear correlations of the input variables in order to select the minimum number or relevant variables for the final training. The linear correlations are shown in Figure 50 for signal and Figure 51 for background in the training samples.

Additionally, a ranking over input variables is obtained by counting how often the variables are used to split decision tree nodes. This counting is done weighting each split occurrence by the separation-gain-squared achieved, and by the number of events in the node. The ranking for the input variables used in the analysis are shown in Table 20.

Based on correlation matrices and variable ranking, we discarded some input variables. Figures 50 and 51, show that the variables *Number of electrons* and the *Minimum* $\Delta R(e, \gamma)$ are highly correlated: 95 % for signal and 93% for background. We discarded the *Number of electrons* variable due to its low ranking. Another set of highly correlated variables is *Max*$(|\eta_{\mu_1}|, |\eta_{\mu_2}|)$ with η_{μ_1} (73% for signal and background) and with η_{μ_2} (73% for signal and 71% for background). In this case, we discarded *Max*$(|\eta_{\mu_1}|, |\eta_{\mu_2}|)$, since it was correlated with two other variables, that are not so correlated among themselves.

With the new set of variables linear correlations greater than 80% disappear in the signal and in the background training samples. This modifies the ranking of the variables. In order to know how much the MVA method differentiates signal from background we look for a BDT that maximizes the area under the ROC curve. The test accuracy can also be obtained as the area under the curve. The greater the area under the curve the more accurate the test. A perfect test has an area under the ROC curve

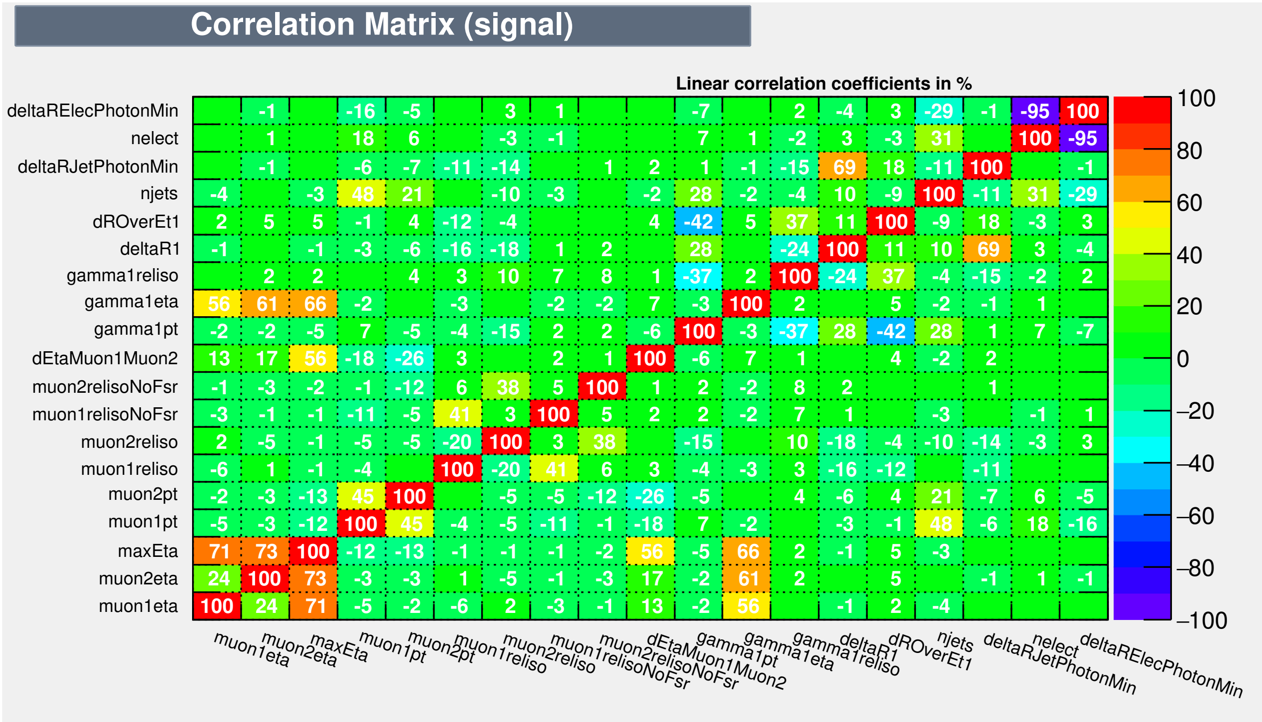

Figure 50: Linear correlation matrix for input variables in the case of the signal training sample.

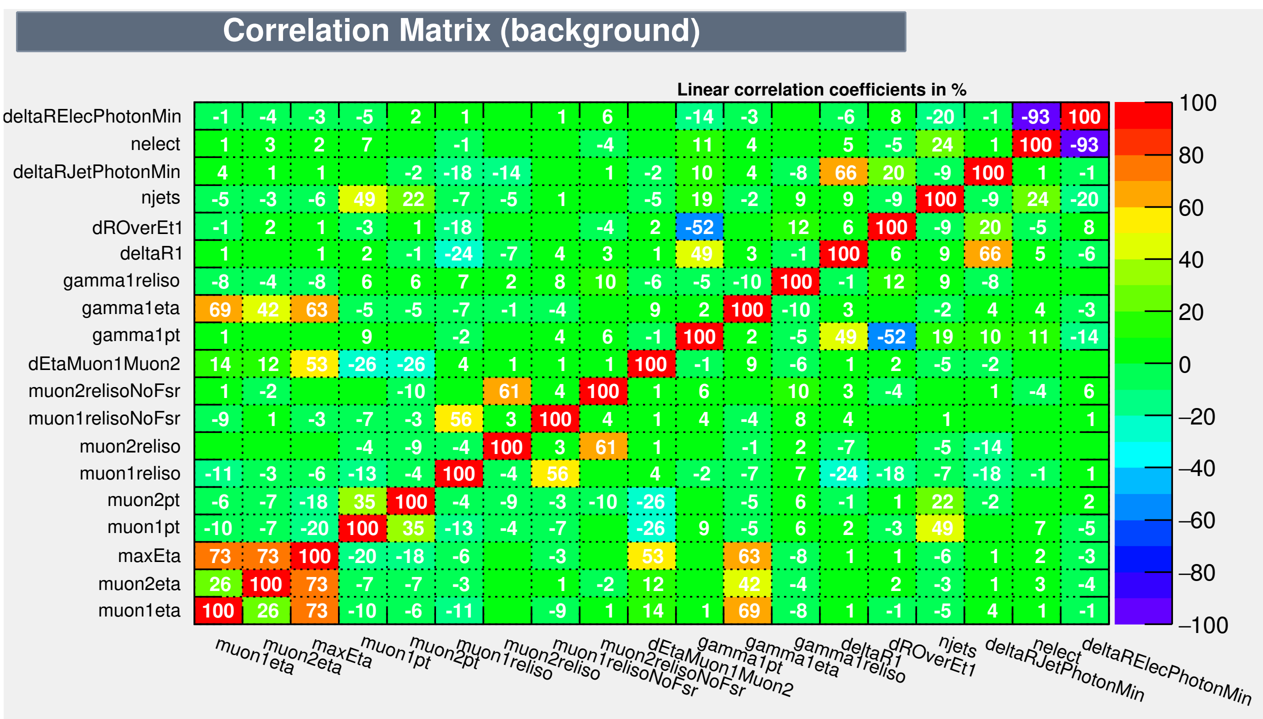

Figure 51: Linear correlation matrix for input variables in the case of the background training sample.

Rank	Variable	Variable Importance
1	$\Delta R/E_T{}^2$	7.830 e-02
2	γ Relative Isolation	7.647e-02
3	$\Delta R(\mu,\gamma)$	7.171e-02
4	η for photon	6.760e-02
5	η for sub-leading muon	6.566e-02
6	η for leading muon	6.483e-02
7	$\Delta\eta(\mu_1,\mu_2)$	6.158e-02
8	Sub-leading muon Relative Isolation	6.143e-02
9	Leading muon Relative Isolation	6.032e-02
10	Max($\lvert\eta_{\mu_1}\rvert, \lvert\eta_{\mu_2}\rvert$)	5.778e-02
11	p_T for sub-leading muon	5.076e-02
12	Relative isolation for sub-leading muon when the photon is excluded	5.014e-02
13	Relative isolation for leading muon when the photon is excluded	4.608e-02
14	p_T for leading muon	4.488e-02
15	Minimum $\Delta R(jet,\gamma)$	4.200e-02
16	p_T for photon	3.934e-02
17	Number of jets	3.079e-02
18	Minimum $\Delta R(e,\gamma)$	2.735e-02
19	Number of electrons	2.988e-03

Table 20: Ranking for the original set of input variables for our BDT.

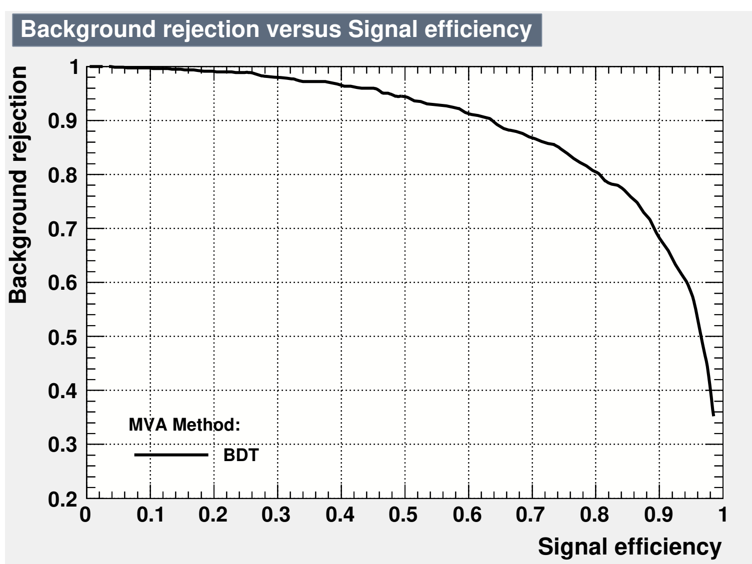

Figure 52: Background rejection versus signal efficiency ("ROC curve").

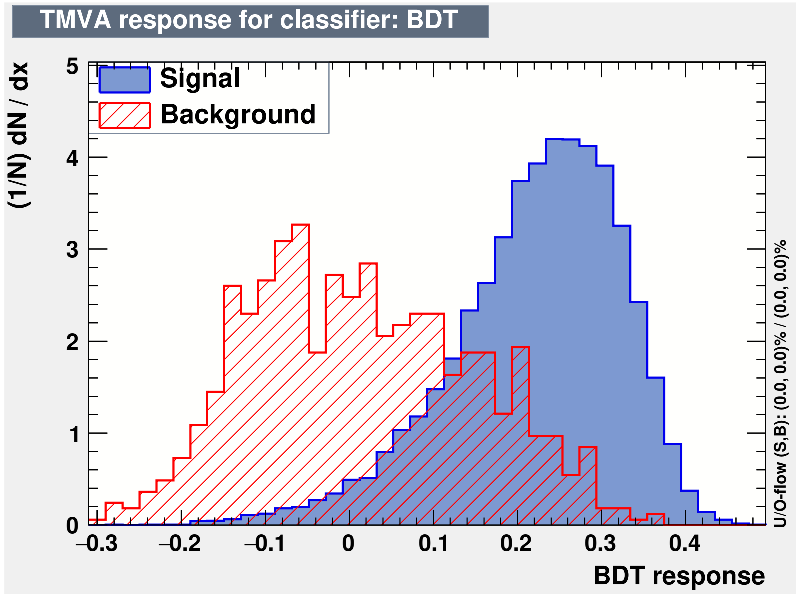

Figure 53: MVA Response for our BDT classifier

of 1, the closer the graph is to the top and right-hand borders, the more accurate the test. Likewise, the closer the graph to the diagonal, the less accurate the test. A good ROC curve gives us an optimal working point for the analysis. Figure 52 shows the ROC curve for our BDT. In our case, the integral of the ROC curve is 0.867.

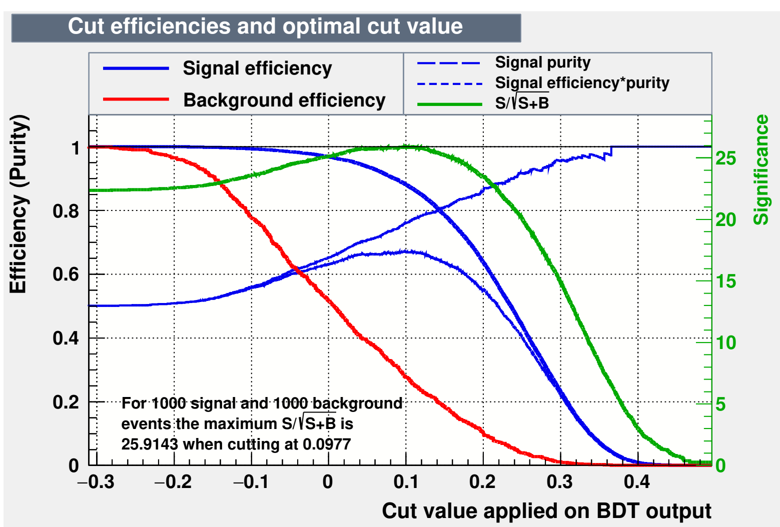

Figure 54: Cut efficiencies for signal and background.

To obtain an optimal working point, we used the BDT response plot. The MVA value that allows to best distinguish between signal and background, is considered a candidate for the working point value. Figure 53 shows the BDT response. A preliminary working point value was chosen to be 0.11. This value permits to discriminate between signal and background. Nevertheless, for a good choice of the working point it is necessary to take into account the efficiency of the MVA value. In Figure 54 we can see the signal and background efficiency as well as the significance $S/\sqrt{S+B}$. In our BDT the working point is 0.0977. This value is more precise but similar to the one that we chose previously with the BDT Response.

If the event has a MVA value greater than the working point it will be identified as signal, else it will be identified as background. The results of the FSR recovery using the MVA value discriminator over Higgs samples will be discussed in Chapter 7, together with the results of the cut-based analysis.

7

RESULTS

In this chapter the main results of this research will be presented stating with the two FSR recovery techniques implemented in order to improve the signal-to-background ratio of the H→ $\mu^+\mu^-$ channel search. In the first section the results of the cut-based technique for FSR Recovery are presented and discussed. In the second section, a similar discussion is presented for the FSR recovery using the MVA technique. These two techniques are discussed in detail in chapters 5 and 6. In the third section, the results of the search for Higgs decaying into muons using the data collected in the second run, at 13 TeV, corresponding to an integrated luminosity of 35.9 fb^{-1}, are presented. This analysis was performed by the CMS H2Mu Analysis group, of which I am a member. This analysis has been submitted for publication to *Physical Review Letters*. In the final section the main conclusions of this dissertation will be presented, starting with the two FSR recovery techniques, which constitute the main contributions of my work for the H→ $\mu^+\mu^-$ analysis. Also, the main conclusions coming from the H→ $\mu^+\mu^-$ search in relation to the SM and BSM physics will be discussed.

7.1 CUT-BASED FSR RECOVERY FOR H→ $\mu^+\mu^-$

The cut-based FSR algorithm was used with the CMS central production Monte Carlo sample for a 125 GeV Higgs produced through gluon fusion. After the muon selection, we studied the effect of the ΔR criterion mentioned in Section 5.2. Figures 55 and 56 show the behavior of the invariant mass when the photon information is used in its reconstruction.

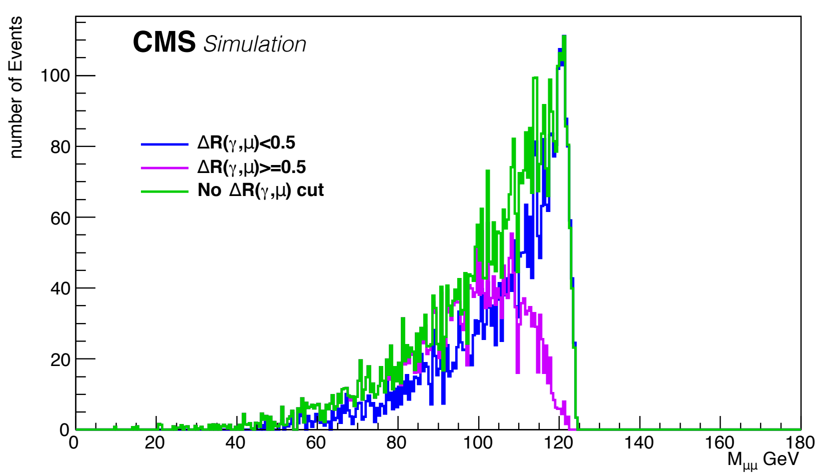

Figure 55: Dimuon invariant mass distribution without FSR corrections and for the case in which the muons pass the relative isolation criterion.

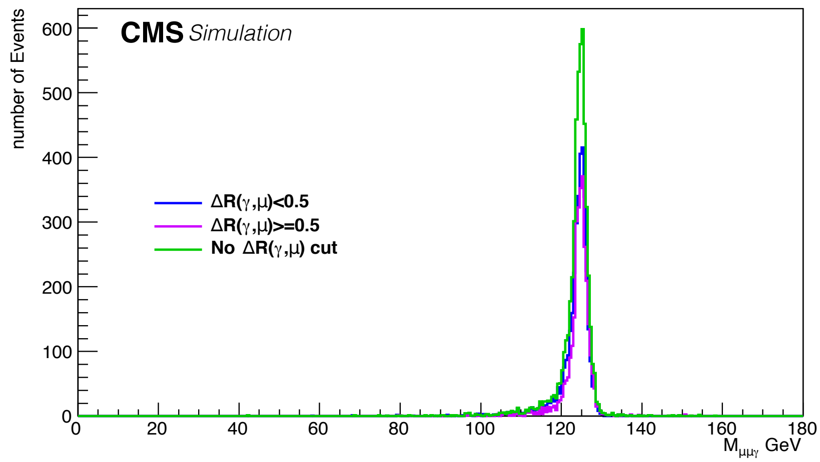

Figure 56: Dimuon invariant mass distribution with FSR corrections, for events with at least one FSR photon, and for the case in which the muons pass the relative isolation criterion.

The plots show that the effect of the $\Delta R(\gamma,\mu)$ <0.5 cut does not improve the mass resolution and diminishes the number of events under the peak, therefore this cut was not included in the selection criteria. In this case, for events with muons that pass the isolation criterion, the improvement in the mass distribution resolution is significant, as will be presented in Section 7.1.2.

7.1.1 EVENT CATEGORIZATION

In order to optimize the sensitivity of the H→ $\mu^+\mu^-$ search, the events are categorized with the same criteria used in the search performed with pp collisions at 7 and 8 TeV [76]. This event categorization is different from the one used with the 13 TeV data, presented in Section 4.4. In order to validate the FSR recovery techniques we have used the first categorization since it is simpler to implement, and the BDT-based categories were still under-development at that time. After event and muon selections, that were presented in Section 5.2, the events were separated in five categories:

- VBF Tight,
- ggF Tight,
- VBF Loose,
- 0,1-jet Tight
- 0,1-jet Loose.

The criteria for each category are described in Table 21. Some of these categories use a preselection criteria that require at least 2 jets with $\Delta R(jet,\mu) > 0.4$, $p_{T jet} > 30$ GeV (with the leading jet $p_{T jet} > 40$ GeV) and $|\eta| < 4.7$. The FSR recovery algorithm was used in each of these event categories and the results are presented in the next section.

Higgs Categories	
VBF Tight	passes preselection use any valid dijet combination dijet mass > 650 GeV dijet $\|\eta\| > 3.5$
ggF Tight	passes preselection dijet mass > 250 GeV dimuon $p_T > 50$ GeV
VBF Loose	passes preselection fails ggF Tight selection
0,1-jet Tight	fails preselection dimuon $p_T \geq 25$ GeV
0,1-jet Loose	fails preselection fails 0,1-jet Tight selection

Table 21: Categorization criteria

7.1.2 FSR RECOVERY RESULTS FOR HIGGS CATEGORIES

The cut-based FSR Recovery was applied for the chosen Higgs event categorization to study the effect over the invariant mass of the dimuon system. The results are presented in Table 22, for the gluon fusion sample: `GluGlu_HToMuMu_M125_13TeV_powheg_pythia8`.

Category	# Events without Recovery	# Events after Recovery	% Recovered events
VBF Tight	2217	2280	2.84 ± 0.35
ggF Tight	4379	4513	3.06 ± 0.26
VBF Loose	11075	11423	3.14 ± 0.16
0/1 Jet Tight	64958	67359	3.70 ± 0.07
0/1 Jet Loose	64814	66089	1.97 ± 0.05

Table 22: Higgs events recovered with the cut-based FSR procedure as a function of event category. The events are counted into a mass window of $110 < m_H < 140$ GeV

These studies show that the FSR recovery procedure provides a significant increase in the number of selected events within the corresponding mass window.

In the same way that the FSR recovery algorithm efficiency was obtained over the geometrical categories, as shown in Section 5.4.3, the efficiency of the algorithm, calculated over the Higgs event categorization, is summarized in Table 23. Compared to the efficiency for geometrical categories (Table 13), the behavior is quite similar. Nevertheless, for the category with biggest statistics, 0/1 Jet Tight, the efficiency improves in around 10%.

Category	Events with an FSR photon at generation level	Reconstructed events with FSR photon identified by the FSR algorithm	FSR reconstruction efficiency (%)
VBF Tight	180	35	19.4 ± 0.08
ggF Tight	374	86	23.0 ± 0.08
VBF Loose	855	214	25.0 ± 0.09
0/1 Jet Tight	4606	1702	37.0 ± 0.10
0/1 Jet Loose	4647	927	19.9 ± 0.08

Table 23: Cut-based FSR recovery algorithm efficiency over Higgs categories.

In order to know if the FSR photons candidates correspond to photons emitted by the muons or by other sources, or if they are the result of misidentification of the photon candidates, we selected the events that have an FSR photon at the reconstruction level that matches a photon at the generation level within a $\Delta R < 0.2$. The purity for the Higgs categories is 10-15% better than that of the geometrical categories (Table 14). Table 24 shows the purity of the FSR recovery algorithm over the Higgs categories.

Category	Reconstructed events with FSR photon (the photon is matched at generation level)	Reconstructed events with FSR photon	FSR reconstruction purity (%)
VBF Tight	35	35	100 ± 0.00
ggF Tight	85	86	98.8 ± 0.02
VBF Loose	211	214	98.6 ± 0.02
0/1 Jet Tight	1681	1702	98.8 ± 0.02
0/1 Jet Loose	917	927	98.9 ± 0.02

Table 24: Cut-based FSR-recovery-algorithm purity obtained for each Higgs category.

In terms of purity, the Higgs categories improve the event selection in comparison to geometrical categories. Nevertheless, the total gain of the FSR Recovery algorithm is reduced in around 1% when the geometrical categories are used.

Another important parameter used to see the effectivity of the method is the significance. The significance, calculated as $S/\sqrt{S+B}$, gives us information about how well we can differentiate between signal and background events. In the case of the H→ $\mu^+\mu^-$ channel the main background sources are DYJetsToLL and TTJets. The significance of the cut-based approach, using these two background sources, is summarized in Table 25. The best significances are obtained for the categories with the higher statistics.

Category	Significance ($S/\sqrt{S+B}$)	
	DYJetsToLL	TTJets
VBF Tight	0.023 ± 0.007	0.123 ± 0.039
ggF Tight	0.084 ± 0.013	0.161 ± 0.025
VBF Loose	0.110 ± 0.006	0.250 ± 0.015
0/1 Jet Tight	0.292 ± 0.003	1.739 ± 0.024
0/1 Jet Loose	0.189 ± 0.002	4.425 ± 0.104

Table 25: Cut-based FSR recovery significance obtained for each Higgs category in the cases in which DYJetsToLL and TTJets samples are used as background sources.

7.2 MVA FSR RECOVERY RESULTS FOR H$\rightarrow \mu^+\mu^-$

In order to improve the percentage of the recovered events, a MVA approach was implemented. The method, previously described in Chapter 6, was used to identify the FSR photons candidates. These candidates were used to correct the final state kinematics of the H$\rightarrow \mu^+\mu^-$ events. The recovery was done over the sample GluGlu_HToMuMu_M125_13TeV_powheg_pythia8 using the categorization presented in Table 22. The results of the recovery are shown in Table 26.

Category	# Events without Recovery	# Events after Recovery	% Recovered events
VBF Tight	2217	2282	2.93 ± 0.35
ggF Tight	4379	4529	3.43 ± 0.27
VBF Loose	11075	11443	3.32 ± 0.17
0/1 Jet Tight	64958	67406	3.77 ± 0.07
0/1 Jet Loose	64814	66112	2.00 ± 0.05

Table 26: Higgs events recovered with the MVA FSR procedure as a function of event category. The events are counted into a mass window of $110 < m_H < 140$ GeV

Comparing to the cut-based results (Table 22), the number of events after the recovery increase in all the Higgs categories. This leads to the percentage of events recovered increasing by 0.03-0.1%.

Figures 57 to 59 show the dimuon-invariant-mass distribution without including the FSR photon contribution (red histogram), the dimuon mass including the FSR photons selected by the cut-based approach (blue histogram), and those selected by the MVA-based approach (green histogram). Also the per-bin gain for both, the cut and MVA-based approaches, is presented for each dedicated Higgs category.

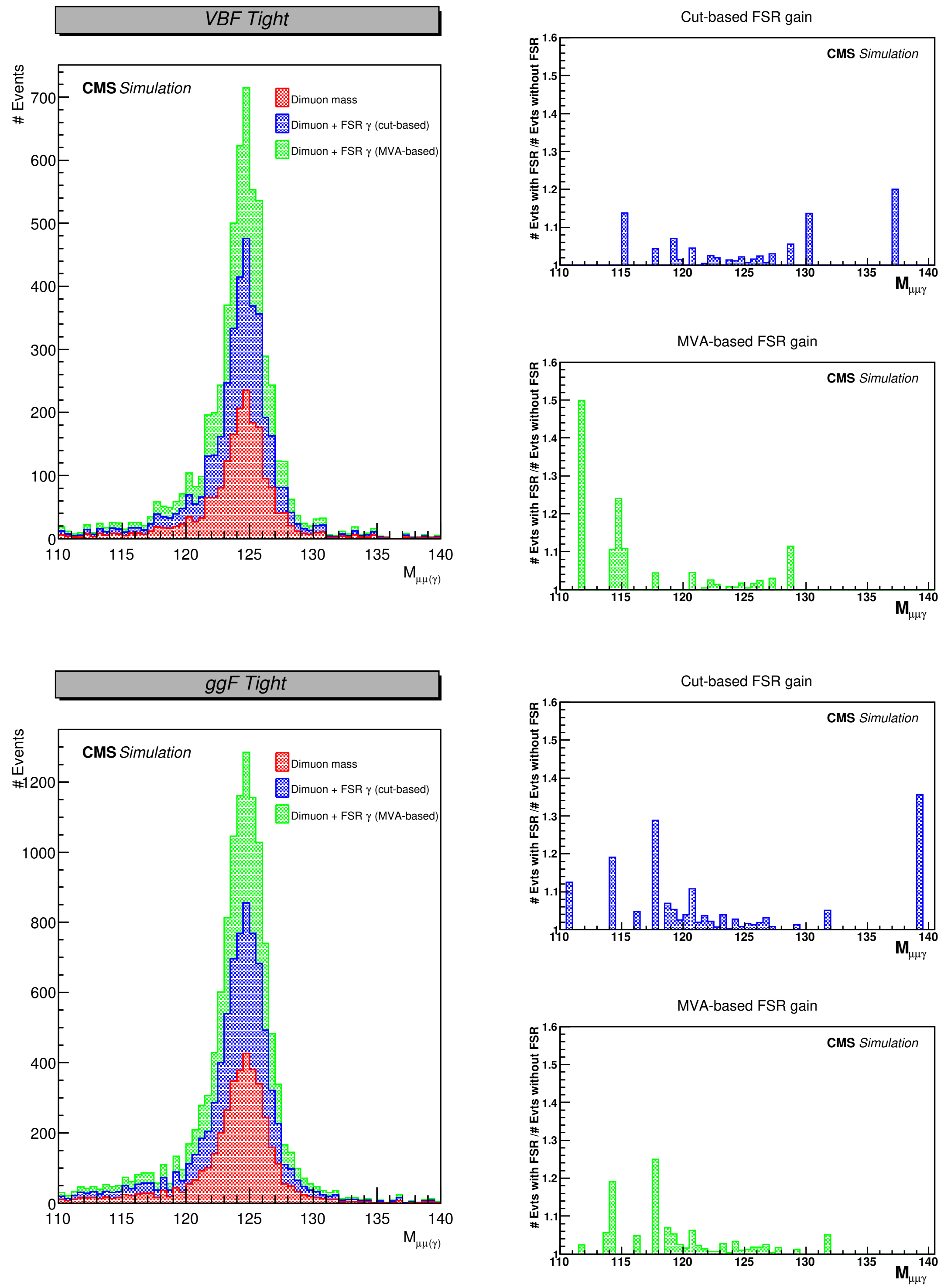

Figure 57: Left: Dimuon mass without added FSR photons (red), adding cut-based selected γ (blue), adding MVA selected γ. Right: Per-Bin gain for cut-based (blue) and MVA (green). Above: VBF Tight category. Below: ggF Tight category

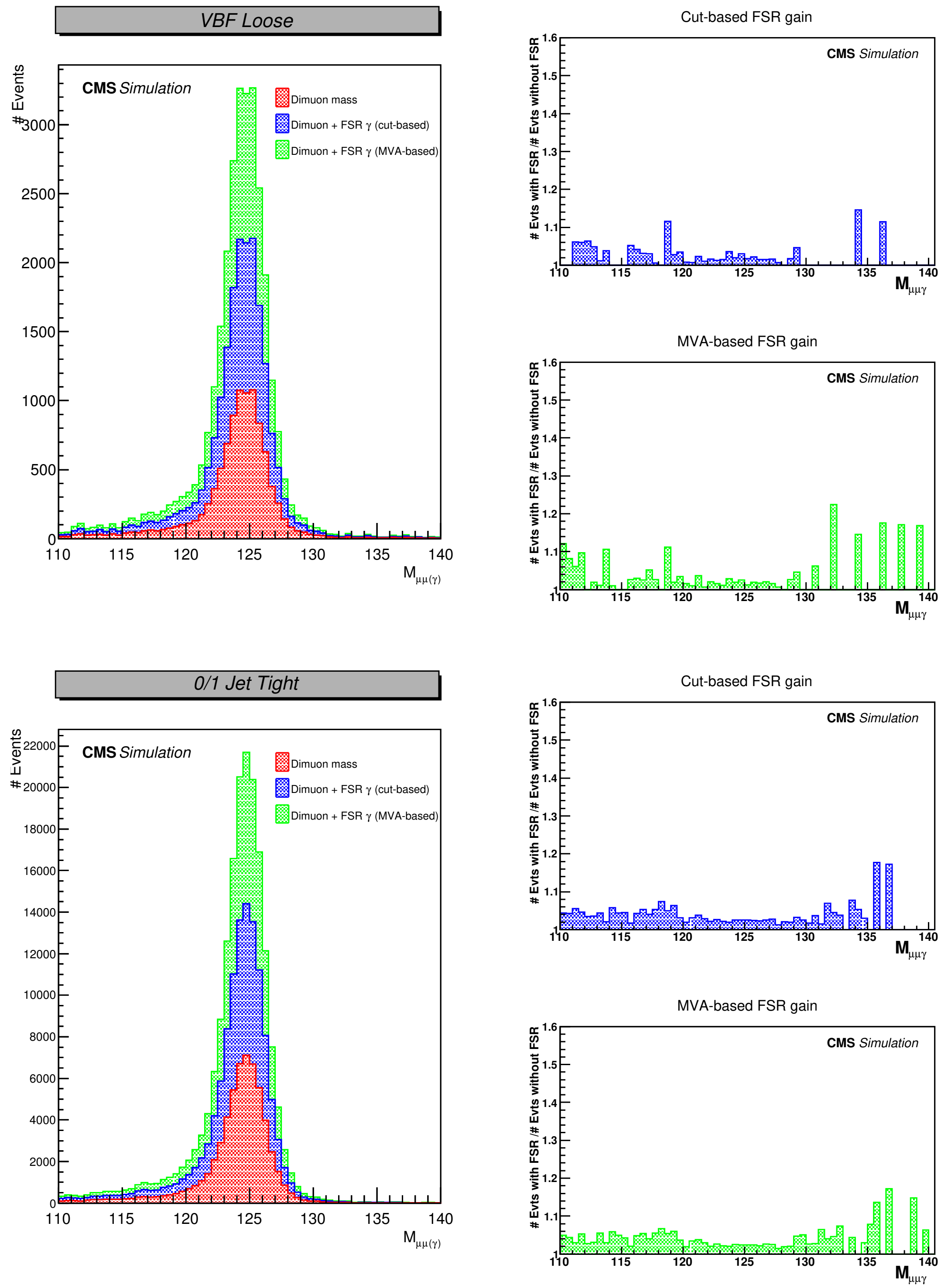

Figure 58: Left: Dimuon mass without added FSR photons (red), adding cut-based selected γ (blue), adding MVA selected γ. Right: Per-Bin gain for cut-based (blue) and MVA (green). Above: VBF Loose category. Below: 0/1 Jet Tight category

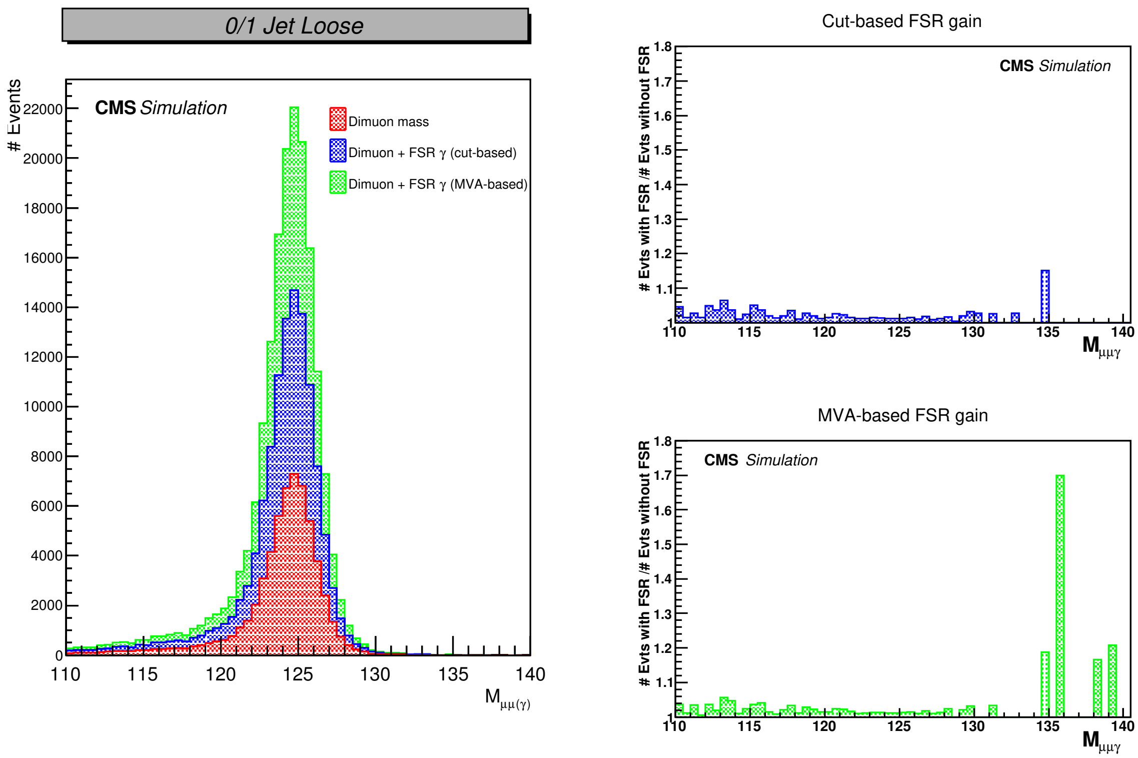

Figure 59: Left: Dimuon mass without added FSR photons (red), adding cut-based selected γ (blue), adding MVA selected γ. Right: Per-Bin gain for cut-based (blue) and MVA (green) for the 0/1 Jet Loose category

In order to look further into the behavior of the MVA approach, the efficiency of the method was also obtained. Table 27 shows the efficiency of the MVA FSR Recovery for all the Higgs categories. The efficiency is calculated in the same way that we described in Section 5.4.3.

Category	Events with an FSR photon at generation level	Reconstructed events with FSR photon identified by the FSR algorithm	FSR reconstruction efficiency (%)
VBF Tight	182	36	19.8 ± 0.09
ggF Tight	384	102	26.6 ± 0.08
VBF Loose	866	233	26.9 ± 0.09
0/1 Jet Tight	4626	1754	37.9 ± 0.10
0/1 Jet Loose	4665	969	20.8 ± 0.08

Table 27: MVA-based FSR recovery algorithm efficiency over Higgs categories.

For all the categories the MVA FSR recovery method is more efficient than the cut-based method (Table 23). The efficiency increases in around 0.4-2.4%. Figure 60 shows the difference between cut-based and MVA approaches.

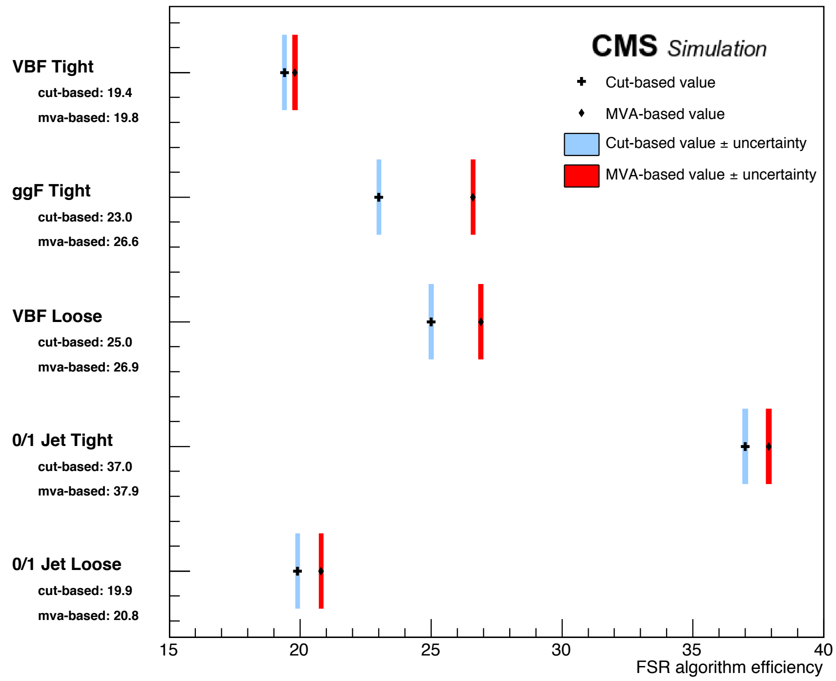

Figure 60: FSR algorithm efficiency for cut-based and MVA approaches.

Also, it is important to know if the MVA approach gives us a similar purity than the cut-based approach. The purity is calculated as described in Section 5.4.4. Table 28 summarizes the purity results for the MVA method. Comparing to the results for the cut-based method (Table 23), we can see that the purity of the MVA approach shows a similar behavior for all the Higgs categories. Figure 61 shows the purity for the two methods.

Category	Reconstructed events with FSR photon (the photon is matched at generation level)	Reconstructed events with FSR photon	FSR reconstruction purity (%)
VBF Tight	35	36	97.2 ± 0.03
ggF Tight	100	102	98.0 ± 0.03
VBF Loose	230	233	98.7 ± 0.02
0/1 Jet Tight	1742	1754	99.3 ± 0.02
0/1 Jet Loose	954	969	98.5 ± 0.02

Table 28: MVA-based FSR-recovery-algorithm purity obtained for each Higgs category.

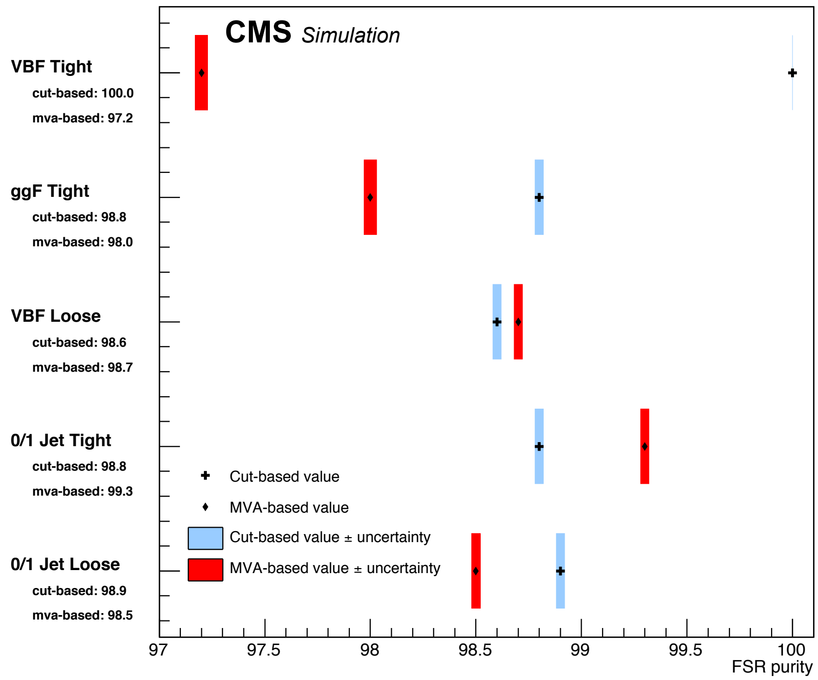

Figure 61: FSR algorithm purity for cut-based and MVA approaches.

Afterwards, it is also important to know how well we can differentiate the signal from background events when using the MVA approach. The significance, calculated using the two main background sources: DYJetsToLL and TTJets, is presented in Table 29. Comparing to the results for the cut-based approach (Table 25) the significance increases for all the Higgs categories. Figure 62 shows the comparison of the significance over DYJetsToLL and TTJets samples for cut-based and MVA methods.

Category	Significance ($S/\sqrt{S+B}$)	
	DYJetsToLL	TTJets
VBF Tight	0.073 ± 0.023	0.124 ± 0.039
ggF Tight	0.086 ± 0.013	0.163 ± 0.026
VBF Loose	0.115 ± 0.007	0.251 ± 0.015
0/1 Jet Tight	0.296 ± 0.003	1.739 ± 0.024
0/1 Jet Loose	0.190 ± 0.002	4.431 ± 0.104

Table 29: MVA FSR recovery significance obtained for each Higgs category in the cases in which DYJetsToLL and TTJets samples are used as background sources.

The MVA approach not only provides an improvement in the total gain of recovered events, but also, gives us a better efficiency and significance. The purity in both approaches is very similar. Even though the MVA method performs better, as compared to the cut-based method, the differences are small. At this stage any of the two methods could be used to increase the number of signal events in the H→ $\mu^+\mu^-$ analysis.

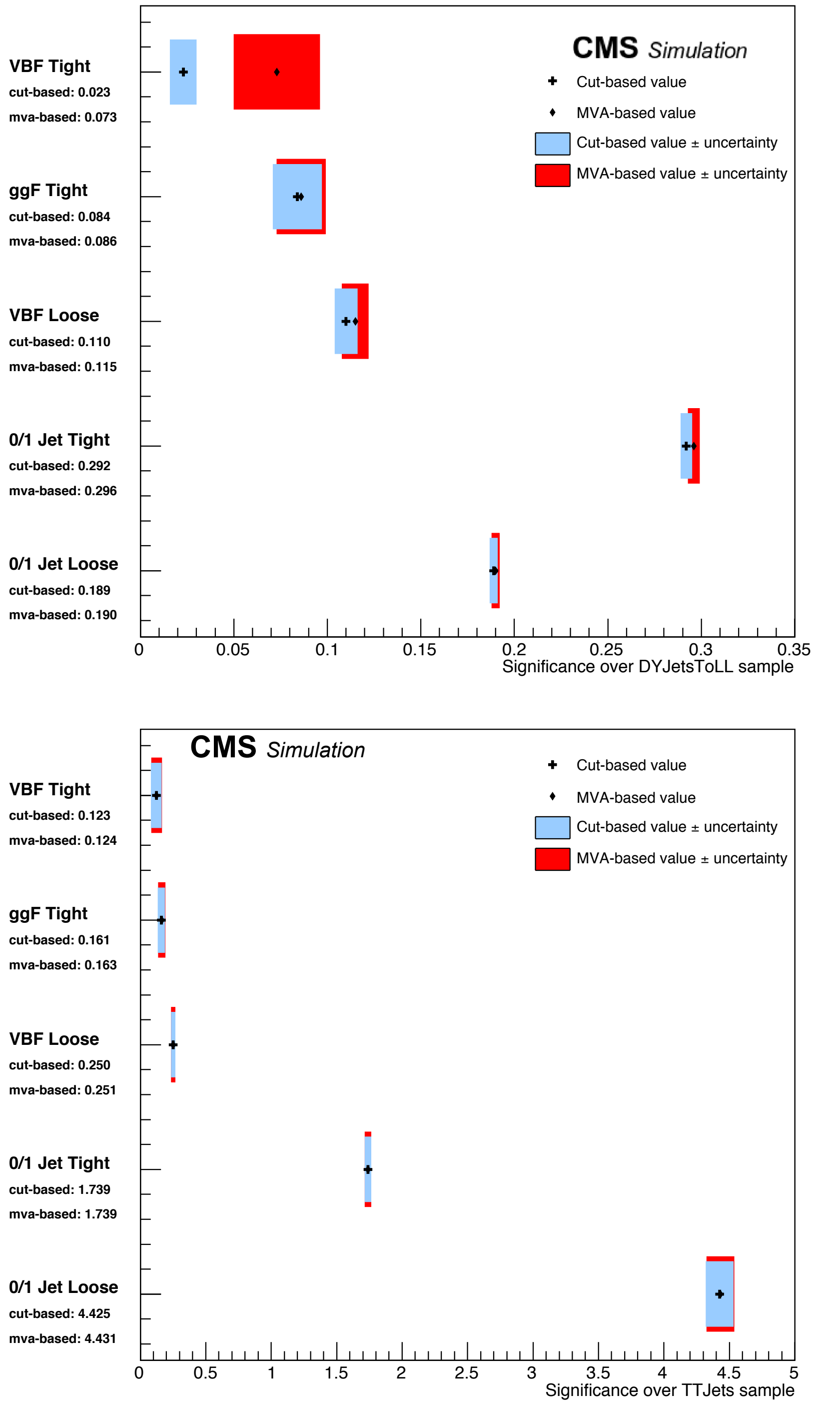

Figure 62: Significance ($S/\sqrt{S+B}$) for `DYJetsToLL` case (Upper) and `TTJets` case (Below) for cut-based and MVA approaches.

7.3 H→ $\mu^+\mu^-$ RESULTS

The CMS H2Mu Analysis Group performed the search for the SM Higgs decaying into two muons, using data sets corresponding to a centre-of-mass energy of 13 TeV, with an integrated luminosity of 35.9 fb^{-1}. Muon identification was performed using the PF technique presented in detail in Section 3.3.1. For both, signal and background studies, the four leading Higgs production processes were included: gluon fusion (ggH), vector boson fusion (VBF), Higgs-strahlung (VH) and associated production with top quarks ($t\bar{t}$H). The Higgs boson simulated samples were generated with POWHEG. The details of these studies were presented in Chapter 4. In the case of the SM Higgs, the main background processes are Drell-Yan (DY) and single and pair top quark production ($t\bar{t}$), followed by various boson production channels. The background samples were generated using POWHEG, MC@NLO, MADSPIN plus PYTHIA8 and GEANT4. The details of these studies were presented in Section 4.2. The information about the event selection criteria and the signal and background models are also presented in Chapter 4.

7.3.1 SYSTEMATIC UNCERTAINTIES

The sources of systematic uncertainties taken into account in this analysis come from:

- shape uncertainties.
- category migration.
- rate uncertainties.

Shape uncertainties affect the signal model in all the categories and production channels. The shape of the peak used to identify the Higgs boson is affected by the muon momentum scale and its resolution. The studies of the propagation of these uncertainties into the invariant mass of the Higgs boson yield 0.05 % variation for the position of the peak and 10% for the value of its width.

The uncertainties resulting from migrations between categories are mostly due to jet energy miscalibrations and migration in and out of the acceptance of the analysis. These uncertainties amount up to 6% of the relative yield. The pile-up effects can also produce migrations that translate in an effect of the categorization of the order of 1%. The uncertainties coming from b-tagging were propagated though the analysis categories resulting in an effect of the order of 1%. Lepton reconstruction and identification uncertainties have an effect of the order of 2%. Uncertainties coming from factorization and renormalization scales, at the Monte Carlo level, have an effect of the order of 6% in category migration. The uncertainties coming from parton distribution functions were estimated to be of the order of 2%.

The uncertainties associated to the rate of the events are due to factorization and normalization scales, parton distribution functions, the branching fraction for the H→ $\mu^+\mu^-$, and from the luminosity measurement. The values of all these uncertainties are summarized in Table 30. The impact of all systematic uncertainties is very small. Therefore the uncertainties of this analysis are dominated by statistics.

	Systematic uncertainties		Percentage (%)
Shape uncertainties	Variations in the position of the peak		0.05
	Changes in width		10
Category migration	Jet energy scale		6
	Jet energy resolution		3
	Pile-up effect		1
	B-tagging		1
	Lepton Reconstruction and identification		2
	Factorization and Renormalization MC scales		6
	Parton Distribution Functions		3
Rate uncertainties	Factorization and Renormalization scales	ggH	3.9
		VBF	0.4
		ZH	3.8
		WH	1.9
		ttH	$\lesssim$10
	Parton Distribution Functions	ggH	3.2
		VBF	2.1
		ZH	1.6
		WH	1.9
		ttH	3.7
	Branching ratio		1.7
	Luminosity		2.5

Table 30: Sources of systematic uncertainties [70]

7.3.2 RESULTS OF THE SEARCH

In summary, the search for H→ $\mu^+\mu^-$ was executed selecting events that passed un-prescaled single muon triggers and classifying them using the criteria presented in section 4.3. A maximum likelihood fit was done over all the categories. The signal strength for a 125 GeV SM Higgs boson is $\hat{\mu}_{125} = 0.7^{+1.1}_{-1.0}$. Figure 63 show the background-only and the signal-plus-background fits done over real data in all the categories.

The expected upper limit, with a 95% CL, for each BDT category, is shown in Figure 64. The upper limits on the strength modifier ($\mu = \frac{\sigma\mathcal{B}_{obs}}{\sigma\mathcal{B}_{SM}}$) that are compatible with the background-only hypothesis are shown in Figure 65 for the 2016 dataset. The observed (expected) upper limit for a 125 GeV Higgs is $\mu < 2.64$ (μ =$2.08^{+0.87}_{-0.59}$). The observed (expected) significance is 0.74 (0.98)σ at 13 TeV.

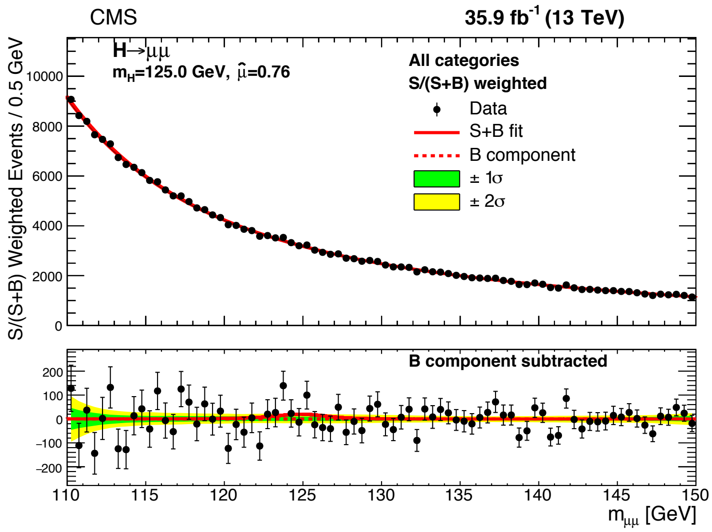

Figure 63: Background-only and signal-plus-background fits summed over all categories. Events are weighted according to the expected signal-to-background ratio of the category to which they belong [64].

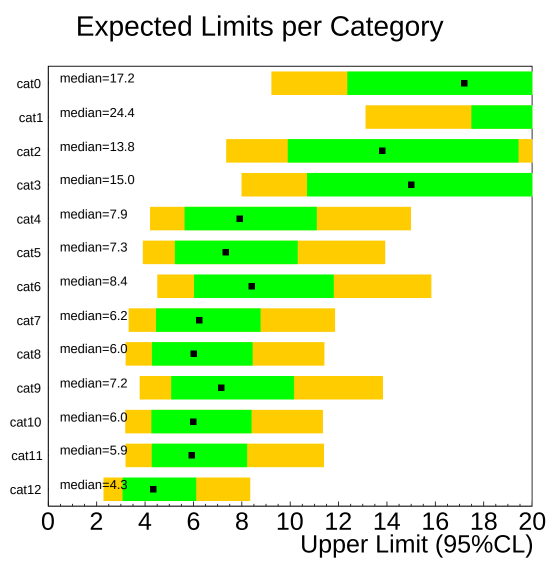

Figure 64: The 95% CL upper limit on the signal strength modifier. Each category is fit independently [64].

Combined limits were obtained using the results of the 7 and 8 TeV searches. The observed (expected) combined significance is 0.74 (0.98 σ). The observed (expected) limit is 2.64 ($1.89^{+0.79}_{-0.54}$) times the standard model limit (See Figure 66).

These limits and significances correspond to an upper limit on the H→ $\mu^+\mu^-$ branching fraction of 5.7×10^{-4}. These are the best results in Higgs coupling to muons up to the current date. A detailed description of this analysis has been presented by the CMS H2Mu analysis group in the Analysis Note CMS AN-17-098 [64] and a paper is been submitted to *Physical Review Letters* [65]. An updated version of this analysis, using the full dataset of the Run 2 (that will finish by the end of 2018) will include, not only much larger statistics (of the order of 100 fb^{-1}) but, other features like the FSR recovery technique that I have developed in this work an presented in this thesis.

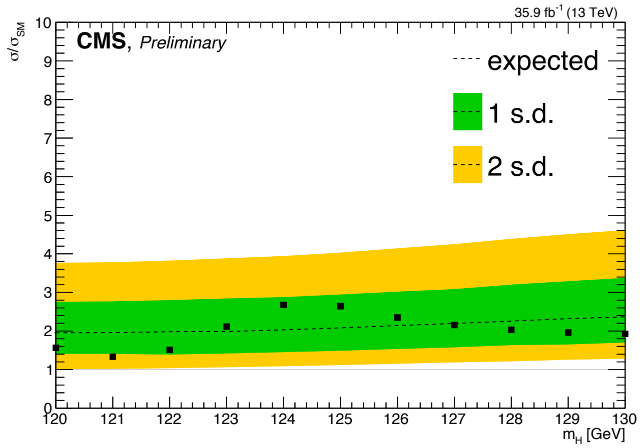

Figure 65: The observed and expected 95% CL upper limit on the signal strength modifier, μ [65].

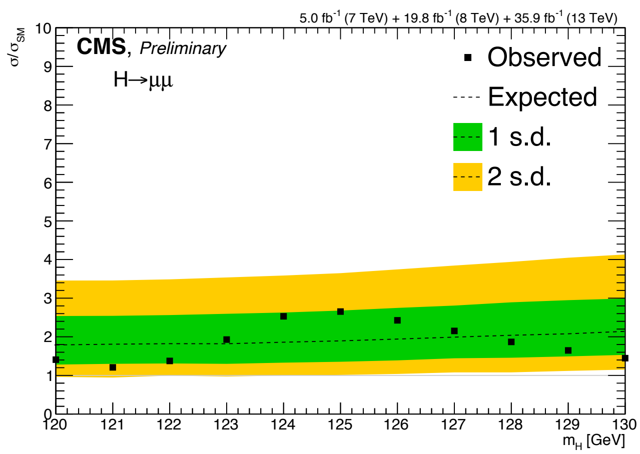

Figure 66: The combined observed and expected 95% CL upper limit on the signal strength modifier, μ, [64]

7.4 CONCLUSIONS

After the discovery of the Higgs boson using the main decay channels, the study of other channels, in which the yield is lower, has become more relevant in order to confirm and complete the SM hypothesis. One of these channels is the one studied in this dissertation: H→ $\mu^+\mu^-$. This channel has been studied with the LHC Run 1 data (7 and 8 TeV) and with the current Run 2 data (13 TeV).The observation of this channel is important to achieve a better understanding of the Higgs properties. The Higgs boson decaying to muons is important because it allows the study of the Higgs coupling to second-generation-leptons, where deviations from the SM could give us hints about physics BSM.

After the first searches with data coming from Run 1, the CMS and ATLAS collaborations placed upper limits of approximately seven times the SM value at the 95% confidence level (CL). ATLAS set its observed (expected) limit to 2.8 (2.9) times the SM expectations when combining Run 1 and Run 2 data. CMS set an observed (expected) upper limit on the production rate of 2.64 (1.89) times the SM value using Run 1 and Run 2 data. Both experiments expected to obtain better limits at the end of Run 2, due to the higher luminosity (of the order of a 100 fb^{-1}).

Due to the very small dimuon branching fraction of the muonic channel (2.18×10^{-4}) for 125 GeV Higgs boson, any improvement in the selection criteria of signal events is desirable. Some of the muons coming from Higgs decays would be discarded in a first event selection since they radiated photons which causes them to be non-isolated. In order to increase the signal yield two different techniques for recovering these events have been implemented in this thesis. These approaches were: a cut-based and a MVA-based FSR recovery techniques. Both give a good and similar improvement in the final signal selection (around $\sim 3\%$).

The cut-based approach was implemented and tested over a Z^0 signal sample using real data and geometrical categories based on the muons pseudorapidity, giving a total gain of around 2%. The recovery method not only increased the number of signal events, but also improved the resolution of the dimuon mass distribution for the Z^0 peak. The same approach was implemented over simulated data for a gluon-fusion Higgs sample. In the case of the same geometrical categories used for the Z^0, the total gain was between 3 and 4%. For the dedicated Higgs categories the gain was between 2 and 4%. When using the Higgs categories, the efficiency of the method and the purity of the recovered events was improved.

The MVA approach was developed as an alternative method, looking for more efficient recovery. The MVA BDT that we implemented was trained using known FSR events from simulated data. The method learned from these events the characteristics of the FSR photons, and created a model that permits choosing events that do not meet all the cuts, but that have the fingerprints of FSR photons. The MVA model used more variables than the cut-based approach in its modeling of the FSR events, so it distinguished between true FSR events and similar phenomena better, giving us also a better significance in the selection. The method was implemented over a simulated data for Higgs sample, using the dedicated Higgs categories. It improved over the cut-based method not only in the total gain but, also in the efficiency and in the rate of discrimination between signal and background. Nevertheless, the gain is not significantly better than the one from the cut-based method. The two methods improve the signal yield similarly, so that the decision to use one or the other depends on CPU time or further studies using BDT-based categories.

The results of these studies show, in the case of real data for Z^0, and in the case of simulated data for the Higgs, that we can recover a significant fraction of events. We expect that the efficiency and gain of these recovery approaches will be improved when tested with the BDT-based Higgs categorization used by the CMS H2Mu group. The two recovery methods could be used with the full data collected during Run 2 when the integrated luminosity will be of the order of 100 fb^{-1} (instead of the current integrated luminosity of 35.9 fb^{-1}). The computational implementation of these two methods, one of the main results of this thesis, are now available to the CMS H2Mu Analysis Group, to be used in future analyses. The code can be accessed at [84]. An analysis note presenting this work is also being submitted to the CMS Collaboration [85].

After Run 2, the Higgs mass will have a precision of ~ 100 MeV, and it will be challenging to improve it. At the expected higher luminosity and collision energy, investigation of CP violating contributions can be undertaken using LHC data. In Run 3 an improvement in the precision on Higgs signal strenght (μ) of around 10 - 15% for the main Higgs decay modes will be achieved with 300 fb^{-1}. Run 3 will start in 2021 and during this period the accelerator will reach the design collision energy of 14 TeV [86].

In 2024, the LHC and its experiments will prepare for the high luminosity LHC (HL-LHC). At this stage, the instantaneous luminosity will be $\mathcal{L} = 5 - 7.5 \times 10^{34}\text{cm}^{-2}\text{s}^{-1}$ and the LHC will deliver the entire combined Run 2 and Run 3 integrated luminosities in one year. One important objective for the physics of the Higgs sector will be to study Higgs pair production and rare decays. ATLAS and CMS will upgraded their detectors so that, with HL-LHC data, will be able to probe SM Higgs couplings deviations with a precision of a few percent. In the case of Higgs couplings, specially the muon coupling, they will be measured with significant precision, and therefore studies of the physics associated with these couplings will be possible. Also, direct searches for heavy Higgs partners and the ttH production mode can be analyzed with the main decay channels [87, 88].

Other types of colliders, like the International Linear Collider (ILC), which will use a e^+e^- collisions, will be an excellent option to achieve a higher precision on most of Higgs properties. Lepton colliders have the advantage of having low backgrounds for Higgs production, making all the decay modes accessible. The production associated with top quarks and Higgs pair is possible in linear colliders because of their energy threshold. A Hadronic collider, like the HL-LHC, on the other hand takes advantage of huge cross-sections and a high-energy, allowing rare decays and heavy final states. Nevertheless, due to their large background not all of the decays would be accessible. Also, both types of colliders can study the BSM Higgs sector in a complementary way. This implies a wide range of scientific possibilities in the search for BSM physics and detail studies of the SM [89].

A

TRIGGER DEVELOPMENT FOR $H \rightarrow Z^0\gamma$ AND $H \rightarrow \gamma\gamma^*$ ANALYSES.

During my PhD studies I worked very closely with the HZG analysis group as part of the CMS service tasks. They work in several analyses, such as searches $H \rightarrow Z^0\gamma$ and $H \rightarrow \gamma\gamma^* \rightarrow \gamma\ell^+\ell^- l$ (Higgs Dalitz) decay channels. After the 7 and 8 TeV results, the group performed some tasks in order to improve the searches at 13 TeV. One of these tasks was the creation of an appropriate trigger. The trigger had to be able to tag a muon and a photon, and be unprescaled with the minimum possible threshold in p_T for the muons. I developed a muon+photon trigger that was approved by the CMS Trigger group and used in the 13 TeV searches.

In order to execute the task, I first studied the composition of the trigger or HLT path, to understand how to create a new HLT path or modify an existing one. After reviewing the possible L1 triggers or L1 seeds, that are available at the L1 menu, we studied the efficiencies, p_T thresholds and sustainability of the L1 seeds (Trigger group decisions). We decided to keep the L1 seed that the group used in the 8 TeV search. Afterwards, we modified the p_T threshold for the photon over the 8 TeV trigger. Once we had found the appropriate HLT path, we tested it over data and MC samples. In the next sections I will present details of that work.

A.1 CMS HLT PATH

In CMS the event selection can be governed by one or by various triggers, which depend on the physics analysis that we wish to undertake. These triggers are contained in one HLT menu. In these menus there are typically around 400 independent paths. A path is a sequence of producer and filter modules. The HLT configuration is stored in a dedicated database called *ConfDB* and is available for lookup and review on the ConfDB website [90].

The first step in one HLT path is normally a level 1 (L1) seed. The L1 seed selects data for Level 1 trigger in order to tag and reduce the data rate. A L1 seed can be used in more than one HLT-path. There are fewer L1 seeds than HLT paths, around 128 for all the trigger groups. After the execution of a L1 seed a pre-scaler can be used. The prescaling discards some of the events in order to reduce the data rate; we preferred to use a non prescaled trigger to maximize the number of events in our analysis.

For the HZG analysis, we used triggers without pre-scaling, a level 2 (L2) module and its respective Reco module and filter. This kind of arrangement can be repeated depending on the HLT path needed for the analysis. After L2, there is a third level which includes tracking, e.g. pixel-only tracks.

A representation of the content of an HLT path is shown on Figure 67. If any filter or module in the path fails, it will be discarded by the path, and the remaining modules will not run. When developing a new HLT path, it is necessary to use an existing one with similar conditions as a base, so that comparisons of their efficiencies can be made.

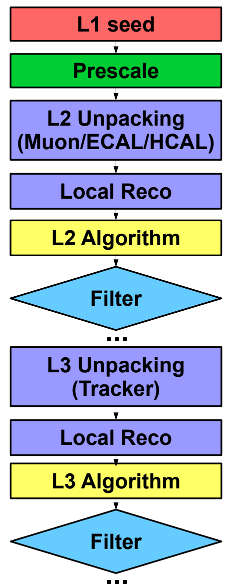

Figure 67: A representation of the content of a CMS HLT path [91].

A.2 PHOTON p_t THRESHOLD AND EFFICIENCIES

The path that I developed as part of my service-work at the CMS experiment, was a muon-photon trigger for the HZG analysis group. It was necessary to create a trigger that allowed for both, a better efficiency and lower event rates, in order for it to be included in the HLT menu. The trigger was developed for the Higgs-Dalitz decay channel (M125, M($\ell\ell$)< 50 GeV).

To determe the necessary threshold for the photon momentum, we used a trigger that had been used for the same analysis at 8 TeV. We kept the same L1 seed (`L1_Mu5_IsoEG18`) and modified only the photon p_T, setting it between 20 and 30 GeV. The base trigger that was used was `HLT_Mu17_-Photon30_CaloIdl_L1ISO`. In Table 31, I summarize the values of the efficiencies and the L1 rate over the most significant background. The efficiency, as a function of momentum, is shown on Figure 68 for all tested p_T values.

A.3 L1 SEED EFFICIENCIES AND L1 RATES

After measuring efficiencies as a function of the photon p_T threshold, we checked the efficiency of the L1 seeds that could identify a muon and a photon on the Dalitz signal. The idea was to reduce the p_T threshold of the photon and to remove the need for photon isolation. We proceeded to calculate the L1 Rate (Hz) for each tested L1 seed. The L1 Rate is measured for the most significant background sample in the analysis, which in the case of the Higgs Dalitz was Drell–Yan (DY→ $\mu^+\mu^-$). The results for these efficiencies and L1 rates are presented on Table 32.

Based on these results, we found that `Mu5_EG20` had the best performance as a L1 seed. This seed selects a muon with a minimum p_T of 5 GeV, and a photon with a p_T> 20 GeV, and no isolation for either the muon or photon. However, due to the L1 seed pre-scaling, we decided to keep on using

Trigger	Efficiency	Rate(Hz) (DYtoMuMu) $\sim$ 40k events
HLT_Mu17_Photon20...	59.5 ± 1.55	0.29 ± 0.03
HLT_Mu17_Photon22...	58.7 ± 1.56	0.25 ± 0.02
HLT_Mu17_Photon24...	58.1 ± 1.56	0.20 ± 0.02
HLT_Mu17_Photon26...	57.7 ± 1.56	0.16 ± 0.02
HLT_Mu17_Photon28...	56.8 ± 1.57	0.13 ± 0.02
HLT_Mu17_Photon30...	56.0 ± 1.57	0.09 ± 0.01

Table 31: Trigger efficiencies for different photon p_T threshold. The p_T threshold for muon remains equal for all the cases. The Rate (Hz) is measured over the most significant background source, the DY sample.

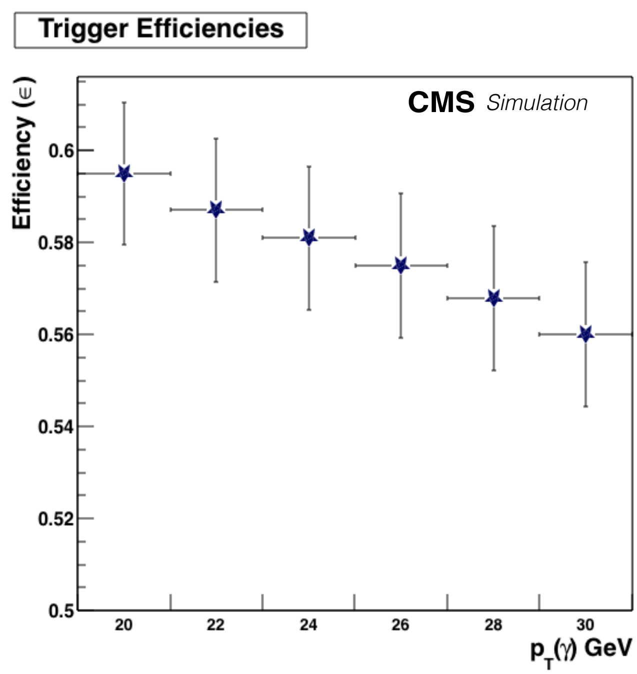

Figure 68: Trigger efficiencies as a function of photon p_T.

L1_Mu5_IsoEG18, which allowed for smaller L1 rates and had an efficiency not so different from the best one (Mu5_EG20).

A.4 FINAL HLT TRIGGER

The trigger developed for the 13 TeV analysis was HLT_Mu17_Photon22_CaloIdL_L1ISO. This HLT path had the best compromise between efficiency and L1 rates, whilst meeting the channel requirements. The trigger must be tested against real data; a known Run is used and the rates are calculated. We used Run 254790, which had an instantaneous luminosity of $7.21 \times 10^{31} \text{cm}^{-2}\text{s}^{-1}$. The certified lumi-sections were used, and the scaling factor for different luminosities was determined. Table 33

L1Muon + Photon Triggers	Efficiency over signal sample (1000 events)	L1 Rate (DYtoMuMu) (∼50 k events) Efficiency L1 Rate (Hz)
L1_Mu12_EG10	72.7 ± 1.4	2.22 ± 0.06
L1_Mu20_EG10	68.4 ± 1.5	1.94 ± 0.06
L1_Mu5_EG15	72.3 ± 1.4	1.17 ± 0.05
L1_Mu4_EG18	70.9 ± 1.4	0.84 ± 0.04
L1_Mu5_EG20	70.1 ± 1.4	0.65 ± 0.03
L1_Mu5_ISOEG18	68.6 ± 1.5	0.62 ± 0.03

Table 32: L1 Muon plus photon trigger (L1 seeds) efficiencies over the Higgs Dalitz signal sample. The L1 Rates are measured over the main background source: DY.

summarizes the characteristics of the run.

Run 254790	
Ins. Luminosity	7.21e32 $cm^{-2}s^{-1}$
Lumi sections	104-143 (40)
Lumi Length	23.3s
Rate L1 seed (L1_Mu5_IsoEG18)	113.21 Hz
Luminosity Scale Factor	For 1.4e34: 19.25 For 5.6e33: 7.76

Table 33: Properties of the Run 254790. This run was used to test the Muon+Photon trigger HLT_Mu17_Photon22_-CaloIdL_L1ISO.

Table 34 presents a summary of the rates for both the original trigger and the trigger that was developed over the tested HLT paths. Table 35 shows the final rates for both triggers at different luminosities.

After the trigger was developed, analyzed, and tested, it had to be approved by the collaboration and included in the HTL menu. For this purpose, after the Trigger Group has found the trigger appropriate, it must be created using ConfDB, a tool developed by CMS; next an integration tests must be executed (hltIntegrationTests). Our HLT path was approved and included in the HLT menu as can be seen in the JIRA ticket [92].

	Events passing L1 Seed	HLT_Mu17_Photon30 (Counts)	HLT_Mu17_Photon22 (Counts)
HLT Physics Part 0	225	2	5
HLT Physics Part 1	217	3	3
HLT Physics Part 2	233	1	2
HLT Physics Part 3	241	0	2
HLT Physics Part 4	235	2	3
HLT Physics Part 5	222	0	3
HLT Physics Part 6	243	0	2

Table 34: Test over HLT Physics Trigger (used for test with data). The number of events that pass the L1 seed, the counts for HLT_Mu17_Photon30 and HLT_Mu17_Photon22 are also presented.

HLT path	Rate (Hz) for 5.6e33	Rate (Hz) for 1.4e34
HLT_Mu17_Photon30	4.34 ± 1.36	10.78 ± 3.81
HLT_Mu17_Photon22	10.86 ± 2.42	26.97 ± 6.03

Table 35: Final rates for HLT_Mu17_Photon30 and HLT_Mu17_Photon22 HLT paths.

B

IMPLEMENTATION OF A MULTIVARIATE ANALYSIS PHOTON IDENTIFICATION FOR H$\rightarrow$Z$^0\gamma$ ANALYSIS

As mentioned in Appendix A, during my doctoral studies and internship at CERN, I collaborated with the CMS HZG analysis group. After a preliminary search for the Higgs boson in the $Z^0\gamma$ channel, using a cut-based photon identification criteria, we explored the usage of MVA photons in order to improve the sensitivity of the search and signal-to-background ratio. In the next section I presented a summary of the work that I performed for that purpose.

B.1 PHOTON PRESELECTION AND CATEGORIZATION

An event preselection is required for the development of the MVA Photon ID studies. Our implemented preselection is summarized in Table 36. The thresholds used were recommended by the EGammaPOG. The procedure for training the MVA model is the same as the one presented in Chapter 6. The weight files (*xml training files*) used to obtain the MVA values were provided by the EGammaPOG group also [55].

Photon Preselection criteria		
SCEta < 1.479	PassedEleSafeVeto OK SigmaIEtaIEta < 0.014	R9 > 0.9 HadOverEm< 0.082 R9 < 0.9 HadOverEm< 0.075
SCEta > 1.479	PassedEleSafeVeto OK SigmaIEtaIEta < 0.034 HadOverEm < 0.075	

Table 36: Preselection criteria for the photons in the H$\rightarrow$Z$^0\gamma$ sample.

After including the preselection criteria, it is necessary to match the photons in the sample with those that come from the Higgs process. To improve the sensitivity, the events were divided in "barrel" and "endcap" categories. For the barrel category, the photons must have $|\eta| \leq 2.4$, while for the endcap category the photons must have $|\eta| > 2.4$. Also four categories that take into account the position of the muons and photons in the detector were devised and used through the analysis. The criteria for these four categories are described in Table 37.

Event Categories	
Category 1	Photon $0 < \|\eta\| < 1.44$ Both leptons $0 < \|\eta\| < 2.1$ and one lepton $0 < \|\eta\| < 0.9$ R9 > 0.9
Category 2	Photon $0 < \|\eta\| < 1.44$ Both leptons $0 < \|\eta\| < 2.1$ and one lepton $0 < \|\eta\| < 0.9$ R9 < 0.9
Category 3	Photon $0 < \|\eta\| < 1.44$ Both leptons $\|\eta\| > 0.9$ or one lepton $2.1 < \|\eta\| < 2.4$
Category 4	Photon $1.56 < \|\eta\| < 2.5$ Both leptons $0 < \|\eta\| < 2.4$

Table 37: Event categories based on lepton pseudo-rapidity.

B.2 MVA VALUES

The MVA values are used as an alternative to Cut-based identification criteria. These values do not match a single physical quantity, but come from statistical correlations between all the relevant physical variables used in the training of the MVA classifier. The chosen physics variables used in the MVA training are those that maximise the identification of signal samples while rejecting the most background events.

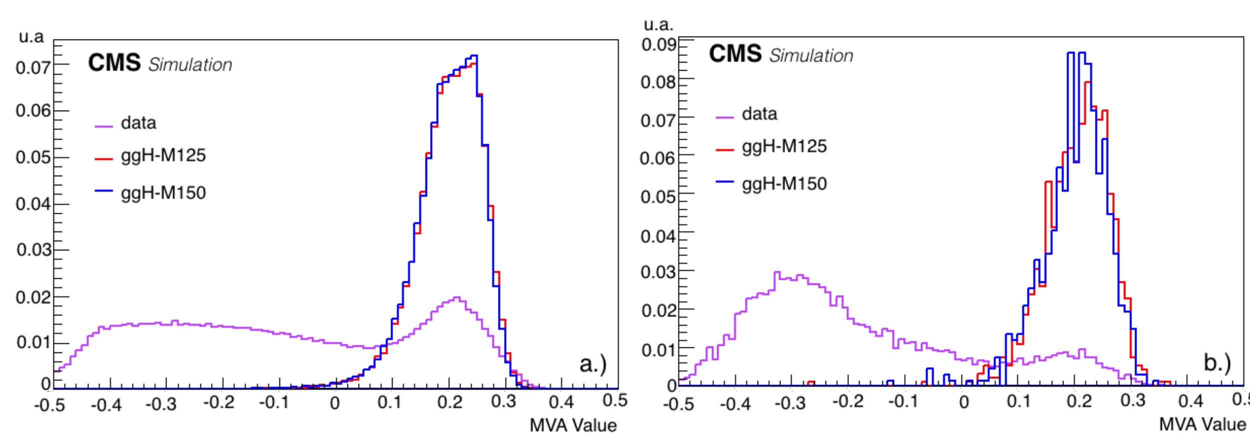

Figure 69: The MVA values for events with photons: a.) in the barrel and b.) in the endcap. The MVA values are obtained for data and signal samples (Higgs Masses 125 and 150 GeV).

The MVA values are calculated for real data and two gluon-fusion signal samples: M125 and M150. The MVA value distribution for barrel and endcap are shown in Figure 69, and the ones for categories 1 through 4 are shown in Figure 70.

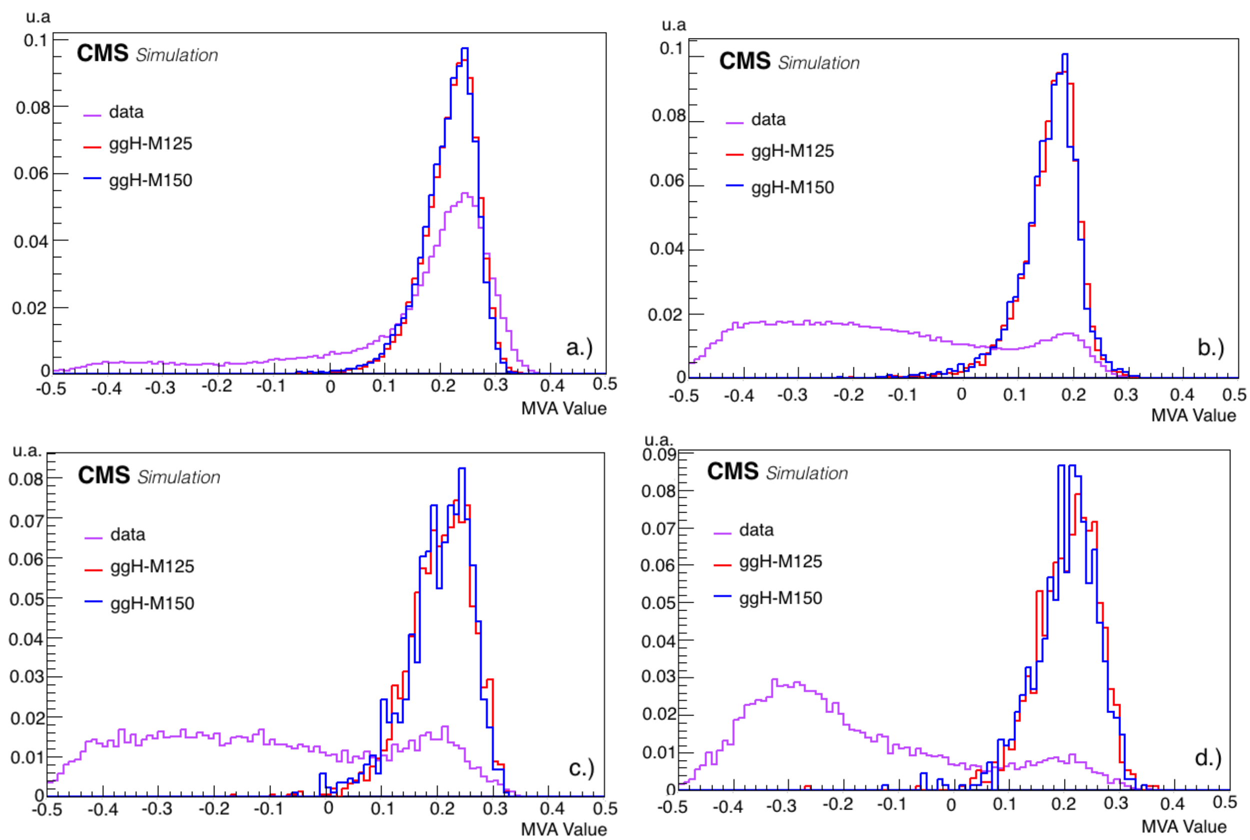

Figure 70: The MVA values for each category: a.) Category 1 b.) Category 2 c.) Category 3 and d.) Category 4.

	MVA Photon ID	
	ggH-125	ggH-150
Barrel	0.079	0.038
Endcap	0.0048	0.003
Category 1	0.069	0.033
Category 2	0.030	0.0138
Category 3	0.0067	0.0032
Category 4	0.0048	0.003

Table 38: Maximum MVA value for each category.

The significance is calculated in each category, in order to find an appropriate MVA value that could reject the most amount of background, and increase the number of signal events. CMS has several ways to calculate the significance, following a suggestion coming from the Higgs Group, we used

Significance $= 2(\sqrt{S+B} - \sqrt{B})$. The MVA values that gave us the best significance are summarized in Table 38.

After carefully analyzing the MVA photon ID implementation, it was widely used by the HZG group in the H$\rightarrow$Z^0 analysis and the Higgs-Dalitz decay at 13 TeV.

BIBLIOGRAPHY

[1] A. Djouadi. The anatomy of electroweak symmetry breaking: Tome I: The Higgs boson in the Standard Model. *Physics Reports*, 457(1- 4):1 – 216, 2008. ISSN 0370-1573. doi: http://dx.doi.org/10.1016/j.physrep.2007.10.004. URL http://www.sciencedirect.com/science/article/pii/S0370157307004334.

[2] CMS Collaboration. Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC. *Physics Letters B*, 716(1):30 – 61, 2012. ISSN 0370-269. doi: 10.1016/j.physletb.2012.08.021. URL http://www.sciencedirect.com/science/article/pii/S0370269312008581.

[3] CMS Collaboration. Observation of a New Boson with Mass Near 125 GeV in pp Collisions at $\sqrt{s} = 7$ and 8 TeV. *Journal of High Energy Physics*, 2013(6):81, 2013. URL https://link.springer.com/article/10.1007/JHEP06(2013)081.

[4] J. Ellis and T. You. Updated global analysis of higgs couplings. *Journal of High Energy Physics*, 2013(6):103, 2013. URL https://link.springer.com/article/10.1007/JHEP06(2013)103.

[5] J. Baglio and A. Djouadi. Higgs Production at the LHC. *Journal of High Energy Physics*, 2011(3):55, 2011. URL https://link.springer.com/article/10.1007/JHEP03(2011)055.

[6] LHC Higgs Cross Section Working Group. Handbook of LHC Higgs Cross Sections: 1. Inclusive Observables. *arXiv preprint arXiv:1101.0593*, 2011. URL https://arxiv.org/abs/1101.0593.

[7] M. Carena et al. Review of Particle Physics, 2014-2015. Status of Higgs boson physics . *Chin. Phys. C*, 38(arXiv:1412.1408):090001. 15 p, 2014. URL http://cds.cern.ch/record/1753419.

[8] CMS Collaboration. Combination of standard model Higgs boson searches and measurements of the properties of the new boson with a mass near 125 GeV. Technical Report CMS-PAS-HIG-12-045, CERN, Geneva, 2012. URL https://cds.cern.ch/record/1494149?ln=es.

[9] CMS Collaboration. A New Boson with a Mass of 125 GeV Observed with the CMS Experiment at the Large Hadron Collider. *Science*, 338(6114):1569–1575. 16 p, Sep 2012. URL http://science.sciencemag.org/content/338/6114/1569.

[10] D. Orestano. Search for the standard model Higgs boson with the ATLAS detector. Technical Report ATL-PHYS-PROC-2013-063, CERN, Geneva, Mar 2013. URL http://cds.cern.ch/record/1522494.

[11] K. Jakobs and M. Schuhmacher. Prospects for higgs boson searches at the lhc. *International Journal of Modern Physics A*, 23(32):5093–5115, 2008. URL https://doi.org/10.1142/S0217751X08042808.

[12] ATLAS Collaboration. Measurements of Higgs boson production and couplings in diboson final states with the ATLAS detector at the LHC. *Phys. Lett.*, B726:88–119, 2013. doi: 10.1016/j.physletb.2014.05.011,10.1016/j.physletb.2013.08.010. [Erratum: Phys. Lett.B734,406(2014)].

[13] S.Y. Choi et al. Identifying the Higgs spin and parity in decays to Z pairs. *Physics Letters B*, 553(1):61–71, 2003. URL https://doi.org/10.1016/S0370-2693(02)03191-X.

[14] CMS Collaboration. Properties of the Higgs-like boson in the decay H to ZZ to 4l in pp collisions at $\sqrt{s} = 7$ and 8 TeV. Technical Report CMS-PAS-HIG-13-002, CERN, Geneva, 2013. URL https://cds.cern.ch/record/1523767.

[15] CMS Collaboration. Evidence for a particle decaying to W+W- in the fully leptonic final state in a standard model Higgs boson search in pp collisions at the LHC. Technical Report CMS-PAS-HIG-13-003, CERN, Geneva, 2013. URL http://cds.cern.ch/record/1523673.

[16] ATLAS Collaboration. Evidence for Higgs Boson Decays to the $\tau^+\tau^-$ Final State with the ATLAS Detector. Technical Report ATLAS-CONF-2013-108, CERN, Geneva, Nov 2013. URL https://cds.cern.ch/record/1632191.

[17] CMS Collaboration. Search for the Standard-Model Higgs boson decaying to tau pairs in proton-proton collisions at $\sqrt{s} = 7$ and 8 TeV. Technical Report CMS-PAS-HIG-13-004, CERN, Geneva, 2013. URL https://cds.cern.ch/record/1528271.

[18] CMS Collaboration. Search for the standard model Higgs boson produced in association with W or Z bosons, and decaying to bottom quarks for LHCp 2013. Technical Report CMS-PAS-HIG-13-012, CERN, Geneva, 2013. URL https://cds.cern.ch/record/1546801.

[19] ATLAS Collaboration. Search for the $b\bar{b}$ decay of the Standard Model Higgs boson in associated $(W/Z)H$ production with the ATLAS detector. Technical Report arXiv:1409.6212. CERN-PH-EP-2014-214, CERN, Sep 2014. URL http://cds.cern.ch/record/1757231.

[20] CMS Collaboration. Search for a Higgs boson decaying into a Z and a photon in pp collisions at $\sqrt{s}$ =7 and 8 TeV. *Physics Letters B*, 726(4):587 – 609, 2013. ISSN 0370-2693. doi: http://dx.doi.org/10.1016/j.physletb.2013.09.057. URL http://www.sciencedirect.com/science/article/pii/S0370269313007880.

[21] ATLAS Collaboration. Search for Higgs boson decays to a photon and a Z boson in pp collisions at $\sqrt{s}$ =7 and 8 TeV with the ATLAS detector. *Physics Letters B*, 732:8 – 27, 2014. ISSN 0370-2693. doi: http://dx.doi.org/10.1016/j.physletb.2014.03.015. URL http://www.sciencedirect.com/science/article/pii/S0370269314001713.

[22] CMS Collaboration. Search for a Higgs boson decaying into $\gamma^*\gamma \to \ell\ell\gamma$ with low dilepton mass in pp collisions at $\sqrt{s} = 8$ TeV. *Phys. Lett. B*, 753(CMS-HIG-14-003. CMS-HIG-14-003. CERN-PH-EP-2015-137):341. 22 p, Jul 2015. URL https://cds.cern.ch/record/2033215. Replaced with published version. Added journal reference.

[23] CMS Collaboration. Search for a standard model-like Higgs boson in the $\mu^+\mu^-$ and e^+e^- decay channels at the LHC. *Phys. Lett. B*, 744(CMS-HIG-13-007. CMS-HIG-13-007. CERN-PH-EP-2014-243):184. 24 p, Oct 2014. URL https://cds.cern.ch/record/1957177.

[24] ATLAS Collaboration. Search for a Standard Model Higgs boson in $H \to \mu\mu$ decays with the ATLAS detector. Technical Report ATLAS-CONF-2013-010, CERN, Geneva, Mar 2013. URL http://cds.cern.ch/record/1523695.

[25] M. Battaglia and A. De Roeck. Determination of the muon yukawa coupling at high energy e+ e-linear colliders. *arXiv preprint hep-ph/0111307*, 2001. URL https://arxiv.org/abs/hep-ph/0111307.

[26] CMS Collaboration. LHC Machine. *JINST*, 3(S08001), 2008. URL http://iopscience.iop.org/1748-0221/3/08/S08001.

[27] M. Spiropulu and S. Stapnes. LHC's ATLAS and CMS detectors. *International Journal of Modern Physics A*, 23(25):4081–4105, 2008. doi: 10.1142/S0217751X08042341.

[28] CMS Public Web. *Public CMS Luminosity Information*, 2014. URL https://twiki.cern.ch/twiki/bin/view/CMSPublic/LumiPublicResults.

[29] CMS Public Web. *CMS Detector Design*, 2012. URL http://cms.web.cern.ch/news/cms-detector-design.

[30] CMS Collaboration. The CMS Experiment at the CERN LHC. *JINST*, 3(S08004), 2008. doi: 10.1088/1748-0221/3/08/S08004. URL http://iopscience.iop.org/1748-0221/3/08/S08004.

[31] CMS Collaboration. Technical proposal for the upgrade of the CMS detector through 2020. Technical Report CERN-LHCC-2011-006. LHCC-P-004, CERN, Geneva, Jun 2011. URL https://cds.cern.ch/record/1355706.

[32] CMS Public Web. *Tracker Detector*, 2012. URL http://cms.web.cern.ch/news/trackerdetector.

[33] CMS public website. *Electromagnetic Calorimeter*, 2012. URL http://cms.web.cern.ch/news/electromagnetic-calorimeter.

[34] CMS Collaboration. CMS Technical Design Report for the Phase 1 Upgrade of the Hadron Calorimeter. Technical Report CERN-LHCC-2012-015. CMS-TDR-10, CERN, Geneva, Sep 2012. URL https://cds.cern.ch/record/1481837.

[35] A. Colaleo et al. CMS Technical Design Report for the Muon Endcap GEM Upgrade. Technical Report CERN-LHCC-2015-012. CMS-TDR-013, CERN, Geneva, Jun 2015. URL https://cds.cern.ch/record/2021453.

[36] CMS Collaboration. *The CMS muon project: Technical Design Report*. Technical Design Report CMS. CERN, Geneva, 1997. URL https://cds.cern.ch/record/343814.

[37] A. Sharma. Muon tracking and triggering with gaseous detectors and some applications. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 666:98 – 129, 2012. ISSN 0168-9002. doi: http://dx.doi.org/10.1016/j.nima.2011.12.001. URL http://www.sciencedirect.com/science/article/pii/S0168900211021607. Advanced Instrumentation.

[38] CMS public website. *Resistive Plate Chambers*, 2012. URL http://cms.web.cern.ch/news/resistive-plate-chambers.

[39] CMS Collaboration. CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems. Technical report, CERN, 2000. URL https://cds.cern.ch/record/706847.

[40] CMS Collaboration. *CMS The TriDAS Project: Technical Design Report, Volume 2: Data Acquisition and High-Level Trigger. CMS trigger and data-acquisition project*. Technical Design Report CMS. CERN, Geneva, 2002. URL https://cds.cern.ch/record/578006.

[41] GL. Bayatian, A. Korablev, et al. CMS Physics: Technical Design Report Volume 1: Detector Performance and Software. Technical report, CMS-TDR-008-1, 2006. URL https://cds.cern.ch/record/922757.

[42] F. Beaudette. The CMS Particle Flow Algorithm. *arXiv preprint arXiv:1401.8155*, 2014. URL https://arxiv.org/abs/1401.8155.

[43] A. Wolfgang, A. Strandlie, R. Frühwirth, and T. Todorov. Reconstruction of electrons with the gaussian-sum filter in the cms tracker at the lhc. *Journal of Physics G: Nuclear and Particle Physics*, 31(9):N9, 2005. URL https://cds.cern.ch/record/815410?ln=en.

[44] CMS Collaboration. Particle-flow reconstruction and global event description with the cms detector. *arXiv preprint arXiv:1706.04965*, 2017. URL https://arxiv.org/abs/1706.04965.

[45] CMS collaboration. Performance of CMS muon reconstruction in pp collision events at *sqrt* =7 TeV. *Journal of Instrumentation*, 7(10):P10002, 2012. doi: https://doi.org/10.1088/1748-0221/7/10/P10002.

[46] CMS MuonPOG. *Baseline muon selections for Run-II*, 2017. URL https://twiki.cern.ch/twiki/bin/viewauth/CMS/SWGuideMuonIdRun2#Muon_Isolation.

[47] CMS collaboration. Performance of electron reconstruction and selection with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV. *arXiv preprint arXiv:1502.02701*, 2015. URL https://arxiv.org/abs/1502.02701.

[48] S. Baffioni et al. Electron reconstruction in CMS. *The European Physical Journal C-Particles and Fields*, 49(4):1099–1116, 2007. doi: 10.1140/epjc/s10052-006-0175-5.

[49] F. Beaudette, D. Benedetti, P. Janot, and M. Pioppi. Electron reconstruction within the particle flow algorithm. In *CMS AN*, volume 34, page 2010, 2010.

[50] CMS EGammaPOG. *Cut Based Electron ID for Run 2*, 2017. URL https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedElectronIdentificationRun2.

[51] CMS EGammaPOG. *Multivariate Electron Identification for Run 2*, 2017. URL https://twiki.cern.ch/twiki/bin/view/CMS/MultivariateElectronIdentificationRun2.

[52] J. Nysten. Photon reconstruction in CMS. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 534(1):194–198, 2004. URL https://doi.org/10.1016/j.nima.2004.07.071.

[53] CMS collaboration. Performance of photon reconstruction and identification with the CMS detector in proton-proton collisions at $\sqrt{s} = 8$ TeV. *arXiv preprint arXiv:1502.02702*, 2015. URL https://arxiv.org/abs/1502.02702.

[54] CMS EGammaPOG. *Cut Based Photon ID for Run 2*, 2017. URL https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedPhotonIdentificationRun2.

[55] CMS EGammaPOG. *General Purpose MVA-based Photon Identification*, 2017. URL https://twiki.cern.ch/twiki/bin/view/CMS/MultivariatePhotonIdentificationRun2.

[56] CMS Collaboration. Performance of Jet Algorithms in CMS. Technical Report CMS-PAS-JME-07-003, CERN, Geneva, 2009. URL https://cds.cern.ch/record/1198227.

[57] D. Dobur. Jets and missing transverse energy reconstruction with CMS. *arXiv preprint arXiv:0904.0391*, 2009. URL https://arxiv.org/abs/0904.0391.

[58] CMS Collaboration. Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET. Technical Report CMS-PAS-PFT-09-001, CERN, Geneva, Apr 2009. URL https://cds.cern.ch/record/1194487.

[59] CMS BTagPOG. *b Tag & Vertexing Physics Object Group*, 2017. URL https://twiki.cern.ch/twiki/bin/view/CMS/BtagPOG.

[60] CMS JetMETPOG. *Jet Identification*, 2017. URL https://twiki.cern.ch/twiki/bin/view/CMS/JetID#RECO_using_RECOed_variables.

[61] CMS BTagPOG. *Recommendation for Using b-tag Objects in Physics Analyses*, 2017. URL https://twiki.cern.ch/twiki/bin/view/CMS/BtagRecommendation.

[62] CMS TauPOG. *Tau Identification Physics Object Group*, 2017. URL https://twiki.cern.ch/twiki/bin/view/CMS/Tau.

[63] CMS JetMETPOG. *Missing Transverse Energy (MET)*, 2017. URL https://twiki.cern.ch/twiki/bin/view/CMS/MissingET.

[64] D. Acosta, P. Bortignon, D. Bourilov, et al. Search for standard model Higgs boson production in the $\mu^+\mu-$ final state at CMS in pp collisions at $\sqrt{s}$ =13 TeV. CMS Analysis Note, 2017.

[65] CMS Collaboration. Search for the standard model Higgs boson decaying to two muons in pp collisions at $\sqrt{s}$ =13 TeV. Physical Review Letters, 2017.

[66] P. Nason. A New Method for Combining NLO QCD with Shower Monte Carlo Algorithms. *JHEP*, 11(hep-ph/0409146. BICOCCA-FT-2004-11):040. 29 p, Sep 2004. URL https://cds.cern.ch/record/792756.

[67] H2Mu Analysis group. *HiggsToMuMu at 13 TeV (2015/16)*, 2017. URL https://twiki.cern.ch/twiki/bin/view/CMS/HiggsToMuMu.

[68] J. Alwall, M. Herquet, F. Maltoni, O. Mattelaer, and T. Stelzer. MadGraph 5: going beyond. *Journal of High Energy Physics*, 2011(6):128, Jun 2011. ISSN 1029-8479. doi: 10.1007/JHEP06(2011)128. URL https://doi.org/10.1007/JHEP06(2011)128.

[69] CMS Physics Validation Group. *Pileup Reweighting Utilities*, 2013. URL https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupMCReweightingUtilities.

[70] CMS Collaboration. Search for standard model Higgs boson production decaying into two muons in pp collisions at $\sqrt{s}$ =13 TeV. CMS Physics Analysis Summaries, 2017.

[71] CMS Muon POG. *Rochester Correction*, 2017. URL https://twiki.cern.ch/twiki/bin/viewauth/CMS/RochcorMuon.

[72] J. Han and Bodek. A. *Rochester Muon Momentum Correction*, 2017. URL https://www-cdf.fnal.gov/~jyhan/cms_momscl/cms_rochcor_manual.html.

[73] D. de Florian et al. Handbook of LHC Higgs Cross Sections: 4. Deciphering the Nature of the Higgs Sector. Technical report, CERN, 2016. URL http://cds.cern.ch/record/2227475.

[74] CMS Public Web. *WorkBookMiniAOD*, 2017. URL https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookMiniAOD#MC_for_Moriond_2017.

[75] CMS Public Web. *WorkBookMiniAOD2015*, 2017. URL https://twiki.cern.ch/twiki/bin/view/CMSPublic/WorkBookMiniAOD2015#Analyzing_MiniAOD.

[76] CMS Collaboration. Search for a standard model-like Higgs boson in the $\mu^+\mu^-$ and e^+e^- decay channels at the LHC. *Phys. Lett. B*, 744(CMS-HIG-13-007. CMS-HIG-13-007. CERN-PH-EP-2014-243):184. 35 p, Oct 2014. URL https://cds.cern.ch/record/1957177. Comments: Submitted to Phys. Lett. B.

[77] CMS Web. *HiggsZZ4l2016*, 2016. URL https://twiki.cern.ch/twiki/bin/viewauth/CMS/HiggsZZ4l2016#Photons_for_FSR.

[78] CMS Collaboration. Studies of Higgs boson production in the four-lepton final state at $\sqrt{s}$ = 13 TeV. Technical Report CMS-PAS-HIG-15-004, CERN, Geneva, 2016. URL http://cds.cern.ch/record/2139978.

[79] W. Verkerke and D. Kirkby. The RooFit toolkit for data modeling. Technical Report physics/0306116, SLAC, Stanford, CA, Jun 2003. URL http://cds.cern.ch/record/622147.

[80] W. Verkerke and D. Kirkby. *RooFit users manual v2. 91*, 2008. URL https://root.cern.ch/download/doc/RooFit_Users_Manual_2.91-33.pdf.

[81] CMS Collaboration. Muon Reconstruction and Identification Improvements for Run-2 and First Results with 2015 Run Data. Technical Report CMS-DP-2015-015, CERN, Geneva, Jul 2015. URL https://cds.cern.ch/record/2037372.

[82] H. Voss, H. Höcker, J. Stelzer, and F. Tegenfeldt. TMVA: Toolkit for Multivariate Data Analysis with ROOT. TMVA, the Toolkit for Multivariate Data Analysis with ROOT. *PoS*, ACAT:040, 2007. URL https://cds.cern.ch/record/1116810.

[83] H. Voss. *Data Analysis With TMVA*, 2010. URL http://lphe.epfl.ch/seminar/extern/Voss_12_04_10.pdf.

[84] LF. Chaparro. *H2Mu FSR studies code*, 2017. URL https://github.com/lchaparr/HtoMuMu.

[85] LF. Chaparro, J. Sanabria, and A. Perieanu. FSR photons in SM Higgs dimuon channel analysis at 13 TeV. CMS Analysis Note, 2017.

[86] A. Hoecker. Physics at the LHC Run-2 and Beyond. In *2016 European School of High-Energy Physics (ESHEP 2016) Skeikampen, Norway, June 15-28, 2016*, 2016. URL https://inspirehep.net/record/1499881/files/arXiv:1611.07864.pdf.

[87] G Apollinari, I. Béjar Alonso, O. Brüning, M. Lamont, and L. Rossi. *High-Luminosity Large Hadron Collider (HL-LHC): Preliminary Design Report*. CERN Yellow Reports: Monographs. CERN, Geneva, 2015. URL https://cds.cern.ch/record/2116337.

[88] G. Apollinari, O. Brüning, T. Nakamoto, and L. Rossi. High Luminosity Large Hadron Collider HL-LHC. *CERN Yellow Report*, (5):1–19, 2015. doi: 10.5170/CERN-2015-005.1.

[89] K. Peters. Prospects for beyond Standard Model Higgs boson searches at future LHC runs and other machines. *PoS*, CHARGED2016:028, 2017.

[90] CMS Web. *HLT Configurations web eXplorer*, 2016. URL https://cmsweb.cern.ch/confdb/.

[91] D. Olivito. *2016 Trigger HATS @ LPC*, 2016. URL https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideHATS2016Trigger.

[92] LF. Chaparro. *Changing pt cuts of muon + photon trigger*, 2015. URL https://its.cern.ch/jira/browse/CMSHLT-590.