Isfahan University of Technology
Physics Department

A thesis presented for the degree of
Doctor of Philosophy

Search for an exotic decay of the Higgs boson to a pair of light pseudoscalars in proton-proton collisions at 13 TeV and Luminosity measurement in Pb-Pb collisions at 5.02 TeV

by

Elham Khazaie

Supervisors:

Dr.Ahmad Shirzad, Dr.Hamed Bakhshian

Advisor:

Dr.Abideh Jafari

Iran, June 2023

CERN-THESIS-2023-320

Isfahan University of Technology
Department of Physics

Search for an exotic decay of the Higgs boson to a pair of light pseudoscalars in proton-proton collisions at 13 TeV and luminosity measurement in Pb-Pb collisions at 5.02 TeV

A Thesis
Submitted in partial fulfillment of the requirements
for the degree of Ph.D.

by
Elham Khazaie

Evaluated and Approved by the Thesis Committee, on June 21, 2023

Supervisor: Dr. A. Shirzad, Assoc. Prof

Supervisor: Dr. H. Bakhshian, Assis. Prof

Advisor: Dr. A. Jafari, Assis. Prof

Examiner: Dr. B. Mirza, Prof

Examiner: Dr. E. Gallo, Prof

Examiner: Dr. M. Mohammadi, Prof

Examiner: Dr. M. Etesami, Assoc. Prof

Department Graduate Coordinateor: Dr. I. Abdolhosseini Sarsari

ABSTRACT

In the first part of the thesis, we report the results of a search for exotic decay of a Higgs boson with $m_H = 125$ GeV to a pair of new light bosons, a_1, where one of the light bosons decays to a pair of muons and the other to a pair of b quarks. Such signatures are predicted in a number of well-motivated extensions of the standard model, including the next-to-minimal supernumerary and generic two-Higgs-doublet models with an additional scalar singlet. A data sample corresponding to an integrated luminosity of 138 fb^{-1} recorded with the CMS detector during 2016-2018 is exploited in $\mu\mu bb$ final state where no statistically significant excess is observed with respect to the standard model backgrounds for different m_{a_1} hypotheses above 15 GeV and below $m_H/2$. Upper limits on $\mathcal{B}(H \rightarrow a_1 a_1 \rightarrow \mu\mu bb)$ are obtained to be $(0.17 - 3.3) \times 10^{-4}$ depending on the m_{a_1} values. The second part of the thesis includes a study of luminosity measurement using lead-lead (PbPb) collisions at 5.02 TeV in 2015.

DEDICATION

In loving memory of my father, this thesis is dedicated to my mother, who has been my pillar of strength and unwavering support throughout my academic journey.

ACKNOWLEDGMENTS

Throughout my doctoral journey, I have been incredibly fortunate to receive exceptional support from numerous individuals.

I would like to express my sincere gratitude to my supervisor, Dr. Ahmad Shirzad, for helping me discover my genuine interest in the experimental side of high energy physics. I am truly thankful for his valuable advice and unwavering support.

To Dr. Hamed Bakhshian, my supervisor, who helped me understand the fundamental principles of experimental particle physics. Throughout the process of developing my research proposal and ultimately completing my thesis, he consistently demonstrated kindness, patience, and unwavering support that positively influenced my academic success.

To Dr. Abideh Jafari for her consistent support, encouragement, and kindness throughout this thesis. Without her assistance and support, this thesis would not have been possible. Working under the supervision of Dr. Hamed Bakhshian and Dr. Abideh Jafari has been an incredibly enjoyable experience, from which I have learned and developed significantly.

I am grateful to Dr. Maryam Zeinali for her assistance and encouragement throughout the process. Furthermore, I would like to acknowledge Dr. Jan Steggemann, Dr. Adinda De Wit, Dr. Stephane Cooperstein, Dr. Ketino Kaadze, and Dr. Philippe Gras for their valuable comments and support during the development of this analysis.

I am also thankful to the members of the jury, Prof. Behrouz Mirza, Dr. Mojtaba Mohammadi, and Dr. Mohsen Etesami, for their invaluable comments. In particular, I would like to express my appreciation to Prof. Elisabetta Gallo for taking the time to read through this thesis and provide me with her valuable feedback.

It was with great pleasure that I was able to work with researchers at Wisconsin Madison University and Princeton University on part of my thesis. I would like to express my gratitude to Dr. Pieter Everaerts, Dr. Isobel Ojalvo, Dr. Cecile Caillol, Dr. Pallabi Das, Stephanie Kwan, and Ho Fung Tsoi.

Finally, I want to convey my sincere appreciation to my family and my friends. The unwavering belief and support of my brothers have been crucial in my achievements. I owe a great deal to my mother for her love and encouragement, without which I would not have been able to accomplish this work.

Elham Khazaie

CONTENTS

Part I

Exotic Higgs decays

CHAPTER

ZERO

INTRODUCTION

Everything is a mathematical trick,
except what you measure in the lab.

Prof. Dr. Armin Scrinzi

In terms of science history, the Standard Model (SM) of fundamental particles is undoubtedly a masterpiece. It includes the work of 13 Nobelists since the 20th century, and delivered accurate evidence of its predictions without a doubt. In particle physics, SM describes the quantum mechanical nature of three fundamental forces of nature, the electromagnetic, strong, and weak forces. Further, it discusses the interactions between force carriers and the fundamental constituents of matter, quarks and leptons. Since 2010, when the Large Hadron Collider (LHC) began colliding protons, one of its primary goals has been to find the last missing piece of the SM: the Higgs boson. In 2012, the CMS and ATLAS experiments [1,2] discovered the Higgs boson and completed the understanding of the SM. Ten years after the discovery of the Higgs boson, we have gained more knowledge about this particle. An overview of the Higgs boson has been published in the journal Nature by two experiments, CMS and ATLAS. Fig. 1 shows the couplings of the Higgs boson with fermions and gauge bosons in terms of particle mass. These measurements are consistent with the predictions of the SM of particle physics, within the uncertainties. So far, in an exceptional degree, SM has been experimentally validated and verified [3, 4].

However, there are several mysteries that cannot be explained by the SM, including Dark Matter, the weakness of the gravitational force, and the hierarchy problem. Therefore, SM is an incomplete theory. There are a number of beyond the standard models (BSM) that can provide answers to these mysteries, while identifying the 125 GeV resonance as part of an extended group of scalar particles. The simple extensions of the SM are two-Higgs-doublet models (2HDM) [5,6], where two doublets of scalar fields are considered leading to five physical states after symmetry breaking. The two-Higgs-doublet model with extra singlet (2HDM+S) [7,8] obtained by adding a complex scalar singlet to already present scalar doublets in 2HDM results into addition of two new bosons. In all these models, one of the scalars is identified as the discovered Higgs boson with a mass of 125 GeV.

A recent measurement of the Higgs boson couplings at the LHC does not rule out the possibility of exotic Higgs decays. ATLAS and CMS experiments, respectively, put upper bounds of 12% and 16% on the branching fraction of the Higgs boson to new particles at 95% confidence levels (CL) using LHC data from Run-II [3,4]. As a result of these bounds, it is important to examine the data for direct evidence of new particles coupling to the Higgs boson, and especially to examine possible extensions of the SM of particle physics.

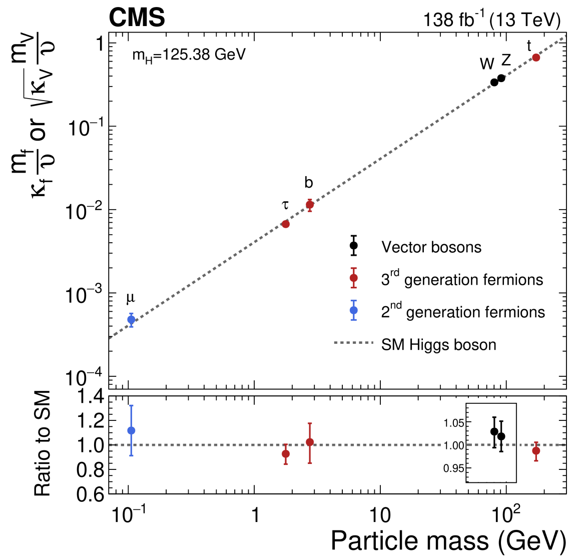


Figure 1: Couplings of the Higgs boson with fermions and gauge bosons in terms of particles mass.

A significant part of the exotic decays of the Higgs boson consists of the Higgs decay into a pair of light (pseudo)scalar bosons and then into a pair of SM particles. Such decays are well-motivated in the context of 2HDM+S. There are four types of such models. All SM particles couple to the first Higgs doublet, Φ_1, in type I models. In type II, up-type quarks couple to Φ_1 while leptons and down-type quarks couple to the second Higgs doublet, Φ_2. Type II models have NMSSM as a special case where the additional singlet is introduced to resolve the so-called μ-problem of the MSSM superpotential. Quarks couple to Φ_1 and leptons couple to Φ_2 in type III and finally in type IV, leptons and up-type quarks couple to Φ_1, while down-type quarks couple to Φ_2. After the electroweak symmetry breaking, the 2HDM models allow for a pair of charged Higgs bosons, $H^{\pm}$, a neutral pseudoscalar, A, and two neutral scalar mass eigenstates, H and h. One can assume that the model is in a decoupling limit such that the lighter scalar eigenstate, H, is the discovered boson with $m_H = 125\,GeV$. Furthermore, to allow for a wider variety of exotic Higgs decay phenomenology, a complex scalar singlet, $S_R + iS_I$ is added to the model which has no direct Yukawa couplings. Hence, it is expected to decay to SM fermions by virtue of mixing with the Higgs sector. This mixing is small enough to preserve the SM-like nature of the H boson. The Higgs boson can decay to fermions through $a_1(s_1)$ where $a_1(s_1)$ is a (pesudo)scalar mass eigenstate mostly composed of $S_R\,(S_I)$.

While the SM-like couplings of the Higgs boson to both fermions and gauge bosons can be preserved, the singlet state of the 2HDM+S may as well be a dark matter candidate that couples to the Higgs boson [9, 10]. The possibility of the Higgs boson decaying into pseudoscalar a_1 is studied in this thesis for the final state of $H \to a_1 a_1 \to \mu\mu bb$, as illustrated in Fig. 2. The gluon fusion production mechanism (ggH) is considered as the dominant Higgs production process with the next-to-leading-order production rate of $\sigma_{ggH} \simeq 48.58 \pm 1.89\,pb$ [11]. The contribution of the vector boson fusion (VBF) Higgs boson production process is also taken into account with the next-to-leading-order production cross section $\sigma_{VBF} = 3.93 \pm 0.08\,pb$ [11]. As a benchmark, a branching fraction 10% is assigned to $H \to a_1 a_1$ where assuming $\tan\beta = 2$, one can obtain $2 \times \mathcal{B}(a_1 \to b\bar{b})\mathcal{B}(a_1 \to \mu^+\mu^-) = 1.7 \times 10^{-3}$ for $m_{a_1} = 30\,GeV$ in type-III 2HDM+S [12].

The search for the exotic a_1 particle is performed with $15 \leq m_{a_1} \leq 62.5\,GeV$. The lower bound is set in order to ignore intermediate states with quarkonia, whose inter-

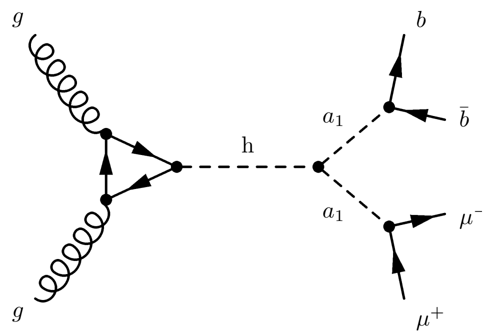


Figure 2: Feynman diagram for the signal topology.

pretation is complicated by non-perturbative QCD effects. Moreover, the sensitivity of the search largely decreases towards $m_{a_1} \approx 15\,GeV$ and lower. The upper bound is imposed by the Higgs mass. The signal selection is optimized for a $H \to a_1 a_1 \to \mu\mu bb$ signal. The analysis is performed using the data collected with the CMS detector during Run-II LHC, corresponding to an integrated luminosity of $138\,fb^{-1}$.

Similar searches are performed at the LHC, at different center-of-mass energies and data samples. The latest analysis by ATLAS [13] has placed a strong bound of $\mathcal{B}(H \to b\bar{b}\mu^+\mu^-) < (0.22 - 4) \times 10^{-4}$ using the entire LHC Run-II data. This was an update on a former search with the partial data sample [14]. The CMS search at $\sqrt{s} = 13\,TeV$ [15] is based on $36fb^{-1}$ of data results in an upper limit of $(1-7) \times 10^{-4}$ for the Higgs boson branching ratio in $\mu\mu bb$ final state. At 8 TeV, the experiment has provided an upper bound of about 2.5 times the expectation for the same branching ratio [16]. The analysis presented in this thesis is an update of Ref. [15] including a more in-depth study of the signal to achieve more stringent limits than what is offered by pure addition of the full Run-II LHC data. Ultimately, these results will be combined with the $\tau\tau bb$ channel to increase sensitivity.

This thesis consists of two parts: study of the exotic decay of the Higgs boson and luminosity measurement. A flowchart of the analysis is presented in Fig. 3 to visualize the steps. First, we will investigate the exotic decay of the Higgs boson in the proton-proton collision, and then we will study the luminosity measurement for the lead-lead

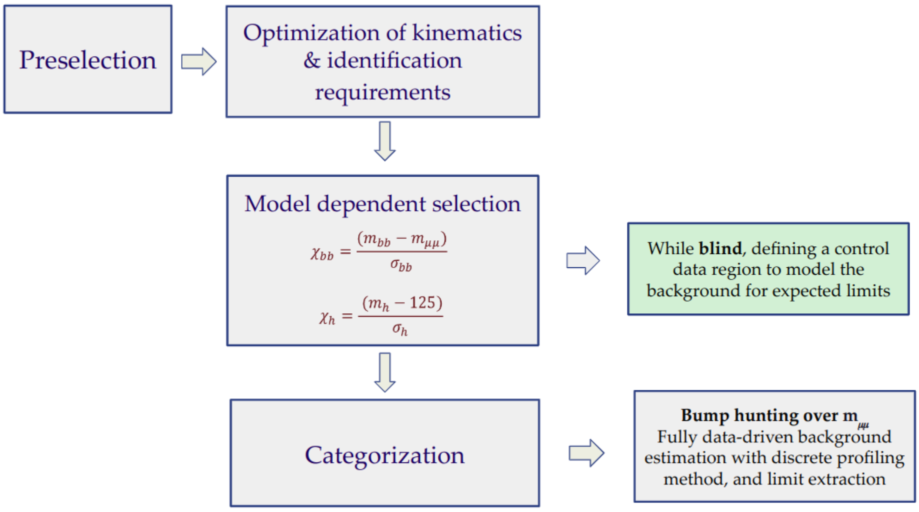


Figure 3: Analysis steps in a nutshell. The green box corresponds to the CR defined in chapter 4 and it is there only for the blind analysis, i.e., will be removed after unblinding.

collision. The first part is arranged as follows: Chapter 1 describes the LHC and the CMS. Chapter 2 deals with a brief presentation of the SM along with its limitations, which made it necessary to introduce new theories. 2HDM+S model is introduced in this chapter as one of the theoretical models beyond the standard model, which enables this analysis. Chapter 3 explains the steps of performing the analysis conducted by the author. This chapter contains a brief presentation of event generation processes, reconstruction of physical objects used in this analysis, simulated samples for signal and background reconstruction along with their correction scale factors, and selection of events. Chapter 4 explain signal and background modeling together with systematic uncertainties. Finally, chapter 5 presents the results of this analysis and search, as well as the combination of this search with another one. The second part of the thesis will be presented in chapter 1, which will study the luminosity measurement using lead-lead (PbPb) collisions at $5.02\,TeV$ in 2015.

CHAPTER

ONE

THE LHC AND THE CMS DETECTOR

The chapter starts by introducing the Large Hadron Collider (LHC), which is the primary apparatus responsible for generating high energy collision events. Section 1.2 then provides an explanation of the Compact Muon Solenoid (CMS) detector, which is responsible for collecting the data produced by the LHC.

1.1 The LHC

The LHC is the largest and most powerful circular particle collider ever constructed. The first collision occurred in 2009 [17]. It is located between the Jura mountain range and Lac Léman, some 100 m underground, and spans the border between France and Switzerland. The reason for locating it underground is that the energetic particles produced by it lose their energy before reaching the surface of the earth, and thus are not a threat to safety and health (radioactive metals). One of the main goals that led to the design and construction of the LHC as well as its multipurpose detectors was to search for the SM scalar boson. Before LHC operation, it was determined from the results of LEP and Tevatron experiments that the mass of the SM scalar boson should be greater than about 114 GeV and less than 1 TeV. The motivations for building the LHC are clearly broader than finding the Higgs boson and may include things such as supersymmetry or the search for dark matter as well. Throughout the tunnel, there are four main collision detectors that are located at different collision points (Fig. 1.1):

CMS[1], ATLAS[2], LHCb[3], and ALICE[4].

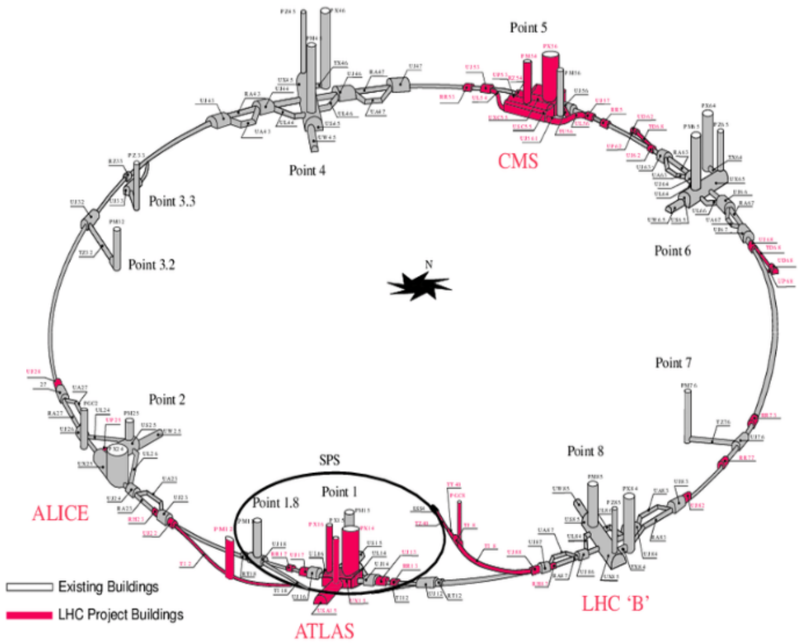


Figure 1.1: The overview of LHC circular underground tunnel with four LHC main experiments located at different collision points of the tunnel: CMS, ATLAS, LHCb and ALICE.

In order to produce beams containing high-energy hadrons, a large acceleration complex is required that can accelerate the hadrons from near rest to speeds close to that of light. The first part of this research focuses solely on proton-proton collisions, so the following description of the CERN accelerator complex is focused on the group of machines that accelerate such particles. Protons that are used in the acceleration process are extracted from molecular hydrogen, which has had its electrons stripped away to create an ionized gas that can be accelerated with an electric field. As a first step to accelerating the protons, LINAC2 [18] is equipped with radiofrequency cavities that are used to charge the cylindrical conductors, through which they reach an energy of 50 MeV. The PSB feeds a larger circular accelerator called the Proton Synchrotron

[1] Compact Muon Solenoid
[2] A Toidal LHC Apparatus
[3] Large Hadron Collider Beauty
[4] A Large Ion Collider Experiment

(PS), which is equipped with approximately 277 electromagnets that ensure the circular movement of the protons along with an energy that is approximately 25 times the rest mass of the proton. The Super Proton Synchrotron (SPS) [19] is the next accelerator in the chain, and the final step before injection into the LHC. The SPS is the second largest machine in the CERN accelerator complex with a circumference of nearly 7 km, which makes it capable of reaching the respectable value of 450 GeV. The Fig. 1.2 presents a comprehensive depiction of the previous description of the accelerator chain, as well as the various experiments conducted at CERN. A proton beam can be accelerated to

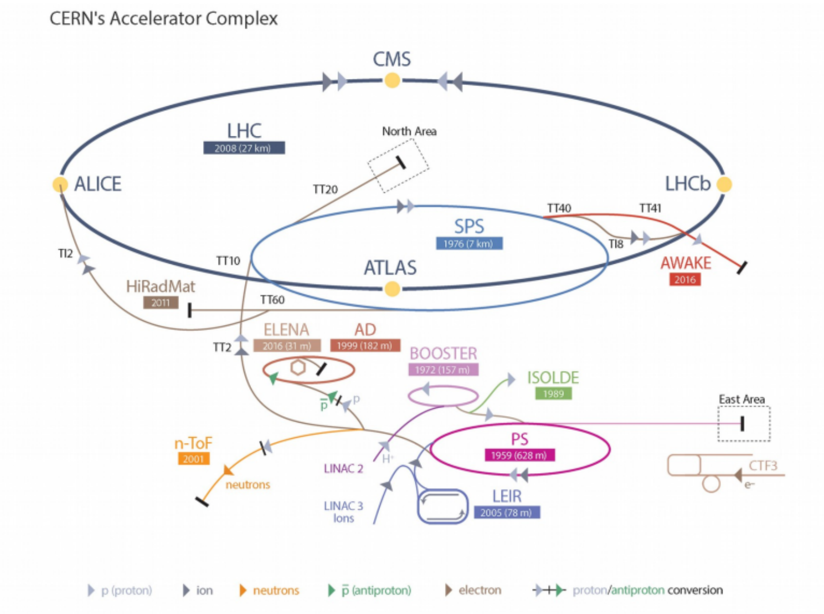


Figure 1.2: Overview of the CERN acceleration complex. To reach their final energy protons successively pass through the LINAC2, the PSB (Proton Synchrotron Booster), the PS (Proton Synchrotron), the SPS (Super Proton Synchrotron) and finaly the LHC (Large Hadron Collider).

a maximum design energy of 7 TeV once it reaches the LHC. The protons are bundled together during the acceleration process into 2808 bunches containing a hundred billion protons separated by 25 nanoseconds rather than having continuous beams. At 6.5 TeV (the operating energy in the main period of 2016-2018), the protons give around 11245 turns in one second along the 27 km ring. This is made possible by radio frequency cavities (eight per beam) inside the LHC, which are cooled to 4.5 K using liquid Helium and can provide an alternating electric field of 400 MHz [20]. Additionally it is crucial for the beams to be kept in a circular path throughout their journey, and in order to ensure that this is achieved, the LHC needs to install 1232 dipole magnets, each about 15 m long and weighing 30 tons, as well as 392 quadrupole magnets, to ensure the beams remain focused. The dipole magnets are superconductors capable of operating at 1.9 K and reaching a magnetic field of 8.33 T. Magnets of higher multipole orders are used to correct smaller imperfections in the field configuration.

1.2 The CMS detector

The central feature of the CMS apparatus is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. Within the solenoid volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron calorimeter (HCAL), each composed of a barrel and two endcap sections. Forward calorimeters extend the pseudorapidity coverage provided by the barrel and endcap detectors. Muons are measured in gas-ionization detectors embedded in the steel flux-return yoke outside the solenoid. Events of interest are selected using a two-tiered trigger system. A brief description of each component of CMS is presented in the following paragraphs, and more detailed description of the CMS detector can be found in Ref. [21]. A schematic representation of the CMS detector and its various components can be found in figure 1.3.

1.2.1 CMS Coordinate System

In the coordinate system used by CMS, the origin is selected as the nominal collision point. The z-axis points in the anticlockwise direction of the proton beam, the x-axis is placed radially towards the center of the LHC, and the y-axis is so that the Cartesian coordinate system is dextrogyral. The equivalent polar coordinate system is defined accordingly, with the ϕ angle measured from the x-axis and the radius r being the radius in the x-y plane. The angle θ identifies the polar angle in the r-z plane (Fig. 1.4).

The use of invariant quantities under Lorentz boosts along the z-axis is essential, since collisions suffer from boosts in the z-axis direction (longitudinal momentum fractions of partons). The quantity called rapidity y has the property that the difference

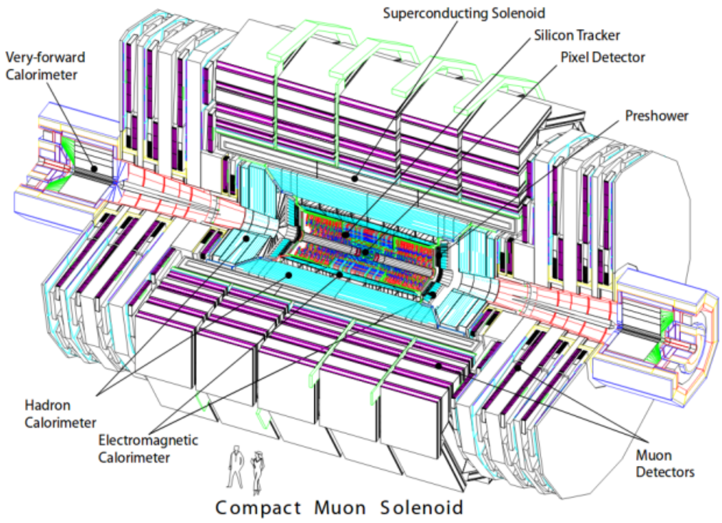


Figure 1.3: General view of the CMS detector structure.

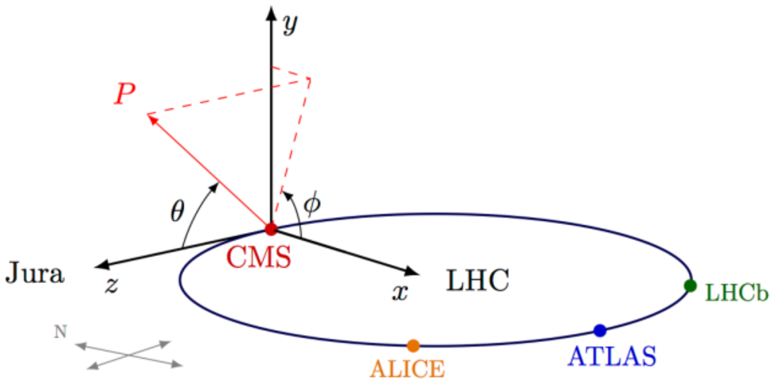


Figure 1.4: The CMS coordinate system.

δy is invariant with respect to such Lorentz boosts [22]. The rapidity is given by

$$y = \frac{1}{2} ln(\frac{E + p_z}{E - p_z}), \tag{1.1}$$

where p_z is the z-component of the three-momentum $\vec{p}$ and E is the energy. Due to the high energy of the collisions taking place at the LHC, the particles in these collisions are typically in the ultrarelativistic approximation ($E \approx |\vec{p}|$), so it is convenient to use, instead, the pseudorapidity η

$$\eta = -ln\tan(\frac{\theta}{2}). \tag{1.2}$$

In the ultrarelativistic limit, the pseudorapidity becomes almost equal to the rapidity and is much easier to measure for highly energetic particles.

1.2.2 Superconducting Magnet

An important component of the CMS detector is the superconducting solenoid magnet. The magnet surrounds all calorimeters and is the largest ever constructed. In this way, all the calorimeters are placed inside the coil, resulting in a compact detector. It has an inner diameter of 5.9 m and a length of 12.9 m. It weighs 12000 tonnes. With its 3.8 T magnetic field, it is approximately 100000 times more powerful than the Earth's.

1.2.3 Inner Tracking System

With the inner tracking system [23] of CMS, the trajectory of charged particles associated with the magnetic field is determined precisely, along with the primary and secondary vertices of the event. The tracker consists of two subsystems; a pixel detector extending radially away from the beamline from 4.4 cm to 10.2 cm, and a silicon strip tracker covering 20 cm to 116 cm in radius. The silicon modules take up a total active area of more than 200 m^2 and cover a pseudorapidity range $|\eta| < 2.5$. The pixel detector is the closest component to the interaction point and consists of three barrel layers and two endcap disks per barrel. The barrel layers (BPIX) are located at radial distances of 4.4 cm, 7.3 cm and 10.2 cm from the beamline, and have a length of 53 cm. The endcap disks (FPIX) extend from 6 to 15 cm in radius and are located at points $z = \pm 34.5$ cm and $z = \pm 46.5$ cm.

During the extended year-end technical stop of the LHC in 2016/2017, the original CMS pixel detector was replaced with an upgraded pixel system (CMS Phase-1 pixel detector). This upgrade was necessary to handle the higher instantaneous luminosities achieved by the LHC following upgrades to the accelerator during the first long shutdown in 2013-2014. Compared to the original pixel detector, the upgraded detector has a better tracking performance and lower mass with four barrel layers and three endcap disks on each side to provide hit coverage up to an absolute value of pseudorapidity of 2.5 [24] (Fig. 1.5).

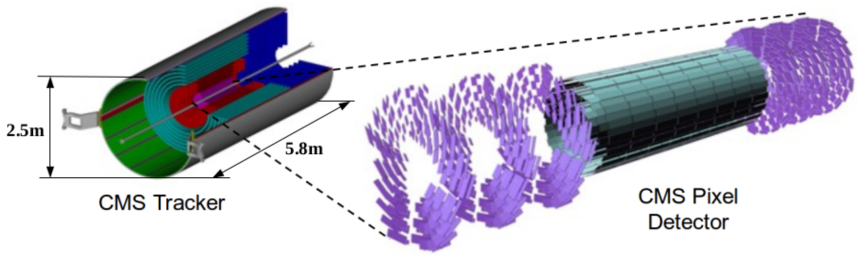


Figure 1.5: The CMS Pixel detector.

1.2.4 Electromagnetic Calorimeter (ECAL)

During 2011 and 2012 data taking, the ECAL played a vital role in the discovery of a new resonance attributed to the SM Higgs boson with a mass of 125 GeV [25]. It measures the energy of electrons and photons, and covers pseudorapidity region of $|\eta| < 3$ (Fig. 1.6). The design of the CMS ECAL was specifically tailored for the search for Higgs boson decays to two photons. The goal was to distinguish the narrow peak of the Higgs boson signal from a large and continuous background. Its good energy resolution and its fine granularity led to a very good energy resolution of the di-photon system by improving the energy and the angle measurement of the two photons.

1.2.5 Hadronic Calorimeter (HCAL)

The HCAL are crucial in physics analyses in which hadron jets or missing transverse energy are present. It plays an important role in the search for new physics [26]. The hadron calorimeter is located around the ECAL and extends between $1.77 < r < 2.95$ m within the solenoid operating at 3.8 T. This constrains the total amount of material which can be put in to absorb the hadronic shower. Materials used in HCAL have short interaction lengths. An outer hadron calorimeter is placed outside the solenoid complementing the barrel calorimeter. Additionally, it should be able to accurately measure the transverse missing energy and extend to large pseudorapidity

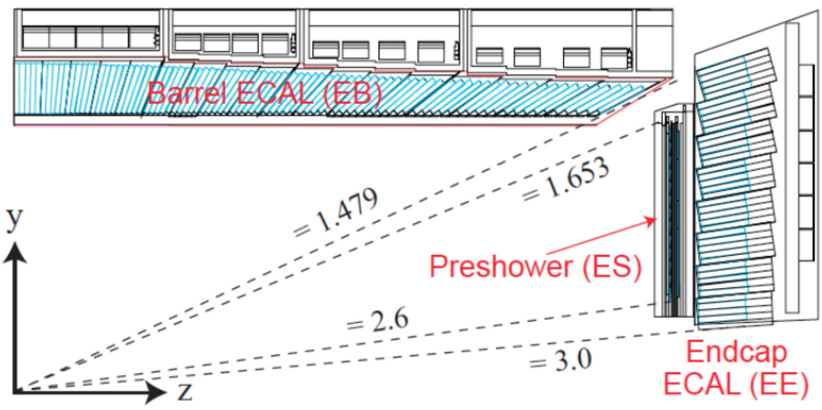


Figure 1.6: Sectional view of ECAL.

values as possible. Therefore, the forward hadron calorimeter is placed at 11.2 m from the interaction point and extends the pseudorapidity coverage to $|\eta| = 5.2$ (Fig. 1.7).

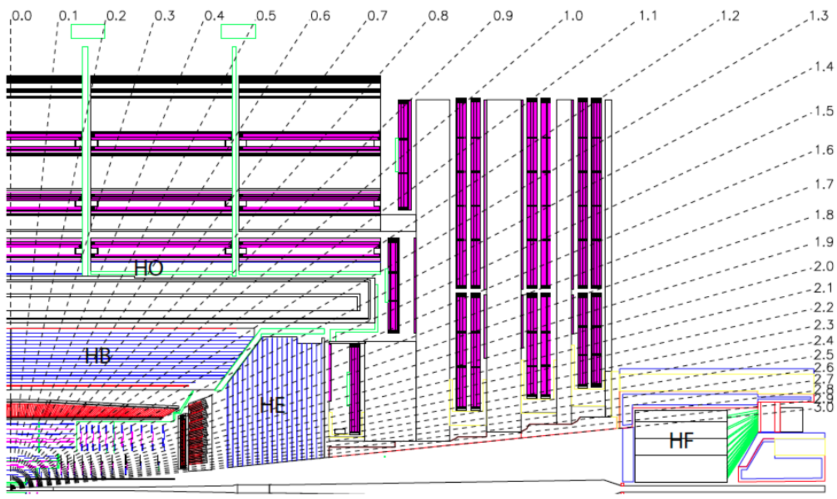


Figure 1.7: Longitudinal view of a quarter of the CMS hadronic calorimeter HCAL.

1.2.6 Muon Systems

In the field of high energy physics, the detection of muons is of the most importance since these particles provide a feasible way to study the second generation of particles in the SM, and they may also be indicative of the decay of possible new particles. Because muons are about 200 times heavier than electrons, do not emit radiation and are observed as minimum ionizing particles (MIPs), which makes it very difficult to detect these particles when they are passing through materials. A muon is a low-interacting particle that is not stopped by any of the calorimeters, and it is capable of penetrating several meters of iron without interfering with it. The muon system is therefore situated at the edge of the experiment where other particles will rarely be able to reach, therefore providing a natural way to discriminate against non-muon backgrounds. A muon system (Fig. 1.8), as part of CMS, performs efficient reconstruction and identification of muons and, along with a tracker, it provides accurate movement measurements of muons with a precision of 1 - 10% within a wide range of kinematics [27].

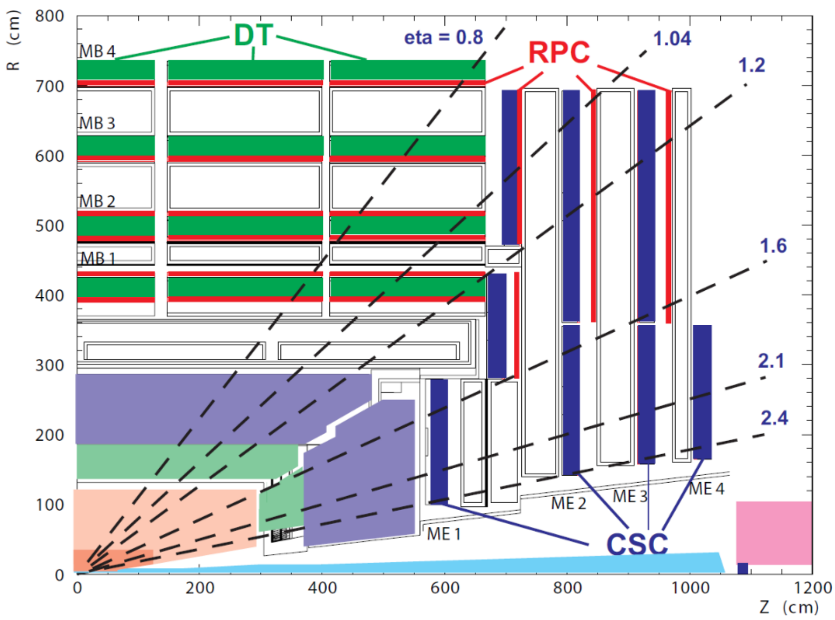


Figure 1.8: Longitudinal view of a quarter of the CMS detector illustrating the layout of the Muon System.

1.2.7 Trigger System

There is a collision rate of about 100 MHz at the LHC, which produces an enormous amount of data, most of which is irrelevant for the CMS physics program and is practically impossible to store. By using the trigger system, CMS is able to select only the small fraction of collision events that are relevant for the project. The CMS trigger system consists of two stages: the Level-1 trigger (L1) [28], uses information from the calorimeters and muon detectors to select events at a rate of around 100 kHz within a fixed latency of about 4μs [29]. The second level, known as the high-level trigger (HLT), consists of a farm of processors running a version of the full event reconstruction software optimized for fast processing, and reduces the event rate to around 1 kHz before data storage [30].

1.3 Chapter summary

The LHC and the CMS detector

The research presented in this thesis was carried out using the experimental data that was delivered by the Large Hadron Collider and collected by the Compact Muon Solenoid detector during the Run-II data taking (2016-2018). CMS detector consists of sub-detectors: tracker, electromagnetic calorimeter, hadron calorimeter and muon system. In this analysis, muon leptons will be used and reconstructed using information from all the subdetectors.

CHAPTER

TWO

THEORETICAL FRAMEWORK

This chapter presents a theoretical framework for the physics search performed in this analysis. Section 2.1 introduces the standard model of particle physics (SM), the most accurate theory for the interaction of fundamental particles. In Section 2.4, we discuss some of the challenges associated with the SM that motivate the need for a deeper physical theory. Section 2.5 present some beyond the standard model theories such as two-Higgs-doublet models extended with a scalar singlet. The chapter ends with a discussion of the phenomenology of the particular $H \to a_1 a_1 \to \mu\mu bb$ signal process chosen for analysis in Section 2.6.

2.1 Standard model

It is obvious that the Standard Model of fundamental particles is a masterpiece in the history of science. Its precise mathematical description is formulated in Quantum Field Theory (QFT) [31], which starts from the idea of describing fundamental particles as quantum fields. In the standard model, fundamental particles are divided into two categories, fermions and bosons, according to their spin number. A brief overview of their characteristics will be provided in the following sections.

2.1.1 Fermions

Fermions are divided into two categories: quarks and leptons. Quarks and leptons are the two basic building blocks of matter. Quarks are the subatomic particles that make up protons and neutrons in the nucleus of an atom, while leptons are particles that do not interact with the strong force, which is responsible for binding the nucleus of an atom.

Gravity affects both quarks and leptons in the same way, as it is a force that acts on all matter. Both quarks and leptons have mass, and thus experience the effects of gravity. In quantum chromodynamics (QCD), a theory similar to quantum electrodynamics for electromagnetic interactions, quarks are attributed an additional type of charge called the "color charge", as they can interact with the strong force. In addition, fermions appear in three generations; their properties remain the same but their mass differs in each generation. In the table 2.1, three distinct generations of quarks and leptons are shown. Additionally, SM predicts the existence of particles with

Table 2.1: The classification of fermions according to their properties.

	Fermions	**Strong interaction**	**Electromagnetic interaction**	**Weak interaction**	**Electric charge**
Quarks	$u \quad c \quad t$	✓	✓	✓	$+2/3$
	$d \quad s \quad b$	✓	✓	✓	$-1/3$
Leptons	$e \quad \mu \quad \tau$	-	✓	✓	-1
	$v_e \quad v_\mu \quad v_\tau$	-	-	✓	0

the same properties as fundamental fermions, but with opposite charges. Paul Dirac discovered this fact when he wrote the Schrödinger wave function for electrons, taking special relativity into account. In total, there are 6 flavors and 3 colors for quarks, resulting in 18 quarks. By adding 6 leptons, a total of 24 fundamental fermions are created. Due to the fact that there is an antiparticle associated with each fermion, a total of 48 fermions are obtainable in SM. To the best of our knowledge, matter in our universe consists only of bound states of first generation quarks, called protons and neutrons, surrounded by electrons that make up atoms.

2.1.2 Bosons

The quantum field theory describes fundamental forces except gravity by introducing gauge bosons which act as force carriers during interactions. Photons carry the electromagnetic force, gluons are strong force carrier particles, while Z and W bosons are weakly interacting particles. The SM interactions are described by three-point vertices, where the gauge bosons are coupled to fermions. A particle can interact through a force only if it carries charges. Similar to electric charge in electromagnetism, strong and weak forces are often characterized by charges called "color charge" and "isospin", respectively. All fermions carry isospin and thus all participate in weak interactions, but quarks carry color charge and can have strong interactions. Finally, these main sources can be compared in terms of power. In Table 2.2, the relative strength of each interaction (including gravity) is shown for a two-body system separated by approximately the proton radius. This is not a comparison with energy [31].

Table 2.2: A comparison of the interaction strengths of the four basic forces.

Force	Strength	Boson	Mass
Strong	1	Gluon	0
Electromagnetic	10^{-3}	Photon	0
Weak	10^{-8}	$W^{\pm}, Z$	$80.4, 91.2\,GeV$
Gravity	10^{-37}	Graviton!	0

The Braut-Englert-Higgs boson, which completed the SM puzzle, is described in QFT as a field that extends through space-time, with a non-zero expected value everywhere, even in a vacuum. Other particles gain mass due to interaction with this field. This boson was detected in 2012 by two detectors at CERN (CMS and ATLAS) with a mass of $124 \pm 0.28\,GeV$ [32]. Unlike other bosons, the Higgs boson is a scalar particle with no spin. The Higgs boson has a very special place in the compatibility of the standard model theory and without the Higgs mechanism this theory will be incomplete and inconsistent. The Higgs mechanism plays the main role in the unity of the two electromagnetic and weak forces, as electroweak. A brief overview of the Higgs mechanism can be found in the following sections.

2.2 Standard model Lagrangian

SM is a gauge theory and is invariant under the $SU(3)_C \times SU(2)_L \times U(1)_Y$ symmetry group. $SU(2)_L \times U(1)_Y$ describes the electroweak interaction, while $SU(3)_C$ describes the strong interaction between fermions. There is a C subscript for color charge, L for weak interactions between the left-handed particles, and Y for weak hypercharge. The most general expression of SM Lagrangian is as follows:

$$\mathcal{L}_{SM} = \mathcal{L}_{gauge} + \mathcal{L}_f + \mathcal{L}_{Yuk} + \mathcal{L}_\phi, \tag{2.1}$$

where each of these represents a gauge, fermion, Yukawa, and scalar term. Generally, the SM Lagrangian is given by:

$$\mathcal{L}_{SM} = i{\psi_i}^\dagger \bar{\sigma}^\mu D_\mu \psi_i - \frac{1}{4} F^{a\mu v} F^a_{\mu v} - Y^{ij} {\psi_i}^\dagger \psi_j H + |D_\mu H|^2 + \mu^2 H^\dagger H - \lambda (H^\dagger H)^2 + h.c \tag{2.2}$$

where $\bar{\sigma}^\mu = (1, -\sigma^i)$, and σ^i is Pauli matrix. $F^a_{\mu v} = \partial_\mu F^a_v - \partial_v F^a_\mu + g f_{abc} F^b_\mu F^c_v$ is Field-strength tensor, where $F = W, B, G$ is a gauge boson field with the corresponding coupling g, and f_{abc} is fine-structure constant. The scalar Lagrangian is as follows:

$$\mathcal{L}_\phi = (D^\mu \phi)^\dagger D_\mu \phi - V(\phi) \tag{2.3}$$

where $V(\phi)$ potential is given as,

$$V(\phi) = \mu^2 \phi^\dagger \phi + \lambda (\phi^\dagger \phi)^2. \tag{2.4}$$

There are two parameters in the equation above: a parameter referred to the Higgs boson mass and a parameter referred to the trilinear self-coupling of the Higgs boson. This potential is the most generalized renormalizable and SU(2) invariant potential. For $\mu^2 > 0$, the minimum energy state is at zero, while for $\mu^2 < 0$, it is not. Hence, the scalar field develops an expectation value (vev) that is not zero. With this parameter choice, the potential has the shape of a Mexican hat (Fig. 2.1). The kinetic energy term in Eq. 2.3 includes the gauge-covariant derivative, which is defined as follows:

$$D_\mu \phi = (\partial_\mu + ig\vec{T}.\vec{W}_\mu + \frac{ig'}{2} B_\mu)\phi. \tag{2.5}$$

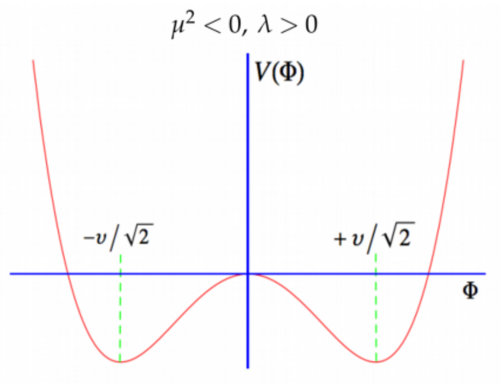


Figure 2.1: A view of the Higgs boson potential if $\mu^2 < 0$ and $\lambda > 0$.

Electroweak symmetry breaking led to the term $\mathcal{L}_\phi$ in Lagrangian, which is actually the mass term that gives mass to fermions and quantum fields [33, 34]

2.2.1 Electroweak symmetry breaking

In the SM, a scalar doublet is introduced as follows:

$$\phi = \frac{1}{\sqrt{2}} \begin{pmatrix} \varphi_1 + i\varphi_2 \\ \varphi_3 + i\varphi_4 \end{pmatrix} \tag{2.6}$$

The minimum of a potential corresponds to a non-zero field, while the maximum is at zero. Based on this minimum, the fields can then be expanded as:

$$\phi = \langle \phi \rangle + \hat{\phi} = \begin{pmatrix} 0 \\ \frac{v}{\sqrt{2}} \end{pmatrix} + \hat{\phi} \tag{2.7}$$

where v is the vacuum expectation value (vev) which is measured to be about $264\,GeV$ and is equal to $\sqrt{\frac{-\mu^2}{\lambda}}$. The covariant derivative of the ϕ field (Eq. 2.5) leads to:

$$|D_\mu < \phi > |^2 = \frac{g^2 v^2}{8} \left((W_\mu^1)^2 + (W_\mu^2)^2 + (-W_\mu^3 + \frac{g'}{g})^2 \right). \tag{2.8}$$

Following are the field transformations for determining the eigenstates of mass:

$$W_\mu^\pm = \frac{1}{\sqrt{2}}(W_\mu^1 \mp W_\mu^2) \tag{2.9}$$

$$Z_\mu = W_\mu^3 \cos\theta_W - B_\mu \sin\theta_W \tag{2.10}$$

$$A_\mu = W_\mu^3 \sin\theta_W + B_\mu \cos\theta_W \tag{2.11}$$

where $\tan\theta_W = \frac{g'}{g}$ and θ_W is the Weinberg angle, substituting Eq. 2.9 into Eq. 2.8 the gauge boson mass term is calculated as follows:

$$m_A = 0, \tag{2.12}$$

$$m_W = \frac{gv}{2}, \tag{2.13}$$

$$m_Z = \frac{v}{2}\sqrt{g^2 + g\prime^2} = \frac{m_W}{\cos\theta_W} \tag{2.14}$$

there is no mass term for the photon, and the masses W and Z are related. Finally, the Higgs boson mass can be calculated as follows:

$$m_h = \sqrt{2}\mu = \sqrt{2\lambda v}. \tag{2.15}$$

The fermion mass terms result from Yukawa interaction in Eq. 2.2. They will be presented as follows:

$$\mathcal{L} = -Y_l^{ij} e_{Ri}^\dagger L_j H - Y_u^{ij} u_{Ri}^\dagger Q_j i\sigma_2 H^* - Y_d^{ij} d_{Ri}^\dagger Q_j H + h.c. \tag{2.16}$$

after electroweak symmetry breaking, mass term expressed in the following way:

$$\mathcal{L} = -\frac{v}{\sqrt{2}} Y_l^{ij} e_{Ri}^\dagger e_{Lj} - \frac{v}{\sqrt{2}} Y_d^{ij} d_{Ri}^\dagger d_{Lj} - \frac{v}{\sqrt{2}} Y_u^{ij} u_{Ri}^\dagger u_{Lj} + h.c. \tag{2.17}$$

Fermion terms are proportional to the vacuum expectation value of the Higgs boson. Fermions have masses $m_f = \frac{v}{\sqrt{2}} Y_f$, where $f = l, u, d$ indicates lepton, up and down quarks, respectively.

All quarks and leptons are coupled to the Higgs field with a strength proportional to their masses, and gauge bosons are coupled with a strength proportional to the square of their masses.

2.3 Higgs production modes and decay channels

2.3.1 SM Higgs production modes

In hadron colliders, the Higgs boson is produced primarily through four mechanisms: the gluon-gluon fusion (ggF), the vector boson fusion (VBF), the associated production with a vector boson (VH), and associated production with heavy quarks like top and bottom quarks ($q\bar{q}H$).

It is depicted in Fig. 2.2 how Feynman diagrams represent the three main production processes of the Higgs bosons. The gluon-gluon fusion (ggF) is the main production mechanism for Higgs particles, and proceeds via a quark loop. It has a production cross section of 49.47 pb for $m_H = 125\,GeV$ at center-of-mass energy 14 TeV. Vector boson fusion (VBF) is the second most dominant production mode and has a cross section of one order of magnitude lower than ggF. It could be an interesting tool to probe BSM physics signature, because of the two high-momentum quarks in the final state. Cross-sections of the other two processes are very small. Fig. 2.3 shows the production cross sections as a function of Higgs mass at center-of-mass energy of 14 TeV.

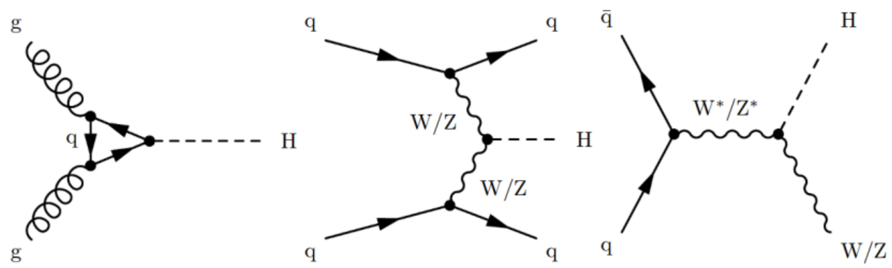


Figure 2.2: Feynman diagrams of Higgs production processes: gluon-gluon fusion (left), vector boson fusion (VBF) (middle), associated production with a vector bosons (VH) (right).

2.3.2 SM Higgs decay channels

The Higgs boson with mass of about 125 GeV can only be observed through its decay products. The branching ratio ($\mathcal{B}$) of the Higgs boson is given by the following equation:

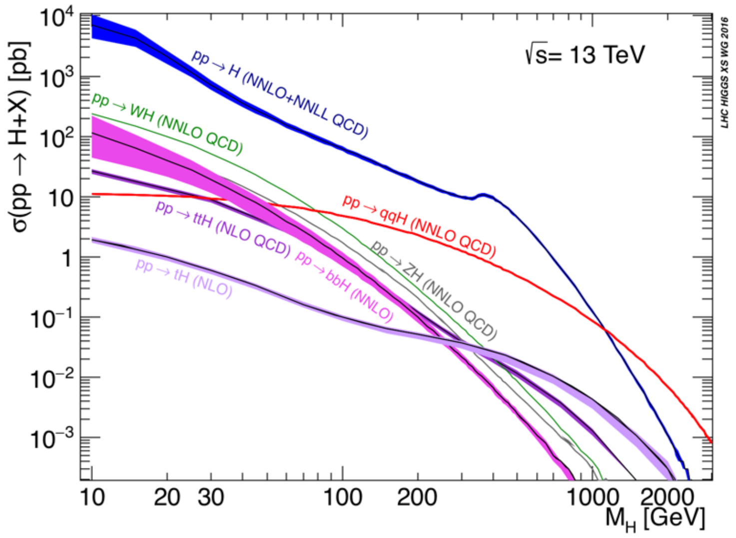


Figure 2.3: Production cross-section for a Higgs boson at the LHC as a function of the Higgs mass.

$$\mathcal{B}(H \to XX) = \frac{\Gamma(H \to XX)}{\sum\limits_{Y \in SM} \Gamma(H \to YY)}. \tag{2.18}$$

The numerator represents the partial decay width particle and the denominator is the sum of the partial decay widths of all possible decay channels. The largest branching ratios are to the more massive particles. For a Higgs boson mass of 125 GeV, the largest branching ratio is to bottom quarks, $\mathcal{B}(H \to bb) = 57.8\%$. Although this channel provides a clear signature in experiments, it suffers from large backgrounds at LHC due to the production of direct bottom quarks. Though, various methods are available for b-tagging that help reduce a variety of backgrounds. Fig. 2.4 shows the branching ratio of the SM Higgs boson as a function of its mass and the different decay channels that account for over 99% of the total width.

2.4 Limitations of the SM

Though there is no doubt regarding the precision of the SM, it is certainly not complete, since there are still questions that it does not resolve. In addition to gravity, the SM also has numerous limitations, such as the identification of dark matter [35], the existence

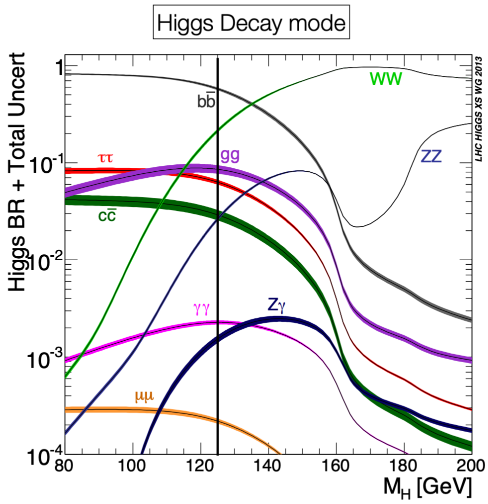


Figure 2.4: Branching ratios of Higgs boson as a function of Higgs mass.

of neutrino masses, the baryon asymmetry of the Universe (BAU) and electroweak Baryogenesis [36], the strong CP problem [37], the "hierarchy problem" between the fundamental forces, and the large number of arbitrary parameters. Some of these issues are shortly described in the next sections.

2.4.1 Dark matter

The term dark matter was first introduced in 1933 by Fritz Zwicky, a Swiss astrophysicist, when he observed the Coma cluster and noticed that its dynamical mass was significantly larger than its luminous mass. This discrepancy indicated the presence of invisible matter within the cluster. However, due to measurement uncertainties, his findings were not widely accepted at that time. It wasn't until the 1970s that the issue resurfaced when Vera Rubin, an American astronomer, studied the rotation curves of galaxies. She discovered that the velocity of stars within galaxies did not decrease as predicted by Newton's laws, suggesting the existence of a massive halo of invisible matter known as dark matter. According to the latest predictions, it is estimated that only 4.5% of the mass content of the universe is represented by the matter we can observe. The remaining 26% is dark matter, while 70% is dark energy. The SM does not provide any predictions for what could be the candidate for dark matter. However, the concept of Supersymmetry theory emerged to address this gap by proposing the Lightest Supersymmetric Particle (LSP) as a potential suitable candidate for dark matter.

2.4.2 Neutrino masses

According to the SM, neutrinos were initially predicted to be massless and to have weak interactions with other particles. However, observations made by collaborations such as the Super-Kamiokande have shown that neutrinos can actually oscillate or change from one flavor to another, which suggests that they cannot be massless. Although the exact masses of neutrinos have not been accurately measured, there are upper limits on their masses and the differences between their masses have been determined.

2.4.3 Baryon asymmetry of the Universe

Although many experiments and investigations are currently underway to address the issue, the standard model of particle physics still lacks a comprehensive explanation for the imbalance between matter and antimatter in the universe. The origins of this imbalance, which should have resulted in equal amounts of matter and antimatter being produced during the Big Bang, remain unresolved. The mechanisms that prevented their total annihilation and led to the predominance of matter are still not fully understood. Efforts are ongoing to discover and understand the underlying causes of this ratio problem.

2.5 Beyond the Standard Model (BSM)

Considering the recent results of ATLAS and CMS experiments and the limitations of the SM, it would seem that there are indications of new physics beyond the SM. A variety of theoretical models have been explored, including supersymmetry (SUSY), the minimal supersymmetric standard model (MSSM), and the two-Higgs-doublet model (2HDM), which extend the SM by incorporating additional particles or fields. This chapter will briefly discuss the SUSY and MSSM models, followed by an examination of the 2HDM. Finally, an extension of the 2HDM model, known as 2HDM+S, will be explored.

2.5.1 Supersymmetry (SUSY)

SUSY is one of the theoretical models that have been proposed to extend the SM and provide answers to many unanswered questions. According to this theoretical model, every fundamental particle has a supersymmetric partner (called a superpartner), which is similar to that particle in every respect except for its spin. Leptons and quarks have spin-0 superpartners, while spin-1 bosons, like photons, have spin-$\frac{1}{2}$ superpartners. Supersymmetric partners of fermions are named by adding the prefix "s" to the fermion name.

To be more precise, each fermion has a bosonic superpartner, and each boson has a fermionic superpartner. Therefore, SUSY introduces the operator "Q" which shows the relationship between fermions and bosons:

$$Q|fermion\rangle = |boson\rangle \tag{2.19}$$

$$Q|boson\rangle = |fermion\rangle \tag{2.20}$$

SUSY has a side effect, which is the violation of lepton and baryon numbers. Therefore, a new quantum number R has been introduced:

$$R = (-1)^{2S+3(B-L)} \tag{2.21}$$

where L, B, and S are the lepton, baryon, and spin numbers, respectively. For all SM particles $R = 1$ and for their superpartners $R = -1$. The presence of R-parity conservation implies that super particles must be produced in pairs in collider experiments and the lightest super particle (LSP) must be absolutely stable. so the lightest symmetric superparticle (LSP) must be stable. This LSP proposed by SUSY is a neutral particle that does not interact via strong or electromagnetic forces, thus making it an excellent candidate for dark matter. In general, LSP is neutral, hence it can pass undetected through the detector material. This is known as "missing energy" and is commonly used to detect the presence of SUSY. In order to confirm supersymmetry, the detection of predicted particles will undoubtedly be necessary. Unfortunately, it is difficult to talk about their mass, because the nature of supersymmetry breaking is unknown and even the simplest and most general model of supersymmetry, called the minimal supersymmetry standard model (MSSM), contains many unknown parameters. However, according to most versions of the theory, including MSSM, superparticles can only be created and destroyed in pairs.

Finally, superparticles are not the only new particles predicted by supersymmetry and grand unification theories, because the Higgs mechanism in such theories is very complex and requires more than the standard model's single Higgs boson.

2.5.2 Minimal supersymmetric extension of the SM (MSSM)

It is well known that SM has 19 free parameters. The minimal supersymmetric standard model (MSSM) contains 124 free parameters. The MSSM is the supersymmetric extension of the SM that contains a minimal number of states and interactions [38]. The model has the gauge symmetry $SU(3)_C \times SU(2)_L \times U(1)_Y$ extended by the supersymmetry to include the supersymmetric partners of the SM fields which have different spins as required by the supersymmetric algebra.

Similar to SUSY, each SM field requires a superpartner field. A fermion's partner is a slepton or a scalar squark, a Higgs fermion's partner is a higgsinos, and a quantized boson's partner is a gaugeino. Higgs boson behave differently because they have no spin, resulting in the largest change in the Higgs sector. In MSSM, instead of a double scalar field, two chiral superfields are substituted, called "higgsino". The presence of the second Higgs double increases the number of physical states to 5, since they now include 8 degrees of freedom instead of 4 degrees of freedom in SM. The five physical states include two charged particles, two neutral CP-even particles and one neutral CP-odd particle. More details will be presented in the next section, which introduces the 2HDM and MSSM as a special case of it.

In summary, MSSM is the simplest supersymmetric extension of the SM, which requires at least two Higgs doublets to give mass to particles by the Higgs mechanism. For this reason, the Higgs mechanism part of this model is a 2HDM model, which includes two Higgs doublets.

2.5.3 Two-Higgs-doublet model (2HDM)

Extensions of the SM scalar sector have been extensively studied in the past [39]. As mentioned before, from a theoretical perspective, there are many reasons to extend the scalar sector of the SM electroweak sector. Similar to MSSM models, 2HDM models introduce two scalar fields that lead to five physical states: two charged scalars ($H^{\pm}$), one pseudoscalar (CP-odd) A, and two neutral scalars (CP-even) h and H. 2HDM model allow a large range of production and decay mode of an individual scalar or a cascade that involves more than one scalars. Hence, the 2HDM allows exotic decays of the SM-like Higgs, which may be one of the neutral scalars h or H. Unfortunately, the 2HDM has been constrained by recent experimental findings which limit the possibility of exotic Higgs decay phenomenology.

There are two ways to avoid these restrictions. One way is to assume that the 2HDM is near or in the decoupling limits where $\alpha \to \beta\pi/2$; as a result, the lightest state of the 2HDM is the observed 125 GeV Higgs state. The other way is to add a scalar complex singlet to the existing model, such as:

$$S = \frac{1}{\sqrt{2}}(S_R + iS_I). \tag{2.22}$$

consequently, a new extended model has been proposed, the 2HDM + S, which predicts the exotic decay channel ($H \to a_1 a_1 \to bb\mu\mu$). The following section discusses this model.

2.5.4 Two-Higgs-doublet model plus a scalar singlet

2HDM+S is a simple extension of 2HDM obtained by adding a mixed single-scalar (Eq. 2.22) to 2HDM. The 2HDM, and hence 2HDM+S, are categorized into four types depending on the interaction of SM fermions with the Higgs doublet structure [16]. All SM particles couple to the first Higgs doublet, Φ_1, in type I models. In type II, up-type quarks couple to Φ_1 while leptons and down-type quarks couple to the second Higgs doublet, Φ_2. Type II models have NMSSM as a special case where the additional singlet is introduced to resolve the so-called μ-problem of the MSSM superpotential. Quarks couple to Φ_1 and leptons couple to Φ_2 in type III, and finally in type IV, leptons and up-type quarks couple to Φ_1, while down-type quarks couple to Φ_2. One of the key motivations for introducing the 2HDM+S model is the fact that the minimal suppersymmetric extension of the SM with additional degrees of freedom (NMSSM) is a special case of this model. NMSSM is a simple extension of MSSM that seeks to solve the "hierarchy problem". In NMSSM, by introducing a mixed singlet-scalar this problem disappears and this is a strong motivation for NMSSM over MSSM. Another motivation comes from the fact that the new scalar particles contribute to the scalar boson mass in NMSSM, which removes the problems in MSSM that originate from the large measured mass of the new particle.

2.6 Exotic decay of the SM Higgs boson

One can assume that the 2HDM+S is in a decoupling limit such that the lighter scalar eigenstate, H, is the discovered boson with $m_H = 125\,GeV$. Furthermore, to allow for a wider variety of exotic Higgs decay phenomenology, a complex scalar singlet, S_R+iS_I is added to the model which has no direct Yukawa couplings. Hence, it is expected to decay to SM fermions by virtue of mixing with the Higgs sector. This mixing is small enough to preserve the SM-like nature of the H boson. The Higgs boson can decay to fermions through $a_1(s_1)$ where $a_1(s_1)$ is a (pseudo)scalar mass eigenstate mostly composed of $S_R\,(S_I)$.

Exotic decay of the $125\,GeV$ Higgs boson into a light pseudoscalar a_1, with a pair of jets b and a pair of leptons μ in the final state seems quite promising. 2HDM+S type III is used to analyze this channel. Figure 2.5 shows the branching fraction $\mathcal{B}(a_1a_1 \to \mu\mu bb)$ for a number of 2HDM+S scenarios. In the type II and for $\tan\beta > 1$, (where NMSSM is a special case), the branching fraction is constant and above 0.01%, while in the type III with $\tan\beta \sim 2$, the branching fraction can be up to 0.1%.

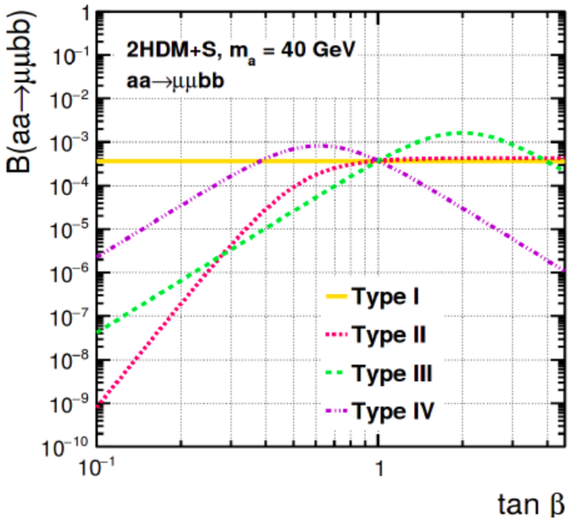


Figure 2.5: Prediction of $\mathcal{B}(a_1 a_1 \to bb\mu\mu)$ for $m_{a_1} = 40\,GeV$ in different types of 2HDM+S, in terms of $\tan\beta$. In general, this picture is the same for all hypothetical values considered for mass in this analysis. Branching fractions were calculated according to the formula in reference [40].

2.7 Chapter summary

Theoretical framework

Despite the fact that the SM accurately describes phenomena within its domain, it remains incomplete. It is possible that the SM is not the final theory of elementary particles but rather the first part of a more comprehensive theory. To address questions within the SM theory, theoretical physicists propose theories beyond the standard model, many of which predict the existence of more than one scalar particle. Among these models is MSSM, which provides a solution to the "hierarchy problem", "unity theory", proposes a dark matter candidate, and introduces a total of five scalar bosons. MSSM is part of a more general class of 2HDM model that include five scalar particles. The 2HDM+S model is an extension of the 2HDM and results in the addition of two new bosons due to the presence of an additional scalar. The purpose of this thesis is to search for this scalar particle.

CHAPTER

THREE

ANALYSIS STRATEGY

The analysis of the $H \to a_1 a_1 \to \mu\mu bb$ channel begins with the selection of data samples and the generation of simulated samples, as described in Section 3.1. Section 3.2 outlines the physics objects reconstruction and preselection. In Section 3.3, events are selected based on different criteria depending on the final state, and the optimization process is implemented. In Section 3.4, in order to separate the signal from the background, we introduce a novel strategy using the variable χ^2_{tot}. In Section 3.5 the events are divided into five different categories depending on the signal over background ratio to increase the sensitivity of the analysis. Finally, the upper limits are obtained on the branching fraction of $H \to a_1 a_1 \to \mu\mu bb$. Generally, this analysis involves the following steps, which will be discussed in greater detail in the following sections and chapters.

- Data samples and simulation
- Physics object reconstruction and preselection
- Event selection and Optimization procedure
- Final selection
- Categorization
- Signal and background estimation
- Systematic uncertainties

- Extracting the upper limits on branching fraction of $H \to a_1 a_1 \to \mu\mu bb$

In order to extract the upper limits on the branching fraction, it is necessary to estimate the background and signal first, followed by calculating the systematic uncertainties. There will be explanations of these steps in chapters 4 and 5.

3.1 Data samples and simulation

3.1.1 Data samples

This analysis is performed using the data from the LHC proton-proton collisions at 13 TeV center-of-mass energy. The data sample, corresponding to an integrated luminosity of 138 fb^{-1} for a mixture of double-muon and single muon triggers, was collected with the CMS detector during Run-II of the LHC. The data correspond to DoubleMuon and SingleMuon primary datasets are used for the signal enriched region as well as a region to validate Drell-Yan background. In addition, MuonEG primary dataset, triggered by the presences of an electron muon pair is used to construct a region for $t\bar{t}$ validation in the $e\mu$ final state. All data have been (re)reconstructed using our knowledge about the physics object reconstructions, energy corrections and additional proton-proton interactions within the same bunch crossing (pileup) in campaigns 17Jul2018, 31Mar2018 and 17Sep2018 for 2016, 2017 and 2018, respectively. Version seven of the so-called NanoAOD event format, produced in 02Apr2020 campaign, is used for both data and simulated samples.

3.1.2 Simulated signal samples

The NMSSMHET model [12] is used to generate signal samples at leading order (LO) with MadGraph5_aMCatNLO [41]. The ggF signal samples are produced for $m_{a_1} = 15$ GeV to $m_{a_1} = 60$ GeV with 5 GeV steps. The VBF signal samples are generated at $m_{a_1} = 20$ GeV, $m_{a_1} = 40$ GeV and $m_{a_1} = 60$ GeV. Signal samples are produced privately using the exact Monte Carlo (MC) matrix element generator setting as that of Ref. [15]. Latest CMS recommendations and guidlines are used for showering, tune, pile up profile, etc. Samples are validated with respect to the central production of Ref. [15]. Figure 3.1 shows some validation distributions.

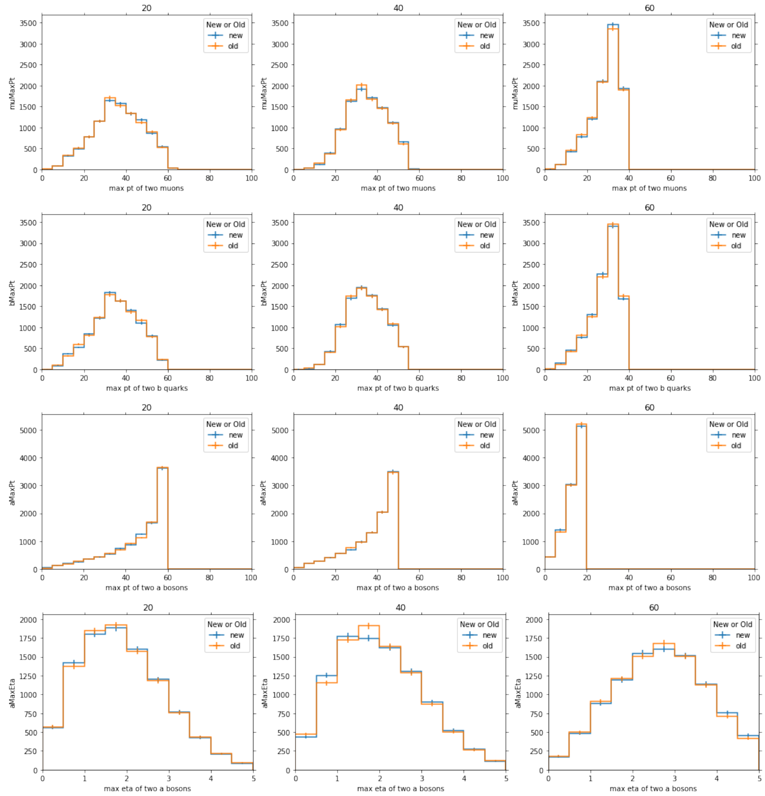


Figure 3.1: Comparison between official (old) and private (new) signal productions for $m_a \in \{20, 40, 60\}\, GeV$: p_T of leading muon (top), p_T of leading b quark (second row), p_T of the leading a_1 boson (third row) and $|\eta|$ of the more forward a_1 boson.

3.1.3 Simulated background samples

The simulated background samples in this analysis are only used for optimization as well as for validation in those steps of selection that yield enough statistics. The final contribution of backgrounds to the selected sample is directly extracted from data with no reference to simulation. The Drell-Yan process, $Z/\gamma^*(\rightarrow ll) + jets$, is modeled with the same event generator at either leading or next-to-leading order (NLO). The matching scheme for the LO samples follows the MLM [42] prescription while the FxFx merging scheme [43] is used at NLO. Based on the m_{ll} threshold at generator level, two Drell-Yan samples are exploited for this analysis, one with $m_{ll} > 50$ GeV and the other with $10 < m_{ll} < 50$ GeV. The high-mass Drell-Yan samples are produced at NLO in all years, exclusive in number of additional partons (up to two).

In low mass, NLO samples are only available in 2016. Furthermore in 2016, an LO sample exclusive in number of additional partons (up to four) exists. The inclusive LO samples are used in all years for low-mass Drell-Yan to harmonize the usage in all years. A k-factor of 1.5 (NLO-to-LO) is extracted from 2016 samples and applied in all years. To account for the limited size of the 2016 samples, an uncertainty of 30% is considered on the k-factor. Figure 3.2 shows NLO-to-LO ratios where both inclusive and jet-binned samples at LO are studied.

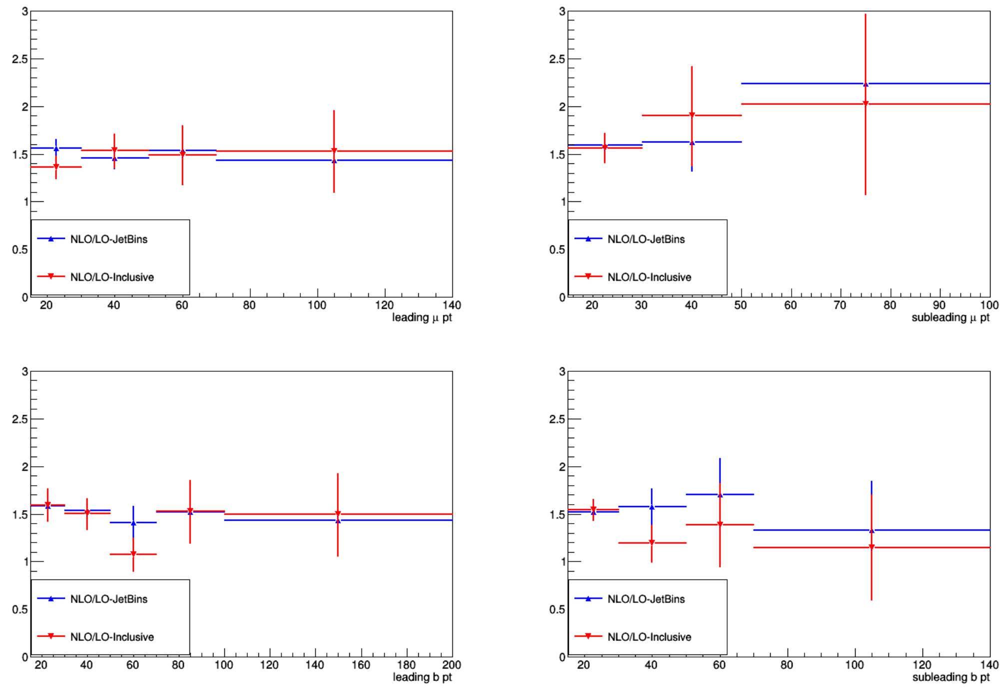


Figure 3.2: The NLO-to-LO k-factors in Drell-Yan low mass for the leading muon p_T (top left), subleading muon p_T (top right), leading b-jet p_T (bottom left) and subleading b-jet p_T (bottom right).

The top quark samples, $t\bar{t}$ and single-top, are produced with POWHEG 2.0 [44] at NLO. Backgrounds from diboson (WW, WZ, ZZ) production are generated with PYTHIA 8.212 [45] at LO. For all samples, PYTHIA with tune CUETP8M1 [46] is used to model the showering. The simulated minimum bias interactions are added to the simulated events to model the effect of pileup. Simulated events are then re-weighted to reproduce the pileup distribution in data. The full CMS detector simulation based on GEANT4 [47] is implemented for all MC generated event samples. Table 3.1 contains the list of signal samples.

3.2 Physics object reconstruction and preselection

Having gathered all the information provided by the sub-detectors of CMS for a particular proton-proton collision, the next step will be to reconstruct and identify all the stable particles in the event, including electrons, muons, photons, charged hadrons, and neutral hadrons. Other physics objects like jets, missing transverse energy, and primary (secondary) vertices are built, identified or reconstructed from the individual elements that compose them.

This section is devoted to details about object reconstruction and preselection. The particle-flow (PF) algorithm [48] is used to reconstruct and identify each individual particle in the event, with an optimized combination of information from the various elements of the CMS detector. The energy of photons is measured in ECAL. The energy of electrons is determined from a combination of the electron momentum at the primary interaction vertex as measured by the tracker, the energy of the corresponding ECAL cluster, and the energy sum of all bremsstrahlung photons spatially compatible with originating from the electron track. The energy of muons is obtained from the curvature of the corresponding track. The energy of charged hadrons is evaluated via a combination of their momentum measured in the tracker and the matching of the ECAL and HCAL energy deposits, corrected for the response function of the calorimeters to hadronic showers. Finally, the energy of neutral hadrons is obtained from the corresponding corrected ECAL and HCAL energies. The primary vertex (PV) is taken to be the vertex corresponding to the hardest scattering in the event, identified using the tracking information alone, as described in Section 9.4.1 of Ref. [49].

In this analysis, the selection requires at least two isolated muons originating from the selected primary vertex with $|\eta| < 2.4$. The leading muon p_T should exceeds a threshold of $17\,GeV$ whereas the sub-leading one is accepted with $p_T > 15\,GeV$. The two muons are required to have an opposite electric charge and to be separated by a minimum $\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} = 0.4$, where ϕ is the azimuthal angle of the particles momentum in the plane perpendicular to the beam line. Muons are identified with the loose criteria defined by the CMS "Muon POG". The isolation variable I_{rel} is

Table 3.1: Privately produced signal samples in all years.

Simulated process	Sample name	Cross section (pb)
RunIISummer16, RunIIFall17, RunIIAutumn18 – NanoAODv7-Nano02Apr2020		
NMSSM $\mu\mu bb$ ggF signal, $m_{a_1} \in \{15, 20, \ldots, 60\}\, GeV$	SUSYGluGluToHToAA_AToMuMu_AToBB_M-m_{a_1}_ TuneCP5_13TeV_madgraph_pythia8_**private**	$\sigma \times BR \approx 0.007$ [12]
NMSSM $\mu\mu bb$ VBF signal, $m_{a_1} \in \{20, 40, 60\}\, GeV$	SUSYVBFToHToAA_AToMuMu_AToBB_M-m_{a_1}_ TuneCP5_13TeV_madgraph_pythia8_**private**	$\sigma \times BR \approx 0.0004$ [12]

calculated by summing the transverse energy deposited by other particles in a cone of size $\Delta R = 0.4$ around the muon, divided by the muon p_T. This quantity is required to be less than 0.15 (0.25) for a tight (loose) selection. In case more muons with those criteria appear in the event, the two with highest p_T are considered. Following the Muon POG recommendations, correction factors for reconstruction and selection efficiencies are applied on simulated events to be comparable with the data.

Jets are reconstructed by clustering the charged and neutral particles using an anti-k_t algorithm [50, 51] with a distance parameter of 0.4. The reconstructed jet energy is corrected for effects from the detector response as a function of the jet p_T and η. Furthermore, contamination from additional interactions (pileup), underlying events, and electronic noise are subtracted [52, 53]. To achieve a better agreement between data and simulation, an extra η-dependent smearing is performed on the jet energy of the simulated events. Events are required to have at least two jets with $|\eta| < 2.4$ and $p_T > 15$, where both jets must be separated from the selected leptons ($\Delta R > 0.4$). The DeepJet and DeepCSV algorithms are tried to identify b-jets where in each case, the two selected jets in the event are required to pass the loose working point (see Section 3.3.1 where the decision is made to use DeepJet). Corrections are applied on simulation to bring the full b-tag discriminator shape in agreement with data. The use of shape-based corrections is motivated by the later categorization of events that relies on the b-tag discriminator value.

The signal events do not have any genuine neutrino and are not expected to produce any imbalance in the event transverse momentum. The missing transverse energy, $\vec{E_T}^{miss}$, is defined as the modulus of $\vec{p_T}^{miss}$, which is the negative vector p_T sum of all reconstructed particles. The jet energy calibration therefore introduces corrections to the $\vec{E_T}^{miss}$ measurement. The optimized threshold for $\vec{E_T}^{miss}$ is found to be $60\,GeV$.

The search for a new scalar, and therefore the final limits, are restricted to $m_{a_1} \in [15, 62.5]$. However for the selection, optimization and eventual background modeling, a slightly wider range is used to provide the full coverage at the boundaries. This means events with $m_{\mu\mu}$ out of $[15, 70]\,GeV$ are discarded. Figure 3.3 shows the distributions of the p_T of the leading and subleading muons and that of leading and subleading jets where b-jets are identified with the DeepCSV algorithm. Similar distributions are provided in Fig. 3.4 using DeepJet b-tagging algorithm is employed.

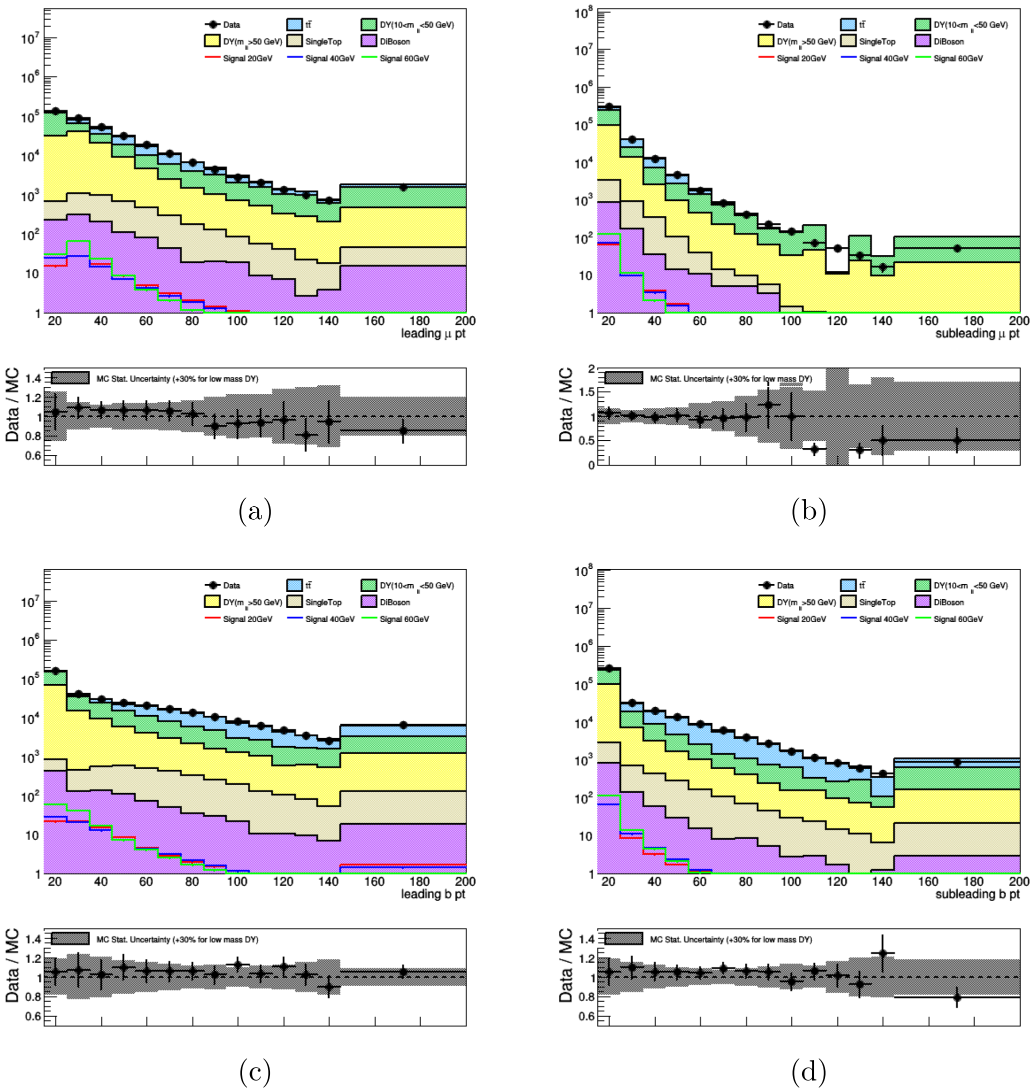
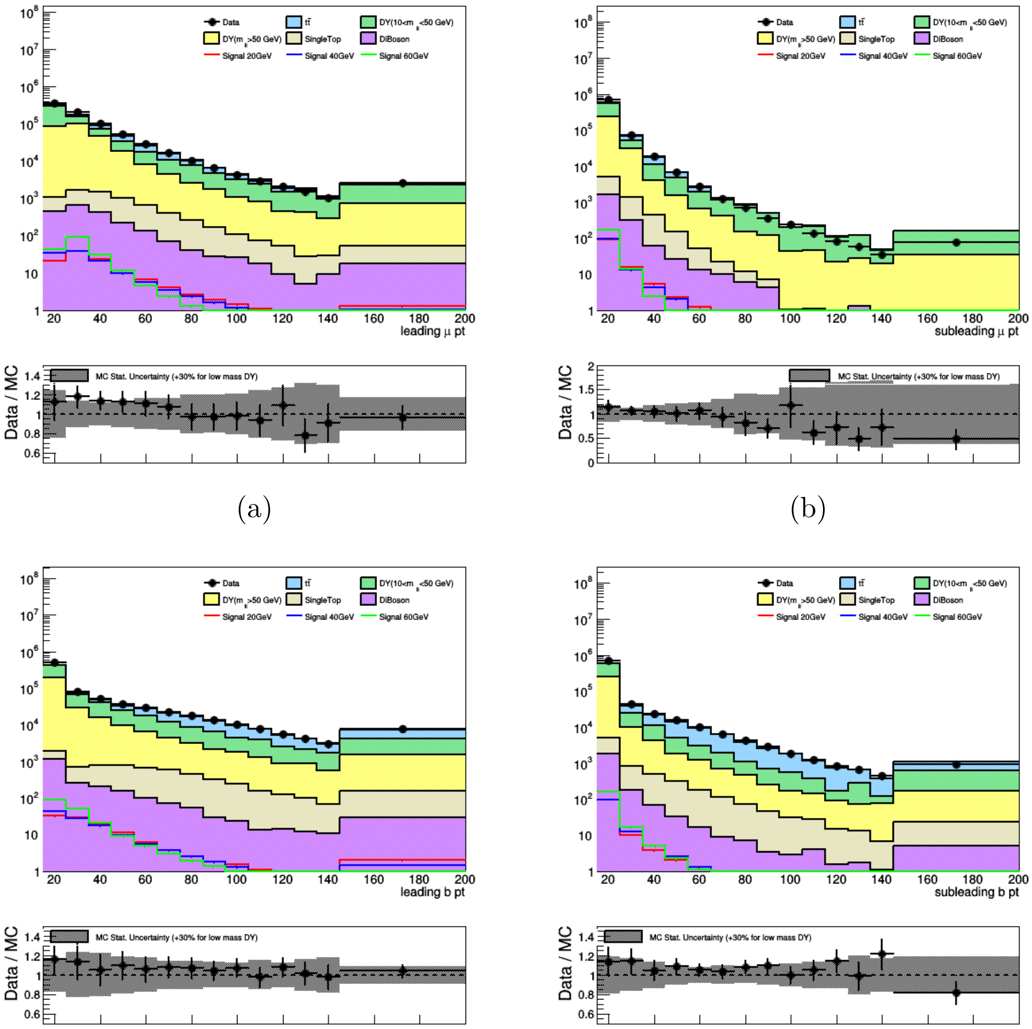


Figure 3.3: The distribution of leading and subleading muon p_T (a,b) and leading and subleading jet p_T (c,d) after asking for two muons and two b-jets with the loose working point of the DeepCSV algorithm. Uncertainty band represents the limited size of MC samples together with 30% uncertainty on low mass Drell-Yan cross-section. Simulated samples are normalized to 138 fb^{-1} with their theoretical cross-sections.



(a) (b)

(c) (d)

Figure 3.4: The distribution of leading and subleading muon p_T (a,b) and leading and subleading jet p_T (c,d) after asking for two muons and two b-jets with the loose working point of the DeepJet algorithm. Uncertainty band represents the limited size of MC samples together with 30% uncertainty on low mass Drell-Yan cross-section. Simulated samples are normalized to 138 fb^{-1} with their theoretical cross-sections.

3.3 Event selection

The $\mu\mu bb$ event candidates are selected based on the requirement that one or two muons are reconstructed at the HLT. Passing the double-muon trigger necessitates two isolated muons with p_T exceeding thresholds of 17 and 8 GeV as listed in Table 3.2. The p_T threshold increases to 24 GeV for the isolated muon in the single-muon trigger path. Accepting events from both single- and double-muon triggers improves the trigger efficiency by including events in which the second muon is not reconstructed at the trigger level.

The final event selection is equivalent to what is summarized in Table 3.5. This selection will result in a sample with few data and simulated events where the statistics for simulation is not enough to draw any conclusion about data-MC agreement. While this is not an issue since the analysis is fully based on data to estimate the background and derive the limits, we show distributions at few steps earlier, with enough statistics, to ensure that we have control on backgrounds and the optimization studies are reliable enough.

Table 3.2: Summary of HLT triggers for different primary datasets used in the signal and Drell-Yan validation regions for different years.

year	primary dataset	HLT path
2016	Double Muon	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL
		HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ
		HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL
		HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ
		HLT_Mu30_TkMu11
	Single Muon	HLT_IsoMu24
		HLT_IsoTkMu24
		HLT_Mu50
2017	Double Muon	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass8
		HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ
		HLT_Mu17_TrkIsoVVL_TkMu8_TrkIsoVVL_DZ_Mass3p8
		HLT_Mu37_TkMu27
	Single Muon	HLT_IsoMu24
		HLT_IsoMu27
		HLT_Mu50
2018	Double Muon	HLT_Mu17_TrkIsoVVL_Mu8_TrkIsoVVL_DZ_Mass3p8
		HLT_Mu37_TkMu27
	Single Muon	HLT_IsoMu24
		HLT_IsoMu27
		HLT_Mu50

events are required to have two muons and at least two b-jets with the kinematics, identification, and isolation criteria detailed in Section 3.2. While the final search in this channel is performed for m_{a_1} between 15 and 62.5 GeV, events are selected with a dimuon invariant mass, $m_{\mu\mu}$, between 14 and 70 GeV. The additional sidebands in $m_{\mu\mu}$ help model the backgrounds at the boundaries. To reduce the background contribution from $t\bar{t}$+jets, events with $\vec{E_T}^{miss} > 60GeV$ are rejected.

3.3.1 Optimization procedure

After applying preselection cuts, in order to reduce background events, the optimization producer is used to select the optimal set of selections and criteria that will maximize the sensitivity of the analysis. In other words, by comparing signal and background yields, we can determine which selections will result in a greater rejection of backgrounds. Various parameters such as b-tagging working points, b-tagging algorithms, and muon object quality identification and isolation are optimized, while aiming for a uniform selection across different mass regions.

Distributions of simulated preselected events agree with the data using loose muon identification and isolation as well as loose and b-tagging requirements. Preselected events in simulation are used to find optimal thresholds of those criteria. The figure of merit is a significance defined as $S/\sqrt{B}$ where S is the expected signal yield for a given m_{a_1} hypothesis and B is the background yield (mainly $t\bar{t}$ and Drell-Yan) in a 5 GeV interval around m_{a_1}. As an example, Fig. 3.5 shows significance as a function of m_{a_1} for different muon isolation scenarios. The identification requirement of muon meets the tight criteria. Optimization is performed at every year, leading to keep the muons loosely isolated.

In Fig. 3.6 (upper row) the difference in significance is shown between DeepJet and DeepCSV algorithms for different years where DeepJet outperforms the other. In lower row, various combinations of working points for the two b-jets are compared. Starting from events with two "loose" b-jets we study if tightening the b-tag criteria on the jets could increase the sensitivity. In each combination, the more stringent criterion is applied on the jet with the higher b-tag score. There is a competition between medium-loose (ML) tight-loose (TL). The TL combination seems to work slightly better.

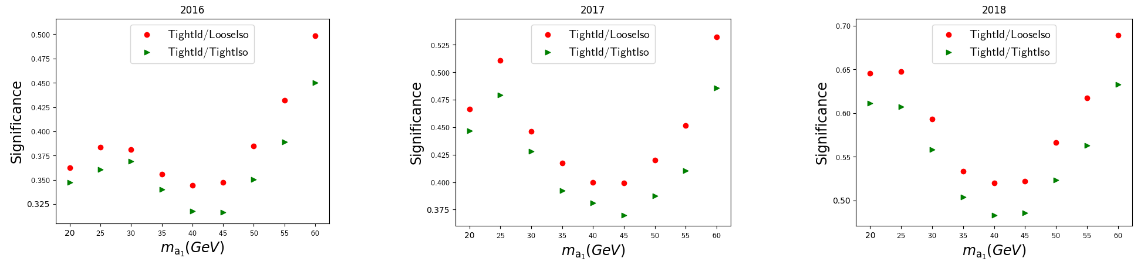


Figure 3.5: Significance as a function of m_{a_1} for different choices of muon isolation in years 2016 (left), 2017 (middle) and 2018 (right).

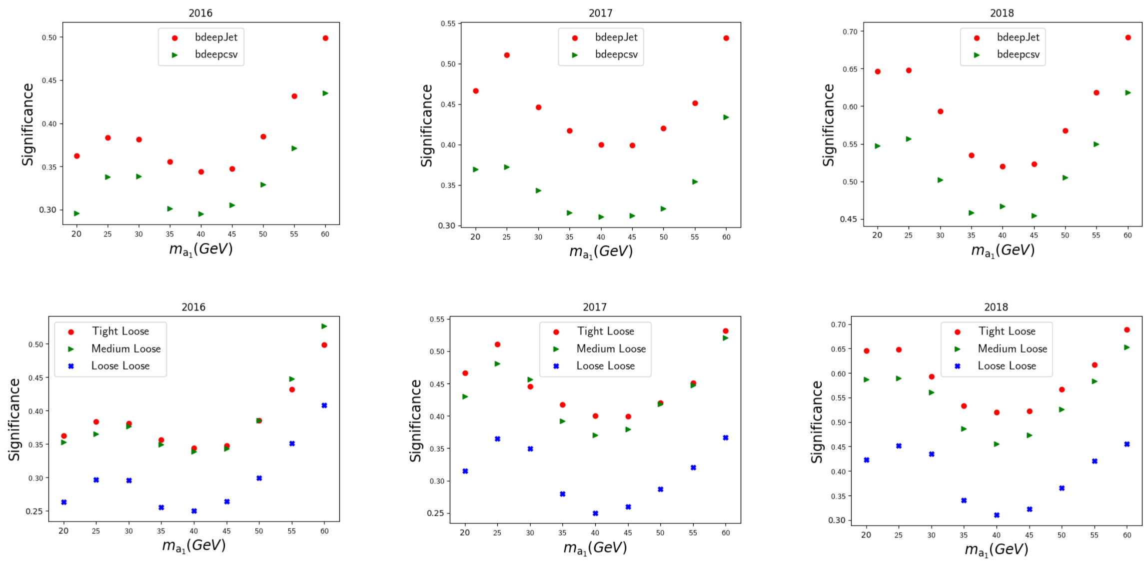


Figure 3.6: Significance as a function of m_{a_1} for different choices of b-algorithms (top) and DeepJet working point combinations (bottom) in years 2016 (left), 2017 (middle) and 2018 (right).

Table 3.3 shows the number of data events together with the expected yield for different backgrounds. The yields in data and simulation are in agreement within 5%. For different m_{a_1} values, expected number of signal events are also provided. It can be seen that for a generic signal cross-section, the signal-to-background ratio is low in the preselected sample and there is room for improvement. In Fig. 3.7, the p_T and the mass of the dimuon system ($m_{\mu\mu}$), the p_T and the mass of the di-b-jet system

(m_{bb}), as well as the p_T and the mass of the Higgs candidate ($m_{\mu\mu bb}$) are presented. The agreement between data and simulation is overall satisfactory. In summary, the optimization at object level leads to tight identification and loose isolation for muons together with tight-loose b-jet identification for the two jets using DeepJet b-tagging algorithm. Optimization continues at event level in the next section where in the end, a summary is presented in Table 3.5.

Table 3.3: Event yields for data and simulated processes using optimized object selection. The expected number of simulated events is normalized to the integrated luminosity of 138 fb^{-1}.
($*$): $\sigma_{\text{sig.}} \times Br \approx 9\, fb$

Process	Yield
$t\bar{t}$	56701 ± 56
Single top	2892 ± 25
Drell-Yan ($10 < m_{\ell\ell} < 50$)	25491 ± 7327
Drell-Yan ($50 < m_{\ell\ell}$)	18523 ± 283
Diboson	267 ± 11
Total expected background	103874 ± 7332
Data	109821

Signal for ggF($*$)		
$m_{a_1} = 20\, GeV$	$m_{a_1} = 25\, GeV$	$m_{a_1} = 30\, GeV$
75.2	70.8	78.0
$m_{a_1} = 35\, GeV$	$m_{a_1} = 40\, GeV$	$m_{a_1} = 45\, GeV$
80.6	69.2	71.6
$m_{a_1} = 50\, GeV$	$m_{a_1} = 55\, GeV$	$m_{a_1} = 60\, GeV$
77.5	82.17	90.9

Signal for VBF		
$m_{a_1} = 20\, GeV$	$m_{a_1} = 40\, GeV$	$m_{a_1} = 60\, GeV$
7.2	7.2	7.5

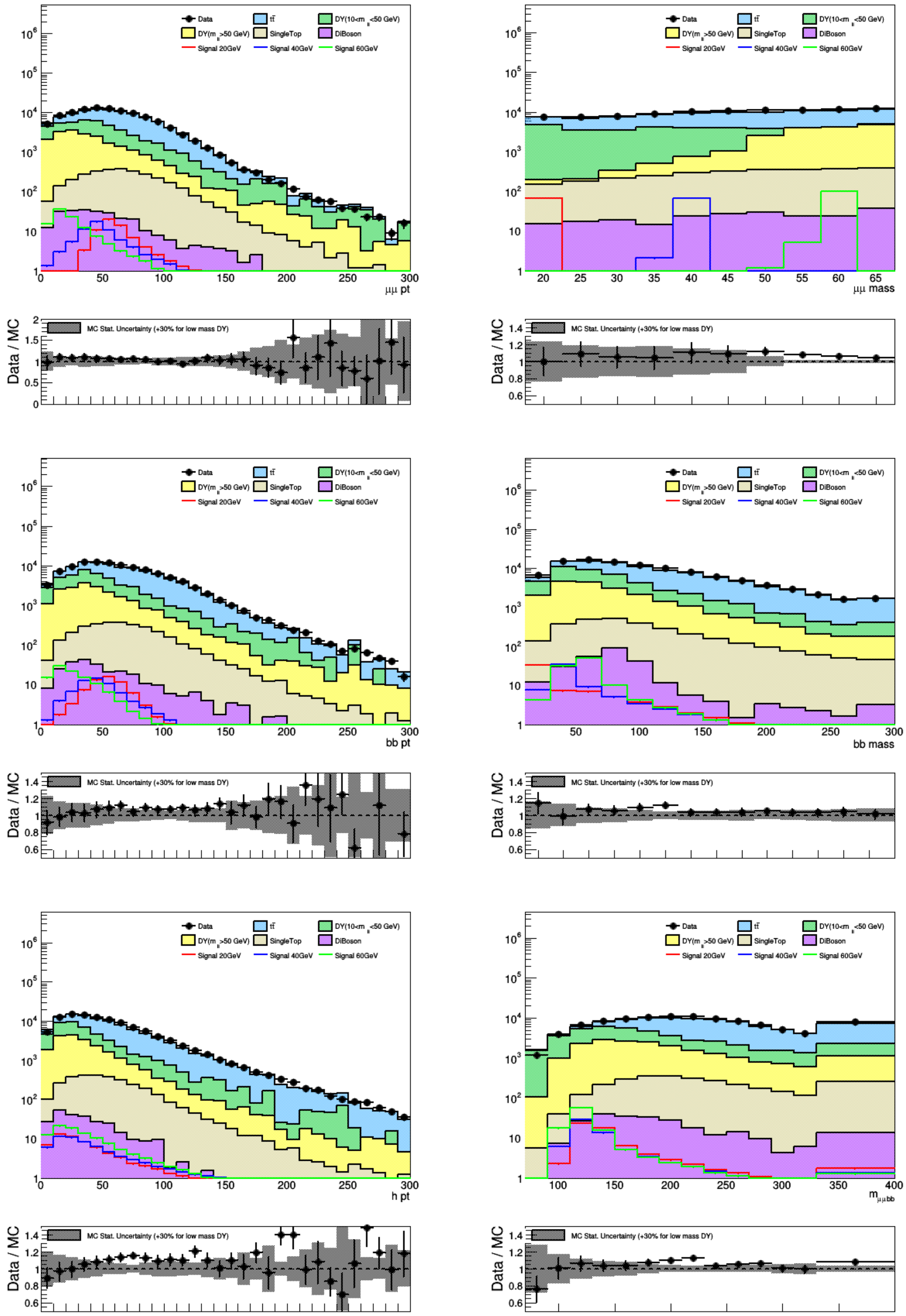


Figure 3.7: The p_T (left) and the invariant mass (right) distributions of the dimuon system (top), di-b-jets system (middle) and the Higgs candidate (bottom) after using the optimized preselection requirements. Simulated samples are normalized to 138 fb^{-1}. The band in the lower panel of each plot represents the MC statistical uncertainty.

3.4 Final selection

In order to reject more backgrounds and improve the sensitivity of the search, a χ^2_{tot} quantity is defined as,

$$\chi^2_{tot} \equiv \chi^2_{bb} + \chi^2_h \tag{3.1}$$

$$\chi_{bb} \equiv \frac{(m_{\mu\mu} - m_{bb})}{\sigma_{bb}} \tag{3.2}$$

$$\chi_H \equiv \frac{(m_{\mu\mu bb} - 125)}{\sigma_H} \tag{3.3}$$

where m_{bb} and $m_{\mu\mu}$ are the masses of di-b-jet and dimuon systems, respectively. The resolution of di-b-jet system is denoted by σ_{bb}, and σ_H is the resolution of the $\mu\mu bb$ system. Resolutions are derived from Gaussian fits to m_{bb} and $m_{\mu\mu bb}$ where objects are matched to the MC truth. Variable χ_{tot} is calculated on an event-by-event, encoding some crucial features of the signal event which are not necessary true for backgrounds. It is expected to be small in signal events since one expects $m_{a_1}(bb) = m_{a_1}(\mu\mu)$ and $m_{\mu\mu bb} = m_H$. For the analysis of 2016 data [15], the study of resolution terms showed that σ_{bb} increases with m_{a_1} while σ_H remains almost unchanged. This feature is re-confirmed using new simulations for different years, as shown in Fig. 3.8. In order to place a single cut on χ_{tot} over the entire m_{a_1} range, σ_{bb} is modeled with m_{a_1} and the dependence is included in Eq. 3.1.

As presented in Fig. 3.9, the χ_{bb} and χ_H are correlated. Also, it looks like the mean value of these quantities is not exactly placed at zero. While the bidimentional distribution of χ_{bb} versus χ_H looks elliptical in signal samples, placing a circular cut, $\chi^2_{tot} < X$, is sub-optimal. Also, such circular cut assumes signal being centered at zero which is not exactly the case. These two effects were ignored in the previous analysis [15]. Here however a detailed study is pursued to improve these shortcomings.

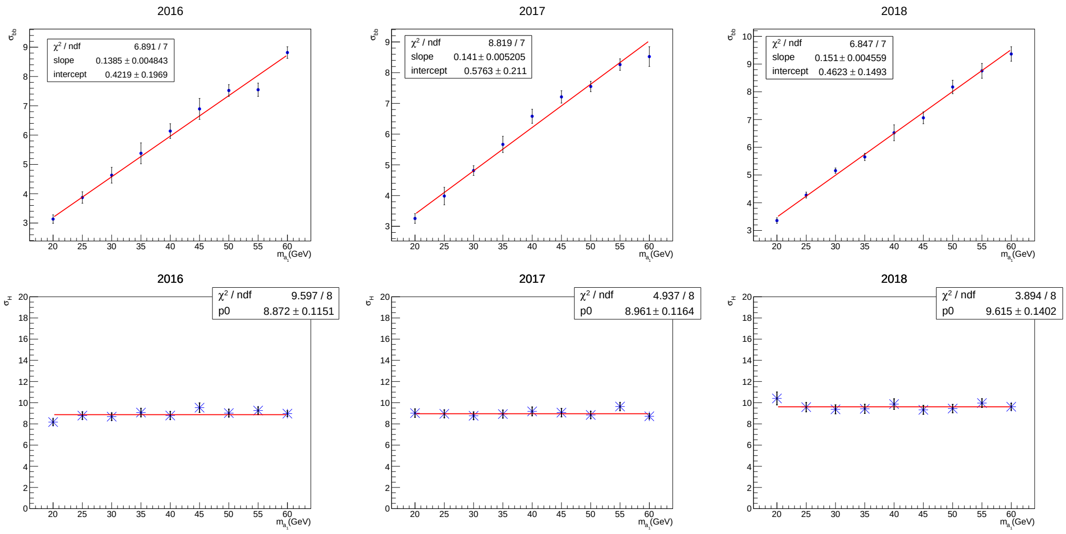


Figure 3.8: Dependence of σ_{bb} (top) and σ_H (bottom) on m_{a_1} in siganl samples of 2016 (left), 2017 (middle) and 2018 (right).

3.4.1 Offset correction

From Fig. 3.9 it is clear that the offset in the mean value changes with m_{a_1}. In order to correct this effect, the difference of m_{bb} and $m_{\mu\mu bb}$ with their nominal values (m_{a_1} and m_H, respectively) are modeled by a line as function of m_{a_1}. Line parameters for each year are obtained by fitting on the simulated mass points. The offset values can be found in Fig. 3.10. Shifted χ values are defined as

$$\chi_{b,s} \equiv \chi_{bb} - (a \times m_{\mu\mu} + b) \tag{3.4}$$

$$\chi_{h,s} \equiv \frac{(m_{\mu\mu bb} - 125)}{\sigma_h} - (c \times m_{\mu\mu} + d) \tag{3.5}$$

where a, b, c and d are the results of the linear fit. The distribution of the $\chi_{b,s}$ versus $\chi_{h,s}$ are shown in Fig. 3.11. It can be seen from the plots that shifted values are more centered around zero.

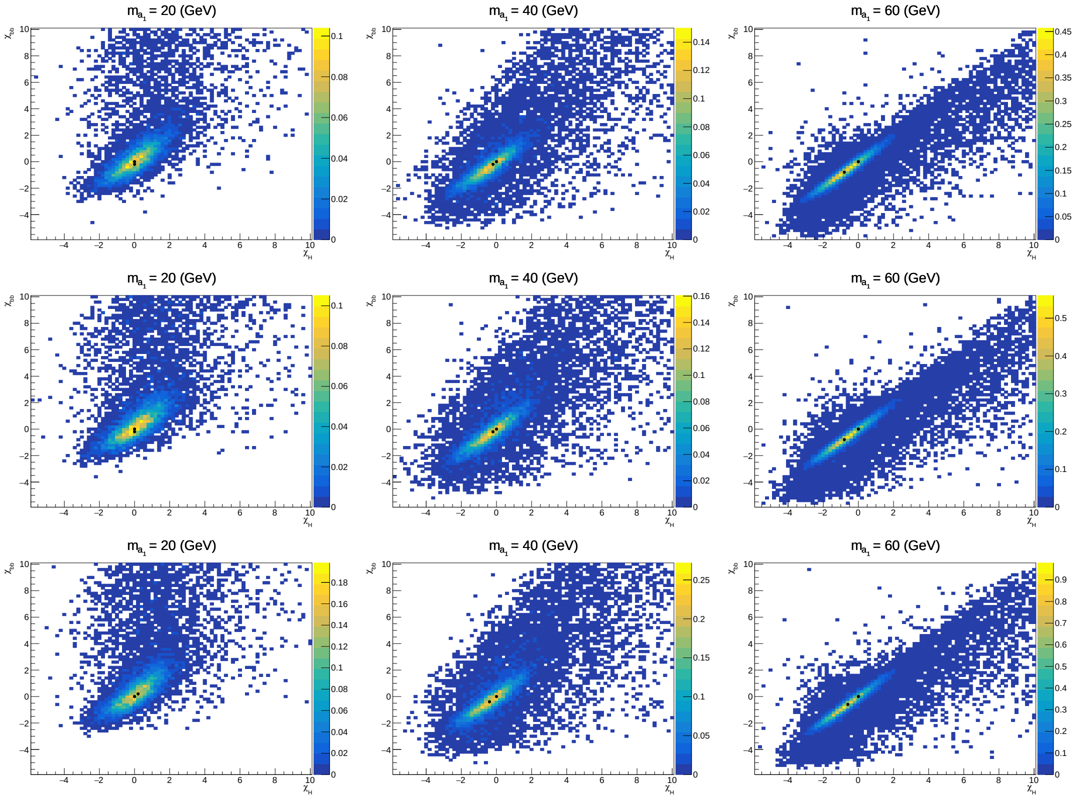


Figure 3.9: Bidimentional distributions of χ_{bb} versus χ_H in signal samples with different m_{a_1} values of 2016 (top), 2017 (middle) and 2018 (bottom).

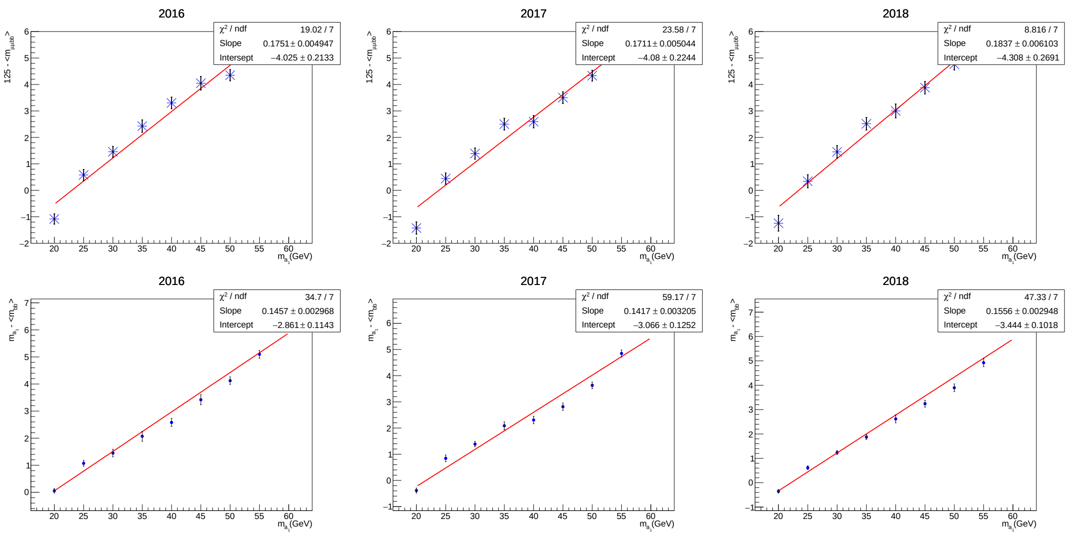


Figure 3.10: Dependence of offset in χ_H (top) and χ_{bb} (bottom) on m_{a_1} in signal samples of 2016 (left), 2017 (middle) and 2018 (right).

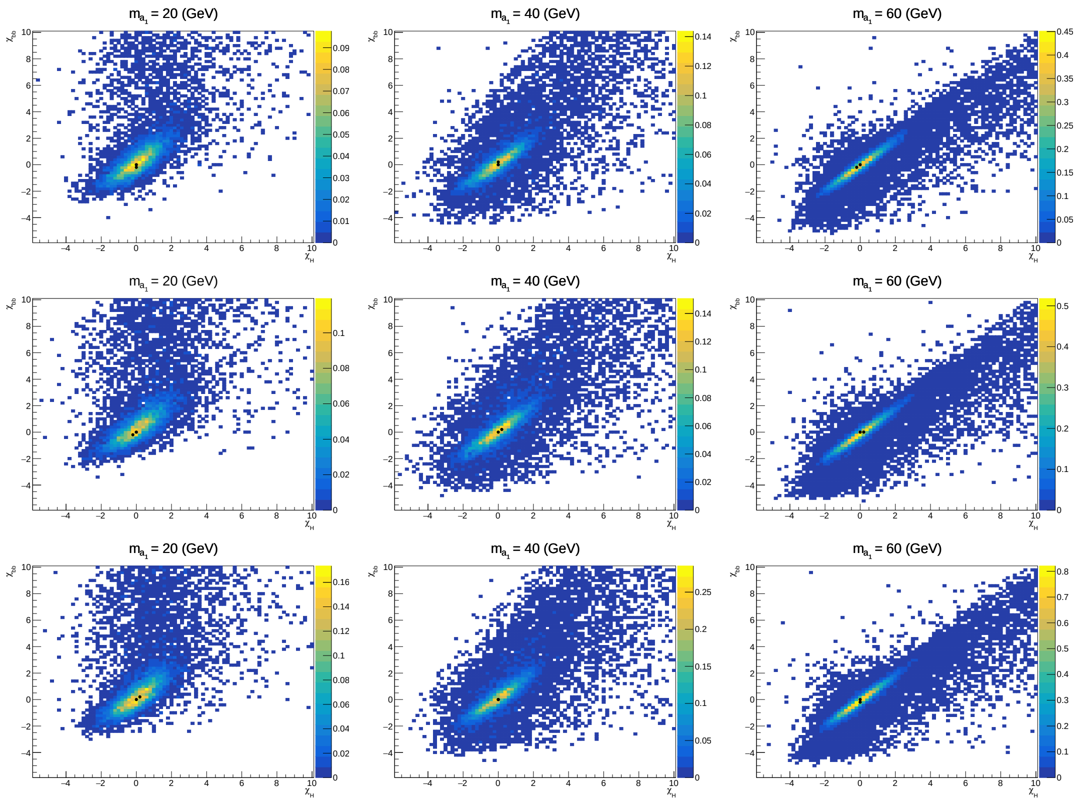


Figure 3.11: Bidimentional distributions of $\chi_{b,s}$ versus $\chi_{h,s}$ in signal samples with different m_{a_1} values of 2016 (top), 2017 (middle) and 2018 (bottom).

3.4.2 Decorrelation of $\chi_{b,s}$ and $\chi_{h,s}$

In order to decorrelate $\chi_{b,s}$ and $\chi_{h,s}$ the principal component analysis (PCA) method is used. The idea behind this method is to use the inverse of the covariance matrix of two variables to define two new uncorrelated variables. The covariance matrix is a 2×2 symmetric matrix, so it can be parameterized by its eigen values (λ_1, λ_2) and the components of one of its eigen vectors $\binom{a}{b}$. It can be shown that the two following variables are decorrelated:

$$\chi_{h,d} \equiv \frac{a}{\sqrt{\lambda_1}}(\chi_{h,s} + \frac{b}{a}\chi_{b,s}) \tag{3.6}$$

$$\chi_{b,d} \equiv \frac{a}{\sqrt{\lambda_2}}(-\frac{b}{a}\chi_{h,s} + \chi_{b,s}) \tag{3.7}$$

$$\chi_d^2 \equiv \chi_{h,d}^2 + \chi_{b,d}^2 \tag{3.8}$$

So this transformation can be parameterized using three parameters: $\frac{a}{\sqrt{\lambda_1}}$, $\frac{a}{\sqrt{\lambda_2}}$ and $\frac{b}{a}$ where we call it $\tan(\alpha)$. Polynomial functions are used to model the dependence of these variables to m_{a_1}. The fitted functions are shown in Fig 3.12 for all the years. To perform the transformation, the transformation matrix is calculated for each at $m_{\mu\mu}$. Correlation of $\chi_{h,d}$ and $\chi_{b,d}$ for $m_{a_1} = 20, 40, 60\, GeV$ is shown in Fig 3.13.

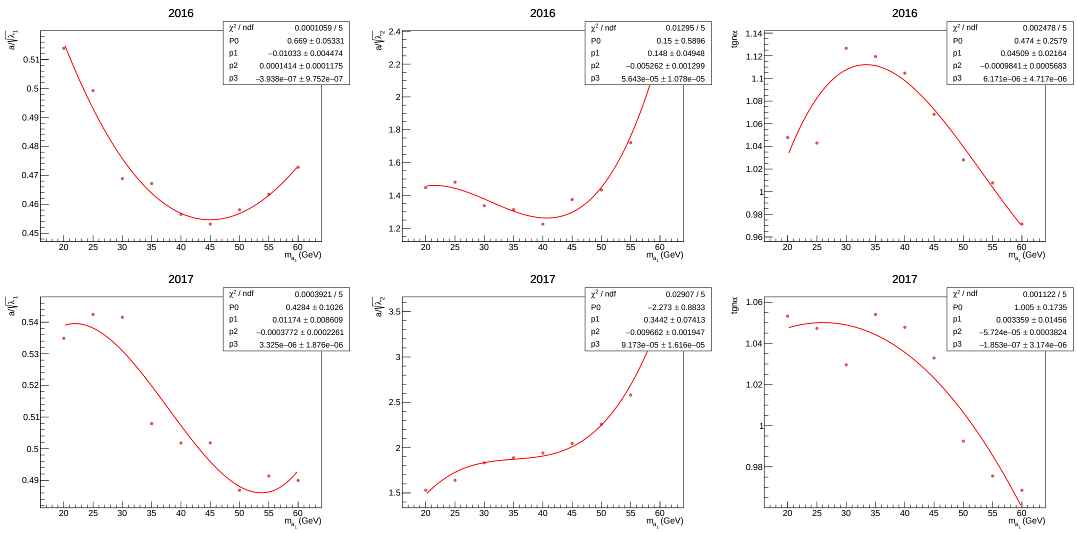


Figure 3.12: Dependence of decorrelation matrix parameters on m_{a_1}, $\frac{a}{\sqrt{\lambda_1}}$ (left), $\frac{a}{\sqrt{\lambda_2}}$ (middle) and $\tan(\alpha)$ (right) in signal samples of 2016 (top), 2017 (bottom).

3.4.3 Optimized χ_d^2

The distribution of the χ_d^2 for signal and backgrounds and a zoomed version of it are shown in Fig 3.14. Data in these plots are blinded for $\chi_d^2 < 3$. A similar procedure to that of Section 3.3.1, the optimal cut is found to be $\chi_d^2 < 1.5$ (Fig. 3.15). Table 3.4 summarizes the expected yields after this requirement. A clear improvement is visible in comparison with Table 3.3. In the final extraction of limits, the signal yield for mass points without simulation is obtained using SPline.

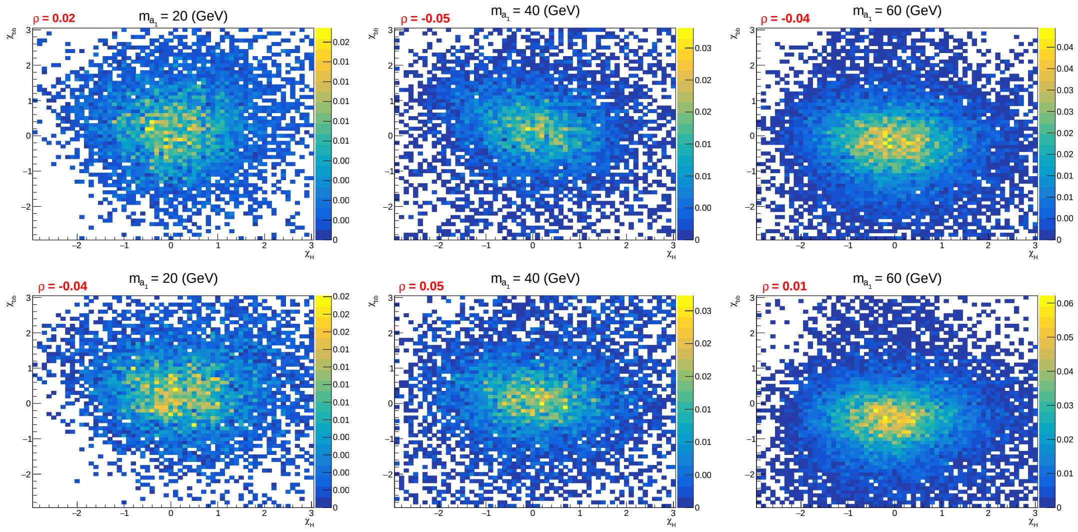


Figure 3.13: Bidimentional distributions of $\chi_{b,d}$ versus $\chi_{h,d}$ in signal samples with different m_{a_1} values of 2016 (top), 2017 (bottom).

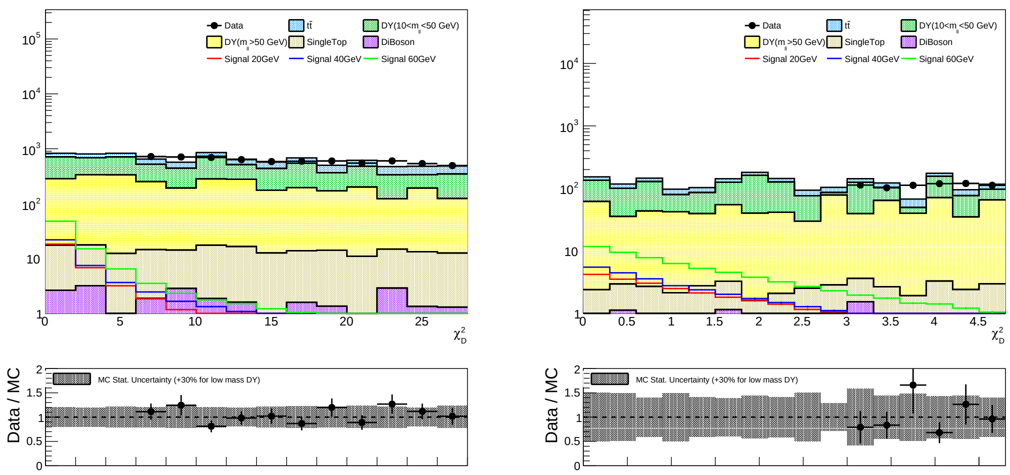


Figure 3.14: Distribution of χ^2_d for the full Run-II dataset in a wide range (left) and zoomed view (right).

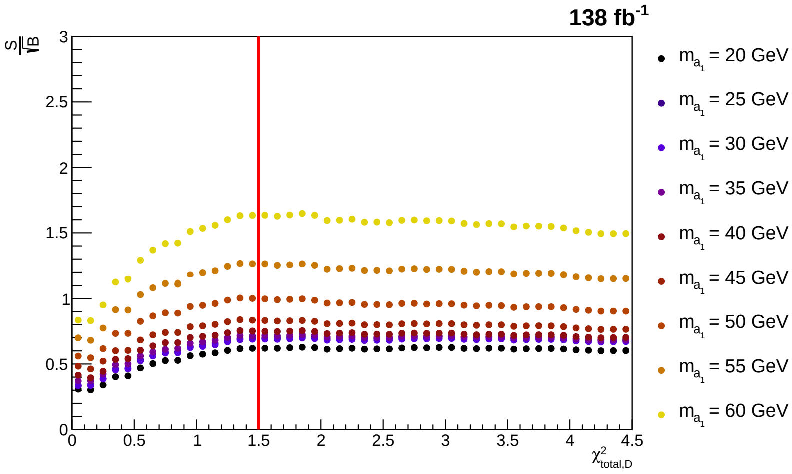


Figure 3.15: Significance as a function of the cut on the χ^2_d value for all m_{a_1} masses. The results are based on the full Run-II simulation.

Table 3.4: Event yields for simulated processes after applying $\chi^2_d < 1.5$ cut. The expected number of simulated events is normalized to the integrated luminosity of $138\,\mathrm{fb}^{-1}$. The type-III parameterization of 2HDM+S with $\tan\beta = 2$ is used to evaluate $\mathcal{B}(a_1 \to ff)$.

Process	Yield
Top	86.3 $\pm$ 2.25
Drell-Yan ($10 < m_{\ell\ell} < 50$)	289.6 $\pm$ 89.5
Drell-Yan ($50 < m_{\ell\ell}$)	200.2 $\pm$ 31.9
Diboson	1.5 $\pm$ 0.9
Single Top	11.4 $\pm$ 1.6
Total expected background	589.05 $\pm$ 95.09
Data	641

Signal for ggH ($\mu\mu bb$)		
$m_{a_1} = 20\,\mathrm{GeV}$	$m_{a_1} = 40\,\mathrm{GeV}$	$m_{a_1} = 60\,\mathrm{GeV}$
15.4 $\pm$ 0.2	18.7 $\pm$ 0.2	40.5 $\pm$ 0.3

Figure 3.16 shows the interpolation of yields for the ggH signals. Figure 3.17 compares the selection performance of the χ^2_d and χ^2_{tot} in terms of efficiencies for signal ($m_{a_1} = 40\,GeV$) and backgrounds. A clear improvement is visible after the offset correction and decorrelation. Events with $\chi^2_d < 1.5$ are selected following optimization studies [54].

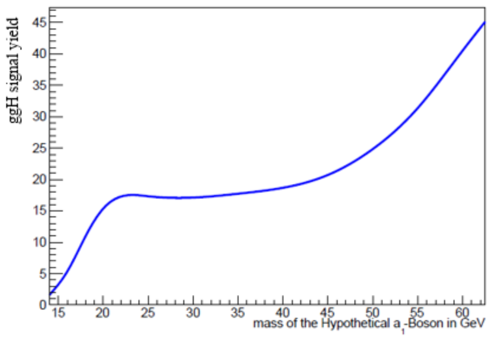


Figure 3.16: Interpolation of the expected signal yield for the ggH Higgs production mechanisms.

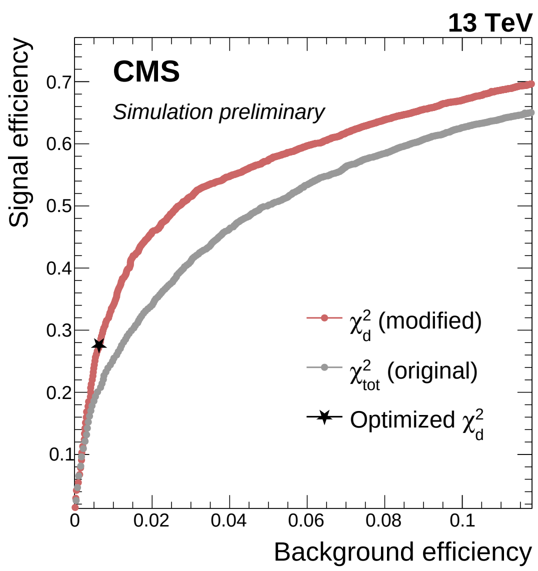


Figure 3.17: Signal ($m_{a_1} = 40\,GeV$) versus background efficiency for different cuts on χ^2_{tot} (gray) and χ^2_d (red) variables. The black star indicates signal efficiency versus that of background for the optimized cut value on χ^2_d.

3.5 Categorization

In order to increase the sensitivity of the analysis, it is important to categorize the events. First, events with at least one jet below 20 GeV are placed in a single category called $LowP_T$. The b-tagging SFs at 20 Gev were applied to these events and their uncertainties were doubled. These events don't categorize based on their B-tag value. This approach was presented at the BTV-WG meeting, and the group was satisfied with the application of SFs and uncertainties in our analysis. See further details in Appendix A. Then, events containing two additional jets events with $|\eta| < 4.7$ and $p_T > 30\,GeV$ are separated from the rest if the invariant mass of the two jet meets the requirement of $m_{jj} > 250\,GeV$. Rest of the events are categorized based on the b-tag discriminator value of the loose b-jet. Three exclusive categories are defined where the loose b-jet is not medium (TL), is medium but not tight (TM), and is tight (TT).

Such categorization relies on the fact that events with real and misidentified b-quark jets distribute differently in those categories. As shown in Fig.3.18, the majority of background events ($\approx 70\%$) fall into the $LowP_T$ category, and for signal(ggH) about 40% of signal events end up in the $LowP_T$ category.

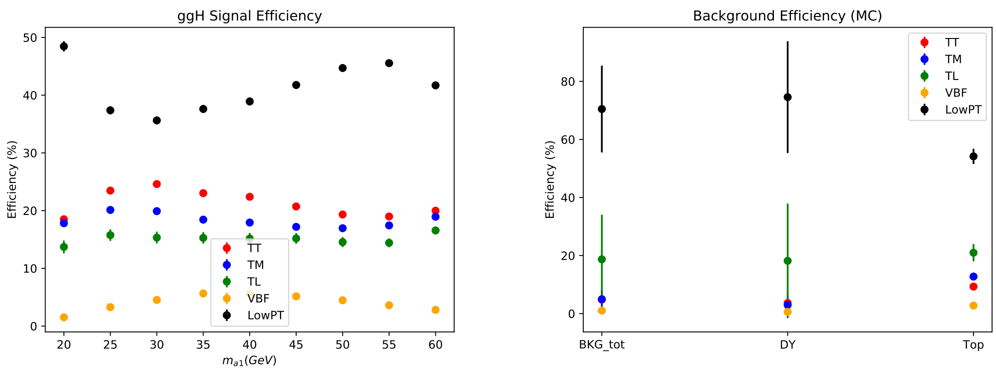


Figure 3.18: Fraction of ggH (left) and background events (right) in each category. For signal samples, fractions are provided for different m_{a_1} values.

Table 3.5 summarizes the final selection and categories of the current analysis, and the yield of expected background and signal samples in each category are summarized in Table 3.6.

Table 3.5: Summary of object and event selection requirements together with categorization, following optimization studies.

Physics objects and event kinematic requirements	
Leading and subleading muon p_T (GeV)	> 17 and 15
Muons identification and isolation	tight-loose
Leading and subleading jet p_T (GeV)	> 15
b-tagging	tight-loose
$\vec{E_T}^{miss}$ (GeV)	< 60
χ^2_d	< 1.5
Categories for selected events	
$LowP_T$	Events with at least one b-jet below 20 GeV
VBF	two add. jets with $p_T > 30\,GeV$, $\lvert\eta\rvert < 4.7$ and $m_{jj} > 250\,GeV$
TL	looser b-jet passes L but fails M
TM	looser b-jet passes M but fails T
TT	looser b jet passes T

Table 3.6: The expected yields for backgrounds and different signal hypotheses in each category. Since the individual numbers were rounded, the total yields are slightly different from the sum of the individual entries.

Category	Signal for ggH ($\mu\mu bb$)			Expected background
	$m_{a_1} = 20\,GeV$	$m_{a_1} = 40\,GeV$	$m_{a_1} = 60\,GeV$	
TL	2.1 ± 0.06	2.8 ± 0.07	6.7 ± 0.11	109 ± 30
TM	2.7 ± 0.06	3.3 ± 0.07	7.7 ± 0.11	27 ± 15
TT	2.8 ± 0.06	4.2 ± 0.08	8.1 ± 0.1	28 ± 11
VBF	0.2 ± 0.02	1.0 ± 0.04	1.1 ± 0.4	5 ± 2
$Lowp_T$	7.4 ± 0.11	7.3 ± 0.11	17 ± 0.17	421 ± 88
Total	15.4 ± 0.2	18.7 ± 0.2	40.5 ± 0.3	589 ± 95

3.6 Chapter summary

Analysis strategy

The analysis process started by selecting physic objects and applying kinematic cuts to ensure only relevant events were included. Then, optimal working points for the b-tagging algorithm were chosen, which helped in identifying jets originating from b-quarks. Additionally, the muon identification and isolation algorithms were further optimized enhance their efficiency. To distinguish the signal events from the background, a variable called χ^2_{tot} was defined. To optimize the cut on this variable, a statistical method called Principal Component Analysis (PCA) was employed. This cut efficiently removed most of the background events and identified the region enriched with the signal. This step greatly improved the sensitivity of the analysis. Then, to further enhance the analysis sensitivity, the events were divided into different categories.

CHAPTER

FOUR

SIGNAL AND BACKGROUND ESTIMATION

In this chapter, we provide an overview of the main components involved in the analysis of $H \to a_1 a_1 \to \mu\mu bb$ signal search, which include the signal and background models. These models serve as the foundation for testing the presence of a signal in the data. Once the events have been chosen, it is now time to estimate both the signal and the background. Signal estimation is elaborated on in Section 4.1, while background estimation is detailed in Section 4.2. Then, Section 4.3 outlines the systematic uncertainties.

4.1 Signal estimation

The procedure of signal estimation has already been validated in Run-I and 2016 Run-II publications [15, 16]. The shape of signal is derived from simulation and is a combination of a Voigt profile (V), and a Crystal ball profile (CB),

$$S(m_{\mu\mu}|f, p_V, p_{cb}) \equiv f \cdot V(m_{\mu\mu}|p_V) + (1 - f) \cdot CB(m_{\mu\mu}|p_{cb}), \tag{4.1}$$

with f being the relative contribution of the Voigt profile, p_V being the set of parameters for the Voigt and p_{cb} being the set of parameters for the Crystal ball profile functions. Figure 4.1 shows the signal distributions together with fitted functions where a reasonable signal description is obtained by the model for simulation for all three years. In the signal model, $S(m_{\mu\mu}|f, p_V, p_{cb})$, there are correlations amongst f,

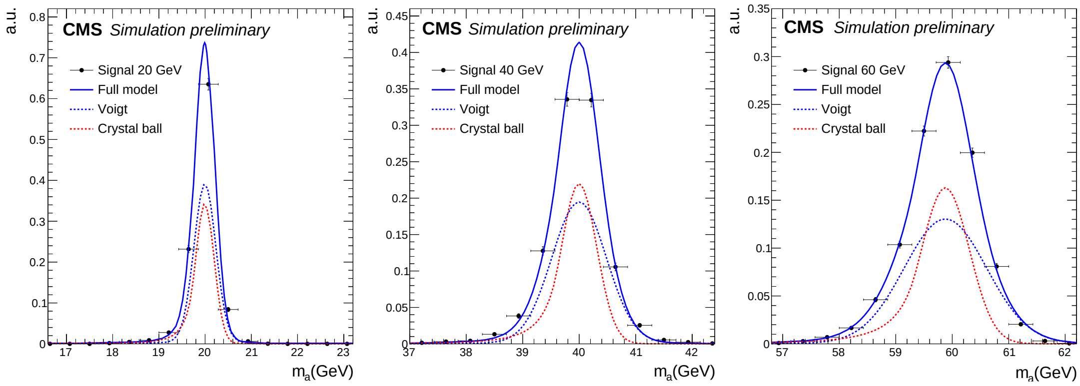


Figure 4.1: The modeling of $m_{\mu\mu}$ distribution in simulated signal event with (left) $m_{a_1} = 20\,GeV$,(middle) $m_{a_1} = 40\,GeV$, and (right) $m_{a_1} = 60\,GeV$. The full model (solid blue) is a weighted sum of the Voigt profile (dashed blue) and the Crystal ball function (dashed red).

p_V and p_{cb} and they could in principal change with m_{a_1}. The correlations are accounted for similarly to previous publications [15, 16], while preserving mass dependencies of relevant parameters. The only difference is that the same exercise is repeated for each year, considering possible differences in reconstructions. There are eventually three free parameters in the signal model: one is a perturbation to the mass of the hypothetical signal, $\widetilde{m}_{a_1} = m_{a_1} + \epsilon_{m_{a_1}}$, and the other two, a_1 and b_1, are used to linearly model the resolutions, σ and σ_{cb}, with mass.

4.2 Background estimation

Similar to the previous analyses [15,16], the background contribution is evaluated using a fully data-driven technique. In this technique, the background contribution is evaluated through a fit to the $m_{\mu\mu}$ distribution in data, without any reference to simulation. It is noteworthy to say that in order to examine the background determination method and estimate the sensitivity of the analysis, a control data region (CR) was defined and a similar method for background estimation was employed using the data in that side-band. The purpose of this is to estimate the sensitivity of the analysis in terms of the expected upper limit on $\mathcal{B}(H \rightarrow a_1 a_1 \rightarrow \mu\mu bb)$, before examining the data in signal region (SR).

4.2.1 Control data region definition

A control data region is identified that is similar to the expected background in SR. By *similar* we mean both the yield and the $m_{\mu\mu}$ distribution should be close enough to the background in SR. The χ^2_d variable is found to have small correlation with $m_{\mu\mu}$, meaning that for backgrounds the shape of the dimuon mass remains stable in different intervals of χ^2_d. Figure 4.2 shows the scattered plot of χ^2_d versus $m_{\mu\mu}$, including the correlations. The signal contamination is negligible for $\chi^2_d > 10$ (Fig. 4.3). A detailed study of simulation suggests a sideband of $12.5 < \chi^2_d < 14.5$ for 2017 together with $13.5 < \chi^2_d < 15$ in 2016 and 2018. These data regions are used to determine input models for backgrounds and the signal extraction fit while being blind to data in SR.

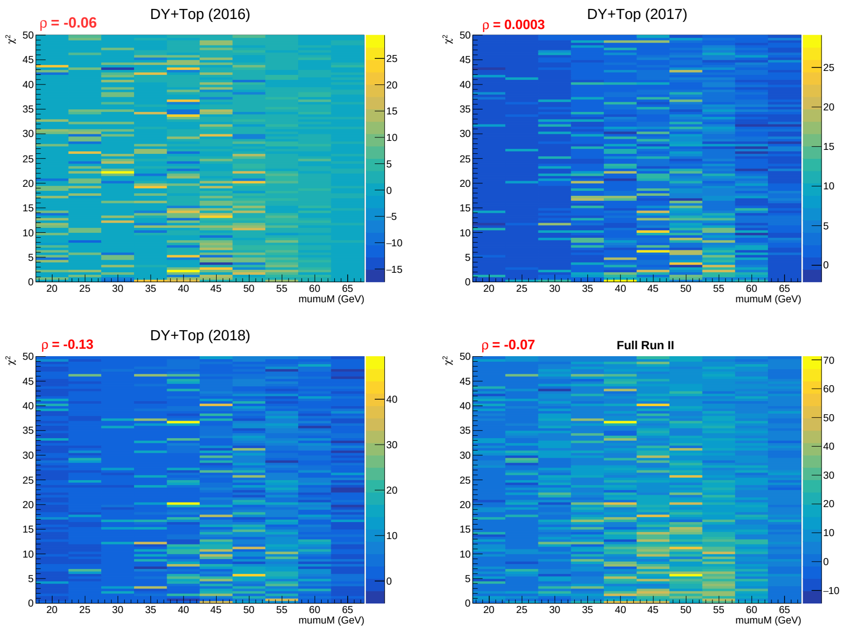


Figure 4.2: Distribution of χ^2_d vs $m_{\mu\mu}$ for simulated backgrounds in 2016 (top left) 2017 (top right) 2018 (bottom left) and full Run-II (bottom right).

4.2.2 Input background models

Different parametrization of polynomials (`RooPolynomial`, `RooBernstein`, `RooChebychev`) as well as $1/P_n(x)$ functions where $P_n(x) \equiv x + \sum_{i=2}^{n} \alpha_i x^i$ are fitted on $m_{\mu\mu}$ distribution to model the shape of the background in each analysis category. The maximum

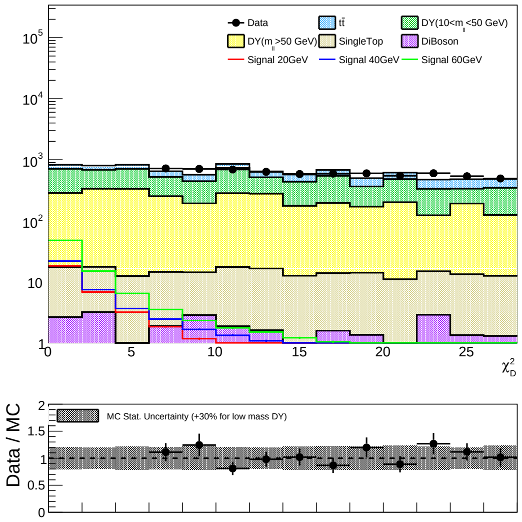
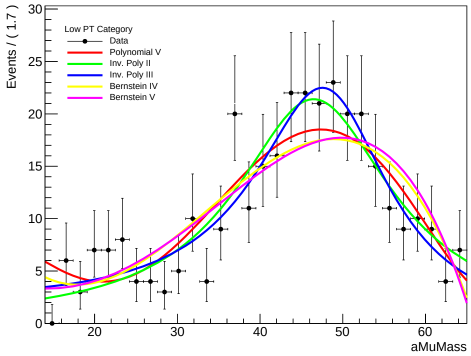


Figure 4.3: Distribution of χ^2_d for the full Run-II datasets.

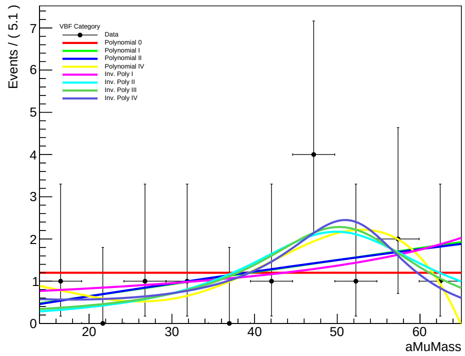
degree of the polynomial in each parametrization is determined through statistical tests. Different categories are cross checked versus each other to make sure that no pdf is repeated in the final pool of background models. The uncertainty associated with the choice of the background model is treated in a similar way to other uncertainties for which there are nuisance parameters in the final fit. Figure 4.4 illustrates the summary of background models found suitable to describe the data in CR. Then input background functions are tried in the minimization of the negative logarithm of the likelihood to extract final fit (appendix B). Upon confirming that the background modeling in CR is accurate and the limits are reliable, a similar approach can be used to estimate the background in SR.

4.2.3 Background estimation in signal region

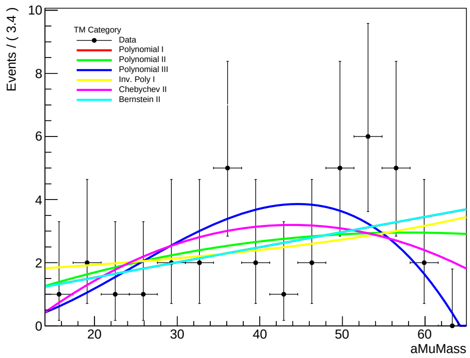
Based on a similar approach used in CR, background in SR is estimated. Figure 4.5 illustrates the summary of background models found suitable to describe the data in SR. The background contributions are extracted in the final fitting procedure described in Chapter 5.



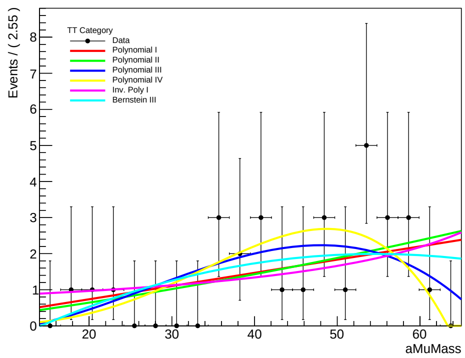
(a)



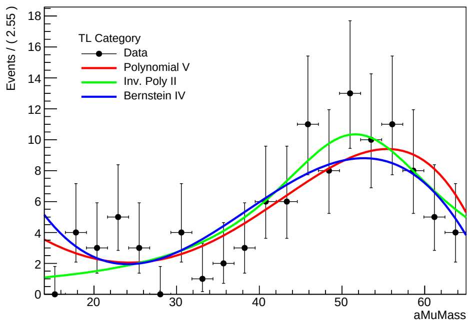
(b)



(c)

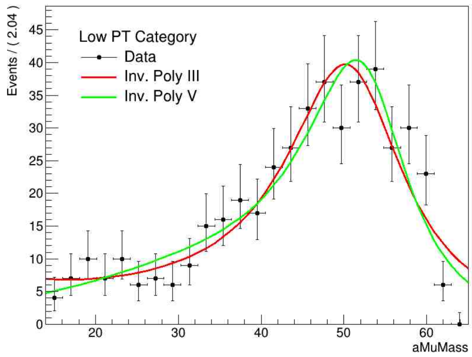


(d)

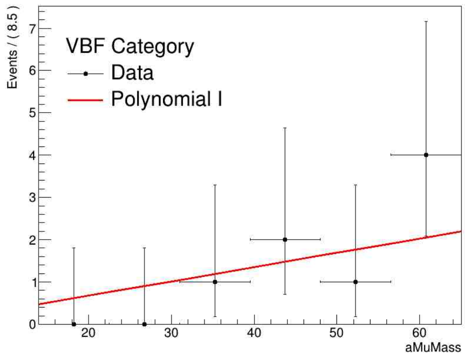


(e)

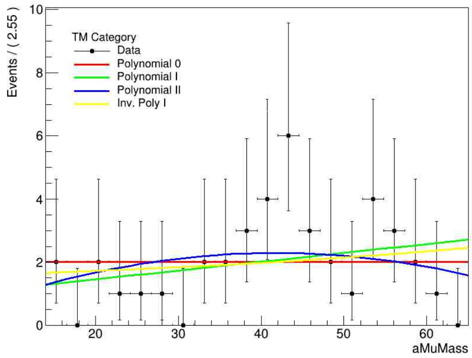
Figure 4.4: Suitable pdfs describing the background from control region in data in the $LowP_T$ (a), VBF (b), TM (c), TT (d), and TL (e) categories.



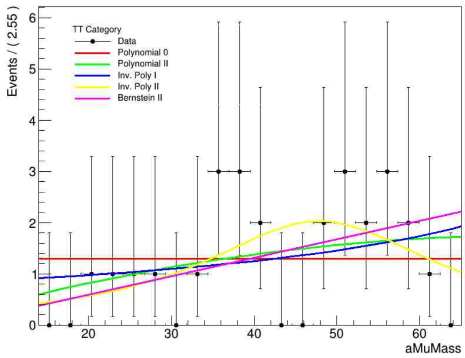
(a)



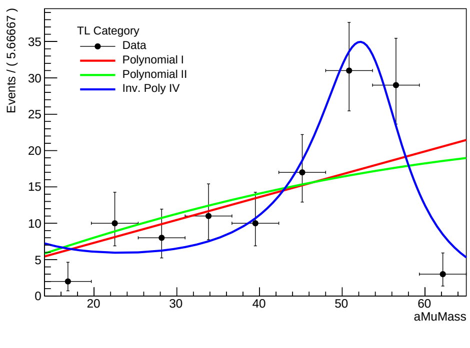
(b)



(c)



(d)



(e)

Figure 4.5: Suitable pdfs describing the background from signal region in data in the $LowP_T$ (a), VBF (b), TM (c), TT (d), and TL (e) categories.

4.3 Systematic uncertainties

The statistical interpretation of the analysis takes into account several sources of systematic uncertainties related to the accuracy in the signal modeling and uncertainties in the signal acceptance. Separate simulations for signal are employed for the data taken in the three years in order to match the varying detector performance and data taking conditions. Therefore the level of the systematic correlations between variations from the three periods is also specified. The imprecise knowledge of the background contributions is taken into account by the discrete profiling method will be described in Chapter 5.

Uncertainties on Higgs production cross section: to evaluate the upper limit on $\mathcal{B}(H \to a_1 a_1 \to \mu\mu bb)$, the Higgs boson production cross section is set to the SM prediction where an uncertainty of 3.6% is considered for the sum of the ggH and VBF production cross sections, accounting for QCD scale, PDF, and α_s uncertainties [55].

Experimental uncertainties: the uncertainty on the integrated luminosity [56–58] has a component that is uncorrelated across the years. It amounts to 1.0%, 2.0% and 1.5% respectively for 2016, 2017 and 2018 periods. Another component is correlated across all three years: 0.6%, 0.9%, and 2.0% respectively for years 2016, 2017 and 2018. Additionally, the luminosity measurements in 2017 and 2018 are accompanied by an additional uncertainty, 0.6% and 0.2% respectively, that is correlated between the two years. The uncertainty in the number of pileup interactions per event is estimated by varying the total inelastic pp cross section by ±4.6% [59], fully correlated across the years. An uncertainty of 2% is assigned to the trigger efficiency. The data-to-simulation correction factors for the muon reconstruction and selection efficiencies are estimated using a "tag-and-probe" method [60] in Drell–Yan data and simulated samples. These uncertainties include the pileup dependence of the correction factors and are considered correlated across the years since common procedural uncertainties are the dominant source. For the jet energy scale (JES), the variations are made according to the η- and p_T-dependent uncertainties and propagated to the p_T^{miss} of the event. These uncertainties include several sources and the correlation among the years is evaluated separately for each source.

An additional uncertainty, arising from unclustered energies in the event, is assessed for p_T^{miss}, considered uncorrelated across the years. For the jet energy resolution, the smearing corrections are varied within their uncertainties [52], uncorrelated across the years. Systematic uncertainty sources that affect the data-to-simulation corrections of the b-tagging discriminator distribution are JES, the contamination from light flavor (LF) jets in the b-jet sample, the contamination from heavy flavor (HF) jets in the light-flavor jet sample, and the statistical fluctuations in data and MC [61]. The JES variations b-tagging are performed consistently with uncertainties on jets and follow the same correlation pattern across the years. The statistical components of the b-tagging Uncertainties are considered uncorrelated while the rest are assumed to be correlated between different periods. During the 2016 and 2017 data taking periods, a gradual shift in the timing of the inputs of the ECAL Level-1 trigger in the forward endcap region ($|\eta| > 2.4$) led to a specific inefficiency. A correction for this effect was determined using an unbiased data sample and is found to be relevant in events with high-p_T jets with $2.4 < |\eta| < 3.0$. This predominantly affect the VBF category of this analysis which has very low statistics. Although the effect on the final result is not expected to be large, a systematic variation of 20% is considered for this correction.

Modeling uncertainties: uncertainties in this category are considered fully correlated across the years. This is because the same model assumptions and procedure are employed in deriving variations. Uncertainties in the choice of the normalization, μ_R, and factorization, μ_F, scales are estimated by doubling and halving μ_R and μ_F simultaneously in the signal sample. For the parton-shower simulation, uncertainties are separately assessed for initial- and final-state radiations, by varying the respective scales up and down by factors of two. Finally, uncertainties arising from the limited understanding of the PDFs [62] are taken into account.

Based on the studies on 2016 data [15] uncertainties have a negligible effect on the shape of the signal. Their effects on the yield are taken into account by introducing nuisance parameters with log-normal distributions into the final fit. In Fig.4.6, the observed impact is presented for two signal hypotheses with $m_{a_1} = 20$ and 35 GeV. For the two masses, it can be seen that the effects of different sources of uncertainty are different. Overall, systematic uncertainties do not play a significant role.

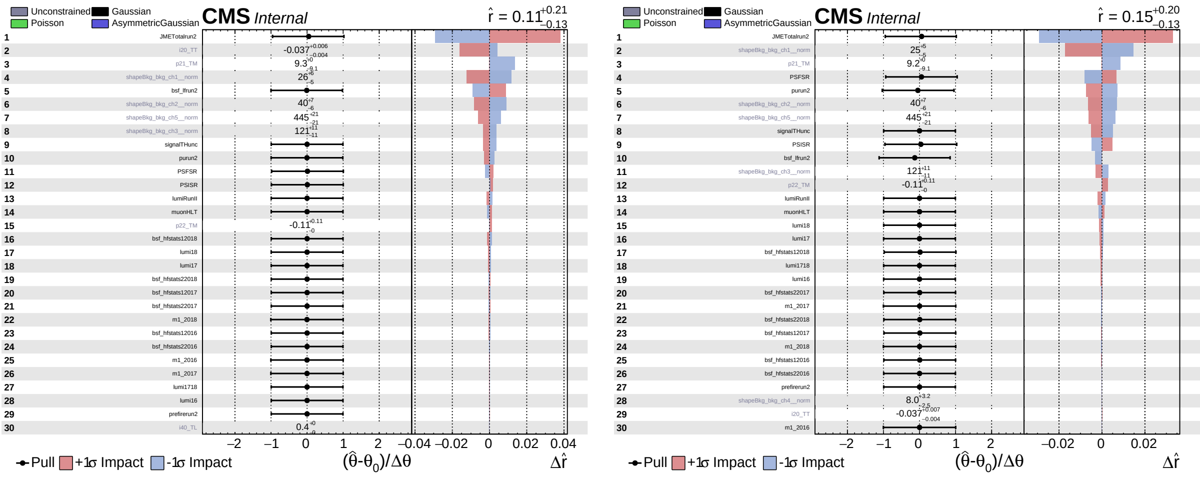


Figure 4.6: Observed impact of systematic uncertainties for a signal with $m_{a_1} = 20\,GeV$ (left) and with a $m_{a_1} = 35\,GeV$ (right).

4.4 Chapter summary

Signal and background estimation

To estimate the signal, a parametric model is fitted to the $m_{\mu\mu}$ distribution of simulated events (MC). Due to the low statistics and systematic uncertainties of the simulated events, the background is estimated from data in the signal region. A combination of signal and background models and systematic uncertainties is then used in final fitting procedure described in Chapter 5.

CHAPTER

FIVE

RESULTS

This chapter provides an overview of the results from the hypothesis testing procedure, which compares the compatibility of the observed data with the background-only and signal plus background hypotheses in Section 5.1. The results of the search are then summarized in terms of upper limits on excluded branching fraction in Section 5.2. Next, in Section 5.3, the current results will be combined with the results from the $\tau\tau bb$ final state, and the interpretation of the combined results will be presented in the context of the 2HDM+S model.

5.1 Signal extraction

A simultaneous unbinned maximum likelihood fit is conducted on the data in all event categories for the $m_{\mu\mu}$ distributions. The fit is performed within the range of 15 to 62.5 GeV, using parametric models for both the signal and background. Specifically, the signal model consists of a combination of a Voigt profile and a Crystal Ball function, with the mean values of the two functions constrained to be the same as discussed in the previous chapter.

To evaluate the background contribution, every selected functional form is treated as a discrete nuisance parameter. In addition, the parameters of each model, as well as the normalization, are part of the background parameter space. A likelihood $\mathcal{L}$ is constructed using the signal and the pool of the background models in all categories, including systematic uncertainties associated with the signal as nuisance parameters.

In the minimization process of the negative logarithm of the likelihood, the discrete profiling method chooses a best-fit background model as the physics parameter of interest, the signal strength, varies. The method incorporates the systematic uncertainty on the background model by taking the envelope of the models provided to the fit. In practice, a penalty term is added to the likelihood to account for the number of free parameters in the final background model. The likelihood ratio for the penalized likelihood function can be written as

$$-2\ln\frac{\widetilde{\mathcal{L}}(\mathrm{data}|\mu,\hat{\theta}_\mu,\hat{b}_\mu)}{\widetilde{\mathcal{L}}(\mathrm{data}|\hat{\mu},\hat{\theta},\hat{b})}, \tag{5.1}$$

where μ is the measured quantity. The numerator is the maximum penalized likelihood for a given μ, at the best-fit values of nuisance parameters, $\hat{\theta}_\mu$ and of the background function, $\hat{b}_\mu$. The denominator is the global maximum for $\widetilde{\mathcal{L}}$, achieved at $\mu=\hat{\mu}$, $\theta=\hat{\theta}$ and $b=\hat{b}$. A confidence interval on μ is obtained with the background function maximizing $\widetilde{\mathcal{L}}$ for any value of μ. This interval is always wider than those evaluated with the fixed functional form from the global best-fit, $b=\hat{b}$ [63].

Finally, to check any possible bias in the real data, we used the pseudo-data (Asimov) and injected the signal with different values (m_{a_1}) to the pseudo-data. No bias was observed in the fitting result (see Appendix C).

The best-fit background models together with their uncertainties are shown in Fig. 5.1 for all event categories after a combined fit of all categories [54, 64].

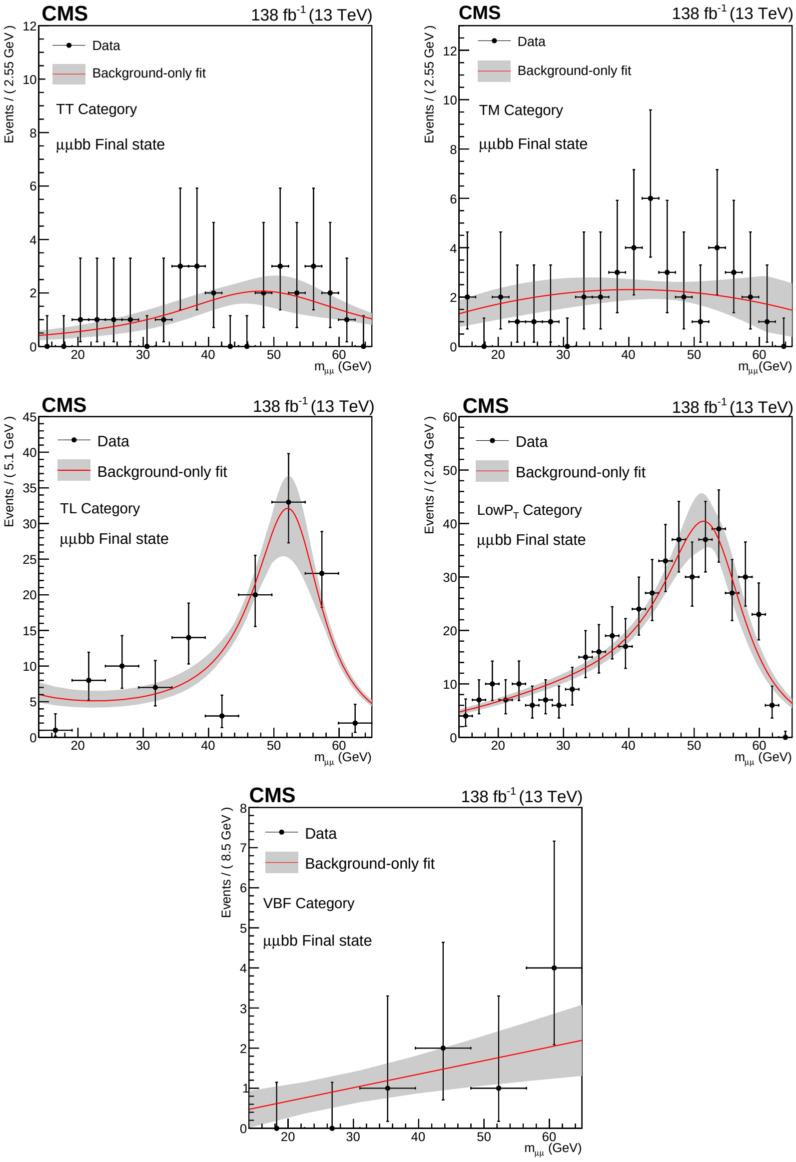


Figure 5.1: The best-fit background models together with 68% CL uncertainty band from the fit to the data under the background-only hypothesis for the (top left) TT category, (top right) TM, (middle left) TL category, (middle right) $LowP_T$ category, and (bottom) VBF category.

The presence of the hypothetical signal on top of the best-fit background model leads to additional uncertainties. Figure 5.2 shows the results of the signal-plus-background fit for two signal hypotheses, $m_{a_1} = 20$ GeV and $m_{a_1} = 40$ GeV, in two given categories, TM and $LowP_T$. One can also see the typical signal width in comparison with background shapes [54]

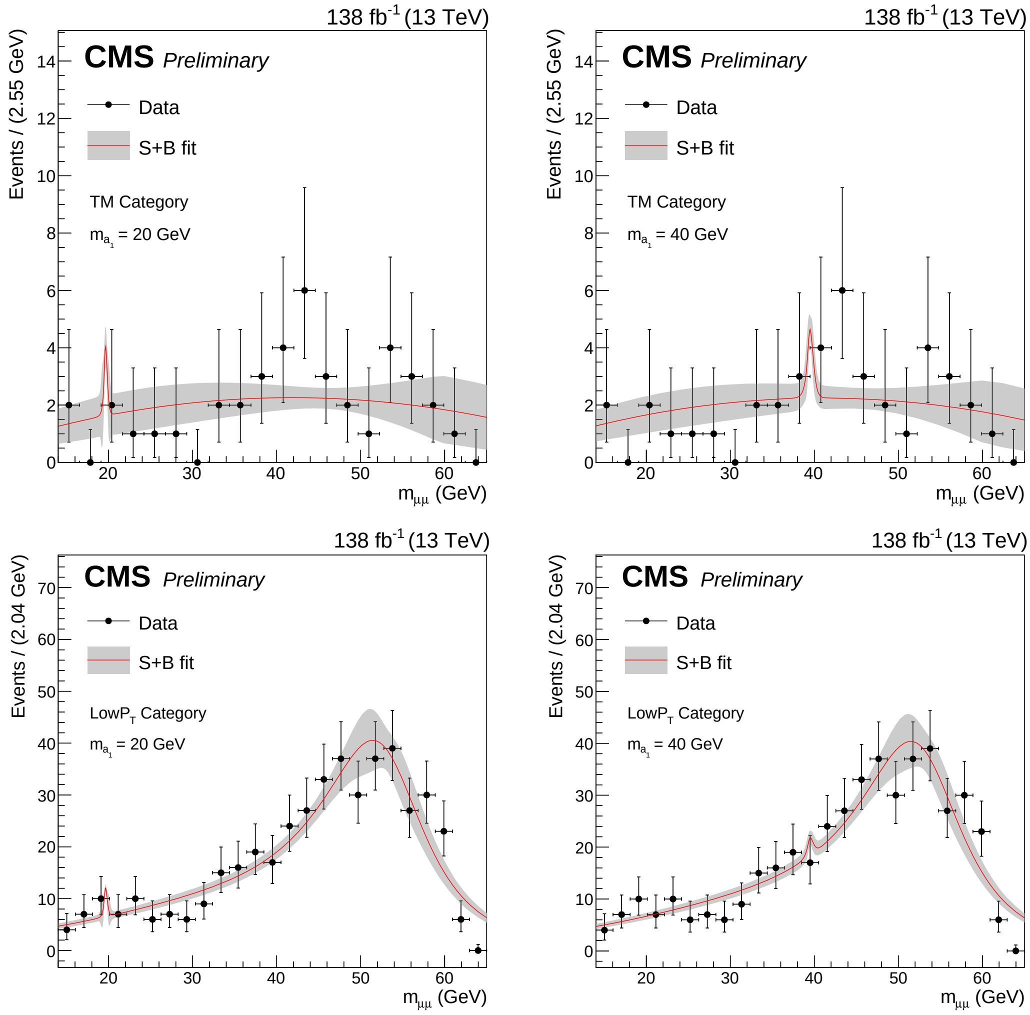


Figure 5.2: The signal-plus-background fit for two signal hypotheses, (left) $m_{a_1} = 20$ GeV and (right) $m_{a_1} = 40$ GeV, in two given categories, (top) TM and (bottom) $LowP_T$.

5.2 Upper limits

No significant excess of events is observed beyond the estimated backgrounds in the $m_{\mu\mu}$ distribution. Using the asymptotic CL_s method, the observed upper limits are set, at 95% confidence level (CL), on $\mathcal{B}(H \to a_1 a_1 \to \mu\mu bb)$, which is shown in Fig. 5.3. The corresponding median expected upper limit (blue dashed curve), along with its $\pm 1\sigma$ (green band) and $\pm 2\sigma$ (yellow band) confidence intervals, are also shown. Current limits are more stringent than those of the CMS analysis using 2016 data [15]. The main improvement comes from the modification of the χ^2_{tot} definition, and the new categorization improved the results for low m_{a_1} values as well. At 95% CL, the observed upper limits on $\mathcal{B}(H \to a_1 a_1 \to \mu\mu bb)$ are $(0.17 - 3.3) \times 10^{-4}$ for the mass range 15 to 62.5 GeV, whilst the expected limits are $(0.35 - 2.6) \times 10^{-4}$. The improvement over the earlier results by CMS [15] goes, by about a factor of two, beyond the expectation from pure increase in the size of the data sample [54].

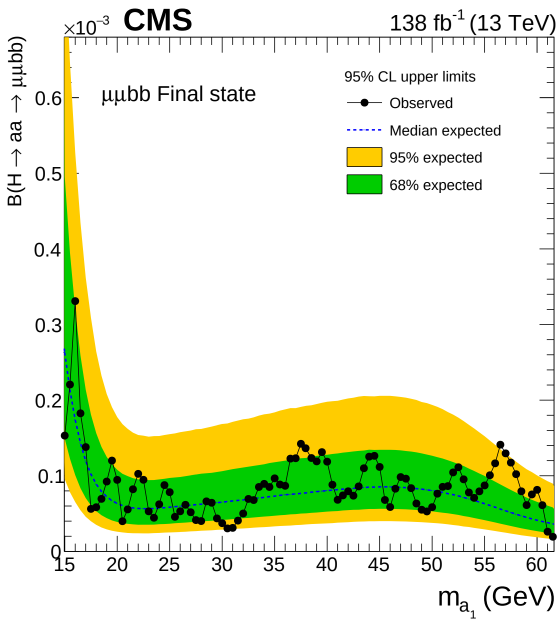


Figure 5.3: Observed limits, at 95% CL, on $\mathcal{B}(H \to a_1 a_1 \to \mu\mu bb)$ at 13 TeV using full Run-II data.

5.3 Combination of $\mu\mu bb$ and $\tau\tau bb$ channels

In order to increase the sensitivity of the analysis, the current results are combined with those from another channel ($\tau\tau bb$) that has been searched by a group at Princeton University and the University of Wisconsin-Madison. The mass of the pseudoscalar is considered to be between 12 and 60 GeV in the $\tau\tau bb$ channel. The $\mu\mu bb$ and $\tau\tau bb$ analyses together introduce a diverse set of final state particles for a $H \to a_1 a_1$ signal. In $\tau\tau bb$, final states with at least one τ lepton decaying to an electron or muon, ie, $e\mu$, $e\tau_h$, and $\mu\tau_h$, are considered. Decays resulting in the same flavor leptons, or in two τ_h, are not included as they bring negligible sensitivity to the analysis. The signal acceptance of $\tau_h\tau_h$ is very low due to high trigger thresholds, whereas ee and $\mu\mu$ final states suffer from low branching fractions. The $\mu\mu bb$ analysis on the other hand relies on the presence of two muons.

Since the results in both channels are statistically limited, the combination mostly benefits from the additional data rather than correlations between common systematic uncertainties. The combined results are still dominated by statistical uncertainties. At $m_{a_1} = 35\,GeV$, all systematic uncertainties amount to about 6% of the total uncertainty. The dominant contributions to systematic uncertainties belong to JES in $\mu\mu bb$, followed by the theory uncertainties on the signal, and the uncertainties on the QCD multijet backgrounds in the $e\mu$ final state of the $\tau\tau bb$ analysis.

Figure 5.4 shows the observed and expected 95% CL exclusion limits on the branching fraction $\mathcal{B}(H \to a_1 a_1 \to \tau\tau bb)$ in % as a function of m_{a_1}. Only the $e\mu$ channel provides sensitivity to the $12\,GeV$ mass point, as in this channel the baseline cut on the ΔR between the two τ candidates is the lowest. For small m_{a_1} values, the decay products appear as boosted and may not be reconstructed as two separate objects. The low ΔR requirement allows to select more signal events where the two τ candidates are close to each other [64]. The DNN categorization brings significant sensitivity improvement compared to the cut-based strategy applied in the earlier analysis using data from 2016 only [65].

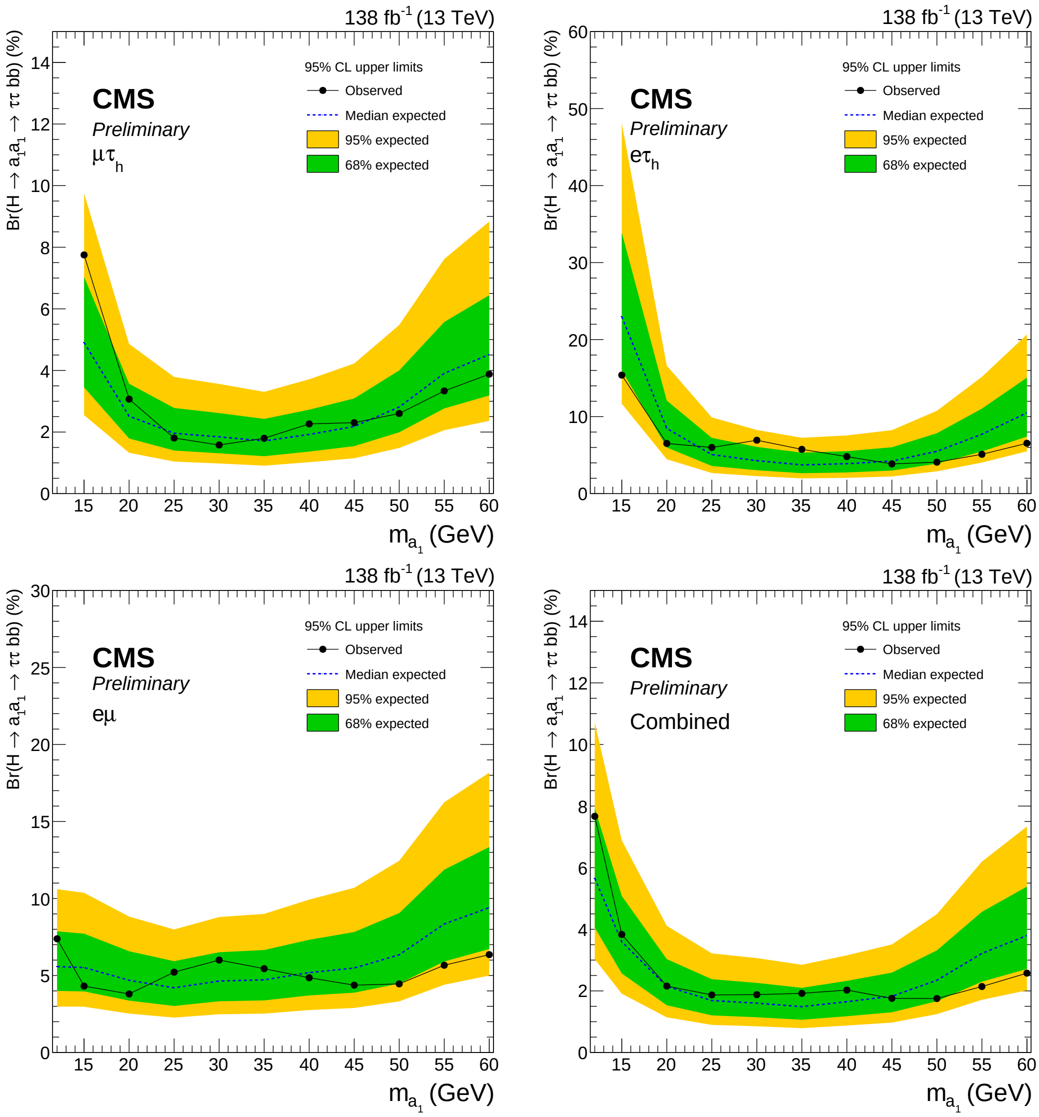


Figure 5.4: Observed and expected 95% CL exclusion limits on $\mathcal{B}(H \to a_1a_1 \to \tau\tau bb)$ in %, for the combination of all years with an integrated luminosity of $138 fb^{-1}$ per channel and the combination. Top left: $\mu\tau$, top right: $e\tau$, bottom left: $e\mu$, and bottom right: combination of all channels is shown.

The $\mu\tau_h$ channel is the most sensitive channel, where limits as low as around 1.8% are set in the intermediate mass range at 35 GeV. Combining all final states in $\tau\tau bb$, expected limits on the branching fraction are found to be in range (1.5–5.6)%, for the pseudoscalar mass between 12 and 60 GeV, corresponding to observed limits in range (1.8–7.7)% at 95% CL.

No excess of events over the expected SM backgrounds is observed in either of the $\mu\mu bb$ and $\mu\mu\tau\tau$ channels. Upper limits are placed, at 95% CL, on the branching ratios $\mathcal{B}(H \rightarrow a_1a_1 \rightarrow llbb)$ as a function of m_{a_1}, with ll being either τ lepton or muon. The two final states are combined to draw the limits on $\mathcal{B}(H \rightarrow a_1a_1)$, assuming fixed decay fractions of a_1 . The branching fractions $\mathcal{B}(a_1 \rightarrow ff)$ depend on the 2HDM+S parameters where f stands for muons, b quarks, and τ leptons.

Figure 5.5 shows the observed and expected limits at 95% CL on $\mathcal{B}(H \rightarrow a_1a_1 \rightarrow llbb)$ as functions of m_{a_1} for 2HDM+S models, independent of the type and $\tan\beta$ parameter. The combined branching fraction $\mathcal{B}(H \rightarrow a_1a_1)$ is obtained upon reinterpretation of the $\mu\mu bb$ and $\tau\tau bb$ results in different types of 2HDM+S and $\tan\beta$ values for $15 \leq m_{a_1} \leq 60$ GeV. Upper limits in range (5–23)% are observed at 95% CL for all Type II scenarios with $\tan\beta > 1.0$. The tightest constraint is obtained for the Type-III model with $\tan\beta = 2.0$. At 95% CL, the expected combined branching fraction is in range (2–10)%, with a similar range for the observed upper limits. For the Type-IV model, the expected upper limits on $\mathcal{B}(H \rightarrow a_1a_1)$ at 95% CL are between about 3% and 11% for $\tan\beta = 0.5$, corresponding to observed limits between 2% and about 12% [64].

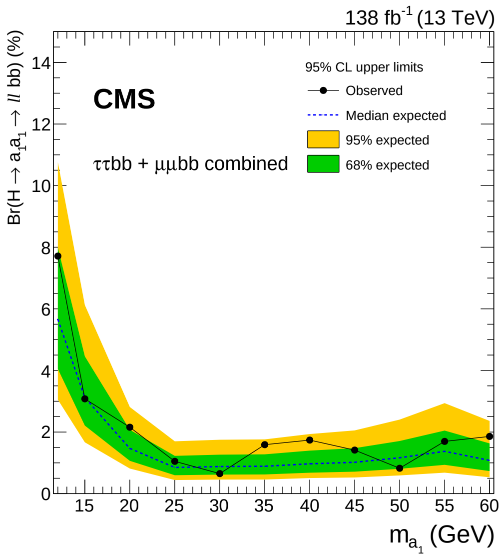


Figure 5.5: Observed and expected 95% CL upper limits on $\mathcal{B}(H \rightarrow a_1 a_1 \rightarrow llbb)$ in %, where ll stands for muons or tau leptons, obtained from the combination of $\mu\mu bb$ and $\tau\tau bb$ channels using the full Run-II integrated luminosity of $138fb^{-1}$. The results are obtained as functions of m_{a_1} for 2HDM+S models, independent of the type and $\tan\beta$ parameter.

The allowed values of $\tan\beta$ and m_{a_1} are shown in Fig. 5.6 in the context of Type-III and Type-IV 2HDM+S. The dashed contours represent the 95% upper limits on Higgs to pseudoscalar decays assuming the branching fraction to be 100% or 16% [64].

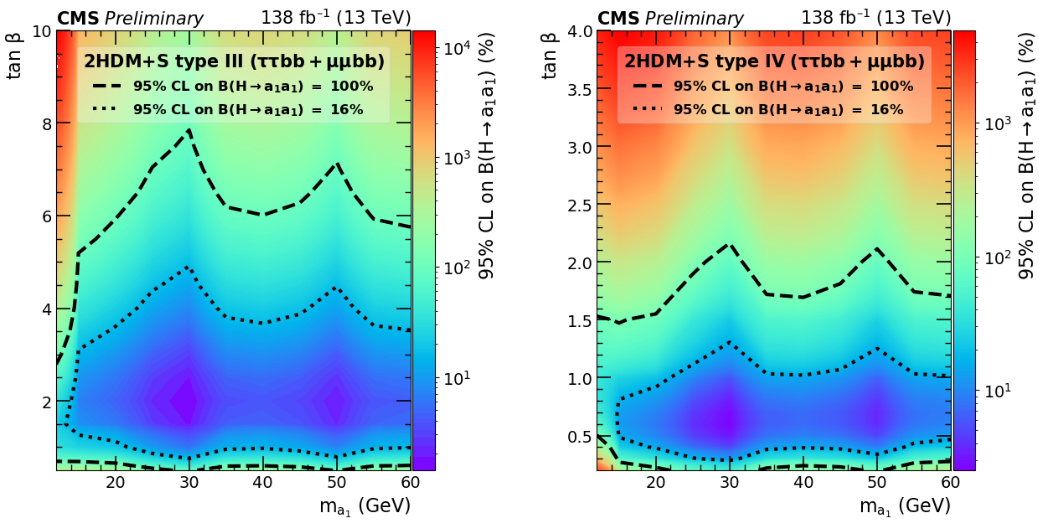


Figure 5.6: Observed 95% CL upper limits on $\mathcal{B}(H \to a_1 a_1)$ in %, for the combination of $\mu\mu bb$ and $\tau\tau bb$ channels using the full Run-II integrated luminosity of $138 fb^{-1}$ for Type-III (left) and Type-IV (right) 2HDM+S in the $\tan\beta$ vs. m_{a_1} phase space. The contours corresponding to branching ratios of 100% and 16% are drawn using dashed lines, 16% corresponding to the previous upper limit on Higgs to BSM particle decays from previous Run-II results [66]. Linear extrapolation has been used between different points on the figures.

5.4 Chapter summary

Results

The search for $H \to a_1a_1 \to \mu\mu bb$ was carried out using a statistical hypothesis test, where the Maximum Likelihood Estimation (MLE) method was employed to create a test statistic. This statistic was used to assess the consistency of the data with the $H \to a_1a_1 \to \mu\mu bb$ scenario. Finally, The CLs metric was used to set upper limits on $\mathcal{B}(H \to a_1a_1 \to \mu\mu bb)$, in the event of no detection.
In the framework of the 2HDM+S model, the results of the two final state searches are combined to determine upper limits on $\mathcal{B}(H \to a_1a_1 \to llbb)$ at 95% CL, with ll being a muon or a tau lepton. Those limits are independent of the type and other parameters of the models considered in this thesis, such as the ratio of the vacuum expectation values of the doublets. For different types of 2HDM+S, upper bounds on the branching fraction $\mathcal{B}(H \to a_1a_1)$ are extracted from the combination of the two channels.

Part II

Luminosity calibration

CHAPTER

ONE

LUMINOSITY MEASUREMENT

1.1 Luminosity and integrated luminosity

In particle physics experiments the energy available for the production of new effects is the most important parameter. The required large center-of-mass energy can only be provided with colliding beams where little or no energy is lost in the motion of the center-of-mass system (cms). Besides the energy the number of useful interactions (events), is important. This is especially true when rare events with a small production cross-section (σ_{eff}) are studied. The quantity that measures the ability of a particle accelerator to produce the required number of interactions is called the luminosity and is the proportionality factor between the number of events per second $R(N/t)$ and the cross-section σ_{eff} :

$$l = \frac{dR}{\sigma_{eff}}. \tag{1.1}$$

The unit of the luminosity is therefore $cm^{-2}s^{-1}$. By integrating the luminosity over time, the integrated luminosity is obtained as follows:

$$L = \int l dt. \tag{1.2}$$

In order to study and accurately measure the interactions between two protons, the amount of integrated luminosity caused by their collisions must be accurately measured. Cross-section measurements will be more accurate if the luminosity is measured with greater accuracy, or with less uncertainties.

1.2 Luminosity measurement

Machine parameters. Using this method, the luminosity can be calculated directly from the beam parameters. In LHC, protons collide with each other in the form of proton bunches. There are about 1.5×10^{11} protons in each bunch, the effective cross-sectional area at the collision point is about $16\mu m$, and the time interval between the collisions of both proton bunches is approximately $25\,ns$.

$$l = \frac{N_1 N_2}{4\pi t \sigma_{eff}^2} = \frac{f N_1 N_2}{4\pi \sigma_{eff}^2} \simeq 10^{34} cm^{-2} s^{-1} \tag{1.3}$$

It is clear that the luminosity depends on factors such as the revolution (and collision) frequency of the proton bunches (f), the number of protons per bunch (N) and the effective cross-sectional area (σ_{eff}) of the proton beam. In this method, the measurement of the effective cross-sectional area at the interaction point has a very large uncertainty (15%), as a result, this uncertainty affects the luminosity measurement.

BRIL project.[1] group is responsible for luminosity measurements at CMS. The group has installed a number of sub-detectors called "luminometers" to measure luminosity and monitor beam conditions. By having several luminometers, if a problem occurs in one of them, the information from other luminometers can be used, and how their behavior changes in different LHC beam conditions can be measured. These luminometers monitor and measure luminosity based on rate measurements of a variety of observables: the CMS silicon pixel detector, the Drift Tubes in the barrel (DT), the Forward Hadronic Calorimeter (HF), the Fast Beam Conditions Monitor (BCM1F), and the Pixel Luminosity Telescope (PLT). The PLT, BCM1F, and HF are characterized by a dedicated data acquisition system for luminosity measurements and thus can provide high statistics luminosity measurements. The silicon pixel detector and DT run at low occupancy and therefore are very linear with varying instant luminosity. In addition these detectors are very stable over time, however are not used in this thesis. The luminometers investigated in this study are PLT, HFOC (HF), and BCM1F.

[1] The Beam Radiation, Instrumentation, and Luminosity

A number of pp interactions occur each time the proton bunches collide. The result of this interaction appears in a specific way in each luminometer according to the characteristics of that luminometer. For example, in a silicon luminometer, if an electron particle is rejected as a result of the pp collision, its effect is seen as the lighting of several pixels, and the set of these pixels is called a "cluster". In these luminometers, luminosity is obtained as:

$$l = \frac{f < N_{observables} >}{\sigma_{visible}} \tag{1.4}$$

which $\sigma_{visible}$ is abbreviated as σ_{vis} and is called "visible cross-section" or "calibration coefficient". The numerator is actually the average observable value in the luminometer, which we denote by μ_{vis}, so the Eq. 1.4 is given by:

$$l = \frac{f\mu_{vis}}{\sigma_{vis}}. \tag{1.5}$$

1.3 Zero-counting algorithm

In the BRIL project, luminometers measure a quantity linearly related to luminosity and therefore to the number of primary interactions. However, luminometers often cannot distinguish between a single hit and multiple hits in the same sensor, prompting the use of the so-called "zero counting" method. Since the number of hits should follow a Poisson distribution, the probability of having k hits is

$$< f_0 >= \sum_{k=0}^{\infty} \frac{\mu^k e^{-\mu}}{k!} p^k = e^{-\mu(1-p)} \tag{1.6}$$

To get the luminosity, we must divide this average value, which is actually μ_{vis}, by σ_{vis} (Eq. 1.5). In the following section, the exact method of obtaining σ_{vis} will be explained.

1.4 Van der Meer method

Luminosity in terms of the probability density distribution of protons for each bunch is given by:

$$l_b = f_r n_1 n_2 \int\int \rho_1(x, y)\rho_2(x, y)dxdy \tag{1.7}$$

where n is the number of particles in each proton bunch and f is the revolution frequency of protons and ρ is the density of particles inside each proton bunch. The total luminosity is obtained from the sum of the total luminosity of the proton bunches [67].

For an ideally linear luminometer, the visible cross-section is independent of the beam characteristics. Therefore, it needs to be determined once every time the detector is installed and setup in a certain way. Once a year, this is performed using what are known as Van der Meer scans, which were discovered by Simon van der Meer in 1968 [67]. The method consists of moving or "scanning" the collision plane (x and y) the two beams across each other in discrete steps of a certain length, as illustrated in Figure 1.1 measurements are averaged over each step. Therefore, in this method

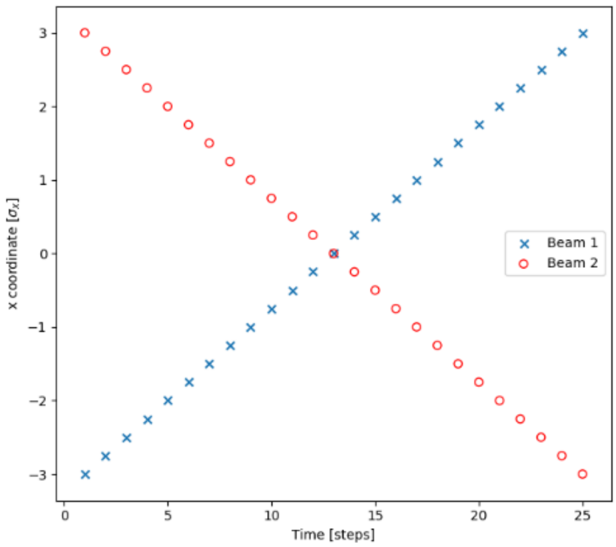


Figure 1.1: Visualization of the movement of the two beams during a Van der Meer scan in the x direction.

Eq. 1.7 becomes:

$$l_b = f_r n_1 n_2 \int \int \rho_{1x}(x, x + \Delta x)\rho_{2x}(x)dx \int \rho_{1y}\rho_{2y}dy, \tag{1.8}$$

Defining the parameter $\sum_x$ as

$$\sum_x = \frac{1}{\sqrt{2\pi}} \frac{\int \rho_{1x}(x+\Delta x)\rho_2(x)dxd(\Delta x)}{\int \rho_{1x}\rho_{2x}dx}, \tag{1.9}$$

and similarly for $\sum_y$, and puting Eq. 1.9 into Eq. 1.8 and equals to Eq. 1.5 the visible cross-section can be calculated as

$$\sigma_{vis} = \mu_{vis} \frac{2\pi n \sum_x \sum_y}{n_1 n_2}. \tag{1.10}$$

σ_{vis} depends on the number of particles of each bunch (n) and $\sum_x$ and $\sum_y$ and is independent of the type of proton density function. If the proton density function is a Gaussian function, then $\sum_x$ and $\sum_y$ are the widths of the Gaussian function.

1.5 Corrections and uncertainties

The previous section makes many assumptions and does not consider all possible effects. After the initial calibration outlined there, a multitude of studies are carried out to account for any possible discrepancies and each one of them yields a correction and/or an additional systematic uncertainty to the measurement.

1.5.1 Bunch current normalization

The bunch current measurement takes into account two possible corrections. The first is from the normalization of the Fast Beam Current Transformers (FBCT) per bunch beam current measurements (whose relative uncertainty is negligible) to the DC current transformers (DCCT) that measure all the charge in the LHC accurately [68].

The second is a correction for "satellite" and "ghost" charges. The former are charges that are in the colliding bunch slot, but not in the correct RF bucket, while the latter are outside any filled bunch slot. They are measured by the DCCT, but they do not contribute to the measured luminosity.

1.5.2 Background estimation

In the luminosity measurement, the backgrounds to be evaluated can originate from three different sources; namely detector noise, beam induced background (BIB) and afterglow. In general the various backgrounds can be present in the same BCID. If the collision of any two empty bunches yields in a non-zero detector rate, this signifies the presence of detector noise given that there is no radiation from detector material. Recording a signal by a luminometer when an unpaired lead bunch passes IP without collision with the opposite bunch, the so-called unpaired bunches, indicate the background due to the BIB source. This is because the elementary charges hit against the residual gas within the empty bunch or interact with the apertures. This source of background is also required to be identified and subtracted from the detector rates. The background due to the afterglow happens when the surrounding material of the detector radiates and the effect is passed to the next bunch crossings. Since the estimation of background correction is luminometer-dependent, the individual study of backgrounds observed in each of three main luminometers is provided in the following subsections.

During a vdM scan data-taking, the LHC is filled with a specific bunch pattern designed to optimize the precision of the luminosity calibration. The pattern for lead bunches is such that generally here are three or more empty bunches between any two consecutive filled slots. This is important in determining the main source of background per detector. Colliding bunches are grouped in so-called trains of bunches and therefore, the empty bunches are classified as either "in trains" or "out of trains".

1.5.3 Orbit drift

Although the LHC beam orbits are generally stable during a fill, even a small variation in the beam positions during a vdM scan can significantly affect the results of the vdM calibration.

1.5.4 Length scale calibration

The length scale (LS) calibration study takes into account the differences in actual beam separation, which is determined by the CMS tracker detector, from the nominal separation set by the LHC magnets. Some dedicated LS scans are performed to estimate the correction factors. During a LS calibration scan, the two incident beams are moved together in one direction while their separation is kept constant. In either of x or y directions, the scans start in a forward orientation and at some point the positions of beams are reversed, still with the same distance as before, and the scans are then taken in a backward direction. The configuration of beam positions for a typical LS scan is depicted in figure 1.2.

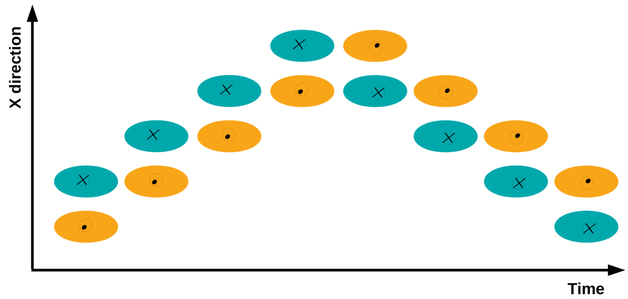


Figure 1.2: The movement of beams in a typical length scale scan. The inward and outward bunches are shown in blue and orange, respectively. As time goes on, the beams are moved together in constant separations. The first four steps contribute to the forward x-scans while the last four steps constitute the backward x-scans. Similar scans are performed in the y direction.

At each scan step, where the beams are separated in a specific distance, the scan is taken for a duration of δt which is usually almost equal to 3 lumi sections (one lumi section 22 seconds). Normally each scan step contains 1 entire lumi section and the other 2 lumi sections overlap with either the start or the end of the scan step. There are two different methods to reconstruct the vertices during each scan step; mainly based on either "timestamp (ts)" or "lumi section (ls)" selection. The ts method uses data between 2 timestamps for each scan step. The timestamps are chosen such that the beams are not moved between them. The ls selection method uses data of an entire lumi section which should be fully contained in a scan step (i.e., the beams do not move during this ls). In principle, the ts method allows data samples of higher statistics being used. Both methods should give consistent results within the errors. The spatial distribution of primary-interaction vertices is fitted using a Gaussian function and the mean value is taken as the beamspot position. To extract the LS calibration coefficient in each direction, the average of the nominal beam positions is compared with the beamspot position measured by the CMS tracker. The slope of the measured versus nominal position, which is fitted with a pol1 function, corresponds to the LS calibration constant.

1.5.5 The x-y correlations

In this correction, we attempt to account for any non-factorizability of beam density distributions in the x-y plane. Two "imaging" scans are performed in each direction, where one beam is scanned across and another remains stationary. In this way, the particle density distribution of the stationary beam can be measured. Data is fitted with a more complex model that includes a third, negative Gaussian function. Multiple Monte Carlo simulations of a VdM scan are conducted using the measured beam densities to derive the correction.

1.6 Luminosity measurement in Pb-Pb collision in 2015 for Fill 4689

For the pp collision in 2015 with a center-of-mass energy of 5.02 TeV using the vdM method, a total uncertainty of 1.1% on σ_{vis} has been measured, also for the data In 2015 and 2016, the total uncertainties in the center-of-mass energy of 13 TeV were measured to be 1.3% and 1.1%, respectively. In addition to pp collision in LHC, heavy ions have also collided in shorter time intervals than pp. For example, in the last two months of 2015, Pb-Pb ions collided at the center-of-mass energy of 5.02 TeV, and then on the third of December, a vdM scan has been performed for Fill 4689. Moving (scanning) the beams in two opposite directions along x and y are done once a year, in 25 steps and for 30 seconds, during the scans, the position of the beams and the values of μ_{vis} is recorded for each scan along x and y. This information is stored in hd5 files. The average value of σ_{vis} for each luminometer is obtained using the vdM framework package. In the following section, one of the systematic uncertainty sources on σ_{vis}, orbit drift, is estimate.d

1.6.1 Orbit drift correction for Fill 4689

In the LHC, beam positions are measured with BPM (Beam Position Measurement) pickup electrodes. Two different readout systems for these BPMs exist. Close to the IP (in the Q1 magnet) a DOROS readout system has been installed during LS1 whereas the BPMs in the ARCs use the readout system available from the beginning of Run-I. The DOROS readout electronics are supposed to deliver data with superior precision compared to the arcBPMs. Both systems are not able to resolve single bunches but in some way deliver an average beam position for all bunches in the beam. The horizontal and vertical coordinates are independently measured for each beam. We consider beam position measurements nominally taken at the nominal head-on position of the beam, before and after each scan. Using the beam position data during these two steps, a slight orbit drift of the beams away from their nominal position is usually observed. The beam positions set by the LHC is shown in Fig. 1.3.

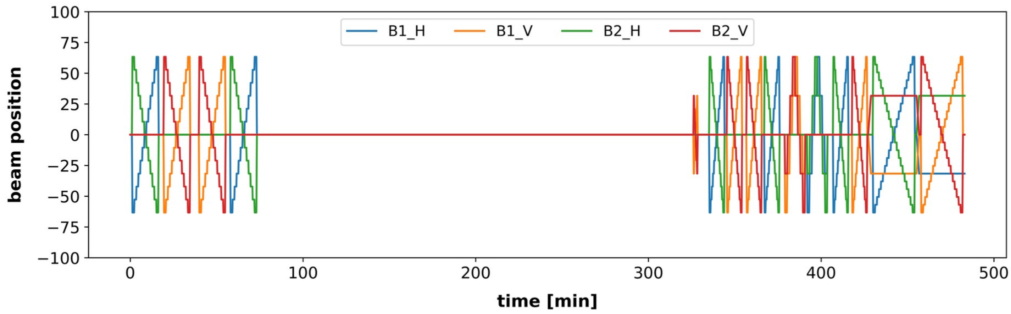


Figure 1.3: The nominal beam positions as designed by LHC for scan program of ions in 2015. The first four individual vdM scans (from 0 to about 80 min) were not properly recorded and then not processed in this note.

1.6.1.1 The DOROS data for fill 4689

The measured beam positions by DOROS during the vdM scans are shown in Fig. 1.4. The deviation from the nominal position can be deduced by comparing this figure with Fig. 1.3. According to Fig. 1.4, there is a large drift of B2 in the horizontal direction during the entire fill, especially within the time of vdM3 scan. The gray vertical lines indicate individual scans. The beam1 (B1) horizontal and vertical positions are shown in blue and orange and the B2 horizontal and vertical positions are shown in green and red, respectively. It is understood that the vertical and also horizontal direction for B1 has changed very little during vdM1 and vdM2 but the B2 horizontal position has changed considerably during vdM3.

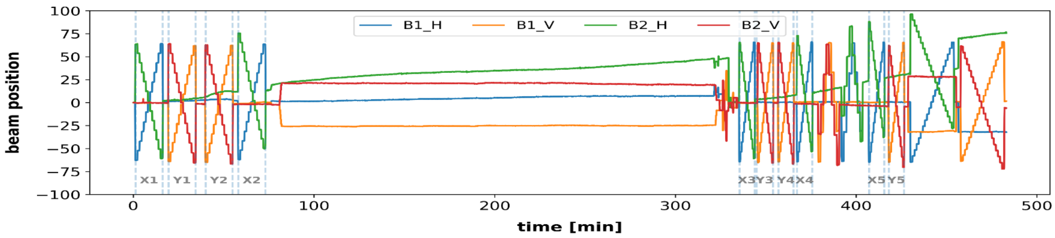


Figure 1.4: The beam positions as measured by the DOROS BPMs as a function of time during the vdM program of fill 4689.

Figure 1.5 shows the difference between the measured orbit drifts of the two beams, from which a large orbit drift as measured by DOROS is indicated by the non-zero values observed. For fill 4689, the DOROS results show a large orbit drift of beam2

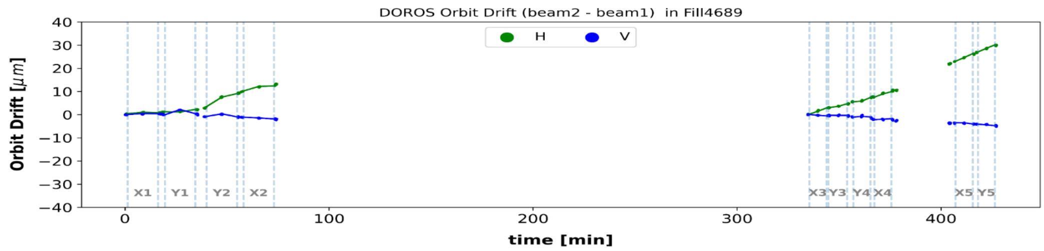


Figure 1.5: The difference between the measured orbit drifts of the two beams measured by DOROS as a function of time during the VdM program of fill 4689.

(B2) in the horizontal direction. Besides, the sign of data from one side of the IP must be reversed to get meaningful results, in addition there are small oscillations of the beam. Lumi group in CMS have made a discussion with LHC in regard to these problems. In 2015 and 2016 the DOROS system was still in the commissioning phase. Having discussed with the LHC experts, it was concluded that the liability of the DOROS data in this period is not completely known. However DOROS data will be used in the following and will be compared to the orbit drifts obtained with the arcBPMs.

This effect is also obvious from the plots in Fig. 1.6 representing the raw data as measured by DOROS and implying that the source of huge horizontal drift comes from one side only (mostly the DOROS on the left side drifts) whereas the one on the right side remains almost constant. Therefore the drift we observe in the DOROS H B2 is instrumental, a problem with that particular DOROS electronics, and hence the DOROS data can not be fully trusted. The impact on σ_{vis} from the orbit drift correction as measured by DOROS is investigated and summarized in Table 1.1. The individual information for different luminometers and vdM scans are presented. The large impact on σ_{vis} observed during vdM3 is visible because the DOROS measurements within that time period is untrustable.

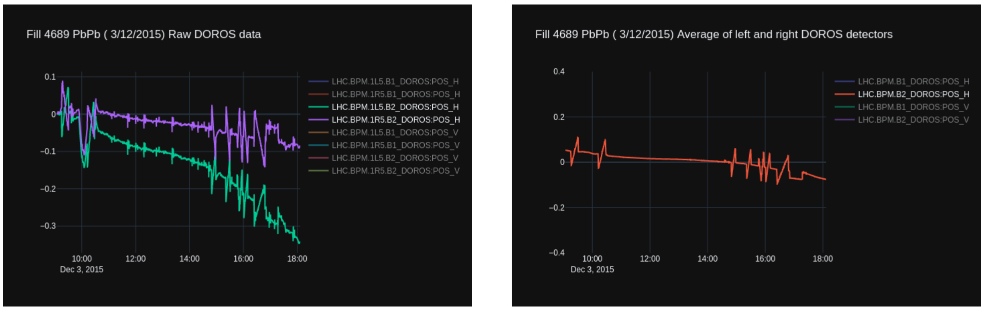


Figure 1.6: Fill 4689 PbPb Left: Raw left and right DOROS data for B2 H and Right: Average of left and right DOROS detectors.

	Fit	**Detector**	**Effect%**		
			vdM1	**vdM2**	**vdM3**
σ_{vis}	**SG**	BCM1F	1.28	1.51	26.86
		PLT	1.28	1.52	27.48
		HFOC	1.42	1.50	26.52

Table 1.1: The impact of the orbit drift correction on values of σ_{vis} using DOROS for all luminometers per scan.

The average correction of 1.3%(1.5%) for σ_{vis} is derived during vdM1(vdM2), while the effect of orbit drift correction as measured by DOROS is too high during vdM3 and reaches a value of about 27%. Since the σ_{vis} is a parameter of the detector and is independent of vdM scans, one can conclude that the DOROS results are not reliable. In the next section, an overview of the arcBPM data system is presented.

1.6.1.2 The arcBPM data for fill 4689

The arcBPM is also used to measure beam position during vdM scans. Figure 1.7 shows orbit drift versus time as measured by arcBPMs. Compared to the similar plot shown in Fig. 1.5, the massive horizontal drift during vdM scans is now disappeared.

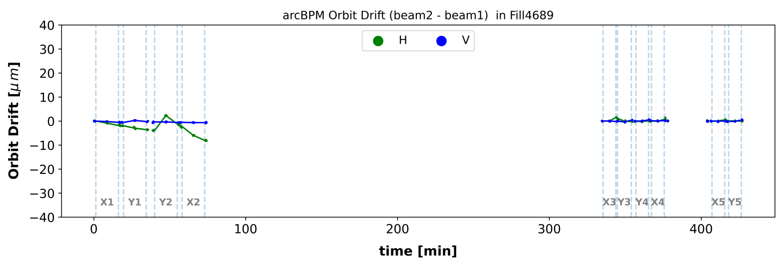


Figure 1.7: Orbit drift measured by arcBPM as a function of time during the VdM program of fill 4689.

The effect of orbit drift correction using arcBPM data is shown in Fig. 1.8. In these plots the measured σ_{vis} values per scan for each detector are presented, when the SG function is used as a fit function. Table 1.2 summarizes the effect of orbit drift correction on σ_{vis} using arcBPM data per scan. Clearly, the impact of the orbit drift correction on σ_{vis} is more reliable than those for DOROS data.

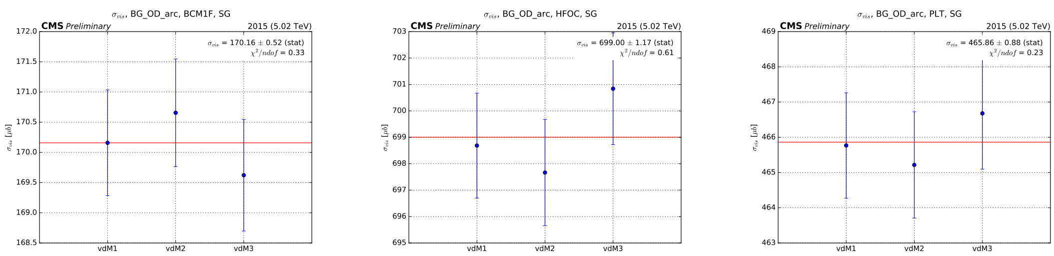


Figure 1.8: Shown are the average values of σ_{vis} for the three luminometers BCM1F, HFOC, and PLT.

	Fit	Detector	Effect%		
			vdM1	vdM2	vdM3
σ_{vis}	**SG**	BCM1F	0.76	0.47	0.43
		PLT	0.76	0.46	0.43
		HFOC	0.76	0.46	0.43

Table 1.2: The impact of the orbit drift correction on values of σ_{vis} using arcBPM for all luminometers per scan.

The average correction of 0.8%, 0.5% and 0.4% for σ_{vis} is derived during vdM1, vdM2 and vdM3 scans, respectively. On average a correction of about 0.6% on σ_{vis} is obtained when SG fit function is used. Since the only available and reliable beam position measurement data are provided by arcBPM, then the entire correction factor, 0.6%, is considered as the systematic uncertainty on the σ_{vis}.

1.6.2 Total luminosity correction and uncertainty

The total value of systematic uncertainties on σ_{vis} is obtained to be about 2.3, and the luminosity for the data collected during the Pb-Pb collision in 2015 is measured to be 519.50 μb^{-1} (506.33 μb^{-1}) for HF (PLT).

1.7 Chapter summary

Luminosity measurment

This part of the project involves the following steps: first, the software package is executed for hd5 files containing information about vdM scans, and then σ_{vis} is derived without taking into account corrections. As a next step, the corrections are applied to σ_{vis}, and then the uncertainty of these corrections is calculated. As a last step, the total uncertainty on the σ_{vis} is calculated, which actually indicates the accuracy of the luminosity measurement for the Pb-Pb collision of 2015 and at the center-of-mass energy of 5.02 TeV.

CONCLUSION

The first part of this thesis present the search for exotic decays of a Higgs boson with $m_H = 125\,GeV$ which is performed in the $\mu\mu bb$ final state with the CMS detector using LHC proton-proton collision data at 13 TeV center-of-mass energy corresponds to an integrated luminosity of $138\,\mathrm{fb}^{-1}$. The decay is assumed to carry out through a pair of new pseudoscalars, $a_1 a_1 \rightarrow ff$, that interact with SM fermions (f) via mixing with the Higgs boson. The mass of the pseudoscalar is considered to be $15 \leq m_{a_1} \leq 62.5\,GeV$. The expected topology of the signal, the mass requirements on the combination of the decay products, is exploited to enhance the sensitivity. The detector resolution effects especially on b-jets are taken into account. Introducing a new basis for the kinematics to decorrelate the di-b-jet and $\mu\mu bb$ systems results in a major improvement with respect to the earlier CMS study based on partial data sample at 13 TeV [15]. For the signal hypotheses with low mass, limits are improved with the addition of a new category, $LowP_T$. In the absence of any significant excess over the SM backgrounds, upper limits $\mathcal{B}(H \rightarrow a_1 a_1 \rightarrow \mu\mu bb) < (0.17 - 3.3) \times 10^{-4}$ are obtained, depending on the m_{a_1} hypothesis.

The latest results combining the two different final states, $\mu\mu bb$ and $\tau\tau bb$, bring additional sensitivity to the search for exotic decays of the Higgs to light pseudoscalars. Combining all final states in $\tau\tau bb$, limits are observed to be in range (1.8–7.7)% for m_{a_1} between 12 and 60 GeV. Integrating over the decays of a_1, and in the context of different parameterization of 2HDM+S, the branching fraction $\mathcal{B}(H \rightarrow a_1 a_1)$ is combined between the two analyses. For m_{a_1} values between 15 and 60 GeV, $\mathcal{B}(H \rightarrow a_1 a_1)$ above 23% are excluded with 95% CL, in most of the Type-II scenarios. In Type-III and IV, upper limits as low as about 2% and 3% are obtained, respectively, for $\tan\beta = 2.0$ and 0.5. The analysis has lower sensitivity for smaller m_{a_1} values, as it cannot resolve boosted daughter particles with the reconstruction methods available. However, improved analysis strategies provide significant gains over previous analyses, in addition to that from the increased integrated luminosity. The second part of this thesis preset the luminosity measurement for the Pb-Pb collision of 2015 and at the center-of-mass energy of 5.02 TeV.

Outlook. As the integrated luminosity increases during the High-Luminosity LHC period, upper limits are expected to be reduced. The search has focused on the mass range between 15 and $62.5\,GeV$, but it can be extended to masses less than $15\,GeV$ in the future.

APPENDIX

A

STUDY OF B-TAGGING BELOW 20 GEV

In this analysis, the minimum b-jet P_T cut is 15 GeV, so the b-tagging in the p_T range between 15 and 20 GeV with scale factors at 20 GeV and doubled uncertainties are used. Under 20 GeV, no significant discrepancy was observed. Fig A.1 shows the sub-leading jet p_T distribution and it can be seen that signal for all masses has a peak below 20 GeV, Background shows a similar trend. It should be emphasized that the uncertainty bands are only MC statistics.

The main concern is, deep jet b-tagging SF's are not available below 20 GeV. Doubling systematic uncertainties may not cover the possible issues. Therefore it motivates the study of events where leading b-jet has p_T below 20 GeV. For this study we use loose-loose b-working point. The questions are:

- Is there any major mis-modeling of b discriminator in our phase space?
- Are current b-tag systematic sufficient?

Events with leading b-jet $p_T < 20\,GeV$ are selected and then samples divided into prompt and non-prompt b-jet (light/gluon jets). From Fig. A.2, it is clear that the statistics from leading b-jet is less than sub-leading b-jet, so leading b-jets are mostly above 20 GeV and not many leading jets are with p_T between 15 and 20 GeV. It can be seen that there is no major mis-modeling within the uncertainties of data and MC statistics.

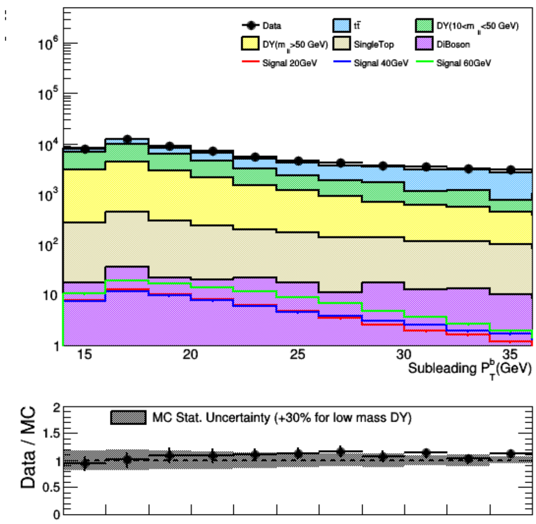


Figure A.1: The sub-leading jet p_T distribution.There is no major discrepancy below 20 GeV. Simulated samples are normalized to $138\,\mathrm{fb}^{-1}$. The band in the lower panel of each plot represents the MC statistical uncertainty.

In Fig. A.3 all different source of systematic uncertainties of bSF's are shown where $p_T < 20\,GeV$ and uncertainties are doubled. The total of them is shown with the green line. Obviously, these systematic uncertainties cover the residuals. It looks safe for $p_T < 20\,GeV$ to apply bSF's at 20 GeV and doubled the uncertainties. Since events are later categorized based on the b-tag score, shape scale factors should be used. It would then need to ensure that shape-based uncertainties cover any mis-modeling that may have taken place.

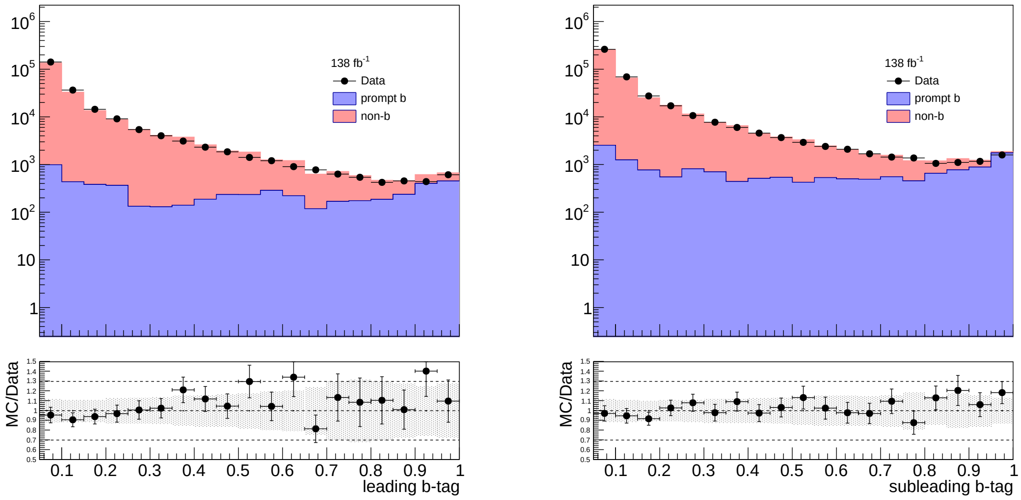


Figure A.2: The distribution of b-tag discriminator for leading b-jet (left) and sub-leading b-jet(right).

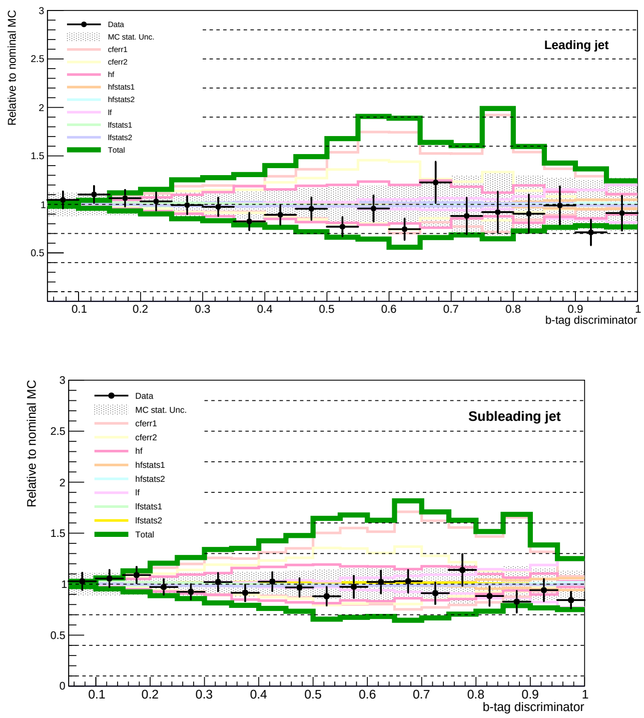


Figure A.3: Relative to nominal MC versus b-tag discriminator for leading b-jet (left) and sub-leading b-jet (right).

APPENDIX

B

EXPECTED RESULTS IN CONTROL REGION

The results here are based on the data in the control region. It can be seen that limits are reliable in comparison with 2016 analysis [15]. This gives the go-ahead to estimate background using the data in the SR. As demonstrated in Fig. B.1, current limits are more stringent than those of the CMS analysis using 2016 data. This goes beyond the increase of the luminosity, thanks to optimized usage of signal properties over the mass range.

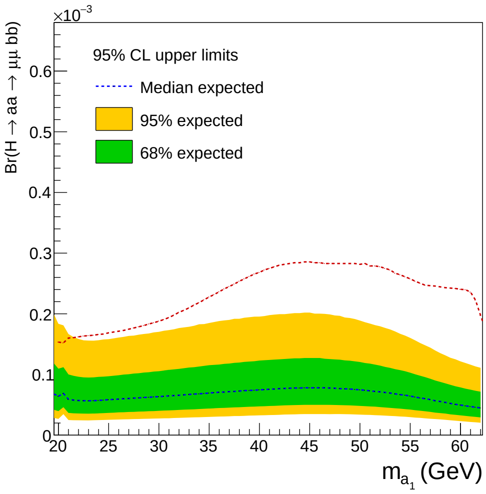


Figure B.1: Comparison of the power of the current analysis with that of 2016 data in terms of expected limits, at 95% CL, on $H \to a_1 a_1 \to \mu\mu bb$.

APPENDIX

C

BIAS STUDY

Figure 5.1 for $LowP_T$ and TL categories show a wide background peak of about 10 GeV, while we are looking for a very narrow peak of about 2 GeV. In CR (Chapter 4), we observed the same behavior (Fig. 4.4), but the statistics are not enough to study it in further detail.

Therefore, to test the bias, the best-fit that describes the background from data in SR is used and fit it to a toy model with an expected signal of 0.1 and 1. Figure C.1 illustrate the r-value as a function of dimuon invariant mass ($> 40\,GeV$) in $LowP_T$ category for a toy model with an injected signal. R-values were changed by about 10^{-4} or less in comparison with input values. The same study is performed for TL category (Fig. C.2). No bias is observed in both categories.

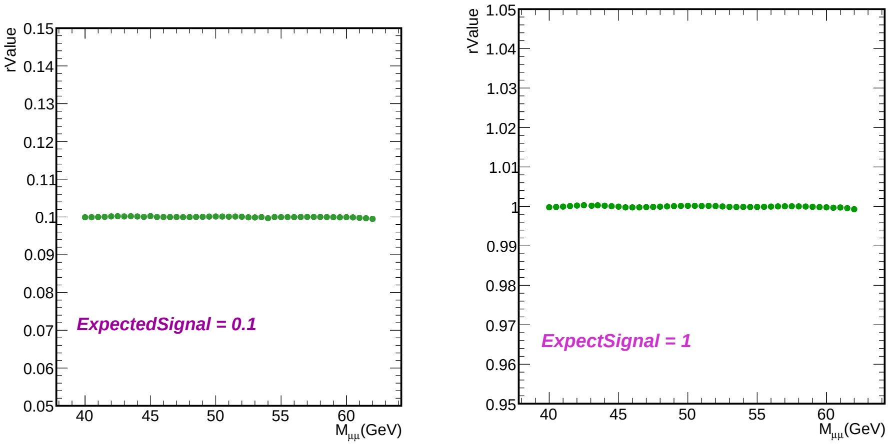


Figure C.1: The r-value as a function of dimuon invariant mass ($> 40\,GeV$) in $LowP_T$ category for a toy model with an injected signal 0.1 (left) and 1 (right).

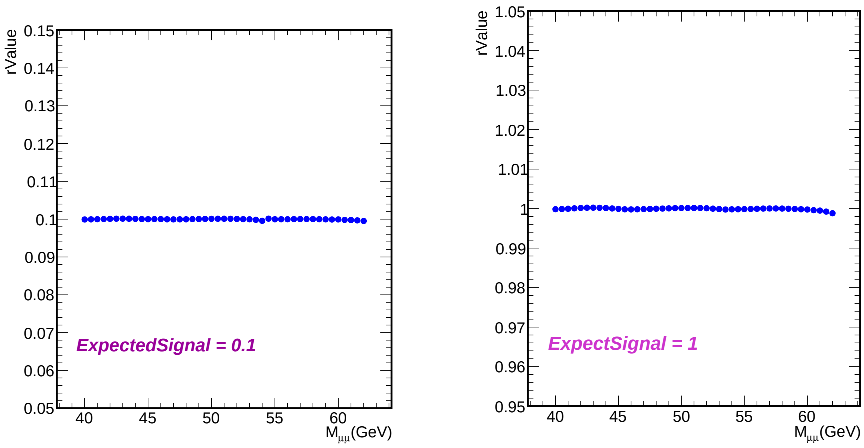


Figure C.2: The r-value as a function of dimuon invariant mass ($> 40\,GeV$) in TL category for a toy model with an injected signal 0.1 (left) and 1 (right).

BIBLIOGRAPHY

[1] G. Aad et al. Observation of a new particle in the search for the standard model higgs boson with the atlas detector at the lhc. *Physics Letters B*, 716, 2012.

[2] S. Chatrchyan et al. Observation of a new boson at a mass of 125 gev with the cms experiment at the lhc. *Physics Letters B*, 716, 2012.

[3] Georges Aad et al. A detailed map of Higgs boson interactions by the ATLAS experiment ten years after the discovery. *Nature*, 607:52, 2022. [Erratum: Nature 612, E24 (2022)].

[4] Armen Tumasyan et al. A portrait of the Higgs boson by the CMS experiment ten years after the discovery. *Nature*, 607:60, 2022.

[5] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, Marc Sher, and Joao P. Silva. Theory and phenomenology of two-Higgs-doublet models. *Phys. Rept.*, 516:1, 2012.

[6] Abdelhak Djouadi. The Anatomy of electro-weak symmetry breaking. II. The Higgs bosons in the minimal supersymmetric model. *Phys. Rept.*, 459:1, 2008.

[7] Tania Robens and Tim Stefaniak. Status of the Higgs Singlet Extension of the Standard Model after LHC Run 1. *Eur. Phys. J. C*, 75:104, 2015.

[8] Tania Robens, Tim Stefaniak, and Jonas Wittbrodt. Two-real-scalar-singlet extension of the SM: LHC phenomenology and benchmark scenarios. *Eur. Phys. J. C*, 80:151, 2020.

[9] B. Grzadkowski and P. Osland. Tempered Two-Higgs-Doublet Model. *Phys. Rev. D*, 82:125026, 2010.

[10] Aleksandra Drozd, Bohdan Grzadkowski, John F. Gunion, and Yun Jiang. Extending two-Higgs-doublet models by a singlet scalar field - the Case for Dark Matter. *JHEP*, 11:105, 2014.

[11] SM Higgs production cross sections at $\sqrt{s} = 13$ TeV (update in CERN Report4 2016) . Technical report, CERN.

[12] David Curtin et al. Exotic decays of the 125 gev higgs boson. *Phys. Rev. D*, 90, Oct 2014.

[13] Search for Higgs boson decays into two spin-0 particles in the $bb\mu\mu$ final state with the ATLAS detector in pp collisions at $\sqrt{s} = 13$ TeV. 3 2021.

[14] ATLAS Collaboration. Search for Higgs boson decays into a pair of light bosons in the $bb\mu\mu$ final state in pp collision at $\sqrt{s} = 13$ TeV with the ATLAS detector. *Phys. Lett. B*, 790:1, 2019.

[15] A.M. Sirunyan et al. Search for an exotic decay of the higgs boson to a pair of light pseudoscalars in the final state with two muons and two b quarks in pp collisions at 13 tev. *Phys. Lett. B*, 795, 2019.

[16] Vardan Khachatryan et al. Search for light bosons in decays of the 125 GeV Higgs boson in proton-proton collisions at $\sqrt{s} = 8$ TeV. 2017.

[17] LHC Machine. *JINST*, 3:S08001, 2008.

[18] E. Boltezar, H. Haseroth, W. Pirkl, G. Plass, T. R. Sherwood, U. Tallgren, P. Tetu, D. Warner, and M. Weiss. Performance of the new cern 50 mev linac. *IEEE Transactions on Nuclear Science*, 26(3):3674, 1979.

[19] N A Tahir, R Schmidt, M Brugger, R Assmann, A V Shutov, I V Lomonosov, A R Piriz, D H H Hoffmann, C Deutsch, and V E Fortov. The cern super proton synchrotron as a tool to study high energy density physics. *New Journal of Physics*, 10(7):073028, jul 2008.

[20] Daniel Boussard and Trevor Paul R Linnecar. The LHC Superconducting RF System. Technical report, CERN, Geneva, 1999.

[21] S. Chatrchyan et al. The CMS Experiment at the CERN LHC. *JINST*, 3:S08004, 2008.

[22] Yogiro Hama. A Note on Lorentz Transformation and Pseudorapidity Distributions. *J. Phys. Soc. Jap.*, 50:21, 1981.

[23] S Chatrchyan, G Hmayakyan, Khachatryan, et al. The CMS experiment at the CERN LHC. The Compact Muon Solenoid experiment. *JINST*, 3:S08004, 2008. Also published by CERN Geneva in 2010.

[24] The Tracker Group of the CMS Collaboration. The cms phase-1 pixel detector upgrade, 2020.

[25] S. Chatrchyan, V. Khachatryan, A.M. Sirunyan, et al. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Physics Letters B*, 716(1):30, sep 2012.

[26] CMS Collaboration. Performance of the cms hadron calorimeter with cosmic ray muons and lhc beam data. *Journal of Instrumentation*, 5(03):3012, mar 2010.

[27] The CMS collaboration. The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC. *Journal of Instrumentation*, 8(11):11002, nov 2013.

[28] G L Bayatyan, N Grigorian, V G Khachatrian, et al. *CMS TriDAS project: Technical Design Report, Volume 1: The Trigger Systems*. Technical design report. CMS.

[29] Albert M Sirunyan et al. Performance of the CMS Level-1 trigger in proton-proton collisions at $\sqrt{s} = 13\,\mathrm{TeV}$. *JINST*, 15:P10017, 2020.

[30] Vardan Khachatryan et al. The CMS trigger system. *JINST*, 12:P01020, 2017.

[31] F.J. Yndurin. Relativistic quantum mechanics and introduction to field theory. *Physics reports*, 1st ed.(ISBN 978-3-540-60453-2), 1996.

[32] The ATLAS collaboration and CMS Collaboration. Combined measurement of the higgs boson mass in pp collisions at s =7 and 8 tev with the atlas and cms experiments. *Physical Review Letters*, 114(19), May 2015.

[33] T. W. B. Kibble. Symmetry breaking in nonAbelian gauge theories. *Phys. Rev.*, 155:1554, 1967.

[34] Peter W. Higgs. Broken Symmetries and the Masses of Gauge Bosons. *Phys. Rev. Lett.*, 13:508, 1964.

[35] Ernest Ma. Verifiable radiative seesaw mechanism of neutrino mass and dark matter. *Physical Review D*, 73(7), apr 2006.

[36] Greg W. Anderson and Lawrence J. Hall. The Electroweak phase transition and baryogenesis. *Phys. Rev. D*, 45:2685–2698, 1992.

[37] R. D. Peccei and Helen R. Quinn. CP Conservation in the Presence of Instantons. *Phys. Rev. Lett.*, 38:1440–1443, 1977.

[38] M. Drees, R. Godbole, and P. Roy. *Theory and phenomenology of sparticles: An account of four-dimensional N=1 supersymmetry in high energy physics.* 2004.

[39] John F. Gunion, Howard E. Haber, Gordon L. Kane, and Sally Dawson. *The Higgs Hunter's Guide*, volume 80. 2000.

[40] David Curtin, Rouven Essig, Stefania Gori, Prerit Jaiswal, Andrey Katz, Tao Liu, Zhen Liu, David McKeen, Jessie Shelton, Matthew Strassler, et al. Exotic decays of the 125 gev higgs boson. *Physical Review D*, 90(7):075004, 2014.

[41] J. Alwall, R. Frederix, S. Frixione, V. Hirschi, F. Maltoni, et al. The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations. *JHEP*, 1407, 2014.

[42] Michelangelo L Mangano et al. Matching matrix elements and shower evolution for top-pair production in hadronic collisions. *JHEP*, 2007, 2007.

[43] Rikkert Frederix and Stefano Frixione. Merging meets matching in MC@NLO. *JHEP*, 12:061, 2012.

[44] Emanuele Re. Single-top Wt-channel production matched with parton showers using the POWHEG method. *Eur. Phys. J.*, C71:1547, 2011.

[45] Torbjörn Sjöstrand, Stephen Mrenna, and Peter" Skands. A brief introduction to PYTHIA 8.1. *Comput.Phys.Commun.*, 178:852–867, 2008.

[46] Peter Skands, Stefano Carrazza, and Juan Rojo. Tuning PYTHIA 8.1: the Monash 2013 Tune. *Eur. Phys. J.*, C74(8):3024, 2014.

[47] S. Agostinelli et al. GEANT4: A simulation toolkit. *Nucl. Instrum. Meth. A*, 506:250, 2003.

[48] A. M. Sirunyan et al. Particle-flow reconstruction and global event description with the CMS detector. *JINST*, 12(10):P10003, 2017.

[49] D Contardo, M Klute, J Mans, L Silvestris, and J Butler. Technical Proposal for the Phase-II Upgrade of the CMS Detector. Technical report, Geneva, 2015. Upgrade Project Leader Deputies: Lucia Silvestris (INFN-Bari), Jeremy Mans (University of Minnesota) Additional contacts: Lucia.Silvestris@cern.ch, Jeremy.Mans@cern.ch.

[50] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. The anti-kt jet clustering algorithm. *JHEP*, 04:063, 2008.

[51] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. FastJet User Manual. *Eur. Phys. J. C*, 72:1896, 2012.

[52] Jet algorithms performance in 13 TeV data. Technical Report CMS-PAS-JME-16-003, CERN, Geneva, 2017.

[53] CMS Collaboration. Determination of jet energy calibration and transverse momentum resolution in CMS. *JINST*, 6:11002, 2011.

[54] Search for exotic decays of the Higgs boson to a pair of new light bosons in the $\mu\mu bb$ final state at $\sqrt{s} = 13$ TeV with the full Run 2 dataset. Technical report, CERN, Geneva, 2022.

[55] D. de Florian et al. Handbook of LHC Higgs cross sections: 4. Deciphering the nature of the Higgs sector. CERN Report CERN-2017-002-M, 2016.

[56] A. M. Sirunyan et al. Precision luminosity measurement in proton-proton collisions at $\sqrt{s} = 13\,TeV$ in 2015 and 2016 at CMS. Submitted to *Eur. Phys. J. C*, 2021.

[57] CMS Collaboration. CMS luminosity measurement for the 2017 data-taking period at $\sqrt{s} = 13\,TeV$. CMS Physics Analysis Summary CMS-PAS-LUM-17-004, 2018.

[58] CMS Collaboration. CMS luminosity measurement for the 2018 data-taking period at $\sqrt{s} = 13\,TeV$. CMS Physics Analysis Summary CMS-PAS-LUM-18-002, 2019.

[59] Albert M Sirunyan et al. Measurement of the inelastic proton-proton cross section at $\sqrt{s} = 13$ TeV. *JHEP*, 07:161, 2018.

[60] Measurements of inclusive w and z cross sections in pp collisions at $\sqrt{s} = 7$ tev. *JHEP*, 01, 2011.

[61] A. M. Sirunyan et al. Identification of heavy-flavour jets with the CMS detector in pp collisions at 13 TeV. *JINST*, 13:P05011, 2018.

[62] Jon Butterworth et al. PDF4LHC recommendations for LHC Run II. *J. Phys. G*, 43:023001, 2016.

[63] P.D. Dauncey, M. Kenzie, N. Wardle, and G.J. Davies. Handling uncertainties in background shapes: the discrete profiling method. 2014.

[64] Search for exotic Higgs boson decays to a pair of pseudoscalars in the $\mu\mu$bb and $\tau\tau$bb final states in proton-proton collisions with the CMS experiment. Technical report, CERN, Geneva, 2023.

[65] "CMS Collaboration". Search for an exotic decay of the Higgs boson to a pair of light pseudoscalars in the final state with two b quarks and two τ leptons in proton-proton collisions at $\sqrt{s} = 13$ TeV. *Physics Letters B*, 785:462, 2018.

[66] A. M. et al. Sirunyan. Combined measurements of Higgs boson couplings in proton-proton collisions at $\sqrt{s} = 13$ TeV. *The European Physical Journal C*, 79(5), May 2019.

[67] S van der Meer. Calibration of the effective beam height in the ISR. Technical report, CERN, Geneva, 1968.

[68] G Anders, Bacchetta, et al. LHC Bunch Current Normalisation for the April-May 2010 Luminosity Calibration Measurements. 2011.