國 立 中 央 大 學

物 理 學 系

碩 士 論 文

Search for heavy resonances decaying into a Z boson and a Higgs boson in the 2l2b final state in pp collisions at $\sqrt{s} = 13$ TeV

研 究 生：童宇軒

指導教授：余欣珊

中 華 民 國 一 〇 五 年 十 一 月

105
國立中央大學
物理學系
碩士論文
Search for heavy resonances decaying into a Z boson and a Higgs boson in the 2l2b final state in pp collisions at $\sqrt{s} = 13$ TeV
童宇軒 撰

國立中央大學圖書館
碩博士論文電子檔授權書

(104 年 5 月最新修正版)

本授權書授權本人撰寫之碩/博士學位論文全文電子檔(不包含紙本、詳備註 1 說明)，在「國立中央大學圖書館博碩士論文系統」。(以下請擇一勾選)

(✓)同意 (立即開放)

(　)同意 (請於西元 ______年_____月_____日開放)

(　)不同意，原因是：______________________________

在國家圖書館「臺灣博碩士論文知識加值系統」

(✓)同意 (立即開放)

(　)同意 (請於西元 ______年_____月_____日開放)

(　)不同意，原因是：______________________________

以非專屬、無償授權國立中央大學、台灣聯合大學系統圖書館與國家圖書館，基於推動「資源共享、互惠合作」之理念，於回饋社會與學術研究之目的，得不限地域、時間與次數，以紙本、微縮、光碟及其它各種方法將上列論文收錄、重製、與利用，並得將數位化之上列論文與論文電子檔以上載網路方式，提供讀者基於個人非營利性質之線上檢索、閱覽、下載或列印。

研究生簽名: ________________　　　　學號：103222037

論文名稱: Search for heavy resonances decaying into a Z boson and a Higgs boson in the 2l2b final state in pp collisions at $\sqrt{s} = 13$ TeV

指導教授姓名：余欣珊

系所：物理 所　☐博士班　☑碩士班

填單日期：2016/11/23

備註：

1. 本授權書之授權範圍僅限**電子檔**，紙本論文部分依著作權法第 15 條第 3 款之規定，採推定原則即預設同意圖書館得公開上架閱覽，如您有申請專利或投稿等考量，不同意紙本上架陳列，須另行加填申請書，詳細說明與紙本申請書下載請至本館數位博碩論文網頁。
2. 本授權書請填寫並**親筆**簽名後，裝訂於各紙本論文封面後之次頁（全文電子檔內之授權書簽名，可用電腦打字代替）。
3. 讀者基於個人非營利性質之線上檢索、閱覽、下載或列印上列論文，應遵守著作權法規定。

Search for heavy resonances decaying into a Z boson and a Higgs boson in the $2l2b$ final state in pp collisions at $\sqrt{s}$ = 13 TeV

by

Yee Shian Henry Tong

Submitted to the Department of Physics
in partial fulfillment of the requirements for the degree of

Master of Physics

at the

NATIONAL CENTRAL UNIVERSITY

November 2016

Author . . . .

Department of Physics
November 29, 2016

Certified by . . . .

Shin-Shan Eiko Yu
Associate Professor
Thesis Supervisor

Accepted by . . . .

Yuan-Hann Chang
Professor
Chairman, Thesis Committee

Search for heavy resonances decaying into a Z boson and a Higgs boson in the $2l2b$ final state in pp collisions at $\sqrt{s}$ = 13 TeV

by

Yee Shian Henry Tong

Submitted to the Department of Physics
on November 29, 2016, in partial fulfillment of the
requirements for the degree of
Master of Physics

Abstract

A search for heavy resonances decaying to a Higgs boson and a Z boson is presented. The analysis is based on the data collected in 2015 with the CMS detector at a center-of-mass energy $\sqrt{s}$ = 13 TeV, corresponding to an integrated luminosity of 2.51 fb^{-1}. The Higgs bosons are reconstructed from high momentum $b\bar{b}$ quark pairs that are detected as a single massive jet, while the Z bosons are reconstructed from electron pairs and muon pairs. The analysis is separated in electron and muon channels, with single and double b-tag categories. A 95% upper limit on the production cross section of $\sigma_X \times \mathcal{B}(X \rightarrow ZH)$ is derived from the combination of four categories with a limit of 0.063 pb to 0.265 pb for m_X from 800 to 4000 GeV.

Thesis Supervisor: Shin-Shan Eiko Yu
Title: Associate Professor

摘要

本篇論文呈現了由新理論模型預測之粒子衰變到一個希格斯粒子和一個 *Z* 玻色子的分析。本分析使用了於 2015 年由大強子對撞機中的緊湊渺子線圈偵測器所記錄之質子-質子對撞總能量為 13 TeV，總亮度為 2.5 fb^{-1} 的數據。高動量的希格斯粒子衰變到一個底夸克和一個反底夸克，在偵測器裡被偵測為一個大質量的噴流。*Z* 玻色子有兩個衰變通道，分別為正反電子通道以及正反渺子通道。本分析將分別探討電子通道和渺子通道，各通道將再細分為單底夸克標記和雙底夸克標記此二類別。通過合併電子通道和渺子通道，以及它們所有的底夸克標記類別，結果顯示質量由 800 GeV 至 4000 GeV 的新粒子於 95%信置區間的生產截面上限為 0.063 *pb* 至 0.265 *pb*。

Contents

List of Figures

List of Tables

Chapter 1

Introduction and Theory Overview

1.1 Introduction

This thesis presents the search of a heavy resonance decaying into a Z boson and a Higgs boson at center-of-mass energy of 13 TeV using 2.51 fb^{-1} proton-proton collision data collected with the CMS detector at the LHC. The Z boson further decays into two charged leptons (electrons or muons), while the Higgs boson decays into two b quarks. The Feymann diagram of the signal production is presented in Figure 1.1.

In this search, the high-momentum Higgs boson is reconstructed as a massive jet, and is identified by a b-tagging algorithm. The leptonic decay of Z is considered in order to discriminate against the large multijet background. The heavy resonance signal appears as an excess in the spectrum of the invariant mass of the jet and the two leptons. This analysis is a part of the search for heavy resonances decaying into one vector boson plus one Higgs boson (VH) [1].

The organization of this thesis is described as follows. In the next section, a brief overview of Heavy Vector Triplets Model described by a simplified phenomenological Lagrangian is presented. A specific explicit model is then introduced, which is the benchmark model in this analysis. In Chapter 2, an overview of the LHC and the CMS detector with its sub-detectors are presented. Chapter 3 reports the data sets and Monte Carlo samples used in this analysis. The reconstruction of physics objects and their selections are also described, and the

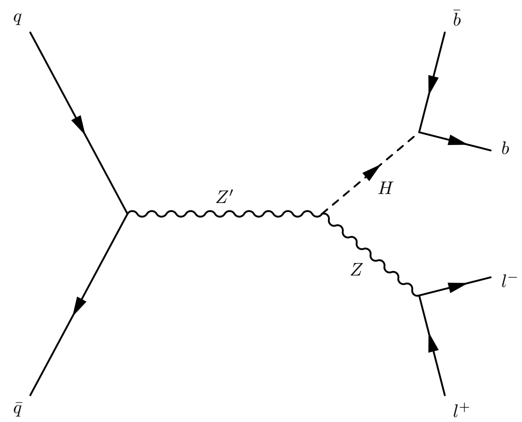

FIGURE 1.1: Feymann diagram of the production of the heavy vector Z', decaying into a Z boson and a Higgs boson. The Z boson further decays into two charged leptons, while the Higgs further decays into two b quarks.

agreement of data sets and Monte Carlo samples are presented. The estimation of backgrounds based on a data driven strategy is presented in Chapter 4. In Chapter 5, various systematic uncertainties are described. In Chapter 6, the results of this analysis are discussed, and a conclusion is summarized.

1.2 Theoretical Motivations

In the Standard Model (SM), the generation of masses for the weak gauge bosons ($W^{\pm}$ and Z) through electroweak symmetry breaking (EWSB) can be explained by the Higgs mechanism [2]. The mechanism was confirmed by the ATLAS and CMS experiments [3–5] at the CERN 50 years after the theory has been proposed.

However, the discovered Higgs boson with mass of 125 GeV is much lighter than the Planck energy, which suggests that the SM may be incomplete. Various theories postulate the existence of new heavy resonances that couple to the SM bosons in an attempt to solve the hierarchy problem or naturalness problem. Some common models include the Little Higgs models [6, 7] and strongly coupled Composite Higgs [8, 9].

Various models can be generalized in the Heavy Vector Triplet (HVT) framework [10], which is a simplified approach based on a phenomenological Lagrangian. In the Simplified Model, only the relevant couplings and mass parameters are retained. The reason for this is that resonant searches are typically not sensitive to all the free parameters of the specific model, but only to those parameters that are related to the resonance mass and the interactions involved in its decay and production.

1.2.1 Heavy Vector Triplet

Consider a heavy vector boson V_μ^a, a = 1,2,3, the simplified Lagrangian is described as

$$\begin{aligned}\mathcal{L}_V = &-\frac{1}{4}D_{[\mu}V_{\nu]}^a D^{[\mu}V^{\nu]a} + \frac{m_V^2}{2}V_\mu^a V^{\mu a} \\ &+ i g_V c_H V_\mu^a H^\dagger \tau^a \bar{D}^\mu H + \frac{g^2}{g_V} c_F V_\mu^a \sum_f \bar{f}_L \gamma^\mu \tau^a f_L \\ &+ \text{quadrilinear terms}\end{aligned} \tag{1.1}$$

The first term[1] in Eq. 1.1 describes the interactions of V with the SM weak bosons. The second term in the equation is the interactions of V with itself, where the mass parameter m_V does not coincide with the physical mass of the resonances. The third and fourth terms of the equation contains the interactions of V with the Higgs current[2] and with the SM left-handed fermionic currents[3], respectively. The quadrilinear terms[4] do not contribute directly to V decays and single production processes, therefore these terms can be disregarded.

[1]The term $D_{[\mu}V_{\nu]}^a = D_\mu V_\nu^a - D_\nu V_\mu^a$, where $D_\mu V_\nu^a = \partial_\mu V_\nu^a + g\varepsilon^{abc}W_\mu^b V_\nu^c$.

[2]The Higgs current term $iH^\dagger\tau^a\bar{D}^\mu H = iH^\dagger\tau^a D^\mu H - iD^\mu H^\dagger \tau^a H$.

[3]The fermionic currents $\sum \bar{f}_L\gamma^\mu\tau^a f_L$ involve the interactions of V to leptons, light quarks and the third quarks family.

[4]The quadrilinear terms can be written as

$$\frac{g_V}{2}c_{VVV}\varepsilon_{abc}V_\mu^a V_\nu^b D^{[\mu}V^{\nu]c} + g_V^2 c_{VVHH}V_\mu^a V^{\nu a}H^\dagger H - \frac{g}{2}c_{VVW}\varepsilon_{abc}W^{\mu\nu a}V_\mu^b V_\nu^c.$$

In Eq. 1.1, besides the $SU(2)_L$ coupling constant g, another coupling constant g_V is introduced to represent the strength of V interactions. In addition, the term c_H describes the V interactions with the SM vector bosons and with the Higgs. Similarly, the term c_F describes the V interactions with fermions. They are expected to be of the order of unity in most models.

Masses

After the EWSB, only the photon stays massless due to the unbroken $U(1)_{EM}$, while the weak bosons acquire a mass and a mixing with heavy vector V. The mass matrix of the (Z, V^0) and the $(W^\pm, V^\pm)$ are

$$\mathcal{M}_N^2 = \begin{pmatrix} \hat{m}_Z^2 & c_H \xi \hat{m}_Z \hat{m}_V \\ c_H \xi \hat{m}_Z \hat{m}_V & \hat{m}_V^2 \end{pmatrix} \tag{1.2}$$

and

$$\mathcal{M}_C^2 = \begin{pmatrix} \hat{m}_W^2 & c_H \xi \hat{m}_W \hat{m}_V \\ c_H \xi \hat{m}_W \hat{m}_V & \hat{m}_V^2 \end{pmatrix} \tag{1.3}$$

respectively, where

$$\hat{m}_Z = \frac{e}{2 \sin\theta_W \cos\theta_W} \hat{v}$$

$$\hat{m}_W = \cos\theta_W \hat{m}_Z$$

$$\hat{m}_V^2 = m_V^2 + g_V^2 c_{VVHH} \hat{v}^2$$

$$\xi = \frac{g_V \hat{v}}{2 \hat{m}_V}$$

Note that $e \approx \sqrt{4\pi/137}$, $\hat{v}$ is the Higgs field Vacuum Expectation Value, which has a value of 246 GeV, and θ_W is the weak mixing angle.

By taking the determinant of the mass matrices in Eq. 1.2 and Eq. 1.3, the relation of the physical masses M between charged and neutral heavy vectors are connected by θ_W.

$$m_W^2 M_\pm^2 = \cos^2\theta_W m_Z^2 M_0^2 \tag{1.4}$$

In the experimental searches, the masses of new vectors should be at or above TeV scale, but the masses of SM bosons $m_{W,Z}$ should be preserved at about 100 GeV. A hierarchy in the mass spectrum is required to have

$$\frac{\hat{m}_{W,Z}}{\hat{m}_V} \sim \frac{m_{W,Z}}{M_{\pm,0}} \ll 1 \tag{1.5}$$

Under the limit in Eq. 1.5, by expanding the determinant of the mass matrices in Eq. 1.2 and Eq. 1.3, a simple approximate expressions for m_W and m_Z are given by

$$m_Z^2 \approx \hat{m}_Z^2(1 - c_H^2\xi^2)$$

$$m_W^2 \approx \hat{m}_W^2(1 - c_H^2\xi^2)$$

Since $\hat{m}_W = \cos\theta_W \hat{m}_Z$, the W-Z mass ratio is given by

$$\frac{m_W^2}{m_Z^2} \simeq \cos^2\theta_W \tag{1.6}$$

Experimentally, the value of $\cos^2\theta_W$ is about 0.77. The charged and neutral Vs are degenerated by

$$M_\pm^2 = M_0^2(1 + \mathcal{O}(\%)) \tag{1.7}$$

It is clear that the mass splitting of charged and neutral states of V is small enough to be ignored. This implies that the two states have comparable production rates.

Decay Widths

Given that the mixing angles between weak bosons and V are small due to the hierarchy in the mass spectrum, the couplings of the neutral and charged V to left- and right-handed fermion chiralities can be written as

$$\begin{cases} g_L^N \simeq \frac{g^2}{g_V}\frac{c_F}{2}, & g_R^N \simeq 0 \\ g_L^C \simeq \frac{g^2}{g_V}\frac{c_F}{\sqrt{2}}, & g_R^C = 0 \end{cases} \tag{1.8}$$

The $(g_{L,R}^{W,Z})_{SM}$ in the Eq. 1.8 is the ordinary SM W and Z couplings with a normalization of $g_L^W = g/\sqrt{2}$. The g is electroweak coupling which has a value of 0.65.

The decay width Γ for fermionic channels can be written as

$$\Gamma_{V_\pm \to f\bar{f}'} \simeq 2\Gamma_{V_0 \to f\bar{f}} \simeq N_c[f](\frac{g^2 c_F}{g_V})^2 \frac{M_V}{48\pi} \tag{1.9}$$

where $N_c[f]$ is the number of colors (3 for di-quarks and 1 for di-leptons). The parameters $c_F = c_l, c_q, c_3$ control the relative branching ratios (BR) to leptons, light quarks and the third family quarks.

The decay width for bosonic channels are

$$\Gamma_{V_0 \to W_L^+ W_L^-} \simeq \Gamma_{V_\pm \to W_L^\pm Z_L} \simeq \Gamma_{V_0 \to Z_L h} \simeq \Gamma_{V_\pm \to W_L^\pm h} \simeq \frac{g_V^2 c_H^2 M_V}{192\pi}[1 + \mathcal{O}(\xi^2)] \tag{1.10}$$

The channels that are not reported in the Eq. 1.9 and Eq. 1.10 are either forbidden or suppressed.

Since the M_V is in the order of TeV scale, the ξ should be very small. In this case, for a given resonance mass, the decay widths are fixed by the couplings $g^2 c_F/g_V$ and $g_V c_H$. The BRs and the production rate are controlled by the two parameters $g^2 c_F/g_V$ and $g_V c_H$.

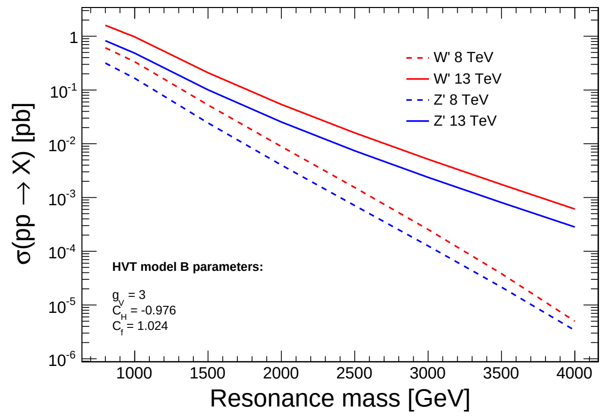

FIGURE 1.2: Theoretical production cross section as a function of resonance mass for HVT Minimal Composite Higgs Model.

1.2.2 Explicit Model

It is now clear that the explicit model can be entirely described in terms of the two couplings $g^2 c_F / g_V$ and $g_V c_H$ and the mass M_V with a good approximation [10].

Consider a strongly coupled scenario, so called Minimal Composite Higgs Model, where the Higgs doublet emerges from the spontaneous symmetry breaking of a global $SO(5)$ symmetry to an $SO(4)$ subgroup. In this scenario, the parameters c_H and c_F are fixed,

$$c_H \sim -1, \quad c_F \sim 1$$

In addition, $g_V \gtrsim 3$ is set to represent the strong coupling. In this case, the dominant BRs are bosonic decays due to $g_V c_H \simeq -g_V$ in Eq. 1.10, while the fermionic decays are extremely suppressed due to $g^2 c_F / g_V \simeq g^2 / g_V$ in Eq. 1.9.

The results of this model is particularly interesting for the present search, since it predicts signal cross sections in the order of fb for resonances up to 2∼3 TeV (Figure 1.2), branching ratios to vector bosons close to the unity (Figure 1.3), and thus being accessible at the LHC Run-II.

If the coupling is very large (for example g_V = 8), the total width will be

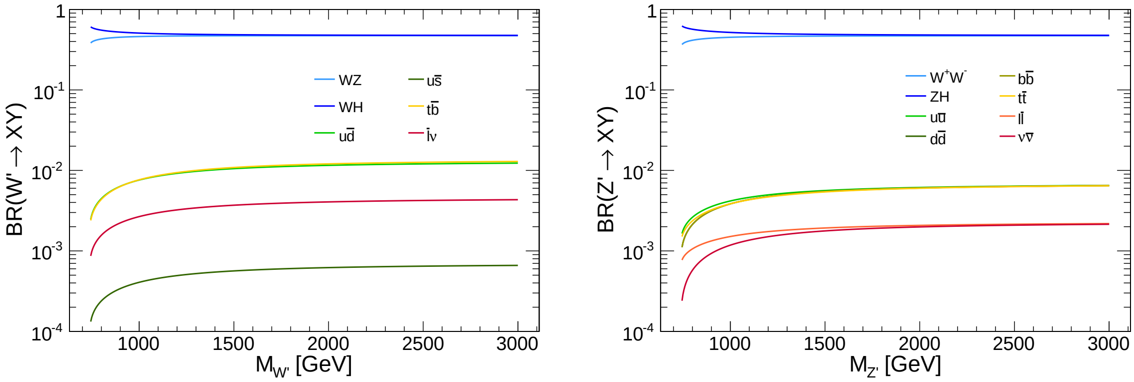

FIGURE 1.3: Branching ratios as a function of the resonance mass for a W′ (left) and Z′ (right) in the HVT Minimal Composite Higgs Model.

increased since the width of decays to dibosons grows with g_V from Eq. 1.10. But a very large coupling leads to an extremely broad resonance, which is not interesting to the experimental searches for a narrow resonance. Therefore, the g_v has been constrained by $g_V \lesssim 4\pi$.

Chapter 2

LHC and the CMS Detector

The analysis is performed using the data collected by the Compact Muon Solenoid (CMS) detector, which is one of the general purpose detectors at Large Hadron Collider (LHC). CMS is designed to search for new physics beyond the Standard Model (BSM), such as extra dimensions, dark matter and so on. In this chapter, a brief overview of the LHC and the CMS detector will be presented.

2.1 Large Hadron Collider

The LHC [11] is the largest and most energetic proton-proton collider in the world, built by the European Organization for Nuclear Research (CERN). The LHC lies in a tunnel 27 km in circumference, and about 100 m in depth beneath the France–Swiss border near Geneva. The layout of the CERN's accelerator complex is presented in Figure 2.1.

The proton source of the LHC is the hydrogen gas. First, the hydrogen gas is injected into a container which is surrounded by an electric field, in order to break down the gas into its constituent protons and electrons. The proton beam is then injected into a chain of accelerators, to increase their energies before entering the LHC: the LINAC2 accelerates the protons to an energy of 50 MeV. The beam is then injected into the Proton Synchrotron Booster (PSB), followed by the Proton Synchrotron (PS), which accelerates the protons to 25 GeV. Protons

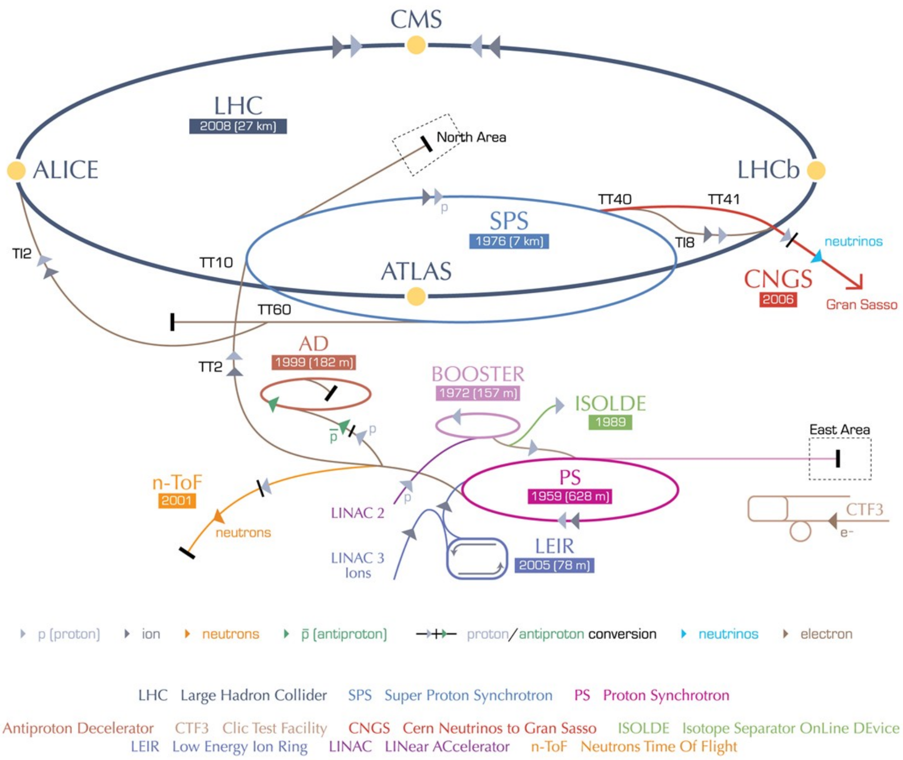

FIGURE 2.1: The CERN's accelerator complex.

are then injected to the Super Proton Synchrotron (SPS) where they are accelerated to 450 GeV.

Finally, the protons are injected to the two beam pipes of the LHC. The proton beams will circulate in two separate vacuum pipes, one clock-wise and the other counter-clockwise. The superconducting coils in the LHC generate two strong antiparallel magnetic fields of 8.4 Tesla to bend the two clock-wise/counter-clockwise proton beams along the same circular path. The proton beams are accelerated to their maximum energy of 6.5 TeV. At Run-II in 2015, the center-of-mass energy[1] $\sqrt{s}$ reached 13 TeV.

The proton beam in LHC consists of 2808 bunches at full intensity, each

[1] The center-of-mass energy can be calculated by the formula

$$E_{CoM} = \sqrt{s} = \sqrt{m_1^2 + m_2^2 + 2(E_1 E_2 + |\boldsymbol{p}_1||\boldsymbol{p}_2|)}$$

If $E_i \gg m_i$, the expression can further be approximated to

$$E_{CoM} \approx \sqrt{4E_1 E_2}$$

bunch containing 1.15×10^{11} protons, and has a bunch-crossing time of 25 ns. It travels at 0.999999991c in the LHC. Once the proton beam reaches its maximum energy, it is brought from one vacuum pipe to the other at four intersection points where the experiments (ATLAS, CMS, ALICE, and LHCb) are located. Focusing magnets are located at the intersection points to reduce the transverse dimensions of the beams, in order to achieve the desired luminosity values.

The design luminosity of the LHC is about 10^{34} cm^{-2}s^{-1}. If the particle distribution in the bunches is Gaussian, the luminosity can be expressed as

$$L = \frac{1}{4\pi}\frac{N_1 N_2 f}{\sigma_x \sigma_y} \tag{2.1}$$

where the N_1 and N_2 are the number of particles per bunch in the beams, σ_x and σ_y are the root-mean-squares of the distribution in the directions transverse to the beam, and f is the the bunch crossing frequency.

2.2 The Compact Muon Solenoid Detector

The CMS experiment [12] is one of the two general-purpose particle detectors built in the LHC. It is designed to explore physics at the TeV scale, to look for evidence of new physics beyond the standard model.

The CMS detector has various advantages on physical research. First, it has good muon identification, momentum and mass resolution, and the ability to determine the muon charge. Second, it has a good charged particle momentum resolution and reconstruction efficiency in the tracking system. Third, it has a good electromagnetic energy resolution and efficient photon and lepton isolation at high luminosity. Lastly, it has a good missing transverse energy and dijet mass resolution.

The CMS detector contains several subdetectors: the innermost layer is a silicon tracking system and is surrounded with an electromagnetic calorimeter;

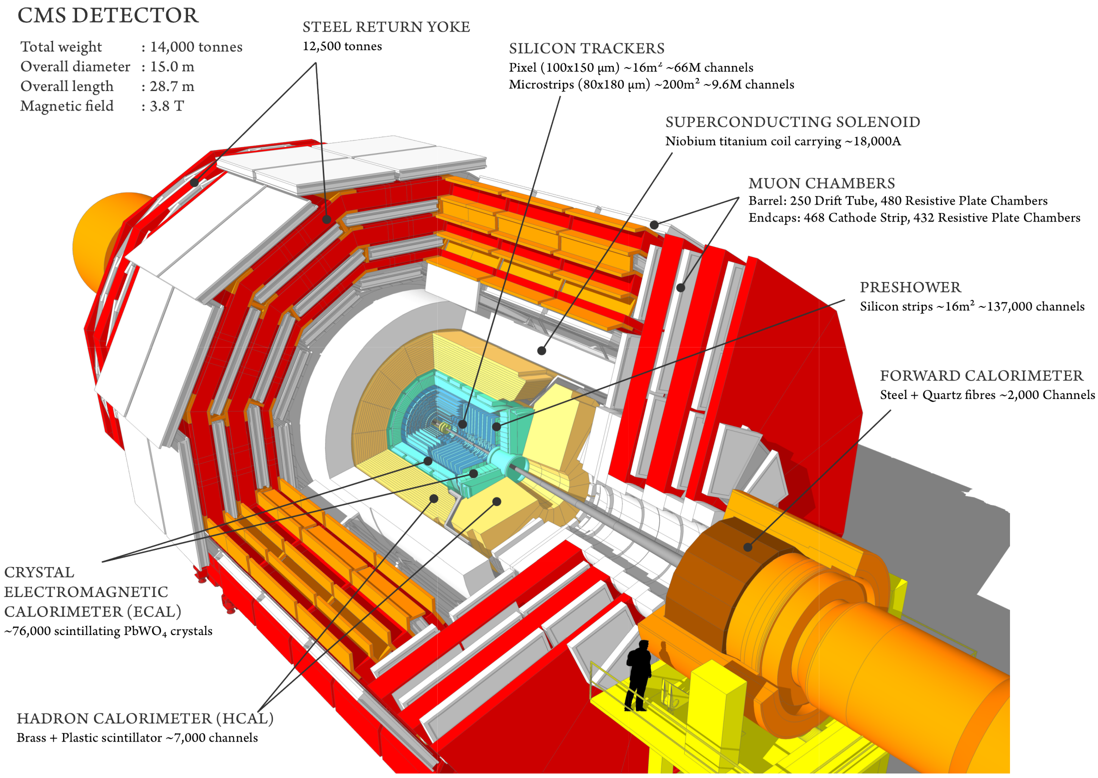

FIGURE 2.2: An overview of the CMS detector.

outside the electromagnetic calorimeter is a sampling calorimeter for hadrons. A magnet system provides a strong magnetic field of 3.8 Tesla in the tracker and the calorimeters. Outside the magnet are the large muon detectors, which are inside the return yoke of the magnet. Overall, the dimensions of the CMS are in 21.6 m in length, 15 m in diameter, and about 14,000 tonnes in weight. The overall layout of the CMS detector is presented in Figure 2.2. Each of the CMS subdetectors are introduced in the following subsections.

2.2.1 Magnet System

In order to have a good momentum resolution for charged particles, the superconducting solenoid [13] is installed at the outer part of the HCAL system. The inner free bore of the solenoid has a diameter of 6m and length of 12.5 m, enclosed inside a 12,000 tonne yoke made of steel. It provides a strong magnetic field of 3.8 Tesla, giving the total stored energy of 2.3 GJ in order to bend the paths of charged particles emerging from high energy collisions. Thus, the momentum and the charge-mass ratio of particles can be measured by tracing its curved path in the tracking system.

The yoke is composed of five three-layered dodecagonal barrel wheels and three endcap disks at each end. In the barrel region, the innermost yoke layer is 295 mm thick while the two outermost ones are 630 mm thick each. The yoke contributes only 8% of the central magnetic flux density; its main role is to increase the field homogeneity in the tracker volume and to reduce the stray field by returning the magnetic flux of the solenoid. In addition, the steel plates play the role of absorber for the four stations of muon chambers in order to reduce fake muon background.

2.2.2 Tracker Detector

The CMS tracker detectors [14] occupy the region around the interaction region of the LHC beams between 4 cm and 110 cm in radius and up to 280 cm along the beam axis and are situated in a 3.8 Tesla magnetic field. It is completely made of silicon, with a sensitive area of over 200 m^2. The layout of the tracker detector is presented in Figure 2.3.

The tracker detectors provide a high precision measurement of the three dimension trajectories of charged particles up to $|\eta| \leq 2.5$. These tracks are used to reconstruct the positions of the primary interaction and secondary decay vertices. Note that η is the pseudorapidity, $\eta \equiv -\ln[\tan(\theta/2)]$, θ is the polar angle of a particle relative to the beam axis.

The tracker detectors consist of two tracking devices: the pixel detector and the strip detector. The pixel detector is located at the innermost part of the tracking system. It has three layers in the barrel region with radii of 4.3 cm, 7.2 cm, and 11 cm, respectively, and two disks on the endcap regions at 34.5 cm and 46.5

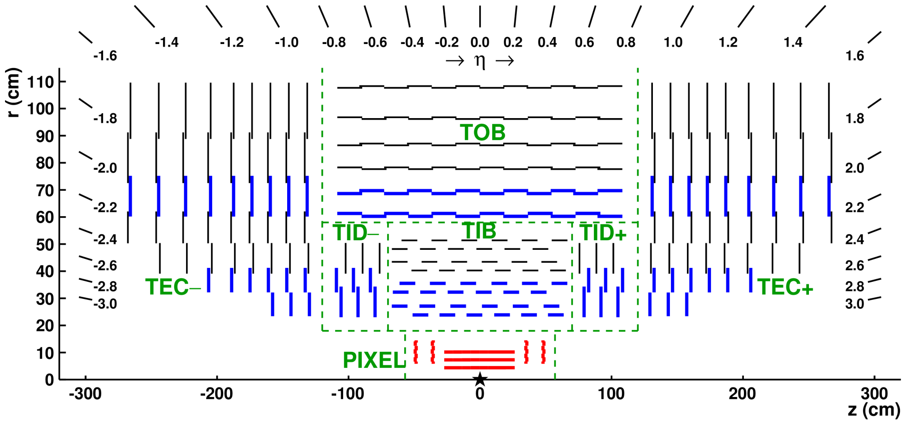

FIGURE 2.3: Longitudinal layout of two quadrant of the Tracker System.

cm from the interaction point. The pixel detector consists of 66 million pixels, covering about 1 m^2 total area.

The pixel detector is surrounded by the strip detector. The strip detector has 10 tracking layers in the barrel region with span radii from 25 cm to 110 cm and along the z axis up to 120 cm: 4 layers in the inner barrel (TIB) and 6 in the outer barrel (TOB). It also has 12 disks in the endcap region with radii up to 110 cm and in z up to 280 cm: 3 inner disks (TID) inside and 9 endcap disks (TEC) outside the TOB.

2.2.3 Electromagnetic Calorimeter

The Electromagnetic Calorimeter (ECAL) [15] measures the energy of EM particles like electrons and photons. The ECAL can identify the EM particles and determine the positions of EM particles with the help of the tracker system. The ECAL has a coverage of $0 \leq |\eta| \leq 1.479$, which is called the barrel region (EB); and a coverage of $1.479 \leq |\eta| \leq 3.0$, which is called the endcap region (EE). There is a preshower system (ES) in the region of $1.479 \leq |\eta| \leq 3.0$, installed in front of the EE. The ES system is used to reject the photons from π^0 decay, thus improving photon reconstruction. The layout of the ECAL is presented in Figure 2.4.

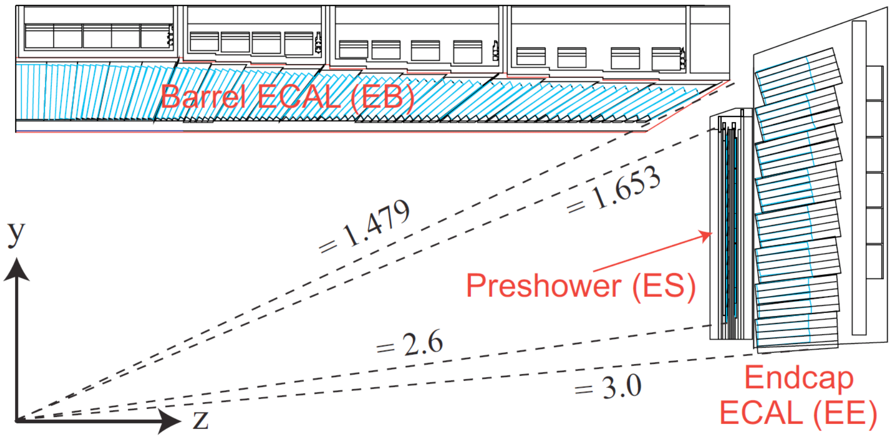

FIGURE 2.4: Longitudinal layout of one quadrant of the ECAL.

The primary detection material in ECAL is made of high density (8.28 g/cm^3) lead-tungstate crystals ($PbWO_4$). The crystal has a radiation length X_0 of 0.89 cm and a Moliére radius R_M of 2.2 cm. Note that a radiation length is the travel distance after which an EM particle looses 63% of its original energy. Having a smaller Moliére radius means that the ECAL has better position resolution and separation for EM showers due to a smaller degree of EM shower overlaps. The EB section contains 61,200 crystals, each of them having a dimension of 22×22 mm of front face and 230 mm (25.8X_0) of length. The EE section contains 7324 crystals, each of them having a dimension of 28.6×28.6 mm of front face and 220 mm (24.7X_0) of length. While the EM particles interact with the crystals by bremsstrahlung, the crystals can emit 80% of scintillation light within 25 ns. This shows that the response time of the ECAL system is faster than the bunch-crossing time of LHC.

Photodetectors are attached onto the back of each crystal. Silicon avalanche photodiodes (APD) are used in the EB section since it can operate in a strong transverse magnetic fields. It has an effective area of 25 mm^2. Vacuum phototriodes (VPT) are used in the EE as it can works under unusual radiation environment. It has an effective area of 280 mm^2.

2.2.4 Hadron Calorimeter

The Hadron Calorimeter (HCAL) [16] measures the energy of hadrons, which is the bound state of quarks and gluons. It also provides indirect measurement of the presence of non-interacting, uncharged particles, which are identified as missing transverse energy. In addition, the HCAL also helps in the identification of EM particles and muons in conjunction with the ECAL and muon detector. The layout of the HCAL is presented in Figure 2.5.

There are four distinct subsystems in HCAL: the barrel (HB), endcap (HE), outer (HO) and forward (HF) calorimeters. The HB has a coverage of $0 \leq |\eta| \leq 1.4$while the HE has a coverage of $1.3 \leq |\eta| \leq 3.0$. A range of $1.3 \leq |\eta| \leq 1.4$ is shared by HB and HE. HB and HE calorimeters are both sampling calorimeters, which consist of plastic scintillators inserted between copper absorber plates. The reason for using copper as an absorption material is that copper is a non-magnetic material. Since both HB and HE are located inside the cryostat of the superconducting solenoid, both experience a strong magnetic field.

The absorber plates are 5 cm thick in the barrel and 8 cm thick in the end-caps. The plastic scintillator is 4 mm thick. When a hadron hits the absorber plates, numerous secondary particles are produced. These secondary particles flow through successive layers of absorbers and interact with those layers, thus producing a bunch of hadronic particles, or hadronic shower. As this shower develops, the particles pass through the scintillator, causing it to emit blue-violet light. These optical signals are read out using wavelength-shifting (WLS) plastic fibers and are converted into fast electronic signals by Hybrid Photodiodes (HPD).

The HO calorimeter is installed outside the superconducting solenoid, in order to catch the energy leakage from HB. In fact, about 5% of all hadrons above 100 GeV energy are deposited in HO.

As a requirement for an optimal missing transverse energy measurement,

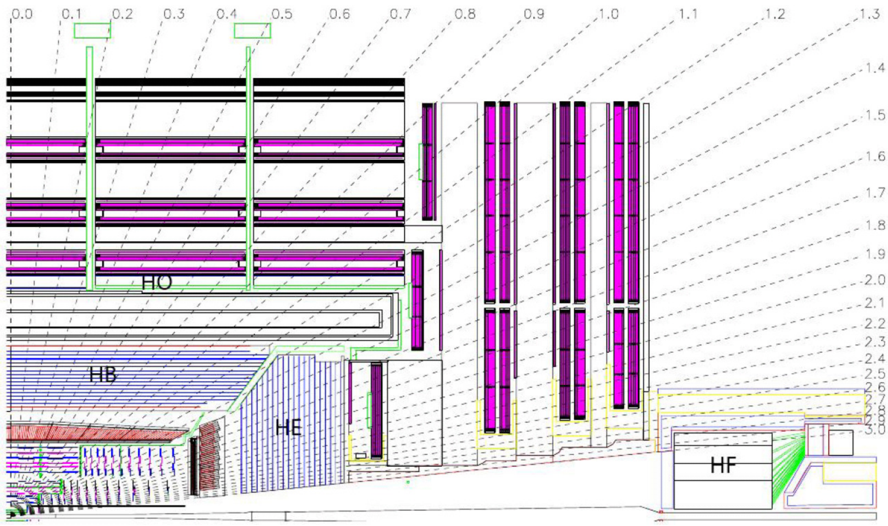

FIGURE 2.5: Longitudinal layout of one quadrant of the HCAL.

the HF calorimeter is installed 6 m downstream from the HE endcaps, and covers the forward region of $3.0 \leq |\eta| \leq 5.2$ to extend the hermeticity of the central hadron calorimeter system. Since the HF is located in a very high rate environment, quartz fibers are used as active element, and are embedded in a copper absorber plates in HF calorimeter.

2.2.5 Muon Detector

The muon detector [17] provides the identification of muons, momentum measurement of muons, as well as muon triggering. Muon detection is one of the most important tasks in CMS, as muons provide clean signatures of interesting physics processes over large kinematic backgrounds. The reason is that the high momentum muons have a relatively longer life-time and can pass through the inner detectors in CMS, which provides good discrimination power from other particles. Moreover, the charged muons bend under strong magnetic fields and leave tracks in the silicon tracker, as well as in the muon detector. Hence, the transverse momentum of muons can be measured precisely. The high momentum resolution of muons is provided by the high spatial resolution of the muon detector, and the high magnetic field of the solenoidal magnet and its flux-return yoke. The layout of the muon detector is presented in Figure 2.6.

The muon detector consists of 3 types of gas-ionization particle detectors: the drift tube chambers (DT), the cathode strip chambers (CSC), and the resistive plate chambers (RPC). To match the cylindrical geometry of the solenoid, the barrel region ($0 \leq |\eta| \leq 1.2$) is instrumented with DT, and the two endcap regions ($0.9 \leq |\eta| \leq 2.4$) with CSC. RPC are interspersed in both the barrel and endcap regions. The muon chambers must detect the traversing track at several points along the track path to utilize the magnet to measure the deflection of muons as they pass through its field. In the barrel region, the chambers should be at several different values of the radial distance r from the beam line, and in the endcap region at several different values of distance along the beam direction z.

The barrel section has low background levels and the magnetic field is mostly uniform. The DT system is installed here since it can measure muon trajectories precisely. They are organized into 12 ϕ-segments per wheel, where ϕ is the azimuthal angle, forming 4 stations at different radii interspersed between plates of the magnet flux return yoke. Each station consists of 8 layers of tubes measuring the position in the bending plane and 4 layers in the longitudinal plane.

The endcap sections have high muon rates and background levels in high luminosity. The CSC system is installed here since it has fast response time and can tolerate the strong, non-uniform magnetic field. Each endcap has 4 stations of chambers mounted on the faces of the endcap steel disks perpendicular to the beam. A CSC consists of 6 layers, each of which measures the muon position in 2 coordinates. The cathode strips inside the CSC run radially outward and provide a precise measurement in the $r - \phi$ bending plane. The wires provide a rough measurement in the radial direction.

Besides these tracking detectors, the RPC system serves as a dedicated triggering detector system with excellent time and spatial resolution. It can provide a fast, independent trigger with a looser transverse momentum threshold over

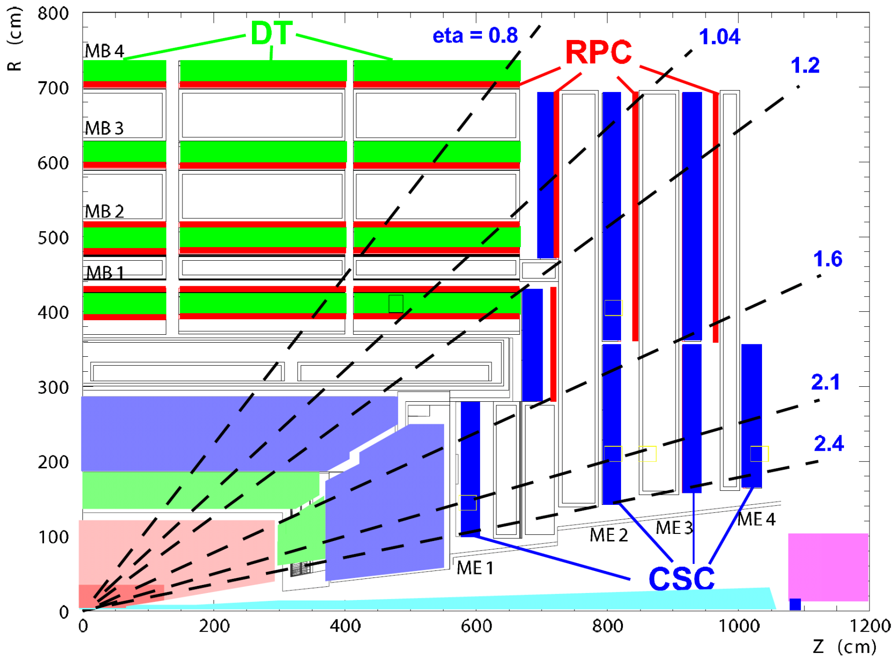

FIGURE 2.6: Longitudinal layout of one quadrant of the Muon Detector.

a large portion of the η range ($|\eta| \leq 1.6$).

2.3 Trigger System

The proton-proton collisions at a center-of-mass energy of 13 TeV have a total cross section of 70 mb, with luminosity of 10^7 $mb^{-1}s^{-1}$. Following the definition of the event rate as the cross section times the luminosity, the event rate is about $7{\times}10^8$ Hz. This means that there are about 10^9 events per second in pp collisions. A typical event occupies $\sim$ 1 MB in the storage system, resulting in a raw data rate of 1 PB per second. It is impossible to have the readout bandwidth or the storage capability to store all of the data produced. Therefore, data reduction is achieved by several trigger systems in CMS, which are the First Level (L1) Trigger [18] and the High Level Trigger (HLT) [19]. The structure of the CMS trigger system is shown in Figure 2.7.

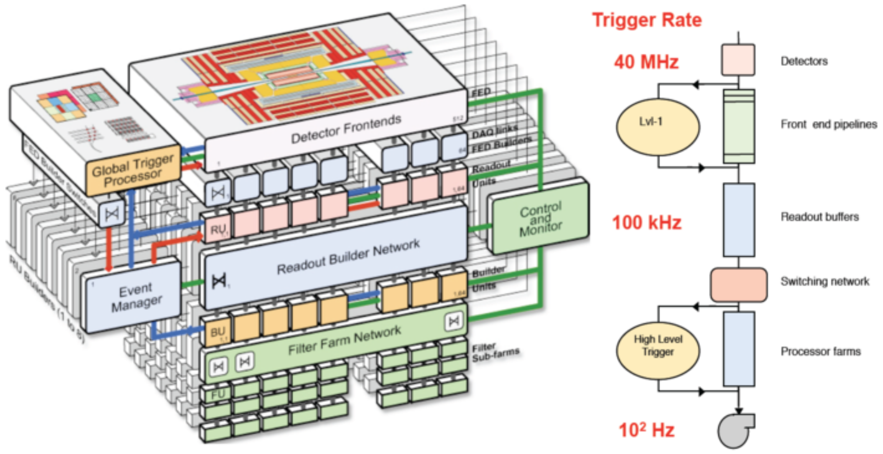

FIGURE 2.7: The CMS trigger and DAQ system.

2.3.1 First Level Trigger

The L1 trigger system processes data at a rate of 40 MHz, with an output rate of 100 KHz. It access the data from calorimeters and the muon system to perform online event selection using electronic systems. The algorithms process data in pipeline fashion, using pattern recognition and fast summing techniques, without introducing dead-time.

2.3.2 High Level Trigger

The HLT is implemented in softwares running on a large computer farm of fast commercial processors. It reduces the L1 output rate to 1 KHz, a sustainable level for storage and physics analysis. The HLT software is similar to the offline reconstruction algorithms, but is more efficient in complying with the strict time requirements of the online selection.

Chapter 3

Analysis Strategy

The purpose of this analysis is to search for a heavy resonance decaying into a Z boson and a Higgs boson. The Z boson is reconstructed from the dielectron or dimuon final state. The $Z \to l^+l^-$ channel is chosen mainly because it has low QCD and $t\bar{t}$ backgrounds. The Higgs boson is reconstructed via the $b\bar{b}$ channel because the branching ratio of $H \to b\bar{b}$ is the largest at m_H = 125 GeV. Note that the b quarks are reconstructed as a single jet with large radius in the CMS detector due to the high transverse momentum of Higgs.

In this chapter, the data sets and Monte Carlo (MC) samples used are presented, and the selection criteria of physics objects are introduced. The final selection efficiency is also discussed and reported. The data and MC comparisons of various kinematic variables are presented to show the agreement between them.

3.1 Data Sets and Monte Carlo Samples

3.1.1 Data Sets

In this analysis, the data collected with CMS during 2015 RunD, with integrated luminosity of 2.51 fb^{-1} are used. The muon and electron data sets are collected with a single-muon or a single-electron trigger, which will be explained in detail

TABLE 3.1: Data sets Run2015D used in this analysis.

Samples	**Luminosity** L **[**fb^{-1}**]**
SingleElectron-Run2015D-05Oct2015-v1	0.877
SingleElectron-Run2015D-PromptReco-v4	1.635
SingleMuon-Run2015D-05Oct2015-v1	0.877
SingleMuon-Run2015D-PromptReco-v4	1.635

TABLE 3.2: Filters used in this analysis.

Filters
CSCTightHaloFilter
eeBadScFilter
HBHENoiseFilter
HBHENoiseIsoFilter

in latter section. All data sets used are listed in Table 3.1. Moreover, a list of filters [20] are applied on data in order to remove problematic or noise-dominated events, are reported in Table 3.2.

3.1.2 Monte Carlo Samples

Signal samples

The signal samples are generated with the `MadGraph5` [21] LO generator. The showering and hadronization have been performed with `PYTHIA8` [22]. A full detector simulation and event reconstruction has been performed with `GEANT4` [23] and `CMSSW`. The simulated signal MC samples are listed in Table 3.3. All signal samples belong to the `RunIISpring15MiniAODv2-74X_mcRun2` campaign with the 25 ns asymptotic conditions. Moreover, the samples are produced assuming the narrow-width approximation, with the resonance width set to 1 MeV.

Background samples

All physics processes yielding final states with one or two leptons in association with one or two b quarks are considered as possible sources of background for the analysis. The list of background samples used is reported in Table 3.4.

The Z+jets background is produced in several samples binned in H_T (the scalar sum of the p_T of the outgoing partons) starting from 100 GeV with the `MadGraph5` LO generator. The contribution of events with H_T less than 100 GeV is found to be negligible after requiring the Z p_T to be greater than 200 GeV. The inclusive $t\bar{t}$ sample has been produced with `POWHEG` [24] interfaced with `PYTHIA8`, including all the possible decays of the W bosons. For the diboson production processes, the WW, WZ, and ZZ are inclusive processes generated with `PYTHIA8`, while the ZH has decay mode same as the decay mode of signal and is generated with `POWHEG`.

The cross sections listed in the table are used to normalize SM backgrounds. The cross sections of Z+jets samples are computed by `MadGraph5` with an NLO/LO electroweak correction [25] (k-factor) derived from the inclusive Z cross section computed by `FEWZ` [26]. The amount of k-factor as a function of the Z p_T is presented in Figure 3.1. The cross section of $t\bar{t}$ samples are obtained from TTbarNNLO group [27], while the VV samples are computed from `MCFM` [28] calculator.

3.2 Trigger

In this analysis, the events are selected online by the L1 trigger and the HLT. For each $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ channel, a single lepton trigger path (a string that identifies a list of selections performed at HLT) is used to select events. The trigger paths used in this analysis are reported in Table 3.5. The same set of trigger paths are applied for both data and simulated events. The single-lepton triggers require at least one lepton.

TABLE 3.3: Signal samples used in this analysis.

Samples	**Number of events**	**Cross section** σ $[pb]$
ZprimeToZhToZlephbb_narrow_M-800_13TeV-madgraph-v1	48400	0.0282665
ZprimeToZhToZlephbb_narrow_M-1000_13TeV-madgraph-v1	50000	0.0153743
ZprimeToZhToZlephbb_narrow_M-1200_13TeV-madgraph-v1	50000	0.00790857
ZprimeToZhToZlephbb_narrow_M-1400_13TeV-madgraph-v1	50000	0.00421385
ZprimeToZhToZlephbb_narrow_M-1600_13TeV-madgraph-v1	50000	0.00233319
ZprimeToZhToZlephbb_narrow_M-1800_13TeV-madgraph-v1	50000	0.00133522
ZprimeToZhToZlephbb_narrow_M-2000_13TeV-madgraph-v1	50000	0.000785119
ZprimeToZhToZlephbb_narrow_M-2500_13TeV-madgraph-v1	50000	0.000227178
ZprimeToZhToZlephbb_narrow_M-3000_13TeV-madgraph-v1	50000	0.000071426
ZprimeToZhToZlephbb_narrow_M-3500_13TeV-madgraph-v1	49800	0.0000235715
ZprimeToZhToZlephbb_narrow_M-4000_13TeV-madgraph-v1	49800	0.00000797489

TABLE 3.4: Background samples used in this analysis.

Samples	**Number of events**	**Cross section σ [pb]**
DYJetsToLL_M-50_HT-100to200_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	2725655	147.4
DYJetsToLL_M-50_HT-200to400_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	973937	40.99
DYJetsToLL_M-50_HT-400to600_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	1018388	5.678
DYJetsToLL_M-50_HT-600toInf_TuneCUETP8M1_13TeV-madgraphMLM-pythia8	998912	2.198
TT_TuneCUETP8M1_13TeV-powheg-pythia8	115928725	831.76
WW_TuneCUETP8M1_13TeV-pythia8	993640	118.7
WZ_TuneCUETP8M1_13TeV-pythia8	978512	47.13
ZZ_TuneCUETP8M1_13TeV-pythia8	996944	16.523
ZH_HToBB_ZToLL_M125_13TeV_powheg_pythia8	1980540	0.05

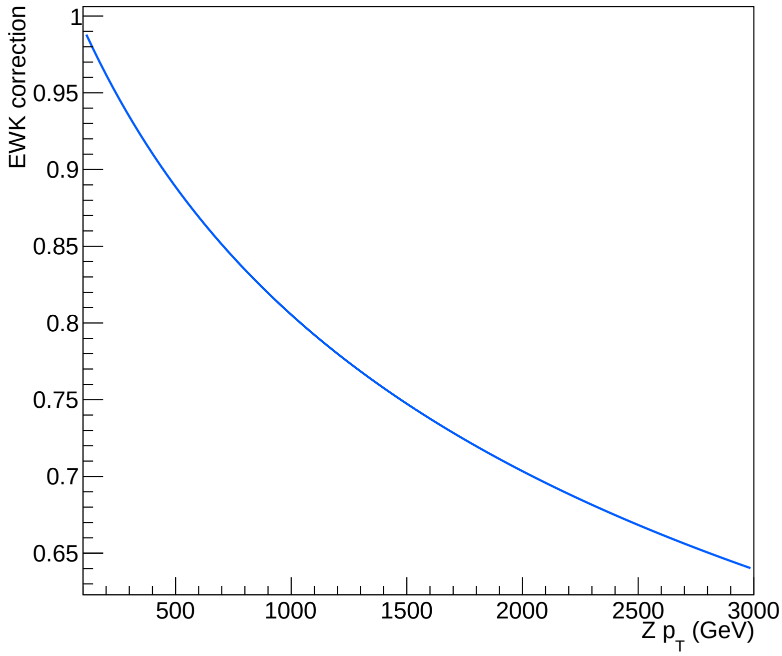

FIGURE 3.1: Electroweak correction for the Z as a function of the transverse momentum.

The trigger efficiencies of both of the trigger paths have been studied by the EGamma and Muon POG [29, 30]. The trigger efficiency is defined as the fraction of events passing offline selection that also satisfies online trigger requirements.

The good leptons are chosen with a tag-and-probe [31] procedure by selecting $Z \to l^+l^-$ events. The tag-and-probe method is a data-driven method to measure efficiency from data. A "tag" is a well identified lepton that satisfies a tight lepton ID requirements. The "probe" is the other lepton that pass very loose selection criteria. The method is used to count the events wherein both "tag" and "probe" exist. The trigger efficiency is then evaluated by studying the probe lepton efficiency as a function of both p_T and η in data.

The study shows that it has more benefits using single lepton triggers without isolation requirement but with higher p_T threshold than non-isolated double-lepton triggers with lower p_T threshold. The single-lepton triggers only require at least one-lepton to pass the loose working point and to have a p_T larger than 135 GeV. As the Z bosons usually have large p_T in this analysis, if the leptons

TABLE 3.5: HLT path used in this analysis.

HLT path
HLT_Ele105_CaloIdVT_GsfTrkIdT
HLT_Mu45_eta2p1

are produced parallel or anti-parallel to direction of Z boson, one of the leptons will have a very low p_T. The single-lepton triggers are safe choices to keep this kind of events.

3.3 Vertex and Pile-up

At the LHC, there is a probability that one single bunch crossing may produce several pp collisions, known as pile-up (PU) events. Due to PU, several primary vertices (PV) are typically reconstructed in an event. The PU contributions are removed and only the most interesting collision event was retained. The primary vertex must pass the following requirements [32]:

- Number of degrees of freedom, $N_{DoF} > 4$
- The distance along the beam axis z of the vertex with respect to the nominal interaction point (0,0,0), $|d_z| < 24$ cm
- The distance perpendicular to the beam axis of the vertex with respect to the nominal interaction point (0,0,0), $d_0 < 2$ cm

The primary vertex associated with the signal in each event is defined as the one with the highest sum of square of the transverse momentum of the tracks associated to it. The `Spring15 v2` MINIAOD MC samples are generated simulating the PU conditions using the 25 ns asymptotic PU scenario. The MC samples are reweighted with the standard CMS PU reweighting technique [33, 34],

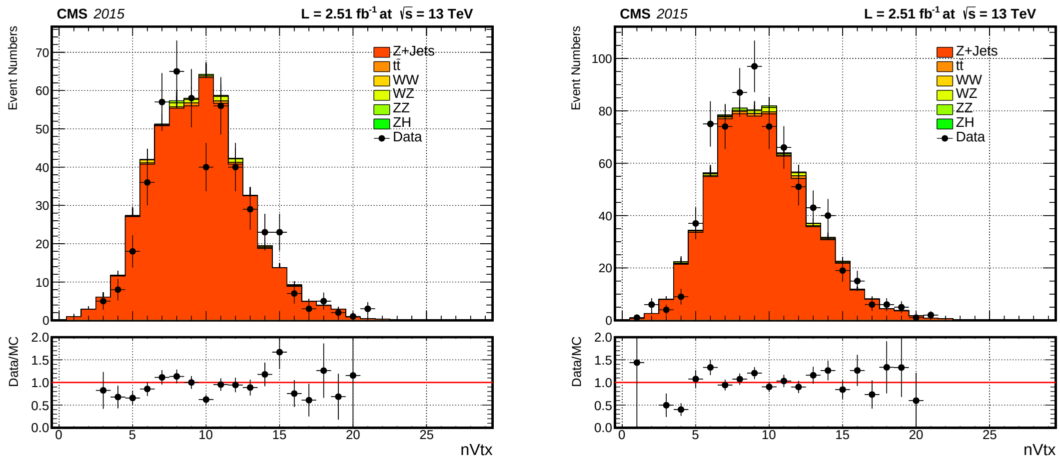

FIGURE 3.2: Distributions of the number of reconstructed primary vertices after reweighting in the electron channel (left) and muon channel (right).

in order to improve the agreement of PU distribution with data. The comparison between the distributions of primary vertices in data and MC after the PU reweighting is shown in Fig. 3.2 for events after full selection.

3.4 Physics Objects

In the CMS experiment, all the physics objects are reconstructed by the Particle Flow (PF) algorithm [35, 36]. This section briefly introduces the reconstruction method of physics objects that are used in this analysis. Their preselections are reported in detail.

3.4.1 Electron Reconstruction

An electron carries electric charge and is involved in electroweak interaction. It can leave tracks in the tracker system and deposit its entire energy in the ECAL. The electrons are reconstructed by associating a track reconstructed in the silicon detector with a cluster of energy in the ECAL.

Energy Clustering

In the CMS detector, the electrons lose their energy via bremsstrahlung before reaching the ECAL. To measure the initial energy of the electrons accurately, it is necessary to collect the energy of the radiated photons that spreads along the ϕ direction due to the bending of the electron trajectory in the magnetic field. Two clustering algorithms, the Hybrid algorithm and the Multi-5×5 algorithm, attempt to cluster the energy of electrons in the barrel and endcap sections of the ECAL [37].

The Hybrid algorithm makes use of the geometry of the EB and properties of the shower shape[1], collecting the energy in a narrow window in η and an extended window in ϕ. The algorithm starts from a seed crystal (crystal with transverse energy E_T above a certain threshold), forming 1×3 or 1×5 crystal dominos, each with their central crystal aligned in η with the seed crystal. The dominos are then clustered in ϕ from the original seed.

Finally, all the non-overlapping clusters in a narrow η window are built to collect the energy radiated by bremsstrahlung that falls outside the main cluster, to form a cluster of clusters, or a superclusters (SC).

The Multi-5×5 algorithm also starts from the most energetic seed then collects the energy in a cluster of 5×5 crystals. These clusters are then grouped into a SC if their total E_T satisfy specific requirements.

Track Reconstruction

Particle tracks are generally constructed by the standard Kalman Filtering (KF) [38] track reconstruction procedure.

The building of trajectories is initiated by looking for a seed in the inner

[1]Since the magnetic field inside the CMS solenoid is purely axial, the clusters belonging to radiation from an electron will be nicely aligned in η, but spread in ϕ.

tracker with the Seed Generator. Then, all the possible trajectories are constructed outward by the given seed with the Trajectory Builder. Among the possible trajectories, the best track candidates are kept with the Trajectory Cleaner. Lastly, the final fit of the track is performed with the Trajectory Smoother, in which all the collected hits are used to estimate the track parameters at each layer.

In the electron tracks reconstruction, the dedicated Gaussian Sum Filter (GSF) [39] is used instead of the KF, which improves the model of bremsstrahlung energy loss distribution of electrons. It turns out that the GSF provides a better momentum resolution of electrons compared to the KF.

Finally, the electrons are reconstructed by matching the SC to the GSF track. The reconstruction efficiency of electrons with E_T greater than 20 GeV is above 95%(90%) in EB(EE).

3.4.2 Electron Selection

The selection criteria of the electrons in this analysis follow the recommendations of EGamma POG [40]. Since the backgrounds are rather low, a loose working point is preferred (average efficiency ∼90%). The electron candidates are selected with $p_T > 20$ GeV and $|\eta| < 2.5$ and at least one of the two electrons must have a p_T higher than 135 GeV. The requirements for electrons are listed in Table 3.7 with the cut values for the ECAL barrel and ECAL endcap. The definition of the electron variables are explained in the following:

- η: The pseudorapidity of the electron track as measured at the inner layer of the tracker and then extrapolated to the interaction vertex.

- η_{SC}: The pseudorapidity of the supercluster of electron, with respect to the point (0,0,0).

- E_T: Defined as $E_T = E_{\text{ele}} \times \sin\theta_{\text{trk}}$, where θ_{trk} is the polar angle of the electron track measured at the inner tracker layer and then extrapolated to the interaction vertex.

- $\sigma_{i\eta i\eta}$: A measure of the spread in η in units of crystals of the electron's energy in a 5×5 block centered on the seed crystal.

- $\Delta\eta_{in}$ and $\Delta\phi_{in}$: The difference in η and ϕ between the track position as measured in the inner tracker layer, extrapolated to the interaction vertex and then extrapolated to the calorimeter and the η and ϕ of the supercluster.

- H/E: The ratio of the hadronic energy collected by the HCAL tower within $\Delta R \equiv \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} < 0.1$ centered on the electron position in the calorimeter to the EM energy of the electron supercluster.

- d_0 and d_z: The distance between the electron trajectory and the PV in the transverse and the longitudinal coordinates.

- $1/E - 1/p$: The difference of the inverse of the energy and the momentum of the electron.

- missing hits: The exception of missing hits at the inner pixel system.

- conversion veto: A flag to exclude electrons produced by photon conversions.

In addition to the above requirements, the electrons are required to be isolated. An isolation requirement is applied to suppress leptons coming from jets due to misidentification or the decays of b hadrons or c hadrons inside the jets.

The isolation I is defined as the scalar sum of the p_T of all the PF candidates, excluding the lepton itself, within a cone of $\Delta R < 0.3$ around the lepton direction, after the contribution from pile-up and other reconstructed leptons are

TABLE 3.6: Effective Area constants for computing pile-up corrections.

$\|\eta\|_{\text{min}}$	$\|\eta\|_{\text{max}}$	**EA**
0.0	1.0	0.1752
1.0	1.479	0.1862
1.479	2.0	0.1411
2.0	2.2	0.1534
2.2	2.3	0.1903
2.3	2.4	0.2243
2.4	5.0	0.2687

removed. The relative isolation, obtained by dividing the simple isolation by the lepton p_T

$$I_{\text{rel}} = I/p_T^{\text{lep}} \tag{3.1}$$

is more frequently used in the analysis.

This analysis uses the relative isolation with effective area (EA) pile-up corrections [37], which is defined as

$$I_{\text{rel}}^{\text{EA}} = I_{\text{chHad}} + I_{\text{NeHad}} + I_\gamma - \rho \times \text{EA} \tag{3.2}$$

where I_{chHad} is the scalar sum p_T of the PF charged hadrons, I_{NeHad} is the scalar sum p_T of the PF neutral hadrons, and I_γ is the scalar sum p_T of the PF photons. The pile up correction term involves the energy density ρ multiplied by η dependent effective area. The effective area constants are listed in Table 3.6.

3.4.3 Muon Reconstruction

The muons are reconstructed independently in the inner tracking system and in the muon system [31]. The muon track in the tracker is called tracker track, while the muon track in the muon system is called standalone muon track. In CMS, there are two reconstruction approaches based on these objects.

TABLE 3.7: Electron identification requirements for Spring15 25 ns conditions. EB: barrel cuts ($|\eta_{SC}| < 1.4442$); EE: endcap cuts ($1.566 < |\eta_{SC}| < 2.5$).

Electrons		**EB**	**EE**
p_T (GeV)	$>$	20	20
$\sigma_{i\eta i\eta}$	$<$	0.0103	0.0301
$\|\Delta\eta_{in}\|$	$<$	0.0105	0.00814
$\|\Delta\phi_{in}\|$	$<$	0.115	0.182
H/E	$<$	0.104	0.0897
relIso(EA)	$<$	0.0893	0.121
$1/E - 1/p$	$<$	0.102	0.126
$\|d_0\|$ (cm)	$<$	0.0261	0.118
$\|d_z\|$ (cm)	$<$	0.41	0.822
missing hits	$\leq$	2	1
conversion veto		yes	yes

In the Global Muon reconstruction, a matching tracker track is found for each standalone muon. Then, the global muon track is fitted using all hits in both tracks by Kalman filter technique. In the Tracker Muon reconstruction, all tracker tracks with specific requirements are extrapolated to the muon system. If at least one muon segment (short track from the DT or CSC hits) matches the extrapolated track, the corresponding tracker track is identified as a Tracker Muon.

As the tracker track efficiency and the reconstructing segments efficiency in the muon system are very high, about 99% of muons from pp collisions which have high momentum are reconstructed either as a Global Muon and/or a Tracker Muon.

3.4.4 Muon Selection

In this analysis, only global muons or tracker muons are used [41]. The muon candidates are selected with $p_T > 20$ GeV and $|\eta| < 2.4$ and at least one of the two muons must have a p_T higher than 55 GeV. In addition, one of the two muons must be reconstructed as a Global Muon. Furthermore, the muons must pass the requirements listed in Table 3.8. The meaning of each muon variable is described as follows:

- muHits: The number of muon-chamber hits included in the global muon track fit. This is required to suppress hadronic punch-through and muons from decays in flight.
- muMatches: Muon segments in the muon stations. This is required to suppress punch-through and accidental track-to-segment matches.
- muTrkLayers: Number of tracker layers with hits. This is required to guarantee a good p_T measurement and also suppress muons from decays in flight.
- muPixelHits: Number of pixel hits. This is required to further suppress muons from decays in flight.
- muTrkPtErr/muTrkPt: The relative p_T error of the muon best track.
- d_{xy}: The transverse impact parameter of the tracker track with respect to the PV. This is required to suppress cosmic muons and further suppress muons from decays in flight.
- d_z: The longitudinal distance of the tracker track with respect to the PV. This is required to suppress cosmic muons, muons from decays in flight and tracks from PU.

For the muon candidates, different isolation criteria are used, based on the ΔR between two muons coming from decays of highly-boosted Z bosons.

The isolation I_{trk} is computed as the scalar sum of the p_T of all the tracks from the leading PV in the event within a cone centered along the muon direction,

TABLE 3.8: Muon identification requirements for Spring15 25 ns conditions.

Muons		Global Muon	Tracker Muon
muHits	>	1	n/a
muMatches	>	2	2
muTrkLayers	>	5	5
muPixelHits	>	1	1
muTrkPtErr/muTrkPt	<	0.3	0.3
$\|d_{xy}\|$	<	0.2	0.2
$\|d_z\|$	<	0.5	0.5
isolation	<	0.2	0.2

normalized by the muon p_T (Eq. 3.3).

$$I_{\text{rel}} = \begin{cases} I_{\text{trk}}/p_T^{\text{lep}} & \text{if} \quad \Delta R > 0.3 \\ (I_{\text{trk}} - p_T^{\text{intrk}})/p_T^{\text{lep}} & \text{if} \quad \Delta R < 0.3 \end{cases} \tag{3.3}$$

In the case when the two muons are close to each other ($\Delta R < 0.3$), the inner track p_T of the neighboring muon p_T^{intrk} is subtracted from the track isolation, in order to reject the other muon inside the isolation cone of the target muon.

3.4.5 Jet Reconstruction

A jet is a collimated bunch of stable hadrons, originating from quarks and gluons after fragmentation and hadronization. Quarks and gluons appear in the final states of many interesting physics processes; due to the quark confinement, they get hadronized to colorless hadrons which can be detected. The jet clustering algorithms collect the hadrons which arise from the hadronization process in a cone around the direction of the parton that characterizes a jet, with a good approximation.

Jet Clustering

There are several algorithms to cluster a jet in the CMS experiment [42]. These algorithms assume that particles within jets have small differences in transverse momenta, thus the particles are grouped based on momentum space, resulting in jets that have fluctuating areas in (η-ϕ) space. The algorithms calculate distance d_{ij} between i and j particles by Eq. 3.4.

$$\begin{aligned} d_{ij} &= \min(p_{ti}^{2a}, p_{tj}^{2a})\frac{\Delta\eta^2 + \Delta\phi^2}{R^2} \\ d_{iB} &= p_{ti}^{2a} \end{aligned} \tag{3.4}$$

where $\Delta\eta^2 + \Delta\phi^2$ is the (η-ϕ) space distance between the two particles, R is the radius parameter which determines the final size of the jet. The variable $d_{iB} = p_{ti}^{2a}$ is the momentum space distance between the beam axis and the particle.

The parameter a corresponding to different algorithms are listed below:

- a = 1: k_t algorithm
- a = 0: Cambridge/Aachen (CA) algorithm
- a = -1: anti-k_t (AK) algorithm

The algorithms begin with finding the minimum of the set $\{d_{ij}$,$d_{iB}\}$. If d_{ij} is the minimum, then particles i and j are combined into one particle (ij) by summing their four-vectors. If d_{iB} is the minimum, then i is labeled a final jet and removed from the list of particles. This process is repeated until all particles are part of a jet with the distance between the jet axes R_{ij} greater than R.

In this analysis, the jet objects are reconstructed by the anti-k_t algorithm [43], which is currently highly recommended within the CMS experiment. From Eq. 3.4, it is clear that the anti-k_t algorithm prefers to cluster hard particles first. Therefore, the algorithm is less sensitive to the underlying event (UE) and pile-up. Although the anti-k_t algorithm is the best at resolving jets, it is the worst for studying jet substructure due to its poor declustering.

Jet Grooming

A grooming algorithm is applied after the jet clustering to reduce the soft and wide-angle radiation components of the jet (quark- and gluon-initiated) while preserving the hard structure that originating from heavy particles. In addition, the grooming algorithm can be used to reduce the effect of pile-up on the jet [44]. There are many different types of jet grooming algorithms. In this section, the pruning algorithm [45] and the soft-drop algorithm [46] are focused on.

The pruning algorithm reclusters the constituents of the jet through the CA clustering algorithm mentioned above (Eq. 3.4), using the same distance parameter. At each step in the clustering algorithm, the soft and wide-angle recombination between the two particles i and j is vetoed when both of the following conditions are met:

$$\begin{aligned} z_{ij} &= \frac{\min(p_{Ti}, p_{Tj})}{p_{Ti} + p_{Tj}} < z_{\text{cut}} \\ \Delta R_{ij} > D_{\text{cut}} &= \frac{2 \times r_{\text{cut}} \times m_J}{p_T} \end{aligned} \tag{3.5}$$

where m_J and p_T are the mass and transverse momentum of the originally-clustered jet, z_{cut} and r_{cut} are parameters of the algorithm. At CMS, the parameter values of z_{cut} and r_{cut} are set to 0.1 and 0.5, respectively.

In the soft-drop algorithm, the constituents in the AK8 jets are first re-clustered with the CA algorithm. Then the CA8 jet is declustered to sub-jets j_1 and j_2. The soft-drop procedure removes the softer constituent unless the following condition is met:

$$\frac{\min(p_{T1}, p_{T2})}{p_{T1} + p_{T2}} > z_{\text{cut}} \times (\frac{\Delta R_{12}}{R_0})^{\beta} \tag{3.6}$$

where R_0 is the distance R of the original jet clustering algorithm, p_{T1} and p_{T2} are the transverse momenta of the constituents with respect to the beam, ΔR_{12} is their distance in the (η-ϕ) plane, and z_{cut} and angular exponent β are parameters of the algorithm. If $\beta > 0$, the soft drop declustering removes soft radiation from a jet while maintaining a fraction of the soft-collinear radiation. If $\beta < 0$, the soft drop declustering can remove both soft and collinear radiations. At CMS, the parameter values of $z_{\rm cut}$ and β are set to 0.1 and zero, respectively.

In this analysis, the mass of Higgs jets are groomed by the pruning algorithm while the sub-jets are obtained from the soft-drop algorithm.

b-Jet Identification

The identification of jets containing b hadrons is important to many analyses in associated with b-quark final state, which may serve as a probe for new physics. The ability to discriminate b jets from other gluons and light-flavor quarks (u,d,s) jet is due to the relatively large masses ($>$ 5.2 GeV) and long lifetimes ($\sim$ 1.5 ps) of the b hadrons. Moreover, b hadrons have higher charged decay multiplicity due to cascade decays.

The b-tagging algorithms takes the impact parameter (IP) significance of the tracks and the position of the secondary vertex (SV) as input parameters. The IP is the distance between the track and the PV at the point of closest approach. It is calculated in three dimensions by taking advantage of the high resolution of the pixel detector. The resolution of the IP depends strongly on the p_T and η of a track. The SV is where the b hadron decays, which can also be reconstructed by the CMS tracking system. The flight distance and direction between primary and secondary vertices are used to discriminate between b and non-b jets.

In this analysis, the b-jets are identified with the Combined Secondary Vertex (CSV) [47, 48] algorithm. This algorithm takes the secondary vertices, together with track-based lifetime information and provides discrimination even when no secondary vertices are found. The output of CSV algorithm is a discriminator value on which the user can cut on to select different regions in the efficiency versus purity phase space.

Reweighting of b-tagging Efficiency

In fact, the b-tagging efficiencies measured in data are different than those predicted by the simulation. In order to take this difference into account, weight factors need to be applied to simulated events. In this analysis, a method with event reweighting using b-tagging efficiency scale factors and MC b-tagging efficiency is used [49]. Note that only the b-tagging efficiency scale factors for b[2] and light flavor[3] jets are provided by BTV POG.

First, the b-tagging efficiencies of signal and background MC samples for each jet flavor are derived according to the jet p_T, which is defined as

$$\epsilon_f(p_T) = \frac{N_f^{\text{b-tagged}}(p_T)}{N_f^{\text{total}}(p_T)} \tag{3.7}$$

where $N_f^{total}(p_T)$ and $N_f^{b-tagged}(p_T)$ are the total number and the number of b-tagged jets, respectively, for each flavor in the p_T bin.

The event reweighting method predicts the correct event yield in data by changing the weight of the selected MC events. Moreover, the same event cannot contribute with different events weights to multiple b-tag multiplicity bins. Note that only the b-tagged jets[4] are considered in the reweighting.

The probability of a given configuration of jets in MC simulation and data is defined as

$$\begin{aligned} P(\text{MC}) &= \prod_{i=\text{tagged}} \epsilon_i \prod_{j=\text{not tagged}} (1-\epsilon_j) \\ P(\text{DATA}) &= \prod_{i=\text{tagged}} \text{SF}_i\epsilon_i \prod_{j=\text{not tagged}} (1-\text{SF}_j\epsilon_j) \end{aligned} \tag{3.8}$$

where ϵ_i is the MC b-tagging efficiency, and SF_i are the b-tagging efficiency scale factors. Note that both the ϵ_i and SF_i are functions of the jet flavor and jet p_T. In the end, the event weight w is calculated as

$$w = \frac{P(\text{DATA})}{P(\text{MC})} \tag{3.9}$$

[2]Scale factors for c jets are taken to be the same as for b jets.
[3]The d-, u-, s-quark and gluon are included in light flavor.
[4]The term "b-tagged jet" refers to a b-tagged jet of any flavor.

TABLE 3.9: Loose jet identification requirements for Spring15 25ns conditions.

PF Jet ID		Loose
Neutral Hadron Fraction	$<$	0.99
Neutral EM Fraction	$<$	0.99
Number of Constituents	$>$	1
Charged Hadron Fraction	$>$	0
Charged Multiplicity	$>$	0
Charged EM Fraction	$<$	0.99

3.4.6 Jet Selection

In this analysis, the jets are clustered by AK algorithm with a cone radius of R = 0.8. Furthermore, jets are required to pass the loose identification criteria [50] listed in Table 3.9.

3.4.7 Boson Reconstruction

The Z candidates are formed from oppositely charged lepton pairs of the same flavor, with invariant mass in the range [70,110] GeV. In addition, the p_T of Z is required to be greater than 200 GeV in this analysis.

The Higgs candidates are selected with pruned jet mass in the range [105,135] GeV and jet $p_T >$ 200 GeV. In order to discriminate against the light-flavored jets, events are classified according to the number of subjets (1 or 2) passing the b-tagging loose selection. A mass threshold for the high mass resonances is also applied with $M_X >$ 750 GeV. A full list of event pre-selection criteria are reported in Table 3.10.

3.5 Data and Monte Carlo Comparison

In this section, all the control plots at the pre-selection level are presented. The data/MC comparison are studied in the inclusive region and all of the weight factors mentioned above have already been applied to MC. The ratio of data/MC are consistent within uncertainty, which means the data and MC have a good agreement.

Figures 3.3, 3.4, and 3.5 show the distribution of ID variables separately for electron in barrel and endcap regions. Figures 3.6 and 3.7 show the distributions of ID variables separately for muon in high p_T and tracker categories. They are studied with $N-1$ cut procedure in order to present the full distribution of identification variables in events more similar to those used in the final analysis and to have an idea about the selection efficiency or this particular variable. Figures 3.8, 3.9, and 3.10 show the kinematic variables of the Z boson candidate and of the corresponding leading and sub-leading leptons, respectively. Figures 3.11 and 3.12 show the kinematic variables, the b-tagger discriminant, and the groomed mass of Higgs boson candidate. Figure 3.13 shows the subjets of the Higgs candidates separately for electron and muon channels.

TABLE 3.10: Summary of the pre-selection criteria in this analysis.

Electrons	$\|\eta\| < 2.5$
	remove $1.4442 < \|\eta\| < 1.566$
	HEEP NoIso ID
	rho-corrected mini Iso < 0.1
	Leading electron $p_T > 115$ GeV
	Sub-leading electron $p_T > 35$ GeV
	Two electrons have opposite charge
Muons	Custom Tracker Muon ID
	One of both pass High p_T Muon ID
	rho-corrected mini Iso < 0.2
	Muon $p_T > 20$ GeV
	Muon $\|\eta\| < 2.4$
	At least one muon $p_T > 50$ GeV
	At least one muon $\|\eta\| < 2.1$
	Two muons have opposite charge
Z bosons	70 GeV $< M_{ll} <$ 110 GeV
	$p_{Tll} > 200$ GeV
AK8 jets	$\|\eta\| < 2.4$
	$p_T > 200$ GeV
	FATjetPassIDLoose
	ΔR between jet and lepton > 0.8
	remove signal region: 65 GeV $<$ corrPRmass $<$ 135 GeV

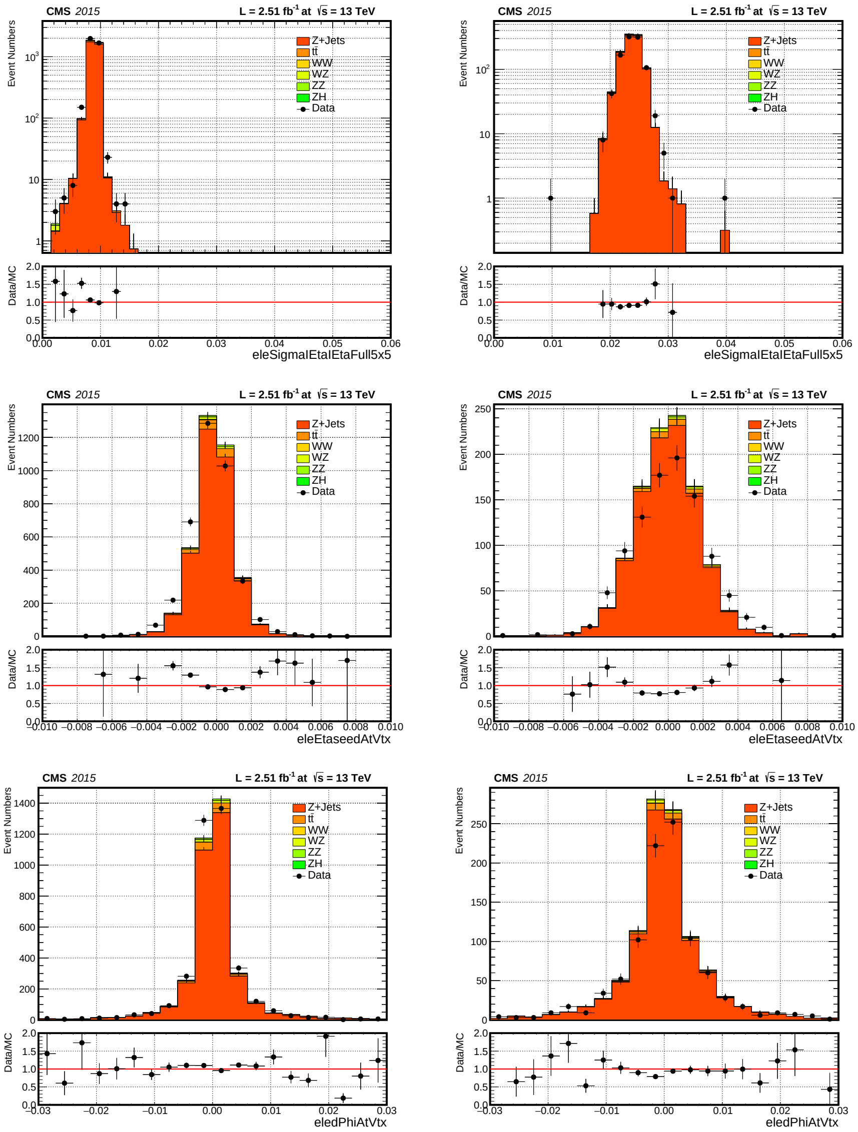

FIGURE 3.3: Distributions of the electron identification variables. From top to bottom: $\sigma_{i\eta i\eta}$, $|\Delta\eta_{in}|$, and $|\Delta\phi_{in}|$ for barrel (left) and endcaps (right).

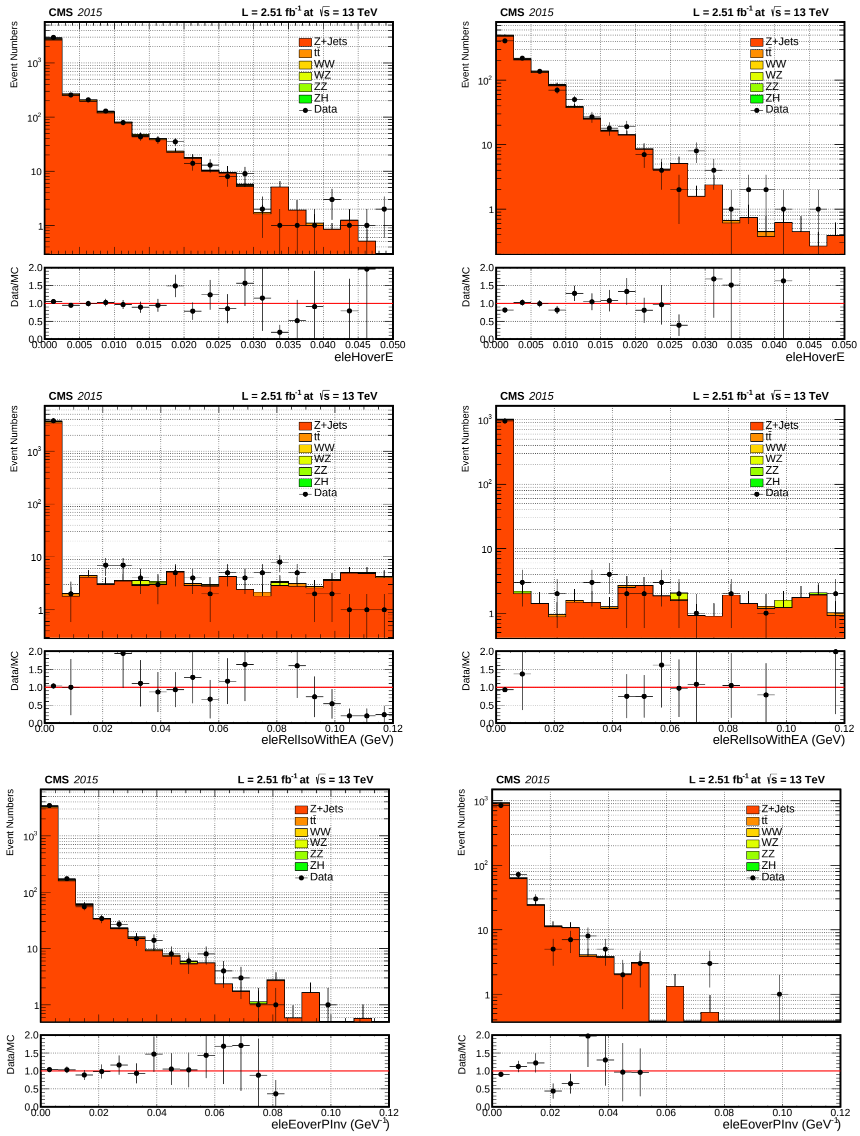

FIGURE 3.4: Distributions of the electron identification variables. From top to bottom: H/E, relIso(EA), and $1/E$-$1/p$ for barrel (left) and endcaps (right).

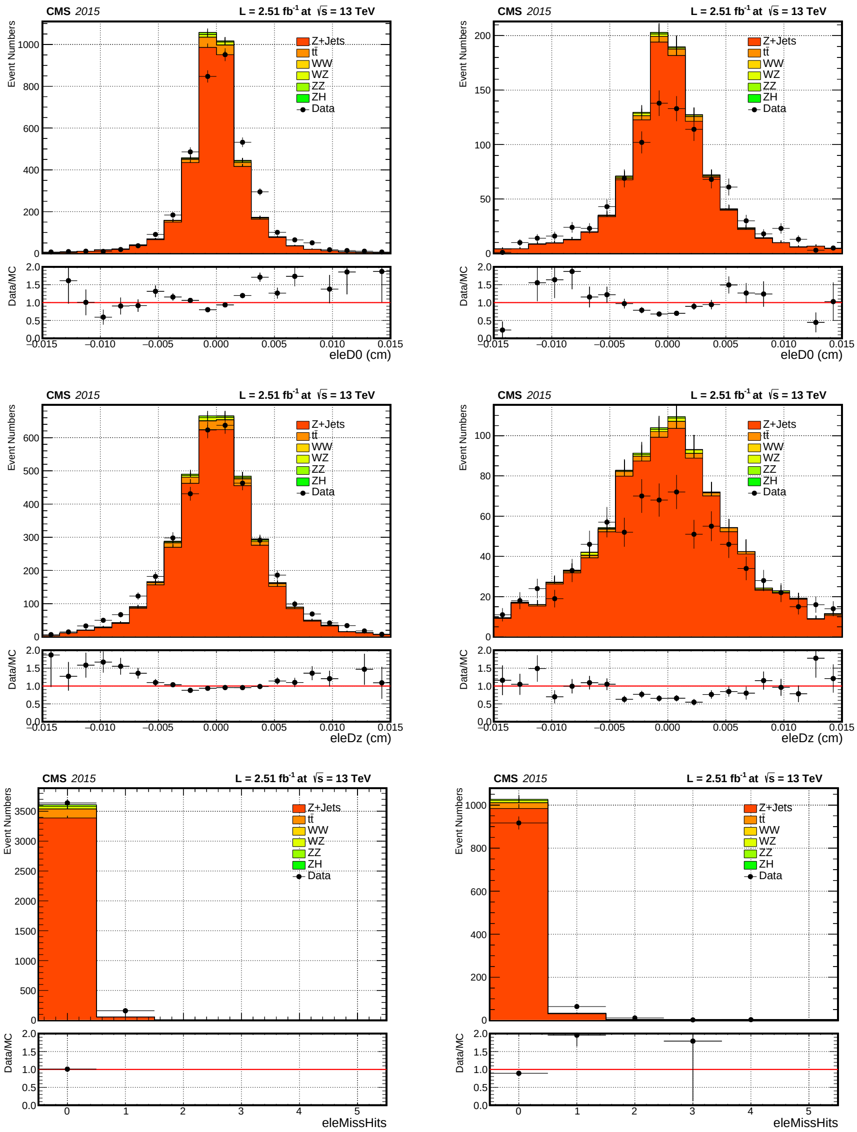

FIGURE 3.5: Distributions of the electron identification variables. From top to bottom: $|d_0|$, $|d_z|$, and number of missing hits in the pixel detector for barrel (left) and endcaps (right).

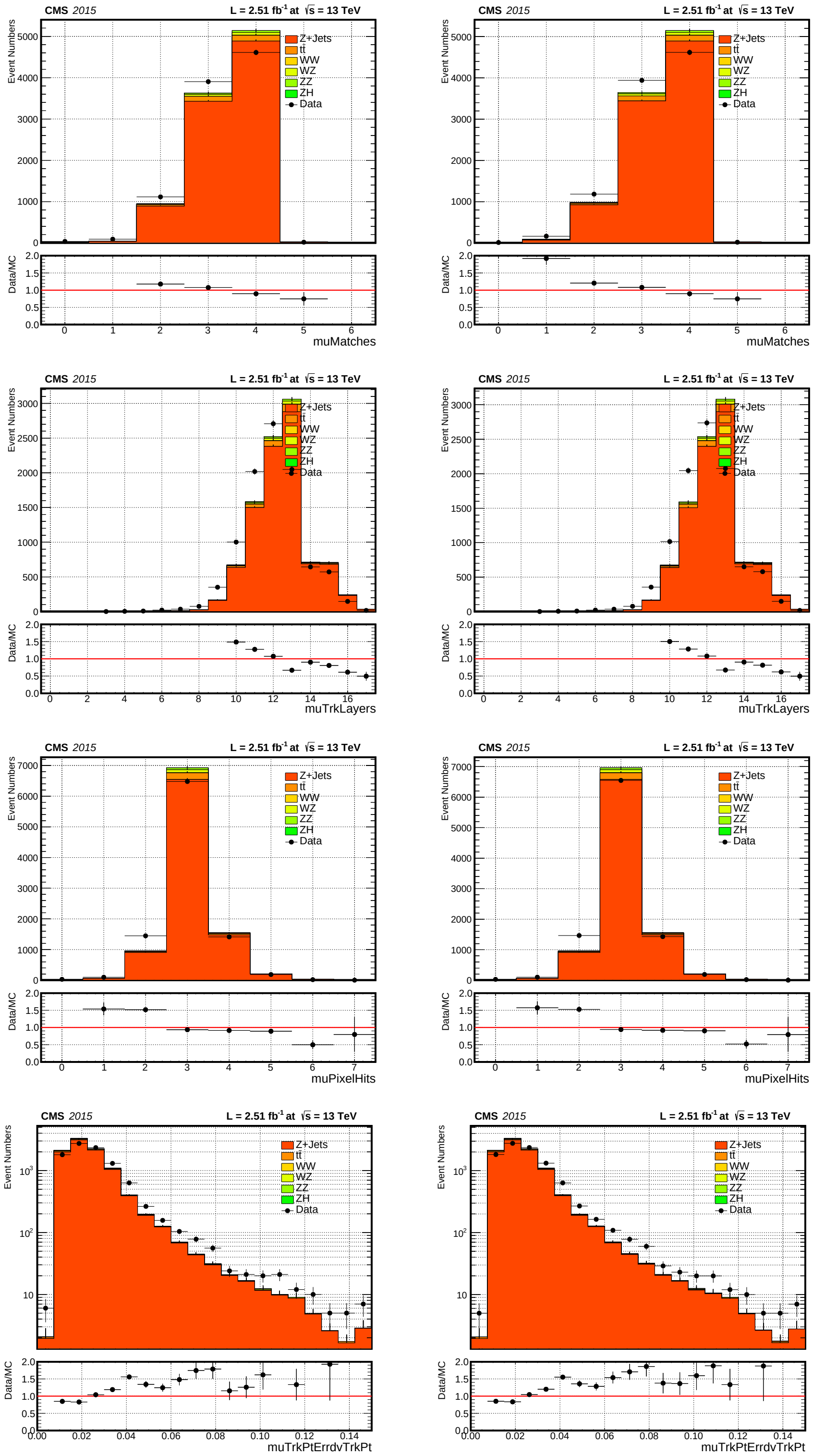

FIGURE 3.6: Distributions of the muon identification variables. From top to bottom: muMatches, muTrkLayers, muPixelHits, and muTrkPtErr/muTrkPt for high p_T category (left) and tracker category (right).

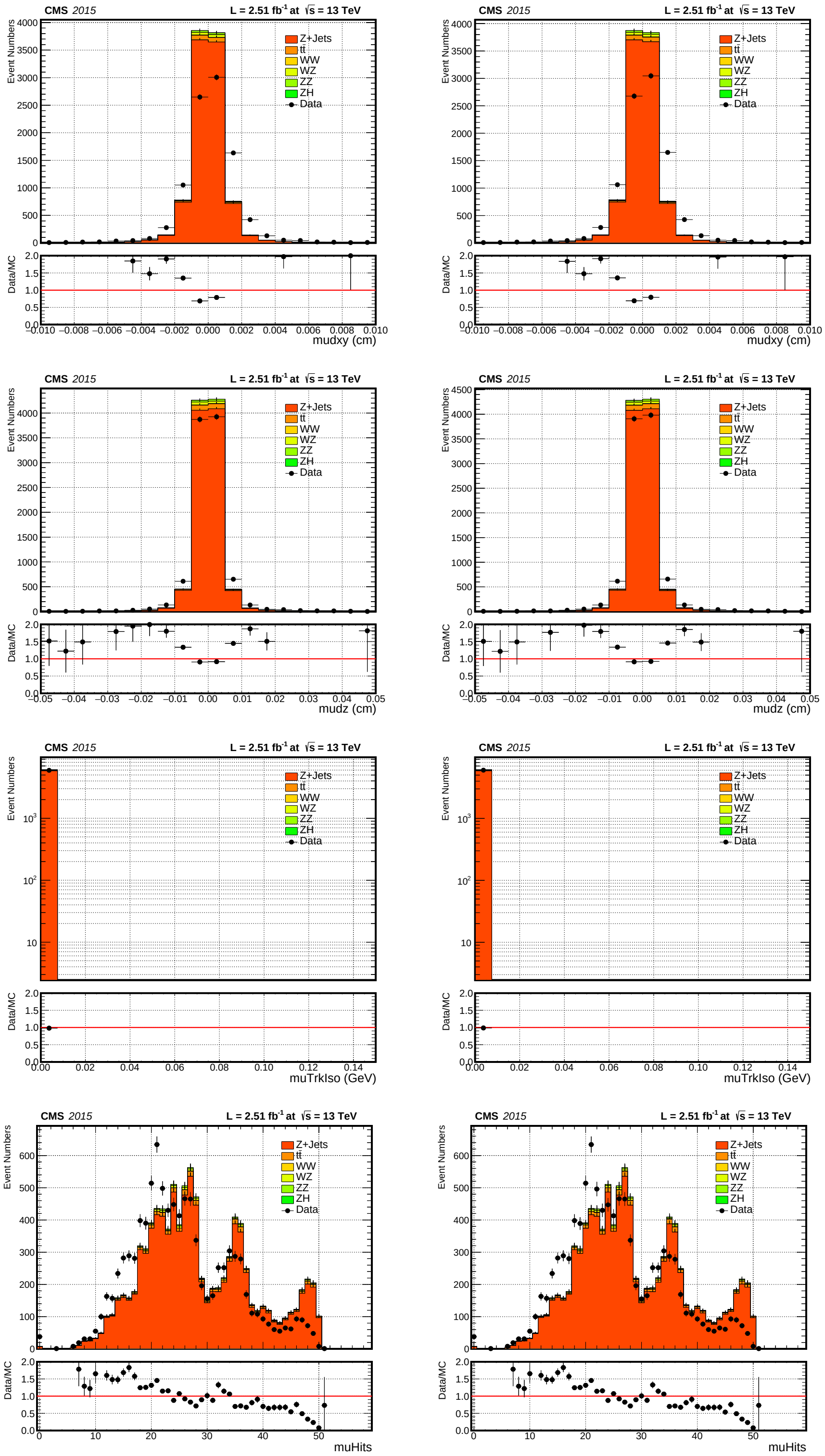

FIGURE 3.7: Distributions of the muon identification variables. From top to bottom: $|d_{xy}|$, $|d_z|$, muTrkIso, muHits for high p_T category (left) and tracker category (right).

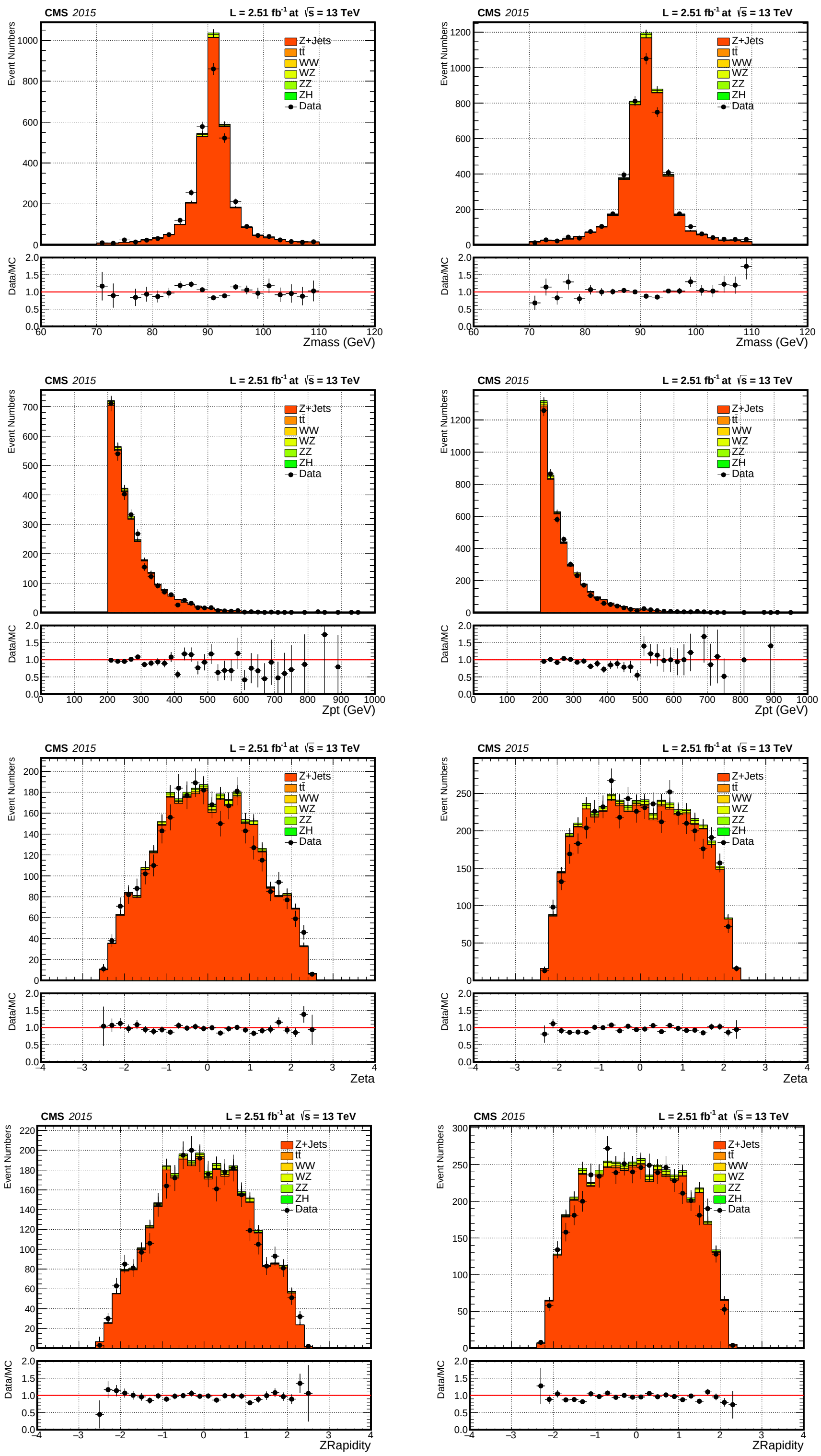

FIGURE 3.8: Distributions of mass, p_T, η, and rapidity variable of Z candidate in electron channel (left) and in muon channel (right).

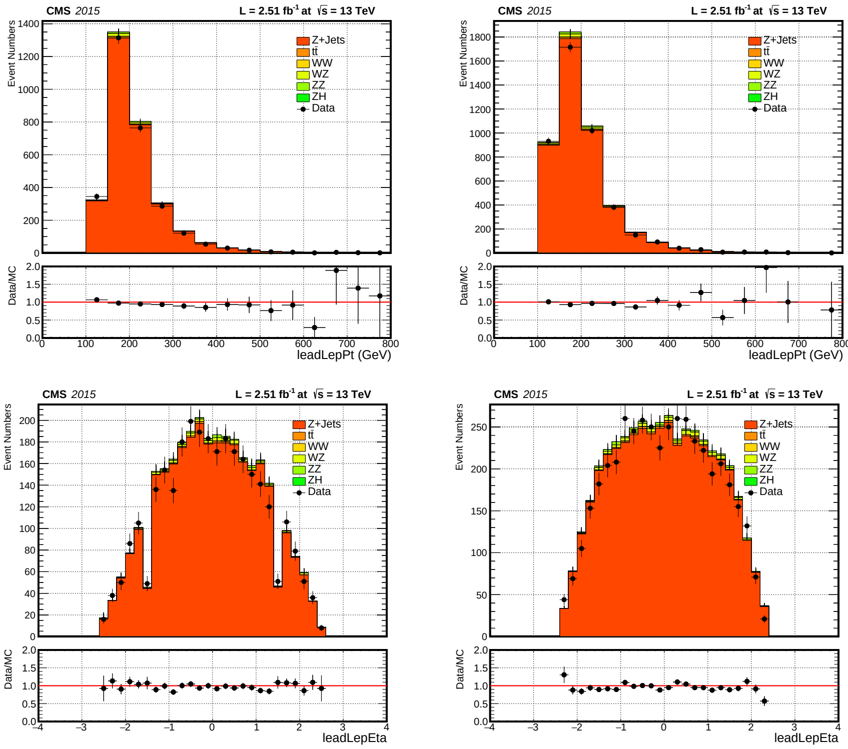

FIGURE 3.9: Distributions of p_T and η variable for the leading leptons of Z candidate in electron channel (left) and in muon channel (right).

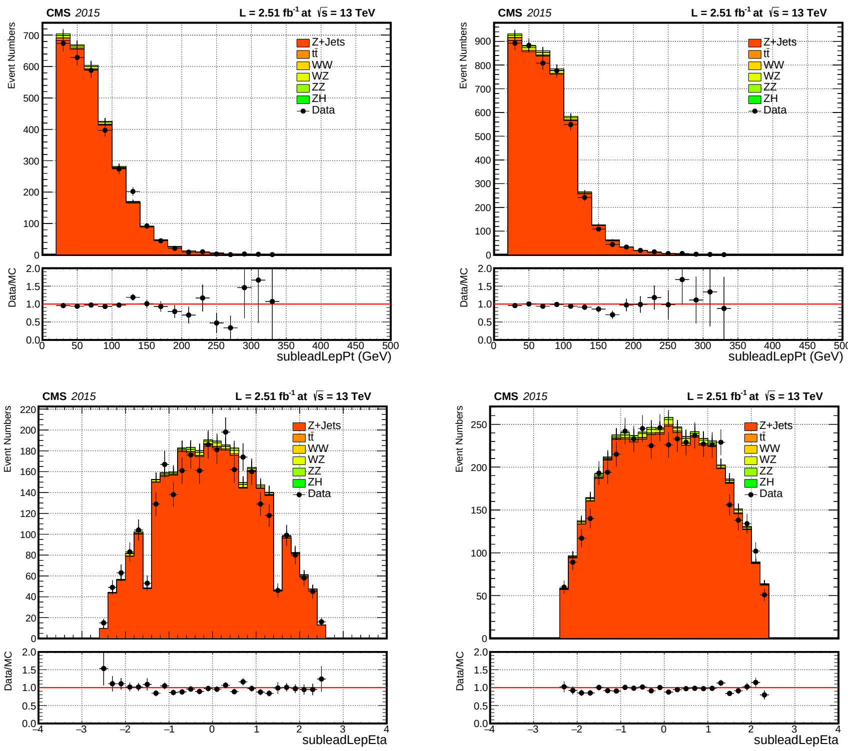

FIGURE 3.10: Distributions of p_T and η variable for the sub-leading leptons of Z candidate in electron channel (left) and in muon channel (right).

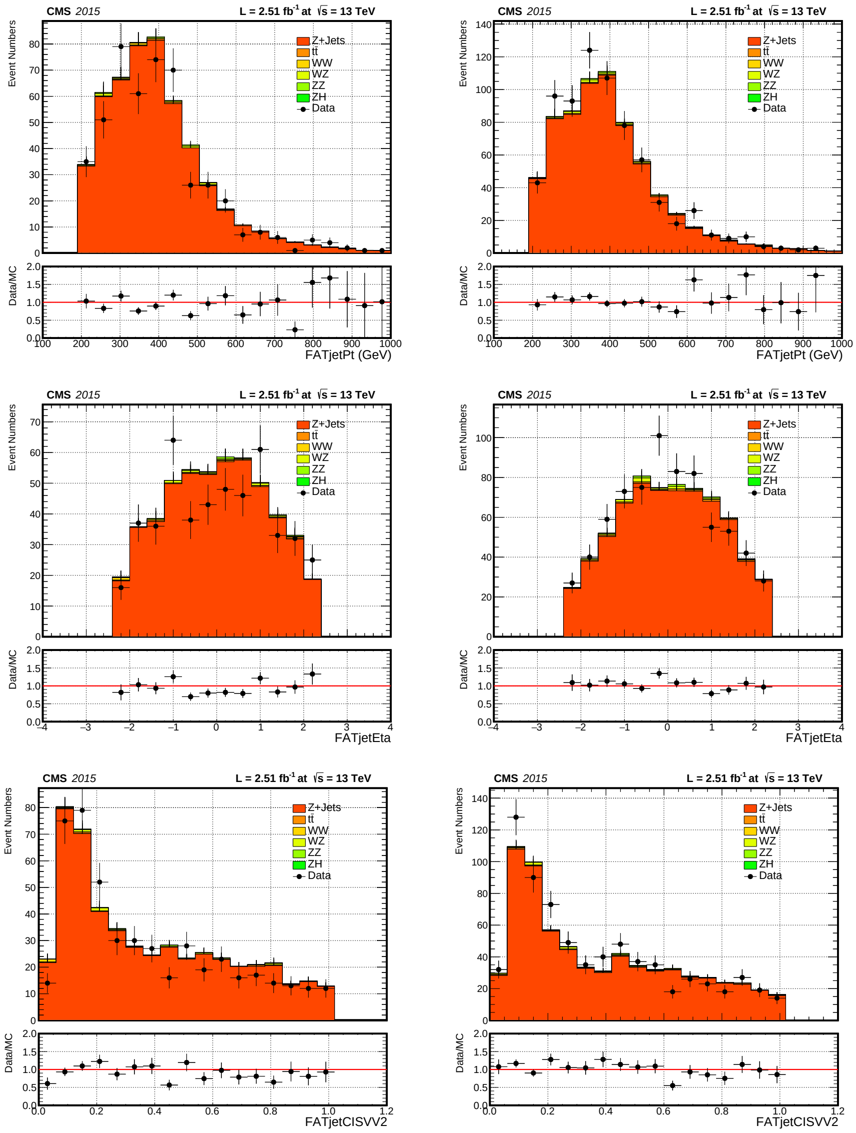

FIGURE 3.11: Distributions of p_T, η, and b-tagging variable of Higgs candidate in electron channel (left) and in muon channel (right).

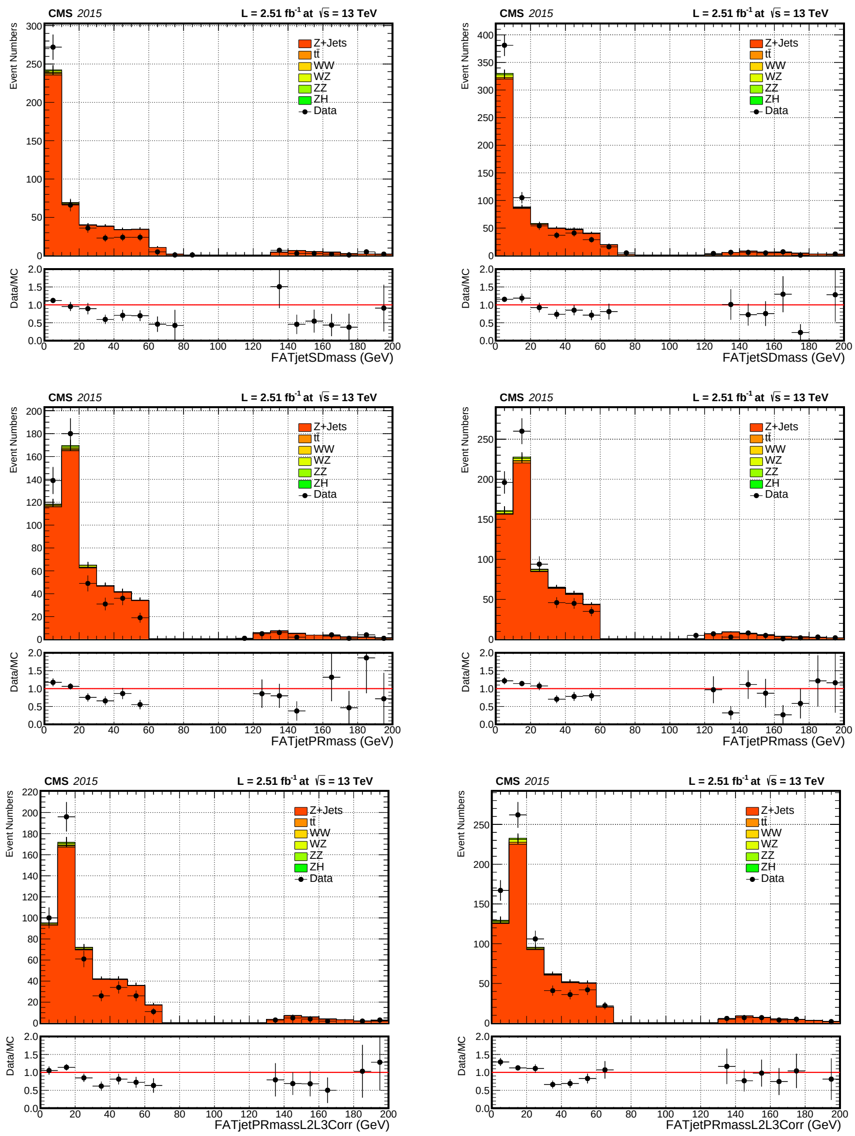

FIGURE 3.12: Distributions of jet mass after performing the softdrop and pruning grooming techniques for the electron channel (left) and the muon channel (right).

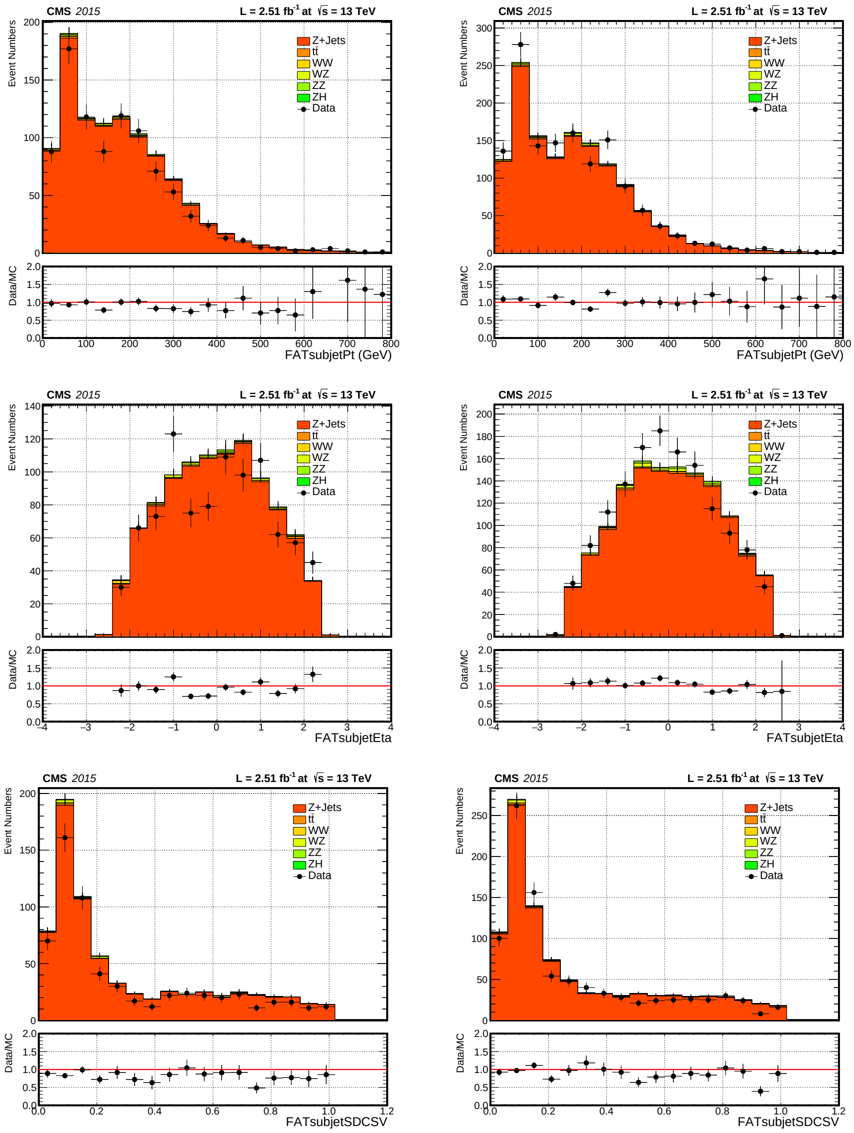

FIGURE 3.13: Distributions of p_T, η, and b-tagging variable of the subjets of Higgs candidate in electron channel (left) and in muon channel (right).

Chapter 4

Background Estimation

In this chapter, a background estimation strategy that helps in searching for localized excesses in the m_{ZH} spectrum is elaborated. The alpha method has been introduced as a more trustworthy strategy, because it is less dependent on the MC simulation. In fact, many systematic uncertainties that are hard to understand and control are believed to approximately cancel in alpha method.

First of all, the signal region and the side band are defined in order to select a signal-enriched or signal-depleted phase space. The corrected pruned jet mass m_j is chosen as the control variable of the two exclusive regions. In this analysis, the signal region (SR) is defined in the m_j range of [105,135] GeV, while the side band (SB) is defined in the m_j range of [30,65] GeV and [135,300] GeV. In addition, the W/Z mass window, in the m_j range of [65,105] GeV, is used neither as side band region nor as signal region in order to respect the blinding policy of the diboson VV searches.

In fact, the background is not purely composed of one physics process either in the signal region or in the side band. The background composition is assumed to be dominated by one single process (Z+jets) whose modeling in simulation is considered not to be trustworthy[1]. For other sub-dominant backgrounds ($t\bar{t}$, VV, ZH), it is safe to assume that if the MC predictions are wrong by 100%, the final effect on the upper limit is small since their contributions are smaller.

[1]The Z+jets samples used in this analysis were produced with PYTHIA8 parton shower, which have worse modeling in jet substructure. Moreover, the samples are LO.

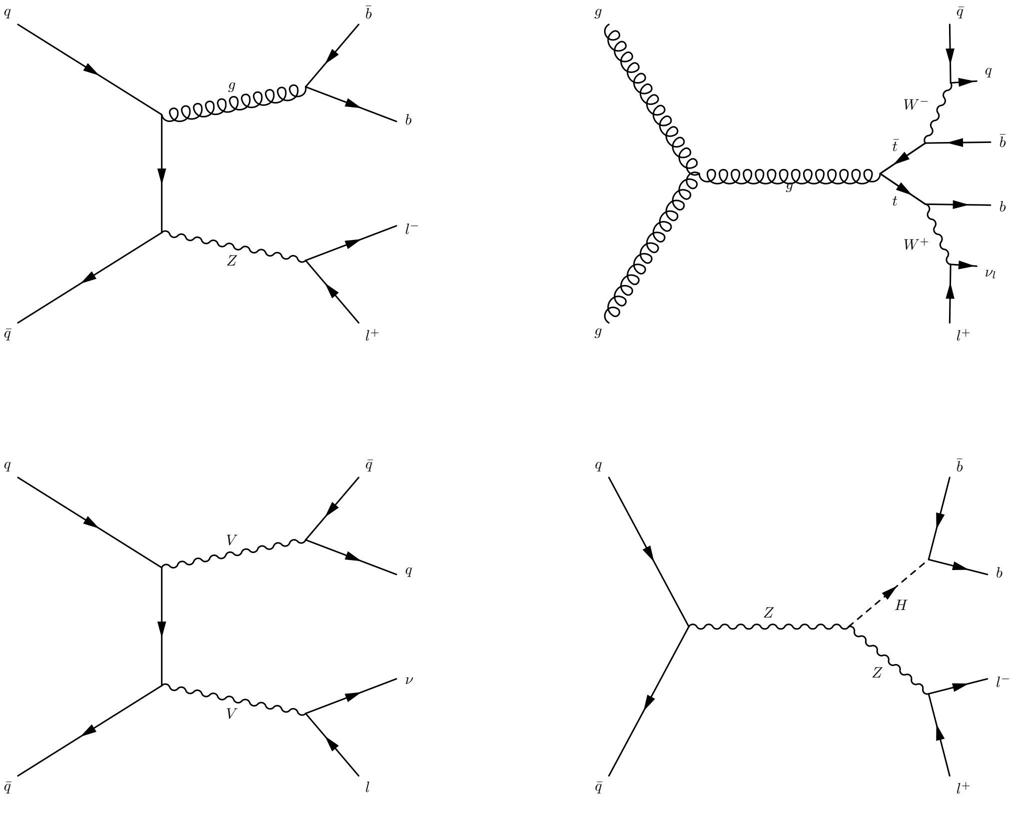

FIGURE 4.1: Representative Feymann Diagrams of Z+jets production (top left), $t\bar{t}$ production (top right), VV (V may represent $W^{\pm}$ or Z boson) production (bottom left), ZH production (bottom right) in SM.

The Feymann diagrams of dominant and sub-dominant processes are shown in Figure 4.1.

The normalization of predicted background in the signal region is interpolated from data in the side band. On the other hand, the shape of predicted background is extrapolated from the data in the side band to the signal region using a transfer function (the α-function) derived from simulation, assuming that the correlation between m_{ZH} and m_j is reasonably well reproduced by the MC.

This analysis is divided into electron and muon channels, with single and double b-tag categories. A different background prediction is derived for each category separately in order to reduce systematic uncertainties.

4.1 The Normalization of Z+jets Background

The number of expected background events in the signal region is determined from data in the side band. The background contributions are described with functional forms determined by fits to the MC, while the parameter values are determined from the fits to data. The dominant Z+jets background has physics properties of a smoothly falling background with no peaks. Hence, a simple exponential function (Eq. 4.1) is chosen to build the jet mass spectrum.

$$f(x) = p_0 \exp\left(-\frac{x}{p_1}\right) \tag{4.1}$$

Nevertheless, a flat model has been taken in double b-tag category in both electron and muon channel due to the lack of events. On the other hand, the sub-dominant backgrounds, which are taken directly from MC prediction, are considered to have 100% systematic uncertainty due to their small contribution.

As a result, the number of expected background events in the signal region is extracted through Eq. 4.2.

$$N_{\mathrm{SR}}^{\mathrm{data}} = N_{\mathrm{SB}}^{\mathrm{data}} \times \left[\frac{N_{\mathrm{SR}}^{\mathrm{fit\ data}}}{N_{\mathrm{SB}}^{\mathrm{fit\ data}}}\right] \tag{4.2}$$

where N are the number of events. Note that the number of expected background events is corrected due to a bias from a fitter test in this analysis, which will be discussed in detail in Section 5.1.

Figure 4.3 presents the m_j distribution in the Z+jets MC samples in the side band of different categories. Figure 4.4 presents the corrected pruned jet mass m_j in data side band in different categories.

4.2 The Shape of Z+jets Background in the m_{ZH} Spectrum

The shape of expected background in the signal region is extracted from the data in the side bands by α-function. The α-function is extracted by fitting the simulated m_{ZH} spectrum in the signal region and side band respectively. The m_{ZH} spectrum is parametrized with a modified exponential function which can model the exponential tails (Eq. 4.3).

$$f(x) = p_0 \exp\left(\frac{-x}{p_1 + p_2 x}\right) \tag{4.3}$$

where p_0, p_1 and p_2 are free parameters. Note that the sub-dominant backgrounds and data are also parametrized with Eq. 4.3.

As a result, the α-function for the dominant Z+jets background can be written as

$$\alpha(m_{ZH}) = \frac{N_{\text{SR}}^{\text{Zjets}}(m_{ZH})}{N_{\text{SB}}^{\text{Zjets}}(m_{ZH})} \tag{4.4}$$

In this manner, the total background prediction is extracted through

$$N_{\text{SR}}^{\text{data}}(m_{ZH}) = \left[N_{\text{SB}}^{\text{data}}(m_{ZH}) - N_{\text{SB}}^{\text{sub}}(m_{ZH})\right] \times \left[\frac{N_{\text{SR}}^{\text{Zjets}}(m_{ZH})}{N_{\text{SB}}^{\text{Zjets}}(m_{ZH})}\right] + N_{\text{SR}}^{\text{sub}}(m_{ZH}) \tag{4.5}$$

From Eq. 4.5, the sub-dominant contributions $N_{\text{SB}}^{\text{sub}}(m_{ZH})$ are subtracted from data in the SB first. The shape is then multiplied with the α-function in order to get the main background expectation in the signal region. By this way an assumption that the extrapolation from the side band to the signal region works in the same way for both data and MC is made. Lastly, the sub-dominant contributions $N_{\text{SR}}^{\text{sub}}(m_{ZH})$ in the signal region are added to the main background estimation.

Figure 4.2 shows the distributions of L2+L3-corrected pruned mass including the signal region. An excess of data over the background prediction at around

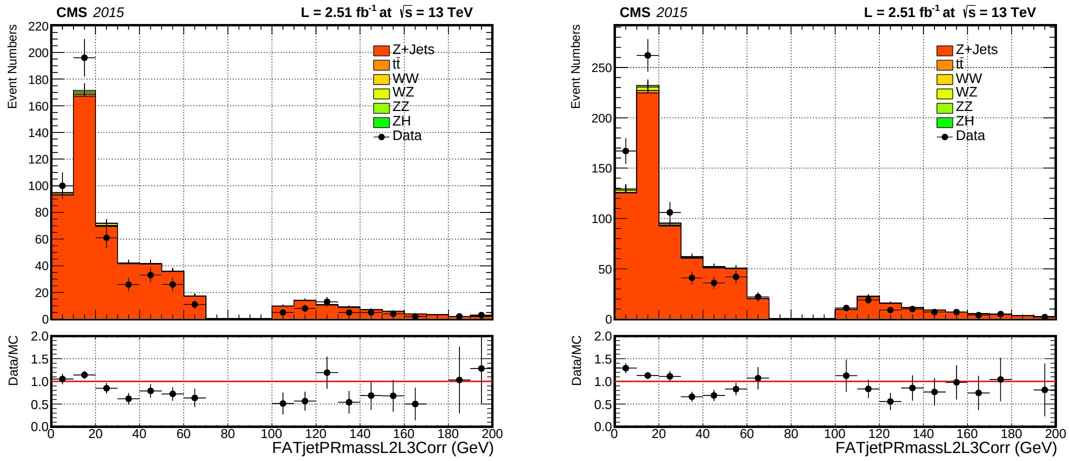

FIGURE 4.2: Distributions of jet mass after performing the pruning techniques including the signal region for the electron channel (left) and in muon channel (right).

125 GeV may indicate a hint of new physics.

Figures 4.5, 4.6, and 4.7 present the m_{ZH} distribution of Z+jets and sub-dominant backgrounds, respectively, in the side bands. Figures 4.8, 4.9, and 4.10 present the m_{ZH} distribution of Z+jets and sub-dominant backgrounds, respectively, in the signal region. Figure 4.11 shows the m_{ZH} distribution of data in the side bands. Figure 4.12 shows the α-function compared to the background shape in both side band and signal regions. Figure 4.13 presents the expected background with the data m_{ZH} distribution in the signal region. Tables 4.1 and 4.2 list the number of events observed in data and predicted background events for different minimum requirements on m_{ZH} .

TABLE 4.1: Number of events observed in data and predicted background events for different minimum requirements on m_{ZH} in electron channel.

	single b-tag		double b-tag	
m_{ZH} (>)	$N_{\mathbf{data}}$	$N_{\mathbf{bkg}}$	$N_{\mathbf{data}}$	$N_{\mathbf{bkg}}$
800	5 ± stat.unc.	8.86 ± syst.unc.	1 ± stat.unc.	5.49 ± syst.unc.
1000	1 ± stat.unc.	5.96 ± syst.unc.	1 ± stat.unc.	2.71 ± syst.unc.
1200	1 ± stat.unc.	4.00 ± syst.unc.	0 ± stat.unc.	1.35 ± syst.unc.
1400	1 ± stat.unc.	2.69 ± syst.unc.	0 ± stat.unc.	0.67 ± syst.unc.
1600	1 ± stat.unc.	1.80 ± syst.unc.	0 ± stat.unc.	0.33 ± syst.unc.
1800	0 ± stat.unc.	1.21 ± syst.unc.	0 ± stat.unc.	0.17 ± syst.unc.
2000	0 ± stat.unc.	0.81 ± syst.unc.	0 ± stat.unc.	0.08 ± syst.unc.

TABLE 4.2: Number of events observed in data and predicted background events for different minimum requirements on m_{ZH} in muon channel.

	single b-tag		double b-tag	
m_{ZH} (>)	$N_{\mathbf{data}}$	$N_{\mathbf{bkg}}$	$N_{\mathbf{data}}$	$N_{\mathbf{bkg}}$
800	15 ± stat.unc.	12.01 ± syst.unc.	1 ± stat.unc.	5.66 ± syst.unc.
1000	10 ± stat.unc.	4.79 ± syst.unc.	1 ± stat.unc.	3.65 ± syst.unc.
1200	8 ± stat.unc.	1.91 ± syst.unc.	0 ± stat.unc.	2.37 ± syst.unc.
1400	4 ± stat.unc.	0.76 ± syst.unc.	0 ± stat.unc.	1.54 ± syst.unc.
1600	1 ± stat.unc.	0.30 ± syst.unc.	0 ± stat.unc.	1.00 ± syst.unc.
1800	0 ± stat.unc.	0.12 ± syst.unc.	0 ± stat.unc.	0.65 ± syst.unc.
2000	0 ± stat.unc.	0.04 ± syst.unc.	0 ± stat.unc.	0.42 ± syst.unc.

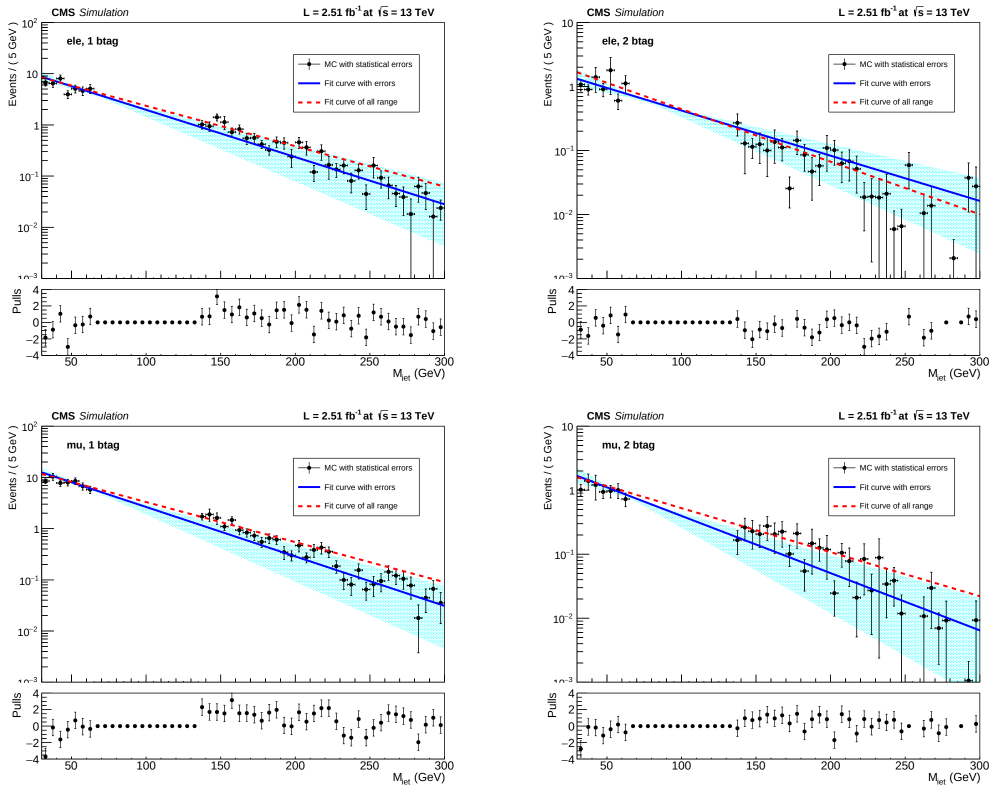

FIGURE 4.3: Fits to simulated m_j in Z+jets in the side band. The red dashed line represented the fit when the signal region is turned on. The blue 1-σ uncertainty bands are coming from the propagation of errors of fit parameters. (Top: electron channel. Bottom: muon channel. Left: single b-tag. Right: double b-tag.)

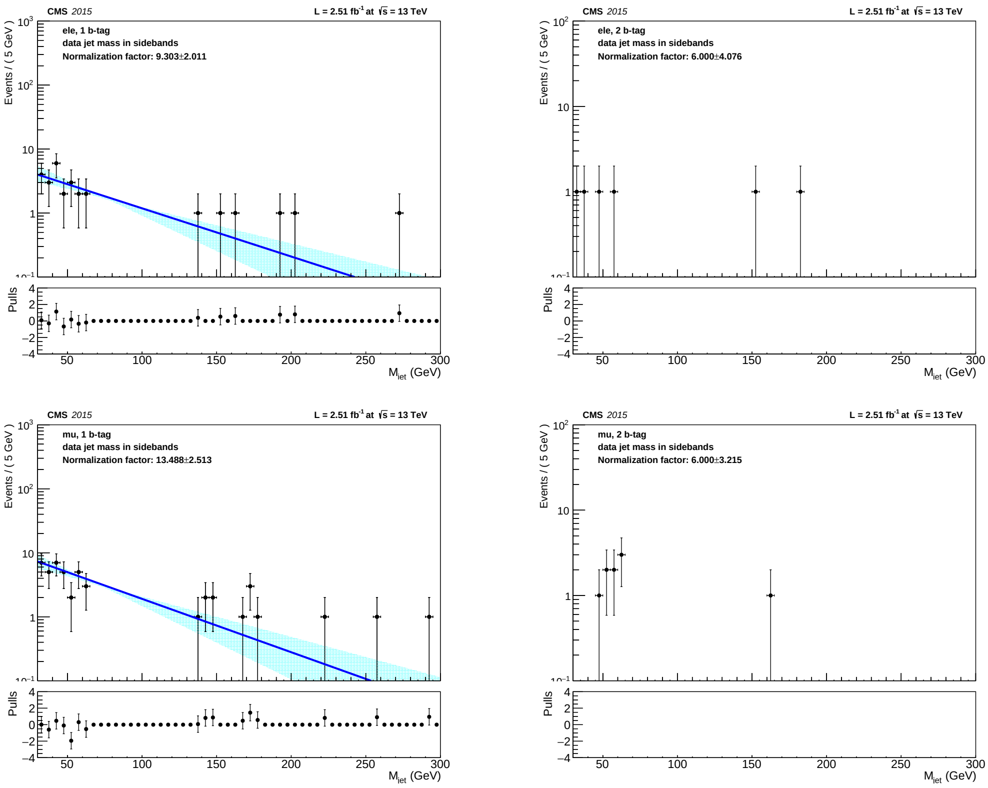

FIGURE 4.4: Fits to data corrected pruned jet mass in the side band. The blue 1-σ uncertainty bands are coming from the propagation of errors of fit parameters.(Top: electron channel. Bottom: muon channel. Left: single b-tag. Right: double b-tag.)

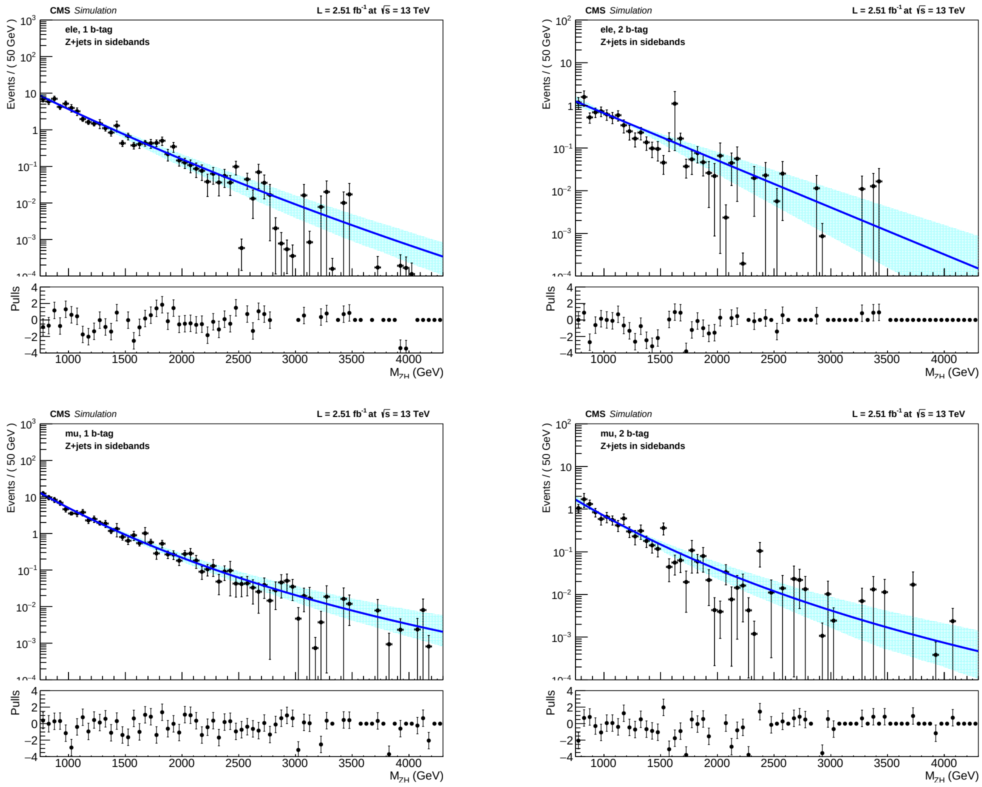

FIGURE 4.5: Fits to simulated m_{ZH} in Z+jets in the side band. The blue 1-σ uncertainty bands are coming from the propagation of errors of fit parameters. (Top: electron channel. Bottom: muon channel. Left: single b-tag. Right: double b-tag.)

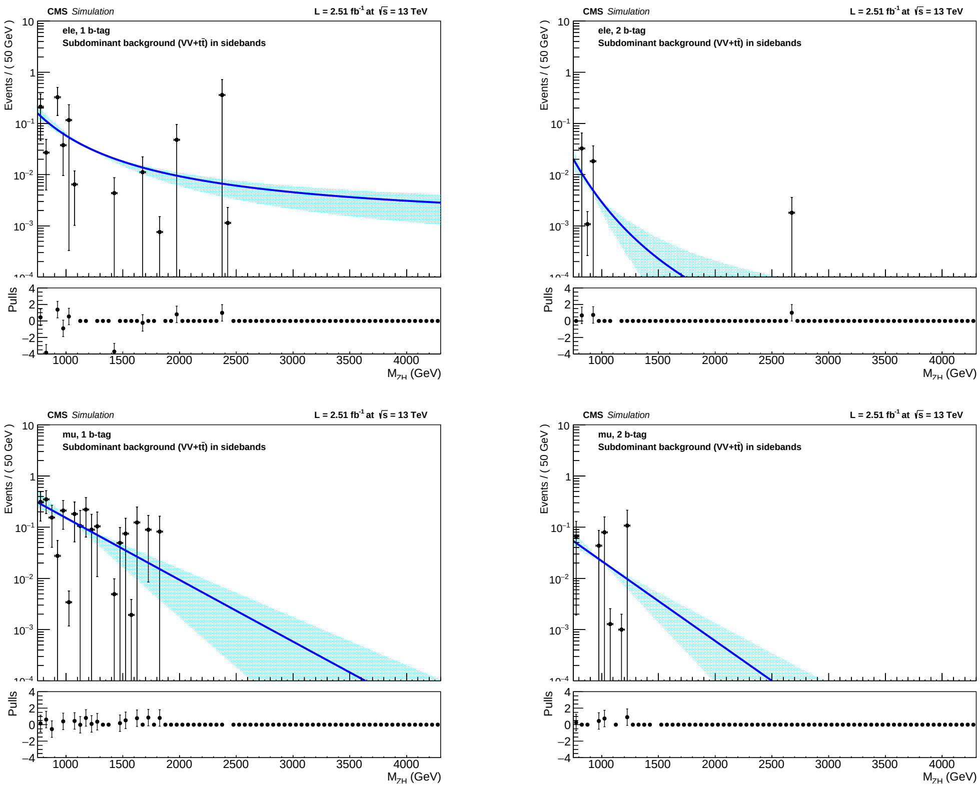

FIGURE 4.6: Fits to simulated m_{ZH} in WW, WZ, ZZ, and $t\bar{t}$ in the side band. The blue 1-σ uncertainty bands are coming from the propagation of errors of fit parameters. (Top: electron channel. Bottom: muon channel. Left: single b-tag. Right: double b-tag.)

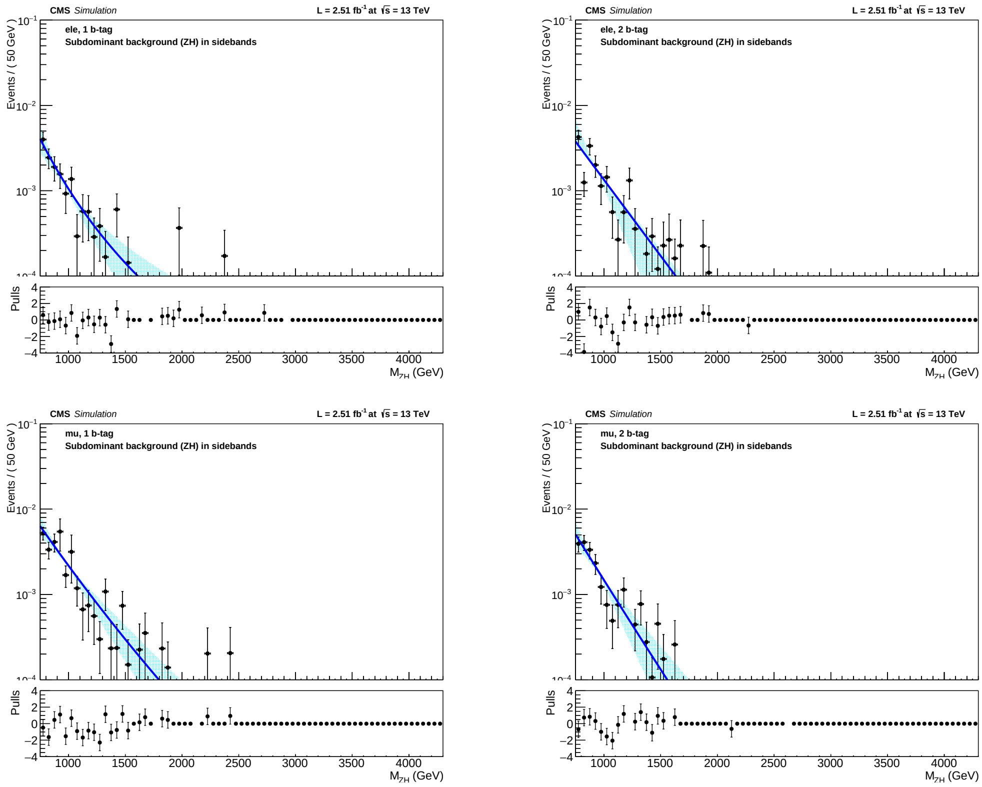

FIGURE 4.7: Fits to simulated m_{ZH} in ZH in the side band. The blue 1-σ uncertainty bands are coming from the propagation of errors of fit parameters. (Top: electron channel. Bottom: muon channel. Left: single b-tag. Right: double b-tag.)

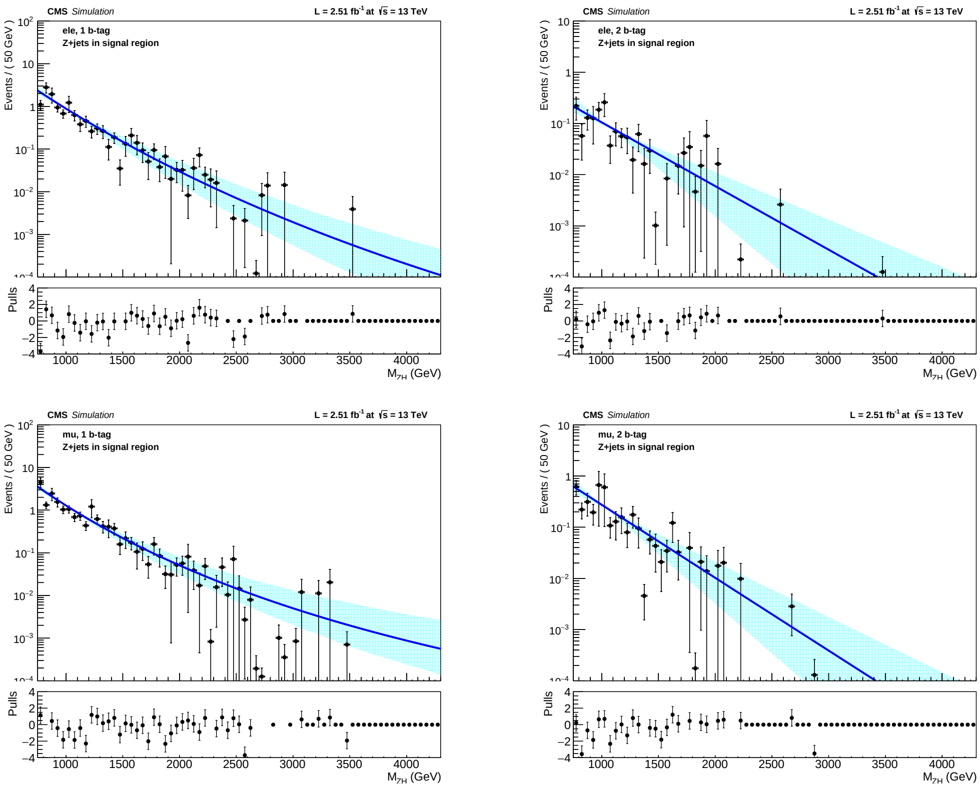

FIGURE 4.8: Fits to simulated m_{ZH} in Z+jets in the signal region. The blue 1-σ uncertainty bands are coming from the propagation of errors of fit parameters. (Top: electron channel. Bottom: muon channel. Left: single b-tag. Right: double b-tag.)

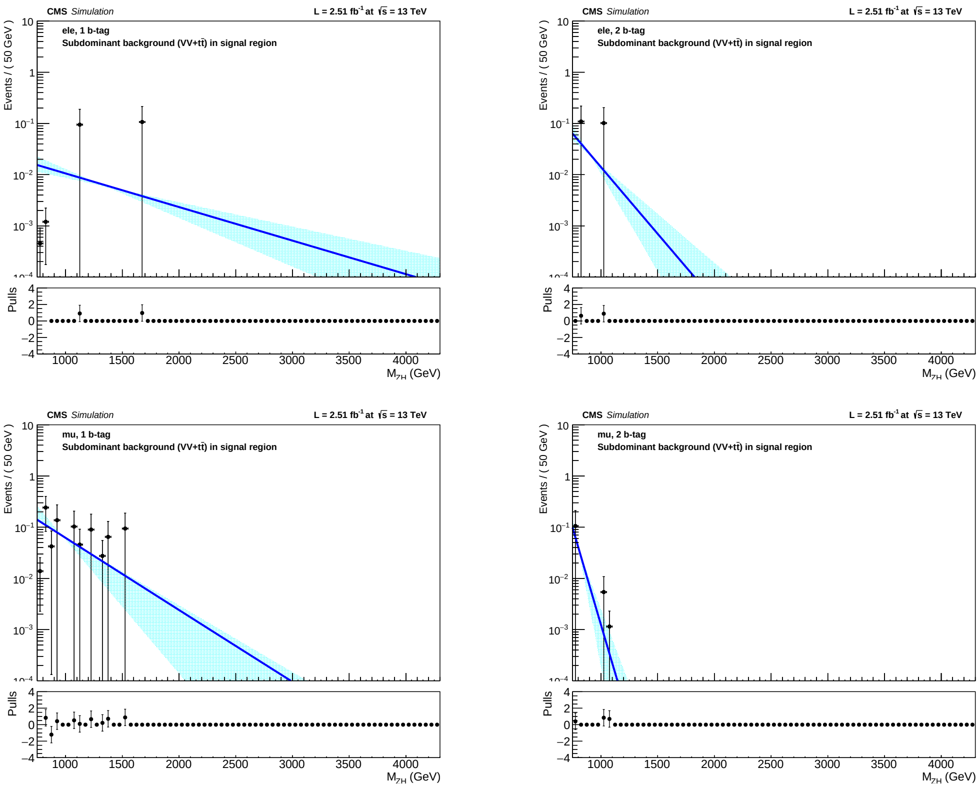

FIGURE 4.9: Fits to simulated m_{ZH} in WW, WZ, ZZ, and $t\bar{t}$ in the signal region. The blue 1-σ uncertainty bands are coming from the propagation of errors of fit parameters. (Top: electron channel. Bottom: muon channel. Left: single b-tag. Right: double b-tag.)

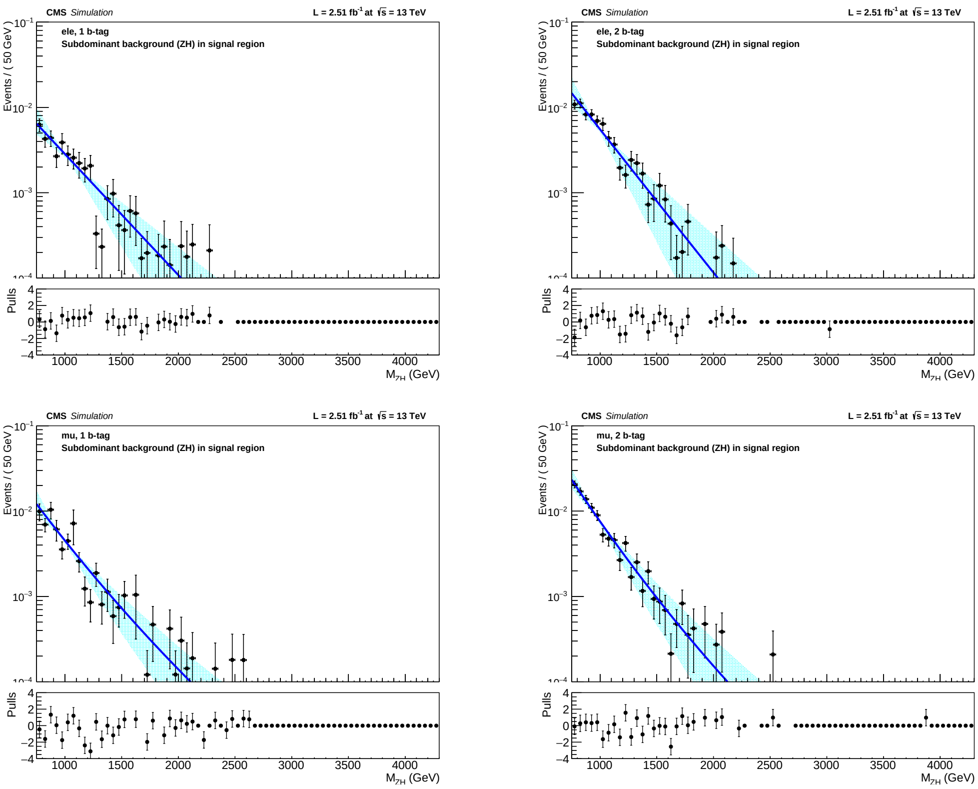

FIGURE 4.10: Fits to simulated m_{ZH} in ZH in the signal region. The blue 1-σ uncertainty bands are coming from the propagation of errors of fit parameters. (Top: electron channel. Bottom: muon channel. Left: single b-tag. Right: double b-tag.)

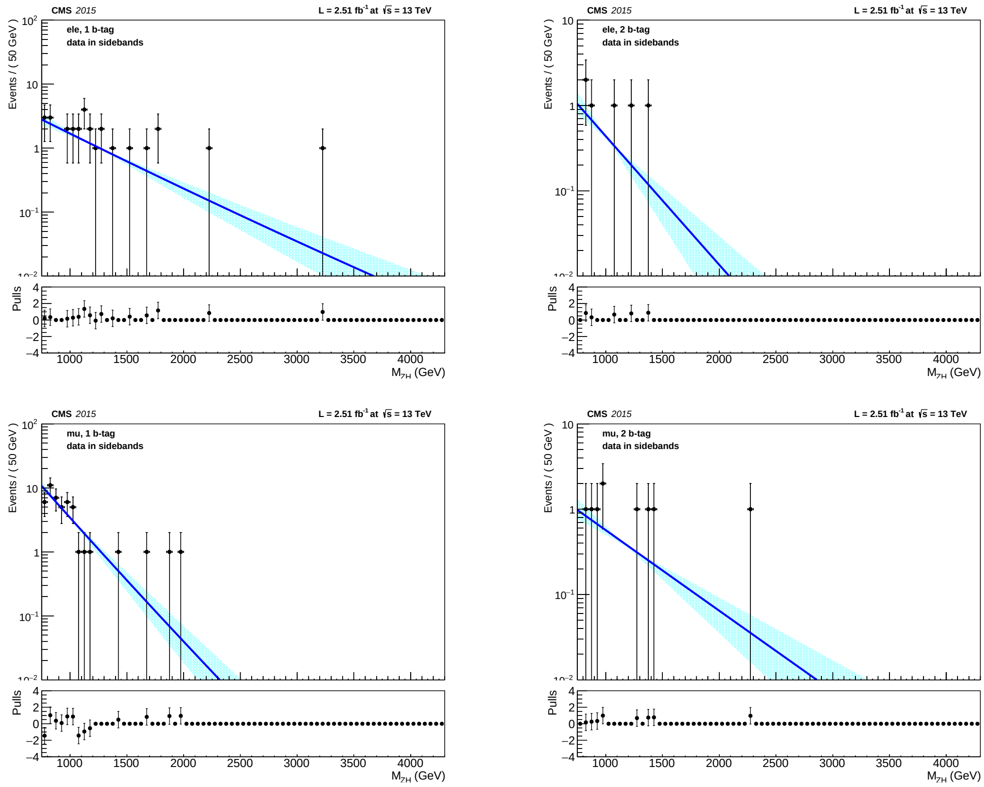

FIGURE 4.11: Fits to data m_{ZH} in the side band. The blue 1-σ uncertainty bands are coming from the propagation of errors of fit parameters. (Top: electron channel. Bottom: muon channel. Left: single b-tag. Right: double b-tag.)

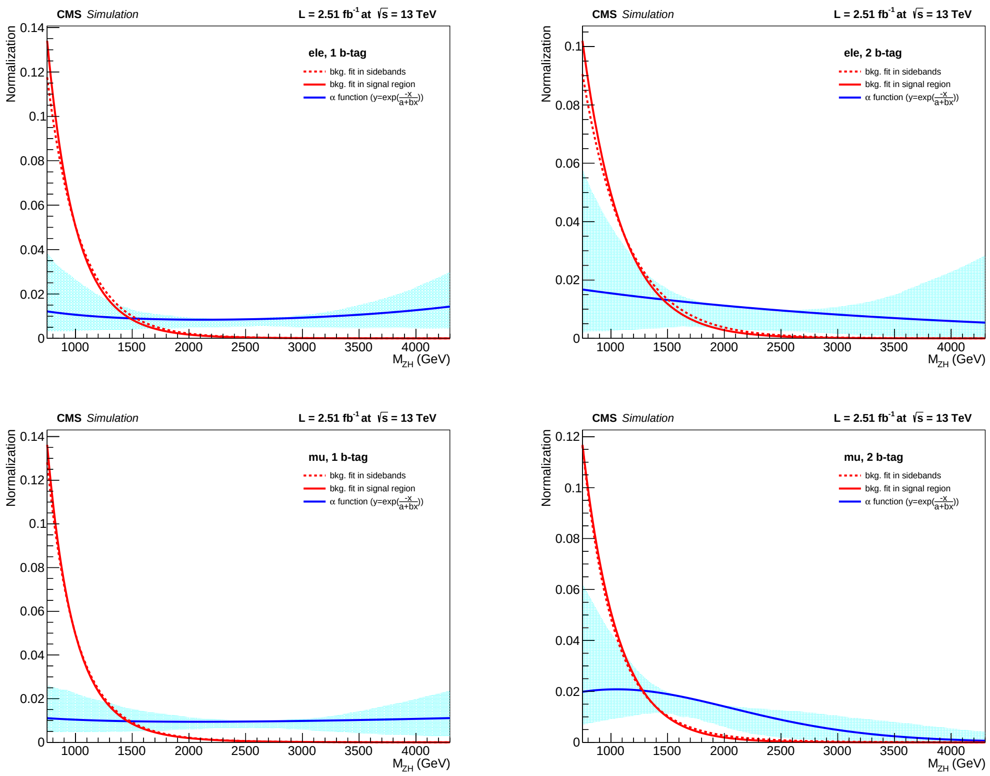

FIGURE 4.12: The α function (blue line) compared to the background shape in both side band and signal region. The red dashed and solid lines represent the estimated background in the side band and signal region respectively. The blue 1-σ uncertainty bands are coming from the propagation of errors of fit parameters. (Top: electron channel. Bottom: muon channel. Left: single b-tag. Right: double b-tag.)

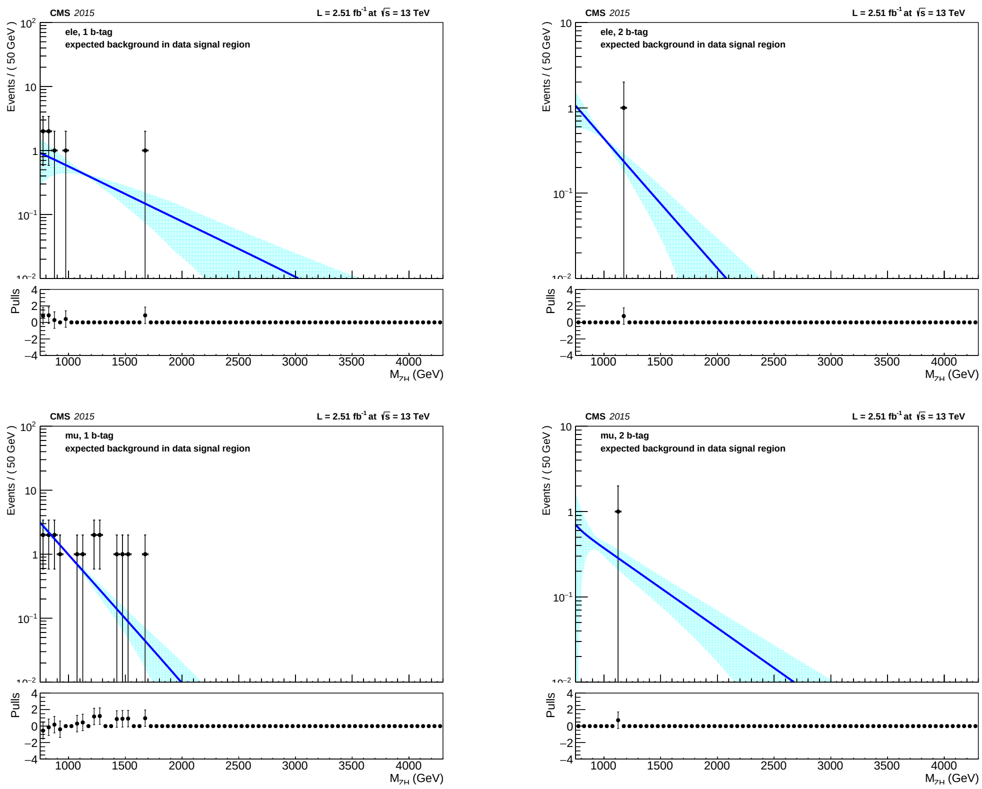

FIGURE 4.13: Expected background with the data m_{ZH} distribution in the signal region. The blue line is the data-driven background estimation, superimpose on the data with error bars. The blue 1-σ uncertainty bands are coming from the propagation of errors of fit parameters. (Top: electron channel. Bottom: muon channel. Left: single b-tag. Right: double b-tag.)

Chapter 5

Systematic Uncertainties

There are several sources of systematic uncertainties that affect the final background prediction (normalization and shape) and signal efficiency. Those uncertainties affect the results of the upper limits on production cross section (Chapter 6). In the following sections, a brief introduction of the sources of systematic uncertainties and the method of estimation are presented.

5.1 Systematics on Background Normalization

The number of predicted background events in the signal region is extracted from the fits to the distribution of corrected pruned jet mass m_j, as described in Section 4.1. Therefore, the propagated error from the fit parameters contributes as a source of systematic uncertainty on normalization. Due to the small sample size in this analysis, the propagation of errors from the fit parameters has the largest contribution among all the sources of systematic uncertainty.

Propagation of errors of fit parameters

The errors from the fit parameters are propagated using the correlation matrix $\mathbf{M}_C$:

$$\sigma_{\text{fit}}^2 = \mathbf{A}^{\mathrm{T}}(\mathbf{M}_C \mathbf{A}) \tag{5.1}$$

where the element $A_{i,1}$ in the column matrix $\mathbf{A}$ is the variation on the number of background events after varying independently the fit parameter p_i by $\pm 1\sigma$.

Study of the fit bias

The unbinned likelihood method provided by the `RooFit` [51] toolkit is chosen as the fit method in this analysis. The toy MC method is used to examine any possible fit bias from the fitter.

In the toy MC method, 1000 randomized distributions of corrected pruned jet mass m_j in side band are generated using a template model of corrected pruned jet mass m_j in MC[1]. Each generated distribution is fitted with the same formula as described in Eq. 4.1. Hence, the number of expected background events from the fitted toy MC is extracted through same formula as described in Eq. 4.2. As a result, the fit bias is determined from the distributions of bias and pull values.

The bias is defined as

$$\text{bias} = \frac{N_{\text{bkg}}^{\text{toyMC}} - N_{\text{bkg}}^{\text{MC}}}{N_{\text{bkg}}^{\text{MC}}} \tag{5.2}$$

where $N_{\text{bkg}}^{\text{toyMC}}$ is the number of background events determined from the fit to toy MC, $N_{\text{bkg}}^{\text{MC}}$ is the number of background events determined from the fit to template MC. The systematic shift due to the imperfection of the fitter is then deduced from the mean value of the bias distribution. In this analysis, the mean of bias in each category is found to be negative, indicating that the expected number of background events is underestimated. Therefore, the bias as a correction is applied to the number of background events.

The pull is defined as

$$\text{pull} = \frac{N_{\text{bkg}}^{\text{toyMC}} - N_{\text{bkg}}^{\text{MC}}}{\sigma_{\text{fit}}} \tag{5.3}$$

[1]The template model is derived from the blue line in Figure 4.3. Then, the SG is removed in each toy distribution.

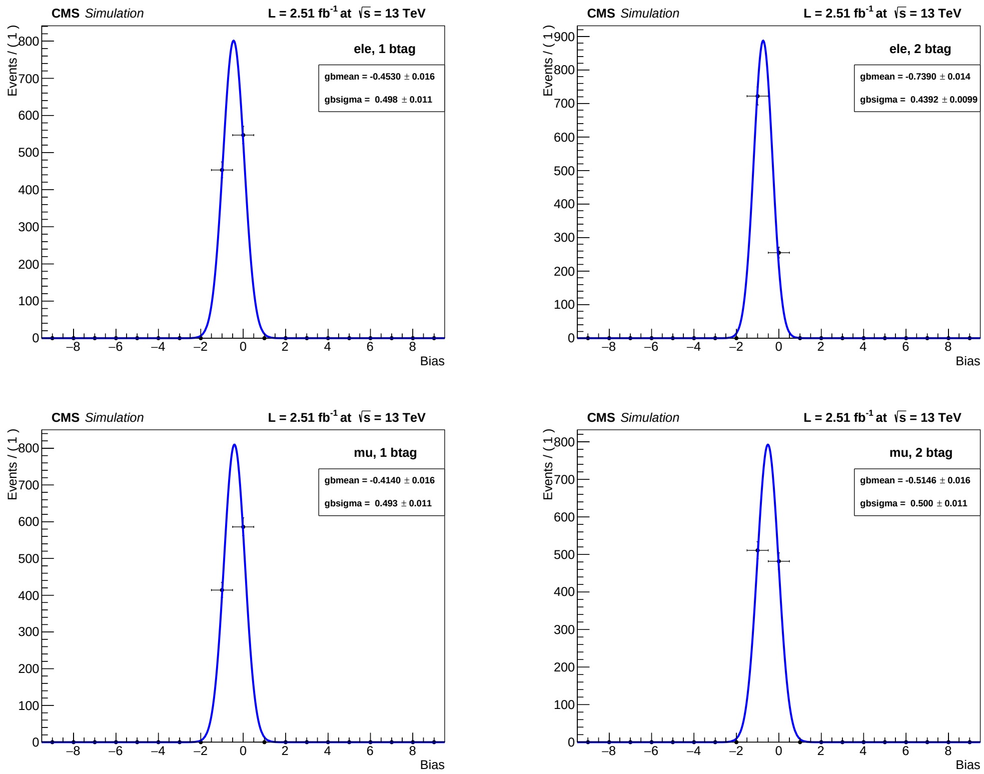

FIGURE 5.1: The distributions of fit bias for the N_{bkg} determined from fit to the toy MC. (Top:electron channel. Bottom: muon channel. Left: single b-tag. Right: double b-tag.)

where the uncertainty σ_{fit} is derived from the propagation of errors of fit parameters. The correctness of the fit errors are then deduced from the width of pull distribution. In this analysis, the width of pull in each category is large than 1, indicating that the fit errors are underestimated. As a result, the width of the pull distribution is added to the propagation of errors of fit parameters on predicted background as an additional uncertainty that affects the normalization.

Figure 5.1 and 5.2 report the bias and pull distribution accordingly in different categories.

Other sources

The jet energy scale factor, which will be described later, is also a source of systematic uncertainty that can affect the background normalization. Apart from this, the sub-dominant backgrounds ($t\bar{t}$, VV, and ZH) are subtracted directly

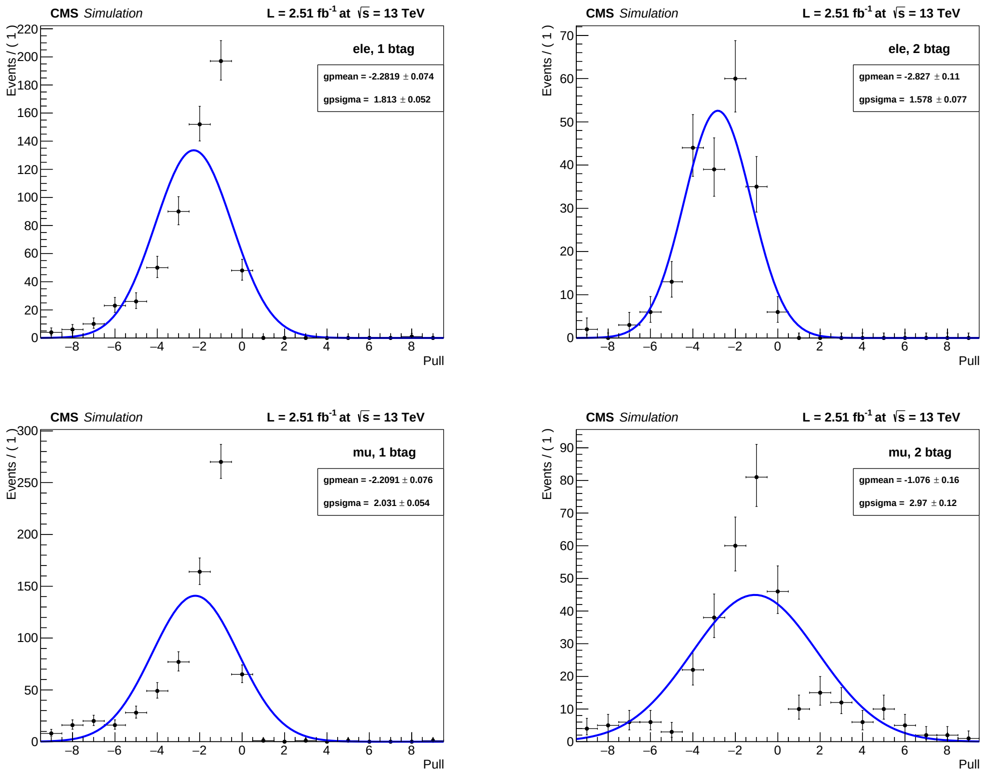

FIGURE 5.2: The pull distributions of N_{bkg} determined from the fit to the toy MC. (Top:electron channel. Bottom: muon channel. Left: single b-tag. Right: double b-tag.)

from data according to MC predictions in this analysis, due to their low statistics. In this case, a 100% uncertainty from the rates of sub-dominant backgrounds is assigned as a systematic uncertainty on total amount of background contribution.

In addition, the uncertainty on the measurement of luminosity in LHC and in CMS also affects the observed cross section limit and the background normalization in MC for the sub-dominant backgrounds. The recommended uncertainty on the integrated luminosity is 2.7% by the Luminosity POG [52].

5.2 Systematics on Background Shape

There are several sources of systematic uncertainty that affect the predicted background shape. By shifting one standard deviation on the sources, the uncertainties are propagated to the α function. As a results, the uncertainties vary as a function of resonance mass.

b-Tagging efficiency scale

The b-tagging scale factor uncertainties are provided by the BTV POG [49] (3.4.5). The uncertainty is evaluated by moving the scale factors up and down by one standard deviation simultaneously for each jet as a function of p_T and η.

Jet energy scale

The jet energy measured by the CMS HCAL can only have an accuracy of 60-70% due to the nature of the hadronic shower. Hence, a calibration of jet energy is performed using γ/Z+jets control samples with the help of high performance ECAL in CMS [53]. The scale factor from the jet energy corrections give errors and propagate to the 4-momentum and the mass of jets both in data and in MC.

QCD factorization and renormalization scale

The QCD factorization scale corresponds to the scale that separates the high-energy perturbative and the low-energy non-perturbative regimes. The QCD renormalization scale is used for dimensional regularization of UV divergence. The QCD factorization scale uncertainties are estimated by varying the corresponding scales up and down by a factor 2. The factorization scale uncertainties are propagated as a shape uncertainty to the final background predictions.

Parton distribution function

The parton distribution function (PDF) is the probability density for finding a particle with a certain longitudinal momentum fraction at a resolution scale. It is determined using results from various experiments with some theoretical assumptions.

Systematic errors from PDF uncertainties are considered following the `PDF4LHC` [54] prescriptions by using the `NNPDF3.0` [55] set. The variation of α-functions resulting from the 100 PDF weights in the `NNPDF3.0` set is propagated as a shape uncertainty on the final background predictions.

5.3 Systematics on Signal Efficiency

The signal efficiencies are affected by b-tagging efficiency scale, jet energy scale, variations of QCD scale and PDF weight (see Section 5.2). Besides, additional lepton efficiency scale factor, trigger data-to-MC scale factor, and pile-up weight are considered as systematic sources that affect signal efficiencies.

Lepton identification and trigger efficiency data-to-MC scale factors

The lepton efficiency scale factors and uncertainties of electrons and muons are provided by Egamma POG and Muon POG respectively [29, 30]. The lepton efficiency scale factors are introduced in order to correct the efficiencies in MC, by measuring scale factors in the $Z \to ee/\mu\mu$ samples with tag-and-probe procedure [31]. The efficiencies and scale factor uncertainties are computed as a function of p_T and η of lepton. Note that in muon channel, additional flat uncertainties of 1% for ID and 0.5% for isolation are considered [30].

The resulting lepton efficiency scale factor uncertainties have a large discrepancy between the electron and muon channels. The scale factors have larger uncertainty at low p_T in the electron channel due to large QCD background, while the scale factors have larger uncertainty at high p_T in the muon channel due to the small sample size. As a result, the scale factor uncertainty is dominated by systematics in electron channel, while limited by statistical uncertainty in muon channel.

The trigger efficiency scale factors and uncertainties are only considered in the muon channel and are provided by Muon POG [30]. The uncertainties are computed as a function of p_T and η of muons. An additional flat uncertainty of 0.5% for muon trigger is considered [30].

Pile-up weight

The systematic source is due to limited knowledge of the total inelastic cross section, which is used to estimate the number of pile-ups in data. A 5% uncertainty is assumed for the default value of 69 mb, and the estimated distributions of the number of pile-ups are varied accordingly [52].

The values of signal efficiency with corresponding systematic uncertainties from each source in each mass point in different categories are reported in Tables 5.1, 5.2, 5.3, and 5.4.

5.4 Summary

In conclusion, the statistical uncertainties associated with the number of events in side band data is the dominant source of uncertainty in this analysis. The sensitivity of this analysis is not limited by systematic uncertainties.

TABLE 5.1: Relative systematic uncertainties on signal efficiency in electron channel, single b-tag category.

m_{ZH}	**signal efficiency**	**b-tagging**	**JES**	**lepton**	**PDF**	**pile-up**	**QCD**
800	0.022	0.086	0.000	0.011	0.040	0.025	0.028
1000	0.033	0.096	0.007	0.010	0.042	0.005	0.041
1200	0.039	0.112	0.009	0.009	0.042	0.009	0.051
1400	0.040	0.112	0.009	0.008	0.046	0.009	0.059
1600	0.041	0.107	0.025	0.008	0.049	0.019	0.068
1800	0.046	0.094	0.014	0.008	0.053	0.009	0.074
2000	0.046	0.091	0.026	0.008	0.053	0.002	0.080
2500	0.054	0.057	0.039	0.008	0.066	0.005	0.094
3000	0.058	0.031	0.035	0.007	0.079	0.012	0.107
3500	0.062	0.015	0.036	0.007	0.100	0.004	0.119
4000	0.056	0.005	0.040	0.007	0.127	0.005	0.129

TABLE 5.2: Relative systematic uncertainties on signal efficiency in electron channel, double b-tag category.

m_{ZH}	**signal efficiency**	**b-tagging**	**JES**	**lepton**	**PDF**	**pile-up**	**QCD**
800	0.040	0.068	0.008	0.010	0.038	0.008	0.026
1000	0.061	0.074	0.011	0.009	0.042	0.004	0.038
1200	0.073	0.078	0.014	0.009	0.043	0.015	0.048
1400	0.076	0.082	0.014	0.009	0.048	0.012	0.057
1600	0.072	0.084	0.011	0.008	0.049	0.007	0.064
1800	0.071	0.087	0.022	0.008	0.051	0.012	0.072
2000	0.071	0.089	0.029	0.008	0.055	0.009	0.079
2500	0.056	0.092	0.036	0.008	0.066	0.002	0.094
3000	0.044	0.093	0.043	0.008	0.081	0.007	0.107
3500	0.034	0.094	0.043	0.008	0.100	0.017	0.118
4000	0.028	0.095	0.050	0.007	0.131	0.003	0.129

TABLE 5.3: Relative systematic uncertainties on signal efficiency in muon channel, single b-tag category.

m_{ZH}	**signal efficiency**	**b-tagging**	**JES**	**lepton**	**PDF**	**pile-up**	**QCD**	**trigger**
800	0.023	0.104	0.023	0.102	0.043	0.026	0.029	0.003
1000	0.041	0.108	0.013	0.104	0.042	0.005	0.042	0.003
1200	0.046	0.108	0.017	0.101	0.043	0.016	0.051	0.003
1400	0.048	0.114	0.018	0.091	0.047	0.007	0.060	0.003
1600	0.051	0.102	0.021	0.092	0.047	0.001	0.068	0.003
1800	0.053	0.093	0.025	0.093	0.051	0.020	0.075	0.003
2000	0.053	0.096	0.027	0.086	0.056	0.012	0.081	0.003
2500	0.060	0.055	0.034	0.082	0.071	0.010	0.095	0.002
3000	0.066	0.028	0.043	0.077	0.079	0.007	0.107	0.002
3500	0.067	0.016	0.040	0.075	0.101	0.005	0.120	0.002
4000	0.066	0.005	0.044	0.075	0.131	0.006	0.130	0.002

TABLE 5.4: Relative systematic uncertainties on signal efficiency in muon channel, double b-tag category.

m_{ZH}	**signal efficiency**	**b-tagging**	**JES**	**lepton**	**PDF**	**pile-up**	**QCD**	**trigger**
800	0.048	0.067	0.006	0.085	0.041	0.017	0.026	0.003
1000	0.077	0.074	0.001	0.094	0.042	0.003	0.039	0.003
1200	0.085	0.078	0.003	0.096	0.045	0.008	0.049	0.003
1400	0.089	0.082	0.010	0.098	0.046	0.006	0.058	0.003
1600	0.083	0.084	0.015	0.098	0.048	0.002	0.066	0.003
1800	0.081	0.087	0.017	0.093	0.048	0.006	0.072	0.003
2000	0.081	0.089	0.023	0.092	0.056	0.003	0.079	0.003
2500	0.064	0.091	0.038	0.089	0.065	0.018	0.094	0.003
3000	0.050	0.093	0.037	0.087	0.078	0.004	0.107	0.003
3500	0.041	0.094	0.054	0.087	0.097	0.008	0.118	0.002
4000	0.031	0.095	0.037	0.081	0.130	0.007	0.130	0.002

Chapter 6

Results and Conclusion

In this chapter, the upper limits on production cross section of a heavy resonance in the electron and muon channels in different b-tag categories are estimated. Moreover, the combination results of electron and muon channel with their two b-tag categories are presented. A conclusion is provided in the end of this chapter.

6.1 Upper Limit on Production Cross Section

Since there is no significant excess above the predicted background in the resonance mass spectrum in data (Figure 4.13), a 95% confidence level upper limit on the signal contribution in the data is estimated in order to quantify the agreement between data and background+signal or background-only hypothesis.

The estimation of sensitivity in the analysis starts from a hypothesis testing [56, 57]. Two hypotheses are made: a signal plus background hypothesis ($N_{\mathrm{S}} + N_{\mathrm{B}}$), and a background-only hypothesis (N_{B}). A log-likelihood ratio $\mathcal{Q}$ from the two hypotheses is defined as

$$\ln \mathcal{Q} = \ln \frac{\mathcal{L}(N_{\mathrm{data}}, N_{\mathrm{S}} + N_{\mathrm{B}})}{\mathcal{L}(N_{\mathrm{data}}, N_{\mathrm{B}})} \tag{6.1}$$

where $\mathcal{L}$ is the likelihood function.

The agreement between data and the two hypotheses is estimated with

$$\begin{cases} CL_{s+b} = P_{s+b}(\ln \mathcal{Q} \leq \ln \mathcal{Q}_{obs}) \\ CL_b = P_b(\ln \mathcal{Q} \leq \ln \mathcal{Q}_{obs}) \end{cases} \tag{6.2}$$

where CL_{s+b} represents the probability to get a result which is less compatible with a signal when the signal hypothesis is true, and CL_b represented the probability to get a result less compatible with the background-only hypothesis than the observed one.

In fact, the CL_s method is frequently used to extract the upper limit. The CL_s is defined as

$$CL_s = \frac{CL_{s+b}}{CL_b} \tag{6.3}$$

By this definition, an exclusion at 95% confidence level ($CL_s < 0.05$) represented the probability to observe more events than what is seen in the data with any of the two hypotheses is less than 5%. Note that an observed limit is obtained with the signal plus background hypothesis, while an expected limit is obtained with the background-only hypothesis. The upper limits in this analysis are obtained with the `Asymptotic` CL_s frequentist criterion, by using the CMS standard `combine` tool [58] recommended by the LHC Higgs Combination Group [59].

The observed 95% CL_s upper limit on effective cross section of heavy resonance mass, as well as the expected limit and its relative 68% and 95% uncertainty bands, in electron and muon channels, with single and double b-tag categories, are reported as a function of resonance mass in Figure 6.1. Note that the effective cross section $\sigma_{\text{effective}}$ is defined as

$$\sigma_{\text{effective}} = \sigma(X) \times \mathcal{B}(X \to ZH) \tag{6.4}$$

where $\sigma(X)$ is the production cross section of the heavy resonance X, and $\mathcal{B}(X \to ZH)$ is the branching ratio of the heavy resonance decaying into ZH channel.

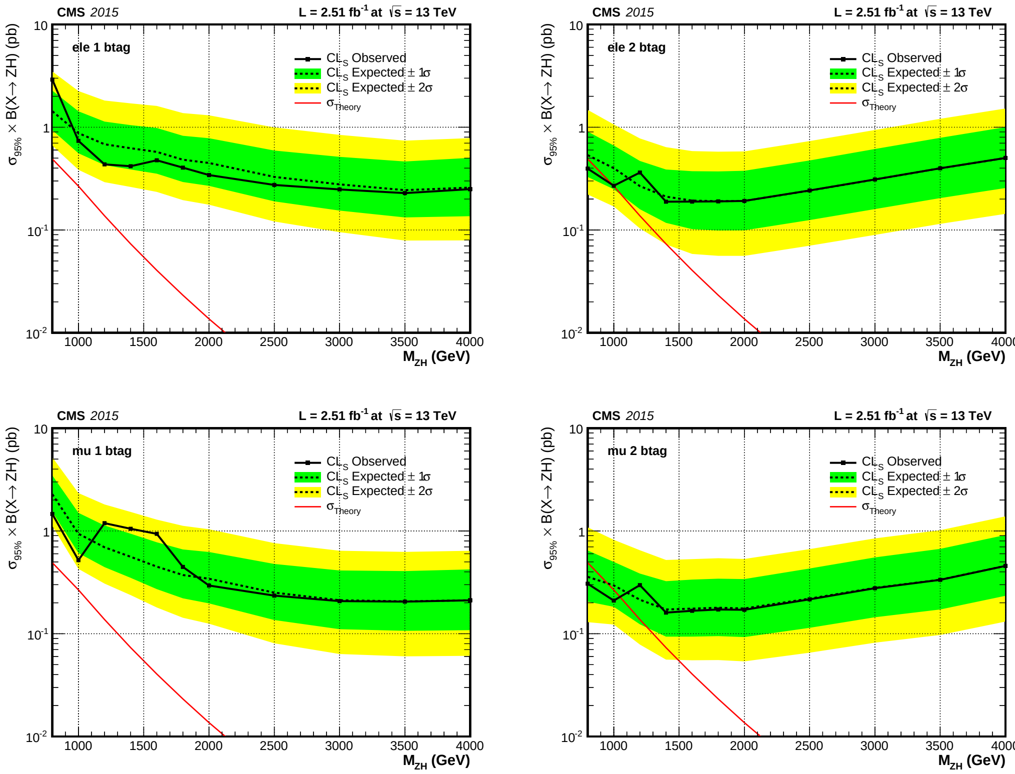

FIGURE 6.1: Observed and expected (with ± 1 and $\pm 2\sigma$ $\pm$ band) 95% CL_s upper limit on $\sigma(X) \times \mathcal{B}(X \rightarrow ZH)$ as a function of resonance mass in each category. (Top: electron channel. Bottom: muon channel. Left: single b-tag. Right: double b-tag.)

The results show that the double b-tag category is more sensitive than the single b-tag category for the same lepton flavor channel.

The trends of the upper limits on single and double b-tag categories can be interpreted by their corresponding signal efficiency ϵ_{signal}, which is defined as

$$\epsilon_{\text{signal}} = \frac{\text{Number of events passed the final selections}}{\text{Number of generated events}} \tag{6.5}$$

The signal efficiency of each lepton and b-tag multiplicity category is presented in Figure 6.2. The signal efficiencies of the single b-tag category are lower than those of the double b-tag category at the low resonance mass region. On the other hand, the signal efficiencies of the double b-tag category are lower than those of the single b-tag category at the high resonance mass region.

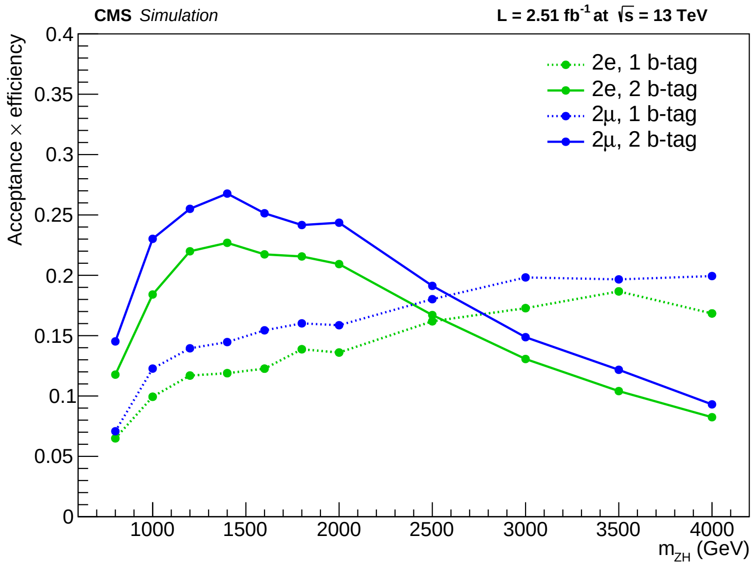

FIGURE 6.2: Signal efficiency separated by final state and b-tagging multiplicity after the signal selections.

The boosted Higgs bosons from the high-mass resonance produce two b-quarks that are very close to each other. The subjet b-tagging fails to associate one secondary vertex to each subjet. Therefore, either two secondary vertices are associated to one subjet or one secondary vertex is associated to only one subjet. As a result, the signal efficiency of the double b-tag category drops in the high resonance mass region.

The upper limit on the effective cross section is proportional to the square root of the number of background events and is inversely proportional to the signal efficiencies and luminosities. Therefore, the downward trend of the upper limits in the low resonance mass region of the single b-tag category corresponds to its upward signal efficiency trend in that region, while the upward trend of the upper limits in the high resonance mass region of the double b-tag category corresponds to its downward signal efficiency trend in that region.

The signal-to-background (S/B) ratio of the double b-tag category is higher than those of the single b-tag category at the low resonance mass region, while the lower signal efficiency in the double b-tag category is recovered by the single b-tag category at the high resonance mass region. The upper limits are improved

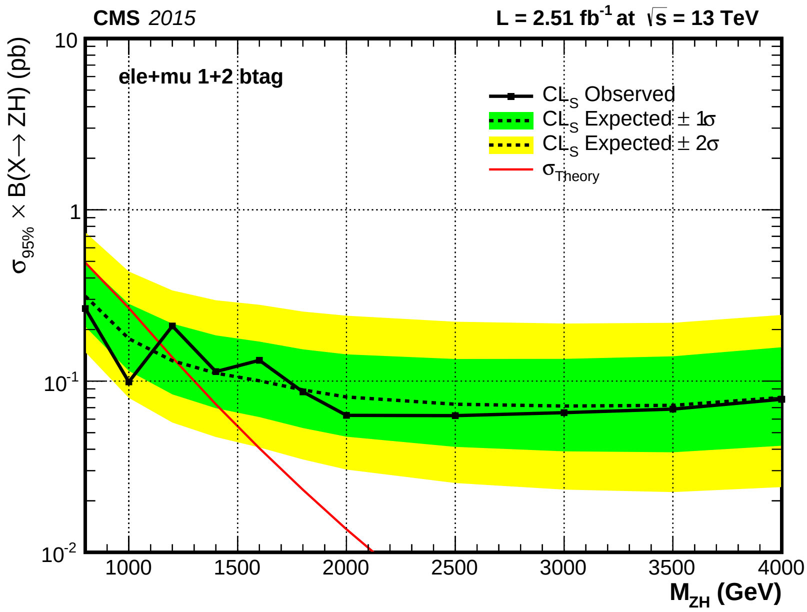

FIGURE 6.3: Observed and expected (with ±1 and ±2σ ± band) 95% CL_s upper limit on $\sigma(X) \times \mathcal{B}(X \rightarrow ZH)$ as a function of resonance mass for the combination of the electron and muon channels in single and double b-tag categories.

after combining the electron and muon channels, with the single and double b-tag categories. An upper limit from 0.063 pb to 0.265 pb for m_{ZH} from 800 to 4000 GeV is obtained from the combination of four categories (Figure 6.3).

In addition, the results are interpreted with the simplified HVT (Section 1.2.1), where the parameters ($g_V = 3$, $c_H \sim -1$, $c_F \sim 1$) are considered in the strongly coupled benchmark model (Section 1.2.2). The cross sections multiplied by the branching ratios predicted in the benchmark model are overlaid with the exclusion limits in Figure 6.3. With the present luminosity, a HVT in strongly coupled benchmark model with $m_{ZH} <$ 1140 GeV is excluded at 95% CL_s upper limit.

6.2 Conclusion

A search for heavy resonances with mass larger than 800 GeV decaying into a Z boson and a Higgs boson, using the data collected at $\sqrt{s}$ = 13 TeV during 2015 with the CMS detector at LHC is presented. The leptonic decay modes of the Z boson are considered, while the Higgs boson decays hadronically into a pair of

b quarks. The luminosity of the available data set is 2.51 fb^{-1}. The backgrounds are predicted using the α method. The observed upper limits on the cross section times the branching ratio of the heavy resonances on various hypothesis mass are presented. There is no evidence of significant deviations from background expectation.

The data collected by CMS during 2015 do not have enough statistics to perform the fit procedure in α method. The estimator can not find a global minimum easily, which causes the functional parameters to be unstable. As a result, a larger uncertainty coming from fitting procedure is introduced. Moreover, the systematic uncertainties are limited by the amount of side band data. The performance of this analysis is expected to improve in the 2016 LHC run, as the luminosity reaches 36 fb^{-1}.

Bibliography

[1] Vardan Khachatryan et al. "Search for heavy resonances decaying into a vector boson and a Higgs boson in final states with charged leptons, neutrinos, and b quarks". In: (2016). arXiv: 1610.08066 [hep-ex].

[2] Peter W. Higgs. "Broken Symmetries and the Masses of Gauge Bosons". In: *Phys. Rev. Lett.* 13 (16 Oct. 1964), pp. 508–509. DOI: 10.1103/PhysRevLett.13.508. URL: http://link.aps.org/doi/10.1103/PhysRevLett.13.508.

[3] Georges Aad et al. "Observation of a new particle in the search for the Standard Model Higgs boson with the ATLAS detector at the LHC". In: *Phys. Lett.* B716 (2012), pp. 1–29. DOI: 10.1016/j.physletb.2012.08.020. arXiv: 1207.7214 [hep-ex].

[4] Serguei Chatrchyan et al. "Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC". In: *Phys. Lett.* B716 (2012), pp. 30–61. DOI: 10.1016/j.physletb.2012.08.021. arXiv: 1207.7235 [hep-ex].

[5] Serguei Chatrchyan et al. "Observation of a new boson with mass near 125 GeV in pp collisions at $\sqrt{s} = 7$ and 8 TeV". In: *JHEP* 06 (2013), p. 081. DOI: 10.1007/JHEP06(2013)081. arXiv: 1303.4571 [hep-ex].

[6] Tao Han et al. "Phenomenology of the little Higgs model". In: *Phys. Rev.* D67 (2003), p. 095004. DOI: 10.1103/PhysRevD.67.095004. arXiv: hep-ph/0301040 [hep-ph].

[7] Maxim Perelstein. "Little Higgs models and their phenomenology". In: *Prog. Part. Nucl. Phys.* 58 (2007), pp. 247–291. DOI: 10.1016/j.ppnp.2006.04.001. arXiv: hep-ph/0512128 [hep-ph].

[8] Roberto Contino et al. "On the effect of resonances in composite Higgs phenomenology". In: *JHEP* 10 (2011), p. 081. DOI: 10.1007/JHEP10(2011)081. arXiv: 1109.1570 [hep-ph].

[9] David Marzocca, Marco Serone, and Jing Shu. "General Composite Higgs Models". In: *JHEP* 08 (2012), p. 013. DOI: 10.1007/JHEP08(2012)013. arXiv: 1205.0770 [hep-ph].

[10] Duccio Pappadopulo et al. "Heavy Vector Triplets: Bridging Theory and Data". In: *JHEP* 09 (2014), p. 060. DOI: 10.1007/JHEP09(2014)060. arXiv: 1402.4431 [hep-ph].

[11] Lyndon Evans and Philip Bryant. "LHC Machine". In: *JINST* 3 (2008), S08001. DOI: 10.1088/1748-0221/3/08/S08001.

[12] S. Chatrchyan et al. "The CMS experiment at the CERN LHC". In: *JINST* 3 (2008), S08004. DOI: 10.1088/1748-0221/3/08/S08004.

[13] V. I. Klyukhin et al. "Measuring the Magnetic Flux Density in the CMS Steel Yoke". In: (2012). DOI: 10.1007/s10948-012-1967-5. arXiv: 1212.1657 [physics.ins-det].

[14] V. Veszpremi. "Operation and performance of the CMS tracker". In: *JINST* 9 (2014), p. C03005. DOI: 10.1088/1748-0221/9/03/C03005. arXiv: 1402.0675 [physics.ins-det].

[15] S Chatrchyan et al. "Performance and Operation of the CMS Electromagnetic Calorimeter". In: *JINST* 5 (2010), T03010. DOI: 10.1088/1748-0221/5/03/T03010. arXiv: 0910.3423 [physics.ins-det].

[16] S Chatrchyan et al. "Performance of the CMS Hadron Calorimeter with Cosmic Ray Muons and LHC Beam Data". In: *JINST* 5 (2010), T03012. DOI: 10.1088/1748-0221/5/03/T03012. arXiv: 0911.4991 [physics.ins-det].

[17] Serguei Chatrchyan et al. "The performance of the CMS muon detector in proton-proton collisions at $\sqrt{s} = 7$ TeV at the LHC". In: *JINST* 8 (2013), P11002. DOI: 10.1088/1748-0221/8/11/P11002. arXiv: 1306.6905 [physics.ins-det].

[18] J. Brooke. "Performance of the CMS Level-1 Trigger". In: *PoS* ICHEP2012 (2013), p. 508. arXiv: 1302.2469 [hep-ex].

[19] Valentina Gori. "The CMS High Level Trigger". In: *Int. J. Mod. Phys. Conf. Ser.* 31 (2014), p. 1460297. DOI: 10.1142/S201019451460297X. arXiv: 1403.1500 [physics.ins-det].

[20] CMS Collaboration. *MET POG - Recommended MET Filters for Run-II*. https://twiki.cern.ch/twiki/bin/view/CMS/MissingETOptionalFiltersRun2?rev=79.

[21] Johan Alwall et al. "MadGraph 5: Going Beyond". In: *JHEP* 06 (2011), p. 128. DOI: 10.1007/JHEP06(2011)128. arXiv: 1106.0522 [hep-ph].

[22] Torbjorn Sjostrand, Stephen Mrenna, and Peter Z. Skands. "A Brief Introduction to PYTHIA 8.1". In: *Comput. Phys. Commun.* 178 (2008), pp. 852–867. DOI: 10.1016/j.cpc.2008.01.036. arXiv: 0710.3820 [hep-ph].

[23] S. Agostinelli et al. "GEANT4: A Simulation toolkit". In: *Nucl. Instrum. Meth.* A506 (2003), pp. 250–303. DOI: 10.1016/S0168-9002(03)01368-8.

[24] Carlo Oleari. "The POWHEG-BOX". In: *Nucl. Phys. Proc. Suppl.* 205-206 (2010), pp. 36–41. DOI: 10.1016/j.nuclphysbps.2010.08.016. arXiv: 1007.3893 [hep-ph].

[25] Stefan Kallweit et al. "NLO QCD+EW predictions for V+jets including off-shell vector-boson decays and multijet merging". In: *JHEP* 04 (2016), p. 021. DOI: 10.1007/JHEP04(2016)021. arXiv: 1511.08692 [hep-ph].

[26] Ryan Gavin et al. "FEWZ 2.0: A code for hadronic Z production at next-to-next-to-leading order". In: *Comput. Phys. Commun.* 182 (2011), pp. 2388–2403. DOI: 10.1016/j.cpc.2011.06.008. arXiv: 1011.3540 [hep-ph].

[27] Vardan Khachatryan et al. "Measurement of the differential cross section for top quark pair production in pp collisions at $\sqrt{s} = 8$ TeV". In: *Eur. Phys. J.* C75.11 (2015), p. 542. DOI: 10.1140/epjc/s10052-015-3709-x. arXiv: 1505.04480 [hep-ex].

[28] John M. Campbell and R. K. Ellis. "MCFM for the Tevatron and the LHC". In: *Nucl. Phys. Proc. Suppl.* 205-206 (2010), pp. 10–15. DOI: 10.1016/j.nuclphysbps.2010.08.011. arXiv: 1007.3492 [hep-ph].

[29] CMS Collaboration. *Electron efficiencies and scale factors*. https://twiki.cern.ch/twiki/bin/view/CMS/EgammaIDRecipesRun2.

[30] CMS Collaboration. *Reference muon id, isolation and trigger efficiencies for Run-II*. https://twiki.cern.ch/twiki/bin/view/CMS/MuonReferenceEffsRun2.

[31] Serguei Chatrchyan et al. "Performance of CMS muon reconstruction in *pp* collision events at $\sqrt{s} = 7$ TeV". In: *JINST* 7 (2012), P10002. DOI: 10.1088/1748-0221/7/10/P10002. arXiv: 1206.4071 [physics.ins-det].

[32] Serguei Chatrchyan et al. "Description and performance of track and primary-vertex reconstruction with the CMS tracker". In: *JINST* 9.10 (2014), P10009. DOI: 10.1088/1748-0221/9/10/P10009. arXiv: 1405.6569 [physics.ins-det].

[33] CMS Collaboration. *Pileup Reweighting*. https://twiki.cern.ch/twiki/bin/viewauth/CMS/PileupReweighting.

[34] CMS Collaboration. *Utilities for Accessing Pileup Information for Data*. https://twiki.cern.ch/twiki/bin/view/CMS/PileupJSONFileforData.

[35] Florian Beaudette. "The CMS Particle Flow Algorithm". In: *Proceedings, International Conference on Calorimetry for the High Energy Frontier (CHEF 2013): Paris, France, April 22-25, 2013*. 2013, pp. 295–304. arXiv: 1401.8155 [hep-ex]. URL: https://inspirehep.net/record/1279774/files/arXiv:1401.8155.pdf.

[36] "Particle-Flow Event Reconstruction in CMS and Performance for Jets, Taus, and MET". In: (2009).

[37] Vardan Khachatryan et al. "Performance of Electron Reconstruction and Selection with the CMS Detector in Proton-Proton Collisions at $\sqrt{s} = 8$ TeV". In: *JINST* 10.06 (2015), P06005. DOI: 10.1088/1748-0221/10/06/P06005. arXiv: 1502.02701 [physics.ins-det].

[38] R. Fruhwirth. "Application of Kalman filtering to track and vertex fitting". In: *Nucl. Instrum. Meth.* A262 (1987), pp. 444–450. DOI: 10.1016/0168-9002(87)90887-4.

[39] Wolfgang Adam et al. "Reconstruction of Electrons with the Gaussian-Sum Filter in the CMS Tracker at the LHC". In: (2005).

[40] CMS Collaboration. *Cut Based Electron ID for Run-II*. https://twiki.cern.ch/twiki/bin/view/CMS/CutBasedElectronIdentificationRun2.

[41] CMS Collaboration. *Baseline muon selections for Run-II*. https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideMuonIdRun2.

[42] Ryan Atkin. "Review of jet reconstruction algorithms". In: *J. Phys. Conf. Ser.* 645.1 (2015), p. 012008. DOI: 10.1088/1742-6596/645/1/012008.

[43] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The Anti-k_t jet clustering algorithm". In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[44] CMS Collaboration. "Pileup Removal Algorithms". In: (2014).

[45] Stephen D. Ellis, Christopher K. Vermilion, and Jonathan R. Walsh. "Recombination Algorithms and Jet Substructure: Pruning as a Tool for Heavy Particle Searches". In: *Phys. Rev.* D81 (2010), p. 094023. DOI: 10.1103/PhysRevD.81.094023. arXiv: 0912.0033 [hep-ph].

[46] Andrew J. Larkoski et al. "Soft Drop". In: *JHEP* 05 (2014), p. 146. DOI: 10.1007/JHEP05(2014)146. arXiv: 1402.2657 [hep-ph].

[47] Serguei Chatrchyan et al. "Identification of b-quark jets with the CMS experiment". In: *JINST* 8 (2013), P04013. DOI: 10.1088/1748-0221/8/04/P04013. arXiv: 1211.4462 [hep-ex].

[48] CMS Collaboration. "Identification of b quark jets at the CMS Experiment in the LHC Run 2". In: (2016).

[49] CMS Collaboration. *Methods to Apply b-Tagging Efficiency Scale Factors*. https://twiki.cern.ch/twiki/bin/view/CMSPublic/BtagRecommendation2010OpenData.

[50] CMS Collaboration. *Jet Identification*. https://twiki.cern.ch/twiki/bin/view/CMS/JetID.

[51] Wouter Verkerke and David P. Kirkby. "The RooFit toolkit for data modeling". In: *eConf* C0303241 (2003). [,186(2003)], MOLT007. arXiv: physics/0306116 [physics].

[52] CMS Collaboration. "CMS Luminosity Measurement for the 2015 Data Taking Period". In: (2016).

[53] Serguei Chatrchyan et al. "Determination of Jet Energy Calibration and Transverse Momentum Resolution in CMS". In: *JINST* 6 (2011), P11002. DOI: 10.1088/1748-0221/6/11/P11002. arXiv: 1107.4277 [physics.ins-det].

[54] Jon Butterworth et al. "PDF4LHC recommendations for LHC Run II". In: *J. Phys.* G43 (2016), p. 023001. DOI: 10.1088/0954-3899/43/2/023001. arXiv: 1510.03865 [hep-ph].

[55] Richard D. Ball et al. "Parton distributions for the LHC Run II". In: *JHEP* 04 (2015), p. 040. DOI: 10.1007/JHEP04(2015)040. arXiv: 1410.8849 [hep-ph].

[56] Thomas Junk. "Confidence level computation for combining searches with small statistics". In: *Nucl. Instrum. Meth.* A434 (1999), pp. 435–443. DOI: 10.1016/S0168-9002(99)00498-2. arXiv: hep-ex/9902006 [hep-ex].

[57] Alexander L. Read. "Presentation of search results: The CL_s technique". In: *J. Phys.* G28 (2002). [,11(2002)], pp. 2693–2704. DOI: 10.1088/0954-3899/28/10/313.

[58] CMS Collaboration. *Documentation of the RooStats-based statistics tools for Higgs PAG*. https://twiki.cern.ch/twiki/bin/view/CMS/SWGuideHiggsAnalysisCombinedLimit.

[59] Vardan Khachatryan et al. "Precise determination of the mass of the Higgs boson and tests of compatibility of its couplings with the standard model predictions using proton collisions at 7 and 8 TeV". In: *Eur. Phys. J.* C75.5 (2015), p. 212. DOI: 10.1140/epjc/s10052-015-3351-7. arXiv: 1412.8662 [hep-ex].