NORTHWESTERN UNIVERSITY

Search for Higgs Boson Decays to a Z Boson and a Photon

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Field of Physics and Astronomy

By

James Bueghly

EVANSTON, ILLINOIS

March 2022

ABSTRACT

Search for Higgs Boson Decays to a Z Boson and a Photon

James Bueghly

Since its discovery in 2012 at the Large Hadron Collider (LHC), efforts have been made to measure and characterize the properties of the Higgs boson. Among these efforts have been searches for rare decays of the Higgs predicted by the Standard Model (SM) of particle physics. One such decay is the process $H \to Z\gamma$, which has an expected branching fraction of $\mathcal{B}(H \to Z\gamma) = (1.57 \pm 0.09) \times 10^{-3}$ in the SM, assuming a Higgs boson mass of $m_H = 125.38\,\mathrm{GeV}$. This decay mode has not yet been experimentally observed, and its observation and measurement remains an important goal of Higgs physics research at the LHC. In addition, a measurement of this decay mode at a rate deviating from the SM prediction would provide indirect evidence of new physics beyond the SM.

This thesis presents a search for $H \to Z\gamma$, where $Z \to \ell^+\ell^-$ with $\ell = e$ or μ. The search is performed using a sample of proton-proton (pp) collision data at a center-of-mass energy of 13 TeV, recorded by the CMS experiment at the LHC, corresponding to an integrated luminosity of 138 fb^{-1}. Events are assigned to mutually exclusive categories, which exploit differences in both event topology and kinematics of distinct Higgs production mechanisms

to enhance signal sensitivity. To detect a potential signal, fits are performed to the distributions of $\ell^+\ell^-\gamma$ invariant mass in each of these categories simultaneously. The signal signal strength μ, defined as the product of the cross section and the branching fraction $[\sigma(\mathrm{pp} \to \mathrm{H})\mathcal{B}(\mathrm{H} \to \mathrm{Z}\gamma)]$ relative to the standard model expectation, is found to be $\mu = 2.4 \pm 0.9$ at a Higgs boson mass of 125.38 GeV. This measurement corresponds to $\sigma(\mathrm{pp} \to \mathrm{H})\mathcal{B}(\mathrm{H} \to \mathrm{Z}\gamma) = 0.21 \pm 0.08$ pb. The statistical significance of the observed excess of events is 2.7 standard deviations. The observed (expected) upper limit at 95% confidence level on μ is 4.1 (1.8). The ratio of branching fractions $\mathcal{B}(\mathrm{H} \to \mathrm{Z}\gamma)/\mathcal{B}(\mathrm{H} \to \gamma\gamma)$ is measured to be $1.5^{+0.7}_{-0.6}$, which agrees with the standard model prediction at the 1.5 standard deviation level.

Acknowledgements

I would like to thank my adviser, Mayda Velasco, for giving me the opportunity to pursue a goal I dreamed up many years ago and for supporting me along the way. I'd also like to thank the many colleagues at Northwestern University, National Central University in Taiwan, CERN, and Fermilab who helped me grow both as a scientist and a person over the last six and a half years.

Special thanks are due to those who significantly impacted the work contained in this thesis: My coanalyzer Ming-Yan Lee, whose discipline and tireless work ethic kept us all going through the many challenges we faced; postdoc Andrew Gilbert, whose expertise and guidance were critical; and the CMS leaders who guided and reviewed our work, including the Higgs conveners, Analysis Review Committee, and our colleague and language editor Jeff Richman. Without all of you, this thesis would not exist.

To my Northwestern colleagues: I will always be grateful for your friendship over these years. Those moments we shared away from the work kept me happy, sane, focused, and balanced. It is impossible to name all of them, but the squash games with Nate Odell, the lunches with Nate, Brian Pollack, Thoth Gunter, and Joseph Cordero Mercado, and the adventures at COFI in San Juan were just as meaningful to me as the work itself.

Last, but not least, I want to thank my parents, sisters, and my partner Anna. You all have shown me so much love, encouragement, and patience throughout this process. For that, I feel extremely fortunate. This is for all of you.

Table of Contents

List of Tables

shape parameters, is given as indicated, with the values averaged over
all event categories. The third column shows the uncertainties that
have a correlated effect across the three data-taking periods.   95

List of Figures

period. The left plot shows the fit to simulated ZH production events, and the right plot shows the fit to simulated WH production events. The blue line shows the total fit function, the green line shows the Crystal Ball function component, and the red line shows the Gaussian function component. 115

CHAPTER 1

Introduction

introductory text

CHAPTER 2

Theory

2.1. The Standard Model

The Standard Model (SM) is currently our best theoretical framework for understanding the nature of fundamental particles. It is rooted in the idea that particles exist as excitations of quantum fields. These fields are constructed so as to obey fundamental symmetries of nature, and they yield the particles of our universe, as well as their interactions. The SM is not only elegant and extensive, but provides a wide variety of measurable observables for the experimentalist to probe. So far, many measurements have been made of the currently known elementary particles, and the predictions of the SM have held up in each case. In this respect, it is a wildly successful theory. In other respects, it is obviously incomplete.

2.2. Electroweak Theory

2.3. Spontaneous Symmetry Breaking (Higgs Mechanism)

The electroweak Lagrangian lacking the Higgs term CITE EQUATION is insufficient. In particular, it does not provide an explanation for massive gauge bosons. Experimental evidence for massive gauge bosons dates back to the early? twentieth century. Beta decay experiments indicated charged current interactions mediated by a massive W boson, and later experiments indicated additional neutral current interactions mediated by

a massive Z boson (CHECK THIS). Later in the twentieth century (GIVE SPECIFIC DATES/REFS) the W and Z bosons were discovered. Indeed, the W and Z masses have been found to be roughly 80 GeV and 91 GeV respectively. An explanation came in the form of the Higgs mechanism (CITE), which spontaneously breaks the SU(2)xU(1) gauge symmetry of the elecroweak Lagrangian. In addition, there arises a real Higgs field accompanied by a massive Higgs boson, and fermion masses are explained via their Yukawa couplings to the Higgs boson. The discovery of the Higgs boson and the subsequent measurements of its properties have provided proof that the Higgs mechanism is indeed a central piece of the Standard Model.

To understand how the Higgs mechanism works, consider the introduction of a complex scalar field Φ, which transforms as a doublet under SU(2).

$$(2.1) \qquad \Phi = \begin{bmatrix} \phi^+ \\ \phi^0 \end{bmatrix}$$

Its contribution to the Lagrangian is given by:

$$(2.2) \qquad \mathcal{L}_\mathcal{H} = (D^\mu \Phi)^\dagger (D_\mu \Phi) - V(\Phi)$$

where the Higgs potential takes the form

$$(2.3) \qquad V(\Phi) = \mu^2 |\Phi^\dagger \Phi| + \lambda \left(|\Phi^\dagger \Phi| \right)^2.$$

D_μ is the covariant derivative

$$(2.4) \qquad D_\mu = \partial_\mu - \frac{ig}{2} \tau \cdot A_\mu - \frac{ig'}{2} B_\mu Y.$$

Consider the case that the parameters of the Higgs potential λ and μ satisfy the conditions $\lambda > 0$ and $\mu^2 < 0$. Then the shape of the potential is shown in **??**. Clearly, there is no minimum of the potential at $\Phi = 0$. Rather, an infinite set of minima lie around a circle in the complex plane. Hence, it is said that Φ has a nonzero vacuum expectation value (VEV). The value of the VEV in terms of μ and λ can be determined by explicitly minimizing the potential:

$$\frac{\partial}{\partial(\Phi^\dagger\Phi)} V(\Phi) = 0$$

$$\mu^2 + 2\lambda\left(|\Phi^\dagger\Phi|\right) = 0$$

(2.5)
$$\mu^2 + 2\lambda\left[(\phi^+)^2 + (\phi^0)^2\right] = 0$$

Clearly, this can be minimized in many ways depending on individual values of ϕ^+ and ϕ^0 in the vacuum. By convention, and without loss of generality, we choose the case in which $\phi^+ = 0$. In this case we obtain the equation

(2.6)
$$\phi^0 = \sqrt{\frac{-\mu^2}{2\lambda}} = \frac{1}{\sqrt{2}}v$$

where we have defined $v \equiv \sqrt{-\mu^2/\lambda}$.

The existence of the Higgs VEV has profound implications. To see this, it is helpful to reparameterize the scalar doublet field Φ as follows:

(2.7)
$$\Phi = \frac{1}{\sqrt{2}}e^{i\frac{\tau^a}{2}\theta_a(x)}\begin{bmatrix} 0 \\ v + h(x) \end{bmatrix}$$

As Φ is invariant under local SU(2) gauge transformations, the prefactor may be rotated away. This is equivalent to setting $\theta(x) = 0$ in equation 2.7. This choice of gauge is known as the unitary gauge, and leads to

$$
(2.8) \qquad \Phi = \frac{1}{\sqrt{2}} \begin{bmatrix} 0 \\ v + h(x) \end{bmatrix}
$$

Given the above form of Φ, we can now evaluate the Higgs Lagrangian (equation 2.2), starting with the kinetic term $(D^\mu \Phi)^\dagger (D_\mu \Phi)$.

$$
\begin{aligned}
(D^\mu \Phi)^\dagger (D_\mu \Phi) &= \left| (\partial_\mu - \frac{ig}{2} \tau \cdot A_\mu - \frac{ig'}{2} B_\mu Y) \Phi \right|^2 \\
&= \frac{1}{2} \left| \begin{bmatrix} \partial_\mu - \frac{i}{2}(gA_\mu^3 + g'B_\mu) & -\frac{ig}{2}(A_\mu^1 - iA_\mu^2) \\ -\frac{ig}{2}(A_\mu^1 + iA_\mu^2) & \partial_\mu + \frac{i}{2}(gA_\mu^3 - g'B_\mu) \end{bmatrix} \begin{bmatrix} 0 \\ v + h(x) \end{bmatrix} \right|^2 \\
&= \frac{1}{2} \left| \begin{bmatrix} -\frac{ig}{2}(A_\mu^1 - iA_\mu^2)(v + h(x)) \\ \partial_\mu h(x) + \frac{i}{2}(gA_\mu^3 - g'B_\mu)(v + h(x)) \end{bmatrix} \right|^2 \\
&= \frac{1}{2} \partial_\mu h(x) \partial^\mu h(x) + \frac{1}{8}(gA_\mu^3 - g'B_\mu)(gA_3^\mu - g'B^\mu)(v + h(x))^2 \\
(2.9) \qquad &+ \frac{g^2}{8}(A_\mu^1 - iA_\mu^2)(A_\mu^1 + iA_\mu^2)(v + h(x))^2
\end{aligned}
$$

With some foreknowledge of the result, let us define the physical gauge fields and their masses as follows:

$$(2.10) \qquad W_\mu^\pm = \frac{1}{\sqrt{2}}(A_\mu^1 \mp iA_\mu^2) \qquad\qquad\qquad m_W = \frac{gv}{2}$$

$$(2.11) \qquad Z_\mu = \frac{1}{\sqrt{g^2 + g'^2}}(gA_\mu^3 - g'B_\mu) \qquad m_Z = \sqrt{g^2 + g'^2}\frac{v}{2}$$

$$(2.12) \qquad A_\mu = \frac{1}{\sqrt{g^2 + g'^2}}(g'A_\mu^3 + gB_\mu) \qquad m_A = 0$$

Then the kinetic term of the Higgs Lagrangian can be recast as

$$
\begin{aligned}
(D^\mu \Phi)^\dagger (D_\mu \Phi) =\ & \frac{1}{2}\partial_\mu h(x)\partial^\mu h(x) \\
& + \frac{1}{2}m_Z^2 Z_\mu Z^\mu + m_W^2 W_\mu^+ W^{-\mu} \\
& + \frac{v}{4}(g^2 + g'^2)Z_\mu Z^\mu h + \frac{1}{8}(g^2 + g'^2)Z_\mu Z^\mu h^2 \\
& + \frac{v}{4}g^2 W_\mu^+ W^{-\mu} h + \frac{1}{8}g^2 W_\mu^+ W^{-\mu} h^2
\end{aligned}
$$

(2.13)

Equation 2.13 provides a great deal of information on the physical ramifications of the Higgs mechanism. The first term is the kinetic term of the physical Higgs boson field. The second and third terms are the mass terms of the Z and W bosons respectively. The fourth and fifth terms show the linear and quadratic couplings of the Z boson to the Higgs boson respectively. Finally, the sixth and seventh terms show the linear and quadratic couplings of the W boson to the Higgs boson respectively. Given the form of equation 2.8, a similar expansion can be carried out on the Higgs potential (equation 2.3). Here, the terms involving the physical Higgs boson are most interesting, so constant terms are

dropped.

$$V(\Phi) = \frac{\mu^2}{2}(v+h)^2 + \frac{\lambda}{4}((v+h)^2)^2$$

(2.14)
$$\rightarrow \lambda v^2 h^2 + \lambda v h^3 + \frac{\lambda}{4}h^4$$

The first term is a Higgs mass term, with $m_H = \sqrt{2\lambda v^2}$. The second and third terms describe the Higgs trilinear and quartic self-coupling respectively.

The Lagrangian including the Higgs doublet Φ can be further extended to incorporate interactions between the Higgs and fermion fields. These interactions, along with the nonzero Higgs VEV, provide a mechanism to generate the fermion masses. The interactions take the form of Yukawa couplings:

(2.15)
$$\mathcal{L}_{Yukawa} = Y_{ij}^d \bar{Q}_L^i \Phi d_R^j + Y_{ij}^u \bar{Q}_L^i \tilde{\Phi} u_R^j + Y_{ij}^e \bar{L}_L^i \Phi e_R^j + h.c.$$

where

(2.16)
$$\tilde{\Phi} \equiv i\tau_2 \Phi^*$$

Plugging in the unitary gauge Φ parameterization of equation 2.8, this evaluates to

(2.17)
$$\mathcal{L}_{Yukawa} = \frac{Y_{ij}^d}{\sqrt{2}}\bar{d}_L^i(v+h)d_R^j + \frac{Y_{ij}^u}{\sqrt{2}}\bar{u}_L^i(v+h)u_R^j + \frac{Y_{ij}^e}{\sqrt{2}}\bar{e}_L^i(v+h)e_R^j$$

It is worth looking closely at the terms in equation 2.17. For a given fermion type, the first term in the parentheses is a fermion mass term. The second term in the parentheses gives the coupling of the fermion to the Higgs boson. With this in mind, we observe that

the fermion mass is given in terms of the couplings and the Higgs vev by

$$(2.18) \qquad m_f = \frac{y_f v}{\sqrt{2}}$$

where y_f is the relevant value taken from the Yukawa coupling matrix. In addition, we see that the strength of the fermion coupling to the Higgs boson is given by $\frac{m_f}{v}$. Thus, fermions couple to the Higgs boson with strength directly proportional to their masses. This result has important consequences in the context of collider experiments, as it regulates the rates of Higgs production and decay related to each fermion-Higgs vertex.

2.4. Higgs Production

2.5. Higgs Decay

2.6. Physics Beyond the Standard Model

CHAPTER 3

Experiment Description

3.1. The Large Hadron Collider (LHC)

The Large Hadron Collider (LHC) is a high energy proton-proton collider that serves several modern particle physics detector experiments. Built at CERN and straddling the French-Swiss border, and first turned on for operation in 2008, it is the most powerful particle collider in history. The LHC was designed to collide protons at a maximum center of mass energy of $\sqrt{s} = 14\,\text{TeV}$ with a maximum instantaneous luminosity of $10^{34} cm^{-2} s^{-1}$. To achieve this level of performance, the 27 km tunnel originally constructed for the Large Electron-Positron Collider (LEP) was repurposed to separately accelerate two couterrotating proton beams. The separate acceleration is made possible by oppositely oriented magnetic dipole fields in the two rings. When the beams achieve the desired energy, they are collided at one of four interaction points.

The protons collided in the LHC are first gathered, bunched, and accelerated by other parts of the CERN accelerator complex before injection into the LHC ring. First, protons are stripped off of hydrogen atoms by a duoplasmatron and accelerated to $50\,\text{MeV}$ by a linear accelerator (LINAC). Next, the protons are further accelerated in a series of synchroton rings of increasing size, the Proton Synchroton Booster (PSB), Proton Synchroton (PS), and Super Proton Synchroton (SPS). The SPS brings the proton beam energy up to $450\,\text{GeV}$. At this point, the beam is injected to the two LHC rings to be
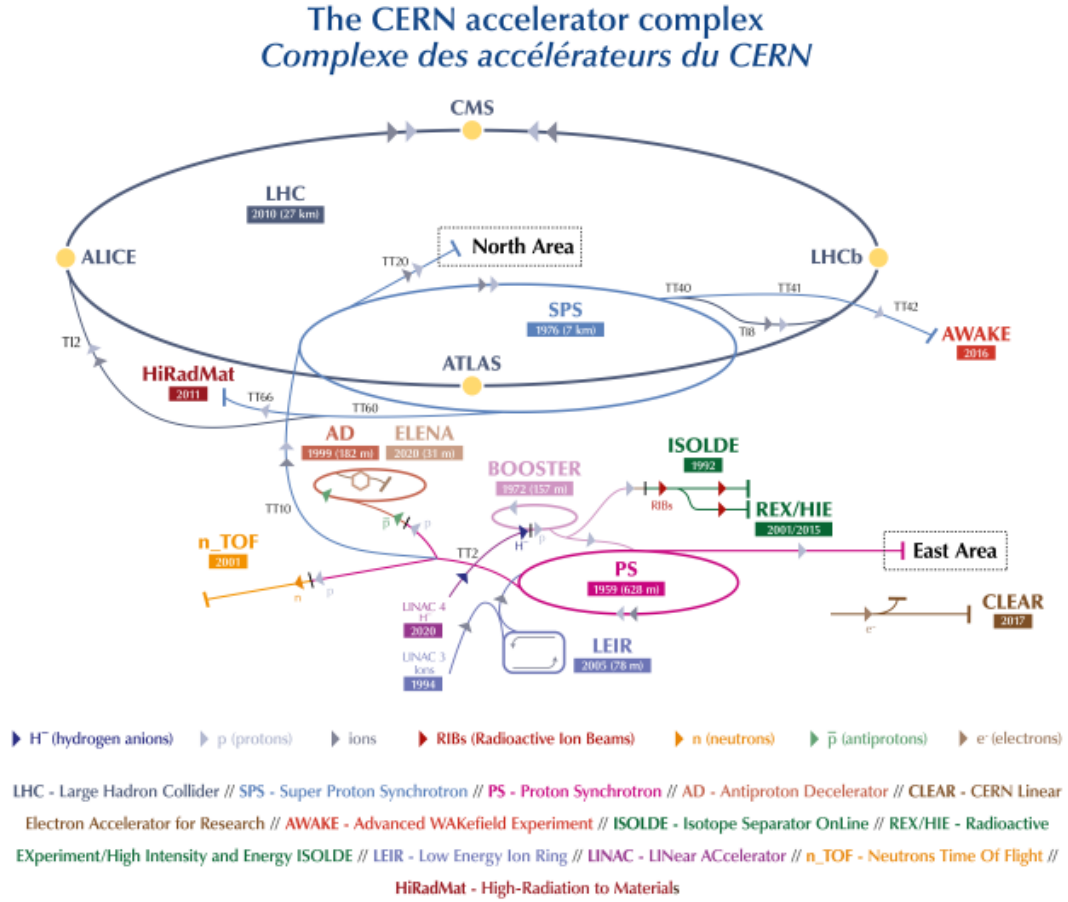
Figure 3.1. CERN accelerator complex.

accelerated up to the final collision energy. Acceleration in the main ring is achieved by a system of thousands of superconducting dipole magnets, along with hundreds of correcting quadrupole magnets. A liquid helium cooling system is used to maintain the magnets at cryogenic temperatures. A full diagram of the CERN accelerator complex, including the parts relevant for the LHC, is shown in Fig. 3.1.

3.2. The Compact Muon Solenoid (CMS)

The CMS apparatus [?] is a multipurpose, nearly hermetic detector, designed to trigger on [?, ?] and identify electrons, muons, photons, and (charged and neutral) hadrons [?, ?, ?, ?]. The central feature of the CMS apparatus is a superconducting solenoid of 6 m internal diameter, providing a magnetic field of 3.8 T. Within the solenoid volume are a silicon pixel and strip tracker, a lead tungstate crystal electromagnetic calorimeter (ECAL), and a brass and scintillator hadron calorimeter (HCAL), each composed of a barrel and two endcap sections. The ECAL consists of 75 848 lead tungstate crystals, which provide coverage in pseudorapidity $|\eta| < 1.48$ in a barrel region (EB) and $1.48 < |\eta| < 3.0$ in two endcap regions (EE). Preshower detectors consisting of two planes of silicon sensors interleaved with a total of 3 radiation lengths of lead are located in front of each EE detector. Forward calorimeters extend the pseudorapidity coverage provided by the barrel and endcap detectors. Muons are measured in gas-ionization detectors embedded in the steel flux-return yoke outside the solenoid. A more detailed description of the detector components and operation is provided below.

3.2.1. Superconducting Magnet

The function of the superconducting magnet within the CMS detector is to bend the trajectories of charged particles. This is crucial for particle identification and momentum measurement. Designed to provide a maximum magnetic field of 4 T, its operating strength during pp collision runs is set to 3.8 T. The bulk of the magnet is composed of NbTi, cooled by liquid helium to 4.5K, which is below the critical temperature allowing for superconductivity. The magnet has a length of 12.5m, diameter of 6.3m, and mass of
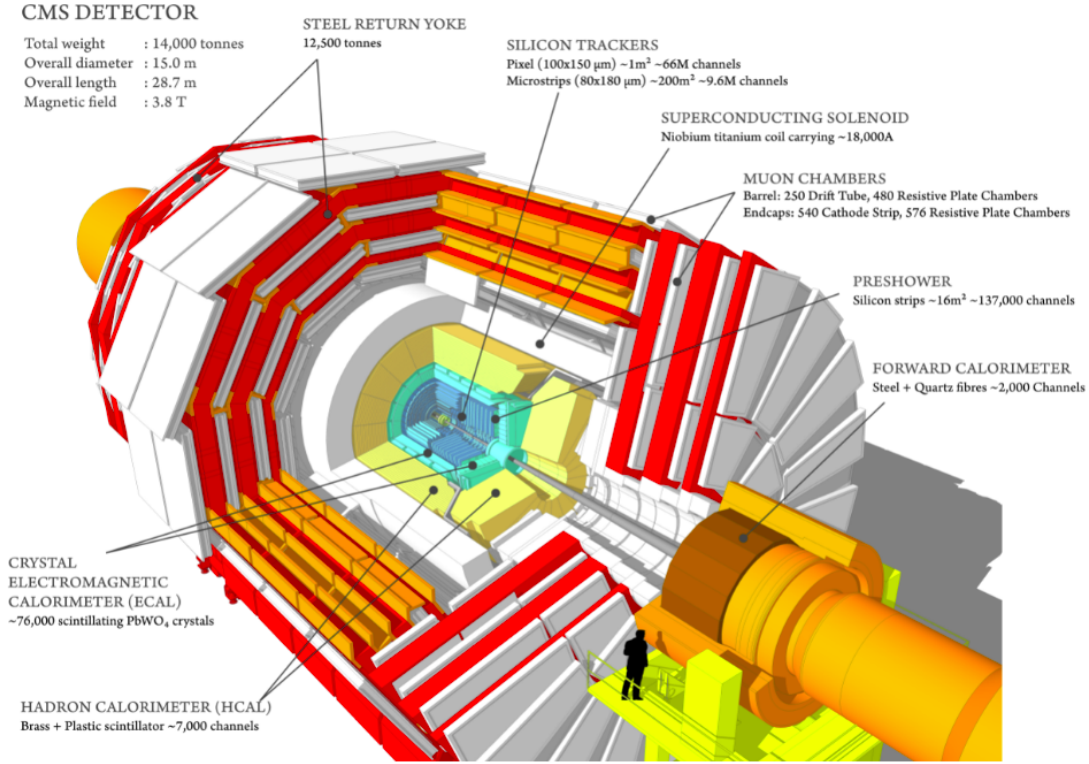
Figure 3.2. The CMS detector apparatus.

220 tons. The solenoid encloses several detector components, including the tracker and the majority of the calorimeters. A steel flux-return yoke is built around the solenoid, and is composed of 5 wheels and two endcaps weighing a total of roughly 10,000 tons.

3.2.2. Inner Tracking System

The inner tracking system is designed to reconstruct charged particle trajectories and vertices arising from particle decays. Comprised of a silicon pixel detector and silicon strip tracker, it covers the pseudorapidity region $|\eta| < 2.5$ and is designed for efficient and precise measurement of charged particles with p_T about about $1\,\mathrm{GeV}$. At the LHC design luminosity for pp collisions, each bunch crossing leads to about 1,000 hits in the
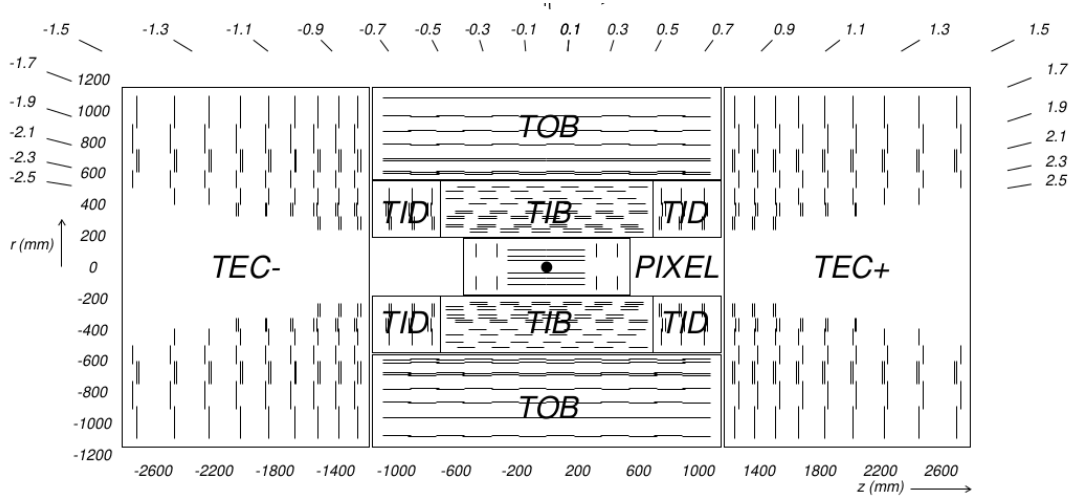
Figure 3.3. Cross section of the CMS inner tracking system in the r-z plane.

inner tracking system. As such, the inner tracking system was designed to be maximally radiation tolerant while maintaining physics performance.

The silicon pixel detector covers the inner region of $r < 10cm$. It is composed of $100 \times 150 \mu m^2$ pixels arranged in three barrel layers and two endcap disks. Due to the extreme radiation environment, the innermost layer was designed to be replaced after at least two years of LHC operation. In response to LHC running conditions and detector degradation, a replacement and upgrade of the full pixel detector was made during the LHC extended year-end technical stop in 2016/2017. The silicon strip tracker covers the region $20 < r < 116cm$. It is divided into three subsystems: the Tracker Inner Barrel (TIB), Tracker Inner Disks (TID), and Tracker Outer Barrel (TOB). The strip thickness is $320\mu m$ ($500\mu m$) in the TIB/TID (TOB). A schematic diagram of the CMS inner tracking system in the r-z plane is shown in Fig. 3.3.

3.2.3. Electromagnetic Calorimeter (ECAL)

The electromagnetic calorimeter (ECAL) is designed to identify and measure electromagnetic particles by inducing and characterizing electromagnetic showers. It is made of 61,200 lead tungstate ($PbWO_4$) crystals in the barrel (EB), which covers $|\eta| < 1.479$, plus 7,324 crystals in each endcap (EE), which fall in the range $1.479 < |\eta| < 3.0$. A preshower system lies in front of the EE. The preshower is a lead and silicon sampling calorimeter designed to aid in the identification of neutral pions and to improve the spatial resolution in the endcap region. The crystal length corresponds to 25.8 (24.7) radiation lengths (χ^0) in the EB (EE). When electromagnetic particles, such as electrons and photons, encounter the ECAL, electromagnetic showers are induced. Light is emitted by the ECAL crystals proportional to the energy of constituent particles in the shower, allowing for reconstruction of the full shower energy. An array of fast and radiation tolerant photodetectors detects the emitted light, and this information is saved off detector. For this purpose, the EB uses Hamatsu avalanche photodiodes, while the EE uses vacuum phototriodes.

3.2.4. Hadronic Calorimeter (HCAL)

Measurements of hadrons, jets, and missing transverse momentum are important to the analysis goals of the CMS physics program. The hadronic calorimeter (HCAL) is instrumental in such measurements. In a nutshell, the HCAL is primarily a sampling calorimeter broken into several subdetectors. The barrel (HB) covers the range $|\eta| < 1.3$, the endcaps (HE) cover the range $1.3 < |\eta| < 3$, the forward calorimeter (HF) covers the range $3 < |\eta| < 5$, and an outer barrel calorimeter (HO) covers $|\eta| < 1.3$ outside of the solenoid magnet.

Layers of brass absorber are interleaved with plastic scintillator in the HB and HE, with a total absorber thickness of 5.82–10.6 interaction lengths in the HB and about 10 interaction lengths in the HE. Scintillation light emitted by the active material is measured and transmitted off-detector using hybrid photodiodes. The design of the HF was motivated by the extreme particle fluxes in the very forward region. The HF is made up of steel plate absorbers, in which quartz fibers are inserted to serve as the active material. Cherenkov light is produced when shower particles above the Cherenkov threshold pass through the fibers, and a calculable fraction of the light is captured by the fibers. The function of the HO is to recover the energy in showers that leak through the HB, which is crucial for the accurate measurement of missing transverse momentum. In the HO, the solenoid coil itself, as well as a thick iron tail catcher, function as absorbers. This extends the effective HCAL depth to 11.8 interaction lengths everywhere except at the barrel-endcap boundary.

3.2.5. Muon Detectors

The CMS muon detector system is comprised of a set of gas ionization detectors embedded in the flux-return yoke outside the solenoid magnet. These detectors cover the pseudorapidity range $|\eta| < 2.4$. The positioning of the detectors outside the magnet takes advantage of the fact that muons deposit minimal energy in the detector materials, such as the calorimeters. Most other types of high energy particles will have deposited their energy before reaching the muon system, so using the combination of muon system and tracker information, CMS is able to reconstruct muons with excellent momentum resolution. The gas ionization detectors in the muon system come in three types: drift tubes

(DTs), cathode strip chambers (CSCs), and resistive plate chambers (RPCs). The DTs cover the barrel region ($|\eta| < 1.2$) and are arranged in four stations, each with 60–70 drift chambers containing a gas mixture of 85% Ar and 15% CO_2. The CSCs are multiwire proportional chambers composed of anode wire planes interleaved with cathode panels. Muons passing through the CSCs ionize a gas mixture of 40% Ar, 50% CO_2, and 10% CF_4, and the electrons flow to the anodes, yielding a detectable avalanche of charge. Covering the region $|\eta| < 2.1$, the RPCs are composed of oppositely charged parallel plates enclosing a gas mixture of primarily $C_2H_2F_4$. One advantage of the RPCs is their good time resolution, which is less than the LHC bunch spacing time of 25ns. This makes them useful for fast muon triggering that matches muon tracks to the relevant bunch crossing. Figure 3.4 shows a cross sectional view of the CMS detector highlighting the muon system components.

3.3. Trigger System

At the LHC, the proton beam crossing interval is 25 ns, which corresponds to a pp collision rate of 40 MHz. Recording the full set of detector information for each collision is unfeasible, as it would lead to far too much data to save to disk. As a result, the CMS experiment employs a two-tiered trigger system designed to preserve information of physics interest while reducing the stored event rate to a more manageable 100 Hz. The first level of the trigger (L1) is a hardware trigger, which reduces the event rate from 40 MHz to about 100 kHz. The second layer is the high level trigger (HLT), a processor farm using software optimized for fast processing.
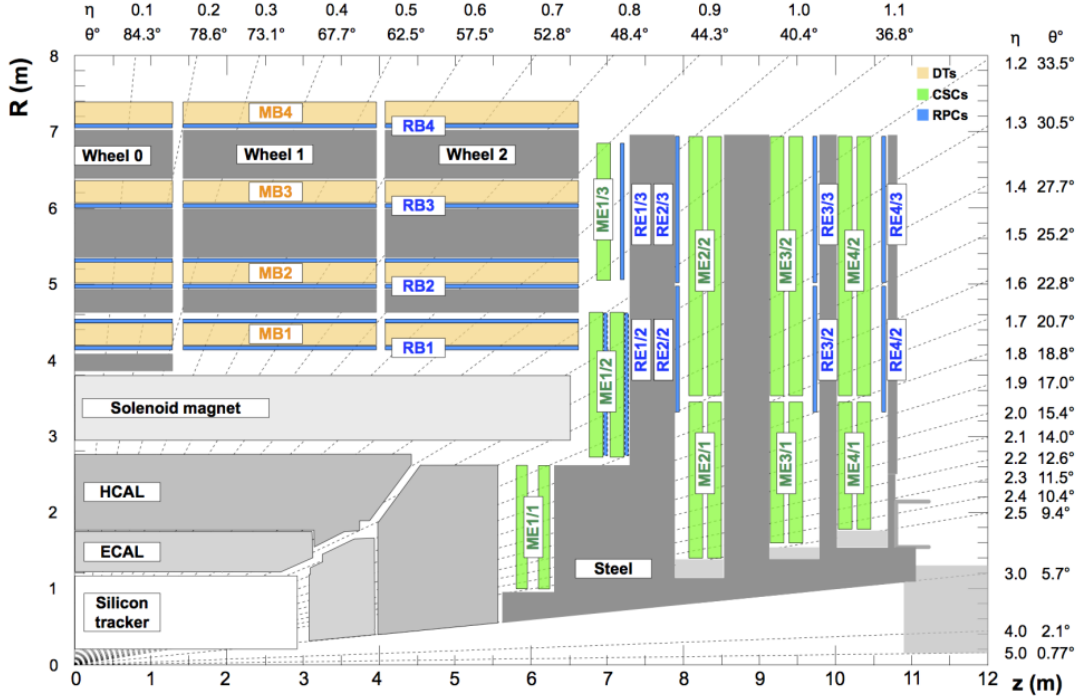
Figure 3.4. Diagram of CMS detector in the r-z plane showing the components of the muon system.

The L1 trigger is implemented using FPGAs and ASICs, and has local, regional, and global components. First, local patterns in the calorimeters, track segments, and hit patterns in the muon chambers are used by the Trigger Primitive Generators (TPGs). The TPGs rank and sort primitive objects corresponding to particle candidates based on energy, momentum, and reconstruction quality. Information from the TPGs is taken as input by regional triggers for the calorimeters and muon systems. These regional triggers further determine candidate physics objects, as well as energy sums and isolation information. Finally, this information is fed to a set of global triggers, which rank trigger objects across the entire detector. The final global trigger must decide whether to accept or reject an event based on information input from the global calorimeter and muon trigger

systems. Algorithms used in this decision can be based on single object p_T thresholds and multiplicity-related thresholds, such as the presence of multiple jets, among other criteria. Events passing the L1 trigger are read out and passed to the HLT for further processing.

In contrast to the L1 trigger, the HLT incorporates the full physics object reconstruction based on the full precision of the detector. This allows the HLT to accept and reject events with algorithms of similar quality to those used by offline analyses. Over 13,000 CPU cores are dedicated to HLT processing. For speed and efficiency, reconstruction and filtering algorithms are applied in increasing order of complexity. If a filter sequence fails, the rest of the reconstruction is skipped. Additionally, processing is done regionally based on the L1 candidates and relevant detector components passed as input to the HLT. Events that pass the HLT are assigned to relevant data streams based on their physics content. For this analysis, the relevant streams are the Double Muon and Double Electron streams, triggered by events with two muons or two electrons passing a set of minimum p_T requirements, among other cuts.

3.4. Object Reconstruction

The global event reconstruction (also called particle-flow (PF) event reconstruction [?]) aims to reconstruct and identify each individual particle in an event, with an optimized combination of all subdetector information. In this process, the identification of the particle type (photon, electron, muon, charged hadron, neutral hadron) plays an important role in the determination of the particle direction and energy.
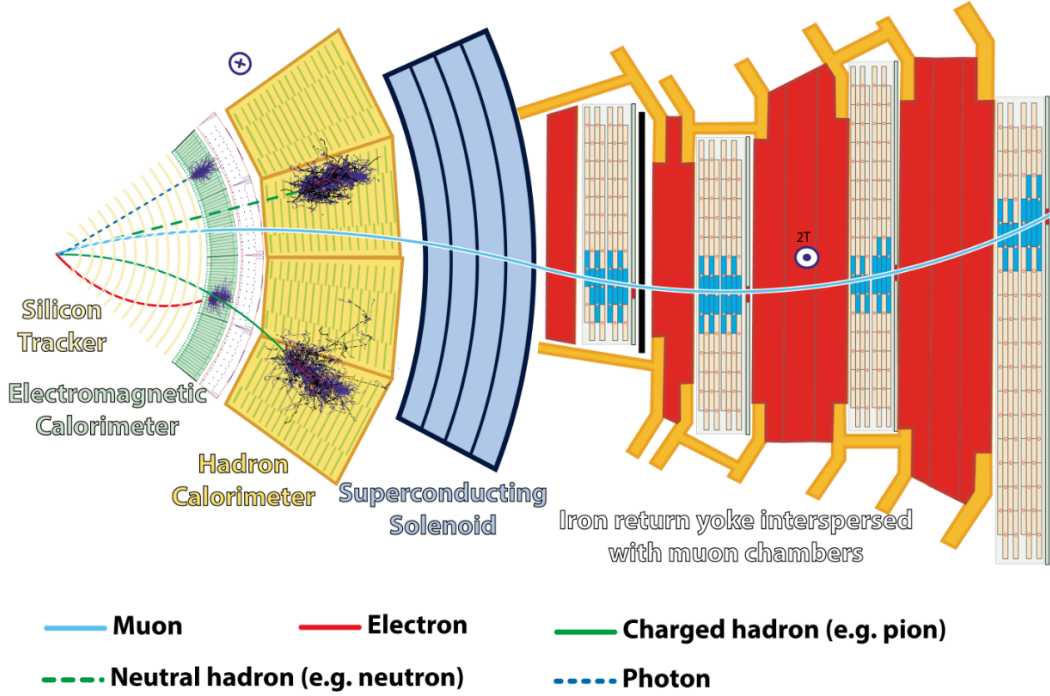
Figure 3.5. Schematic of different particles interacting with the CMS detector.

3.4.1. Photon and Electron Reconstruction

Photons and electrons interact with the material in the ECAL, depositing the majority of their energy before reaching the HCAL. As they interact, photons convert into electron-positron pairs and electrons radiate bremsstrahlung photons. This leads to the formation of an electromagnetic shower. Because of this, the original particle energy is split into multiple energy deposits, which must be combined to reconstruct the original energy. Deposits in individual ECAL crystals are combined into clusters, and these clusters are in turn combined into superclusters. Two algorithms are used to generate superclusters: the so-called "mustache" algorithm, and the so-called "refined" algorithm. The mustache algorithm defines a seed cluster with energy above a certain threshold, and then combines it with other clusters within a region of the eta-phi plane centered at the seed location.

The refined algorithm uses the mustache superclusters as well as tracking information to extrapolate bremsstrahlung and conversion tracks to decide whether a cluster should belong to the supercluster.

The distinction between photons and electrons is made using tracking information, where photons are associated with no tracks and electrons with tracks. The Gaussian Sum Filter (GSF) [REF] track fitting algorithm is used to identify and characterize tracks that might be associated with an electron. It first begins with a hit pattern in the tracker, which is used as a seed. This seed can either be tracker-driven, coming from the collection of generic tracks tested for mutual compatibility, or it can be ECAL-driven, where a mustache supercluster is compared in location with a collection of tracker hit patterns to determine if the supercluster is consistent with the trajectory indicated by the track. Electron seeds are then converted into reconstructed electron tracks. In the absence of any GSF electron tracks, a photon candidate is obtained. Additional separation between photons and electrons is obtained through further selection requirements. The measured energy resolution for electrons produced in Z boson decays in pp collision data ranges from 2–5%, depending on electron pseudorapidity and energy loss through bremsstrahlung in the detector material [?].

3.4.2. Muon Reconstruction

Muon reconstruction utilizes information from the muon detectors and the tracker. First, detector hits in the CSCs, DTs, and RPCs are used to build standalone tracks using a Kalman-filter technique. Subsequently, these standalone muon tracks are combined with tracker information via two algorithms. So-called "tracker muons" are reconstructed

using an "inside-out" algorithm, which starts from tracker tracks and matches them to DT or CSC segments. So-called "global muons" are reconstructed with an "outside-in" approach which starts from standalone muon tracks and matches them to tracker tracks using a Kalman-filter technique. In the case where both algorithms reconstruct a muon sharing the same tracker tack, the two outputs are merged into a single muon candidate. In general, the tracker muon algorithm is more efficient in the region of low muon p_T, while the global algorithm is efficient at high p_T.

The energy of muons is obtained from the corresponding track momentum. Matching muons to tracks measured in the silicon tracker results in a p_T resolution, for muons with p_T up to $100\,\text{GeV}$, of 1% in the barrel and 3% in the endcaps. The p_T resolution in the barrel is better than 7% for muons with p_T up to $1\,\text{TeV}$ [**?**].

3.4.3. Hadrons

Charged hadrons are identified as charged particle tracks that are neither identified as electrons nor as muons. Neutral hadrons are identified as HCAL energy clusters not linked to any charged hadron trajectory, or as a combined ECAL and HCAL energy excess with respect to the expected charged hadron energy deposit.

3.4.4. Jets

For each event, hadronic jets are reconstructed as clusters of PF particles, including photons, electrons, muons, and hadrons. For clustering, the infrared and collinear-safe anti-k_T algorithm [**?**, **?**] is used. The goal of jet clustering algorithms is to combine PF particles together into a single jet object by merging them together based on their relative geometric
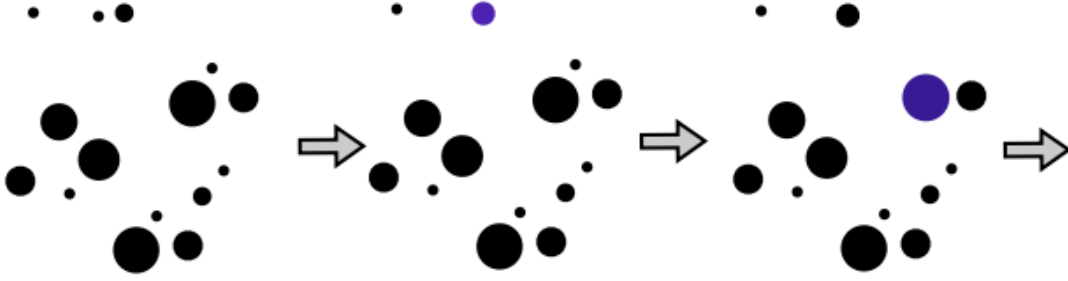
Figure 3.6. An example of the first few steps of the anti-k_T jet clustering algorithm.

distances and transverse momenta (k_T). In the case of the anti-k_T algorithm, clustering proceeds from smallest geometric distance to largest, weighted by squared inverse k_T. This inverse weighting preferentially combines PF particles with higher momenta before lower momenta, for a given distance. The advantage of this approach is that low momentum particles are more likely to be clustered with high momentum counterparts, rather than with other low momentum particles. It follows that these low momentum particles will not significantly modify the jet shape, whereas the more important, high-momentum particles, will. A precise definition of the anti-k_T algorithm is defined in equation 3.1.

$$(3.1) \qquad d_{ij} = min(k_{T,i}^{2p}, k_{T,j}^{2p})\frac{(\eta_i - \eta_j)^2 + (\phi_i - \phi_j)^2}{R^2}, \quad d_i = k_{T,i}^{2p}$$

In equation 3.1, the indices i and j refer to two PF particles, the parameter p is taken to be -1, and the radius parameter R is parameter that can be tuned. For standard jets in CMS, as used in this analysis, R is set to 0.4. Particles are clustered from smallest d_{ij} to largest until there is no d_{ij} smaller than k_{ti}^{-2}, at which point the final jet is defined. A visual representation of the anti-k_T algorithm is shown in Fig. 3.7.
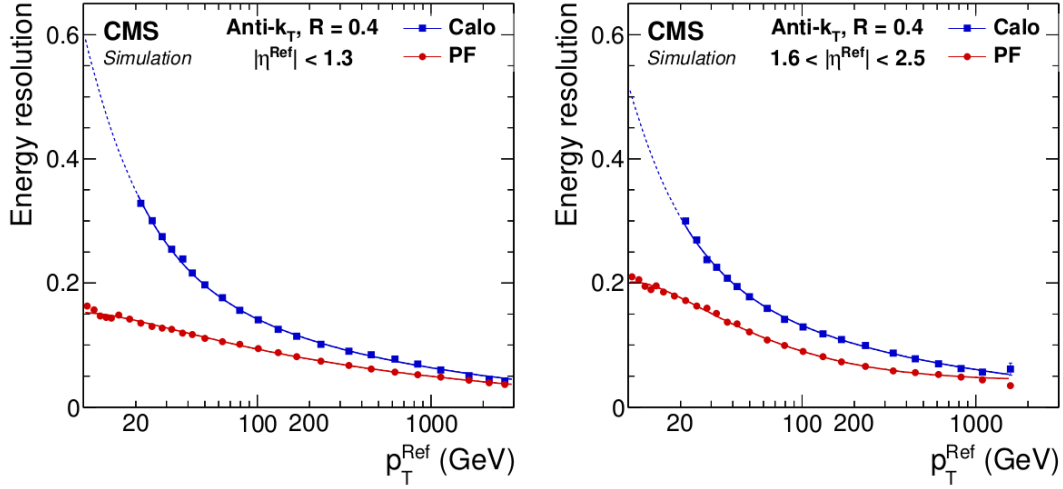
Figure 3.7. Jet energy resolution as a function of simulated jet p_T in the barrel (left) and in the endcap (right) regions. Calo refers to the sum of ECAL and HCAL energy deposits, while PF corresponds to jets clustered from PF particles using the anti-k_T algorithm.

Jet momentum is determined as the vectorial sum of all particle momenta in the jet, and is found from simulation to be, on average, within 5–10% of the true momentum over the entire p_T spectrum and detector acceptance. Additional pp interactions within the same or nearby bunch crossings (pileup) can contribute additional tracks and calorimetric energy depositions to the jet momentum. To mitigate this effect, charged particles identified to be originating from pileup vertices are discarded and an offset correction is applied to correct for remaining contributions [?]. The jet energy resolution typically amounts to 15–20% at 30 GeV, 10% at 100 GeV, and 5% at 1 TeV [?].

CHAPTER 4

Overview of Analysis Strategy

Before discussing the full details of the analysis procedure and results, it is worth summarizing the broader strategy taken in our search for H $\rightarrow$ Zγ. As two prior CMS results have been published with Run 1 data and 2016 data, we will emphasize the ways in which our analysis overlaps and differs from these previous approaches. Very broadly, the current search is similar to the previous analyses in trigger, object, and basic event selection. However, it is significantly more advanced in three body mass reconstruction, event categorization, and background modeling. We will show that the innovations in these areas have significantly improved the expected sensitivity and statistical robustness of the search with respect to the past CMS analyses.

Chapter 5 provides a detailed description of the data and Monte Carlo simulation used in our analysis. Standard dimuon and dielectron trigger streams are used for 2016, 2017, and 2018 LHC data. The full dataset corresponds to an integrated luminosity of 137 fb^{-1}. In other words, we use the full CMS Run 2 dataset at 13 TeV center of mass energy. Simulated signal samples are used to determine the expected signal yields and three body mass shape. Simulated background samples are used for MVA training and category optimization. However, the background shape and normalization in the final result is determined by fitting the data and does not rely on any simulation.

Chapter 6 describes the basic object and event selection used in the analysis, and Chapter 7 describes further selection and categorization using MVAs. As mentioned, the

basic selection requirements are fairly similar to previous CMS analyses. Muons are selected with a loose cut-based ID, while electrons and photons are selected with loose MVA IDs. Loose ID requirements are chosen in order to maximize signal efficiency. Background is then suppressed through a combination of basic cuts and MVA methods. Basic kinematic cuts on isolation, mass, and photon energy variables are able to significantly reduce backgrounds from initial and final state radiation, while the MVA methods are able to strongly disciminate against backgrounds from jets misreconstructed as photons. Finally, a kinematic fit procedure in the dilepton mass is able to significantly improve the signal mass resolution. The kinematic fit is an innovation of the current analysis and contributes to its improved sensitivity.

Chapter 8 details the approach to signal and background modeling of the three body mass spectrum. As in previous CMS H $\to$ Zγsearches, the signal shape is determined via an analytic fit to simulation. A resonant background contribution from $H \to \mu^+\mu^-$is modeled similarly. The nonresonant background contribution is taken from a fit to data in the range of 105 to 170 GeV. This background model includes both the turn-on arising from the real Z boson peak as well as the falling spectrum at higher mass. We note that the turn-on was fit in the Run 1 analysis as well, but was dropped in the 2016 analysis. A more thorough discussion of the merits of fitting the turn-on will be described in chapter 8.

Chapter 9 describes the systematic uncertainties relevant for the analysis, and Chapter 10 gives a basic overview of the statistics used to arrive at the final results. It is worth noting that given the current integrated luminosity, the analysis is dominated by statistical

uncertainty. Chapter 11 gives the full set of results, including best fit signal strength, limits, and comparisons with $H \to \gamma\gamma$.

CHAPTER 5

Data and Simulated Samples

The data sample used for the H $\to$ Zγ search corresponds to a total integrated luminosity of $138\,\mathrm{fb}^{-1}$ and was collected over a data-taking period spanning three years: $36.3\,\mathrm{fb}^{-1}$ in 2016, $41.5\,\mathrm{fb}^{-1}$ in 2017, and $59.8\,\mathrm{fb}^{-1}$ in 2018 [?, ?, ?]. To be considered in the analysis, events must satisfy the high-level trigger requirements for at least one of the dielectron or dimuon triggers. The dielectron trigger requires a leading (subleading) electron with $p_\mathrm{T} > 23\,(12)\,\mathrm{GeV}$, while the dimuon trigger requires a muon with $p_\mathrm{T} > 17\,(8)\,\mathrm{GeV}$. The corresponding CMS data streams are listed in Table 5.1. To ensure data quality, luminosity masks are applied based recommendations from the Physics Performance and Dataset group. Table 5.2 shows the luminosity masks applied for each of the three data-taking years.

Signal samples for ggH, VBF, VH, and t$\bar{\mathrm{t}}$H production, with H $\to$ Zγ and Z $\to \ell^+\ell^-$ ($\ell =$ e, μ, or τ), are generated at next-to-leading order (NLO) using POWHEG v2.0 [?, ?]. Samples are produced for m_H of 120, 125, and 130 GeV. The SM Higgs boson production cross sections and branching fractions recommended by the LHC Higgs Working Group [?] are considered for each mass point.

The dominant backgrounds, $\mathrm{Z}/\gamma^*(\to \ell^+\ell^-)+\gamma$ and $\mathrm{Z}/\gamma^*(\to \ell^+\ell^-)+$jets, are generated at NLO using the MADGRAPH5_aMC@NLO v2.6.0 (v2.6.1) generator [?] for 2016 (2017 and 2018) samples. Events arising from t$\bar{\mathrm{t}}$ production [?] are a relatively minor background and are generated at NLO with POWHEG v2.0 [?, ?]. The background from vector

Table 5.1. Summary of data samples used.

data set name	channel
/DoubleEG/Run2016B-17Jul2018_ver2-v1/MINIAOD	$Z\gamma \to ee + \gamma$
/DoubleEG/Run2016C-17Jul2018-v1/MINIAOD	$Z\gamma \to ee + \gamma$
/DoubleEG/Run2016D-17Jul2018-v1/MINIAOD	$Z\gamma \to ee + \gamma$
/DoubleEG/Run2016E-17Jul2018-v1/MINIAOD	$Z\gamma \to ee + \gamma$
/DoubleEG/Run2016F-17Jul2018-v1/MINIAOD	$Z\gamma \to ee + \gamma$
/DoubleEG/Run2016G-17Jul2018-v1/MINIAOD	$Z\gamma \to ee + \gamma$
/DoubleEG/Run2016H-17Jul2018-v1/MINIAOD	$Z\gamma \to ee + \gamma$
/DoubleEG/Run2016H-17Jul2018-v1/MINIAOD	$Z\gamma \to ee + \gamma$
/DoubleEG/Run2017B-31Mar2018-v1/MINIAOD	$Z\gamma \to ee + \gamma$
/DoubleEG/Run2017C-31Mar2018-v1/MINIAOD	$Z\gamma \to ee + \gamma$
/DoubleEG/Run2017D-31Mar2018-v1/MINIAOD	$Z\gamma \to ee + \gamma$
/DoubleEG/Run2017E-31Mar2018-v1/MINIAOD	$Z\gamma \to ee + \gamma$
/DoubleEG/Run2017F-31Mar2018-v1/MINIAOD	$Z\gamma \to ee + \gamma$
/EGamma/Run2018A-17Sep2018-v2/MINIAOD	$Z\gamma \to ee + \gamma$
/EGamma/Run2018B-17Sep2018-v1/MINIAOD	$Z\gamma \to ee + \gamma$
/EGamma/Run2018C-17Sep2018-v1/MINIAOD	$Z\gamma \to ee + \gamma$
/EGamma/Run2018D-22Jan2019-v2/MINIAOD	$Z\gamma \to ee + \gamma$
/DoubleMuon/Run2016B-17Jul2018_ver2-v1/MINIAOD	$Z\gamma \to \mu\mu + \gamma$
/DoubleMuon/Run2016C-17Jul2018-v1/MINIAOD	$Z\gamma \to \mu\mu + \gamma$
/DoubleMuon/Run2016D-17Jul2018-v1/MINIAOD	$Z\gamma \to \mu\mu + \gamma$
/DoubleMuon/Run2016E-17Jul2018-v1/MINIAOD	$Z\gamma \to \mu\mu + \gamma$
/DoubleMuon/Run2016F-17Jul2018-v1/MINIAOD	$Z\gamma \to \mu\mu + \gamma$
/DoubleMuon/Run2016G-17Jul2018-v1/MINIAOD	$Z\gamma \to \mu\mu + \gamma$
/DoubleMuon/Run2016H-17Jul2018-v1/MINIAOD	$Z\gamma \to \mu\mu + \gamma$
/DoubleMuon/Run2016H-17Jul2018-v1/MINIAOD	$Z\gamma \to \mu\mu + \gamma$
/DoubleMuon/Run2017B-31Mar2018-v1/MINIAOD	$Z\gamma \to \mu\mu + \gamma$
/DoubleMuon/Run2017C-31Mar2018-v1/MINIAOD	$Z\gamma \to \mu\mu + \gamma$
/DoubleMuon/Run2017D-31Mar2018-v1/MINIAOD	$Z\gamma \to \mu\mu + \gamma$
/DoubleMuon/Run2017E-31Mar2018-v1/MINIAOD	$Z\gamma \to \mu\mu + \gamma$
/DoubleMuon/Run2017F-31Mar2018-v1/MINIAOD	$Z\gamma \to \mu\mu + \gamma$
/DoubleMuon/Run2018A-17Sep2018-v2/MINIAOD	$Z\gamma \to \mu\mu + \gamma$
/DoubleMuon/Run2018B-17Sep2018-v1/MINIAOD	$Z\gamma \to \mu\mu + \gamma$
/DoubleMuon/Run2018C-17Sep2018-v1/MINIAOD	$Z\gamma \to \mu\mu + \gamma$
/DoubleMuon/Run2018D-PromptReco-v2/MINIAOD	$Z\gamma \to \mu\mu + \gamma$

boson scattering (VBS) production of $Z/\gamma^* + \gamma$ pairs, with the Z boson decaying to a pair of leptons, is simulated at leading order using the MADGRAPH5_aMC@NLO generator.

Table 5.2. JSON files for luminosity masks used in the analysis.

run period	JSON file	luminosity (fb^{-1})
2016	Cert_271036-284044_13TeV_ReReco_07Aug2017_Collisions16_JSON.txt	36.3
2017	Cert_294927-306462_13TeV_EOY2017ReReco_Collisions17_JSON_v1.txt	41.5
2018	Cert_314472-325175_13TeV_17SeptEarlyReReco2018ABC_PromptEraD_Collisions18_JSON.txt	59.8

The decay $H \rightarrow \mu^+\mu^-$ is considered as a resonant background and is generated for the ggH, VBF, VH, and $t\bar{t}H$ production mechanisms. The ggH production cross section is computed at next-to-next-to-NLO precision in QCD and at NLO in electroweak (EWK) theory [?]. The cross sections for Higgs boson production in the VBF [?] and VH [?] mechanisms are calculated at next-to-NLO in QCD, including NLO EWK corrections, while the $t\bar{t}H$ cross section is computed at NLO in QCD and EWK theory [?].

All simulated events are interfaced with PYTHIA v8.226 (v8.230) [?] with the CUETP8M1 [?] (CP5 [?]) underlying event tune for 2016 (2017–2018) for the fragmentation and hadronization of partons and the internal bremsstrahlung of the leptons. The NLO parton distribution function (PDF) set, NNPDF v3.0 [?] (NNPDF v3.1) [?], is used to produce these samples in 2016 (2017–2018). The response of the CMS detector is modeled using the GEANT4 program [?]. The simulated events are reweighted to correct for differences between data and simulation in the number of additional pp interactions, trigger efficiencies, selection efficiencies, and efficiencies of isolation requirements for photons, electrons, and muons.

5.1. Z+jets Overlap Removal

In the simulation of the Z+jets sample, PYTHIA adds ISR and FSR photons to the events. However, these events overlap with those in SM Zγ sample. In addition, the ISR and FSR photons in the SM Zγ sample are simulated at the matrix element level, and

the interference between the ISR and FSR Z𝛾 diagrams is better computed in Madgraph. Therefore the overlapping events are removed from the Z+jets sample and retained in the SM Z𝛾 sample. Events are identified as overlapping if they contain at least one photon which is either flagged as "fromHardProcessFinalState" nor "isPromptFinalState" with $\Delta R < 0.1$, $\Delta R = \sqrt{\Delta\eta^2 + \Delta\phi^2}$, with respect to the selected reconstructed photon.

5.2. Photon Internal Conversion

It was noticed by the H $\to$ 𝛾𝛾 group that Pythia 8 showering automatically includes a component in which H $\to$ 𝛾𝛾 photons are "internally converted" into H $\to$ 𝛾𝛾∗ $\to$ ff𝛾𝛾 (with ff mostly ee). This also affects the H $\to$ Z𝛾 and SM Z𝛾 samples. We lose some events with γ^* decay to fermions since they do not pass our selection of 2 leptons plus a photon. Since the normalization of the samples does not include the internal conversion component, the cross section has to be adjusted accordingly. To determine the number by which the cross section is adjusted, we measure the internal conversion rate in MC. We define a ratio where the numerator is the number of photons flagged as **fromHardProcessFinalState** and the denominator is total number of events in the MC. Subtracting this ratio from 1 gives the internal conversion rate. In our analysis, the internal conversion rate is 3.1%(2016) and 3.0%(2017). We therefore divide our cross sections by 0.969 (0.97) for 2016 (2017 and 2018).

5.3. $ll\gamma$ p_{T} Reweighting

For the 2016 MC, the underlying event tuning was developed in Run-1 with 7 TeV data. To get a more appropriate underlying description, the GEN group provided a new tuning for 2017 and 2018 MC production, updating from CUETP8M1 (2016) to CP5

(2017/2018). However, a discrepency was found in the $\ell\ell\gamma$ p_{T} spectrum which had an obvious and significant impact in the SM $Z\gamma$ sample. Since $p_{\mathrm{T}}^{ll\gamma}/m_{ll\gamma}$ is an important variable in the kinematic MVA training, we correct for this issue by reweighting the $p_{\mathrm{T}}^{ll\gamma}$ from simulation to data in 2017 and 2018 using a sideband region. We define the sideband region as $115GeV < m_{ll\gamma} < 120GeV$ and $130GeV < m_{ll\gamma} < 135GeV$, which is chosen close to signal region to avoid differences in kinematics. We evaluate this weight for the electron and muon channels for 2017 and 2018.
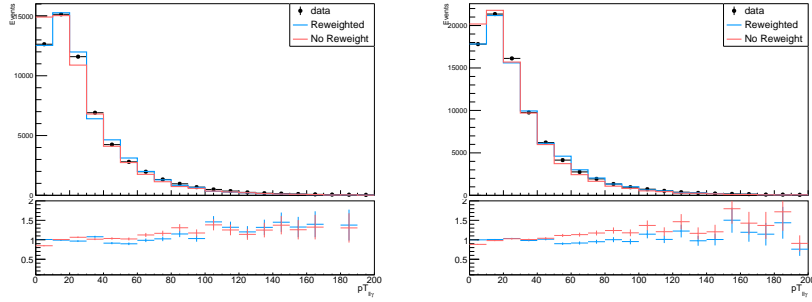
Figure 5.1. Simulation before and after the correction to the $\ell\ell\gamma$ p_{T}. The black points show the data. Left:2017, right: 2018.

5.4. Pileup Reweighting

The simulation includes an accurate distribution of the number of interactions taking place in each bunch crossing. Although the Deterministic Annealing primary vertex reconstruction [?] has been shown to be efficient and well-behaved up to the observed levels of pileup, the final distribution for the number of reconstructed primary vertices is still sensitive to differences between data and MC in the primary vertex reconstruction and underlying event. Additionally, there is a potentially larger effect where the distribution

for the number of reconstructed vertices can be biased by the offline event selection criteria and even by the trigger. In order to factorize these effects, instead of reweighting the MC by the number of reconstructed Primary Vertices, we reweight the number of pileup interactions in the simulation (as stored in the PileupInfo collection in the MC). The pileup distribution for data is derived by using the per-bunch-crossing-per-luminosity-section instantaneous luminosity from the LumiDB together with the total pp inelastic cross section. The total pp inelastic cross section is taken to be 66 mb, 72.4 mb, and 80 mb for 2016, 2017, and 2018 respectively. This yields an expected pileup distribution, correctly weighted by the per-bunch-crossing-per-luminosity-section integrated luminosity over the entire data-taking period. The inelatistic cross-sections are choosing the Figure 5.2 shows the number of primary vertices distribution after pile-up reweighting in the *ee*γ channel.
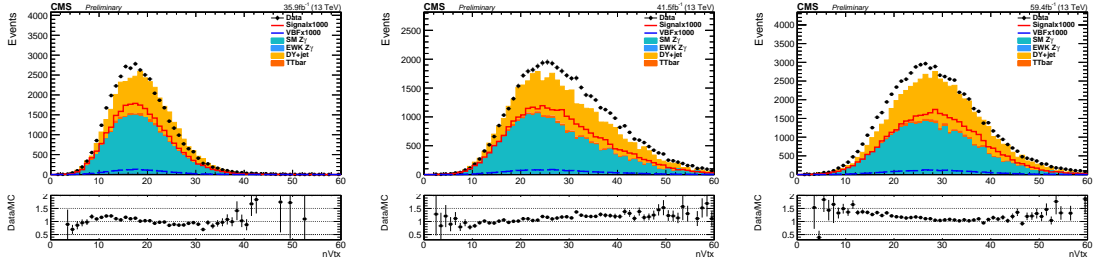
Figure 5.2. Distribution of the number of primary vertices in data and simulation shown for the electron channel after 2 leptons and 1 photon are selected. Left to right: 2016, 2017, 2018. The pileup cross section is taken to be 66 mb, 72.4 mb, and 80 mb for 2016, 2017, and 2018 respectively.

5.5. L1 Prefiring Problem

In 2016 and 2017, a gradual timing shift of the ECAL was not properly propagated to L1 trigger primitives (TP) resulting in a significant fraction of high eta TP being mistakenly associated to the previous bunch crossing. Since Level 1 rules forbid two consecutive bunch crossings to fire, an unpleasant consequence of this (in addition to not finding the TP in the bx 0) is that events can self veto if a significant amount of ECAL energy is found in the region of $2. < |\eta| < 3$. This effect is not described by the simulations [**?**].

The JETMET twiki provides a recipe to compute the probability for an event not to prefire, that can then be applied to the simulations. This is achieved with an EDProducer that runs over all offline photons and jets found in the event and assign them a prefiring probability.

The final event weight is obtained as the product of the non prefiring probability of all objects (measured using unprefirable events), namely:

$$(5.1) \qquad \omega = 1 - P(Prefiring) = \prod_{i=photon,jets} (1 - \epsilon_i^{pref}(\eta, p_{\mathrm{T}}^{EM}))$$

The impact of the L1 prefiring for electrons (muons) is a 1.4 (0.8)% loss in signal yield for 2016 and 2.4 (1.1)% loss in signal yield for 2017.

CHAPTER 6

Object and Event Selection

6.1. Triggers

The topology and basic kinematics of the H $\to$ Zγ process guide the choice of triggers used in this analysis. As this is a three body decay, the p_T of the photon tends to be less than in other channels like $H \to \gamma\gamma$. Consequently, the CMS photon trigger p_T thresholds are too great to make them viable options. Instead, we trigger on the leptons arising from the decay of the Z boson, which tend to have larger p_T. The best approach to maximize signal efficiency is to use the double lepton triggers. The double muon trigger has p_T thresholds of 17 and 8 GeV, and the double electron trigger has thresholds of 23 and 12 GeV. These are the lowest unprescaled double lepton triggers generally available for CMS Run 2 analysis. Events passing both the double muon and double electron triggers are treated as double muon events.

The triggers are applied to both data and simulation. Trigger efficiencies and scale factors are measured using the simulation samples corresponding to each data-taking year. These measurements use a tag and probe method [CITE]. The tag and probe method takes advantage of the high purity of $Z \to \ell^+\ell-$ events near the Z mass peak. One lepton functions as the tag, and satisfies a set of tight trigger, identification, isolation, p_T requirements. The second lepton, the probe, must pass a looser selection and is used to measure the efficiency in question. Using this approach, trigger efficiencies for each

leg of a given double lepton trigger are measured in both data and simulation. Then a corrective scale factor, defined as the ratio of data efficiency to simulation efficiency, is applied to the simulation. Scale factors are measured and applied in bins of p_T and $|\eta|$.

For the double electron trigger efficiency measurements, the tag electron must satisfy the following requirements. It must pass the single electron trigger, pass a tight cut-based identification, have $p_T > 30\,(35)$ in 2016 (2017/2018), and have $|\eta| < 2.5$. The probe electron must pass the loose electron MVA identification (with isolation) requirement. The efficiencies for each leg of the trigger are measured separately, so in each case, the probe electron must match the trigger leg being measured. The efficiencies and scale factors of each double electron trigger leg for 2016, 2017, and 2018 are shown in Figures [FIGS].

For the double muon trigger efficiency measurements, the tag muon must satisfy the following requirements. It must pass the single muon trigger, pass a tight cut-based identification, have $p_T > 26\,(29)$ GeV for 2016 (2017/2018) and satisfy $|\eta| < 2.4$. The probe muon must pass the $H \to ZZ$ identification and isolation cuts. The details of the $H \to ZZ$ muon identification will be described later. The efficiencies for each leg of the trigger are measured separately, so in each case, the probe muon must match the trigger leg being measured. The efficiencies and scale factors of each double muon trigger leg for 2016, 2017, and 2018 are shown in Figures [FIGS].

6.2. Muon Selection

A loose cut-based muon identification is used in the analysis. This identification was originally developed by the $H \to ZZ$ analysis [REF] and is well-suited for H $\to$ Zγ due

to the similarity of the multiple, potentially soft, muons in the final state. All muons are first required to pass a set of common cuts, followed by a separate set of cuts for high p_T (greater than 200 GeV) and low p_T muons. All muons must satisfy $p_T > 5$ GeV, $|\eta| < 2.4$, $|d_{xy}| < 0.5$ cm, and $|d_z| < 1$ cm, where d_{xy} and d_z are impact parameters defined with respect to the primary vertex of the interaction using the best muon track. Additionally, the three dimensional impact parameter (analogously defined) must have a magnitude less than four times its uncertainty. Muons must either be reconstructed as global muons or tracker muons. Muons with standalone tracks (tracks only in the muon system) are rejected. Muons must pass a particle flow-based isolation requirement, where the relative particle flow isolation within a $\Delta R = 0.3$ cone is defined as

$$\text{(6.1)} \qquad \mathcal{I} \equiv \left(\sum p_T^{\text{charged}} + \max \left[0, \sum p_T^{\text{neutral}} + \sum p_T^{\gamma} - p_T^{\text{PU}}(\ell) \right] \right) / p_T^{\ell}.$$

Muons must satisfy $\mathcal{I} < 0.35$.

For muons with $p_T < 200$ GeV, muons satisfying the common requirements and identified by the particle flow identification algorithm are selected. For muons with $p_T > 200$ GeV, muons are selected if they pass the particle flow identification or if they pass a set of high-p_T requirements. These requirements are the following: the muon must be matched to segments in at least two muon stations; satisfy $\frac{p_T}{\sigma_{p_T}} < 0.3$, $|d_{xy}| < 0.2$ cm, $|d_z| < 0.5$ cm, have at least one pixel hit, and have tracker hits in at least six tracker layers.

In 2016, data-taking was affected by a problem in which the level one trigger sent only one candidate per $60°$ sector instead of up to three [REF?]. As a result, when two muons in the same endcap had a low $\Delta\phi$ separation, only one would fire the trigger. To account

for this, 2016 data events containing identified muons with $\Delta\phi < 70°$ in the same endcap region are rejected.

Identification efficiencies and scale factors corresponding to the muon identification above are measured and provided by the $H \to ZZ$ analysis working group, and are shown in Figure [FIG].

6.3. Electron Selection

Electrons are identified using a boosted decision tree multivariate (MVA) discriminator trained on Drell-Yan plus jets simulation with prompt electrons matched to generator-level objects as signal and unmatched and non-prompt electrons as background. The features used in the training include p_T, supercluster η, shower shape variables, ratio of hadronic to electromagnetic energy, track and pixel hit variables, and isolation variables [REF]. These features are sensitive to bremsstrahlung along the electron trajectory, momentum-energy matching between electron trajectory and ECAL cluster, shower shape, and electrons from photon conversions. Since isolation features are included in the training of the discriminator, there is no need for separate isolation cuts. Electrons pass the identification requirement of the H $\to$ Zγ analysis if their discriminator score is higher than a loose working point value, corresponding to 98% signal efficiency. In addition to the MVA cut, electrons must have $|d_{xy}| < 0.5$ cm and $|d_z| < 1$ cm with respect to the primary vertex. Electrons with $p_T \leq 7$ GeV are rejected.

Electron identification efficiencies and scale factors are measured using a tag and probe method on dielectron events near the Z boson peak. These electrons must pass the single electron trigger with a p_T threshold of 27 (32) GeV for 2016 (2017/2018) data.

The dielectron mass must be between 60 and 120 GeV. The tag electron must pass a tight cut-based electron identification and have $p_T >$ 30 (35) GeV for 2016 (2017) and $|\eta| <$ 2.5. The probe electron must pass the loose MVA identification cut and associated impact parameter and p_T cuts described above. Identification efficiencies and scale factors are measured and applied in bins of p_T and supercluster η. The scale factors for each data-taking year are shown in Figures [FIGS].

6.4. Photon Selection

Photons are identified using a boosted decision tree MVA discriminator trained on photon plus jets simulation. The features used in the training include supercluster kinematics, isolation variables, and shower shape variables [REF]. To reject electrons faking photons, a conversion-safe electron veto is applied. In order to improve agreement between simulation and data in the H $\rightarrow$ Zγ analysis, shower shape corrections are taken from the Higgs to diphoton analysis [REF] and applied to simulated events. The original MVA provided by the EGamma POG is then reevaluated after these corrections. The following features are corrected: R_9, defined as the ratio of energy in the 5x5 array of ECAL crystals to the supercluster energy; S_4, defined as the ratio of the maximum energy 2x2 array to the energy of the 5x5 array; the energy weighted shower widths σ_η and σ_ϕ; the energy weighted widths by crystal index $\sigma_{i\eta i\eta}$ and $\sigma_{i\eta i\phi}$; photon isolation, charged isolation with respect to the primary vertex, and charged isolation with respect to the worst vertex choice. The validity of the shower shape corrections is checked using tag and probe procedures for $Z \rightarrow e^+ e^-$ events where the probe electron mimics a photon and $Z \rightarrow \mu^+ \mu^-$ events with an FSR photon. A comparison of the agreement between

uncorrected and corrected simulation with data is shown in Figures [FIGS]. Simulation comparisons with data for individual shower shape features before and after correction can be found in Appendix [APPENDIX], Figures [FIGS].

After deriving the corrected photon MVA discriminator, 90% signal efficiency working point cuts for barrel and endcap photons are determined for the H $\rightarrow$ Zγanalysis. The working points are defined based on real photons from SM $Z\gamma$ simulation, and correspond to discriminator scores greater than -0.4 (-0.59) for barrel (endcap) photons. A comparison of these working points with the standard EGamma POG 90% efficiency working points, plotted on the receiver operator characteristic (ROC) curve for SM $Z\gamma$ and Z plus jets simulation, is shown in Figure [FIG]. The efficiency of the photon identification is measured with $Z \rightarrow e^+e^-$ data using a tag and probe technique. The tag electron must pass the single electron trigger with $p_T$threshold 27 (32) GeV in 2016 (2017/2018), tight cut-based identification, have $p_T>$ 30 (35) GeV for 2016 (2017/2018), and have $|\eta| <$ 2.5. The probe electron must pass the photon MVA identification with shower shape corrections described above. The scale factors, measured and applied in bins of $p_T$and supercluster η, are shown in Figure [FIG]. The efficiencies and scale factors for the conversion-safe electron veto are measured by the EGamma POG using $Z\mu^+\mu^-$ events with an FSR photon and applied in the H $\rightarrow$ Zγanalysis. The scale factors are cataloged in Table [TAB].

6.5. Jet Selection

Jets are selected in order to categorize events coming from potential VBF Higgs production, but no jet multiplicity requirement is present for the H $\rightarrow$ Zγselection. In fact,

the majority of simulation and data events selected in the analysis have no jets. However, identifying and selecting jets to categorize VBF events can still significantly improve the sensitivity of the search. Jets are required to pass a loose cut-based identification in 2016 and a tight cut-based identification in 2017 and 2018. These sets of identification cuts are determined and provided by the JetMET POG. Additionally, jets must satisfy $p_T > 30$ GeV, $|\eta| < 4.7$, and $\Delta R > 0.4$ with respect to each lepton and the photon selected in the analysis. An issue with noise in the ECAL endcap in 2017 caused an artificial increase in jet multiplicity in data within a specific kinematic phase space [REF]. To mitigate this, jets are rejected if they have raw $p_T < 50$ GeV and $2.65 < |\eta| < 3.139$. This cut reduces the efficiency to reconstruct dijet pairs by 12% in the specified region. To tag VBF events, we are interested in dijet pairs. To this end, if there are more than two jets satisfying the above criteria, only the two jets with highest p_T are selected.

6.6. Object Corrections

Several standard corrections are applied to the physics objects selected in the analysis. Rochester muon momentum scale and resolution corrections are applied to both data and simulation [REF]. Energy and momentum scale and resolution corrections for electrons and photons are provided by the EGamma POG and applied in the analysis. Jet momentum scale and resolution correctors are provided by the JetMET POG and applied in the analysis.

An additional muon momentum correction is obtained using an FSR photon recovery procedure, and is based on the procedure used by the $H \rightarrow ZZ$ analysis [REF]. FSR photons are not considered during the standard CMS muon reconstruction. As the p_T of FSR

photons from radiating muons is generally very low, particle flow photons are considered, in contrast with the fully reconstructed photons used in the main analysis selection. A particle flow photon must pass a set of cuts in order to be identified as an FSR photon associated with one of the muons selected by the analysis. The photon must have $p_T >$ 2 GeV, $|\eta| < 2.4$, and relative particle flow isolation less than 1.8. It must also satisfy $\Delta R(\gamma, \mu)/p_{T,\gamma}^2 < 0.012$ and $\Delta R(\gamma, \mu) < 0.4$. If multiple particle flow photons pass these requirements, the photon with the smallest $\Delta R(\gamma, \mu)/p_{T,\gamma}^2 < 0.012$ is chosen. Then, the four momentum of the FSR photon is added back to the four momentum of the muon, and the muon kinematics reevaluated. In simulation, we find that the selected FSR photon matches the generator level FSR photon with 93% efficiency. Figure [FIG] shows the dilepton and three body invariant mass among FSR photon-containing signal events with and without applying the FSR recovery correction. The procedure yields a 1% improvement on the three body mass resolution in the muon channel.

6.7. Event Selection

Events are required to have at least one good primary vertex with a reconstructed longitudinal position within 24 cm of the geometric center of the detector and a transverse position within 2 cm of the nominal beam collision point. The vertex with the largest value of summed physics-object p_T^2 is taken to be the primary interaction vertex. The physics objects used in this calculation are derived from information obtained from the charged-particle tracking detectors only. These objects include jets reconstructed by clustering charged particle tracks; the associated missing transverse momentum, defined as the negative vector sum of the p_T of those jets.

Events with two same-flavor opposite sign leptons (e or μ) and a photon are selected. The leading muon (electron) is required to have p_T greater than 25 (20) GeV, and the trailing lepton must have p_T greater than 15 (10) GeV. Electrons (muons) must have $|\eta|$ less than 2.5 (2.4). The photon must have p_T greater than 15 GeV and must satisfy $0 < |\eta| < 1.4442$ or $1.566 < |\eta| < 2.5$. This avoids the calorimeter transition region, in which photon reconstruction is more difficult. The invariant mass of the dilepton system is required to be greater than 50 GeV. In events with multiple dilepton pairs, the pair with mass closest to the nominal Z boson mass [REF] is selected.

Events are required to have a photon satisfying $p_T^\gamma/m_{\ell\ell\gamma} > 0.14$, which suppresses the Z plus jets background without significantly reducing signal efficiency and without significantly shaping the three body mass spectrum. Each lepton must have $\Delta R > 0.4$ with respect to the photon. To reject events with final-state radiation from Drell-Yan processes, we require $m_{\ell\ell\gamma} + m_{\ell\ell} > 185$ GeV. Finally, the three body mass is required to satisfy $105 < m_{\ell\ell\gamma} < 170$ GeV.

6.8. Kinematic Fit

A kinematic fit technique is used to improve the dilepton mass resolution. This constrains the dilepton mass based on the true Z boson lineshape while accounting for the known detector resolution. The position information of the leptons has a negligible impact on the fit, so only the lepton transverse momenta and dilepton mass are included in the fit. The procedure is based on previous studies by the $H \to ZZ$ analysis [REF]. We do not carry out an analogous kinematic fit for the $m_{\ell\ell\gamma}$ invariant mass in order to

avoid any potential bias due to reshaping the mass distribution. The kinematic fit is a maximum likelihood fit defined by the likelihood function below:

(6.2)

$$\mathcal{L}(P_T{}^1, P_T{}^2 | P_T{}^{reco1}, \sigma_{P_T}{}^1, P_T{}^{reco2}, \sigma_{P_T}{}^2)$$

$$= Gauss(P_T{}^{reco1} | P_T{}^1, \sigma_{P_T}{}^1) \cdot Gauss(P_T{}^{reco2} | P_T{}^2, \sigma_{P_T}{}^2) \cdot \mathcal{L}(m_{12} | m_Z)$$

Here, p_T^{reco1} and p_T^{reco2} are the reconstructed transverse momenta of the two leptons, $\sigma_{p_T}^1$ and $\sigma_{p_T}^2$ are the per-lepton transverse momentum resolutions, p_T^1 and p_T^2 are the parameters to optimize, and m_{12} is the invariant mass calculated from p_T^1 and p_T^2. $\mathcal{L}(m_{12} | m_Z)$ is the likelihood given the true Z mass lineshape. The outputs of the fit are p_T^1 and p_T^2. Their uncertainties are saved and used to calculate a per-event uncertainty for the mass measurement. To optimize this procedure, we determine the true generator level Z lineshape from a gluon-gluon fusion H $\rightarrow$ Zγsample. Generator level Z lineshapes for the dielectron and dimuon final states are fit with a single-sided Crystal Ball function plus three gaussian functions. For each event, the likelihood is maximized and the $p_T$information of the refit leptons is updated. Then, the dilepton mass and three body mass are recalculated based on the refit leptons. The mass distributions for before and after the kinematic fit are shown in Figure [FIG]. The level of improvement in the dilepton mass resolution, as measured by σ_{eff}, is shown in in Table 6.1.

We also perform a closure test on the background as shown in Figure [FIG]. The dilepton mass should be constrained, while the three body mass should not be distorted by the kinematic constraint. Indeed, this is what is observed.

Table 6.1. Percent improvement in dilepton and three-body mass resolution (measured by σ_{eff}) after the kinematic fit.

	electron channel		muon channel	
	m_{ee}	$m_{ee\gamma}$	$m_{\mu\mu}$	$m_{\mu\mu\gamma}$
2016	20%	20%	17%	12%
2017	28%	27%	21%	11%
2018	24%	24%	20%	10%

CHAPTER 7

Event Categorization

The sensitivity of the H → Zγ search can be significantly improved by leveraging the differences between signal events arising from different Higgs boson production modes with different final state topologies and kinematics. This is achieved by defining mutually exclusive categories targeting these different types of signal events. In previously published CMS searches for H → Zγ [REFS], cut-based approaches were used to define the categories. Broadly, the presence of at least one additional lepton was used to tag VH and ttH production, a dijet system was used to tag VBF production, a boosted jet was used to tag gluon-gluon fusion events recoiling off of a jet, and properties of the photon were used to divide the remaining events into kinematically distinct untagged categories. The specific definitions of the categories used in the 2016 search at 13 TeV are shown in Figure [FIG].

The categorization of the present search is inspired by the previous CMS searches, but is significantly more sophisticated and better optimized than those searches. The lepton tag category remains, as do the concepts of dijet and untagged categories. However, the dijet and untagged categories are determined via the training of two BDTs, with category boundaries optimized using the resulting BDT scores. Moreover, the boosted category is studied and found to yield no improvement over the categories determined by the BDTs, so the boosted category is dropped in the present analysis. We proceed with a description of the BDT procedures for the untagged and dijet cases. Then, with our BDTs in hand,

we describe the procedure for defining an optimal set of categories to be used in the analysis. We show that this categorization procedure is more optimal than what has been done in past CMS searches.

7.1. Kinematic BDT

The kinematic BDT is used to separate H $\rightarrow$ Zγsignal events from background events. Its primary purpose is to define the untagged categories, which are distinguished primarily by kinematics and the quality of the final state photon. Its secondary use is as an input to the dijet BDT (described below) in order to increase signal to background discrimination in the dijet BDT training. The kinematic BDT is trained on H $\rightarrow$ Zγsignal events from all Higgs production modes and background events from SM $Z\gamma$, Z plus jets, and $t\bar{t}$. The training is restricted to use half of the simulated events, with the other half reserved for category optimization and signal modeling. All training events are required to pass the basic object and event selection described in Chapter 6. Events from 2016, 2017, and 2018 simulation samples in both the muon and electron channels are combined for the training. These samples are weighted by their respective cross sections and weighted by the luminosity of each year.

The following features are used in the training of the kinematic BDT: photon MVA score, photon energy resolution, η of each lepton and the photon, smallest and largest of the two ΔR values of the photon with respect to each lepton, p_T/m for the three body system, and three theoretically motivated decay angles. The angular quantities are defined in Figure [FIG] and their theoretical, generator-level, and reconstructed distributions are shown in Figure [FIG]. The signal and background distributions used in the

training for all kinematic BDT features are shown in Figure [FIG]. Table [TAB] ranks the kinematic BDT training features in order of importance for the final discriminator. We observe that the most discriminating feature is the photon MVA identification score which mainly serves distinguish real photons from misreconstructed jets. A comparison of the BDT distributions in the training and test samples allows us to check for any possible overtraining of the model. This is shown in Figure [FIG], where no significant overtraining is observed.

The output of the kinematic BDT in data and simulation is shown in Figure [FIG]. We observe good agreement between data and simulation. For additional validation of the method, two control regions are defined. A background-enriched control region is defined by $m_{\ell\ell} + m_{\ell\ell\gamma} < 185$ GeV or $p_T^\gamma < 15/110$, and an irreducible background-enriched (SM $Z\gamma$) control region is defince by $m_{\ell\ell} + m_{\ell\ell\gamma} < 185$ GeV and $80 < m_{\ell\ell} < 100$ GeV. The data and simulation distributions for the kinematic BDT for these control regions are shown in Figures [FIGS]. Again, we observe good agreement between data and simulation.

7.2. Dijet BDT

The dijet BDT is used to discriminate signal from VBF Higgs production from other dijet signal and background events. The training and evaluation of the dijet BDT is only carried out on events which have two jets selected according to the requirements in section [SEC]. The training is performed using VBF H $\rightarrow$ Zγevents as signal and gluon-gluon fusion H $\rightarrow$ Zγ, SM $Z\gamma$, Z plus jets, and $t\bar{t}$ events as background. VH and ttH H $\rightarrow$ Zγevents are neglected, as their contribution is negligibly small. Only about 65% of the signal jets correspond to the true VBF jets in which we are interested. Because of this,

we perform an additional matching procedure for these jets. In order for an event to be used in the training, both reconstructed jets must be matched to generator-level partons, which must also be matched to the generator-level p_T values of the true VBF partons. As with the kinematic BDT, only half of the simulated events are used for the training procedure, with the remainder used for category optimization and signal modeling. Again, events from all data-taking years in both the muon and electron channels are combined for the training. The samples are weightd by their respective cross sections and weighted by the luminosity of each year.

The input features to the dijet BDT training include $\Delta\eta$(j, j), $\Delta\phi$(j, j), $\Delta\phi(Z\gamma$, jj), $\Delta R(j, \gamma)$, $p_T^{j1,2}$, p_T^t (defined as $\frac{2|p_{Tx}^Z p_{Ty}^\gamma - p_{Ty}^Z p_{Tx}^\gamma|}{p_T^H}$), system p_T balance (defined as $|\frac{\sum_{Z,\gamma,j_1,j_2} \vec{p}_{transverse}}{\sum_{Z,\gamma,j_1,j_2} p_T}|$), photon Zeppenfeld variable (defined as $|\eta_\gamma - \frac{\eta_{j1}+\eta_{j2}}{2}|$), and kinematic MVA score. The angular quantities provide discrimination between VBF and non-VBF events, while the kinematic MVA score and p_T^t provide discrimination between signal and background events. The p_T^t [REF] is the component of the transverse momentum of the $Z\gamma$ system that is perpendicular to the difference of the 3-momenta of the Z boson and the photon candidate.

7.3. Categorization Procedure

The procedure used for event categorization is described below.

(1) Events with at least one additional electron or muon with $p_T > 7$ (5) GeV are assigned to the lepton-tagged category.

(2) Events not assigned to the lepton-tagged category, but which contain two jets satisfying the selection requirements described in Section 6, are classified as dijet events, indicative of possible VBF production. If multiple dijet pairs exist within

an event, the two jets with highest p_T are considered. The subdivision of dijet events into a set of three dijet categories is described later in this section. The VBF BDT classifier is trained to separate VBF signal events from ggH+jets and background events. The distribution of $\mathcal{D}_\mathrm{VBF}$ is shown in Fig. 7.1 (left) for both simulated event samples and data.

(3) Events not assigned to the lepton-tagged or dijet categories are classified as un-tagged events. The subdivision of untagged events into a set of four untagged categories is described later in this section. The kinematic BDT classifier is trained to distinguish signal events from background events based on the kine-matics of the leptons and photon in the $Z\gamma$ candidate system, as well as on the measured properties of these physics objects. The distribution of $\mathcal{D}_\mathrm{kin}$ is shown in Fig. 7.1 (right) for both simulated samples and data.

The subdivision of dijet and untagged events into categories is based on the VBF BDT and kinematic BDT discriminants. Category boundaries are defined as mutually exclusive regions of $\mathcal{D}_\mathrm{VBF}$ and $\mathcal{D}_\mathrm{kin}$. The locations of the boundaries defining the categories are optimized by iterating over all possible combinations of boundaries using $\sum_i^n S_i^2/B_i$ as a figure-of-merit. The variables S_i and B_i represent the number of expected signal and background events in the ith category, and n is the total number of categories. We consider categories with boundaries corresponding to signal efficiencies between 0–100% in 10% increments. The optimization procedure results in three dijet categories for the VBF BDT and four untagged categories for the kinematic BDT. The lowest $\mathcal{D}_\mathrm{kin}$ boundary corresponds to the 10% point in integrated signal efficiency, and events below the 10% point are excluded from the analysis to preserve the stability of the background model.
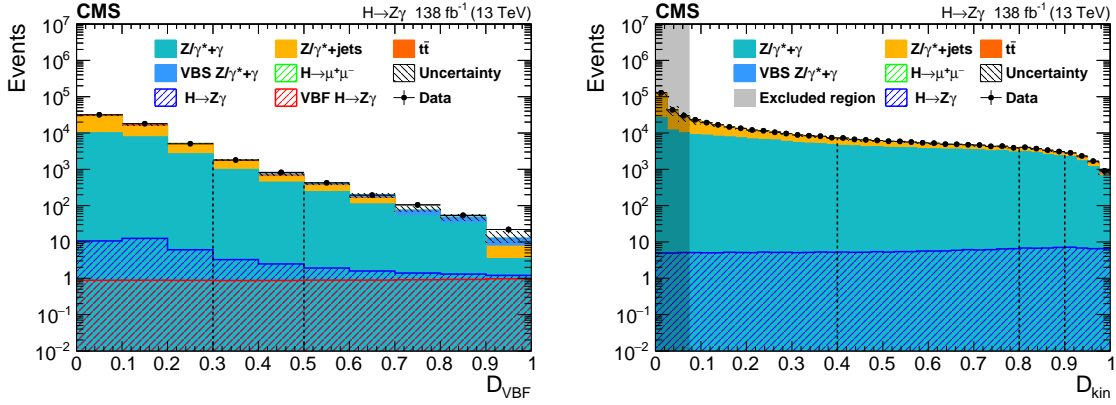
Figure 7.1. The $\mathcal{D}_{\text{VBF}}$ (left) and $\mathcal{D}_{\text{kin}}$ (right) distributions for signal, simulated background, and data. The $\mathcal{D}_{\text{VBF}}$ distribution includes only dijet-tagged events, and the $\mathcal{D}_{\text{kin}}$ distribution includes only untagged events. The sum of contributions from all signal production mechanisms is shown by the blue line, while the contribution from only the VBF mechanism is shown by the red line. The uncertainty band incorporates all statistical and systematic uncertainties in the expected background. The dashed lines indicate the boundaries for the dijet and untagged categories. The gray shaded region in the $\mathcal{D}_{\text{kin}}$ distribution is excluded from the analysis.

Table 7.1. Summary of the category definitions. The lepton-tagged category requires at least one additional electron or muon. Dijet categories are defined by regions of $\mathcal{D}_{\text{VBF}}$ and untagged categories are defined by regions of $\mathcal{D}_{\text{kin}}$.

Lepton	Dijet 1	Dijet 2	Dijet 3	Untagged 1	Untagged 2	Untagged 3	Untagged 4
≥ 1 e, μ	\multicolumn{3}{}{$\mathcal{D}_{\text{VBF}}$ selection}		$\mathcal{D}_{\text{kin}}$ selection		
	0.5–1.0	0.3–0.5	0.0–0.3	0.9–1.0	0.8–0.9	0.4–0.8	0.1–0.4

The full categorization and optimization procedure results in the following eight mutually exclusive categories: one lepton-tagged category, three dijet categories, and four untagged categories. The category definitions are summarized in Table 7.1.

Table 7.2 lists the event categories used in the analysis, along with the expected event yields for an $m_{\text{H}} = 125.38\,\text{GeV}$ signal arising from ggH, VBF, VH, and $t\bar{t}$H production, as well as a 6% resonant background contribution from FSR from $H \rightarrow \mu^+\mu^-$. The

Table 7.2. Yields and approximate significance ($S/\sqrt{B}$) for each category, where S and B are the expected number of signal and background events in the narrowest $m_{\ell^+\ell^-\gamma}$ interval containing 95% of the expected signal distribution. Also shown is the $m_{\ell^+\ell^-\gamma}$ resolution, computed using the narrowest interval containing 68% of the expected signal distribution.

$138\,\mathrm{fb}^{-1}$	Lepton		Dijet 1	Dijet 2	Dijet 3	Untagged 1	Untagged 2	Untagged 3	Untagged 4
SM signal yield									
ggH	0.51	e^+e^-	1.10	1.62	9.44	6.89	7.35	29.8	22.5
		$\mu^+\mu^-$	1.41	2.05	12.1	8.52	9.17	38.0	29.0
VBF	0.09	e^+e^-	1.94	0.76	1.13	0.71	0.35	0.92	0.51
		$\mu^+\mu^-$	2.40	0.97	1.43	0.89	0.43	1.18	0.65
$\mathrm{VH}+\mathrm{t\bar{t}H}$	1.84	e^+e^-	0.04	0.13	1.89	0.31	0.17	0.45	0.27
		$\mu^+\mu^-$	0.05	0.16	2.36	0.39	0.21	0.57	0.33
SM resonant background									
$\mathrm{H}\to\mu^+\mu^-$	0.14	$\mu^+\mu^-$	0.27	0.27	0.43	0.62	0.49	2.02	1.78
Mass resolution (GeV)	2.12	e^+e^-	1.91	2.06	2.15	1.80	1.97	2.12	2.33
		$\mu^+\mu^-$	1.52	1.61	1.72	1.37	1.42	1.62	1.83
Data yield	1485		168	589	11596	1485	1541	2559	17608
$\boldsymbol{S/\sqrt{B}}$	0.06		0.54	0.24	0.26	0.45	0.35	0.53	0.30

expected event yield from $\mathrm{H}\to\tau^+\tau^-$ is estimated to be below the 1% level relative to the $\mathrm{H}\to\mathrm{Z}\gamma$ yield and is neglected. The dominant contribution to the signal yield is generally from ggH production, except in the lepton-tagged category, in which VH and $\mathrm{t\bar{t}H}$ events dominate, and in the dijet 1 category, in which VBF events dominate. The categorization procedure increases the sensitivity of the analysis by 24% with respect to an inclusive event selection. The product of signal acceptance and efficiency for $\mathrm{pp}\to\mathrm{H}\to\mathrm{Z}\gamma\to\ell^+\ell^-\gamma$ for $m_\mathrm{H}=125.38\,\mathrm{GeV}$ is 23 (29)% in the electron (muon) channel.

7.4. Dropping of Boosted Category

In the previous 2016 analysis, a boosted Higgs category was used. This category is defined by a simple cut on the p_T of the $\ell\ell\gamma$ system. However, for the current MVA-based analysis, the Higgs kinematics are already included in the BDT training. In this case, it stands to reason that the boosted category may no longer be useful for improving the analysis senstivity. We therefore test the possibility of constructing a boosted category before constructing the untagged categories to see whether the combined significance is increased or decreased. The left plot of Figure 7.2 shows the p_T spectrum of the $\ell\ell\gamma$ system for various signal MC samples. We see that the p_T spectrum for the gluon-gluon fusion process is similar to the main backgrounds, while the other signal processes are more boosted. The right plot shows the upper limit on σ/σ_{SM} as a function of a p_T cut on the $\ell\ell\gamma$ system, where the background estimation is taken from data. The upper limit for the boosted category (shown in blue) worsens as the p_T cut increases. This is expected since we lose a substantial amount of gluon-gluon fusion signal. However, the combined limit of the untagged categories improves. The overall combined limit (shown in black) is rather flat as a function of the p_T of the $\ell\ell\gamma$ system. For the overall combined limit, we observe fluctuations beyond a p_T cut of 60 GeV, so we choose 60 GeV as the cut defining a potential boosted category. Comparing the combined significance of the boosted category plus untagged categories to the untagged categories alone, we see that the combined significance is substantially higher for the untagged categories alone. An explanation for this behavior is as follows. In the current analysis flow, dijet events are selected before boosted events. We loosen the dijet selection requirements, which causes the boosted

events to migrate to the dijet categories and reduces the significance of boosted category.

Therefore, we choose not to use a boosted category in this version of the analysis.
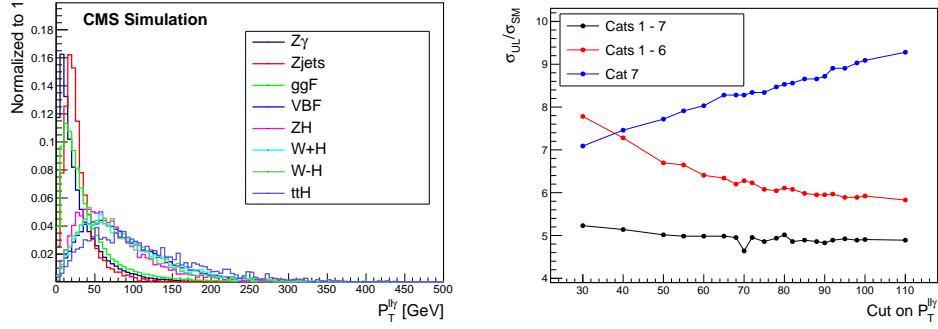
Figure 7.2. Left: $\ell\ell\gamma$ p_{T} for different signal MC processes. Right: optimization of the boosted category by scanning the expected limit as function of $\ell\ell\gamma$ p_{T}.
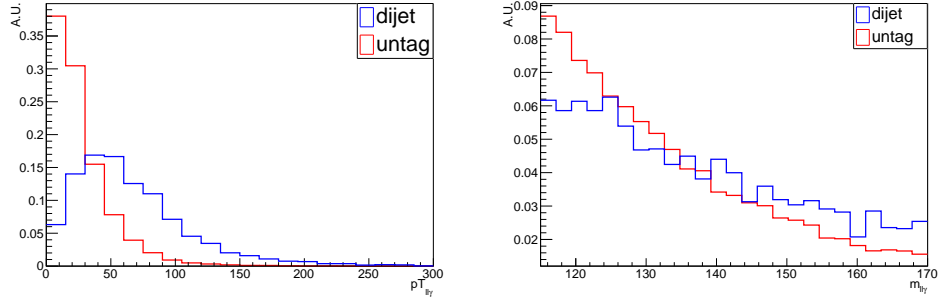
Figure 7.3. Left: $\ell\ell\gamma$ p_{T} for untagged events and dijet events. Right: The $m_{\ell\ell\gamma}$ spectrum of untagged events and dijet events.

CHAPTER 8

Signal and Background Modeling

In each category, we carry out a shape analysis to search for a signal peak in the $m_{\ell\ell\gamma}$ spectrum. The signal and background mass shapes are modeled using parametric functions.

8.1. Signal Modeling

The signal model is defined as the sum of Crystal Ball [**?**] and Gaussian functions. The signal shape parameters are determined by fitting this model to simulated signal events in each category. To account for differences in mass resolution, these fits are performed separately for the event samples used to model each data-taking year, as well as for muon and electron channel events. This results in six signal models that are summed to give the total signal expectation in a given category. Separate sets of parameter values are found by fitting simulated events with m_H of 120, 125, and 130 GeV. Using linear interpolation, parameter values are also determined at 1 GeV intervals in m_H from 120–130 GeV, as well as at 125.38 GeV. In the fit to data, the mean and resolution parameters are allowed to vary subject to constraints from several systematic uncertainties, described in Section 9, while the remaining parameters are held fixed.

Figure [REF] (REF) shows the signal fits for the $m_H = 125$ GeV simulated samples in the electron (muon) channel for the 2016 data-taking period. Figure [REF] shows the signal fits for the lepton-tagged category for the 2016 data-taking period, in which electron

and muon channel events are combined. In the lepton-tagged category, the signal shape is modeled separately for ZH and WH production in order to account for differences in potential lepton mispairing. Signal fits for the 2017 and 2018 data-taking periods are shown in Figs. [REF] in Appendix [REF].

8.2. Resonant Background Modeling

The process $H \to \mu^+\mu^-$, where at least one muon radiates an FSR photon, is expected to contribute at the 6% level relative to the $H \to Z\gamma$ signal yield. Because of this, we treat it as a resonant background. To model the $H \to \mu^+\mu^-$ shape, we perform fits to the $m_{\ell^+\ell^-\gamma}$ distributions in simulated event samples, following the same procedure used for the signal fits described above. As for the signal, the analytic model is taken to be the sum of Crystal Ball and Gaussian functions. The resulting fits are shown in Figs. [REF].

8.3. Nonresonant Background Modeling

As in previous versions of the analysis, we use a parametric fit to the $m_{\ell^+\ell^-\gamma}$ spectrum in data to estimate the background. The background model in each category is obtained from the data using the discrete profiling method [?]. This technique accounts for the systematic uncertainty associated with choosing an analytic functional form to fit the background. The background function is chosen from a set of candidate functions via a discrete nuisance parameter in the fit. These functions are derived from the data in each category, with muon and electron events from all data-taking years combined. The discrete profiling method is described in further detail later in this section.

In Run 1 [?], the background fitting started from $100\,\text{GeV}$. At the lower end of this fit range, below about 115 GeV, there is a notable turn-on in the mass distribution that
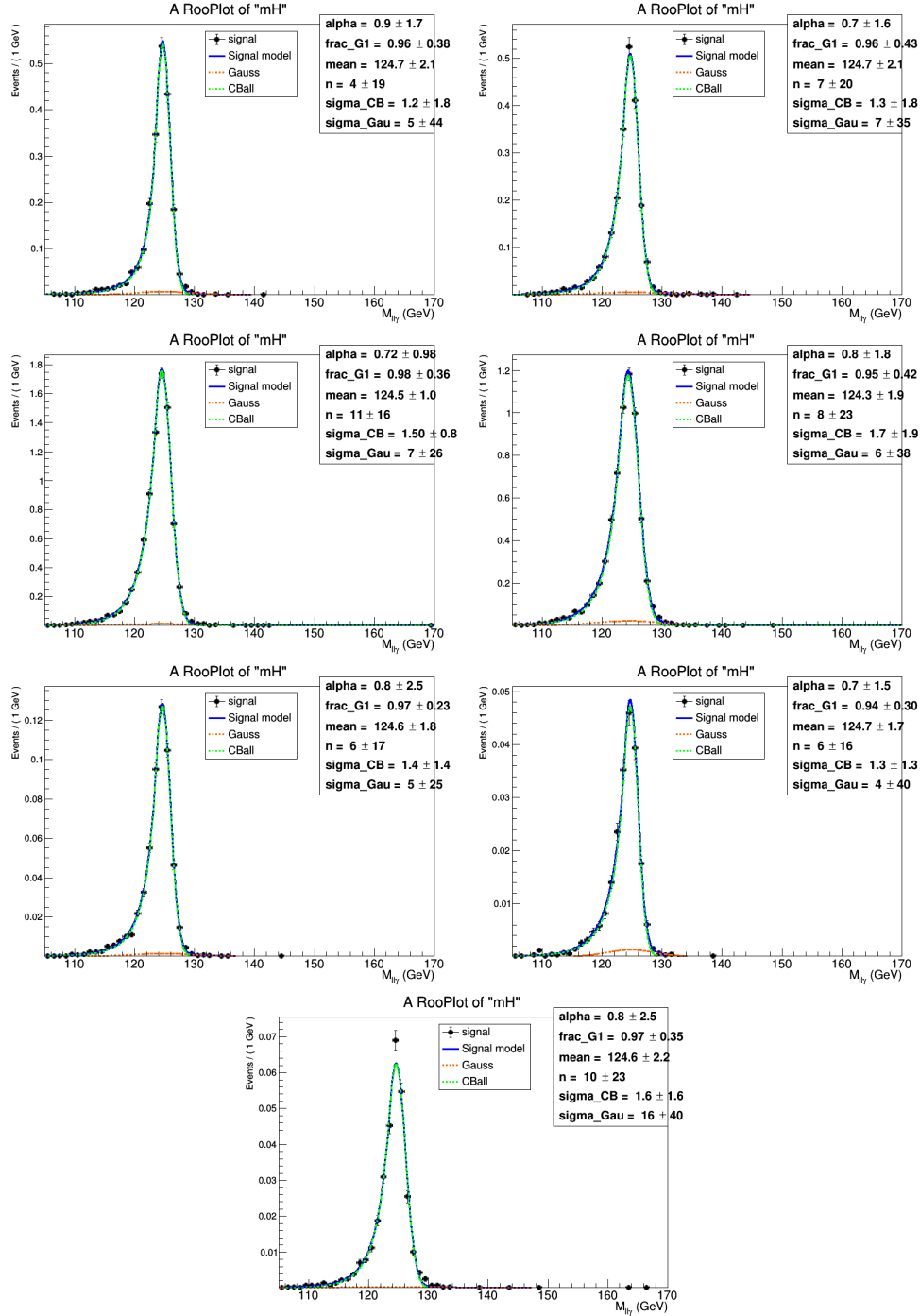
Figure 8.1. Fits to simulated $m_{\ell^+\ell^-\gamma}$ signal distributions in the electron channel for $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ for the 2016 data-taking period. The blue line shows the total fit function, the green line shows the Crystal Ball function component, and the red line shows the Gaussian function component. The top four plots correspond to the untagged categories, and the bottom three plots correspond to the dijet categories.

Figure 8.2. Fits to simulated $m_{\ell^+\ell^-\gamma}$ signal distributions in the muon channel for $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ for the 2016 data-taking period. The blue line shows the total fit function, the green line shows the Crystal Ball function component, and the red line shows the Gaussian function component. The top four plots correspond to the untagged categories, and the bottom three plots correspond to the dijet categories.

Figure 8.3. Fits to simulated $m_{\ell^+\ell^-\gamma}$ signal distributions in the electron and muon channels combined in the lepton-tagged category for $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ for the 2016 data-taking period. The left plot shows the fit to simulated ZH production events, and the right plot shows the fit to simulated WH production events. The blue line shows the total fit function, the green line shows the Crystal Ball function component, and the red line shows the Gaussian function component.

arises due to the presence of the on-shell Z boson. As such, the Run 1 analysis used PDFs composed of a Gaussian turn-on component convoluted with a falling spectrum component. In the first Run 2 analysis using only the 2016 data, this procedure was significantly changed. The lower bound of the fit range was updated to 115 GeV in order to avoid the turn-on. This meant that the functional forms used to model the background could be simplified, as they no longer required the turn-on component. Now, in the full Run 2 analysis with 2016, 2017, and 2018 data, we have returned to the Run 1 approach of fitting the turn-on. The reason for this is that the turn-on can have an impact on the mass distribution, even above 115 GeV. As a consequence, avoiding the turn-on component of the fit leads to some mismodeling in the low mass region. Moreover, this mismodeling can be enough in some cases to lead to significant bias in the signal strength measurement. The solution has been to set the lower bound at 105 GeV such that the turn-on is roughly Gaussian and can be modeled and fit properly.
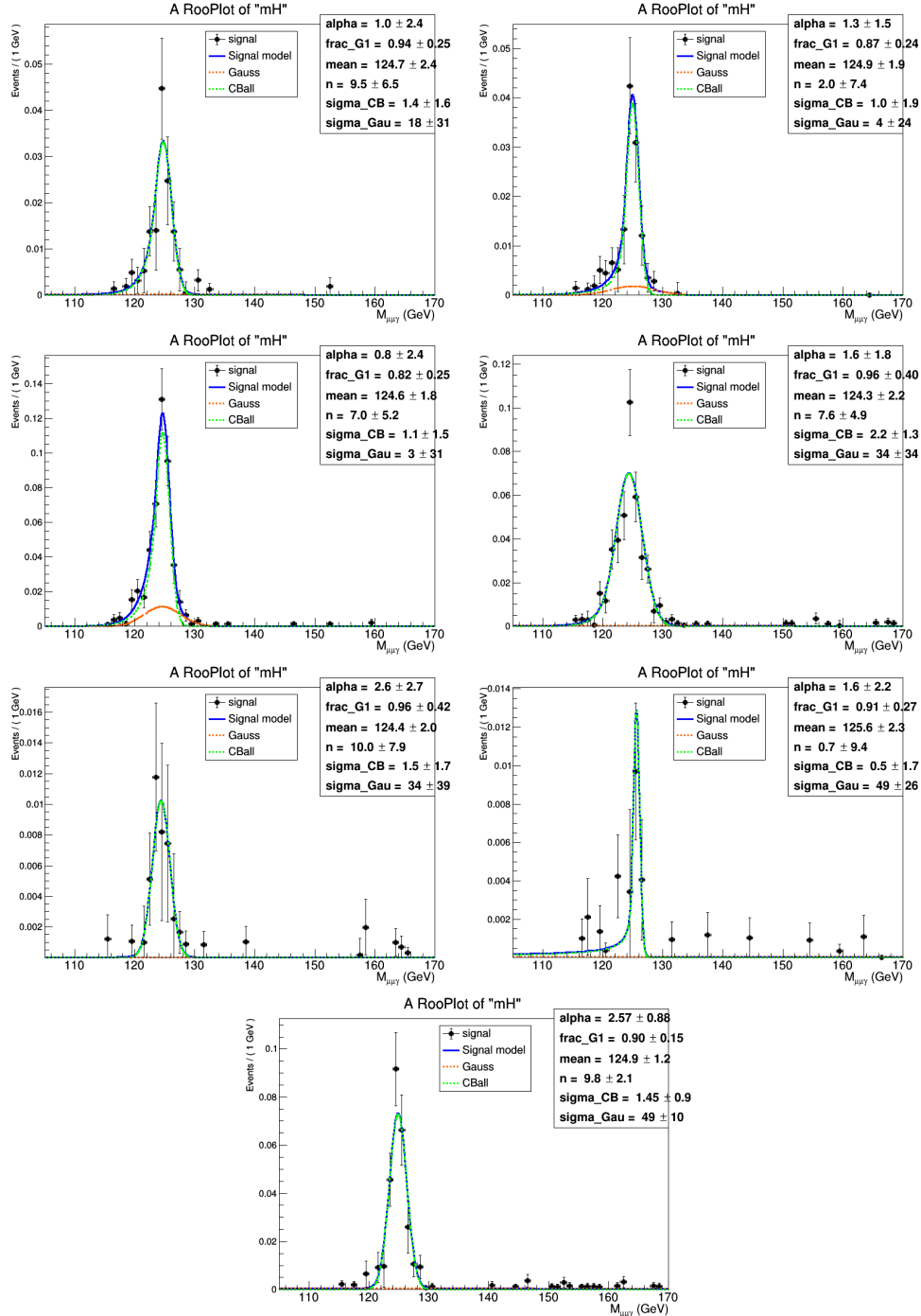
Figure 8.4. Fits to simulated $m_{\mu^+\mu^-\gamma}$ resonant background distributions from $H \to \mu^+\mu^-$ for $m_H = 125\,\text{GeV}$ for the 2016 data-taking period. The blue line shows the total fit function, the green line shows the Crystal Ball function component, and the red line shows the Gaussian function component. The top four plots correspond to the untagged categories, and the bottom three plots correspond to the dijet categories.

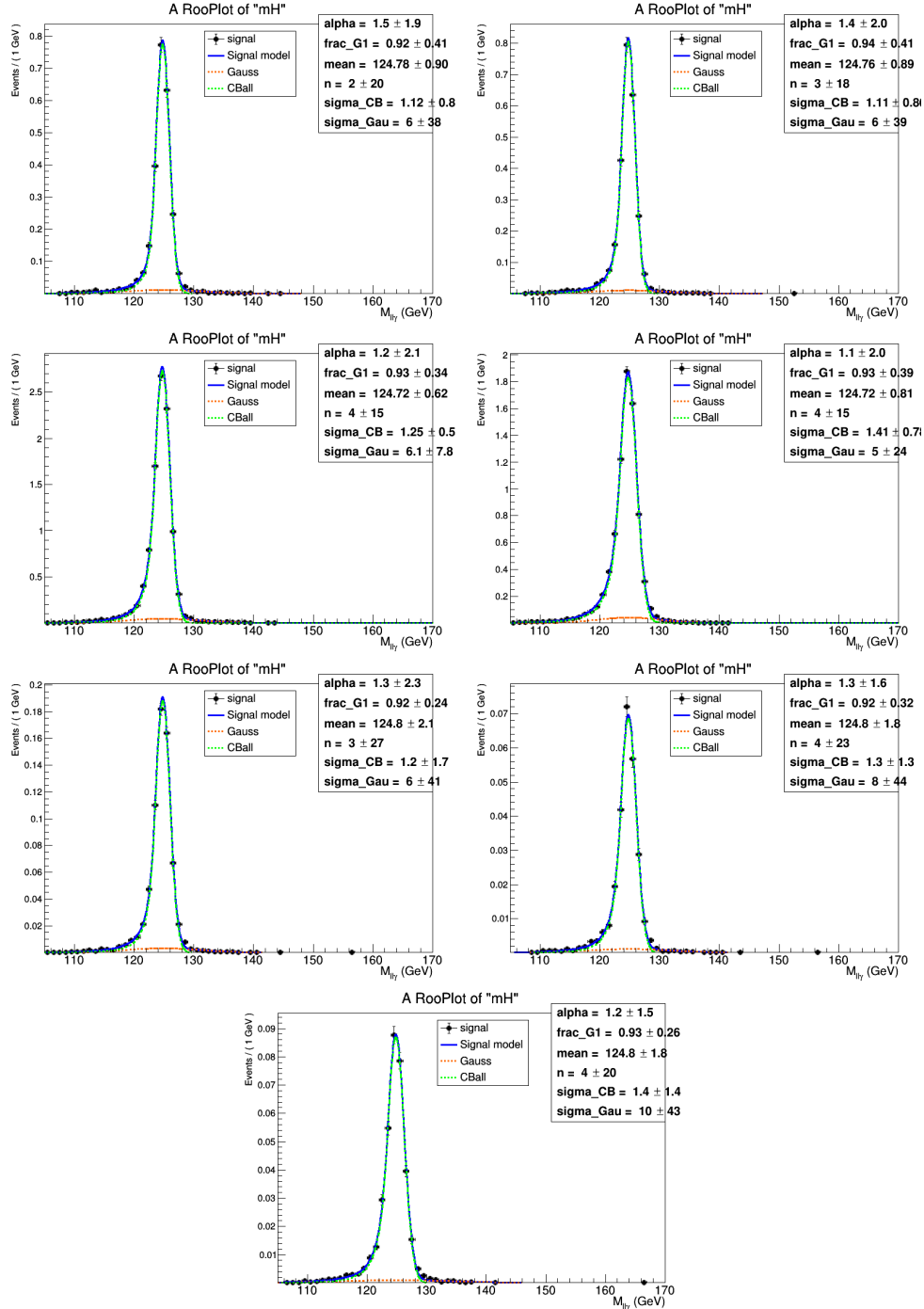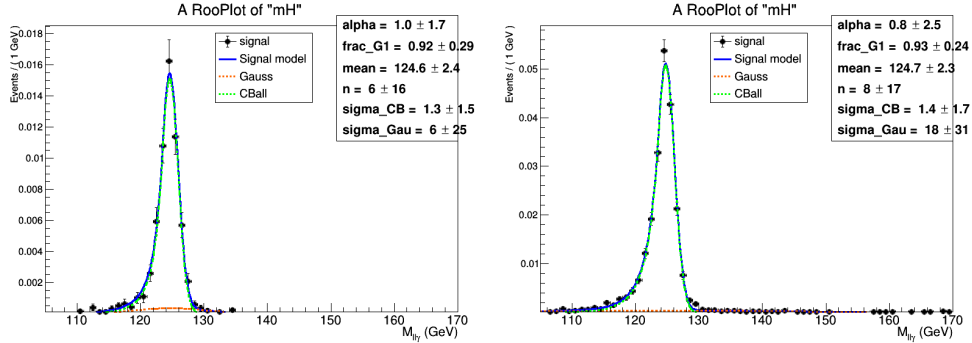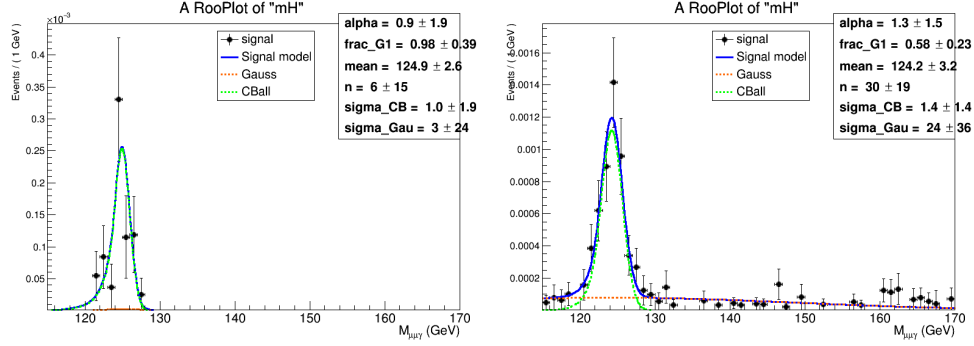Figure 8.5. Fits to simulated $m_{\ell^+\ell^-\gamma}$ resonant background distributions from H $\rightarrow \mu^+\mu^-$ in the electron and muon channels combined in the lepton-tagged category for $m_H = 125\,\text{GeV}$ for the 2016 data-taking period. The left plot shows the fit to simulated ZH production events, and the right plot shows the fit to simulated WH production events. The blue line shows the total fit function, the green line shows the Crystal Ball function component, and the red line shows the Gaussian function component.

8.3.1. General Form of PDFs

The $m_{\ell^+\ell^-\gamma}$ spectrum consists of a turn-on peak around 110–115 GeV and a falling spectrum in the high-mass tail, where the turn-on peak is driven by the photon p_T selection. These features are modeled by the convolution of a Gaussian function with a step function multiplied by one of several falling spectrum functions. The complete function has the general form:

$$(8.1) \qquad \mathcal{F}(m_{\ell^+\ell^-\gamma}; \mu_G, \sigma_G, s, \vec{\alpha}) = \int_{105}^{170} \mathcal{N}(m_{\ell^+\ell^-\gamma} - t; \mu_G, \sigma_G)\Theta(t; s)f(t; \vec{\alpha})dt,$$

where t is the integration variable for the convolution, $\mathcal{N}(m_{\ell^+\ell^-\gamma} - t; \mu_\mathrm{G}, \sigma_\mathrm{G})$ is the Gaussian function with mean μ_G and standard deviation σ_G, $\Theta(t; s)$ is the Heaviside step function with step location s, and $f(t; \vec{\alpha})$ is the falling spectrum function with shape parameters $\vec{\alpha}$. The total number of parameters in the fit is equal to three plus the number of parameters in the falling spectrum component.

8.3.2. Falling Spectrum Component Families

The choice of falling spectrum component to model the background in a given category is not a priori known. For this reason, we consider several families of PDFs for the falling spectrum. The families tested include Bernstein polynomials, exponential series, power law series, and Laurent series. These families are described in more detail as follows:

- exponential series of order N:

$$(8.2) \qquad \mathrm{Exp_N}(m_{\ell\ell\gamma}) = \sum_{i=1}^{N} f_i e^{p_i\, m_{\ell\ell\gamma}}$$

  with $2N$ free parameters: $p_i < 0$ and f_i. The lowest order considered has $N = 1$, i.e. one term. The next order has 2 exponential terms, but 3 parameters.

- power law series of order N:

$$(8.3) \qquad \mathrm{Pow_N}(m_{\ell\ell\gamma}) = \sum_{i=1}^{N} f_i m_{\ell\ell\gamma}^{p_i}$$

  with $2N$ free parameters: $p_i < 0$ and f_i. The lowest order considered has $N = 1$, i.e. one term.

- Bernstein polynomial of order N:

$$(8.4) \qquad \mathrm{Bern_N}(m_{\ell\ell\gamma}) = \sum_{i=0}^{N} f_i b_{i,N}(m_{\ell\ell\gamma})$$

$$(8.5) \qquad \mathrm{b_{i,N}}(m_{\ell\ell\gamma}) = \binom{N}{i} m_{\ell\ell\gamma}^{i}(1 - m_{\ell\ell\gamma}^{N-i})$$

with N free parameters f_i. The lowest order considered has $N = 1$.

- Laurent series with 2, 3, 4, 5, or 6 terms, where N + 1 equals the number of free parameters:

$$(8.6) \qquad \mathrm{Lau_1}(m_{\ell\ell\gamma}) = f_2 m_{\ell\ell\gamma}^{-4} + f_3 m_{\ell\ell\gamma}^{-5}$$

$$(8.7) \qquad \mathrm{Lau_2}(m_{\ell\ell\gamma}) = f_1 m_{\ell\ell\gamma}^{-3} + f_2 m_{\ell\ell\gamma}^{-4} + f_3 m_{\ell\ell\gamma}^{-5}$$

$$(8.8) \qquad \mathrm{Lau_3}(m_{\ell\ell\gamma}) = f_1 m_{\ell\ell\gamma}^{-3} + f_2 m_{\ell\ell\gamma}^{-4} + f_3 m_{\ell\ell\gamma}^{-5} + f_4 m_{\ell\ell\gamma}^{-6}$$

$$(8.9) \qquad \mathrm{Lau_4}(m_{\ell\ell\gamma}) = f_1 m_{\ell\ell\gamma}^{-2} + f_2 m_{\ell\ell\gamma}^{-3} + f_3 m_{\ell\ell\gamma}^{-4} + f_4 m_{\ell\ell\gamma}^{-5} + f_5 m_{\ell\ell\gamma}^{-6}$$

$$(8.10) \qquad \mathrm{Lau_5}(m_{\ell\ell\gamma}) = f_1 m_{\ell\ell\gamma}^{-2} + f_2 m_{\ell\ell\gamma}^{-3} + f_3 m_{\ell\ell\gamma}^{-4} + f_4 m_{\ell\ell\gamma}^{-5} + f_5 m_{\ell\ell\gamma}^{-6} + f_6 m_{\ell\ell\gamma}^{-7}$$

8.3.3. Discrete Profiling Method

In each category, there is ambiguity about which PDF to choose to model the background. In previous versions of this analysis, a detailed bias study was performed for each falling

spectrum component family, and the family with the least bias was chosen. However, no uncertainty was assigned due to the choice of background PDF. In principle, the choice of PDF is a source of systematic uncertainty in the measurement. Therefore, the discrete profiling method was since developed as a way to assess this uncertainty in a formal way. In this method, the choice of PDF for the background fit is included as a discrete nuisance parameter in the full likelihood function. All reasonable families of functions should be considered, and additionally, it is found that certain orders within the same family should be considered. The set of functions profiled in a given category is referred to as an *envelope.* The exact composition of the envelope in each category is based on a set of selection requirements that account for goodness of fit, an F-test procedure, and an assessment of bias. The details of this selection will be described in the following section. When fitting the background, all functions in the envelope are tried, with a penalty term added to the likelihood to account for the number of free parameters in the fit. When a measurement is made of a parameter of interest, for example the signal strength, the PDF with the smallest negative log likelihood, profiled as a function of the parameter of interest, is used. The profiling allows the best fit function to change for different values of the parameter of interest, so we can say that we carry out the fit with the full envelope of functions. The envelope yields a profile likelihood curve that is broader than the profile likelihood curve obtained from any individual function. This increase in breadth reflects the uncertainty associated with the choice of background PDF.

8.3.4. Envelope Selection

Individual functions are added to the envelope in a given category based on a series of selection requirements. In general, the requirements are related to three main factors: goodness of fit, F-test, and bias. The simplest requirement is a basic goodness of fit cut. For each function considered, the chi-squared per number of degrees of freedom is calculated. This is then converted into a p-value, or chi-squared probability. If the chi-squared probability is greater than 0.01, we consider the quality of the fit good enough for the envelope. Any function failing this requirement is thrown away, as it cannot be considered a reasonable description of the data.

The second consideration is the result of an F-test. The F-test is a way to compare functions within a specific family of falling spectrum component PDFs. It compares the fit of a lower order function in the family to that of a higher order function. While it is expected that the higher order function will always provide a better fit, the question is whether the higher order fit is better in a statistically significant way. If it is not significantly better, there is a motivation to stick to the lower order function and exclude the higher order function from the envelope. The details of the F-test procedure are as follows. First, the lowest order PDF in a given family is fit to the data in a given category. Then, the next highest order PDF is fit to the same data. The difference between twice the negative log likelihood between the two fits, $2\Delta NLL_{N+1} = 2(NLL_N - NLL_{N+1})$, can be used to determine whether the data are better supported by the higher order PDF. More precisely, $2\Delta NLL_{N+1}$ is distributed as a χ^2 PDF with M degrees of freedom, where M is the difference in the number of free parameters between the order $N+1$ PDF and the order N PDF. For example, in the case of the first two orders of the exponential family,

$M = 4 - 2 = 2$, and for the Bernstein polynomial family, $M = 3 - 2 = 1$. After carrying out the fits and computing the negative log likelihoods, a p-value is then calculated as

$$(8.11) \qquad \qquad \text{p} = P(2\Delta NLL > 2\Delta NLL_{N+1} | \chi^2(M)).$$

If the p-value is less than 0.05, we state that the higher order PDF is supported by the data, and the procedure is repeated for the next highest order PDF. Alternatively, if the p-value is greater than 0.05, the higher order PDF is assumed to be too flexible given the data, and the F-test ends, having found the highest order PDF supported by the data. This procedure is done to find the highest-order PDF supported by the data for each PDF family. In this analysis, we consider fits up to order 5 for each falling spectrum family. In each family, we identify the highest order function with an F-test probability below 0.05 and include it in the envelope. We then also include all lower order functions within the family that pass the chi-squared goodness of fit criterion.

The final consideration in selecting functions for the envelope is bias and coverage. After constructing an initial envelope in a given category based on the chi-squared and F-test results, an envelope bias study is performed. The full details of the envelope bias studies for each category can be reviewed in Section 8.3.6. In a nutshell, toys are generated from each PDF contained in the envelope, and these toys are subsequently fit with the full envelope. The collection of toy fits allows us to compute the bias and coverage of the signal strength. Should the bias and coverage be at an acceptable level, the envelope can be used in the analysis. However, in the presence of large bias or low coverage, we can consider modifying the composition of the envelope in order to reduce the bias. In practice, this is often achievable by adding higher order functions that may have been excluded by the

F-test, but nevertheless fit the data well. These more flexible functions are often able to reduce the bias and improve the coverage of the envelope. This is considered category by category. The fits of the chosen envelope functions in each category are shown in Figure 8.6.

8.3.5. Comparison with Previous Approach

Previously, in the 2016 analysis and in earlier versions of the full Run 2 analysis, the background was modeled on the three body mass range 115 to 170 GeV. After extending the range down to 105 GeV and studying the fits in each category, it is worth checking how the new modeling compares to the old. One simple way to compare is by overlaying the fit plots using the best fit function in each range. This is shown in Figure 8.7. From the plots, it is clear that modeling the turn-on properly yields a better fit around 115 GeV, the previous lower bound of the fit.

8.3.6. Envelope Bias Studies

Upon the selection of an initial envelope of PDFs to model the background in a given category of the analysis, a bias study is performed for that category. The principal goal of the bias study is to measure the bias and coverage of the full model, the envelope of functions, for different choices of the true underlying PDF. Of course, we do not know the true underlying PDF a priori, so we rely on the chi-squared and F-test probabilities used to construct the envelope to identify likely truth functions. In practice, this means that each individual PDF in the envelope is considered separately as a possible truth function in our bias studies. There are several possible outcomes of the bias studies in
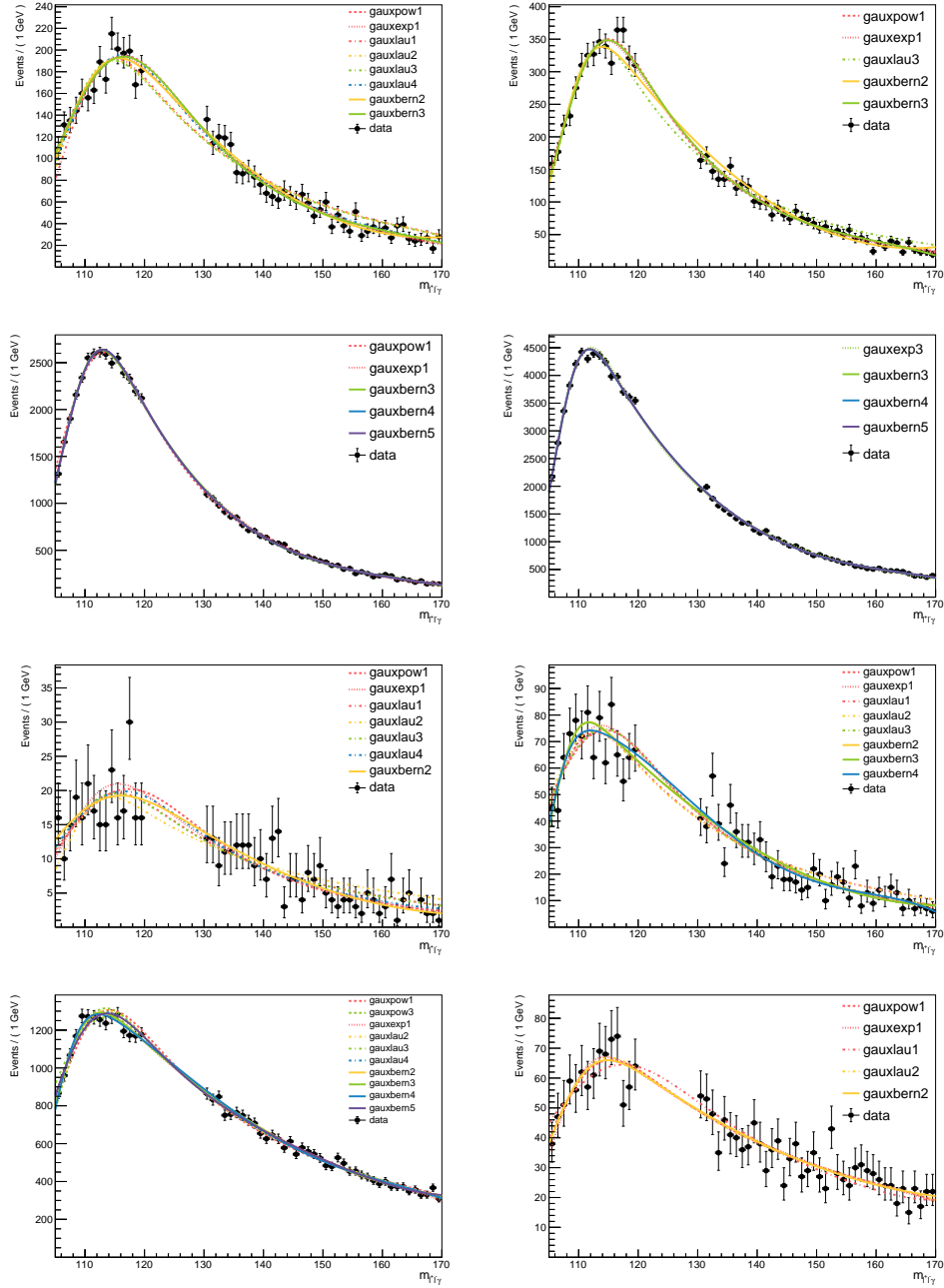
Figure 8.6. Background functions determined by the discrete profiling method. The top four plots correspond to the untagged categories, and the bottom four plots correspond to the dijet categories and lepton tag category.
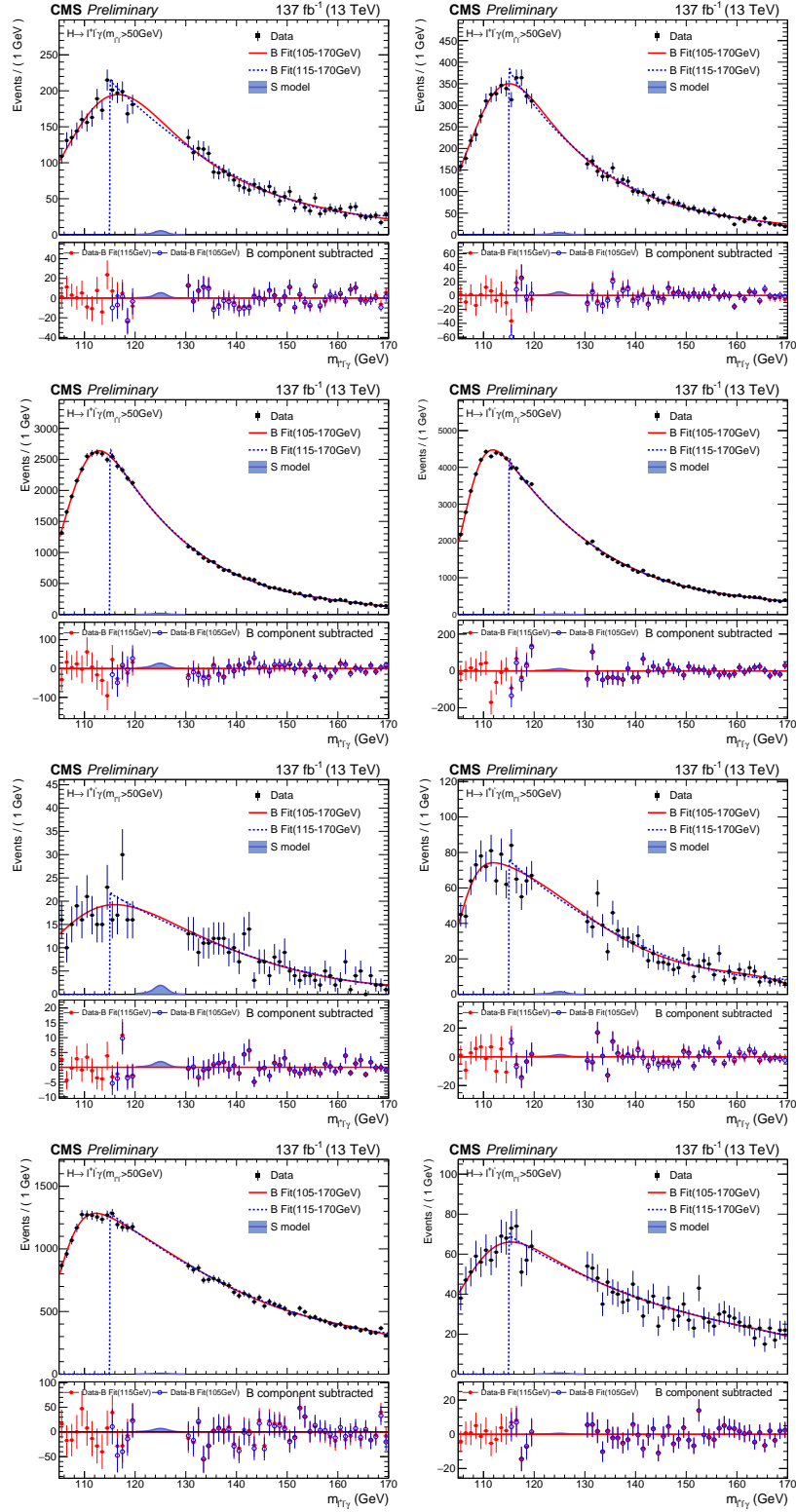
Figure 8.7. Comparison of background fits with and without modeling the turn-on. The top four plots correspond to the untagged categories, and the bottom four plots correspond to the dijet categories and lepton tag category.

each category. One outcome is that the bias is minimal and the coverage good for all truth functions. In this case, we accept the composition of the envelope as good for the analysis. Another possible outcome is that one or more truth functions is associated with a large (significantly greater than 14%) bias or low (significantly less than 68%) coverage. In this case, we can consider adding more functions to the envelope to increase the flexibility of the model, and then retry the bias study using the extended envelope.

A more precise description of the bias study procedure follows. First, we generate a set of 1,000 toy datasets for a given truth function. The parameters of the truth function are fixed by fitting the data. Then, this PDF is used to generate the toys, with the same level of statistical uncertainty as the data and assuming a true signal strength of zero (background only). Next, each toy is fit using the full envelope, where the index of the best fit function is profiled with the signal strength. In general, this means that different toys may be fit by different constituent PDFs within the envelope, and these PDFs may be different from the generating PDF. For each toy fit, we obtain a best fit signal strength $\hat{r}$ and a 68% confidence interval $(\hat{r}_{down}, \hat{r}_{up})$. We define the pull for each toy as $pull = (\hat{r} - r_{true})/\sigma_{\hat{r}}$. In the background only case, $r_{true} = 0$. The uncertainty in the denominator accounts for the assymetric error bars coming from the confidence interval. The pull distribution is expected to be a unit Gaussian. We fit the pull distribution with a Gaussian, and the best fit mean value is taken to be the bias. The coverage is defined as the frequency with which zero falls within the confidence interval $(\hat{r}_{down}, \hat{r}_{up})$. It is expected to be near 68%. If overcoverage is observed, this implies the uncertainty in the signal strength may be overly conservative. Within a few percent of 68%, this is acceptable. If undercoverage of a significant amount is observed, this implies a problem with the model,

and may mean that the individual functions are biased or that the envelope should be made more flexible.

The bias and coverage results for the final envelope in each category are shown in Tables 8.1a through 8.1h. Note that in all categories except untagged 3, untagged 4, and dijet 3, the envelope functions are exactly those identified by the chi-squared and F-test probabilities previously described. However, in untagged 3, untagged 4, and dijet 3, it was observed that the initial functions chosen by the chi-squared and F-test procedure showed large biases (greater than 20%) and low coverages (less than 60%) for certain truth functions. Therefore, the envelope was extended to include higher order Bernstein functions up to order 5 for each of these categories. The bias tables for these categories reflect the results for the final extended envelope.

8.3.7. Interpretation of Bias Study Results

We now turn to the interpretation of the initial envelope bias studies. First, it can be seen that most categories show no signs of significant bias or undercoverage for any potential truth functions. Specifically, we see that the envelopes for categories untagged 1, untagged 4, and dijet 1 can immediately be accepted as good for the analysis. In untagged 2, untagged 3, dijet 2, dijet 3, and the lepton tag category, specific truth functions deserve a bit more scrutiny.

In the untagged 2, dijet 2, dijet 3, and lepton tag categories, we observe slightly larger bias for the Laurent polynomial truth functions. From the heat map of the chosen functions, we can see the toys generated with Laurent functions are most often fit with Bernstein polynomials. Additionally, from the initial chi-squared probability results in

Table 8.1. Bias and coverage results for each category.

Generating PDF	Exp1	Pow1	Lau1	Lau2	Lau3	Lau4	Bern2	Bern3
Bias	-0.03	0.02	0.02	0.06	0.04	-0.09	-0.07	0.00
	±0.03	±0.04	±0.04	±0.04	±0.04	±0.04	±0.03	±0.04
Coverage	0.68	0.66	0.71	0.68	0.71	0.68	0.65	0.70

(a) Untagged 1 bias and coverage results from the final envelope bias study.

Generating PDF	Exp1	Pow1	Lau3	Bern2	Bern3
Bias	-0.10	-0.09	-0.27	-0.07	0.07
	±0.04	±0.05	±0.04	±0.03	±0.04
Coverage	0.67	0.66	0.67	0.67	0.65

(b) Untagged 2 bias and coverage results from the final envelope bias study.

Generating PDF	Exp1	Pow1	Bern3	Bern4	Bern5
Bias	-0.05	-0.11	0.03	-0.02	0.27
	±0.04	±0.03	±0.04	±0.03	±0.04
Coverage	0.67	0.67	0.70	0.67	0.66

(c) Untagged 3 bias and coverage results from the final envelope bias study.

Generating PDF	Exp3	Bern3	Bern4	Bern5
Bias	-0.08	0.07	0.16	0.10
	±0.03	±0.03	±0.03	±0.03
Coverage	0.68	0.67	0.67	0.68

(d) Untagged 4 bias and coverage results from the final envelope bias study.

Generating PDF	Exp1	Pow1	Lau1	Lau2	Lau3	Lau4	Bern2
Bias	0.03	0.09	-0.02	-0.06	-0.02	0.02	-0.04
	±0.03	±0.03	±0.04	±0.04	±0.03	±0.03	±0.03
Coverage	0.64	0.66	0.66	0.68	0.67	0.68	0.66

(e) Dijet 1 bias and coverage results from the final envelope bias study.

Generating PDF	Exp1	Pow1	Pow3	Lau2	Lau3	Lau4	Bern2	Bern3	Bern4	Bern5
Bias	-0.12	-0.29	-0.14	-0.22	-0.16	-0.18	-0.06	-0.05	0.05	-0.13
	±0.04	±0.04	±0.04	±0.04	±0.04	±0.04	±0.04	±0.03	±0.03	±0.03
Coverage	0.67	0.65	0.68	0.66	0.66	0.67	0.68	0.64	0.67	0.65

(f) Dijet 2 bias and coverage results from the final envelope bias study.

Generating PDF	Exp1	Pow1	Pow3	Lau2	Lau3	Lau4	Bern2	Bern3	Bern4	Bern5
Bias	-0.12	-0.29	-0.14	-0.22	-0.16	-0.18	-0.06	-0.03	0.05	-0.13
	±0.03	±0.04	±0.04	±0.04	±0.03	±0.03	±0.03	±0.05	±0.03	±0.03
Coverage	0.67	0.65	0.68	0.67	0.66	0.67	0.68	0.64	0.67	0.65

(g) Dijet 3 bias and coverage results from the final envelope bias study.

Generating PDF	Exp1	Pow1	Lau1	Lau2	Bern2
Bias	-0.04	0.05	0.27	0.03	-0.03
	±0.03	±0.03	±0.04	±0.03	±0.03
Coverage	0.71	0.69	0.66	0.70	0.70

(h) Lepton tag bias and coverage results from the final envelope bias study.

Table **??**, the Laurent functions give lower probabilities than the other functions in the envelope. This means that they are quite unlikely to be the actual truth function. From the likelihood scan with $\mu = 1$ in Figure **??**, we also observe that the Laurent functions are not involved in the likelihood scan between 0 and 1. This means that the Laurent family is irrevelant in the final background model. Given this consideration, along with the fairly good coverage for these Laurent functions, we expect the impact of these biases on the analysis to be negligible. A similar line of logic also applies to the case of the power law 1 truth function in categories dijet 2 and dijet 3. Again, we expect the impact of these biases to be negligible.

In untagged 3, we observe a small, but somewhat significant bias of 27% for the truth function with falling spectrum component Bernstein 5. In this case, we see that the coverage of 66% is still quite good, only slightly below 68%. Looking at the fit frequency heat map for this category, we see that this bias arises when the fit function is a lower order Bernstein polynomial. This type of behavior is expected. Generating with a higher order and fitting with a lower order within the same function family will usually yield some level of bias. This is less of a concern than a case in which one truth family shows a large bias when fit with a different family. Given the small size of the bias for a single truth function, and the fact that the behavior is somewhat expected for lower order fits within the Bernstein family, we expect the overall impact on the analysis to be negligible. Therefore, the untagged 3 envelope is also deemed fine for the analysis.

CHAPTER 9

Systematic Uncertainties

The uncertainties associated with the choice of background shape are incorporated into the fit to the data through the use of the discrete profiling method. They are, therefore, reflected in the statistical uncertainties obtained from the fit. The systematic uncertainties, affecting either the normalization or the shape of the signal expectation, are listed below, and the numerical values are summarized in Table 9.1, which also indicates whether the effect is correlated between the data-taking periods.

- Theoretical cross section calculations: These include the effects of the choice of PDFs, the value of the strong coupling constant (α_S), and the effect of missing higher orders in the perturbative cross section calculations, evaluated from variations of the renormalization and factorization scales (μ_R, μ_F) [?, ?, ?]. The uncertainties are treated as independent for each Higgs boson production mechanism. The uncertainty in $\mathcal{B}(H \to Z\gamma)$ is also considered [?].

- Underlying event and parton shower modeling: The uncertainty associated with the choice and tuning of the generator is estimated with dedicated samples which are generated by varying the parameters of the tune used to generate the original signal samples. The uncertainties are treated as correlated for the 2017 and 2018 samples, which use the CP5 tune [?], while being uncorrelated with the 2016 sample, which uses the CUETP8M1 tune [?].

- Integrated luminosity: The integrated luminosities for the 2016, 2017, and 2018 data-taking years have uncertainties of 1.2%, 2.3% and 2.5% [**?, ?, ?**], respectively, with an overall uncertainty for the 2016–2018 period of 1.6%, the improvement in precision reflecting the (uncorrelated) time evolution of some systematic effects.

- L1 trigger: During the 2016 and 2017 data-taking periods, a gradual shift in the timing of the inputs of the ECAL L1 trigger in the $|\eta| > 2.4$ region led to a specific inefficiency. A correction of approximately 1% is applied to the simulation along with the corresponding uncertainty in the inefficiency measurement.

- Trigger: There are uncertainties in the corrections applied to the simulation to match the trigger efficiencies measured in data with $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ events.

- Photon identification and isolation: There are uncertainties in the corrections applied to the simulation to match the selection efficiencies in data measured with $Z \to e^+e^-$ events.

- Lepton identification and isolation: There are uncertainties in the corrections applied to the simulation to match electron and muon selection efficiencies in data measured with $Z \to e^+e^-$ and $Z \to \mu^+\mu^-$ events.

- Pileup modeling: The uncertainty in the description of the pileup in the signal simulation is estimated by varying the total inelastic cross section by $\pm4.6\%$ [**?**].

- Kinematic BDT: The uncertainties in the photon and lepton energy and the correction of the photon MVA discriminant are propagated to $\mathcal{D}_{\mathrm{kin}}$. Changes in $\mathcal{D}_{\mathrm{kin}}$ cause the migration of signal events across category boundaries.

- VBF BDT: The uncertainties in the jet energy and the uncertainty in $\mathcal{D}_{\text{kin}}$ are propagated to $\mathcal{D}_{\text{VBF}}$. Changes in $\mathcal{D}_{\text{VBF}}$ cause the migration of signal events across category boundaries.

- Photon energy scale and resolution: The photon energy in the simulation is varied due to the ECAL energy scale and resolution uncertainties, and the effects on the signal mean and resolution parameters are propagated to the fits.

- Lepton momentum scale and resolution: The lepton momentum in the simulation is varied due to the lepton momentum scale and resolution uncertainties, and the effects on signal mean and resolution parameters are propagated to the fits.

In the $\mathcal{B}(\text{H} \to \text{Z}\gamma)/\mathcal{B}(\text{H} \to \gamma\gamma)$ measurement, the common sources of theoretical and systematic uncertainty in the two analyses are treated as correlated in the fit. These are the theoretical uncertainties in the Higgs production cross section calculations, and the systematic uncertainties in the underlying event and parton shower modeling, the integrated luminosity, and the L1 trigger inefficiency. The remaining uncertainties are treated as uncorrelated.

Table 9.1. Sources of systematic uncertainty affecting the simulated signal. The normalization effect on the expected yield, or the effect on the signal shape parameters, is given as indicated, with the values averaged over all event categories. The third column shows the uncertainties that have a correlated effect across the three data-taking periods.

Sources	Uncertainty (%)	Year-to-year correlation
Normalization		
Theoretical	—	✓
– $\mathcal{B}(\mathrm{H} \to \mathrm{Z}\gamma)$	5.7	
– ggH cross section (μ_F, μ_R)	3.9	
– ggH cross section (α_S)	2.6	
– ggH cross section (PDF)	1.9	
– VBF cross section (μ_F, μ_R)	0.4	
– VBF cross section (α_S)	0.5	
– VBF cross section (PDF)	2.1	
– WH cross section (μ_F, μ_R)	$+0.6 \atop -0.7$	
– WH cross section (PDF)	1.7	
– ZH cross section (μ_F, μ_R)	$+3.8 \atop -3.1$	
– ZH cross section (PDF)	1.3	
– WH/ZH cross section (α_s)	0.9	
– $\mathrm{t\bar{t}H}$ cross section (μ_F, μ_R)	$+5.8 \atop -9.2$	
– $\mathrm{t\bar{t}H}$ cross section (α_s)	2.0	
– $\mathrm{t\bar{t}H}$ cross section (PDF)	3.0	
Underlying event and parton shower	3.7–4.4	Partial
Integrated luminosity	1.2–2.5	Partial
L1 trigger	0.1–0.4	—
Trigger	—	—
– Electron channel	0.9–1.9	
– Muon channel	0.1–0.4	
Photon identification and isolation	0.2–5.0	✓
Lepton identification and isolation	—	✓
– Electron channel	0.5–0.7	
– Muon channel	0.3–0.4	
Pileup	0.4–1.0	✓
Kinematic BDT	2.5–3.7	✓
VBF BDT	5.9–14	✓
Shape		
Photon energy and momentum		✓
– Signal mean	0.1–0.4	
– Signal resolution	3.1–5.9	
Lepton energy and momentum		✓
– Signal mean	0.007	
– Signal resolution	0.007–0.01	

CHAPTER 10

Statistical Analysis

CHAPTER 11

Results and Interpretation

Figures 11.1 and 11.2 show the $m_{\ell^+\ell^-\gamma}$ distributions of the data events in each category. The expected SM H $\rightarrow$ Zγ distributions, scaled by a factor of 10, are also shown. Figure 11.3 shows the signal-plus-background fit to the data and the corresponding distribution after background subtraction for the sum of all categories. Each category is weighted by the factor $S/(S+B)$, where S is the signal yield and B is the background yield in the narrowest mass interval containing 95% of the signal distribution.

The best fit value of the signal strength is $2.4^{+0.8}_{-0.9}$ (stat) $^{+0.3}_{-0.2}$ (syst) at $m_{\mathrm{H}} = 125.38\,\mathrm{GeV}$. The corresponding measured value of $\sigma(\mathrm{pp} \rightarrow \mathrm{H})\mathcal{B}(\mathrm{H} \rightarrow \mathrm{Z}\gamma)$ is $0.21^{+0.07}_{-0.08}$ (stat) $^{+0.03}_{-0.02}$ (syst) pb. This measurement is consistent with the SM prediction of 0.09 ± 0.01 pb at the 1.6 standard deviation level. Figure 11.4 shows the signal strengths obtained for each category separately, corresponding to the fit results shown in Figs. 11.1 and 11.2, as well as from simultaneous fits to the dijet categories, the untagged categories, and all categories combined. Among the eight categories, dijet 1 is the most sensitive. A category compatibility p-value, under the hypothesis of a common signal strength in all categories, is calculated from the likelihood ratio between the nominal combined fit, in which all categories have the same signal strength parameter, and a separate fit, in which each category has its own signal strength parameter. This p-value is found to be 0.02, corresponding to 2.3 standard deviations, and is driven by the dijet 3 category, which has a signal strength of $\hat{\mu} = 12.3^{+3.7}_{-3.5}$. The observed (expected) local significance is $2.7\,(1.2)$ standard
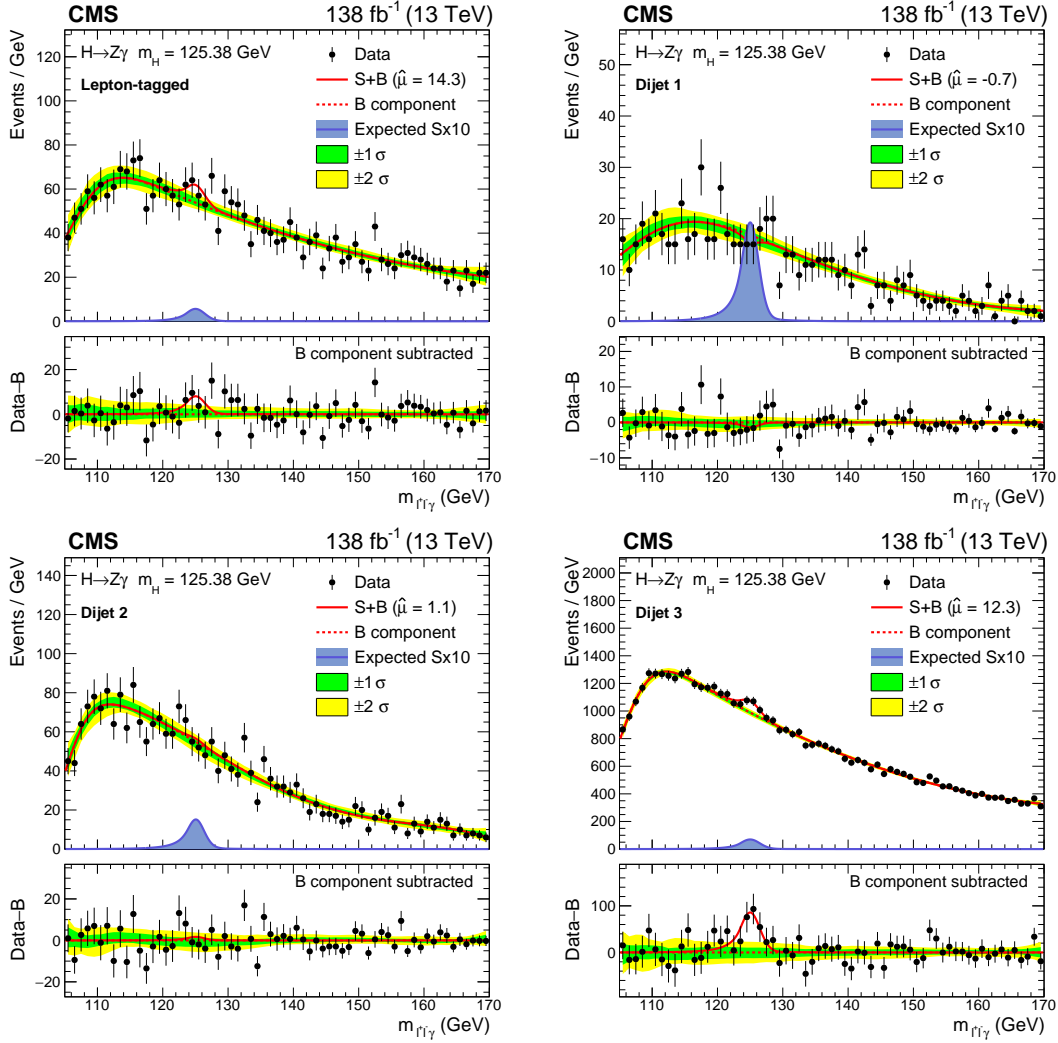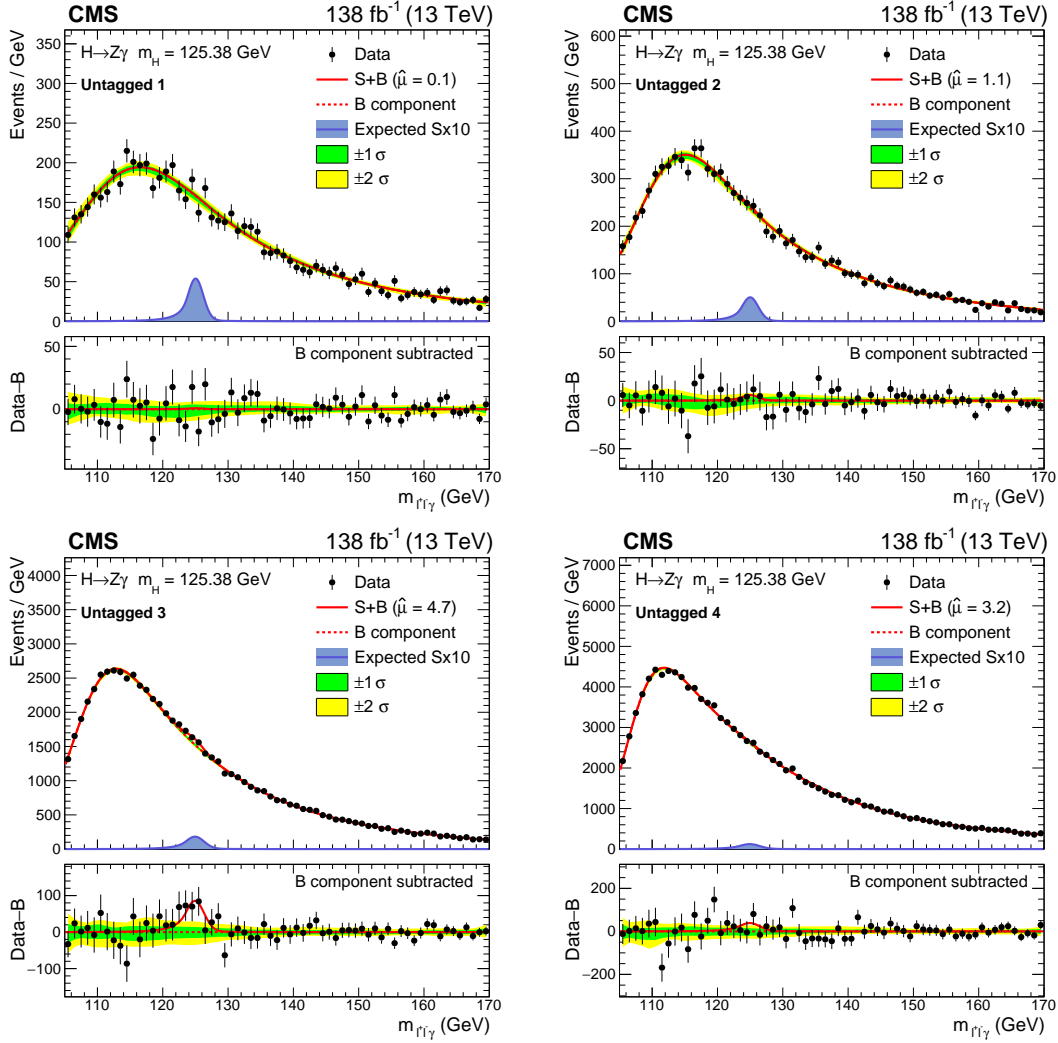
Figure 11.1. Fits to the $m_{\ell^+\ell^-\gamma}$ data distribution in the lepton-tagged (upper left), dijet 1 (upper right), dijet 2 (lower left), and dijet 3 (lower right) categories. In the upper panel, the red solid line shows the result of a signal-plus-background fit to the given category. The red dashed line shows the background component of the fit. The green and yellow bands represent the 68 and 95% CL uncertainties in the fit. Also plotted is the expected SM signal, scaled by a factor of 10. In the lower panel, the data minus the background component of the fit is shown.

Figure 11.2. Fits to the $m_{\ell^+\ell^-\gamma}$ data distribution in the untagged 1 (upper left), untagged 2 (upper right), untagged 3 (lower left), and untagged 4 (lower right) categories. In the upper panel, the red solid line shows the result of a signal-plus-background fit to the given category. The red dashed line shows the background component of the fit. The green and yellow bands represent the 68 and 95% CL uncertainties in the fit. Also plotted is the expected SM signal, scaled by a factor of 10. In the lower panel, the data minus the background component of the fit is shown.
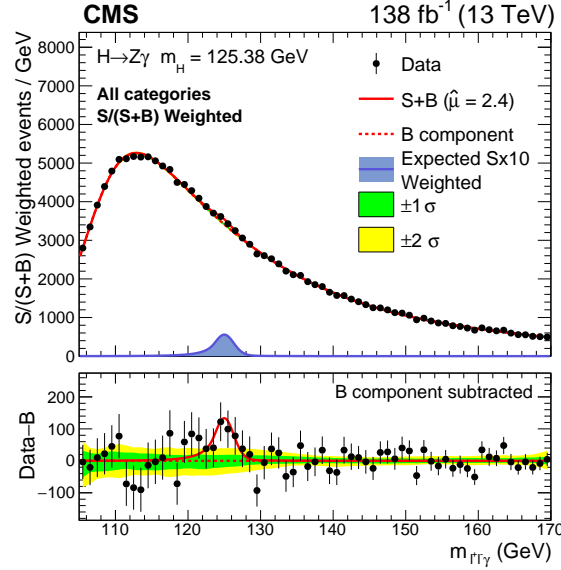
Figure 11.3. Sum over all categories of the data points and signal-plus-background model after the simultaneous fit to each $m_{\ell^+\ell^-\gamma}$ distribution. The contribution from each category is weighted by $S/(S + B)$, as defined in the text. In the upper panel, the red solid line shows the signal-plus-background fit. The red dashed line shows the background component of the fit. The green and yellow bands represent the 68 and 95% CL uncertainties in the fit. Also plotted is the expected SM signal weighted by $S/(S+B)$ and scaled by a factor of 10. In the lower panel, the data minus the background component of the fit is shown.

deviations. Upper limits on μ are calculated at $1\,\mathrm{GeV}$ intervals in the mass range of $120 < m_{\ell^+\ell^-\gamma} < 130\,\mathrm{GeV}$ and at $m_\mathrm{H} = 125.38\,\mathrm{GeV}$, as shown in Fig. 11.5. The observed (expected) limit at 95% CL relative to the SM expectation for $m_\mathrm{H} = 125.38\,\mathrm{GeV}$ is 4.1 (1.8). The measured value of $\mathcal{B}(\mathrm{H} \to \mathrm{Z}\gamma)/\mathcal{B}(\mathrm{H} \to \gamma\gamma)$ from the combined fit with the $\mathrm{H} \to \gamma\gamma$ analysis is $1.5^{+0.7}_{-0.6}$. This measurement is consistent with the SM prediction for the ratio at the 1.5 standard deviation level.
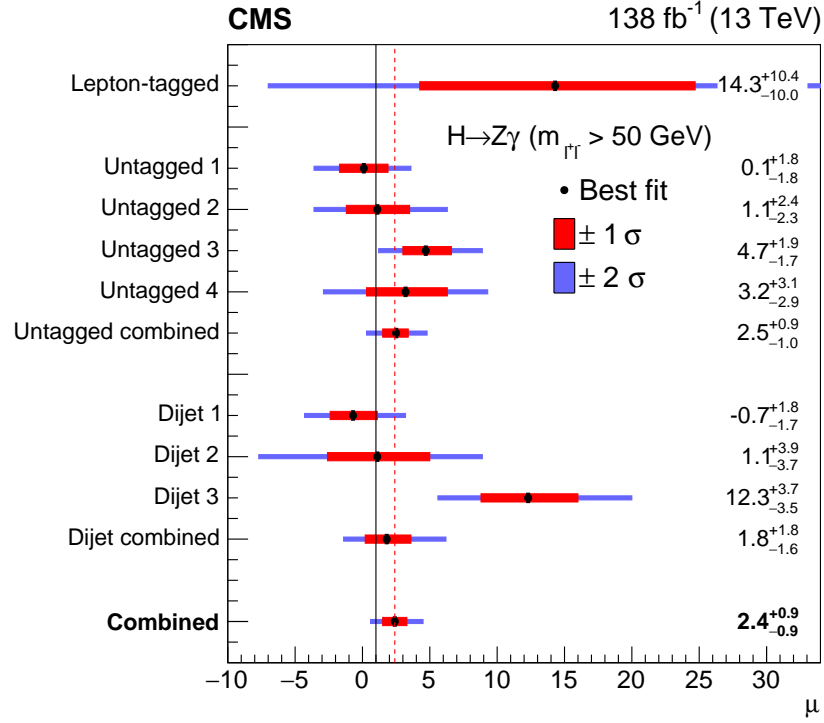
Figure 11.4. Observed signal strength (μ) for a SM Higgs boson with $m_{\mathrm{H}} = 125.38\,\mathrm{GeV}$. The labels "untagged combined," "dijet combined," and "combined" represent the results obtained from simultaneous fits of the untagged categories, dijet categories, and full set of categories, respectively. The black solid line shows $\mu = 1$, and the red dashed line shows the best fit value $\hat{\mu} = 2.4 \pm 0.9$ of all categories combined.
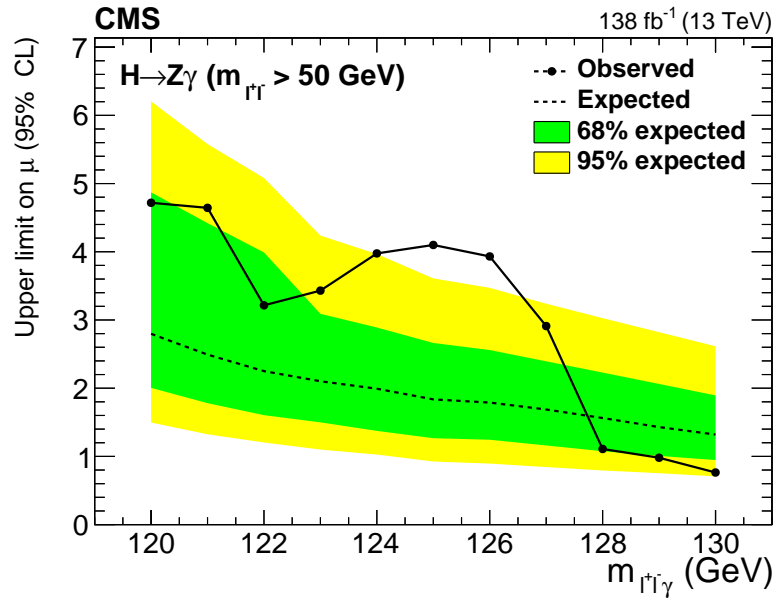
Figure 11.5. Upper limit (95% CL) on the signal strength (μ) relative to the SM prediction, as a function of the assumed value of the Higgs boson mass used in the fit.

CHAPTER 12

Conclusion

References

[1] A bibliographic item. A bibliographic item. A bibliographic item. A bibliographic item.

[2] Another bibliographic item.

[3] Yet another bibliographic item.

APPENDIX A

Signal and Resonant Background Fits

A.1. 2017 Signal Fits

A.2. 2018 Signal Fits

A.3. 2017 Resonant Background Fits

A.4. 2018 Resonant Background Fits
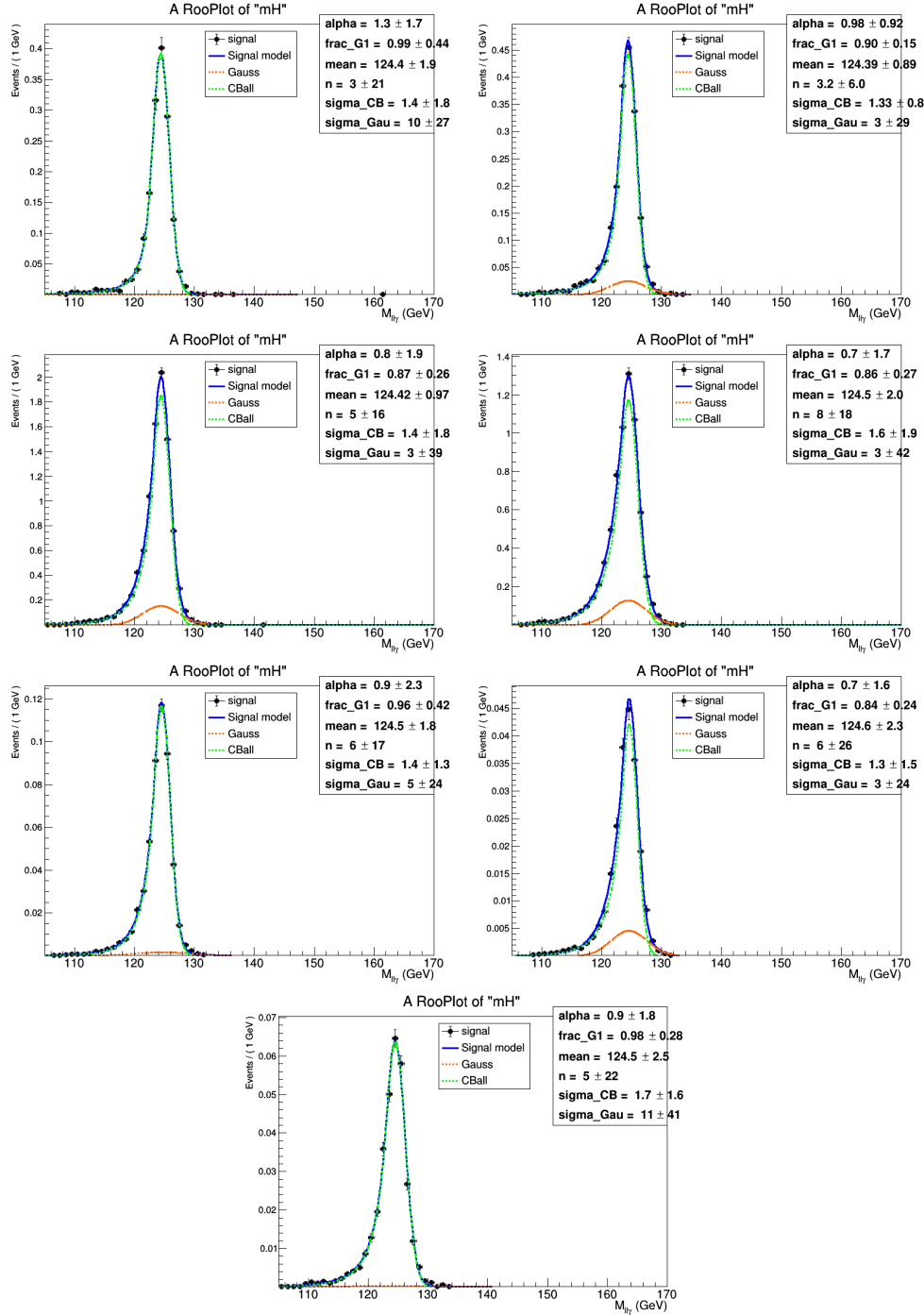
Figure A.1. Fits to simulated $m_{\ell^+\ell^-\gamma}$ signal distributions in the electron channel for $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ for the 2017 data-taking period. The blue line shows the total fit function, the green line shows the Crystal Ball function component, and the red line shows the Gaussian function component. The top four plots correspond to the untagged categories, and the bottom three plots correspond to the dijet categories.
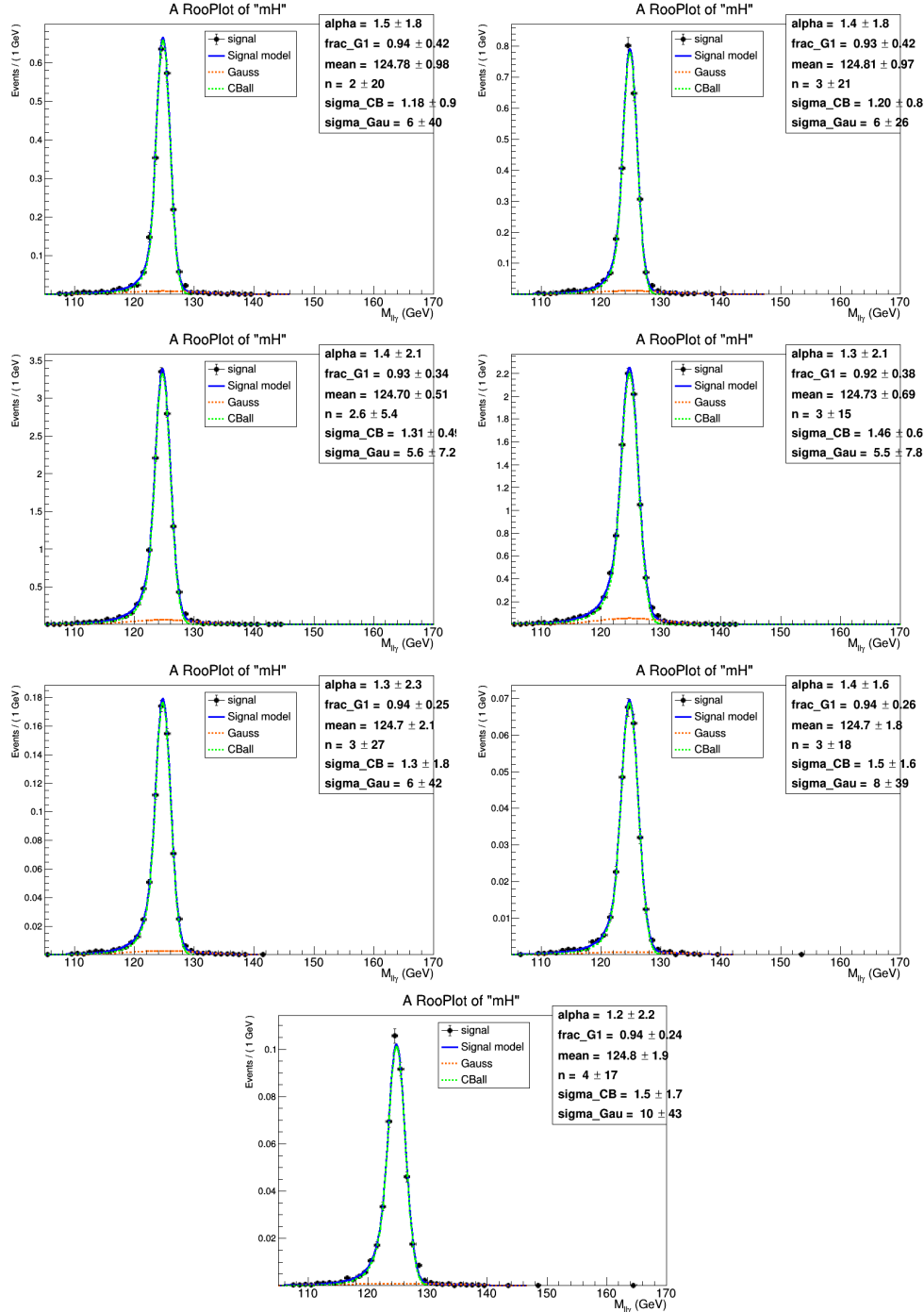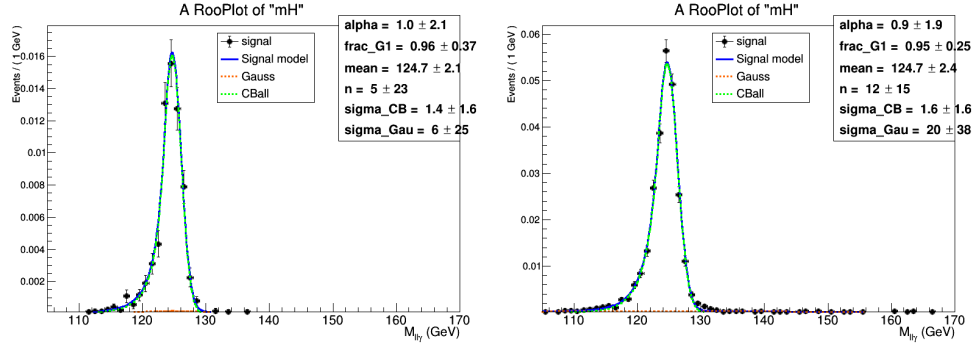
Figure A.2. Fits to simulated $m_{\ell^+\ell^-\gamma}$ signal distributions in the muon channel for $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ for the 2017 data-taking period. The blue line shows the total fit function, the green line shows the Crystal Ball function component, and the red line s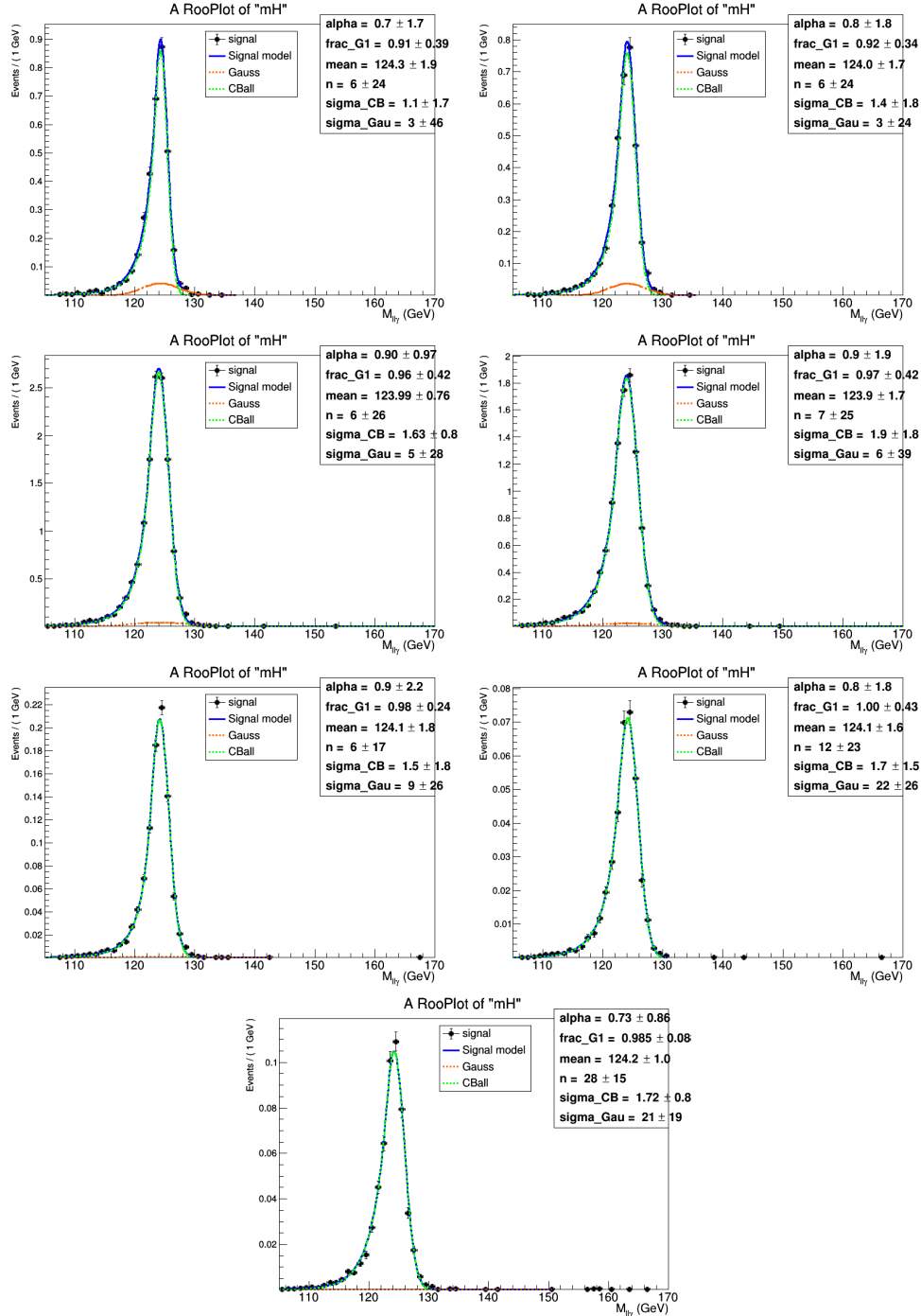hows the Gaussian function component. The top four plots correspond to the untagged categories, and the bottom three plots correspond to the dijet categories.
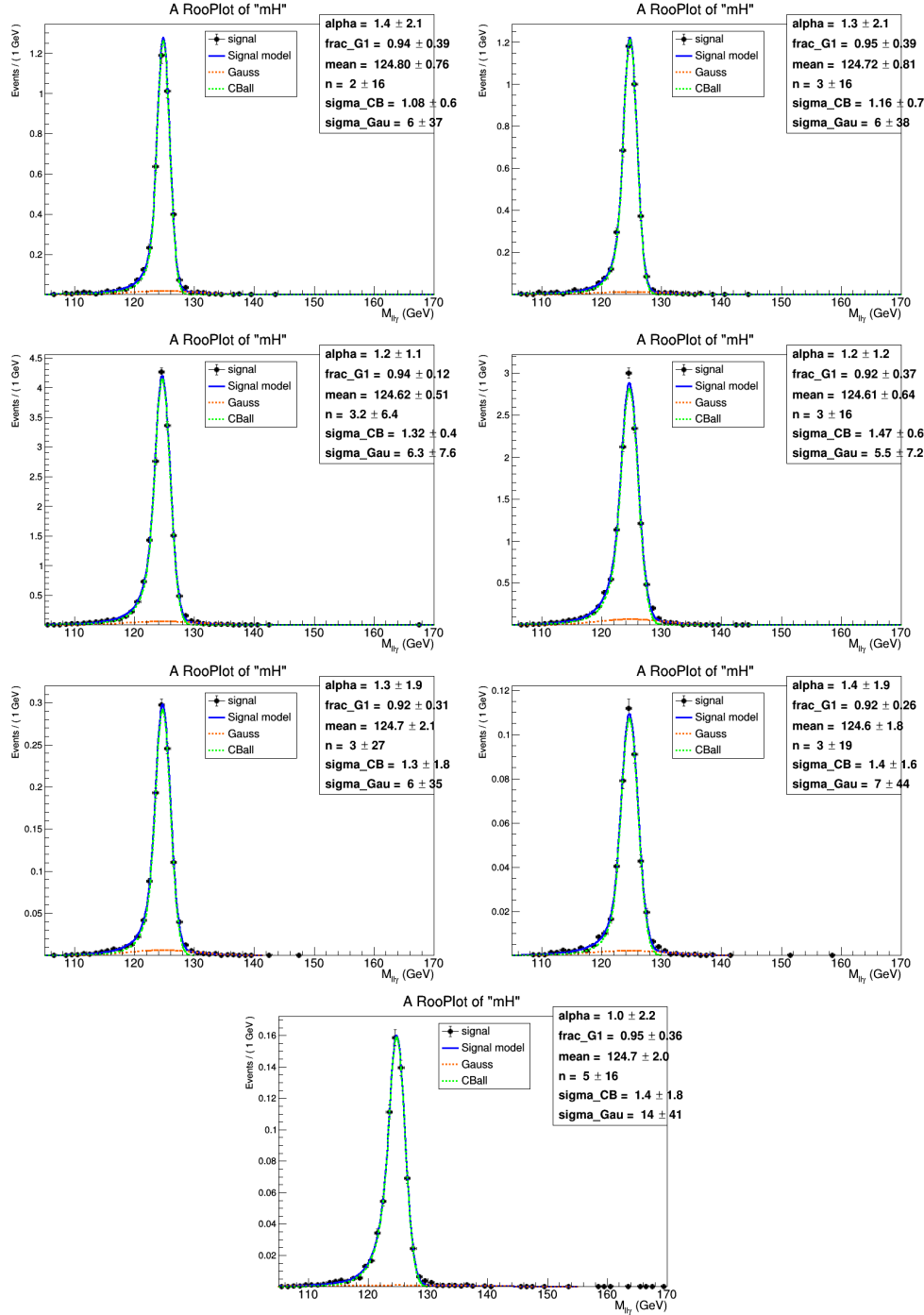
Figure A.3. Fits to simulated $m_{\ell^+\ell^-\gamma}$ signal distributions in the electron and muon channels combined in the lepton-tagged category for $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ for the 2017 data-taking period. The left plot shows the fit to simulated ZH production events, and the right plot shows the fit to simulated WH production events. The blue line shows the total fit function, the green line shows the Crystal Ball function component, and the red line shows the Gaussian function component.

Figure A.4. Fits to simulated $m_{\ell^+\ell^-\gamma}$ signal distributions in the electron channel for $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ for the 2018 data-taking period. The blue line shows the total fit function, the green line shows the Crystal Ball function component, and the red line shows the Gaussian function component. The top four plots correspond to the untagged categories, and the bottom three plots correspond to the dijet categories.

Figure A.5. Fits to simulated $m_{\ell^+\ell^-\gamma}$ signal distributions in the muon channel for $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ for the 2018 data-taking period. The blue line shows the total fit function, the green line shows the Crystal Ball function component, and the red line shows the Gaussian function component. The top four plots correspond to the untagged categories, and the bottom three plots correspond to the dijet categories.
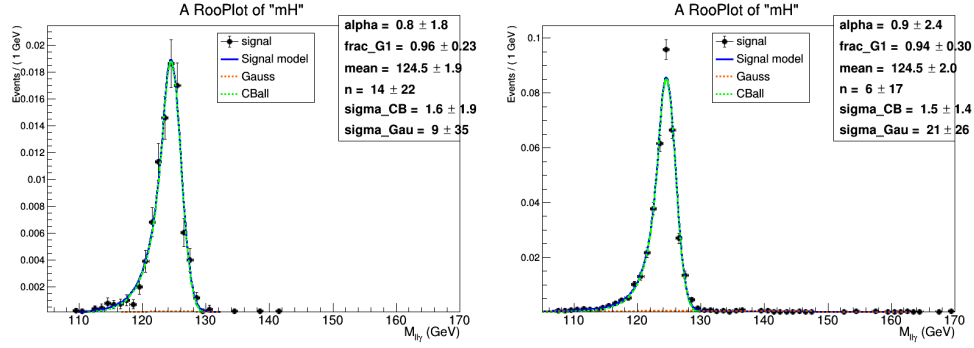
Figure A.6. Fits to simulated $m_{\ell^+\ell^-\gamma}$ signal distributions in the electron and muon channels combined in the lepton-tagged category for $m_\mathrm{H} = 125\,\mathrm{GeV}$ for the 2018 data-taking period. The left plot shows the fit to simulated ZH production events, and the right plot shows the fit to simulated WH production events. The blue line shows the total fit function, the green line shows the Crystal Ball function component, and the red line shows the Gaussian function component.
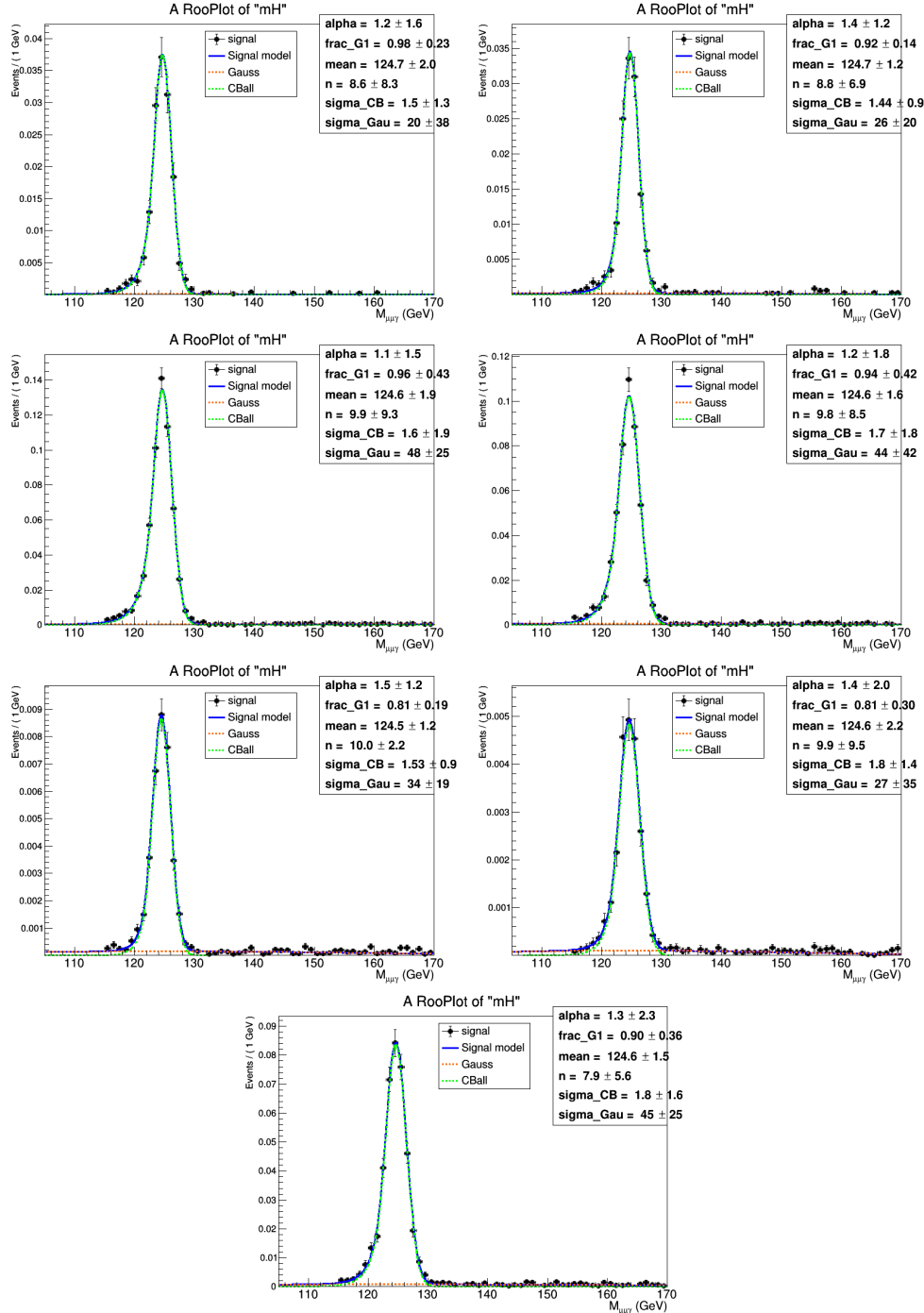
Figure A.7. Fits to simulated $m_{\mu^+\mu^-\gamma}$ resonant background distributions from H $\to \mu^+\mu^-$ for $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ for the 2017 data-taking period. The blue line shows the total fit function, the green line shows the Crystal Ball function component, and the red line shows the Gaussian function component. The top four plots correspond to the untagged categories, and the bottom three plots correspond to the dijet categories.
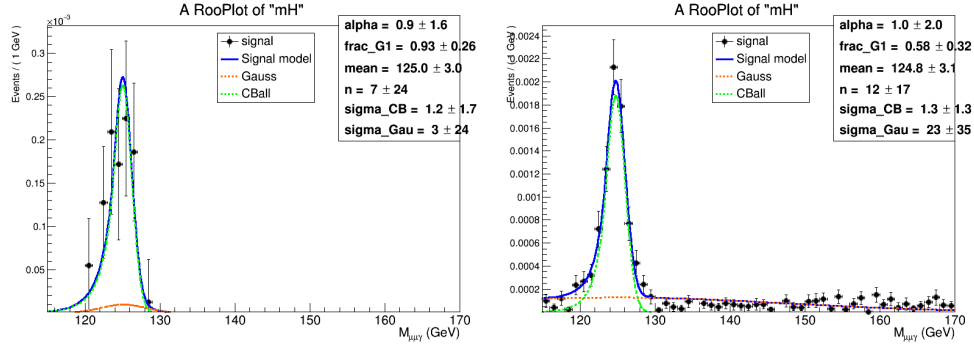
Figure A.8. Fits to simulated $m_{\ell^+\ell^-\gamma}$ resonant background distributions from $H \to \mu^+\mu^-$ in the electron and muon channels combined in the lepton-tagged category for $m_H = 125\,\text{GeV}$ for the 2017 data-taking period. The left plot shows the fit to simulated ZH production events, and the right plot shows the fit to simulated WH production events. The blue line shows the total fit function, the green line shows the Crystal Ball function component, and the red line shows the Gaussian function component.
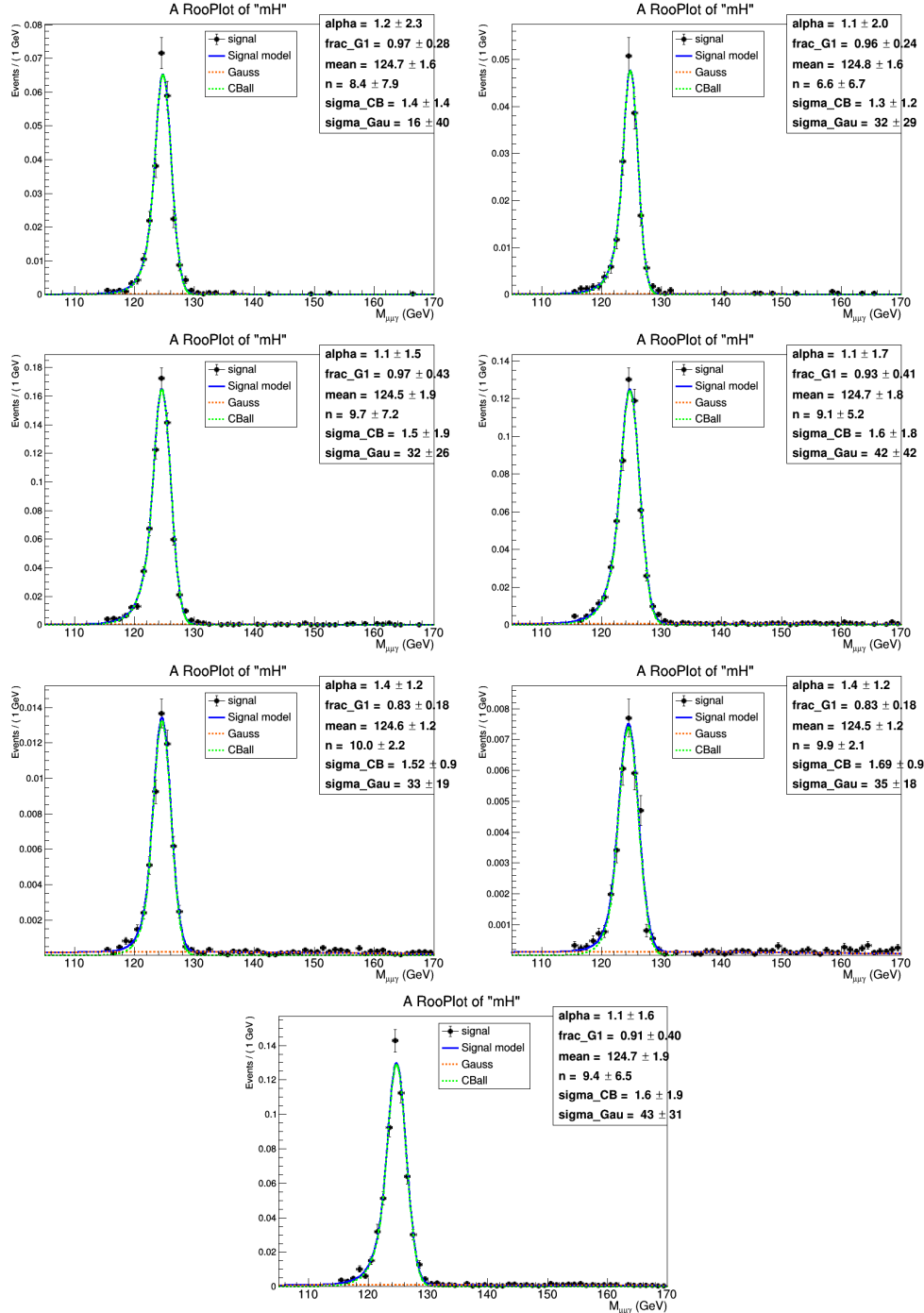
Figure A.9. Fits to simulated $m_{\mu^+\mu^-\gamma}$ resonant background distributions from H $\to \mu^+\mu^-$ for $m_\mathrm{H} = 125\,\mathrm{GeV}$ for the 2018 data-taking period. The blue line shows the total fit function, the green line shows the Crystal Ball function component, and the red line shows the Gaussian function component. The top four plots correspond to the untagged categories, and the bottom three plots correspond to the dijet categories.
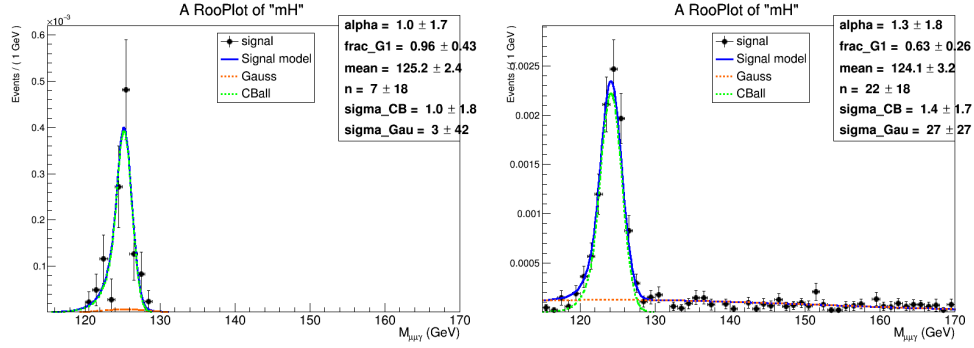
Figure A.10. Fits to simulated $m_{\ell^+\ell^-\gamma}$ resonant background distributions from $H \to \mu^+\mu^-$ in the electron and muon channels combined in the lepton-tagged category for $m_H = 125\,\text{GeV}$ for the 2018 data-taking period. The left plot shows the fit to simulated ZH production events, and the right plot shows the fit to simulated WH production events. The blue line shows the total fit function, the green line shows the Crystal Ball function component, and the red line shows the Gaussian function component.