NORTHWESTERN UNIVERSITY

The Search for the Higgs Boson Decaying into 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

June 2022

ABSTRACT

The Search for the Higgs Boson Decaying into 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 properies 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 $H \rightarrow Z\gamma$, which is predicted by the SM to have a branching fraction of XX. An observation of this decay mode at a rate deviating from the SM prediction would provide indirect evidence of new physics beyond the SM. Previous searches for $H \rightarrow Z\gamma$ were carried out using proton-proton collision data from Run 1 of the LHC. Run 1 exclusion limits on the process were placed at roughly ten times the SM expectation by the Compact Muon Solenoid (CMS) experiment.

This thesis describes the search for $H \rightarrow Z\gamma$ in the $\ell\ell\gamma$ final state with the CMS detector using LHC Run 2 proton-proton collision data. The process is not observed, and exclusion limits are placed on the production cross section times branching fraction at XX times the SM expectation.

Acknowledgements

Text for acknowledgments.

Table of Contents

List of Tables

List of Figures

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 $(D^\mu\Phi)^\dagger(D_\mu\Phi)$.

$$(D^\mu\Phi)^\dagger(D_\mu\Phi) = \left|(\partial_\mu - \frac{ig}{2}\tau \cdot A_\mu - \frac{ig'}{2}B_\mu Y)\Phi\right|^2$$

$$= \frac{1}{2}\left| \begin{bmatrix} \partial_\mu - \frac{i}{2}(gA_\mu^3 + g'B_\mu) & -\frac{ig}{2}(A_\mu^1 - iA_\mu^2) \\ -\frac{ig}{2}(A_\mu^1 + iA_\mu^2) & \partial_\mu + \frac{i}{2}(gA_\mu^3 - g'B_\mu) \end{bmatrix} \begin{bmatrix} 0 \\ v + h(x) \end{bmatrix} \right|^2$$

$$= \frac{1}{2}\left| \begin{bmatrix} -\frac{ig}{2}(A_\mu^1 - iA_\mu^2)(v + h(x)) \\ \partial_\mu h(x) + \frac{i}{2}(gA_\mu^3 - g'B_\mu)(v + h(x)) \end{bmatrix} \right|^2$$

$$= \frac{1}{2}\partial_\mu h(x)\partial^\mu h(x) + \frac{1}{8}(gA_\mu^3 - g'B_\mu)(gA_3^\mu - g'B^\mu)(v + h(x))^2$$

$$(2.9) \qquad + \frac{g^2}{8}(A_\mu^1 - iA_\mu^2)(A_\mu^1 + iA_\mu^2)(v + h(x))^2$$

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

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

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

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

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

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

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

Experimental Apparatus

apparatus text

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 $H \to Z\gamma$. 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 $H \rightarrow Z\gamma$ searches, the signal shape is determined via an analytic fit to simulation. A resonant background contribution from $H \rightarrow \mu^+\mu^-$ 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 $H \rightarrow \gamma\gamma$.

CHAPTER 5

Data and Simulation

CHAPTER 6

Object and Event Selection

6.1. Triggers

The topology and basic kinematics of the $H \to Z\gamma$ 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 $H \to Z\gamma$ 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

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

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 $H \to Z\gamma$ 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 $H \to Z\gamma$ 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 $Z \to e^+e^-$ events where the probe electron mimics a photon and $Z \to \mu^+\mu^-$ 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 $H \rightarrow Z\gamma$ analysis. 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 $H \rightarrow Z\gamma$ analysis. 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 $H \rightarrow Z\gamma$ selection. 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 $H \rightarrow ZZ$ 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 $H \rightarrow Z\gamma$ sample. 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 $H \to Z\gamma$ 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 $H \to Z\gamma$ [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 $H \rightarrow Z\gamma$ signal 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 $H \rightarrow Z\gamma$ signal 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 $H \to Z\gamma$ events as signal and gluon-gluon fusion $H \to Z\gamma$, SM $Z\gamma$, Z plus jets, and $t\bar{t}$ events as background. VH and ttH $H \to Z\gamma$ events 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 $\Delta\eta(\text{j, j})$, $\Delta\phi(\text{j, j})$, $\Delta\phi(Z\gamma, \text{jj})$, $\Delta R(j, \gamma)$, $p_T{}^{j1,2}$, 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

7.4. Dropping of Boosted Category

7.5. Comparison to Previous Approach

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 $H \rightarrow \mu^+\mu^-$, where at least one muon radiates an FSR photon, is expected to contribute at the 6% level relative to the $H \rightarrow Z\gamma$ signal yield. Because of this, we treat it as a resonant background. To model the $H \rightarrow \mu^+\mu^-$ 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
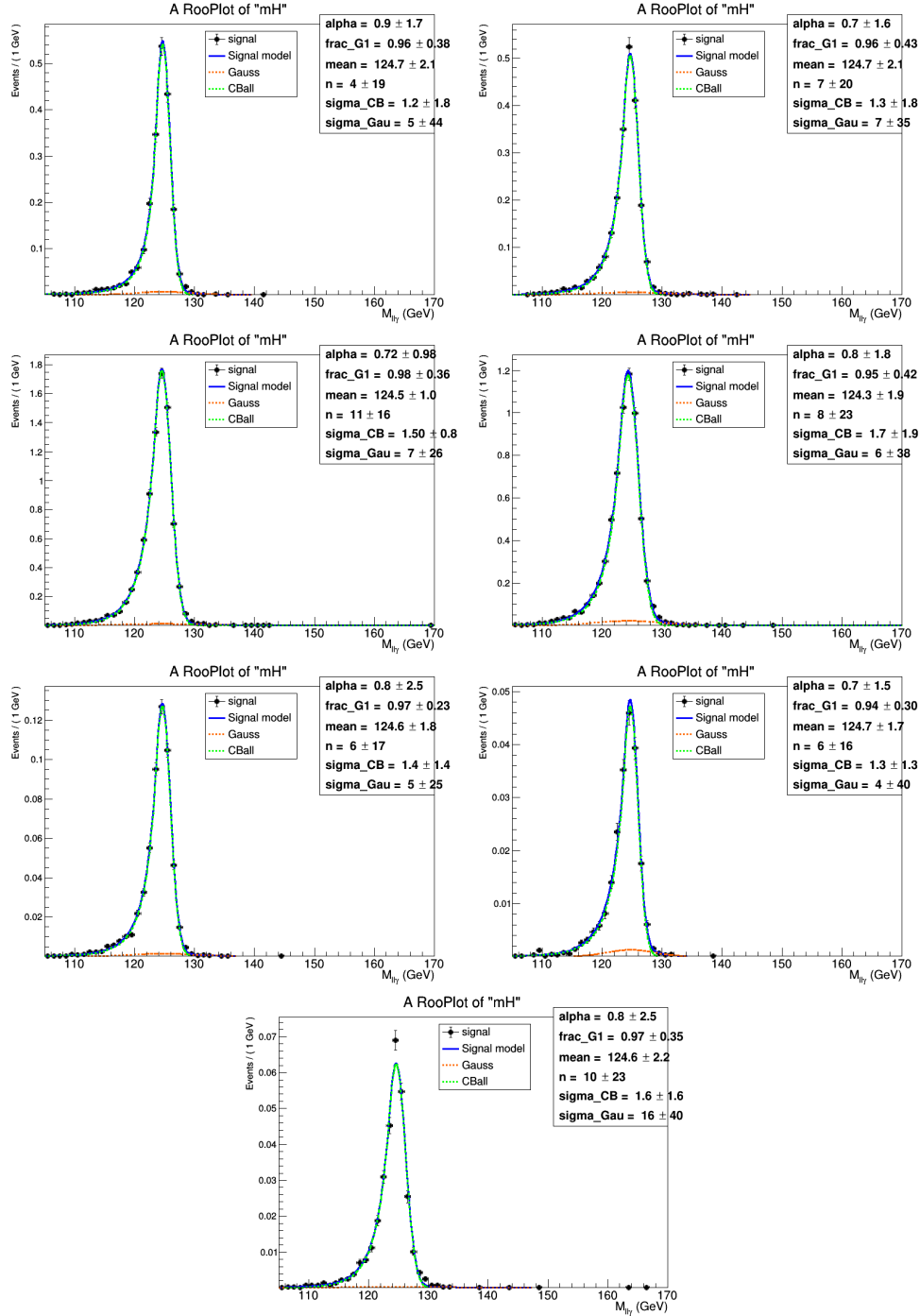
Figure 8.1. Fits to simulated $m_{\ell^+\ell^-\gamma}$ signal distributions in the electron channel for $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ for the 2016 data-taking period. The blue line shows the total fit function, the green line shows the Crystal Ball function component, and the red line shows the 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.
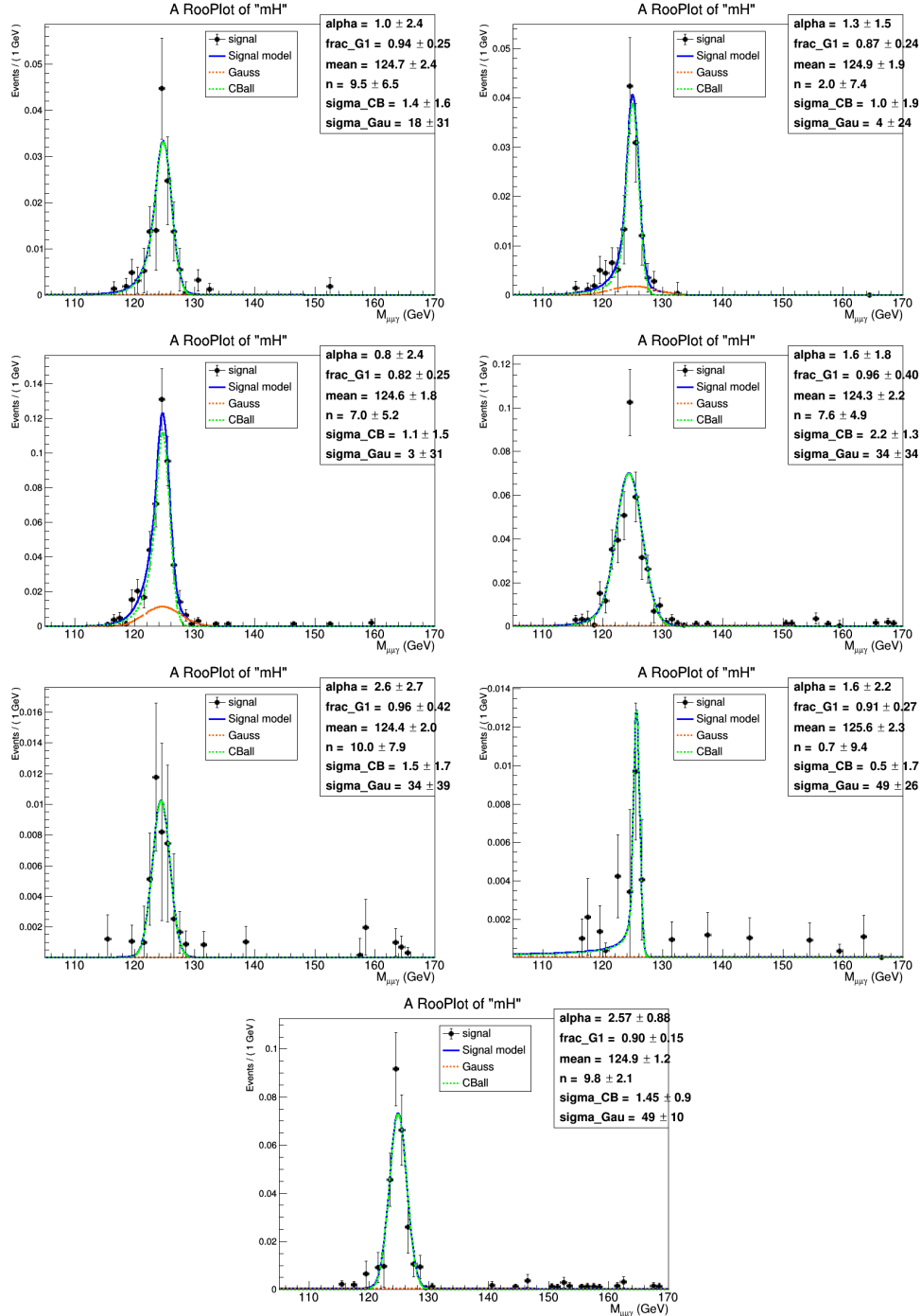
Figure 8.4. Fits to simulated $m_{\mu^+\mu^-\gamma}$ resonant background distributions from H $\to \mu^+\mu^-$ for $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ for the 2016 data-taking period. The blue line shows the total fit function, the green line shows the Crystal Ball function component, and the red line shows the Gaussian function component. The top four plots correspond to the untagged categories, and the bottom three plots correspond to the dijet categories.
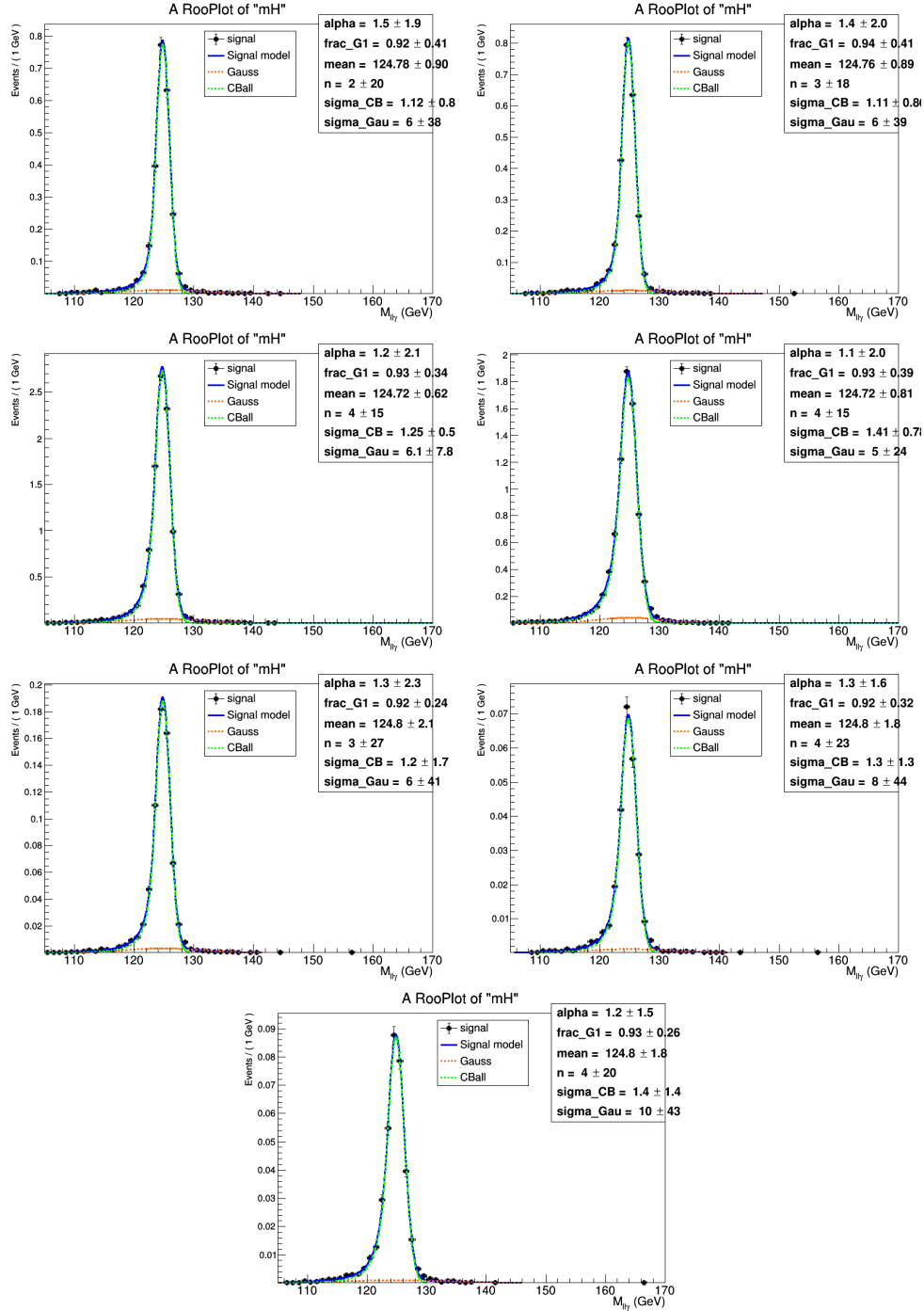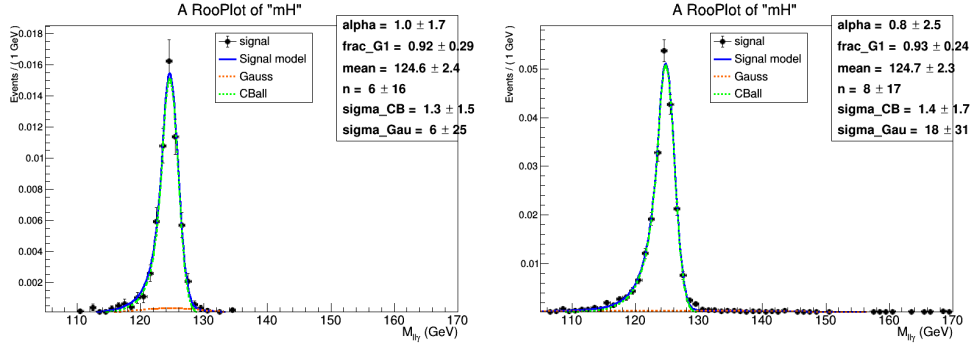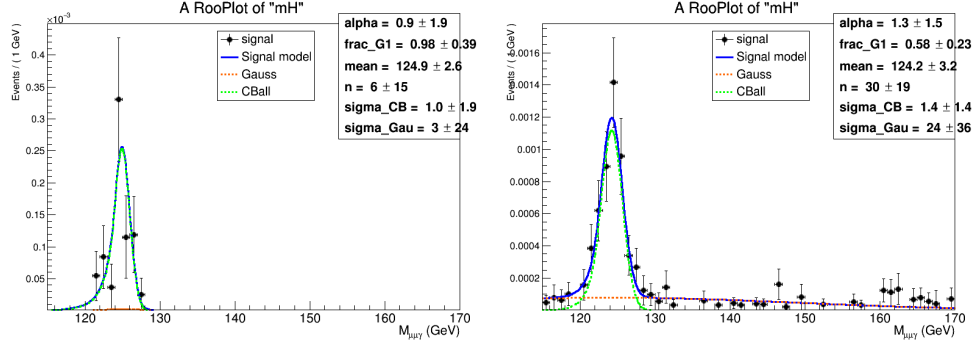
Figure 8.5. Fits to simulated $m_{\ell^+\ell^-\gamma}$ resonant background distributions from $H \to \mu^+\mu^-$ in the electron and muon channels combined in the lepton-tagged category for $m_H = 125\,\text{GeV}$ for the 2016 data-taking period. The left plot shows the fit to simulated ZH production 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:

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

  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:

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

  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:

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

$$(8.5) \qquad \text{b}_{i,N}(m_{\ell\ell\gamma}) = \binom{N}{i} m_{\ell\ell\gamma}^{i}(1 - m_{\ell\ell\gamma}^{N-i})$$

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:

$$(8.6) \qquad \text{Lau}_1(m_{\ell\ell\gamma}) = f_2 m_{\ell\ell\gamma}^{-4} + f_3 m_{\ell\ell\gamma}^{-5}$$

$$(8.7) \qquad \text{Lau}_2(m_{\ell\ell\gamma}) = f_1 m_{\ell\ell\gamma}^{-3} + f_2 m_{\ell\ell\gamma}^{-4} + f_3 m_{\ell\ell\gamma}^{-5}$$

$$(8.8) \qquad \text{Lau}_3(m_{\ell\ell\gamma}) = f_1 m_{\ell\ell\gamma}^{-3} + f_2 m_{\ell\ell\gamma}^{-4} + f_3 m_{\ell\ell\gamma}^{-5} + f_4 m_{\ell\ell\gamma}^{-6}$$

$$(8.9) \qquad \text{Lau}_4(m_{\ell\ell\gamma}) = f_1 m_{\ell\ell\gamma}^{-2} + f_2 m_{\ell\ell\gamma}^{-3} + f_3 m_{\ell\ell\gamma}^{-4} + f_4 m_{\ell\ell\gamma}^{-5} + f_5 m_{\ell\ell\gamma}^{-6}$$

$$(8.10) \qquad \text{Lau}_5(m_{\ell\ell\gamma}) = f_1 m_{\ell\ell\gamma}^{-2} + f_2 m_{\ell\ell\gamma}^{-3} + f_3 m_{\ell\ell\gamma}^{-4} + f_4 m_{\ell\ell\gamma}^{-5} + f_5 m_{\ell\ell\gamma}^{-6} + f_6 m_{\ell\ell\gamma}^{-7}$$

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
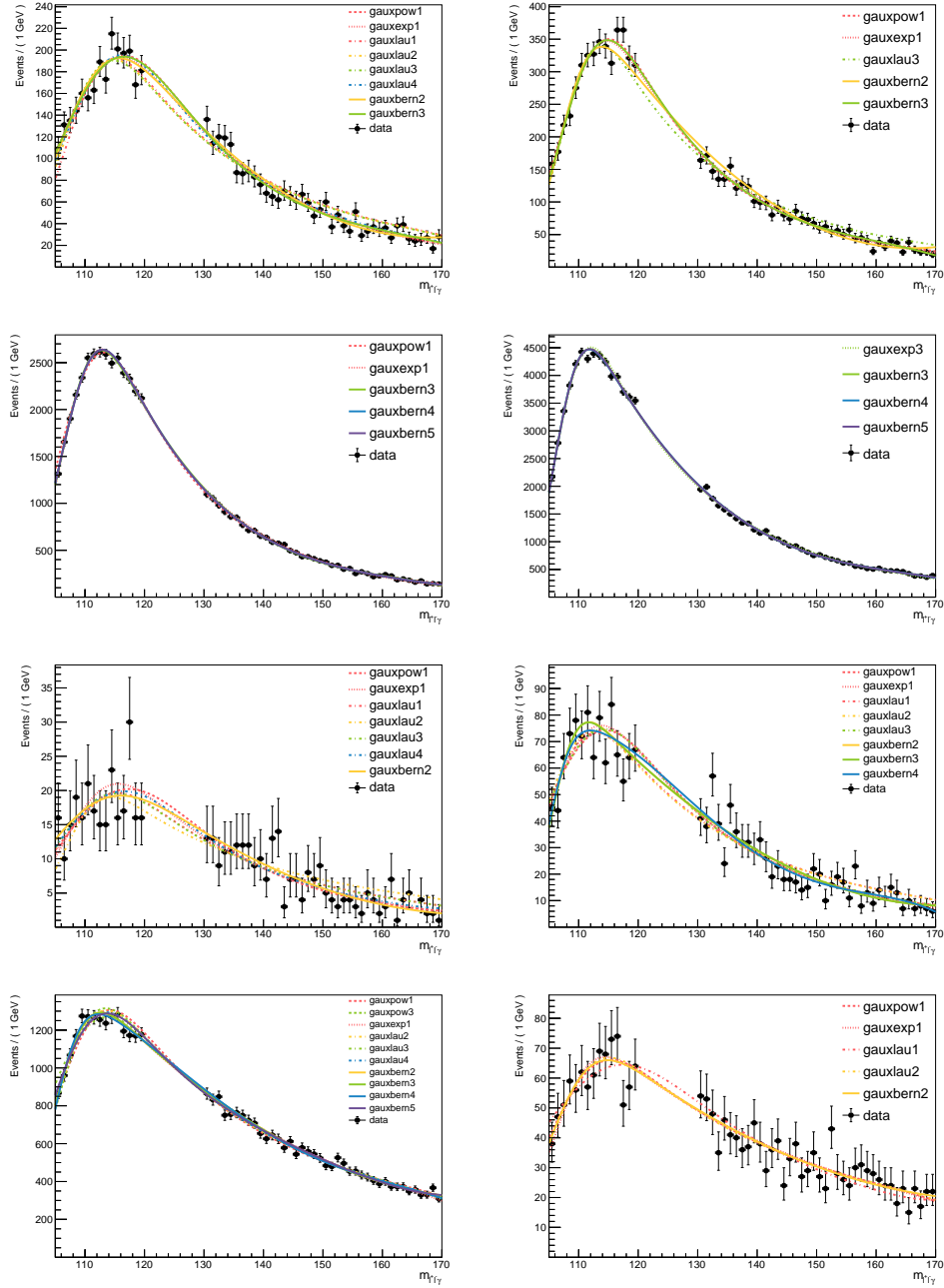
Figure 8.6. Background functions determined by the discrete profiling method. The top four plots correspond to the untagged categories, and the bottom four plots correspond to the dijet categories and lepton tag category.
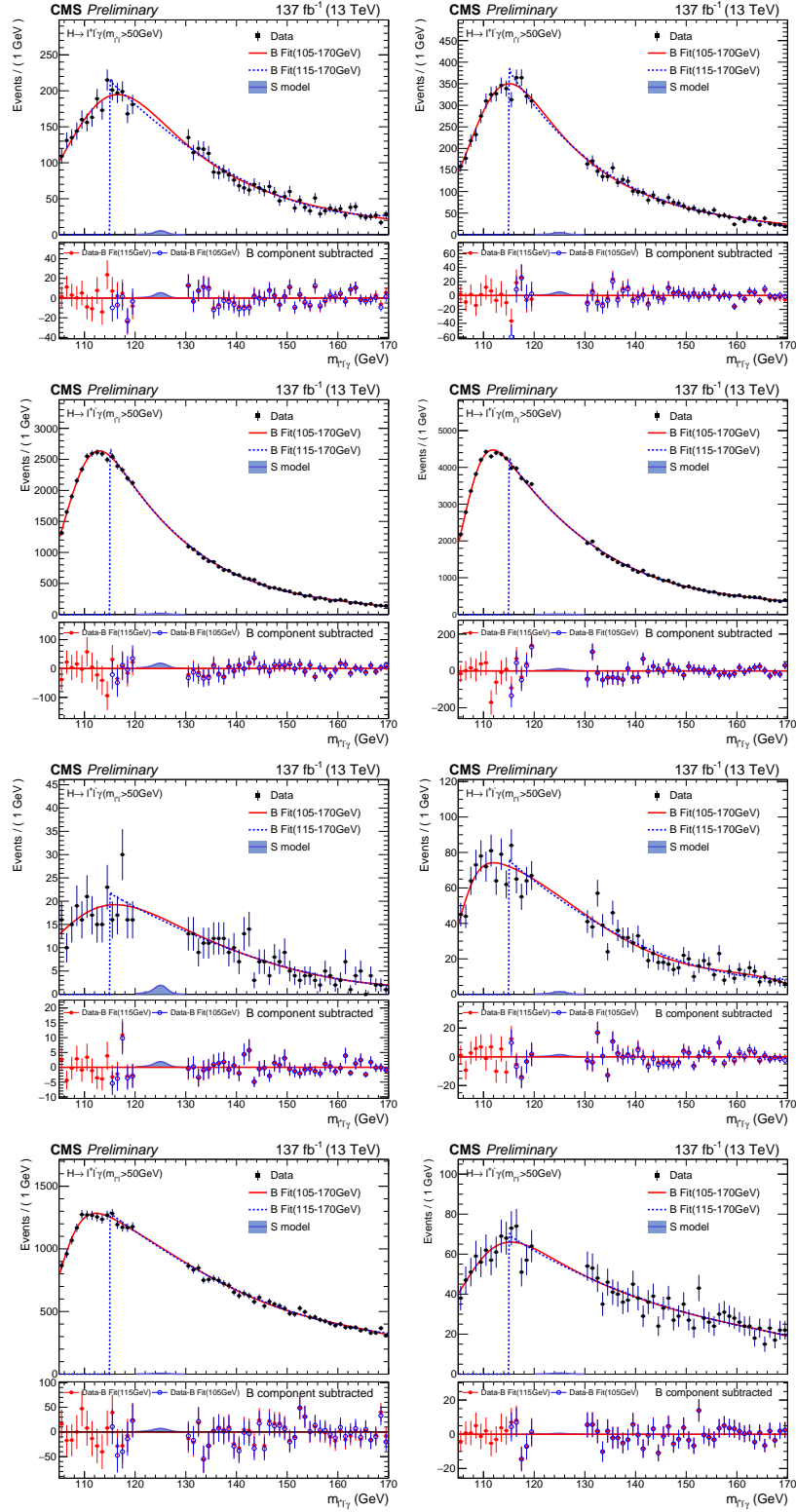
Figure 8.7. Comparison of background fits with and without modeling the turn-on. The top four plots correspond to the untagged categories, and the bottom four plots correspond to the dijet categories and lepton tag category.

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

CHAPTER 10

Statistical Analysis

CHAPTER 11

Results and Interpretation

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

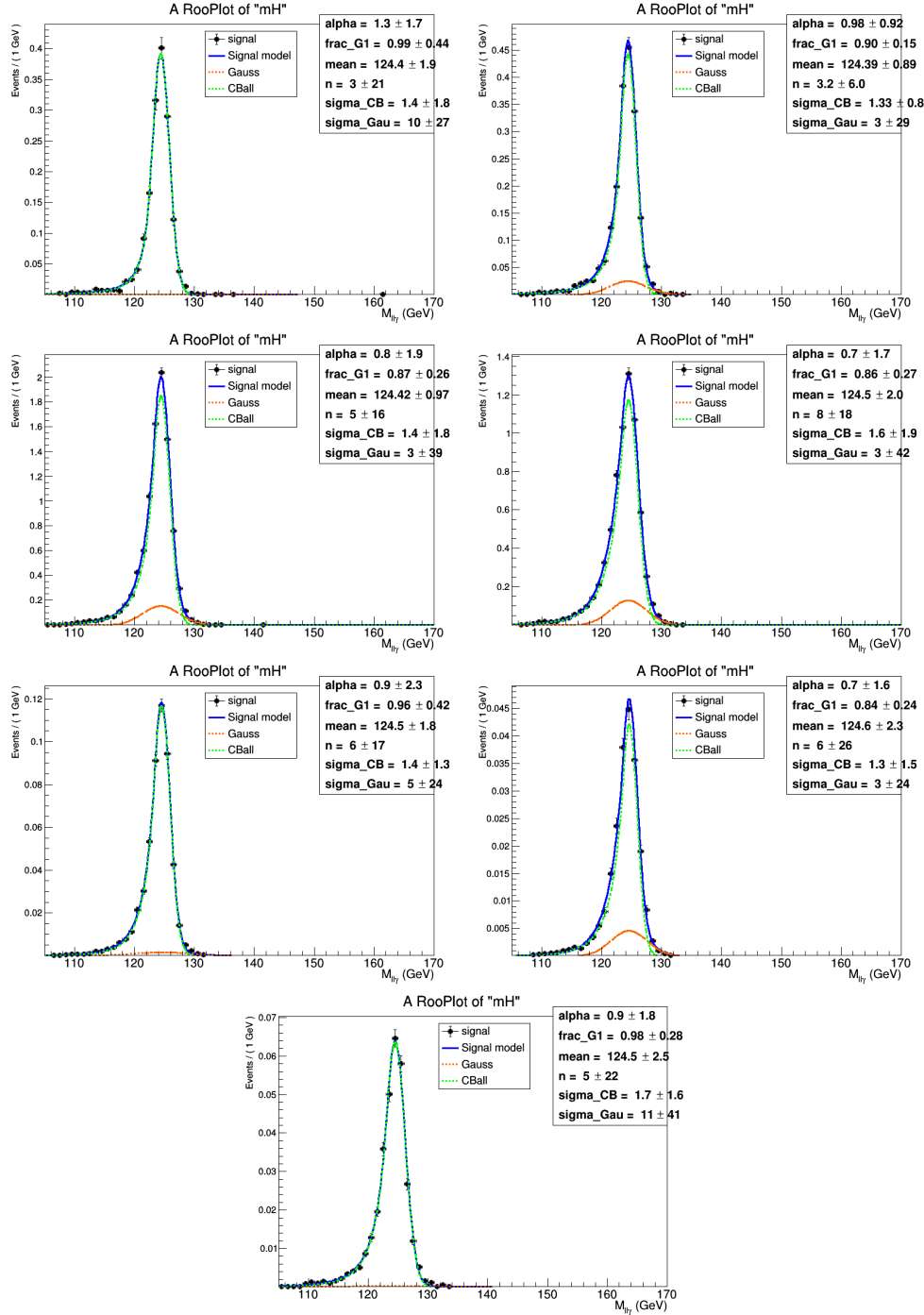### A.3. 2017 Resonant Background Fits

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.
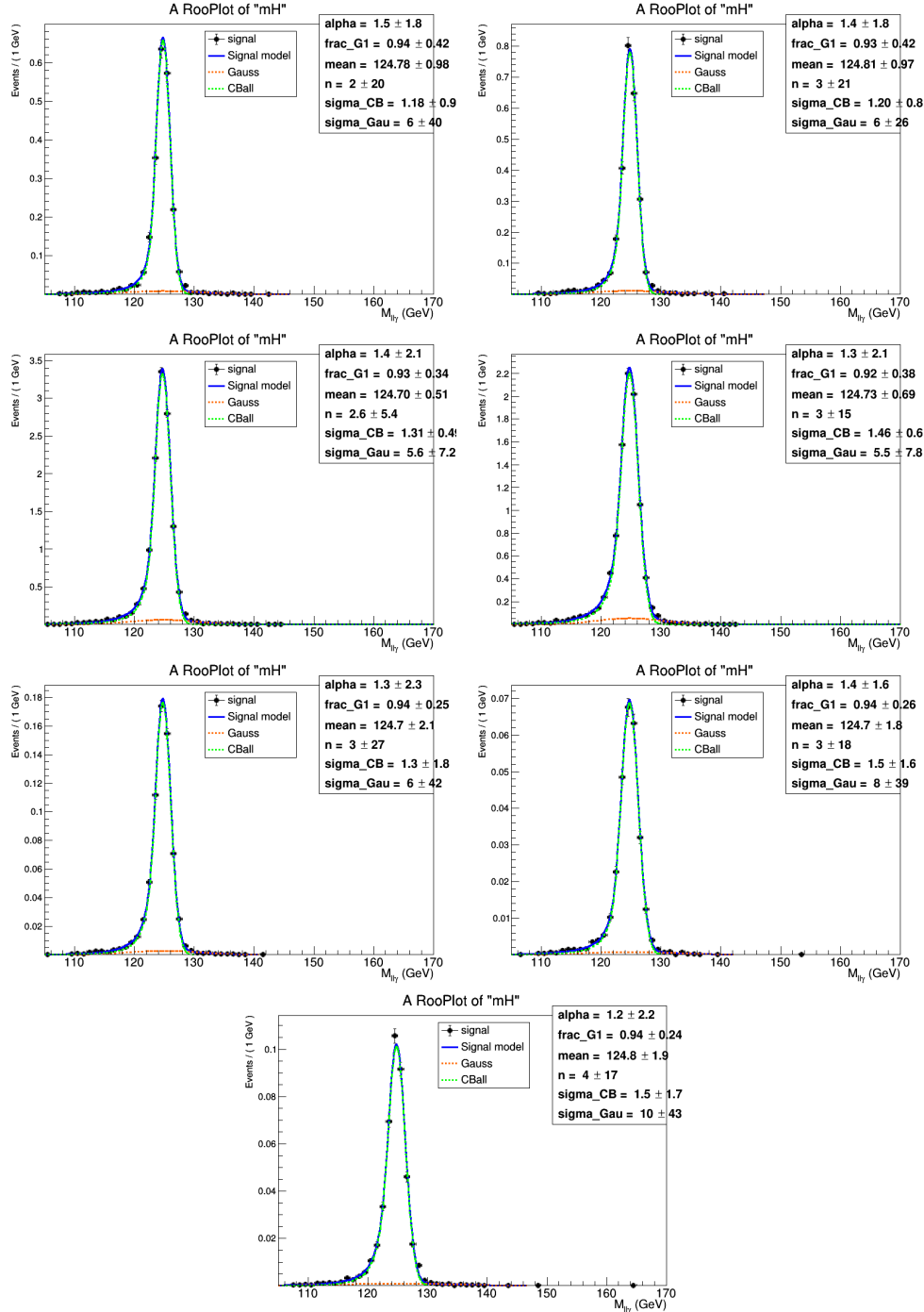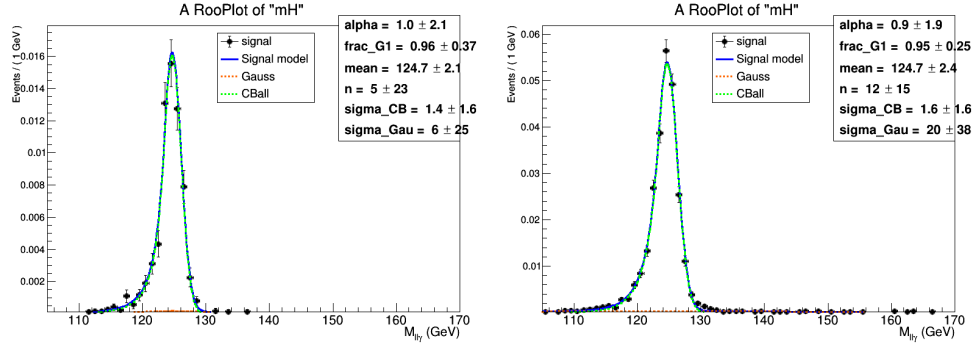
Figure A.2. Fits to simulated $m_{\ell^+\ell^-\gamma}$ signal distributions in the muon channel for $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ for the 2017 data-taking period. The blue line shows the total fit function, the green line shows the Crystal Ball function component, and the red line s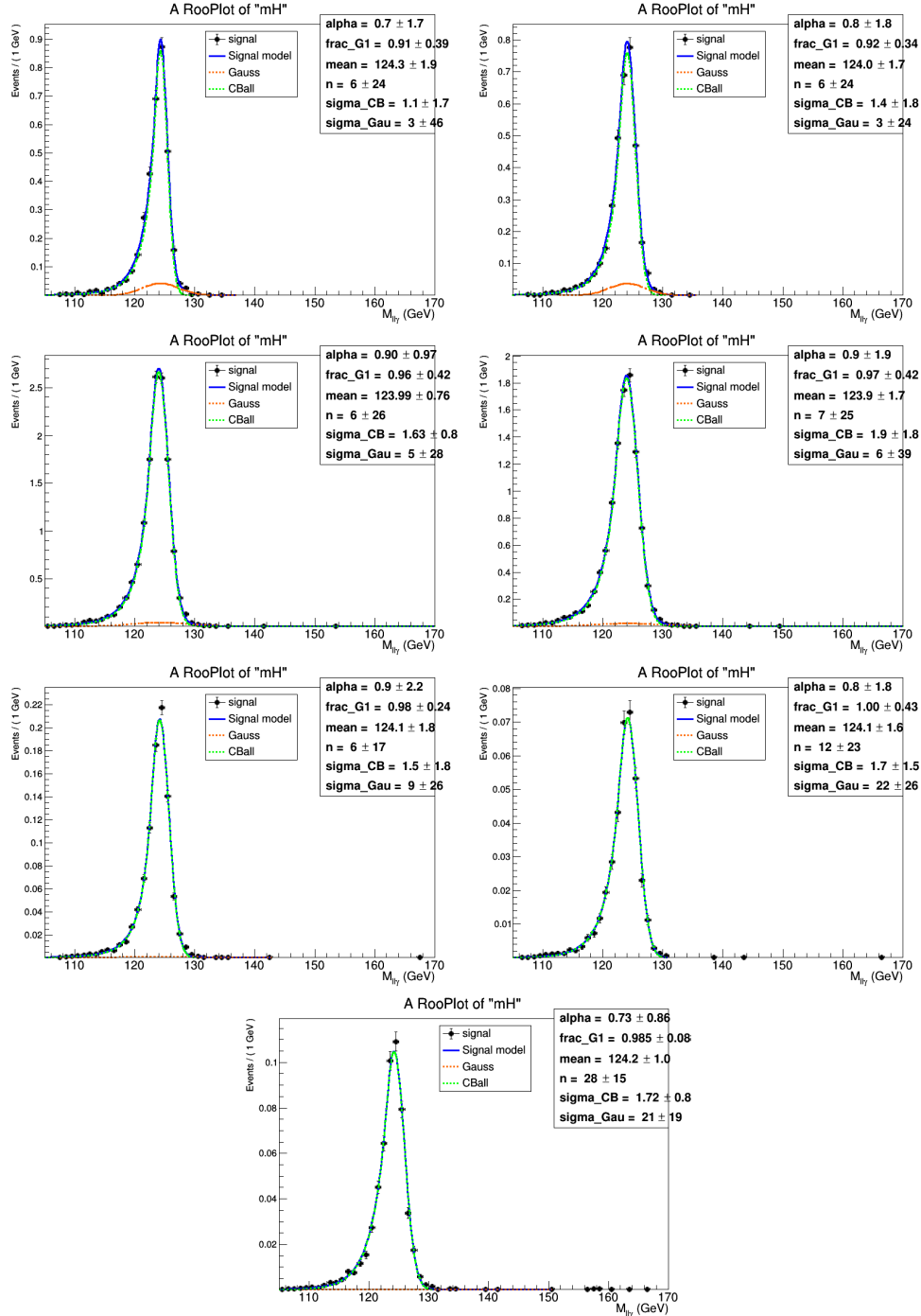hows the Gaussian function component. The top four plots correspond to the untagged categories, and the bottom three plots correspond to the dijet categories.
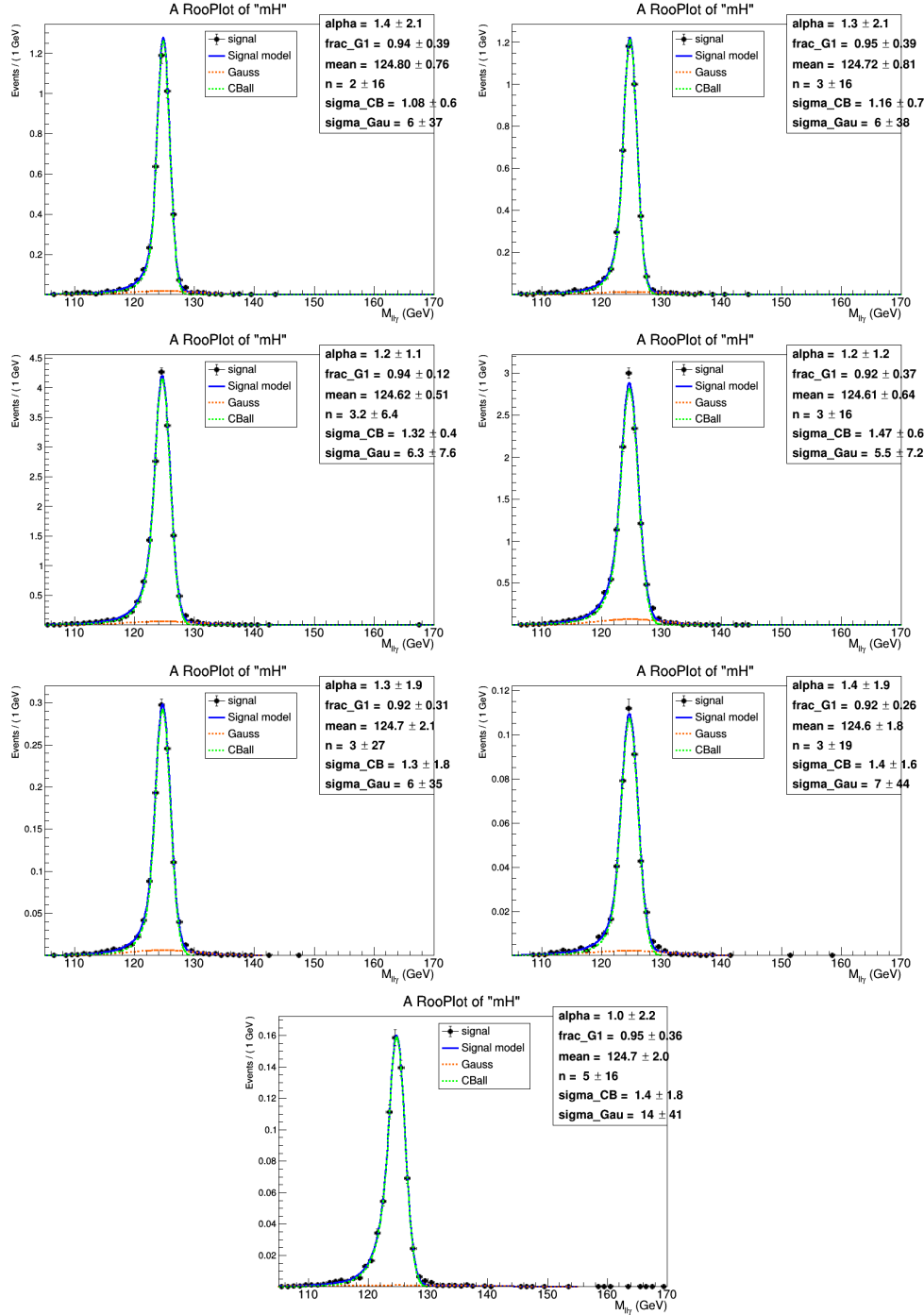
Figure A.3. Fits to simulated $m_{\ell^+\ell^-\gamma}$ signal distributions in the electron and muon channels combined in the lepton-tagged category for $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ for the 2017 data-taking period. The left plot shows the fit to simulated ZH production events, and the right plot 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 $m_H = 125\,\text{GeV}$ 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.
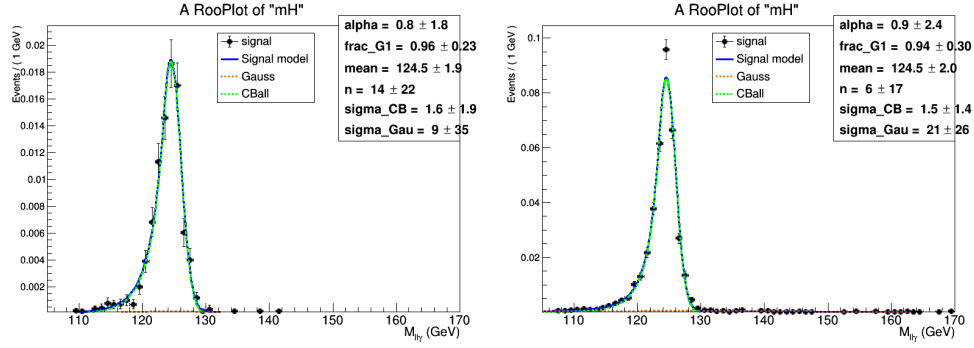
Figure A.6. Fits to simulated $m_{\ell^+\ell^-\gamma}$ signal distributions in the electron and muon channels combined in the lepton-tagged category for $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ for the 2018 data-taking period. The left plot shows the fit to simulated ZH production events, and the right plot 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.
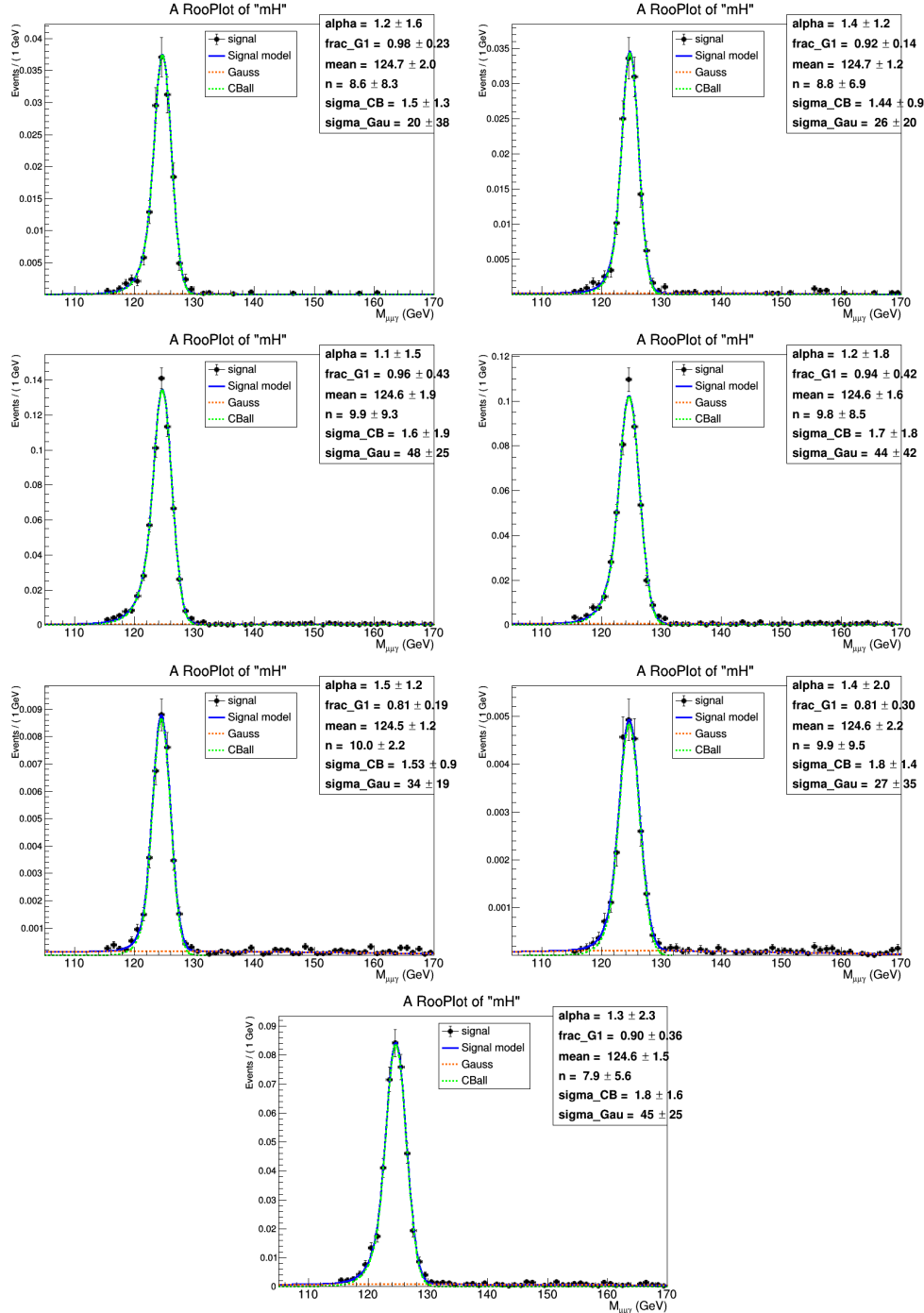
Figure A.7. Fits to simulated $m_{\mu^+\mu^-\gamma}$ resonant background distributions from H $\to \mu^+\mu^-$ for $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ for the 2017 data-taking period. The blue line shows the total fit function, the green line shows the Crystal Ball function component, and the red line shows the Gaussian function component. The top four plots correspond to the untagged categories, and the bottom three plots correspond to the dijet categories.
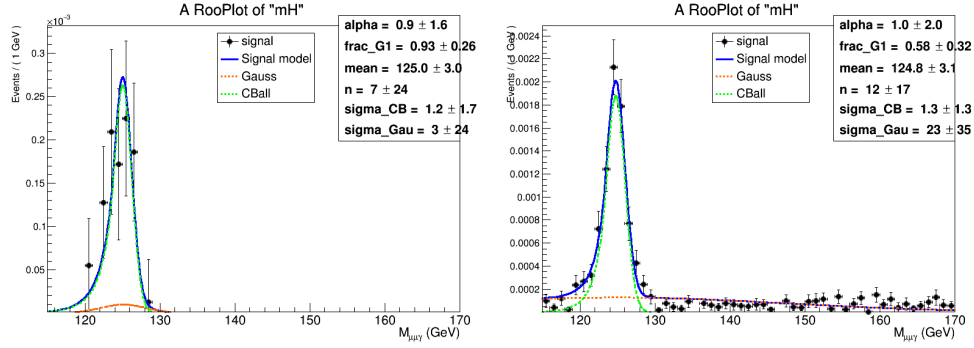
Figure A.8. Fits to simulated $m_{\ell^+\ell^-\gamma}$ resonant background distributions from H $\to$ $\mu^+\mu^-$ in the electron and muon channels combined in the lepton-tagged category for $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ for the 2017 data-taking period. The left plot shows the fit to simulated 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.
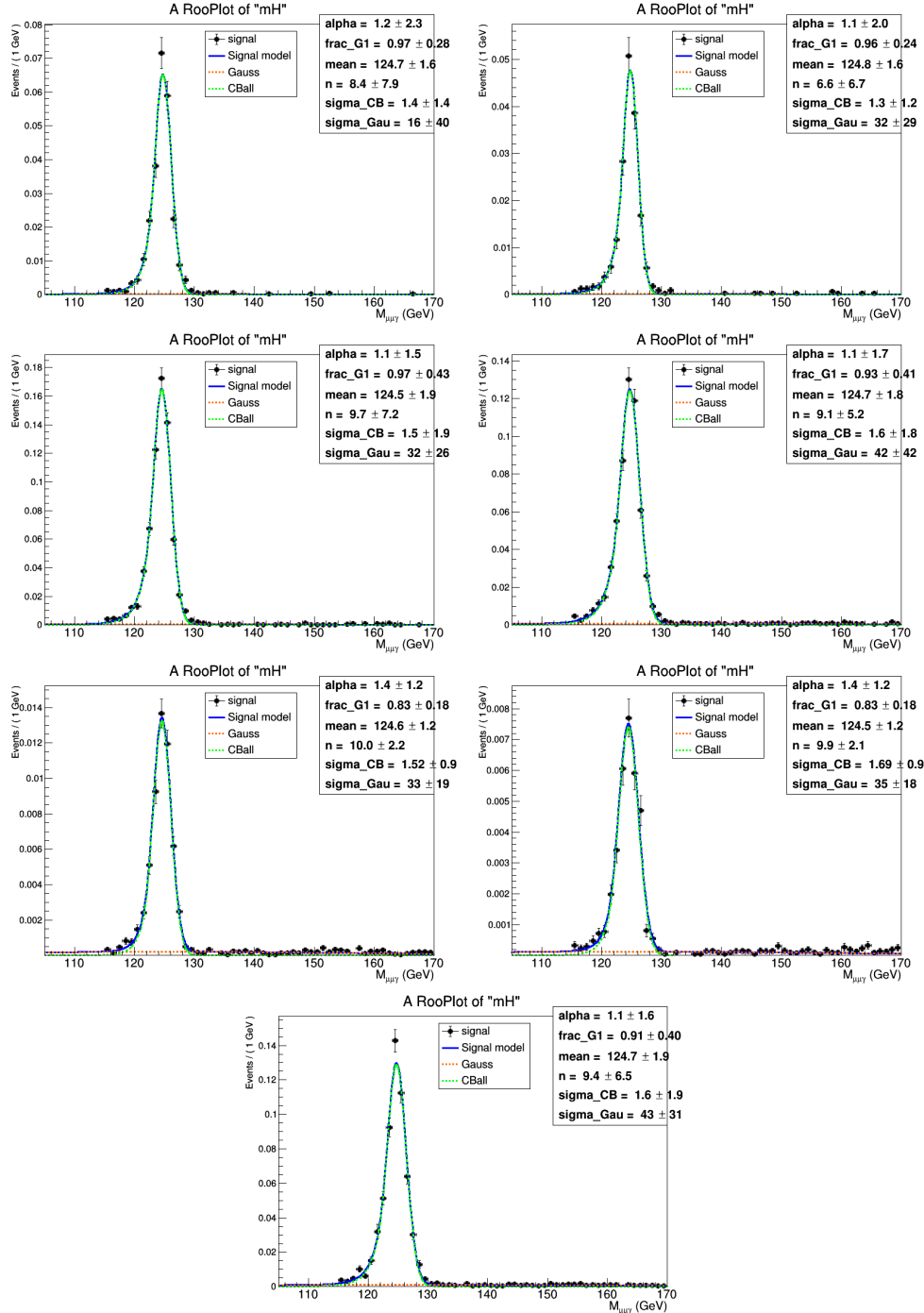
Figure A.9. Fits to simulated $m_{\mu^+\mu^-\gamma}$ resonant background distributions from H $\to$ $\mu^+\mu^-$ for $m_H = 125\,\text{GeV}$ for the 2018 data-taking period. The blue line shows the total fit function, the green line shows the Crystal Ball function component, and the red line shows the Gaussian function component. The top four plots correspond to the untagged categories, and the bottom three plots correspond to the dijet categories.
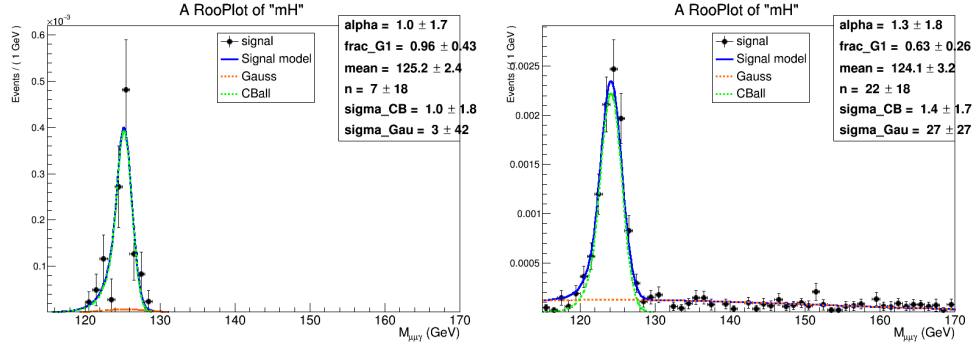
Figure A.10. Fits to simulated $m_{\ell^+\ell^-\gamma}$ resonant background distributions from H $\to \mu^+\mu^-$ in the electron and muon channels combined in the lepton-tagged category for $m_{\mathrm{H}} = 125\,\mathrm{GeV}$ for the 2018 data-taking period. The left plot shows the fit to simulated 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.