Multi-Model Inference Ranking and Applications to Physics at the Large Hadron Collider

Thesis by

Valère Lambert

In Partial Fulfillment of the Requirements

for the Degree of

Bachelors of Science

California Institute of Technology

Pasadena, California

2014

(Submitted May 9, 2014)

Acknowledgements

I would like to express my deepest gratitude to my thesis advisor, Maria Spiropulu, whose faith, determination and enthusiasm not only opened the door for my research in High Energy Physics but also inspired me to continue on in the field.

I would also like to extend my gratitude to Harvey Newman for his inspiration and support, not only for myself but for the entire undergraduate research population.

I want to express my whole-hearted appreciation for my co-mentor, Alex Mott, who has been my role model throughout this experience. Without his guidance and patience, this thesis would surely not have been possible.

I would also like to thank Adi Bornheim, Cristian Peña, Jan Veverka, Si Xie and the other members of the Caltech High Energy Physics group for helping, inspiring and shaping me throughout my research over the past few years.

Finally, I thank the wonderful people who raised me, whose love and support were integral in my journey here.

Abstract

In the measurement of the Higgs Boson decaying into two photons the parametrization of an appropriate background model is essential for fitting the Higgs signal mass peak over a continuous background. This diphoton background modeling is crucial in the statistical process of calculating exclusion limits and the significance of observations in comparison to a background-only hypothesis. It is therefore ideal to obtain knowledge of the physical shape for the background mass distribution as the use of an improper function can lead to biases in the observed limits. Using an Information-Theoretic (I-T) approach for valid inference we apply Akaike Information Criterion (AIC) as a measure of the separation for a fitting model from the data. We then implement a multi-model inference ranking method to build a fit-model that closest represents the Standard Model background in 2013 diphoton data recorded by the Compact Muon Solenoid (CMS) experiment at the Large Hadron Collider (LHC). Potential applications and extensions of this model-selection technique are discussed with reference to CMS detector performance measurements as well as in potential physics analyses at future detectors.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

Scientific modeling involves the generation of conceptual or mathematical models to explain relations revealed through experimental data. Traditional methods for model testing consist of defining a single alternative hypothesis to a null hypothesis. Test statistics are then used to assign probabilistic values for the null hypothesis that define statistical significance. However, these methods are inferentially limited and do not include the uncertainty associated with model selection [17]. In terms of the test statistics, if the probability of the null hypothesis is considered to be low then one concludes that the alternative is a better choice by default. However, the alternative hypothesis may never be tested and the probabilities of the null and alternative hypotheses are uncertain. In effect, the traditional pair-wise hypothesis testing only provides a measure of how good one hypothesis is in relation the second, yet neither is confirmed to fit the data well [9].

A measure of agreement in a global model is needed in order to determine a more rigorous method of model testing. The desired system would involve evaluating the relative worth of alternative hypotheses other than just a one-sided measure of the probability. The conventional methods for model testing can be replaced with Information-Theoretic methods, which provide a strict measure of evidential strength for both the null and alternative hypotheses [17]. These IT methods determine an approach for producing an a priori set of hypotheses and means for quantifying the data based on evidence as well as a ranking of each hypothesis. Such methods use tools such as Kullback-Leibler information, which represents the information lost as a model g is used to represent full reality. In effect, one

wishes to minimize the K-L information and select the model that is closest to representing reality [11].

The primary goal is be to determine a method of avoiding the bias of pair-wise hypothesis testing in order to produce more robust inference results. Given a data set and multiple models corresponding to a phenomenon one wishes to be able to determine a measure of separation or entropy from which to assign a distance for each model from the data [11]. A ranking for the models can be produced according to this index and then an appropriate weight may be assigned to each model. One would potentially be able to isolate data signal forms using the model weights to determine a fundamental expression for that phenomenon [9].

In this study, we develop a multi-model inferencing (MMI) scheme using directional separation as a method for calculating and assigning weights for models in relation to the data. A substantial portion of the Caltech High Energy Physics group with Compact Muon Solenoid (CMS) Collaboration has been highly involved in the analysis of the Higgs boson decaying into two photons. Due to degraded conditions at the Large Hadron Collider (LHC) from increased pile up, extensive work must be done to select a correct modeling of the Standard Model background [1, 22]. We apply the MMI scheme towards better modeling the background of the $H \rightarrow \gamma\gamma$ decay channel in order to characterize the $H \rightarrow \gamma\gamma$ decay. This tool may then be used to better train the photon regression in order to improve the photon tagging efficiency within the CMS electromagnetic calorimeter and aid in further studying properties of the Higgs boson.

We also examine potential future Higgs studies at the LHC as well as applications of the MMI scheme for improved measurements. The recent discovery of a Higgs-like boson with a mass around 125 GeV by the ATLAS and CMS collaborations at the LHC provides unequivocal evidence of some mechanism of spontaneous electroweak symmetry breaking and of the generation of the masses of fundamental particles [1]. One of the most important subsequent measurements for reconstructing the scalar potential of the Higgs doublet field

is the measurement of the Higgs self-couplings. The Higgs self-couplings can be measured by studying the production rates and kinematics of double and triple Higgs boson production at the LHC. These processes are highly suppressed, therefore a large amount of integrated luminosity is required in order for them to be observed and measured. With these measurements in mind, CERN is considering the proposed Phase II upgrades for the LHC in 2022 to the High-Luminosity LHC (HL-LHC), which is expected to be able to deliver a total integrated luminosity of 3 ab^{-1} [22].

The HL-LHC is expected to run with increased pile up at an average of 140 simultaneous pile up events [1]. The most optimistic channel for the di-Higgs analysis is that of one Higgs boson decaying to a pair of photons, and the other Higgs boson decaying to a pair of b-quarks, $HH \rightarrow b\bar{b}\gamma\gamma$. As the photon identification and b-jet tagging efficiencies in the CMS detector degrade with increasing pile up there is the need to re-optimize the algorithms for particle identification to differentiate the signal and backgrounds. Seeing as the predominant background for this channel involves the mistagging of two light jets as b-jets, the potential improvements to the measurements of the $H \rightarrow \gamma\gamma$ channel as well as to the b-tagging efficiencies of the detector may prove highly beneficial to the Higgs self-coupling measurements. The overall increase in data and its complexity in future physics analyses, such as at the HL-LHC, suggests a need for more sophisticated computing methods [22]. The use of multi-model inference may prove to be a valuable contribution to these efforts.

This thesis intends to provide a thorough explanation of model selection and the motivation for multimodel inference while using the CMS $H \rightarrow \gamma\gamma$ analysis at the LHC as a guide for potential current and future applications of the model selection scheme.

Chapter 2 introduces the notion of model selection and expands upon using an Information Theoretic approach with Kullback Leibler Information. It formally explains our selection and use of the Akaike Information Criterion (AIC) as an inference tool. Finally, it covers the multi-model inference and model averaging procedure that has been designed using AIC as a method to select an appropriate model for the $H \rightarrow \gamma\gamma$ Standard Model back-

ground.

Chapter 3 provides an explanation of the Large Hadron Collider, Compact Muon Solenoid experiment, $H \rightarrow \gamma\gamma$ analysis, and the composition of the related Standard Model background. It covers our analysis applying the multi-model inference technique to 2013 $H \rightarrow \gamma\gamma$ data from the LHC. We discuss the results of the selection as well as the potential bias to the Higgs signal region based on the resultant fit model.

Chapter 4 presents the dependence of the $H \rightarrow \gamma\gamma$ analysis on the detector performance of the CMS electromagnetic calorimeter. It covers the measurement of the photon energy resolution and scale as well as an explanation of the analysis methodology using $Z \rightarrow \mu\mu\gamma$ events. Finally, we discuss potential improvements to the detector performance analysis, including applications of the MMI method.

Chapter 5 then provides an overview of potential measurements of the Higgs Self-Couplings at a future CMS detector under conditions at the proposed High-Luminosity Large Hadron Collider as a motivation for future detector design. We discuss the expected detector performance results with current detector technology and propose areas of improvement for future analyses.

Chapter 2

Model Selection

2.1 Model Selection

Traditional model testing often involves the comparison of two differing hypothesis, in which one is selected over the other without confirmation that either fit the data well. The concept of null hypothesis testing only provides arbitrary dichotomies and can often end in unsubstantial results where the null hypothesis is false on a priori grounds [17] It is typical, when selecting a model function for an analysis, to choose one function for measurement and try to prove that it is acceptable. A process which is inheritantly biased in its design. Under classical model selection , one shifts more towards model-based inference [9].

In the context of model selection, it is assumed that data and a set of models exist and that statistical inference is model based. There is no certainty as to which model should be used but there is the assumption that there is a single correct model, or at least, a best model. Finally, it is presumed that the selected ”best” model suffices as the sole model that fits the data and from which all inferences about data may be made [10]. The pitfall of this is that the uncertainty related to the model selection itself is ignored, something that seems justified as the single ”best” model has been found [9]. It is difficult to construct an adequate model based the information of a finite set of observations [2]. In practice the models are only approximations and are different from reality, therefore the hypothesis testing procedure is not properly formulated as a procedure of approximation [2].

The progression to data-driven model selection which seeks to determine appropriate models and parameters from the data itself still contain these limitations in which the uncertainty linked to the final model selection is ignored [17]. We are in need of of a new criterion that provides quantitative information to judge the strength of evidence. This criterion must be estimable for each fitted model from the data and must be incorporated in a general statistical inference framework. In short, the model selection must be justified and operate within either or both a likelihood or Bayesian framework [10]. The difficulties in designing rigorous model selection procedures reside in the method for selection an appropriate set of possible models, for which there remains no systematic way of generating [14]. The realistic aim should not be on seeking a final truth but rather using common sense to make useful predictions [17] If these prediction prove useful then that is confirmation that the hypothesis space is acceptable a this time, with the possibility of expanding or revising it later.

2.2 Kullback-Leibler Information

Traditional hypothesis testing is limited in the range of potential models examined and the robustness of the inference technique used to determine how well the model describes the data. We wish to move away from the dichotomy of rejecting or not rejecting individual hypotheses and provide a quantitative probability of agreement. We consider the use of information-theoretic approaches to provide a formal measurement of the evidence for a model given the data, and introduce the concept of Kullback-Leiber (K-L) Information which represents the information lost when a model is used to represent ”reality,” or the data [11]. Denote f as full reality or truth, having no parameters and being invariant with sample size. Now denote g as the approximating model which represents a probability distribution of how likely one will observe reality and is dependent on sample size. The information that is lost when one uses g to model f can then be expressed as a ratio of the two in Equation (2.1).

$$I(f,g) = \sum_{i=1}^{k} f(x) log\left(\frac{f(x)}{g(x)}\right) dx \tag{2.1}$$

Through logarithmic relations this can be separated to show the K-L Information as the difference between reality and the expectation of the approximating model.

$$I(f,g) = C - E_f[log(g(x|\theta))] \tag{2.2}$$

Where, $C = \int f(x) log(f(x)) dx$. The information lost, I(f, g), cannot be calculated itself seeing as it requires knowledge of full reality. However, one is able to estimate the relative K-L information, $E_f[log(g(x|\theta))]$, of competing approximating models [11].

2.3 Akaike Information Criterion

Our selected approach is to use Akaike Information Criterion (AIC), which is a test statistic that minimizes the K-L information and provides a measure of relative support for each model to the data [11]. The standard test statistic used for deterring how well a model fits the data is the Log-Likelihood of the fit. The use of the maximum likelihood for statistical model provides a method for estimating the free parameters of a model given a specified dimension and structure [2]. For a parametric candidate model, the likelihood function the estimates the conformity of the model to the data. When the complexity of this models increased, it is able to conform to various additional characteristics of the data. Therefore, the selection of the fitted model that maximizes the likelihood of the fit with undoubtedly determine the most complex model in the model set. To correct for this, Hirotugo Akaike extended this method to consider the approach where the model dimension is also unknown, and therefore the model estimation and selection are simultaneously derived from data [2].

Akaike determined a relationship between likelihood theory and K-L information, in particular that the Log-Likelihood statistic is a biased estimate of the relative K-L information, $E_x E_y[log(g(x|\theta(y)))]$. Moreover, it was determined that the biasing factor is asymptotically related to the number of estimable parameters in the model.

$$AIC = -2log\left(L(m|Data)\right) + 2k + \frac{2k(k+1)}{(n-k-1)} \tag{2.3}$$

This form of the criterion is more accurate, particularly for settings where the data set, n, is small and the number of free parameters, k, is relatively large. Note that for large values of n in comparison to k, the AIC penalization factor asymptotically converges to 2k, twice the number of estimable parameters..

$$AIC = -2log\left(L(m|Data)\right) + 2k \tag{2.4}$$

The relative likelihood for each model is based on the separation of each model's AIC value from that of the minimum AIC value in the set, Δ_i. These likelihoods can then determine each model's respective Akaike weight which can be used to compare the models as well as produce a combined model from the constituent set. [11],

$$w_i = \frac{e^{-\frac{1}{2}\Delta_i}}{\sum_{r=1}^{n} e^{-\frac{1}{2}\Delta_r}} \tag{2.5}$$

The resulting benefit of using AIC over the standard Log-Likelihood involves the balancing of under and over-fitting the data. A model with too few parameters results in a poor fit of the data, biases the expected measurements and will therefore have a poor Log-Likelihood. However, by increasing the number of parameters in the approximating model and the degrees of freedom for the fit, one has the potential to fit any shape well without deriving any relative mathematical relation. Therefore, a model with too many parameters may have a superior Log-Likelihood but also suffers from poor precision and may identify spurious effects in the data [30]. To balance these issues AIC introduces this penalization factor to the Log-Likelihood based on the number of estimable parameters the model uses. The optimal fitted model is then identified as that which has the minimum AIC value, however all values are considered for assessing the criterion preferences [2]. The use of AIC then provides a more rigorous method to calculate the separation for each model from the data and select the model that is closest to representing reality.

A benefit of AIC is that it is asymptotically efficient. In the event that the generating model is of finite dimension and this model is within the candidate set, a consistent criterion will asymptotically select the correct structure with probability one. However, in the practical sense where the generating model is of infinite dimension and therefore lies outside of the candidate collection, an asymptotically efficient criterion selects fitted candidates that minimize the mean squared error of prediction [2]. The application of the Akaike Information Criterion therefore, does not rely on the assumption that one of the candidate models is the "true" model. Another substantial advantage of information-theoretic criteria, such as the AIC, is that they are valid for non-nest models whereas traditional ratio tests are defined only for nested models, thus limiting their use in hypothesis testing for model selection [9].

2.3.1 AIC as a Bayesian result

In consideration of determining the proper inference tool for model selection, it is appropriate to compare our choice of the Akaike Information Criterion with other IT criteria. The most popular alternative to AIC in data-based model selection is called Bayesian Information Criterion (BIC), which at first glance seems to have a similar construct.

$$BIC = -2ln(L) + klog(n) \tag{2.6}$$

However, the use of "Bayesian" for BIC may be considered a misnomer as it is not related to information theory and can be derived as a non-Bayesian results [10]. For very large samples, the model selected by BIC is the best for to use for inference. However, as the sample sizes become more moderate, the BIC-selected model becomes more sparing than model g, particularly if this model is the most general in the set. Concern arises for realistic sample sizes as the BIC-selected model tends to under fit at the given n as it approaches the target model from below as n increases. Due to the assumption in the derivation of BIC that there was a true model, independent of n, that generated the data, the target model is also not dependent on the sample size [9]. The derivation implies that the true model will be in the model set and that this will be the target model for BIC selection. Unfortunately this limits the selection of the true model only in the limit that the sample size is very large [10].

Despite BIC being a misnomer as a Bayesian result, it has been shown by Burnham and Anderson [9] that AIC can be justified as Bayesian with the use of a "savvy" prior on models that is a function of both sample size and the number of free parameters. The formulation of Akaike Information Criterion is built on the minimization of K-L Information. For K-L Information, there is no concept of a true model implied and no assumption is made that the models must be nested [10]. Akaike found that an approximately unbiased estimator of the $E_y E_x[log(g(x|\hat{\theta}(y)))]$, the expected K-L Information for a model given the data. This finding allows for the combination of estimation and model selection under a unified optimization framework [2]. The asymptotic bias correction term is in no way arbitrary and allows for the values of the AIC to be dependent on sample size themselves [10].

The determination of AIC as a Bayesian result is actually derived from BIC [9]. The BIC model selection comes about in the context of large-sample approximations to the Bayes factor, along with assuming equal priors on models [10]. The Bayesian posterior model probability is approximated as,

$$Pr(g_i|data) = \frac{exp(-\frac{1}{2}\Delta BIC_i)q_i}{\sum_{r=1}^{R} exp(-\frac{1}{2}\Delta BIC_r)q_r} \tag{2.7}$$

This posterior depends not just on the data, but also on the model set and the prior distribution on those models. Akaike weights can easily be obtained by using the model prior,

$$q_i \propto e^{(\frac{1}{2}\Delta BIC_i)}\dot{e}^{(-\frac{1}{2}\Delta AIC_I)} \tag{2.8}$$

It is clearly shown then that,

$$e^{(-\frac{1}{2}\Delta BIC_i)}\dot{e}^{(\frac{1}{2}\Delta BIC_i)}\dot{e}^{(-\frac{1}{2}\Delta AIC_i)} = e^{(-\frac{1}{2}\Delta AIC_i)} \tag{2.9}$$

Hence, with the implied prior probability distributions on models, we get,

$$p_i = \frac{e^{(-\frac{1}{2}\Delta BIC_i)}q_i}{\sum_{r=1}^{R} exp(-\frac{1}{2}\Delta BIC_r)q_r} = \frac{e^{(-\frac{1}{2}\Delta AIC_i)}}{\sum_{r=1}^{R} e^{(-\frac{1}{2}\Delta AIC_r)}} = w_i \tag{2.10}$$

This corresponds with the Akaike weight for model g_i. This prior probability on models can be expressed in a simple form as,

$$q_i = C\dot{e}^{(\frac{1}{2}k_i log(n) - k_i)} \tag{2.11}$$

where $C = \frac{1}{\sum_{r=1}^{R} e^{(\frac{1}{2}k_r log(n) - k_r)}}$ Therefore, one may determine that, for large samples, the Akaike weights from AIC are Bayesian posterior model probabilities for this model prior [9].

Given a model, the prior distribution on the data should not depend on the model set size, R. The BIC approach assumes a prior that would not depend on sample size nor the number of parameters which is neither necessary nor reasonable [10]. There is limited information in a sample, so the more parameters that are used for estimates, the poorer the average precision will be. The prior $q_r = 1/R$ is neither reasonable nor innocent as it implies that the target model is reality rather than the "best" approximating model, given that the parameters are to be estimated. Seeing that the errors on individual parameters depend on sample size, it is reasonable that the appropriate model would as well [9].

Given that AIC can be derived from the BIC approximation to the Bayes factor, this distinction between the two cannot be based on a Bayes versus frequentist view. The distinction is more focused on the prior model, $q = 1/R$ for BIC and K-L for AIC [10]. Seeing that too few parameters wastes information and too many leads to imprecise results, the latter prior's dependency on the number of estimable parameters and sample size makes it a reasonable choice. In summary, BIC corresponds to the measurement of the odds of a model being the true model given the data whereas AIC is an estimator of the information lost when approximating the truth [17]. In consideration of the departure from the selection of one "best" model, the choice of AIC is deemed as an appropriate measure of the relative support for each model in a set given the data.

2.4 Multi-Model Inference

The concept we wish to introduce with regards to multi-model inference is that of an approach to model selection through multi-model averaging. The approach supports the notion that more than one ”best” model may be supported by the data and it permits the evaluation of certain model selection uncertainty [17]. The approach begins where a set of plausible models are defined a priori, taking in account the sample size and previous knowledge of influential parameters. AIC is used to evaluate the empirical support for each model from the data, expressed as a weight corresponding to the probability of the model being the best approximating model given the model set. The estimated probabilities sum to 1 across the model set such that they may be used to rank models, quantify the extent of evidence in favor of each model and evaluate multi-model averaged effects of the results [10, 17]. The averaging process allows us to define a ”composite” model which we define as a model built of the sum of each independent model in the model set, weighted by the probability assigned for the model being the best approximating model. The use of composite models weighted by the empirical support for each model has been shown to be superior to constructing inferences for the relative importance of variable based on only one ”best” model, particularly when the second and third best models are similarly supported by the data [10].

Consider the case where two models have probabilities 0.5 and 0.45 with the other models in the set have much lower probabilities. In accounting for the uncertainty linked to model selection, the two models may be deemed to represent the majority of the evidence together and there is no true reason to select one over the other, particularly as the ”best” model only has a probability of 1.1 times that of the next best model. In addition to losing information, there remains the possibility that another sample would yield another best model [17].

The approach is therefore not oriented to testing a specific hypothesis but rather determining a model that is closest to representing the data. We use the example of multi-model averaging in industrial hygiene literature by Lavoue et al. in which models are compared to determine a statistical link between measured exposure levels and environmental vari-

ables [17]. The results of the analysis lead to a subset of the variables being identified as determinants that are viewed as non-influential based on their presence, or lack thereof, in the final chosen model. The method leading to the selection of the final model has a substantial impact on the conclusions drawn from the analysis as there is no specific aim to test a particular hypothesis but rather identify influent variables in a set of plausible candidates.

2.4.1 MMI Method

In order to expand the range of potential models used in determining the appropriate characterization of the Standard Model background we implement a multi-model inference method. One begins by defining a set of all plausible models, R. This set is still limited in that the only models tested are those added to this set. For each model in R, we fit the data and determine the likelihood for that model with respect to the data, $L(m|Data)$. Then we calculate the AIC for the model,

$$AIC = -2\log\left(L(m|Data)\right) + 2k + \frac{2k(k+1)}{(n-k-1)} \tag{2.12}$$

where k represents the number of floating model parameters and n denotes the sample size. The model with the lowest AIC values is then selected and one uses the AIC distances, Δ_i, to determine the relative probability that the ith model minimizes the information lost when representing the data.

$$\Delta_i = AIC_i - AIC_{min} \tag{2.13}$$

This already provides a ranking metric based on the empirical support for each model, as can be seen in Table 2.1.The model corresponding with the lowest AIC value is deemed as the ”best” model in the set for representing the data. Models with very weak relative likelihoods ($\Delta_i \sim 10$) may be omitted from further investigation [11]. For the remaining models one has the following considerations:

1. More data may be acquired to help distinguish between the models.

2. One may conclude that the data is insufficient to select one model over another.

3. One may alternatively combined the models with their weighted averages.

The first consideration is impractical as one is often limited in their given data set from which they wish to draw inferences. The second leads to no conclusion, however the third provides a solution derived from the previous consideration. One may not be able to prefer one model objectively over another, however, one may effectively expand the original model set to include every possible combination of the original models by using their weighted averages.

Δ_i	**Level of Empirical Support for Model i**
0-2	Substantial
4-7	Considerably Less
10 +	Essentially no support

Table 2.1: Interpretation of empirical support for AIC distances [11]

From the relative likelihoods for each model one may determine their respective Akaike weight [11],

$$w_i = \frac{e^{-\frac{1}{2}\Delta_i}}{\sum_{r=1}^{n} e^{-\frac{1}{2}\Delta_r}} \tag{2.14}$$

With the remaining models in the investigation one then produce a composite "realistic" model:

$$M_{realistic} = \sum_{i=1}^{r} w_i M_i \tag{2.15}$$

This approach allows one to further expand the pool of tested models by including not just the set of individual models but every possible weighted combination.

2.4.2 Preliminary Closure Tests

As a closure test of the MMI method that we have in place we test two theoretical cases with generated data from known truth functions. In the first study we set the truth model, or ”reality,” to be a simple exponential and generate toy data sets of 10,000 events based off of this known model. We then implement the MMI method with our set of plausible models in Table 2.2. The second study follows the same process but with a power law truth model.

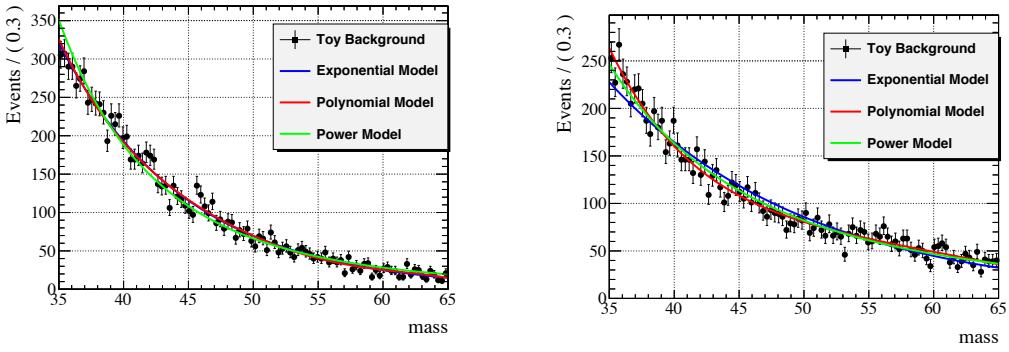
Plausible Model Set	
0. Single Exponential	$c_1 e^{\alpha x}$
1. Double Exponential	$c_1 e^{\alpha x} + c_2 e^{\beta x}$
2. Triple Exponential	$c_1 e^{\alpha x} + c_2 e^{\beta x} + c_3 e^{\gamma x}$
3. Modified Exponential	$c_1 e^{\alpha x + \eta}$
4. 5th-Order Polynomial	$c_5 x^5 + c_4 x^4 + c_3 x^3 + c_2 x^2 + c_1 x + c_0$
5. Single Power Law	$c_1 x^{\alpha}$
6. Double Power	$c_1 x^{\alpha} + c_2 x^{\beta}$

Table 2.2: Model set used for closure tests.

In Figure 2.1, we see that it is difficult to make an ad-hoc decision between the models within the set when considering how they characterize either toy data set. By comparing the results in Table 2.3 we see that all four of the exponential models had the same Log-Likelihood value but ,with the penalization factor included in the AIC measure, the truth model of the single exponential was selected with the most significant weight of 0.62. The only other model with substantial support is the modified exponential which is inherently a single exponential with an added constant. The polynomial and power law models are all ”rejected”as their Δ_i values exceed the cut-off.

The difference in the AIC method is also seen Table 2.4, where both power models shared the same Log-Likelihood with the polynomial and triple exponential models. In this case

the penalization factor for the added parameters also selected for the truth model, giving it a weight of 0.81 and leaving the only other semi-substantially supported model, the double power law, with a weight of 0.14. These closure tests illustrate the fall-backs of the Log-Likelihood method alone and the potential of the AIC method to compensate for the overfitting bias.



(a) Single Exponential Truth Model

(b) Single Power Law Truth Model

Figure 2.1: Comparisons of fits for exponential and power truth models.

	AIC Results for Single Exponential Truth Model						
Models	Single Exponential	Double Exponential	Triple Exponential	Modified Exponential	5th-Order Polynomial	Single Power	Double Power
Log(L)	**741977**	741977	741977	741977	741974	741789	741789
AIC	**-1483950**	-1483946	-1483942	-1483948	-1483936	-1483574	-1384570
AIC Δ_i	**0**	4.00	8.00	1.53	14.6	377	381
AIC w_i	**0.62**	0.08	0.01	0.29	4.e-4	1.e -82	1.e -83

Table 2.3: MMI closure results for Single Exponential Truth Model

	AIC Results for Single Power Law Truth Model						
Models	Single Exponential	Double Exponential	Triple Exponential	Modified Exponential	5th-Order Polynomial	Single Power	Double Power
Log(L)	725200	725291	725296	725292	725296	**725296**	725296
AIC	-1450396	-1450574	-1450580	-1450578	-1450580	**-1450588**	-1450584
AIC Δ_i	191	12.7	7.5	9.4	7.3	**0**	3.5
AIC w_i	3.e -42	0.001	0.02	0.008	0.02	**0.81**	0.14

Table 2.4: MMI closure results for Single Power Law Truth Model

Chapter 3

Standard Model Background

To fit the Higgs signal mass peak over a continuous background we wish to determine the most appropriate model for the Standard Model Background. This background is crucial in the statistical procedure for calculating exclusion limits and the significance of observations in comparison to a background-only hypothesis [7]. The current method involves the use of a 5th-Order polynomial which fits many shapes with many parameters, but does not necessarily reflect reality. It is ideal to obtain knowledge of the physical shape for the background mass distribution as the use of an incorrect function can lead to biased observation limits. By implementing this MMI method on $H \rightarrow \gamma\gamma$ data we wish to build a composite model which best represents the data, from many physics motivated models.

3.1 Standard Model

The Standard Model of particle physics is a unified theory meant to describe the interactions among elementary particle physics. The non-abelian gauge field theory is based on the symmetry group $SU(3) \otimes SU(2) \otimes U(1)$ and has 12 generators with non-trivial commutators. The $SU(2) \otimes U(1)$ component describes the electroweak interactions , Quantum Electrodynamics (QED), which unify electric and magnetic forces as the electromagnetic force, along with the weak force [3]. The electroweak force describes interactions among all particles with the exception of quarks, which are best described by the strong force [31]. The SU(3) group corresponds with the color group of the theory of strong interactions, Quantum Chromodynamics (QCD) [3]. Though the theory is not complete, the Standard

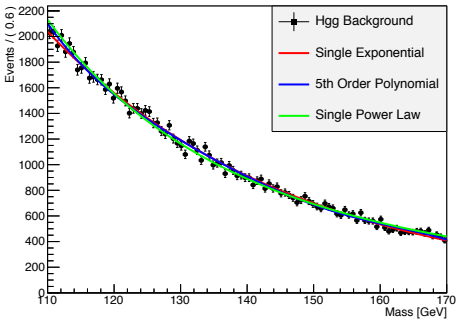
Model has been tested experimentally to unprecedented precision, with the Higgs boson being the last particle, predicted by the Standard Model, to be confirmed [31].

The study of the Higgs boson and its properties is a key topic in particle physics as it is not only responsible for spontaneous electroweak symmetry break which attributes mass to particles, but also holds prospects in isolating other studies on physics beyond the Standard Model. The primary production mechanism for the Higgs at the LHC is through gluon fusion with addition minor contributions from vector boson fusion (VBF) as well as production with a W or Z boson, or a $t\bar{t}$ pair [28]. The most promising channel for the measurement of the SM Higgs boson is that of its decay into two photons, $H \rightarrow \gamma\gamma$. The channel has a very small branching ratio, varying between 0.14% and 0.23% between 100 and 150 GeV, as well as a large diphoton continuum background [31]. However, due to the optimized reconstruction and high energy resolution for photons at the CMS detector, the channel provides a clean signature with a well-defined peak on a smooth, continuous background.

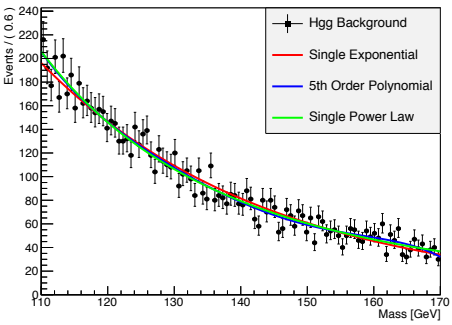
In order to extract the Higgs mass signal it is necessary to be able to isolate the signal peak from the continuous Standard Model background. The $H \rightarrow \gamma\gamma$ backgrounds are dominated by QCD processes and arise from irreducible prompt diphoton production, as well as reducible $pp \rightarrow \gamma + jet$ and $pp \rightarrow jet + jet$ where one or more photons are mistagged jets [23, 28]. The kinematic distributions of these backgrounds are not precisely modeled due to their generation at leading order and the complex nature of jets being misidentified as photons [31].

In order to fit the Higgs signal mass peak over a continuous background, one must first designate an appropriate model parameterization for the background. This model then serves as a fully-differential prediction of the mean expected diphoton mass distribution for the background-only hypothesis and is therefore essential for the statistical procedure that defines exclusion limits or the significance of an observation [7]. It is desirable to have an exact understanding of the functional form of the background mass distributions, or at

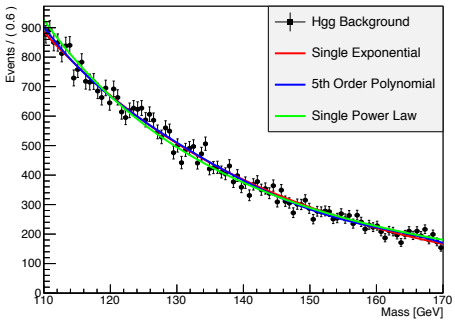
least be able to express the limited knowledge of its shape in a finite set of parameters. The use of an inappropriate function for the background model can lead to undesirable effects in the extraction of the signal yield, such as biases in the observed limits [15].



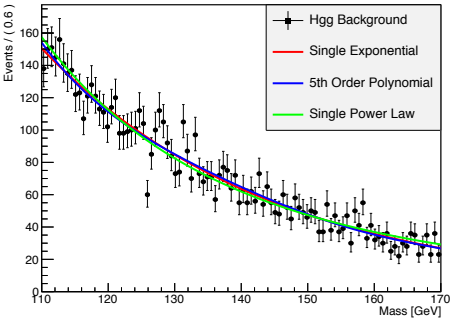
(a) Inclusive Data



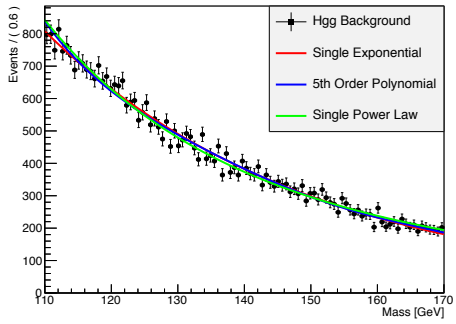
(b) Cat 0: Barrel High R9



(c) Cat 1: Barrel Low R9



(d) Cat 2: Endcaps High R9



(e) Cat 3: Endcaps Low R9

Figure 3.1: Fits of various background model families to 2013 $H \rightarrow \gamma\gamma$ data categories. One sees that there is no ad-hoc method for determining which family of functions is appropriate for modeling the background.

In the previous $H \rightarrow \gamma\gamma$ analysis, the data was divided into four categories based on the location of the photons in the detector and the energetic spread of the reconstructed photons. A single analytical fit function was then chosen for each of these four classes after a study of the potential bias for the estimated background was performed. The potential bias using the chosen function on various truth functions was required to be negligible and then the number of degrees-of-freedom for the fit was increased until the bias became negligible in comparison to the uncertainty of the fits. This criterion for the bias to be negligible was determined such that it should be five times smaller than the statistical uncertainty in the number of fitted events within the mass window corresponding to the full width at half maximum for the corresponding signal model [15].

The results for the background selections were developed in an attempt to account for the systematic uncertainty associated with the choice of the function for the model. The families of models considered for the background analysis were exponentials, power-law functions, polynomials in Bernstein basis and Laurent series [15]. When comparing the models by minimizing twice the negative logarithm of the likelihood, all functions had an added penalty term to account for the number of free parameters in the fitted function such that the likelihood function look as:

$$q = -2ln\mathcal{L} + cN_p \tag{3.1}$$

where N_p is the number of free parameters in the fitting function. Two values of c were used, $c = 1$ and $c = 2$, which are justified by the χ^2 p-value and Akaike Information Criterion respectively. For each class, the functions from each family were then fit to the data and the degrees of freedom were increased until there was no significant improvement to the likelihood. The function with N degrees of freedom was then retained from each family to compare int he study of the expected bias to the signal region [15].

Seeing that the true functional form of the background is unknown, the realistic goal is to find one that minimizes the bias to the fitted signal strength. The model determined

to have the smallest bias was a 5th order polynomial, with a 4th order polynomial being comparable within the fit range of 100 to 160 GeV [31]. A comparison of the sensitivity loss on the exclusion limits was performed where a positive loss meant that the chosen fit model would lend less sensitive results than the truth, giving more conservative results. Negative loss on the other hand would result in overly optimistic signal strength expectations. The models in general provided positive sensitivity loss with the exception of the 4th order polynomial which the truth model was a 5th order polynomial. The chosen model for the previous $H \rightarrow \gamma\gamma$ analysis was concluded to be a 5th order polynomial as it reduced sensitivity loss [31].

The use of a 5th order polynomial for the Standard Model background is convenient as polynomials may fit many shapes with many parameters. However, the limitation on the selection of this model tends around the assumption that this was the only model within the set that could fit the data. As previously mentioned, the restriction to one ”best” model in model selection often results in discarding valuable information, which is not necessarily valid when the theory alone does not provide enough motivation for the structure of the model parameterization. The use of model averaging allows for more information to be conserved and therefore provides less sensitivity to statistical variations between data samples. It is therefore desired o produce quantitative estimates of the empirical support for each model, given the data, and then build a composite model from many physics motivated models to achieve a model that better reflects reality.

3.2 Compact Muon Solenoid Experiment

The data used in this analysis is collected from the Compact Muon Solenoid experiment (CMS) from proton-proton collisions at the Large Hadron Collider (LHC). The LHC is a circular particle accelerator with a circumference of 27 km that is located 50 to 175 meters underground at CERN, on the border between Switzerland and France.The design center of mass energy ($\sqrt{s}$) of the LHC for pp collisions is 14 TeV, however, in 2012 this measured as 8 TeV [31].

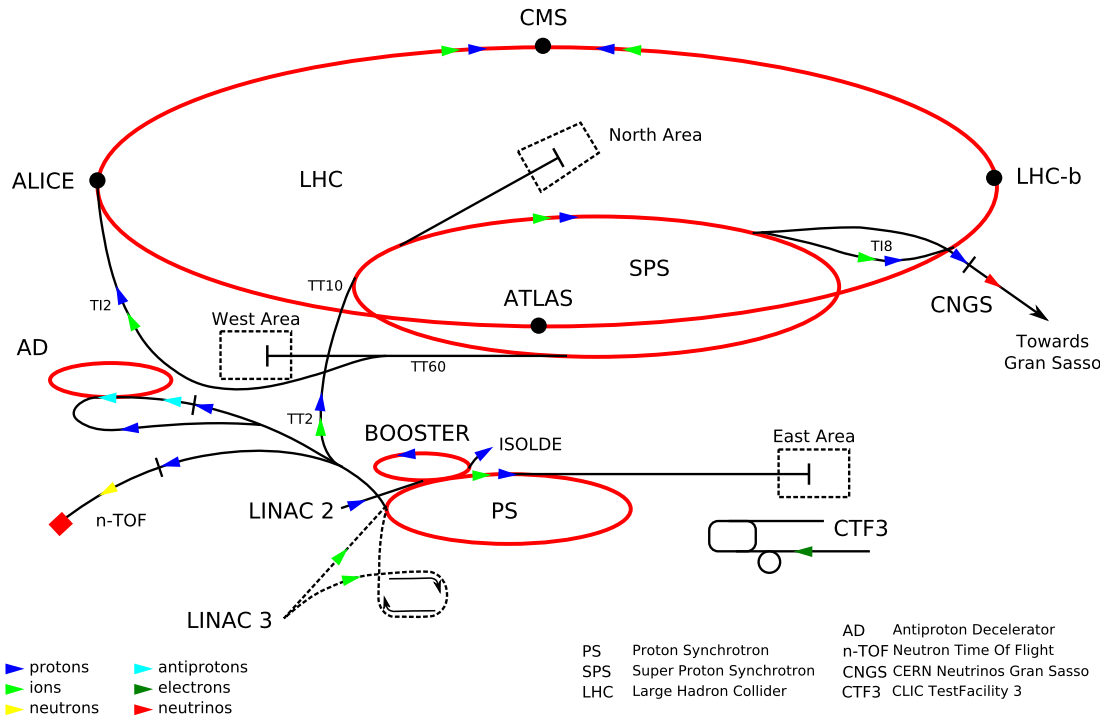


Figure 3.2: A schematic diagram of the LHC accelerator complex. Protons are initially accelerated in the LINAC linear accelerator and then injected into the PS Booster to reach a kinetic energy of 1.4 GeV. They then enter the Proton Synchrotron ring, are accelerated to 25 GeV, and then further accelerated to 450 GeV in the Super Proton Synchrotron (SPS). They are finally accelerated to the maximum energy after being injected into the LHC [31].

The Compact Muon Solenoid experiment is a general purpose detector designed to measure a variety of potential physics studies beyond the Standard Model. It is a nearly hermetic detector, allowing energy balance measurements in the plane transverse to the direction of the beam [27]. The central feature of the detector is a superconducting solenoid of 13 meters in length and 6 meters in diameter, capable of producing an axial field of 3.8 Tesla. The center of the solenoid contains layered detection systems and the steel return yoke outside of the solenoid holds gas chambers for muon identification [8]. Collisions

occur along the central beam line and charged particle trajectories are measured by a silicon pixel and strip tracker with a full azimuthal coverage of ϕ from 0 to 2π within $|\eta| < 2.5$. Here the pseudorapidity η is defined as,

$$\eta = -ln(\frac{\theta}{2}) \tag{3.2}$$

where θ is the polar angle of the trajectory with respect to the positive end of the z-axis [31]. A lead tungstate crystal electromagnetic calorimeter (ECAL) surrounds the tracking volume and covers a barrel region of $|\eta| < 1.48$ and two endcaps that extend up to $|\eta| = 3$. A lead/silicon-strip pre-shower detector is located in from of the ECAL end cap and a steel/quartz-fibre Cherenkov forward calorimeter extends to cover $|\eta| < 5.0$. A brass/scintillator hadronic calorimeter (HCAL) encompasses the ECAL behind the crystal layer [6]. Global event reconstruction under particle flow reconstruction consists in identifying each single particle with an optimized combination of all sub detector information [15]

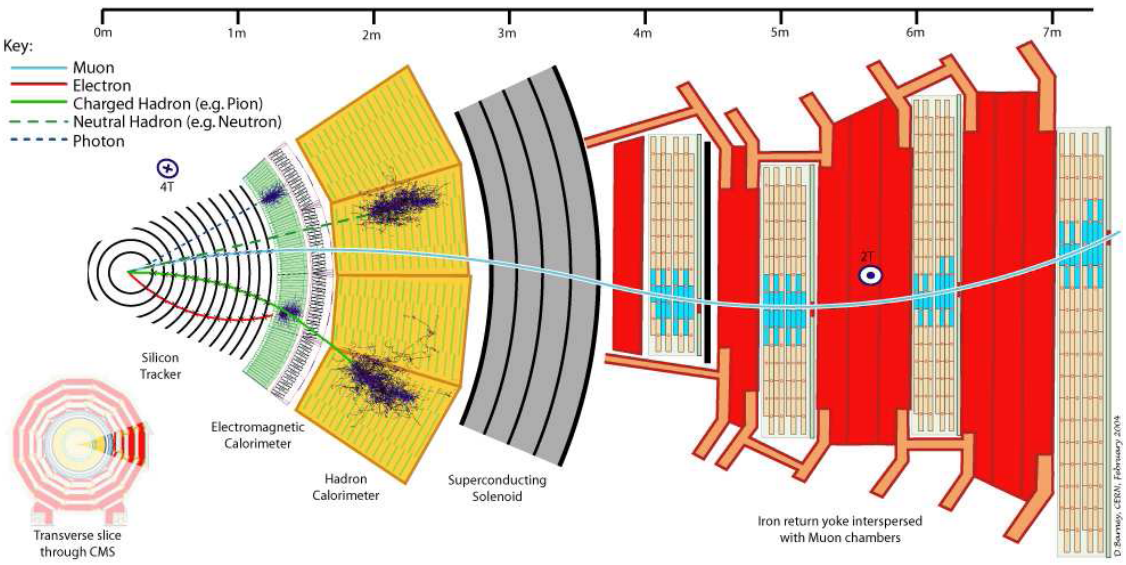


Figure 3.3: A slice diagram depicting the various layers of the CMS detector [31].

The ECAL is optimized for high resolution measurements of electrons and photons. Within the (η, ϕ) plane, the ECAL is composed of 5x5 crystal arrays with fewer crystals in the end caps [29]. An important kinematic variable for reconstructed particles is the extent shower spread R_9 defined as the energy sum of the $3x3$ crystals centered on the most energetic crystal in a supercluster, divided by the total energy of that supercluster [15].

$$R_9 = \sqrt{\Delta\eta^2 + \Delta\phi^2} \tag{3.3}$$

The crystal transparency deteriorates due to radiation during the LHC runs and is monitored continuously and corrected for using a light injected from a laser and LED monitoring system. ECAL calibrations are also performed with photons from $\pi^0 \rightarrow \gamma\gamma$ and $\eta \rightarrow \gamma\gamma$ decays and electrons from $W \rightarrow e\nu$ and $Z \rightarrow e^+e^-$ decays [15]. Comparisons of data and simulation results for $Z \rightarrow e^+e^-$ and $Z \rightarrow \mu^+\mu^-\gamma$ events are used to apply corrections for signal invariant mass shape in $H \rightarrow \gamma\gamma$ and other analyses. The jet energy measurement is also calibrated to correct for detector effects using dijet, $\gamma + jet$ and the Z + jets events [28].

3.3 Diphoton selection for 2013 $H \rightarrow \gamma\gamma$ Analysis

The data used for this analysis consisted of diphoton events front he 2013 CMS $H \rightarrow \gamma\gamma$ analysis (DoublePhoton PD – 22Jan2013 CiC Selection). A detailed description of the processing and photon selection may be found in Reference [15]. In general, the data consists of events determined with diphoton triggers and corresponding to an integrated luminosity of 5.1 fb^{-1} at 7 TeV and 19.6 fb^{-1} at 8 TeV. The diphoton triggers have asymmetric transverse energy, E_T thresholds and complementary photon selections. The selection requires loose calorimetric identification based on the electromagnetic shower shape and loose isolation requirements of the photon candidate. Other selections require that the photon candidate has a high value of the R_9 shower shape variable [15].

Photons with high R_9 values correspond to primarily unconverted photons with better

energy resolution and higher signal-to-background ratios. Photons that are reconstructed within the barrel also have both a better energy resolution and higher signal-to-background ratio than those found in the end caps [15]. The four event classes from the previous $H \rightarrow \gamma\gamma$ analysis are used with an additional class to represent the inclusive results: 0: both photons are in the barrel and have $R_9 > 0.94$

1: both photons are in the barrel and at least one fails the requirement of $R_9 > 0.94$

2: at least one photon is in the end cap and both photons have $R_9 > 0.94$

3: at least one photon is in the end cap and at least one of them fails the requirement $R_9 > 0.94$

4: inclusive category for all selected diphotons

3.4 AIC Results

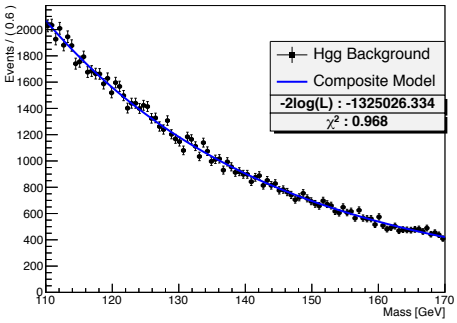
In order to select an appropriate model for the Standard Model background we apply the MMI method to diphoton events (DoublePhoton PD – 22Jan2013 CiC Selection) from the LHC and calculate the AIC weights for each fit model in the model set. The results of this study are displayed in Table 3.2 with the "best" model for each category in bold. One finds that the model that is best support most frequently is the modified exponential, as seen in the inclusive, Low R9 Barrel and High R9 End Cap categories. The modified exponential is also well supported in the other two categories though is preceded by the single power law within the Barrel at High R9 and single exponential in the End Caps at Low R9.

A significant observation to note is that the polynomial fit model was not substantially supported by the data in any of the five categories. This is contrary to the results of the previous $H \rightarrow \gamma\gamma$ background bias study which chose the 5th-Order polynomial as the preferred background model [15, 20]. In Figure 3.4 we display the combined composite model fit to the data for each category along with their likelihoods and the χ^2 value for each fit. We see that each composite model seems to provide an acceptable level of agreement within each category. In Table 3.3 we display the rankings of each model based on twice the negative log likelihoods alone. As expected, the higher-order functions of the

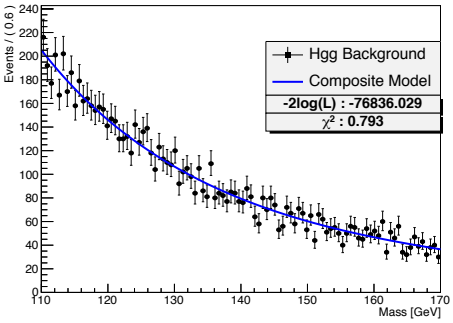
5th order polynomial, double and triple exponential all ranked generally high among the models. However, the discrepancies among the models from the lowest minimized negative log likelihood reveals that the variations among the few models following the "best" model by likelihood standard is relatively small. The composite model is included in this table and ranks consistently around the third or fourth preferred model, with a difference in $-2Log(L) < 2.5$ for each category. At the very least the composite model for each category provides relatively similar results with the other preferred models.

Plausible Model Set	
0. Single Exponential	$c_1 e^{\alpha x}$
1. Double Exponential	$c_1 e^{\alpha x} + c_2 e^{\beta x}$
2. Triple Exponential	$c_1 e^{\alpha x} + c_2 e^{\beta x} + c_3 e^{\gamma x}$
3. Modified Exponential	$c_1 e^{\alpha x + \eta}$
4. 5th-Order Polynomial	$c_5 x^5 + c_4 x^4 + c_3 x^3 + c_2 x^2 + c_1 x + c_0$
5. Single Power Law	$c_1 x^{\alpha}$
6. Double Power	$c_1 x^{\alpha} + c_2 x^{\beta}$

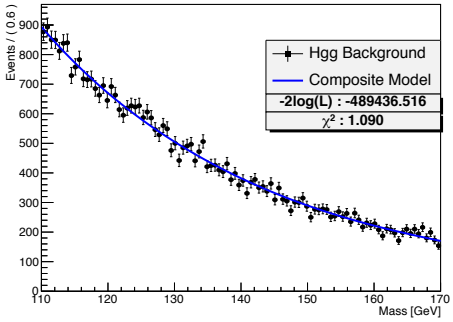
Table 3.1: Model set of plausible functions for MMI process.



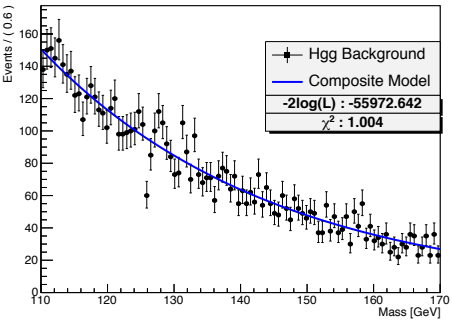
(a) Inclusive Data



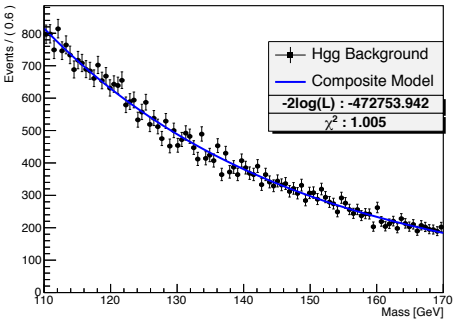
(b) Cat 0: Barrel High R9



(c) Cat 1: Barrel Low R9



(d) Cat 2: Endcaps High R9



(e) Cat 3: Endcaps Low R9

Figure 3.4: Fits of ”composite” background model to 2013 $H \rightarrow \gamma\gamma$ data categories.

	AIC Results for $H \to \gamma\gamma$ Background						
Models	Single Exponential	Double Exponential	Triple Exponential	Modified Exponential	5th-Order Polynomial	Single Power	Double Power
Inclusive							
AIC Δ_i	5.9	1.2	5.2	**0**	13.5	24.5	28.5
AIC w_i	0.03	0.33	0.04	**0.60**	0.001	3. e-6	4. e-7
Cat 0: High R9, Barrel							
AIC Δ_i	6.4	3.9	7.8	2.1	5.9	**0**	4.0
AIC w_i	0.02	0.08	0.01	0.20	0.03	**0.57**	0.08
Cat 1: Low R9, Barrel							
AIC Δ_i	0.5	1.2	5.2	**0**	3.1	12.5	16.5
AIC w_i	0.30	0.21	0.02	**0.38**	0.08	0.001	0.0001
Cat 2: Low R9, End Caps							
AIC Δ_i	**0**	4.0	8.0	1.9	7.4	6.4	10.4
AIC w_i	**0.62**	0.08	0.01	0.24	0.02	0.03	0.003
Cat 3: High R9, End Caps							
AIC Δ_i	0.4	2.0	6.0	**0**	12.9	8.4	12.4
AIC w_i	0.36	0.16	0.02	**0.44**	0.001	0.007	0.001

Table 3.2: AIC results and weights for Standard Model background in 2013 $H \to \gamma\gamma$ data.

The resulting selected ”best” models for each category are in bold.

	Category				
Ranking	Cat 0	Cat 1	Cat 2	Cat 3	Inclusive
1	Poly. (0)	Poly. (0)	Poly. (0)	Trip. Exp. (0)	Trip. Exp. (0)
2	Trip. Exp. (1.91)	Doub. Exp. (2.08)	Mod. Exp. (0.50)	Doub. Exp.(0.02)	Doub. Exp. (0.04)
3	**Composite (1.93)**	Trip. Exp. (2.08)	**Composite (0.57)**	Mod. Exp. (0.02)	**Composite (0.47)**
4	Doub. Exp (1.94)	**Composite (2.24)**	Sing. Exp. (0.61)	**Composite (0.30)**	Mod. Exp. (0.80)
5	Sing. Pow. (2.10)	Mod. Exp. (2.90)	Doub. Exp. (0.61)	Sing. Exp. (2.41)	Poly.(8.27)
6	Doub. Pow. (2.10)	Sing. Pow. (5.40)	Trip. Exp. (0.61)	Poly. (6.91)	Sing. Exp. (8.70)
7	Mod. Exp. (2.22)	Sing. Pow. (17.38)	Sing. Pow. (7.00)	Sing. Pow. (10.46)	Sing. Pow. (27.34)
8	Sing. Exp. (6.48)	Doub. Pow. (17.38)	Doub. Pow. (7.00)	Doub. Pow. (10.46)	Doub. Pow. (27.34)

Table 3.3: Ranking of models for each cut category based on value of $-2Log(L)$ where rank 1 is assigned to the model with the minimized negative log likelihood. The difference between $-2Log(L)$ for each model and the minimized value is shown in parentheses next to the corresponding model.

3.4.1 Stability of Model Selection

Want to examine the stability of the MMI process and the consistency of the composite model shape that is produced. In order to assess the stability of the composite model produced by the AIC analysis to statistical variations between experimental data sets, we perform a "boot-strapping" procedure. In this process the $H \rightarrow \gamma\gamma$ data is randomly re-sampled to create new toy data sets from the original data set with the same total number of

events. The random selection of events and filling of the new data simulates the statistical variations from one experimental run to another. The MMI method is then run over each sample and a composite model is produced for each sample.

We evaluate the systematic uncertainty for this composite production by calculating the number of background events within the signal range from 123 GeV to 127 GeV, normalized by the total number of events for each trial, over 800 toy samples. We then fit this distribution with a gaussian to determine the systematic uncertainty of the composite model production. The measured uncertainty amounts to 0.09 ± 0.0002 which is well below the statistical uncertainty of 0.31, quoted from the mass measurement of the Higgs boson in the previous $H \rightarrow \gamma\gamma$ analysis [15]. Therefore we determine the consistency of the composite model generation from the $H \rightarrow \gamma\gamma$ data to be within acceptable uncertainty.

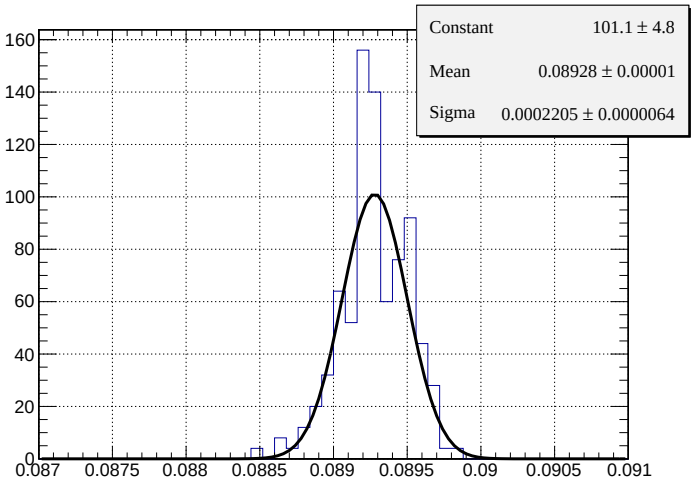


Figure 3.5: Distribution of background events within signal region for composite models derived from 800 toy trials of randomized $H \rightarrow \gamma\gamma$ samples

3.4.2 Bias Analysis

The significance of the model selection for characterizing the Standard Model Background is that an inappropriate function may lead to biased observed signal yields when comparing the $H \rightarrow \gamma\gamma$ data to a background-only hypothesis. Our purpose is to determine the model

for the $H \rightarrow \gamma\gamma$ Standard Model background that minimizes the bias to the signal region. In the previous $H \rightarrow \gamma\gamma$ analysis, the method used for selecting a background model was to compare the maximal potential bias that was introduced by each fit model while characterizing possible truth-models. The bias was determined to be, "the deviation from zero of the median deviation of the fitted number of mean background events between the truth and the fit-model in a mass window corresponding to the full-width-half-maximum of the signal model" [7].

$$Bias(m_H) := median\left(N_{true}^{FWHM} - N_{fit}^{FWHM}\right) \tag{3.4}$$

One then compares this statistical uncertainty, the value to the uncertainty of the fitted number of events within this region, and define the bias to be negligible when,

$$\left| median\left(\frac{N_{true}^{FWHM} - N_{fit}^{FWHM}}{\Delta N_{fit}^{FWHM}}\right)\right| < 0.2 \tag{3.5}$$

In this manner one determines the extent to which each potential fit model alters the expected background event yield within the Higgs signal mass window.

To compare the results from the AIC analysis with this previous study and consider the potential bias of the composite model, we are replicating this bias study by generating toy events for each possible truth model in the model set and fitting them again with all of the potential models, including the composite model. By integrating the truth and fit probability density functions within the mass range of 123 to 127 GeV/c^2, weighted by number of total events in the toy set, we determine the expected number of background events for the truth and fit models within the bias region. The difference between these is then divided by the uncertainty of the fitted number events within the region. The resulting bias comparisons are displayed in Tables 3.4 to 3.11

Truth Type	**Single Exponential**							
Fit Type	Sing. Exp.	Doub. Exp.	Trip. Exp	Mod. Exp.	Poly.	Sing. Pow.	Doub. Pow.	Composite
cat0	0.0033	0.0040	0.0044	0.0041	0.0006	0.0219	0.0219	0.0042
cat1	0.0036	0.0036	0.0036	0.0013	0.0107	0.0214	0.0213	0.0007
cat2	0.012	0.0123	0.0123	0.0233	0.0163	0.0066	0.0066	0.0180
cat3	0.0023	0.0068	0.0054	0.0066	0.0123	0.0176	0.0176	0.0090
cat4	0.0031	0.0031	0.0032	0.0038	0.0004	0.0139	0.0139	0.0033

Table 3.4: Estimated bias of from the various fit models to the signal region of 123 to 127 GeV for the single exponential truth model.

Truth Type	**Double Exponential**							
Fit Type	Sing. Exp.	Doub. Exp.	Trip. Exp	Mod. Exp.	Poly.	Sing. Pow.	Doub. Pow.	Composite
cat0	0.0221	0.0043	0.0043	0.0076	0.0094	0.0042	0.0005	0.0045
cat1	0.0079	0.0018	0.0039	0.0022	0.0041	0.0093	0.0093	0.0035
cat2	0.0084	0.0084	0.0084	0.0022	0.0178	0.0260	0.0261	0.0074
cat3	0.0033	0.0065	0.0065	0.0071	0.0019	0.0183	0.0183	0.0053
cat4	0.0082	0.0029	0.0028	0.0008	0.0036	0.0082	0.0082	0.0016

Table 3.5: Estimated bias of from the various fit models to the signal region of 123 to 127 GeV for the double exponential truth model.

Truth Type	**Triple Exponential**							
Fit Type	Sing. Exp.	Doub. Exp.	Trip. Exp	Mod. Exp.	Poly.	Sing. Pow.	Doub. Pow.	Composite
cat0	0.0140	0.0048	0.0049	0.0009	0.0118	0.0043	0.0053	0.0036
cat1	0.0100	0.0028	0.0029	0.0073	0.0042	0.0077	0.0078	0.0075
cat2	0.0125	0.0126	0.0126	0.0122	0.0035	0.0319	0.0319	0.0125
cat3	0.0053	0.0025	0.0031	0.0024	0.0053	0.0103	0.0103	0.0041
cat4	0.0001	0.0078	0.0027	0.0041	0.0601	0.0163	0.0163	0.0053

Table 3.6: Estimated bias of from the various fit models to the signal region of 123 to 127 GeV for the triple exponential truth model.

Truth Type	**Modified Exponential**							
Fit Type	Sing. Exp.	Doub. Exp.	Trip. Exp.	Mod. Exp.	Poly.	Sing. Pow.	Doub. Pow.	Composite
cat0	0.0268	0.0079	0.0079	0.0122	0.0111	0.0088	0.0022	0.0088
cat1	0.0050	0.0021	0.0018	0.0000	0.0020	0.0119	0.0119	0.0013
cat2	0.0084	0.0086	0.0084	0.0004	0.0044	0.0276	0.0276	0.0048
cat3	0.0126	0.0058	0.0084	0.0064	0.0050	0.0031	0.0030	0.0072
cat4	0.0114	0.0074	0.0073	0.0057	0.0544	0.0052	0.0052	0.0064

Table 3.7: Estimated bias of from the various fit models to the signal region of 123 to 127 GeV for the modified exponential truth model.

Truth Type	**5th Order Polynomial**							
Fit Type	Sing. Exp.	Doub. Exp.	Trip. Exp.	Mod. Exp.	Poly.	Sing. Pow.	Doub. Pow.	Composite
cat0	0.0196	0.0069	0.0069	0.0051	0.0254	0.0018	0.0018	0.0142
cat1	0.0024	0.0002	0.0004	0.0010	0.0120	0.0153	0.0153	0.0118
cat2	0.0008	0.0016	0.0017	0.0017	0.0105	0.0194	0.0194	0.0024
cat3	0.0003	0.0093	0.0027	0.0041	0.0102	0.0149	0.0149	0.0057
cat4	0.0071	0.0022	0.0022	0.0002	0.0001	0.0097	0.0097	0.0007

Table 3.8: Estimated bias of from the various fit models to the signal region of 123 to 127 GeV for the 5th order polynomial truth model.

Truth Type	**Single Power Law**							
Fit Type	Sing. Exp.	Doub. Exp.	Trip. Exp.	Mod. Exp.	Poly.	Sing. Pow.	Doub. Pow.	Composite
cat0	0.0137	0.0044	0.0075	0.0009	0.0121	0.0043	0.0044	0.0047
cat1	0.025 5	0.0114	0.0119	0.0113	0.0056	0.0083	0.0083	0.0091
cat2	0.0105	0.0114	0.0117	0.0038	0.0168	0.0071	0.0074	0.0062
cat3	0.0265	0.0152	0.0159	0.0145	0.0140	0.0120	0.0120	0.0129
cat4	0.0149	0.0059	0.0059	0.0027	0.0048	0.0014	0.0014	0.0035

Table 3.9: Estimated bias of from the various fit models to the signal region of 123 to 127 GeV for the single power law truth model.

Truth Type	**Double Power Law**							
Fit Type	Sing. Exp.	Doub. Exp.	Trip. Exp.	Mod. Exp.	Poly.	Sing. Pow.	Doub. Pow.	Composite
cat0	0.0126	0.0035	0.0039	0.0026	0.0111	0.0058	0.0059	0.0040
cat1	0.0174	0.0040	0.0043	0.0037	0.0070	0.0009	0.0009	0.0040
cat2	0.0129	0.0097	0.0092	0.0088	0.0091	0.0045	0.0046	0.0088
cat3	0.0193	0.0041	0.0015	0.0072	0.0055	0.0047	0.0033	0.0026
cat4	0.0181	0.0052	0.0039	0.0049	0.0099	0.0021	0.0022	0.0014

Table 3.10: Estimated bias of from the various fit models to the signal region of 123 to 127 GeV for the double power law truth model.

Truth Type	**Composite Model**							
Fit Type	Sing. Exp.	Doub. Exp.	Trip. Exp.	Mod. Exp.	Poly.	Sing. Pow.	Doub. Pow.	Composite
cat0	0.1338	0.1265	0.1239	0.1241	0.1166	0.1177	0.1175	0.1229
cat1	0.0726	0.0697	0.0705	0.0660	0.0759	0.0561	0.0560	0.0713
cat2	0.0090	0.0299	0.0298	0.0145	0.0552	0.0278	0.0277	0.0304
cat3	0.0222	0.0205	0.0208	0.0205	0.0167	0.0070	0.0070	0.0215
cat4	0.1914	0.1878	0.1879	0.1696	0.1854	0.1776	0.1776	0.1868

Table 3.11: Estimated bias of from the various fit models to the signal region of 123 to 127 GeV for the $H \rightarrow \gamma\gamma$ truth model.

3.4.3 Bias in Background Shape and Signal Yield

An additional concern in the model selection is how the shape of the background distribution effects the final location of the signal peak. Variations in the slope of the background distribution may lead to biases in the position of the measured signal peak where the mean mass for the Higgs may be shifted to higher or lower values. We examine the variations among the slopes of the background model fit distributions around the Higgs mass to de-

termine if there is any substantial variation introduced by the models. Similar to the bias study, the slope for the fit background distribution is determined for each potential fit model in the set, as well as the composite model, for each potential truth background type. The results for the slopes are displayed in Tabels 3.12 to 3.19.

The composite model, double power law, double and triple exponential models maintain relatively consistent fit slopes among truth types as well as in relation with each other. The single exponential and polynomial fit models, however, show the greatest variation in calculated slope around the Higgs mass, not only from the other fit models but also among their fits to the various truth models. The deviation in the slope for the polynomial is more substantial than would be expected and we cannot explain these outcomes. Similarly the values for the single power law fit functions did not converge, despite the fits converging, which is why they are omitted from the results. Certain categories for the modified exponential fits also failed to converge and are designated by a (*). We therefore proceed with caution in deriving strong conclusions from these results but acknowledge the consistency of the slope resulting from the composite model with the other fit models.

Truth Type	**Single Exponential**						
Fit Type	Sing Exp	Doub Exp	Trip Exp	Mod Exp	Poly	Doub Pow	Composite
cat0	-0.0007871	-0.0006567	-0.0006584	-0.0006306	-0.0176181	-0.0006935	-0.0006566
cat1	-0.0008828	-0.0006200	-0.0006200	-0.0021350	0.0105560	-0.0006572	-0.0006119
cat2	-0.0007696	-0.0006615	-0.0006615	-0.0036701	-0.0164910	-0.0006984	-0.0006373
cat3	-0.0011159	-0.0005591	-0.0005609	-0.0002608	-0.0154558	-0.0005848	-0.0005314
cat4	-0.0009333	-0.0006031	-0.0006032	-0.0010931	-0.0165993	-0.0006406	-0.0006026

Table 3.12: Estimated slope of various fit models at $m_H = 125$ GeV for a single exponential truth model.

Truth Type	**Double Exponential**						
Fit Type	Sing Exp	Doub Exp	Trip Exp	Mod Exp	Poly	Doub Pow	Composite
cat0	-0.0007432	-0.0007344	-0.0007343	*	-0.0184917	-0.0007241	-0.0007176
cat1	-0.0008796	-0.0006107	-0.0006357	-0.0001200	-0.0167992	-0.0006599	-0.0006253
cat2	-0.0009059	-0.0006122	-0.0006122	-0.0026094	-0.0159813	-0.0006491	-0.0006024
cat3	-0.0011501	-0.0005521	-0.0005521	-0.0003344	-0.0158394	-0.0005750	-0.0005452
cat4	-0.0009583	-0.0006157	-0.0006157	-0.0000411	-0.0170691	-0.0006335	-0.0006147

Table 3.13: Estimated slope of various fit models at $m_H = 125$ GeV for a double exponential truth model.

Truth Type	**Triple Exponential**						
Fit Type	Sing Exp	Doub Exp	Trip Exp	Mod Exp	Poly	Doub Pow	Composite
cat0	-0.0007281	-0.0007234	-0.0007234	*	-0.0179595	-0.0007212	-0.0007183
cat1	-0.0008349	-0.0006336	-0.0006334	-0.0003936	0.0066703	-0.0006757	-0.0006380
cat2	-0.0007349	-0.0006754	-0.0006753	-0.0008127	-0.0184208	-0.0007141	-0.0006751
cat3	-0.0010634	-0.0005706	-0.0005713	-0.0004647	-0.0163894	-0.0006000	-0.0005660
cat4	-0.0009810	-0.0005907	-0.0005983	-0.0002325	0.0131302	-0.0006257	-0.0005945

Table 3.14: Estimated slope of various fit models at $m_H = 125$ GeV for a triple exponential truth model.

Truth Type	**Modified Exponential**						
Fit Type	Sing Exp	Doub Exp	Trip Exp	Mod Exp	Poly	Doub Pow	Composite
cat0	-0.0007269	-0.0007446	-0.0007447	*	-0.0181357	-0.0007377	-0.0007198
cat1	-0.0009121	-0.0006221	-0.0006220	-0.0001656	-0.0172350	-0.0006486	-0.0006155
cat2	-0.0007736	-0.0006599	-0.0006599	-0.0029800	-0.0163527	-0.0006971	-0.0006474
cat3	-0.0010368	-0.0005863	-0.0005879	-0.0000857	-0.0164109	-0.0006085	-0.0005837
cat4	-0.0009441	-0.0006156	-0.0006156	-0.0001215	0.0118232	-0.0006379	-0.0006144

Table 3.15: Estimated slope of various fit models at $m_H = 125$ GeV for a modified exponential truth model.

Truth Type	**5th Order Polynomial**						
Fit Type	Sing Exp	Doub Exp	Trip Exp	Mod Exp	Poly	Doub Pow	Composite
cat0	-0.0007736	-0.0007063	-0.0007063	*	-0.0191395	-0.0007014	-0.0006957
cat1	-0.0008453	-0.0006412	-0.0006443	-0.0003173	-0.0176346	-0.0006719	-0.0006110
cat2	-0.0007726	-0.0006624	-0.0006626	-0.0006327	-0.0172462	-0.0006992	-0.0006634
cat3	-0.0011208	-0.0005376	-0.0005582	-0.0002546	-0.0156649	-0.0005834	-0.0005423
cat4	-0.0009250	-0.0006246	-0.0006244	-0.0000558	-0.0169317	-0.0006445	-0.0006194

Table 3.16: Estimated slope of various fit models at $m_H = 125$ GeV for a 5th order polynomial truth model.

Truth Type	**Single Power Law**						
Fit Type	Sing Exp	Doub Exp	Trip Exp	Mod Exp	Poly	Doub Pow	Composite
cat0	-0.0007600	-0.0006322	-0.0006541	*	-0.0175902	-0.0007067	-0.0006462
cat1	-0.0008132	-0.0006880	-0.0006881	*	-0.0174916	-0.0006858	-0.0006845
cat2	-0.0007969	-0.0006830	-0.0006822	*	-0.0172260	-0.0006927	-0.0006874
cat3	-0.0011147	-0.0005855	-0.0005853	*	-0.0164403	-0.0005865	-0.0005848
cat4	-0.0009329	-0.0006380	-0.0006380	*	-0.0170663	-0.0006427	-0.0006363

Table 3.17: Estimated slope of various fit models at $m_H = 125$ GeV for a single power law truth model.

Truth Type	**Double Power Law**						
Fit Type	Sing Exp	Doub Exp	Trip Exp	Mod Exp	Poly	Doub Pow	Composite
cat0	-0.0007298	-0.0007145	-0.0007138	*	-0.0178325	-0.0007185	-0.0007147
cat1	-0.0009039	-0.0006272	-0.0006495	*	-0.0166073	-0.0006528	-0.0006183
cat2	-0.0008567	-0.0006417	-0.0006382	-0.0002461	-0.0171010	-0.0006679	-0.0006389
cat3	-0.0010964	-0.0005824	-0.0005740	*	-0.0163575	-0.0005958	-0.0005888
cat4	-0.0009438	-0.0006387	-0.0006266	*	-0.0168969	-0.0006455	-0.0006391

Table 3.18: Estimated slope of various fit models at $m_H = 125$ GeV for a double power law truth model.

Truth Type	$H \rightarrow \gamma\gamma$ **Composite Model**						
Fit Type	Sing Exp	Doub Exp	Trip Exp	Mod Exp	Poly	Doub Pow	Composite
cat0	-0.0007443	-0.0007012	-0.0006988	-0.0000010	-0.0174774	-0.0007121	-0.0007002
cat1	-0.0008409	-0.0006421	-0.0006432	-0.0003506	0.0079754	-0.0006735	-0.0006384
cat2	-0.0007597	-0.0006484	-0.0006483	-0.0001201	-0.0163126	-0.0007051	-0.0006552
cat3	-0.0010887	-0.0005604	-0.0005606	-0.0006709	-0.0158355	-0.0005923	-0.0005575
cat4	-0.0009096	-0.0006275	-0.0006274	-0.0000923	-0.0169268	-0.0006496	-0.0006243

Table 3.19: Estimated slope of various fit models at $m_H = 125$ GeV for the AIC composite model from Inclusive 2013 data.

3.5 Summary

We have applied the multi-model inference ranking to 2013 $H \rightarrow \gamma\gamma$ data to examine the relative strength of the various potential background models within our model set. We see that the model selected by the previous analysis, a 5th Order Polynomial, is not well supported using Akaike Information Criterion for this data set. In fact, the Modified Exponential function is well supported in all five categories and is selected as the ”best” model in three of the five categories.

We also investigated the multi-model averaging approach by build a composite model using the Akaike weights to weight each model within the model set by their relative support from the data. We see an acceptable agreement in the fit for the composite model to the data and that the potential bias to the size and shape of the background distribution in the signal region is at least not substantially larger than any of the other models in the set. The composite model is also found to be calculated in a robust manner such that the systematic uncertainty for its determination is relatively small at 0.09 in comparison to the statistical uncertainty of the Higgs mass measurement at 0.31.

The discrepancy in the selected model and the lack of support for the currently-used 5th

Order Polynomial model suggest further need to examine the techniques used for modeling the Standard Model background. The use of a composite model produced by multi-model averaging may prove beneficial for future analyses as the resulting function contains more information obtained from the data itself than a single ”best” model and therefore is a more robust approximation of the data.

Chapter 4

Detector Performance

Documentation on the 2012-2013 Photon energy resolution and scale analysis with $Z \rightarrow \mu\mu\gamma$ events and instructions for modifying the PHOSPHOR fit code has been compiled on: *https://twiki.cern.ch/twiki/bin/view/CMS/PhosphorZmmg*

4.1 ECAL Energy Resolution

The Compact Muon Solenoid is a general purpose detector, designed to examine a wide variety of new physics . It includes the world's largest high resolution crystal electromagnetic calorimeter (ECAL) which is comprised of 76,000 lead tungstate (PbWO4) crystals that cover a solid angle of almost 4π [8]. For the experimental environment set at the LHC, the ECAL must not only have a high resolution but also be reliable, fast, have high granularity and be radiation resistant. This motivates the use of lead tungstate, which has a short radiation length (0.89 cm) and a small Molliere radius (2.19 cm), that permits a design with both high granularity and compactness [8]. The ECAL is a hermetic, homogeneous calorimeter with a two-fold purpose of measuring the time, location and amount of energy deposits related to electromagnetically interacting particles, and fully containing the electrons and photons within its volume. This allows for their identification, as well as precise estimates of their momenta and the ability to trigger on highly energetic electrons and photons and large missing transverse energy [29].

The optimization for the ECAL design was based on the potential to discover a Higgs bo-

son in the mass region below 130 GeV through the decay scheme of $H \rightarrow \gamma\gamma$ [8]. Due to the small expected intrinsic decay width of the decay channel, the measured width, which is crucial for the significance of signal over the expected background, is entirely dependent on the ECAL energy resolution. The goal for the energy resolution of the detector is better than 0.5% for photons above 100GeV [8].

The measurement of the ECAL resolution is produced by fitting the line shape of the Z boson in its decay to two electrons, $Z \rightarrow ee$. Due to its large mass, the lifetime of the Z boson is very short and the particle decays again via the same coupling that leads to its production [8]. The nature of the final state is random among fermion flavors and generations but the probabilities of the different fermion flavors are fixed by theory and have been experimentally confirmed to high precision [6]. Due to the design of the ECAL, the decays to electrons and muons are particularly easy to identify in the detector and separate from other processes with similar signatures [6].

To fit the Z peak and determine the energy resolution of the ECAL detector we use a convolution of Crystal Ball (CB) and Breit Wigner (BW) functions. The width of the Z mass peak is a combination of the contributions from the ECAL resolution as well as the Z boson's physical mass width. The Breit Wigner is fixed to a mass of 91.2 GeV/c^2 and width of 2.5 GeV/c^2 to represent the physical process of the Z $\rightarrow$ ee decay. The Crystal Ball parameters are left to float to represent the detector effects. The Δm_{CB} represents the difference in the mass peak for the fit, σ_{CB} estimates the absolute mass resolution and α_{CB} is the exponential parameter for the Crystal Ball function.

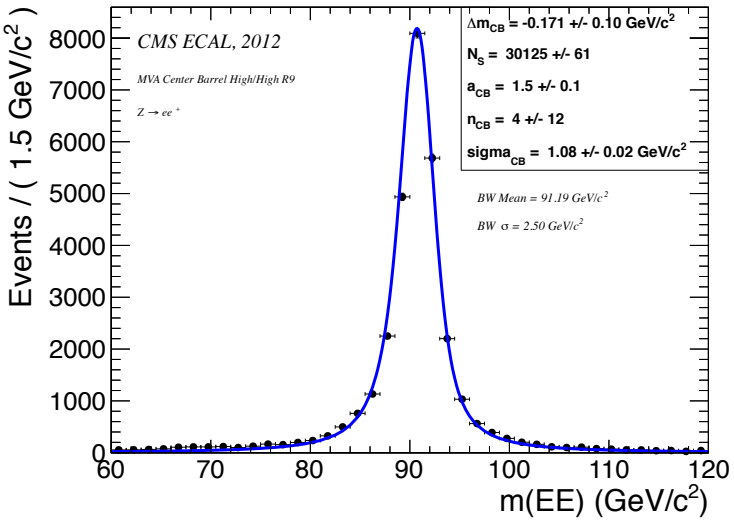


Figure 4.1: Fit of invariant mass peak for $Z \rightarrow ee$ decay for extraction of mass resolution, σ_{CB}, and scale, Δm_{CB} [16].

The energy resolution of the detector is related to the mass resolution measured from the $Z \rightarrow ee$ peak by assuming a Gaussian resolution for the core of the distribution and equal sharing of the energy. Then the energy resolution can be determined as:
[6]

$$\frac{\sigma_E}{E} = \sqrt{2}\left(\frac{\sigma_{CB}}{M_Z}\right) \tag{4.1}$$

The mass distributions produced by Data have less pronounced peaks with greater spread in comparison to peaks produced by Monte Carlo simulation. We examine the strength of our simulation by determining the extent to which the Monte Carlo needs to be smeared to match the Data . Using the measured resolution, σ_{CB}, from Data and Monte Carlo the smearing numbers for the Monte Carlo are calculated as:

$$smear = \sqrt{2\left(\left(\frac{\sigma_{CB}data}{peak_{CB}}\right)^2 - \left(\frac{\sigma_{CB}MC}{peak_{CB}}\right)^2\right)} \tag{4.2}$$

The energy resolution and smearing numbers are used to correct the reconstructed energy

for the $H \rightarrow \gamma\gamma$ simulation to match performance measured in data [6].

The ECAL is comprised of a barrel and two end caps, as well as two pre shower detectors. Due to the presence of the tracker material, as well as the strong magnetic field, electrons and positrons from pp collisions radiate bremsstrahlung photons before being detected by the ECAL. On the other hand, photons tend to convert into electron-positron pairs. The ECAL is compartmentalized into superclusters consisting of multiple contiguous clusters of crystal which correspond to individual particles entering the ECAL [8]. Due to the strong magnetic field and presence of the tracker material, electrons and positrons from pp collisions tend to radiate bremsstrahlung photons before they reach the ECAL. Photons, on the other hand, tend to convert into electron-positron pairs. Therefore, superclusters consist of one or more spatially contiguous clusters of ECAL crystals. These clusters correspond to individual particles entering the ECAL [8]. Photons with high R_9 value correspond to predominantly unconverted particles and have both a better energy resolution and a higher signal-to-background ratio. Similarly, photons that are detected within the barrel tend to have both a better energy resolution and higher signal-to-background ratio than end cap photons due to several factors, including the amount of material in front of the calorimeter and the less precise single channel calibration [15]. For these reasons, resolution and scale measurements are separated into categories dependent on the detector location as well as the corresponding R_9 values of the detected particles.

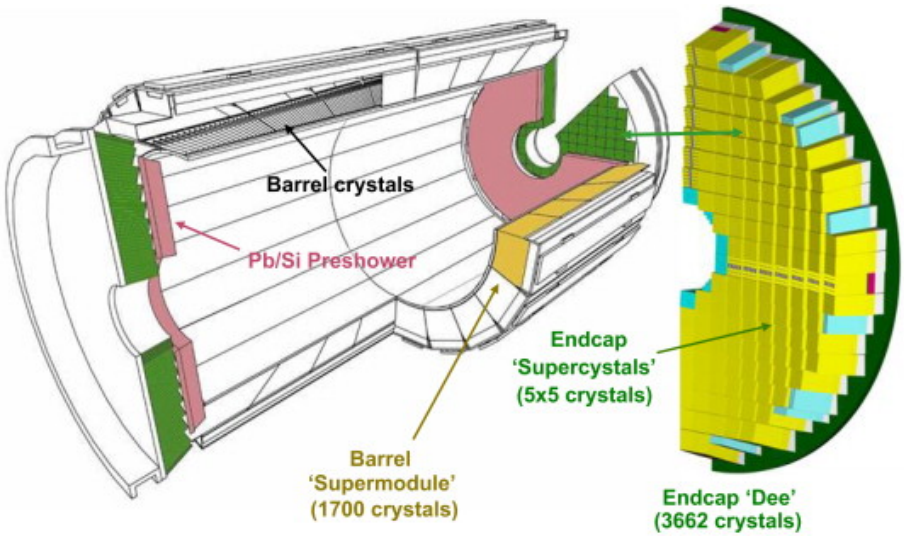


Figure 4.2: Diagram of CMS electromagnetic calorimeter (ECAL) exhibiting compartmentalization of the Barrel and Endcaps [13].

ECAL Detector Regions	
EB: Barrel ($\|\eta\| < 1.44$)	EE: End Caps ($1.5 < \|\eta\| < 2.5$)
R9 Classification	
High: $R9 > 0.94$	Low: $R9 < 0.94$

4.2 PHOSPHOR Fit

The Photon energy Scale and Photon energy Resolution (PHOSPHOR) Fit method is a measurement scheme developed by Jan Veverka, a recent graduate of the Caltech High Energy Physics group, that extracts the photon energy scale and resolution in-situ measurement using radiative muonic Z boson decays [29]. The measurement is based on an unbinned maximum likelihood fit of a model for the invariant mass of $Z \rightarrow \mu\mu\gamma$. Similar to the $Z \rightarrow ee$ process, the width and peak position of the Z boson can be utilized to measure the photon energy resolution and determine the photon energy scale from data. The $Z \rightarrow \mu\mu\gamma$ process is a desirable measurement channel for CMS as muons are very well reconstructed in the detector and the decay process is a relatively clean source of high energy photons in op-collisions [29].

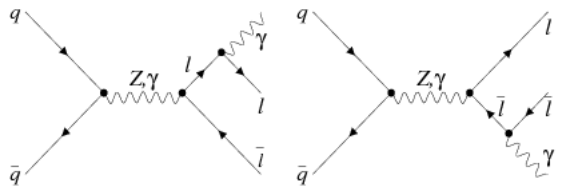


Figure 4.3: Feynman Diagram for Z boson radiating into two leptons and a photon.

Due to their relatively low mass, electrons tend to radiate bremsstrahlung photons in the tracker material at a significantly higher rate than other charged particles [18]. The in-situ measurement of electrons considers this electron showering and photon conversions within the tracker material which lead to a spread of electromagnetic showers due to the bending of electrons and positrons by the magnetic field [16]. Due to the spreading energy over a larger angular range, the likelihood that part of the shower goes into the inter-crystal, inter-module and inter-supermodule cracks is increased, resulting in more energy being lost on average. Fluctuations in these losses contribute to the detector resolution and the electronic noise is increased for showering electrons relative to non-showering electrons due to the energy spread across a larger array of crystals [16].

The reconstruction of muons is optimized to a very high efficiency for the CMS detector. Muons are reconstructed as tracks in the inner silicon tracker and are matched to signals in the outer muon chambers. The signals resulting in the outer systems can either be isolated hits or reconstructed tracks [16]. Signals that do not originate from primary isolated muons, such as hadronic punch-through, decay in flight, accidental track-to-segment matches, or cosmic muons, may be minimized through isolations and kinematic cuts [16].

The ultimate goal of the PHOSPHOR fit is the characterize the in-situ ECAL response for photons by the photon energy scale, s, and the energy resolution, r. These parameters are components extracted from the signal modeled are allowed to float in the fit, together with the signal purity, fS, and two other background components. The ECAL response is

better realized as a probability density function with an infinite number of degrees of freedom considering full generality [29]. To deal with this, we avoid maintaining generality and restrict the results to a subset of considered responses. The simulation is the starting point for the response-density phase space, with the location and scale of its kernel density estimator allowed to vary. The mode and effective sigma of the location and scale parameters are then selected and interpreted as the photon energy scale and resolution [29]. In these terms we consider the photon energy resolution as the effective sigma, "the half of the minimum interval of the E_γ response distribution that contains 68.3% of the area under the curve, the same as the area under the Gaussian distribution around its mean" [29].

In consideration of the determination of signal and background events, the signal is defined by matching generator level topology for any event with a photon radiated off of one of the legs of the Z by the electromagnetic component of the parton shower with enough p_T to pass the predefined selection criteria. Any other event the passes the preselection criteria is considered background [29]. The various background contributions for the decay channel come from QCD fakes, W+jet production, $t\bar{t}$ production and other non-resonant processes. Another resonant background from Z+jets is considered separately as the additional peak biases the measurement of the photon energy scale and resolution [29].

The input Monte Carlo is divided to be used to both be fitted and create the fitting model as it is more robust and precise to use the sample of the simulation in the model creation. The model is smoothed with a kernel density estimator in order to obtain a non-parametric estimate of the photon energy response density [29]. From this the MC truth represents the benchmark estimate of the photon energy scale and resolution in the simulation and the comparison of the MC truth and MC fit corresponds to a closure test for the fit method. The final data fit is then used as an estimate of the photon energy scale and resolution in data [29].These measurements are used to apply corrections to the photon momenta in both data and Monte Carlo and allow the simulation performance to match that of the data.

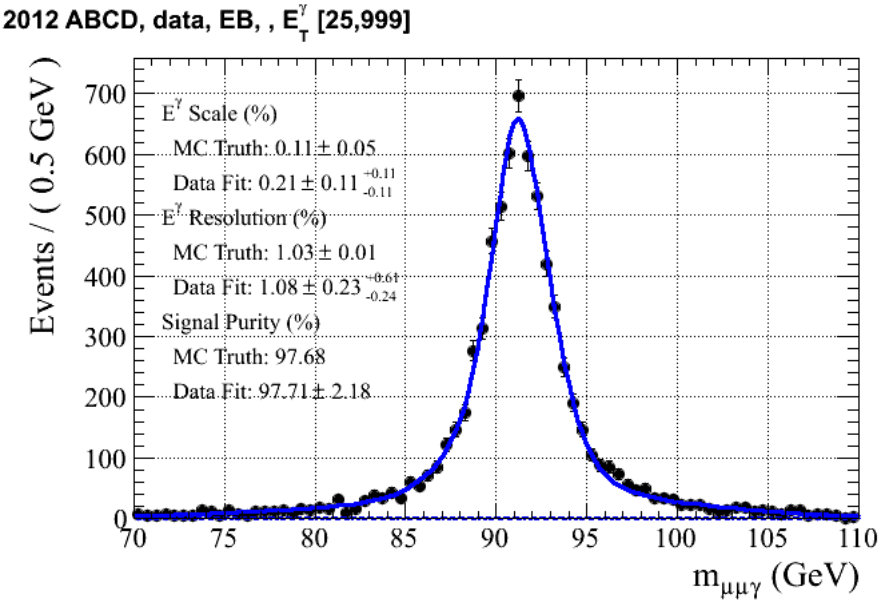


Figure 4.4: An example of the reconstructed $\mu\mu\gamma$ peak and fit with extracted resolution and scale values from PHOSPHOR Fit.

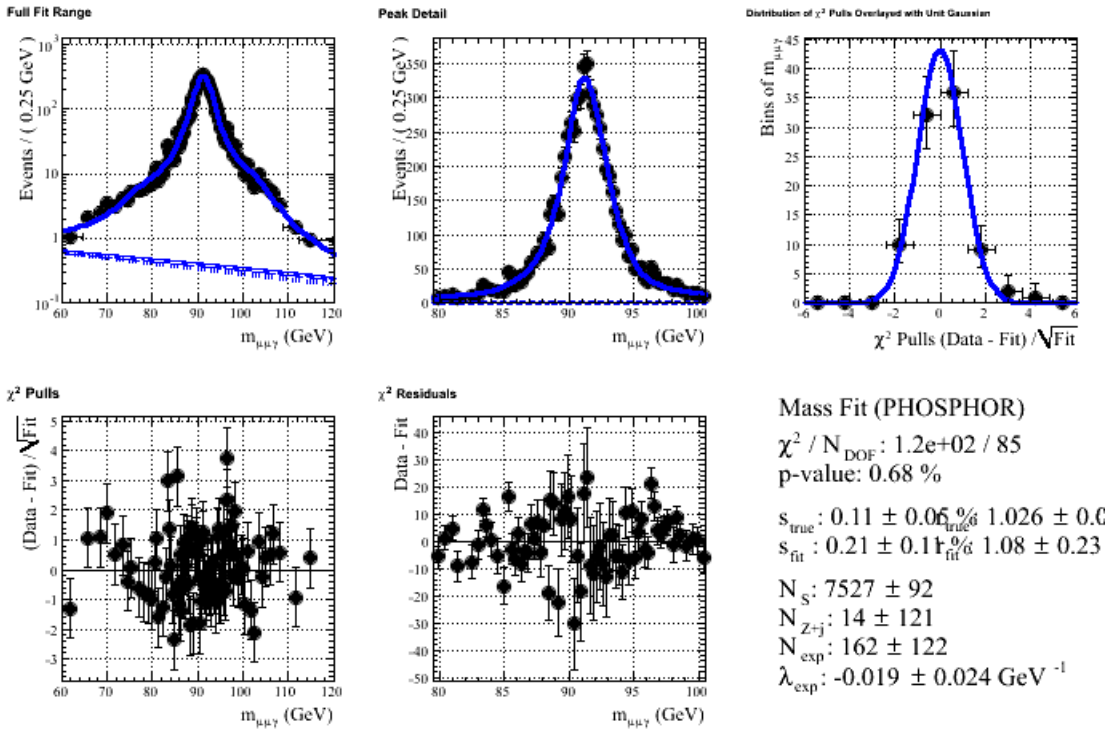


Figure 4.5: An example of modeling of E_γ response peak by Gaussian (Top Left). The data and the fit function are displayed on a logarithmic y-axis scale (Top Middle). The data corresponds to photons in the ECAL barrel with $E_T > 25$ and $R_9 > 0.94$.

4.3 2013 Resolution Measurements

This section contains the PHOSPHOR fit for photon scale and resolution measurements using $Z \rightarrow \mu\mu\gamma$ events from 2013 rereco data at the LHC. The Data and Monte Carlo samples used in this analysis are displayed in Table 4.1. The data and MC samples are processed under the Caltech Vecbos regression and then run through an interface to translate the ntuples to a format compatible with the PHOSPHOR machinery as well as to implement 2012 Rochester muon momentum corrections and reweigh the Z mass distributions for the Monte Carlo with respect to the data pile up distribution.

As the detector resolution varies based on the detector region, R_9 classification, degree

2013 Caltech Vecbos Data
Hgg_53X/Caltech/Reduced/DoubleMuParked_Run2012B_22Jan2013 Hgg_53X/Caltech/Reduced/DoubleMuParked_Run2012C_22Jan2013 Hgg_53X/Caltech/Reduced/DoubleMuParked_Run2012D_22Jan2013 Hgg_53X/Caltech/Reduced/DoubleMu_Run2012A_22Jan2013
Monte Carlo
Hgg2012/MC/DYToMuMu_M_20_FSRFilter_8_TuneZ2star_8TeV_pythia6_GEN_SIM_v2/veverka-Summer12-PU_S7_START52_V9_sttep3_RAW2DIGI_L1Reco_RECO_PU_v2-90a3c643a4855c1621ba3bfcbef2e742_VecBosV20-5_2_X/

Table 4.1: Data and Monte Carlo used in the study of $Z \rightarrow \mu\mu\gamma$ events

of pile up, and the transverse momentum, p_T, of the detected particles, we measure the detector performance in distinct categories, displayed in Table 4.3.

ECAL Detector Regions	
EB: Barrel ($\|\eta\| < 1.44$)	EE: End Caps ($1.5 < \|\eta\| < 2.5$)
R9 Classification	
High: $R9 > 0.94$	Low: $R9 < 0.94$
Pile Up	
High: $NV > 18$	Low $NV < 18$

Monte Carlo Resolution ($\frac{GeV}{c^2}$)		
Pt Bin ($\frac{GeV}{c}$)	High R9	Low R9
Barrel		
10 to 12	1.93	6.16
12 to 15	1.79	5.9
15 to 20	1.51	4.7
20 to 25	1.25	3.34
25 to 999	1.03	2.29
End Caps		
10 to 12	3.22	7.88
12 to 15	2.8	6.8
15 to 20	2.33	5.73
20 to 25	1.99	4.57
25 to 999	1.83	3.91

Data Resolution ($\frac{GeV}{c^2}$)		
Pt Bin ($\frac{GeV}{c}$)	High R9	Low R9
Barrel		
10 to 12	3.37	9.48
12 to 15	3.22	8.32
15 to 20	1.5	6.4
20 to 25	1.28	4.47
25 to 999	1.42	3.14
End Caps		
10 to 12	6.5	12.01
12 to 15	6.83	11.08
15 to 20	4.93	9.38
20 to 25	4.51	8.23
25 to 999	4.35	7.61

Monte Carlo Scale ($\frac{GeV}{c^2}$)		
Pt Bin ($\frac{GeV}{c}$)	High R9	Low R9
Barrel		
10 to 12	0.46	1.31
12 to 15	0.29	0.81
15 to 20	0.28	0.64
20 to 25	0.12	0.62
25 to 999	0.11	0.45
End Caps		
10 to 12	1.14	3.32
12 to 15	1.14	1.63
15 to 20	0.97	2.15
20 to 25	0.47	1.8
25 to 999	0.46	0.94

Data Scale ($\frac{GeV}{c^2}$)		
Pt Bin ($\frac{GeV}{c}$)	High R9	Low R9
Barrel		
10 to 12	-0.18	1.4
12 to 15	-1.37	0.28
15 to 20	-0.38	-0.12
20 to 25	-0.55	0.29
25 to 999	0.19	0.4
End Caps		
10 to 12	-0.97	5.49
12 to 15	0.52	1.62
15 to 20	2.49	3.33
20 to 25	0.92	2.74
25 to 999	1.71	3.01

In previous studies, the impact of pile up on the ECAL energy resolution has been found to be mild and overshadowed by the dependency on the gaussian fraction for any given fit [6]. In order to consider the effects of pile up, we examine the difference in energy resolution and scale measurements with relation to the number of vertices in the measured event. For both data and MC one sees that the resolution deteriorates for higher pile up conditions, the number of measured vertices begin greater than 18. Similarly, the scale measurements reveal a greater deviation in the fit mass for increased pile up. Pile up remains a key topic in consideration as future detector conditions move to higher luminosity environments where increased pile up may cause severe degradation of the detector performance.

Monte Carlo Resolution ($\frac{GeV}{c^2}$) **: Inclusive**			**Data Resolution** ($\frac{GeV}{c^2}$)**: Inclusive**		
Pt Bin ($\frac{GeV}{c}$)	High Pile Up	Low Pile Up	Pt Bin ($\frac{GeV}{c}$)	High Pile Up	Low Pile Up
Barrel			Barrel		
10 to 12	4.82	4.06	10 to 12	8.12	5.97
12 to 15	4.6	3.78	12 to 15	6.2	5.57
15 to 20	3.49	3.04	15 to 20	5.41	3.83
20 to 25	2.45	2.19	20 to 25	3.5	2.56
25 to 999	1.66	1.54	25 to 999	1.78	2.04
Endcaps			End Caps		
10 to 12	7.56	6.0	10 to 12	11.19	9.66
12 to 15	5.83	5.17	12 to 15	8.28	8.69
15 to 20	4.67	4.26	15 to 20	6.86	7.4
20 to 25	3.65	3.2	20 to 25	6.61	5.66
25 to 999	3.0	2.64	25 to 999	5.51	5.58

Monte Carlo Scale ($\frac{GeV}{c^2}$) **: Inclusive**			**Data Scale**($\frac{GeV}{c^2}$) **: Inclusive**		
Pt Bin ($\frac{GeV}{c}$)	High Pile Up	Low Pile Up	Pt Bin ($\frac{GeV}{c}$)	High Pile Up	Low Pile Up
Barrel			Barrel		
10 to 12	1.0	0.8	10 to 12	0.13	0.35
12 to 15	0.8	0.46	12 to 15	-1.12	-0.7
15 to 20	0.63	0.46	15 to 20	-1.58	-0.19
20 to 25	0.45	0.29	20 to 25	0.29	-0.39
25 to 999	0.28	0.28	25 to 999	0.67	0.22
End Caps			End Caps		
10 to 12	3.5	1.82	10 to 12	2.42	1.78
12 to 15	1.47	1.3	12 to 15	1.95	0.53
15 to 20	0.66	1.14	15 to 20	1.66	1.8
20 to 25	0.64	0.97	20 to 25	2.54	1.14
25 to 999	1.11	0.62	25 to 999	2.61	2.09

4.4 Resolution Contributions

The ECAL resolution is sensitive to various factors along the chain of the detection and readout system. Electromagnetic showers of photons and electrons develop within the crystal, stopping all energetic particles in crystal volume. The resulting scintillation light is collected by photodetectors attached to back of the crystals where the signal is amplified by avalanche photodiodes [8]. The crystal light output is dependent on varying temperature and crystal color change. A laser monitoring system is used to continuously track and correct these changes in order to maintain the detector energy resolution [8]. The energy dependence of the energy resolution itself may be parameterized as the quadratic sum of three terms [8],

$$\frac{\sigma_E}{E_0} = \frac{A}{\sqrt{E_0}} \oplus \frac{B}{E_0} \oplus C \tag{4.3}$$

The first term refers the the stochastic contribution due to fluctuations based on the physical development of the signal shower. For homogeneous calorimeters, intrinsic fluctuations are small as the energy is deposited in the active volume of the detector by a monochromatic incident beam of particles which is constant from event to event [13].

The second ”noise” term contains contributions not only from the noise of the readout electronics but also from pile up [8, 29]. The significance of the noise term depends on the detector technique for the readout circuit. For scintillator-based sampling or homogeneous calorimeters, the level of noise can be limited if the electronic chain begins with a photo-sensitive device which provides a high-gain multiplication of the original signal with no noise [13]. The noise term is larger in detectors in which the signal is collected as charge because the first element in in the readout chain is a preamplifier [29]. The noise contribution to the energy resolution of the detector increases with decreasing energy of the incident particles and at energies of just a few GeV or lower, may become the dominant term [13].

The third ”constant term” arises from several effects that do not depend on the energy of the particle such as, but not limited to, calibration fluctuations, non-uniformity of the longitudinal light collection, shower leakage from the back of the crystal, geometric effects,

Term	Contribution	Barrel ($\eta = 0$)	Endcaps ($\|\eta\| = 2$)	Unit
Noise at low luminosity	Electronics noise (at start up), Leakage current noise, Pileup noise	16	77	% GeV
Noise at high luminosity	Electronics noise (at start up), Leakage current noise, Pileup noise	21	92	% GeV
Stochastic	Containment, Photo-statistics, Preshower sampling	2.7	5.7	% $\sqrt{GeV}$
Constant	Inter-calibration, Longitudinal non-uniformity, Others	0.55	0.55	%

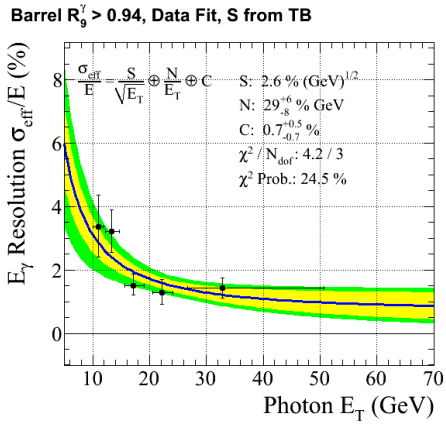
Table 4.2: Expected contributions to the CMS ECAL energy resolution [29].

and imperfections or damage in calorimeter structure and readout systems [8, 13, 29]. The primary goal for the design of a calorimeter is to determine an optimal compromise among the contributions from these terms. For a high resolution electromagnetic calorimeter, the energy resolution at high energies is often dominated by the constant term [8].

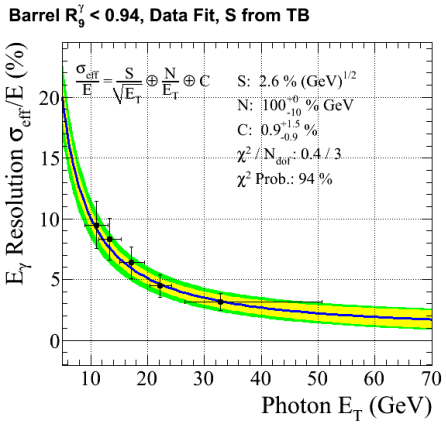
By fitting the photon resolution from the $Z \rightarrow \mu\mu\gamma$ events as a function of the detected photon energy, E_T, we can extract the stochastic, noise and constant terms to the detector energy resolution. In Figure 4.8, one can see how the resolution varies under the effects of pile up. Note that the stochastic and constant terms remain consistent for both barrel and end cap categories but the noise contribution is clearly more pronounced in the high pile up category, rising from 64% to 77% GeV in the barrel and 95% to 99.2% GeV in the End Caps.

4.5 Future Improvements

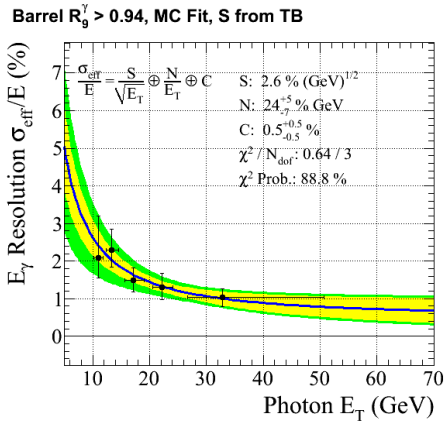
Current limitations to the PHOSPHOR Fit method using $Z \rightarrow \mu\mu\gamma$ events is primarily due to certain fit categories failing due to limited statistics in either data or Monte Carlo. Seeing that the Monte Carlo sample is divided in order to both create the fit models and be fitted itself, large samples refined for $Z \rightarrow \mu\mu\gamma$ events are ideal to ensure enough statistics



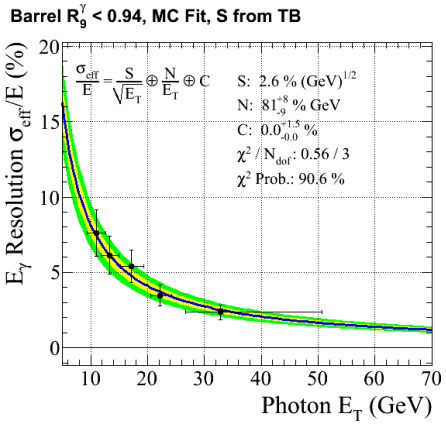
(a) Data High R9: Barrel



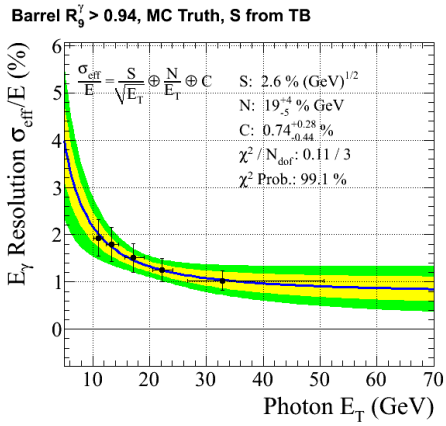
(b) Data Low R9: Barrel



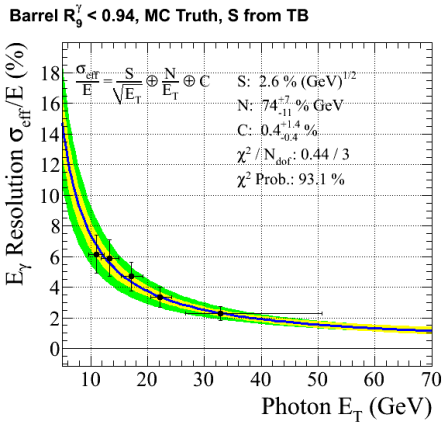
(c) MC High R9: Barrel



(d) MC Low R9: Barrel

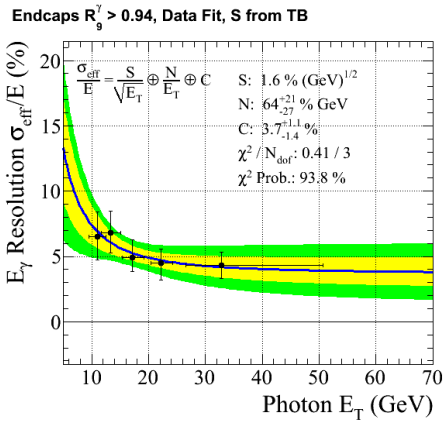


(e) MC Truth High R9: Barrel

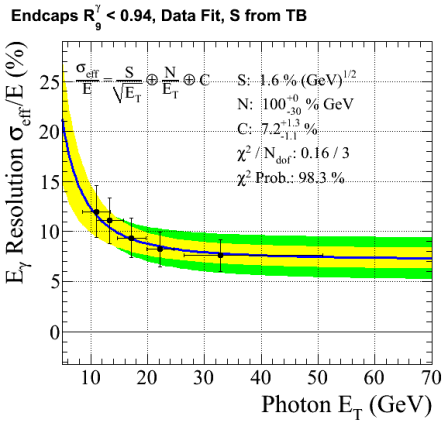


(f) MC Truth Low R9: Barrel

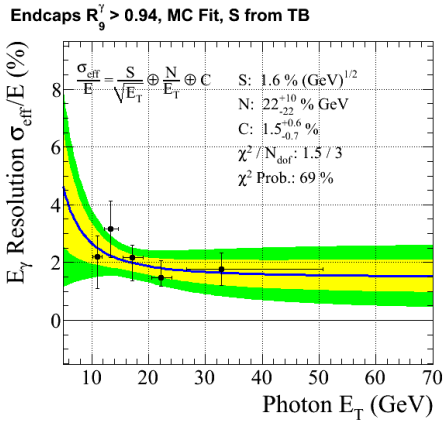
Figure 4.6: Barrel resolution contributions with Data fits on top, Monte Carlo fits in the mid row and Monte Carlo Truth on bottom.



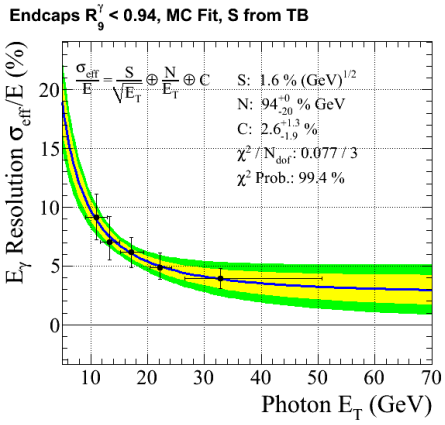
(a) Data High R9: End Caps



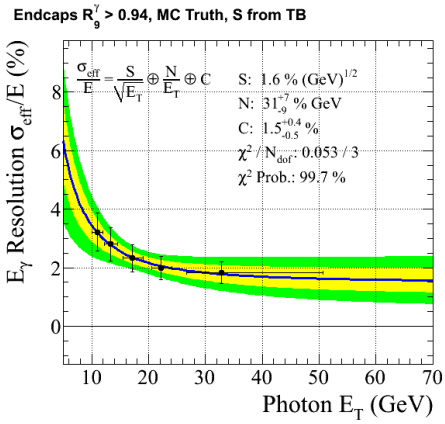
(b) Data Low R9: End Caps



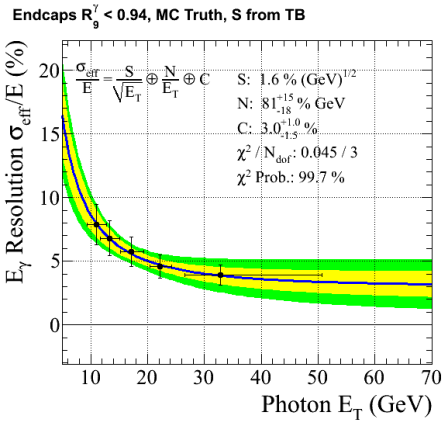
(c) MC High R9: End Caps



(d) MC Low R9: End Caps

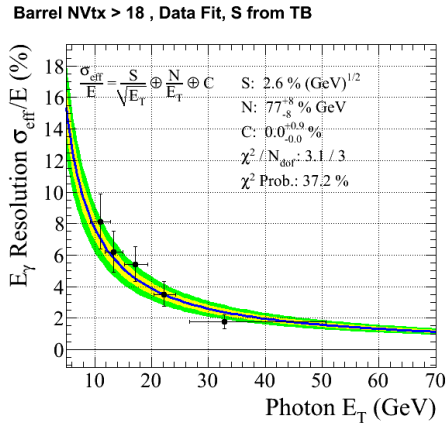


(e) MC Truth High R9: End Caps

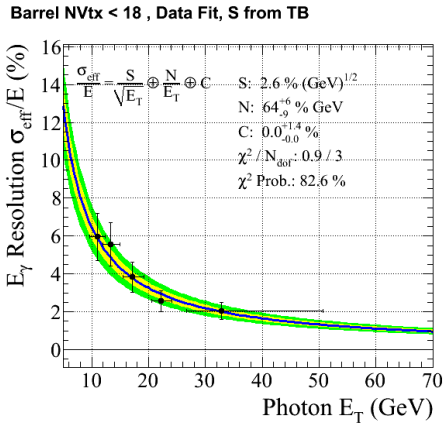


(f) MC Truth Low R9: End Caps

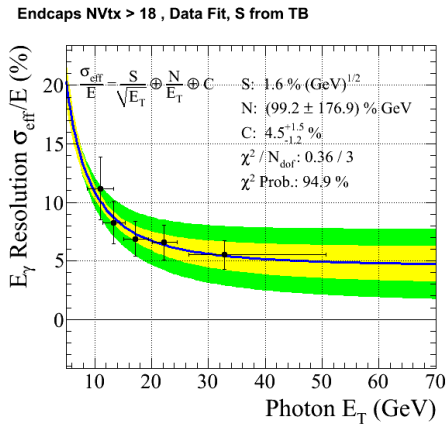
Figure 4.7: End Caps resolution contributions with Data fits on top, Monte Carlo fits in the mid row and Monte Carlo Truth on bottom.



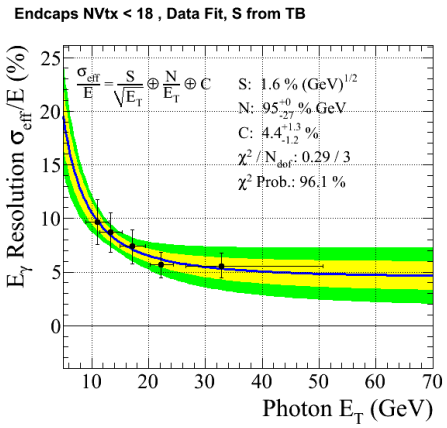
(a) Data High Pile Up: Barrel



(b) Data Low Pile Up: Barrel



(c) Data High Pile Up: End Caps



(d) Data Low Pile Up: End Caps

Figure 4.8: Resolution Contributions in bins of pile up and detector location

for the fits to converge. In addition to this, a limitation in the fit method itself has been identified related to the creation of the background models.

The PHOSPHOR Fit method creates three separate models for the signal, non-resonant background and resonant background based around the peak from Z+jets. The resonant background model is meant to specifically fit the Z+jets contribution, which is normally dominant over all other resonant backgrounds, and does not factor in any other contributions. This however may not be properly motivated as additional background from the combination of muons and a photon produced from bremsstrahlung of a muon smears the shape of the resonant background peak and shifts the mean to a lower mass. Usually this muon bremsstrahlung background is overwhelmed by Z+jets but in certain categories is has been noted to be substantial enough to cause the fit to fail, particularly in categories that are already sensitive.

Another issue arises in the background modeling as the Monte Carlo does not consistently model the non-resonant background properly. During the minimization process for the fit it appears that the model discounts the exponential feature, though the data distribution clearly supports the existence of this feature. In the model construction, it is assumed that this background can be appropriately modeled by an exponential function which may prove to be invalid. The optimal approach for improving the fit method and increasing the efficiency for the fit convergence is to determine an appropriate modeling of the background contributions and produce a combined background model based on the independent fits weighted by the relative strength of the backgrounds. In consideration of this, an appropriate method for future modeling may be the implementation of a multi-model averaging scheme, considering the scenarios of either combining independently modeled resonant and non-resonant backgrounds or attempting to build a composite model of both features. Using data-driven model selection for these backgrounds may assist not only in improving the fit convergence efficiency but also the overall performance measurement by increasing the signal to background discrimination.

Proposed alternatives to improving the fit convergence are derived from cut-based motives including enhancing the Z+jets background by releasing cuts on the di-muon and reconstructed $\mu\mu\gamma$ masses or implementing even stricter cuts to reduce the muon bremsstrahlung background and non-resonant contribution. The outcome of the first alternative has been attempted and resulted in unfavorable results with further degradation in efficiency for the fit convergence. Unfortunately for the latter alternative, applying additional kinematic restrictions poses additional challenges in skewing the shapes of the fits and also further limiting the statistics for the fit.

Chapter 5

Higgs Self-Coupling Measurement

This chapter summarizes a recent analysis of the potential for future measurements of the Higgs self-couplings at an upgraded CMS detector along with proposed upgrades to the High-Luminosity Large Hadron Collider in the near future. It concludes with a discussion of the expected errors in future measurements as well as focused areas of improvement for detector designs and particle identification algorithms. More information on the analysis can be found in references [22] and [1].

5.1 Introduction

The recent discovery of a Higgs-like boson with a mass around 125 GeV by the ATLAS and CMS collaborations at the LHC provides unequivocal evidence of some mechanism of spontaneous electroweak symmetry breaking and of the generation of the masses of fundamental particles [1]. This discovery concludes an era of Higgs discovery and begins a new focused on precision measurements of the properties of this particle. Before the identification of the Standard Model Higgs boson can be completely establish we need more precise measurements of its coupling and evidence that its spin is zero [3]. One of the most important subsequent measurements is to reconstruct the scalar potential of the Higgs doublet field that is responsible for electroweak symmetry breaking through measurements of the Higgs self-couplings [4].

$$V(\eta_H) = \frac{1}{2}m_H^2\eta_H^2 + \lambda\nu\eta_H^3 + \frac{1}{4}\tilde{\lambda}\eta_H^4 \tag{5.1}$$

Where λ refers to the trilinear self-coupling, $\tilde{\lambda}$ the quartic self-couplings, $\nu = (\sqrt{2}G_F)^{-1/2}$ the vacuum expectation value, η_H the physical Higgs field, and G_F the Fermi constant [5]. The Higgs self-couplings can be measured by studying the production rates and kinematics of double and triple Higgs boson production at the LHC. These processes are highly suppressed, therefore a large amount of integrated luminosity is required in order for them to be observed and measured. With these measurements in mind, CERN is considering the proposed Phase II upgrades for the LHC in 2022 to the High-Luminosity LHC (HL-LHC), which is expected to be able to deliver a total integrated luminosity of 3 ab^{-1} [22].

A significant challenge at high luminosity is the increase of pile up events, multiple collisions per bunch crossing. The HL-LHC is expected to run with increased pile up at an average of 140 simultaneous pile up events [1]. Increased pile up poses challenges on the objectID and trigger to preserve their performance in this environment [25]. In order to address issues in maintaining detector performance, CMS will install an upgraded pixel tracker in the Phase1 upgrade period during the shutdown targeted for 2018, as well as several possible upgrades to the forward calorimetry systems and the silicon tracker currently under study for the Phase2 upgrade period [1].

The measurement of the Higgs boson self-coupling provides an important benchmark for evaluating the utility and performance of the different proposed designs. The dominant production mechanism for double Higgs boson production at the LHC is the gluon fusion process, and the production cross section for this process at a center of mass energy of 14 TeV has been calculated to next-to-leading-order to be 33.89 fb [4]. Double Higgs boson production can occur, not only through diagrams involving the Higgs boson self-coupling, but also through loop induced diagrams involving two $t\bar{t}H$ couplings. These processes will interfere destructively, such that the cross section is approximately a factor of two larger than in the Standard Model when lHHH is set to zero. The most optimistic channel for the di-Higgs analysis is that of one Higgs boson decaying to a pair of photons, and the other Higgs boson decaying to a pair of b-quarks, $HH \rightarrow b\bar{b}\gamma\gamma$, due to its relatively large branching fraction with a comparatively low background and good mass resolution of the

$H \rightarrow \gamma\gamma$ channel [24]. This channel has been studied in the past and shown to have reasonably good sensitivity at the LHC [4, 22].

In this chapter we present a comprehensive study of the Higgs pair production process in the $b\bar{b}\gamma\gamma$ decay channel at the HL-LHC using the CMS detector. Through a combination of generator level, fast simulation and full simulation studies we estimate the sensitivity of the di-Higgs production cross section measurement and examine the dependencies of the measurement sensitivity on various detector scenarios such as the high pile up environment as well as the photon and jet energy resolutions.

5.2 Object Selection

The signal process for this analysis is the production of two Higgs bosons, one decaying to a pair of b quarks and the other decaying to a pair of photons. The relevant backgrounds are broadly categorized into resonant backgrounds which contain a Higgs boson decaying to two photons and non-resonant backgrounds which do not contain a Higgs boson. Due to the excellent photon energy resolution of the CMS ECAL, the background involving the Higgs boson decaying to two b-jets is suppressed to negligible levels.

Our signal and background Monte Carlo samples are produced from generator-level events either from full simulation Monte Carlo samples of the current Run 1 CMS detector or Madgraph samples, undergoing parton showering and hadronization using Pythia and weighted with tagging efficiencies for photon and b-jets. Photons are selected using a relatively robust cut-based approach, applying requirements on the electromagnetic shower width, the hadronic to electromagnetic energy ratio and isolation. The efficiencies for photons to pass these selection criteria is evaluated using the GEANT-based full simulation of the current Run 1 CMS detector. We determine the tagging efficiency as the number of reconstructed objects passing ID cuts over the total number of generator-level objects. These efficiencies are calculated as a function of η and p_T and then are used to weight each event as a probability of each constituent particle. A similar process is used to determine the tagging

efficiencies of b-jets by searching for the presence of secondary vertices.

For this study we produced samples for the numerous reducible and irreducible backgrounds for the $HH \rightarrow b\bar{b}\gamma\gamma$ process. The irreducible backgrounds are estimated using full simulation samples and include resonant backgrounds from a Higgs boson produced in association with a Z boson, ZH, a top and anti-top quark pair, $t\bar{t}H$, and a b and anti-b quark pair, $b\bar{b}H$. There is also non-resonant irreducible background including QCD production of $b\bar{b}\gamma\gamma$, QCD production of $jj\gamma\gamma$ with light jets mistagged as b-jets, QCD production of $b\bar{b}j\gamma$ and $b\bar{b}jj$ with one and two jets misidentified as photons respectively, and QCD production of four jets with two jets mis-identified as photons and two jets mistagged as b-jet, $jjjj$, dominated by mistagged charm jets, $ccjj$.

To produce samples for the reducible backgrounds from fake photons and mistagged b-jets we calculated similar mistagging efficiencies for photons and b-jets. I calculated the efficiencies for jet-induced fake photons using full simulation of QCD di-jet samples, displayed in Figure 5.1. We note that the average fake rate for photons is about 5x10^{-4} for gluon jets and about 2x10^{-3} for quark jets. The mistag rate for light jets and charm jets as b-jets are also evaluated. The average mistag rate for light jets is about 1% and for charm jets is about 15%.

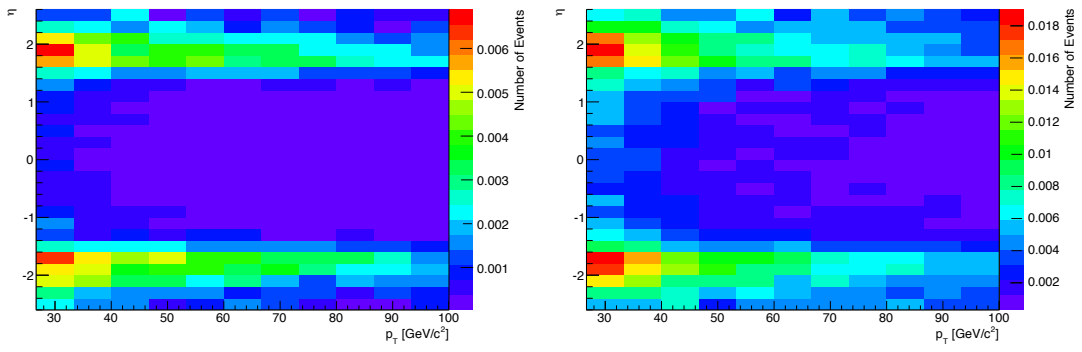


Figure 5.1: Photon mistagging efficiences due to gluons faking photons, on the left, and quarks faking photons, on the right.

In the selection process for reducible background samples we produce generator-level events only containing the constituent particles matching the respective background components. For the case of two fake photons and two real b-jets, $b\bar{b}jj$, I reject all events that contain real generator-level photons and that do not contain two generator-level b-jets, selected with PDG (Particle Detection Group) Identification. To increase the sample size we produce all potential fake photon pairs by promoting all unique pairs of remaining non-b-jets that have $p_T > 20$ and $|\eta|$ <2.5 as photons. The tagging efficiencies for each corresponding fake photon and real b-jet is then used to weight the individual events. Similar processes are used to determine the other reducible background $jj\gamma\gamma$, $ccjj$ and $jjjj$.

Using a combination of the full simulation of the CMS Phase1 detector with an average of 140 pileup events and the Delphes fast simulation, we extrapolate the photon and b-tagging efficiency to pileup conditions expected for the high luminosity LHC. The standard reconstruction has been used for the high pile up full simulation sample, and no attempt was made to re-optimize the algorithms. Therefore, significant improvements on these performances are likely possible. Based on these estimates, we reduce the b-jet tagging efficiency by a relative fraction of about 20% per b-jet and we reduce the photon selection efficiency by a relative fraction of about 20% per photon to extrapolate to the high pileup scenario of the HL-LHC. These object selection efficiency reductions result in an overall decrease of the signal efficiency by about 55%.

5.3 Cut Analysis

After the object selections for each signal and background sample we select events containing two photons with pT greater than 25 GeV and $|\eta| < 2.5$, and two b-tagged jets with pT greater than 30 GeV and $|\eta| < 2.4$. One of the two photons is required to have pT > 40 GeV. To suppress $t\bar{t}H$ background events, we require that there are no electrons or muons passing the veto selection and that the number of jets with $|\eta| < 2.5$ is less than 4. In order to determine the expected event yields for each sample the Monte Carlo is normalized to the respective sample cross section multiplied by the branching ratio of 0.089 fb. By inte-

grating over the normalized, weighted composite $b\bar{b}\gamma\gamma$ mass distributions we calculate the event yields for 3 ab^{-1} of integrated luminosity.

I investigated a number of different additional kinematic requirements in order to improve the signal to background ratio. We apply angular requirements on the ΔR between photons and b-jets. We denote this as "Selection A," requiring that $\Delta R_{\gamma\gamma} < 2.0$ and $\min(\Delta R_{\gamma b}) > 1.0$. I then examined additional requirements for further signal to background optimization by tightening these angular restrictions such that $\min(\Delta R_{\gamma b}) > 1.5$ and introducing a requirement on the ΔR of the two b-jets, $\Delta R_{b\bar{b}} < 2.0$. These cuts, in addition, to the previous selections, are denoted as "Selection B." Furthermore, I studied the performance of an alternative kinematic selection, where cuts are applied on the transverse momenta of the diphoton, di-bjet and di-Higgs system rather than the angular variables. I introduced "Selection C" requirements with $10 < P_{Tb\bar{b}\gamma\gamma} < 110$ GeV and $(P_{Tb\bar{b}} + P_{T\gamma\gamma}) > 260\ GeV$. Finally, I attempted to improve the signal to background further by tightening the cuts on the angular variables such that $\Delta R_{\gamma\gamma} < 1.6$, $\Delta R_{b\bar{b}} < 1.6$ and $\min(\Delta R_{\gamma b}) > 1.5$. We denote these cuts as "Selection B Tight."

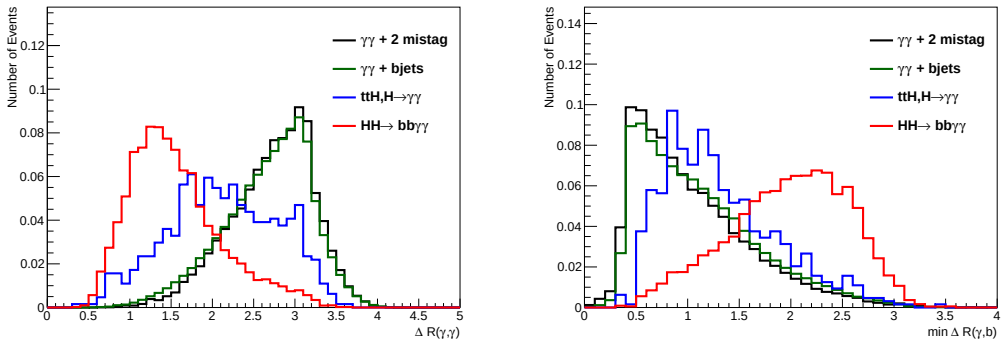


Figure 5.2: Normalized distributions of $\Delta R_{\gamma\gamma}$ and $min(\Delta R_{\gamma b})$ are shown for the di-Higgs signal, the $t\bar{t}H$ background and the QCD non-resonant backgrounds

In Table 5.1, we compare the expected signal and background event yields within the resonant mass windows for the four selection schemes that we studied. We observe that "Selection B" and "Selection C"" achieve similar signal to background ratios, and both im-

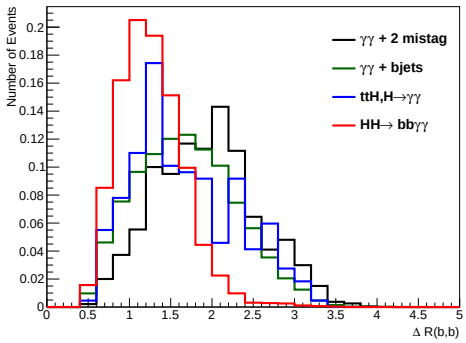


Figure 5.3: Normalized distributions of ΔR_{bb} after the "Selection A"requirements are shown for the di-Higgs signal, the $t\bar{t}H$ background, and the QCD non-resonant backgrounds.

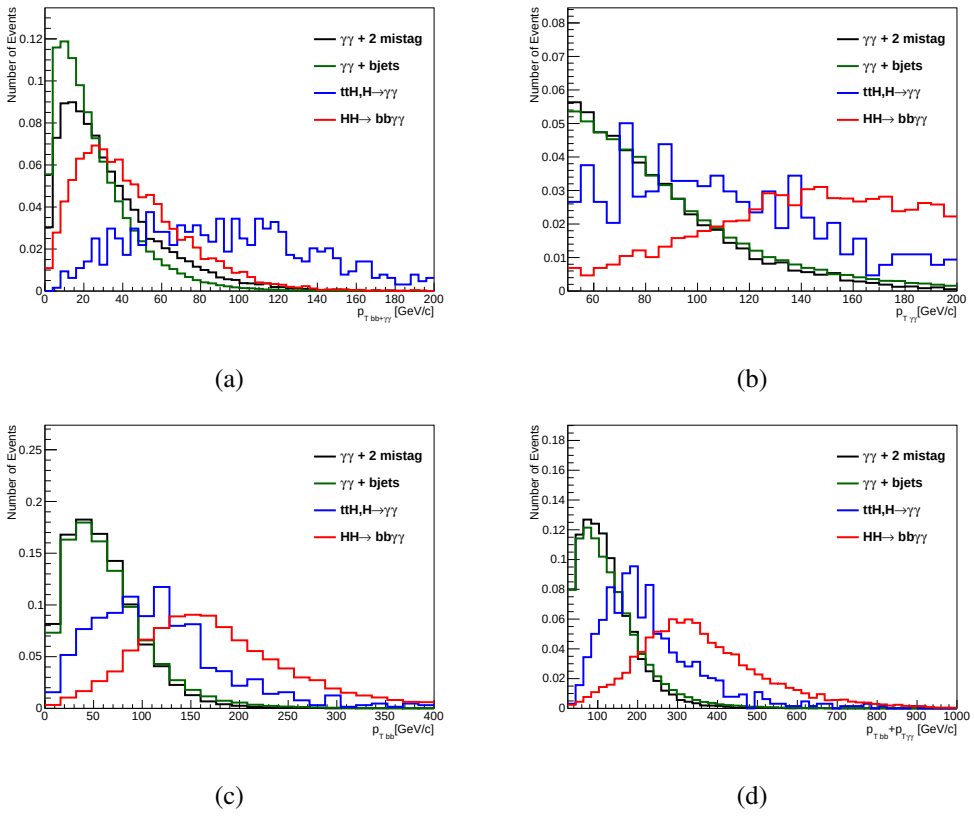


Figure 5.4: Normalized distributions of the p_T of the di-Higgs system (a), the diphoton system (b), and the di-bjet system (c) are shown for the di-Higgs signal, the $t\bar{t}H$ background, and the QCD non-resonant backgrounds. The distribution of the sum of the p_T of the diphoton system and the p_T of the di-bjet system is shown in (d).

prove slightly upon the "Selection A". "Selection B Tight" achieves significantly improved signal to background, but at a cost of about 25% in signal efficiency. Attempts at achieving

better signal to background ratio by tightening cuts in the "Selection C" scenario resulted in significantly larger reduction in signal yield compared to the "Selection B Tight" scenario. I also studied the degree to which the different kinematic requirements alter the shape of

Expected Events: M_{bb}: (105,145) GeV, $M_{\gamma\gamma}$: (120,130) GeV				
Sample Type	Selection A	Selection B	Selection C	Selection B Tight
$HH \rightarrow b\bar{b}\gamma\gamma$	11.9	10.6	10.3	7.8
$ZH \rightarrow b\bar{b}\gamma\gamma$	3.6	2.3	2.4	2.0
$ttH, H \rightarrow \gamma\gamma$	2.9	1.7	1.5	1.1
$bbH, H \rightarrow \gamma\gamma$	1.2	0.9	0.8	0.6
$bb\gamma\gamma$	15.3	9.3	8.7	3.7
$jj\gamma\gamma$	16.0	7.0	9.7	2.7
$jjjj$, $ccjj$, $bbjj$	0.2	0.02	0.02	0.0
$t\bar{t}$	0.8	0.1	0.1	0.0
S/B Ratio	0.30	0.50	0.44	0.77
S/$\sqrt{B}$ (Signal Significance)	1.9	2.3	2.1	2.4

Table 5.1: Expected event yields for $3ab^{-1}$ of integrated luminosity for the di-Higgs signal and background under various event selection schemes

the diphoton and di-bjet mass distributions. In order to extract the signal from the non-resonant background we perform a maximum likelihood fit on the diphoton and di-bjet mass distributions. For this analysis method, it is critical that any of the kinematic selection cuts that are applied preserve the smooth and monotonic behavior of the non-resonant background. Any peak-like feature that is introduced in the non-resonant background would bias the determination of the background as well as the cross section measurement.

In Figure 5.5, we show the distributions of the diphoton and di-bjet mass after the "Selection A", "Selection B", and "Selection C" requirements for the non-resonant $b\bar{b}\gamma\gamma$ background. We note that the Selection A and the Selection B cuts do preserve the general exponentially decaying trend in the mass distributions, however the Selection C cuts appear to significantly change the shape of the mass distributions. They still appear smooth and monotonic, but are significantly more flat, with a more severe inflection in the case of the diphoton mass.Therefore, we prefer the "Selection B" cuts and continued the remaining analysis with the "Selection B" requirements.

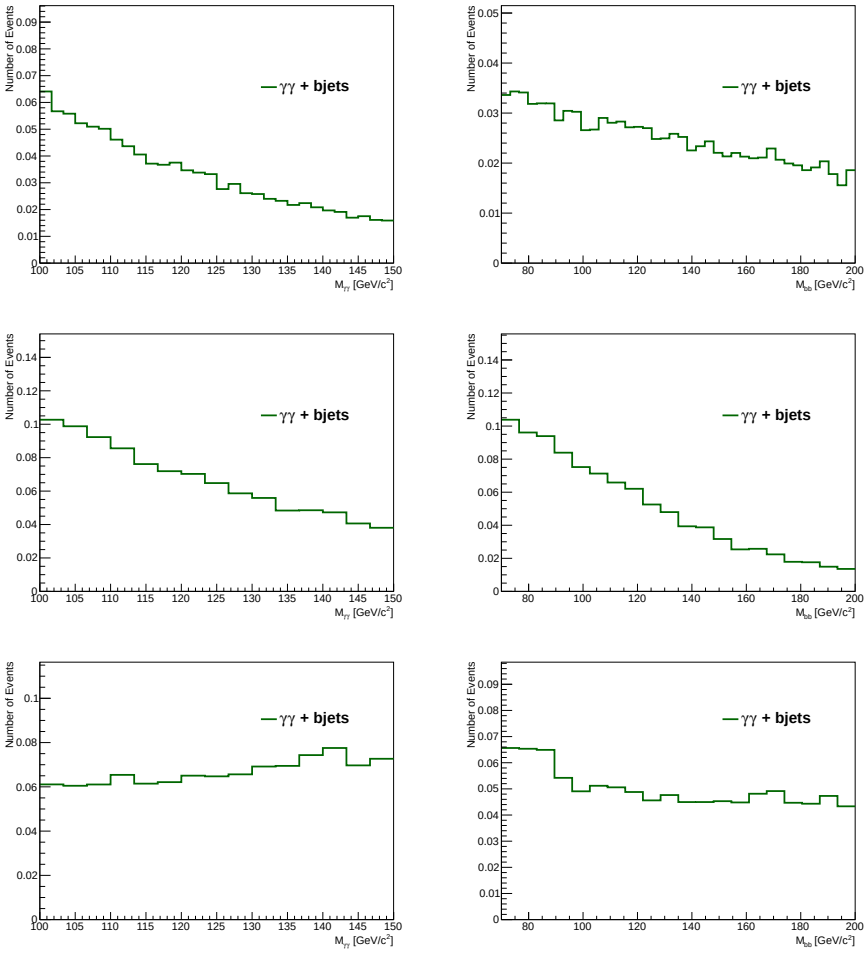


Figure 5.5: The distribution of the diphoton and di-bjet mass for the QCD $b\bar{b}\gamma\gamma$ background process, after the various event selection schemes. The diphoton mass distributions are shown on the left and the di-bjet mass distributions are shown on the right. The selections are in descending order of A, B and C.

5.4 Signal Extraction

Two methods are utilized to determine the expected $HH \rightarrow b\bar{b}\gamma\gamma$ signal yield. The first involves a cut and count, where we apply event selection requirements as mentioned in the previous section and count the event yield, subtracting the expected background. The second analysis applies ”Selection B” kinematic selections and performs a two dimensional maximum likelihood fit on the diphoton and di-bjet masses. We derive probability

density functions (PDFs) for the diphoton mass, $M_{\gamma\gamma}$, and the dibjet mass, M_{bb}, distributions for the signal, the resonant background and the non-resonant background by fitting the distributions from the Monte Carlo simulation samples to particular parametrization of the line-shape for $M_{\gamma\gamma}$ and M_{bb}. The diphoton mass distribution for the signal and resonant background is fitted to a Gaussian distribution. The non-resonant backgrounds, being dominated by ZH where the Z bosons decay into a pair of b-jets, is fitted to a Crystal Ball distribution. The non-resonant background for the di-bjet mass is fitted to a decaying exponential. Figure 5.6 shows the diphoton and dibjet mass distributions and PDFs used to fit the model for the line-shape for the signal, resonant and non-resonant background with MC samples under low pile up conditions. To model the degraded jet energy resolution under the HL-LHC pile up conditions, the jet energy resolution parameter is then appropriately increased.

The PDFs for the signal and backgrounds are weighted with the number of events within the fit mass window determined by the cut analysis in order to build the full model for the signal region sample. Then toy MC experiments are randomly drawn from the full model and two-dimensional fits are performed where the yields for the signal and backgrounds are floated. We also evaluate the 95% confidence level upper limit on the di-Higgs production cross section, as an alternative measure of the sensitivity and find an upper limit of 78fb corresponding to 2.3 times the Standard Model cross section. Using the dependence of the cross section on the coupling parameter provided by [4], the projected cross section uncertainty translates into an uncertainty on λ_{HHH} of $+386\% - 98\%$.

We show also the distribution of the fit uncertainty on the signal yield where we observe that the average statistical uncertainty of the cross section measurement using the two dimensional fit is about 45%, representing an improvement on the cut-based analysis of about 10%. Based on the dependence of the cross section on the coupling parameter, this cross section uncertainty translates into an uncertainty on λ_{HHH} of $+77\%/-48\%$.

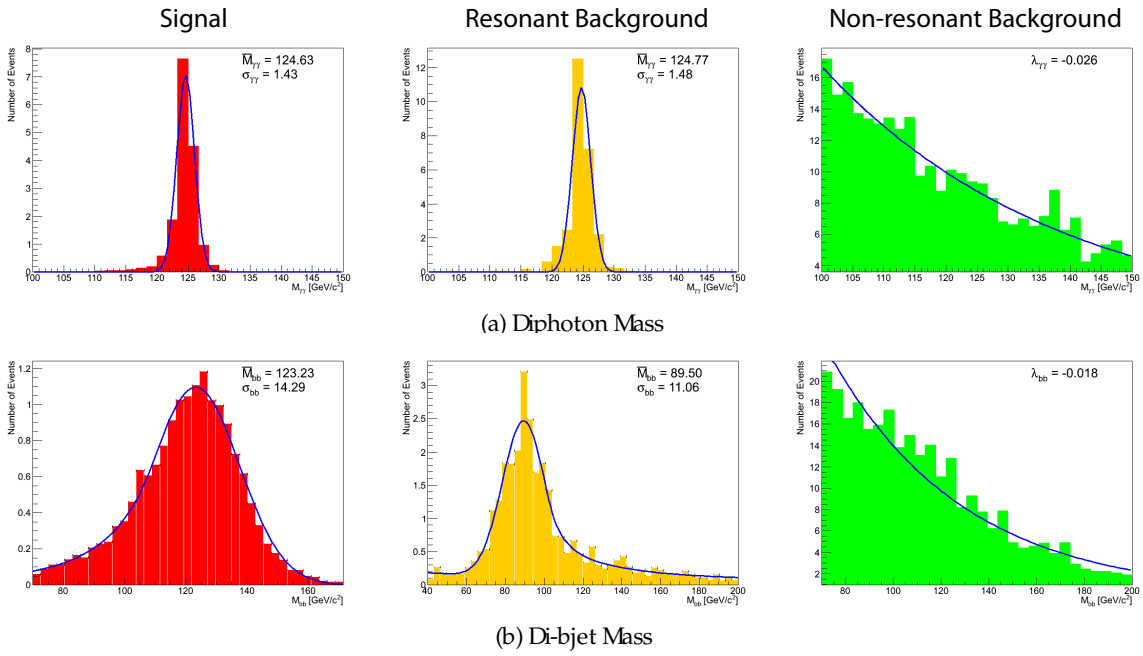


Figure 5.6: Initial fits to the diphoton and di-bjet mass distributions for the signal (left), the resonant background (middle) and the non-resonant background (right) to define the probability density functions. The PDF models for the diphoton mass are shown on top and the models for the di-bjet mass are on the bottom. [1]

5.5 Systematic Uncertainties

At an integrated luminosity of 3000 fb^{-1} the systematic uncertainties are dominated by statistical uncertainties. The photon selection efficiency systematic uncertainty is dominated by the systematic efficiency of the electron veto requirement. With the jet energy resolution measured to better than 10% and the photon energy resolution to within 15%, the average bias induced on the cross section measurement are about 2% and 9% for the jet and photon energy resolutions, respectively. Systematic uncertainties for the photon selection efficiency and b-tagging efficiency are negligible in comparison to statistical uncertainty. The systematic uncertainty on the non-resonant background is determined where the assumed non-resonant background model used is the product of an exponential and a fourth degree polynomial fitted to the expected non-resonant background distribution with an exponential fit used to perform the fit. This results in an average bias of about 12%. The dominant systematic uncertainties on the coupling measurement arises from the theoretical uncertainties of the double Higgs boson production cross section. The total systematic uncertainty of

34% which corresponds to the cross section measurement as a coupling measurement in a total systematic uncertainty of $+52\%/-38\%$ on the coupling parameter [1].

5.6 Upgrade Scenarios

The degradation in object selection performance under the HL-LHC, pile up conditions results in a decrease in the signal efficiency of more than a factor of two. As the measurement is primarily limited by the number of selected signal events, improving the object selection efficiency is the most important aspect. To explore the effect of the object selection efficiencies and to provide a general goal for the detector upgrade. We show in figure 5.7 the sensitivity as a function of the relative improvement on the photon selection efficiency and the b-tagging efficiency, over the current performance estimate under the HL-LHC pile up conditions.

With more sophisticated multivariate techniques, improvements to the jet energy resolution may be possible. We examine the dependence of the di-Higgs cross section measurement sensitivity on the photon and jet energy resolution. We vary the widths of the input model PDF for the di-photon mass and the di-bjet mass, generating a corresponding set of toy experiments, and performing the 2D mass fit analysis for each such set. For the di-photon mass, we scan widths ranging from 0.9 GeV to 3.9 GeV, corresponding to average photon energy resolutions ranging from 1% to 4.5%. For the dibjet mass, we scan widths ranging from 10 GeV to 20 GeV, corresponding to average jet energy resolutions ranging from 11% to 22% . For a goal precision of 50% on the di-Higgs production measurement, we find that the jet energy resolution must improve by roughly 25% over the current performance estimated with HL-LHC pile up conditions [1]. Looking at the sensitivity of the cross section measurement on the total integrated luminosity in Figure 5.8, with two experiments both collecting an integrated luminosity of 3 ab^{-1}, we expect a relative uncertainty of about 35%. This translates into an uncertainty on λ_{HHH} of $+54\%/-39\%$.

The current fake photon background in the end cap is prohibitively large. To study the

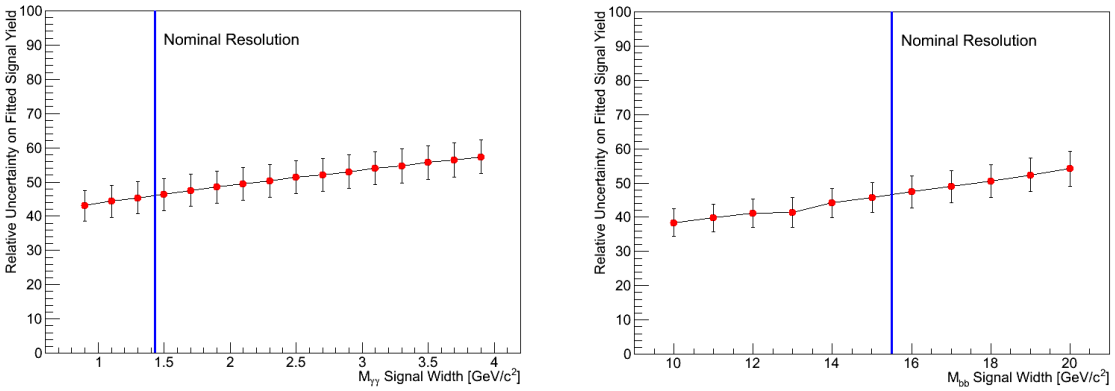


Figure 5.7: The average expected relative uncertainty on the di-Higgs cross section measurements are shown as a function of the assumed widths of the diphoton (left) and di-bjet (right) mass distributions for the signal. [1]

effect of an upgraded end cap calorimeter with possibly improved signal to background discrimination for photons, we examine the assumed photon fake rate in the end cap and observe that a reduction in the end cap photon fake rate of a factor of 10 , yielding a corresponding reduction in the fake photon non-resonant background, yields an improvement in the cross section measurement uncertainty of about 4%. Finally, we study the analysis sensitivity on the total integrated luminosity. We find that in order to achieve a relative uncertainty on the cross section measurement of 50% with the one experiment, we must achieve a relative improvement in the photon selection efficiency and b-tagging efficiency of about 25% and an improvement in the jet energy resolution of about 25%. These performance parameters correspond roughly to what is achieved under Run1 pile up conditions. These establish clear performance goals for the detector upgrades and algorithm improvements. A significant part of the required improvement can be achieved from improved b-tagging and photon identification algorithms, while the rest may be achievable through an upgraded silicon tracker with possibly an extended coverage to forward regions, and an upgraded forward calorimeter with improved photon-pion discrimination.

We investigate the sensitivities for the cut and fit-based analyses with potential detector scenarios. First, we extrapolated to the HL-LHC pileup scenario by combining the full simulation of the CMS Phase 1 detector with an average of 140 pileup events and the

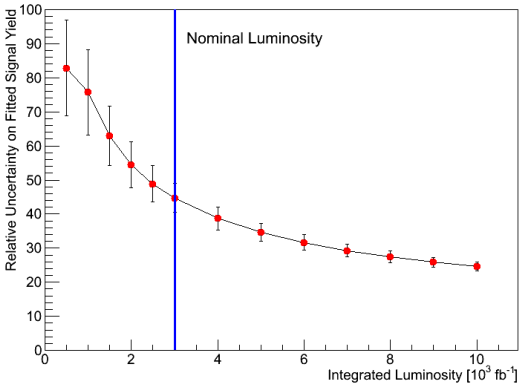


Figure 5.8: The average expected relative uncertainty on the di-Higgs cross section measurements are shown as a function of the integrated luminosity [1].

Delphes fast simulation and then extracting the degraded photon and b-tagging efficiencies for the higher luminosity conditions. We then reweight the samples and produce the expected event yields. From these expected event yields, we observe that under this very pessimistic detector performance scenario, the di-Higgs cross section measurement sensitivity would degrade by about 50%, yielding a statistical uncertainty of about 77% for a cut-based analysis. The two dimensional mass fit analysis yields statistical uncertainty of about 68% under this scenario.

We also explore the scenario where the size of the non-resonant background is increased by a factor of 2. This may occur due to degraded object selection performance resulting in poorer signal to background discrimination, or simply due to a systematic mis-estimate of the QCD backgrounds from the current Monte Carlo simulation. We also explore the scenario where the size of the non-resonant background is decreased by a factor of 2, which may occur if detector hardware or algorithmic improvements can significantly improve the b-tagging performance such that the mistag background from the $jj\gamma\gamma$ is reduced to negligible levels. The results for the cut-based analysis and the two dimensional mass fit analysis are summarized in Table 5.3.

Expected Event Yield with Average 140 Pile Up: Selection B			
Sample Type	Fit Window $M_{b\bar{b}}$: (70,200) $M_{\gamma\gamma}$: (100,150)	Conservative Window $M_{b\bar{b}}$: (105,145) $M_{\gamma\gamma}$: (120,130)	Optimistic Window $M_{b\bar{b}}$: (110,140) $M_{\gamma\gamma}$: (122,128)
$HH \rightarrow b\bar{b}\gamma\gamma$	7.5	5.5	4.3
$ZH \rightarrow b\bar{b}\gamma\gamma$	8.9	1.3	0.6
$ttH, H \rightarrow \gamma\gamma$	2.2	0.8	0.5
$bbH, H \rightarrow \gamma\gamma$	1.5	0.4	0.3
$bb\gamma\gamma$	69.0	4.4	2.1
$jj\gamma\gamma$	73.4	5.1	2.6
$jjjj$, $ccjj$, $bbjj$	11.1	0.0	0.0
$t\bar{t}$	5.3	0.2	0.05
S/B Ratio	0.04	0.45	0.70
S/$\sqrt{B}$ (Signal Significance)	0.57	1.6	1.7

Table 5.2: Expected event yields for $3ab^{-1}$ of integrated luminosity estimated for high pile up scenario expected for the HL-LHC. The event yield reflect "Selection B" requirements within three different mass windows.

5.7 Conclusion

Through the combination of generator level, fast simulation, and full simulation studies we have estimated the sensitivity to the di-Higgs production cross section measurement to be around 56% for 3 ab^{-1} of integrated luminosity at the HL-LHC. The expected fit uncertainty is around 104% for the nominal object selection performance scenario, the experimental systematic uncertainty is about 13%. The expected statistical uncertainty is around 45% and the systematic uncertainty, dominated by theoretical uncertainties, is about 34%. This translates into statistical and systematic uncertainties on the self-coupling parameter of $+77\% - 48\%$ and $+52\%/ - 38\%$, respectively. The approximate impact of the high pileup environment expected for the HL-LHC has been evaluated and, though the degradation is rather severe using the naive and un-tuned algorithms, substantial recovery of the current detector performance may be achieved by re-optimizing the reconstruction algorithms. Finally, we have reported dependencies of the cross section measurement sensitivity on the eventual performance of the detector on photon and jet energy resolutions,

Scenario	Cut-Based Analysis	2D Mass Fit Analysis
Nominal Result	50%	45%
High Pile Up	77%	68%
Non-Resonant Bkgd x 2	61%	52%
Non-Resonant Bkgd / 2	43%	41%

Table 5.3: The average expected relative uncertainties on the di-Higgs cross section measurement are shown for various detector scenarios [22].

as well as the total integrated luminosity [1].

Accounting for the additional theoretical uncertainty on the predicted cross section of 30%, we obtain statistical and systematic uncertainties on the Higgs trilinear self-coupling parameter λ_{HHH} of $+385\% - 98\%$ and $+50\% - 37\%$, respectively under the nominal object selection performance scenario. To achieve the goal of measuring the di-Higgs production cross section to 50% precision (35% with two identically performing experiments), we must achieve 25% improvement on the photon selection efficiency, b-tagging efficiency and jet energy resolution over the current performance estimates for the HL-LHC pile up condition. If this performance is achieved, the Higgs trilinear self-coupling parameter can be determined to a precision of $+93\% - 52\%$ with one experiment and $+55\% - 39\%$ with two identically performing detectors [22].

Significant degradation in cross-section measurements is expected if progress is not made to improve particle tagging efficiencies and discrimination between signal and background. A key feature to the $HH \rightarrow b\bar{b}\gamma\gamma$ channel is the diphoton decay component of the Higgs boson which hosts a narrow peak over a smooth background. Improvements to the photon energy resolution would result in a narrower mass peak which improves the overall differentiation of signal to background [25]. Photon resolution and identification efficiency at HL-LHC will be significantly influence by the pile-up conditions. Photon isolation and shower shapes play a central role in identification. CMS employs a particle flow algorithm to determine the photon isolation, optimally combining the information from all subdetectors to determine the energy flow in the isolation cone around the photon. Photon

reconstruction, energy measurement, and identification all depend on sophisticated algorithms which are tuned on large samples of fully simulated events with proper pile-up conditions. It is crucial that these algorithms are enhanced to compensate for future conditions [1].

A predominant background for $HH \rightarrow b\bar{b}\gamma\gamma$ was from non-resonant QCD $b\bar{b}\gamma\gamma$ production. As previously mentioned, the non-resonant background was evaluated with an assumed background model using a product of an exponential and fourth degree polynomial fitted with an exponential function. In this case where there is no theory-motivated model to describe this non-resonant background, the use of multi-model averaging may prove beneficial in producing a data-driven model that can better describe the background and provide better results for the signal yield. Improved model selections for both the $H \rightarrow \gamma\gamma$ SM background as well as the non-resonant background of the $HH \rightarrow b\bar{b}\gamma\gamma$ could yield significant improvements in future di-Higgs signal measurements. Overall, the increase in data and its complexity in future physics analyses, such as at the HL-LHC, suggests a need for more sophisticated computing methods.

Chapter 6

Discussion

The aim of this study was to apply a multi-model inference ranking method using Akaike Information Criterion to determine an appropriate model for the Standard Model background that best represents the data. The current results reveal that the previous selection of a 5th-order polynomial as the background model is not well supported by the 2013 diphoton data based on the AIC measurements. Seeing that the improper modeling of the Standard Model background has a substantial effect on the modeling and measurement of the $H \rightarrow \gamma\gamma$ signal, a review of the model selection is necessary.

In addition to the applications of MMI technique to the $H \rightarrow \gamma\gamma$ background analysis, we have presented updated results on the CMS detector performance using $Z \rightarrow \mu\mu\gamma$ events as well as the expected di-Higgs cross section measurement uncertainties for the future analysis of the Higgs self-couplings at the future High-Luminosity Large Hadron Collider. We have discussed not only the current state of these measurements under present detector conditions but also motivate potential improvements in both the detector design and algorithmic development. It is apparent that the increase in pile up is a substantial factor for the detector energy resolution as well as the success of the trigger and particle identification systems. The inevitable future outlook of increased luminosity conditions with an increase in pile up to even an order of magnitude truly poses a challenge to future analyses. In the case of the HL-LHC, significant degradation in measurements are expected if efforts are not made to improve detector hardware and particle identification algorithms.

In terms of advancements in computational methods, we have also highlighted potential applications of the MMI method to both the PHOSPHOR Fit method as well as the di-Higgs cross section measurement. The issue of background model selection and its application to improving signal to background differentiation is apparent in many areas of study. The use of a multi-model averaging technique to properly model the $Z \rightarrow \mu\mu\gamma$ backgrounds may prove beneficial in improving the overall detector resolution measurements which are directly applicable to the $H \rightarrow \gamma\gamma$ and $HH \rightarrow b\bar{b}\gamma\gamma$ measurements. Producing data-driven models for these backgrounds may assist in surpassing the statistical limits of the relevant fitting processes.

Overall, the significance of model selection with applications to fit-based analyses for signal extraction is quite clear. The ability to properly extract relevant information from combined signal and background models rather than relying on cut-based approaches to isolate signal has multiple statistical benefits. The issues of relying on kinematic selections for signal discrimination is that the overall amount of statistics for the final fit is limited which may bias the results and get rid of valuable information. As seen previously in the Higgs self-coupling analysis, it is also difficult to determine appropriate cuts that isolate solely the background contributions and that do not significantly alter the shape of the signal and fitted background distributions. This analysis also revealed the overall improvement in the statistical uncertainty on the signal yield from fit-based analyses over cut-based analyses. Ultimately it is advantageous to be able to determine a proper model function for the shape of the background distribution given in order to minimize the bias in the final signal extraction. In the event that our theory does not provide enough motivation for this model, a realistic goal is to determine a data-driven model which best approximates this distribution and minimizes the bias to the signal region. Multi-model averaging may prove beneficial for future analyses as the resulting function contains more information obtained from the data itself than a single ”best” model and therefore is a more robust approximation of the data.

Bibliography

[1] Abdallah, H., An, D., Christian, A., Klute, M., Lambert, V. and Xie S. Measurement of the Higgs Pair Production Cross Section at 14 TeV in the Decay Channel to Two Photons and Two B-Jets. *CMS Analysis Note*, AN-2013/050, 2013.

[2] Akaike, H. Information Theory and an Extension of the Maximum Likelihood Principle. *Proc. 2nd Int. Symp. Information Theory*, pages 267–281, 1973.

[3] Altarelli, G. Collider Physics within the Standard Model: a Primer. *CERN-PH-TH/2013-020*, 2013.

[4] Baglio, J., Dijouadi, A., Grober, R. et al. The measurement of the Higgs self-coupling at the LHC: theoretical status. *JHEP*, 1304, 2013.

[5] Baur, U., Plehn, T. and Rainwater, D. Probing the Higgs self-coupling at hadron colliders using rare decays. *Phys. Rev. D*, 69, 2003.

[6] Bornheim, A., Veverka, J., Ma, Y., Timciuc, V., et al. Measurement of the CMS ECAL scale and resolution performance from $Z \rightarrow ee$ decays and determination of the signal line shape for $H \rightarrow \gamma\gamma$. *CMS Analysis Note*, AN-2011/204, 2011.

[7] Bornheim, A., Veverka, J., Ma, Y., Timciuc, V., et al. Updated measurements of the Higgs-like boson at 125 GeV in the two photon decay channel. *CMS Analysis Note*, AN-2013/008, 2013.

[8] Brown, R.M., and Cockerill, D.J.A. Electromagnetic calorimetry, Nuclear Instruments and Methods in Physics Research. *Phys. Lett. B*, A:47–79, 2012.

[9] Burnham, K. and Anderson, D. *Model Selection and Multimodel Inference, A Practical Information-Theoretic Approach*. Springer, 2nd edition, 2004.

[10] Burnham, K., and Anderson, D. Multimodel Inference, Understanding AIC and BIC in Model Selection. *Sociological Methods and Research*, 3(2):261–304, November 2004.

[11] Burnham, K., Anderson, D. and Huyvaert, K. . AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. *Behavioral Ecology Sociobiology*, 65:23–25, 2011.

[12] CMS Collaboration. Search for a Higgs boson decaying into two photons in pp collisions recorded by the CMS detector at the LHC. *CMS Analysis Note*, AN-2011/426, 2011.

[13] Fabjan, C. and Gianotti, F. . Calorimetry for particle physics. *Reviews of Modern Physics*, 75, October 2003.

[14] Gull, S. Bayesian Inductive Inference and Maximum Entropy. *Maximum-Entropy and Bayesian Methods in Science and Engineering*, 1:53–74, 1988.

[15] $H \rightarrow \gamma\gamma$ Working Group. Updated Measurements of the new Higgs-like boson at 125 GeV in the two photon decay channel. *CMS Physics Analysis Summary*, HIG-2013/001, 2013.

[16] Lambert, V., Xie, S., Bornheim, A., Di Marco, E., Newman, H. and Spiropulu, M. CMS ECAL Performance with Z Dielectron Events in 2012 Data. *CMS Analysis Note*, AN-2012/408, 2012.

[17] Lavoué, J., and Droz, P. . Multimodel Inference and Multimodel Averaging in Empirical Modeling of Occupational Exposure Levels. *In Annuals of Occupational Hygiene*, 53, 2009.

[18] Peña, C., Bornheim, A., Mott, A., Veverka, J., Xie, S., and Yang, Y. Measurement of Photon Energy Scale and Resolution with $Z^0 \rightarrow \mu^+\mu^-\gamma$ decays in 2011 and 2012 Data. *CMS Analysis Note*, AN-2012/382, 2012.

[19] Pearl, J. *Causality*. Cambridge University Press, 2nd edition, 2009. UCLA.

[20] The CMS Collaboration . Search for the standard model Higgs boson decaying into two photons in pp collisions at $\sqrt{s} = 7$ TeV. *Physics Letters B*, 710:403–425, April 2012.

[21] The CMS Collaboration. Observation of a new boson at a mass of 125 GeV with the CMS experiment at the LHC. *Phys. Lett. B*, 716:3061, 2012.

[22] The CMS Collaboration. A Study of the Higgs Boson Pair Production Cross Section at 14 TeV in the Decay Channel to Two Photons and Two B-Jets. *CMS Physics Analysis Summary*, FTR-2013/001, 2013.

[23] The CMS Collaboration. Properties of the observed Higgs-like resonance decaying into two photons. *CMS Physics Analysis Summary*, HIG-2013/016, 2013.

[24] The CMS Collaboration. Search for the resonant production of two Higgs bosons in the final state with two photons and two bottom quarks. *CMS Physics Analysis Summary*, HIG-2013/032, 2013.

[25] The CMS Collaboration. Study of cross-section and Higgs property measurements in the $\gamma\gamma$ decay mode with an upgraded CMS detector at the high luminosity LHC. *CMS Physics Analysis Summary*, FTR-2013/004, 2013.

[26] The CMS Collaboration. Observation and measurement of the two photon decay of the 125 GeV Higgs boson. *CMS Physics Analysis Summary*, HIG-2014/001, 2014.

[27] The CMS Collaboration. Search for a Higgs boson decaying into $\gamma\gamma \rightarrow \mu\mu\gamma$ with dilepton mass below 20 GeV in pp collisions at $\sqrt{s} = 8$ TeV. *CMS Physics Analysis Summary*, HIG-2014/003, 2014.

[28] The CMS Collaboration. Search for the resonant production of two Higgs bosons in the final state with two photons and two bottom quarks. *CMS Physics Analysis Summary*, HIG-2014/032, 2014.

[29] Veverka, J. *Studies of $Z\gamma$ Production and Constraints on Anomalous Triple Gauge Coupling in pp Collisions at $\sqrt{s} = 7$ TeV*. PhD thesis, California Institute of Technology, 2013.

[30] Weinstein, M. and Horn, D. Dynamic quantum clustering: a method for visual exploration of structures in data. *Physics Review E*, 80:06617, 2009.

[31] Yang, Y. *Search for a Standard Model Higgs Boson Decaying to Two Photons in CMS at the LHC*. PhD thesis, California Institute of Technology, 2013.