2018

A Search For New Low-Mass Diphoton Resonances At Atlas And An Investigation Into Using Gaussian Process Regression To Model Non-Resonant Two-Photon Standard Model Backgrounds

Rob Roy Macgregor Fletcher
University of Pennsylvania, robroy.fletcher@gmail.com

Recommended Citation

Fletcher, Rob Roy Macgregor, "A Search For New Low-Mass Diphoton Resonances At Atlas And An Investigation Into Using Gaussian Process Regression To Model Non-Resonant Two-Photon Standard Model Backgrounds" (2018). *Publicly Accessible Penn Dissertations*. 3115.
https://repository.upenn.edu/edissertations/3115

A Search For New Low-Mass Diphoton Resonances At Atlas And An Investigation Into Using Gaussian Process Regression To Model Non-Resonant Two-Photon Standard Model Backgrounds

Abstract

The Standard Model of particle physics has been tested over many years with many ex- periments and has predicted experimental results with remarkable accuracy. In 2012, the last piece of the Standard Model, the Higgs boson, was discovered by the experiments ATLAS and CMS at the Large Hadron Collider (LHC). Although this completes the Standard Model, this by no means completes our picture of the physics that describes the observable universe. Several phenomena and measurements remain unexplained by the Standard Model including gravity, dark matter, the baryon-antibaryon asymmetry of the universe and more. One of the primary goals of the LHC and the ATLAS experiment are to search for extensions and modifications to the Standard Model that could help to explain these phenomena. This the- sis presents three areas where I made major contributions. The first is in the identification of prompt electrons in ATLAS using a likelihood method both in the online trigger system and in offline data analysis. Prompt electrons are ubiquitous in the signatures of electroweak physics, one of the cornerstones of the ATLAS physics program. Next I present a search for new physics in low-mass (65-110 GeV) diphoton events. This is a model independent search that is motivated by several extensions to the Standard Model including the two Higgs doublet model where new scalars can appear as lighter versions of the Standard Model Higgs. No evidence for a new narrow resonance is found, so limits ranging from 30 to 101 fb are set on the production cross section of such a resonance, assuming that its branching fraction to two photons is 100 percent. The sensitivity of these results are limited by the systematic uncertainties due to the potential spurious signals introduced by the two-photon non-resonant Standard Model background. My third contribution was an initial investigation of a new method to model this background using Gaussian Process Regression.

Degree Type

Dissertation

Degree Name

Doctor of Philosophy (PhD)

Graduate Group

Physics & Astronomy

First Advisor

I. J. Kroll

Keywords

Diphoton

Subject Categories

Physics

A SEARCH FOR NEW LOW-MASS DIPHOTON RESONANCES AT ATLAS AND AN INVESTIGATION INTO USING GAUSSIAN PROCESS REGRESSION TO MODEL NON-RESONANT TWO-PHOTON STANDARD MODEL BACKGROUNDS

Rob Roy MacGregor Fletcher

A DISSERTATION
in
Physics and Astronomy

Presented to the Faculties of The University of Pennsylvania
in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy
2018

I. Joseph Kroll, Professor, Physics
Supervisor of Dissertation

Joshua Klein, Professor, Physics
Graduate Group Chairperson

Dissertation Committee
Elliot Lipeles, Associate Professor, Physics
Cullen Blake, Assistant Professor, Physics
Evelyn Thomson, Associate Professor, Physics
Justin Khoury, Professor, Physics
I. Joseph Kroll, Professor, Physics

A SEARCH FOR NEW LOW-MASS DIPHOTON RESONANCES AT ATLAS AND AN INVESTIGATION INTO USING GAUSSIAN PROCESS REGRESSION TO MODEL NON-RESONANT TWO-PHOTON STANDARD MODEL BACKGROUNDS

ABSTRACT

A SEARCH FOR NEW LOW-MASS DIPHOTON RESONANCES AT ATLAS AND AN INVESTIGATION INTO USING GAUSSIAN PROCESS REGRESSION TO MODEL NON-RESONANT TWO-PHOTON STANDARD MODEL BACKGROUNDS

Rob Roy MacGregor Fletcher

I.J. Kroll

The Standard Model of particle physics has been tested over many years with many experiments and has predicted experimental results with remarkable accuracy. In 2012, the last piece of the Standard Model, the Higgs boson, was discovered by the experiments ATLAS and CMS at the Large Hadron Collider (LHC). Although this completes the Standard Model, this by no means completes our picture of the physics that describes the observable universe. Several phenomena and measurements remain unexplained by the Standard Model including gravity, dark matter, the baryon-antibaryon asymmetry of the universe and more. One of the primary goals of the LHC and the ATLAS experiment are to search for extensions and modifications to the Standard Model that could help to explain these phenomena. This thesis presents three areas where I made major contributions. The first is in the identification of prompt electrons in ATLAS using a likelihood method both in the online trigger system and in offline data analysis. Prompt electrons are ubiquitous in the signatures of electroweak physics, one of the cornerstones of the ATLAS physics program. Next I present a search for new physics in low-mass (65-110 GeV) diphoton events. This is a model independent search that is motivated by several extensions to the Standard Model including the two Higgs doublet model where new scalars can appear as lighter versions of the Standard Model Higgs. No evidence for a new narrow resonance is found, so limits ranging from 30 to 101 fb are set on the production cross section of such a resonance, assuming that its branching fraction to two photons is 100 percent. The sensitivity of these results are limited by the systematic uncertainties due to the potential spurious signals introduced by the two-photon non-resonant Standard Model background. My third contribution was an initial investigation of a new method to model this background using Gaussian Process Regression.

Contents

Preface

This thesis describes some of the work I did as a graduate student but does not attempt to be a complete and detailed discussion of any topics. My hope is that someone will find this useful, especially the last chapter which describes my attempts to refine how background modeling in the low mass diphoton group is done. This work will likely be carried on by new students and this may be a good starting point to see what I have tried in the past, what work I was doing at the end, and what studies were outstanding. This will also be a source of references that describe in much more detail the topics contained in here.

This thesis will be organized as follows: Chapter 1 contains an overview of my time in graduate school, with very brief descriptions of work I performed that are not the focus of later sections. This section is largely historical and is intended to document projects for my own benefit. Chapter 2 gives a short introduction to the Standard Model of particle physics as well as a few shortcomings of the theory. As the main analysis described in this thesis is motivated by extensions to the scalar sector of the Standard Model, emphasis is placed on how additional scalars can help to address gaps in our understanding of the fundamental particles and their interactions. Chapter 3 describes the Large Hadron Collider and the ATLAS experiment. This chapter skips over a tremendous amount of detail, but references give complete descriptions of both the LHC and ATLAS. Each collaborator on the ATLAS experiments earns the right to be an author on physics publications by performing a qualification task. My task was performed in a so-called e/gamma group, a Combined Physics and Performance group that focuses on the identification of electrons and photons whose experimental signatures are very similar in the ATLAS detector. I worked on the identification of prompt electrons produced in the decay of electroweak processes such as $W \to e\nu$, $Z \to e^+e^-$. This work is described in Chapter 4. After qualifying for authorship, I began the work that would make up the bulk of this thesis. Experimental signatures with two photons in the final state provide clean processes with relatively small backgrounds which makes them a good choice to search for new physics. A search for low mass diphoton resonances below the Standard Model Higgs mass is described in Chapter 5. The analysis on diphoton events uses a method known as spurious signal to test for functional forms to describe the background distribution. This method requires large quantities of simulated data which are prohibitively expensive to generate. Chapter 6 describes some initial investigation with the goal to replace this method with a more advanced technique based on Gaussian Process Regression. This method was ultimately not used in this round of analysis, however I think is a very promising way forward for not only the diphoton group, but many other groups with similar regression tasks.

> ”Sometimes science is more art than science, Morty. A lot of people don’t get that.”
>
> – Rick Sanchez

Rob Roy MacGregor Fletcher

University of Pennsylvania, July 2018

Chapter 1

Introduction

The discovery of new particle consistent with the Higgs boson was announced on July 4th, 2012. During the announcement, I was driving a truck through western Pennsylvania on my way to start graduate school at Penn. This was an exciting time in particle physics, and Penn had played a big part. The summer before classes began for me, I worked with Doug Schaefer, one of the Penn graduate students at that time, on the WW decay channel of the Higgs boson. Although I did not work on this for too long, it got me excited to finish classes and dig into more analysis.

This chapter will briefly describe some of the work that I did on ATLAS that does not fit well into other sections or does not merit its own section. This work is primarily being documented for my own benefit.

1.1 Electron ID

My first real work on ATLAS was in the Egamma group which is a combined pysics and performance group that focuses on the identification of prompt electons and photons. My focus was the identification of electrons. This was to be my service work on ATLAS. When I joined the group, the electron ID had been reworked by a few Penn students and had been changed from a cuts-based system, to a naive Bayes method called electron likelihood. At this time the likelihood approach was only used in the offline analysis; the cuts-based analysis had been used in the trigger. There was interest in moving the trigger over to an electron likelihood as well. This would remove some inefficiencies that arose from using 2 different IDs. Some electrons that passed the cuts ID would not pass the likelihood and vice versa. Work began to modify the likelihood to work in the trigger. For me this culminated in 2 sleepless weeks with another Penn student Joey Reichert and I trying to finalize new versions of the likelihood before the deadline to get them into the trigger.

1.2 Large Extra Dimension Models

After my service work was done, I transitioned into the exotic diphoton group. This group focuses on searches for exotic particles whose final states include two photons. In the beginning I was interested in graviton models that could be measured in the diphoton channel. The analysis I had started on was a search for Arkani-Hamed, Dimopoulos and Dvali (ADD) type gravitons [1] and Randall-Sundrum type gravitons [2]. These models are also sometimes known as Large Extra Dimension (LED) models as they propose an additional spatial dimension, x^5. If we consider a periodic extra dimension,

$$x^5 \sim x^5 + 2\pi R \tag{1.1}$$

where R is the radius of extra dimension, the momentum of a particle propagating in this dimension would be

$$p^5 \sim \frac{n}{R}, \quad n \in \mathbb{Z}. \tag{1.2}$$

Then we can write the equation motion of such a particle as,

$$\delta_\mu \delta^\mu \phi^n(x^\mu) = \frac{n^2}{R^2} \phi^n(x^\mu) \tag{1.3}$$

with mass,

$$m^2 = \frac{n^2}{R^2}. \tag{1.4}$$

Theories of this type with an extra dimension are known as Kaluza-Klein theories after Theodore Kaluza [3] and Oscar Klein[4]. These theories have been tested down to extremely small scales with current limits at $R \lesssim 10^{-21}$ meters. Extra dimensions of this size are pretty much hopeless to find. Arkani-Hamed, Dimopoulos and Dvali proposed that maybe this extra dimension is only accessible to gravity and, this causes the size of the extra dimension to only be bounded by tests of gravity, which means $R \lesssim 1$mm. One very interesting feature of theories like this is that since only gravity can propagate through this extra dimension, it can have the effect of "diluting" the effect of gravity in the 3 other space dimensions. This can possibly explain why there is such a large difference between the strength of the weak force and gravity. This difference is known as the hierarchy problem.

What does gravity look like in more than the 3 dimensions we are used to? The gravitational potential we all learn in our beginning physics classes can be written,

$$V(r) = G\frac{m_1 m_2}{r} \tag{1.5}$$

where G is the gravitational constant and r is the distance between masses m_1 and m_2. Using the relationship $M_{pl} \sim \sqrt{1/G}$ (where I'm leaving out factors of $\hbar c$ and π for simplicity) where M_{pl} is the Planck mass, we can write this as

$$V(r) \sim \frac{1}{M_{pl}^2}\frac{m_1 m_2}{r} \tag{1.6}$$

Now, if we add in n extra dimensions[1], we can use Gauss's law, considering $r << R$, to write the gravitational potential as,

$$V(r) \sim \frac{1}{M_D^{2+n}}\frac{m_1 m_2}{r^{1+n}} \tag{1.7}$$

We can see that if $n > 0$ for relatively small scales we should measure the potential deviating from the familiar $1/r$ law. What about at scales where gravity has been well measured ($r \gg R$)? In this case the potential can be written as,

$$V(r) \sim \frac{1}{M_D^{2+n} R^n} \frac{m_1 m_2}{r} \tag{1.8}$$

Comparing equations 1.6 and 1.8 we can see that this relates the 3-dimensional Planck mass that we measure, to the n-dimensional one where gravity is allowed to propagate in more dimensions.

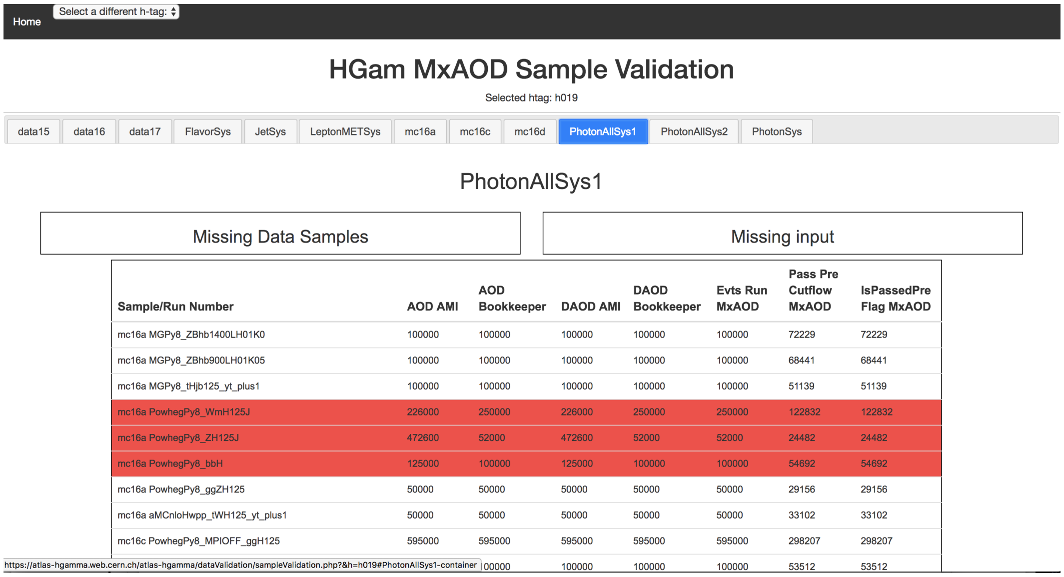
$$M_{pl}^2 \sim M_D^{2+n} R^n \tag{1.9}$$

Using this relationship we can begin to try to explain the hierarchy problem by setting the n-dimensional Planck mass to the weak scale ($\sim$ 1 TeV). This is saying that there really is no difference in the forces when considering n-dimensions, and we only see an apparent difference because we can only observe the 3-dimensional version of the Planck mass. What does this mean for the radius of the extra dimensions R? For $n = 1$ we get that $R \sim 10^{13} cm$. This is on the order of the size of the solar system! This definitely does not agree with experiment. For $n = 2$ we get that $R \sim 1mm$ which we said was around the smallest scale tests of gravity have been performed. So tests of gravity do not rule out extra dimension models with a number of extra dimensions $n = 2$ or greater.

The way that just adding extra spatial dimensions could very simply address the hierarchy problem was very interesting to me. I joined the search for large extra dimensions using data from the early part of Run2 at ATLAS. Results published on 2015 data showed a 3.4σ deviation from Standard Model backgrounds at 750 GeV, with CMS (the competing experiment to ATLAS) also showing a 3.9σ deviation [5, 6]. Although the threshold for discovery used in high energy physics is 5σ, both experiments seeing an excess at $\sim$ 750 GeV was enough to spark intense interest from the theoretical community. Around 200 papers were published within a month of the results being announced. During this time I participated in cross checks of the results [7].

1.3 HGam Data Validation

Early in my time with the exotics diphoton group, it was decided that this group should be merged with the Higgs to gamma gamma group (HGam) as their analysis was very similar and the background samples were largely the same. My first large project in the HGam group was to work on a framework that could perform validation on smaller datasets derived from the larger raw datasets collected by ATLAS. The validations primary goal was to ensure that no data was missing or incorrectly calculated in the derived data. Another Penn student, Tony Thompson, had begun this project and had produced very rough skeleton of a framework when he left the experiment. My main contribution to this was to write a script that automatically checked whether the proper data sets were created and if the information they contained matched the raw data. This was done in python and the outputs were written to JSON. The outputs were then used to build a dynamic website that contained easy to read tables with all of the relevant information so that it could be used as a reference for anyone in the group. Figure 1.1 shows a screen shot of the website. When the script completed running over the datasets it would also email a report summarizing any issues it found to the people responsible for creating the datasets. Although this project was not directly related to Higgs physics, it was very fun and one of the first times I had built a complete website that functioned dynamically based on underlying data.



HGam MxAOD Sample Validation

Selected htag: h019

PhotonAllSys1

Sample/Run Number	AOD AMI	AOD Bookkeeper	DAOD AMI	DAOD Bookkeeper	Evts Run MxAOD	Pass Pre Cutflow MxAOD	IsPassedPre Flag MxAOD
mc16a MGPy8_ZBhb1400LH01K0	100000	100000	100000	100000	100000	72229	72229
mc16a MGPy8_ZBhb900LH01K05	100000	100000	100000	100000	100000	68441	68441
mc16a MGPy8_tHjb125_yt_plus1	100000	100000	100000	100000	100000	51139	51139
mc16a PowhegPy8_WmH125J	226000	250000	226000	250000	250000	122832	122832
mc16a PowhegPy8_ZH125J	472600	52000	472600	52000	52000	24482	24482
mc16a PowhegPy8_bbH	125000	100000	125000	100000	100000	54692	54692
mc16a PowhegPy8_ggZH125	50000	50000	50000	50000	50000	29156	29156
mc16a aMCnloHwpp_tWH125_yt_plus1	50000	50000	50000	50000	50000	33102	33102
mc16c PowhegPy8_MPIOFF_ggH125	595000	595000	595000	595000	595000	298207	298207
[illegible]	[illegible]	00000	100000	100000	100000	53512	53512

https://atlas-hgamma.web.cern.ch/atlas-hgamma/dataValidation/sampleValidation.php?&h=h019#PhotonAllSys1-container

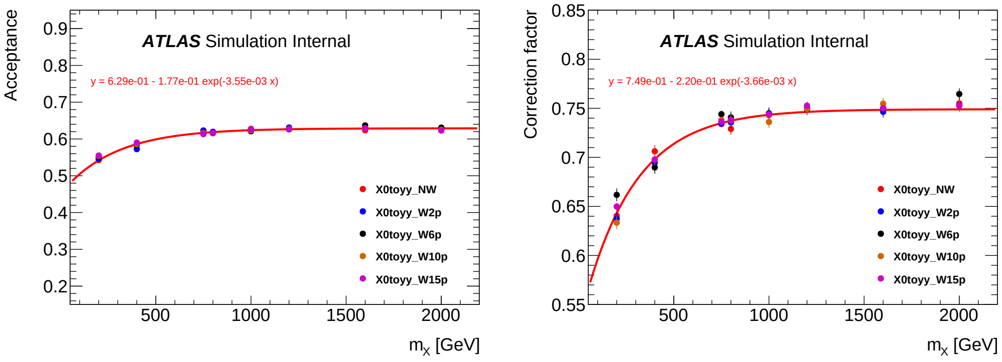
Figure 1.1: A screen shot of the HGam validation monitoring website. Every dataset is dynamically split on the page into tabs and rows of samples which can be colored red when an issue in that sample is detected.

1.4 High-Mass Diphoton

The high-mass diphoton search in the HGam group included searches for new scalar particles, non-resonant and resonant spin-2 signals in the mass range 200 GeV - 2.7 TeV [8, 9]. In order for the results to be reported in a way that is useful to theorists and for comparison to other experiments, they must be corrected to account for various detector and selection effects. For example, if we measure the cross-section of a particle by counting how many times we observe that particle and comparing that with the integrated luminosity collected, this will give incorrect results as some of the particles produced may follow a path that takes it through gaps in the detector, or regions of the detector that are used for cabling or support structure. Or maybe that particle had some kinematic configuration that meant it was rejected by our selection, or by our trigger due to technical or computing limitations. If a theorist wanted to compare our measurement to his or her model, they would need to know the actual cross-section, not just the part of it we can see. In order to correct these values we calculate acceptance and correction factors by using Monte Carlo (MC) simulation. To do this, we take generated MC and find the ratio of particles that pass our analysis selection, or that pass through the active volume of the detector to the total number generated. These ratios are shown in equation 1.10

$$C_X = \frac{N_{selection}}{N_{fiducial}} \qquad A_X = \frac{N_{fiducial}}{N_{total}} \tag{1.10}$$

Here $N_{selection}$ is the number of particles passing the applied selection cuts, N_{total} is the total number of particles generated in Monte Carlo, and $N_{fiducial}$ is the number of particles generated in Monte Carlo that pass through the fiducial volume of the detector. The definition of the fiducial volume of the detector is something that can take a lot of work on the part of analysts to accurately describe. For the sake of simplicity this can be thought of as a region of the detector where its response is well understood and where all components are functioning properly. In the high-mass diphoton search paper released in 2017 [8], I was responsible for calculating these so-called AxCx factors and their associated uncertainties. The factors and the fits used to model them are shown in figure 1.2 To calculate the uncertainties on these factors several modified samples are produced. These samples have some quantity modified away from its nominal value up 1σ and down 1σ. The AxCx factors are recalculated for each of these and their difference is fit with a function. This function is then used to model the uncertainty on the factors due to imperfect knowledge of the quantity being varied. Figure 1.3



(a) (b)

Figure 1.2: The acceptance factor (a) and the correction factor (b) for several resonance widths and at several masses of a scalar particle. The first part of the label Xtoyy refers to MC produced from a new particle X decaying to two photons. The last part of the labels are used to designate the decay width of the particles used in the MC samples, NW being the nominal narrow width sample.

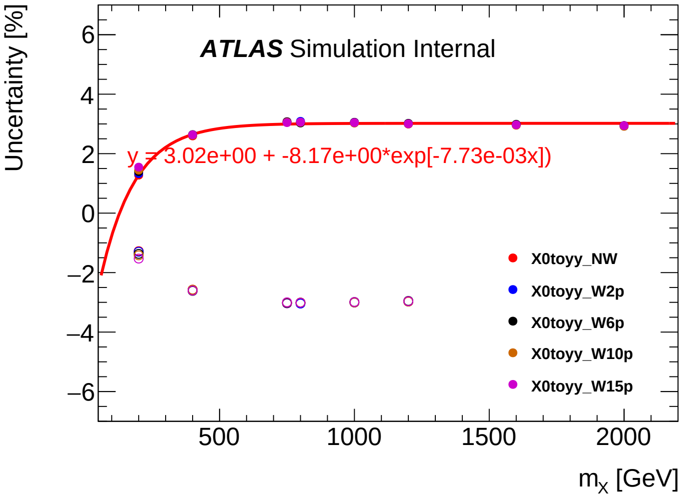


Figure 1.3: The uncertainty and fit function used for the photon ID efficiency. The closed circles represent variation of the nominal value up by 1σ. The open circles represent the variations down by 1σ.

Chapter 2

The Standard Model

The Standard Model of particle physics is the name given to theory used to describe matter and how it interacts. It has taken form over the course of several decades and began with separate theories. The first steps toward a unified picture of the fundamental particles and forces came in 1961 when Sheldon Glashow unified the weak and electromagnetic forces [10]. This led to the current description of the Standard Model that began to solidify in the 1970s with the experimental confirmation of quarks in deep inelastic scattering experiments conducted at the Stanford Linear Accelerator [11].

2.1 The Standard Model

The Standard Model as it is now described is a type of quantum field theory known as a gauge theory whose local transformations obey the symmetry groups,

$$SU(3)_C \times SU(2)_L \times U(1)_Y \tag{2.1}$$

when arranged in particular multiplet representations. Interactions between all of the matter particles are described by gauge bosons which appear as fields that must be added into the field content of the model in order to ensure local invariance under these transformations. This is done by the modification of the derivative to the covariant derivative which depends on these new gauge fields. The generators of the respective symmetry groups provide the description of the gauge fields responsible for the fundamental forces. The 8 generators of $SU(3)$ are the 8 gluon fields G_μ of the strong interaction, with the 3 generators of the $SU(2)$ group combined with the single generator of the $U(1)$ group giving the W_μ and B_μ fields responsible for the electroweak interactions.

The symmetry groups as written require that all particle content in the Standard Model be massless which is contrary to experiment where the fermions and weak bosons are observed to be massive. The solution to this problem was proposed through the Higgs mechanism developed by three groups: Peter Higgs [12, 13], Robert Brout and Francois Englert [14], and Gerald Guralnik, Carl R. Hagen, and Tom Kibble [15]. This mechanism breaks the electroweak symmetry through the addition of a scalar field with non-trivial vacuum.

$$SU(2)_L \times U(1)_Y \rightarrow U(1)_{EM} \tag{2.2}$$

After symmetry breaking the scalar field acquires a vacuum expectation value in the low-energy regime and the interaction with the fermionic fields and the gauge fields gives them mass. The W_μ and B_μ fields mix to give the physical $W^\pm$ and Z massive gauge bosons which are the mediators of the weak force. The $U(1)_{EM}$ symmetry is responsible for an orthogonal mixing to give the massless gauge field A_μ of the photon responsible for mediating the electromagnetic force.

The last piece of the standard model was added in the 1970s which described the interactions of the strong interaction in the theory of quantum chromodynamics (QCD). Quarks experience a phenomena known as confinement due to the fact that their interaction becomes stronger and larger distances. This posed a problem for experiment as it predicted that quarks could never be detected on their own but must exist in bound states with other quarks. Politzer, Wilczek and Gross [16, 17, 18] in 1973 showed the asymptotic freedom of

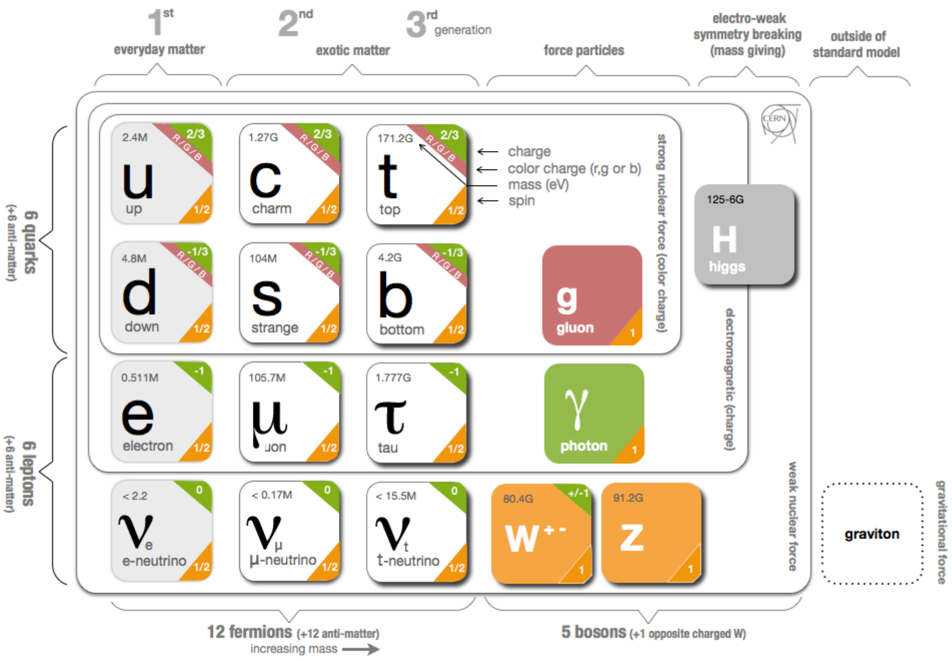


Figure 2.1: An illustration of the fields present in the Standard Model [19].

quarks in which their coupling gets weaker at high energies and thus QCD can be described perturbations.

The combination of the electroweak theory and quantum chromodynamics are what became known as the Standard Model of Particle Physics (SM). The field content of the SM is illustrated in figure 2.1.

2.2 Beyond the Standard Model

The predictions made by the Standard Model have agreed remarkably well with experiment. However, we know that the SM cannot be the only component in our description of physics. One of the most glaring shortcomings of the SM is the complete lack of any description of gravity. There have been many attempts at a quantum theory of gravity and unifications with the 3 other forces in the SM, but as of yet none have been completely successful. There are several other places where the SM fails to give any specific reason for some effect we know exists, or provides no prediction on an experimentally observed quantity. Endeavors to expand or modify the Standard Model are collectively known as Beyond the Standard Model.

2.2.1 Baryon Asymmetry

In all of the observable universe we see almost entirely matter. We do not seem to see hardly any antimatter. In the early universe these two should have been created in equal parts and experiment seems to tell us that they annihilate in equal parts. Something must have caused the universe to prefer one over the other. In 1967 Sakharov [20] proposed three conditions that must be met in order to have an asymmetry in baryonic matter in the universe. These are:

1. A baryon number violating process
2. Charge (C) and Charge-Parity (CP) violation
3. Interactions outside of thermal equilibrium

Baryon number (B) is obviously a requirement as if we assume initially $B = 0$ and in the later we observe B to be a large number, then this number cannot be conserved. There is no known process in the Standard Model that violates baryon number. C and CP violation both exist in the weak interactions but the amount of violation is not large enough to account for the remnants of the early universe that we observe today. Several extensions to the Standard Model address these in different ways. The two Higgs doublet model adds a second scalar doublet to the SM and allows for more places where CP violation can occur. Other models address the asymmetry in a more direct way, for example Darkogenesis [21] where the baryon asymmetry is generated through interaction with a "dark" sector of matter that can in effect transfer baryon number to the visible sector.

2.2.2 Dark Matter

Dark matter is another area where the Standard Model remains silent. We know from observations of galactic rotation that the luminous matter cannot account for all of the mass in a galaxy [22]. It is also observed that large regions of space bend light in an effect called gravitational lensing [23]. The standard model does not contain any particle that could exhibit all of the properties of dark matter, the most important of which is that it does not interact with photons (hence the name 'dark').

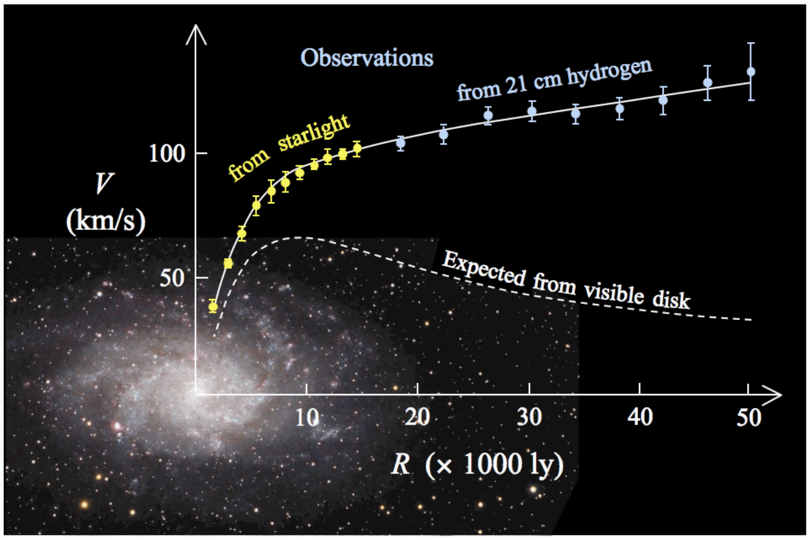


Figure 2.2: Rotation curve of the spiral galaxy M33. The yellow and blue points are the measured curve, while the dotted line is the one predicted from luminous matter only [24]

Extensions to the Standard Model need to add a particle that does not couple to photons, has a very small cross section to any other SM particle, has mass and is stable. Some extensions to the scalar sector of the SM contain what are called inert fields that do not interact with the rest of the SM. These massive bosons could then be a candidate for dark matter. Supersymmetric models also contain particles that can fill this role.

2.2.3 Supersymmetry

There are several search programs looking for new particles beyond the SM. The most popular of these programs is the search for Supersymmetry (SUSY), a type of space-time symmetry that supposes that all fermions have bosonic super parters, and all bosons have fermionic super partners [25]. This approximately doubles the particle content in the SM. In addition, at least one more scalar doublet field must be added. One of Supersymmetry's biggest draws is that it solves the Hierarchy Problem where divergent terms in the SM cancel against terms in SUSY. This also solves the problem of the unification of forces. Figure 2.3 shows the running couplings of the 3 forces in the SM, and with SUSY. As can be seen the SM couplings never completely unify as all three forces don't converge at a single point. However when including SUSY, the running couplings are modified and the forces do all unify at a point. Supersymmetry can also provide a candidate for dark matter as the couplings of SUSY particles to the SM are tightly constrained and so can have relatively stable light particles [26]. These particles could easily have couplings to the photon and other SM particles smaller than observed coupling limits on dark matter.

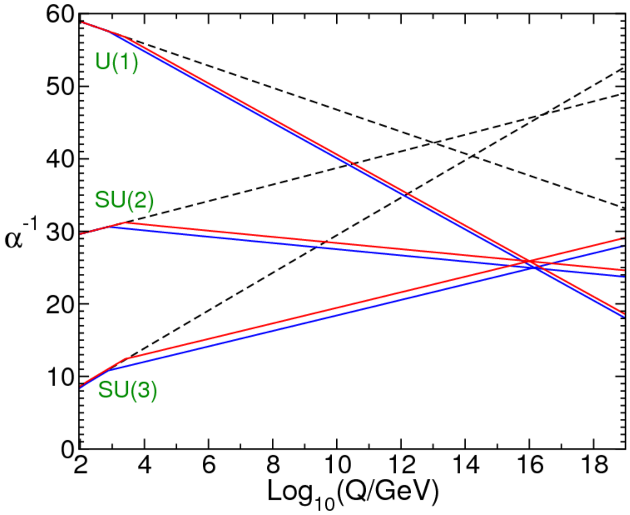


Figure 2.3: The running couplings of the forces in the SM shown with the dotted lines. With the addition of Supersymmetry, the forces have a point of unification at high energy scales, shown in red and blue [25].

Chapter 3

The LHC and ATLAS Detector

The Large Hadron Collider (LHC) is one of the largest, most complicated machines ever built by man. It has fascinated me for years and is one of the reasons that I chose to study high energy physics. The detectors built to take advantage of the collisions produced in the LHC are no less impressive and represent the cumulative work of thousands of physicists and engineers. This chapter will briefly introduce the LHC and the ATLAS detector. This is by no means a complete description. For more detail on the LHC see references [27, 28] and for ATLAS see references [29, 30, 31, 32].

3.1 The Large Hadron Collider

The LHC is a particle collider consiting of about 10,000 superconducting magnets. It accelerates protons around the 27 kilometer ring and collides them with a center of mass energy of $\sqrt{s} = 13$ TeV. The facility is built at European Organization for Nuclear Research (CERN) about 100 meters underground across the French-Swiss border. The first collisions were recorded in 2010 with a center of mass energy of $\sqrt{s} = 7$ TeV. Protons collide at four interaction points along the beamline where the four experiments sit. These are ATLAS and CMS, which are general purpose detectors, ALICE and LHCb which are detectors designed

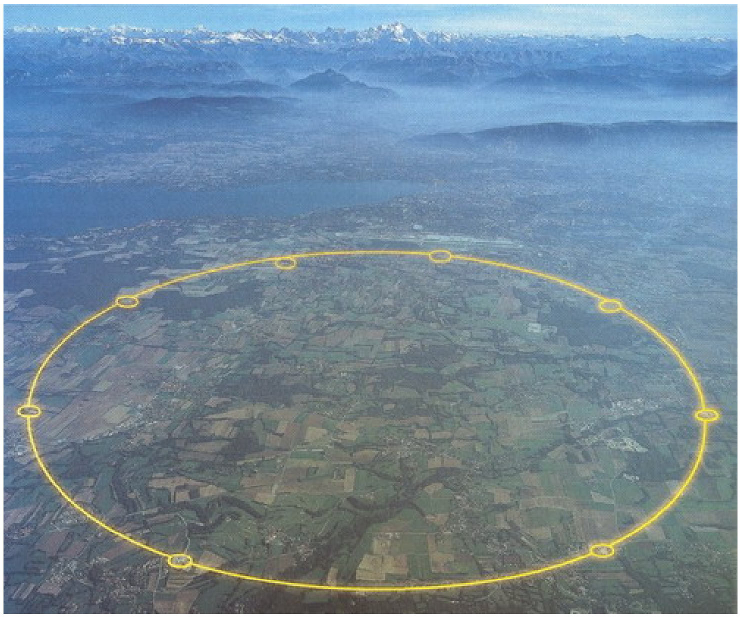


Figure 3.1: The location of the LHC tunnel at the French-Swiss border near Geneva. (Photo: CERN)

for heavy ion collisions and b-quark physics respectively. The LHC is the final stage of a series of accelerators that raise the beam energy until it can be directed into the main beamline. The CERN accelerator complex is shown in figure 3.2

Protons in the LHC travel around the ring in bunches of 1×10^{11} protons per bunch. They circulate at a rate of 50 Mhz (25ns bunch spacing) and each time bunches cross there are approximately 30 interactions under the current run conditions. A quantity called instantaneous luminosity is used to measure the rate that data is taken. Figure 3.3 shows the instantaneous luminosity per day for data taking in 2017. The integrated luminosity is how we measure the amount of data taken and is quoted in units of inverse femtobarns (fb^{-1}). These units make it easy to estimate the number of events we can expect for a given process with cross-section σ (measured in femtobarns). In the data taking period from 2015 to 2017 $\sim 80fb^{-1}$ were recorded with the ATLAS detector. Figure 3.4 shows the luminosity recorded by ATLAS per month for the years 2011 through 2018. Although such high luminosities come

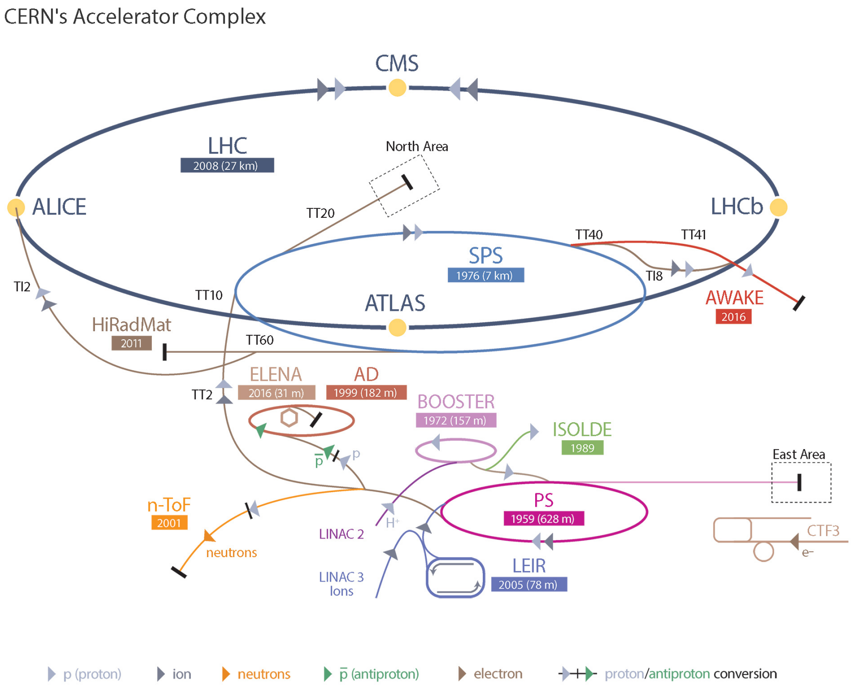


Figure 3.2: A schematic drawing of the CERN accelerator complex responsible for feeding protons into the main LHC beamline. (Image: CERN)

with challenges due to the rate of interactions, observation of extremely rare processes require that we run at these high rates.

3.2 The ATLAS Detector

The LHC houses two general purpose detectors; ATLAS and CMS. These detectors have been designed to accomodate a very broad physics program and serve as cross checks of eachother. In this section I will give a brief introduction to the design of the ATLAS detector with a focus on the detector subsystems that are relevant to electron identification.

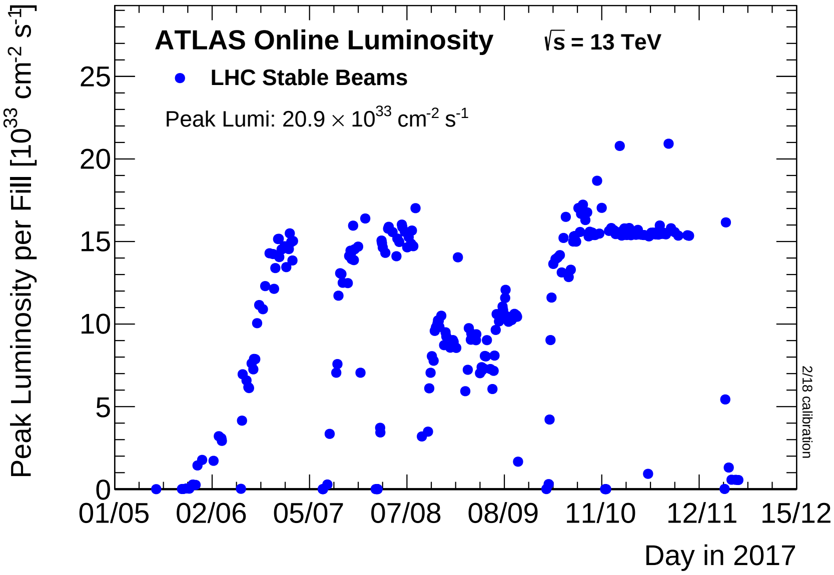


Figure 3.3: Peak luminosity by day for data taking in 2017 [33].

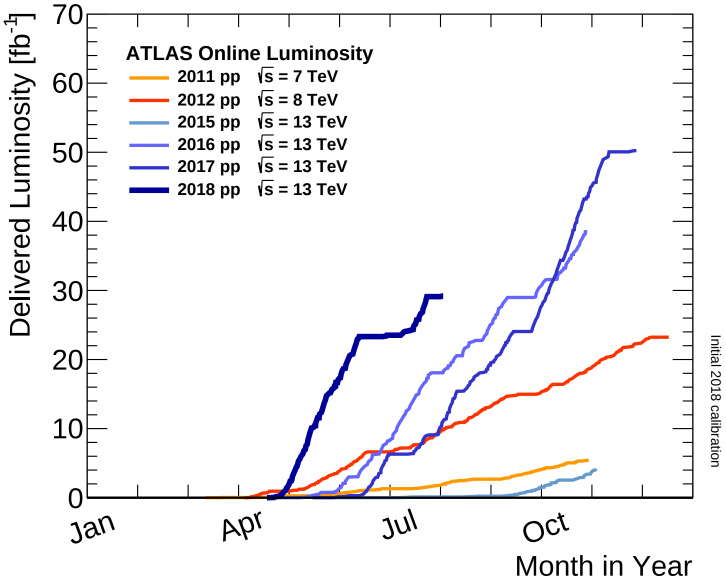


Figure 3.4: The luminosity per month for years 2011-2018 recorded by ATLAS [34].

3.2.1 Overview

The ATLAS detector is situated at one of the interaction points along the LHC that is on the CERN campus. It is located approximately 100m underground, is 25m high, 44m long and weighs around 7000 tons. Proton interactions take place at the center of the detector and are recorded as the particles created in the interaction escape outward. Figure 3.5 The ATLAS detector is composed of several subsystems that are designed to make differnt kinds of measurements. These can generally be split into 3 systems; trackers (collectively know as the inner detector), calorimeters, and muon spectrometer. These systems are arranged in a concentric pattern around the interaction point, and is made up of a barrel shaped section that is approximately uniform in the plane orthogonal to the beamline, and end cap sections that are circular and located on either end of the barrel.

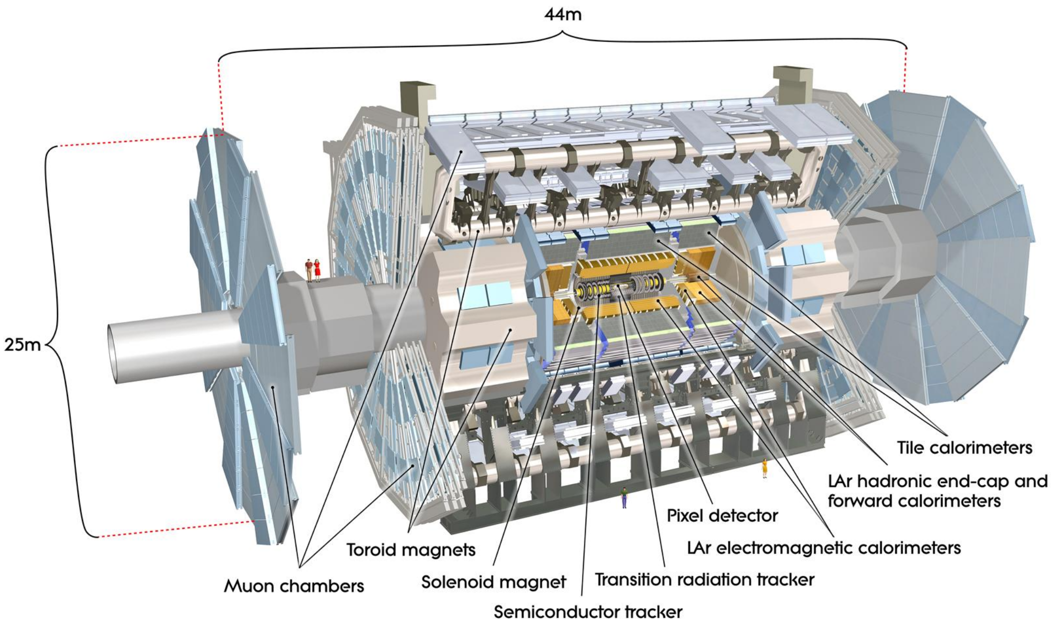


Figure 3.5: A cutaway view of the ATLAS detector. (Image: CERN)

3.2.2 Inner Detector

The inner detector is responsible for tracking particle paths as the pass through the detector. The inner detector is composed of 3 seperate subsystems; the pixel detector, silicon strip detector (SCT), and the transition radiation tracker (TRT). The arrangement of these subsystems is shown in figures 3.6 and 3.7.

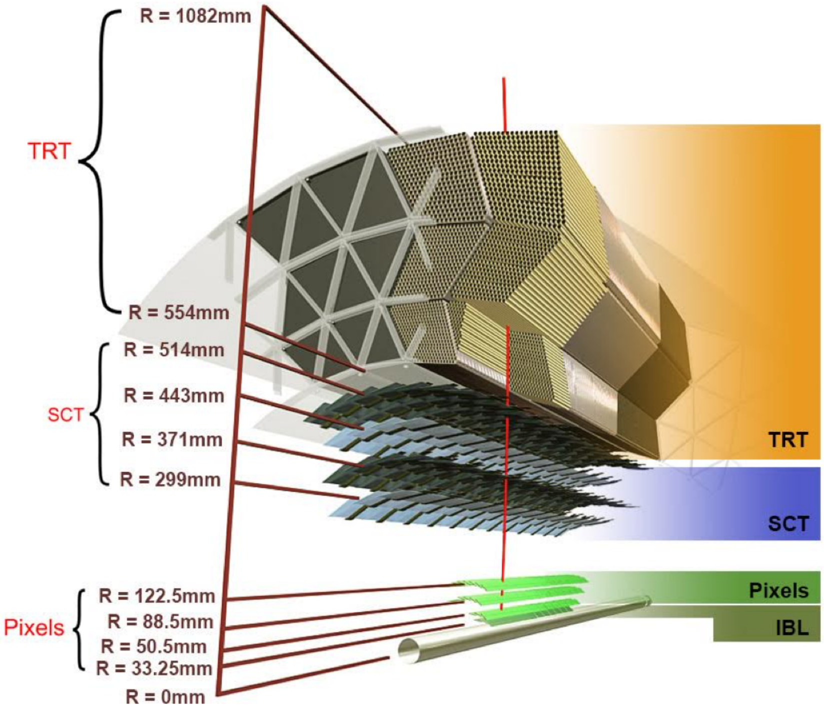


Figure 3.6: A cross section of the inner detector showing all layers of the subdetector. (Image: CERN)

The pixel detector consists of 80 million channels in 3 layers. Just outside of this is the SCT microstip detector with 4 layers. These two detectors both use silicon based tracking where charged particles passing through the layers causes current to flow and a corresponding point in space is recorded in data. The third layer works a bit differently to the previous two. The transition radiation tracker consists of straw tubes filled with xenon (or argon in

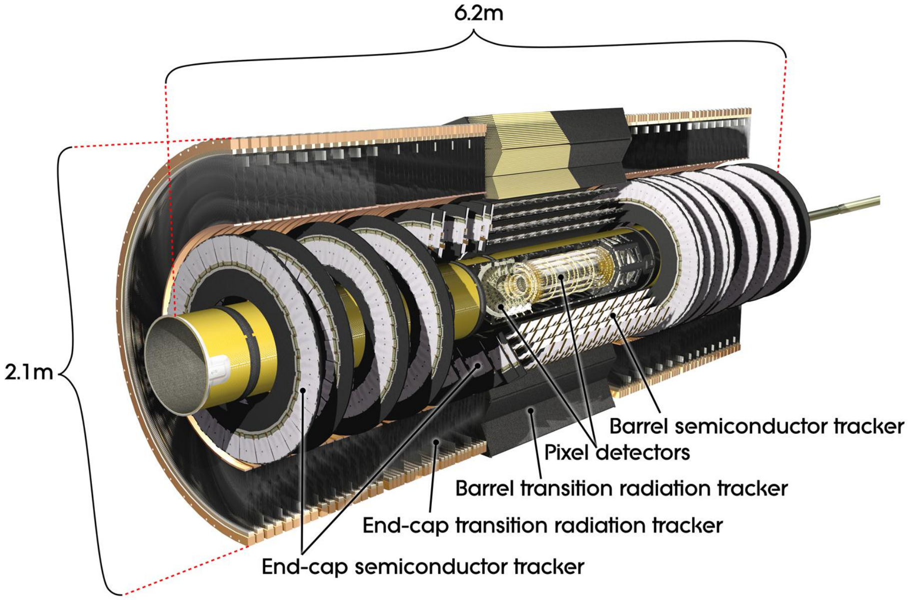


Figure 3.7: A cutaway of the inner detector showing the arrangement of all subsystems in relation to each-other. (Image: CERN)

some sections) and a wire suspended in the middle. When charged particles pass through the straw tubes the gas inside is ionized and the ions drift to the walls of the straw tube where the total charge collected is recorded. This can serve a tracking function as there are 50,000 straws in the barrel section and the path of the particle can be reconstructed by finding the line of straws that have signal in them. The reason for this design comes from the transition radiation (TR) emitted by particles interacting with detector. Lighter charged particles will emit more transition radiation than heavier ones. This allows the use of the TR to help distinguish electrons from other particles and is a valuable input for electron identification. These three trackers are housed inside of a solenoid magnet with 2 Tesla field. This field causes charged particle tracks to bend and by calculating the radius, the momentum and charge of the particles can be determined.

3.2.3 Electromagnetic Calorimeter

Calorimeters are used to measure the energy deposited in them as particles pass through, or stop in them. The first of these is the Liquid Argon Calorimeter (LAr) also known as the electromagnetic calorimeter (EMcal) [35]. It uses liquid argon as a scintillator and accordion shaped electrodes arranged in 3 layers. The layers are segmented differently to measure the shape of electromagnetic showers in the material. The first layer is very finely segmented in the η direction, the second and deepest layer is segmented evenly in the η and ϕ direction. The third layer is the most shallow and has larger segmentation than the other two. A slice of the EMcal is shown in figure 3.8. The total depth of the EMcal is ~ 22 radiation lengths so most showers should be entirely contained within this subsystem. This calorimeter is the primary system used in electron and photon identification.

3.2.4 Hadronic Calorimeter

Located just outside of the EMcal is the Hadronic Calorimeter, also called the tile calorimeter (tileCal). Each of 64 modules in the barrel are composed of alternating layers of iron plates and scintillating tiles. As the name suggests, this calorimeter is designed to measure the energy of hadronic particles. The tiles sample the hadronic shower and based scintillating measured the total energy in the shower can be calculated.

3.2.5 Muon Spectrometer

Along with neutrinos, muons are one of the only particles that will normally completely escape from the detector. While other particles are completely stopped in the calorimeters, muons will pass through the outer layers. The muon spectrometer consists of chambers that measure the momentum of muons. The system uses a series of 4 Tesla toroid magnets.outer

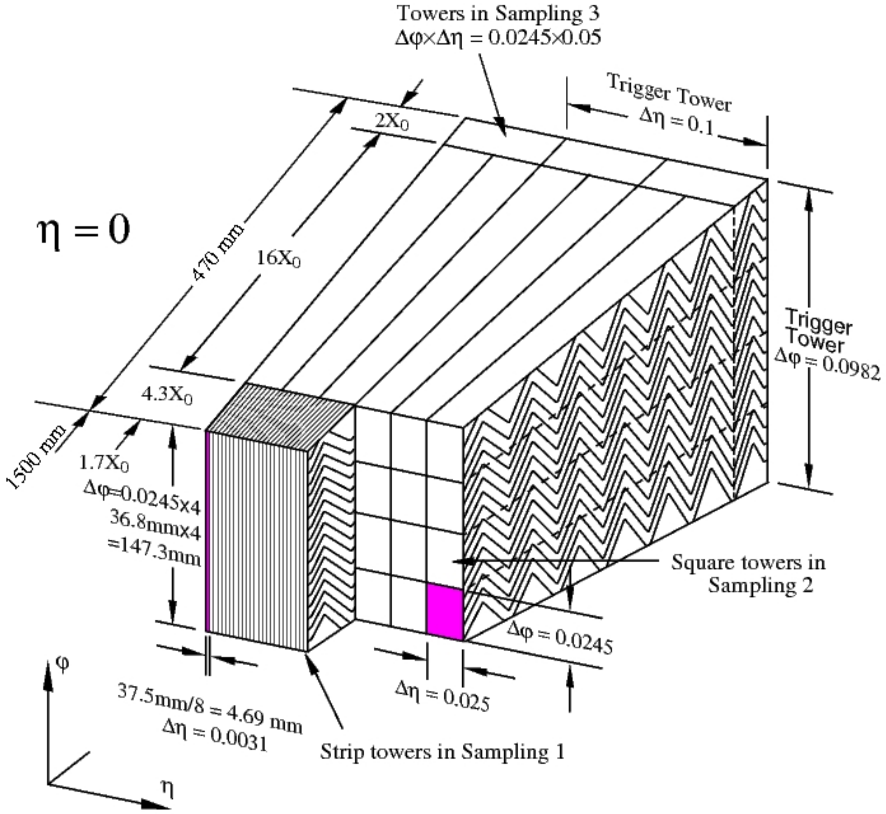


Figure 3.8: A small slice of the ATLAS Electromagnetic Calorimeter system showing how each of the 3 layers is segmented [36].

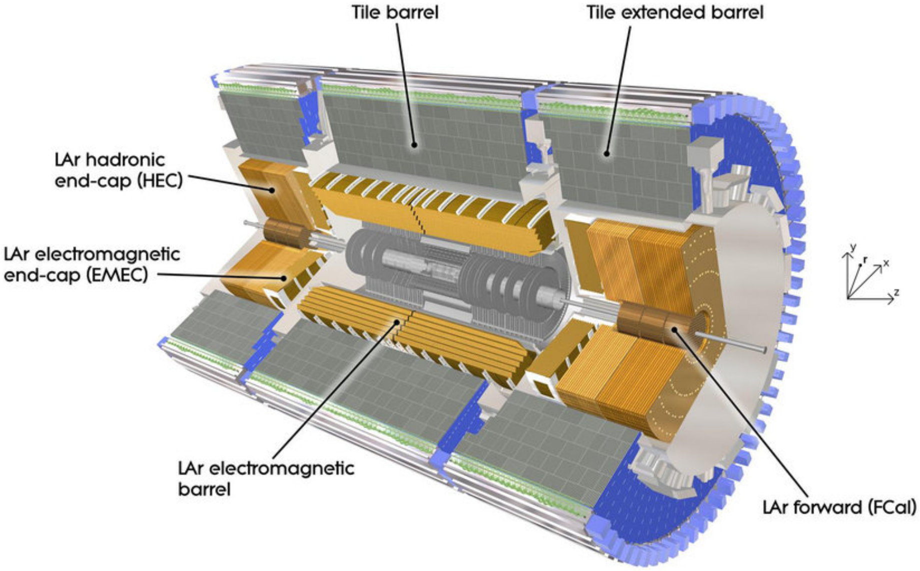


Figure 3.9: A cutaway view of the calorimeter system in the ATLAS detector. The EM calorimeter and Hadronic calorimeter are shown [37].

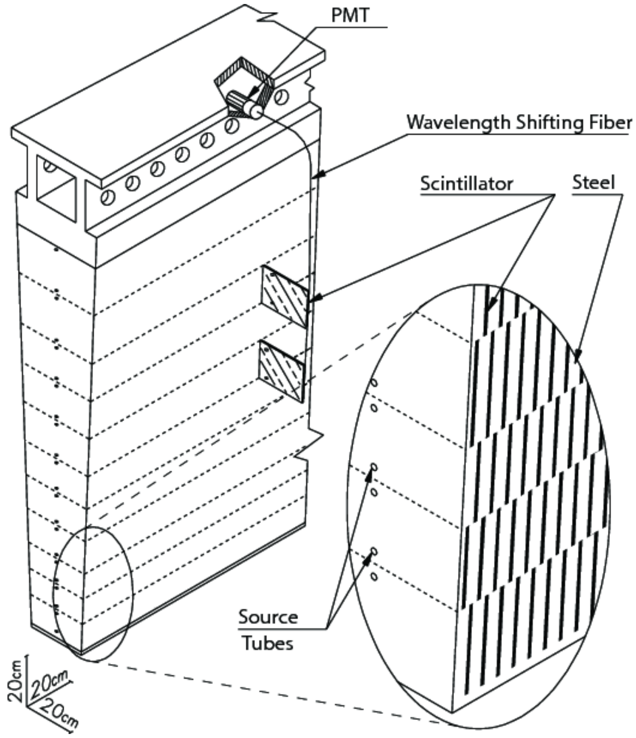


Figure 3.10: One barrel module of the hadronic calorimeter showing the arrangement of iron and scintillating tiles [38].

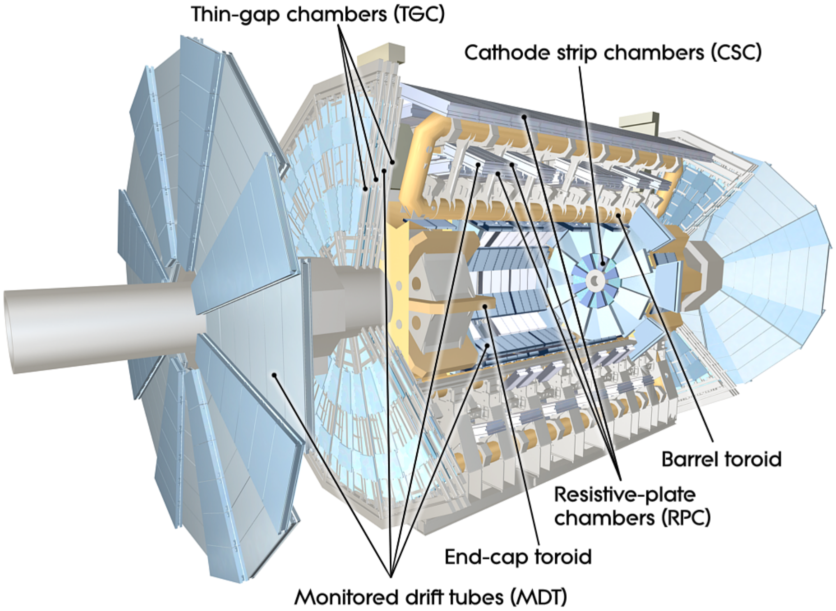


Figure 3.11: A cutaway view of the ATLAS muon spectrometer [39].

Chapter 4

Electron Identification

Electrons are an important part of many ATLAS analyses. Whether directly looking for decays involving electrons or identifying them to veto events, electron ID finds it way into use in a large number of groups. The primary instruments used in ID are the pixel and silicon-strip tracking systems, transition radiation tracker, EM calorimeter and to a lesser extent the hadronic calorimeter. A number of quantities (variables) are measured in each of these subsystems are used to differentiate electrons and photons from other particles. In the past a cuts-based ID was used where each variable is subdivided into an accept and reject region. If an electron candidate lies in the accept region, this process is repeated for the next variable. This continues until all of the variable distributions have been cut on. The candidates that pass all of these cuts are then considered to have passed some level of electron identification. The line dividing the accept and reject regions can be adjusted to control the efficiency of identification for what we would call an operating point. Although this method is simple and provided a fairly small rate of false positive, with the increasing demands placed on the detector due to higher luminosity, a better method was sought. The Electron Likelihood was selected to replace the cuts based method [40] [41].

My work on ATLAS began in the Egamma group with the Electron Likelihood. While working on electron ID, I contributed almost all areas of the Likelihood ID, but a majority of my work involved preparing tunes, or the set of inputs used to define the ID operating points for the ID. At this time, due to the success of the Likelihood used on offline data, it was also decided to be adapted for use in the trigger ID algorithms as well. I was involved in both the software changes and modifications to the method required by running in the online environment. In this chapter, I will briefly discuss electron reconstruction and the electron likelihood method.

4.1 Electron Reconstruction

Electron reconstruction is the process of creating electron candidates from the raw information collected by the triggering system. This is performed in several steps.

- **Seed-cluster reconstruction:** Energy deposits in the calorimeter are grouped into clusters. This is done with a sliding window algorithm summing the transverse energies in three layers of the EM calorimeter within an $\eta \times \phi$ region of 0.075×0.125.
- **Track reconstruction:** The path through the tracking volume is calculated creating a track.
- **Electron specific track refit:** Tracks are matched to a cluster. The η and ϕ of the track are extrapolated into the second layer of the EM calorimeter. If this track is close enough to a cluster it is refit with a more advanced track fitter.
- **Electron candidate reconstruction:** The track matching of the previous step is redone with stricter criteria.

Because photons will look the same as electrons in the EM calorimeter, tracking criteria must be used to remove photons. Electron candidates that do not have a track associated to them are removed and considered photons. Electrons whose tracks are reconstructed to a secondary vertex are also considered to be photons and are removed. Some electron candidates, based on their E/p, their p_T and if there is an associated pixel hit can be classified as ambiguous, or unambiguous. This determines if it will be considered as only an electron, or if it will also be included in the photon container.

After all of these steps, some additional quantities are computed, such as the four-momentum, and the electron candidates have their energies calibrated.

4.2 Electron ID

4.2.1 Variables used for Electron ID

The various subsystems in the ATLAS detector can be used to make many measurements. The measurements used in the electron and photon ID are referred to as variables. These are described in table 4.1. A schematic drawing of the path of an electron through a section of the detector is depicted in figure 4.1

The b-layer, silicon strips (SCT) and TRT all provide space point hits for tracking while the TRT also provides a number of high threshold hits that can be used to distinguish electrons from other particles that have similar detector signatures like pions. Tracking variables d_0 and $|d_0/\sigma_{d_0}|$ are measurements of the impact parameter, or the distance away from the beamline the particle originates from. These help to distinguish electrons from b-jets and c-jets that tend to have secondary vertices far from the beamline.

The TRT helps to discriminate electrons from heavier particles through the use of transition radiation. Lighter particles that have higher γ-factors will produce more photons when passing through detector material than heavier particles. This is measured with the variable F_{HT} which is the ratio of high threshold hits to the total number of hits in the TRT. This tends toward higher values for electrons than for heavier particles.

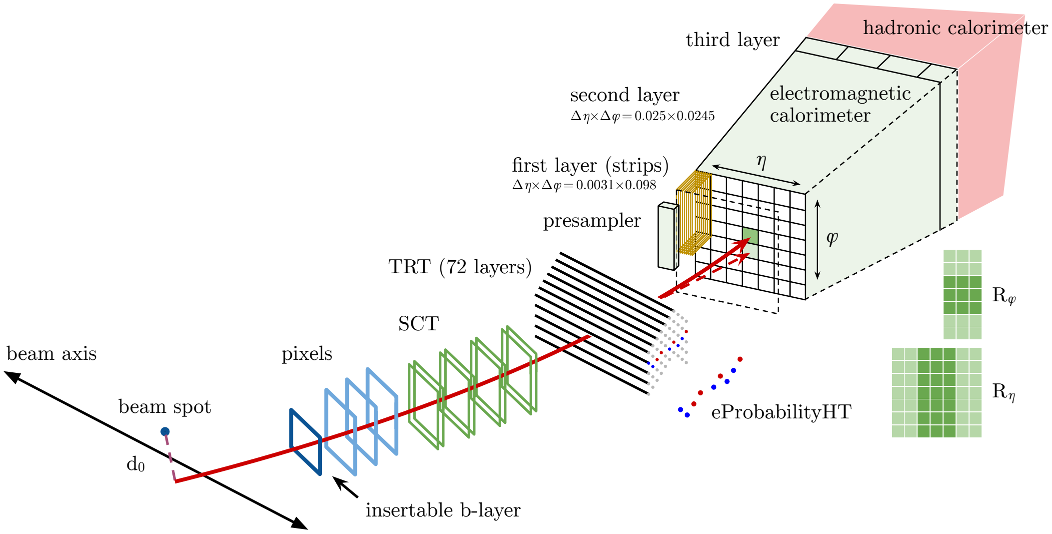


Figure 4.1: A schematic drawing of a section of the ATLAS detector. The red line shows the path an electron takes through each subsystem [41].

The three main layers of the EM calorimeter provide shower shape information at different granularities. EM objects should not have showers longer than the EM calorimeter, so any energy deposited behind it into the hadronic calorimeter can be used to discriminate between electrons and light/heavy flavor objects. This is measured with the R_{had} variable which is the ratio of transverse energy in the hardronic calorimeter to the EM calorimeter. The f_1 and f_3 variables measure the ratio of the energy in the first and third layers respectively to the total energy in the EM calorimeter. These allow transverse characteristics of the shower to be measured. The width the of the showers are measured with variables w_{stot}, w_η, R_η and R_ϕ. Electrons tend to have very narrow showers compared with heavier hadrons.

4.2.2 Electron Likelihood

The Electron Likelihood method developed to replace the cuts based ID is a naive Bayes classification method using probability density functions (PDFs) to determine the probability that a candidate is an electron. PDFs for both signal and background are created and are used

Type	Description	Name	Rejects			Use
			LF	γ	HF	
Hadronic leakage	Ratio of E_{T}in the first layer of the hadronic calorimeter to $E_{\mathrm{T}}$of the EM cluster. (used over the range $\lvert\eta\rvert < 0.8$ or $\lvert\eta\rvert > 1.37$)	R_{had1}	x		x	$\mathcal{L}$
	Ratio of E_{T}in the hadronic calorimeter to $E_{\mathrm{T}}$of the EM cluster. (used over the range $0.8 < \lvert\eta\rvert < 1.37$)	R_{had}	x		x	$\mathcal{L}$
Back layer of EM calorimeter	Ratio of the energy in the back layer to the total energy in the EM accordion calorimeter. This variable is only used for $p_{\mathrm{T}} < 80$ GeV due to known inefficiencies at high p_{T}, and is also removed from the LH for $\lvert\eta\rvert > 2.37$, where it is poorly modeled by the MC.	f_3	x			$\mathcal{L}$
Middle layer of EM calorimeter	Lateral shower width, $\sqrt{(\Sigma E_i \eta_i^2)/(\Sigma E_i) - ((\Sigma E_i \eta_i)/(\Sigma E_i))^2}$, where E_i is the energy and η_i is the pseudorapidity of cell i and the sum is calculated within a window of 3×5 cells	w_{η}	x	x		$\mathcal{L}$
	Ratio of the energy in 3×3 cells over the energy in 3×7 cells centered at the electron cluster position	R_{ϕ}	x	x	x	$\mathcal{L}$
	Ratio of the energy in 3×7 cells over the energy in 7×7 cells centered at the electron cluster position	R_{η}	x	x	x	$\mathcal{L}$
Strip layer of EM calorimeter	Shower width, $\sqrt{(\Sigma E_i (i - i_{\max})^2)/(\Sigma E_i)}$, where i runs over all strips in a window of $\Delta\eta \times \Delta\phi \approx 0.0625 \times 0.2$, corresponding typically to 20 strips in η, and $i_{\max}$ is the index of the highest-energy strip	w_{stot}	x	x	x	C*
	Ratio of the energy difference between the maximum energy deposit and the energy deposit in a secondary maximum in the cluster to the sum of these energies	δE_{max}	x	x		$\mathcal{L}$
	Ratio of the energy in the strip layer to the total energy in the EM accordion calorimeter	f_1	x			$\mathcal{L}$
Track conditions	Number of hits in the innermost pixel layer; discriminates against photon conversions	n_{Blayer}		x		C
	Number of hits in the pixel detector	n_{Pixel}		x		C
	Number of total hits in the pixel and SCT detectors	n_{Si}		x		C
	Transverse impact parameter with respect to the beam-line	d_O		x	x	$\mathcal{L}$
	Significance of transverse impact parameter defined as the ratio of d_O and its uncertainty	$\lvert d_0/\sigma_{d_0}\rvert$		x	x	$\mathcal{L}$
	Momentum lost by the track between the perigee and the last measurement point divided by the original momentum	$\delta p/p$	x			$\mathcal{L}$
TRT	Likelihood probability based on transition radiation in the TRT	eProbHT	x			$\mathcal{L}$
Track-cluster matching	$\Delta\eta$ between the cluster position in the strip layer and the extrapolated track	δ_{η_1}	x	x		$\mathcal{L}$
	$\Delta\phi$ between the cluster position in the middle layer of the calorimeter and the momentum rescaled track extrapolated from the perigee	$\delta\phi_{res}$	x	x		$\mathcal{L}$
	Ratio of the cluster energy to the track momentum	E/p	x	x		C*

Table 4.1: Definitions of electron discriminating variables, the types of backgrounds the variables help to discriminate against, and if a variable is used as a likelihood PDF ($\mathcal{L}$) or used as a rectangular cut (C). The * refers to the fact that the E/p and w_{stot} variables are only used for electrons with $p_{\mathrm{T}} > 150$ GeV for the *Tight* identification operating point (in software release 20.7), and are not used for the looser operating points.

in combination to calculate a discriminant value. A cut is then placed on this discriminant to determine a desired signal efficiency or background rejection.

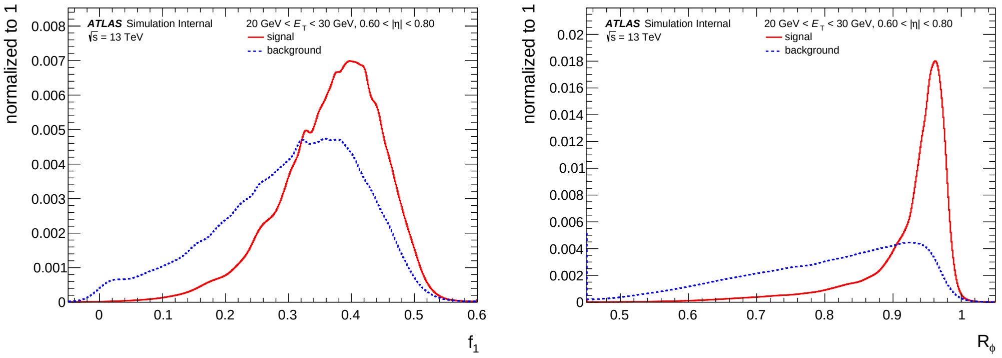
The likelihood is constructed by making a set of PDFs from a sample of electrons (signal) and electron fakes (background) for each of the n variables to be used. Each electron candidate to be classified has a set of the same n variables which we will call x. The value from each signal (background) PDF associated with the variables x is calculated and their product is taken to get the signal (background) likelihood value. These are then combined to get the likelihood discriminant value according to equation (4.1).

$$d_{\mathcal{L}} = \frac{\mathcal{L}_S}{\mathcal{L}_S + \mathcal{L}_B}, \qquad \mathcal{L}_S(x) = \prod_{i=1}^{n} P_{S,i}(x_i) \tag{4.1}$$

This distribution for signal (background) tends to have a compressed peak at 1 (0) and so for finer control over the operating point a transformation is applied to the discriminant (4.2). This transformation broadens the peaks so the shape is more defined but still maintains signal and background separations.

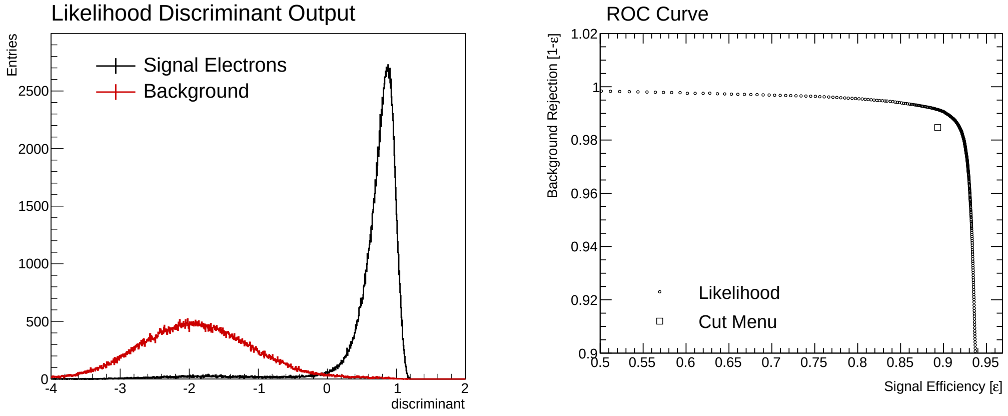
$$d'_{\mathcal{L}} = -\tau^{-1} d_{\mathcal{L}} ln(d_{\mathcal{L}} - 1), \qquad \tau = 15 \tag{4.2}$$

More advanced methods such as the electron likelihood have the advantage that they take into consideration multiple distributions at once, unlike the cuts method that looks at only one variable at a time. Figure 4.2 shows the signal and background distributions for the variables f_1 and R_ϕ. The large overlap of the signal and background causes any cut made on these variable to have poor background rejection. In some situations the variable might not be used at all as it adds no power to the method. With a likelihood, so long as there is a difference in the shape, it can contribute to the power of the classification. Figure 4.3 shows the output discriminant of an example likelihood after transformation and the ROC curve corresponding to the continuum of operating points representing cuts on the discriminant. The good separation between signal and background can be seen.



(a) (b)

Figure 4.2: Distribution of the variables f_1 and R_ϕ for signal(red) and background(blue). The large overlap of the signal and background of variables like this cause the cuts method to have poor background rejection.



(a) (b)

Figure 4.3: An example likelihood discriminant output (a), after having been transformed by Equation 2, for data signal and background distributions. The corresponding ROC curve is shown in (b), illustrating the continuum of operating points. A cut-based menu is plotted for comparison.

A "tuned" likelihood consists of a set of PDFs, for each variable used in the likelihood and for each *eta* and *phi* bin used, and values of a cut on the discriminant for each desired operating point. The shape of the PDFs vary as a function of p_T and η so they are divided into 10 p_T and 9 η PDFs each with their own set of discriminant cuts. Tables 4.2 and 4.3 show the binning of the PDFs and discriminant cuts in p_T and η. In addition to the discriminant cuts, one or more hits in the pixel tracker as well as seven or more SCT hits are required.

	Bin boundaries in p_{T}[GeV]										
PDFs	4.5	7	10	15	20	30	40	∞			
Discriminant cut values	4.5	7	10	15	20	25	30	35	40	45	∞

Table 4.2: Electron transverse energy binning used for the electron likelihood PDFs and discriminant cut values.

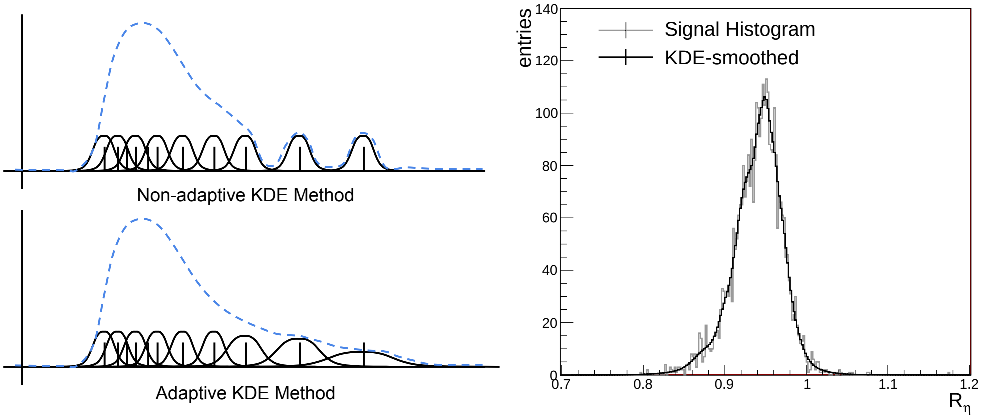
Bin boundaries in $\lvert\eta\rvert$									
0.0	0.6	0.8	1.15	1.37	1.52	1.81	2.01	2.37	2.47

Table 4.3: Electron pseudorapidity binning used for the electron likelihood PDFs and discriminant cut values.

The discriminant values chosen are set to roughly match the efficiency of the 3 operating points (loose, medium, tight) used in the Run 1 offline likelihood. A fourth operating point was added (veryloose) with relaxed criteria compared to the loose operating point. This can be used to to study backgrounds that are similar to signal electrons.

4.2.3 Probability Density Functions

PDFs can be obtained by making histograms of the variables used in the likelihood from signal and background samples. Finite statistics and the division of samples by p_T and η mean that the PDFs produced will not be smooth. This can lead to undesirable behavior in the likelihood as electrons in adjacent bins could have large differences in their respective discriminants. In the tails of the PDF we also tend to have bins with 0 entries. In order to avoid undefined results, the PDFs should also be greater than zero everywhere. To correct this, we smooth the PDFs using an adaptive kernel density estimator (KDE). This takes each bin in the histogram and replaces it with a Gaussian distribution (the kernel) whose height and width are determined by the content of the bin. The difference between a non-adaptive and adaptive KDE is that Gaussian widths are variable in the adaptive case. This has the effect of narrowing the Gaussian kernel with increasing bin content as bins with high statistics have less uncertainty. The KDE technique is illustrated in figure 4.4 along with an example PDF before and after smoothing.



(a) (b)

Figure 4.4: An example likelihood discriminant output (a), after having been transformed by Equation 2, for data signal and background distributions. The corresponding ROC curve is shown in (b), illustrating the continuum of operating points. A cut-based menu is plotted for comparison.

4.2.4 Correcting MC variable mis-modeling

One big advantage of the electron likelihood is that it can be tuned using data. A sample of electrons are gathered using the tag-and-probe method detailed in [42]. This sample will be a relatively unbiased, pure sample of data electrons with which to build PDFs. At the beginning of Run 2, however, no data samples were available for the run conditions and center of mass energy the LHC was preparing for. This meant that we were forced to use MC to build the PDFs and tune the likelihood. These samples have some known mis-modeling effects in the form of shifts and in a few cases width scaling. This can be corrected however by comparing the MC variable distributions to $\sqrt{s} = 8$ TeV data and shifting or modifying the width of the distributions. Because of the difference in run conditions, the 8 TeV data will not exactly represent the variable shapes for 13 TeV data, but the shifts and widths will be close enough to use until data PDFs can be made with 13 TeV data.

The shifts are implemented as constant offsets of the mean position of the distributions. A comparison of data and MC showed that the variables which needed shifting corrections were f_1, f_3, R_η, w_{η^2} and R_ϕ. Width correction is performed by modifying the full width at half max (FWHM), which the variables $\Delta\eta_1$, $\Delta\phi_{res}$ and R_{had} were found to need.

The operation to apply the shifts is performed on an electron-by-electron basis using $v^*_{MC} = v_{MC} - a$ for some offset a. To modify the widths the operation $v^*_{MC} = (v_{MC} - \bar{v}_{MC}) * w + \bar{v}_{MC}$ for a width parameter w and where $\bar{v}_{MC}$ is the mean of the distribution. The parameter a is found by minimizing a χ^2 test statistic,

$$\chi^2 = \sum_{bins} \frac{(n_{data} - n_{MC})^2}{\sigma_{data} + \sigma_{MC}}. \tag{4.3}$$

The parameter w is selected by taking the ratio of the FWHM of data to MC. Figure 4.5 shows the MC distributions before and after the corrections compared with the data distributions used.

While working in the Egamma group I was responsible for calculating all parameters of the shifts used to correct the MC. To make the corrections available to the entire group, I wrote a tool that could be used to automatically correct all relevant distributions based on matrices of parameters that were kept updated in the tool. This was integrated into the official Egamma software package.

4.2.5 Signal and Background PDFs

The results of the KDE smoothing and MC correction procedure are shown in figures 4.6 and 4.7 for signal and background. The PDFs shown are taken from the $20GeV < p_T < 30GeV$, $0.6 < |\eta| < 0.8$ bin.

4.3 Pileup Correction

The LHC is continually trying to increase the luminosity of the collider. This allows more data to be collected in a shorter amount of time. The increase in luminosity means that the average number of interactions per bunch crossing also increases. In 2016 the average was ∼25 interaction, up from ∼12.5 in 2015. The additional vertices in each event contribute to what is known as pileup. Particles from completely different interaction vertices can overlap, depositing energy in the same regions of the detector. Some ID variables are particularly effected by this and their shapes, in general, become broader. R_η and R_{had} have the strongest dependence on pileup as seen in figure 4.8. When these distributions become wider, their PDFs become more background like. For events with a lot of pileup, this causes a drop in the efficiency to identify electrons. R_η and R_{had} are among the variables with the highest discriminating power as can be seen in figure 4.9 and so they cannot just be removed from the likelihood.

We would like to correct for this drop in efficiency. We can do this by making the discriminant cut linearly dependent on the number of primary vertices (n_{vtx}), $d(n_{vtx}) = d_{\mathcal{L}} + a\dot{n}_{vtx}$.

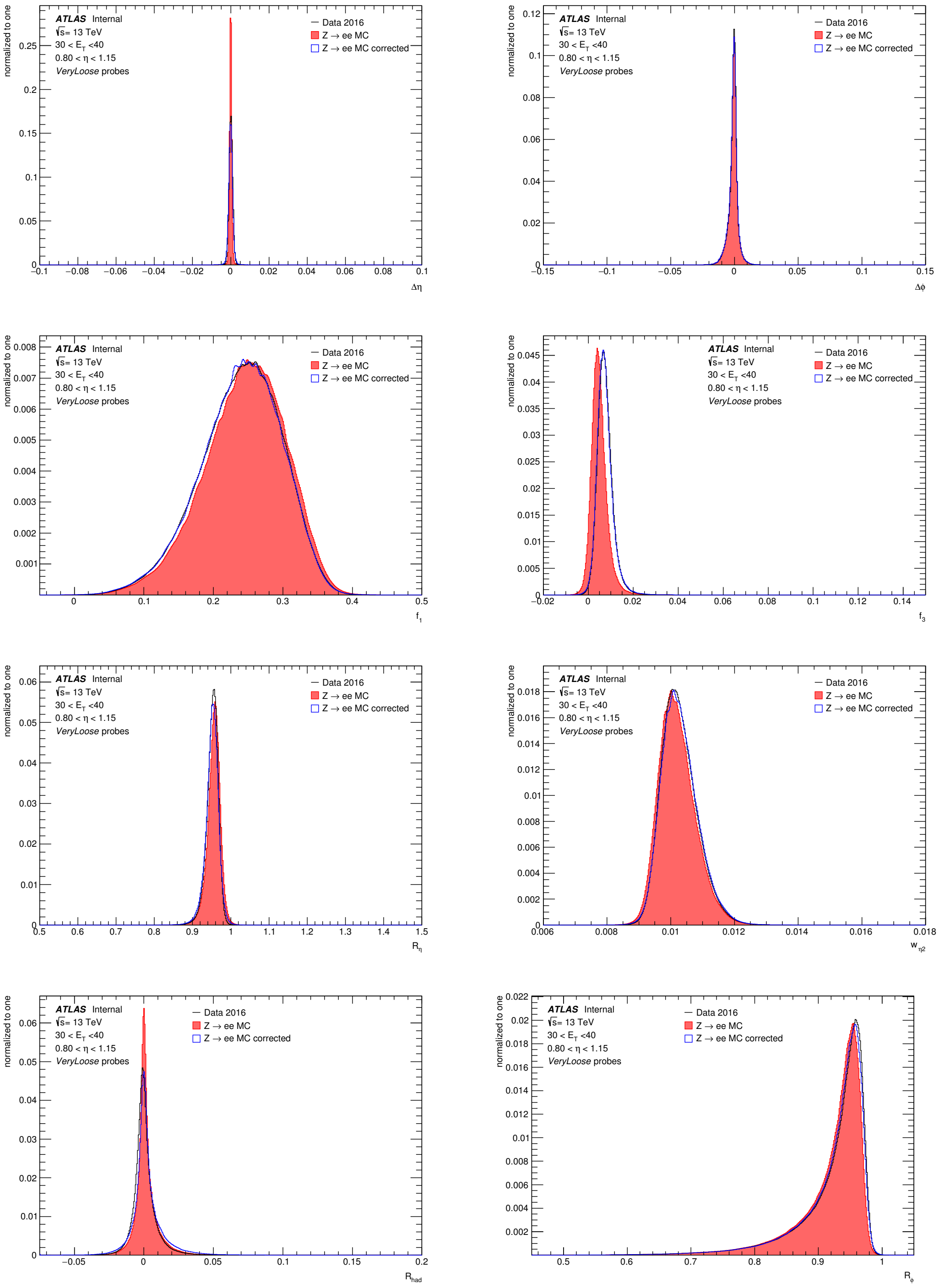


Figure 4.5: Data and MC electron variable distributions obtained using the $Z \to ee$ tag-and-probe method, and in the $30 \leq E_{\mathrm{T}}$¡ 40, $0.80 \leq \eta$¡ 1.15 bin. MC is shown before and after applying the constant shift and width corrections described in the text. KDE smoothing has been applied.

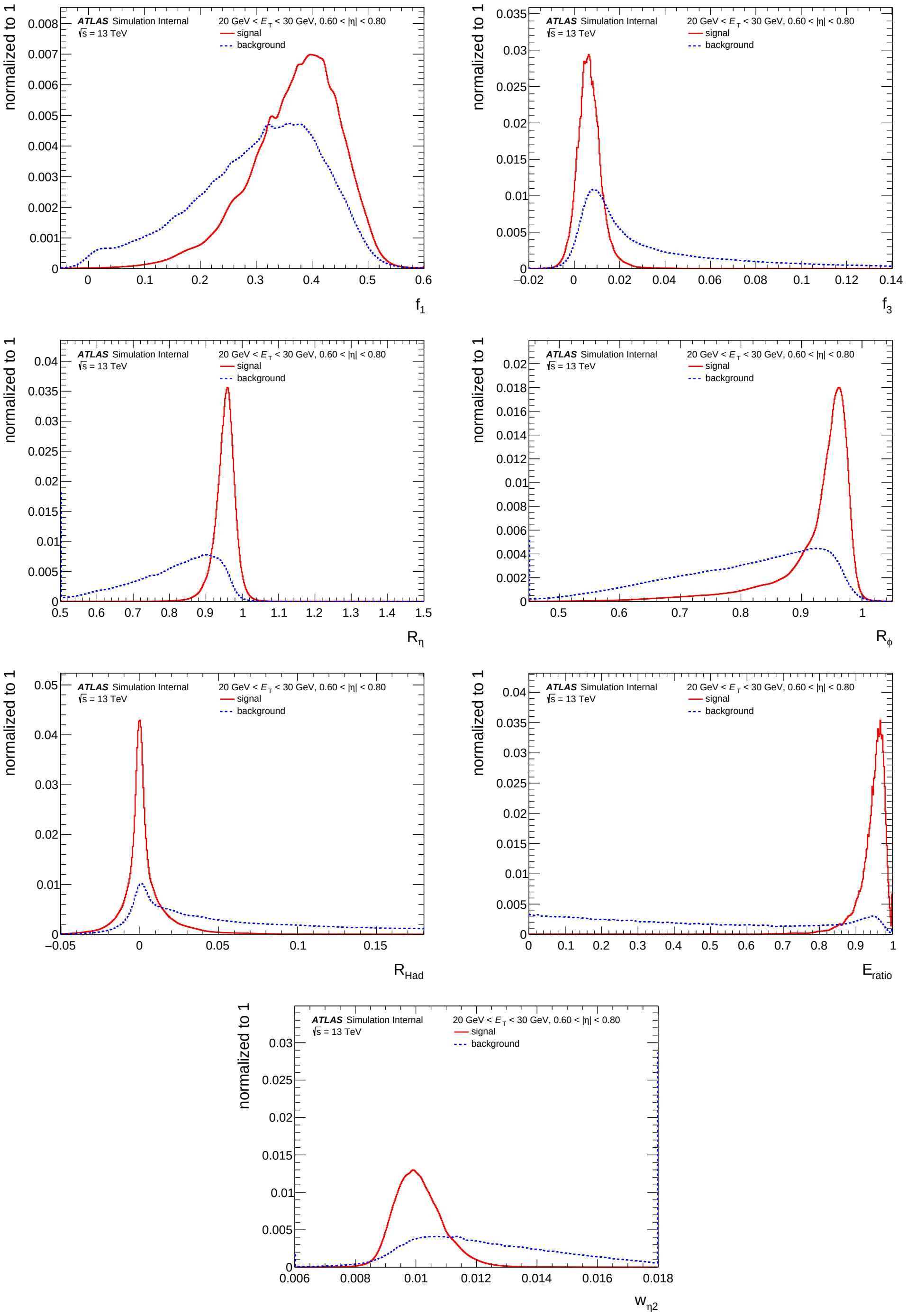


Figure 4.6: MC-based signal and background PDFs used in the electron LH for software release 20.7 (with MC shower shape shifts and widths applied), shown for the variables f_1, f_3, R_η, R_ϕ, R_{had}, E_{ratio}, and w_η, after KDE smoothing has been applied. Only electron candidates with 20 GeV $< E_\mathrm{T} <$ 30 GeV and $0.60 < |\eta| < 0.80$ are shown. As the MC was used to construct these PDFs, the probes from $Z \to ee$were additionally required to pass a truth matching requirement, while the background electron candidates from a dijet sample were required to fail the electron truth matching.

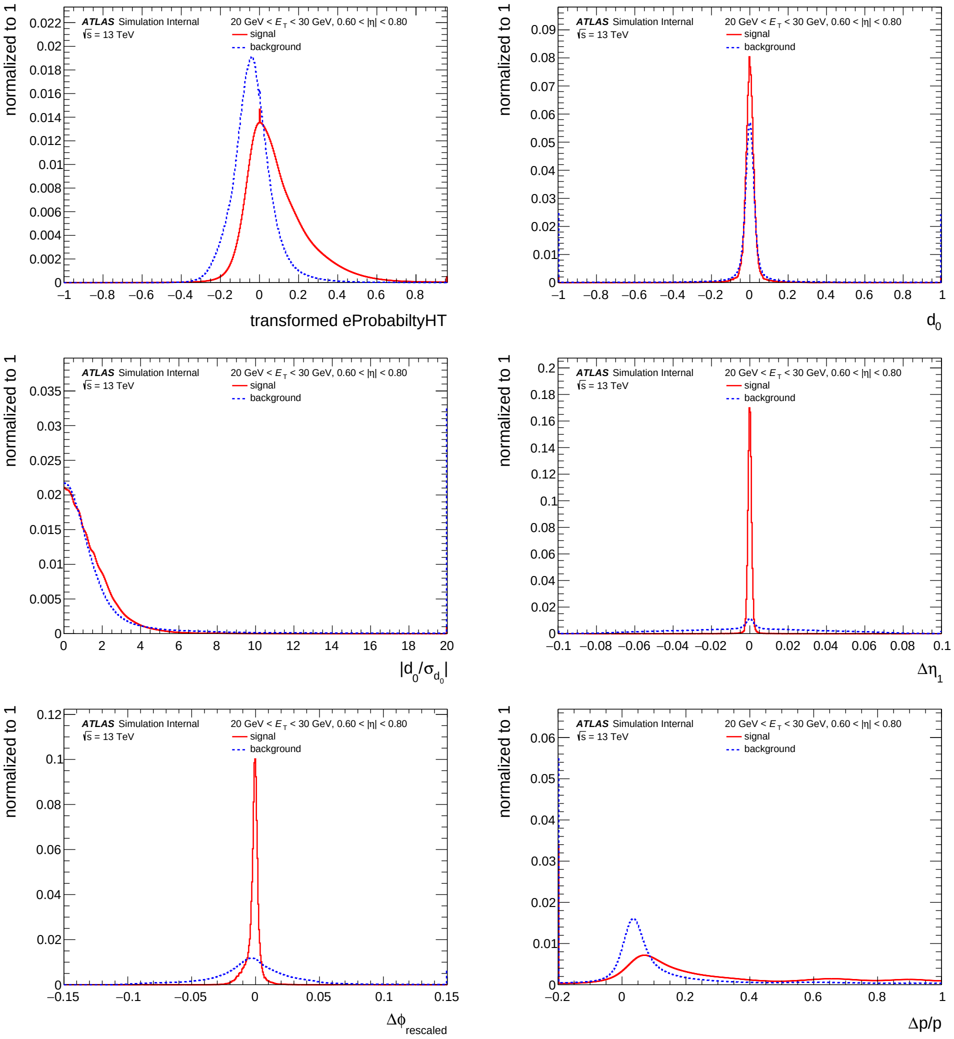
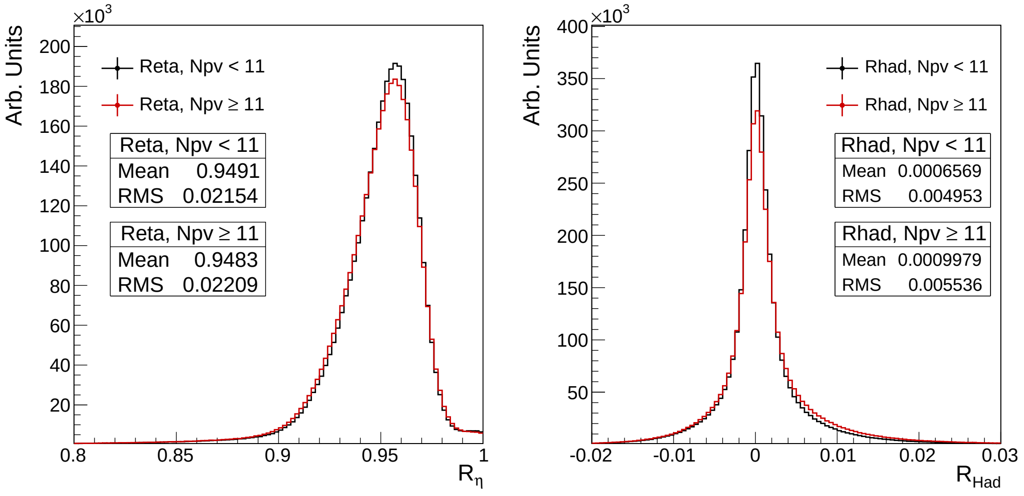


Figure 4.7: MC-based signal and background PDFs used in the electron LH for software release 20.7 (with MC shower shape shifts and widths applied), shown for the variables eProbabilityHT, d_O, $|d_O/\sigma_{d_0}|$, $\Delta_{\eta 1}$, $\Delta\phi_{res}$, and $\Delta p/p$, after KDE smoothing has been applied. Only electron candidates with 20 GeV $< E_\mathrm{T} <$ 30 GeV and $0.60 < |\eta| < 0.80$ are shown. As the MC was used to construct these PDFs, the probes from $Z \rightarrow ee$were additionally required to pass a truth matching requirement, while the background electron candidates from a dijet sample were required to fail the electron truth matching.



(a) (b)

Figure 4.8: R_η is shown in (a) integrated for Nvtx< 11 and Nvtx> 11. The same is shown for R_{had} in (b).

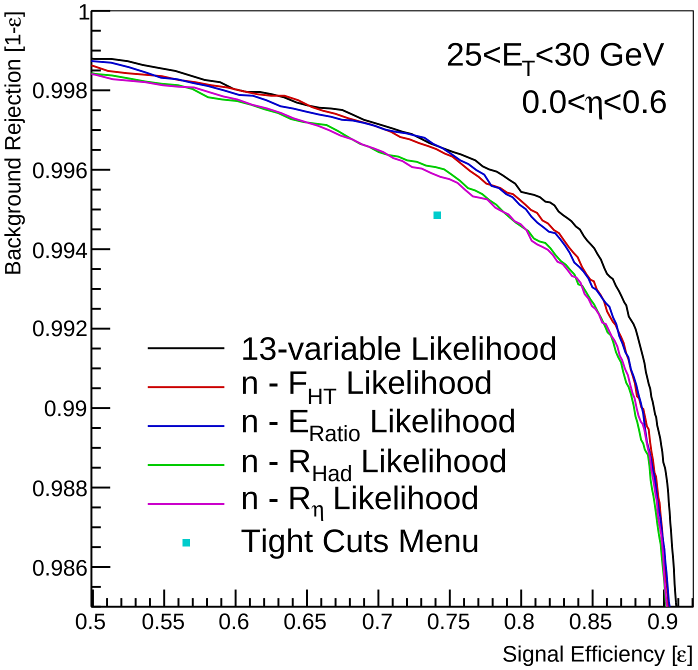


Figure 4.9: The $n-1$ method used to optimize the choice of variables to use in the electron likelihood. Individual variables are removed from the nominal list of likelihood variables, and the likelihood recalculated to assess the relative power of each variable. The example shows the importance of F_{HT}, E_{ratio}, R_{had} and R_η; the performance of the likelihood decreases when each is removed. The `Tight` cut-based operating point is shown for comparison.

This can correct for the negative slope in the signal efficiency, but because we are relaxing the cut at high n_{vtx} we are also letting more background pass the cut. This situation is illustrated in figure 4.10.

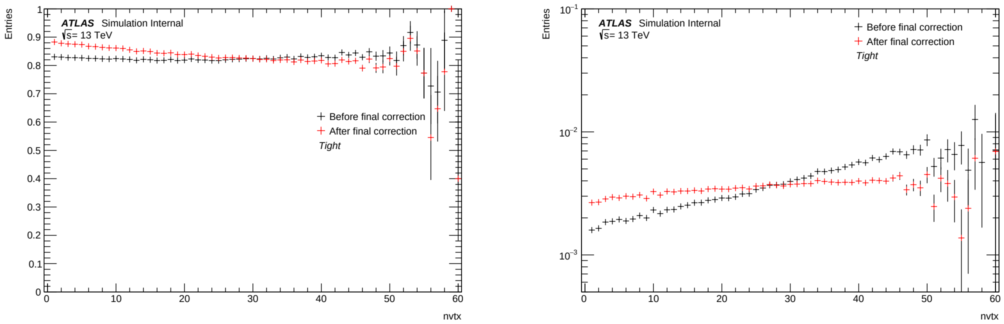


Figure 4.10: In black are the efficiencies corresponding to the signal efficiency tuning being linearly corrected as a function of nvtx. We observe how the background blows up for this first order correction. In red we see the second correction in where the signal efficiency is re-injected with some slope in order to stabilize the background.

This correction as a function of n_{vtx} is applied to only the `Medium` and `Tight` operating points as the pileup dependence for `Loose` and `VeryLoose` are too small to have a noticeable effect. In the plane defined by the likelihood discriminant as a function of n_{vtx} this correction would correspond to tilting the line defining the cut giving it non-zero slope. However, since the `Loose` and `VeryLoose` operating points are not adjusted, there are situations where at high pileup the line defining the `Medium` operating point will cross the looser ones. The electrons passing the `Medium` likelihood is then no longer strictly a subset of the `Loose` likelihood. Figure 4.11 shows how this situation can arise.

The remedy for this situation we decided to make was to keep the discriminant cuts flat as a function of n_{vtx} and instead correct the discriminant output as a function of the pileup. This will ensure that all tighter operating points are subsets of looser ones by construction.

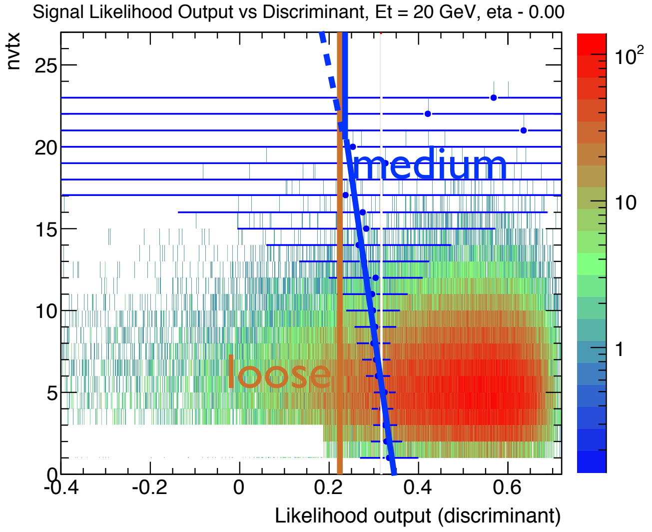


Figure 4.11: Example of the pileup correction used in Run 1 where the z-axis shows the number of entries in each bin. The blue dashed line shows that with large pileup, the `Medium` operating point could be looser than the `Loose` operating point, so they are not perfect subsets. The solid blue line shows what is desired, namely, for `Medium` to always be a subset of `Loose`.

The correction to the discriminant is done in the following way:

$$d_{\mathrm{new}}(n_{vtx}) = \begin{cases} d & \text{if } d < d_{\mathrm{ref}} \\ d_{ref} + (d - d_{ref})\frac{d_{\mathrm{tight,old}} - d_{ref}}{d_{\mathrm{tight,new}} - d_{\mathrm{ref}}} & \text{if } d_{\mathrm{ref}} \leq d < d_{\mathrm{tight,new}} \\ d_{\mathrm{tight,old}} + (d - d_{\mathrm{tight,new}})\frac{d_{\mathrm{max}} - d_{\mathrm{tight,new}}}{d_{\mathrm{max}} - d_{\mathrm{tight,new}}} & \text{if } d_{\mathrm{tight,new}} \leq d < d_{\mathrm{max}} \\ d & \text{if } d_{\mathrm{max}} \leq d \end{cases} \tag{4.4}$$

Here d_{ref} is a reference discriminant which defines the lowest point where pileup corrections are to be done. This is set to the `VeryLoose` operating point as we do not require it to be pileup corrected. Similarly d_{max} is the largest value of the discriminant that should be corrected. This method was found to have performance similar to the Run 1 scheme of correcting the cut values themselves, but ensures that the tighter points remain subset of the looser.

4.3.1 Conclusion

The electron likelihood used offline in Run 1 has been extremely successful. Its improvements over the cuts based ID made it a good candidate for operation in the trigger in Run 2 where the demands on the ID would be higher. The operating points implemented for online and offline running closely match the signal efficiencies of the cuts based ID, but with better background rejection. Although the luminosity of the LHC is steadily increasing, the likelihood is able to adjust around this and will continue to deliver steady performance through the end of Run 2.

Chapter 5

The Search for Low Mass Diphoton Resonances

Most of the activities on Atlas can be divided into measurements and searches. Measurement groups find a numerical value for some parameter of a previously discovered particle or process. An example of this would be the Higgs cross section group which makes ever more accurate measurements of the cross section and branching ratios of the Higgs Boson. In order to make these measurements, the particle whose parameters you will be assessing need to be discovered. This is where the search groups come in. Search groups have the goal of finding new particles and processes. These measurements are simply to determine if there is a statistically significant deviation from currently known physics. Where I spent a majority of my time on Atlas was in a group interested in searching for new particles that decay to two photons and appear in the data as resonances, or narrow bumps.

In the following I will present a search for low mass diphoton resonances during 2017 and 2018 [43]. This analysis used pp collision data with a total integrated luminosity of 80.4 $fb^{-}1$ and a center of mass energy of $\sqrt{S} = 13$TeV collected in 2015, 2016 and 2017. This search is an update of the Run 1 results with $20.3 fb^{-}1$ of data at $\sqrt{s} = 8$TeV [44] The search is performed for narrow resonances with spin 0. Experimental signatures such as these are predicted by models involving an extended Higgs sector and some Supersymmetric/Axion models.

This analysis represents the work of many people over many months. This chapter closely follows the low mass diphoton support note [43] meant to supplement the conference note[45] both of which I spent the last year or so working on. Although I was responsible for many support and production tasks such as data validation, and signal Monte Carlo production, my main focus was on the non-resonant background modeling and the spurious signal test described in section 5.6.3.

5.1 Introduction

Several theoretical models predict new resonances in the diphoton invariant mass spectrum. One of the more prominent are extensions to the Higgs sector of the standard model. The Standard Model Higgs sector contains one complex scalar doublet field that gives rise to the standard model Higgs boson. This is the minimal scalar sector that can be added to the Standard Model to give mass to give mass to particles through the Higgs mechanism. The scalar sector does not necessarily need to be minimal however.

The simplest extension is to add one additional complex scalar doublet field. This is known as the two Higgs doublet model (2HDM) [46] [47] [48]. The scalar potential for two doublets can be written,

$$\begin{aligned} V = m_{11}^2\ \Phi_1^\dagger\Phi_1 + m_{22}^2\ \Phi_2^\dagger\Phi_2 - m_{12}^2(\Phi_1^\dagger\Phi_2 + \Phi_2^\dagger\Phi_1) + \frac{\lambda_1}{2}(\Phi_1^\dagger\Phi_1)^2 + \frac{\lambda_2}{2}(\Phi_2^\dagger\Phi_2)^2 \\ +\lambda_3\ \Phi_1^\dagger\Phi_1\Phi_2^\dagger\Phi_2 + \lambda_4\ \Phi_1^\dagger\Phi_2\Phi_2^\dagger\Phi_1 + \frac{\lambda_5}{2}\left[(\Phi_1^\dagger\Phi_2)^2 + (\Phi_2^\dagger\Phi_1)^2\right]. \end{aligned}$$

The minimization of this potential is,

$$\langle\Phi_1\rangle_0 = \begin{pmatrix} 0 \\ \frac{v_1}{\sqrt{2}} \end{pmatrix}, \qquad \langle\Phi_2\rangle_0 = \begin{pmatrix} 0 \\ \frac{v_2}{\sqrt{2}} \end{pmatrix}. \tag{5.1}$$

With two complex scalar doublets, there will be 8 fields in total. As in the SM Higgs, 3 of these fields get "eaten" to give mass to the $W^\pm$ and Z bosons. This leaves 5 physical scalar fields; two charged scalars, two neutral scalars, and one pseudoscalar.

The Yukawa couplings in the 2HDM [47] in general, will be,

$$\mathcal{L}_y = y^1_{ij}\ \bar{\psi}_i \psi_j \Phi_1 + y^2_{ij}\ \bar{\psi}_i \psi_j \Phi_2, \tag{5.2}$$

with mass matrix,

$$M_{ij} = y^1_{ij} \frac{v_1}{\sqrt{2}} + y^2_{ij} \frac{v_2}{\sqrt{2}} \tag{5.3}$$

In the Standard Model, the mass matrix and Yukawa matrices diagonalize meaning that there are no flavor changing neutral currents (FCNC). With the addition of an additional scalar doublet, the Yukawa matrix is not necessarily guaranteed to be diagonalized if the mass matrix is. This means that the 2HDM can have FCNC at tree level which have very strong constraints from experiment [49]. To avoid this problem an additional $\mathbb{Z}_2$ symmetry can be enforced. This can be applied in a few different ways; the type I where all fermions only couple to one of the doublets, and the type II where the up-type quarks couple to one doublet, with the down-type quarks and leptons coupling to the other. Two other models are possible (type III and type IV) which do allow tree level FCNC but constrain them in other ways.

The two neutral scalars are h and H with h the lighter of the two. Either one of these two could be the SM Higgs and so the second neutral scalar could be found above or below 125 GeV. Just like in the SM the branching ratio of bosons to photons in the 2HDM are not predicted to be large, but the clean experimental signature and high resolution of the ATLAS detector make the diphoton channel very good at searching for an extended scalar sector.

5.2 Data and Simulation

5.2.1 Data

In 2015 the ATLAS experiment began the Run 2 data collection period where the center of mass energy was increased from $\sqrt{s} = 8$ TeV to near its design energy of $\sqrt{s} = 13$ TeV. For this analysis, data from 2015, 2016 and 2017 were used. Events were recorded by requiring a diphoton trigger with a transverse energy (E_T) of 20 GeV on both the leading and subleading photon candidates. The high level trigger (HLT) then requires that both photons pass tight reconstruction criteria in addition to being isolated for the 2017 data only. The data used after these requirements corresponds to $80 fb^{-1}$ of total integrated luminosity [43].

5.2.2 Signal Simulation

Monte Carlo simulation (MC) is used to produce both signal and background samples that are used for various modeling tasks as well as to optimize the search strategy. Due to the difference in run conditions between the 2015/2016 and the 2017 data collection periods, simulation must be generated separately for each. The mc16a samples correspond to the 2015/2016 and the mc16d samples to the 2017 data.

The different MC signal samples are detailed in Table 5.1. The default samples are the ggF ones, generated using the effective-field-theory approach implemented in MadGraph5_aMC@NLO [50] version 2.3.3 at next-to-leading order (NLO) in quantum chromodynamics (QCD). From the Higgs characterization framework [51], CP-even dimension-five operators coupling the new resonance to gluons and photons were included. Samples were generated with the NNPDF3.0 NLO parton distribution functions (PDFs) [52], using the A14 set of tuned parameters (tune) of Pythia 8.186 [53, 54] for the parton-shower and hadronization simulation. Narrow Width Approximation is assumed for all samples (a 4 MeV width is used in practice).Interference effects between the resonant signal and the background processes are neglected. Mass points outside of the search region (65-110 GeV) were generated in order to have a better interpolation of the signal modeling (cf Section 5.4). But in the end the 40 and 50 GeV mass points are not used because the shape is biased by the trigger cut.

Since the goal of the analysis is to perform a model-independent search, the properties of the fiducial volume and correction factors (detailed in Section 5.8) are also studied for other Higgs-like production modes: vector boson fusion (VBF), associated production with a vector boson (WH, ZH) and associated production with a top quark pair ($t\bar{t}H$). All those samples are simulated with the Powheg [55, 56] generator both interfaced to Pythia8 [53] for parton showering and hadronization.

Process	Generator	Mass [GeV]	Nevents ($\times 10^3$)		width
			mc16a	mc16d	
ggF	MadGraph	40-50-60-70-80-90-100-110-120-140-160-180	30	40	NWA
		60	30	40	5%
ggF	PowHeg+Pythia8	40-60-80-100-120	30	40	NWA
VBF	PowHeg+Pythia8	40-60-80-100-120	30	40	NWA
WH	Pythia8	40-60-80-100-120	30	40	NWA
ZH	Pythia8	40-60-80-100-120	30	40	NWA
$t\bar{t}$H	PowHeg+Pythia8	40-60-80-100-120	30	40	NWA

Table 5.1: Signal samples of the five Higgs production modes, available mass points and whether the NWA width is used or not [43].

5.2.3 Background Simulation

MC samples of prompt diphoton events are used to validate the functional form used to model the background diphoton invariant mass spectrum.

Events containing two photons are also generated using the Sherpa 2.2.2 generator, with a 20 GeV cut on the photon transverse energy. Matrix elements are calculated with up to 2 partons at LO and merged with the Sherpa parton shower [57] using the ME+PSatLO

Generator	$m_{\gamma\gamma}$ range [GeV]	Cross section [pb]	Filter efficiency	Nevents ($\times 10^6$)		Reconstruction
	–			mc16a	mc16d	
Sherpa LO	0 – 55	$1.4088{\cdot}10^{+2}$	$2.4335{\cdot}10^{-1}$	1.0	1.0	FS
	55 – 100	$1.4778{\cdot}10^{+2}$	$4.5670{\cdot}10^{-1}$	1.0	1.0	FS
	100 – 160	$3.9728{\cdot}10^{+1}$	$4.9730{\cdot}10^{-1}$	1.0	1.0	FS
Sherpa NLO	50 – 90	$1.3904{\cdot}10^{+2}$	1	91.7	130.0	AFII
	90 – 175	$5.1823{\cdot}10^{+1}$	1	114.7	162.7	AFII

Table 5.2: Prompt diphoton background samples.

prescription [58]. The CT10 PDF set is used in conjunction with dedicated parton shower tuning developed by the Sherpa authors. Those samples are used for optimisation of the analysis and for cross-checks.

A recent implementation of Sherpa, version 2.2.4 [59], is also used. It consistently combines parton-level calculations of varying jet multiplicity up to NLO (The $\gamma\gamma$ and $\gamma\gamma$+1 parton processes are generated at NLO accuracy, while the $\gamma\gamma$+2 partons and $\gamma\gamma$+3 partons processesare generated at LO. Charm and bottom quarks are included in these matrix elements in the massless approximation.) with parton showering while avoiding double-counting effects. The NNPDF3.0 NNLO PDFs [60] are used in conjunction with the corresponding Sherpa default tuning. Those samples with high-statistics are used for the background modelling studies described in Section 5.5.

Details on those samples are given in Table 5.2. In order to maximize the available statistics over the mass range of interest, the simulation is separately performed in exclusive slices of diphoton invariant mass. Those samples are fully reconstructed and are used for optimization of the analysis.

Samples with one photon and up to 2 jets are simulated with Sherpa and are generated in slices of the photon transverse energy. They are described in the general HGam supporting note [61] and are used for the study of the isolation cuts.

Monte-Carlo samples of $Z \to ee$ decays are simulated using POWHEG-BOX V2 interfaced

Generator	m_{ee} range [GeV]	Filter	Cross section [pb]	Filter efficiency	Nevents ($\times 10^6$)	
	–				mc16a	mc16d
POWHEG	–		$1.9012 \cdot 10^{+3}$	1	61.2	79.3
SHERPA	0 – 70	c-veto, b-veto	$1.9810 \cdot 10^{+3}$	0.82143	8.0	
	0 – 70	c-filter, b-veto	$1.9816 \cdot 10^{+3}$	0.11407	5.0	
	0 – 70	b-filter	$1.9821 \cdot 10^{+3}$	0.06576	8.0	
	70 – 140	c-veto, b-veto	$1.1063 \cdot 10^{+3}$	0.69432	6.0	
	70 – 140	c-filter, b-veto	$1.1045 \cdot 10^{+2}$	0.18697	2.0	
	70 – 140	b-filter	$1.1043 \cdot 10^{+2}$	0.11605	6.0	
	140 – 280	c-veto, b-veto	$4.0711 \cdot 10^{+1}$	0.61632	5.0	
	140 – 280	c-filter, b-veto	$4.0683 \cdot 10^{+1}$	0.23302	3.0	
	140 – 280	b-filter	$4.0671 \cdot 10^{+1}$	0.15319	12.4	
	280 – 500	c-veto, b-veto	$8.6711 \cdot 10^{+0}$	0.56328	2.0	
	280 – 500	c-filter, b-veto	$8.6597 \cdot 10^{+0}$	0.26640	2.0	
	280 – 500	b-filter	$8.6793 \cdot 10^{+0}$	0.17638	2.0	
	500 – 1000		$1.8096 \cdot 10^{+0}$	1	3.0	
	1000 – ∞		$1.4875 \cdot 10^{-1}$	1	1.0	

Table 5.3: $Z \rightarrow ee$ samples used to study the Drell-Yan background.

to the PYTHIA8 version 8.186 parton shower model. The CT10 parton density function set is used in the matrix element. The AZNLO set of tuned parameters is used, with PDF set CTEQ6L1, for the modeling of non-perturbative effects. Sliced samples generated with Sherpa 2.2.1 are also used for computing systematic uncertainties. Details on those samples are given in Table 5.3.

The data and MC samples are processed through the HGam framework, and the mini-xAODs version h021 are used unless stated otherwise. Weights are applied to correct the simulation: Scale Factor for the photon identification, pile-up, choice of the diphoton vertex, and MC weight when applicable (eg Sherpa samples).

5.3 Event Selection

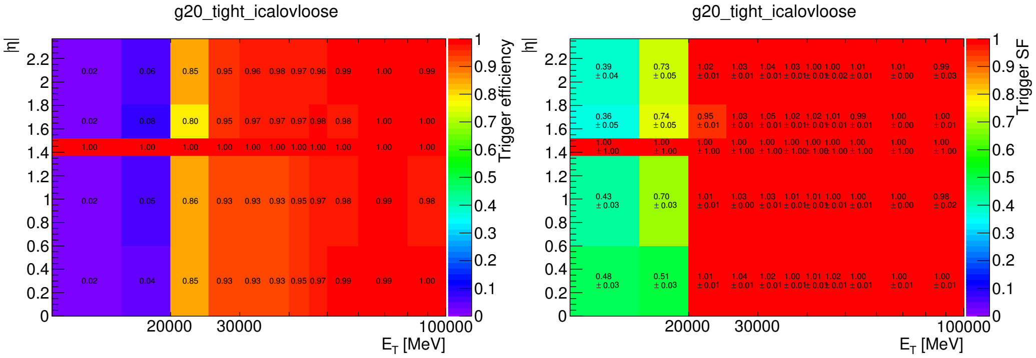
5.3.1 Trigger Selection

In general we want to be able to perform this search as low as possible in diphoton mass. One of the limits we run into in this regard is the threshold at which triggers become prescaled. At very low transverse energies (E_{T}), the rate that events occur in the detector is much higher than the detector is able to record at. This means that low E_{T}, the triggers must be prescaled, or have their rate limited. We would like to use triggers that are unprescaled as they will collect a higher fraction of the events produced above their threshold. For 2015 and part of 2016 running the trigger 2g20_tight was used, which corresponds to a diphoton trigger where both the leading and subleading photons are required to have $E_{\mathrm{T}}>$ 20 GeV and both pass tight identification. For the later part of 2016 running, the threshold was raised to 22 GeV for the lowest unprescaled trigger (2g22_tight) due to the increase in luminosity of the LHC. In 2017, the threshold was lowered back to 20 GeV by adding the additional requirement that the photons pass the very loose calorimeter isolation (2g20_tight_icalovloose). The triggers used are summarized in table 5.4.

Year	2015	2016 up to D3	2016 from D3	2017
HLT item	`2g20_tight`	`2g20_tight`	`2g22_tight`	`2g20_tight_icalovloose`
luminosity [fb^{-1}]	3.2 fb^{-1}	11.5	21.5	43.6 fb^{-1}

Table 5.4: Lowest unprescaled HTL items, depending on the data-taking period, and associated integrated luminosity [43]

The use of 20 GeV triggers means we can use an offline cut of 22 GeV on the photon E_{T}. This causes a kinematic turn on in the invariant mass distribution which peaks at around 60 GeV. Although this turn on makes background modeling more difficult (see section 5.6.3), it allows us to fit the distribution down to 60 GeV. The efficiency of the triggers are studied by the Egamma group's photon trigger experts and are shown in figure 5.1a. These efficiencies have good agreement with simulation and are shown in figure 5.1b as the ratio of data to MC.



(a) (b)

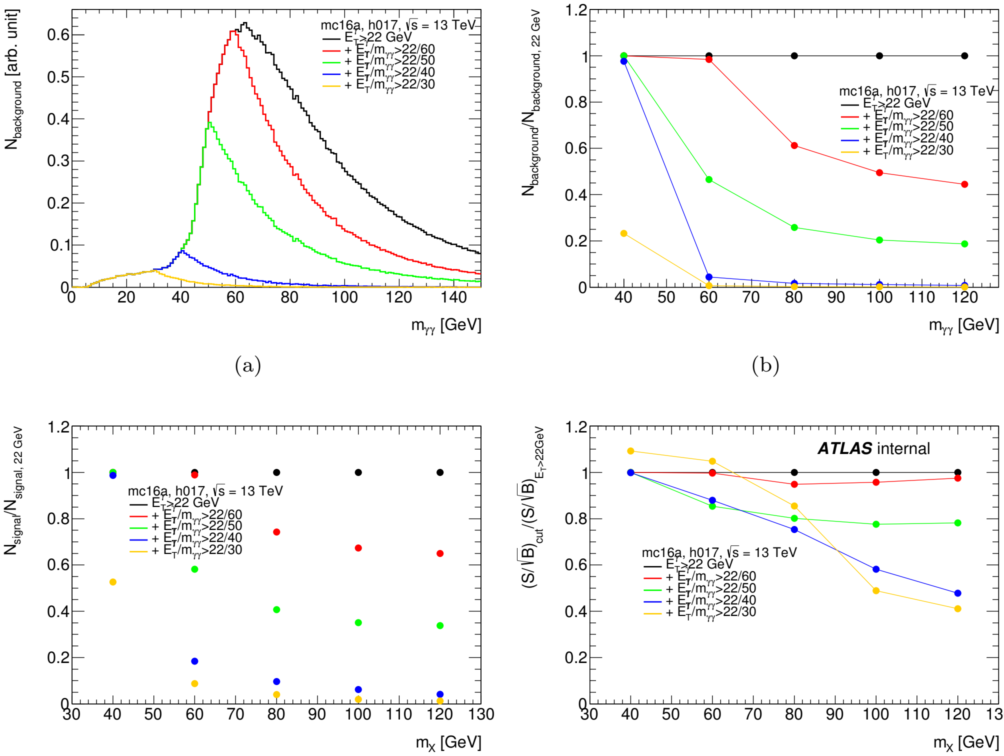
Figure 5.1: `2g20_tight_icalovloose` trigger efficiencies (a) and data/MC ratios (b) measured on 2017 data.

5.3.2 Kinematic Selection

As mentioned before we use an offline cut of $E_{\mathrm{T}}>$ 22 GeV. We want to keep this as low as possible for two reasons: to be able to start the search as low as possible and to keep the kinematic turn on caused by the cut far away from the Drell-Yan peak. Changes to this cut were not considered.

Also considered were relative cuts where the absolute E_{T}cut is also divided by the invariant mass of the pair of photon candidates. These relative cuts are used in the SM $H \rightarrow \gamma\gamma$analysis [62] as well as the high-mass diphoton analysis [8]. Figure 5.2a shows a comparison of the absolute and relative cuts defined as $E_{\mathrm{T}} > 22GeV/XGeV$. You can see that for the relative cuts the kinematic turn on follows the absolute cut until the value of $XGeV$ where it falls off. To measure the effect on the background we use background only SHERPA MC samples described in section 5.2. The number of background events are measured in a window centered at m_X, the mass of the signal point, with a width of $2 * \sigma_{CB}$ where σ_{CB} is the width of the double sided crystal ball function used in the signal parameterization described in section 5.4. Figure 5.2b shows the comparison of the tested cuts on the number of background events normalized to the $E_{\mathrm{T}} > 22\mathrm{GeV}$ cut. The same is done in figure 5.2c for the number of signal events measured on the ggF signal MC samples, again normalized to the absolute cut.

In order to test the cuts, we use the significance $Z = S/\sqrt{B}$. This analysis is process independent and so we have no expected number of events to compare. The ratio of $S/\sqrt{B}$ for the relative cut divided by $S/\sqrt{B}_{E_{\mathrm{T}}>22GeV}$ for the default 22GeV cut is shown in figure 5.2d. The significance is much lower at high masses for the relative cuts so it was concluded that the default absolute cut $E_{\mathrm{T}} > 22\mathrm{GeV}$ will be used for this analysis.



(a) (b)

(c) (d)

Figure 5.2: (a) $M_{\gamma\gamma}$ distribution of the diphoton background for various sets of kinematic cuts applied on the photon transverse energies. (b) Ratio between the number of background events in $m_X \pm 2 * \sigma_{CB}$ with the tested cuts and with the 22 GeV cuts. (c) Ratio between the number of signal events in $m_X \pm 2 * \sigma_{CB}$ with the tested cuts and with the 22 GeV cuts. (d) Ratio between $S/\sqrt{B}$ with the tested cuts and $S/\sqrt{B}$ with the 22 GeV cuts.

5.3.3 Isolation Selection

In addition to using an isolation requirement on the 2017 trigger, we also use isolation in the offline selection of events. Isolation ensures that there is not too much activity in the detector around the photon from both other interaction vertices but also from other decay product originating from the primary vertex. Non-isolated candidates have a higher probability of being misconstructed as photons. Two types of isolation variables are used; track-based and calorimeter-based. Track-based isolation, *ptcone*20, is defined as the scalar sum of the transverse momenta of all tracks with $p_T > 1\text{GeV}$ in the cone of size $\Delta R = \sqrt((\Delta\eta)^2 + (\Delta\phi)^2) = 0.2$ around each photon candidate. Only tracks that are reconstructed as originating from the primary vertex are used, and tracks associated with conversions are also removed. The calorimeter-based isolation uses only the topological clusters in the calorimeter. *topoetcone*40 is the sum of all positive energy in clusters within a cone of $\Delta R = 0.4$ after having subtracted the contribution from the photon candidate itself.

Similar to how the photon ID is done, there are three different isolation cuts that are available as working points:

- FixedCutTightCaloOnly: $topoetcone40 < 0.022 p_{\mathrm{T}} + 2.45$ [GeV], aiming for SM measurements;
- FixedCutTight: $topoetcone40 < 0.022 p_{\mathrm{T}} + 2.45$ [GeV] and $ptcone20/p_{\mathrm{T}} < 0.05$, aiming for high-$E_{\mathrm{T}}$ photons;
- FixedCutLoose: $topoetcone20 < 0.065 p_{\mathrm{T}}$ and $ptcone20/p_{\mathrm{T}} < 0.05$, designed for the SM $H \to \gamma\gamma$ analysis.

In order to choose which working point to use, we look at the significance for each cut. We will measure the significance relative to the FixedCutLoose working point by taking the ratio

Generator of signal samples	Mass point [GeV]	
MadGraph	$M_X = 40, 50, 60, 70, 80, 90, 100, 110$	$M_X = 120, 140, 160, 180, 200$
	$M_X - 5 < m_{\gamma\gamma} < M_X + 5$ GeV	$M_X - 10 < m_{\gamma\gamma} < M_X + 10$ GeV
PowHeg+Pythia8	$M_X = 40, 60, 80, 100, 120$	$M_X = 160$
	$M_X - 10 < m_{\gamma\gamma} < M_X + 10$ GeV	$M_X - 20 < m_{\gamma\gamma} < M_X + 20$ GeV

Table 5.5: $m_{\gamma\gamma}$ cuts applied on signal and background samples with respect to the different mass of resonance points.

of the two. This gives us the relative efficiency $\epsilon_{relative} = N_{testcut}/N_{FixedCutLoose}$ where $N_{testcut}$ ($N_{FixedCutLoose}$) is the number of events passing the cut we are testing (number of events passing FixedCutLoose). The gain or loss in significance as measured relative to FixedCutLoose is written as $Z_{testCut}/Z_{FixedCutLoose} = \epsilon_S/\sqrt{\epsilon_B}$ where ϵ_S and ϵ_B are the relative efficiencies for the signal and background respectively.

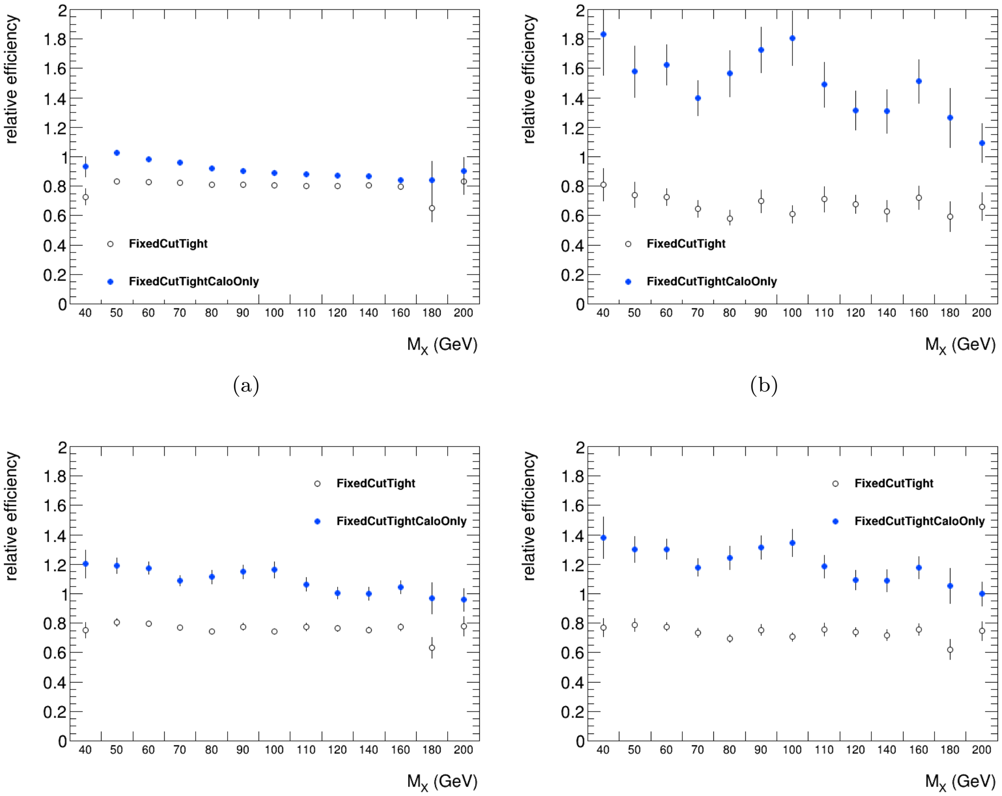
The samples primarily used for the isolation study are the MadGraph samples, however, samples produced with PowHeg+Pythia8 are used as well to check the effect of the working points on different production modes. These samples are detailed in section 5.2.2. For the background samples we use the Sherpa leading order samples in addition to the Sherpa γ jet samples. The γ jet component ranges from about 50% to 30% of the total background events according to the study in section 5.6.1. This means we need to account for this contribution as jets have a different isolation distribution from photons. Isolation working points are tested for two fraction of γ jet contribution, 50% $\gamma\gamma$: 50% γ jet and 70% $\gamma\gamma$: 30% γ jet. Jet-jet components are neglegted as they are small ($\sim$ 6% section 5.6.1).

The events that are used for testing are required to pass the basic kinematic selection of 2g20_tight trigger, tight identification and $E_{\mathrm{T}} > 22$ GeV. There is then an addition cut placed on the events invariant mass in a window around the mass of the signal being tested. These cuts range from ±5GeV for lower masses to ±20 GeV for higher masses in the PowHeg samples. These $M_{\gamma\gamma}$ cuts are shown in table 5.5.

Figure 5.3 shows the relative isolation efficiencies for the background samples. The binning in this plot corresponds to the mass points in the MadGraph signal sample so that they may be easily compared and combined. Likewise, figure 5.4 shows the same background samples, but with the binning corresponding to the mass points in the PowHeg signal samples. As can be seen, the ratios for the FixedCutTight isolation are below 1, showing that this working point has a higher rejection than the FixedCutLoose. This is expected as the tight cuts are designed to have a greater rejection (or lower efficiency). The relative efficiency for the FixedCutTightCaloOnly is greater than or close to 1, showing that its rejection power is lower than that of the track-based FixedCutLoose working point.

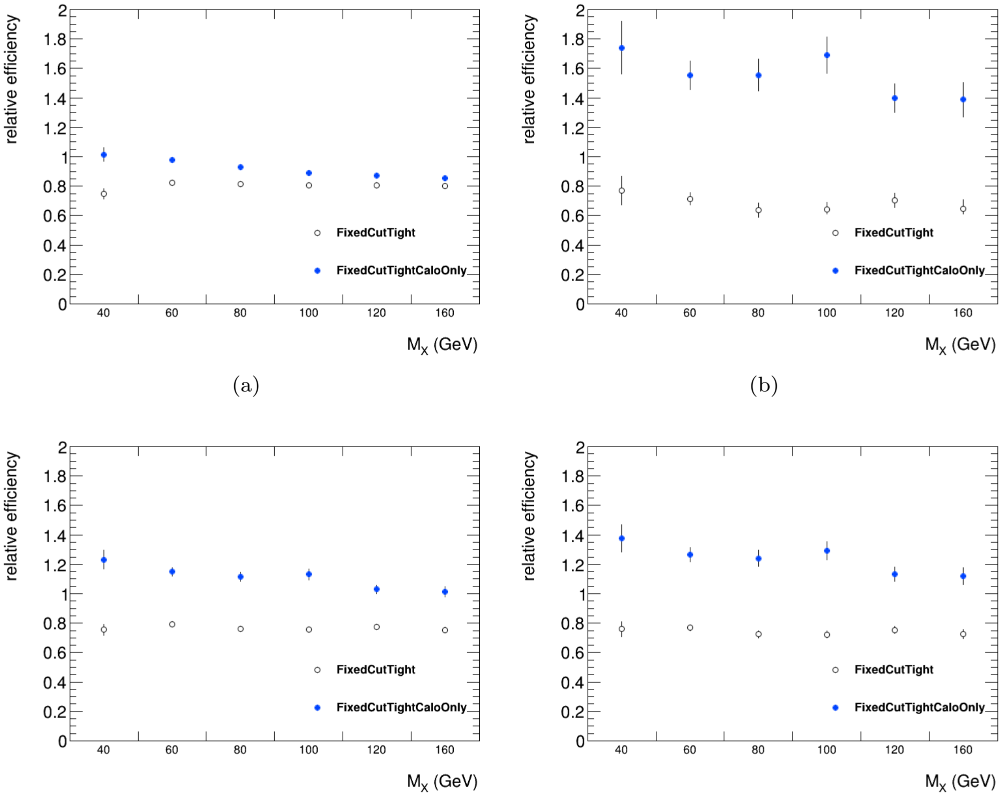
The relative efficiency for the MadGraph signal samples are shown in firgure 5.5. We find a similar situation to the background efficiency with the FixedCutTightCaloOnly being more efficient than FixedCutLoose, and FixedCutTight being slightly less. Figure 5.6 shows the significance relative to the FixedCutLoose working point for the two different fractions of γ-jet. In both cases, the FixedCutTight working point has a higher significance than the calorimeter only version, but both are lower and 1 in general. Although the FixedCutTight isolation is very close to one for a few mass points, it is on average lower.

Figure 5.7 shows the relative efficiencies for the PoweHeg + Pythia8 signal samples with the ggH, ttH, VBF, WH, and ZH production modes. The significance for each of these production modes is shown in figure 5.8 for the 70% : 30% γ-jet fraction, and the same in figure 5.9 for the 50% : 50% γ-jet fraction. We see similar results for the PowHeg + Pythia8 samples as for the MadGraph samples. The FixedCutLoose working point tends to have the highest significance across the invariant mass range and so it is used for the analysis.



(a) (b) (c) (d)

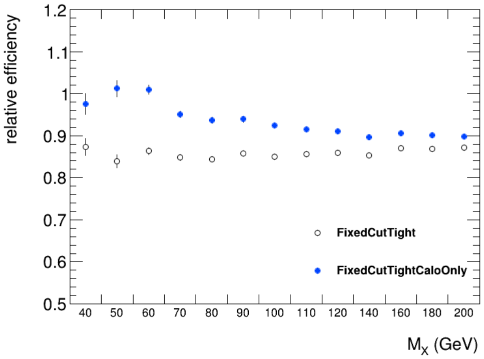
Figure 5.3: Isolation efficiency relative to the FixedCutLoose working point for background samples: (a) $\gamma\gamma$. (b) γjet. (c) combined background sample with $\gamma\gamma : \gamma jet = 70\% : 30\%$. (d) $\gamma\gamma : \gamma jet = 50\% : 50\%$. Efficiencies for the FixedCutTight (respectively FixedCutTightCaloOnly) working point are represented with full (respectively open) circles.



(a) (b)

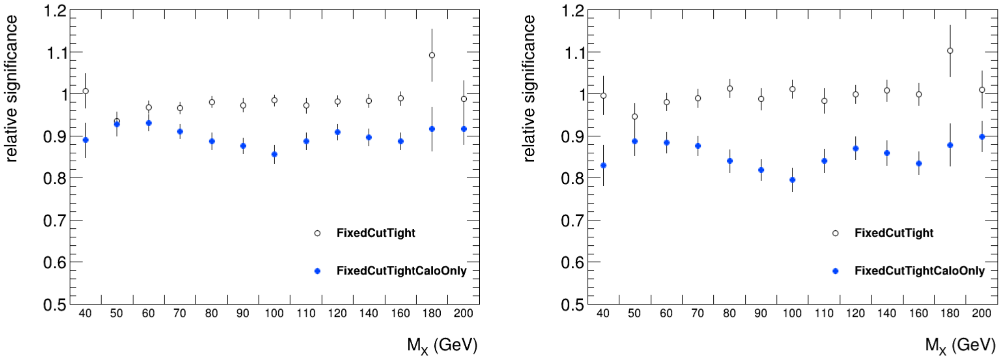
(c) (d)

Figure 5.4: Isolation efficiency relative to the FixedCutLoose working point for background samples: (a) $\gamma\gamma$. (b) γjet. (c) combined background sample with $\gamma\gamma : \gamma jet = 70\% : 30\%$. (d) $\gamma\gamma : \gamma jet = 50\% : 50\%$. Efficiencies for the FixedCutTight (respectively FixedCut-TightCaloOnly) working point are represented with full (respectively open) circles.



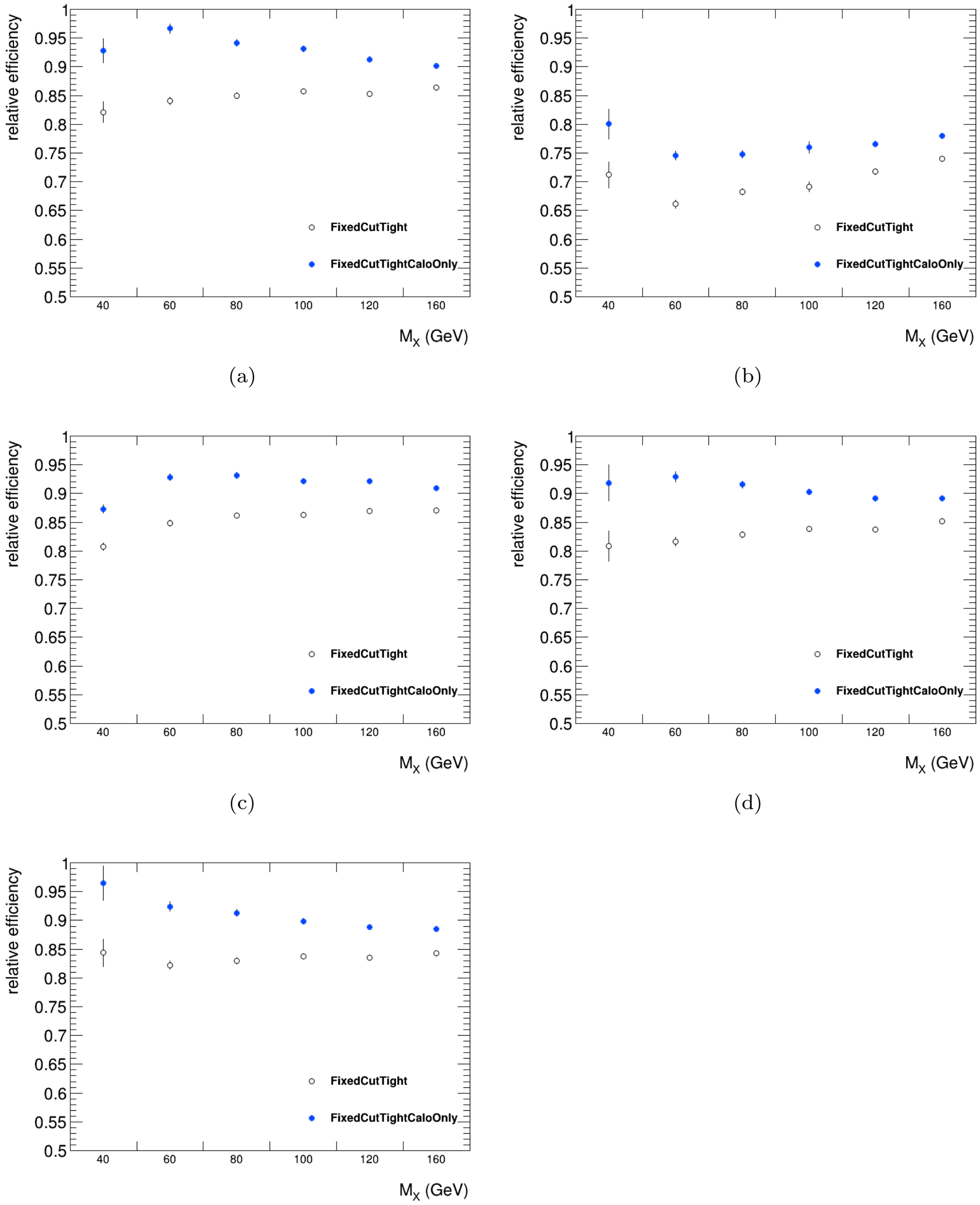
(a)

Figure 5.5: Isolation efficiency relative to the FixedCutLoose working point for MadGraph signal samples. Black open circles represents that FixedCutTight isolation requirement is applied, and blue dots represents FixedCutTightCaloOnly.



(a) (b)

Figure 5.6: Significance relative to the FixedCutLoose working point with MadGraph signal sample and background contains: (a) $\gamma\gamma : \gamma jet = 70\% : 30\%$. (b) $\gamma\gamma : \gamma jet = 50\% : 50\%$. Significances for the FixedCutTight (respectively FixedCutTightCaloOnly) working point are represented with full (respectively open) circles.



(a)

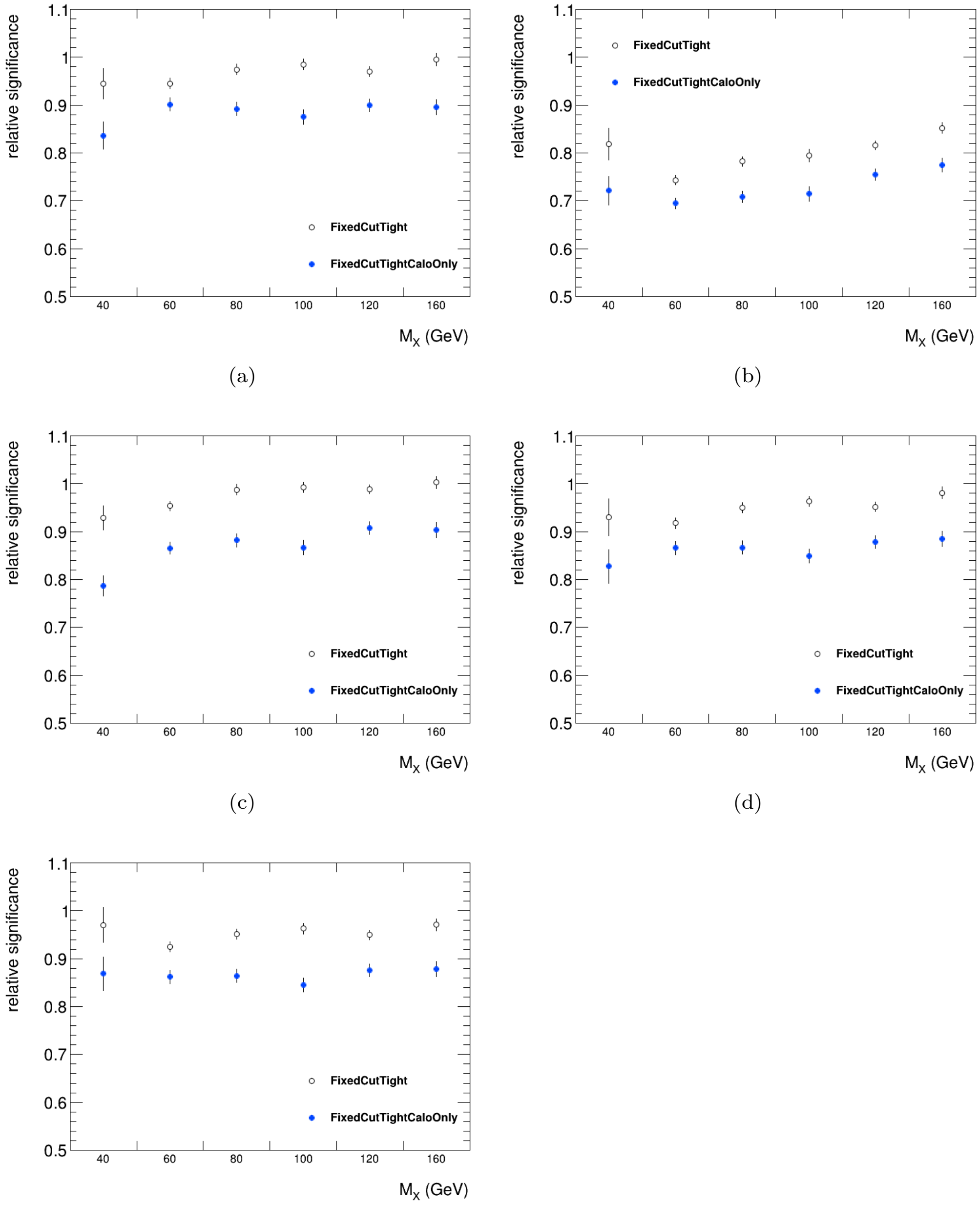
(b)

(c)

(d)

(e)

Figure 5.7: Isolation efficiency relative to the FixedCutLoose working point for signal samples: (a) ggH. (b) ttH. (c) VBFH. (d) WH. (e) ZH. Efficiencies for the FixedCutTight (respectively FixedCutTightCaloOnly) working point are represented with full (respectively open) circles.

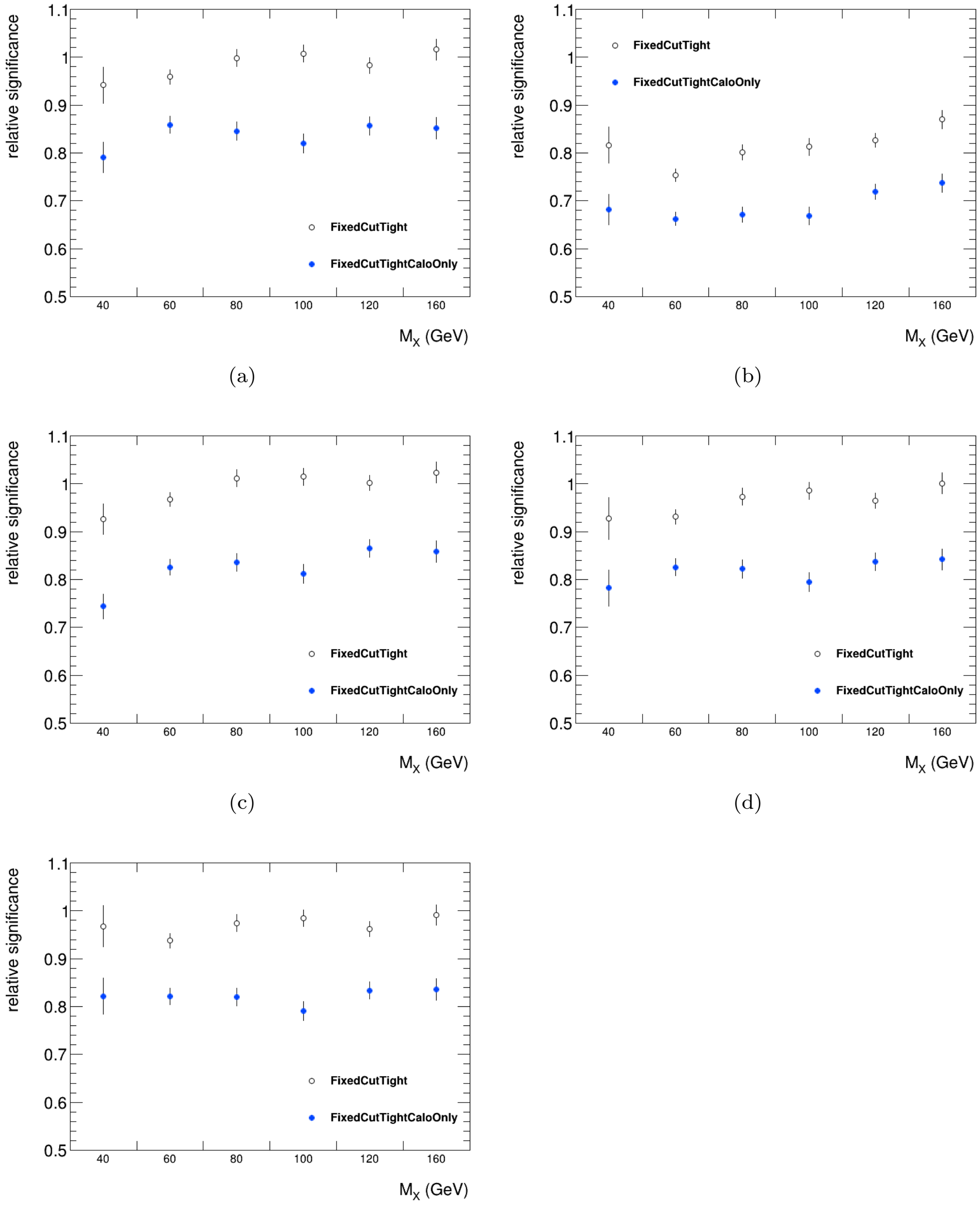


(a) (b)

(c) (d)

(e)

Figure 5.8: Significance relative to the FixedCutLoose working point with signal sample as: (a) ggH. (b) ttH. (c) VBFH. (d) WH. (e) ZH. The background contains $\gamma\gamma : \gamma jet = 70\% : 30\%$. Significances for the FixedCutTight (respectively FixedCutTightCaloOnly) working point are represented with full (respectively open) circles.



(a)

(b)

(c)

(d)

(e)

Figure 5.9: Significance relative to the FixedCutLoose working point with signal sample as: (a) ggH. (b) ttH. (c) VBFH. (d) WH. (e) ZH. The background contains $\gamma\gamma : \gamma jet = 50\% : 50\%$. Significances for the FixedCutTight (respectively FixedCutTightCaloOnly) working point are represented with full (respectively open) circles.

5.3.4 Categorization

The determination of which and how many categories to use is important for this analysis as we expect to have different numbers of Drell-Yan events for different conversion status of the photon candidates. Conversions happen when photons in material convert into e^+e^-pairs. Conversion status is split into 3 categories. Unconverted, converted with 1 track, and converted with 2 tracks. The two conversion categories are then further subdivided by which detector sub systems (silicon tracker or TRT) the tracks appear in.

- unconverted: unconverted photon
- conv 1-track: single track converted photon, divided in:
  - singleSi: one track only, with Si hits
  - singleTRT: one track only, no Si hits (TRT only)
- conv 2-track: double track converted photon, divided in:
  - doubleSi: two tracks, both with Si hits
  - doubleTRT: two tracks, none with Si hits (TRT only)
  - doubleSiTRT = two tracks, only one with Si hits

Three different categorization schemes are considered where the leading and subleading photons conversion status are treated separately.

- 4 Categories - Checking only whether the photon is converted or not for leading and subleading.
- 9 Categories - As above but splitting the conversions into 1-track and 2-track varieties.

- 36 Categories - Treating all of the above conversion states divide by subdetector separately for leading and subleading.

To determine whether categorization is necessary or which to use as similar approach is taken to the previous sections. The significance is tested with each categorization scheme and compared with no categorization. For these tests the signal used is the $X \rightarrow \gamma\gamma$ process and with $Z \rightarrow ee$ as background. The significance is defined as $Z = N_{X\rightarrow\gamma\gamma}/\sqrt{N_{Z\rightarrow ee}}$ where $N_{X\rightarrow\gamma\gamma}$ and $N_{Z\rightarrow ee}$ are the number of events passing all selection criteria and falling within a window of $2 * \sigma_{CB}$. For the $X \rightarrow \gamma\gamma$ signal samples, only the mass point $m_X = 90GeV$ is used. The significance defined for a single category i can be written as,

$$Z_i = f_i^{X\rightarrow\gamma\gamma} * N_{X\rightarrow\gamma\gamma}/\sqrt{f_i^{Z\rightarrow ee} * N_{Z\rightarrow ee}}. \tag{5.4}$$

Here we have used the f_i as the fraction of events in each conversion category as measured in the signal or background MC. The sum in quadrature of all conversion categories is the overall significance for that categorization scheme shown in eq. 5.5

$$Z_{categories} = \sqrt{\Sigma_i Z_i^2} = Z_{nocategories}\sqrt{\Sigma_i \left(\frac{f_i^{X\rightarrow\gamma\gamma}}{\sqrt{f_i^{Z\rightarrow ee}}}\right)^2} \tag{5.5}$$

This allows us to use part of this equation as a figure of merit: $FoM = \sqrt{\Sigma_i (f_i^{X\rightarrow\gamma\gamma}/\sqrt{f_i^{Z\rightarrow ee}})^2}$. The figure of merit tells us the change in significance we expect for a particular scheme of categorization relative to having no categorization. If this number is greater than 1, the categorization is better than no categories. Tables 5.6, 5.7, and 5.8 show the results for the schemes with 4, 9 and 36 categories respectively. The FoM for each scheme is 1.71, 1.78, and 1.84 respectively. We can see from the table 5.8 that a majority of the converted photons in both leading and subleading are of the doubleSi type. Even thought the figure of merit for this scheme is the highest, the majority of conversion belonging to one type means we are safe to consider fewer categories. Similarly, in the case of 4 categories, the case where there is one converted photon, and one unconverted photon has very similar fractions for both the leading and subleading photons being converted. Because they are so close we can merge these categories as well. So, instead of 4 categories we would have 3: unconverted/unconverted, (converted/unconverted + unconverted/converted), and converted/converted labeled UU, UC, and CC respectively. It was decided ultimately to use this 3 category scheme for the analysis as the difference in the FoM is not large and using less categories simplifies many other parts of the analysis.

$f^{cat}_{Z\to ee}(\%)$	*subleading* γ	
	unconverted	converted
leading γ		
unconverted	10.2	23.6
converted	17.3	48.9
$f^{cat}_{H\to\gamma\gamma}(\%)$	*subleading* γ	
	unconverted	converted
leading γ		
unconverted	50.5	21.1
converted	19.9	8.4
$f_i^{X\to\gamma\gamma}/\sqrt{f_i^{Z\to ee}}$	*subleading* γ	
	unconverted	converted
leading γ		
unconverted	1.58	0.44
converted	0.48	0.12

Table 5.6: Percentage of events containing photons reconstructed as unconverted or converted in $Z \to ee$ and $H \to \gamma\gamma$ simulated events, and $\frac{f_i^{X\to\gamma\gamma}}{\sqrt{f_i^{Z\to ee}}}$ per category. The Figure of Merit is 1.71.

$f^{cat}_{Z\to ee}(\%)$	*subleading* γ		
	unconverted	single-track conv	double-track conv
leading γ			
unconverted	10.2	2.8	20.8
conv 1 track	3.4	1.2	6.2
conv 2 tracks	13.9	5.4	35.5
$f^{cat}_{H\to\gamma\gamma}(\%)$	*subleading* γ		
	unconverted	single-track conv	double-track conv
leading γ			
unconverted	50.5	8.3	12.9
conv 1 track	8.3	1.2	2.3
conv 2 tracks	11.7	2.0	3.0
$f_i^{X\to\gamma\gamma}/\sqrt{f_i^{Z\to ee}}$	*subleading* γ		
	unconverted	single-track conv	double-track conv
leading γ			
unconverted	1.58	0.50	0.28
conv 1 track	0.45	0.11	0.09
conv 2 tracks	0.31	0.08	0.05

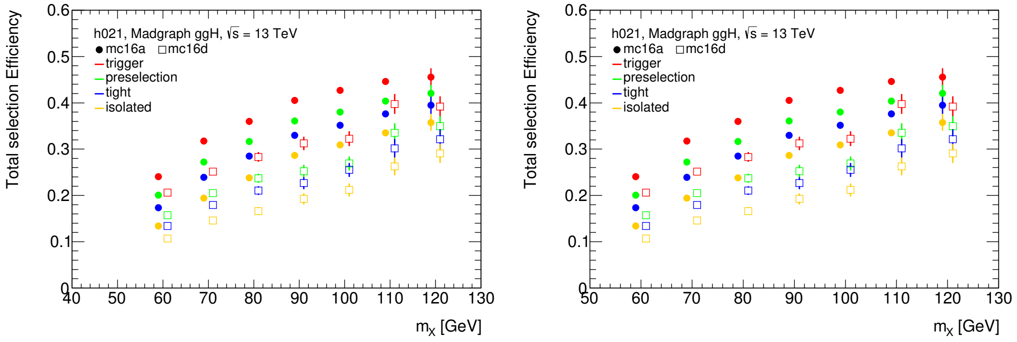
Table 5.7: Percentage of events containing photons reconstructed as unconverted, single-track converted or double-track converted in $Z \to ee$ and $H \to \gamma\gamma$ simulated events, and $\frac{f_i^{X\to\gamma\gamma}}{\sqrt{f_i^{Z\to ee}}}$ per category. The Figure of Merit is 1.78.

$f^{cat}_{Z\rightarrow ee}(\%)$	*subleading* γ					
	unconverted	singleSi	singleTRT	doubleSi	doubleTRT	doubleSiTRT
leading γ						
unconverted	10.2	1.4	1.4	19.7	0.5	0.6
singleSi	1.7	0.1	0.3	3.3	0.0	0.0
singleTRT	1.8	0.3	0.4	3.3	0.1	0.1
doubleSi	12.7	2.4	2.6	31.2	0.8	0.8
doubleTRT	0.8	0.1	0.1	0.8	0.0	0.0
doubleSiTRT	0.4	0.1	0.1	0.8	0.0	0.0
$f^{cat}_{H\rightarrow\gamma\gamma}(\%)$	*subleading* γ					
	unconverted	singleSi	singleTRT	doubleSi	doubleTRT	doubleSiTRT
leading γ						
unconverted	50.5	2.1	6.2	10.5	1.8	0.6
singleSi	2.0	0.1	0.3	0.5	0.1	0.0
singleTRT	6.3	0.3	0.5	1.4	0.3	0.1
doubleSi	9.3	0.4	1.1	1.9	0.3	0.1
doubleTRT	1.7	0.1	0.3	0.5	0.1	0.0
doubleSiTRT	0.6	0.0	0.1	0.2	0.0	0.0
$f_i^{X\rightarrow\gamma\gamma}/\sqrt{f_i^{Z\rightarrow ee}}$	*subleading* γ					
	unconverted	singleSi	singleTRT	doubleSi	doubleTRT	doubleSiTRT
leading γ						
unconverted	1.58	0.18	0.53	0.24	0.25	0.08
singleSi	0.16	0.01	0.05	0.03	0.16	0.01
singleTRT	0.48	0.06	0.08	0.08	0.08	0.02
doubleSi	0.26	0.03	0.07	0.03	0.03	0.02
doubleTRT	0.19	0.01	0.09	0.05	0.08	0.01
doubleSiTRT	0.10	0.00	0.03	0.02	0.00	0.02

Table 5.8: Percentage of events containing photons reconstructed as unconverted or one of the five flavors of conversion in $Z \rightarrow ee$ and $H \rightarrow \gamma\gamma$ simulated events, and $\frac{f_i^{X\rightarrow\gamma\gamma}}{\sqrt{f_i^{Z\rightarrow ee}}}$ per category. The Figure of Merit is 1.84

5.3.5 Selection Efficiency on Signal

The total and relative selection efficiency is summarized in Figure 5.10 where the pileup conditions corresponding to 2015+2016 and 2017 are compared. Most of the efficiency loss comes from the trigger selection (cf Table 5.4), in particular the 20 GeV or 22 GeV online cuts. The preselection cuts (i.e. asking for two loose photons with a tranverse energy higher than 22 GeV) also have a relative efficiency growing from 60 to 85% with increasing E_{T}. The offline *tight* efficiency is around 90%, and flat with E_{T}as there is already a *tight* requirement on the trigger. The overall efficiency goes from 14%/10% for $m_X = 60$ GeV (the limit will start at 65 GeV) to 32%/24% for $m_X = 100$ GeV, for 2016 and 2017 respectively. The difference between the two years comes from the increasing pile-up which degrades the photon identification and isolation efficiencies (already included at trigger level). The total efficiency is only of 2% for $m_X = 40$ GeV, and the trigger efficiency of 4%, meaning we would need to use a different trigger with lower E_{T}cuts in order to search below 60 GeV.



(a) (b)

Figure 5.10: (a) Absolute and (b) relative selection efficiency for signal, with mc16a (2015+2016) and mc16d (2017) samples for different mass points.

The fraction of events in each category is shown in Figure 5.11. It is rather flat with the mass, with average fractions of 50%, 42% and 9% for the **UU** , **UC** and **CC** respectively.

Those fractions are parameterized with second order polynomial functions:

$$\begin{aligned} f_{UU} &= 5.81186.10^{-01} - 1.31921.10^{-03} * x + 4.04665.10^{-06} * x^2 \\ f_{UC} &= 3.51254.10^{-01} + 9.60341.10^{-04} * x - 2.99837.10^{-06} * x^2 \\ f_{CC} &= 1 - f_{UU} - f_{UC} \end{aligned} \tag{5.6}$$

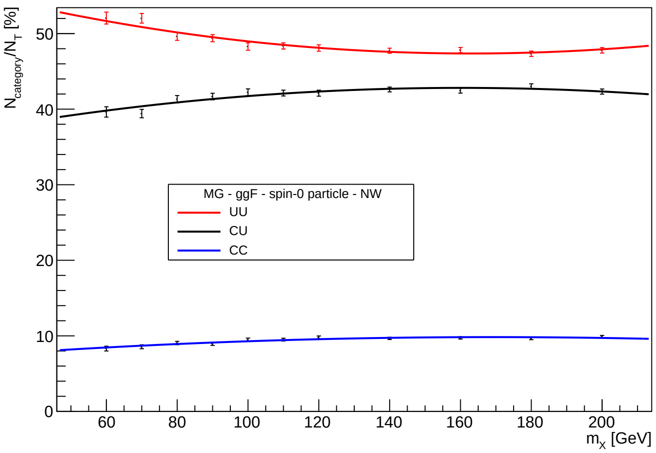


Figure 5.11: Fraction of conversion categories in terms of diphoton invariant mass. A second order polynomial fit is performed. The statistical errors are binomial.

5.3.6 Selection Efficiency on Data

The number of selected events in data is summarized in Table 5.9. The total number of selected events is 3414135. The use of an isolation requirement in the trigger for 2017 causes fewer numbers of events to pass the preselection in 2017 compared with 2016 where no isolation was used in preselection, despite the similar integrated luminosity. After isolation is applied in the full selection the numbers are comparable.

Year	2015	2016	2017
preselection	638392	6720477	4101517
tight	391529	4041009	2838686
isolated	152850	1564423	1696862
UU	49.9%	52.5%	47.8%
UC	41.0%	39.4%	42.1%
CC	9.1%	8.1%	9.9%

Table 5.9: Cut flow measured in data, in the $60 < M_{\gamma\gamma} < 120$ GeV mass window, for 2015, 2016, and 2017, as well as the fraction of events in each category.

5.4 Signal Modeling

5.4.1 Signal Parameterization

To model the signal shape, we use a functional form that is Gaussian near the center, but has tails that are non-Gaussian. The reason for this is that energy leakage effect primarily the lower tail. This causes the distribution to be asymmetric. To model this behavior we choose a double sided crystal ball function (DSCB). This function is used several $H \rightarrow \gamma\gamma$ type analyses. Because we use the narrow width approximation (NWA) where the width of the resonance is narrower than the energy resolution of the calorimeter, this function can work for different narrow resonance models.

The double-sided Crystal Ball function is defined as

$$N \cdot \begin{cases} e^{-t^2/2} & \text{if } -\alpha_{Low} \geq t \geq \alpha_{High} \\ \frac{e^{-0.5\alpha_{Low}^2}}{\left[\frac{\alpha_{Low}}{n_{Low}}\left(\frac{n_{Low}}{\alpha_{Low}}-\alpha_{Low}-t\right)\right]^{n_{Low}}} & \text{if } t < -\alpha_{Low} \\ \frac{e^{-0.5\alpha_{High}^2}}{\left[\frac{\alpha_{High}}{n_{High}}\left(\frac{n_{High}}{\alpha_{High}}-\alpha_{High}+t\right)\right]^{n_{High}}} & \text{if } t > \alpha_{High}, \end{cases} \tag{5.7}$$

where $t = \Delta m_X/\sigma_{CB}$, $\Delta m_X = m_X - \mu_{CB}$, N is a normalisation parameter, μ_{CB} is the peak of the Gaussian distribution, σ_{CB} represents the width of the Gaussian part of the function, α_{Low} (α_{High}) is the point where the Gaussian becomes a power law on the low (high) mass side, n_{Low} (n_{High}) is the exponent of this power law. A schematic representation of the DSCB is shown in figure 5.12.

Four of the six parameters are found to be mass dependent. n_{low} and n_{high} are not mass dependent and so will be modeled with constants. The remaining parameters will be modeled with a first order polynomial in the mass m_X. To set the parameters we use the MadGraph

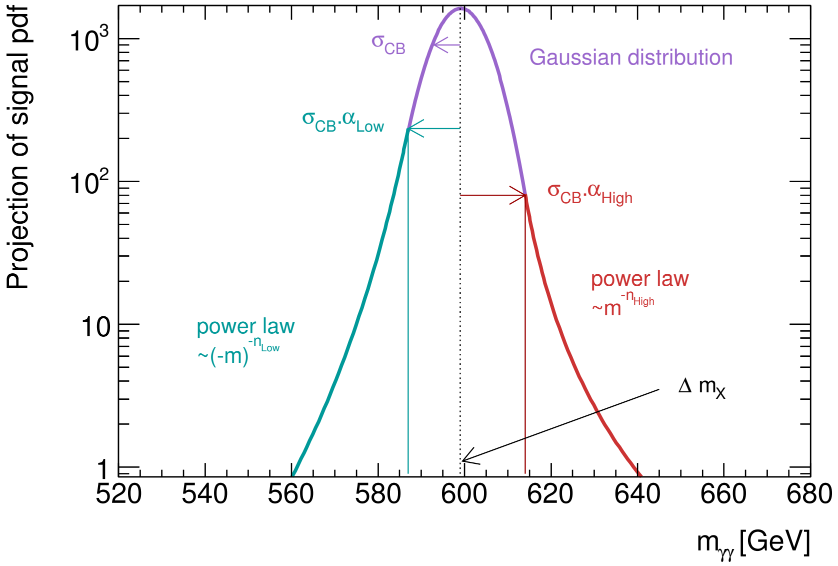


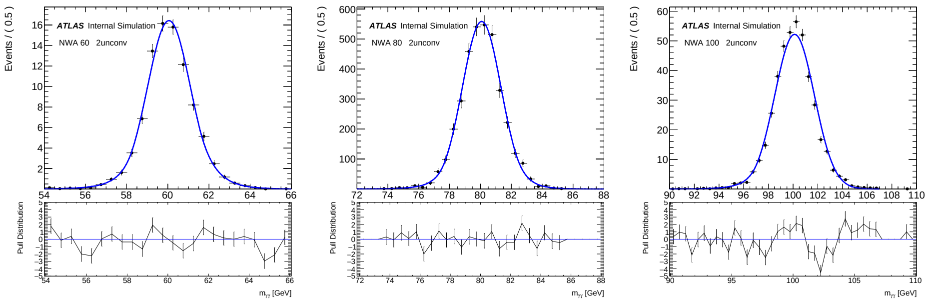
Figure 5.12: A schematic representation of the double-sided crystal ball function with a signal mass of m_X.

ggF MC signal sample. These samples are categorized according to the 3 categories UU, UC, and CC described in section 5.3.4.

To find the parameterizations as a function of m_X for the DSCB we use two fits: a single point fit, and a multiple, simultaneous fit. The single point fit involves taking each mass point signal sample separately and fitting with a DSCB. The DSCB parameters for each fit are then fit with a first order polynomial, or constant in the case of n_{low} and n_{high} to get the parameterizations of the DSCB as a function of m_X. The multiple, simultaneous fit takes all of the signal mass points and fits them at once using the first order polynomials directly in the fit. The multiple fit method will better pick up the correlations between some fit parameters and so this is used as the main method to obtain the parameterization. The single mass point fits are used as a validation.

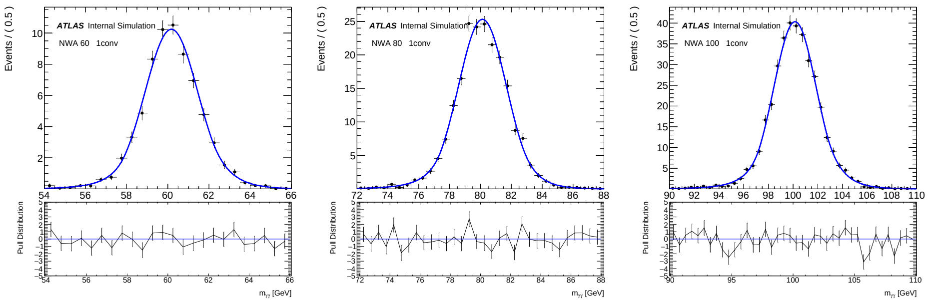
The single mass point fits for $m_X = 60$, 80, and 100 are shown in figures 5.13, 5.14, and 5.15

for the three conversion categories **UU** , **UC** , and **CC** respectively.



(a) (b) (c)

Figure 5.13: Fit of the $m_{\gamma\gamma}$ distributions for ggF samples at $m_X = 60$ (left), 80 (middle) and 100 GeV (right) to a double-sided Crystal Ball function for **UU** category. The bottom insets show the pulls in each bin (the difference between the simulated point and the fit, divided by the statistical uncertainty in each bin).

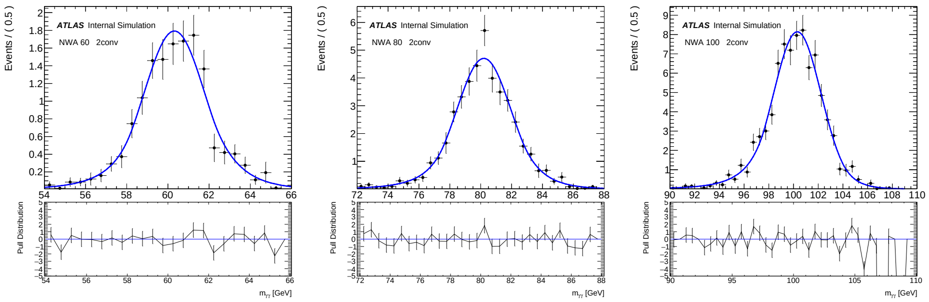


(a) (b) (c)

Figure 5.14: Fit of the $m_{\gamma\gamma}$ distributions for ggF samples at $m_X = 60$ (left), 80 (middle) and 100 GeV (right) to a double-sided Crystal Ball function for **UC** category. The bottom insets show the pulls in each bin (the difference between the simulated point and the fit, divided by the statistical uncertainty in each bin).

The results of the multiple fit are shown in table 5.10. The parameterization is shown as a function of $m_{nX} = \frac{m_X - 100}{100}$. These are the parameters that are used for the analysis. Figures 5.16-5.18 show the results of the parametrization of the multiple fit compared with the single fit points as well as the fitted line to the single fit mass points.

The correlation between all parameters in the multiple fit are shown in figure 5.19. The



(a) (b) (c)

Figure 5.15: Fit of the $m_{\gamma\gamma}$ distributions for ggF samples at $m_X = 60$ (left), 80 (middle) and 100 GeV (right) to a double-sided Crystal Ball function for **CC** category. The bottom insets show the pulls in each bin (the difference between the simulated point and the fit, divided by the statistical uncertainty in each bin).

strong correlation between several of the parameters shows that the multiple fit method should be used to make sure these relationships are fully captured by the fit.

Parameter	Parameterization	Coefficient	**UU**	**UC**	**CC**
Δm_X	$a + bm_{nX}$	a	0.06 ± 0.03	0.14 ± 0.04	0.20 ± 0.10
		b	0.01 ± 0.06	0.05 ± 0.078	0.03 ± 0.19
σ_{CB}	$a + bm_{nX}$	a	1.43 ± 0.03	1.70 ± 0.05	1.96 ± 0.11
		b	0.77 ± 0.06	0.81 ± 0.09	0.88 ± 0.20
α_{Low}	$a + bm_{nX}$	a	1.65 ± 0.12	1.58 ± 0.13	1.51 ± 0.23
		b	-0.03 ± 0.14	0.06 ± 0.13	-0.09 ± 0.26
n_{Low}	a	a	9.85 ± 3.75	6.74 ± 2.34	8.04 ± 5.91
α_{High}	$a + bm_{nX}$	a	1.45 ± 0.09	1.48 ± 0.11	1.51 ± 0.23
		b	0.16 ± 0.15	-0.05 ± 0.16	-0.09 ± 0.26
n_{High}	a	a	1468 ± 12943	1624811 ± 8566690	652 ± 17070

Table 5.10: Parameterizations of the double-sided Crystal Ball function describing the signal shape, result of the multiple mass point fit at low-mass, for the three categories. m_{nX} is defined as $m_{nX} = \frac{m_X - 100}{100}$.

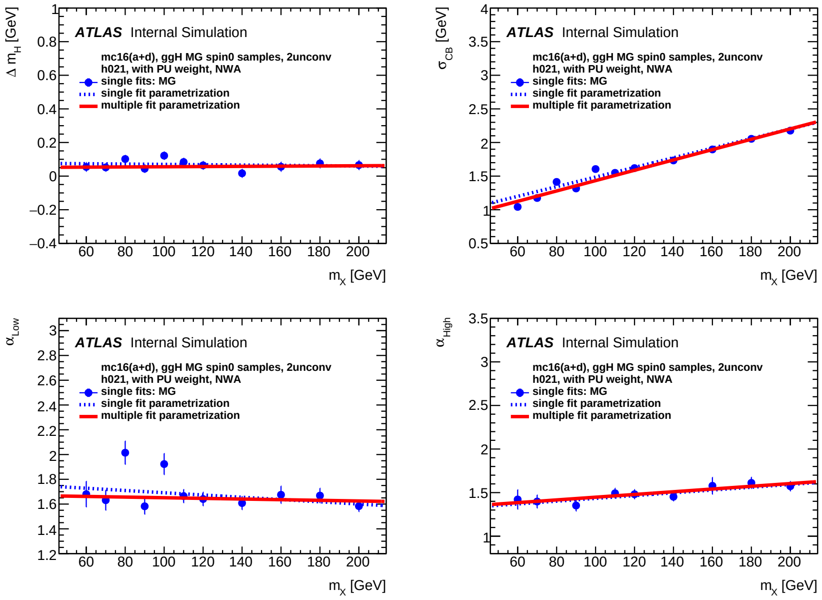


Figure 5.16: Result of the simultaneous fits of the $m_{\gamma\gamma}$ distributions for ggF samples from 40 to 200 GeV for the **UU** category. The four plots correspond to the parameters Δ_{m_H}, σ_{CB}, α_{low} and α_{high}. The red line corresponds to the result of the multiple fit while the dashed lines correspond to the linear fit of the individual fits.

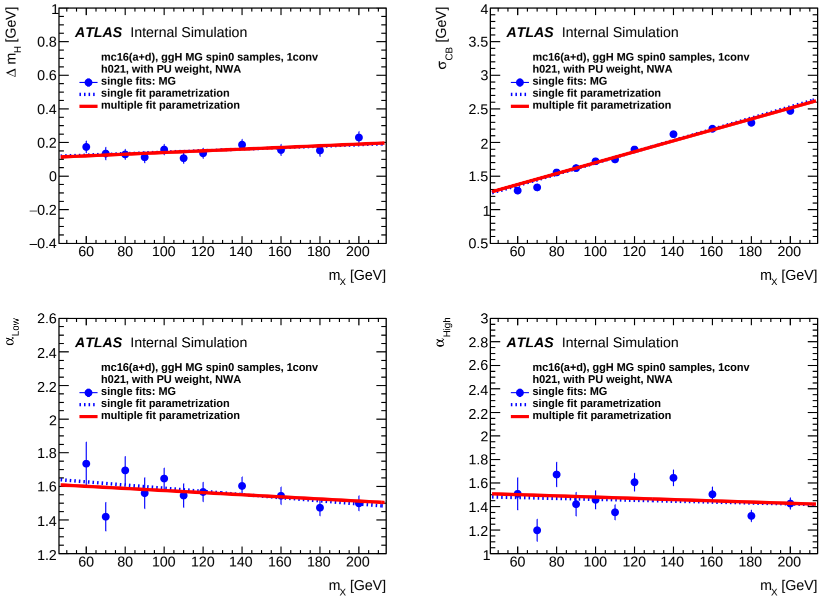


Figure 5.17: Result of the simultaneous fits of the $m_{\gamma\gamma}$ distributions for ggF samples from 40 to 200 GeV for the **UC** category. The four plots correspond to the parameters Δ_{m_H}, σ_{CB}, α_{low} and α_{high}. The red line corresponds to the result of the multiple fit while the dashed lines correspond to the linear fit of the individual fits.

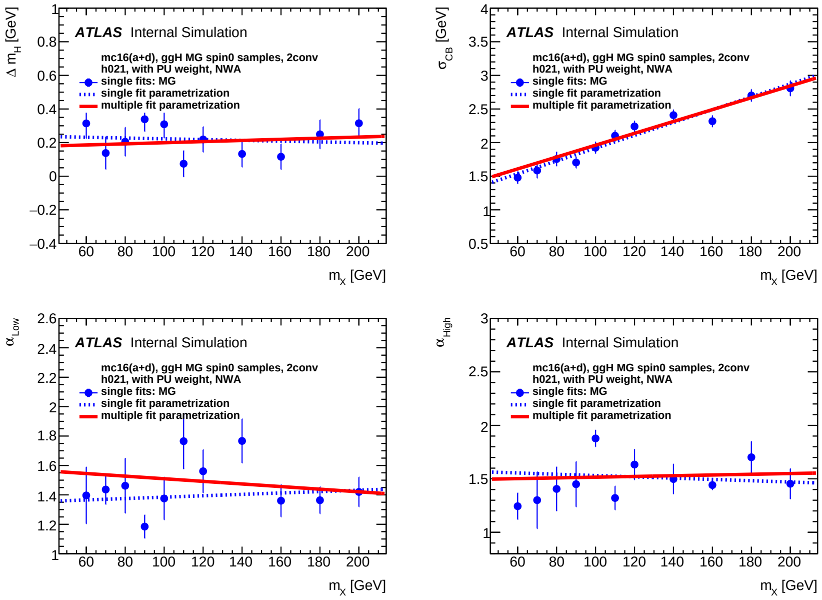


Figure 5.18: Result of the simultaneous fits of the $m_{\gamma\gamma}$ distributions for ggF samples from 40 to 200 GeV for the **CC** category. The four plots correspond to the parameters Δ_{m_H}, σ_{CB}, α_{low} and α_{high}. The red line corresponds to the result of the multiple fit while the dashed lines correspond to the linear fit of the individual fits.

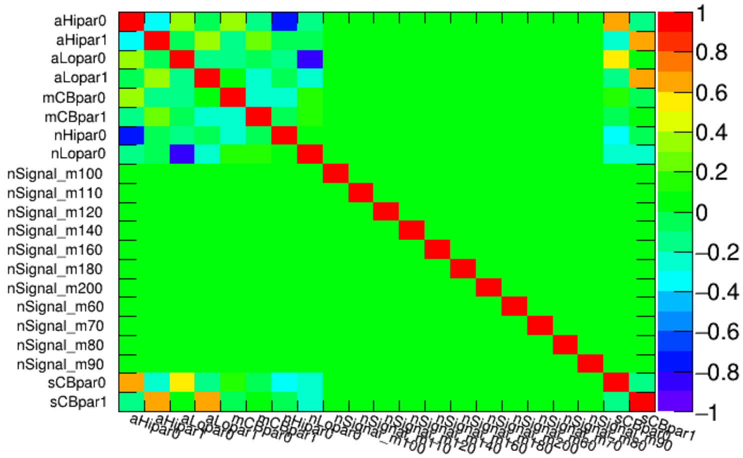


Figure 5.19: Correlation between different parameters of the global fit for ggF samples. The index "0" indicates the constant part of the linear polynomial.

5.5 Background Modeling

One of the most important and first steps in performing any analysis is modeling the backgrounds. ATLAS sees many thousands of interactions per second but most of these are processes that we are not interested in. However, since the experimental signature can look identical to the process that we do actually care about, it is often impossible to separate the uninteresting and interesting events. This means that we will have large backgrounds in our datasets and searching for new physics often involves looking for small signals on top of the background processes.

We expect to see two main types of background, non-resonant backgrounds and resonant ones. The non-resonant backgrounds come from QCD production of photon pairs ($\gamma\gamma$) shown in figure 5.20, photon and jet pairs (γj), and jet pairs (jj). This background appears as a smoothly falling distribution across the entire invariant mass range and is modeled with a fit function.

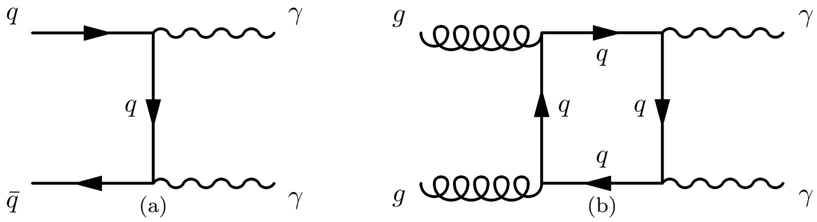


Figure 5.20: SM background processes that contribute to the non-resonant background. (a) is the $q\bar{q}$ process and (b) is the dominant ggF process.

The resonant background is due to the Drell-Yan process $Z \to ee$ where both of the electrons are misidentified as photons shown in figure 5.21.

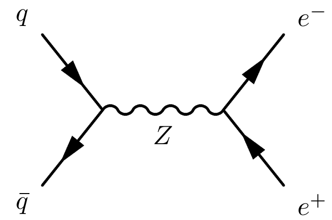


Figure 5.21: The Drell-Yan process $Z \rightarrow ee$. This process gives the resonant background where the electrons are misidentified as photons.

5.6 Non-Resonant Background

To model the non-resonant part of the background, we construct templates from background MC samples for the $\gamma\gamma$ part, and take the γjet component from control regions in data. These two components are added together according to their respective fractions as measured in the next section.

5.6.1 Background Composition

In order to properly build the templates used to find an approprite fit function we need to determine the relative contributions in the data from $\gamma\gamma$, γjet, and jetjet events. This will allow us to combine the appropriate samples with the correct fraction. The **2x2D** sideband method is used which is described in detail in [63]. This method involves extrapolating from background control regions in the side bands of the isolation and identification variables.

To begin, a sample is prepared from the data using a modification to the Tight ID criteria where all cuts on variable measured in the first layer of the EMCal are removed except for w_{stot}. This working point is known as **Loose'**. This sample is referred to as $L'L'$. In this sample the total yield of all events, $W_{tot}^{L'L'}$, will be the sum of all $\gamma\gamma$, γj, and jj events.

$$W_{tot}^{L'L'} = W_{\gamma\gamma}^{L'L'} + W_{\gamma j}^{L'L'} + W_{j\gamma}^{L'L'} + W_{jj}^{L'L'}. \tag{5.8}$$

The yields on the right of equation 5.8 are unknown. This sample is then divided into 16 orthogonal subsamples by whether each photon passes or fails the tight ID and isolation requirement. The signal region is defined by both the leading and subleading photon passing the tight ID and isolation which we denote by $TITI$. The remaining 15 control regions are where one or both photon candidates are required to fail either the ID or isolation or both. For example the region defined by the leading photon failing the tight ID and isolation is denoted $\bar{T}\bar{I}TI$ and corresponds to a part of the $W_{j\gamma}^{L'L'}$ component of equation 5.8. The total yield for any of the subsamples can be written as a function of the efficiencies to pass the tight and isolation requirement for photons, as a function of the fake rate for a jet to missidentified as a photon. There are also non-trivial correlations between the fake rates for jj events and so a factor for this must also be put in. The correlation between the other components is neglected. For the $TITI$ signal region, this is,

$$\begin{aligned} N_{\text{TITI}} &= W_{\gamma\gamma}^{L'L'}\epsilon_{T1}\epsilon_{I1}\epsilon_{T2}\epsilon_{I2} \\ &+ W_{\gamma j}^{L'L'}\epsilon_{T1}\epsilon_{I1}f_{I2}f_{T2} \\ &+ W_{j\gamma}^{L'L'}f_{I1}f_{T1}\epsilon_{T2}\epsilon_{I2} \\ &+ W_{jj}^{L'L'}f'_{I1}f'_{T1}f'_{I2}f'_{T2}\xi_{Ijj}\ , \end{aligned} \tag{5.9}$$

where

- $\epsilon_{I1}$and $\epsilon_{I2}$are the efficiencies of the isolation criteria of one of the six analysis under study for the leading and subleading photons respectively. They are determined from the diphoton simulation.
- $\epsilon_{T1}$and $\epsilon_{T2}$are the Tight identification efficiencies for the leading and subleading photons respectively, also determined from the diphoton simulation.
- $f_{I1}$and $f_{I2}$are the isolation fake rates for the γj and $j\gamma$ events, fitted directly on data.

- $f_{T1}$and $f_{T2}$are the Tight identification fake rates for the γj and $j\gamma$ events, fitted directly on data.
- $f'_{I1}$and $f'_{I2}$are the isolation fake rates for the jj events, fitted directly on data.
- $f'_{T1}$and $f'_{T2}$are the Tight identification fake rates for the jj events, fitted directly on data.
- $\xi_{Ijj}$is the isolation correlation factor between the jets in the jj events, fitted directly on data.

As an example of one of the control regions, equation 5.10 shows $TI\bar{T}\bar{I}$, is the subsample where the leading photon is required to pass both tight ID and isolation, but the subleading is required to fail both. The remaining 14 equations are listed in appendix A.1.

$$\begin{aligned} N_{TI\bar{T}\bar{I}} = {} & W_{\gamma\gamma}^{L'L'}\epsilon_{I1}\epsilon_{T1}\left(1-\epsilon_{I2}\right)\left(1-\epsilon_{T2}\right) \\ & + W_{\gamma j}^{L'L'}\epsilon_{I1}\epsilon_{T1}\left(1-f_{I2}-f_{T2}+f_{I2}f_{T2}\right) \\ & + W_{j\gamma}^{L'L'}f_{I1}f_{T1}\left(1-\epsilon_{I2}\right)\left(1-\epsilon_{T2}\right) \\ & + W_{jj}^{L'L'}f'_{I1}f'_{T1}\left(1-\xi_{Ijj}f'_{I2}-f'_{T2}+f'_{I2}f'_{T2}\xi_{Ijj}\right) \end{aligned} \tag{5.10}$$

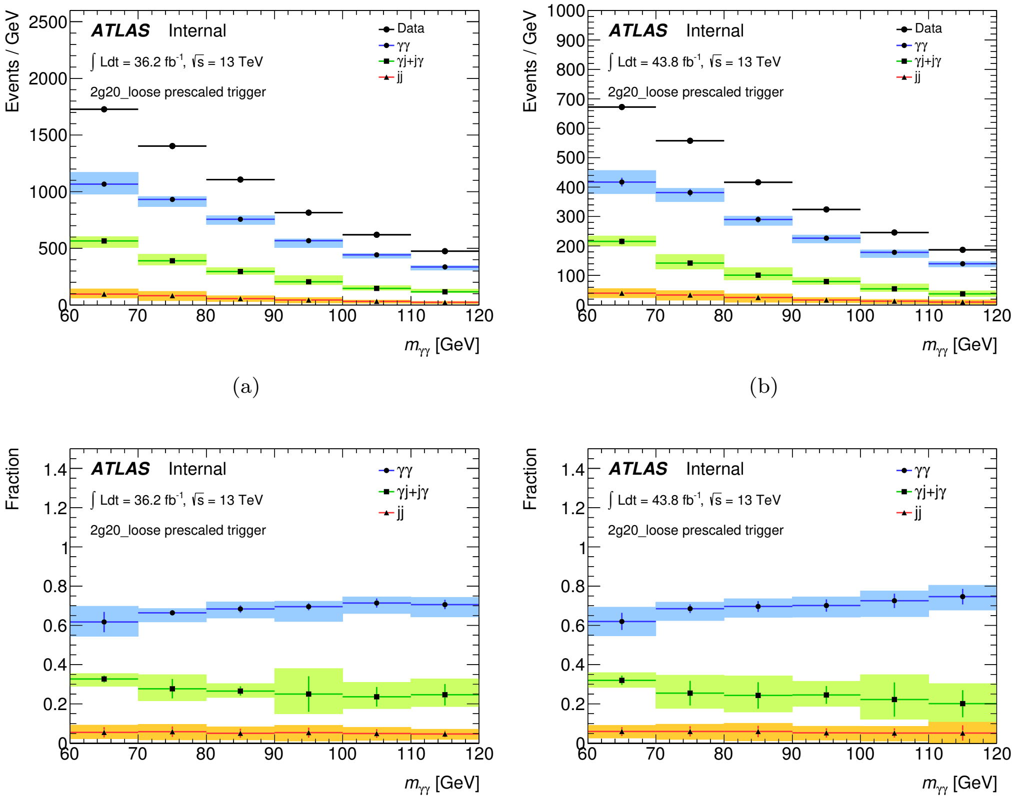
With the efficiencies and isolations measured directly on the Sherpa $\gamma\gamma$ MC sample the remaining unknowns are the fractions ($W_{\gamma\gamma}^{L'L'}$, $W_{\gamma j}^{L'L'}$, $W_{j\gamma}^{L'L'}$, $W_{jj}^{L'L'}$), the fake rates (f_{T1}, f_{T2}, f_{I1}, f_{I2}, f'_{T1}, f'_{T2}, f'_{I1}, f'_{I2}), and the jj correlation (ξ_{Ijj}). These are all determined by a fit to data. Table 5.11 shows the efficiencies measured on MC used as inputs to the fit. The results of the fits are shown in table 5.12 for each of the categories **UU** , **UC** , **CC** as well as the inclusive sample. The fraction of each component are shown as a function of invariant mass in figures 5.22-5.24.

Category	Year	Isolation		Identification	
Inclusive	2015+2016 (mc16a)	ϵ_{I1}	0.9291 ± 0.0004	ϵ_{T1}	0.9520 ± 0.0003
		ϵ_{I2}	0.8664 ± 0.0005	ϵ_{T2}	0.9289 ± 0.0004
	2017 (mc16d)	ϵ_{I1}	0.9043 ± 0.0006	ϵ_{T1}	0.9446 ± 0.0004
		ϵ_{I2}	0.8290 ± 0.0007	ϵ_{T2}	0.9219 ± 0.0005
UU	2015+2016 (mc16a)	ϵ_{I1}	0.9423 ± 0.0005	ϵ_{T1}	0.9480 ± 0.0005
		ϵ_{I2}	0.8829 ± 0.0007	ϵ_{T2}	0.9288 ± 0.0005
	2017 (mc16d)	ϵ_{I1}	0.9174 ± 0.0008	ϵ_{T1}	0.9406 ± 0.0006
		ϵ_{I2}	0.8464 ± 0.0009	ϵ_{T2}	0.9210 ± 0.0007
UC	2015+2016 (mc16a)	ϵ_{I1}	0.9198 ± 0.0006	ϵ_{T1}	0.9548 ± 0.0005
		ϵ_{I2}	0.8557 ± 0.0008	ϵ_{T2}	0.9286 ± 0.0006
	2017 (mc16d)	ϵ_{I1}	0.8965 ± 0.0009	ϵ_{T1}	0.9469 ± 0.0006
		ϵ_{I2}	0.8196 ± 0.0011	ϵ_{T2}	0.9220 ± 0.0007
CC	2015+2016 (mc16a)	ϵ_{I1}	0.9000 ± 0.0015	ϵ_{T1}	0.9608 ± 0.0010
		ϵ_{I2}	0.8275 ± 0.0018	ϵ_{T2}	0.9314 ± 0.0012
	2017 (mc16d)	ϵ_{I1}	0.8741 ± 0.0021	ϵ_{T1}	0.9545 ± 0.0013
		ϵ_{I2}	0.7864 ± 0.0024	ϵ_{T2}	0.9256 ± 0.0015

Table 5.11: Isolation and identification efficiencies for true photons used as input to the 2x2D sideband method for the measurement of the diphoton purity in the $[60, 120]$ GeVmass range of the sample passing the full selection, for the years 2015+2016 (mc16a) and 2017 (mc16d). The efficiencies are determined with respect to the leading and subleading photon candidates of true diphoton events that pass the full selection except the isolation and tight identification criteria, which are removed; the photons must pass instead the (looser) L' identification requirements. The uncertainty arises from the MC statistics.

Category	component	fraction	
		2015+2016	2017
Inclusive	$\gamma\gamma$	$0.666 \pm 0.004^{+0.022}_{-0.034}$	$0.673 \pm 0.006^{+0.027}_{-0.041}$
	γj	$0.183 \pm 0.002^{+0.027}_{-0.019}$	$0.178 \pm 0.004^{+0.039}_{-0.030}$
	$j\gamma$	$0.096 \pm 0.002^{+0.016}_{-0.013}$	$0.086 \pm 0.003^{+0.029}_{-0.024}$
	jj	$0.056 \pm 0.001^{+0.024}_{-0.029}$	$0.063 \pm 0.001^{+0.031}_{-0.037}$
UU	$\gamma\gamma$	$0.673 \pm 0.006^{+0.011}_{-0.036}$	$0.688 \pm 0.010^{+0.018}_{-0.047}$
	γj	$0.179 \pm 0.003^{+0.030}_{-0.018}$	$0.175 \pm 0.005^{+0.037}_{-0.026}$
	$j\gamma$	$0.094 \pm 0.003^{+0.010}_{-0.006}$	$0.080 \pm 0.005^{+0.027}_{-0.016}$
	jj	$0.054 \pm 0.001^{+0.022}_{-0.025}$	$0.057 \pm 0.002^{+0.023}_{-0.028}$
UC	$\gamma\gamma$	$0.658 \pm 0.006^{+0.035}_{-0.035}$	$0.661 \pm 0.009^{+0.028}_{-0.033}$
	γj	$0.186 \pm 0.004^{+0.020}_{-0.017}$	$0.181 \pm 0.005^{+0.029}_{-0.025}$
	$j\gamma$	$0.099 \pm 0.003^{+0.018}_{-0.015}$	$0.093 \pm 0.005^{+0.031}_{-0.025}$
	jj	$0.057 \pm 0.001^{+0.020}_{-0.025}$	$0.065 \pm 0.002^{+0.029}_{-0.034}$
CC	$\gamma\gamma$	$0.649 \pm 0.012^{+0.042}_{-0.031}$	$0.654 \pm 0.017^{+0.042}_{-0.021}$
	γj	$0.184 \pm 0.008^{+0.022}_{-0.024}$	$0.179 \pm 0.011^{+0.050}_{-0.043}$
	$j\gamma$	$0.108 \pm 0.007^{+0.022}_{-0.010}$	$0.093 \pm 0.010^{+0.034}_{-0.047}$
	jj	$0.059 \pm 0.003^{+0.022}_{-0.031}$	$0.074 \pm 0.004^{+0.043}_{-0.050}$

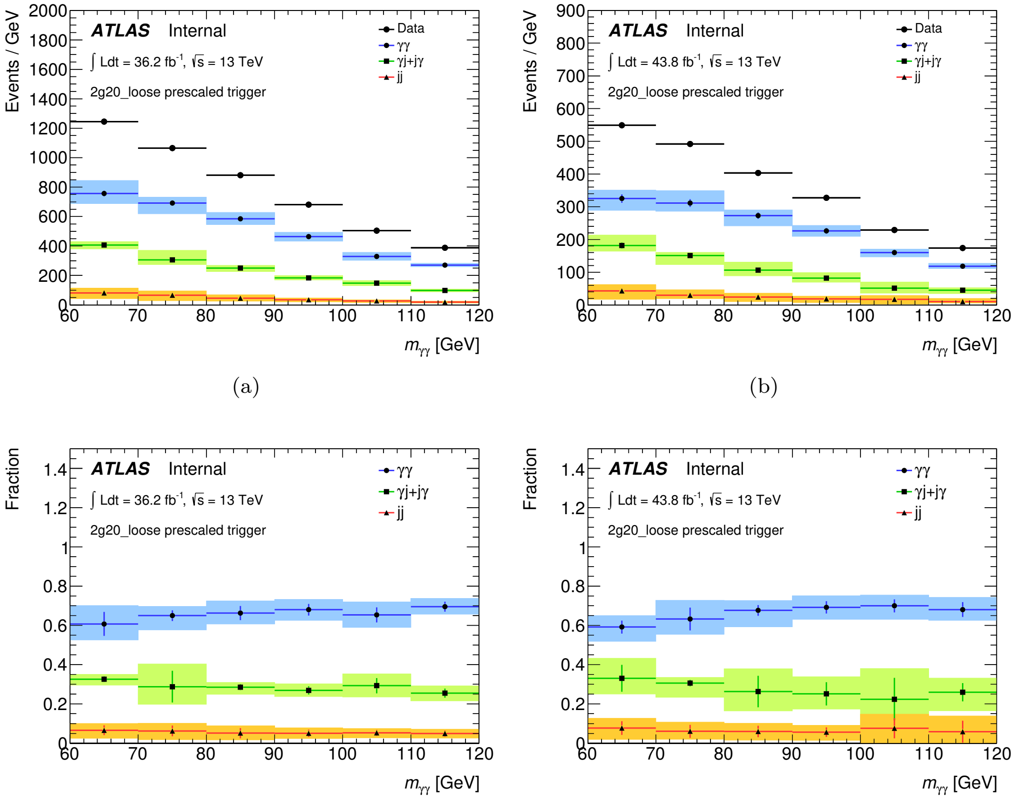
Table 5.12: Signal and background fraction in the mass range $[60, 120]$ GeV, obtained with the 2x2D sideband method, for the year 2015+2016 and 2017 inclusively and in conversion category.



(a) (b)

(c) (d)

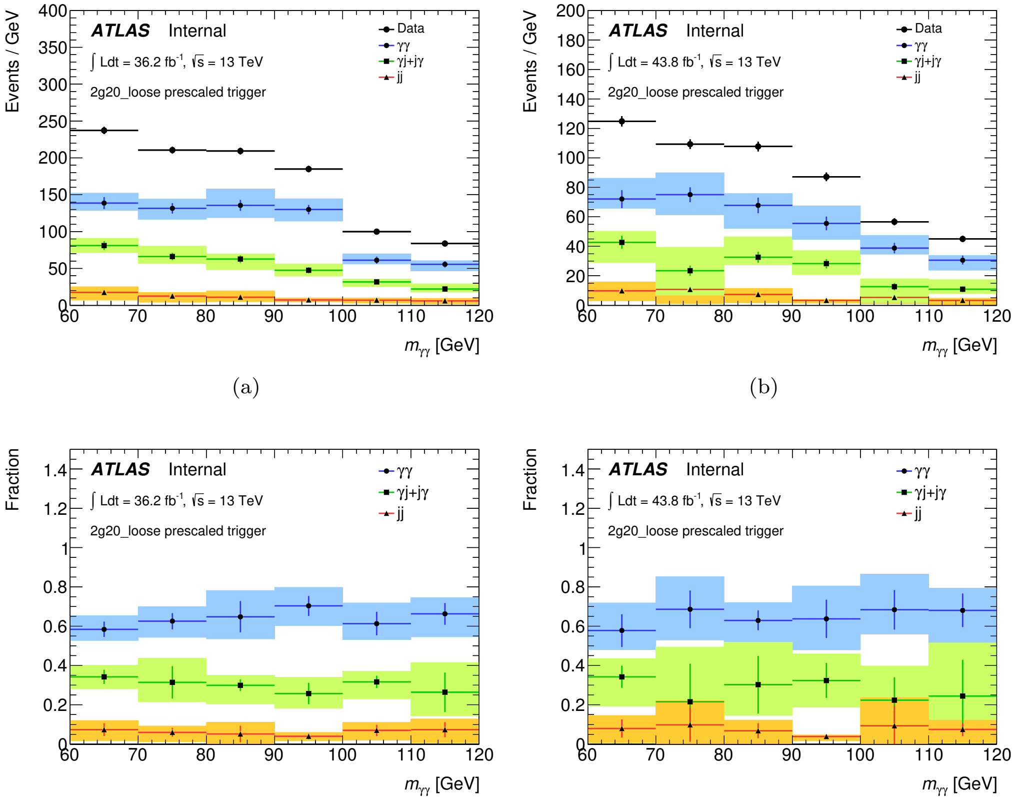
Figure 5.22: Data and $\gamma\gamma$, γj and jj yields for the **UU** category, determined by the 2x2D sideband method as a function of the diphoton mass, for the years 2015+2016 (a) and 2017 (b). The resulting fractions are shown below and for 2015+2016 (c) and 2017 (d).



(a) (b)

(c) (d)

Figure 5.23: Data and $\gamma\gamma$, γj and jj yields for the **UC** category, determined by the 2x2D sideband method as a function of the diphoton mass, for the years 2015+2016 (a) and 2017 (b). The resulting fractions are shown below and for 2015+2016 (c) and 2017 (d).



(a) (b)

(c) (d)

Figure 5.24: Data and $\gamma\gamma$, γj and jj yields for the **CC** category, determined by the 2x2D sideband method as a function of the diphoton mass, for the years 2015+2016 (a) and 2017 (b). The resulting fractions are shown below and for 2015+2016 (c) and 2017 (d).

5.6.2 Background Samples and Templates

Modeling the diphoton backgrounds we find in this resonance search requires the use of Monte Carlo samples in addition to samples taken from regions of data outside of our search region. Modeling the background with data only can be tricky. Before performing our analysis we cant know whether signal exists in our selection or not. Modeling using data directly runs the risk of absorbing the signal into what we are classifying as background. For this reason, the main SM $\gamma\gamma$ backgrounds are modeled using Monte Carlo where we can be certain that the sample contains only background processes. With about 35% of the non-resonant background coming from jets faking photons, we must also account for the γj and $j\gamma$ background. The jj background is small compared to the others and so will be ignored. Following a strategy to section 5.6.1, data is used to model the jet components of the backgrounds.

The $\gamma\gamma$ component of the background comes from the Sherpa NLO sample listed in table 5.2. Although this is a high-statistics sample, the MC/data ratio is not as high as it has been in past iterations of this analysis. This means that there is more statistical fluctuation in the sample relative to the data than we would like. The weights used in this sample can contribute to this fluctuation as well. The four main weights used for this MC sample are shown in fig 5.25. IF there are weights that have large fluctuations we would like to be able to not apply them in order to minimize the uncertainty as much as possible. The photon ID SF and MC weight both have a strong dependence on $M_{\gamma\gamma}$ so we must retain these or risk sculpting the invariant mass distribution. The other two weights however, are relatively flat in $M_{\gamma\gamma}$ and what little dependence they do have is smaller than the statistical uncertainty on the MC sample (see section 5.6.3 for the distributions). These two weights can safely be removed.

While performing the spurious signal test described in the next section, it was noticed that

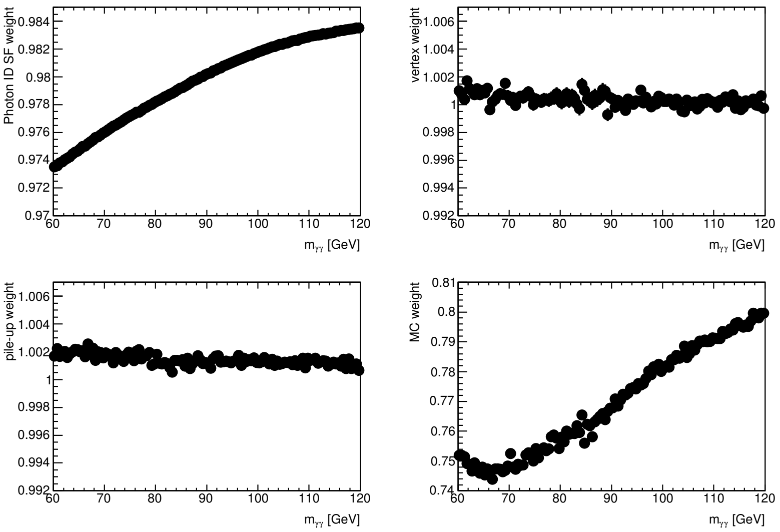
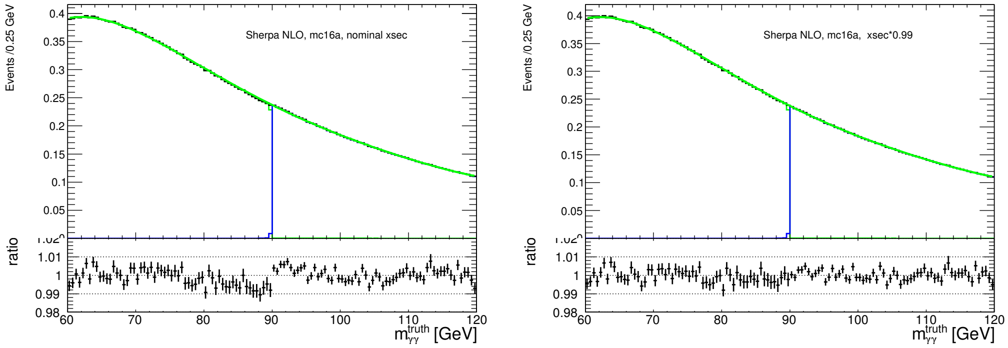


Figure 5.25: Profile of the weights applied to the Sherpa NLO diphoton simulation, as a function of $M_{\gamma\gamma}$. Only the preselection is applied. The four plots correspond to the Photon ID, vertex, pile-up and MC weights.

the normalization between the [50,90]GeV and [90,175]GeV samples was slightly different. The spurious signal test tends to be very sensitive to these kinds of shifts. Sherpa experts confirmed that the uncertainty on the cross-section of the samples is on the order of 1%. To fix this normalization, a 5th order polynomial was fit to the distribution with different weights applied to one slice of the sample. The χ^2/ndf was compared for each correction to find the best normalization between the two slices.

- nominal cross-section: $\chi^2/ndf = 241.3/114$;
- 0.99 correction: $\chi^2/ndf = 134.596/114$;
- 0.98 correction: $\chi^2/ndf = 143.65/114$.

Based on these results we apply a correction of 0.99 to the higher mass slice. The $M_{\gamma\gamma}$ distributions before and after the correction to the cross-section is applied is shown in figure 5.26



(a) (b)

Figure 5.26: Invariant diphoton mass distribution computed with the Sherpa NLO sample before (a) and after (b) the correction of cross-section, fitted with a fifth order polynomial function. The bottom panels show the ratio between the histograms and the fits.

In order to get the γj and $j\gamma$ components of the templates, we must find a control region in data that allows us to extract jet enriched samples. In the following, only 2017 data is used to construct the templates and it is assumed that the difference in shape of the 2015 and 2016 data in the invariant mass distribution is negligible.

Three control regions are tested:

- CR1: the two photons pass the *tight* identification cuts, one photon passes the isolation cuts and one fails it. It uses the trigger of the analysis (2g20_tight_icalovloose, corresponding to 43.0 fb^{-1}). It has the advantage of being very close from the signal region and the bias from inverting the cuts is minimal. But it has a high contamination from the diphoton background: the estimated fraction of $\gamma\gamma$ from MC studies is of 25%.

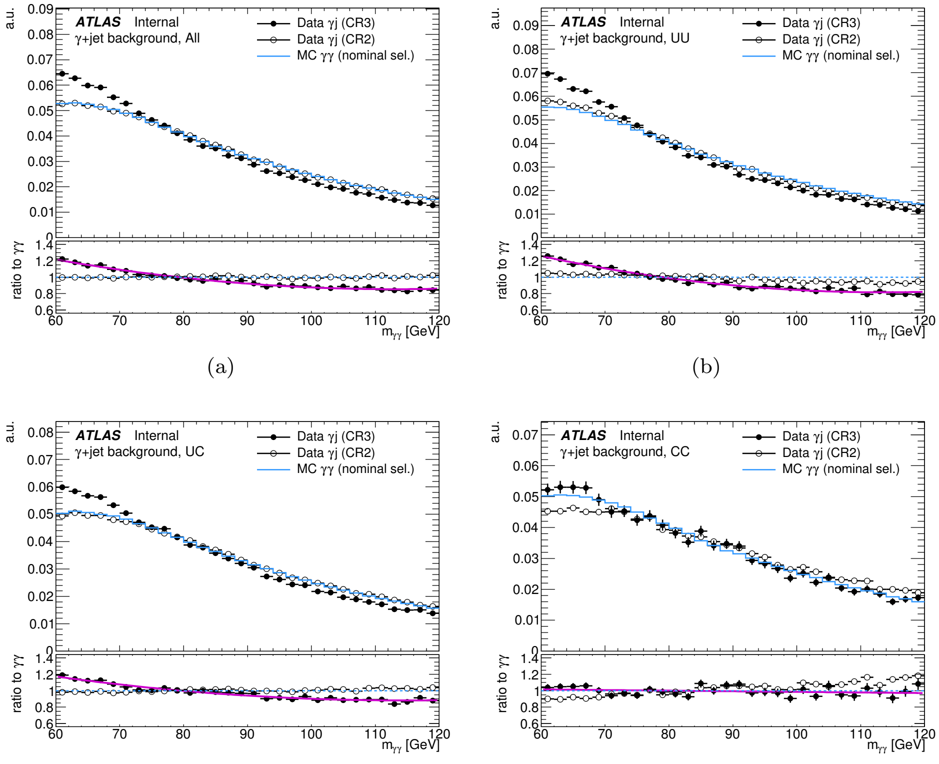
- CR2: one photon passes the *tight* identification and the isolation cuts, while the other one passes the *loose* identification but fails both the *tight* identification and the isolation cuts. It uses the 2g20_loose trigger since the online and offline *tight* identifications are very close, corresponding to 1.54 fb^{-1}. It has the advantage of a minimal contamination from the diphoton background (less than 10%). But this CR is far from the signal region so the shape could be biased by the inversion of cuts.

- CR3: the two photons pass the isolation cuts, one passes the *tight* identification, the other one passes the *loose* identification but fails the *tight* identification. It uses data from the 2g20_loose trigger and has the advantage of being close to the signal region. But the diphoton contamination is high, around 25%.

The first control region, CR1, is very close to the signal region in data which causes it to have a very high contamination of Drell-Yan background. This would make if very difficult to accurately model the γj shape without first removing this background. Removing the Drell-Yan would require an entire modeling study on this control region which is not feasible. This control region is not considered further.

CR3 is the closest to the signal region as it has the closest cuts. This will be used as the base line for the templates. Because CR3 uses the 2g20_loose trigger its statistics are small compared with the high-stats Sherpa $\gamma\gamma$ sample. We are unable to use CR3 to directly sum with the MC as this would cause the fluctuations in the resulting template to be very large. Instead, a ratio of the $\gamma\gamma$ MC and CR3 are taken. This ratio is then fit with a second order polynomial for the **UU** and **UC** categories, and fit with a first order polynomial for the **CC** category. This fit can then be used to reweight the $\gamma\gamma$ MC sample in order to reproduce the γj shape in high-statistics. This γj template will then be summed with the nominal $\gamma\gamma$ MC sample according to their respective purity fractions found in section 5.6.1. Figures 5.27 and 5.28 show the invariant mass shapes for the CR2 and CR3 control regions compared with the Sherpa $\gamma\gamma$ sample as well as the ratio of the control region to MC for each category.

Several templates are also created with variations. These are to asses the impact of the choice of fit function as well as the selection of the control region. No systematic uncertainties are directly applied to the templates.

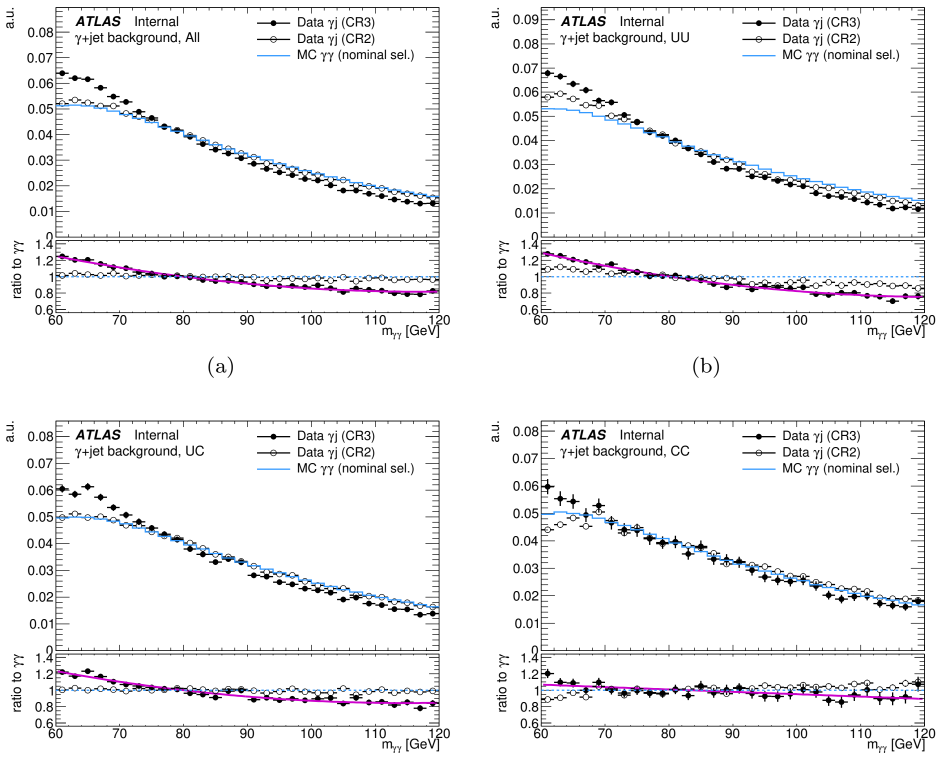


(a) (b)

(c) (d)

Figure 5.27: $M_{\gamma\gamma}$ shapes for the data-driven CR2 and CR3 and from the diphoton MC sample for the three categories, made with 2015+2016 data and mc16a diphoton MC for inclusive(a), **UU** (b) , **UC** (c), and **CC** (d). All histograms are normalized to the same area. The bottom panels correspond to the ratio to the diphoton MC shape. The pink line corresponds to a first/second order polynomial fit to the CR3/$\gamma\gamma$ MC ratio.

- the $\gamma\gamma$ purity is increased by 5% (conservative since the relative uncertainty on the measured purity varies between 0.6 and 2.6%).
- the $\gamma\gamma$ purity is decreased by 5% (conservative since the relative uncertainty on the measured purity varies between 0.6 and 2.6%).
- the smoothing of the reducible background is changed: the Sherpa $\gamma\gamma$ / CR3 ratios fitted by second(first) order polynomial are fitted by first(second) order polynomial.
- the CR2 is used instead of the CR3 for the shape of the reducible background.



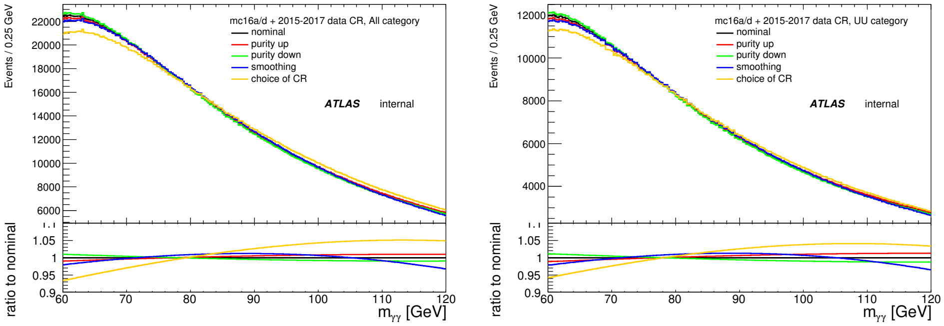
(a) (b)

(c) (d)

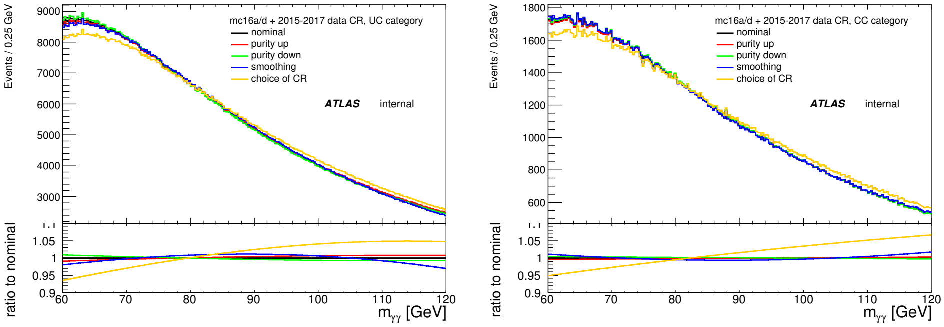
Figure 5.28: $M_{\gamma\gamma}$ shapes for the data-driven CR2 and CR3 and from the diphoton MC sample for the three categories, made with 2017 data and mc16d diphoton MC for inclusive (a), **UU** (b) , **UC** (c), and **CC** (d). All histograms are normalized to the same area. The bottom panels correspond to the ratio to the diphoton MC shape. The pink line corresponds to a first/second order polynomial fit to the CR3/$\gamma\gamma$ MC ratio.

The templates with those different variations are showed in Figure 5.29, as well as the ratio to the nominal template. The impact of the fraction of irreducible and reducible backgrounds and the choice of the function for the smoothing of the reducible background is small (2% at most), while using a different CR can give a difference up to 5%. The Spurious Signal method will be used on the five samples in order to test the robustness of the chosen function.

A closure test is also performed on the resulting templates to validate them against the signal region in data. The 2g20_loose trigger is used which corresponds to about 10% of the



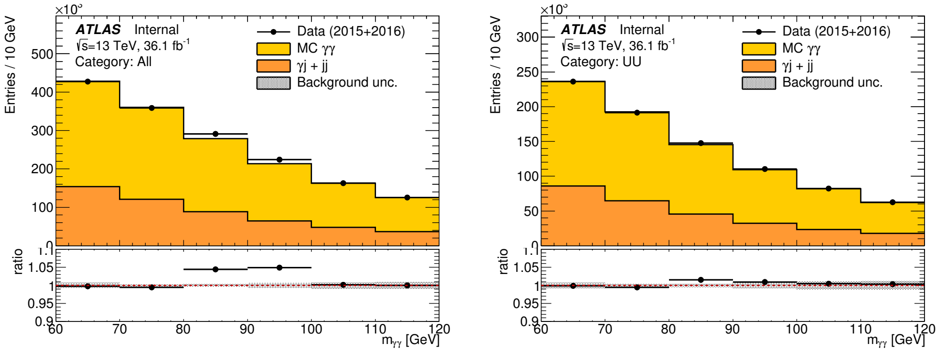
(a) Inclusive (b) **UU** category



(c) **UC** category (d) **CC** category

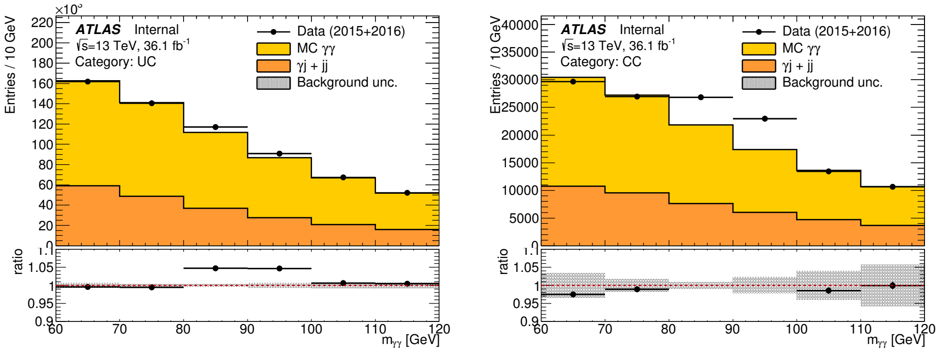
Figure 5.29: Non-resonant background templates built from the irreducible and reducible components for the inclusive case and the three categories, made with the mc16a+mc16d diphoton MC and data-driven components from 2015-2017 data. The black lines correspond to the nominal template and the other ones to variations of the templates. The bottom panels show the ratios to the nominal template.

total statistics in data. A wide binning of 10GeV in $M_{\gamma\gamma}$ is also used in order to obscure any signal that may exist in this data. Figures 5.30 and 5.31 shows the templates compared with the signal region data. The Drell-Yan peak is clearly visible in the data especially in the **CC** category. As these templates do not model the resonant background we only check the agreement outside of the peak region. The agreement with data is good.



(a) Inclusive

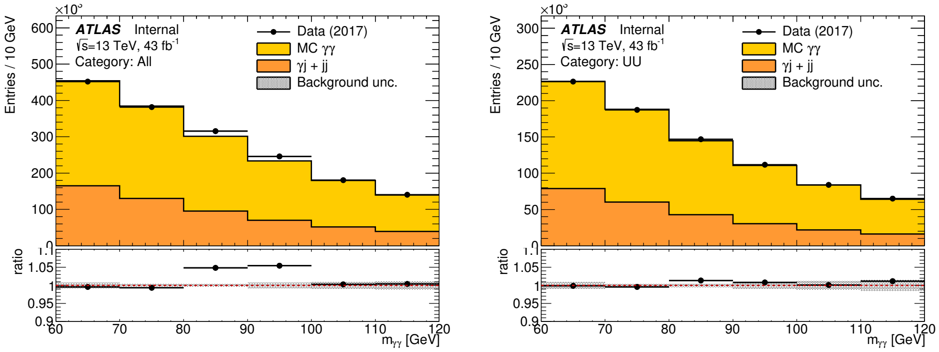
(b) **UU** category



(c) **UC** category

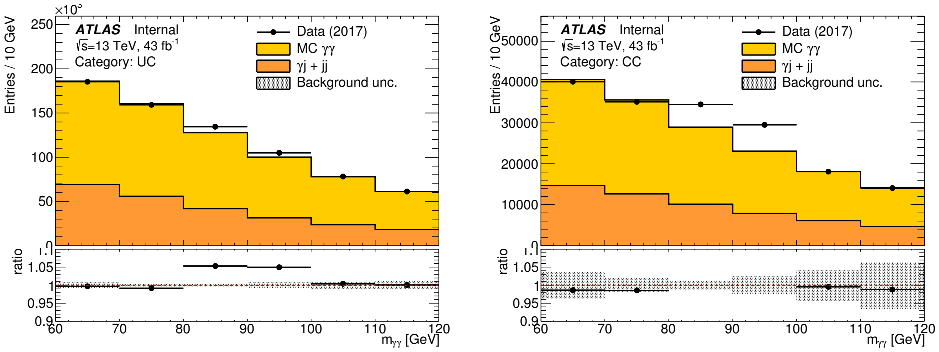
(d) **CC** category

Figure 5.30: Invariant mass distribution from 2015+2016 data (black points), and the non-resonant template. The grey bands correspond to the envelope of all variations.



(a) Inclusive

(b) **UU** category



(c) **UC** category

(d) **CC** category

Figure 5.31: Invariant mass distribution from 2017 data (black points), and the non-resonant template. The grey bands correspond to the envelope of all variations.

5.6.3 Spurious Signal Tests

In order to fit a function to a distribution, we first must select a function. This is not always as easy as it seems since most times in physics we don't know what the underlying process is generating the distribution. Without an exact functional form, we must choose one that we think approximates the distribution well. A common test in ATLAS for the suitability of a function to model the data is the spurious signal test. This involves fitting your chosen background function plus a function for your signal model on a Monte Carlo sample that contains only background events. Because we have limited resources with which to generate the MC samples, there will be statistical fluctuation in them. This will cause the signal part of the fit to report non-zero values. Because we are using a sample that has no signal in this, we call this the Spurious Signal. Obviously when we go to search for real signal in our data we would hope that our chosen functions do not give us any spurious results and so we would like them to have low values of the Spurious Signal when tested on background only samples. One way we can try to select the right function to describe the background is to test several functions against each-other, then select the one that has the lowest spurious signal, or that meets some other criteria. Which functions to test are ad-hoc and several functions that have the general behavior we see in the background distribution are chosen.

The procedure for choosing a fit function with the spurious signal test follows from the procedure used in the $H \rightarrow \gamma\gamma$ cross section analysis [62]. I will start by detailing the nominal spurious signal test that has been used in past analyses and then discuss changes made because of changes in the run 2 analysis.

- Prepare a high statistics background only sample of MC simulation.
- Fit this distribution with a signal plus background function and measure the yield from the signal part $N_{spur,i}$.

- For each mass point in 1 GeV steps, repeat this process.
- Find the maximum value of $N_{spur,i}$ for all mass points in the range. This is called the size of the spurious signal N_{spur}

Normally a criteria is adopted to determine which function passes the test. In the run 1 low mass analysis, the criteria was that the size of the spurious signal N_{spur} must be lower than 20% of the background statistical uncertainty. In the standard model Higgs search, they also require that the size of N_{spur} be lower than some percentage of the expected signal, however, since for this analysis there is no expected signal, this criteria was ignored.

Several fit functions may satisfy the criteria and so the one with the fewest degrees of freedom is chosen. The size of the spurious signal will also be used later in the statical model for the search as a Gaussian constraint using a width of N_{spur}.

The fit range for the low-mass search is $60 < m_{\gamma\gamma} < 120$ GeV. This range is selected to search below the SM Higgs mass but above the kinematic turn on caused by the 20 GeV and 22 GeV triggers used.

The MC sample used is a $\gamma\gamma$ only background sample produced with Sherpa. Several files are produced in ranges of $m_{\gamma\gamma}$ and are then summed and weighted by their cross section. This sample is then re-weighted by the shape extracted from γ+Jet data described in section 5.6.2

The fit functions chosen to test are:

- Fermi distribution: $f(x) = exp(-x/\mu)/(1 + exp(-(x - m_0)/\sigma))$
- Bernstein polynomials: $f(x) = \Sigma_{i=0}^{n} p_i u^i (1 - u^i)$, where $u = (x - x_{min})/(x_{max} - x_{min})$ and n=4 to n=7
- A Landau function
- The sum of a Landau and exponential function.

For each conversion category the function is fit in the range [60, 120] GeV and the spurious signal is measured in the range [65,110] GeV. For the run 1 analysis, the function that passes the above criteria and has the fewest degrees of freedom is chosen to model the background.

For the run 2 analysis most of the spurious signal test remains unchanged, however some modifications were necessary due to a limit on the amount of Monte Carlo we were able to produce.

In the run 1 analysis because ATLAS had collected $20fb^{-1}$ of data [64] compared to the $80fb^{-1}$ collected in run 2, in order to keep the spurious signal numbers near what we saw in run one, much more data would have to be generated. Limitations on the number of MC events able to be generated meant that we would have to perform this test with a much lower MC to data ratio. The standard model Higgs group uses a sample produced with a diphoton mass between 90 GeV and 175 GeV. This sample detailed above was used as the upper slice of our MC. We also needed to simulate data down to 50 GeV, so another production had to be done. Because we were the only group using this particular slice, we were not able to request the same statistics in the sample. In the end we had a MC to data ratio of 15 : 1 and 50 : 1 for the low and high slices respectively. A lower MC to data ratio will result in larger spurious signal numbers as the relative size of the statistical fluctuations will be higher.

The results of the spurious signal fits on the $\gamma\gamma + \gamma J$ templates discussed above are shown for the categories UU fig:5.32, UC fig:5.33 and CC fig:5.34. The information is summarized in table:5.13, table:5.14 and table:5.15. The signal function and parameterization used for these fits is the DSCB function described in section 5.4

Because the MC to data ratio is so low we are not able to find any functions that meet the nominal criteria, so we instead define a "relaxed criteria" defined by equation 5.6.3. This is used to check if the tested functions are statistically compatible with the maximum allowable value for $S/\delta S$.

$$\zeta_{spur} = \begin{cases} N_{spur} + n * \Delta_{MC} & N_{spur} + n * \Delta_{MC} < 0 \\ N_{spur} - n * \Delta_{MC} & N_{spur} - n * \Delta_{MC} > 0 \\ 0 & otherwise \end{cases}$$

We use $n = 2$. In other words we are allowing the 2σ error bars to fall within the original range of 30% background uncertainty. Relaxing the criteria is necessary in the case of limited statistics in MC. Figure 5.39 shows ζ_{spur} compared with the nominal spurious signal result for each category. The relaxed criteria is included in the results tables.

Even with the case of requiring the relaxed criteria, we still do not find functions passing in each category. This inspired trying to instead perform the spurious signal test on the $M_{\gamma\gamma}$ spectrum using a relative E_t (etRel) cut instead of the absolute one. The etRel is defined by dividing the photon E_t by the diphoton invariant mass of the photon pair in the event. The advantage of moving to relative cuts are that the spectrum no longer shows the turn on curve from the trigger cuts which allows much easier selection of fit function. This however comes with the downside of reduced data statistics in our $M_{\gamma\gamma}$ spectrum. You can see from figure 5.35 the $M_{\gamma\gamma}$ spectrum for absolute E_t cuts and various relative E_t cuts. Selecting the cut $E_t > 22/58GeV$ completely removes the trigger turn on at the beginning of the fit range. Figure 5.36 show a comparison of the absolute $E_t > 22$ GeV and relative $E_t > 22/55$ GeV cuts for the full fit range in the **UU** category. Integrating these two spectra we get that the loss in statistics is ~60% going from absolute to relative cuts. Even though this seems like a large loss, the change in significance is ~10% (fig. 5.37). We add new fit functions to the test as we are now able to consider functions that do not need to model the trigger turn on:

- Dijet: $m_{\gamma\gamma}^{(b0+b1log(m_{\gamma\gamma})+b2log(m_{\gamma\gamma}^{2}))}(1-m_{\gamma\gamma}^{1.5})^{a}$
    - Dijet1: $b1 = b2 = 0$

  - Dijet2: $b2 = 0$
  - Dijet3: All coefficients included

- Exponential Poly2: $e^{(p0+p1*m_{\gamma\gamma}+p2*m_{\gamma\gamma}^2)}$

The results of the spurious signal test for the relative E_t distributions are shown in figures 5.40 - 5.42 and tables 5.16 - 5.18 for each of the categories **UU** , **UC** and **CC** . Comparing the functions that we would select to use for the background, we can see that the relative E_t cuts do perform much better as we expect (table 5.19).

The final decision on whether to use the absolute or relative cuts came down to the effect on the expected limits that could be set in each case. Figure 5.38 shows a comparison of the expected limits using the absolute E_t cuts and the relative E_t cuts. In the case of absolute cuts, the functions used are the ones selected and indicated in the results tables. Two different choices of functional forms are used in the relative E_t limit indicated on the figure. The 5%-10% change in limit is what we would expect to see considering the similar change in significance we see. The decision was made to use the absolute E_t cuts as they have slightly better limits, and don't require changing any other parts of the analysis.

The functions chosen to model the background are the one in each category with the lowest value of ζ_{spur}. This is a Landau + Exponential for the **UU** and **UC** categories, and a fifth order Bernstein polynomial for the **CC** .

Category	function	$S/\delta S\%$	$\zeta_{spur}/\delta S\%$	$N_{spurious}$	Free parameters	
UU	Landau	-137	-86.1	-549	2	
	Fermi	1.16e+06	7.46e+05	-1.24e+03	3	
	LandauExp	-128	-76.7	-604	4	⇐
	Bern5	125	-83.4	701	5	
	Bern6	-131	-88.8	-689	6	
	Bern7	-142	-100	-730	7	

Table 5.13: Spurious signal results for each of the tested functions in the UU category.

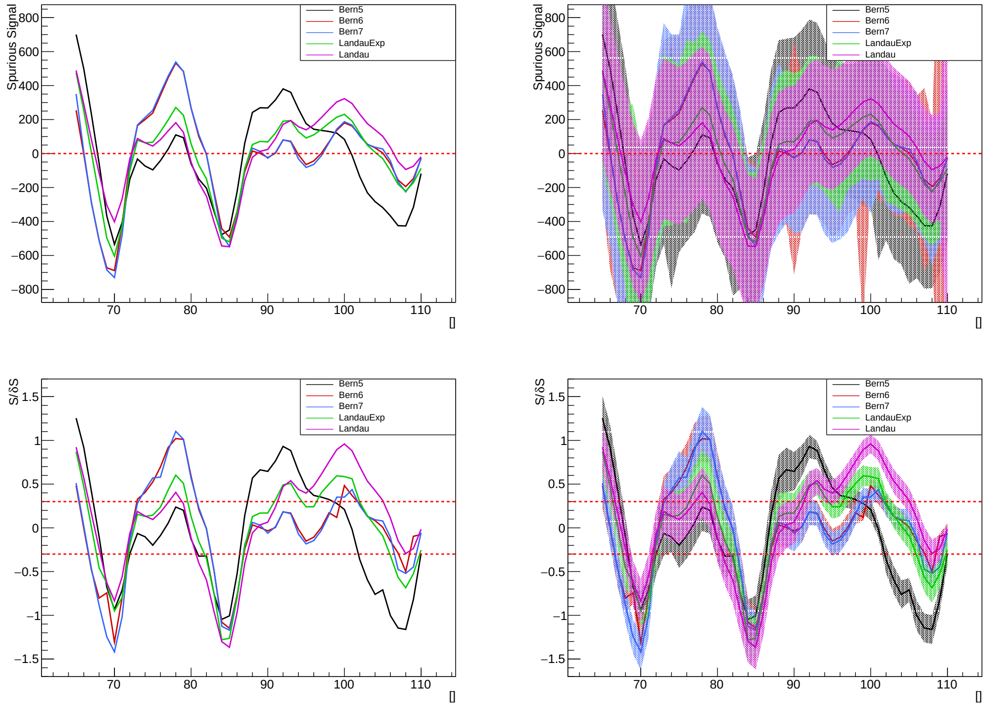


Figure 5.32: The spurious signal and the spurious signal divided by the background uncertainty with dashed lines representing 30% background uncertainty. Figures on the right include uncertainty in the UU category. The figures on the left are the same as those on the right but with no error bars shown.

Category	function	$S/\delta S\%$	$\zeta_{spur}/\delta S\%$	$N_{spurious}$	Free parameters	
UC	Landau	198	138	974	2	
	Fermi	-217	-187	-1.22e+03	3	
	LandauExp	-104	-61.4	-496	4	⇐
	Bern5	-119	-77.2	-577	5	
	Bern6	-105	-75.2	-703	6	
	Bern7	-136	-107	-884	7	

Table 5.14: Spurious signal results for each of the tested functions in the UC category.

Category	function	$S/\delta S\%$	$\zeta_{spur}/\delta S\%$	$N_{spurious}$	Free parameters	
CC	Landau	-121	-75.2	-244	2	
	Fermi	-85	-37.2	-170	3	
	LandauExp	-128	-80.9	-251	4	
	Bern5	-78.7	-38.3	-181	5	⇐
	Bern6	-84.2	-45.5	-202	6	
	Bern7	-77.6	-36.3	-204	7	

Table 5.15: Spurious signal results for each of the tested functions in the CC category.

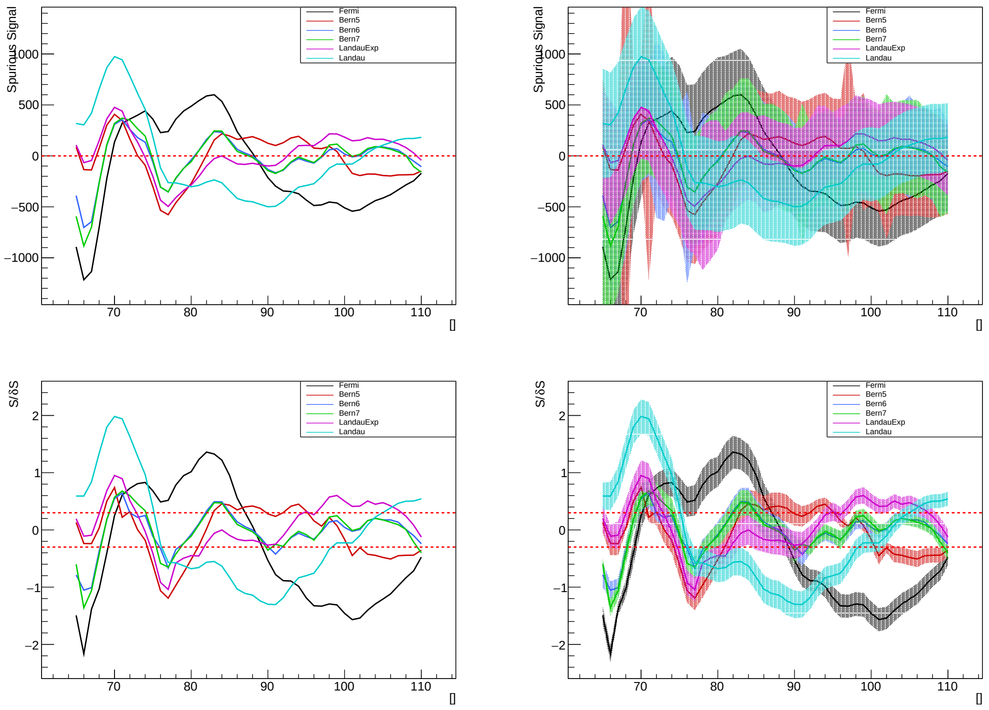


Figure 5.33: The spurious signal and the spurious signal divided by the background uncertainty with dashed lines representing 30% background uncertainty. Figures on the right include uncertainty in the UC category. The figures on the left are the same as those on the right but with no error bars shown.

Category	function	$S/\delta S\%$	$\zeta_{spur}/\delta S\%$	$N_{spurious}$	Free parameters	
UU	Dijet2	-99.3	-25.2	337	3	⇐
	Bern4	-88.8	-15	-285	4	
	Bern5	-64	6.17	238	5	
	ExpPoly2	-96.5	48	326	2	
	Landau	-307	-237	-977	2	
	Dijet1	-531	-482	1.64e+03	2	
	Bern3	-307	-258	-903	3	
	LandauExp	77.3	34.3	304	4	
	Dijet3	-114	-30.9	343	4	

Table 5.16: Spurious signal results for each of the tested functions in the UU category for relative E_t cuts. The arrow represents the function selected to model the background in this category.

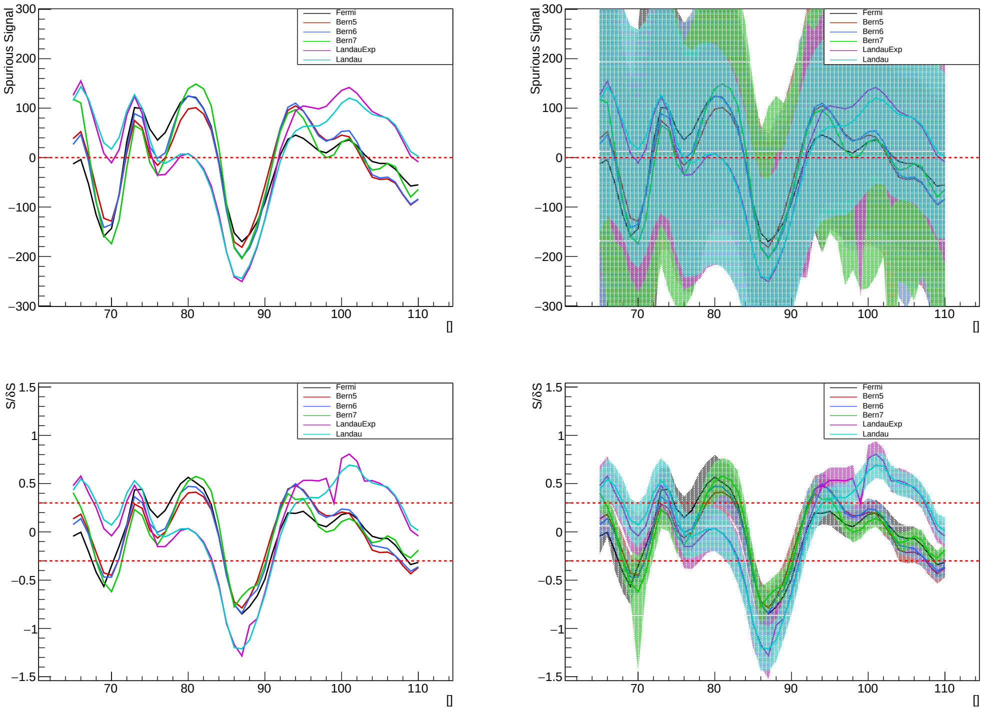


Figure 5.34: The spurious signal and the spurious signal divided by the background uncertainty with dashed lines representing 30% background uncertainty. Figures on the right include uncertainty in the CC category. The figures on the left are the same as those on the right but with no error bars shown.

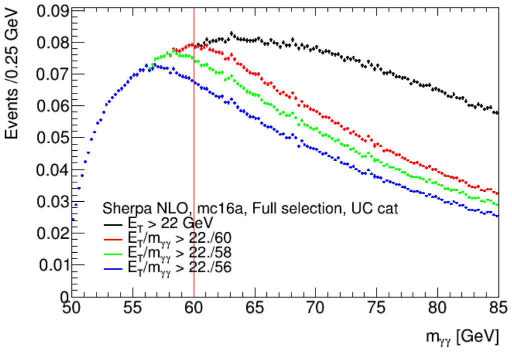


Figure 5.35: Comparison of the diphoton spectrum for absolute and relative E_t cuts. The vertical line shows the beginning of the non-resonant background fit range.

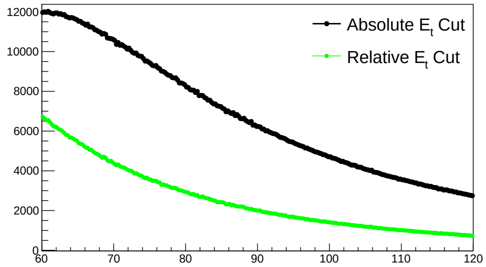


Figure 5.36: Comparison of the diphoton spectrum for absolute and $E_t > 22/55GeV$ cuts for the full fit range.

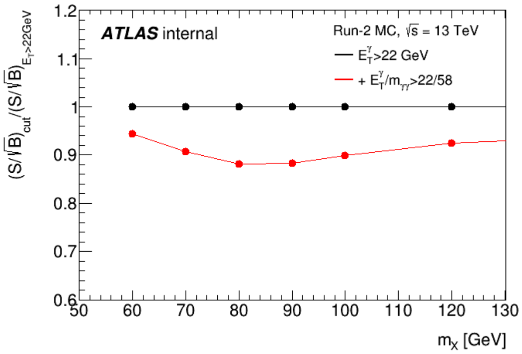


Figure 5.37: Comparison of significance for absolute and relative E_t cuts.

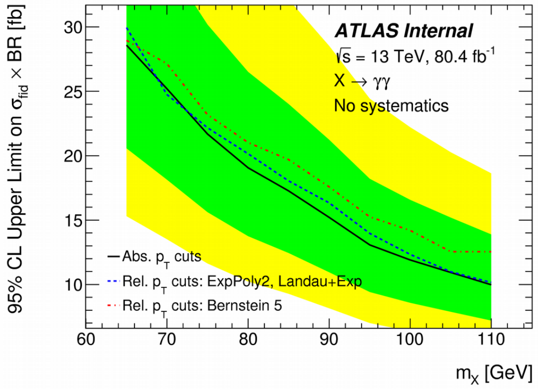


Figure 5.38: Comparison of the expected limits for absolute and relative E_t cuts.

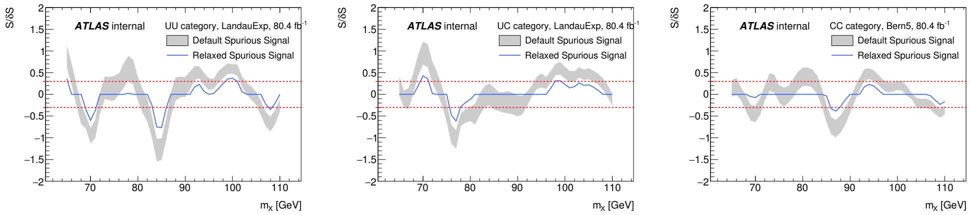


Figure 5.39: The spurious signal S divided by the background uncertainty in gray, and the relaxed spurious signal ζ divided by the background uncertainty in blue as a function of M_X for the **UU** (a), **UC** (b) and **CC** (c) categories [43].

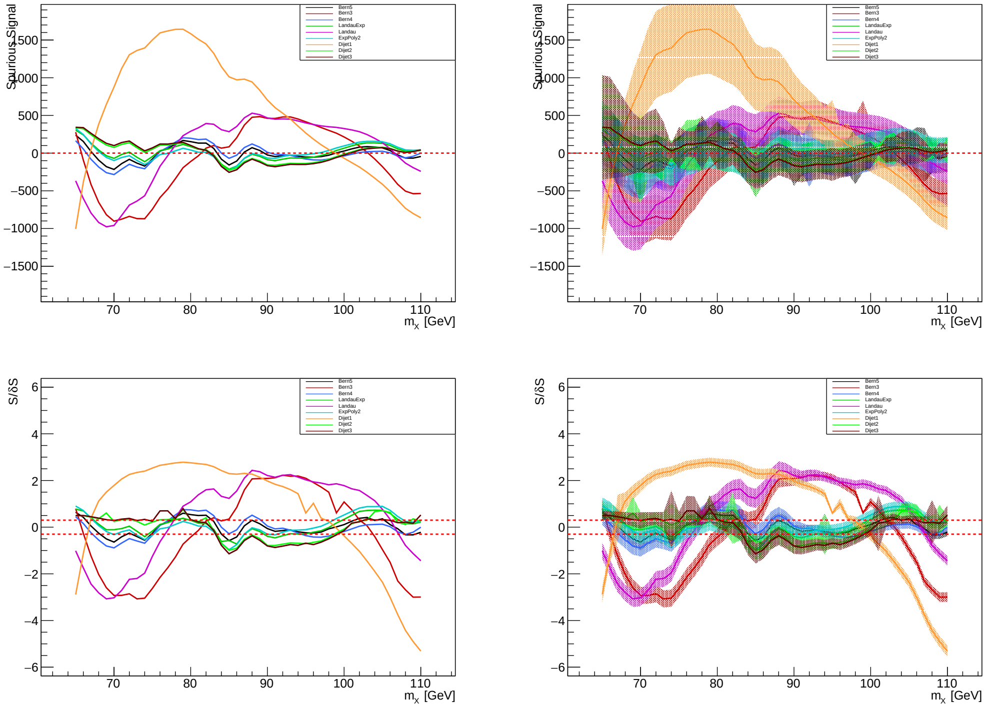


Figure 5.40: The spurious signal and the spurious signal divided by the background uncertainty for relative E_t cuts with dashed lines representing 30% background uncertainty. Figures on the right include uncertainty in the UU category. The figures on the left are the same as those on the right but with no error bars shown.

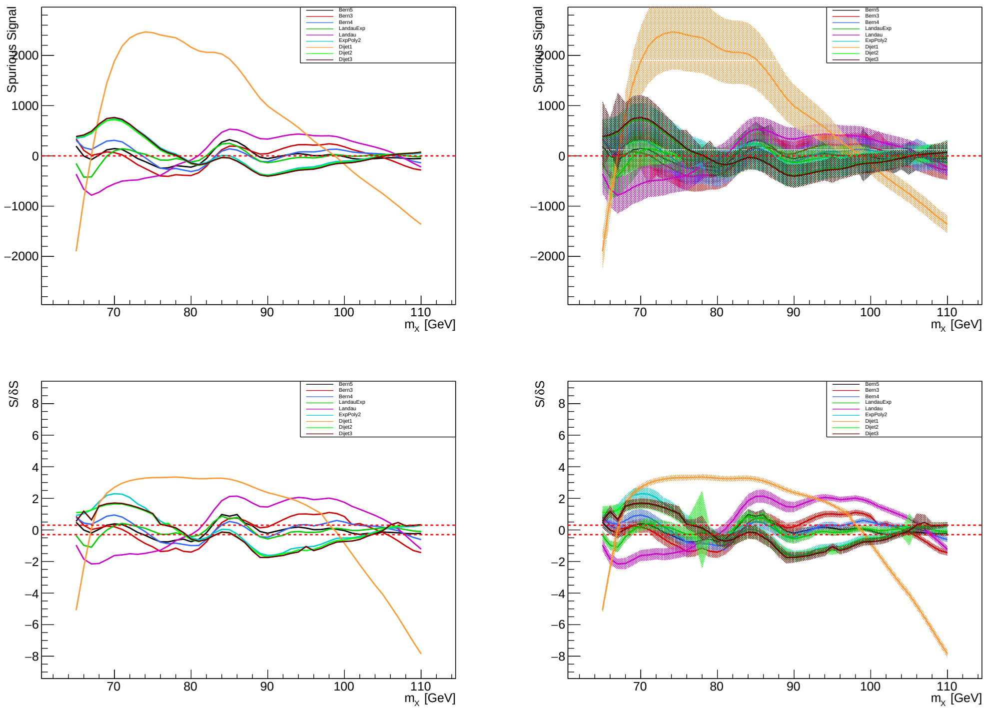


Figure 5.41: The spurious signal and the spurious signal divided by the background uncertainty for relative E_t cuts with dashed lines representing 30% background uncertainty. Figures on the right include uncertainty in the UC category. The figures on the left are the same as those on the right but with no error bars shown.

Category	function	$S/\delta S\%$	$\zeta_{spur}/\delta S\%$	$N_{spurious}$	Free parameters	
UC	ExpPoly2	229	146	735	2	
	Landau	-215	164	-783	2	
	Dijet1	-785	-731	2.47e+03	2	
	Bern3	-143	-104	-407	3	
	Dijet2	-171	-105	723	3	
	Bern4	-99.2	-33.3	-312	4	⇐
	LandauExp	-110	-44.1	-424	4	
	Dijet3	-175	111	764	4	
	Bern5	97.7	33.4	322	5	

Table 5.17: Spurious signal results for each of the tested functions in the UC category for relative E_t cuts. The arrow represents the function selected to model the background in this category.

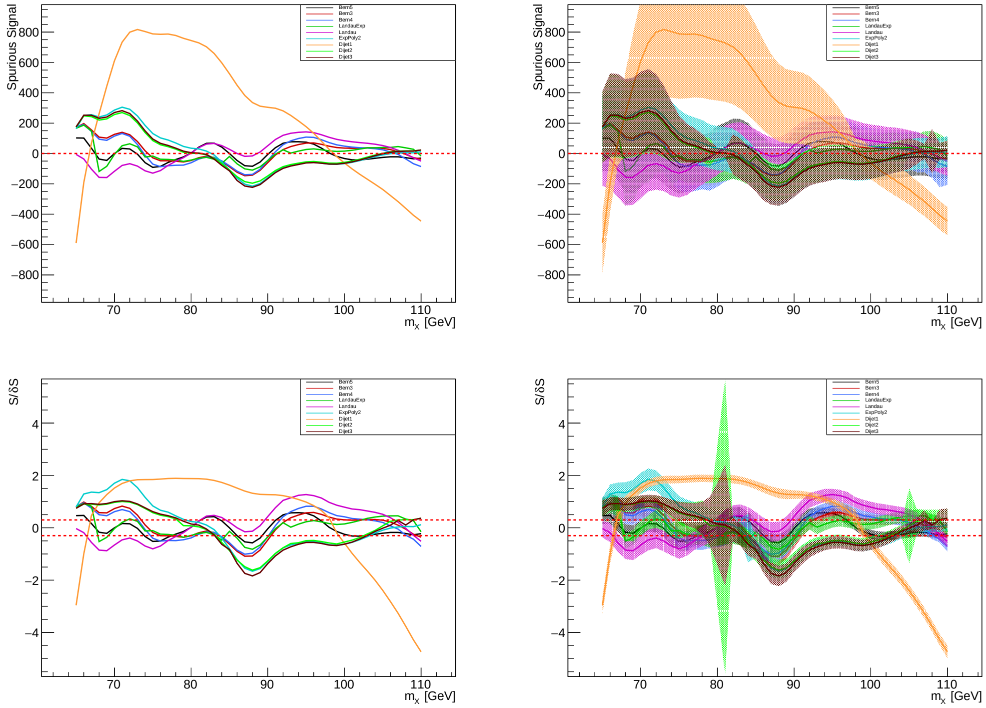


Figure 5.42: The spurious signal and the spurious signal divided by the background uncertainty for relative E_t cuts with dashed lines representing 30% background uncertainty. Figures on the right include uncertainty in the CC category. The figures on the left are the same as those on the right but with no error bars shown.

Category	function	$S/\delta S\%$	$\zeta_{spur}/\delta S\%$	$N_{spurious}$	Free parameters	
CC	LandauExp	92.9	24.1	187	4	⇐
	Bern5	57.8	20.1	102	5	
	Landau	127	83.6	-158	2	
	ExpPoly2	185	101	305	2	
	Dijet1	-474	-424	817	2	
	Bern3	-109	-37.2	197	3	
	Dijet2	-162	-81.4	270	3	
	Bern4	-101	44.6	191	4	
	Dijet3	-184	-105	282	4	

Table 5.18: Spurious signal results for each of the tested functions in the CC category for relative E_t cuts. The arrow represents the function selected to model the background in this category.

Category	Variable	LandauExp (Abs Cut)	Dijet2 (Rel Cut)
UU	$S/\delta S\%$	-128	-99.3
	$\zeta_{spur}/\delta S\%$	-76.6	-25.2
	$N_{spurious}$	-604	337
	Free Parameters	4	3

Category	Variable	LandauExp (Abs Cut)	Bern4 (Rel Cut)
UC	$S/\delta S\%$	-104	-99.2
	$\zeta_{spur}/\delta S\%$	-61.4	-33.3
	$N_{spurious}$	-496	-312
	Free Parameters	4	4

Category	Variable	Bern5 (Abs Cut)	LandauExp (Rel Cut)
CC	$S/\delta S\%$	-78.7	92.9
	$\zeta_{spur}/\delta S\%$	-38.3	24.1
	$N_{spurious}$	-181	187
	Free Parameters	5	4

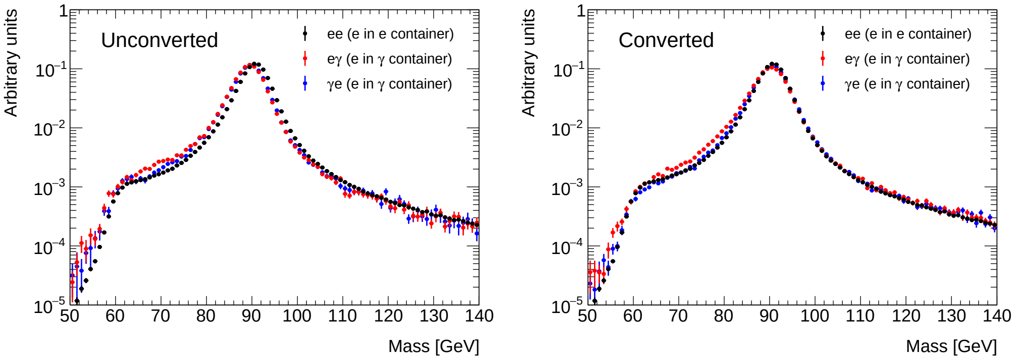
Table 5.19: Comparison of selected function for the absolute and relative E_t cuts. The column labeled 'Abs Cut' (Rel Cut) shows the criteria for the selected function with absolute cuts (relative cuts).

5.7 Resonant Background

The resonant background comes from the Drell-Yan process, $Z \rightarrow ee$ where both electrons are misidentified as photons. As this background is a signal like bump right in the middle of our search range, it is important to have a good description of this background especially in the **CC** category where this effect is largest. The background shape will be modeled with a template built from $Z \rightarrow ee$ MC. This template can then be summed with the others to create a complete background description.

The Drell-Yan background is estimated using a sample of dielections that are reconstructed as electrons and required to pass the tight electron ID. No isolation is applied but the E_{T}cut used in the photon selection is used. A template is constructed from this sample and is normalized to the number of electrons faking photons we expect by using the fake rates.

The distribution of electrons faking photons has shifted kinematics compared with $Z \rightarrow ee$ events reconstructed as electrons. This is because a majority of the electrons reconstructed as photons have undergone large bremsstrahlung. This shift must be modeled for the final templates. Figure 5.43 shows the invariant mass distribution of ee and $e\gamma$. To do this we use a $Z \rightarrow ee$ MC sample made of only electrons. From this sample we use events where two photons are reconstructed and that pass all selection criteria for the diphoton analysis detailed in section 5.3. These reconstructed photons must also be ΔR matched to truth electrons originating from a Z decay to ensure that the photon candidates are not from final state radiation. The dielectron data distribution is then transformed so that its kinematics are closer to that of the fake photons obtained from the MC sample.



(a) (b)

Figure 5.43: Invariant mass distributions (normalized to unity) of electrons pairs (black) and pairs made of a leading truth-matched electron reconstructed as an electron and subleading truth-matched electron reconstructed as a photon (red) or pairs made of a leading truth-matched electron reconstructed as photon and subleading truth-matched electron reconstructed as an electron (blue), for unconverted (a) and converted (b) photons, obtained on a $Z \to ee$ simulation sample.

5.7.1 Transformation of Dielectron Events

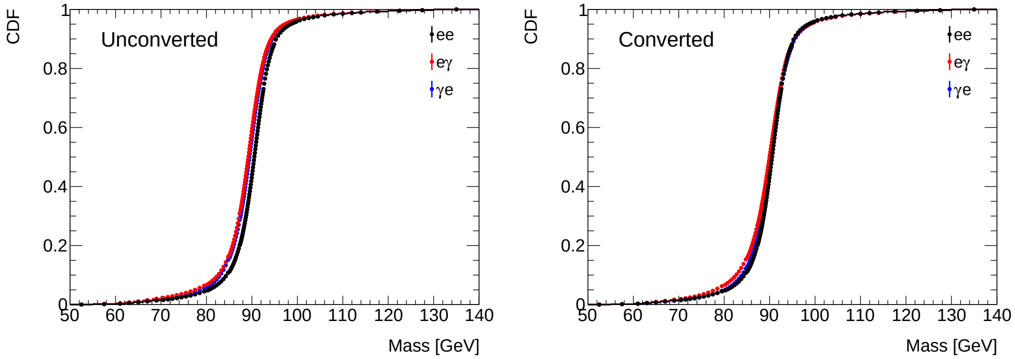
The transformation is done with a Smirnov, or inverse transform [65]. This method determines the transform $\phi(x)$ needed to go from a distribution $F(x)$ to $G(x)$. First we calculate the cumulative distributions (CDF) for $F(x)$ and $G(x)$ denoted $CDF_F(x)$ and $CDF_G(x)$. If the unknown transform $\phi(x)$ takes you from $F(x) \to G(x)$ then we can write

$$CDF_F(x) = CDF_G(\phi(x)). \tag{5.11}$$

Taking the inverse of $CDF_F(x)$ gives the definition of the transform,

$$\phi(x) = CDF_G^{-1}(CDF_F(x)). \tag{5.12}$$

So, using the CDFs for each function defines a transformation to go from one to the other. We can use this to transform our electron distributions into photon distributions in the MC samples. Figure 5.43 shows the CDFs for ee, $e\gamma$, and γe samples obtained from $Z \to ee$ MC. These are then used to determine the transforms to take an electron distribution to the photon distribution for both leading and subleading objects. In this case the transforms work out to be simple shifts along the $M_{\gamma\gamma}$ axis between the two CDFs. So for a particular mass point, the transform is just a horizontal shift with a magnitude of the distance from the electron CDF, to the photon CDF. We write the transforms δ_1 and δ_2 for the leading and subleading object respectively.



(a) (b)

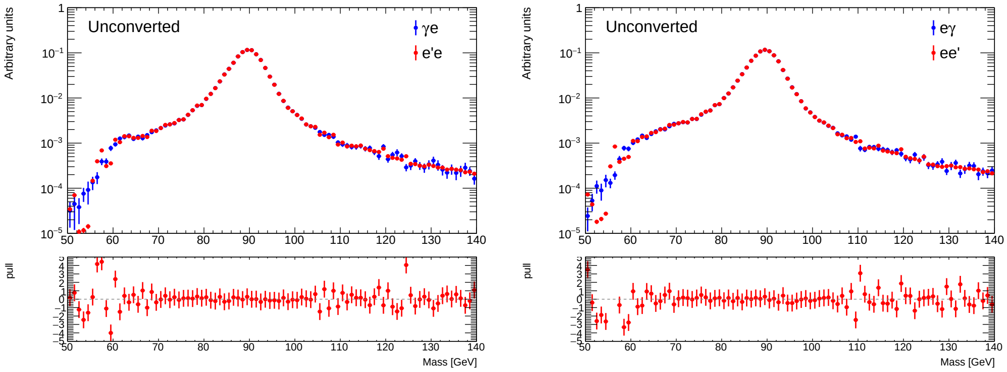
Figure 5.44: Cumulative distribution functions of discrete invariant mass values computed using binned distributions of electrons pair (black) and electron with fake subleading photon (red) or fake leading photon (blue) pairs in unconverted (a) and converted (b) photon category, obtained on a $Z \to ee$ simulation sample.

Figures 5.45 and 5.46 show the effect of this transformation to reproduce fake unconverted and converted photons, respectively. Transformed electrons in dielectron pairs here after are noted e'.

In order to obtain the invariant mass spectrum of both transformed electrons, one need to sum up shift values derived for leading and subleading transformations at the given dielectron mass value,

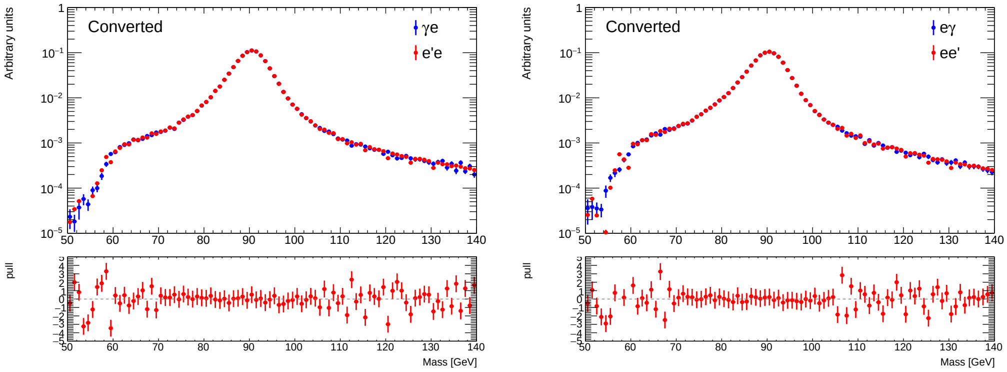
$$m_{e'e'} = m_{ee} + \delta_1 + \delta_2 = m_{e'e} + m_{ee'} - m_{ee} \ .$$

Once these transformations are determined, the dielectron data can be shifted so the kinematics look like the resonant background in the diphoton $M_{\gamma\gamma}$ spectrum. Figure 5.47 shows the sum result of the transformation on both the leading and subleading electrons.



(a) (b)

Figure 5.45: Invariant mass distributions (normalised to unity) of electron with fake photon (blue) and transformed electron (red) pairs in leading (a) and subleading (b) cases for unconverted photons, obtained on a $Z \rightarrow ee$ simulation sample.

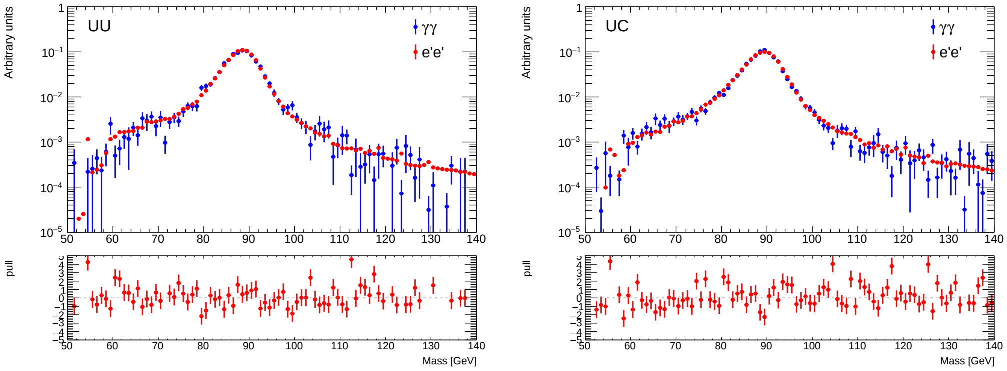


(a) (b)

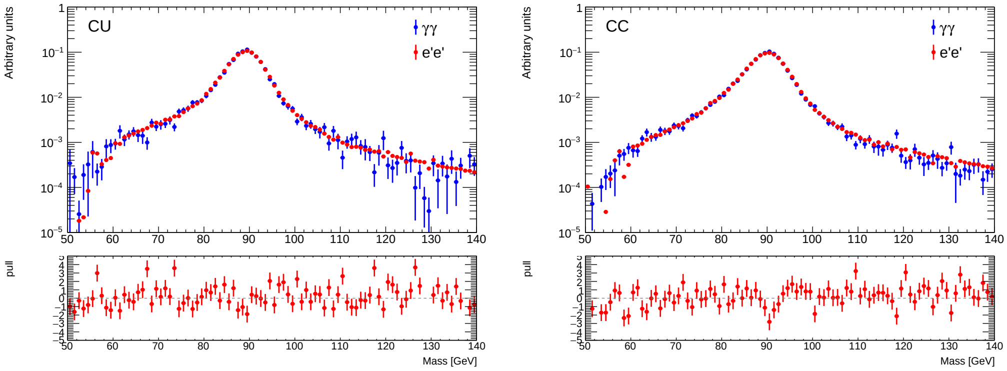
Figure 5.46: Invariant mass distributions (normalised to unity) of electron with fake photon (blue) and transformed electron (red) pairs in leading (a) and subleading (b) cases for converted photons, obtained on a $Z \rightarrow ee$ simulation sample.

5.7.2 Electron to Photon Fake Rates

After the shapes of the leading and subleading electrons have been corrected to the photon distributions, the fake rates must be calculated. To build a template with the correct



(a) **UU** (b) **UC**



(c) **CU** (d) **CC**

Figure 5.47: Invariant mass distributions (normalised to unity) of diphoton pair (blue) and transformed electrons pairs obtained with summed up shifts (red) in **UU** (a), **UC** (b), **CU** (c) and **CC** (d) categories, obtained on a $Z \to ee$ simulation sample.

normalizations, we must know the proper number of fakes that we can expect to occur.

To measure the electron to photon fake rates, a dielectron and a electron-photon sample are used. The fake rate is measured separately for leading (ρ_1) and subleading (ρ_2) objects. These fake rates are measured in a window around the Z-peak with a width of $6 * \sigma_{CB}$ which is 6 times the width of a double sided crystal ball fit to the data. As this is not a MC sample, the non-resonant background must be subtracted away to give a good measure of the fake rates. This is done with a signal + background fit, where the signal portion is a double sided crystal ball and a Fermi × power law for the continuum portion of the Drell-Yan. The background uses a Fermi × exponential function. Figure 5.48 shows the results of the fits in the four conversion categories.

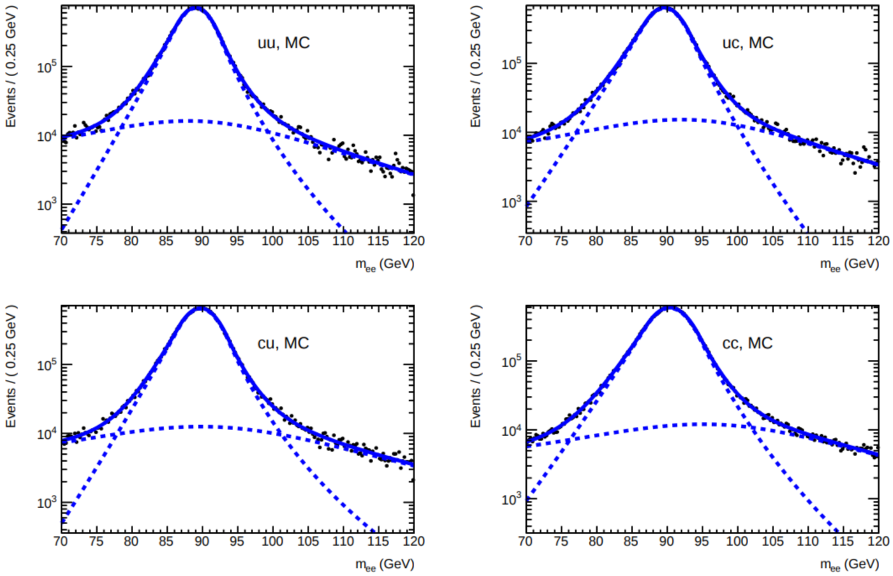


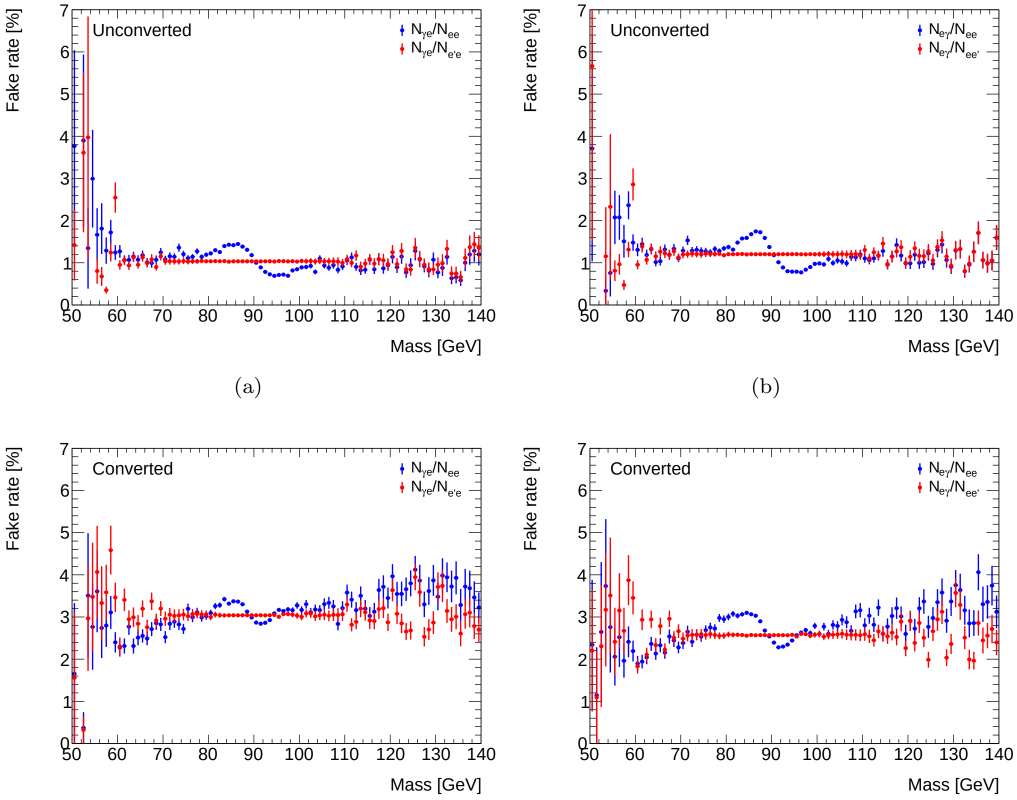
Figure 5.48: Non-resonant background subtraction from the transformed $e'e'$ invariant mass distributions performed before calculation of leading and subleading fake rates in data in four conversion categories.

The electron to photon fake rates are calculated as:

$$\begin{aligned} \rho_1 &= \frac{N_{\gamma e}}{N_{e'e}}, \\ \rho_2 &= \frac{N_{e\gamma}}{N_{ee'}} \end{aligned} \tag{5.13}$$

where as before, transformed electrons are denoted e'. The fake rates are first calculated on a $Z \to ee$ MC sample, shown in figure 5.49. We can see that after the transformation, the fake rate is flat in the invariant mass. This means we can use a single number for the rate and do not have to parameterize it in $M_{\gamma\gamma}$.

The final fake rates as measure on the data samples are shown in figure 5.50. The red lines represent a fit to the data inside the mass window around the Z-peak.



(a) (b)

(c) (d)

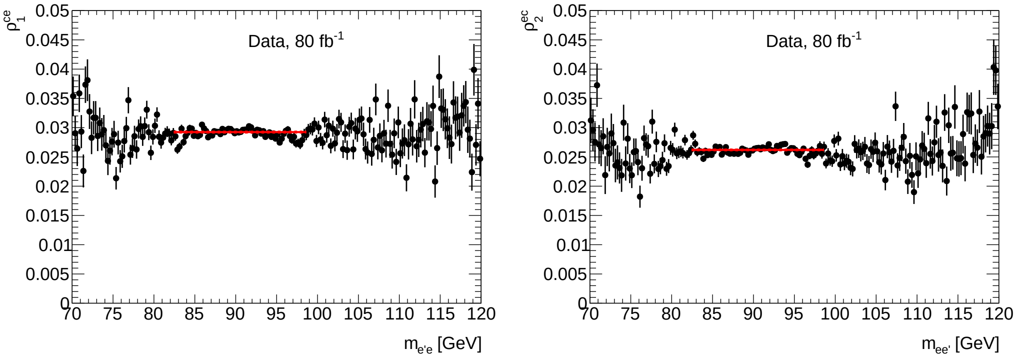
Figure 5.49: Invariant mass dependence of the leading (a) and subleading (b) **unconverted** and the leading (c) and subleading (d) **converted** photon fake rates as measured in $Z \to ee$ simulation, before (blue) and after (red) applying the Smirnov transformation.

The total number of fake diphoton events coming from the Drell-Yan process is then

$$N_{\gamma\gamma}^{DY} = \alpha \rho_1 \rho_2 N_{ee}^{DY} \phi(m_{ee}) \ ,$$

where $\phi(m_{ee})$ represents the Smirnov transformation of mass distribution from section 5.7.1, and α is a correction factor introduced to take into account the reconstruction efficiency differences between single photons and pairs of photons. This correction factor is evaluated from $Z \to ee$ MC events and is defined as:

$$\alpha = \frac{1}{\rho_1^{MC} \rho_2^{MC}} N_{\gamma\gamma}^{MC} / N_{e'e'}^{MC} \ ,$$



imagesratio

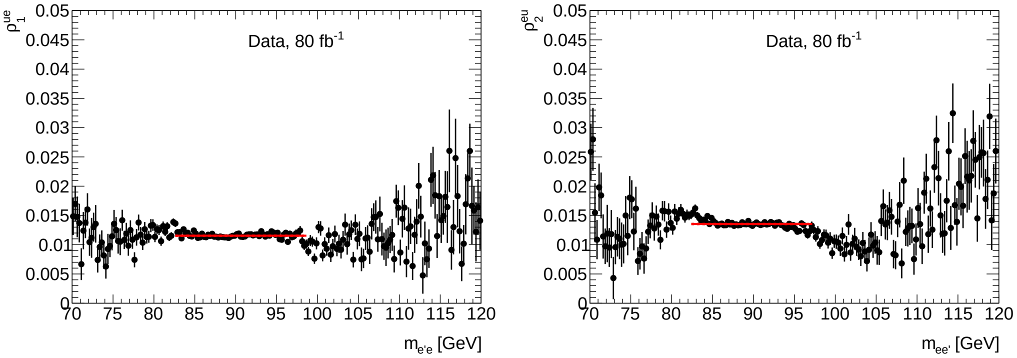
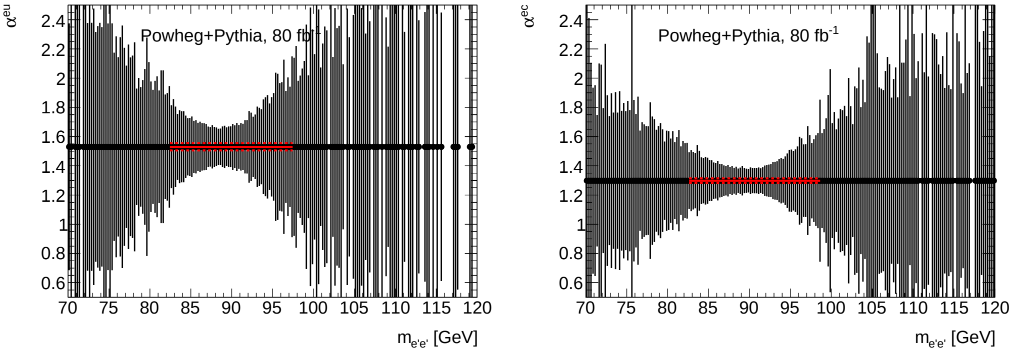


Figure 5.50: Invariant mass dependence of the leading (left) and subleading (right) unconverted (top) and converted (bottom) photon fake rates, as measured in data. The red lines represent the fit in the mass window.

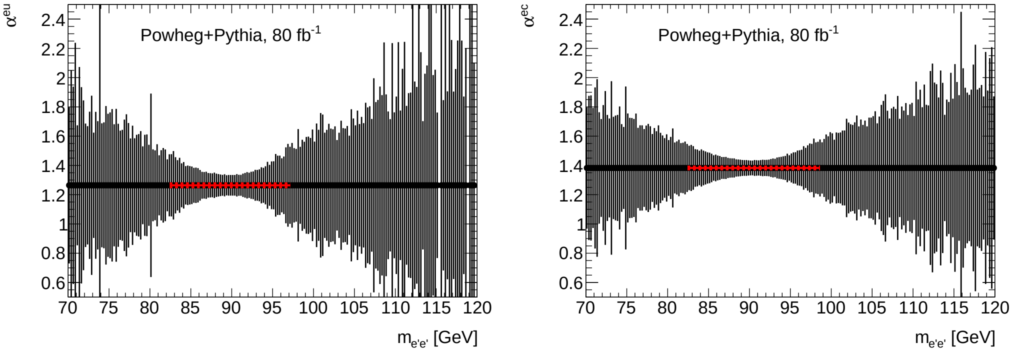
where two transformed electrons $e'e'$ spectrum is obtained using mapping to $\gamma\gamma$ distribution directly. Because the distributions after transformation don't depend on the $M_{\gamma\gamma}$, the fake rate in nominator also doesn't have mass dependence in the measured correlation factor as seen in Fig. 5.51. This factor is expected to be around 1 if the photon reconstruction does not depend on the reconstruction of the other candidate. However, since the track isolation uses the information from both photon candidates in the event to determine the correct primary vertex, there is some correlation.

Resulting normalization of the transformed distributions is obtained by rescaling $e'e'$ spectra to the levels of the expected $\gamma\gamma$ background in data using the normalization factors



(a) **UU** (b) **UC**

images



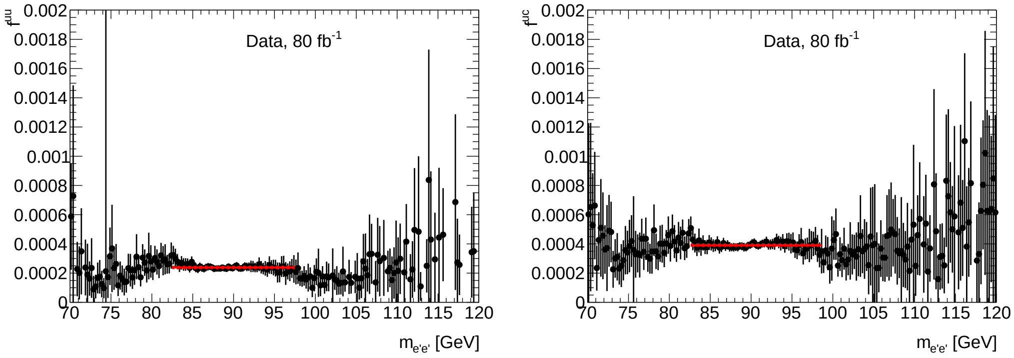
(c) **CU** (d) **CC**

Figure 5.51: Invariant mass dependence of correlation factor α in **UU** (a), **UC** (b), **CU** (c) and **CC** (d) categories, obtained on a $Z \rightarrow ee$ simulation sample. The red lines represent the fit in the mass window.

$f = \alpha\rho_1\rho_2$ shown in Fig. 5.52.

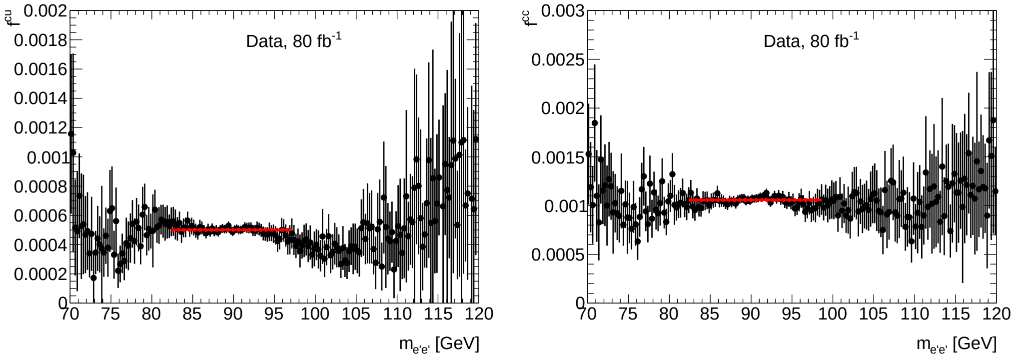
5.7.3 Resonant Background Templates

The resonant background templates are constructed from the $Z \rightarrow ee$ MC samples after kinematics transformation. These are then normalized using the fraction $f = \alpha\rho_1\rho_2$ as described in the previous section. The resulting templates are shown in figure 5.53. Note that since we are using a 3 category scheme for this analysis, the **UC** and **CU** categories are merged in the end.



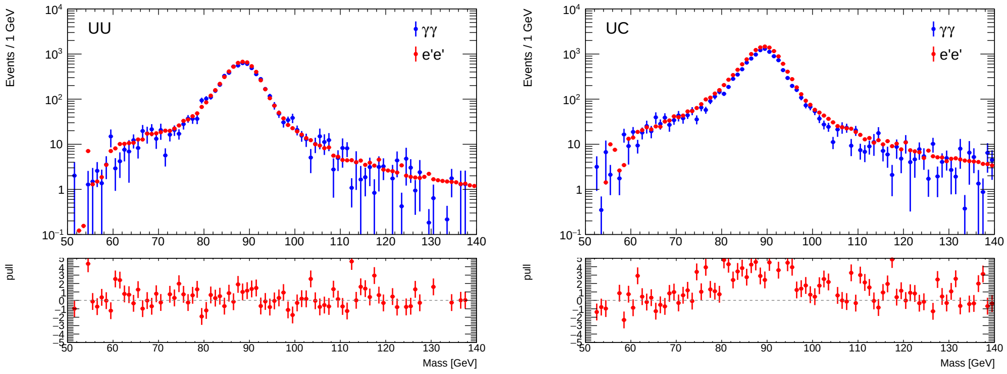
(a) **UU** (b) **UC**

imagesnorm



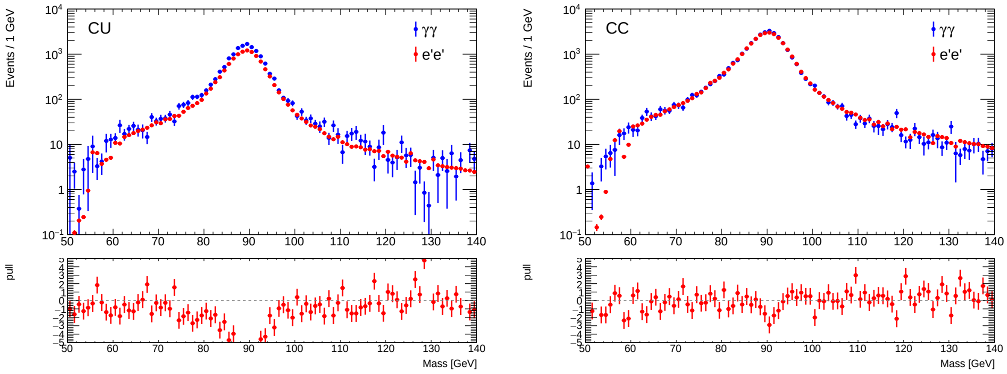
(c) **CU** (d) **CC**

Figure 5.52: Invariant mass dependence of normalisation factor f in **UU** (a), **UC** (b), **CU** (c) and **CC** (d) categories, used to obtain expected number of fake photons in data. The red lines represent the fit in the mass window.



(a) **UU** (b) **UC**

images



(c) **CU** (d) **CC**

Figure 5.53: Drell-Yan invariant mass templates (red) describing the expected diphoton background component (blue) in the **UU** (a), **UC** (b), **CU** (c) and **CC** (d) categories, obtained on a $Z \to ee$ simulation sample.

5.7.4 Standard Model Higgs Background

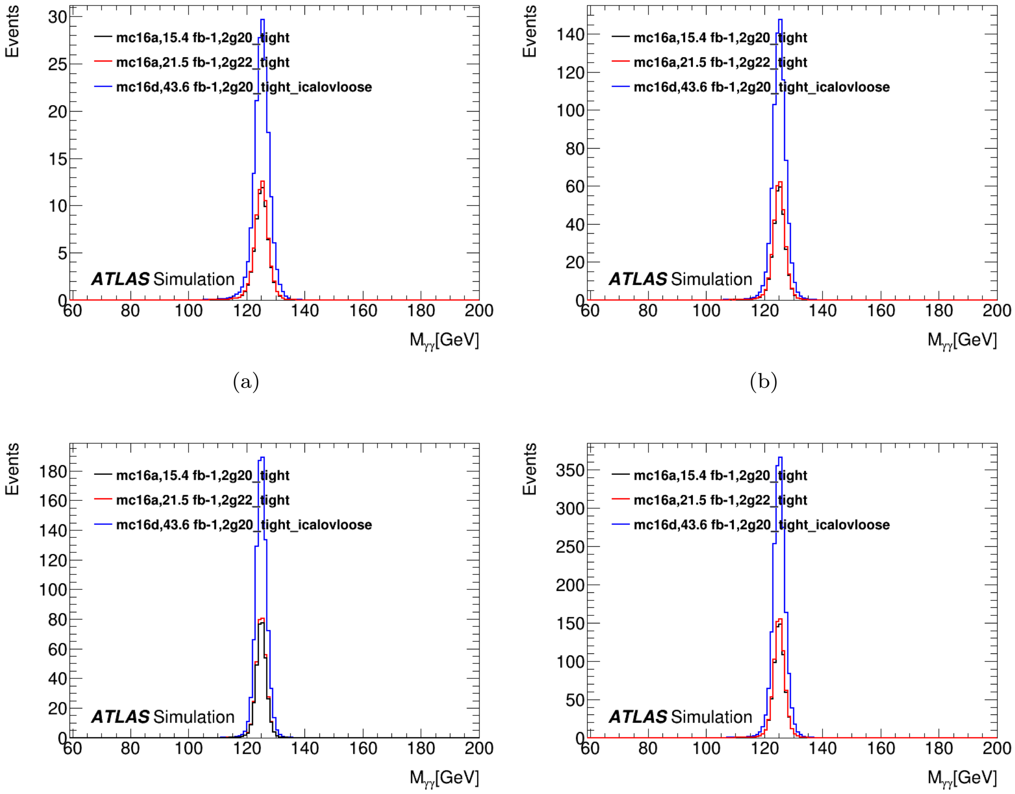
The standard model Higgs represents a very signal like resonant background. With its mass as $M_{\gamma\gamma} = 125$ GeV, it is outside our search window of [65, 110] GeV, however the lower tail of the Higgs peak could contribute to the background. PowHeg+Pythia8 ggF Monte Carlo is used to estimate the expected contribution to the backgrounds.

The selection described in Sec. 5.3 is applied. Table 5.20 shows the number of selected SM Higgs events for each trigger configuration as used in the data and each conversion type category. The $m_{\gamma\gamma}$ distributions are shown in Fig. 5.54 (full range) and Fig. 5.55 (mass range (60, 120) GeV).

Sample/luminosity(fb^{-1})	trigger	CC	CU	UU	total
mc16a/15.4	2g20_tight	2.31	6.86	2.74	11.69
mc16a/21.5	2g22_tight	2.47	6.64	2.86	12.19
mc16d/43.6	2g20_tight_icalovloose	6.57	18.57	9.09	34.23

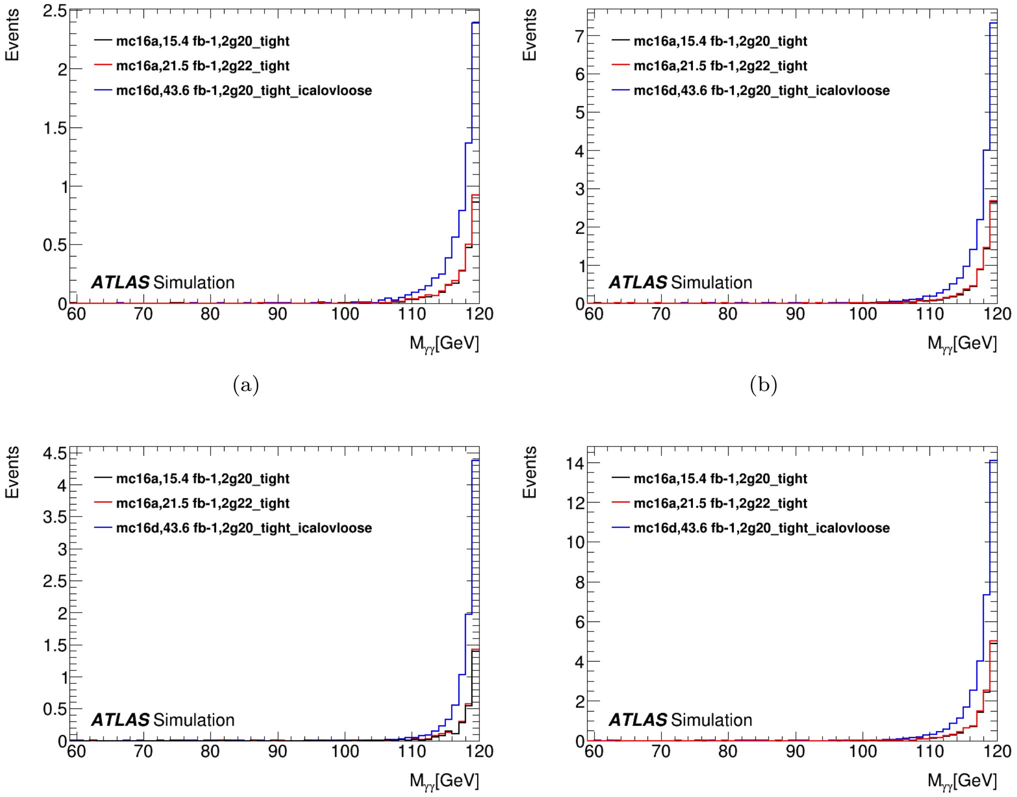
Table 5.20: The number of SM Higgs with $60 < M_{\gamma\gamma} < 120$ GeV after normalizing to the corresponding luminosity. The results are shown in three categories according to the conversion types of the photons.

Adding the Standard Model Higgs component to the predicted $\gamma\gamma + \gamma jet$ background described in Sec. 5.5, $N_{Higgs}/\sqrt{N_{bkg}}$ and $N_{Higgs}/(N_{Higgs}+N_{bkg})$ are calculated, where N_{Higgs} and N_{bkg} are the numbers of Standard Model Higgs and $\gamma\gamma + \gamma jet$ backgrounds respectively. Only the mc16d sample is used in this check, and $Z \rightarrow ee$ background is not considered. Results are shown in Fig. 5.56 and Fig. 5.57. One can see that the Standard Model Higgs only represents a small fraction (0.35% in the highest $M_{\gamma\gamma}$ bin, without categorization in conversion types) of the total background. Therefore, the SM Higgs contribution to the backgrounds is ignored when the final background template is built.



(a) (b) (c) (d)

Figure 5.54: $M_{\gamma\gamma}$ distribution of three Standard Model Higgs MC samples. (a) CC. (b) CU. (c) UU. (d) Sum of three categories.



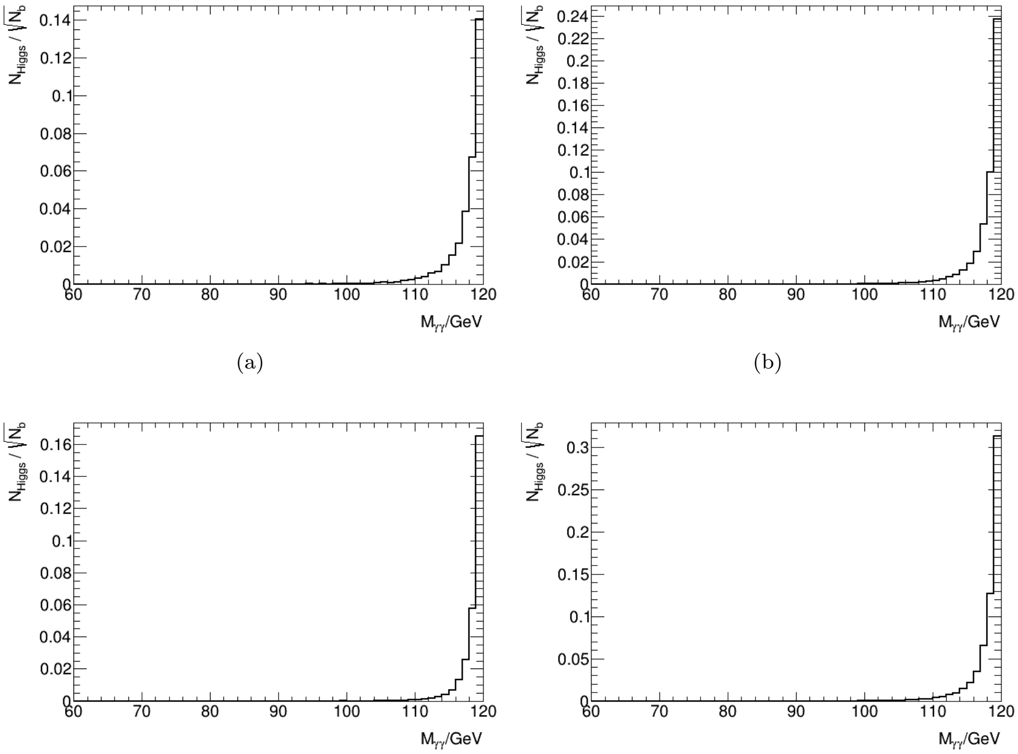
(a)

(b)

(c)

(d)

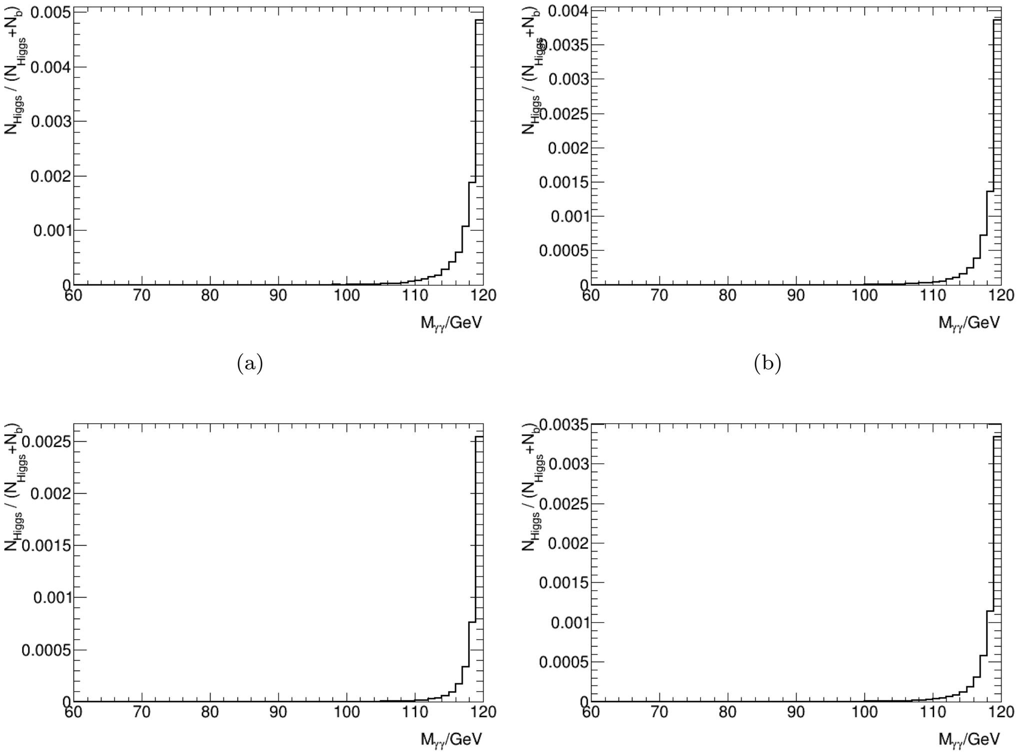
Figure 5.55: $M_{\gamma\gamma}$ distribution of three Standard Model Higgs MC samples with $60 < M_{\gamma\gamma} < 120$ GeV. (a) CC. (b) CU. (c) UU. (d) Sum of three categories.



(a) (b)

(c) (d)

Figure 5.56: Backgrounds bias due to Standard Model Higgs. (a) CC. (b) CU. (c) UU. (d) Sum of three categories.



(a) (b)

(c) (d)

Figure 5.57: Fraction of Standard Model Higgs in the total background. (a) CC. (b) CU. (c) UU. (d) Sum of three categories.

5.8 Fiducial Volume and Cx Factors

5.8.1 Fiducual Volume Definition

In order to quote results in a model independent way, we must correct measurements of the cross-section. This allows the comparison with other exeriments as well as allowing theorists to use the results. The correction factor will correct for detector effects such as reconstruction, identification and selection efficiencies. It is defined as the number of reconstructed signal events passing all analysis requirements, $N_{selection}$, divided by the number of signal events generated in the fiducial volume, $N_{acceptance}$.

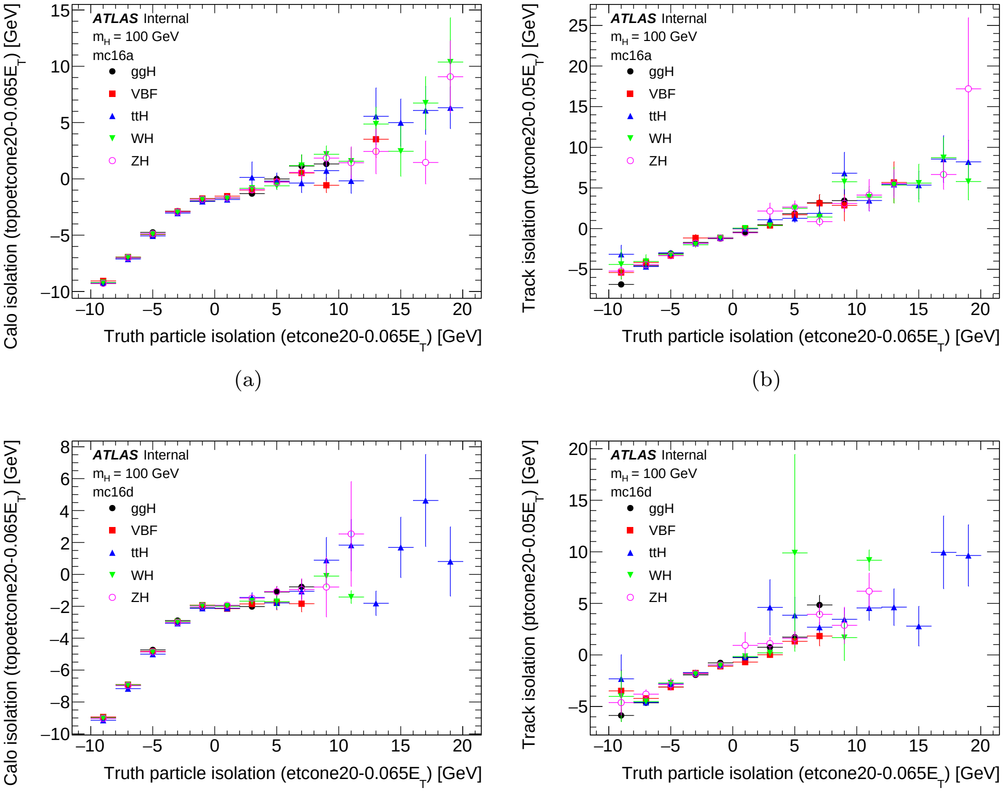
$$C_X = \frac{N_{selection}}{N_{acceptance}} \tag{5.14}$$

The correction factor is measured on MC truth values. In order for this correction factor to provide a model independent result, it must use a definition of the fiducial volume that is does not depend on the final state.

In order to choose the fiducial volume, a study on MC is done. Using samples with different production modes (ggF, VBH, WH, ZH, ttH), a wide range of kinematics, as well as different final states, we can optimize the definition of the fiducial volume.

All cuts applied on reconstructed photons are applied at truth level. Both photons should be within the detector volume, $|\eta| < 2.37$ and not in the transition region between the barrel and endcaps, $1.37 < |\eta| < 1.52$. The standard E_{T}cut is also applied, $E_{\mathrm{T}} > 22$ GeV.

Photons that are produced in final states that also have final state jets (ZH, WH, ttH) will have a much higher isolation and therefore a much lower selection efficiency. We would like the correction factor be be as stable in efficiency as possible so adding an isolation cut is required. The cut is applied on the truth calorimeter isolation variable as $etcone20 - 0.65 * E_{\mathrm{T}} < 0$. Figure 5.58 shows the calorimeter isolation values in relation to the track isolation that is used in the offline kinematic selection for the analysis. Figure 5.59 shows the correction factors for different production modes with and without the truth isolation cut.



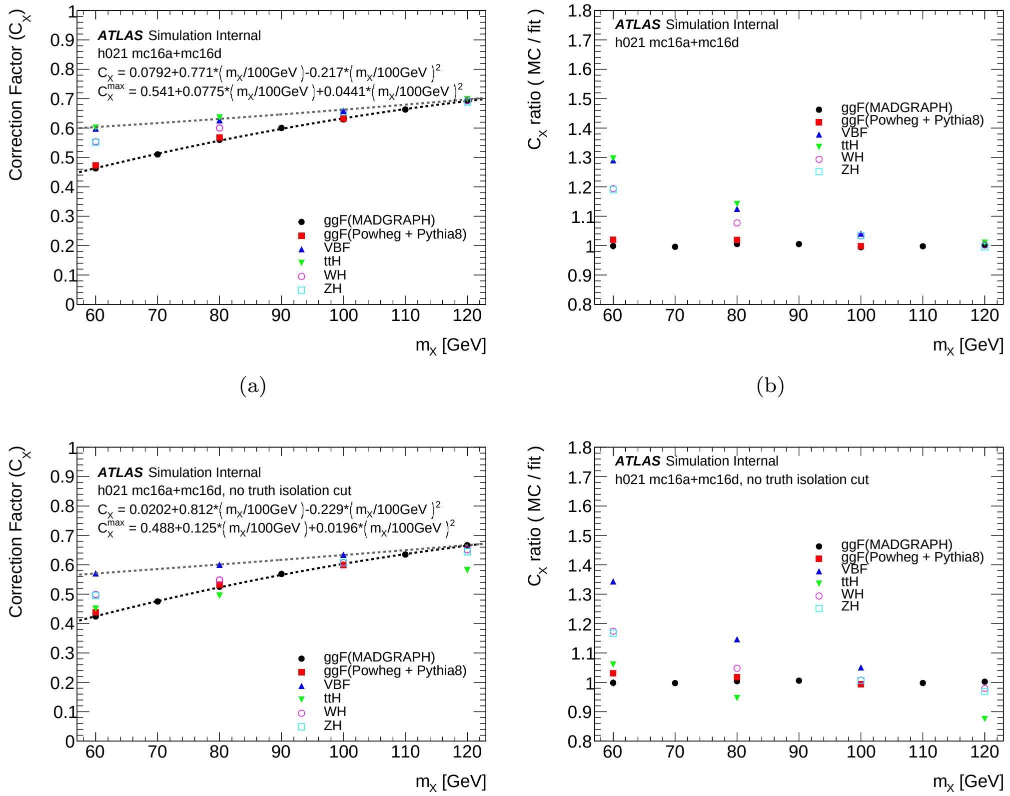
(a) (b)

(c) (d)

Figure 5.58: Reconstructed calorimeter isolation (a) and reconstructed track isolation (b) as a function of the truth particle isolation, for mc16a MC samples for all Higgs production modes and m_X=100 GeV. Similarly, Figures(c) and (d) show the same for mc16d MC samples.

Also supplied for the purposes of reinterpretation of the results is the acceptance factor. Figure 5.60 shows the fiducial acceptance using the cuts described above. The parameterization of the acceptance factor as a function of m_X is given as:

$$A_X = 0.5729 - 2.069 \cdot e^{-3.13 \cdot m_X/(100 \quad \text{GeV})}. \tag{5.15}$$



(a) (b)

(c) (d)

Figure 5.59: C_X factors computed for five different production modes, as a function of m_X. In Figures (a) and (b) an isolation cut is also applied at truth level. In Figures (c) and (d), the fiducial volume is defined with photons E_{T} and η cuts only.The black dotted line shows the 2nd order polynomial function used for the parameterization of C_X as a function of m_X. The parameters of this function is obtained by a fit to the ggF MadGraph MC samples. The gray dotted line shows the 2nd order polynomial function used for the systematic error, which the parameters are obtained by a fit to the MC points with the maximum C_X values. Figures (b) and (d) show the ratio of the fit and the MC points. These ratios are taken as a systematic error due to the different final states.

5.8.2 Cx Factors in Production Modes

As it is seen in Figure 5.59b, there is a 30% difference in C_X factors between ggF and ttH production modes for $m_X = 60$ GeV. This difference is mainly due to the difference in the photon E_{T} distributions. The tight ID and photon isolation both have a dependence on E_{T}. This causes differences in the ID and isolation efficiency. The photon E_{T} distribution

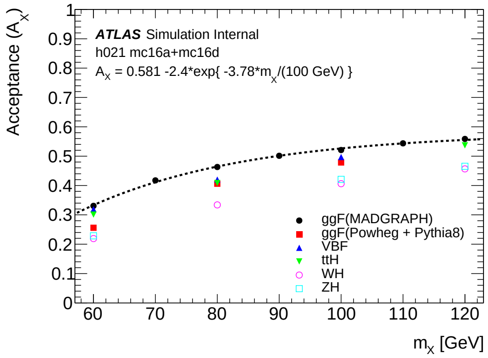
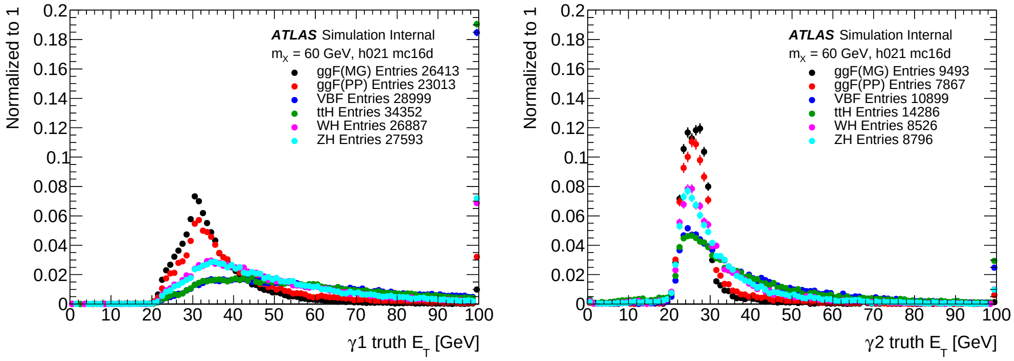


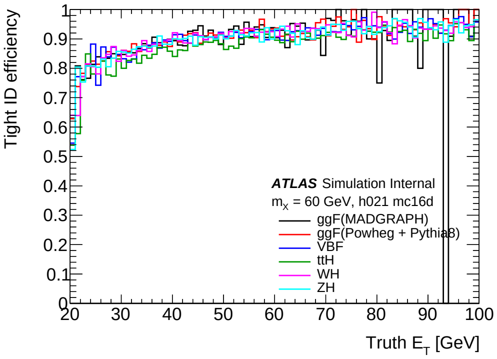
Figure 5.60: Acceptance factors A_X for different production modes, as a function of m_X.

is shown in Figure 5.61. Photons from the ttH production mode have larger E_{T} compared to those from the ggF mode. Figure 5.62 presents the Tight ID and isolation efficiencies as functions of photon E_{T}. These efficiencies increase for larger E_{T}, therefore the ttH signals have higher selection efficiencies. The contribution of the isolation efficiency to the production mode differences is small compared to that of the Tight ID. This is because the fiducial volume is defined so that the production mode differences due to isolation efficiency is reduced, as it is described above. Thus, the Tight photon ID is the main cause for the difference in C_X factors for ttH and ggF signals.



(a) (b)

Figure 5.61: Truth E_{T} distributions for photons from different signal production modes. They are shown for (a) leading photons and (b) subleading photons which pass the pre-selection. The photons from ttH signals have larger E_{T} compared to those from ggF signals.

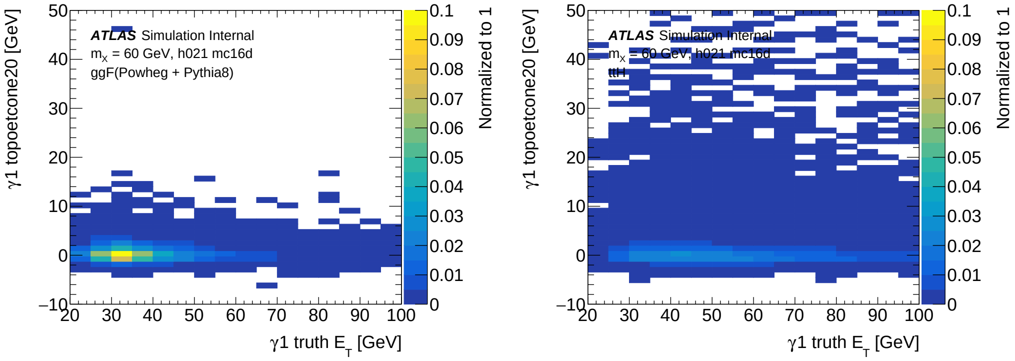


(a)

(b)

Figure 5.62: Selection efficiencies of photons with respect to truth photon E_{T}. They are shown for leading photons passing the pre-selection. (a) Tight photon ID efficiency, and (b) isolation (track and calorimeter) isolation efficiency. These efficiencies increase for larger photon E_{T}.

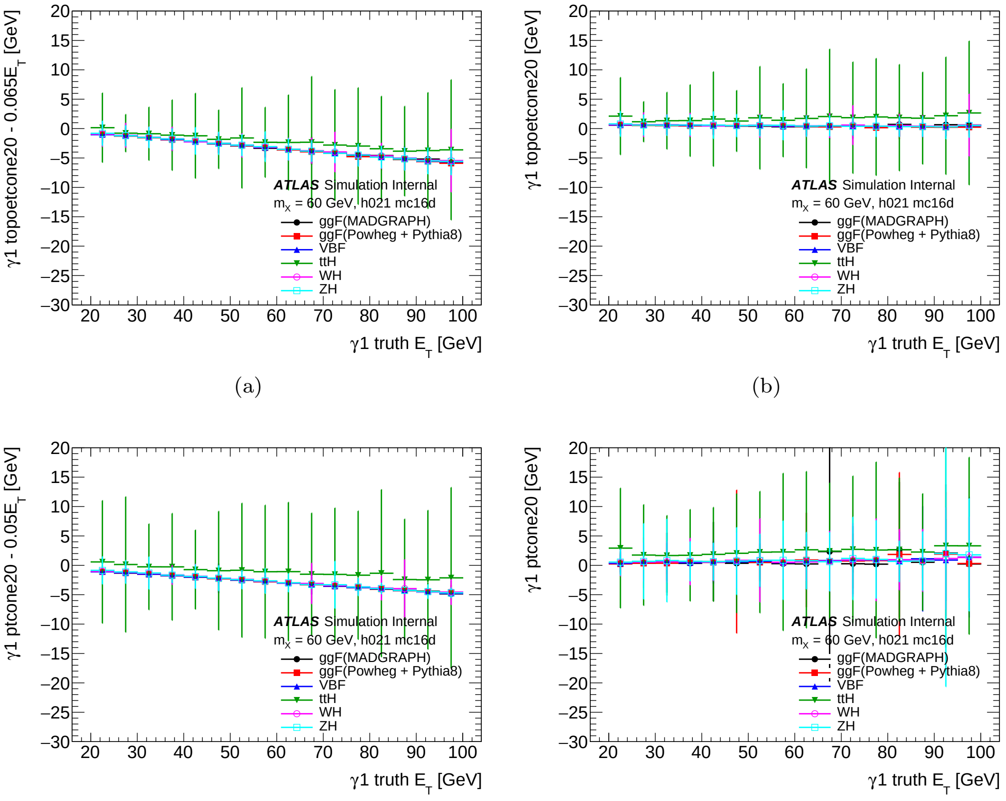
Figures 5.64 and 5.63 show the profile plots and correlation plots of the isolation variables with respect to the truth E_{T} of photons. These figures demonstrate how the isolation efficiencies increase with respect to photon E_{T}.



(a)

(b)

Figure 5.63: Correlation plots of topoetcone20 and truth E_{T} for leading photons passing the pre-selection. (a) ggF (Powheg+Pythia sample) and (b) ttH signals.



(a) (b)

(c) (d)

Figure 5.64: Profile plots of isolation distributions with respect to the truth E_{T}, for leading photons passing the pre-selection. The markers show the average, and the bars show the standard deviation for each E_{T} bin. (a) Calorimeter isolation, (b) topoectone20, (c) track isolation, and (d) ptcone20. Note that the large bars for ttH signal is because some of its events can take large topoetcone20 values, as it is seen in Figure 5.63b.

5.9 Systematic Uncertainties

5.9.1 Uncertainties on the Cx Factors

5.9.1.1 Production Mode

The uncertainty on the correction factor is taken as the envelope between the correction factors for the production modes farthest apart. As shown in the section 5.8, the production modes with the biggest difference in the C_X factor is the nominal ggF sample and the ttH sample. The uncertainty can be parameterized as the ratio of the fits between the two samples.

$$\text{variation up} = \frac{0.5436 + 0.06644 * M_{\gamma\gamma}/100 + 0.05077 * M_{\gamma\gamma}/100 * M_{\gamma\gamma}/100}{0.08858 + 0.7405 * M_{\gamma\gamma}/100 - 0.1989 * M_{\gamma\gamma}/100 * M_{\gamma\gamma}/100} - 1 \tag{5.16}$$

This uncertainty decreases from 25% at 65 GeV to 2.5% at 110 GeV.

5.9.1.2 Experimental Uncertainties

There are several uncertainties we need to account for on the calculation of C_X factors associated with experimental effects. These are:

- Trigger Scale Factor
- Uncertainty from pileup
- Photon identification efficiency
- Isolation efficiency
- Photon energy resolution
- Energy scale

We saw in section 5.8 that the correction factor depends on the photon ID efficiencies as well as the isolation efficiency. These two also both depend on E_T, so any uncertainty on the energy resolution will also need to be accounted for.

The uncertainty on C_X due to the trigger scale factors is illustrated in Figure 5.65a and parameterized as a function of m_X as:

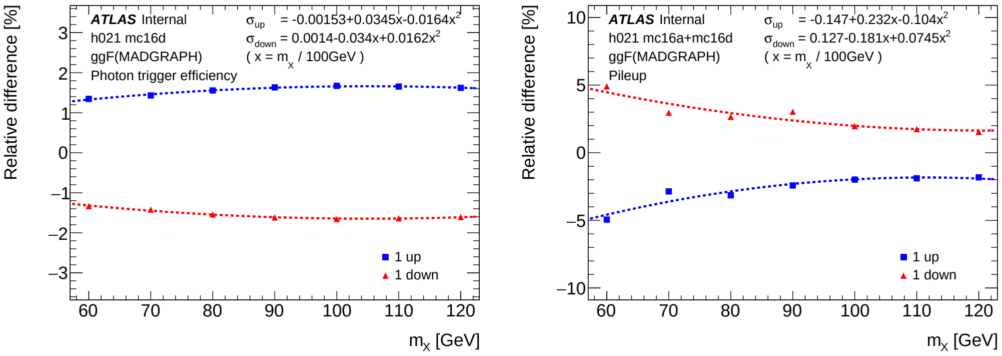
$$\begin{aligned}\text{variation up} &= -0.00153 + 0.03453 * M_{\gamma\gamma}/100 - 0.01643 * M_{\gamma\gamma}/100 * M_{\gamma\gamma}/100 \\ \text{variation down} &= 0.001404 - 0.03403 * M_{\gamma\gamma}/100 + 0.01619 * M_{\gamma\gamma}/100 * M_{\gamma\gamma}/100\end{aligned} \tag{5.17}$$

This uncertainty increases from 1.4% at 65 GeV to 1.7% at 110 GeV.

The uncertainty on C_X due to the pileup is illustrated in Figure 5.65b and parameterized as a function of m_X as:

$$\begin{aligned}\text{variation up} &= 0.1266 - 0.1812 * M_{\gamma\gamma}/100 + 0.07446 * M_{\gamma\gamma}/100 * M_{\gamma\gamma}/100 \\ \text{variation down} &= -0.1475 + 0.2319 * M_{\gamma\gamma}/100 - 0.1041 * M_{\gamma\gamma}/100 * M_{\gamma\gamma}/100\end{aligned} \tag{5.18}$$

This uncertainty decreases from 4.1% at 65 GeV to 1.8% at 110 GeV.



(a) (b)

Figure 5.65: Systematic uncertainties associated with to the trigger efficiency (a) and pile-up (b) as a function of m_X, computed with the Madgraph ggF sample.

The uncertainty on C_X due to the photon identification efficiency is illustrated in Figure 5.66a and parameterized as a function of m_X as:

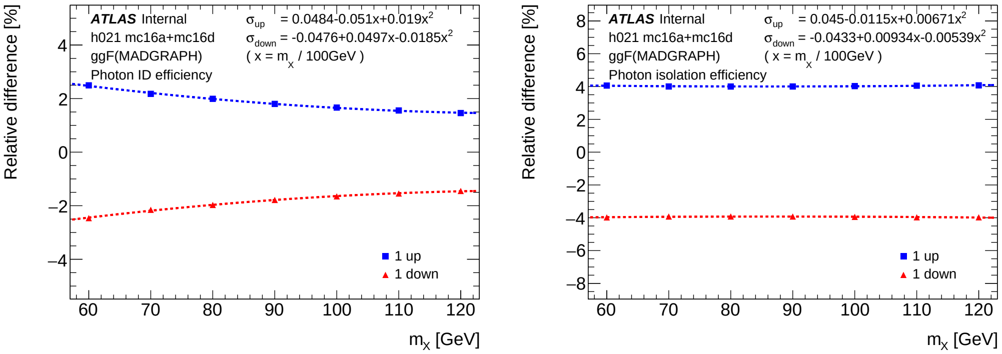
$$\begin{aligned}\text{variation up} &= 0.04845 - 0.05095 * M_{\gamma\gamma}/100 + 0.01904 * M_{\gamma\gamma}/100 * M_{\gamma\gamma}/100 \\ \text{variation down} &= -0.0476 + 0.04974 * M_{\gamma\gamma}/100 - 0.01854 * M_{\gamma\gamma}/100 * M_{\gamma\gamma}/100\end{aligned} \tag{5.19}$$

This uncertainty decreases from 2.3% at 65 GeV to 1.5% at 110 GeV.

The uncertainty on C_X due to the photon isolation efficiency is illustrated in Figure 5.66b and parameterized as a function of m_X as:

$$\begin{aligned}\text{variation up} &= 0.04495 - 0.01151 * M_{\gamma\gamma}/100 + 0.006713 * M_{\gamma\gamma}/100 * M_{\gamma\gamma}/100 \\ \text{variation down} &= -0.04398 + 0.01106 * M_{\gamma\gamma}/100 - 0.00645 * M_{\gamma\gamma}/100 * M_{\gamma\gamma}/100\end{aligned} \tag{5.20}$$

This uncertainty is of %, flat as a function of m_X.



(a) (b)

Figure 5.66: Systematic uncertainties associated with to the photon identification (a) and isolation (b) efficiencies as a function of m_X, computed with the Madgraph ggF sample.

The uncertainty on C_X due to the photon energy resolution is illustrated in Figure 5.67a and parameterized as a function of m_X as:

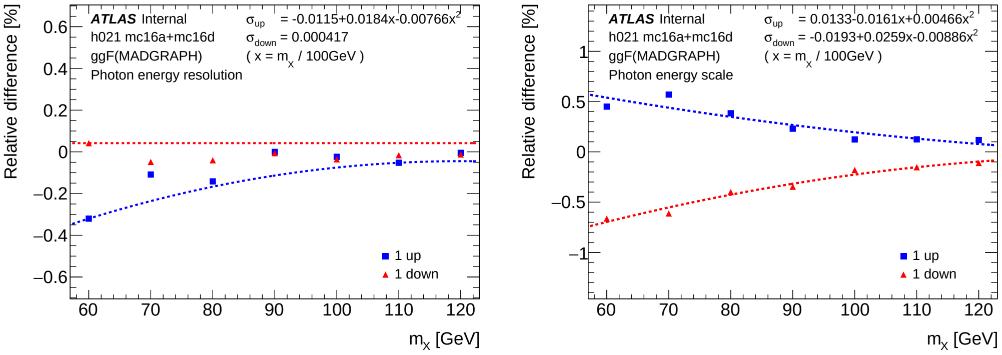
$$\begin{aligned}\text{variation up} &= 0.0004173 \\ \text{variation down} &= -0.01147 + 0.01837 * M_{\gamma\gamma}/100 - 0.007655 * M_{\gamma\gamma}/100 * M_{\gamma\gamma}/100\end{aligned} \tag{5.21}$$

This uncertainty decreases from 0.28% at 65 GeV to 0.05% at 110 GeV.

The uncertainty on C_X due to the photon energy scale is illustrated in Figure 5.67b and parameterized as a function of m_X as:

$$\begin{aligned} \text{variation up} &= 0.01334 - 0.01605 * M_{\gamma\gamma}/100 + 0.004656 * M_{\gamma\gamma}/100 * M_{\gamma\gamma}/100 \\ \text{variation down} &= -0.01934 + 0.02594 * M_{\gamma\gamma}/100 - 0.008859 * M_{\gamma\gamma}/100 * M_{\gamma\gamma}/100 \end{aligned} \tag{5.22}$$

This uncertainty decreases from 0.49% at 65 GeV to 0.13% at 110 GeV.



(a) (b)

Figure 5.67: Systematic uncertainties associated with to the photon energy resolution (a) and photon energy scale (b) as a function of m_X, computed with the Madgraph ggF sample.

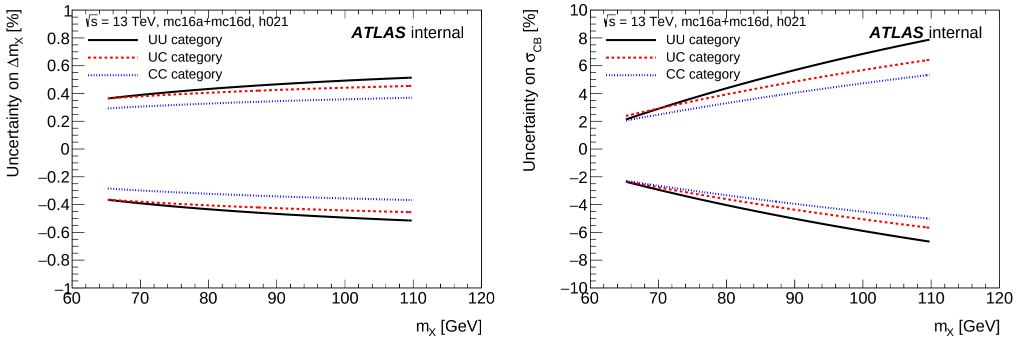
5.9.2 Signal Model

5.9.2.1 Energy Scale and Resolution

In this analysis the narrow width approximation is used which means that the shape of the signal is determined almost entirely by the energy resolution and the energy scale of the detector. The uncertainty on these two will cause some uncertainty in deviation on the mean ΔM_X and the width σ_{CB} of the double-sided crystal ball function. The uncertainty for the energy resolution and scale are estimated by varying them up and down. The parameters ΔM_X and σ_{CB} are refit for each variation with the multiple mass point method detailed in section 5.4 keeping all other parameters fixed to their nominal values. The uncertainty is then calculated as the ratio of the result of the varied sample to the nominal one. The uncertainties on Δm_X vary between 0.3 and 0.5% and the ones on the width σ_{CB} vary between 2 and 8%. These are summarized in table 5.21 and figure 5.68

Parameter	Parameterization	Coefficient	**UU**	**UC**	**CC**
Energy scale up					
Δm_X	$a + bm_{nX}$	a [GeV]	0.55	0.58	0.56
		b	0.74	0.64	0.51
Energy scale down					
Δm_X	$a + bm_{nX}$	a [GeV]	-0.44	-0.30	-0.16
		b	-0.73	-0.54	-0.46
Energy resolution up					
σ_{CB}	$a + bm_{nX}$	a [GeV]	1.53	1.797	2.0543
		b	0.98	0.99292	1.0513
Energy resolution down					
σ_{CB}	$a + bm_{nX}$	a [GeV]	1.35	1.61	1.87
		b	0.61	0.66	0.74

Table 5.21: Parameterizations of the double-sided Crystal Ball function describing the signal shape, result of the multiple mass point fit at low-mass, for the three categories, when the photon energy scale and resolutions are varied up and down. m_{nX} is defined as $m_{nX} = \frac{m_X - 100}{100}$.

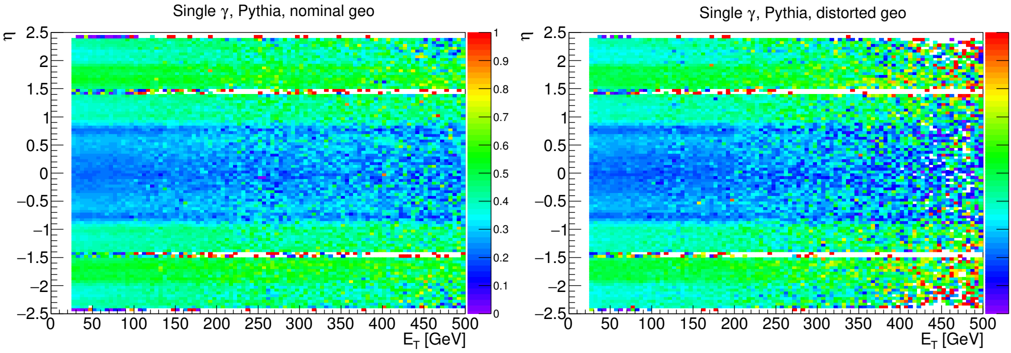


(a) (b)

Figure 5.68: Systematic uncertainties associated with to the photon energy scale (a) and photon energy resolution (b) as a function of m_X, computed with the Madgraph ggF sample.

5.9.2.2 Conversion Categories

Photon conversions happen in the material of the tracking system. The uncertainty on the number of photons that migrate from one conversion category to another is assessed by using samples that have been reconstructed with a distorted detector geometry. The Egamma group produced samples with a distorted geometry containing γ+jets that are sliced in bins of photon E_{T} from 17 GeV to 5000 GeV. Because this is not a standard diphoton sample, we cannot use the default selection and categorization. We instead use the fraction of conversions in a particular E_{T} and η bin, $f(E_{\mathrm{T}}, \eta)$. This fraction is shown in figure 5.69.



(a) (b)

Figure 5.69: fraction of converted photons in bins of E_{T} and η for the (a) nominal and (b) distorted geometries.

We may then use this fraction $f(E_{\mathrm{T}}, \eta)$ for each of the leading and subleading photons in the signal MC samples to get the fraction of events in each category of our signal MC adjusted for the distorted geometry. This is done by looping over all events in the signal MC and summing the fraction for the leading and subleading photons for their particular value of E_{T} and η. For example, in the **CC** category we have:

$$N_{\mathbf{CC}} = \sum_{allevents} f(E_{\mathrm{T}}^{\gamma 1}, \eta^{\gamma 1}) * f(E_{\mathrm{T}}^{\gamma 2}, \eta^{\gamma 2}) \tag{5.23}$$

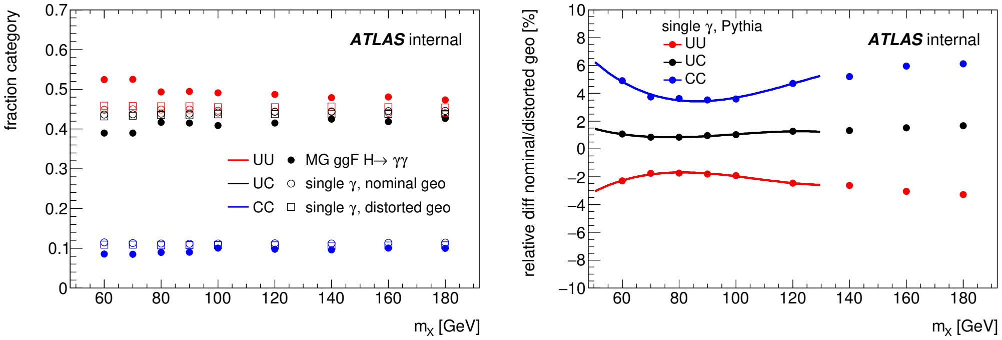
And, then for the **UU** category we have:

$$N_{\mathbf{UU}} = \sum_{allevents} (1 - f(E_{\mathrm{T}}^{\gamma 1}, \eta^{\gamma 1})) * (1 - f(E_{\mathrm{T}}^{\gamma 2}, \eta^{\gamma 2})) \tag{5.24}$$

The fraction of events in each category is then just the number in that category over the total number of events in the signal sample, $f_{\mathbf{CC}} = N_{\mathbf{CC}} / N_{total}$ and $f_{\mathbf{UU}} = N_{\mathbf{UU}} / N_{total}$. The third category **UC** is then just $f_{\mathbf{CU}} = 1 - f_{\mathbf{UU}} - f_{\mathbf{CC}}$. The fraction in each category computed with the Single Photon samples and with the $X \rightarrow \gamma\gamma$ signal samples are compared in Figure 5.70a. One can see that the fractions recomputed with the fraction in the Single Photon samples do not match the numbers computed directly with the signal sample. That is most probably due to the correlation between the two photons. But assuming that this correlation is independent from the geometry one can estimate systematic uncertainties by comparing the numbers computed with the nominal and distorted geometries, as illustrated in Figure 5.70b. Those migration uncertainties are slightly mass-dependent and this dependence with modeled by third order polynomials:

- **UU** category: $f(m_X) = -15.5734 + 0.430247 * m_X - 0.00427721 * m_X^2 + 1.33562e{-}05 * m_X^3$ %, with an average of +4%.
- **UC** category: $f(m_X) = 8.57602 - 0.254114 * m_X + 0.00268446 * m_X^2 - 8.955e - 06 * m_X^3$ %, with an average of +1%.
- **UC** category: $f(m_X) = 25.4582 - 0.624664 * m_X + 0.00560123 * m_X^2 - 1.53142e{-}05 * m_X^3$ %, with an average of -2%.

The dependence of the fraction of events in each category on the production modes was checked and is illustrated in Figure 5.71. One can see in the ratios to the ggF process that the fractions agree within statistical uncertainties. But there is a small systematic effect



(a) (b)

Figure 5.70: (a) Fraction of events in each category computed with the signal samples (plain markers) and using the conversion fractions from Single Photon samples (open markers). (b) Ratio between the default geometry and distorted geometry.

and the expected limit was recomputed using more conservation values of -7.5% and +6.5% for the **UU** and **UC** categories respectively, instead of -2% and +1% on average. The expected limits with the two migration uncertainties are compared in Figure 5.72. The effect of this increased uncertainty is negligible, probably because the results are dominated by the **UU** category. Since the C_X factors are computed inclusively the difference on the migration between category have no impact on a possible re-interpretation.

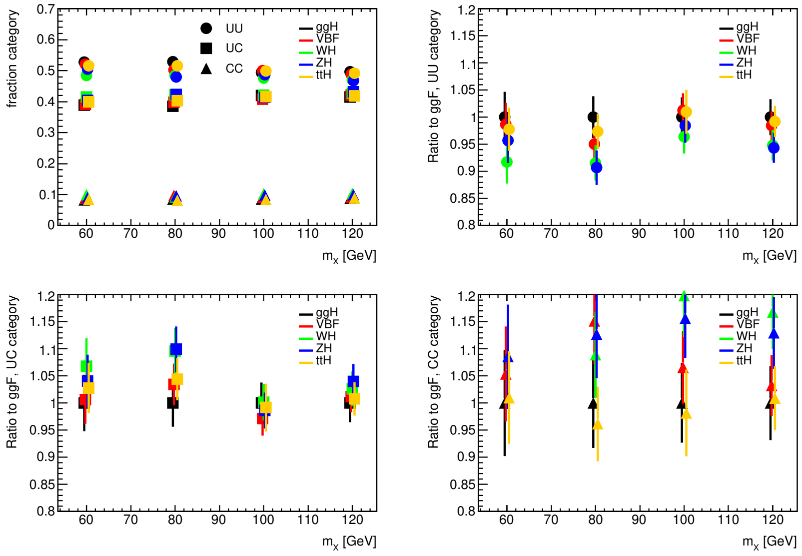


Figure 5.71: (a) Fraction of events in each category computed with the signal samples for different production modes. (b) Ratio of the fraction in each category to the ggF process for the **UU** category (c) for the **UC** category (d) for the **CC** category.

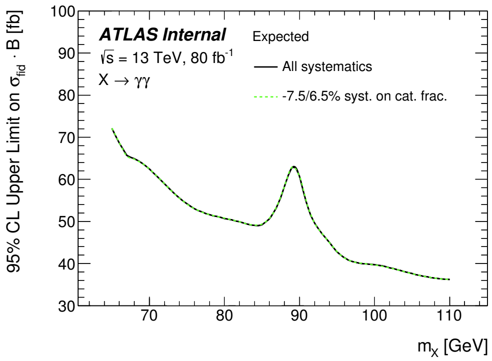


Figure 5.72: Expected limits on the fiducial production cross-section $\sigma_{fid}.BR(X \to \gamma\gamma)$ of a narrow resonance as a function of the resonance mass m_X in the 65-110 GeV mass range. The black line corresponds to the default systematics uncertainties, and the green line to more conservative migration uncertainties.

5.9.3 Systematics summary table

Table 5.22: Summary of the main sources of systematic uncertainty on the limit on σ_{fid}.

Source	Uncertainty	Remarks
Signal yield, C_X factors		
Luminosity	±2%	
Trigger	±1.4 − 1.7%	m_X-dependent
Photon identification	±1.5 − 2.3%	m_X-dependent
Isolation efficiency	±4%	m_X-dependent
Photon energy scale	±0.13 − 0.49%	m_X-dependent
Photon energy resolution	±0.053 − 0.28%	m_X-dependent
Pile-up	±1.8 − 4.1%	m_X-dependent
Production mode	±2.4 − 25%	m_X-dependent
Signal modelling		
Photon energy scale	±0.3 − 0.5%	m_X and category–dependent
Photon energy resolution	±2 − 8%	m_X and category-dependent
Migration between categories		
Material	−2.0 / +1.0 / +4.1%	category-dependent
non-resonant Background		
Spurious Signal	604 / 496 / 181 events	m_X and category-dependent
DY Background modelling		
Peak position	±0.1 − 0.2%	category-dependent
Template shape	±2 − 3%	category-dependent
Normalisation	±9 − 21%	category-dependent

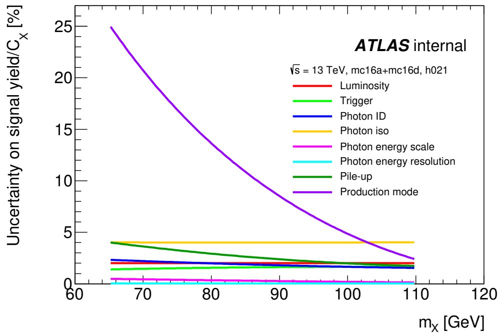


Figure 5.73: Summary of the systematic uncertainties affecting the signal yield or the C_X factors.

5.9.4 Impact of the expected limit

The impact of the different groups of systematics uncertainties (signal, DY background, Continuous background) is tested by computing the expected limit on the fiducial production cross-section, in Figure 5.74. The largest impact comes from the Spurious Signal systematics, making this analysis systematics-limited.

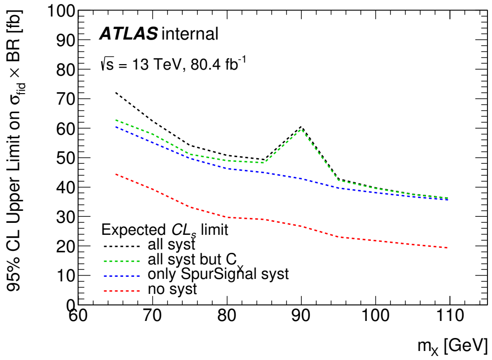


Figure 5.74: Expected limits on the fiducial production cross-section $\sigma_{fid}.BR(X \rightarrow \gamma\gamma)$ of a narrow resonance as a function of the resonance mass m_X in the 65-110 GeV mass range, for different hypotheses: in red no systematic uncertainties are considered, in blue only the Spurious Signal uncertainties are considered, in green all uncertainties are considered except for the one on the C_X factor, in black all uncertainties (including the ones on the signal yield and the DY background) are considered.

5.10 Results

5.10.1 Statistical Model

The data are described using an extended PDF expressed as

$$\mathcal{L} = \prod_{c=1}^{n_c} e^{-N_c^{total}} \prod_{i=1}^{n_c^{data}} \mathcal{L}_c(M_{\gamma\gamma}(i,c)) \tag{5.25}$$

where $n_c = 3$ is the number of categories, n_c^{data} and N_c^{total} are respectively the number of data events and the sum of the fitted numbers of events in each component in category c, and $M_{\gamma\gamma}(i,c)$ is the $M_{\gamma\gamma}$ value for event i of category c. The per-event term is expressed as

$$\begin{aligned} \mathcal{L}_c(M_{\gamma\gamma}; \sigma_{fid}, m_X, N_{uu,c}, N_{uc,c}, N_{cu,c}, N_{cc,c}, N_{bkg,c}, \mathbf{c}_c, \theta) = & \\ & N_{X.c}(\sigma_{fid}, m_X, \theta_{N_X}, \theta_{SS}) f_X(M_{\gamma\gamma}, m_X, \mathbf{x}_X(m_X), \theta_\sigma) \\ + & \; N_{uu,c}(\theta_{N_{uu,c}}) f_{uu,c}(M_{\gamma\gamma}, \mathbf{x}_{uu,c}, \theta_{uu,c}) \\ + & \; N_{uc,c}(\theta_{N_{uc,c}}) f_{uc,c}(M_{\gamma\gamma}, \mathbf{x}_{uc,c}, \theta_{uc,c}) \\ + & \; N_{cu,c}(\theta_{N_{cu,c}}) f_{cu,c}(M_{\gamma\gamma}, \mathbf{x}_{cu,c}, \theta_{cu,c}) \\ + & \; N_{cc,c}(\theta_{N_{cc,c}}) f_{cc,c}(M_{\gamma\gamma}, \mathbf{x}_{cc,c}, \theta_{cc,c}) \\ + & \; N_{bkg,c} f_{bkg,c}(M_{\gamma\gamma}, \mathbf{c}_c) \end{aligned} \tag{5.26}$$

where σ_{fid} is the fiducial production cross-section of the new resonance of mass m_X; $N_{uu,c}$, $N_{uc,c}$, $N_{cu,c}$, and $N_{cc,c}$ are the number of Drell-Yan background events identified respectively as unconverted-unconverted (contributing to the **UU** category), unconverted-converted, converted-unconverted (both contributing to the **UC** category) and converted-converted (contributing to the **CC** category). $N_{bkg,c}$ is the fitted number of background events and $\mathbf{c}_c$ collectively refers to the background parameters used to describe its shape; finally, θ collectively designates the nuisance parameters used to describe the systematic uncertainties. The nuisance parameters are listed below:

- θ_{lumi} : uncertainty on the integrated luminosity of the data sample.

- θ_{trig}, θ_{ID}, θ_{isol}, θ_{CX}, θ_{PU} : experimental uncertainties on the signal yield.

- θ_{ERS}, θ_{ESS} : experimental uncertainties on the signal yield and shape arising from energy resolution and scale.

- θ^{mat}_{mig}: migration uncertainty between categories.

- $\theta_{bkg,UU}$,$\theta_{bkg,CU}$,$\theta_{bkg,CC}$, : spurious signal systematics for each category.

- $\theta^{stat}_{norm,uu}$,$\theta^{stat}_{norm,uc}$, $\theta^{stat}_{norm,cu}$,$\theta^{stat}_{norm,cc}$: uncertainty on the normalization of each Drell-Yan component due to limited statistics in the computation of the template. Since this computation is done independently for each component, the errors are assumed to be uncorrelated. It takes also into account the systematic effect of the variation of the mass window used to compute the electron fake rate around the Zmass.

- $\theta^{mat}_{norm,uu}$,$\theta^{mat}_{norm,uc}$, $\theta^{mat}_{norm,cu}$,$\theta^{mat}_{norm,cc}$: uncertainty on the normalization of each Drell-Yan component due to systematic effects of the material. In principle those nuisance parameters could be considered as correlated since coming from the same source. But since part of this systematics could be of statistical origin due to the limited size of the samples, it was considered safer to consider them as uncorrelated.

- $\theta^{gen}_{norm,uu}$,$\theta^{gen}_{norm,uc}$, $\theta^{gen}_{norm,cu}$,$\theta^{gen}_{norm,cc}$: uncertainty on the normalization of each Drell-Yan component due to systematic effects of using the Sherpa generator instead of Powheg. The errors are assumed to be uncorrelated.

- $\theta^{BS}_{peak,uu}$,$\theta^{BS}_{peak,uc}$, $\theta^{BS}_{peak,cu}$,$\theta^{BS}_{peak,cc}$, $\theta^{BS}_{sigma,uu}$,$\theta^{BS}_{sigma,uu}$, $\theta^{BS}_{sigma,uc}$, $\theta^{BS}_{sigma,cu}$,$\theta^{BS}_{sigma,cc}$, $\theta^{BS}_{ALo,uu}$,$\theta^{BS}_{sigma,uc}$, $\theta^{BS}_{ALo,cu}$,$\theta^{BS}_{ALo,cc}$, $\theta^{BS}_{AHi,uu}$,$\theta^{BS}_{AHi,uc}$, $\theta^{BS}_{AHi,cu}$: uncertainty on the parameters of the Double-sided Crystal-Ball used as model of each Drell-Yan component computed with a bootstrap method (see Section 5.7). Since this computation is done independently for each component, the errors are assumed to be uncorrelated.

- $\theta^{mat}_{peak,DY}$, $\theta^{mat}_{sigma,DY}$: uncertainty on the peak position and width of each Drell-Yan component due to systematic effects of the material description. These effects are assumed to be correlated between Drell-Yan components, so are described as a function of a single nuisance parameters

- $\theta^{gen}_{peak,DY}$, $\theta^{gen}_{sigma,DY}$: systematic uncertainty on the shape of the Drell-Yan components of using the Sherpa generator instead of Powheg. These effects are assumed to be correlated between Drell-Yan components.

The quantity $N_{X,c}$ represents the number of events for the new resonance, described as the product of the total number of events (proportional to σ_{fid}) and the category fractions. This allows σ_{sig} to be directly extracted from the fit.

The number of Drell-Yan events in each category is defined as

$$N_i = N_i^0 \; \exp(\sigma^{stat}_{norm,i}\theta^{stat}_{norm,i}) \; \exp(\sigma^{mat}_{norm,i}\theta^{mat}_{norm,DY}) \; \exp(\sigma^{gen}_{norm,i}\theta^{gen}_{norm,DY})$$

for each component ($i = uu, uc, cu, cc$), where N_i^0 is the normalization of the DY template and $\sigma^{stat}_{norm,i}$, $\sigma^{mat}_{norm,i}$ and $\sigma^{gen}_{norm,i}$ are the systematic uncertainties on the template normalization due to statistical and systematic effects (material description and alternative generator). The nuisance parameters are described above.

The PDF of each Drell-Yan component is described with double-sided Crystal Ball shape, $f_i(M_{\gamma\gamma}, \mathbf{x}_i(\theta_i))$, where the $\mathbf{x}_i(\theta_i)$ are the parameters of the PDF and the θ_i refer to the nuisance parameters on the Drell-Yan peak position and shape. The nominal values of the PDF parameters are obtained by fitting the nominal templates obtained in Section 5.7. The shape uncertainties have three sources: the uncertainty coming from the combined effect of limited statistics in the MC and data samples and the background subtraction used to get the template (referred as the “Bootstrap” BS, uncertainty) ; the uncertainty coming from the detector material description ; the uncertainty on the MC generator. The material uncertainties on the shape are obtained by fitting the up/down distorted templates for each of the two shape uncertainties and linearly interpolating the PDF parameters between the "down", "nominal" and "up" values (corresponding respectively the values -1, 0, and 1 of the $\theta^{mat}_{peak,DY}$ and $\theta^{mat}_{sigma,DY}$ parameters). The effect of the two shape variations are assumed to be independent. The generator uncertainties on the shape are similarly described with the $\theta^{gen}_{peak,DY}$ and $\theta^{gen}_{sigma,DY}$ parameters.

The continuum background PDF $f_{bkg}(M_{\gamma\gamma}, \mathbf{c})$ is described by the function chosen for each category in Section 5.6.3.

5.10.2 Partial Unblinding

It was decided during the approval process that the background model, especially the DY one, would be validated using 10% of the data, corresponding to an integrated luminosity of 8 fb^{-1}, where the events are picked up randomly. A background-only fit and a p-value scan are run and shown in Figure 5.76a and the pulls of the fits are checked carefully, as shown in Figure 5.77.

A small excess with a local significance of 2.4σ is observed at a mass of 90 GeV. This corresponds to a slight deviation of the nuisance parameter $\theta^{mat}_{norm,uc}$, in the systematic on the normalization of the DY background, arising from the detector material description, for uc events. After several checks nothing wrong could be found. In particular most of the excess comes from the *cu* events, as can be seen in Figure 5.76b.

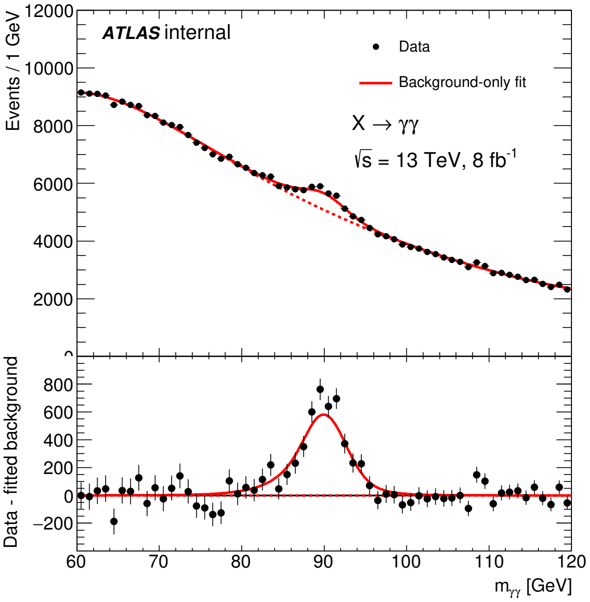
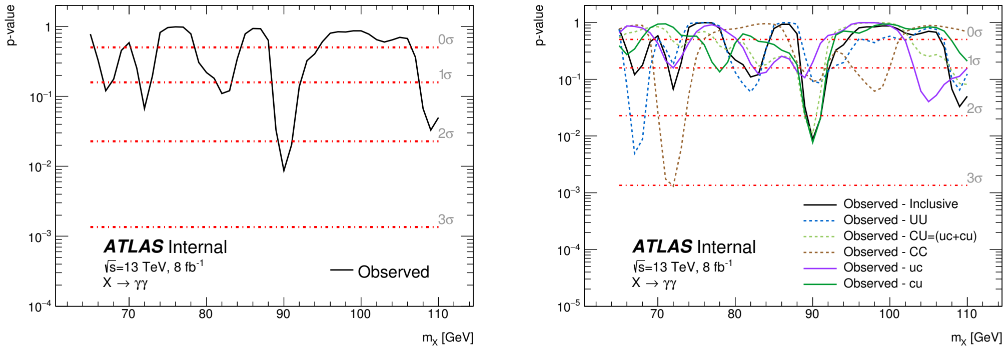
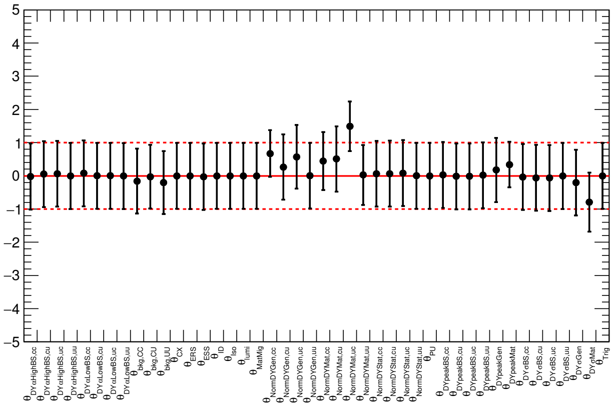


Figure 5.75: Data distributions of $M_{\gamma\gamma}$ overlaid with background-only fits, computed with 10 % of the total dataset.

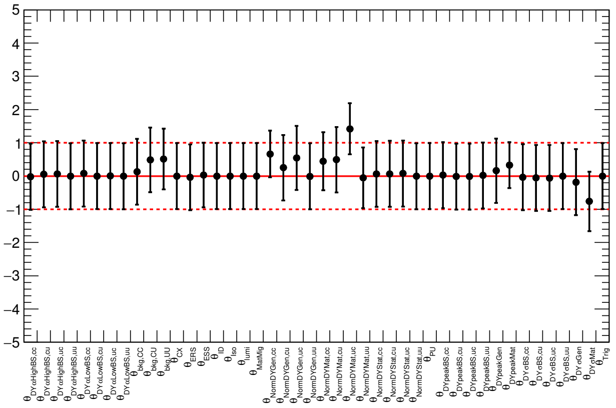


(a) (b)

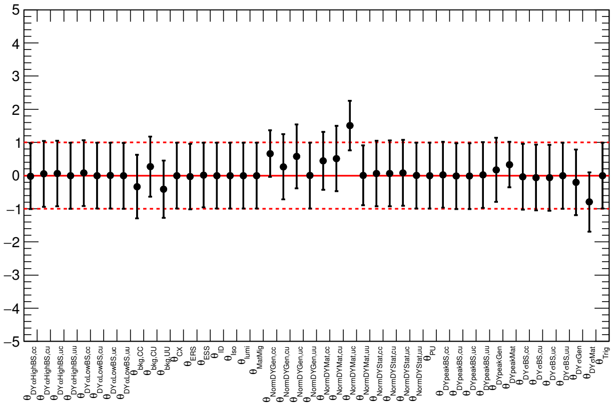
Figure 5.76: Observed p-value for the background-only hypothesis, p_0, as a function of the resonance mass m_X, computed with 10 % of the total dataset, (a) with the combined fit (b) by fitting the categories one by one.



(a) $m_X = 65$ GeV



(b) $m_X = 90$ GeV



(c) $m_X = 105$ GeV

Figure 5.77: Pulls of the constrained nuisance parameters obtained for a signal mass hypothesis of (a) 65 GeV, (b) 90 GeV, (c) 105 GeV, computed with 10 % of the total dataset.

5.10.3 Distributions of Invariant Mass in Data

5.10.3.1 Prefit distributions

The data distribution with the full dataset is compared to the background template described in Sections 5.6.2 and 5.7 before any fits are applied in Figure 5.78. One can see that before the fit there is already a good agreement between data and the background template within uncertainties. A small excess of data can still be noticed in the DY region, but covered by the systematic uncertainties on the normalization of the DY background.

5.10.3.2 Postfit distributions

The data distribution with the full dataset is shown in Figure 5.79, as well as the bkg-only fits. The DY peak is clearly visible, and no structure is seen in the residuals. The small excess of events in the DY region of the **UC** category is still visible post-fit This is related to the pull distributions in Figure 5.82b discussed below.

5.10.4 Discovery p-value

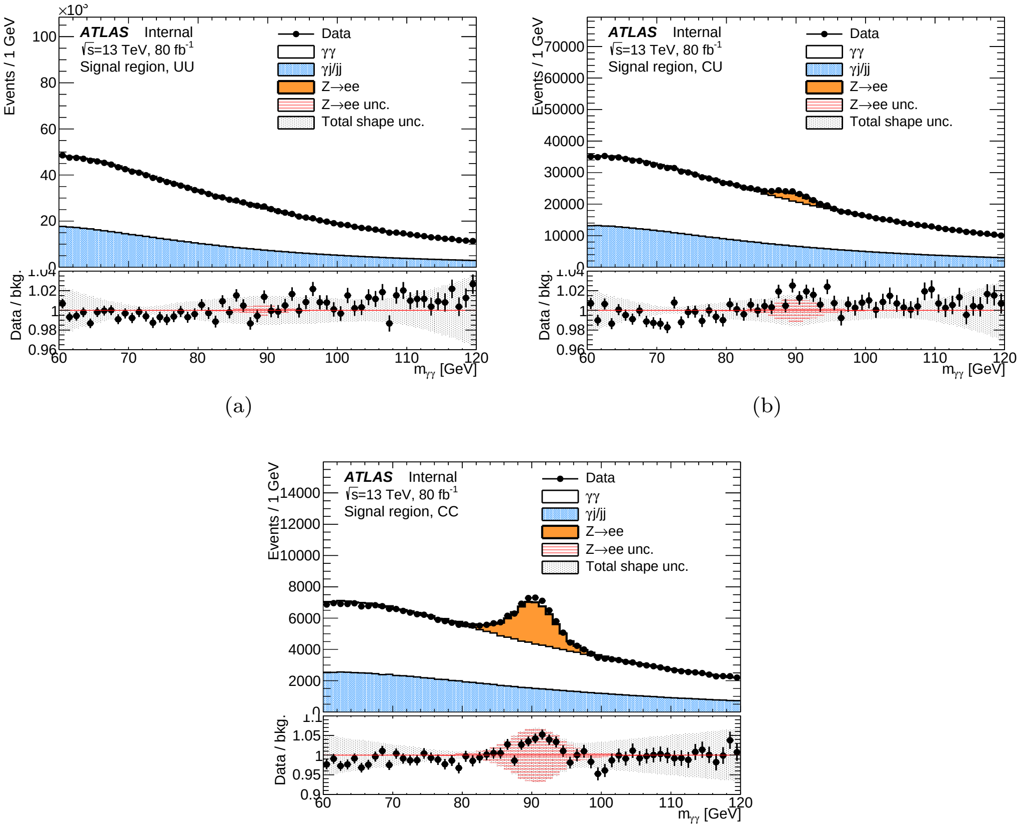
The p-value scan with the full dataset is shown in Figure 5.80. No excess is seen on data.

In order to compute the fluctuation we could except in the search range, one can use the approximate formulas to compute the global significance, since the average peak fluctuation is given by finding the value of the local significance Z_{local} which gives $Z_{global} = 0$:

$$\begin{aligned} p_{global} &= 1 - (1 - p_{local})^{N_{trials}} \\ N_{trials} &= 1 + \sqrt{\pi/2} \cdot N_{indep} \cdot Z_{local} N_{indep} = \frac{\text{scan range}}{\text{peak width}} \end{aligned} \tag{5.27}$$

which gives the following formula for $Z_{global} = 0$:

$$(1 - \Phi(Z_{local}))^{1+\sqrt{\pi/2} \cdot N_{indep} \cdot Z_{local}} = 0.5 \tag{5.28}$$



(a)

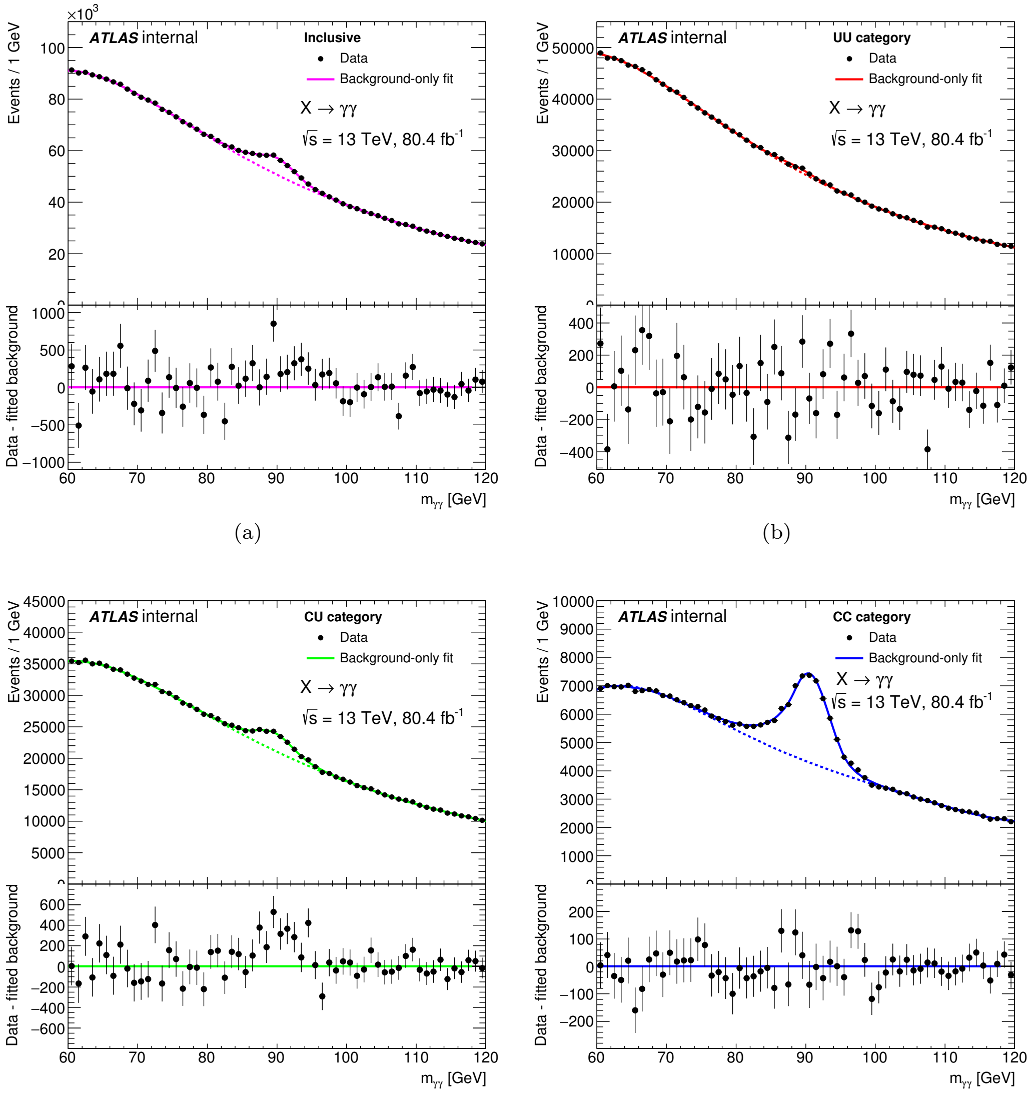
(b)

(c)

Figure 5.78: Data distributions of $M_{\gamma\gamma}$ overlaid with the background template, for (a) the **UU** category, (b) the **UC** category and (c) the **CC** category. The $\gamma\gamma$ contribution is taken from the Sherpa MC samples, the $\gamma j/jj$ contribution from data-driven control regions, and the $Z \to ee$ contribution from the data-driven template. The bottom panels correspond to the ratio between data and the fit, with the total uncertainty in grey and the DY normalization uncertainty in red.

The search range is 45 GeV while the width the signal is around 1.5 GeV, which gives that we expected a 2.4σ fluctuation on average.

Figures 5.81 and 5.82 show the pulls of the constrained nuisance paramters. They do appear to have any particular structure. The pulls of the DY normalization are a bit pulled by almost 1σ.



(a)

(b)

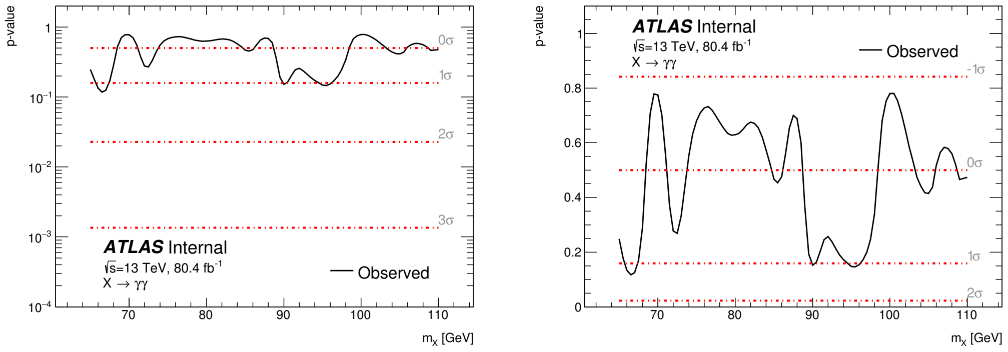
(c)

(d)

Figure 5.79: Data distributions of $M_{\gamma\gamma}$ overlaid with background-only fits, for (a) the inclusive selection (b) the **UU** category, (c) the **UC** category and (d) the **CC** category. The bottom panels correspond to the difference between data and the the fit.

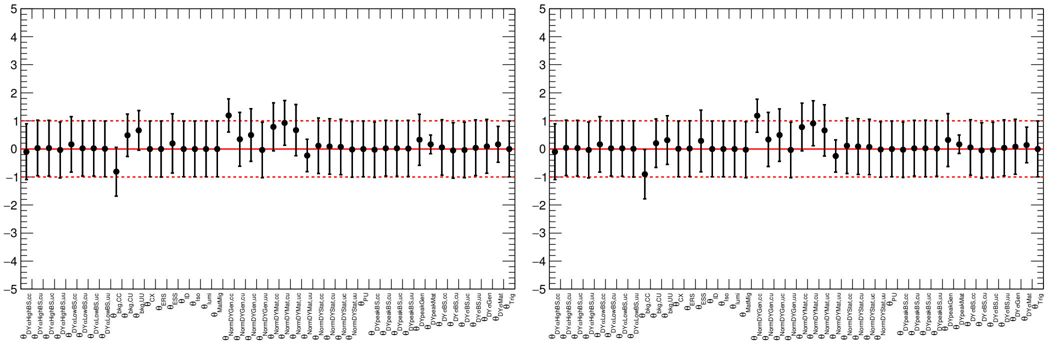
5.10.5 Limit on the fiducial cross-section

The expected and oberved limits are shown in Figure 5.83.



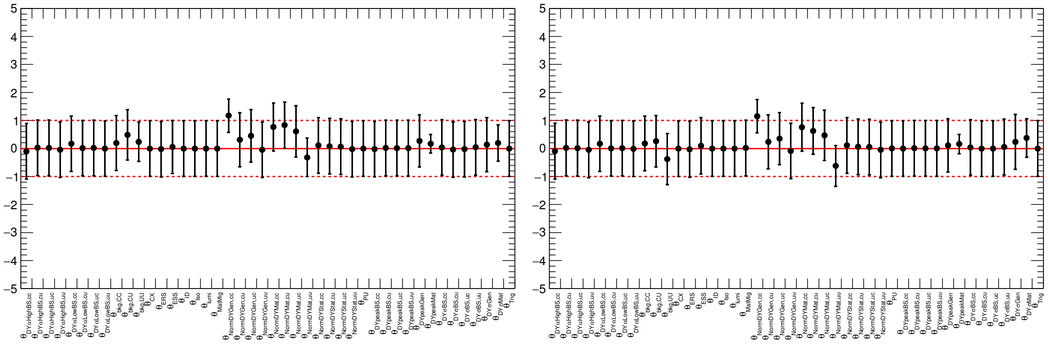
(a) (b)

Figure 5.80: Observed p-value for the background-only hypothesis, p_0, in (a) logarithmic and (b) linear scale, as a function of the resonance mass m_X.



(a) (S+B) fit, m_X = 65 GeV (b) B-only fit, m_X = 65 GeV

Figure 5.81: Pulls of the constrained nuisance parameters obtained for a signal mass hypothesis of 65 GeV, for the (a) S+B fit and (b) the background-only fit.



(a) (S+B) fit, m_X = 90 GeV (b) B-only fit, m_X = 90 GeV

Figure 5.82: Pulls of the constrained nuisance parameters obtained for a signal mass hypothesis of 90 GeV, for the (a) S+B fit and (b) the background-only fit.

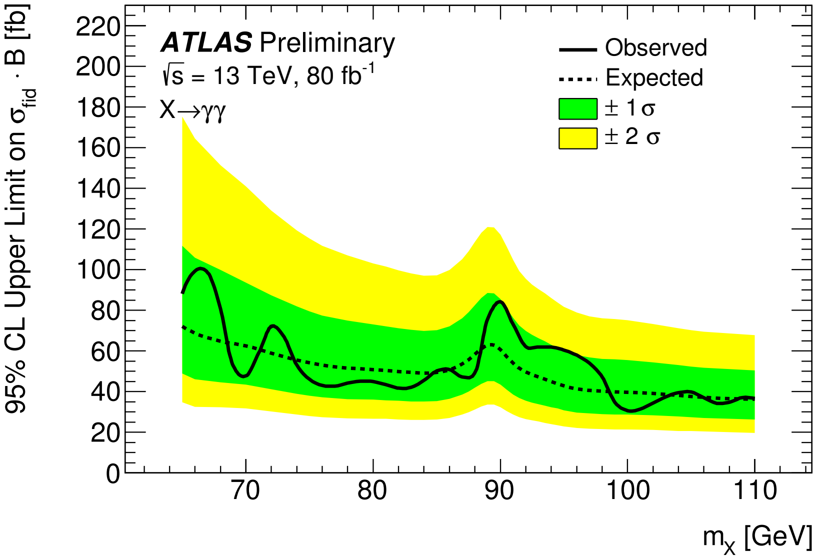


Figure 5.83: Expected and observed limits on the fiducial production cross-section $\sigma_{fid}.BR(X \rightarrow \gamma\gamma)$ of a narrow resonance as a function of the resonance mass m_X in the 65-110 GeV mass range. The green and yellow bands show the ±1σ and ±2σ uncertainties on the expected limit.

5.11 Conclusion

In this analysis we performed a search for narrow diphoton resonances in the range 65-110 GeV. No significant excesses were and a limit is set on the fiducial cross section between 80 fb and 40 fb. This analysis is limited by systematics and limited by the spurious signal systematic particularly. It is clear from figure 5.74 that the largest improvements to be made to this analysis can be made in the non-resonant background description. The spurious signal test requires large amounts of MC in order to perform well and it is no longer feasible for these samples to be produced. A new method needs to be found before the next round of analyses.

Chapter 6

Gaussian Process For Modeling Backgrounds

Many of the analysis and statistical methods used on ATLAS, and high energy experiments in general, are based on methods used in previous experiments where computing resources were far more limited than they are now. Although many newer methods have found their way into ATLAS, the breadth of the so called ”machine learning” techniques remain relatively unexplored when compared with computer science and the data science industry. This is not completely surprising as we in the physics field tend to make much more precise measurements, and need a far greater understanding of systematic uncertainties and biases, than does the data science industry. With more complicated methods also comes more difficulty in explaining, in understanding, and in using these methods. However, the ever increasing size of the ATLAS dataset, and the requirements for better and better modeling of the data, are forcing us to increasingly turn toward these more complicated and computationally intensive methods to meet our needs.

In this chapter I will discuss Gaussian Process Regression (GPR) and an investigation into how it might be used in the low-mass diphoton analysis. Because of the very short timescale I am on to graduate, I will not be able to see GPR used in the results of the analysis, so my intention is to document my work exploring this method. I will attempt to make this chapter as useful as possible to a student looking for an introduction to using GPR for a modeling task in their analysis. In my experience with GPR, conceptual understanding of the method has been difficult despite the quite simple mathematics of actually applying the method. Although there are many others that are much more expert in GPR than I am, I hope to be able to give a good introductory explanation that can help the reader develop a solid intuition for the mechanics of the technique. The algorithmic simplicity of GPR along with its robustness against assumptions put into a model, I think, make it a great candidate to be used for many regression problems that analysis groups face on ATLAS.

Much of this chapter is drawn from Gaussian Process for Machine Learning by Rasmusen and Williams, which is available online for free [66].

6.1 Gaussian Process Regression

There are two main viewpoints that are taken to describe the Gaussian Process, the weight-space view, and the function-space view. I find the function space view to be more intuitive and also more in line with the way we think about modeling in ATLAS, so I will concentrate on this. I am going to start with a review of some properties of multivariate Gaussian distrbutions and some definitions before moving into their relationship to GPR.

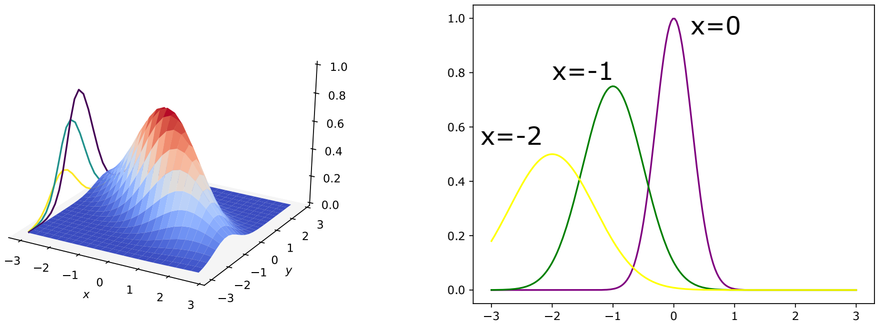
6.1.1 Introduction to GPR

A multivariate Gaussian (MVG) is simply a high dimension representation of a Gaussian or normal distribution. A one-dimensional Gaussian is given by equation 6.1 where μ is the mean of the distribution and σ is the width, or standard deviation. Often, this is written as

$\mathcal{N}(\mu, \sigma)$.

$$\mathcal{N}(\mu, \sigma) = \frac{1}{\sqrt(2\pi\sigma^2)} e^{-\frac{(x-\mu)^2}{\sigma^2}} \tag{6.1}$$

In the multivariate case, we just extend this to more dimensions. The x and μ become vectors of dimension D which is also the dimension of the MVG, and σ becomes a matrix of size $D \times D$ (often called the covariance matrix). We then write this as $\mathcal{N}(\mathbf{x}, \mathbf{\Sigma})$, with bold letters representing vectors and bold capitals representing matrices. MVGs have some very convenient properties, the most important of which is that the marginal distribution of a Gaussian is a Gaussian. This means that if we were to take a 2-D Gaussian and slice it along a line in the 2-D plane, the distribution along the slice would be a 1-D Gaussian. This is also true of high dimension MVGs where slices produce a Gaussian distribution of lower dimension. Figure 6.1 shows a multivariate Gaussian with $D = 2$ and three different marginal distributions corresponding to slicing the 2-D distribution along constant values of the x axis.



(a) (b)

Figure 6.1: A 2-D Gaussian distribution with the marginalized 1-D Gaussian distributions shown projected on the left plot (a) and on their own in the right plot (b). The purple, green and yellow lines correspond to marginalization along the y axis given x values of 0, -1, -2, respectively.

Notice that depending on where we slice the 2-D distribution we get 1-D distributions with different values of μ and σ. We would like to be able to calculate the μ and σ for the marginal distributions given the parameters for the 2-D distribution. Luckily for Gaussians this is easy. Continuing with the setup shown in figure 6.1, we make make some more concrete definitions for the 2-D Gaussian in eq.6.2.

$$\begin{bmatrix} f_x \\ f_y \end{bmatrix} = \mathcal{N}\left(\begin{bmatrix} \mu_x \\ \mu_y \end{bmatrix}, \begin{bmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{bmatrix}\right) \tag{6.2}$$

Now we have defined this Gaussian as having dimension 2 with axes labeled x and y. We would like to find the mean and width of the marginal distributions along one axis, lets say y, given a value of the other axis, x, written as $\mu_{y|x}$ and $\Sigma_{y|x}$ respectively. Eq.6.3 shows these results without proof. This is a well known result and can be found in many standard texts.

$$\begin{aligned} \mu_{y|x} &= \mu_y + \Sigma_{xy}\Sigma_{xx}^{-1}(x - \mu_x) \\ \Sigma_{y|x} &= \Sigma_{yy} - \Sigma_{yx}^{T}\Sigma_{yy}^{-1}\Sigma_{yx} \end{aligned} \tag{6.3}$$

μ_y here is the y-component of the MVG mean. Often, this will be taken as 0 when used in GPR (more on this later). We will take the mean to be 0 for now ($\mathcal{N}(\mu, \sigma) \rightarrow \mathcal{N}(0, \sigma)$)

One more point we need to make before moving on is the difference between a continuous function represented by a functional form, and a binned function that can be represented by a vector. We are used to referring to functions defined as a functional form e.g. $f(x) = x^2 + 1$. This is defined for all values of x in $\mathcal{R}^1$. When talking about binned data, it can be more convenient to describe a function as a vector $\mathbf{f}_n = \{x_0, x_1, ..., x_n\}$. Any functional form can be described this way for a specified number of bins, but this has the advantage that any arbitrary function, including one that may not have a closed functional form, may also be represented this way. This can be used interchangeably with the functional form when considering binned distributions. Instead of a continuous dependent variable, the index of the vector is used in its place.

With the tools in equation 6.3 we can begin to discuss how this can be used in GPR. As a simple example we can consider the set of data points with no uncertainty as shown in figure 6.2. To start with GPR we need to construct a multivariate Gaussian distribution where the number of dimensions is the number of data points in your data set (6 in the case of figure 6.2).

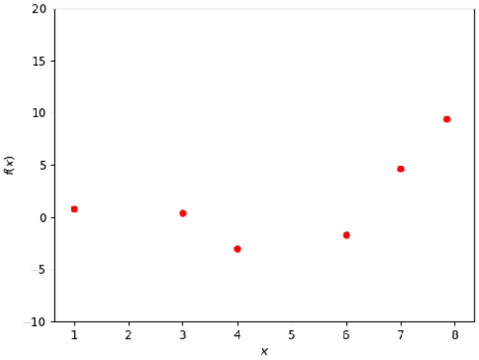


Figure 6.2: Example data points with no uncertainty.

For this early example we will take the mean of this MVG, μ, to be 0. For the covariance matrix we will need a 6×6 matrix. To fill out this matrix we will use a covariance function $k(x, x')$, where x and x' are two points on the x-axis. The function $k(x, x')$ is also called the 'kernel'. With 5 data points the covariance matrix will have dimension 6×6. The role the covariance function plays is to tell the MVG distribution how each of its axes, and in turn each data point, are related to one another. The behavior that we would like this function to have is that the closer two data points are together on the x-axis, the closer we would like them to be along the y-axis. Loosely, this gives us some sense of smoothness of the data points. The covariance function that we will use is the squared exponential shown in equation 6.4. This is often called the Radial Basis Function (RBF) in machine learning litterature. This kernel has 2 parameters, σ_f and l. Because these are parameters of the model and not the function or distribution they are commonly called hyperparameters. These allow us to adjust the behavior of the covariance matrix. Roughly speaking, σ_f is the maximum allowed covariance between close points, and l is the distance at which points along the x axis become decorrelated. I will go into more detail on hyperparameters and selecting their values later. For now, let's assume we have chosen reasonable values for σ and l.

$$k(x, x') = \sigma_f \exp\left(-\frac{(x - x')^2}{2l^2}\right) \tag{6.4}$$

I will denote the matrix of these $k(x, x')$ values as $\mathbf{K}$. We now have everything we need to construct our MVG distribution, $\mathcal{N}(0, \mathbf{K})$. We can now begin to make predictions at values along the x axis. This is done by marginalizing the distribution along the axis corresponding to the x value we would like to predict at (remember there is one dimension per point on the x-axis). At this point, this isn't very interesting as the only x values we have in our matrix are the 5 points in our data. Since we are not considering any uncertainty on the points yet, this marginalization will just return the y value of the point in question (it's easy to check this for yourself). We would like to be able to make predictions at new points on the x axis. Extending the covariance matrix to 6 dimensions will allow us to do this. The new entries in the matrix will be denoted $\mathbf{K}_* = [k(x_*, x_1), k(x_*, x_2), ..., k(x_*, x_n)]$ where n is the number of data points and $\mathbf{K}_{**} = k(x_*, x_*)$ with the * entries being the new x location and are added to the MVG as in equation 6.5.

$$\begin{bmatrix} \mathbf{y} \\ y_* \end{bmatrix} = \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} \mathbf{K} & \mathbf{K}_*^T \\ \mathbf{K}_* & \mathbf{K}_{**} \end{bmatrix}\right) \tag{6.5}$$

Repeating this process for many x values, we can construct a high dimensional MVG with fine enough binning along the x axis to look like a smooth function. For example, if we were to use 1000 equally spaced x values we will get a 1000 dimension MVG where each point in this space defines one curve in our $x-y$ plane. This is the prior distribution. Figure 6.3 shows 10 random points drawn from the prior MVG distribution.

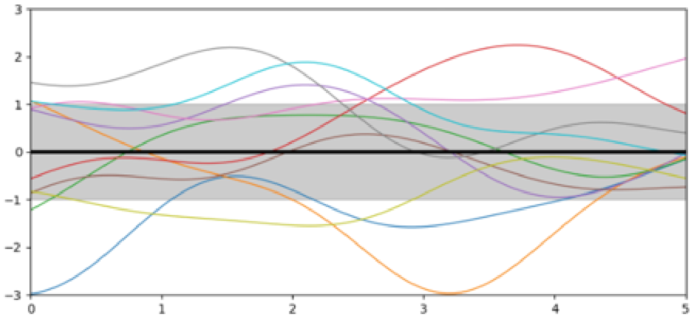


Figure 6.3: 10 random draws from the prior distribution construced with the squared exponential kernel.

It should now be clear that every point in this MVG space represents a function whose functional form is not explicitly specified (its represented as a vector of numbers). This also means that this MVG can be interpreted as a prior probability distribution over functions. To perform regression, what we really want is the posterior distribution. This will be a distribution over likely functions that describe the data that the MVG is conditioned on. To get the posterior distribution, we can return to equation 6.3. This is easy to extend to more than two dimensions to cover our example case. This is shown in equation 6.6 for the mean and variance of the posterior distribution along the x_* axis.

$$\begin{gathered} y_* = \mathbf{K}_* \mathbf{K}^{-1} \mathbf{y} \\ var(y_*) = \mathbf{K}_{**} - \mathbf{K}_* \mathbf{K}^{-1} \mathbf{K}_*^T \end{gathered} \tag{6.6}$$

Now we are able to get the marginalized distribution along an axis of the MVG corresponding to any value of the x axis in our data. To say this another way, at any x value in our data, the MVG defines a Gaussian distribution that the value of $y(x)$ is distributed along. Figure 6.4 shows 10 random draws from the posterior distribution that have been conditioned on the example data points using 1000 points along the x axis.

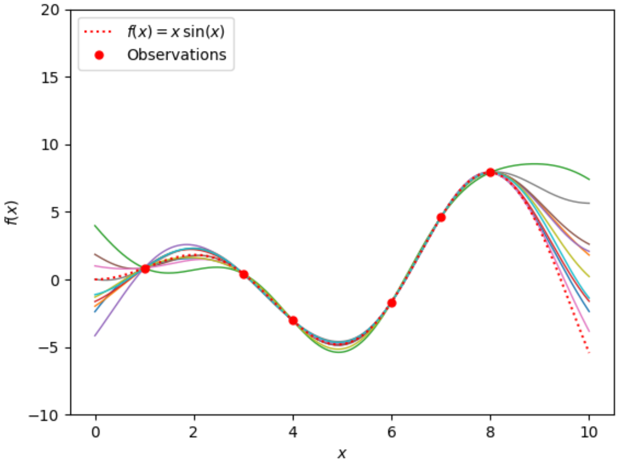


Figure 6.4: 10 random draws from the posterior distribution constructed with the squared exponential kernel and conditioned on the example data points. The red dashed line is the function used to sample the data from.

Figures 6.3 and 6.4 show that the Gaussian Process can be thought of a distribution over functions whose correlations are defined by the kernel $\mathbf{K}(x, x')$. When we then consider some data, this has the effect of down-weighting parts of the posterior distribution that are not consistent with this data. In the end we are not really interested in the distribution of functions around the data, but a prediction for the most likely value that a function describing the data will take. Plotting the mean and 2σ variance at each point for the data gets us figure 6.5

The function that the data points are taken from is shown by the red dashed line. We can see that our prediction for the mean is quite close to this function despite only having a few data points to fit. It is worth noting a few things here. No assumptions were made on the functional form or the shape we expect in the data beyond using the squared exponential

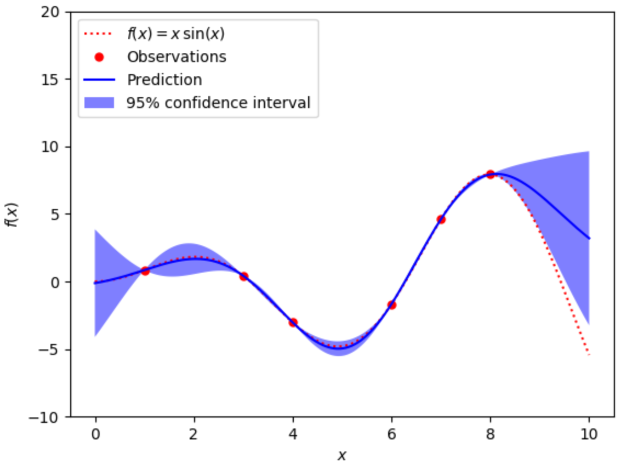


Figure 6.5: The mean (blue line) and 2σ variance (blue shaded region) of the posterior distribution constructed with the squared exponential kernel and conditioned on the example data points. The red dashed line is the function used to sample the data from.

kernel. We can also see that the variance on our distribution grows when we get further away from data.

In this example we have also not considered any uncertainty on the data points. This is reflected in the prediction made by the GP which you can see because the mean passes exactly through the data points and the variance goes to 0 at the points as well. We are unlikely to run into a situation where there is no uncertainty on our data points. For GPs the uncertainty is easy to model. In the case of the squared exponential kernel, we simply need to add the uncertainty on our measurements to the diagonal of the covariance matrix. This will make the kernel look like equation 6.7

$$k(x, x') = \sigma_f^2 \exp\left(-\frac{(x-x')^2}{2l^2}\right) + \sigma_n^2 \delta(x, x') \tag{6.7}$$

where $\delta(x, x')$ is the Kronecker delta. As you can see, this will mean that the diagonal of the matrix will be $\sigma_f^2 + \sigma_n^2$. To illustrate this situation, figure 6.6 shows the same generating function as the above plots but with some random noise put in. For this simple example, a random shift away from the function is added as well as an error bar. The uncertainty on the points represented by the error bars are then added into the covariance function as in equation 6.7. More points have been added to this plot as it makes the example more clear. These points are still taken from the same underlying distribution.

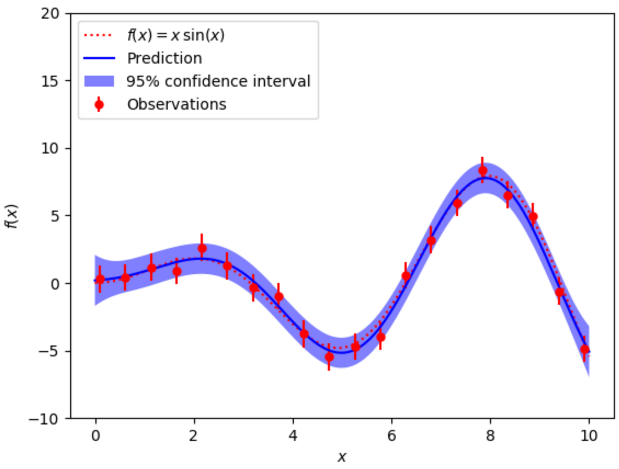


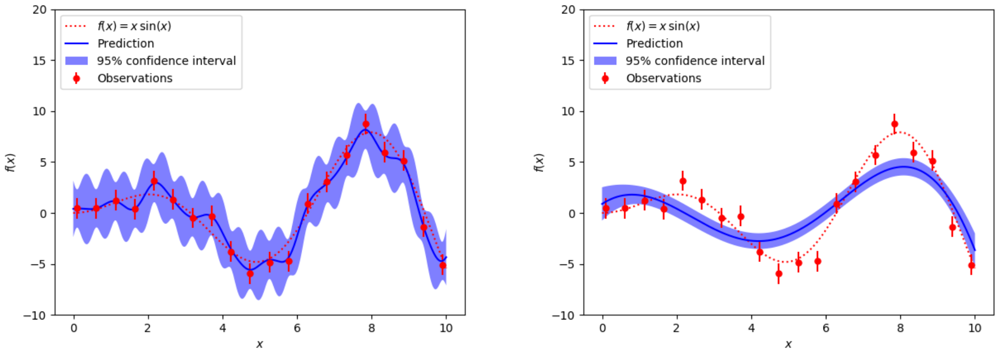
Figure 6.6: The mean (blue line) and 2σ variance (blue shaded region) of the Gaussian Process prediction to noisy data. The points have been randomly shifted away from their generating function (red dashed line) and uncertainty has been added.

The mean now is not required to pass exactly through the data points and the variance on the prediction also reflects this by staying non-zero, even close to measured values.

6.1.2 Kernel Hyperparameters

Just about any kernel you choose will have some number of hyperparameters in them. Our example of the squared exponential kernel had two, σ_f, the amplitude, and l, the length scale. Previously we had just assumed that we had reasonable choices made for these two, but now we will discuss the ways that we can select values for the hyperparameters.

To illustrate the effect that the length scale can have on the GP prediction, figure 6.7 shows the data from the previous example with l fixed to a value that is fixed to values away from the optimal value (more on what optimal means later). Setting the length scale too short results in the prediction being allowed to fluctuate up and down a lot in between the data, where as too long gives an oversmoothed prediction.



(a) (b)

Figure 6.7: GP prediction on the data with $l = 0.3$ (a) and $l = 6$ (b).

These examples highlight the need to select appropriate values for these hyperparameters. One way to do this is to maximize the probability $P(\boldsymbol{\theta}|\mathbf{x}, \mathbf{y})$ where $\boldsymbol{\theta}$ are our hyperparameters. In the case of two hyperparameters we have $\boldsymbol{\theta} = \{\sigma_f, l\}$, but these could be any hyperparameters associated with your chosen kernel. Now, Bayes theorem tells us that maximizing this corresponds to maximizing $P(\mathbf{y}|\mathbf{x}, \boldsymbol{\theta})$. Because we are making the assumption that our prior probability distributions are Gaussian distributed, this is exactly our definition of the MVG space we are using for the GP (normalized to unity to make it a probability of course). The multivariate form of a normal distribution written out is

$$\mathcal{N}(\boldsymbol{\mu}, \mathbf{K}) = \frac{1}{\sqrt{(2\pi)^n |\mathbf{K}|}} \exp\left(-\frac{1}{2}(\mathbf{y} - \boldsymbol{\mu})^T \mathbf{K}^{-1} (\mathbf{y} - \boldsymbol{\mu})\right) \tag{6.8}$$

where n is the dimension (the number of data points) and $\boldsymbol{\mu}$ is the mean of the MVG which we have been setting to 0. For convenience we can take the logarithm of this distribution, and we arrive at what is called the log marginal likelihood (LML),

$$logP(\mathbf{y}|\mathbf{x},\boldsymbol{\theta}) = -\frac{1}{2}\mathbf{y}^T\mathbf{K}^{-1}\mathbf{y} - \frac{1}{2}log|\mathbf{K}| - \frac{n}{2}log2\pi \tag{6.9}$$

where I have set $\boldsymbol{\mu} = 0$. The contour plot of the log marginal likelihood for our example data set is shown in fig.6.8. We can then use our favorite minimizer to find where this distribution is extremized, and this will give us good starting values for our hyperparameters. We can see that the first term in equation 6.9 is the only one that depends on the input data points. This term will decrease the LML with increasing length scale as the GP model becomes less and less flexible and therefore less able to describe the data. The second term is often referred to as the complexity penalty. This term only depends on the covariance kernel and will increase the LML with increasing length scale. Just as the name implies, this term penalizes the likelihood for selecting more complex models with shorter length scale. See [66] for a detailed discussion of hyperparameter selection.

The values obtained this way are a great place to start, but there are a few considerations we should make when ultimately determining what values to use. One of the big advantages of the GP is that we can use the kernel and the hyperparameters to inject knowledge of our model into the prediction. As an example, let's consider the length scale hyperparameter. Even before we go through the process of maximizing the log marginal likelihood, we may have some understanding of what we expect the behavior of our data to be. Even though our data might fluctuate up and down quite a bit, which in some cases would cause the maximization to tend toward a shorter length scale, we may know that we only expect bumps or structure beyond some characteristic length scale. We can then set a minimum value in

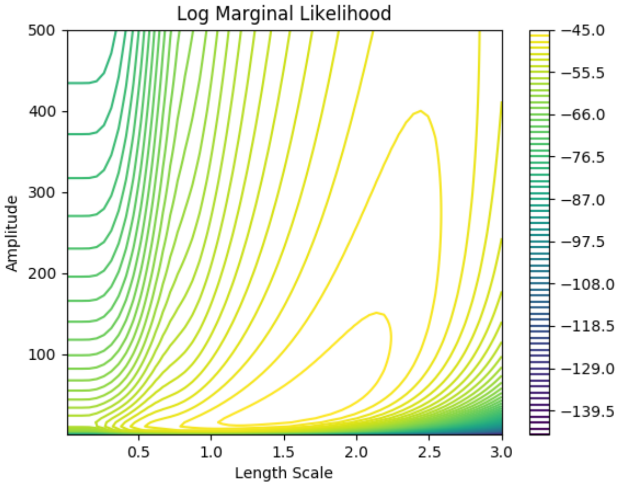


Figure 6.8: The log marginal likelihood contours as function of the amplitude σ_f, and the length scale l.

the optimization of this hyperparameter to ensure that the GP is fitting with some level of smoothness that we expect. This is extremely valuable in cases where you know ahead of time some of the features about the physical process you are modeling, or perhaps the response of your measurement apparatus. Selecting the correct hyperparameters can allow a lot of control over the way your GP performs while still remaining flexible enough to let the data speak for itself.

In addition to hyperparameter selection the choice of kernel can also have a similar impact on the performance of the Gaussian Process. In cases where you would like to model specific behavior, you may need to choose a kernel that has more hyperparameters, or one whose covariance has specific characteristics. As an example, consider a RBF kernel like we have used here, but instead of a single value for the length scale l, we could allow the length scale to vary along the x-axis $l(x)$. In the case where the length scale is a linear function, you get a covariance known as the Gibbs kernel. The process of choosing or designing a kernel is often referred to as kernel engineering. Reference [66] contains a lot of useful detail and discussion on several common kernels used for various regression and classification tasks.

6.2 Gaussian Process in Low Mass Diphoton

As described in Chapter 5, there are some challenges to modeling the backgrounds in the diphoton resonance search. One major challenge is the non-resonant background shape. This is currently modeled with a function that is selected by using the spurious signal test on a high statistics sample of Monte Carlo. In order to get good results from the spurious signal test, the amount of Monte Carlo that needs to be produced must increase with the amount of data collected. Already the number of events needed has become infeasible due to the high luminosity of the LHC. In addition, the turn-on curve at low invariant mass caused by the trigger requirements make appropriate functional forms difficult to find. They are also limited to only being able to model a small portion of this shape that can be described by the relatively simple functions that can be used. As we collect more data, it is becoming more clear that we are not able to easily select a non-biased description of the background. These are areas where the Gaussian Process (GP) can improve our background description.

When we first began discussing using a GP to model the diphoton non-resonant background, we took an approach similar to what another group on ATLAS uses to smooth and interpolate distributions. The dijet+ISR group looks for resonances in the dijet invariant mass spectrum in events with large initial state radiation. This group has recently begun to use GP to smooth a distribution, known as a transfer factor that is used to estimate a number of events in a signal region of the data based on a control region [67]. This factor is defined on a 2-D plane. Points of this 2-d plane are calculated in a grid and are then fit using a 2-D GP. In their case, to avoid the GP fitting any signal that might exist in the sample, they remove points in a window around where they are currently fitting for the transfer factor. This allows them to fit across the window and not be biased by anything in the window.

For our first attempt at using the GP in diphoton, we removed a window around the search mass from the background distribution in data, similar to the dijet+ISR group's strategy. This allowed us to fit across the window. If a signal bump exists in data, then we don't run the risk of absorbing that into our background description since it would be completely removed. Figure 6.9 shows an example of a diphoton background distribution with a 10 GeV window removed around 125 GeV where a signal is injected. Note, this distribution has different limits than the low-mass diphoton analysis discussed earlier. During this early stage of testing there were no low-mass samples yet available. In the bottom panel the ratio of the data points to the fit shows that the signal bump injected into the sample appears to be recovered as the GP is fit only on the sidebands of the distribution where there is no signal. Although this method recovers the injected signal, it could possibly induce biases if there are effects that could cause one side of the window to move away from the background. For example, if there were two signal bumps near each other. This could cause one side of the window to be dragged upward, and even though the window is centered on a real signal, the GP would absorb a portion of it, causing a loss in sensitivity. Having two as yet unknown signals near each other in this way is very unlikely (and the kind of problem physicists would love to have), but we decided to try to leverage some of the properties of the GP to eliminate this problem.

As mentioned in the previous section, the hyperparameters can give you quite a bit of input to how you would like you GP to behave. In particular the length scale parameter of the squared exponential covariance function in equation 6.7 will be useful here. It is useful to note at this point that the length scale has units of the x-axis of the distribution you are fitting which in this case is GeV. This hyperparameter allows for the control of the scale at which structure in the GP is fit. This means that if we set an appropriate lower limit on this parameter, we should be able to cause the GP to ignore narrow bumps. In this case we have

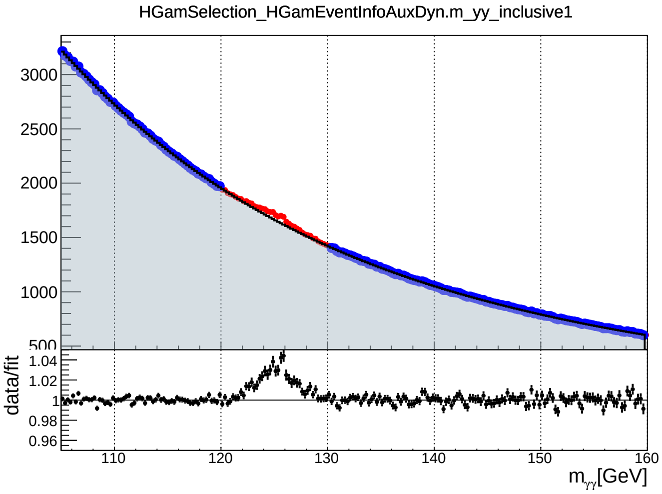
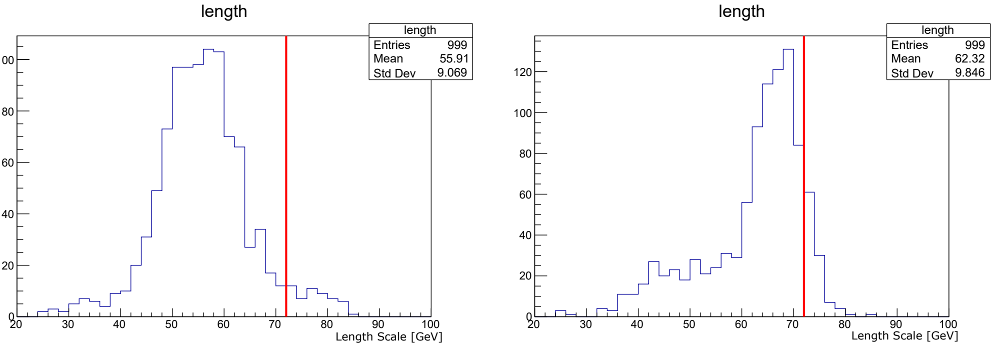


Figure 6.9: An example of removing a window around a search point for GP fitting. The red points are diphoton MC with DSCB signal injected at 125 GeV, the blue points are the same MC with a 10 GeV window removed centered at 125 GeV, the black points are the GP fit to the blue points. The bottom panel shows the ratio of the red data points over the GP fit.

a narrow resonance of ∼ 8 GeV for the signal and a broad, smoothly falling background. As long as the length scale is set longer than 8 GeV, we should not need to worry about the signal being absorbed in the fit. This is an example of one large advantage of using GPs. Something as simple as bounding the length scale can encode quite a bit of relevant information about the types of signals you expect to see. The procedure that we settled on in the end was to optimize the hyperparameters using the method described in Section 6.1.2 where the input used is background-only Monte Carlo. Since there is no signal in this sample, the length scale will tend toward longer values associated with the broad, smooth background. These hyperparameters can then be used in the kernel in order to perform a GP fit directly on data.

As one of the biggest problems with the spurious signal test is Monte Carlo statistics; we would like to check that this procedure of obtaining hyperparameters from MC is robust against statistics. Figure 6.10 shows distributions of the length scale obtained from optimizing on 1000 toy data sets with different numbers of events. Figure 6.10a corresponds to 1 million events per toy and figure 6.10b to 10 million events per toy. The vertical red line at length scale∼ 72 GeV corresponds to the value of the length scale obtained from the optimization on the full statistics MC sample of ∼ 100 million events. As can be seen, with lower statistics the length scale tends to optimize to a lower value. This is due to the larger fluctuations. Even with ∼ 1% of the statistics we are still able to get reasonable numbers for the length scale and we should not have to worry about signal being absorbed into the background description.



(a) (b)

Figure 6.10: The length scale distribution found through LML optimization for 1000 toys. 6.10a has 1 million events per toy, 6.10b has 10 million events per toy. The red line at $l \sim 72$ GeV is the value obtained when optimized on the full MC sample with ∼ 100 million events.

6.2.1 RooGP Software Development

Every analysis on ATLAS needs a statistical framework to calculate results of measurements. There are many frameworks in place, but they are all based upon the RooFit and RooStats packages [68][69]. In order be able to use Gaussian Process in the context of searches on ATLAS, a toolkit must be built that is able to interact with these frameworks. The RooFit package uses models based on classes that represent probability distribution functions (PDFs) and provides a language for constructing, composing, and minimizing parameters of them. How this is normally done is that a PDF for the signal portion and background portions are constructed separately and then summed into a full-model PDF. In the case of a search, one can then fit this PDF to the data and extract the normalization for the signal component of the PDF (we will call this nSig). Because with a GP the prediction needs to be completely refit for every value of the signal normalization that is taken, we need to treat these PDFs slightly differently than the normal RooFit treatment.

A large portion of the work I have contributed to implementing GPs in the low-mass search has been in the development of the `RooGP` tool that allows RooFit to use GPs in its fitting procedure. This tool is largely designed to be an interface between the RooFit package and the Gaussian Process Regression library in the scikit-learn python package [70] using the Python language. Because of the way that RooFit performs minimization and the fact that we need to completely refit the GP at every step, the best way to perform a signal plus background (S+B) fit is to directly incorporate the signal model into RooGP. The fit procedure that the package follows is this:

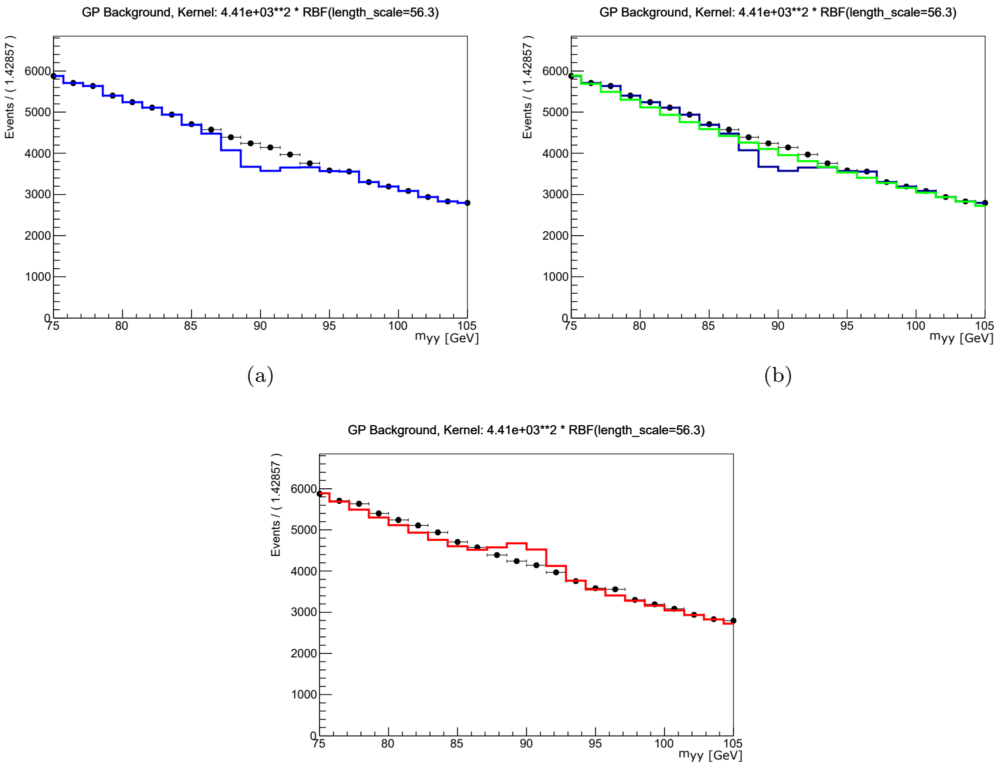
- For each step in the minimization process (tested value of nSig):
  - Construct signal function containing nSig events (a DSCB in the low-mass diphoton case)
  - Subtract this signal from the data distribution.
  - Fit the GP on this signal subtracted data.
  - Add the signal function for nSig back into the result of the GP fit (modified GP).
  - Calculate the negative log likelihood ($-log(\mathcal{L})$) of the modified GP given the data.
- Repeat until a minimum of the $-log(\mathcal{L})$ is found for nSig. This corresponds to the best fit value of the number of signal events.

Figure 6.11 illustrates this procedure on a background only MC sample. For these plots only a single step of the minimization process is shown. The nSig value chosen is 2000 events which is a number far from where we would expect the result because this is a background only sample (we expect close to 0). This choice is simply to make the process easy to see. Figure 6.11c shows the result of the process. You can see that when a value for nSig that is far from 0 is chosen, the fit does not agree well with the data. The minimizer will then move to the next step in the process, moving down the gradient in $-log(\mathcal{L})$. If we imagine the case where the signal portion of the fit is set to have 0 events, (its normalization is 0), then we subtract nothing in the first step, add nothing in the last step and it amounts to fitting the MC sample.

The RooGP class works through a Python class which inherits from `RooFit::RooAbsPdf`. The base class needs only to have overloaded a constructor responsible for setting all parameters, and a function `RooAbsPdf::evaluate` which calculates the value of the PDF given values of its parameters. The `RooGP` class operates on histograms and handles all selection of hyperparameters internally. Training and fit histograms can be supplied separately in order to tune hyperparameters on a background only sample. Currently the signal model used internally by the library only supports a double-sided Crystal Ball function (DSCB) (see section 5.4) as mainly this is being developed for low-mass diphoton. Future improvements can add support for arbitrary function or the use of template histograms.

6.2.2 Tests with Gaussian Process

The `RooGP` class also supports use of a prior mean. This can be passed to the class in the form of a `ROOT::TF1` or a histogram. In all of the previous examples, the mean was taken to be 0. In general the GP is very robust against this choice. Taking the prior mean to be 0 means that the prior probability distribution has a central value corresponding to a line at 0

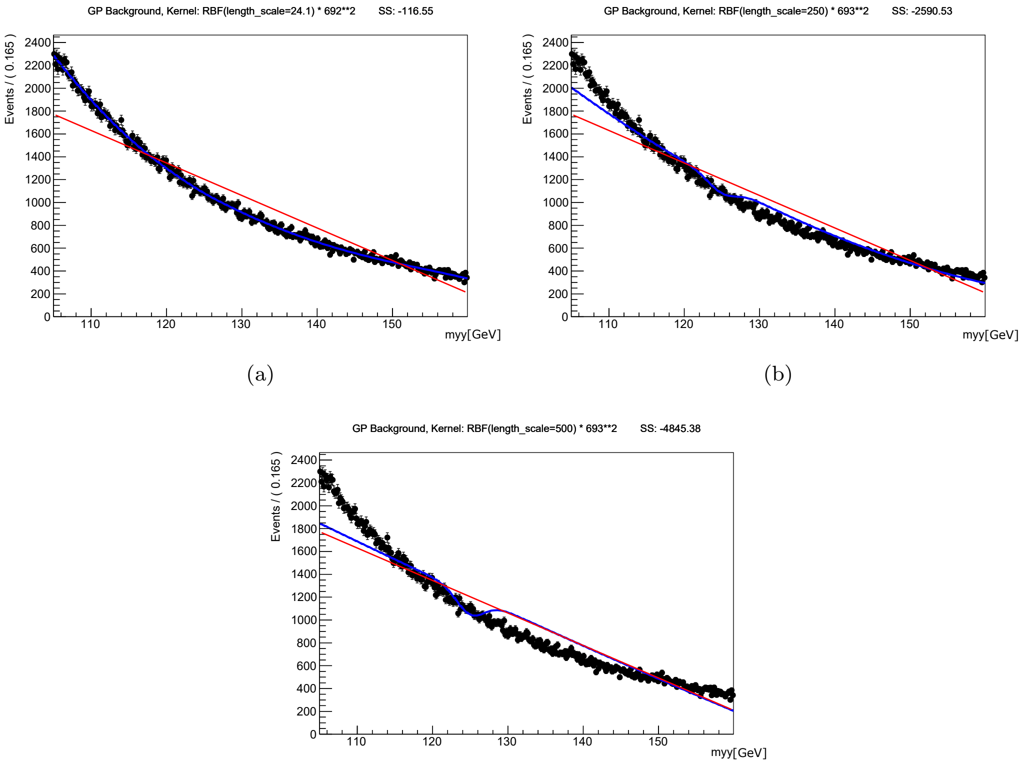


(a) (b)

(c)

Figure 6.11: Illustration of the procedure for fitting nSig with RooGP. These show one step of the minimization process with $nSig = 2000$. Black points are background only MC. (a) The blue line is the MC with the DSCB signal function subtracted. (b) The green line is the GP fit to the blue line. (c) The red line is the GP fit with the DSCB added back in. The red line and black points do not agree well indicating that the point $nSig = 2000$ is far from the minimum of $-log(\mathcal{L})$. Here m_{yy} is the diphoton invanriant mass.

on the fit axes, or equivalently a MVG with mean 0. When the GP is conditioned on data, this line is pulled away from 0 to the posterior prediction. In some cases the GP will perform better if the prior distribution is closer to the data distribution that will be fit. An interesting feature of the prior mean in the GP is that with increasing length scale, the prediction of the GP approaches the prior mean. In equation 6.7 and 6.3 we can see that as l gets larger, the components of the covariance matrix get smaller and the term depending on the kernel gets smaller leaving just the prior mean. Figure 6.12 shows this effect by using a first order polynomial as the prior mean. The three plots correspond to an length scale optimized on the data, a length scale longer than that, and one that is much longer. You can see that the prediction tends toward the prior mean. The downward bump in the fit is due to the signal portion of the fit.



(a)

(b)

(c)

Figure 6.12: The GP prediction plus a DSCB function shown in blue using a first order polynomial as the prior mean shown in red. The black points are the data being fit. (a) is the fit with the length scale optimized on the data. (b) has the length scale fixed to 250. (c) has the length scale fixed to 500. Here myy is the diphoton invariant mass.

Figure 6.13 shows just how flexible the Gaussian Process is for fitting complicated shapes. This is the full-range low-mass distribution showing the complicated trigger turn-on curve. In the low-mass analysis the fit is started at 60 GeV due to the difficulty in describing the turn-on shape. As you can see, the GP can easily fit complicated shapes such as this with

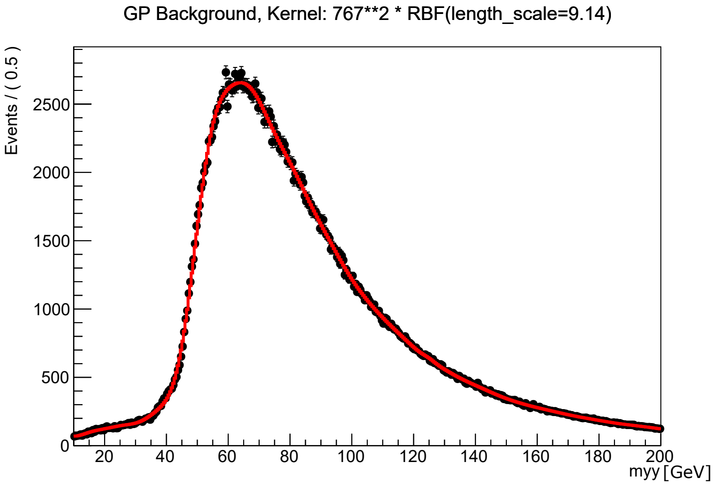


Figure 6.13: A fit of the GP is shown with the red line, to the full low mass invariant mass spectrum shown with black points.

almost no prior assumptions. This fit is done by taking the prior mean as 0 and allowing the hyperparameters to be optimized on the data.

A logical comparison to make when considering this property of the prior mean is to compare it to the results of the functional form used in the spurious signal test from Chapter 5.6.3. For the **UU** category, the functional form selected was a Landau + exponential. If we just use this function as the prior mean, the GP will effectively model the residuals of the data around this function, thereby improving the prediction. If we increase the length scale used in the covariance matrix, we should see the GP approach the Landau + exponential prior. Because with the functional form we test the spurious signal, that will be a good comparison here. Figure 6.14 shows the spurious signal at 90 GeV as a function of length scale of a background-only sample fit with a GP using a prior mean function. The dashed line corresponds to the value of the spurious signal obtained using a Landau+Exp functional form fit. The sample used is the MC background template used for the low-mass analysis in the **UU** category. You can clearly see that as the length scale increases the spurious signal value approaches the

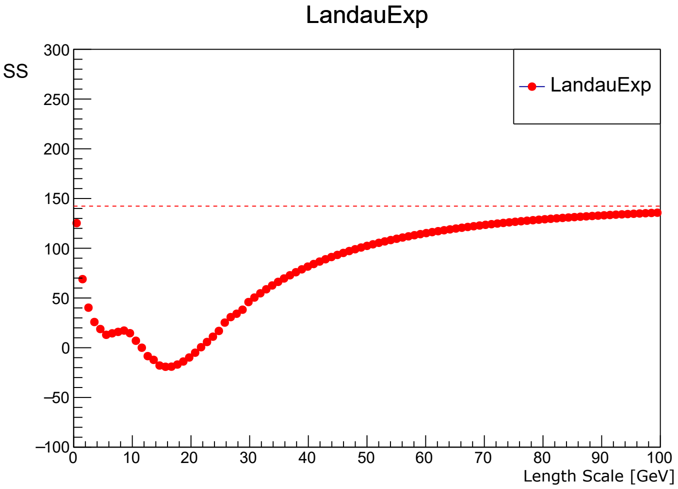
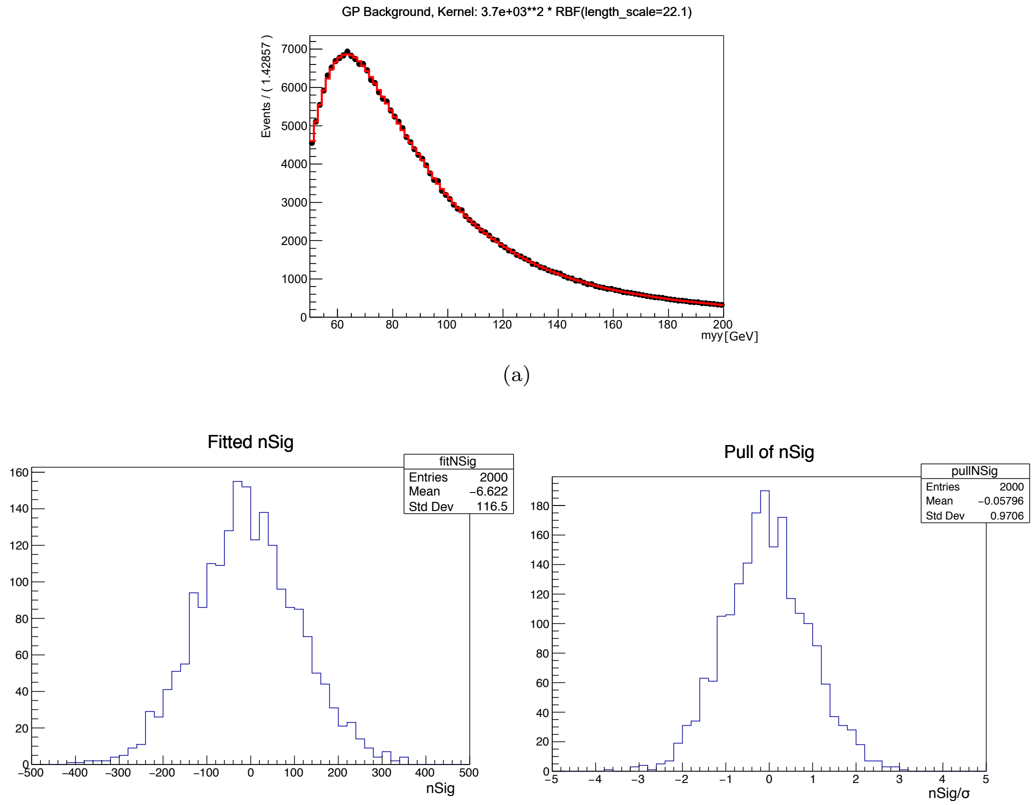


Figure 6.14: Spurious signal (SS) of the GP prediction on background only MC as a function of the length scale with a signal mass at 90 GeV. The red points correspond to a GP with a Landau+exp prior mean with the red dashed line showing the spurious signal value for the Landau+exp functional form.

result obtained with the functional form. This means that as long as our length scale is long enough to not absorb signal and not so long as to be near this asymptote, the GP outperforms a functional form fit when measured with a spurious signal.

When choosing a functional form to describe the background distribution, we know that the function does not represent the underlying function that is generating the data. The function will have some bias in it. The purpose of the spurious signal test is to try to get an estimate on the size of the bias. GPs are much more flexible than a functional form and should be much less biased. We can test the bias by generating toy distributions and evaluating the distribution of spurious signal calculated for each. 2000 toys are generated and the spurious signal calculated for each in a background MC sample with a GP fit from 50 GeV to 200 GeV. The spurious signal is extracted at 125 GeV.

When there is no signal present in the data, we know that a GP can fit the shape well. Using the RooGP class we would like to test whether or not it can accurately extract the signal



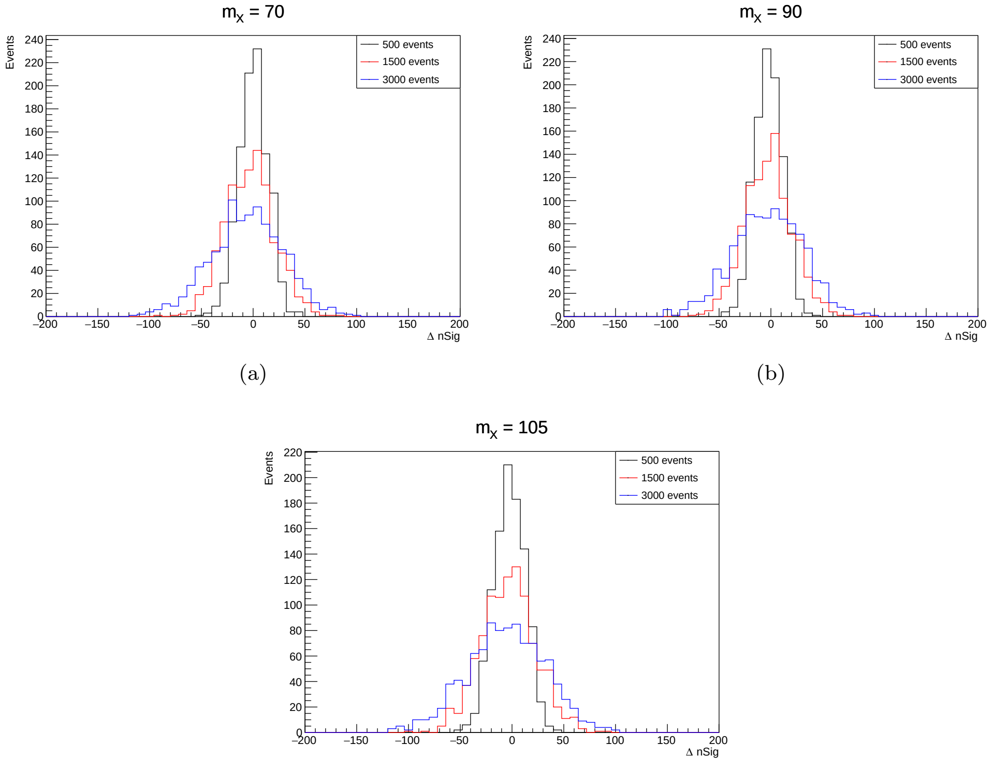
(a)

(b)

(c)

Figure 6.15: The bias of the GP is tested by fitting the GP on 2000 toys generated from background MC in the range 50-200 GeV (a). The spurious signal is extracted for each toy at 125 GeV and is shown in figure (b). (c) shows the pull distribution of the spurious signal. The bias in the spurious signal for the GP is $\sim 6\%$ of the statistical uncertainty and the width of the pull distribution is ~ 1.

portion of a fit if there were signal. We can do this with signal injection tests. For this I take the background templates used to model the background in the low-mass diphoton analysis. On top of these templates I inject a signal by randomly generating events according to a DSCB distribution with parameterization described in section 5.4. I then use the RooGP class and fit the signal + background GP to it and extract the signal part of the fit. Figure 6.16 shows the results of 1000 signal toys where the number of injected signal events is subtracted from the number of extracted signal events for three different signal sizes of 500 events, 1500 events, and 3000 events. The three plots in figure 6.16 correspond to different mass points that the



(a)

(b)

(c)

Figure 6.16: Signal is injected to the **UU** background template used in low-mass diphoton. $\Delta nSig$ here is the number of signal events extracted by the GP minus the injected number of signal events. The black, red and blue lines are the results for 500, 1500, and 3000 injected events, respectively. (a) Corresponds to a signal injected at 70 GeV, (b) at 90 GeV, and (c) at 105 GeV.

signal is injected at, 70 GeV, 90 GeV, and 105 GeV. This is to test the signal extraction across the background range. These distributions all have mean near 0 which is what we expect if the GP is able to extract close to the correct number of signal events. The RMS for these test are approximately 25 for the case with 500 events and approximately 50 for the case with 3000 events which correspond to $\sim 5\%$ and $\sim 1\%$ error respectively. This shows that the GP is not significantly biased when fitting a signal + background distribution.

6.3 Conclusions

Gaussian Process is a method for regression that has found wide use in the data science and machine learning industries. It is now beginning to see use in the ATLAS community as some simpler methods of modeling are no longer accurate enough, or computationally feasible. GP's ability to take very broad assumptions about the data that is to be modeled while still remaining flexible enough to be unbiased make it a useful tool to have for various tasks. I have begun work on implementing a GP based package that can help to solve the problems presented by the spurious signal technique currently used on the low-mass diphoton analysis. It is now impossible for the previous methods to be used on this analysis and I believe GP presents a way to move forward.

There are still several studies and refinements to be done to fully validate and prepare the GP for use in the low-mass diphoton analysis. Investigation into different kernels could be a particularly useful study. Although the GP should be a very unbiased estimation method, the use of a different kernel, one that is possibly more motivated by the actual underlying physics or detector response, could be less likely to be biased or, at the very least, can have its motivation tied directly to the physics making its use easier to justify.

Checking the bias of the GP fitting on the background distribution in the way that I have done it is very time intensive. It requires many toy MC samples to be created and the GP fit to each. I have done this for a few mass points, but in order to fully understand the performance of the method, the bias should be tested at every mass point used in the low-mass search (65-110 GeV). It may be that toys are the best way to do this, but a faster method would be very useful in the long run.

Many analyses on ATLAS make use of statistics packages derived from RooFit and RooStats but do not directly use either of these. In the case of low-mass diphoton, a package called

HistFitter is used. This package is built in C++ and uses input datacards to define the statistical models to be used in the analysis. At the time of writing this, there was only one C++ based Gaussian Process Regression library that I could find and its functionality was quite limited. The python language, on the other hand, has several high quality libraries with GP implementations. This language mismatch represents a technical hurdle to be overcome in fully implementing the a GP in an ATLAS analysis. Also, at the time of writing this, a group of students in the Diana-HEP group (http://diana-hep.org/) have been writing a python implementation of the HistFactory statistics package called PyHF [71]. This package can be used to fully implement the statistical model used in the low-mass diphoton group

Appendix A

A.1 2x2D Sideband Method

Each photon candidate is classified as belonging to a category TI, $T\bar{I}$, $\bar{T}I$ or $\bar{T}\bar{I}$ depending on whether it fails or passes the identification and isolation criteria. A pair of candidates therefore can belong to one out of 16 categories, labeled from $TITI$ to $\bar{T}\bar{I}\bar{T}\bar{I}$, based on the categories of both photon candidates. Neglecting the photon-photon and photon-jet correlations, the number of events in each category can be written as:

$$\begin{aligned} N_{TITI} &= W_{\gamma\gamma}^{L'L'}\epsilon_{I1}\epsilon_{T1}\epsilon_{I2}\epsilon_{T2} \\ &+ W_{\gamma j}^{L'L'}\epsilon_{I1}\epsilon_{T1}f_{I2}f_{T2} \\ &+ W_{j\gamma}^{L'L'}\epsilon_{I2}\epsilon_{T2}f_{I1}f_{T1} \\ &+ W_{jj}^{L'L'}f'_{I1}f'_{T1}f'_{I2}f'_{T2}\xi_{Ijj} \end{aligned} \tag{A.1}$$

$$\begin{aligned} N_{TIT\bar{I}} &= W_{\gamma\gamma}^{L'L'}\epsilon_{I1}\epsilon_{T1}\left(1-\epsilon_{I2}\right)\epsilon_{T2} \\ &+ W_{\gamma j}^{L'L'}\epsilon_{I1}\epsilon_{T1}\left(1-f_{I2}\right)f_{T2} \\ &+ W_{j\gamma}^{L'L'}f_{I1}f_{T1}\left(1-\epsilon_{I2}\right)\epsilon_{T2} \\ &+ W_{jj}^{L'L'}f'_{I1}f'_{T1}\left(1-f'_{I2}\xi_{Ijj}\right)f'_{T2} \end{aligned} \tag{A.2}$$

$$\begin{aligned} N_{T\bar{I}TI} &= W^{L'L'}_{\gamma\gamma}\left(1-\epsilon_{I1}\right)\epsilon_{T1}\epsilon_{I2}\epsilon_{T2} \\ &+ W^{L'L'}_{\gamma j}\left(1-\epsilon_{I1}\right)\epsilon_{T1}f_{I2}f_{T2} \\ &+ W^{L'L'}_{j\gamma}\left(1-f_{I1}\right)f_{T1}\epsilon_{I2}\epsilon_{T2} \\ &+ W^{L'L'}_{jj}\left(1-f'_{I1}\xi_{Ijj}\right)f'_{T1}f'_{I2}f'_{T2} \end{aligned} \tag{A.3}$$

$$\begin{aligned} N_{T\bar{I}T\bar{I}} &= W^{L'L'}_{\gamma\gamma}\left(1-\epsilon_{I1}\right)\left(1-\epsilon_{I2}\right)\epsilon_{T1}\epsilon_{T2} \\ &+ W^{L'L'}_{\gamma j}\left(1-\epsilon_{I1}\right)\left(1f_{I2}\right)\epsilon_{T1}f_{T2} \\ &+ W^{L'L'}_{j\gamma}\left(1-\epsilon_{I2}\right)\left(1-f_{I1}\right)\epsilon_{T2}f_{T1} \\ &+ W^{L'L'}_{jj}\left(1-f'_{I1}-f'_{I2}+f'_{I1}f'_{I2}\xi_{Ijj}\right)f'_{T1}f'_{T2} \end{aligned} \tag{A.4}$$

$$\begin{aligned} N_{TI\bar{T}I} &= W^{L'L'}_{\gamma\gamma}\epsilon_{I1}\epsilon_{T1}\epsilon_{I2}\left(1-\epsilon_{T2}\right) \\ &+ W^{L'L'}_{\gamma j}\epsilon_{I1}\epsilon_{T1}f_{I2}\left(1-f_{T2}\right) \\ &+ W^{L'L'}_{j\gamma}f_{I1}f_{T1}\epsilon_{I2}\left(1-\epsilon_{T2}\right) \\ &+ W^{L'L'}_{jj}f'_{I1}f'_{T1}f'_{I2}\left(1-f'_{T2}\right)\xi_{Ijj} \end{aligned} \tag{A.5}$$

$$\begin{aligned} N_{\bar{T}ITI} &= W^{L'L'}_{\gamma\gamma}\epsilon_{I1}\left(1-\epsilon_{T1}\right)\epsilon_{I2}\epsilon_{T2} \\ &+ W^{L'L'}_{\gamma j}\epsilon_{I1}\left(1-\epsilon_{T1}\right)f_{I2}f_{T2} \\ &+ W^{L'L'}_{j\gamma}f_{I1}\left(1-f_{T1}\right)\epsilon_{I2}\epsilon_{T2} \\ &+ W^{L'L'}_{jj}f'_{I1}\left(1-f'_{T1}\right)f'_{I2}f'_{T2}\xi_{Ijj} \end{aligned} \tag{A.6}$$

$$
\begin{aligned}
N_{TI\bar{T}\bar{I}} &= W_{\gamma\gamma}^{L'L'}\epsilon_{I1}\epsilon_{T1}\left(1-\epsilon_{I2}\right)\left(1-\epsilon_{T2}\right) \\
&+ W_{\gamma j}^{L'L'}\epsilon_{I1}\epsilon_{T1}\left(1-f_{I2}-f_{T2}+f_{I2}f_{T2}\right) \\
&+ W_{j\gamma}^{L'L'}f_{I1}f_{T1}\left(1-\epsilon_{I2}\right)\left(1-\epsilon_{T2}\right) \\
&+ W_{jj}^{L'L'}f'_{I1}f'_{T1}\left(1-\xi_{Ijj}f'_{I2}-f'_{T2}+f'_{I2}f'_{T2}\xi_{Ijj}\right) \qquad \text{(A.7)}
\end{aligned}
$$

$$
\begin{aligned}
N_{\bar{T}\bar{I}TI} &= W_{\gamma\gamma}^{L'L'}\left(1-\epsilon_{I1}\right)\left(1-\epsilon_{T1}\right)\epsilon_{I2}\epsilon_{T2} \\
&+ W_{\gamma j}^{L'L'}\left(1-\epsilon_{I1}\right)\left(1-\epsilon_{T1}\right)f_{I2}f_{T2} \\
&+ W_{j\gamma}^{L'L'}\left(1-f_{I1}-f_{T1}+f_{I1}f_{T1}\right)\epsilon_{I2}\epsilon_{T2} \\
&+ W_{jj}^{L'L'}\left(1-f'_{I1}\xi_{Ijj}-f'_{T1}+f'_{I1}f'_{T1}\xi_{Ijj}\right)f'_{I2}f'_{T2} \qquad \text{(A.8)}
\end{aligned}
$$

$$
\begin{aligned}
N_{T\bar{I}\bar{T}I} &= W_{\gamma\gamma}^{L'L'}\left(1-\epsilon_{I1}\right)\epsilon_{T1}\epsilon_{I2}\left(1-\epsilon_{T2}\right) \\
&+ W_{\gamma j}^{L'L'}\left(1-\epsilon_{I1}\right)\epsilon_{T1}f_{I2}\left(1-f_{T2}\right) \\
&+ W_{j\gamma}^{L'L'}\left(1-f_{I1}\right)f_{T1}\epsilon_{I2}\left(1-\epsilon_{T2}\right) \\
&+ W_{jj}^{L'L'}\left(1-f'_{I1}\xi_{Ijj}\right)f'_{T1}f'_{I2}\left(1-f'_{T2}\right) \qquad \text{(A.9)}
\end{aligned}
$$

$$
\begin{aligned}
N_{\bar{T}IT\bar{I}} &= W_{\gamma\gamma}^{L'L'}\epsilon_{I1}\left(1-\epsilon_{T1}\right)\left(1-\epsilon_{I2}\right)\epsilon_{T2} \\
&+ W_{\gamma j}^{L'L'}\epsilon_{I1}\left(1-f_{I2}\right)\left(1-\epsilon_{T1}\right)f_{T2} \\
&+ W_{j\gamma}^{L'L'}f_{I1}\left(1-\epsilon_{I2}\right)\left(1-f_{T1}\right)\epsilon_{T2} \\
&+ W_{jj}^{L'L'}f'_{I1}\left(1-f'_{T1}\right)\left(1-f'_{I2}\xi_{Ijj}\right)f'_{T2} \qquad \text{(A.10)}
\end{aligned}
$$

$$\begin{aligned} N_{\bar{T}I\bar{T}I} &= W_{\gamma\gamma}^{L'L'}\epsilon_{I1}\left(1-\epsilon_{T1}\right)\epsilon_{I2}\left(1-\epsilon_{T2}\right) \\ &+ W_{\gamma j}^{L'L'}\epsilon_{I1}\left(1-\epsilon_{T1}\right)f_{I2}\left(1-f_{T2}\right) \\ &+ W_{j\gamma}^{L'L'}f_{I1}\left(1-f_{T1}\right)\epsilon_{I2}\left(1-\epsilon_{T2}\right) \\ &+ W_{jj}^{L'L'}f'_{I1}f'_{I2}\left(1-f'_{T1}-f'_{T2}+f'_{T1}f'_{T2}\right)\xi_{Ijj} \end{aligned} \tag{A.11}$$

$$\begin{aligned} N_{T\bar{I}\bar{T}\bar{I}} &= W_{\gamma\gamma}^{L'L'}\left(1-\epsilon_{I1}\right)\epsilon_{T1}\left(1-\epsilon_{I2}\right)\left(1-\epsilon_{T2}\right) \\ &+ W_{\gamma j}^{L'L'}((1-\epsilon_{I1})\,\epsilon_{T1}\left(1-f_{I2}-f_{T2}+f_{I2}f_{T2}\right) \\ &+ W_{j\gamma}^{L'L'}\left(1-f_{I1}\right)f_{T1}\left(1-\epsilon_{T2}\right)\left(1-\epsilon_{I2}\right) \\ &+ W_{jj}^{L'L'}f'_{T1}\left(1-f'_{I2}-f'_{T2}-f'_{I1}+f'_{I2}f'_{T2}+f'_{T2}f'_{I1}+f'_{I2}f'_{I1}\xi_{Ijj}-f'_{I1}f'_{I2}f'_{T2}\xi_{Ijj}\right) \end{aligned} \tag{A.12}$$

$$\begin{aligned} N_{\bar{T}\bar{I}T\bar{I}} &= W_{\gamma\gamma}^{L'L'}\left(1-\epsilon_{I1}\right)\left(1-\epsilon_{T1}\right)\left(1-\epsilon_{I2}\right)\epsilon_{T2} \\ &+ W_{\gamma j}^{L'L'}\left(1-\epsilon_{I1}\right)\left(1-\epsilon_{T1}\right)\left(1-f_{I2}\right)f_{T2} \\ &+ W_{j\gamma}^{L'L'}\left(1-f_{I1}-f_{T1}+f_{I1}f_{T1}\right)\left(1-\epsilon_{I2}\right)\epsilon_{T2} \\ &+ W_{jj}^{L'L'}\left(1-f'_{I1}-f'_{T1}-f'_{I2}+f'_{I1}f'_{T1}+f'_{T1}f'_{I2}+f'_{I1}f'_{I2}\xi_{Ijj}-f'_{I1}f'_{T1}f'_{I2}\xi_{Ijj}\right)f'_{T2} \end{aligned} \tag{A.13}$$

$$\begin{aligned} N_{\bar{T}I\bar{T}\bar{I}} &= W_{\gamma\gamma}^{L'L'}\epsilon_{I1}\left(1-\epsilon_{T1}\right)\left(1-\epsilon_{I2}\right)\left(1-\epsilon_{T2}\right) \\ &+ W_{\gamma j}^{L'L'}\epsilon_{I1}\left(1-\epsilon_{T1}\right)\left(1-f_{I2}-f_{T2}+f_{I2}f_{T2}\right) \\ &+ W_{j\gamma}^{L'L'}f_{I1}\left(1-f_{T1}\right)\left(1-\epsilon_{I2}\right)\left(1-\epsilon_{T2}\right) \\ &+ W_{jj}^{L'L'}f'_{I1}\left(\left(1-f'_{T1}\right)\left(1-f'_{I2}\xi_{Ijj}\right)-f'_{T2}\left(1-f'_{T1}\right)\left(1-f'_{I2}\xi_{Ijj}\right)\right) \end{aligned} \tag{A.14}$$

$$\begin{aligned}
N_{\bar{T}\bar{I}\bar{T}I} &= W_{\gamma\gamma}^{L'L'}\epsilon_{I2}\left((1-\epsilon_{T2})\,(1-\epsilon_{I1})-\epsilon_{T1}\,(1-\epsilon_{T2})\,(1-\epsilon_{I1})\right) \\
&+ W_{\gamma j}^{L'L'}f_{I2}\left((1-\epsilon_{I1})\,(1-f_{T2})-\epsilon_{T1}\,(1-\epsilon_{I1})\,(1-f_{T2})\right) \\
&+ W_{j\gamma}^{L'L'}\epsilon_{I2}\left((1-\epsilon_{T2})\,(1-f_{I1})-f_{T1}\,(1-f_{I1})\,(1-\epsilon_{T2})\right) \\
&+ W_{jj}^{L'L'}\left((1-f'_{T2})\,(1-f'_{I1}\xi_{Ijj})-f'_{T1}\,(1-f'_{T2})\,(1-f'_{I1}\xi_{Ijj})\right)f'_{I2}
\end{aligned} \tag{A.15}$$

$$\begin{aligned}
N_{\bar{T}\bar{I}\bar{T}\bar{I}} &= W_{\gamma\gamma}^{L'L'}[1-\epsilon_{T1}-\epsilon_{T2}+\epsilon_{T1}\epsilon_{T2}+\epsilon_{I2}(1-\epsilon_{T1}-\epsilon_{T2}+\epsilon_{T1}\epsilon_{T2}) \\
&+ \epsilon_{I1}(-(1-\epsilon_{I2})(1-\epsilon_{T1})+\epsilon_{T2}(1-\epsilon_{I2})(1-\epsilon_{T1}))] \\
&+ W_{\gamma j}^{L'L'}[1-\epsilon_{T1}-f_{T2}+\epsilon_{T1}f_{T2}+f_{I2}(1-\epsilon_{T1}-f_{T2}+\epsilon_{T1}f_{T2}) \\
&+ \epsilon_{I1}((1-\epsilon_{T1})(1-f_{I2})+f_{T2}(1-f_{I2})(1-\epsilon_{T1}))] \\
&+ W_{j\gamma}^{L'L'}[1-\epsilon_{T2}-f_{T1}+\epsilon_{T2}f_{T1}+\epsilon_{I2}(1-f_{T1}-\epsilon_{T2}+\epsilon_{T2}f_{T1}) \\
&+ f_{I1}((1-\epsilon_{I2})(1-f_{T1})+\epsilon_{T2}(1-\epsilon_{I2})(1-f_{T1}))] \\
&+ W_{jj}^{L'L'}[1-f'_{T1}-f'_{T2}+f'_{T1}f'_{T2}+f'_{I2}(1-f'_{T1}-f'_{T2}+f'_{T1}f'_{T2}) \\
&+ f'_{I1}(-(1-f'_{I2}\xi_{Ijj})(1-f'_{T1})+f'_{T2}(1-f'_{I2}\xi_{Ijj})(1-f'_{T1}))]
\end{aligned} \tag{A.16}$$

In those 16 equations we have 19 unknowns, six of which are inputs or are fixed:

- ϵ_{T1} and ϵ_{T2} are the TIGHT identification efficiencies for the leading and subleading photons respectively. Their values are determined from Monte Carlo simulation of the diphoton signal.

- ϵ_{I1} and ϵ_{I2} are the isolation efficiencies for the leading and subleading photons respectively, also determined with Monte-Carlo.

The other 13 unknowns are outputs from the extended 2D sideband method, namely:

- $W_{\gamma\gamma}^{L'L'}$, $W_{\gamma j}^{L'L'}$, $W_{j\gamma}^{L'L'}$ and $W_{jj}^{L'L'}$ are the yields for the four categories of event after the LOOSE' preselection.

- f_{T1} and f_{T2} are the TIGHT identification fake rates for the leading and subleading jets in γ-jet and jet-γ events.

- f'_{T1} and f'_{T2} are the TIGHT identification fake rates for the leading and subleading jets in jet-jet events. Those fake rates are fixed to be equal to f_{T1} and f_{T2} respectively, as the difference between the prime and non-prime identification fake rates are much smaller than their precision.

- f_{I1} and f_{I2} are the isolation fake rates for the leading and subleading jets in γ-jet and jet-γ events.

- f'_{I1} and f'_{I2} are the isolation fake rates for the leading and subleading jets in jet-jet events.

- ξ_{Ijj} is the correlation between the isolation of the jets in the jj component.

List of Tables

List of Figures

Bibliography

[1] N. Arkani-Hamed, S. Dimopoulos, and G. R. Dvali, *The Hierarchy problem and new dimensions at a millimeter*, Phys. Lett. **B**429 (1998) 263–272, arXiv:hep-ph/9803315 [hep-ph]. 1.2, 1.2

[2] L. Randall and R. Sundrum, *A Large mass hierarchy from a small extra dimension*, Phys. Rev. Lett. **8**3 (1999) 3370–3373, arXiv:hep-ph/9905221 [hep-ph]. 1.2

[3] T. Kaluza, *Zum unittsproblem in der physik*, Sitzungsber. Preuss. Akad. Wiss. Berlin. **3**3 (1921) 996–972. 1.2

[4] O. Klein, *Quantentheorie und fünfdimensionale Relativitätstheorie*, Zeitschrift fur Physik **3**7 (Dec., 1926) 895–906. 1.2

[5] A. Collaboration, *Search for scalar diphoton resonances with 15.4 fb*$^{-1}$ *of data collected at* $\sqrt{s}$*=13 TeV in 2015 and 2016 with the ATLAS detector*, Tech. Rep. ATLAS-COM-CONF-2016-056, CERN, Geneva, Jul, 2016. `https://cds.cern.ch/record/2199338`. 1.2

[6] *750 GeV diphoton excess on Wikipedia*, `https://en.wikipedia.org/wiki/750_GeV_diphoton_excess`. Accessed: 2018-08-19. 1.2

[7] J. Beacham, R. R. M. Fletcher, K. Grevtsov, T. Guillemin, S. M. Mazza, B. R. Mellado Garcia, G. D. Carrillo-Montoya, J. Kroll, Y. Okumura, R. Reed, R. Turra, Z. Wang, I. Wingerter-Seez, and A. Vallier, *Cross-checks and diphoton event activity studies: Search for new phenomena in diphoton events with the ATLAS detector at* $\sqrt{s} = 13$ *TeV*, Tech. Rep. ATL-COM-PHYS-2016-759, CERN, Geneva, Jun, 2016. `https://cds.cern.ch/record/2162756`. 1.2

[8] ATLAS Collaboration, M. Aaboud et al., *Search for new phenomena in high-mass diphoton final states using 37 fb*$^{-1}$ *of proton–proton collisions collected at* $\sqrt{s} = 13$ *TeV with the ATLAS detector*, Phys. Lett. **B**775 (2017) 105–125, arXiv:1707.04147 [hep-ex]. 1.4, 1.4, 5.3.2

[9] L. Barak, N. Berger, T. Carli, L. Carminati, M. Delmastro, M. Dyndal, K. Grevtsov, W. Guan, A. Hard, Y. Huang, R. Hyneman, S. Jin, S. Laplace, B. Lenzi, F. Malek,

S. Manzoni, G. Marchiori, S. M. Mazza, T. Meideck, C. Meyer, C. Peng, E. Petit, L. Roos, T. A. Schwarz, F. Siegert, M. Silva Jr., J. Stark, S. Sun, K. Tackmann, G. Unal, M. White, I. Wingerter-Seez, S. L. Wu, H. Yang, Y. C. Yap, C. Zhou, R. Di Nardo, X. Ju, L. Lee, F. Wang, S. Han, G. D. Carrillo-Montoya, R. R. M. Fletcher, J. Kroll, and S. Raspopov, *Search for new phenomena in diphoton events with the ATLAS detector at* $\sqrt{s} = 13$ *TeV : Supporting documentation for the publication based on the 2016 data*, Tech. Rep. ATL-COM-PHYS-2017-148, CERN, Geneva, Feb, 2017. `https://cds.cern.ch/record/2252642`. 1.4

[10] S. Glashow, *Partial symmetries of weak interactions*, Nucl.Phys. **22** (1961) 579–588. 2

[11] V. Gribov and L. Lipatov, *Deep inelastic e-p scattering in perturbation theory*, Sov.J.Nucl.Phys. **15** (1972) 438–450. 2

[12] P. Higgs, *Broken symmetries, massless particles and gauge fields*, Physics Letters **12** (1964) 132–133. 2.1

[13] P. W. Higgs, *Broken symmetries and the masses of gauge bosons*, Phys. Rev. Lett. **13** (1964) 508–509. 2.1

[14] F. Englert and R. Brout, *Broken symmetry and the mass of gauge vector mesons*, Phys. Rev. Lett. **13** (1964) 321–323. 2.1

[15] G. S. Guralnik, C. R. Hagen, and T. W. B. Kibble, *Global conservation laws and massless particles*, Phys. Rev. Lett. **13** (1964) 585–587. 2.1

[16] H. D. Politzer, *Reliable perturbative results for strong interactions?*, Phys. Rev. Lett. **30** (1973) 1346–1349. 2.1

[17] D. J. Gross and F. Wilczek, *Asymptotically free gauge theories. I*, Phys. Rev. D **8** (1973) 3633–3652. 2.1

[18] D. J. Gross and F. Wilczek, *Asymptotically free gauge theories. II*, Phys. Rev. D **9** (1974) 980–993. 2.1

[19] CERN, *The Standard Model*, `http://united-states.cern/physics/standard-model-and-beyond`. [Online; accessed September, 2018]. 2.1, A.1

[20] A. D. Sakharov, *Violation of CP Invariance, C Asymmetry, and Baryon Asymmetry of the Universe*, Soviet Journal of Experimental and Theoretical Physics Letters **5** (Jan., 1967) 24. 2.2.1

[21] J. Shelton and K. M. Zurek, *Darkogenesis: A baryon asymmetry from the dark matter sector*, Phys. Rev. **D**82 (2010) 123512, arXiv:1008.1997 [hep-ph]. 2.2.1

[22] E. Corbelli and P. Salucci, *The Extended Rotation Curve and the Dark Matter Halo of M33*, Mon. Not. Roy. Astron. Soc. **3**11 (2000) 441–447, arXiv:astro-ph/9909252 [astro-ph]. 2.2.2

[23] R. Massey, T. Kitching, and J. Richard, *The dark matter of gravitational lensing*, Rept. Prog. Phys. **7**3 (2010) 086901, arXiv:1001.1739 [astro-ph.CO]. 2.2.2

[24] Deluca, Stefania, *Extended Rotation Curve of M33*, 2014. `https://commons.wikimedia.org/wiki/File:M33_rotation_curve_HI.gif`. [Online; accessed July, 2018]. 2.2, A.1

[25] S. P. Martin, *A Supersymmetry primer*, arXiv:hep-ph/9709356 [hep-ph]. [Adv. Ser. Direct. High Energy Phys.18,1(1998)]. 2.2.3, 2.3, A.1

[26] C. Munoz, *Models of Supersymmetry for Dark Matter*, EPJ Web Conf. **1**36 (2017) 01002, arXiv:1701.05259 [hep-ph]. 2.2.3

[27] L. Evans and P. Bryant, *LHC Machine*, Journal of Instrumentation **3** (2008) no. 08, S08001. `http://stacks.iop.org/1748-0221/3/i=08/a=S08001`. 3

[28] LHC Study Group Collaboration, T. S. Pettersson and P. Lefvre, *The Large Hadron Collider: conceptual design*, Tech. Rep. CERN-AC-95-05-LHC, Oct, 1995. `https://cds.cern.ch/record/291782`. 3

[29] ATLAS Collaboration, *ATLAS detector and physics performance: Technical Design Report*. CERN, Geneva, 1999. 3

[30] *Atlas Computing: technical design report*. CERN, Geneva, 2005. 3

[31] ATLAS Collaboration, *The ATLAS experiment at the CERN Large Hadron Collider*, JINST **3** (2008) S08003. 3

[32] ATLAS Collaboration, *Expected performance of the ATLAS experiment: detector, trigger and physics*. CERN, Geneva, 2008. arXiv:0901.0512. 3

[33] ATLAS, *Peak Luminosity per Fill in 2017*, `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2`. [Online; accessed September, 2018]. 3.3, A.1

[34] ATLAS, *Delivered Luminosity versus time for 2011-2018 (p-p data only)*, `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2`. [Online; accessed September, 2018]. 3.4, A.1

[35] ATLAS Collaboration Collaboration, *ATLAS liquid-argon calorimeter: Technical Design Report*. Technical Design Report ATLAS. CERN, Geneva, 1996. `https://cds.cern.ch/record/331061`. 3.2.3

[36] ATLAS, *Liquid Argon Calorimeter*, `https://twiki.cern.ch/twiki/bin/view/AtlasPublic/LuminosityPublicResultsRun2`. [Online; accessed September, 2018]. 3.8, A.1

[37] J. Pequenao, *Computer Generated image of the ATLAS calorimeter*, Mar, 2008. 3.9, A.1

[38] for the ATLAS collaboration Collaboration, A. M. Henriques Correia, *The ATLAS Tile Calorimeter*, Tech. Rep. ATL-TILECAL-PROC-2015-002, CERN, Geneva, Mar, 2015. `https://cds.cern.ch/record/2004868`. 3.10, A.1

[39] K. Ntekas, T. Alexopoulos, and S. Zimmermann, *The ATLAS New Small Wheel Upgrade Project*, Tech. Rep. ATL-MUON-PROC-2014-004, CERN, Geneva, Jun, 2014. `http://cds.cern.ch/record/1711717`. 3.11, A.1

[40] A. Collaboration, *Electron Reconstruction and Identification in the ATLAS Experiment using the 2015 and 2016 LHC proton-proton collision data*, Tech. Rep. ATL-COM-PHYS-2018-436, CERN, Geneva, Apr, 2018. `https://cds.cern.ch/record/2315291`. 4

[41] C. Anastopoulos, L. Aperio Bella, K. Brendlinger, E. Benhar Noccioli, C. Diaconu, V. Ellajosyula, L. M. Flores, R. R. M. Fletcher, R. Gugel, S. Heim, F. Hubaut, S. J. Kahn, J. I. Kroll, E. Monnier, J. Reichert, K. Tackmann, S. J. Thais, E. Thomson, C. Wang, and R. Zhang, *Support Note for Electron ID: electron reconstruction and identification*, Tech. Rep. ATL-COM-PHYS-2017-1012, CERN, Geneva, Jul, 2017. `https://cds.cern.ch/record/2273328`. 4, 4.1, A.1

[42] ATLAS Collaboration Collaboration, A. Collaboration, *Electron efficiency measurements with the ATLAS detector using 2012 LHC proton-proton collision data. Electron efficiency measurements with the ATLAS detector using 2012 LHC proton-proton collision data*, Eur. Phys. J. C **7**7 (Dec, 2016) 195. 64 p. `https://cds.cern.ch/record/2237544`. 65 pages in total, author list starting page 49, 24 figures, 3 tables. Published in EPJC. All figures including auxiliary figures are available at http://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/PAPERS/PERF-2016-01. 4.2.4

[43] R. R. M. G. Fletcher, A. Hadef, Y. Kano, C. Meyer, I. Nomidis, E. Petit, S. Raspopov, R. Turra, A. R. L. Vallier, Y. Wang, L. Roos, Y. Liu, N. Berger, R. Lafaye, R. C. Edgar, D. E. Amidei, and J. Kroll, *Search for new phenomena in low-mass diphoton events with the ATLAS detector at* $\sqrt{s} = 13\ TeV$, Tech. Rep. ATL-COM-PHYS-2018-193, CERN, Geneva, Mar, 2018. `https://cds.cern.ch/record/2306589`. 5, 5.2.1, 5.1, 5.4, 5.39, A.1, A.1

[44] ATLAS Collaboration, G. Aad et al., *Search for high-mass diphoton resonances in pp collisions at* $\sqrt{s} = 8\ TeV$ *with the ATLAS detector*, Phys. Rev. **D**92 (2015) no. 3, 032004, arXiv:1504.05511 [hep-ex]. 5

[45] A. Collaboration, *Search for resonances in low-mass diphoton final states using 80* fb^{-1} *of pp collisions collected at* $\sqrt{s} = 13\ TeV$ *with the ATLAS detector*, Tech. Rep. ATL-COM-PHYS-2018-504, CERN, Geneva, May, 2018. `https://cds.cern.ch/record/2316944`. 5

[46] T. D. Lee, *A Theory of Spontaneous* T *Violation*, Phys. Rev. D **8** (Aug, 1973) 1226–1239. `https://link.aps.org/doi/10.1103/PhysRevD.8.1226`. 5.1

[47] G. C. Branco, P. M. Ferreira, L. Lavoura, M. N. Rebelo, M. Sher, and J. P. Silva, *Theory and phenomenology of two-Higgs-doublet models*, Phys. Rept. **5**16 (2012) 1–102, arXiv:1106.0034 [hep-ph]. 5.1, 5.1

[48] G. Bhattacharyya and D. Das, *Scalar sector of two-Higgs-doublet models: A minireview*, Pramana **8**7 (2016) no. 3, 40, arXiv:1507.06424 [hep-ph]. 5.1

[49] ATLAS Collaboration, G. Aad et al., *Search for flavour-changing neutral current top quark decays* $t \to Hq$ *in pp collisions at* $\sqrt{s} = 8\ TeV$ *with the ATLAS detector*, JHEP **1**2 (2015) 061, arXiv:1509.06047 [hep-ex]. 5.1

[50] J. Alwall et al., *The automated computation of tree-level and next-to-leading order differential cross sections, and their matching to parton shower simulations*, JHEP **07** (2014) 079, arXiv:1405.0301 [hep-ph]. 5.2.2

[51] P. Artoisenet et al., *A framework for Higgs characterisation*, JHEP **11** (2013) 043, arXiv:1306.6464 [hep-ph]. 5.2.2

[52] R. D. Ball et al., *Parton distributions for the LHC Run II*, JHEP **04** (2015) 040, arXiv:1410.8849 [hep-ph]. 5.2.2

[53] T. Sjöstrand, S. Mrenna, and P. Z. Skands, *A brief introduction to PYTHIA 8.1*, Comput. Phys. Commun. **178** (2008) 852, arXiv:0710.3820 [hep-ph]. 5.2.2

[54] ATLAS Collaboration, *Summary of ATLAS Pythia 8 tunes*, ATL-PHYS-PUB-2012-003. `https://cds.cern.ch/record/1474107`. 5.2.2

[55] S. Alioli, P. Nason, C. Oleari, and E. Re, *NLO Higgs boson production via gluon fusion matched with shower in POWHEG*, JHEP **0904** (2009) 002, arXiv:0812.0578 [hep-ph]. 5.2.2

[56] P. Nason and C. Oleari, *NLO Higgs boson production via vector-boson fusion matched with shower in POWHEG*, JHEP **1002** (2010) 037, arXiv:0911.5299 [hep-ph]. 5.2.2

[57] S. Schumann and F. Krauss, *A Parton shower algorithm based on Catani-Seymour dipole factorisation*, JHEP **0803** (2008) 038, arXiv:0709.1027 [hep-ph]. 5.2.3

[58] S. Höche, F. Krauss, S. Schumann, and F. Siegert, *QCD matrix elements and truncated showers*, JHEP **0905** (2009) 053, arXiv:0903.1219 [hep-ph]. 5.2.3

[59] F. Siegert, *A practical guide to event generation for prompt photon production with Sherpa*, Journal of Physics G: Nuclear and Particle Physics **44** (2017) no. 4, 044007. `http://stacks.iop.org/0954-3899/44/i=4/a=044007`. 5.2.3

[60] R. D. Ball, V. Bertone, S. Carrazza, C. S. Deans, L. Del Debbio, S. Forte, A. Guffanti, N. P. Hartland, J. I. Latorre, J. Rojo, and M. Ubiali, *Parton distributions for the LHC run II*, Journal of High Energy Physics **2015** (Apr, 2015) 40. `https://doi.org/10.1007/JHEP04(2015)040`. 5.2.3

[61] T. Min, C. Meyer, R. Turra, H. Wang, and Y. Zhang, *Selection and performance for the $H \to \gamma\gamma$ 2018 analyses using 2015+16+17 pp collision data at $\sqrt{s} = 13$ TeV with the ATLAS detector*, . `https://cds.cern.ch/record/2309522`. 5.2.3

[62] ATLAS Collaboration, *Measurement of fiducial, differential and production cross sections in the $H \to \gamma\gamma$ decay channel with 13.3 fb^{-1} of 13 TeV proton-proton collision data with the ATLAS detector*, Tech. Rep. ATLAS-COM-CONF-2016-068, CERN, Geneva, Jul, 2016. `https://cds.cern.ch/record/2200362`. 5.3.2, 5.6.3

[63] L. Carminati, M. Delmastro, M. Kataoka, M. Kuna, R. Lafaye, S. Laplace, J. Leveque, K. Liu, G. Marchiori, I. Nikolic, J. Ocariz, L. Roos, J. Schaarschmidt, M. Schwoerer, and L. Yuan, *Measurement of the isolated di-photon cross section in 4.9 fb^{-1} of pp collisions at $\sqrt{s} = 7$ TeV with the ATLAS detector*, Tech. Rep. ATL-COM-PHYS-2012-592, CERN, Geneva, May, 2012. `https://cds.cern.ch/record/1450063`. 5.6.1

[64] HSG1 group, *Supporting note*
Limit on fiducial production cross-section of a narrow resonance decaying into two photons with the ATLAS detector using 20 fb^{-1} of proton-proton collision data, Tech. Rep. ATL-COM-PHYS-2013-1324, CERN, Geneva, Sep, 2013.
`https://cds.cern.ch/record/1599990`. 5.6.3

[65] ATLAS Collaboration, M. Aaboud et al., *Measurement of the photon identification efficiencies with the ATLAS detector using LHC Run-1 data*, Eur. Phys. J. **C**76 (2016) no. 12, 666, arXiv:1606.01813 [hep-ex]. 5.7.1

[66] C. E. Rasmussen and C. K. I. Williams, *Gaussian Processes for Machine Learning.* MIT Press, Cambridge, Massachusetts, 2006. `www.GaussianProcess.org/gpml`. 6, 6.1.2, 6.1.2

[67] J. Alison, F. Dias, L. S. Kaplan, Y.-K. Kim, K. Krizka, Y. W. Y. Ng, A. Sogaard, C. O. Shimmin, K. Whalen, D. Whiteson, K. Pachal, C. Leonidopoulos, and S. L. Wu, *Search for low mass di-jet resonances using proton-proton collisions at $\sqrt{s} = 13$ TeV with the ATLAS detector.*, Tech. Rep. ATL-COM-PHYS-2016-1786, CERN, Geneva, Dec, 2016.
`https://cds.cern.ch/record/2238709`. 6.2

[68] *ROOT*, http://root.cern.ch, Accessed July 2018. 6.2.1

[69] L. Moneta, K. Belasco, K. Cranmer, A. Lazzaro, D. Piparo, G. Schott, W. Verkerke, M. Wolf, K. Belasco, K. Cranmer, A. Lazzaro, D. Piparo, G. Schott, W. Verkerke, and M. Wolf, *The RooStats Project*, PoS **A**CAT2010 (Sep, 2010) 057.
`https://cds.cern.ch/record/1289965`. Comments: 11 pages, 3 figures, ACAT2010 Conference Proceedings. 6.2.1

[70] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, *Scikit-learn: Machine Learning in Python*, Journal of Machine Learning Research **1**2 (2011) 2825–2830. 6.2.1

[71] *PyHF - pure-python fitting/limit-setting/interval estimation HistFactory-style*, https://diana-hep.org/pyhf/, Accessed July 2018. 6.3