Practicing Background Subtraction of the Higgs Boson with the 13 TeV Dataset*

Dane Kealii Gollero
Department of Physics and Astronomy, University of Utah.
(ATLAS Collaboration)†
(Dated: August 26, 2021)

This data analysis experiment is an exercise in various data analysis techniques implemented by Particle Physicists at CERN in the 13 TeV Open Data set [3]. The purpose is to repeat and tweak the analyses done in 2012 to confirm the mass of the Higgs Boson and to analyze the effect of the various cuts employed to search for it. Several standard cuts are performed in the H $\rightarrow \gamma\gamma$ channel with reduced chi square values computed. Removing various cuts in the di-photon channel obscured the signal completely when none were performed and furthermore, peak shifting and signal broadening effects from excess events not attributed to a Higgs Boson decay were seen. In addition, we attempt to combine results from the H $\rightarrow$ ZZ channel to the di-photon via simple averaging procedure. A combined mass measurement of the Higgs Boson was observed: 124.5 ± 0.85 GeV which is in good agreement with the accepted value: 125.09 ± 0.32 GeV

I. INTRODUCTION

A. Theory (Brout-Englert-Higgs mechanism)

In the Standard Model of Particle Physics the Higgs Boson is an elementary particle of spin 0 whose interaction gives the other elementary particles their mass. Here we will cover a brief description of the mechanism.

Many important properties in particle physics are encoded in the so-called Lagrangian Density (Usually called a Lagrangian). Symmetries of the Lagrangian come out in the quantum-field theoretic calculations and have significant consequences on the real world.

All elementary particles obey certain symmetries and the predictions born out by the standard model are some of the most accurate in Human History. Despite this, before the theoretical discovery of the Higgs Boson the Standard Model predicted that all particles (electrons, quarks, W and Z Bosons, etc.) were massless. For the photons and the Gluons this was fine but for the W and Z bosons in particular this spelled death for the theory. During this time a particular field in the lagrangian was used so that all massive particles we find in nature acquired their mass via an interaction with it, dubbed the Higgs Field.

B. Higgs Generation and Contaminating Background

I'll describe two important processes which produce the Higgs Boson relevant to the Large Hadron Collider (LHC) as well as the contaminating background we reduce via cuts.

C. Jets

In many particle collisions collimated sprays of particles called Jets are produced by Quarks or Gluons coming out of a collision which then radiate a Gluon which then subsequently splits into more Gluons and Quark-AntiQuark pairs. This eventually forms Hadrons and these particles constitute the collimated spray. Jets can contaminate the Higgs Boson signal by annihilating into other products like photons. Here is an animation of a Jet [16]:

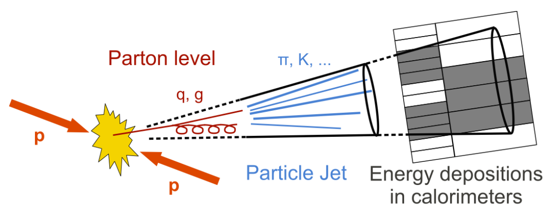

FIG. 1. Animation of a Jet taken from [16]

D. Gluon-Gluon Fusion (ggH)

One of the main ways the Higgs Boson is produced is via Gluon-Gluon fusion. At a very basic level, a quark is exchanged between two Gluons when then leads to a Quark and Anti-Quark annihlating to produce a Higgs Boson. The process at leading order is shown below:

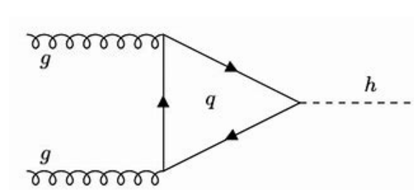

FIG. 2. Feynman diagram Gluon-Gluon Fusion[18]

---

* A footnote to the article title
†

E. Vector-Boson Fusion (VBF)

Vector Boson Fusion (VBF) produces a Higgs Boson via two quarks from their respective incoming proton beams radiating out a heavy vector-boson who then interact and produce a Higgs Boson. This process is accompanied by jets originating from the scattered Quarks.

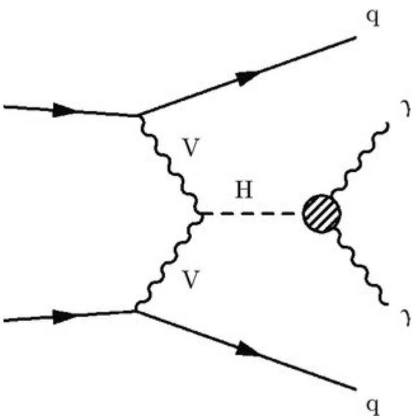

FIG. 3. Feynman diagram representing vector boson fusion[17]

In this short paper, we repeat some of the analyses used to discover the particle and reconstruct its mass. This is done using the 13 TeV Open Data set published by CERN. All of this is done to practice real-world techniques in experimental particle physics. In particular we highlight the effects of the removal of various backgrounds in the two channels:

- H $\rightarrow \gamma\gamma$
- H $\rightarrow ZZ$

Additionally, we combine the masses and uncertainties in each channel via a crude weighted averaging process to report a combined mass and uncertainty.

II. METHODS

All data in the 13 TeV dataset [3] were collected in the ATLAS detector and pre-processed for Educational Purposes at the Undergraduate and High School level. Despite this it is more than sufficient for the purposes of this paper. In the following subsections we provide a list of common terms, briefly describe the layout of the ATLAS detector, mention the relevant pieces of it for this analysis, describe the code needed to analyze the data and generate our plots and describe the rudimentary statistical tests used to evaluate them.

A. Terminology

(courtesy of ATLAS)

- Jets - Collimated Sprays of Particles [7]
- Luminosity - The number of particles passing down the beam line per unit time, per unit area [7]
- Barnes - Unit of area used by Physicists: $1b = 10^{-24}(cm)^2$
- Cross Section - The cross-section is a measure of the probability of a process occurring upon collision of particles.
- Branching Fractions - Ratio of number of channel decays to total decays [6]
- Pseudorapidity - A convenient quantity to measure in a particle physics experiment. Defined as: $\eta = -\ln\tan\frac{\theta}{2}$ where θ is the polar angle with respect to the direction along the beam-line.
- QCD - Quantum Chromodynamics, this is the Quantum Field Theory Describing the Strong Nuclear Force.
- Hadrons - A hadron is a subatomic particle that contains quarks, antiquarks, and gluons, and so experiences the strong force. [10]
- Leptons - A class of particles including the Electron.

B. ATLAS Detector

The ATLAS Detector is a cylindrical particle detector capable of measuring many different properties of elementary particles. The features of it most important to this paper are the inner detector (ID), the electromagnetic calorimeter (EMCAL), the Hadronic Calorimeter (HCAL), and the Muon Spectrometer. The ID gives position measurements of charged particles, EMCAL measures the energy and position of electromagnetic showers, while the HCAL and Muon Spectrometer measure Hadrons and Muons flying out of the collision. Below are pictures/visualizations of each respective piece:

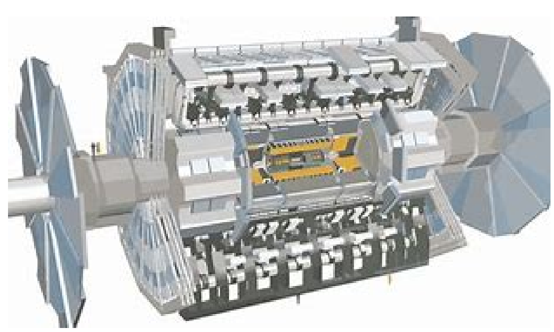

FIG. 4. visualization of ATLAS detector

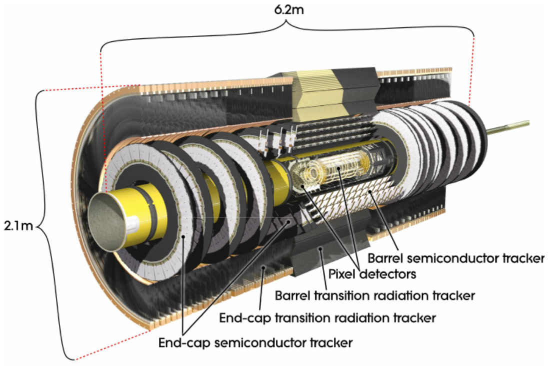

FIG. 5. Inner Detector (taken from [12])

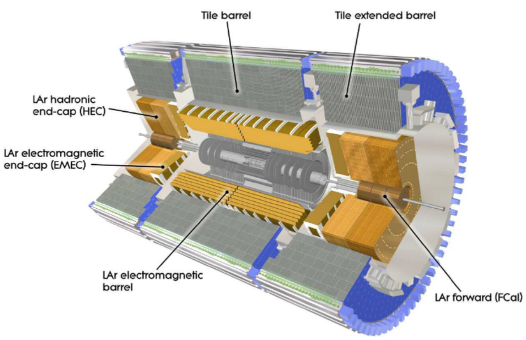

FIG. 6. Electromagnetic Calorimeters (taken from [11])

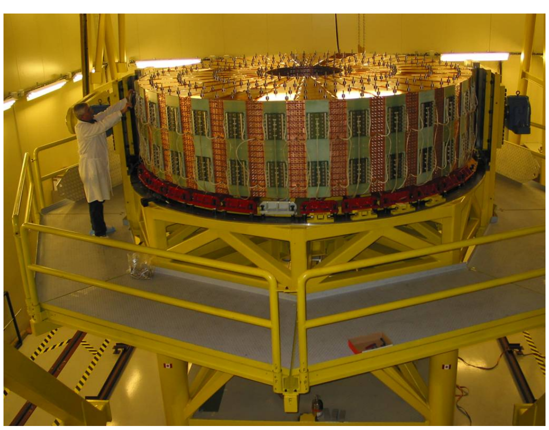

FIG. 7. Hadronic Calorimeter (taken from [13])

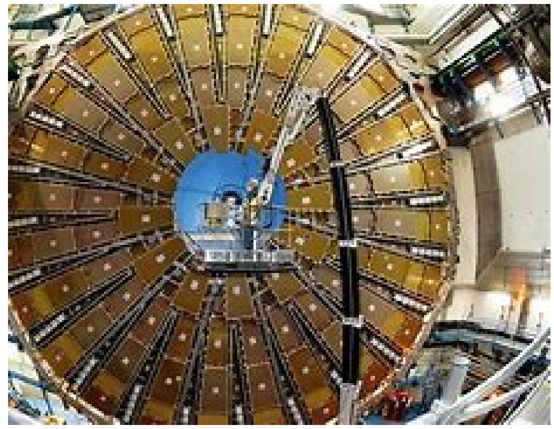

FIG. 8. Muon Spectrometer(taken from [14])

C. Data Processing

1. *Cuts for $H \to \gamma\gamma$*

For the Higgs Decaying into two photons, In order to isolate the two photons coming from a Higgs Boson decay we must sift through the data and cut away events in our detector not attributed to the Higgs. The following cuts are made:

- Energy Isolation in Calorimeter: This reduces the background from Hadron Jets decaying into photons. Specifically, we require that the transverse energy (which is the fraction energy of the photon going along the beam line direction) of each of the two photons are less than 4 GeV.
- Well re-Constructed Photons: The 13 TeV dataset includes information about whether or not the photons being analyzed have passed specific criteria relating to the electromagnetic shower shapes and whether or not there is energy leakage in the HCAL.
- Energy Threshold for two photons: There is a Transverse energy threshold requirement for the leading and sub-leading photons of > 40GeV and > 30GeV respectively.
- Optimal detection region cut: We exclude photon candidates in the region $1.37 < |\eta| < 1.52$.

D. Statistical Analysis

The codes to produce the plots in this paper are a mixture of modified versions of the OpenData set Jupyter Notebooks ran in Pycharm, and the actual Notebooks themselves. The data were fit with a third-order polynomial for the background and the signal portion of the model was a Gaussian with mean and width as free parameters. The error bars on the plots are calculated by assuming Poisson Statistics in each bin and calculating the square-root of the number of events. A reduced chi-square fit of the background + signal model (b+s) to the data, in addition to a reduced chi-square fit of just the background model (b) were carried out. From these reduced chi-square values a probability of an event as least as likely as the observed reduced chi-square was calculated with Scipy's stats package [20] to demonstrate the efficacy of the (b+s) model over the (b) model. The Jupyter Notebooks provided by ATLAS were used to produce figures (16) and (17) to show another crucial decay channel for the Higgs Boson. To extract the mass of the Higgs Boson from these plots two strategies were employed:

1. First, we use the mean of the fitted Gaussian in the (b+s) fit in the $H \to \gamma\gamma$ channel as the mass of the Higgs Boson, with the associated width as the uncertainty.

2. Second, in the $H \to ZZ$ channel we estimate the mean, and width by hand by wiping away extraneous information not directly attributed to the signal. The standard deviation is obtained by halving the width estimate.

Combining these two measurements was done using a simple averaging procedure.

Finally, using only a fraction of the data available (due to computer run-time constraints) we demonstrate the effect of the various cuts on the width and position of the signal peak in the data.

III. RESULTS

First we present the evidence of the Higgs Boson in two separate channels: The $H \to \gamma\gamma$ and the $H \to ZZ$. Next, the effect of dropping all di-photon channel cuts listed in section C of Methods is shown. Finally we further demonstrate the impact in both the width and position of the signal peak when various cuts are left out as well as on the number of events kept. We conclude with the averaged value of the Higgs Boson mass with its uncertainty as estimated from the di-photon and ZZ channels.

A. Evidence for the Background + Signal Hypothesis

The background + signal model used in the $H \to \gamma\gamma$ channel was extremely successful in all plots produced. Conversely, so was the background only model which fit nearly as well, although the better fit went to the b+s model indicating a signal. In addition another signal was observed in the $H \to ZZ$ channel at a similar mass.

Below is a plot of events per 2 GeV vs. di-photon invariant mass. A clear excess of events is shown near a mass of 123.98 GeV. Formally, giving us the measurement:

$$123.98 \pm 0.42 GeV \tag{1}$$

The top of the plot is all of the data and the b+s fit whilst the bottom is the signal fit to the data with the estimated background taken out. Both the reduced chi-squared for b+s and background only fits are shown at the top along with the degrees of freedom in the fitted model and associated probabilities based off of the appropriate chi-squared distribution. Note that the probability of b+s is not actually 1 in any of the plots in this section and is a relic of rounding.

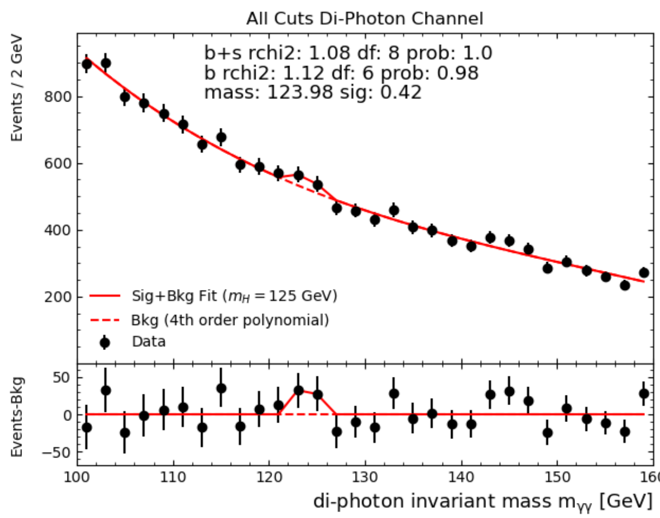

FIG. 9. Full GamGam data set (A,B,C,D) with standard ATLAS cuts,fraction (1)

Below is a plot of collision backgrounds from Monte-Carlo simulations as well as the data, and histograms of the data. Note in particular the cyan histogram near a mass of 125 GeV.

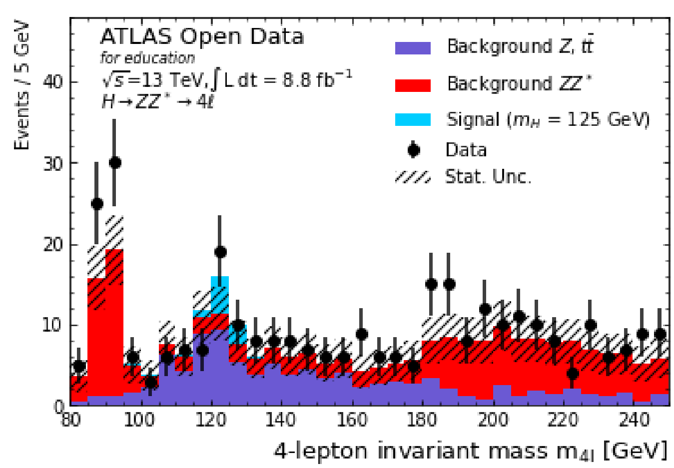

FIG. 10. ZZ channel plot generated with Open Data Set's Jupyter Notebook

B. Cut Effects

To illustrate the importance of subtracting backgrounds in an experiment we show a plot of the di-photon channel data without any of the cuts outlined above. Note that a discernable signal is difficult to distinguish below even in the signal-only portion in the bottom. Had none of the cuts been made this would have left considerable doubt as to whether or not the Higgs Boson was actually observed despite the good fit of both the background only and signal + background models.

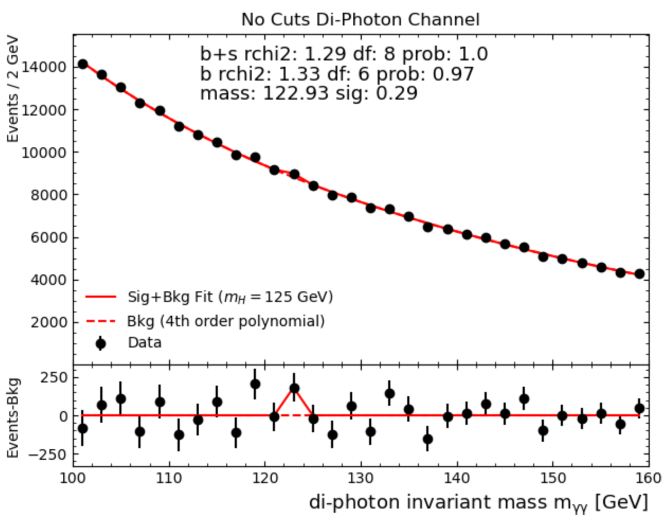

FIG. 11. Full GamGam data set (A,B,C,D) with no ATLAS cuts, fraction(1)

C. Peak Broadening and Displacement

The plots below were generated with less than a tenth of the data used above to account for slow run times. Despite this we can see the impact of the various cuts in how they shift the Higgs signal in different directions and significantly broaden it.

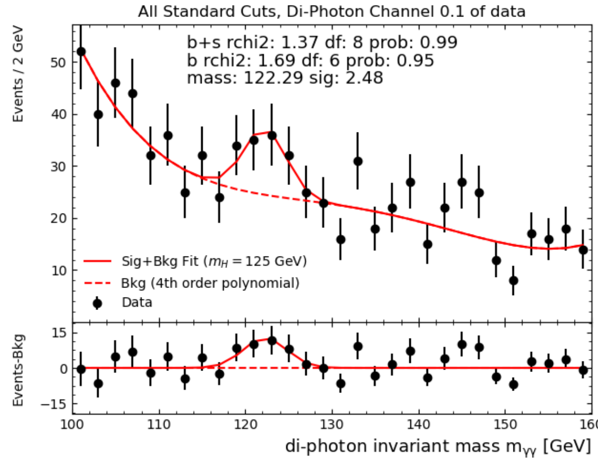

FIG. 12. All Cuts

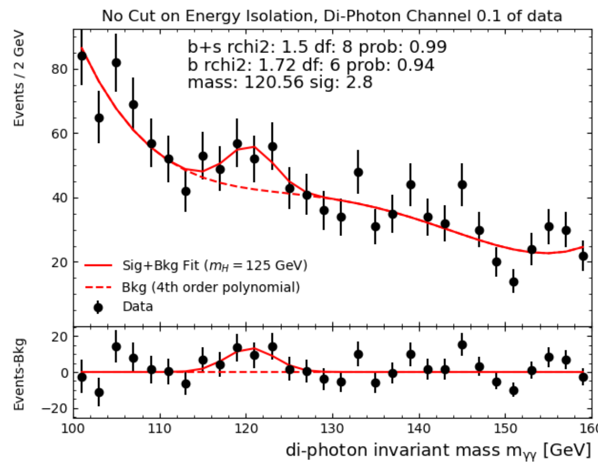

FIG. 13. No cut on energy isolation requirements

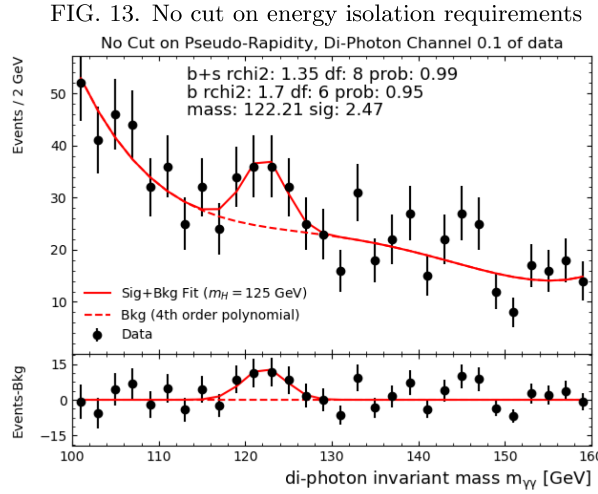

FIG. 14. No cut on pseudo-rapidity

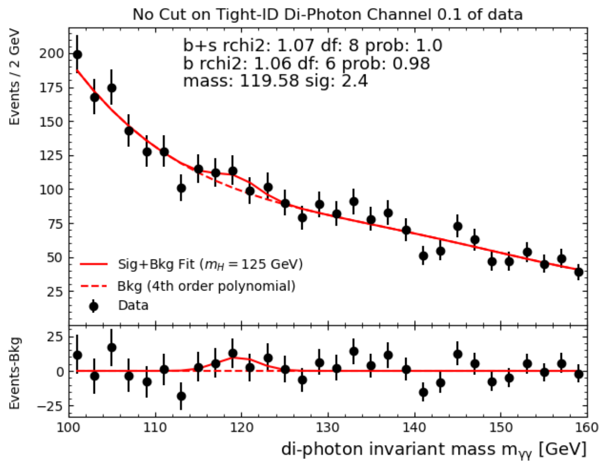

FIG. 15. No cut for well-constructed photons

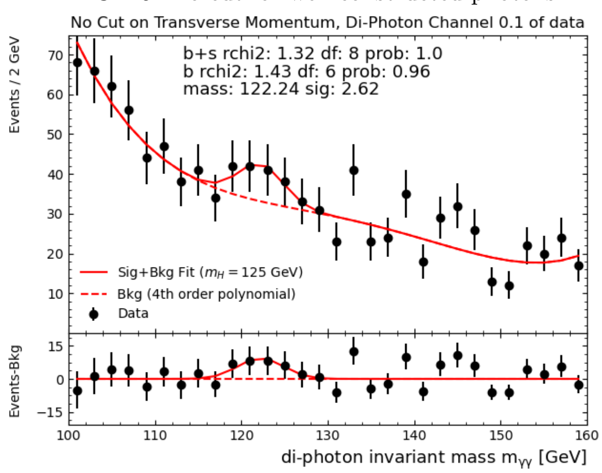

FIG. 16. No cut on transverse momentum

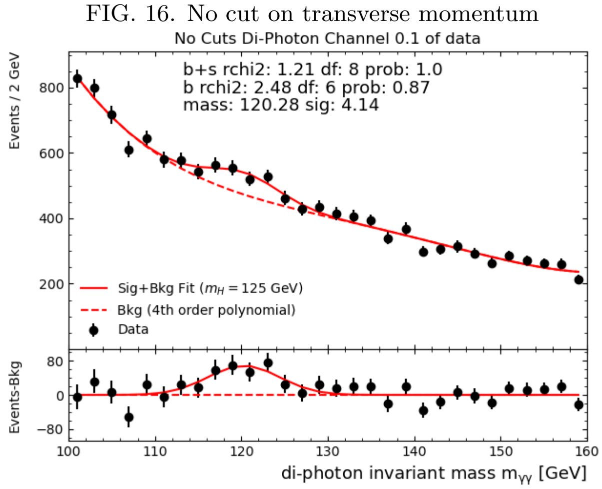

FIG. 17. No Cuts

D. Combining Uncertainties and Masses

In order to combine the mass measurements of the two channels above we needed to extract the mass from figure (10). To do this an ad-hoc procedure to estimate the mass and uncertainty was used. We wiped away all data but the signal from figure (10), took out the bottom of the histogram boosting it up, and eye-balled its peak and width. This yielded a mass and uncertainty of:

$$125 \pm 1.28 GeV \tag{2}$$

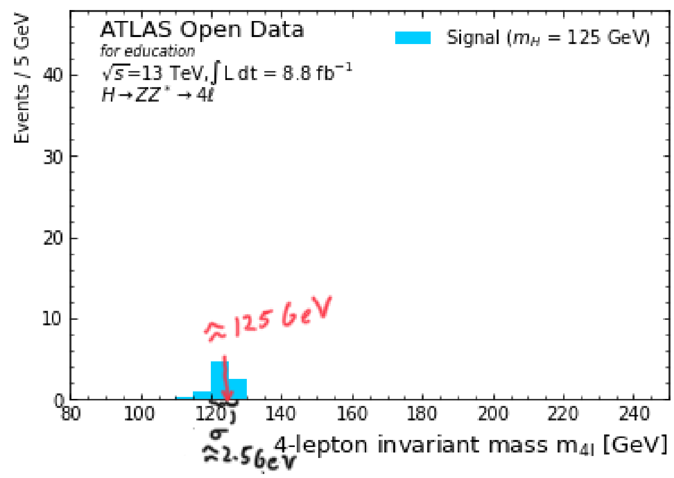

FIG. 18. ZZ channel plot with purely isolated signal and no bottom

Finally, we average (8) and (9) to report a final mass measurement of:

$$124.5 \pm 0.85 GeV \tag{3}$$

Which is in agreement with the accepted value of [4]:

$$125.09 \pm 0.32 GeV \tag{4}$$

IV. DISCUSSION

A. Background + Signal Comments

In the di-photon channel the goodness of fit of the background only model is strange at first but not expected to be a problem because of the way the fit was calculated. The background model was constructed by fitting the data directly and minimizing that fit, therefore the high probability of explaining the data observed is really a relic of how the background model was obtained and not an error in the code. The background + signal model was obtained similarly with a Gaussian in addition to the background polynomial and is superior to the background model across all plots shown.

B. Cut Effects

Cuts were necessary as shown in figure (11) in order to observe a significant signal and to ensure isolation of events coming from a Standard Model Higgs Boson. Further cut effects were observed in figures 12-17. These included broadening of the signal and a shift in the peak. All of these are due to excess events not attributed to a theoretical Higgs Boson.

C. Rigor of Higgs Observation

Although the background + signal model is remarkeably well-fit to the di-photon data, the number of data points in excess after the background has been subtracted is not much and there is still some ambiguity based on the results of the presence of the Higgs Boson. Similarly, the ZZ Channel measurements suffer from the same issues. Strangely, when using only a tenth of the data from only one portion of the OpenData Set, a better excess of events is seen in Figure (13) in comparison to figure (9). It remains to be seen if this is from transferring the code from the Jupyter Notebooks to the Author's home machine or if it comes from the events in the data set itself. It is also possible that not applying every cut present in [1] to the data could have been the issue as well as only the standard cuts present in the Jupyter Notebooks were successful. Further cuts were attempted but failed.

D. Statistical Analysis

The statistical analysis could be significantly improved by employing the Likelihood analysis of [1]. This will be pursued in future work, building off of this one.

V. CONCLUSION

Using the 13 TeV OpenData set provided by the ATLAS collaboration an excess of events in both the $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ$ channels were observed. Taking the average of the mass measurements and uncertainties from both channels using the width of the signal modelled as a Gaussian lead to a combined mass measurement of:

$$124.5 \pm 0.85 GeV \tag{5}$$

Which is in good agreement with the accepted value for the Higgs Boson mass:

$$125.09 \pm 0.32 GeV \tag{6}$$

Although the background + signal model in the diphoton channel fit the data extremely well, and an excess of events in the four-lepton channel was observed there were not enough events above the background in either case to claim a discovery. Despite this, the impact of the various cuts in the $\gamma\gamma$ channel were seen in the form of signal-broadening and peak shifting. Furthermore it is clear from comparing figure (9) to figure (11) the necessity of subtracting out background events to see a possible signal.

ACKNOWLEDGMENTS

We wish to Thank Wayne Springer for his tremendous efforts and help with this project, without whom this would not have been possible for the Author. In addition, a thank you to the ATLAS Collaboration is in order as well for providing the tools to do this.

VI. REFERENCES

1. Chatrchyan, S, et Al. , "Observation of a New Boson at a Mass of 125 GeV with the CMS Experiment at the LHC." Physics Letters. B 716.1 (2012): 30-61. Web.

2. Cousins, Robert D. "What is the Likelihood Function and How is it Used in Particle Physics?" (2020). Web.

3. ATLAS Collaboration, "Review of the 13 TeV ATLAS Open Data release", https://cds.cern.ch/record/2707171/files/ANA-OTRC-2019-01-PUB-updated.pdf, (accessed 4/18/2021)

4. Allport, P. P, et AL, "Measurements of Higgs Boson Properties in the Diphoton Decay Channel with 36 Fb-1 of Pp Collision Data at S=13 TeV with the ATLAS Detector." Physical Review. D 98.5 (2018): Physical Review. D, 2018-09, Vol.98 (5). Web.

5. Aaboud, M, Aad, G, Abbott, B, Abdinov, O, Abeloos, B, Abhayasinghe, D.K, Alderweireldt, S.C, Caron, S, Colasurdo, L, Groot, N. De, Fabiani, V, Filthaut, F, Igonkina, O, Ilic, N, Klein, M, König, A.C, Nektarijevic, S, Peaza Diaz, L, Schouwenberg, J.F.P, and Zwalinski, L. "Measurement of the Photon Identification Efficiencies with the ATLAS Detector Using LHC Run 2 Data Collected in 2015 and 2016." The European Physical Journal. C, Particles and Fields 79.3 (2019): 1-41. Web.

6. Griffiths, D. "Introduction to Elementary Particles (2nd edition)"

7. Schwartz, M. "Quantum Field Theory and the Standard Model"

8. Peskin, S. SChroeder, D. "An Introduction to Quantum Field Theory"

9. Lancaster, T. Blundell, S. "Quantum Field Theory for the Gifted Amateur"

10. CERN Homepage,https:home.cern/science/accelerators/large-hadron-collider, (accessed 4/21/2021)

11. Design and implementation of the Front End Board for the readout of the ATLAS liquid argon calorimeters, (Accessed 4/21/2021)

12. The detector control system of the ATLAS Semi-Conductor Tracker during macro-assembly and integration (needs url,author,etc)

13. https://www.atlas-canada.ca/hec2.html, (accessed 4/21/2021)

14. https://fineartamerica.com/featured/atlas-detector-muon-spectrometer-claudia-marcelloni-cernscience-photo-library.html,(accessed 4/21/2021)

15. https://atlas.cern/updates/feature/dark-matter

16. https://phys.org/news/2012-07-jets-cms-energy-scale.html

17. https://www.researchgate.net/publication/344422492_Search_f
proton_collisions_at_sqrts_13_TeV

18. https://www.researchgate.net/publication/343743976

19. https://www.theorie.physik.uni-muenche.de/lsfrey/teaching/archiv/sose_09/rng/higgs_mechan

20. https://www.scipy.org/