

# Measurements of the Higgs-boson mass in the $H \rightarrow \gamma\gamma$ channel using $10 \text{ fb}^{-1}$ of $pp$ collisions at $\sqrt{s} = 13 \text{ TeV}$ from ATLAS Open Data

H. Alida F. Hardersen  
*Fysisk Institutt, Universitetet i Oslo*  
(Dated: June 7, 2020)

In this report analysis of the diphoton sample in the 13 TeV ATLAS Open Data is preformed in the search for the Higgs boson mass in the decay channel  $H \rightarrow \gamma\gamma$ . The data is collected at ATLAS from  $pp$ -collisions in the LHC with  $\sqrt{s} = 13 \text{ TeV}$  and the integrated luminosity  $10\text{fb}^{-1}$ .

## I. INTRODUCTION

The leading theory to describe how the fundamental particles interact is the Standard model(SM) of particle physics. In the standard model is the mechanism that describes the origin of mass for the fundamental particles, the Brout-Englert-Higgs mechanism, named after the physicists that discovered it. With this mechanism also came the prediction of another fundamental particle, the scalar spin-0 Higgs boson, but it does not give any predictions on its mass. One of the main goals when building the Large Hadron Collider (LHC) at CERN was the discovery(or exclusion) of the Higgs boson. The Higgs boson mass is a fundamental constant of nature, and its measurement is required for precise calculations of high-energy processes. It is also thought to play a role for the stability of our universe with respect to quantum fluctuations.

The two beams circulating in the LHC are made up of many proton bunches, each containing  $10^{11}$  protons, spaced 25 ns apart. The measure of how many collisions that can be produced in the detector per area per second is the Luminosity, and this depends on the number of protons in each bunch, the number of bunches and the beam size at the collision point. The integral of luminosity over time reflects the amount of data taken.

A Higgs boson candidate was discovered in 2012 by the ATLAS and CMS experiments at CERN using data from the 7 TeV and 8 TeV runs of  $pp$ -collisions at the LHC. Measurements of its properties are consistent with the expected SM Higgs boson, and its mass is determined to be  $m_H = 125.09 \pm 0.21 \pm 0.11 \text{ GeV}$ . The ATLAS collaboration announced[1] the 95% exclusion of the Higgs in the mass range 111–559 GeV, except for a narrow region 122–131 GeV where an excess of events with significance  $5.9\sigma (p_0 = 1.7 \times 10^{-9})$  was observed (In the entire mass range of 110 – 600 GeV, the global significance of the excess was  $5.1\sigma$ ).

## II. ATLAS DETECTOR

One of the four major experiments at the LHC is the ATLAS (A Toroidal LHC ApparatuS) experiment[2]. ATLAS is a general-purpose detector designed to exploit

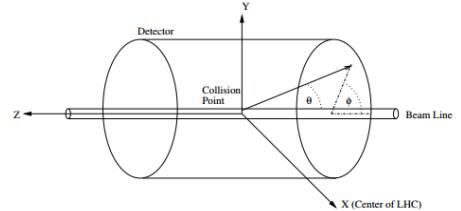


FIG. 1: Illustration of the coordinate system used in ATLAS. [3]

the full discovery potential and the large range of physics opportunities that the LHC provides. Because we are unable to store all the data from all the collisions, the ATLAS detector has a triggering system that filters what events are interesting and what stem from well known processes. Interesting processes are rare, so we want to get as many as possible. Figure 3 shows an illustration of the detector and its different components.

### A. Detector geometry

The ATLAS detector uses a right handed coordinate system with an origin at the interaction point (IP) in the center of the detector, the x-axis pointing towards the center of the LHC ring, the y axis pointing up and the z-axis pointing along the beam line as shown in figure 1. Cylindrical coordinates  $(r, \phi)$  are used in the transverse plane, where  $\phi$  is the azimuthal angle around the beam line, ranging between  $-\pi, \pi$  with respect to the x-axis. The polar angle  $\theta$  is measured from the z-axis and is in the range  $[0, \pi]$ .

The collisions take place in the center-of-mass(COM) frame of the  $pp$  system, which is not the COM frame of the colliding partons. The momentum and energy component of the partons along the z-axis is unknown, so we define the transverse components in terms of their projections on the x-y plane,

$$E_T = E \sin \theta, \quad p_T = E \sin \theta \quad (1)$$

$$p_x = p_T \cos \phi, \quad p_y = p_T \sin \phi, \quad p_z = p \cos \theta \quad (2)$$

It is useful to express the angles in terms of rapidity,  $y$ , which is invariant under a boost along the beam line,  $\Delta y = \Delta y'$ . For high energy jets, the jet mass is small compared to the jet energy and  $p_z \approx E \cos \theta$ .

$$y = \frac{1}{2} \ln \left( \frac{E + p_z}{E - p_z} \right) \approx \frac{1}{2} \ln \left( \frac{1 + \cos \theta}{1 - \cos \theta} \right) = \frac{1}{2} \ln \left( \cot^2 \frac{\theta}{2} \right) \quad (3)$$

Therefore the pseudorapidity,  $\eta$ , can be used in place of the rapidity at high energy

$$\eta = -\ln \left( \tan \frac{\theta}{2} \right) \quad (4)$$

The distance in  $\eta - \phi$  space is defined as,

$$\Delta R = \sqrt{(\Delta\eta)^2 + (\Delta\phi)^2} \quad (5)$$

### 1. Invariant mass

The invariant mass of a diphoton system can be found using the transverse momentum of the photons, the azimuthal angle  $\phi$  between the photons and their values of pseudorapidity  $\eta$ ,

$$m_{\gamma\gamma} = \sqrt{2p_{T,1}p_{T,2}(\cosh \Delta\eta - \cos \Delta\phi)} \quad (6)$$

where  $\Delta\eta = \eta_1 - \eta_2$  and  $\Delta\phi = \phi_1 - \phi_2$

## B. Detector components

The inner detector (ID)[4] consists of three detector components immersed in a 2T solenoidal magnetic field parallel to the beam axis, and measures the direction, and momentum of charged particles. The innermost component is a Pixel detector consisting of three silicon pixel layers. The innermost layer (b-layer) is designed for reconstruction of particles with longer lifetimes that will travel a short distance from the primary vertex before decaying. Following the pixel detector is the SemiConductor Tracker (SCT) consisting of silicon microstrips which surrounds the pixel detector layers. The two first silicon detectors provide precision tracking within the pseudorapidity range  $|\eta| < 2.5$ . The last component is a Transition Radiation Tracker (TRT) which is composed of gaseous proportional counter-straws embedded in a radiator material. The straws are filled with a 70% Xe, 27% CO<sub>2</sub> and 3% O<sub>2</sub> gas mixture. The TRT offers discrimination between electrons and charged hadrons for  $|\eta| < 2.0$  [5].

The ID is surrounded by a lead-liquid argon (LAr) electromagnetic calorimeter (ECAL) with accordionshape

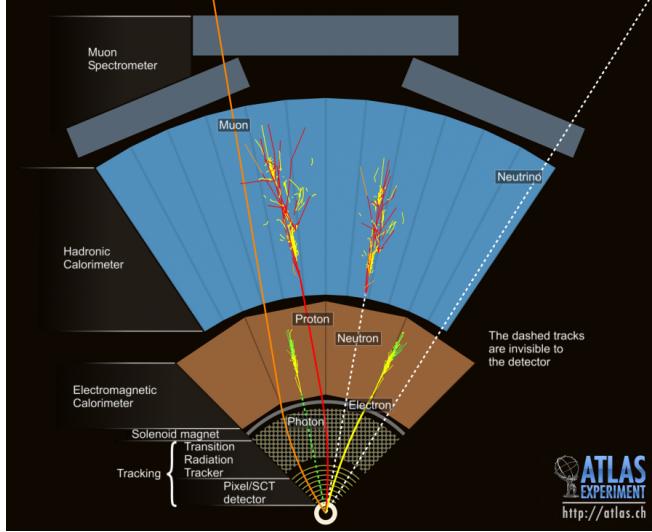
absorbers and electrodes. The ECAL consists of a barrel section that covers the pseudorapidity region  $|\eta| < 1.475$ , and two end-cap regions, one covering  $1.375 < |\eta| < 2.5$  and the other covering  $2.5|\eta| < 3.2$ . In between the barrel and end-cap region ( $1.37 < |\eta| < 1.52$ ) is a section of mostly cables and cryostat walls that give bad precision. In the region matching the ID pseudorapidity, the ECAL is segmented into three layers. First there is a layer with a fine  $\eta$  segmentation in order to discriminate the isolated photons from neutral hadrons decaying to pairs of close-by photons such as  $\pi^0 \rightarrow \gamma\gamma$ . The second layer is where most of the energy is collected, and together with the first layer it is possible to measure the shower direction without making any assumptions on the photons production point. The third layer is less segmented and collects the tails of the energy distribution.

The next component of ATLAS is the hadron calorimeter (HCAL), which consists of three parts, 1) an iron-scintillator tile calorimeter in the barrel, covering the central region  $|\eta| < 1.7$ . 2) the liquid-argon hadronic end-cap calorimeters covering the range  $1.5 < |\eta| < 3.2$ , and 3) the liquid-argon forward calorimeter which covers the range  $3.1 < |\eta| < 4.9$ . The three systems are closely coupled together giving a hermetic design which minimises the energy losses in cracks and limits the backgrounds that might reach the muon spectrometer (MS) that surrounds the calorimeters. The MS consists of separate trigger and precision tracking chambers in the magnetic field provided by three large toroids. The momentum of charged particles are measured in the pseudorapidity ranges  $|\eta| < 2.4$  and  $|\eta| < 2.7$  for the trigger and precision tracking chambers respectively. Figure 2 illustrates the interaction of particles with different components of the detector.

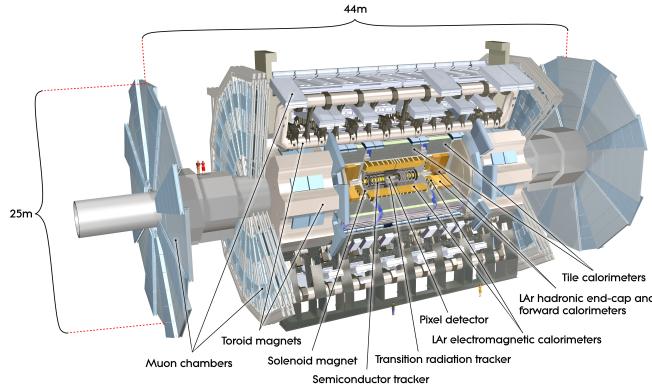
## III. THE HIGGS BOSON PRODUCTION AND DECAY

### A. Higgs production

The main production mechanisms for the Higgs boson at the LHC are gluon-gluon fusion (ggH), vector-boson fusion (VBF), production with a gauge boson (ZH/WH) and production associated with two top quarks (ttH). The leading feynman diagrams for these processes are shown in 4. The most dominant production mechanism at the LHC is the gluon-gluon fusion mediated by the exchange of a virtual top quark, with a contribution of 88.2% of the total cross section at  $\sqrt{s} = 13\text{TeV}$ [6]. The second most dominating process will be the VBF where scattering of two quarks produces a Higgs boson, with a contribution to the total cross section of 6.8%. The associated processes WH, ZH and ttH have a contribution to the cross section of 2.5%, 1.6%, and 0.9% respectively. The cross section as a function of center of mass energy



**FIG. 2:** Illustration of particle interactions in the ATLAS detector. The dotted lines indicate that the particle does not interact. (Copyright CERN)

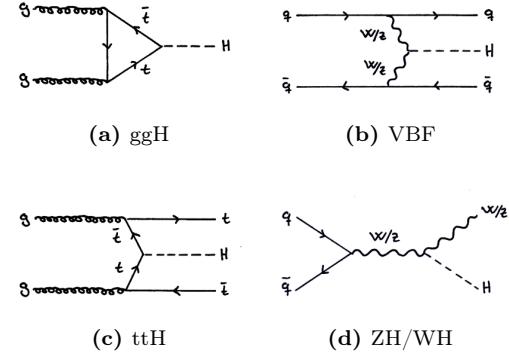


**FIG. 3:** The ATLAS detector and its different components. The people at the bottom and to the left of the detector gives a good impression of how large the detector really is. (Copyright CERN)

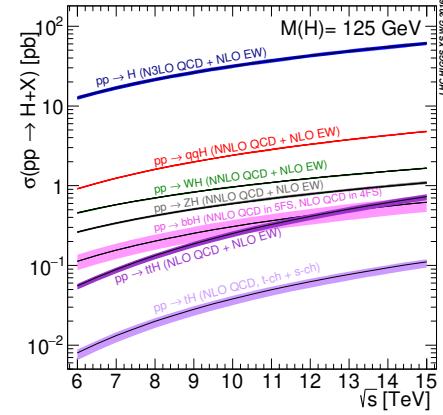
is shown in 5

### B. The $H \rightarrow \gamma\gamma$ decay

The higgs boson has multiple decay channels, this analysis considers the decay  $H \rightarrow \gamma\gamma$  in which the Higgs boson decays to a pair of photons through a quantum loop process involving massive charged particles, mainly the W-boson or top quark as illustrated in figure 6. This decay channel is rare, with a branching ratio of  $2.27 \times 10^{-3}$  for  $m_H = 125\text{GeV}$ , [6], but because the final state particles can be precisely measured the resolution of the reconstructed  $m_H$  is very good.



**FIG. 4:** Leading order Feynman diagrams for the main Higgs production mechanisms

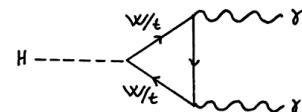


**FIG. 5:** The SM Higgs boson production cross sections as a function of the center of mass energy,  $\sqrt{s}$ , for  $pp$  collisions at the LHC [6]

## IV. EVENT SELECTION

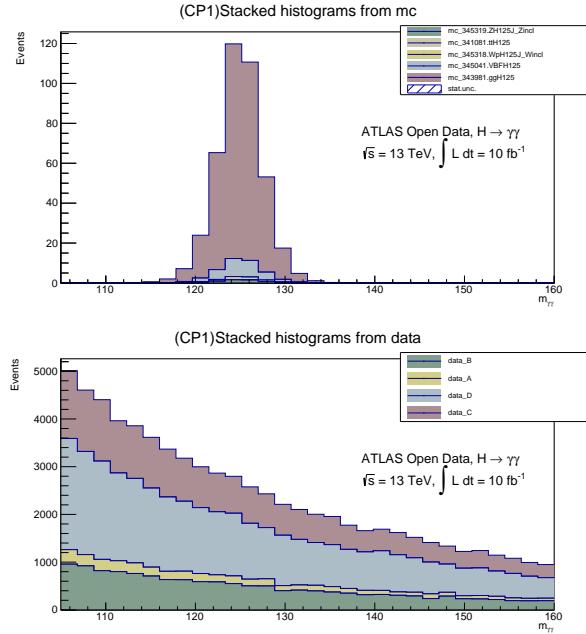
### A. Preselection

Events are preselected<sup>1</sup> using a diphoton trigger which requires the presence of two clusters of energy deposited in the ECAL with  $E_T > 25\text{ GeV}$  each. The two photon candidates must be in the fiducial region of



**FIG. 6:** Feynman diagram of the decay  $H \rightarrow \gamma\gamma$

<sup>1</sup> Selection applied before treating the photons separately



**FIG. 7:** Contribution to the data from each of the data-sets. Top is stacked histograms for for the different Higgs production processes, and bottom is the data. This is only included as an interesting plot.

the calorimeter with  $|\eta| < 1.37$  or  $1.52 < |\eta| < 2.37$  so the barrel-endcap transition regions where the precision is low are excluded. The photons then have to pass the tight identification criteria which are used to reduce the contamination from the background, primarily associated with neutral hadrons in jets decaying into photon pairs. These criteria adds information from the finely segmented strip layer of the ECAL described in section II B. These tight criteria are optimised separately for unconverted and converted photon candidates, and are described in [7]. For the ALTAS Open data, information of whether the photons have passed these criteria is stored in a boolean variable.

An isolation requirement is then applied in order to further suppress the background photons from the prompt photons. Background photons are often accompanied by surrounding hadron activity and therefore have poor isolation. This isolation is defined based on the transverse energy flow in an isolation cone around the photon with  $R = \sqrt{\eta^2 - \phi^2}$ . The energy flow is characterised by 1) Calorimeter isolation, which is the sum of the energy of the clusters in a cone  $R = 0.2$ , and 2) Track isolation, which is the scalar sum of transverse energy of the tracks within a cone  $R = 0.3$ . In this analysis, candidates with a calorimeter isolation larger than 5% of the photon transverse momentum are rejected, and candidates with a track isolation larger than 5% of the photon transverse momentum are rejected.

**TABLE I:** Cut flow of the analysis event selection shows the number of events in data( $N_D$ ) and MC( $N_{mc}$ ) samples after each cut, and the fraction of events after the cut. The number of MC events are not scaled.

Cut	$N_D$	%	$N_{mc}$	%
Total events	7798424	100.0	2473335	100.0
Diphoton trigger	7798424	100.0	2473335	100.0
Two photons	7772074	99.7	2434618	98.4
Preselection	335772	4.3	1584334	64.1
Mass window	88446	1.1	1582958	64.0
Kinematic selection	71587	0.9	1456541	58.9
CP1 (inclusive)	71587	0.9	1456541	58.9
CP2	13536	0.2	376687	15.2
CP3	8113	0.1	222146	9.0
CP4	5423	0.1	154541	6.2

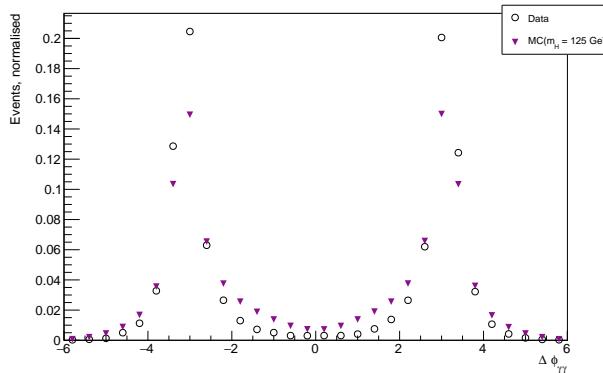
## B. Kinematic selection

After the preselection, the photon candidate with the highest  $p_T$  is labelled *leading* and the other is labelled *subleading*, the leading photon are required to have  $E_T > 35 \text{ GeV}$  and  $E_T > 25 \text{ GeV}$  respectively. A variable  $p_T$  cut is then applied to the photons, where the leading photon is required to have  $p_T/m_{\gamma\gamma} > 0.35$  and the subleading photon is required to have  $p_T/m_{\gamma\gamma} > 0.25$ . Using variable  $p_T$  thresholds that are scaled by mass prevents distortions in the low end of the invariant mass spectrum [8]. The effect of this  $p_T$  cut on the inclusive samples can be seen in the two top plot of 13. The invariant mass of the diphoton system is found using equation 6, with  $p_{T,1}$  and  $p_{T,2}$  as the transverse momentum of the leading and subleading photon candidate respectively. This analysis considers a mass window of  $105 \text{ GeV} \leq m_{\gamma\gamma} \leq 160 \text{ GeV}$ , and the number of events and fraction of the original number is summerized in table I. The pseudorapidity of the photons are also shown in figure 13, where we can clearly see the cut on the barrel-endcap region.

The difference in azimuthal angle  $\Delta\phi_{\gamma\gamma}$  for the leading and subleading photons is shown in figure 8. From momentum conservation it can be expected that the photons are produced back-to back with  $\Delta\phi = \pm\pi$  which corresponds well with the figure, indicating that the Higgs boson was produced without a lot of motion orthogonal to the beam direction.

## C. Event categories

After the event selection is done, the events are further separated into three categories defined by weather the photons are in the central region of the detector and if both photons are converted or unconverted. The categories used are:



**FIG. 8:** The difference in azimuthal angle between the leading and subleading photon candidate. The top figure is the data, and the bottom figure is the MC samples. [6]

**TABLE II:** Expected signal events for each category from the MC samples with  $m_H = 125$  GeV. CP1 is the inclusive category, and SM are the expected fraction of signal events from the SM, found in the review of particle physics. [6]

	Total	ggH [%]	VBF[%]	WH[%]	ZH[%]	ttH[%]
CP1	407.05	89.771	7.516	1.392	1.319	0.002
CP2	101.83	89.803	7.723	1.227	1.245	0.003
CP3	60.27	89.768	7.745	1.241	1.243	0.003
CP4	41.56	89.854	7.692	1.205	1.247	0.003
SM		88.2	6.86	2.48	1.59	0.9

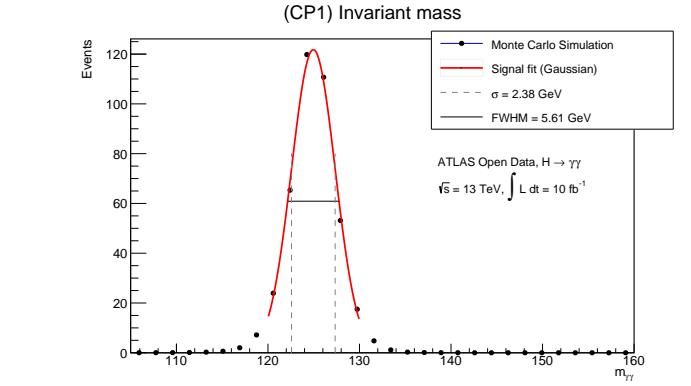
CP1 Inclusive category, both unconverted and converted photons in the fiducial region of the detector.

CP2 Central Inclusive, both unconverted and converted photons in the central region  $|\eta| < 0.75$ .

CP3 Central unconverted, both photons are unconverted and in the central region of the detector.

CP4 Central converted, both photons are converted and in the central region of the detector.

The number of events in each category after all cuts is listed in I. The signal yield for each of the production processes described in section III A are listed in table II. The signal yield for ggH and VBF is somewhat higher than the expected signal yield, while for WH, ZH and the yield from MC is slightly lower. The fraction from the ttH signal sample is a lot lower than expected.



**FIG. 9:** Gaussian fit to signal MC samples for the inclusive category with  $m_H = 125$  GeV. The vertical gray lines shows the interval containing 68.3% of the distribution and the horizontal line shows the full width at half maximum.

## V. SIGNAL AND BACKGROUND MODELLING

The invariant mass of the Higgs boson signal shape is modelled by fitting a Gaussian function<sup>2</sup> to monte carlo samples simulating the Higgs boson production mechanisms in the mass window  $\pm 5\text{GeV}$  around  $125\text{GeV}$ . The simulations are weighted to account for pileup, photon efficiency and trigger efficiency, as well as scaling the to the correct cross section and integrated luminosity, Figure 9 shows the signal model fit for the inclusive category.

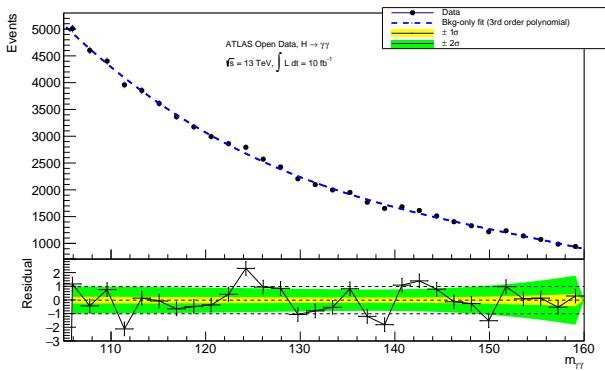
The background is estimated from data with a 3rd order polynomial fit in the full invariant mass range. The result of a background only fit is shown in figure 10. The lower inset shows the residuals of the data after subtracting the background fit.

The signal+background is modelled using a combination of the Gaussian from the signal fit and a 3rd order polynomial with free parameters. This function is fitted to data in the full invariant mass range. The fit is shown in 11, where the lower inlay shows the residuals after subtracting the fit from the data.

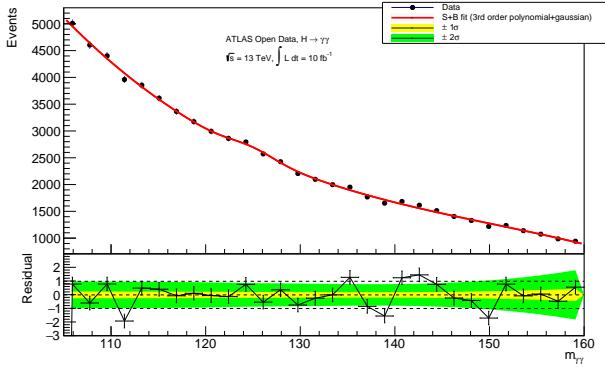
## VI. RESULTS

The distributions of the invariant mass for the diphotons in the inclusive category is shown in figure 12. A small excess can be seen around  $m_{\gamma\gamma} = 125$  GeV. The lower figure of figure 12 shows the residuals after the background component is removed. Figure 18 shows this plot for the remaining categories. An estimate of the

<sup>2</sup> Gaussian distribution:  $P(x) = Ce^{-\frac{(x-\mu)^2}{2\sigma^2}}$  where  $\mu$  and  $\sigma^2$  is the mean and variance of the distribution

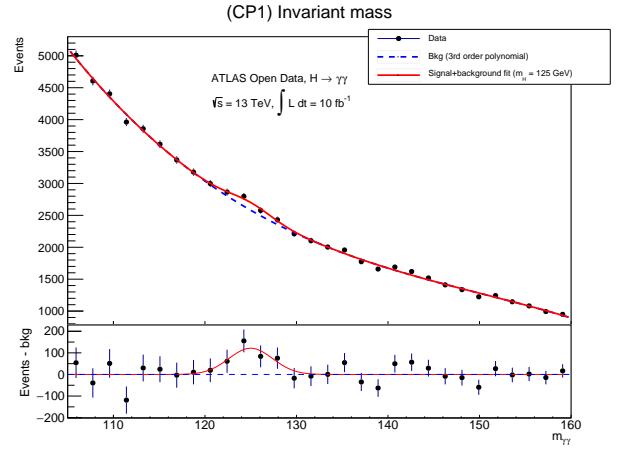


**FIG. 10:** Background only fit to the inclusive data sample using a 3rd order polynomial in the mass region 105 – 160GeV. The lower inlay shows the residuals of the data wrt the fit, where the green and yellow standard deviation bands include statistical uncertainties in the background component of the fit.



**FIG. 11:** Combined signal+background fit to the data in the mass region 105 – 160GeV. The lower inlay shows the residuals of the data wrt the fit, and the green and yellow standard deviation bands include statistical uncertainties in the background component of the fit.

significance <sup>3</sup> for the signal + background hypothesis in a mass window 120 – 130 GeV by counting the number of expected background events, expected number of signal events and the actual number of observed events yields an expected p-value of  $9.13 \times 10^{-4}$  which translates into an expected significance of  $3.12\sigma$ . The observed p-value in this region is  $1.26 \times 10^{-3}$ , which translates to  $Z_{obs} = 3.02\sigma$ . For the background only hypothesis, the  $p_0$ -value in the signal region 120 – 130 GeV is  $1.01 \times 10^{-1}$  which would translate to an observed significance  $1.28\sigma$ .



**FIG. 12:** Diphoton invariant mass spectrum. The red line shows the signal+background model, and the blue is the background component. The inlay shows the residuals of the data and the s+b model after background is subtracted.

Modelling the signal using a Gaussian distribution renders a  $\chi^2/ndf = 1207/3$ , with a very low probability  $P(\chi^2 \geq 1207) \approx 0\%$ , which is a lot lower than the 5% minimum for a decent fit. The results of the signal+background fit and the background only fit has a  $\chi^2/ndf$  much closer to 1 (0.76 and 1.08 respectively).

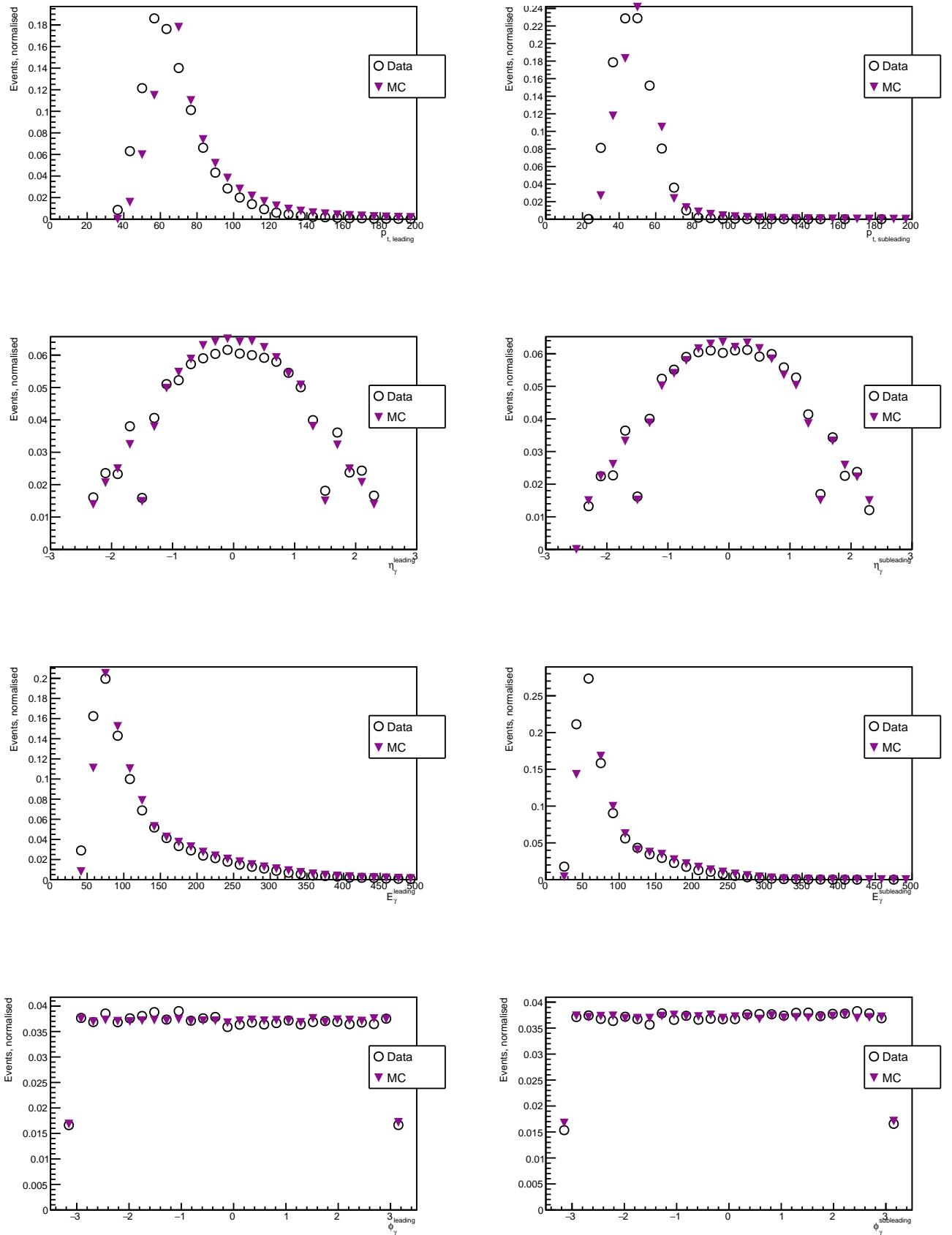
## VII. CONCLUSION/SUMMARY

An analysis of the 13TeV ATLAS Open data for  $H \rightarrow \gamma\gamma$  is preformed, using MC samples to model a Higgs boson signal, while background is estimated from data. An excess can be seen around 125 GeV, however, in order to determine if this excess is more than a statistical fluctuation in the background, a proper statistical analysis needs to be preformed.

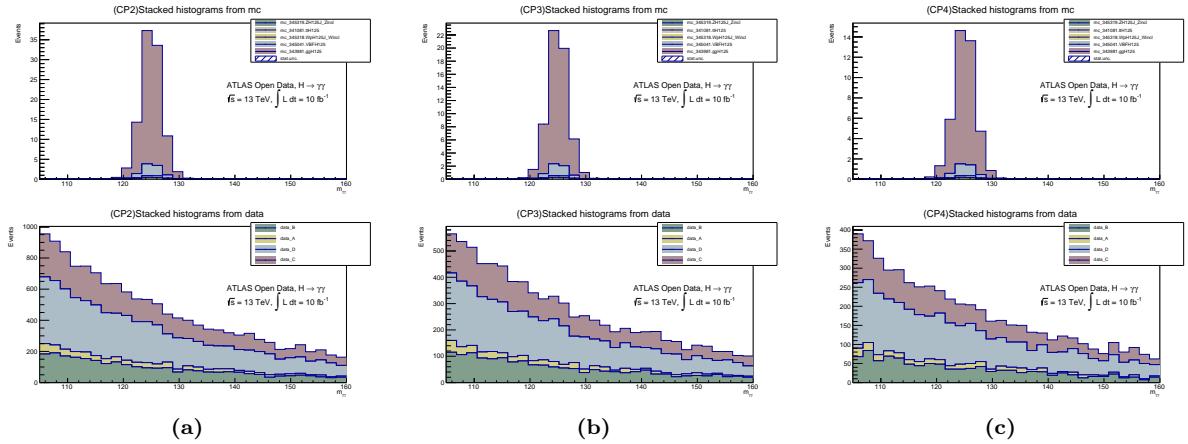
<sup>3</sup>  $Z_{exp} = \sqrt{2 \left[ (s + b) \ln \left( 1 + \frac{s}{b} \right) - s \right]}$  and  
 $Z_{obs} = \sqrt{2 \left[ n_{obs} \ln \frac{n_{obs}}{b} - n_{obs} + b \right]} [9]$

- 
- [1] G. Aad, T. Abajyan, B. Abbott, J. Abdallah, S. Abdel Khalek, A. Abdelalim, O. Abdinov, R. Aben, B. Abi, M. Abolins, et al., Physics Letters B **716**, 1–29 (2012), ISSN 0370-2693, URL <http://dx.doi.org/10.1016/j.physletb.2012.08.020>.
- [2] T. A. Collaboration, G. Aad, E. Abat, J. Abdallah, A. A. Abdelalim, A. Abdesselam, O. Abdinov, B. A. Abi, M. Abolins, H. Abramowicz, et al., Journal of Instrumentation **3**, S08003 (2008), URL <https://doi.org/10.1088%2F1748-0221%2F3%2F08%2Fs08003>.
- [3] M. Schott and M. Dunford, The European Physical Journal C **74** (2014), ISSN 1434-6052, URL <http://dx.doi.org/10.1140/epjc/s10052-014-2916-1>.
- [4] E. Stanecka (ATLAS), in *32nd International Symposium on Physics in Collision* (2013), pp. 383–388, 1303.3630.
- [5] M. Aaboud, G. Aad, B. Abbott, O. Abdinov, B. Abeelooos, S. H. Abidi, O. S. AbouZeid, N. L. Abraham, H. Abramowicz, H. Abreu, et al. (ATLAS Collaboration), Phys. Rev. D **98**, 052005 (2018), URL <https://link.aps.org/doi/10.1103/PhysRevD.98.052005>.
- [6] M. Tanabashi, K. Hagiwara, K. Hikasa, K. Nakamura, Y. Sumino, F. Takahashi, J. Tanaka, K. Agashe, G. Aielli, C. Amsler, et al. (Particle Data Group), Phys. Rev. D **98**, 030001 (2018), URL <https://link.aps.org/doi/10.1103/PhysRevD.98.030001>.
- [7] M. Aaboud, G. Aad, B. Abbott, O. Abdinov, B. Abeelooos, D. K. Abhayasinghe, S. H. Abidi, O. S. AbouZeid, N. L. Abraham, and et al., The European Physical Journal C **79** (2019), ISSN 1434-6052, URL <http://dx.doi.org/10.1140/epjc/s10052-019-6650-6>.
- [8] A. M. Sirunyan, A. Tumasyan, W. Adam, F. Ambrogi, E. Asilar, T. Bergauer, J. Brandstetter, E. Brondolin, M. Dragicevic, and et al., Journal of High Energy Physics **2018** (2018), ISSN 1029-8479, URL [http://dx.doi.org/10.1007/JHEP11\(2018\)185](http://dx.doi.org/10.1007/JHEP11(2018)185).
- [9] O. Behnke, K. Kröninger, T. Schörner-Sadenius, and G. Schott, eds., *Data analysis in high energy physics: A practical guide to statistical methods* (Wiley-VCH, Weinheim, Germany, 2013), ISBN 978-3-527-41058-3, 978-3-527-65344-7, 978-3-527-65343-0.

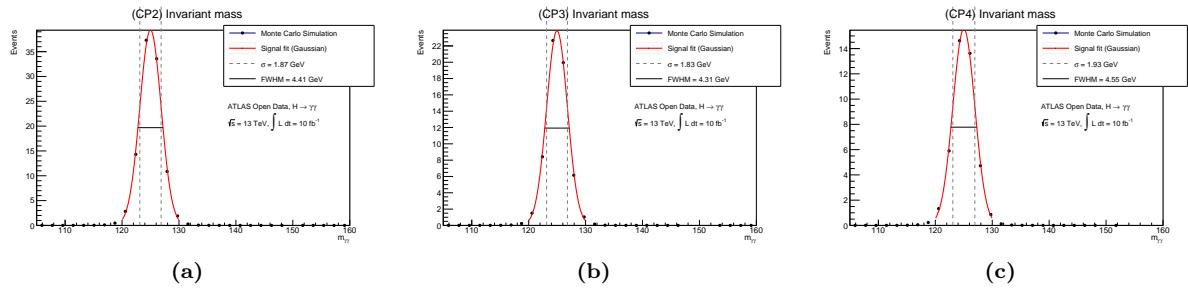
## VIII. APPENDIX



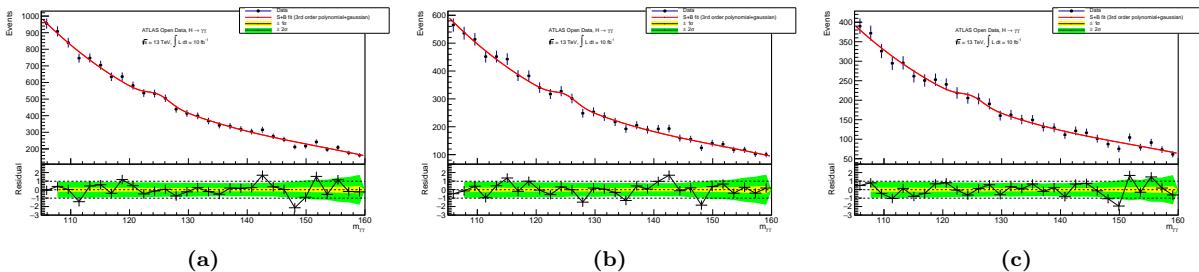
**FIG. 13:** Distributions of  $p_{t,\gamma}$ ,  $\eta_\gamma$ ,  $E_\gamma$  and  $\phi_\gamma$  for the leading(left) and subleading(right) photons. Data is represented by white circles and MC samples for  $m_H = 125\text{GeV}$  is represented by pink triangles.



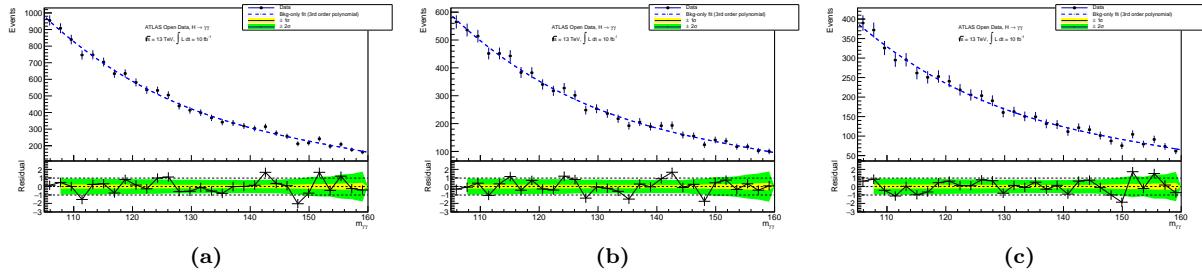
**FIG. 14:** Contribution to the data from each of the data-sets. Top is stacked histograms for for the different Higgs production processes, and bottom is the data.



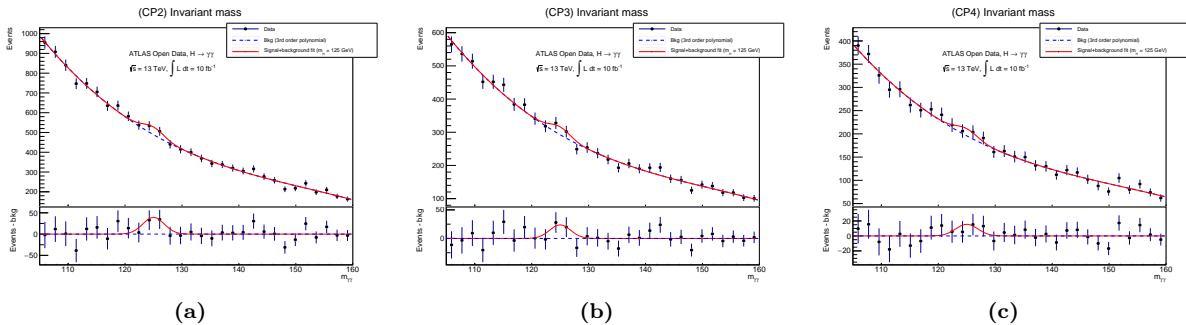
**FIG. 15:** Gaussian fit to signal MC samples for the remaining categories with  $m_H = 125 \text{ GeV}$ . The vertical gray lines shows the interval containing 68.3% of the distribution and the horisontal line shows the full width at half maximum.



**FIG. 16:** Combined signal+background fit to the data for the remaining categories in the mass region 105 – 160GeV. The lower inlay shows the residuals of the data wrt the fit, and the green and yellow standard deviation bands include statistical uncertainties in the background component of the fit.



**FIG. 17:** Background only fit to the remaining categories using a 3rd order polynomial in the mass region  $105 - 160\text{GeV}$ . The lower inlay shows the residuals of the data wrt the fit, where the green and yellow standard deviation bands include statistical uncertainties in the background component of the fit.



**FIG. 18:** Diphoton invariant mass spectrum. The red line shows the signal+background model, and the blue is the background component. The inlay shows the residuals of the data and the s+b model after background is subtracted.