Analysis Based on Data from ATLAS Experiment in Search of Higgs Boson*

Linxuan Zhu, Xueyi Tang, and Yinrui Liu
(Dated: May 16, 2019)

Based on the data from ATLAS experiment in 2011 and 2012, a rough data analysis process is performed in order to repeat the discovery of standard model Higgs boson, on the Higgs to diphoton channel. The analysis proves its existence on a basis of 3.4σ and 3.6σ for data in 2011 and 2012 respectively, and determines the invariant mass of Higgs boson to be $126.5^{+1.3}_{-1.2}$ GeV, where the uncertainty is statistical only. The results are in line with those in the article[1] published by ATLAS collaboration.

I. INTRODUCTION

The Higgs boson is a particle predicted in the standard model theory. It is named after a physicist Peter Higgs, who contributed to the proposition of the mechanism suggesting the existence of this particle.

In 2012, the discovery of standard model Higgs boson was announced by two collaborations of LHC — ATLAS collaboration and CMS collaboration. The statistics significance announced by ATLAS is over five sigma. This article performs data analysis in the channel $h \to \gamma\gamma$, using data from ATLAS experiment in 2011 and 2012, with the total amount of entries 23788 and 35251 respectively.

II. IMPLEMENTATION

The observable in this analysis process is the invariant mass of diphoton m_H. Only m_H ranging from 100 GeV to 160 GeV survives the selection of detectors, and there are 23788 and 35251 entries for further selection for year 2011 and 2012 separately. The distribution of m_H is approximately of exponential decay. The distribution of signal is assumed to be a linear combination of a Crystal Ball function and a Gaussian function, which is decided by Monte-Carlo simulation. Because of the fluctuation of background, it is difficult to recognize the signal directly. Therefore, here we employ a method of categorization, based on the signal resolution and signal-to-background ratios. More specifically, the entries are divided into ten categories in each year. It is defined by where the photons are detected (central region, rest region, or transition region), whether the photons have converted into electrons, and the value of their transverse momentum p_{Tt}. Also, a category containing two jets is also included. The ten categories in each year as well as the number of entries are presented in Table I.

#	Category	year N	2011 N_S	year N	2012 N_S
1	Unconv. cetral, low p_{Tt}	2054	10.5	2945	14.2
2	Unconv. cetral, high p_{Tt}	97	1.5	173	2.5
3	Unconv. rest, low p_{Tt}	7129	21.6	12136	30.9
4	Unconv. rest, high p_{Tt}	444	2.8	785	5.2
5	Conv. cetral, low p_{Tt}	1493	6.7	2015	8.9
6	Conv. cetral, high p_{Tt}	77	1.0	113	1.6
7	Conv. rest, low p_{Tt}	8313	21.1	11099	26.9
8	Conv. rest, high p_{Tt}	501	2.7	706	4.5
9	Conv. transition	3591	9.5	5140	12.8
10	Two-jet	89	2.2	139	3.0
	Inclusive	23788	79.6	35251	110.5

TABLE I. The ten categories in each year as well as their contents. N represents the total entries in each category, while N_S is the expected number of signals given by standard model theories in each category, also considering the detector performance. These values are also given in article[1].

The major idea of this analysis is a hypothesis test. We will compare the statistics significance of the null hypothesis (background-only model), where we assume there is no signal, with the alternative hypothesis (signal-plus-background model), where we assume there is signal of expected shape. The larger the difference between the two significance, the more likely the signal we expected exists. Considering the physical information in each category, the background models are slightly different. For the even number categories, the background model is exponential distribution. For category 1,5,9, it is an exponential of second-order polynomial. For category 3 and 7, it is a fourth-order Bernstein polynomial function.

Because we are proving the existence of standard model Higgs boson, rather than a random particle with certain invariant mass, the signal in each category should follow the certain ratio as shown in the table. Therefore, what we should do here is the simultaneous fit. The parameters of signal's distribution are determined by Monte-Carlo simulation, as a function of m_H. The background parameters and the numbers of signals in each category are set to be free. Also, we set a parameter μ as a scale factor to describe the relative size of signals. The expression can be written as Equation(1).

$$\mu \begin{pmatrix} sig_1 \\ sig_2 \\ ... \\ sig_{10} \end{pmatrix} + \begin{pmatrix} bkg_1 \\ bkg_2 \\ ... \\ bkg_{10} \end{pmatrix} = \begin{pmatrix} N_1 \\ N_2 \\ ... \\ N_{10} \end{pmatrix} \tag{1}$$

When μ is 0, it corresponds to the background-only

* We appreciate tremendous help from Prof. Yury Kolomensky and Prof. Haichen Wang. The data from ATLAS were kindly provided by Prof. Wang.

model; when μ is 1, it will obtain what exactly standard model theories expect. In the signal-plus-background model, μ is a free parameter. Then we can perform the simultaneous fit on each model and calculate their negative-log-likelihood (NLL). The relationship between statistics significance and NLL is shown in Equation(2).

$$Z = \sqrt{2(NLL_0 - NLL_S)} \tag{2}$$

For each m_H, we can calculate the significance of the existence of Higgs boson of this invariant mass. Therefore, we can utilize a loop to obtain the relation between Z and m_H, so as to determine the invariant mass of Higgs boson.

III. ANALYSIS

A. Simultaneous Fitting

Because there are many free parameters in the simultaneous fitting, we need to give a reasonable initial set of values. Therefore, we perform fitting in each category separately, in order to get the proper ranges of parameters for each background model, and do the fitting in background-only model and signal-plus-background model respectively.

Next, we plug into the values as mean we get from separate fit, and give a reasonable range for them to vary. In background-only model, μ is set to be 0, and the number of background entries is also fixed. In signal-plus-background model, μ is variable, and the number of background entries is assumed to be $N - \mu \cdot N_S$.

As reported in article[1], the most significant m_H is 126.0 GeV and 127.0 GeV for data in 2011 and 2012 respectively. Therefore, we perform the process decribed above at this m_H first, which is to say, the mean mass of signals, as well as other parameters of signal model, are fixed. The fitted distribution of signal-plus-background model using simultaneous fit is shown in Figure I, taking category 3 in 2011 for example. The other figures are similar.

Using the fitting result and formula(2), we calculate the significance of the existence of Higgs boson with invariant mass 126.0 GeV is 3.49, using data in 2011; that with invariant mass 127.0 GeV using data in 2012 is 3.35. The two values in article[1] is 3.4 and 3.2 respectively. The relative difference is about 2.7% in 2011 and 4.7% in 2012, which is within our expectation.

The parameter of interest μ is fitted to be 2.2 ± 0.6 for data in 2011 and 1.7 ± 0.5 for data in 2012. In article[1], these values are 2.2 ± 0.7 and 1.5 ± 0.6 respectively. Combining the two results, the value of μ is determined to be 1.9 ± 0.4.

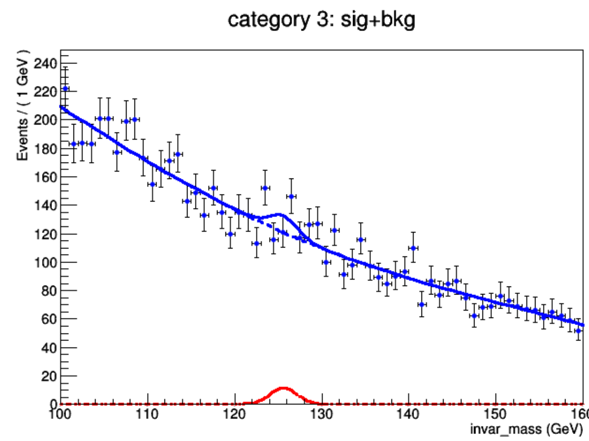

FIG. 1. The distribution of the invariant mass of diphoton based on a signal-plus-background model simultenous fitting. The data in this figure comes from category 3 in run 2011 only. The data points with their error bars are raw data; the blue solid line is the combined model; the blue dashed line is the backgorund; the red line is the signal.

B. Weakened Version Fittings

because doing simultaneous fitting is rather time-consuming, and what we do here is just a rough analysis. That is to say, we have ignored some factors, so even we obtain very perform accurate fitting, the small uncertainty of result will be enlarged significantly. Therefore, there is no harm to weaken our fitting process to a less accurate, but faster version. Here we performed two weakened fitting — binned fitting and simplified-model fitting.

In binned fitting, we decrease the number of data point so as to shorten the calculation time in simultaneous fit. According to the number of entries in each category, we respectively set 15 bins for category 2,6,10; 30 bins for category 1,4,5,8; and 60 bins for category 3,7,9. In this case, there is only one data point in each bin, thus we set the weight of each bin to be its original amount of contents. We run the binned fitting and find the results do not change a lot, but the running time is still too long.

Then we try to decrease the free parameter of fitting instead of the number of data points. Therefore, we simplified the background model. The model of most parameters is the Berstein polynomial, which corresponds to category 3 and 7. We try to use the other two types of model to fit them. Finally we decided to use an exponential distribution for category 3 and an exponential of second-order polynomial for category 7. We perform the simplified-model fitting and obtain the results quickly enough. Also, the results do not change a lot compared to the original simultaneous fitting.

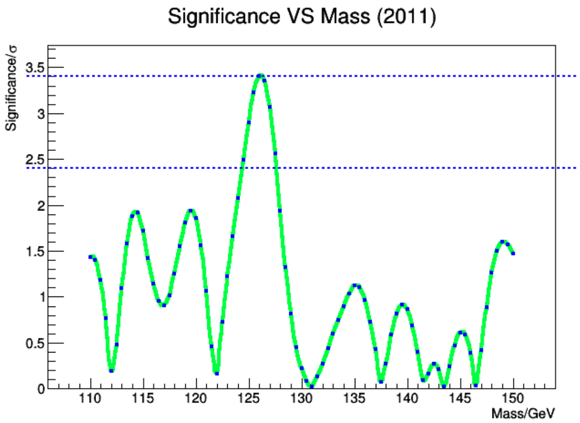

FIG. 2. The plot of significance versus mass for data in 2011.The black dots are calculated points and green line is the plot based on dots. The upper blue line correspond to the significance of $m_H = 126.0$ GeV, while the lower one is that minus one, thus gives the one-sigma range of the determined mass.

C. Determining the Invariant Mass of Signals

Using a loop, we can perform simplified-model fitting for a list of m_H, and calculate their significances. The parameters for signals are fitted as polynomial functions of m_H.

We calculate significance of m_H ranges from 110 GeV to 150 GeV. Figure 2 is for data in 2011. We can see there is a most significant peak with m_H around 126.0GeV. Then we perform finer scanning in its neighborhood, as shown in Figure 3. Therefore, based on data in 2011, the invariant mass of standard model Higgs boson is determined to be $126.1^{+1.6}_{-1.8}$ GeV.

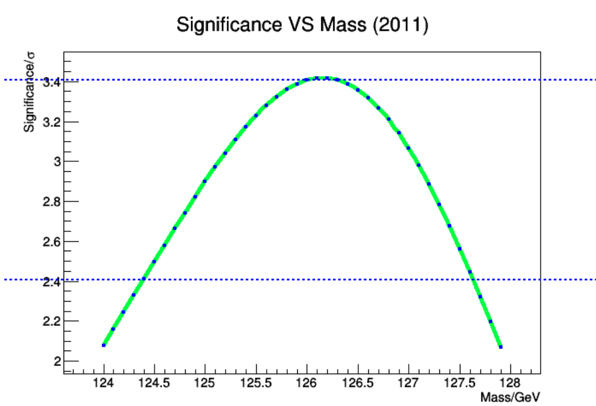

FIG. 3. The finer scanning of the plot in Figure 2. The elements in this figure are similar to those in Figure 2. The mass is set to be accurate to one decimal in GeV. Therefore, we determine the mean value of invariant mass to be 126.1 GeV. The lower and upper limits are determined to be 124.3 GeV and 127.7 GeV.

Similarly, we determined the invariant mass based on data in 2012 to be $127.0^{+2.1}_{-1.7}$ GeV. Combining the results for each year together, the determined value of the invariant mass is $126.5^{+1.3}_{-1.2}$ GeV.

IV. CONCLUSION

Although we have performed a rather rough data analysis, the results we have obtained are in accord with those in article[1]. Our final results are $\mu = 1.9 \pm 0.4$, $m_H = 126.5^{+1.3}_{-1.2}$ GeV, and proof of the existence of standard model Higgs boson on a basis of over three sigma. However, the uncertainty here is statistical only, because we do not have access to the data used to determine systematic uncertainties in ATLAS experiment. Considering those uncertainties as well as the systematic uncertainty from our simplified model, the error of the values may be obviously larger, but the mean values will not change too much since these uncertainties are not biased significantly.

Therefore, one way to improve our data analysis is to include the uncertainty of the simplified model, in order to give a more convincing and complete result. The other improvement may lies in that we do not need to simplify the model, but employ some optimizations in algorithm to speed up the calculation. Moreover, although we lack theoretical knowledges about standard model, it concerns us the physical meaning of the sub-peaks as shown in Figure 2. Maybe further analysis can spot some certain particles there.

V. REFERENCES

[1] ATLAS Collaboration, Phys.Lett.B 716, 1-29(2012)