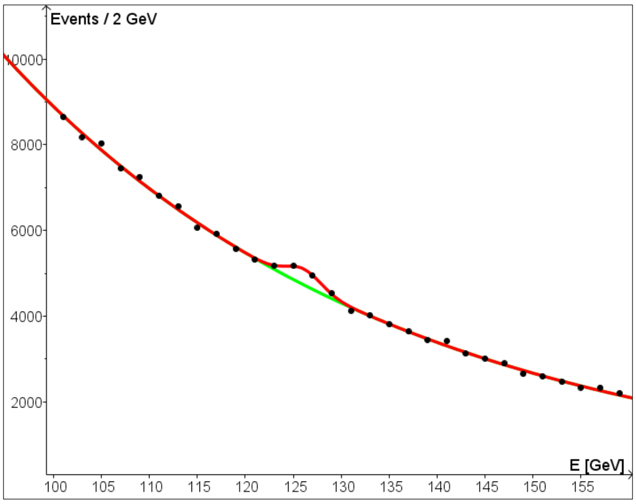
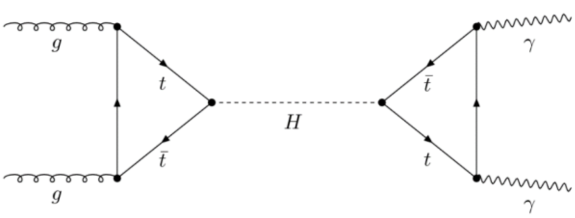
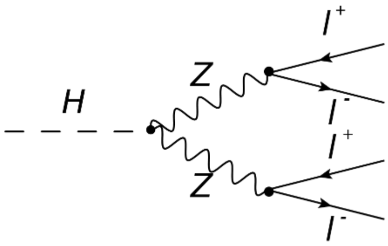
Analysis of CERN's 2012 Boson Signal







Author:
Benjamin Estermann, 6b
Im Hubel 4
6210 Sursee

Supervisor:
Dr. Jerzy Sromicki
Kirchstrasse 9
8107 Buchs, ZH

1 Index

2 Introduction

In July 2012, the CMS experiment and the ATLAS experiment announced discovery of a new particle which is consistent with the Standard Model (SM) Higgs boson[1]. Although different researchers developed the mechanism at the same time, it was named after Peter Higgs.

In March 2013, during the "Rencontres de Moriond" in Italy, both experiments published their latest results produced with 2.5 times more data. Therefore, there is a unique situation and I wanted to take this opportunity to work on it. What I wanted to do was to analyse this data with own statistic methods which are less complex than the ones from the CMS and ATLAS experiments. I also wanted to learn something about the functioning of the detectors and the process of the experiments, but not as first priority. Because of this, I joined two workshops of International physic Master classes at the University of Zurich (UZH) and at the University of Bern to get into the two different experiments and extend my knowledge about particle physics. It is clear to me that I cannot reproduce the enormous experiments at the CERN, but I want to reproduce the last steps of the data analysis in an own way and compare my results with the official results of the two experiments.

3 Higgs boson[1]

The Higgs boson, also called "the God particle" in some press reports, was postulated 1964 by three groups of physicists (inter alia Peter Higgs, François Englert and Robert Brout) and its discovery was awarded 2013 with the Nobel Prize in Physics.
The Higgs boson is part of the Standard Model of particle physics. It is the interacting particle of the Higgs field. This mechanism explains why some fundamental particles have mass. It also explains why the weak force has a much shorter range than the electromagnetic force. Even if the Standard Model predicts the Higgs boson, it does not say how heavy it is. However, for a given Higgs mass, the Standard Model predicts how it decays into known particles and how often it is produced in particle collisions.

Because of its relatively high mass, the Higgs boson has a very short lifetime and travels only a very tiny distance. Thus there is no chance to detect the Higgs boson directly, but it is possible to detect its decay products. To find this boson, the Large Hadron Collider was built, able to create Higgs bosons and other particles by high-energy proton-proton collisions for observation and study.

To detect the decay products of the Higgs boson and other particles, two huge detectors of large scientific collaborations (ATLAS and CMS) were developed and placed underground at the accelerator ring. With its different concepts behind the construction of the detectors, they track the particles and measure the energy, electric charge and momentum of them. That allows conclusions about the properties of the original particles, such as a Higgs boson.

2012 CERN announced discovery of a new Higgs like particle, in March 2013, they confirmed that it is really likely to be a SM Higgs boson, such it fulfils fundamental attributes like having a positive parity and spin zero, however there is more data needed to find out if the new particle really matches al predictions of the Standard Model.

4 Data collection

Because there was no possibility to get the exact data of the experiments directly, I had to measure them out of the plots by a simple method. I printed the plots on a paper in A3 size and then I measured the points and calculated the real worth in GeV of the points.

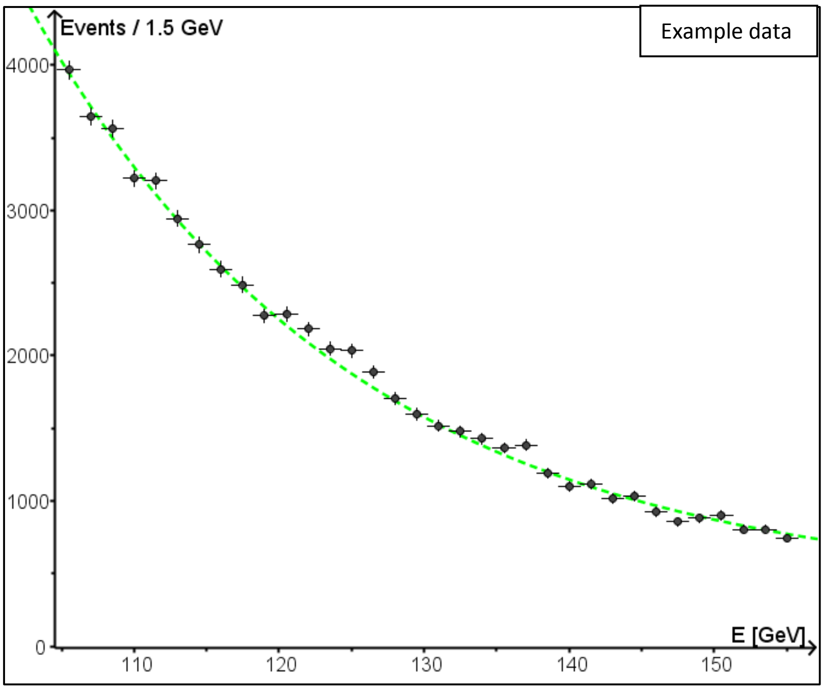


Fig. 1 This Plot shows the events per 1.5 GeV, the green line is the fitted background. The points here are all recalculated from the official CMS Plot[2].

To prove that there is no big difference between my recalculated data and the official data, see Fig. 2 and Fig 3.

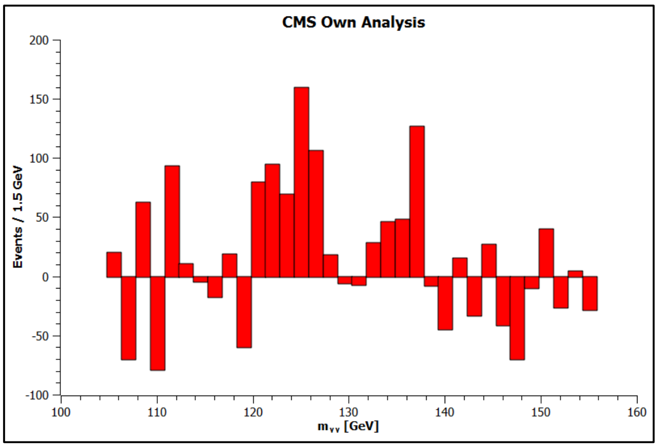


Fig. 2 The background residuals from my own analysis. The events were subtracted by the own background fit. (Graphic made with SciDAVis)

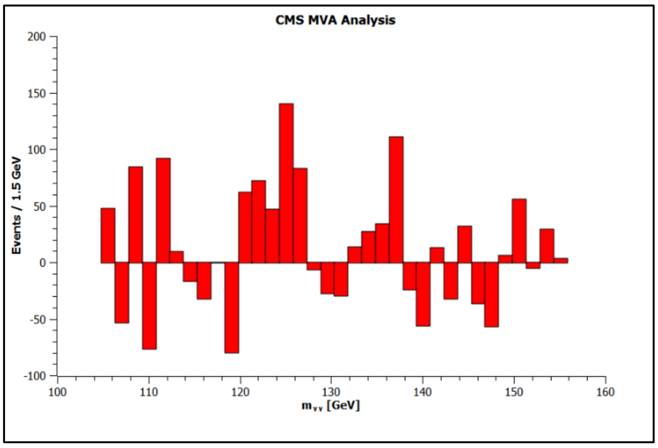


Fig. 3 The background residuals from the official CMS MVA analysis data. (Graphic made with SciDAVis)

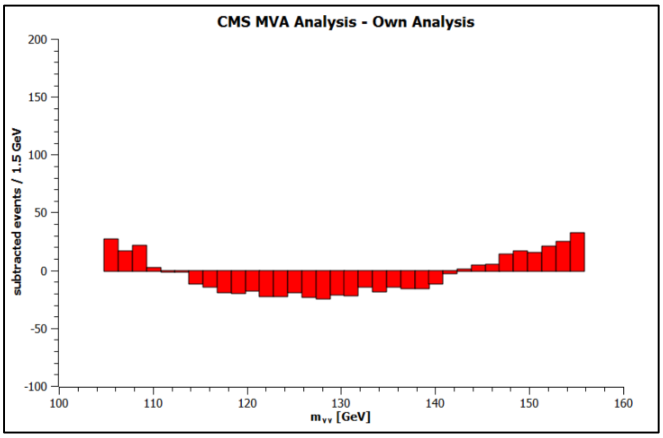


Fig. 4 CMS MVA analysis – own analysis. (Graphic made with SciDAVis)

It is visible in Fig. 4 that there is no big difference between the official analysis of the CMS experiment and my own analysis. The main part of the deviations is caused by the different background fit method, but even this part is very small. Only the small uncertainties are caused by the reading procedure, but these differences are really far less than one percent of the original value. This leads to the conclusion that the "reading-out procedure" has no influence to the final results.

5 Methods

5.1 Fitting procedure

5.1.1 Background fit

To reproduce the background, a three parameter exponential function was used in the form:

$$N = N_0 \exp\left(-\frac{E - E_0}{\tau}\right) + C_0$$

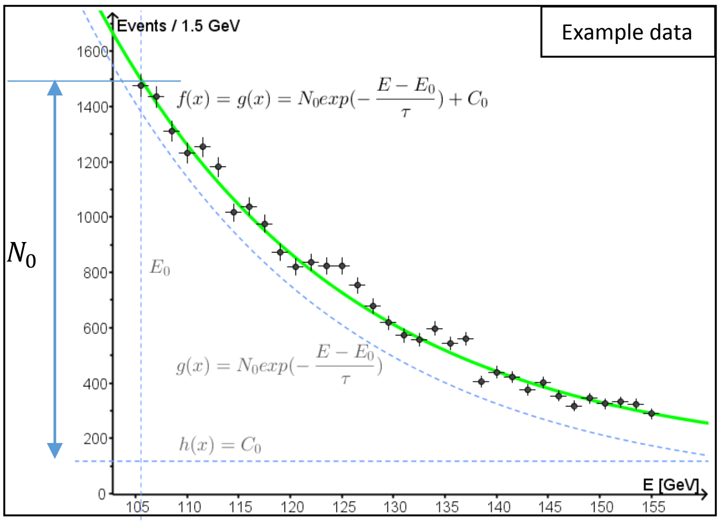


Fig. 5 Splitting the background fit function in its different parts

Fitting the background with only three parameters results in an advantage to the CMS and ATLAS experiments who use at least 4 parameters. E_0 is not a parameter itself, it is defined by the energy of the first point.

5.1.2 Background residuals - Gaussian fit

To continue the background was subtracted and it was worked with the background residuals. To describe the distribution of the measured values, an adapted form of the Gaussian normal distribution was used. The Gaussian normal distribution is normally used to describe the distribution which is caused by many random factors and the random dispersion of measuring uncertainties[3]. The natural mass width of the Higgs boson itself is predicted to be only around 100 MeV, but the mass resolution of the detectors of the two experiments is not that exact. For this reason, the response function of the detector plays an important part in the normal distribution and sets the width of the adapted Gaussian curve. The height of the curve has to be adapted too, it is given by the number of counts and has to be fitted.

The normal Gaussian Curve:

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right)$$

The used function is:

$$g(E) = A\exp\left(-\frac{(E-E_0)^2}{2\sigma^2}\right)$$

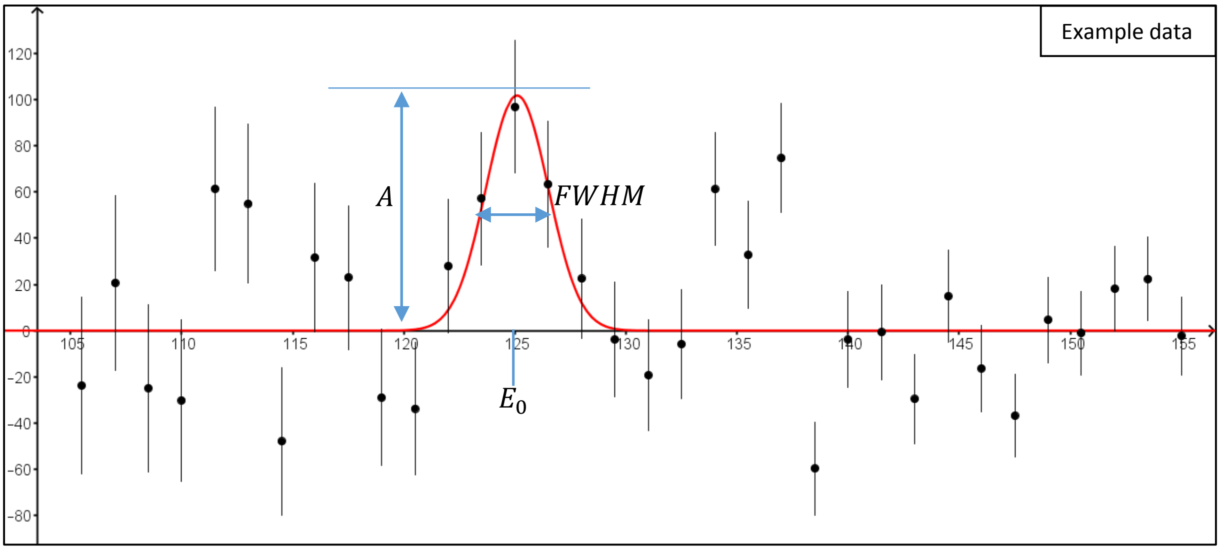


Fig. 6 The different parameters of the adapted Gaussian curve. The Full Width at Half Maximum (FWHM) is = 2.35 σ.

5.1.3 Least squares method

To test how good the fit is, the least squares method, also known as the chi-square distribution[4], was used. To perfect a fit, one has to minimize chi-square.

$$\chi^2 = \sum_{i=1}^{n} \left(\frac{(x_i - \mu_i)^2}{\sigma_i^2} \right)$$

x_i = number of events
μ_i = number of events predicted by the function which has to be fitted
σ_i = standard deviation

To know about the quality of the fit, chi square per degree of freedom was used[5]. This value takes into account the number of parameters and prevents a too good description of the data.

$$\frac{\chi^2}{\nu} = \chi^2_{DF}$$

ν = n – m (n = independent data points, m = used parameters)

For a good fit, χ^2_{DF} should be close to 1, what means that in average every point is not more than a standard deviation away from the fit. But one shouldn't be always concentrated on χ^2_{DF} and χ^2, it should be concentrated on the physics, and this two values allow to know where to focus.

5.1.4 Partly automation of the fitting procedure

To accelerate the fitting procedure I wrote a program. This program works with the background residuals. It calculates the best-fitting amplitude using the least squares method and also the corresponding χ^2 per degree of freedom (χ^2_{DF}) as function of the mass. The width of the adapted Gaussian curve was fixed, because it is given by the response function of the detector. The program was written in the programming language "Processing"[6]. The code of the whole program is available in the appendix (see chap. 13.2). Because of problems with GeoGebra, every value had to be calculated without decimal places. The decimal places were calculated back in GeoGebra afterwards.

```
for(x0 = xwerte[0]; x0 <= xwerte[xwerte.length-1]; x0 += 0.1){
  HilfsHi = 100;
  HilfsA = 300;
  Hilfsx0 = 0;
  for(A = -150; A < 500; A += 0.1){

    if( hirechnung(A, x0) < HilfsHi){
      HilfsHi = hirechnung(A, x0);
      HilfsA = A;
      Hilfsx0 = x0;
    }

  }
  As = append(As, HilfsA);
  His = append(His, HilfsHi);
  x0s = append(x0s, Hilfsx0);
}
```

Fig. 7 In this two for-loops, the best-fitting χ^2_{DF} and amplitude for every position of the curve were calculated and saved in an array.

5.2 Statistical significance

To see how significant a certain excess is or if it is just randomly created, one can calculate the number of standard deviations. It is the ratio between the number of counts and its standard uncertainty. It is a sign how reliable the signal is. 3 standard deviations is called an evidence, 5 a discovery.

$$S = \frac{C}{\sigma} = \ Number\ of\ standard\ deviations\ of\ the\ excess$$

C = number of counts
σ = standard deviation of the counts.

5.2.1 Simple estimation

In case of high background (for example γγ channel in CERN's experiments), the excess of counts C is much smaller than the background level B. Since C arises as a difference between total, T, and background counts, the number of background events B dominates the standard deviation of C. Therefore, $\sigma \approx \sqrt{B}$. This allows for a simple estimation of the significance of the signal by setting left/right markers defining the limits of the excess of the counts and calculating:

$$S \approx \frac{T - B}{\sqrt{B}} = \frac{C}{\sqrt{B}}, \qquad C \ll B, T$$

However, disadvantage of this simple method is its sensitivity to the marker setting. This is particularly acute in case when it is not clear if certain bin at the edge of the signal should be attributed to the peak or not.

In order to avoid such ambiguities I decided to work with counts, which were determined by integrations over the full available spectrum. This was achieved by fit procedures discussed below.

5.2.2 Integrated counts

The number of counts is the integral of the Gaussian curve divided by the bin width. The integral of the normal Gaussian curve is 1:

$$\int_{-\infty}^{\infty} f(x)dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(x - x_0)^2}{2\sigma^2}\right) dx \ = 1$$

Our curve is $A\sqrt{2\pi}$ times higher and σ times wider than the normal Gaussian curve:

$$A\sqrt{2\pi}\,\sigma * f(x) = A\ \exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right)$$

Therefore:

$$\int_{-\infty}^{\infty} A\ \exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right)dx = A\sqrt{2\pi}\,\sigma * \int_{-\infty}^{\infty} f(x)\,dx$$

Consequently:

$$\int_{-\infty}^{\infty} A\ \exp\left(-\frac{(x-x_0)^2}{2\sigma^2}\right)dx = A\sqrt{2\pi}\,\sigma * 1$$

So to calculate the number of counts the following formula was used:

$$N = \frac{A\sqrt{2\,\pi} * \sigma}{B}$$

B = bin width

This method was also tested by adding the different contributing bins. This leaded to an excellent agreement with the integral method.

5.2.3 Uncertainty of number of counts

A special method was used to get the uncertainty of number of counts. The amplitude was increased or decreased to increase chi square + 1. With this method, you get the uncertainty of the amplitude and therefore the uncertainty of the number of counts. The idea behind this method is to take into account the fluctuations of the signal. If you then increase chi square by 1, you actually get the standard deviation of the whole excess.

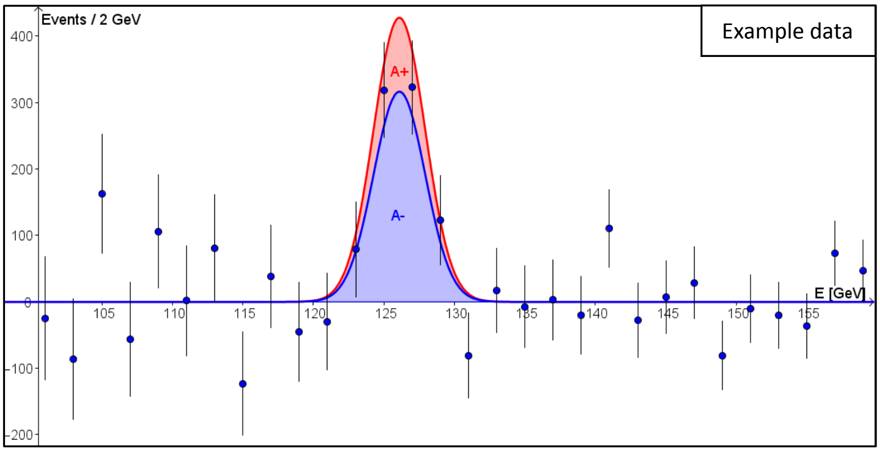


Fig. 8 The blue area is the integral of the decreased amplitude, the red area is the integral of the increased amplitude

There is also an uncertainty of the background which influences also the number of counts. For this reason the background level was also increased and decreased to increase chi square + 1. With this method you also take into account the fluctuations of the whole background. If the background fluctuates strongly, this will increase the uncertainty of the background and consequent the significance of the excess.

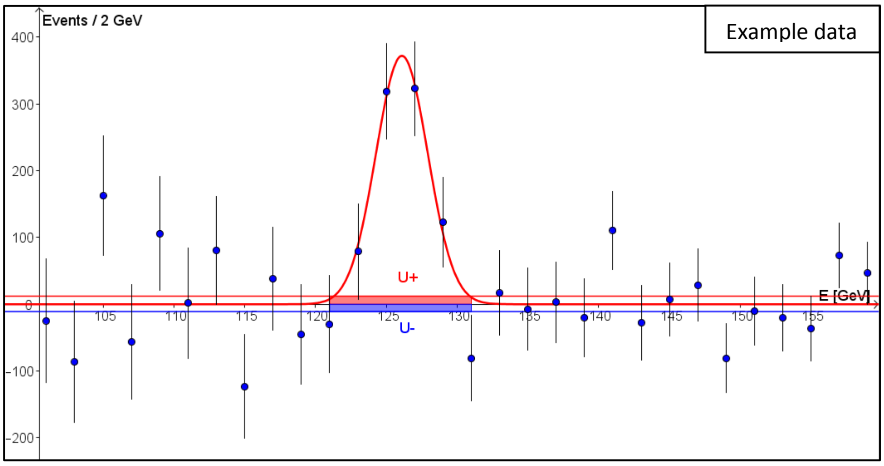


Fig. 9 The blue rectangle is the decreased area, the red rectangle is the increased area

This two deviations are then added quadratically to get the uncertainty of the number of counts:

$$\sqrt{\sigma_1^2 + \sigma_2^2} = \sigma$$

It turned out that the background uncertainty has a much smaller influence to the total uncertainty than the one of the amplitude.

5.3 Mass of the boson and its uncertainty

To calculate the mass of the boson and its uncertainty nearly the same method was used as for the uncertainty of the background and the amplitude. The mass is given by E_0 of the best fitting Gaussian curve. To find the uncertainty E_0 was increased and decreased to increase chi square by 1.

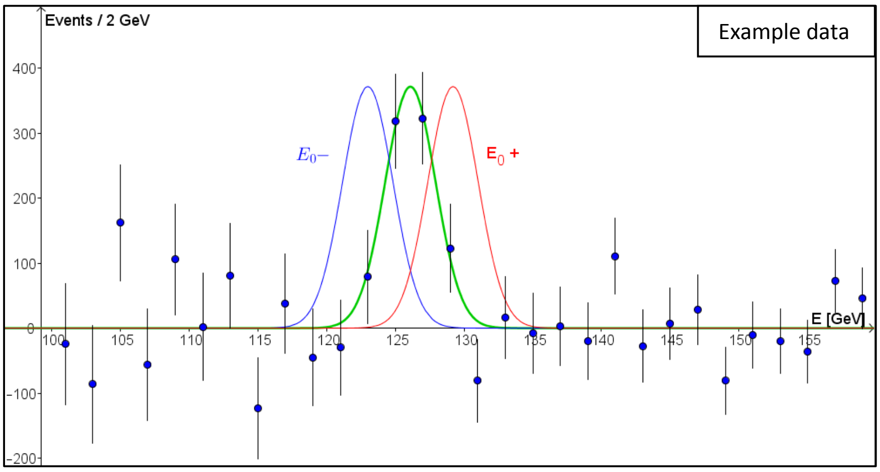


Fig. 10 The blue curve shows the decreased E_0, the red curve shows the increased E_0. For better visibility, E_0 was increased and decreased 4 times more than statistically necessary.

6 Gamma+Gamma Channel

The Higgs boson can decay in very different particles[7], also into two photons. It seems ironic that only 0.2% of the created Higgs bosons will decay into two photons, but in this channel, it is a lot easier to separate Higgs events from other events. There rests a big background but also a relatively strong signal. This decay channel is important for the analysis too, because it has a very good mass resolution.

6.1 Structure of detectors

Most important for this decay channel is mainly the electromagnetic calorimeter and its high mass resolution to recalculate the mass of the photons and consequent of the Higgs boson. Different parts like the hadronic calorimeter and the myon chambers are not that important in this decay channel.

6.2 Atlas New Data

Basis for this analysis are the data from the Atlas experiment with 25.5 * 10^{12} proton-proton collisions (25.5 fb^{-1}). These data were presented in March 2013 at the "Rencontres de Moriond"[8] and published shortly after that.

6.2.1 Background

To fit the background the exponential function was used with the following parameters:

N_0 = 8592.4
T_0 = 40.76
C_0 = 87.3

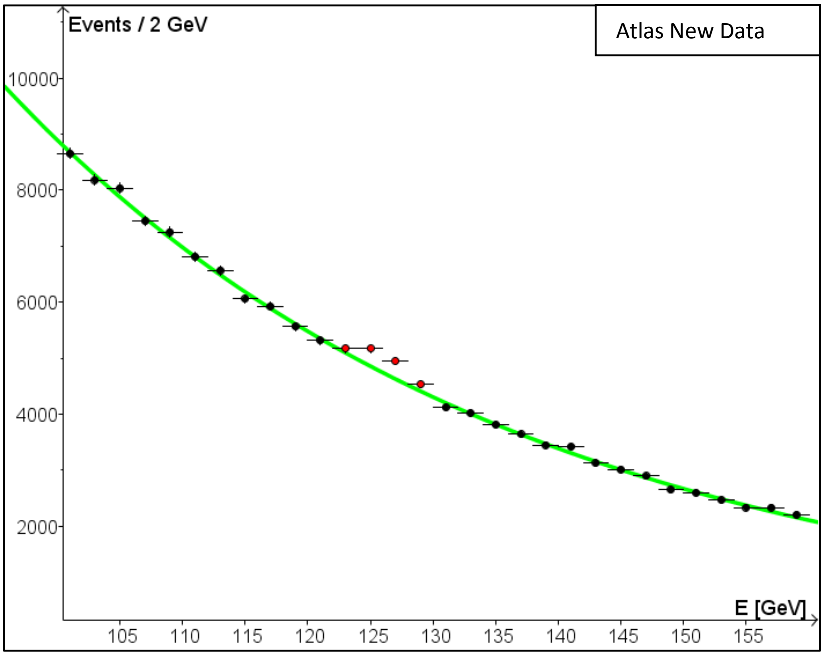


Fig. 11 The exponential background fit function. The red points were excluded for the fit to don't weak the boson signal. If these points include a signal, they wouldn't be part of the background.

This fit with χ^2_{DF} = 2.52 shows that there must be a signal that does not fit with the background. The biggest deviations are in the area of 125 to 129 GeV. Outside this area, there are no big deviations and no hint for a signal outside the above-named area.

6.2.2 Background residuals - Gaussian fit

For the next step of the fitting procedure, the background was subtracted and the background residuals were used to continue.

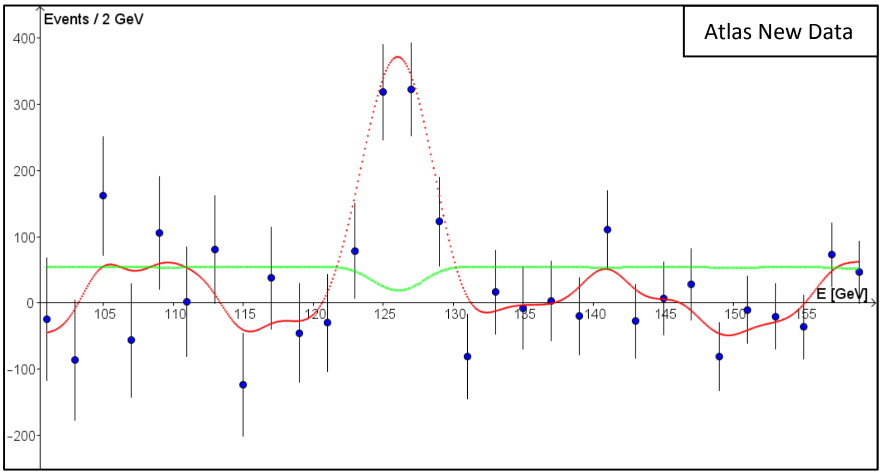


Fig. 12 The red points show the height/amplitude of a Gaussian curve as function of the energy, fitted with the least squares method too. The green line shows χ^2_{DF} of the Gaussian curve as function of the energy.

Visible in Fig. 12, there is a clear improvement of χ^2_{DF} visible in the area between 125 and 127 GeV. Also visible in this area is a clear excess of events. The parameters of the adapted Gauss curve, which describes this excess, placed at the best-fitting position, are:

A = 371.8 Events
E_0 = 126.1 GeV
σ = 1.83 GeV

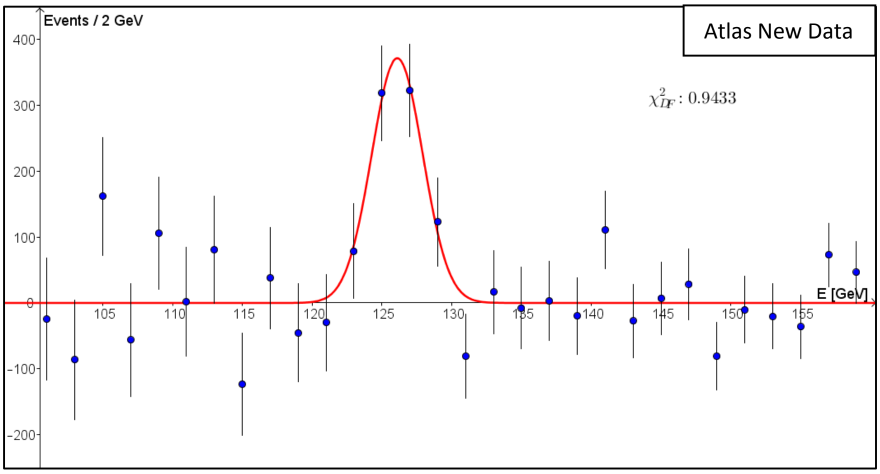


Fig. 13 The best fitting Gaussian curve describes the data perfectly with χ^2_{DF} = 0.94.

With this function χ^2_{DF} could be decreased from 2.52 to 0.94 what means that the fit describes the data perfectly and there are no signs to think about a different placement of the Gaussian curve.

6.2.3 Statistical significance

Simple estimation (see 4.2.1) leads to $C = 844$ and $B = 18987$. Therefore estimated significance $S = 6.1$. This is a beautiful, strongest evidence of the Higgs boson observed up to now (Fig. 14).

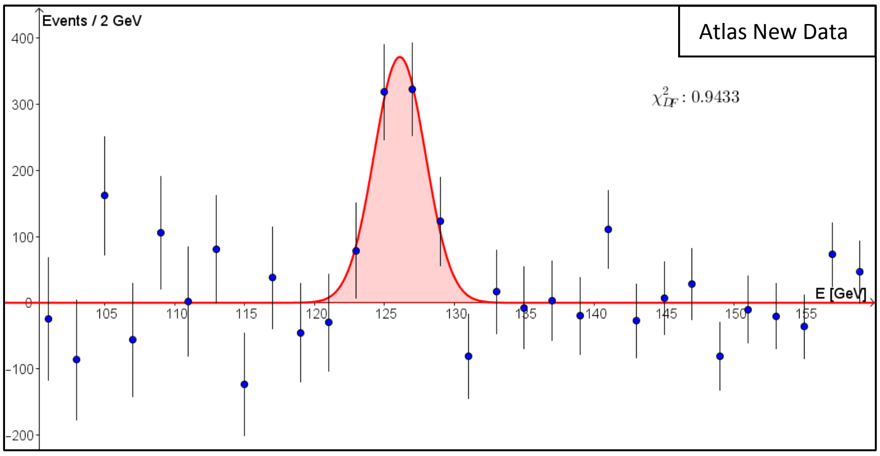


Fig. 14 The integrated function leads the number of counts.

There are $\frac{A\sqrt{2\pi}*\sigma}{B} = 853$ counts calculated with a statistical uncertainty of the amplitude ±128 counts and of the background ± 60 counts.

So there are 853 ± 141 possible boson counts what leads to a statistical significance of $\frac{853}{141} = 6.0$ standard deviations or a probability of $^1/_{5*10^8}$ that this excess was created randomly. This is a really strong result and a good improvement compared to the low statistic data of the Higgs boson discovery. There is no doubt that a particle exists which causes this excess.

6.2.4 Mass of the boson

The mass of the boson according to this experiment is calculated as 126.1 ± 0.4 GeV.

6.3 CMS Old Data

These data are the discovery data of the Higgs boson, published in July 2012[9]. This low statistic data contain 10.4 fb^{-1} Luminosity.

6.3.1 Background fit

The exponential function to describe the background has the following parameters:

N_0 = 1383
T_0 = 23.84
C_0 = 117

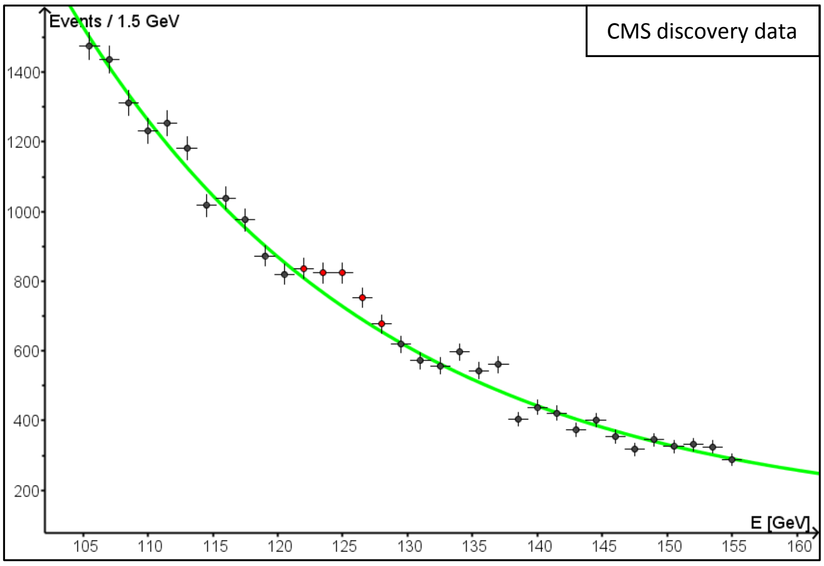


Fig. 15 The background fit function. The red points were excluded for the fitting procedure because they most likely contain a signal which is not part of the background.

The fit has a χ^2_{DF} of 2.39, what shows that there must be something what isn't described by this fit. The biggest deviations are in the area from 122 GeV to 128 GeV. This low statistic data are more instable than high statistic data, but the excess in the above-named area is clearly discernible.

6.3.2 Background residuals - Gaussian fit

The Gaussian fit with the background residuals showed one big excess mainly in the area of 125 GeV. Also only in this area an improvement of χ^2_{DF} is visible.

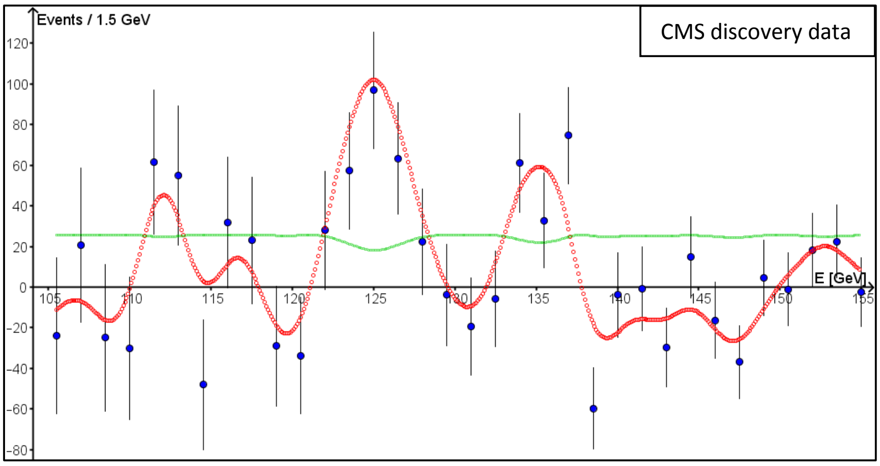


Fig. 16 The red points show the amplitude of a fitted Gaussian curve as a function of the energy. The green points show χ^2_{DF} as function of the energy.

The parameters of the best-fitting Gaussian curve to describe he excess has the following parameters:

A = 101.8 Events
E_0 = 125.1 GeV
σ = 1.43 GeV

σ is not fitted, it is given by the response function of the detector, so it was fixed during the fitting procedure.

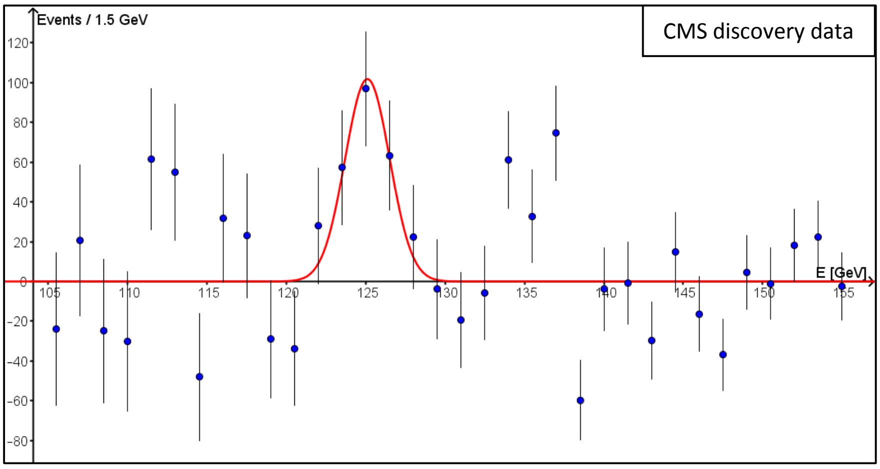


Fig. 17 The best-fitting Gaussian curve for this data.

χ^2_{DF} could be decreased from 2.4 to 1.8. That is a clear improvement, but there must be some deviations in the data which cause this χ^2_{DF}.

6.3.3 Statistical significance

Simple estimation (see 4.2.1) leads to $C = 269$ and $B = 3648$. Therefore estimated significance $S = 4.4$.

Because of the poor χ^2_{DF}, χ^2 had to be increased by 1.8 and not as usually by 1. If a curve fits the data not perfect, it also has a bigger uncertainty and this uncertainty has to be considered in the calculation to take into account the fluctuations.

The number of counts (See appendix: Fig. 33) were calculated as $\frac{A\sqrt{2\pi} * \sigma}{B} = 243.3$ counts. They have a statistical uncertainty of the amplitude of ±70.5 counts and of the background ± 23 counts.

Altogether there are 243 ± 74 possible boson counts, which are all part of the excess. This leads to a full statistical significance of 3.3 Standard Deviations or a probability of 0.1 % that this excess was created randomly.

6.3.4 Mass of the boson

The mass of the boson was then calculated as 125.1 ± 0.6 GeV, what is in very good agreement with the mass published in 2012 by the CMS experiment.

6.4 CMS CIC Analysis New Data

6.4.1 Background fit

Unfortunately there was no possibility for me to find the original background data in the given time for the CMS cut based analysis (CIC), so this analysis carried out only with the background residuals[10] and no own background fit. The biggest difference between the mass fit analysis (MVA) and the CIC analysis is that the CIC analysis contains 1.2 times more data than the MVA analysis, but the different datasets overlap partly[11].

6.4.2 Background residuals - Gaussian fit

The automatic fitting of the adapted Gaussian curve showed only one excess with a clear improvement of χ^2_{DF} in the area between 123 and 126 GeV. In the other mass spectrum there were no excesses with a clear improvement of χ^2_{DF} (See appendix: Fig. 34).

The curve with the best χ^2_{DF} is centred at 124.6 GeV with an amplitude A = 177.4 counts and σ = 1.5 GeV.

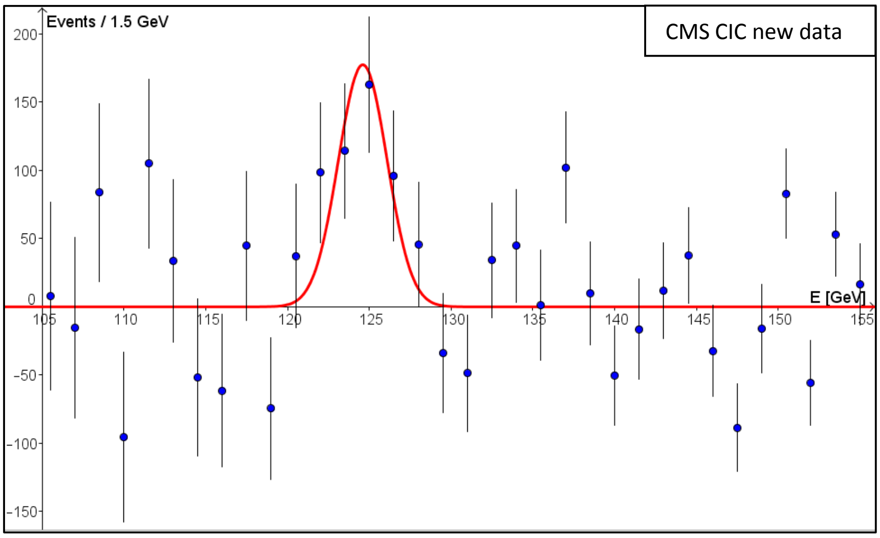


Fig. 18 The best fitting curve for the background residuals.

χ^2_{DF} of the fit without the Gaussian curve was 2.8, which is really bad and shows that there must be some excesses or signals in the data. With the adapted Gaussian curve, χ^2_{DF} could be decreased to 1.7. This is a clear improvement, even the fit doesn't describe the data perfect. It indicates that there are some other disturbances in the data. It could be based on other particles, what is really unlikely, or just be the result of bad statistics. With more data, it could be possible to erase this problem of the fluctuating background.

6.4.3 Statistical significance

The simple method could not be applied here because of unavailable background spectrum. The further analysis was performed with the available background-subtracted spectrum (Fig. 18).

Because of the poor χ^2_{DF} ($\chi^2_{DF} = 1.7$), χ^2 had to be increased by 1.7 and not as usually by 1 (see also chap 5.2.3).

The number of boson counts was calculated as $\frac{A\sqrt{2\,\pi}*\sigma}{B} = 442$ counts (See appendix: Fig. 35). They have a statistical uncertainty of the amplitude of ±121 counts and of the background ± 45.3 counts.

The statistical uncertainty of the background was very one-sided, what suggests that there could be a different weighting of the events. Because I have no specific information about this problem, I trust the official background fit from the CMS CIC analysis.

Overall there are 441.7 ± 129.2 boson counts, which leads to a statistical significance of 3.4 standard deviations or a probability of a little bit less than 0.1 % that this excess was created randomly. This statistical significance is pretty strong, but not as strong as expected.

6.4.4 Mass of the boson

The mass of the boson was calculated as 124.6 GeV ± 0.7 GeV. This is in excellent agreement by the published results of the CMS experiment for the CIC analysis (statistical significance = 3.9 std. dev. and m_H = 124.5 GeV[12])

6.5 CMS MVA Analysis New Data

The official published data of the CMS MVA analysis (24.7 fb^{-1}) served as basis for this analysis. [13]

6.5.1 Background fit

The parameter for the background fit function are:

N_0 = 3564.3
T_0 = 22.4
C_0 = 381.8

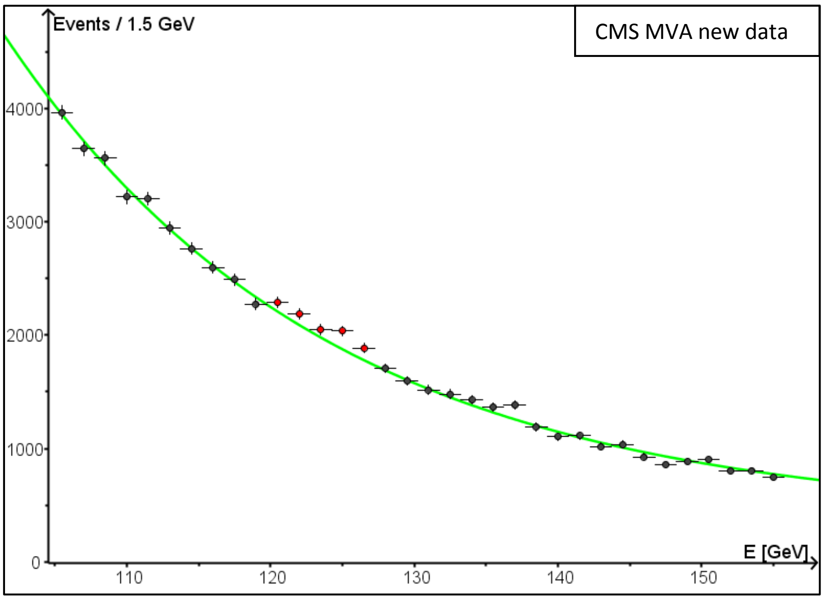


Fig. 19 The 'best fit' background function. The red points were excluded for the fitting procedure because they most likely contain a signal which is not part of the background.

χ^2_{DF} for this background fit is 2.2, what is really poor and shows that there must be some excess in the data. The biggest deviations are found in the area between 120 GeV to 126 GeV.

6.5.2 Background residuals - Gaussian fit

To continue the work the background was subtracted. The automatic analysis of the background residuals leads to the following result:

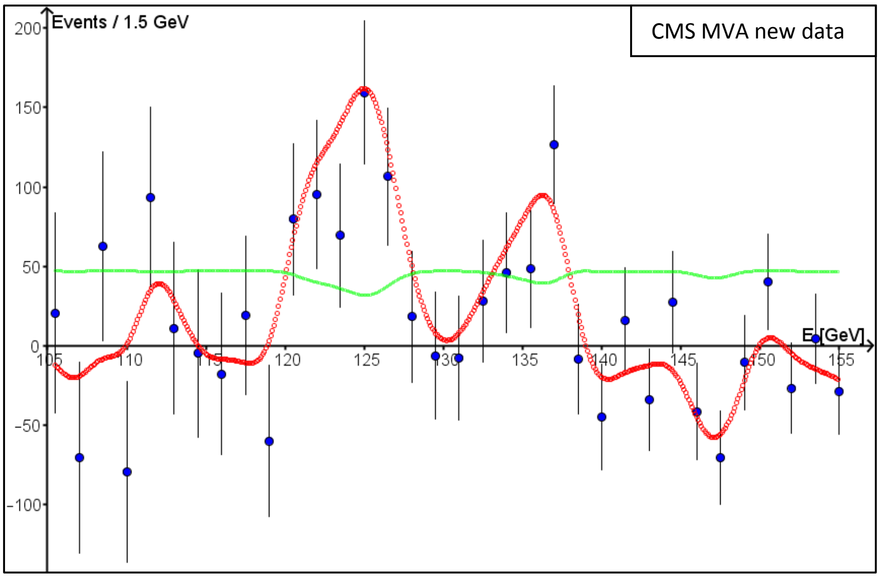


Fig. 20 Visible here is the amplitude (red points) and χ^2_{DF} (green points) as function of the energy. An improvement of χ^2_{DF} is visible in the area of 125 GeV

The best χ^2_{DF} is at 125.1 GeV. The adapted Gaussian curve has an amplitude A = 161.9 and a width σ = 1.43 GeV.

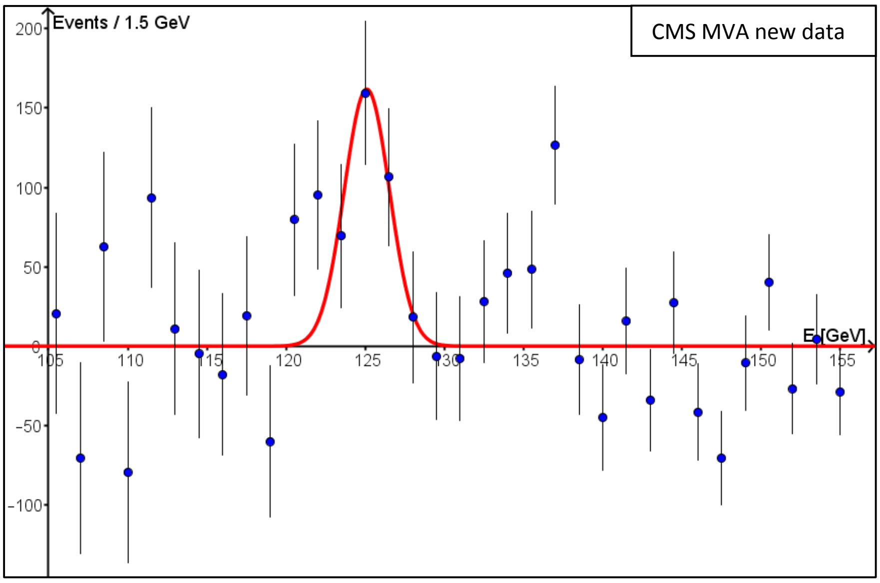


Fig. 21 The best fitting Gaussian curve, placed at 125.1 GeV with a χ^2_{DF} of 1.6.

With this Gaussian curve, an improvement of χ^2_{DF} could be achieved from 2.2 to 1.6. This is a clear improvement although there are other deviations in the data.

6.5.3 Statistical significance

Simple estimation (see 4.2.1) leads to $C = 354$ and $B = 7325$. Therefore estimated significance $S = 4.1$.

The number of counts were calculated as $\frac{A\sqrt{2\pi}*\sigma}{B} = 387$ counts. Because of the relatively poor χ^2_{DF} ($\chi^2_{DF} = 1.6$), χ^2 had to be increased by 1.6 (see also chap. 5.2.3).

The uncertainty of the amplitude is ± 103 counts and the uncertainty of the background is ± 39 counts. This leads to the final result of 387 ± 110 counts and a statistical significance of 3.5 σ. This would be a probability of under 0.1 % that this excess was created randomly.

6.5.4 Mass of the boson

In the CMS MVA analysis the mass of the boson was calculated as 125 GeV ± 0.6 GeV.

6.6 Overview of the statistical significances

The table below summarises the most important data concerning discovery of the Higgs particle (di-photon channel). I judge overall agreement of the results obtained with different data sets and methods as good. The combined statistical significance for all data in this channel amounts to 7 to 8 standard deviations. Here all methods are in very good agreement.

	$\chi^2 + 1$ **Method**	$\boldsymbol{S/\sqrt{B}}$	**Official results**
ATLAS new data	6.0	6.0	6.1
CMS old data	3.3	4.4	4.1
CMS new data MVA	3.5	4.1	3.2
CMS new data CIC	3.4	Not available	3.9

Table 1 Statistical significances of the different experiments calculated with different methods.

6.7 Extensions of the CMS MVA Analysis

For a description of the structure between 122 GeV and 128 GeV under assumption of Gaussian response function two solutions were found: One with FWHM = 3.4 GeV and the other with FWHM = 5.2 GeV (see Fig. 18). The values of χ^2_{DF} are 1.59 and 1.52, respectively. The statistical significance of the second solution is therefore slightly better, however at present I have no information if the corresponding increase of the width lies within the limits, which are allowed from the studies of the resolution of the electromagnetic calorimeter.

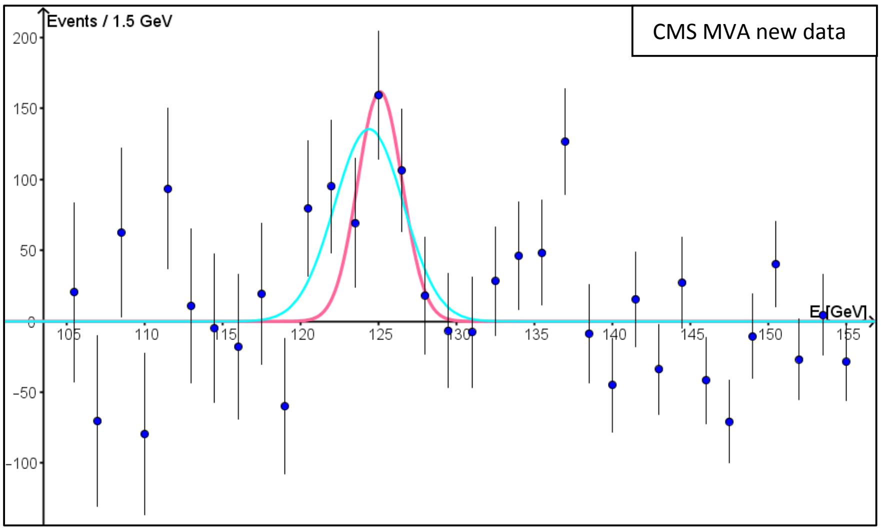


Fig. 22 The pink curve has FWHM = 3.4 GeV, the blue has FWHM = 5.2 GeV. The blue curve is a little bit transformed to the left side, it corresponds slightly better with the left energy shoulder of the peak.

6.8 Multiple Higgs signal?

At this point I take the freedom of speculating about possible attractive future developments of the experiments. They could be associated with theoretical scenarios 'beyond the Standard Model'. Extensions of the SM, e.g. Super symmetry, were considered within last decades in multitude of theoretical papers. In many of these models multiplets of Higgs bosons are predicted. I ask therefore a question: Do the present high statistics 25 fb^{-1} data contain any hint of more than one Higgs particle?

It can be noticed that the description of the background residuals of the CMS experiment by one boson hypothesis in the whole energy range (105.5 – 155 GeV) is not quite satisfying (see Fig. 23, χ^2_{DF} = 1.59).

By inspection of the data above (see Fig. 20) some excess of events around 135 GeV and an improvement of χ^2_{DF} in this area can be noticed. This feature is particularly pronounced in MVA data set. Is this excess strong enough to be a second Boson signal? What is the statistical significance of this excess?

6.8.1 Hypothesis: Two bosons

In order to investigate these questions, I make a hypothesis of presence of two new Bosons. With such a hypothesis, the background was fitted excluding the two expected Boson areas (Fig. 23).

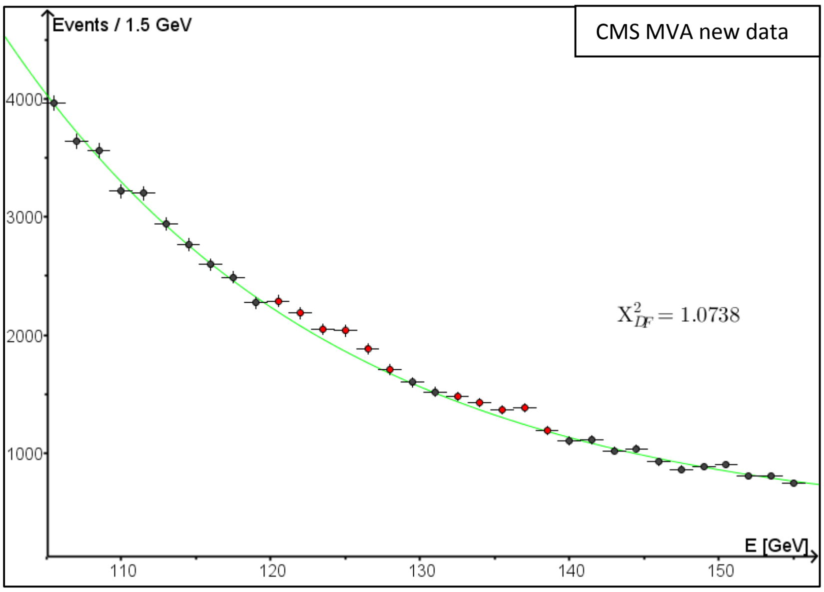


Fig. 23 The background fit excluded the red points. χ^2_{DF} of the background fit is 1.1 so there is no excess outside the excluded Boson areas.

An attempt to describe the data with two boson hypothesis (minimizing procedure described above, chap. 4) leads to the result which is shown in Fig. 24. A significant improvement in the quality is obtained (χ^2_{DF} has decreased from 1.59 to 1.26).

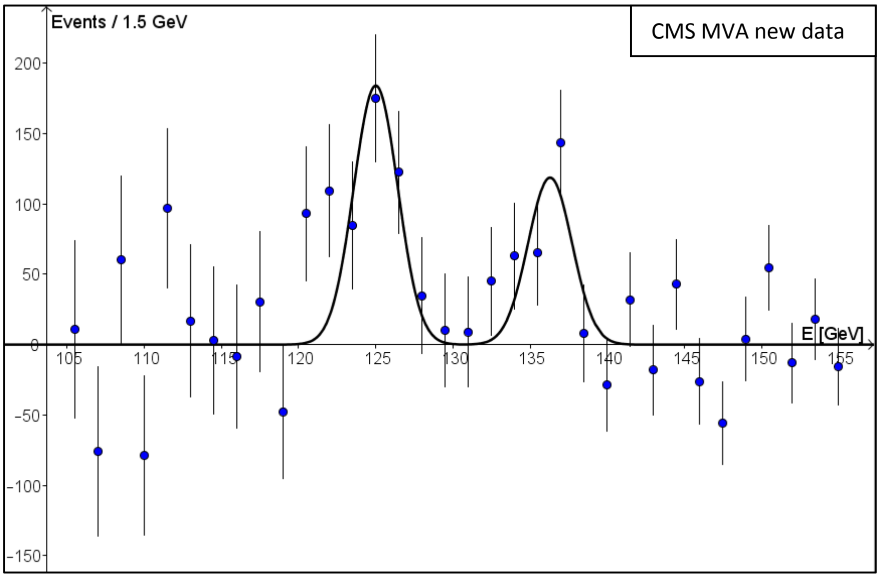


Fig. 24 Solid line: optimal description of the background residual data with two boson hypothesis. Amplitudes and positions of the two signals were varied to find the 'best fit'.
The width of both peaks was kept at the same value as in the CMS MVA analysis

The statistical evidence of the second peak located at 135 GeV is only about 2 standard deviations. This is a really weak evidence and the probability that this excess was created randomly is high. In the new data only the signal is hardly present in contrast to the older low statistics data.

In fact, the similar behaviour of the data could be already seen in the low statistics 10 fb^{-1} CMS data (CMS "discovery, July 4th 2012" data). Next picture shows the result of the corresponding analysis of this data set. χ^2_{DF} has improved from 1.81 to 1.50.
It is streaking that the ratio of the areas under both peaks have stayed approximately the same, even by increasing the data amount by factor 2.5.

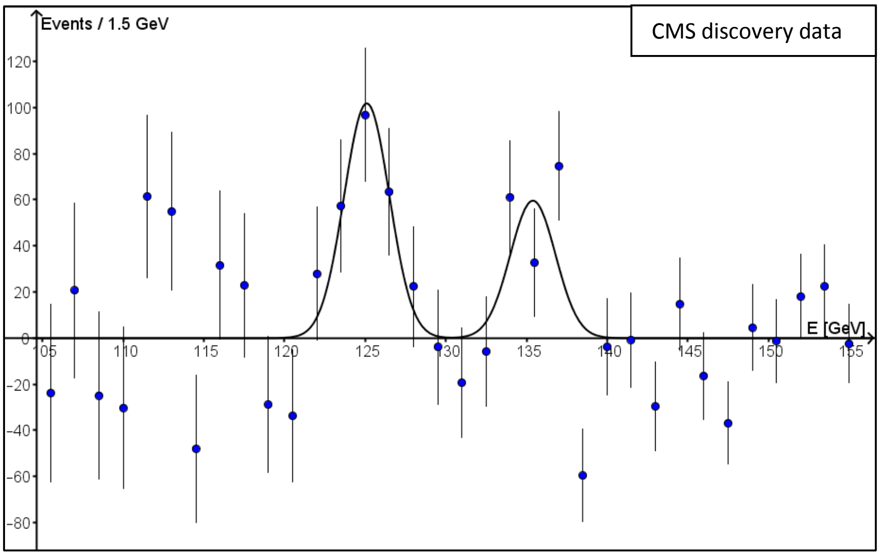


Fig. 25 Solid line: optimal description of the background residual data with two boson hypothesis. Amplitudes and positions of the two signals were varied to find the 'best fit'. The width of both peaks was kept at the same value as in the CMS "discovery data" analysis.

The statistical evidence of this peak located at 135 GeV is only about 2 standard deviations too. Because the statistical evidence of the signal in the high statistics data of the CMS MVA analysis is not higher, it is very likely that this excess will disappear in newer data.

6.8.2 Hypothesis: Three bosons

The description of these residual CMS MVA data can be further improved by speculating about a presence of a third boson signal, at the low energy shoulder of the 125 GeV peak. This leads to the result presented in Fig. 26. An excellent description of the data with χ^2_{DF} 1.10 is obtained

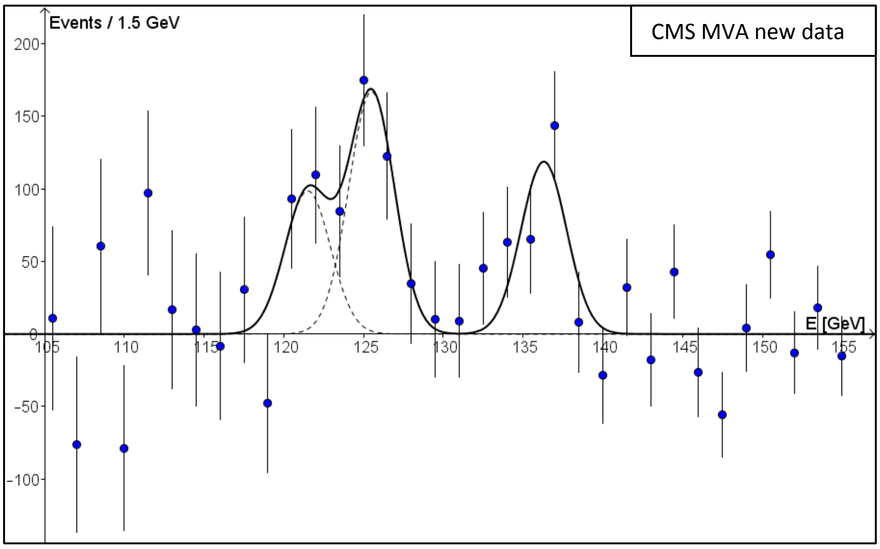


Fig. 26 CMS MVA experiment data with three boson hypothesis. The main peak is located at 125.5 GeV, the right peak at 136.3 GeV and the low energy shoulder of the main peak at 121.5 GeV.

However, one has to recognize that "several boson" hypothesis considered above is not compatible with for me available Atlas data, as well as with CMS CIC analysis. Also, the two satellite boson data have obviously less statistical significance than the main, 125 GeV peak. Therefore the speculations above should be taken with appropriate portion of precaution.
Nevertheless, from the point of view of further development of the experiments even a moderate improvement in the energy resolution of the photon detectors or in the resolution of the photon re-construction in the data analysis would be very interesting.

7 ZZ/4l Channel

The Higgs boson doesn't decay only in two photons, it can also decay in two Z bosons which decay then into four leptons. In this decay channel there is a small background but also a small signal. The knowledge about this background is really big and all the background is theoretically calculated. For this reason I can't do an own background analysis, but I can do the last step of the calculation and calculate the statistical significance and the mass of the boson.

7.1 Atlas experiment

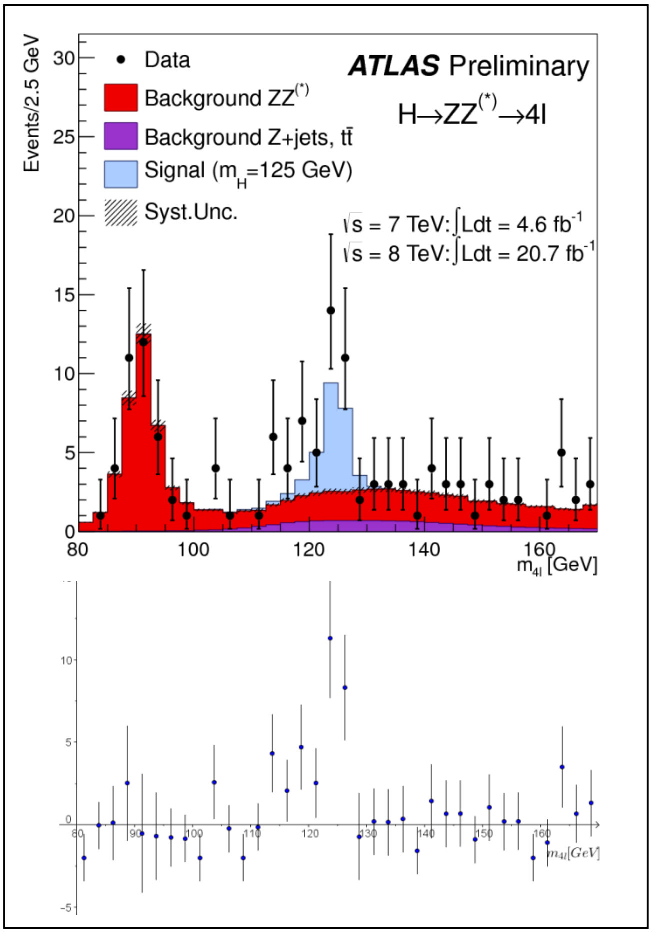


Fig. 27 Original plot from the Atlas experiment with 25.3 fb^{-1} data[14]. The peak on the left side is part of the background and helps to adjust the mass resolution. Below the background subtracted spectrum. The background was theoretically calculated.

7.1.1 Statistical significance

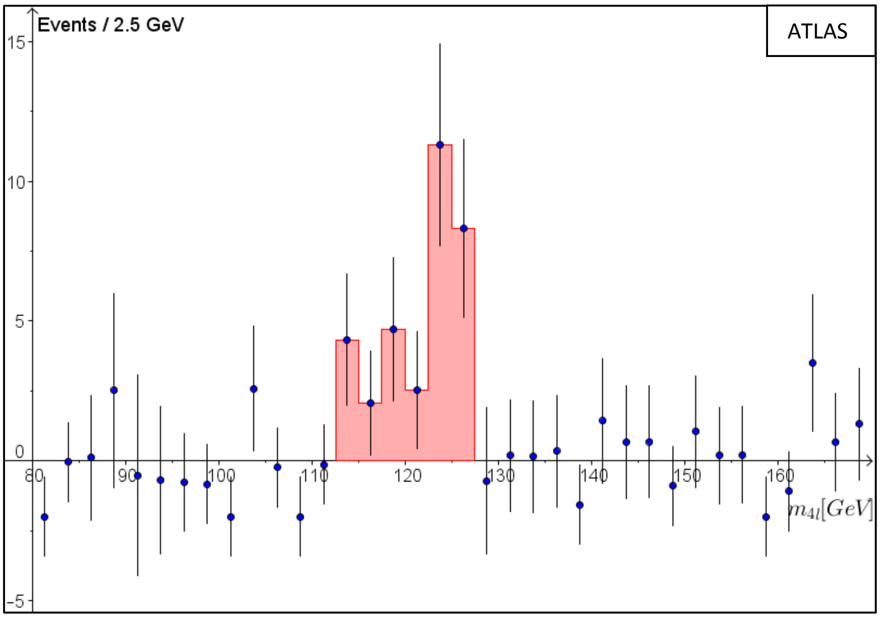


Fig. 28 Integrated histogram in the limits 112 to 128 GeV

The integrated histogram amounts to 33 events. The error of this was calculated by quadratic addition of the errors of 6 contributing bins. The statistical significance amounts to 5 standard deviations, which scales very god with 3.5 standard deviations obtained in the discovery announcement, which was obtained by 2.5 times less data.

Outside the limit, signal - background is = 0, what means that there is a good understanding of the sources of background events.

7.1.2 Mass of the boson

The following formula was used to calculate the mass:

$$< E > = \frac{\sum N_i \, E_i}{\sum N_i}$$

N_i = number of counts
E_i = energy

The mass of the boson was calculated as 123.7 GeV ± 0.6 GeV. The excess on the left side leads the mass smaller.

7.2 CMS experiment

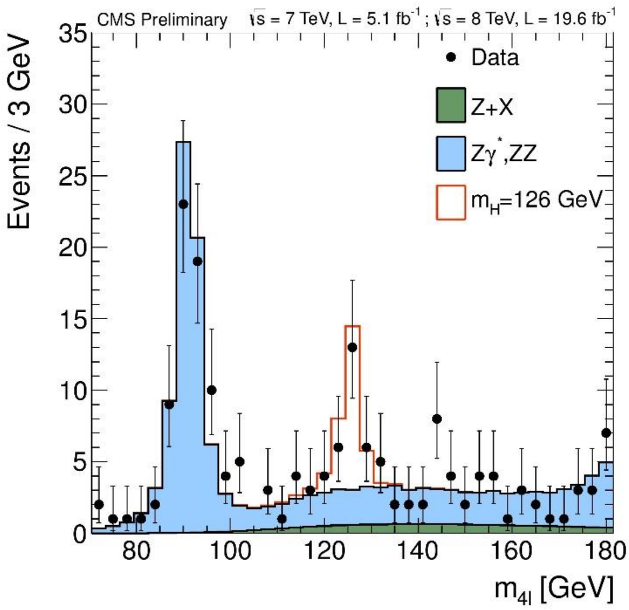


Fig. 29 Original plot from the CMS experiment of the mass range from 70 to 180 GeV[15].

Basis of this analysis are 25.5 fb^{-1} data, 2.5 times more than the “discovery data”.

7.2.1 Statistical significance

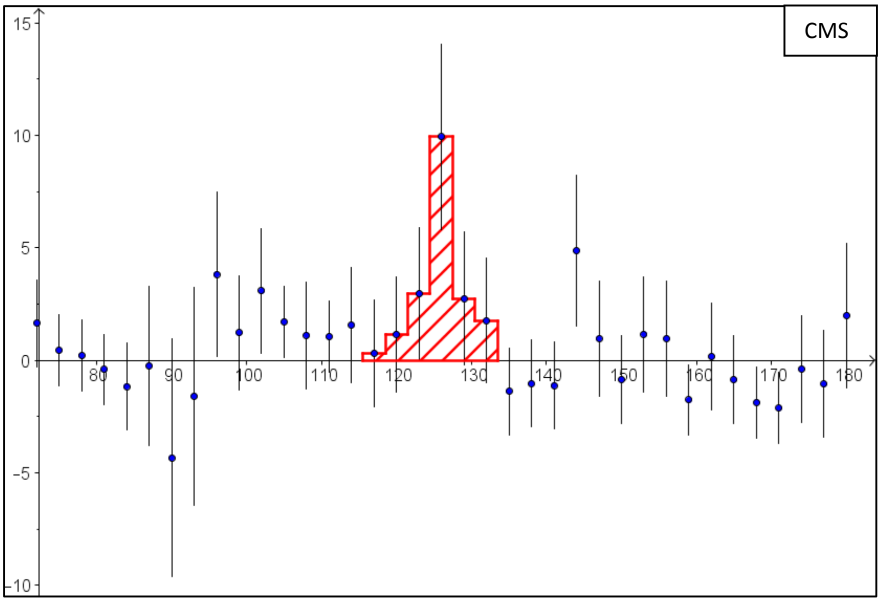


Fig. 30 Background subtracted data. The integrated histogram between 117 to 132 GeV shows the Higgs events.

The integrated histogram counts 19 events with an uncertainty of ± 6 events. This leads to a statistical significance of 3.3 standard deviations. This result is relatively weak. As opposed to the Atlas experiment, which had 1.7 times more events than predicted by the standard model, the CMS experiment had 10% less events than predicted by the standard model. For this reason, the statistical evidence of the CMS experiment is less strong than the one of the Atlas experiment in the ZZ/4l analysis.

7.2.2 Mass of the boson

The mass was calculated (see chap. 7.1.2) as 126.2 GeV ± 0.7 GeV, what corresponds perfectly with the official CMS announcement.

8 Combination of the channels

After all, I take the freedom to combine the results of the different channels to get one overall result.

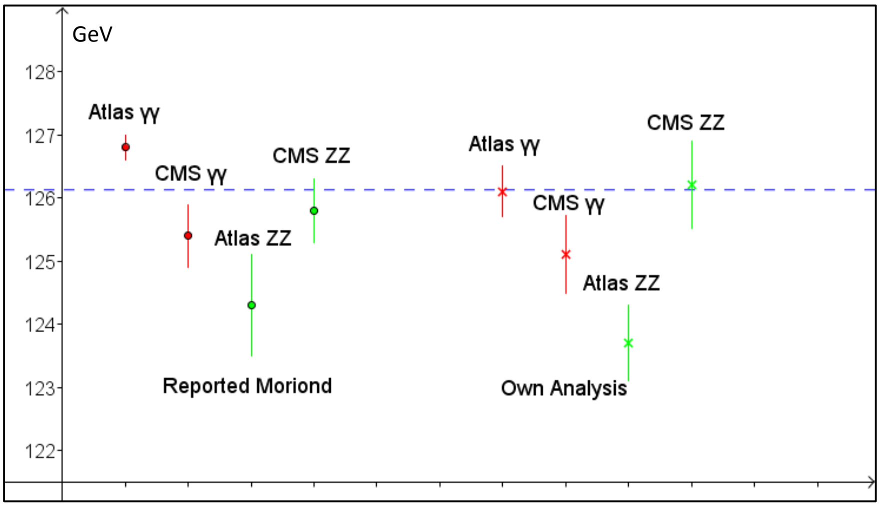


Fig. 31 Combination of the different plots with only statistical errors. Blue line shows the average mass.

The biggest difference comes from the Atlas analysis of the γγ-channel which has a small statistical error, but is not in acceptance with the own ZZ-channel analysis. All other different analysis are mainly in agreement with each other.

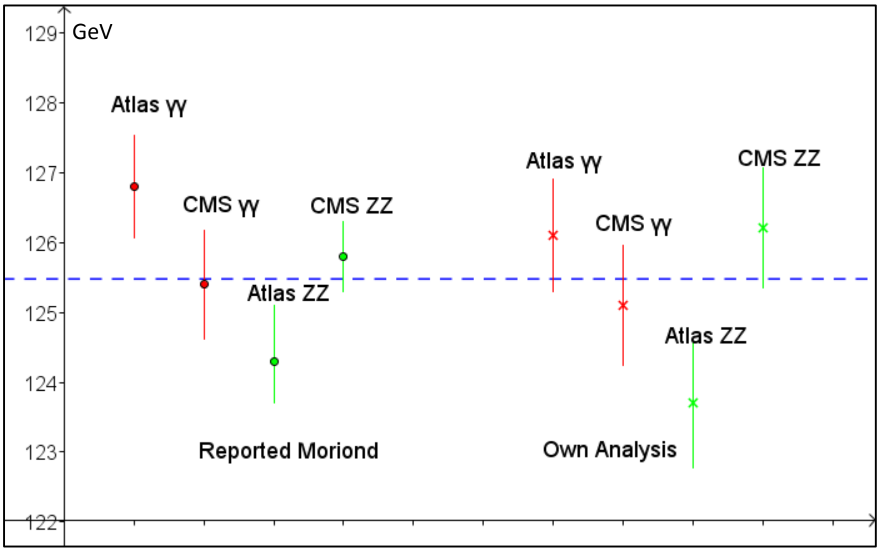


Fig. 32 Combination of the different plots with statistical and systematic errors included. Blue line shows the average mass.

The calculation of pure statistical errors points to discrepancies between the experiments
Atlas is not in agreement with itself in γγ or ZZ/4l channel.
However, after taking into account systematic uncertainty claimed in the summary papers by both experiments, the agreement between all results is acceptable. Systematic uncertainties could be caused by errors in the absolute calibration of the energy scale.
Of course, my analysis can only taking into account statistical uncertainties, including systematic errors reported by the experiments. The results of my analysis are in very good agreement with the overall mass determination by CERN experiments.

With included systematic errors, the mass of the Higgs boson of my own analysis is calculated as

125.3 GeV ± 0.6 GeV

The overall statistical evidence of all experiments and decay channels combined leads to

9 σ

This is a very good result which corresponds perfectly with actual announced results of the mass of the Higgs boson. The statistical evidence is overwhelming and leaves no doubt about the presence of a new particle.

9 Conclusions

Overall it is visible that the Atlas and the CMS experiment are in a good agreement with each other. The Atlas data look cleaner and have smaller background fluctuations than the CMS data sets, they also contain a stronger signal. It is unquestionable that the two experiments found a new particle, most likely the Higgs boson. With the present amount of data it is obscure if it is the standard model Higgs boson or not. This question should be possible to solve with new high energy data, for this reason the detector and the accelerator ring is being rebuilt at the moment.

My results correspond better than expected with the official results. It is very interesting to see that it is possible to get nearly the same results with lot easier and different statistical methods, even the statistical uncertainty of the official results is slightly better than mine. I reached my goal to perform an simple and independent data analysis. During a little improvement of my work, I obtained with the approximate method $S/\sqrt{B}$ confirming results.

A next step in this project could be to fully automate the data analysis, what would give the opportunity to see different things and analyse more data easier and with less expenditure of time. Another opportunity would be to analyse some different decay channels to know even more about the Higgs boson and its different decays and to get a statistically stronger result.

10 Thanks

I would like to thank:

- Dr. Jerzy Sromicki for his brilliant support and his big knowledge about particle physics and experimental physics, which was very important for this project.
- Dr. Christoph Grab for the very interesting conversation. He showed me the different possibilities for my project and gave a critical feedback for several ideas.
- Prof. Dr. Michele Weber for his support on behalf of “Schweizer Jugend forscht”
- My family for supporting me and helping me to solve different little problems and opening my mind.

11 Declaration

„Ich erkläre hiermit,

- dass ich die vorliegende Arbeit selbständig und nur unter Benutzung der angegebenen Quellen verfasst habe,
- dass ich auf eine eventuelle Mithilfe Dritter in der Arbeit ausdrücklich hinweise,
- dass ich vorgängig die Schulleitung und die betreuende Lehrperson informiere, wenn ich diese Maturaarbeit, bzw. Teile oder Zusammenfassungen davon veröffentlichen werde, oder Kopien dieser Arbeit zur weiteren Verbreitung an Dritte aushändigen werde.“

Sursee, 13. Oktober 2013 Benjamin Estermann
Überarbeitet 24. März 2014

12 References

For all calculations and all graphics, unless otherwise noted, the open source software GeoGebra (www.geogebra.org) was used.

Front page images:
http://kjende.web.cern.ch/kjende/en/zpath_hevents.htm (3.10.2013)
http://commons.wikimedia.org/wiki/File%3AHiggs_to_4leptons.png (3.10.2013)

[1] http://www.20min.ch/wissen/dossier/cern/story/-Der-Spass-hat-gerade-erst-begonnen--30754605 (3.10.2013)
http://en.wikipedia.org/wiki/Higgs_boson (24.3.2014)
http://atlas.physicsmasterclasses.org/en/zpath_hboson.htm (24.3.2014)

[2] https://twiki.cern.ch/twiki/pub/CMSPublic/Hig13001TWiki/mvamvasbweightedmass_1_5GeV.png (3.10.2013)

[3] http://en.wikipedia.org/wiki/Gaussian_distribution (11.10.2013)

[4] W.R. Leo, Techniques for Nuclear and Particle Physics Experiments, ISBN 3-540-17386-2 Springer-Verlag.

[5] W.R. Leo, Techniques for Nuclear and Particle Physics Experiments, ISBN 3-540-17386-2 Springer-Verlag. 4.7 Curve Fitting

[6] www.processing.org

[7] http://atlas.physicsmasterclasses.org/zpath_files/img/higgs_torte_DE.png (10.10.2013)

[8] https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2013-012/fig_03.png (3.10.2013)

[9] https://twiki.cern.ch/twiki/pub/CMSPublic/Hig12028TWiki/fig3.png (3.10.2013)

[10] https://twiki.cern.ch/twiki/pub/CMSPublic/Hig13001TWiki/ciccicsbweightedmassres1_5GeV.png (3.10.2013)

[11] https://twiki.cern.ch/twiki/pub/CMSPublic/Hig13001TWiki/Overlap.png (3.10.2013)

[12] http://cds.cern.ch/record/1530524/files/HIG-13-001-pas.pdf (10.10.2013)

[13] https://twiki.cern.ch/twiki/pub/CMSPublic/Hig13001TWiki/mvamvasbweightedmass_1_5GeV.png (3.10.2013)

[14] https://atlas.web.cern.ch/Atlas/GROUPS/PHYSICS/CONFNOTES/ATLAS-CONF-2013-013/fig_04a.png (7.10.2013)

[15] http://cds.cern.ch/record/1523767/files/Figure_007-a.pdf

13 Appendix

13.1 Graphics

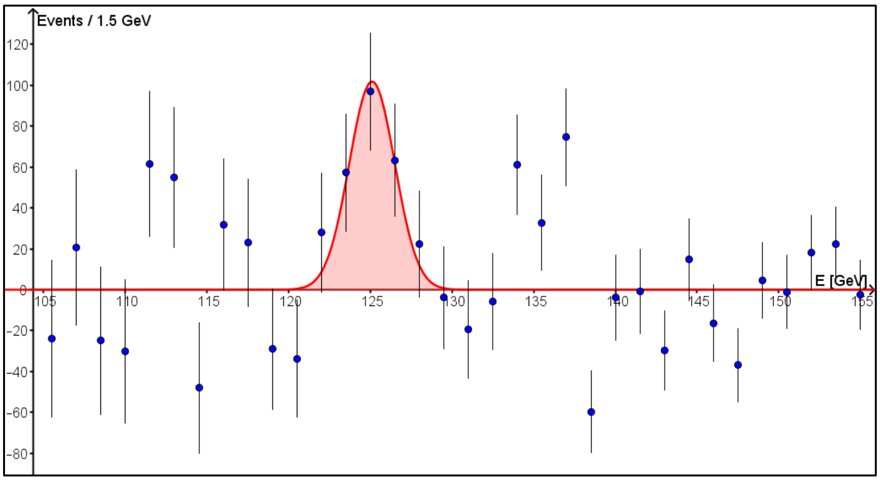


Fig. 33 Integral from the CMS experiment with the “discovery data”

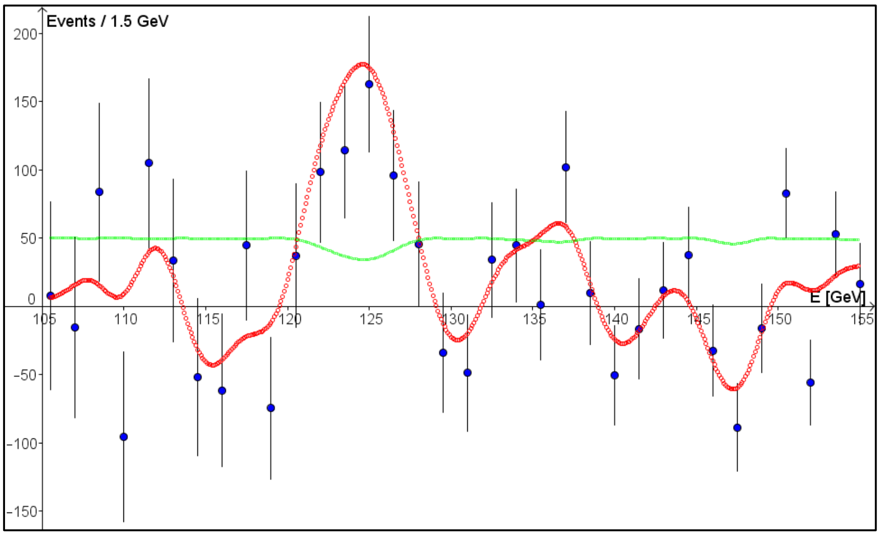


Fig. 34 Results of the automatic fitting procedure with the background residuals for the CMS CIC analysis. The amplitude (red points) and the χ^2_{DF} as function of the energy. Only in one area is an improvement of χ^2_{DF} visible. This area is the expected boson area.

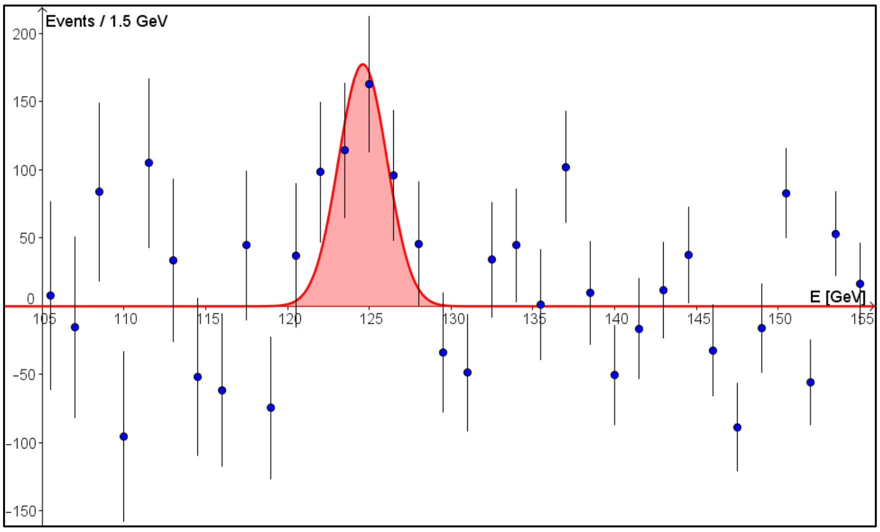


Fig. 35 The integral of the number of counts for the CMS CIC analysis

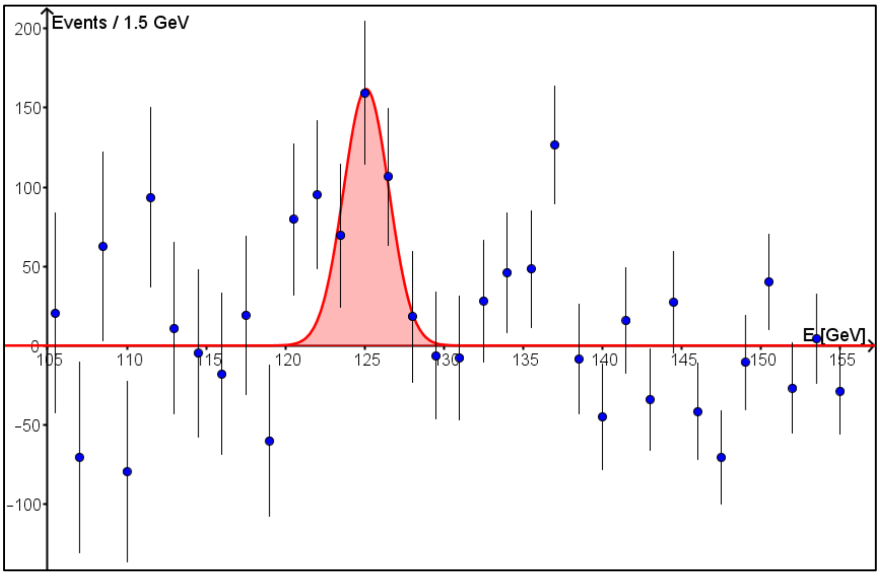


Fig. 36 The red area is the integral of the Gaussian curve for the CMS MVA analysis.

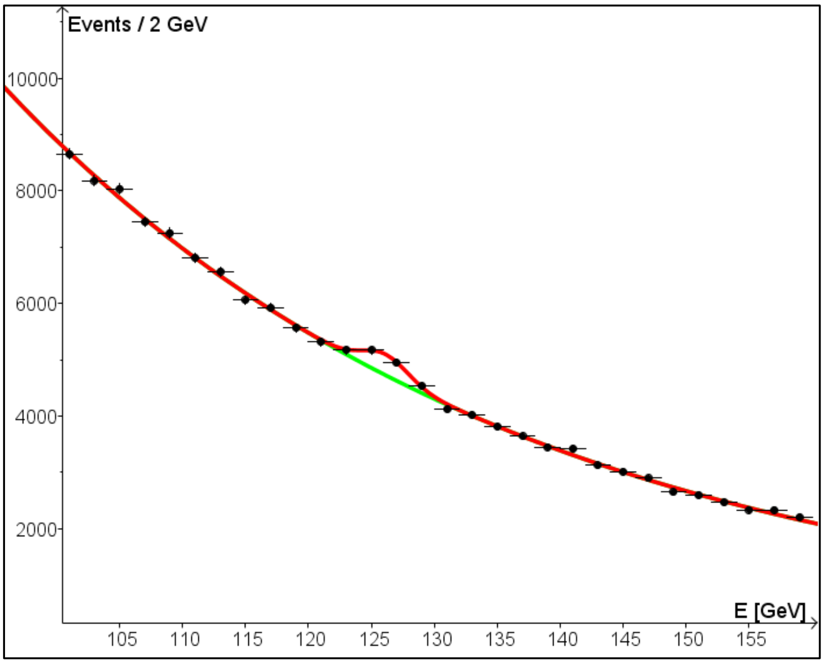


Fig. 37 Plot from the Atlas experiment with 25.5 fb^{-1} data. The green curve shows the background function. The red curve shows the fitted peak with the adapted Gaussian curve.

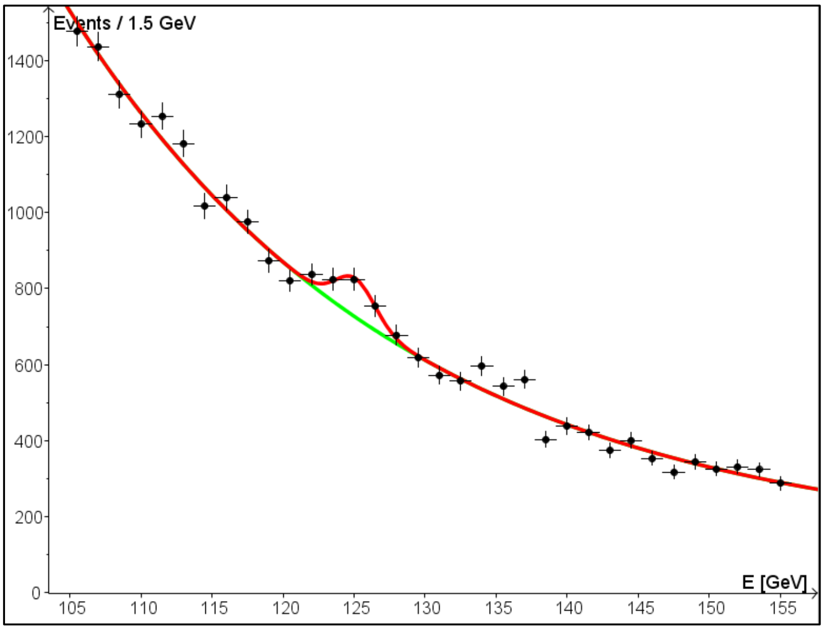


Fig. 38 Plot from the CMS experiment with the low statistics "discovery data". The green curve shows the background function. The red curve shows the fitted peak with the adapted Gaussian curve.

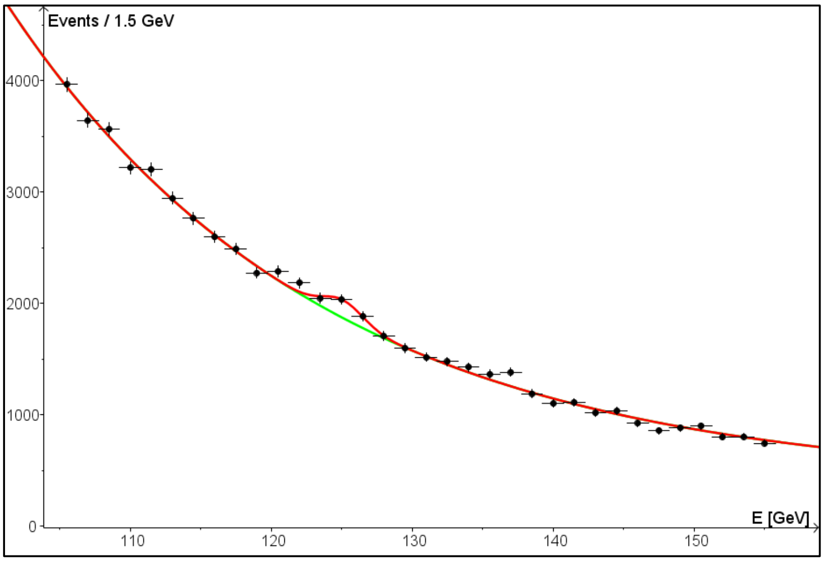


Fig. 39 Plot from the CMS MVA analysis with the high statistics 24.7 fb^{-1} data. The green curve shows the background function. The red curve shows the fitted peak with the adapted Gaussian curve.

13.2 Source code of the program

```
float[] counts = {};
float[] werte = {};
float[] xwerte = {};

float[] As = {};
float[] His = {};
float[] x0s = {};

int[] A1 = {};
int[] H1 = {};
int[] X1 = {};

float A = 0;
float x0 = 0;
float Sold = 3.5;
float Hi = 0;
float S = 0;

void setup(){
  S = Sold/2.35;
  float HilfsHi = 100;
  float HilfsA = 0;
  float Hilfsx0 = 0;
  for(x0 = xwerte[0]; x0 <= xwerte[xwerte.length-1]; x0 += 0.1){
    HilfsHi = 100;
    HilfsA = 300;
    Hilfsx0 = 0;
    for(A = -150; A < 400; A += 0.1){

      if( hirechnung(A, x0) < HilfsHi){
        HilfsHi = hirechnung(A, x0);
        HilfsA = A;
        Hilfsx0 = x0;
      }
```

```
    }
    As = append(As, HilfsA);
    His = append(His, HilfsHi);
    x0s = append(x0s, Hilfsx0);
  }

/*for(int i = 0; i < As.length; i++)
  println("x0: "+x0s[i]+", A: "+As[i]+", Hi: "+His[i]); */
for(int i = 0; i < As.length; i++){
  int a, b, c;
  As[i] = As[i] * 100000;
  His[i] = His[i] * 100000;
  x0s[i] = x0s[i] * 100000;

  a = int(As[i]);
  A1 = append(A1, a);

  b = int(His[i]);
  H1 = append(H1, b);

  c = int(x0s[i]);
  X1 = append(X1, c);
}

println("x0: ");
for(int i = 0; i < x0s.length; i++)
  println(X1[i]);
println();
println("A: ");
for(int i = 0; i < As.length; i++)
  println(A1[i]);
println();
println("Hi: ");
for(int i = 0; i < His.length; i++)
  println(H1[i]);
println();

}

float glocke(float x0, float b, float x) {
  float y = b*exp((-1 * ((x - x0)*(x - x0)))/(2*S*S));
  return y;
}

float hirechnung(float A, float x0) {
  float k = 0;
  for(int i = 0; i < counts.length; i++)
    k += (((counts[i] - glocke(x0, A, xwerte[i]))*(counts[i] - glocke(x0, A, xwerte[i])))/ werte[i]);
  k = k / (counts.length - 5);
  return k;
}
```