

Statistical Methods to Assess the Combined Sensitivity of the ATLAS Detector to the Higgs Boson in the Standard Model

Kyle Cranmer, Bruce Mellado, William Quayle, Sau Lan Wu

*University of Wisconsin - Madison
Department of Physics*

Abstract

We discuss statistical methods used to assess the combined sensitivity to the Standard Model Higgs boson using the ATLAS detector. The combination is performed using the likelihood ratio as a test statistic. Combined results are calculated using the results of a recent scientific note. Difficulties in calculating the significance using likelihood ratio techniques in the high event rate and high significance environment of the LHC are discussed, and software packages which implement solutions to these problems are presented. We also provide a brief discussion about different ways to incorporate systematic error and introduce the statistical notion of power to the ATLAS Higgs search.



1 Introduction

The search for the Standard Model Higgs boson is one of the major goals of the ATLAS detector. Several studies have been performed to assess the observability of different channels (combinations of production and decay modes of the Standard Model Higgs boson). Combining these different channels in a statistically and computationally rigorous way is the focus of this note.

Previously, it was shown that the migration of the statistical techniques that were used in the LEP Higgs Working group to the LHC environment is not as straightforward as one might naïvely expect [1, 2]. After a brief overview in Section 2, those difficulties and their ultimate solution are discussed in Section 3. Our group has developed two independent software solutions (both in C++; both with FORTRAN bindings; one ROOT based and the other standalone) which can be found at: <http://www-wisconsin.cern.ch/software>

In Section 4, we discuss the incorporation of systematic errors and compare a few different strategies. In Section 5, we document exactly which channels have been used and their expected systematic uncertainty on the background rates.

In Section 6, we present the combined significance to the Standard Model Higgs boson using the ATLAS detector. In Section 7, we present and discuss the “luminosity plot” which shows the luminosity required for discovery as a function of the Higgs mass, m_H . We note that this plot must be interpreted with some care. In Section 8, we discuss the statistical notion of *power* (which is related to the probability of Type II error – the probability we accept the “background-only hypothesis” in the presence of the signal). Lastly, in Section 9 we present and discuss the LEP-style $-2 \ln Q$ vs. m_H plot.

2 Formalism

Until recently, the ATLAS combined significance was calculated by adding the significances of the individual channels in quadrature [3]. This quadrature approach is a convenient approximation; however, recent work has focused on migrating the more rigorous techniques that were used at LEP to the LHC environment [1, 2]. We refer the interested reader to [4] for an introduction to the fundamentals, to [5] for why the likelihood ratio has been chosen as a test statistic, to [6] for a Monte Carlo approach to the calculation, and to [7] for the analytic calculation using Fast Fourier Transform (FFT) techniques. For completeness we introduce the basic approach below using the notation found in [1].

For a counting experiment where we expect on average b background events and s signal events, we consider two hypotheses: The null (or background-only) hypothesis, H_0 , in which the number of expected events, n , is described by a Poisson distribution

$P(n; b)$ and the alternate (or signal-plus-background) hypothesis, H_1 , in which the number of expected events is described by a Poisson distribution $P(n; s + b)$. Here the number of events serves the purpose of a *test statistic*: a real number which quantifies an experiment.

It is possible to include a discriminating variable x which has a some probability density function (pdf) for the background, $f_b(x)$, and some pdf for the signal, $f_s(x)$, both normalized to unity. Given an observation at $\hat{x}$ we can construct the likelihood ratio $Q = e^{-s}(sf_s(\hat{x}) + bf_b(\hat{x}))/bf_b(\hat{x})$. With several independent observations $\{\hat{x}_i\}$ we can consider the combined likelihood ratio $Q = \prod Q_i$. It is possible, and in some sense optimal, to use Q (or in practice $q = \ln Q$) as a test statistic.

The computational challenge of using the log-likelihood ratio in conjunction with a discriminating variable x is the construction of the log-likelihood ratio *distribution* for the background-only hypothesis, $\rho_b(q)$, and for the signal-plus-background hypothesis $\rho_{s+b}(q)$. In this case there are not only the Poisson fluctuations of the number of events, but also the continuously varying discriminating variable x . In particular, for a single background event the log-likelihood ratio distribution, $\rho_{1,b}(q)$, must incorporate all possible values of x . From these single event distributions we can build up the expected log-likelihood ratio distribution by repeated convolution. This is most effectively done by using a Fast Fourier Transform (FFT) where convolution can be expressed as multiplication in the frequency domain (denoted with a bar). In particular we arrive at:

$$\overline{\rho_b(q)} = e^{b[\overline{\rho_{1,b}(q)} - 1]} \quad (1)$$

and

$$\overline{\rho_{s+b}(q)} = e^{(s+b)[\overline{\rho_{1,s+b}(q)} - 1]}. \quad (2)$$

From the log-likelihood distribution of the two hypotheses we can calculate a number of useful quantities. Given some experiment with an observed log-likelihood ratio q^* we can calculate the background-only confidence level, CL_b

$$CL_b(q^*) = \int_{q^*}^{\infty} \rho_b(q') dq' \quad (3)$$

In the absence of an observation we can calculate the expected CL_b given the signal-plus-background hypothesis is true. To do this we first must find the median of the signal-plus-background distribution $\bar{q}_{s+b}$. From these we can calculate the expected CL_b by using Equation 3 evaluated at $q^* = \bar{q}_{s+b}$.

Finally, we can convert the expected background confidence level into an expected Gaussian significance, $N\sigma$, by finding the value of N which satisfies

$$CL_b(\bar{q}_{s+b}) = \frac{1 - \text{erf}(N/\sqrt{2})}{2}. \quad (4)$$

where $\text{erf}(N) = (2/\sqrt{\pi}) \int_0^N \exp(-y^2) dy$ is a function readily available in most numerical libraries.

3 Numerical Difficulties

The methods described in the previous section have been applied to the combined ATLAS Higgs effort with some caveats related to numerical difficulties [1]. In particular, in the extreme tails of $\rho_b(q)$, the probability density is dominated by numerical noise. This numerical noise is an artifact of round-off error in the double precision numbers used in the Fast Fourier Transform. The noise is on the order of 10^{-17} (for double precision floating point numbers), which translates into a limit on the significance of about 8σ .

For particular values of the Higgs mass, ATLAS has an expected significance well above 8σ with only 10 fb^{-1} of data. In order to produce significance values above the 8σ limit, various extrapolation methods were used in [1]. We now introduce a definitive solution to this problem based on arbitrary precision floating point numbers.

First it should be made clear that the numerical precision problem is not because CL_b is so small that the evaluation of the integral in Equation 3 cannot be treated with double precision floating point numbers. Instead, the numerical precision problem is due to the many (approximately 2^{20}) Fourier modes which must in total produce a number very close to 0. In order to rectify this problem, we have implemented the Fast Fourier Transform with the arbitrary-precision floating point numbers provided in the CLN library¹ [8].

Figure 1 demonstrates the problem and its solution; it shows the expected significance versus the number of expected signal events for a number-counting experiment with 5 expected background events. For a single-channel number-counting analysis, the CL_b can be calculated from the Poisson distribution $P(n; b)$ directly, and no FFT need be performed (the black curve). However, the calculation can also be done with likelihood ratio techniques, and the results should agree exactly. The red curve was obtained from a likelihood ratio calculation performed with double-precision numbers using the FFTW library² for the FFT. From the figure, it is clear that it agrees very well with the exact calculation until the significance approaches about 8σ , where the numerical noise starts to dominate. The green and blue curves show the results of the same calculation performed with 32 digit and 64 digit CLN numbers, respectively. The result is clear and unsurprising: using higher precision numbers to calculate the likelihood ratio probability distribution reduces the numerical noise and makes the calculation of the confidence level (and significance) reliable to much more extreme values.

One might protest that above 5σ we are not interested in the precise value of the significance and that this exercise is purely academic. We refer the interested reader to Section 8 for a different summary of the ATLAS discovery potential based on the notion of power.

¹CLN is available at www.ginac.de

²FFTW is available at www.fftw.org

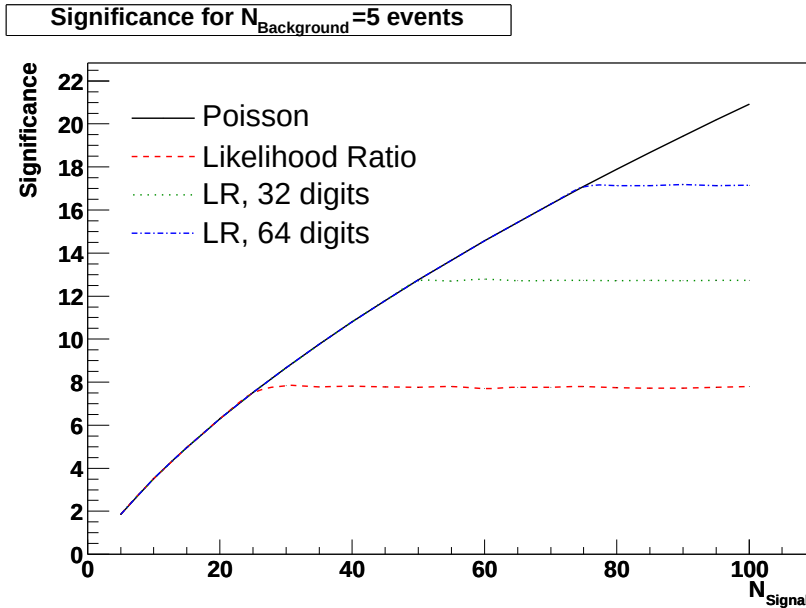


Figure 1: Comparison of significance calculations for a number-counting experiment with 5 background events; the horizontal axis is the number of signal events. The solid curve shows the exact result; the dashed, dotted, and dash-dotted curves show the results of a Likelihood ratio calculation using ordinary double precision numbers, 32-digit CLN numbers, and 64-digit CLN numbers, respectively.

4 Incorporation of Systematic Error

One encounters both philosophical and technical difficulties when one tries to incorporate uncertainty on the predicted values s and b found in Eq. 1. In a frequentist formalism the unknown s and b become nuisance parameters. In a Bayesian formalism, s and b can be marginalized by integration over their respective priors. At LEP the practice was to smear ρ_b and ρ_{s+b} by integrating s and b with a multivariate normal distribution as a prior. This smearing technique is commonly referred to as the Cousins-Highland Technique, and it has some Bayesian aspects.

4.1 A Purely Frequentist Technique

Once an analysis has been frozen, the effective cross-section for the background (equivalently, the expected background, b) in that phase-space region is also fixed. While the true background might be unknown, in nature it assumes a unique true value, b_t . To incorporate this background uncertainty into a frequentist calculation requires the addition of a *nuisance parameter*. One can not refer to a probability measure on the

nuisance parameter; one can only refer to the probability of an auxiliary measurement, M , given some value of the nuisance parameter (denoted $P(M|b)$). In practice, the auxiliary measurement is a side-band measurement or a control-sample used to normalize a related Monte Carlo prediction.

The logic for the frequentist method is to simultaneously consider the auxiliary measurement, M , and the test statistic, q , for each value of the nuisance parameter. One performs the Neyman construction: *i.e.* builds $N\sigma$ acceptance regions in the $M-q$ space for the background-only hypothesis. If the measurements M^*, q^* do not fall in the acceptance region, the background-only hypothesis is not consistent with the data at the $N\sigma$ level – which is equivalent to the condition for discovery when $N = 5$.

This technique has been applied to High-Energy Physics in Refs. [9, 10] when the distribution of M and q are both Poissonian. This technique was extended to arbitrary distributions $P(q, M|b)$ at the PhysStat2003 conference [11]. This method relies on the full Neyman construction and uses a likelihood ratio similar to the profile method as an ordering rule. In this formalism, channels with few events are more severely impacted by a systematic uncertainty at the level of 10% than when they are treated with the Cousins-Highland technique. This method is considerably more difficult to implement, and no general-purpose software has been developed. Until these technical difficulties are overcome, the Cousins-Highland technique is the most accepted method currently implemented.

4.2 The Cousins-Highland Technique

The incorporation of systematic errors into frequentist statistical techniques is a debatable topic [12]; however, Cousins and Highland provided a formalism which has become rather accepted [13, 14, 6, 7]. The idea is to consider an ensemble of experiments in which the number of expected events, $\mathbf{u}$, is described by some probability density function. When considering multiple channels, $\mathbf{u}$ is a vector quantity and u_i is the number of expected events in the i^{th} process/channel for some particular experiment in the ensemble.³ In general we need a multivariate probability density function to $P(\mathbf{u})$ to accommodate correlated systematic uncertainty between the channels. For instance, if our b-tagging has some uncertainty, then that effect will propagate to the various channels which use b-tagging in a correlated fashion.

The problem with the Cousins-Highland technique is largely philosophical: $P(\mathbf{u})$ is meaningless in a frequentist formalism: there is no ensemble in nature with which to measure $P(\mathbf{u})$. In a Bayesian formalism one can obtain $P(\mathbf{u}|M)$ by considering $P(M|\mathbf{u})$ and inverting it with the use of Bayes's theorem and the *a priori* likelihood for $\mathbf{u}$. The prior on $\mathbf{u}$ is subjective and quantifies ones “degree of belief” in the value $\mathbf{u}$. Typically, by $P(\mathbf{u})$ users of this technique mean $P(\mathbf{u}|M)$ and implicitly invert $P(M|\mathbf{u})$ with a uniform prior on $\mathbf{u}$.

³The variable b in the previous section, corresponds to some component of $\mathbf{u}$.

If we let $\rho_{\mathbf{u}}(q)$ be a generic distribution of the log-likelihood (or any other test-statistic) when we expect u_i events in the i^{th} process/channel, then the general form of the Cousins-Highland approach to incorporating systematic error is given by⁴:

$$\rho(q) = \int_{u_i \geq 0} \rho_{\mathbf{u}}(q) P(\mathbf{u}) d\mathbf{u} \quad (5)$$

The most common form of $P(\mathbf{u})$ is a Gaussian distribution. If we include a correlated error matrix $S_{ij} = \langle u_i - \langle u_i \rangle \rangle \langle u_j - \langle u_j \rangle \rangle$ then Equation 5 takes the form:

$$\rho(q) = \int_{u_1 \geq 0} \dots \int_{u_N \geq 0} \rho_{\mathbf{u}}(q) \left(\frac{1}{\sqrt{2\pi}} \right)^N \left(\frac{1}{\sqrt{|S|}} \right) e^{\sum_{i,j=1}^N -\frac{1}{2}(u_i - \langle u_i \rangle) S_{ij}^{-1} (u_j - \langle u_j \rangle)} du_i \quad (6)$$

Reference [7] provides an analytic expression for the resulting log-likelihood ratio distribution including a correlated error matrix; however, this equation was obtained with an integration over negative numbers of expected events and does not hold.

While the Gaussian form is quite popular, it is not necessarily the most justified. In particular, we impose that the expected number of events satisfies $u_i \geq 0$ while for a Gaussian distribution there is a finite probability of $u_i < 0$. Furthermore, these errors/uncertainties may not be normally distributed even near their mean. Further complicating matters is the fact that if we want to make statistical statements at the 5σ level, then in some way we need to be able to describe (or be independent of) the distribution $P(\mathbf{u})$ at the 5σ level as well.

In this note, we assume that the systematic error distribution has a Gaussian shape out to 3σ and is 0 elsewhere (properly normalized). We perform the integral over $\mathbf{u}$ in Equation 6 with Monte Carlo techniques; however the likelihood ratio convolution is still performed with an FFT. In other words, for each experiment in the ensemble we make an exact calculation with the FFT; however, the ensemble is based on Monte Carlo.

4.3 The “Plus-Minus” Technique

Another method that has been used to incorporate systematic error in a frequentist setting is to simply increase the background expectation by some amount and decrease the signal expectation by some amount. For lack of a better name, this method will be referred to as the “plus-minus” method.

There are a few variants on this procedure: *e.g.* not varying the signal expectation. The “plus-minus” and Cousins-Highland methods are truly distinct. Neither method is “more conservative” in general – depending on the number of events and the systematic

⁴This integral is often referred to as a convolution, however $\rho_{\mathbf{u}}(q)$ is also a function of $\mathbf{u}$ so it is not. For a general $P(\mathbf{u})$, this integral can be performed with Monte Carlo techniques.

error either method may produce a lower significance. A comparison of the background confidence level, CL_b , for the two methods in two different scenarios is presented in Figure 2.

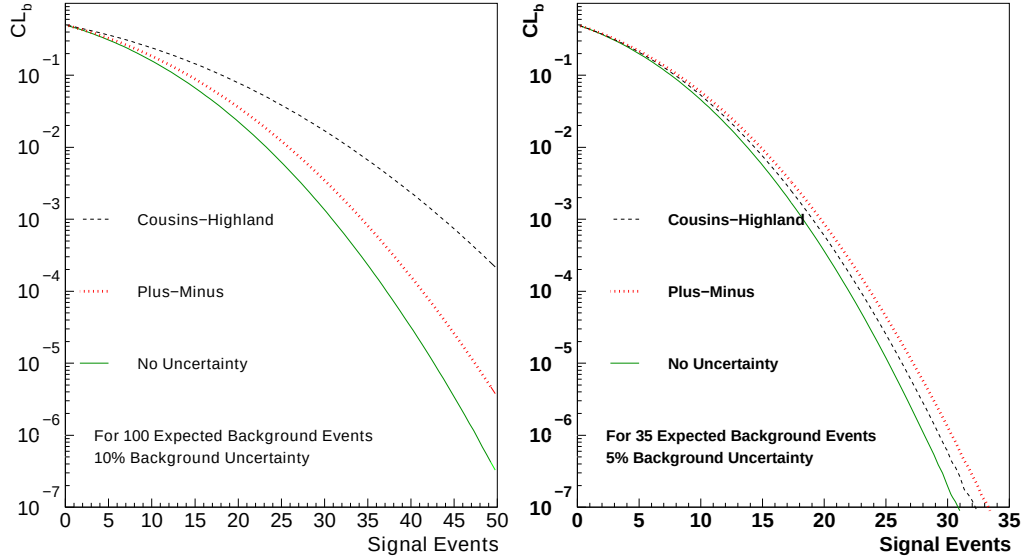


Figure 2: Comparison of the background confidence level, CL_b , as a function of the number of signal events for different experiments and different methods of incorporating systematic error. The left plot corresponds to an experiment with 100 background events and 10% background uncertainty. The right plot corresponds to an experiment with 35 background events with a 5% background uncertainty. The curves show that without systematic error (green solid line) the CL_b is the lowest (highest significance). Depending on the experiment, the use of the Cousins-Highland technique (black dashed line) can produce a lower or higher CL_b than the “plus-minus” method (red dotted line).

4.4 An Example

Let us consider the case of 100 expected background events, 50 expected signal events, and a 5% uncertainty on the background normalization. The interpretation of these quantities and the number of events needed to claim a discovery, x , is different for each of the three techniques.

The Frequentist Technique

By “5% uncertainty on the background normalization” we mean that $P(M|b)$ is a Gaussian centered at b with standard deviation Δb (where $\Delta = 5\%$).⁵ In this method,

⁵We use the symbols Δ and α to distinguish between the standard deviation in the measurement

each value of b is considered, so the interpretation of “100 expected background events” means a measurement of M for which the maximum likelihood estimator is $b = 100$. In our Gaussian case, this implies that we make a hypothetical measurement at $M^* = 100$. The signal expectation is used when constructing the likelihood ratio, and affects the Neyman construction and the numerical value of the test statistic.

The question now is how many events, x , must we observe to claim a discovery?⁶ As was presented in [11], one must observe 167 events to claim a discovery.

The Cousins-Highland Technique

What is meant by “5% uncertainty on the background normalization” in this formalism? If we mean that $P(M|b)$ is a Gaussian centered at b with standard deviation Δb (where $\Delta = 5\%$), then with the implicit Bayesian step (invert $P(M|b)$ into $P(b|M)$ using a uniform prior) we obtain $P(b|M)$. This results in a Gaussian $P(b|M)$ centered at $b = 100$ with standard deviation αb , where $\alpha = 5\%$. The signal expectation is used when constructing the likelihood ratio, and the numerical value of the test statistic.

With the code described in Section 1, it is easy to calculate that one must observe 161 events to claim a discovery.

Initially, one might think that 161 is quite close to 167; however, they differ at the 4% level and the methods are only considering a $\Delta = 5\%$ effect. Still worse, is the probability that the two methods lead to different conclusions about the data. If H_0 is true (say $b_t = 100$) and one can claim a discovery with the Cousins-Highland method ($x_0 > 161$), the chance that one could not claim a discovery with the fully frequentist method ($x_0 < 167$) is $P(x < 167|x > 161, H_0) \approx 95\%$. Similarly, if H_1 is true and one can claim a discovery with the Cousins-Highland method, the chance that one could not claim a discovery with the fully frequentist method is $P(x < 167|x > 161, H_1) \approx 50\%$. Even practically, there is quite a difference between these two methods.

The Plus-Minus Technique

In this technique, background uncertainty is taken into account by simply scaling up the background expectation by αb . If we construct a Poisson distribution for x given 105 expected events, then we can calculate the discovery condition without a signal expectation. With the code described in Section 1, it is easy to calculate that one must observe 160 events to claim a discovery.

In this case, the “plus-minus” and Cousins-Highland results are similar; however, this is not always the case (see Figure 2).

M and the standard deviation in the expected background b , respectively. Note, only Δ is well defined in a frequentist formalism.

⁶In practice, one would measure both x^* and M^* , then ask “have we made a discovery?”. For the sake of explanation, we have broken this process into two pieces.

4.5 Saturation of Significance

An important feature of the incorporation of systematic error with the Cousins-Highland technique is the saturation of signal significance. What is meant by “saturation” is that the significance reaches an asymptotic value as more data are collected (holding the systematic error fixed). Consider a channel in which the background normalization has a relative uncertainty α . In the limit of large integrated luminosity, the natural statistical variation becomes negligible compared to the systematic error, and the signal significance approaches a constant σ_∞ .

The Cousins-Highland Technique

In the Cousins-Highland technique, one can calculate the saturation exactly. As more events are collected, the Gaussian approximation of a Poisson distribution is valid, thus

$$\sigma_\infty^{CH} = \lim_{L \rightarrow \infty} \frac{s}{\sqrt{b(1 + \alpha^2 b)}} = \frac{s/b}{\alpha}. \quad (7)$$

The above equation is somewhat misleading because there are implicit restrictions on the Cousins-Highland approach with a Gaussian $P(\mathbf{u})$. To claim an observation is inconsistent with the background at the $N\sigma$ level, we must be able to describe the background at the $N\sigma$ level. If the $N\sigma$ contours of our background description include the unphysical prediction $b < 0$, we know our background description is failing. Thus the Cousins-Highland approach is internally inconsistent for $\alpha \gtrsim 1/N\sigma > 1/\sigma_\infty^{CH}$. In particular, the background must be known within 20% to achieve a 5σ effect.

The Frequentist Technique

The frequentist significance calculation is very difficult; however, the limiting behavior of σ_∞ can be derived geometrically. Using the likelihood-ratio as an ordering rule, observing that this produces approximately similar tests, and observing that the contours of this likelihood have a simple form one arrives at:

$$\frac{s}{M} = \frac{\sigma_\infty^F \Delta}{1 - \sigma_\infty^F \Delta}. \quad (8)$$

The frequentist method shows explicitly that as $\Delta \rightarrow 1/\sigma_\infty$, the required $s/M \rightarrow \infty$. This equation can be rewritten as

$$\sigma_\infty^F = \frac{s}{\Delta(s + M)} = \frac{s/M}{\Delta(1 + s/M)}. \quad (9)$$

Figure 3 compares the contours of σ_∞ for the Cousins-Highland and fully frequentist methods.

The Plus-Minus Technique

It is worth noting that this saturation feature is not present in the “plus-minus” method. In that case,

$$\sigma_{\infty}^{\pm} = \lim_{L \rightarrow \infty} \frac{s(1 - \alpha)}{\sqrt{b(1 + \alpha)}} = \infty. \quad (10)$$

This behavior is not surprising, the “plus-minus” method is equivalent to assuming one knows the background exactly (it just happens to be more than one originally expected).

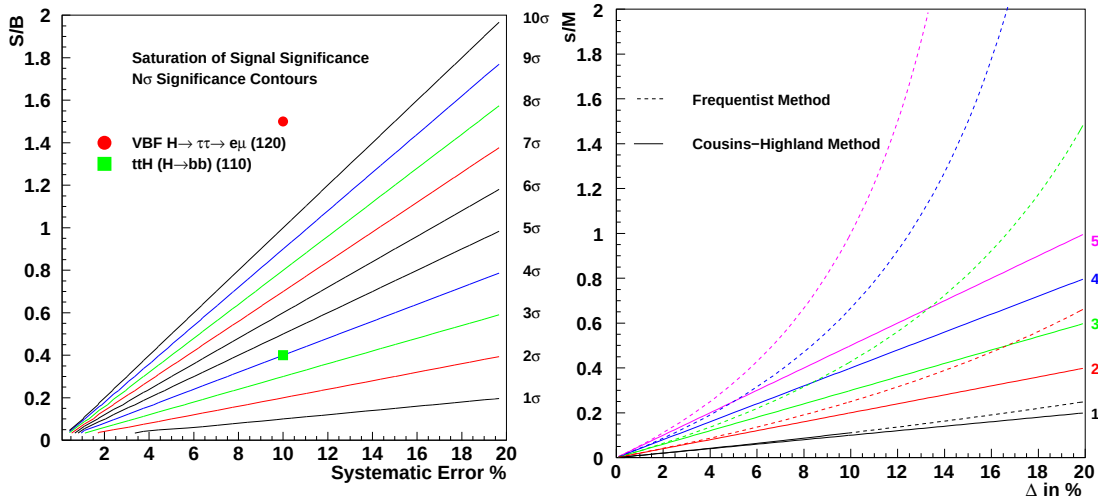


Figure 3: Left: Contours of σ_{∞}^{CH} in the plane of signal-to-background ratio vs. the systematic error α in percent. Two search channels are marked in the plane for reference. Right: Comparison of the contours for the Cousins-Highland and frequentist techniques. The frequentist variables M and Δ are used to distinguish between their Bayesian equivalents b and α .

4.6 Strategy for Systematics

Until a general-purpose code is available for the frequentist calculation, the Cousins-Highland technique remains the most accepted method currently implemented. We feel that the Cousins-Highland method is a superior method to the “plus-minus” method, and that criticism of the Gaussian shape for systematic errors should be directed to the choice of $P(\mathbf{u})$ and not to the formalism seen in Equation 5.

5 Channels Considered

For the combinations presented below, we use the results of a the recent ATLAS scientific note [15]. We have not used more recent results intentionally, in order to focus attention on the combination procedure. For completeness we provide a brief description and relevant references to the channels considered in this combination.

- Vector Boson Fusion $H \rightarrow WW^{(*)}$: This is the dominant discovery channel across most of the mass range considered because of its large cross-section (relative to other VBF processes) and its high signal-to-background ratio. However, because of the presence of two neutrinos in the final state, bona fide mass reconstruction is not possible. We consider a 10% systematic uncertainty on the background normalization [15]. Due to the complementarity of the event selection for this channel and the inclusive $H \rightarrow WW^{(*)} \rightarrow ll\cancel{p}_T$, the overlap of events selected by the two channels has been found to be negligible.
- Vector Boson Fusion $H \rightarrow \tau^+\tau^-$: This is a very powerful channel for masses around 110-140 GeV. In contrast to the VBF $H \rightarrow WW$ channel, mass reconstruction can be performed here with a mass resolution on the order of 10%. We consider a 10% systematic uncertainty on the background normalization.
- $t\bar{t}H(H \rightarrow b\bar{b})$: This analysis is very important near the LEP exclusion limit where the branching ratio of $H \rightarrow b\bar{b}$ is very large. Based on reference [15], we apply a uniform 5% uncertainty on this channel when calculating the expected significance, and not the larger systematic found in [16]. If the results in Ref. [16] are used, it is impossible to reach a 5σ significance level with 10% background uncertainty due to the relatively low signal-to-background ratio (see Figure 3).
- $H \rightarrow \gamma\gamma$: This analysis requires excellent understanding of the electromagnetic energy scale and a low systematic error on the background due to the very low signal-to-background ratio. The systematic error in this channel is considered to be negligible.
- $H \rightarrow WW^* \rightarrow \nu l \nu$: This analysis is complementary to the $H \rightarrow ZZ^* \rightarrow 4l$ analysis near a Higgs mass of 170 GeV. In contrast to the Vector Boson Fusion analysis, the production mode for the inclusive analysis is dominated by gluon-gluon fusion. The complementary jet requirements are responsible for removing potential overlap with the Vector Boson Fusion analysis. We use the a 5% systematic uncertainty on the background normalization.
- $H \rightarrow ZZ^{(*)} \rightarrow 4l$: Sometimes referred to as the “golden channel”, this channel has been the dominant discovery mode for ATLAS across a very large range of masses. Though it no longer has the highest expected significance, this channel offers a stunning mass peak and will be pivotal to the discovery of a Higgs with a mass above 200 GeV. No systematic uncertainty in the background was included.

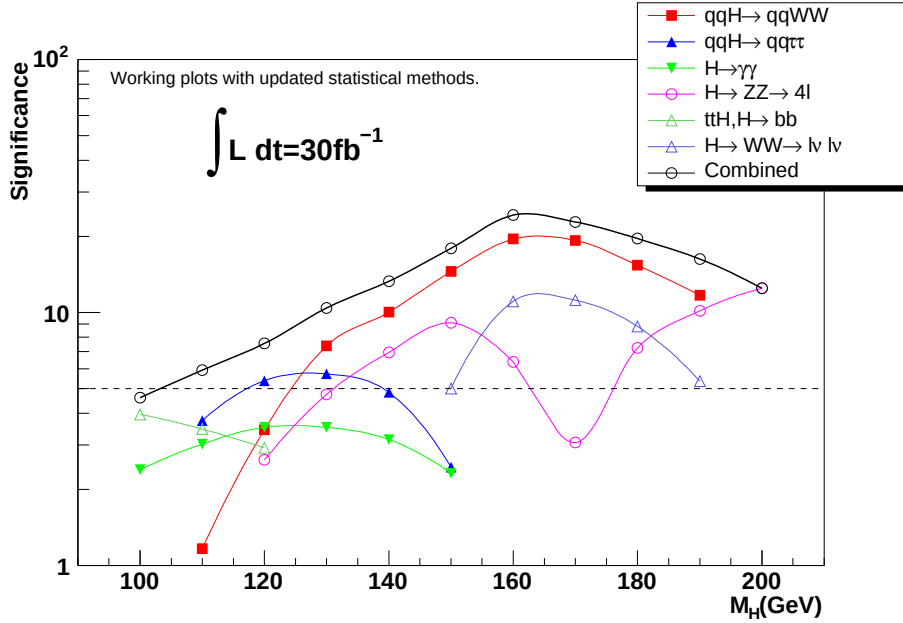


Figure 4: Individual and combined significance versus the Higgs mass hypothesis.

6 Combined Significance

In this section we present the combined significance of the ATLAS detector (see Figure 4). These combinations were made with a consistent treatment which uses the likelihood ratio as a test statistic and the Cousins-Highland formalism to incorporate systematic errors. The combination corresponds to 30 fb^{-1} of integrated luminosity.

The combined significance for 30 fb^{-1} of integrated luminosity is expected to be above 5σ for $m_H \gtrsim 105 \text{ GeV}$ – below the LEP limit [14]. The combined significance is dominated by $ttH(H \rightarrow bb)$ at low masses, VBF $H \rightarrow WW$ for intermediate masses, and $H \rightarrow ZZ$ for higher masses. Near the LEP limit, several channels are required and available to observe the Higgs.

Recall that an expected significance of 5σ means that there is only a 50% chance to observe an effect in excess of 5σ if the Higgs is indeed there (see Section 8). In the 50% of cases in which the effect is less than 5σ , we can be quite confident that the effect will still be in excess of 3 or 4σ (the exact probability is a function of s/b).

7 Luminosity Plots

In this section we present the “Luminosity Plot” (see Figure 5). We define the discovery luminosity $L^*(m_H)$ to be the amount of luminosity necessary for the expected significance at a Higgs mass m_H to reach 5σ . The discovery luminosity is an informative quantity; however, it must be interpreted with some care:

- Collecting an integrated luminosity equal to the nominal discovery luminosity $L^*(m_H)$ does not guarantee that a discovery will be made if the Higgs is indeed present at the corresponding mass m_H . Instead, with $L^*(m_H)$ of data, the median of the expected signal-plus-background will be at the 5σ level – which corresponds to a 50% chance of discovery. See Section 8 for more details.
- In practice an analysis’ cuts, systematic error, and signal and background efficiencies are luminosity-dependent quantities. When we make the “luminosity” plot, we treat the analysis as constant. We must interpret the discovery luminosity with some care and realize that beyond 30 fb^{-1} we move from a low- to a high-luminosity environment. We also must realize, though this is fairly obvious, that a discovery luminosity of 1 fb^{-1} does not mean the first 1 fb^{-1} of data, but at least 1 fb^{-1} of well-understood data that is consistent with the analysis’ assumptions.

It is also worth pointing out that there does not always exist a luminosity for which the significance reaches 5σ (see Section 4.5). For instance, the $t\bar{t}H(H \rightarrow b\bar{b})$ analysis is not a discovery channel if we hold fixed the signal-to-background ratio and the systematic error (at 10%) as we increase the luminosity. The background uncertainty in this channel includes statistical fluctuation in the control sample that is used to determine the background [16]. As more data are accumulated, the background uncertainty will reduce. Thus for Figure 5 we have assumed a 5% systematic error for the $t\bar{t}H(H \rightarrow b\bar{b})$ analysis.

8 The Power of a 5σ Test

The traditional plot which is used to summarize the ATLAS discovery potential is the combined significance shown in Section 6; however, as was noted in Section 3 and in [1], this plot becomes very difficult to make in a consistent way when the significance goes beyond about 8σ . Furthermore, the plot itself starts to lose relevance when the significance is far above 5σ . The Luminosity Plot shown in Section 7 is another possible way of showing the ATLAS discovery potential, but as was discussed it must be interpreted with some care. In this section we introduce a third illustration of the ATLAS discovery potential which is related to the probability of a “false-negative” or Type II error: the *power*.

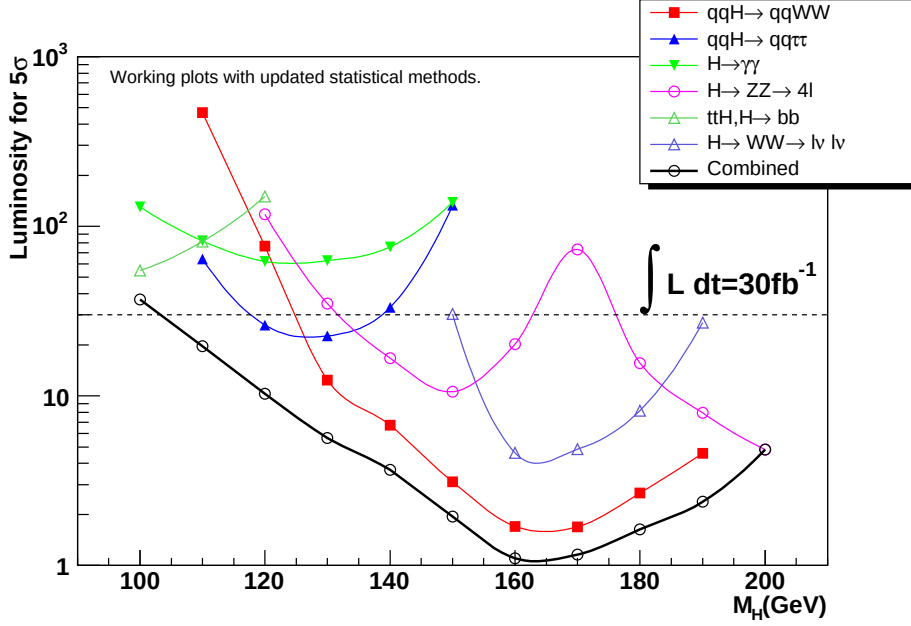


Figure 5: The integrated luminosity required for the expected significance to reach 5σ as a function of the Higgs mass, m_H , in the enhanced scenario.

First, it should be noted that the significance plot measures the separation between the *medians* of the background-only and signal-plus-background hypotheses. Thus, when we see the significance curve cross the 5σ line (at some mass m_H^*) there is only a 50% chance that we would observe a 5σ effect if the Higgs does indeed exist at that mass. In practice, we claim a discovery if the observed data exceeds the 5σ critical region, and do not claim a discovery if it doesn't. The meaning of the 5σ discovery threshold is a convention which sets the probability of a “false-positive” or Type I error to be $2.85 \cdot 10^{-7}$. With that in mind, the idea that the significance is 12σ at $m_H = 300$ GeV is irrelevant. What is relevant is the probability that we will claim discovery of the Higgs if it is indeed there: that quantity is called the power. The power is defined as $1 - \beta$, where β is the probability of Type II error: the probability that we reject the signal-plus-background hypothesis when it is true [4].

Consider Figure 6 with a background expectation of 100 events. The black vertical arrow denotes the 5σ discovery threshold. The red curve shows the distribution of the number of expected events for a signal-plus-background hypothesis with 150 events. Normally we would say the expected significance is 5σ for this hypothesis; however, we can see that only 50% of the time we would actually claim discovery. The blue curve shows the distribution of the number of expected events for a signal-plus-background hypothesis with 180 events. Normally we would say the expected significance is 8σ for this hypothesis; however, a more meaningful quantity - the power - is the probability

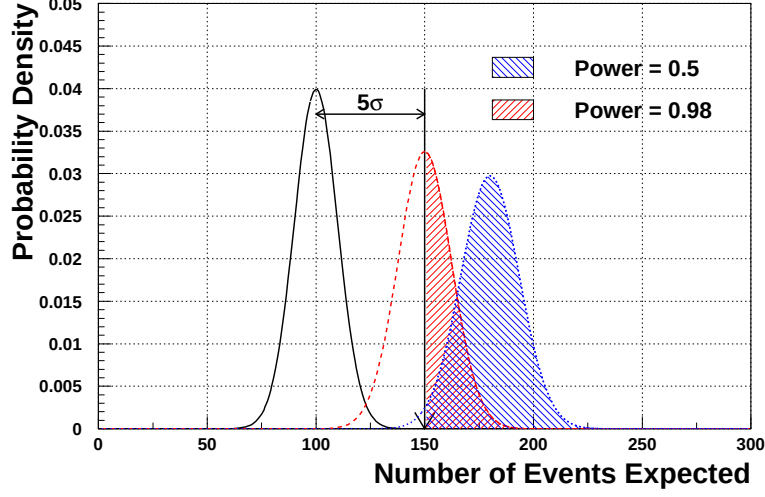


Figure 6: Examples of power for two different signal-plus-background hypotheses with respect to a single background-only hypothesis with 100 expected events (black).

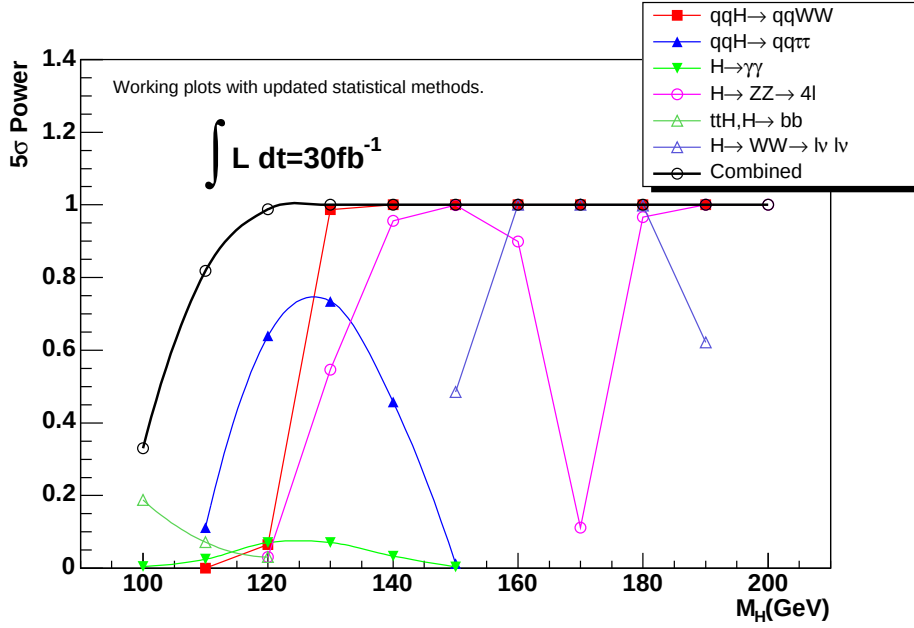


Figure 7: The power (evaluated at 5σ) of ATLAS as a function of the Higgs mass, m_H , for 30 fb^{-1} with and without systematic errors for the enhanced scenario.

we would claim discovery which, in this case, is about 98%.

When we use the likelihood ratio as a test statistic, the abscissa in Figure 6 is no longer a number of expected events; but instead it is the log-likelihood ratio q . Figure 7 shows the power of ATLAS' combined Higgs searches as a function of mass for the enhanced scenario.

9 LEP-Style $-2\ln Q$ vs. m_H Plots

The last plot which we present to summarize the ATLAS discovery potential is the LEP-style $-2\ln Q$ vs. m_H plot. For each Higgs mass, one can imagine the two distributions of the log-likelihood ratio, $\rho_b(q)$ and $\rho_{s+b}(q)$. Figure 8 traces out the median of both of these distributions, and shows the 3σ and 5σ contours around the background only median. From this plot, one obtains the same general information as the significance plot. This plot would also need bands around the signal-plus-background curve for the power to be deduced. This plot is much easier to make than the significance plot, but is considerably harder to interpret for non-experts.

The reason that the ordinate is $-2\ln Q$, instead of $\ln Q$, is that $-2\ln Q$ is approximately equal to the difference in χ^2 when the data configuration is compared to the background-only and signal-plus-background hypotheses. The line at $-2\ln Q = 0$ corresponds to a data configuration that is ambivalent between the two hypotheses.

10 Conclusions

We have presented four summaries of the ATLAS combined discovery: significance, discovery luminosity, power, and $-2\ln Q$. We have compared three methods to incorporate background uncertainty, and we advocate the usage of a purely frequentist method once available. To perform these combinations, we used the log-likelihood ratio, q , as a test statistic, constructed background-only and signal-plus-background distributions of q with an arbitrary precision Fast Fourier Transform, and incorporated background uncertainty with the Cousins-Highland technique.

Assuming the signal-plus-background hypothesis, ATLAS has at least a 50% chance to claim a 5σ discovery if $m_H \gtrsim 105$ GeV with just 30 fb^{-1} of data (see Figure 4). If the Higgs is heavier than 120 GeV, we can be roughly 95% confident that ATLAS will be able to claim a 5σ discovery with 30 fb^{-1} (see Figure 8). Multiple channels are available for almost all values of m_H , allowing for a robust discovery and the potential for coupling measurements.

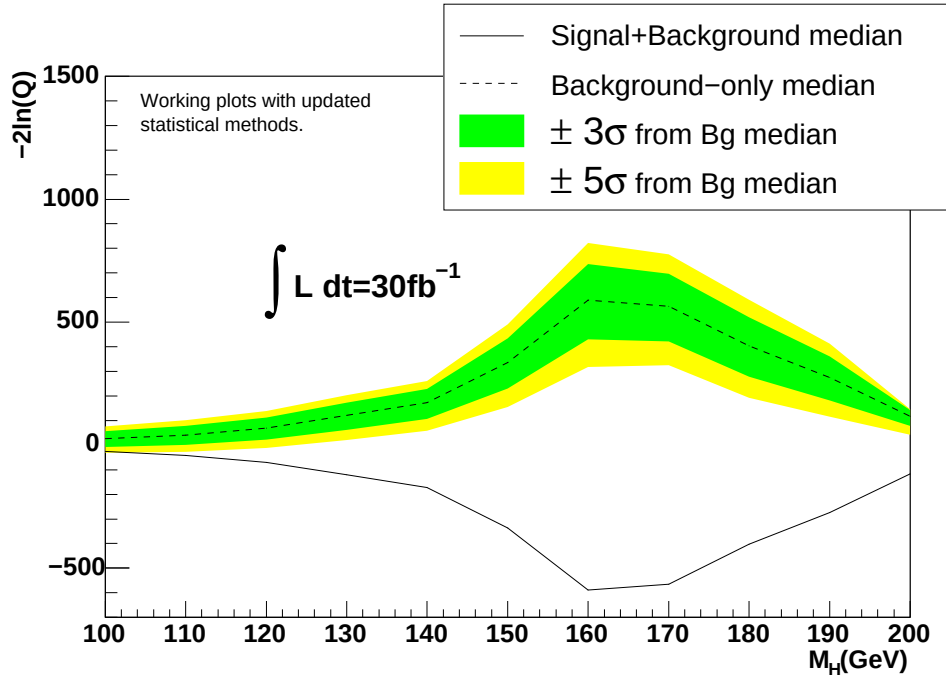


Figure 8: A plot of $-2 \ln Q$ vs. m_H for 30 fb^{-1} of integrated luminosity in the reference (top) and enhanced (bottom) scenarios.

References

- [1] K. Cranmer, P. McNamara, B. Mellado, W. Quayle, Sau Lan Wu, Confidence Level Calculations for $H \rightarrow W^+W^- \rightarrow l^+l^- \cancel{p}_T$ for $115 < M_H < 130$ GeV Using Vector Boson Fusion, ATLAS Note ATL-PHYS-2003-008 (2003).
- [2] K. Cranmer, B. Mellado, W. Quayle and Sau Lan Wu, Challenges of Moving the LEP Higgs Statistics to the LHC, PhyStat2003 physics/0312050 (2003).
- [3] ATLAS Collaboration, Detector and Physics Performance Technical Design Report (Volume II), CERN-LHCC/99-15 (1999).
- [4] A. Stuart, J.K. Ord, and S. Arnold, Kendall’s Advanced Theory of Statistics, Vol 2A (6th Ed.) (Oxford University Press, New York, 1994).
- [5] A.L. Read, Optimal Statistical Analysis of Search Results based on the Likelihood Ratio and its application to the search for the MSM Higgs Boson at $\sqrt{s} = 161$ and 172 GeV, DELPHI note 97-158 PHYS 737 (1997).
- [6] T. Junk, Nucl. Instrum. Meth. **A434** (1999) 435–443.
- [7] H. Hu, J. Nielsen, Analytic Confidence Level Calculations Using the Likelihood Ratio and Fourier Transform, “Workshop on Confidence Limits”, Eds. F. James, L. Lyons and Y. Perrin, CERN 2000-005 (2000), p. 109.
- [8] C. Bauer, A. Frank, and R. Kreckel, J. Symbolic Computation **33** (2002) 1–12.
- [9] James, F. and Roos, M., Nucl. Phys. **B172** (1980) 475.
- [10] Cousins, R. D., Nucl. Instrum. Meth. **A417** (1998) 391–399.
- [11] K. Cranmer, Frequentist Hypothesis Testing with Background Uncertainty, Phy-Stat2003 physics/0310108 (2003).
- [12] F. James (ed.), Y. Perrin Y.(ed.) and L. Lyons (ed.), Confidence limits. Proceedings, 1st Workshop, Geneva, Switzerland, January 17-18, 2000, CERN-2000-005.
- [13] R.D. Cousins and V.L. Highland, Nucl. Instrum. Meth. **A320** (1992) 331–335.
- [14] LEP Higgs Working Group, Phys. Lett. **B565** (2003) 61–75.
- [15] S. Asai et al., Eur. Phys. J. **C3252** (2004) 19–54.
- [16] J. Cammin and M. Schumacher, The ATLAS Discovery Potential for the Channel $t\bar{t}H$, H to $b\bar{b}$, ATLAS Note ATL-PHYS-2003-024 (2003).