Statistical Combination of Several Important Standard ² Model Higgs Boson Search Channels

Abstract

3

In this note we describe statistical procedures for combination of results from Δ independent searches for the Higgs boson. Here only the Standard Model 5 Higgs is considered, although the methods can easily be extended to non-6 standard Higgs models as well as to other searches. The methods are applied to Monte Carlo studies of four important search channels: $H \rightarrow \tau^+ \tau^-$, 8 $H \to W^+W^- \to e\nu\mu\nu, H \to \gamma\gamma$ and $H \to ZZ^{(*)} \to 4$ leptons. The statistical 9 treatment relies on a large sample approximation that is expected to be valid 10 for an integrated luminosity of at least 2 fb^{-1} . Results are presented for the 11 expected statistical significance of discovery and expected exclusion limits. 12

13 **1 Introduction**

Higgs searches will exploit a number of statistically independent decay channels. One wishes to combine
all of the information from them to provide a single measure of the significance of a discovery or limits
on Higgs production. The approach taken in this paper is based on frequentist statistical methods, where
effects of systematic uncertainties are incorporated by use of the profile likelihood ratio.

¹⁸ The statistical procedures used for establishing discovery and setting limits are described in Section 2. ¹⁹ These methods are very general and can be applied to the combination of results of essentially any search ²⁰ that will be carried out at the LHC. Section 3 summarizes the four search channels for the Standard Model ²¹ Higgs boson considered in this note: $H \rightarrow \tau^+\tau^-$, $H \rightarrow W^+W^- \rightarrow e\nu\mu\nu$, $H \rightarrow \gamma\gamma$ and $H \rightarrow ZZ^{(*)} \rightarrow$ ²² 4 leptons.

The statistical treatment requires knowledge of the distribution of a test statistic based on the profile likelihood ratio. To determine these distributions by Monte Carlo so as to establish discovery at a high level of significance would require an enormous amount of simulated data, which is not practical at present. Therefore the distributions have been estimated using the functional form expected to hold in the large sample limit. Investigations shown in Section 3 indicate that this approximation should be reliable for an integrated luminosity above 2 fb⁻¹.

In Section 4 we show the result of the combination. For different values of the integrated luminosity and hypothesized Higgs mass, we present the signal significance expected assuming the Standard Model Higgs production rate, as well as expected upper limits on the Higgs production cross section, under the hypothesis of no Higgs signal.

The channels considered here focus on the search for a Higgs boson in the low-mass range. It is planned to include other channels in the future, e.g., further final states from the W^+W^- and ZZ modes. This will improve sensitivity especially at higher Higgs mass values.

36 2 Statistical methods

³⁷ In this section we describe the general statistical model and likelihood function, first for a single channel

and then generalized to multiple channels. In Section 2.2 we give the procedure used to establish discov-

³⁹ ery based on a frequentist significance test, where the effects of systematic uncertainties are incorporated

⁴⁰ by use of the *profile likelihood ratio*. Section 2.3 covers the corresponding methods for setting limits. For

⁴¹ both discovery and exclusion one requires the sampling distribution of the statistic used in the test; this

1 is described in Section 2.4. Section 2.5 discusses a series of approximations used to determine expected

² values of the discovery significance and exclusion limits.

³ The approach taken in this note is to carry out tests for discovery and exclusion for fixed values of

4 the Higgs mass m_H . In principle the entire procedure is then repeated for all masses, resulting in limits

5 on or a measurement of m_H . In practice, an interpolation is made between finite steps in m_H .

6 2.1 The statistical model and likelihood function

First we consider the case of a single search channel. The measurement results in a set of numbers of events found in kinematic regions where signal could be present. These typically correspond to a histogram of a variable such as the mass of the reconstructed Higgs candidate, with the numbers of entries denoted by $\mathbf{n} = (n_1, \dots, n_N)$. In some cases one may consider a histogram with only one bin, i.e., the measured outcome is simply a number of candidate events found. The number of entries in bin *i*, n_i ,

12 is modeled as a Poisson variable with mean value

13

14

15

16

17

$$E[n_i] = \mu L \varepsilon_i \sigma_i \mathscr{B} + b_i \equiv \mu s_i + b_i , \qquad (1)$$

where *L* is the integrated luminosity, ε_i , σ_i and \mathscr{B} are the signal efficiency, Higgs cross section, and branching ratio, and b_i is the expected number of background events. Here μ is a signal strength parameter defined such that $\mu = 0$ corresponds to the absence of a signal; $\mu = 1$ gives the signal rate s_i expected from the Standard Model. If we consider a fixed Higgs mass m_H , the only parameter of interest is μ . All other adjustable parameters needed to specify the model are called *nuisance parameters*.

In principle the expected background values b_i can be predicted using Monte Carlo models for Standard Model processes. In the measurements considered here, however, the systematic uncertainty in the Standard Model prediction is in many cases quite large, and this would severely limit the sensitivity of the search. Therefore data regions where one expects only a very small amount of signal (*control regions*)

²² are used to constrain the background in the signal region (see also below).

For the *i*th bin of a histogram of a discriminating variable x, the expected signal and background can be written

$$s_i = s_{\text{tot}} \int_{\text{bin}\,i} f_s(x;\theta_s) \, dx \,,$$
 (2)

$$b_i = b_{\text{tot}} \int_{\text{bin}\,i} f_b(x;\theta_b) dx , \qquad (3)$$

where s_{tot} and b_{tot} are the total expected numbers of events in the histograms, $f_s(x;\theta_s)$ and $f_b(x;\theta_b)$ are the probability density functions (pdfs) of x for signal and background, and θ_s and θ_b represent sets of *shape parameters*.

The parametric forms of the pdfs $f_s(x; \theta_s)$ and $f_b(x; \theta_b)$ are determined from Monte Carlo simulations or data control samples. In the following we will use $\theta = (\theta_s, \theta_b, b_{tot})$ to refer to all of the nuisance parameters. The signal normalization s_{tot} here is not an adjustable parameter, but rather is fixed equal to the Standard Model prediction.

In addition to the measured histogram **n**, some search channels also make use of a set of subsidiary measurements $\mathbf{m} = (m_1, \dots, m_M)$ in control regions where one expects mainly background events. These can be modeled as being Poisson distributed with mean values

$$E[m_i] = u_i(\theta) , \qquad (4)$$

where the u_i are calculable quantities depending on a set of parameters, at least some of which are the

same as those entering into the predictions for s_i and b_i above. In practice the subsidiary measurements

are constructed so as to provide information on the background normalization b_{tot} and sometimes also on its shape.

³ If the measurement is based on counting events in a given kinematic region, i.e., without using the

shape of a distribution, in the formalism above the histograms have a single bin. The value $s = s_{tot}$ is

⁵ then the Standard Model prediction for the signal and $b = b_{tot}$ is the (unknown) expected background.

⁶ There are then no shape parameters, and b itself plays the role of θ as the single nuisance parameter. In

 τ this case the subsidiary measurement *m* is made in a control region where signal is absent (or can to good

⁸ approximation be neglected), and has an expectation value

$$E[m] = u = \tau b , \qquad (5)$$

⁹ where τ is a scaling constant whose value can be estimated from a Monte Carlo simulation.

¹⁰ The likelihood function is the product of Poisson probabilities for all bins:

$$L(\mu,\theta) = \prod_{j=1}^{N} \frac{(\mu s_j + b_j)^{n_j}}{n_j!} e^{-(\mu s_j + b_j)} \prod_{k=1}^{M} \frac{u_k^{m_k}}{m_k!} e^{-u_k} .$$
(6)

11 Equivalently the log-likelihood is

$$\ln L(\mu, \theta) = \sum_{j=1}^{N} \left(n_j \ln(\mu s_j + b_j) - (\mu s_j + b_j) \right) + \sum_{k=1}^{M} \left(m_k \ln u_k - u_k \right) + C,$$
(7)

where C represents terms that do not depend on the parameters and thus can be dropped. Here and in (6)

the parameters θ enter through Eqs. (2), (3), and (4).

In the case where the presence of signal in the histogram **n** gives a peak sitting on a smooth background, one does not need a subsidiary measurement **m**. Rather, as long as the number of parameters in the models for the signal and background distributions is smaller than the total number of bins measured, one can determine the strength parameter μ from the histogram **n** alone. Here the regions away from the peak (the sidebands) play the role of the subsidiary measurement by providing information on the background level. Of course if an additional subsidiary measurement is available, this will improve the accuracy of the background determination, which will increase the sensitivity of the analysis.

In the case of several independent search channels, the method described above is generalized in a straightforward manner. For each channel *i* there is a likelihood function $L_i(\mu, \theta_i)$. Its general form is given by Eq. (6), except that all quantities carry an additional index *i* to label the channel except the global strength parameter μ , which is assumed to be the same for all channels. Since the channels are statistically independent, the full likelihood function is given by the product

$$L(\mu, \theta) = \prod_{i} L_{i}(\mu, \theta_{i}) , \qquad (8)$$

where θ here represents all of the nuisance parameters.

Systematic uncertainties are effectively included in the analysis through the nuisance parameters θ . The model must be sufficiently flexible, i.e., it must contain enough parameters, so that for at least some point in its parameter space it can be regarded as representing the truth. One must exercise some restraint in achieving this, however, as an increasing number of nuisance parameters leads to a decrease in sensitivity to the parameters of interest. Some of the components of θ may be common among different channels, e.g., parameters relating to uncertainty in the integrated luminosity. These then represent a common (correlated) systematic uncertainty.

As an example, consider the signal efficiency ε that enters in the relation between the cross section and expected number of signal events. Suppose the efficiency has been estimated to have a value $\hat{\varepsilon}$ and systematic uncertainty $\sigma_{\hat{\varepsilon}}$. To incorporate this uncertainty into the model, we can regard the measured 1 value $\hat{\varepsilon}$ as a random variable whose true value ε is treated as a nuisance parameter. For the pdf $f_{\varepsilon}(\hat{\varepsilon};\varepsilon,\sigma_{\hat{\varepsilon}})$

² one could use, e.g., a Gaussian distribution centred about ε , or for a quantity such as the efficiency which

³ must lie in the range $0 \le \varepsilon \le 1$ one could use a pdf that automatically satisfies this constraint (e.g., a beta

⁴ distribution). For whatever choice is deemed appropriate, the likelihood (6) is multiplied by $f_{\varepsilon}(\hat{\varepsilon}; \varepsilon, \sigma_{\hat{\varepsilon}})$,

⁵ evaluated with the best estimate $\hat{\varepsilon}$, and the parameter ε is included in the set of nuisance parameters θ .

⁶ To test a hypothesized value of μ we construct the profile likelihood ratio,

$$\lambda(\mu) = \frac{L(\mu, \hat{\theta})}{L(\hat{\mu}, \hat{\theta})} \,. \tag{9}$$

⁷ Here $\hat{\theta}$ in the numerator denotes the value of θ that maximizes *L* for the specified μ , i.e., it is the ⁸ conditional maximum-likelihood estimator (MLE) of θ (and thus is a function of μ). The denominator ⁹ is the maximized (full) likelihood function, i.e., $\hat{\mu}$ and $\hat{\theta}$ are the MLEs. The presence of the nuisance ¹⁰ parameters broadens the profile likelihood ratio as a function of μ relative to what one would have if ¹¹ their values were fixed. This reflects the loss of information about μ due to the systematic uncertainties. ¹² The likelihood ratio (9) and procedures for incorporating systematic uncertainties applied here differ ¹³ somewhat from those used for the searches carried out at LEP. Some of these differences are discussed ¹⁴ further in Appendix A.

From the definition of the profile likelihood ratio one can see that $0 \le \lambda \le 1$, with $\lambda(\mu) = 1$ implying

¹⁶ good agreement between the data and the hypothesized value of μ . Equivalently it is convenient to work

17 with the quantity

$$q_{\mu} = -2\ln\lambda(\mu) , \qquad (10)$$

so that high values of q_{μ} correspond to poor agreement between the data and the hypothesized μ . The 18 statistic q_{μ} will have a sampling distribution $f(q_{\mu}|\mu')$. Here μ refers to the strength parameter used 19 to define the statistic q_{μ} , entering in the numerator of the likelihood ratio, and μ' is the value used to 20 define the data generated to obtain the distribution (i.e., the 'true' value). For the special case $\mu' = \mu$ and 21 for a sufficiently large data sample, the pdf $f(q_{\mu}|\mu)$ approaches a limiting form related to the chi-square 22 distribution, discussed further in Section 2.4. For $\mu' \neq \mu$, the distribution of q_{μ} is shifted to higher values, 23 reflecting the decreased agreement between the data generated with μ' and the hypothesis tested by q_{μ} , 24 as indicated in Fig. 1. The two cases of particular interest are $\mu = 0$, the background-only hypothesis, 25 and $\mu = 1$, the hypothesis of background plus signal present at the Standard Model rate. 26

The level of compatibility between data that give an observed value $q_{\mu,\text{obs}}$ for q_{μ} and a hypothesized value of μ is quantified by giving the *p*-value

$$p_{\mu} = \int_{q_{\mu,\text{obs}}}^{\infty} f(q_{\mu}|\mu) dq_{\mu} .$$
⁽¹¹⁾

This is the probability, under the assumption of μ , of seeing data with equal or greater incompatibility, as measured by q_{μ} , relative to the data actually obtained. This is illustrated in Fig. 1, where the shaded area indicates the *p*-value of the hypothesized μ . The figure also indicates the median value of q_{μ} under the assumption of a different value of the strength parameter μ' used to generate the data. For μ and μ' values that are increasingly different, the median med $[q_{\mu}|\mu']$ moves further to the right. An observed value of q_{μ} at this median would give a correspondingly small *p*-value for μ .

35 2.2 Establishing discovery

To establish discovery we try to reject the $\mu = 0$ (background-only) hypothesis, i.e., that there is no Higgs signal present. To do this we use the statistic $q_0 = -2\ln\lambda(0)$. One expects to find a low value



Figure 1: Illustration of the determination of the *p*-value of a hypothesized value of μ . The left-hand curve indicates the pdf of q_{μ} for data generated with the same value of μ as was used to define the statistic q_{μ} ; this is used to determine the *p*-value of μ , shown as the shaded region. The right-hand curve indicates the pdf of q_{μ} for data generated with a different value of the strength parameter, μ' .

of $\lambda(0)$ (high q_0) if the data include signal. Here even though one is testing the hypothesis that the Higgs does not exist, the definition of q_0 depends on the hypothesized Higgs mass m_H . It enters through the denominator of the likelihood ratio (9), which contains the maximum-likelihood estimator $\hat{\mu}$ for the strength of a Higgs signal at the mass m_H . By defining the test statistic in this way one maximizes the probability of rejecting the $\mu = 0$ hypothesis if the Higgs boson exists at the specified mass. This search procedure is then carried out for all values of m_H (in practice an interpolation is carried out between finite

⁷ steps in m_H).

⁸ A given data set will result in an observed value $q_{0,obs}$ of q_0 . The level of compatibility between the

9 data and the no-Higgs hypothesis is quantified by giving the *p*-value

$$p_0 = \int_{q_{0,\text{obs}}}^{\infty} f(q_0|0) \, dq_0 \,. \tag{12}$$

This is the probability, under the assumption of $\mu = 0$ (background only), of seeing data as signal-like or more so relative to the data actually obtained. A small value is interpreted as evidence against $\mu = 0$, i.e., a discovery of the signal.

One can define the *significance* corresponding to a given p-value as the number of standard deviations Z at which a Gaussian random variable of zero mean would give a one-sided tail area equal to p. That is,

the significance Z is related to the p-value by

$$p = \int_{Z}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-x^{2}/2} dx = 1 - \Phi(Z) , \qquad (13)$$

where Φ is the cumulative distribution for the standard (zero mean, unit variance) Gaussian. Equivalently one has

$$Z = \Phi^{-1}(1-p), \qquad (14)$$

where Φ^{-1} is the quantile of the standard Gaussian (inverse of the cumulative distribution). In (13) and (14) the subscript 0 was dropped as these relations hold for all *p*-values, not only those of the $\mu = 0$ hypothesis. The relation between *Z* and *p* is illustrated in Fig. 2.

A significance of Z = 5 corresponds to $p = 2.87 \times 10^{-7}$. For a sufficiently large data sample, one would obtain a *p*-value of 0.5 for data in perfect agreement with the expected background. With the definition of Z given above, this gives Z = 0. If the data fluctuate below the expected background, Z becomes negative.

²⁵ Note that according to the definition (14), a *p*-value of 0.05 corresponds to Z = 1.64. This should not ²⁶ be confused with a 1.96 σ fluctuation of a Gaussian variable that gives 0.05 for the two-sided tail area.

The significance of a discovery Z depends on the data obtained. To quantify our ability to discover a

²⁸ hypothesized signal in advance of seeing the data, we report the *median* significance under the assump-

tion that the signal is present at the Standard Model rate, $\mu = 1$. Since Z is a monotonic function of p_0 , and p_0 is also a monotonic function of q_0 , we have for the median significance, HIGGS – STATISTICAL COMBINATION OF SEVERAL IMPORTANT STANDARD MODEL HIGGS ...



Figure 2: Illustration of the correspondence between the significance Z and a p-value.

$$Z_{\text{med}} = \Phi^{-1}(1 - p_{0_{\text{med}}}) = \Phi^{-1}(1 - p_0(q_{0_{\text{med}}})).$$
(15)

This can be obtained from the median value of q_0 found using data generated under the assumption of $\mu = 1$.

³ A complete evaluation of the median significance is computationally difficult, as it requires a large

⁴ number of repeated simulations of the full set of experimental outputs and determination of $Z(q_0)$ from

5 the combination of all channels. Therefore in this note we have used the approximate methods described

6 in Section 2.5, which allow one to estimate quickly the median significance.

7 2.3 Setting limits

In addition to establishing discovery by rejecting the $\mu = 0$ hypothesis, we can consider the alternative hypothesis of some non-zero μ and try to reject it. A *p*-value is computed for each μ , and the set of μ values for which the *p*-value is greater than or equal to a fixed value 1 - CL form a confidence interval for μ , where typically one takes a *confidence level* CL = 95%. The upper end of this interval μ_{up} is the upper limit (i.e., $\mu \le \mu_{up}$ at 95% CL).

To compute the *p*-value for a hypothesized μ we first consider again the test statistic $q_{\mu} = -2 \ln \lambda(\mu)$ as initially defined in (9) and (10). For purposes of computing limits, we introduce a modification to this definition as described below.

If the data are incompatible with the hypothesized μ , one expects a large value of $-2\ln\lambda(\mu)$, i.e., $\lambda(\mu)$ close to zero. If a data set generated according to the hypothesis μ gives a large value of $-2\ln\lambda(\mu)$, this can be the result of either an upward or downward fluctuation in $\hat{\mu}$ relative to μ . This is illustrated in the scatterplot of $\hat{\mu}$ versus $-2\ln\lambda(\mu)$ shown in Fig. 3(a), which is from a toy Monte Carlo study with $\mu = 0.8$. The projection of the points on the $\hat{\mu}$ axis is shown in Fig. 3(b). Note that $\hat{\mu} \ge 0$ is imposed; the reasons for and consequences of this requirement are discussed in Section 2.4.

For purposes of setting an upper limit, however, we want to determine the smallest μ such that there is a fixed small probability (one minus the confidence level) to find data as compatible with that value of μ or less, relative to the degree of compatibility found with the real data. Therefore the data with upward fluctuations in $\hat{\mu}$ are not counted when computing the *p*-value, because they would be compatible with some larger μ . Therefore for purposes of computing limits we redefine q_{μ} to be¹

$$q_{\mu} = \begin{cases} -2\ln\lambda(\mu) & \hat{\mu} \le \mu ,\\ 0 & \text{otherwise.} \end{cases}$$
(16)

The distribution of the new q_{μ} thus corresponds to the lower branch only of the U-shaped scatterplot shown in Fig. 3(a).

¹Equivalently, one could retain the definition $q_{\mu} = -2\ln(L(\mu, \hat{\theta})/L(\hat{\mu}, \hat{\theta}))$ by placing an upper bound on $\hat{\mu}$ equal to μ , i.e., by imposing $0 \le \hat{\mu} \le \mu$. In this way, when $\hat{\mu} = \mu$ then one has $q_{\mu} = 0$ just as in the case of discovery when testing $\mu = 0$.

HIGGS - STATISTICAL COMBINATION OF SEVERAL IMPORTANT STANDARD MODEL HIGGS ...



Figure 3: (a) Scatterplot of $\hat{\mu}$ versus $-2\ln\lambda(\mu)$; (b) the distribution of $\hat{\mu}$ (see text).

Using the new definition (16), the *p*-value is given by the integral of $f(q_{\mu}|\mu)$ from the observed value $q_{\mu,\text{obs}}$ to infinity as in Eq. (11) and as illustrated in Fig. 1. The *p*-value is computed in this manner for all values of μ , and the upper limit μ_{up} at 95% confidence level is the largest value of μ for which the *p*-value is at least 0.05.

The result can be summarized by giving the upper limit on μ as a function of the Higgs mass m_H . Specifically, if we can reject the hypothesis $\mu = 1$ at a certain confidence level, then the corresponding value of m_H is regarded as excluded for a Standard Model Higgs. The lowest mass value not excluded is the lower limit m_{lo} .

One is also interested in the median limit under the assumption that there is no Higgs. As in the case of the discovery significance, a full calculation of the median limit is difficult as it requires a large number of repeated simulations based on the full profile likelihood ratio. For purposes of this note, therefore, we use the approximation techniques described in Section 2.5.

13 2.4 Sampling distribution of the likelihood ratio

¹⁴ To determine the *p*-values required for both discovery and exclusion we need the sampling distribution, ¹⁵ assuming data generated according to a given value of μ , of the statistic q_{μ} , i.e., $f(q_{\mu}|\mu)$. For the case of ¹⁶ discovery significance we use $q_0 = -2\ln\lambda(0)$, and for setting limits we use $q_{\mu} = -2\ln\lambda(\mu)$ for $\hat{\mu} \le \mu$ ¹⁷ and $q_{\mu} = 0$, otherwise.

¹⁸ To claim discovery we require *p*-values for $\mu = 0$ down to around 10^{-7} , and therefore to do this with ¹⁹ a Monte Carlo simulation requires an extremely large number of simulated measurements. In practice ²⁰ this is only carried out for simple test cases. Even for setting limits at 95% confidence level, it is often ²¹ not practical to use Monte Carlo.

²² Under a set of regularity conditions and for a sufficiently large data sample, *Wilks' theorem* says that ²³ for a hypothesized value of μ , the pdf of the statistic $-2\ln\lambda(\mu)$ approaches the chi-square pdf for one ²⁴ degree of freedom [2]. More generally, if there are *n* parameters of interest, i.e., those parameters that do ²⁵ not get a double hat in the numerator of the likelihood ratio (9), then $-2\ln\lambda(\mu)$ asymptotically follows ²⁶ a chi-square distribution for *n* degrees of freedom. A proof and details of the regularity conditions can ²⁷ be found in standard texts such as [3].

In the searches considered here, the data samples are generally large enough to ensure the validity of the asymptotic formulae for the likelihood-ratio distributions. In our case, however, the distributions are modified because of constraints imposed on the expected number of events.

³¹ Usually when searching for a new type of particle reaction one regards the mean number of events ³² contributed to any bin from any source, signal or background, to be greater than or equal to zero. In 1 some analyses it could be meaningful to consider a new effect that suppresses the expected number

² of events, e.g., the presence of a new decay channel could mean that the number of decays to known

channels is reduced. Here, however, we will regard any contribution to an expected number of events as
 non-negative.

5 Assuming only non-negative event rates, the maximum-likelihood estimators for the parameters are

⁶ constrained, e.g., $\hat{\mu} \ge 0$. As a consequence, if the observed number of events is below the level predicted

⁷ by the background alone, then the maximum of the likelihood occurs for $\mu = 0$, i.e., negative μ is not

⁸ allowed. We can consider the effect of having $\hat{\mu} = 0$ on the distribution of q_{μ} for two cases: $\mu = 0$ and

9 $\mu > 0.$

For $\mu = 0$, i.e., when computing the discovery significance, if $\hat{\mu} = 0$ one has (see (9)),

$$\lambda(0) = \frac{L(0,\hat{\theta})}{L(\hat{\mu},\hat{\theta})} = \frac{L(0,\hat{\theta})}{L(0,\hat{\theta})} = 1,$$
(17)

since $\hat{\mu} = 0$ and therefore $\hat{\theta} = \hat{\theta}$. The statistic $q_0 = -2\ln\lambda(0)$ is therefore equal to zero. This can be seen in the scatterplot of q_0 versus $\hat{\mu}$ in Fig. 4(a). Figure 4(b) shows the corresponding q_0 distribution with the peak visible at $q_0 = 0$. The superimposed curve is a chi-square distribution multiplied by one half, corresponding to the half of the events with $\hat{\mu} > 0$.



Figure 4: (a) Scatterplot of $\hat{\mu}$ versus q_{μ} from a Monte Carlo study with $\mu = 0$; (b) the distribution of q_0 (see text).

From Fig. 4(b) one can see that except for the spike at $q_0 = 0$ (when $\hat{\mu} = 0$), the pdf of q_0 can be well approximated by the chi-square pdf. Assuming a fraction *w* for the cases with $\hat{\mu} > 0$ one has the pdf

$$f(q_0|0) = w f_{\chi_t^2}(q_0) + (1-w)\delta(q_0) .$$
(18)

¹⁷ In the usual case where upward and downward fluctuations of $\hat{\mu}$ are equally likely we have w = 1/2. The

¹⁸ *p*-value of the background-only hypothesis given an observation $q_{0,obs}$ greater than zero is therefore

$$p = \int_{q_{0,\text{obs}}}^{\infty} w f_{\chi_1^2}(q_0) dq_0 = w(1 - F_{\chi_1^2}(q_{0,\text{obs}})) , \qquad (19)$$

where $F_{\chi_1^2}$ is the cumulative chi-square distribution for one degree of freedom.

The second case to consider is $\mu > 0$, e.g., when one wants to set an upper limit on μ . Under the

hypothesis μ , one obtains $\hat{\mu} > \mu$ and $\hat{\mu} \le \mu$ with approximately equal probability. Figure 5 shows the distributions of q_{μ} for both cases $\hat{\mu} > \mu$ and $\hat{\mu} \le \mu$ obtained from the scatterplot Fig. 3(a), from a Monte

²³ Carlo study with $\mu = 0.8$.

HIGGS - STATISTICAL COMBINATION OF SEVERAL IMPORTANT STANDARD MODEL HIGGS ...



Figure 5: Distributions of $-2\ln\lambda(\mu)$ for (a) $\hat{\mu} > \mu$ and (b) $\hat{\mu} \le \mu$. The superimposed curves are chi-square distributions for one degree of freedom normalized to half the number of entries in the original distribution (see Fig. 3(a)).

From Fig. 5(a) one can see that for $\hat{\mu} > \mu$, the data follow the chi-square pdf quite accurately. This portion of the distribution is ignored, however, when setting upper limits on μ , because of the modified definition of q_{μ} (20) used for limits,

$$q_{\mu} = \begin{cases} -2\ln\lambda(\mu) & \hat{\mu} \le \mu ,\\ 0 & \text{otherwise.} \end{cases}$$
(20)

⁴ Suppose now $\hat{\mu} \leq \mu$ with a probability *w*; in practice this is close to one half. (Note for the case of q_0 , ⁵ *w* is the probability of $\hat{\mu} > \mu$. The different definitions of *w* are used so as to give similar forms for ⁶ $f(q_0|0)$ and $f(q_{\mu}|\mu)$.) Thus for $\hat{\mu} > \mu$ one has from (20) $q_{\mu} = 0$, and therefore the distribution has a ⁷ delta function at $q_{\mu} = 0$ with weight 1 - w. The pdf of $f(q_{\mu}|\mu)$ can therefore be written

$$f(q_{\mu}|\mu) = wf(q_{\mu}|\mu, \hat{\mu} \le \mu) + (1 - w)\delta(q_{\mu}).$$
(21)

⁸ where $f(q_{\mu}|\mu, \hat{\mu} \leq \mu)$ is the conditional pdf for q_{μ} given $\hat{\mu} \leq \mu$.

For $\hat{\mu} \le \mu$, one may sometimes find $\hat{\mu}$ equal to zero, i.e., the lower edge of the allowed range, as can be seen in the scatterplot of $\hat{\mu}$ versus q_{μ} shown in Fig. 3(a). Although for the case $\mu = 0$ this gave a peak at $q_0 = 0$, here it gives

$$\lambda(\mu) = \frac{L(\mu, \hat{\theta})}{L(\hat{\mu}, \hat{\theta})} = \frac{L(\mu, \hat{\theta})}{L(0, \hat{\theta})}, \qquad (22)$$

which in contrast to (17) is not equal to unity. The effect of having $\hat{\mu} = 0$ on the distribution of q_{μ} is therefore more complicated than was the case for q_0 .

In general for $\hat{\mu} \le \mu$, the distribution of q_{μ} falls off more steeply than the chi-square distribution. This is seen in Fig. 5(b). Therefore a *p*-value based on the chi-square formula will be larger than the true *p*-value, and the corresponding significance Z will be smaller. The upper limits obtained for μ are therefore larger, i.e., a smaller set of μ values is excluded.

If μ is sufficiently large, then $\hat{\mu}$ is very rarely pushed to zero and $f(q_{\mu}|\mu, \hat{\mu} \le \mu)$ approaches a chi-square distribution for one degree of freedom. For purposes of the present study, the chi-square approximation is adequate, but gives somewhat conservative limits. That is, we take the distribution of q_{μ} to be

²¹ q_{μ} to be

$$f(q_{\mu}|\mu) = w f_{\gamma^{2}}(q_{\mu}) + (1 - w)\delta(q_{\mu})$$
(23)

and use w = 0.5. One has therefore the same pdf for q_{μ} using the modified definition (20) as was found in (18) for q_0 based on the original definition, $q_0 = -2 \ln \lambda(0)$.

- To summarize the result above, the pdf of q_{μ} can be approximated by a mixture of a chi-square pdf for one degree of freedom with weight *w* and a delta function at zero with weight 1 - w. This holds both
- 5 for discovery ($\mu = 0$) and setting limits ($\mu > 0$).
- 6 Consider now the variable

$$u = \sqrt{q_{\mu}} = \sqrt{-2\ln\lambda(\mu)} , \qquad (24)$$

7 which has the pdf

$$f(u) = \Theta(u)w\sqrt{\frac{2}{\pi}}e^{-u^2/2} + (1-w)\delta(u), \qquad (25)$$

⁸ where $\Theta(u) = 1$ for $u \ge 0$ and is zero otherwise. The second term in (25) follows from the fact that the

9 values $q_0 = 0$ and u = 0 occur with equal probability, 1 - w. Furthermore if a variable *x* follows the 10 standard Gaussian distribution, then one can show x^2 follows a chi-square distribution for one degree of 11 freedom. Therefore if x^2 follows a χ^2 distribution, then $\sqrt{x^2}$ follows a Gaussian scaled up by a factor of 12 two for x > 0 so as to have a total area of unity.

¹³ The *p*-value of the hypothesis μ for a non-zero observation $q_{\mu,\text{obs}}$ is therefore

$$p = P(q_{\mu} \ge q_{\mu,\text{obs}}) = P(u \ge \sqrt{q_{\mu,\text{obs}}}) = 2w \int_{\sqrt{q_{\mu,\text{obs}}}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du = 2w(1 - \Phi(\sqrt{q_{\mu,\text{obs}}})) .$$
(26)

¹⁴ Combining this with Eq. (14) for the significance Z gives

$$Z = \Phi^{-1}(1 - 2w(1 - \Phi(\sqrt{q_{\mu,\text{obs}}}))) .$$
(27)

In the usual case where the weights of the chi-square and delta-function terms are equal, i.e., w = 1/2, Eq. (27) reduces to to the simple formula

¹⁶ Eq. (27) reduces to to the simple formula

$$Z = \sqrt{q_{\mu,\text{obs}}} . \tag{28}$$

17 2.5 Approximate methods

¹⁸ To determine the discovery significance or to set limits using a given data set, one must carry out the ¹⁹ global fit described above. For this one needs first to combine the likelihood functions for the individual ²⁰ channels into the full likelihood function containing a single strength parameter μ , and use this to find ²¹ the profile likelihood ratio. It is possible, however, to find approximate values for the *median* discovery ²² significance and limits in a way that only requires as input the separate profile likelihood ratio values ²³ from each of the channels. This is very useful especially in the planning phase of a search that combines ²⁴ multiple channels.

The procedure relies on two separate approximations. First, we estimate the median value of the pro-25 file likelihood ratio $\lambda(\mu)$ by evaluating the likelihood function with a single, artificial data set in which 26 all statistical fluctuations are suppressed, as described in Section 2.5.1. Second, to determine the signif-27 icance values from the likelihood ratios, we use the asymptotic form of the distribution of $-2\ln\lambda(\mu)$ 28 valid for sufficiently large data samples. This is described in Section 2.5.2, and its validity is checked 29 for the individual channels in Section 3. Here the limitations of the approximation are investigated and 30 for one case where it is found to be insufficiently accurate (the discovery significance for the channel 31 $H \rightarrow W^+W^-$ plus no jets), an alternate procedure is followed. 32

1 2.5.1 Approximation for the median likelihood ratio

² To find the median discovery significance and limits, the median likelihood ratios $\lambda_i(\mu)$ are first found ³ for each channel separately, and then combined to give the full median likelihood ratio. One can estimate ⁴ the median $\lambda_i(\mu)$ by the value of the likelihood ratio evaluated with a single artificially constructed data ⁵ set, in which all statistical fluctuations are suppressed and the data values **n** and **m** are replaced by their ⁶ expectation values for a given integrated luminosity and a hypothesized strength parameter μ_A . We refer ⁷ to this as an 'Asimov' data set.² It replaces having to simulate a large number of experiments from which ⁸ one would determine the median.

As before, $\mu_A = 0$ is the background only hypothesis and $\mu_A = 1$ corresponds to background plus signal present at the Standard Model rate. The median referred to thus pertains to what one would obtain with a large number of experiments generated under the assumption of μ_A . The approximation is in fact more accurate if one uses noninteger values for numbers of events in the log-likelihood (the factorial terms are in any case absent) so that the Asimov likelihood L_A is found by substituting

$$n_j = \mu_A s_j + b_j \tag{29}$$

$$m_k = u_k, \qquad (30)$$

into the likelihood function (6) for each channel. Here for s_j , b_j and u_k , one needs in principle the 14 expectation values, i.e., these quantities should have no statistical errors. In practice they are estimated 15 using a Monte Carlo sample corresponding to an integrated luminosity substantially larger than what is 16 considered for the data. The numbers of signal and background events are then scaled to the desired 17 luminosity. The other nuisance parameters such as shape parameters are estimated as would be done 18 with any other data set; we refer below to the resulting values as θ_A . Because the Asimov data set has 19 no statistical fluctuations, the θ_A are simply the values one would derive from a very large Monte Carlo 20 data sample. 21

The estimate of the median likelihood ratio used for the *i*th channel is therefore

$$\lambda_{A,i}(\mu) = \frac{L_{A,i}(\mu,\hat{\theta})}{L_{A,i}(\hat{\mu},\hat{\theta})} \approx \frac{L_{A,i}(\mu,\hat{\theta})}{L_{A,i}(\mu_{A},\theta_{A})}, \qquad (31)$$

where $L_{A,i}$ denotes the likelihood function (6) evaluated with the Asimov data values (29) and (30). The approximation used for the final step in (31) exploits the fact that ML estimate of $\hat{\mu}$ is very close to the input value μ_A when the likelihood function is constructed using the Asimov data set.

Note that if the likelihood functions for the individual channels were to be constructed with data containing statistical fluctuations rather than with the artificial Asimov data, then the ML estimate of the strength parameter, $\hat{\mu}$, would in general be different for each channel. The full likelihood function (8) used for the combination, however, contains a single global μ . We can now exploit the fact that for the Asimov data one has $\hat{\mu} \approx \mu_A$ for all of the channels and thus obtain the median likelihood ratio for the combination as the product of the individual $\lambda_{A,i}(\mu)$,

$$\lambda_A(\mu) = \prod_i \lambda_{A,i}(\mu) . \tag{32}$$

³² Monte Carlo studies show that Eq. (32) provides an excellent approximation to the median value one ³³ would find from data generated with μ_A as the strength parameter.

²The name of the Asimov data set is inspired by the short story *Franchise*, by Isaac Asimov [1]. In it, elections are held by selecting a single voter to represent the entire electorate.

For purposes of quantifying how likely we are to discover the Higgs if it exists, we report the significance obtained from the *p*-value of $\mu = 0$ with an Asimov data set that corresponds to $\mu_A = 1$,

$$\lambda_{s+b}(0) = \prod_{i} \frac{L_{s+b,i}(0,\hat{\theta})}{L_{s+b,i}(1,\theta_{\rm A})} \,. \tag{33}$$

³ Here the subscript s+b refers to the Asimov data set; the argument 0 denotes the value of μ being tested. ⁴ That is, Eq. (33) approximates what one would obtain with data generated with signal and background ⁵ for the median value of $\lambda(0)$, which is used to test the background-only ($\mu = 0$) hypothesis. Equation

6 (33) provides the median $q_{0_{\text{med}}} = -2 \ln \lambda_{s+b}(0)$, and from this the *p*-value and significance Z are found 7 using equation (15).

⁸ To determine the limits on μ that we expect to set if the Higgs does not exist (or is beyond our ⁹ reach), we find the *p*-value of a hypothesized μ using the likelihood ratio $\lambda(\mu)$ based on Asimov data ¹⁰ for background only ($\mu_A = 0$),

$$\lambda_b(\mu) = \prod_i \frac{L_{b,i}(\mu,\hat{\theta})}{L_{b,i}(0,\theta_{\rm A})} \,. \tag{34}$$

¹¹ That is, $\lambda_b(\mu)$ approximates the median value of $\lambda(\mu)$ one would obtain from data generated according ¹² to the background-only hypothesis. The value of $\lambda_b(\mu)$ is used to determine the median $q_{\mu_{med}}$, which is ¹³ used to find the median *p*-value, $p_{\mu_{med}}$. This is computed for all μ and the point where $p_{\mu_{med}} = 0.05$ gives ¹⁴ the 95% CL upper limit.

Because $\hat{\mu} \approx \mu_A$ holds for each channel individually when using Asimov data, it is possible to determine the values of the likelihood ratio entering into (32) separately for each channel, which simplifies greatly the task of estimating the median significance that would result from the full combination. It should be emphasized, however, that the discovery significance or exclusion limits determined from real data require one to construct the full likelihood function containing a single parameter μ , and this must be used in a global fit to find the profile likelihood ratio.

Furthermore, some systematic errors, e.g., the uncertainty in the integrated luminosity, are common to all channels and correspond to a common nuisance parameter. When using Asimov data, the values of such parameters will be fitted to the same values in all channels. Thus the correlations between common systematics are taken into account just as they would be in a global fit of all channels.

A limitation of the procedure with Asimov data is that it only provides an estimate of the median likelihood ratio. To obtain an uncertainty band on the expected (median) discovery and exclusion sensi-

tivities as a function of m_H one would have to simulate a large number of experiments.

28 2.5.2 Approximate relation between likelihood ratio and significance

To compute the *p*-values we need the distribution $f(q_{\mu}|\mu)$ of $q_{\mu} = -2\ln\lambda(\mu)$. For a sufficiently large data sample the pdf of q_{μ} takes on a well defined limiting form related to the chi-square distribution, as discussed in Section 2.4. Assuming this form, the *p*-value of the hypothesis μ is found to be

$$p_{\mu} \approx 1 - \Phi(\sqrt{q_{\mu}}) , \qquad (35)$$

³² and the significance Z is given by the formula

$$Z = \Phi^{-1}(1 - p_{\mu}) \approx \sqrt{-2\ln\lambda(\mu)} . \tag{36}$$

For estimating the median discovery significance we use equations (35) and (36) together with the equation (33), the likelihood ratio based on Asimov data containing signal and background. To find the median limit on μ , we use again equations (35) and (36) but with the likelihood ratio based on Asimov

² background data only, equation (34).

³ The validity of this approximation is investigated for each channel by generating distributions of

⁴ q_{μ} for $\mu = 0, 1$ using a fast Monte Carlo simulation and comparing the resulting histograms with the

5 expected asymptotic form. These comparisons are shown in Section 3.

6 2.6 Consequences of testing many Higgs mass values

⁷ The statistical significance of a potential discovery is quantified by giving the *p*-value of the no-Higgs ⁸ hypothesis, i.e., the probability, under the assumption of background only, that that one would see data ⁹ with equal or less compatibility with this hypothesis relative to the data obtained. Finding this proba-¹⁰ bility below a specified threshold (e.g., the 5σ threshold, or $p < 2.87 \times 10^{-7}$) corresponds to claiming ¹¹ discovery of a Higgs boson.

The approach taken in this analysis is to compute the *p*-value of the no-Higgs hypothesis separately as a function of the Higgs mass. The threshold *p*-value is thus the false discovery rate for Higgs boson of a given mass. Further one should also estimate the probability, under the assumption of background only, that this *p*-value will fall below the discovery threshold for *any* mass within the range considered. By searching for the Higgs within a broad range of hypothetical masses, one increases the probability of observing what appears to be a signal at some mass, and so the effective significance of the discovery is reduced. In HEP this is sometimes referred to as the "look-elsewhere effect".

To first approximation the effective increase in the false-discovery rate is given by the number of statistically separate mass ranges explored. If a certain data set would give a *p*-value of $m_{\rm H}$ below the discovery threshold, then the same data would in general also indicate discovery for other masses very close by. Roughly speaking, the mass range in which a given data set would indicate discovery is set by the mass resolution for the Higgs candidate. So the factor by which the *p*-value is inflated is given by the mass range explored divided by the average mass resolution. Monte Carlo studies can be used to validate and refine this approximation; this approach is planned for future analyses.

An alternative to considering fixed Higgs masses is to treat both the strength parameter μ and the 26 Higgs mass $m_{\rm H}$ as free parameters in the likelihood ratio. For example, to establish discovery one com-27 putes the *p*-value of the no-Higgs hypothesis. As before, this is the probability, under the assumption of 28 no Higgs, of finding data with equal or lesser compatibility with $\mu = 0$ relative to the data obtained. In 29 contrast to the fixed-mass case, however, "less compatible" here means having a lower likelihood ratio 30 for any allowed value of the Higgs mass; the lowest value comes when the denominator contains the 31 fitted maximum-likelihood estimator $\hat{m}_{\rm H}$. In practice the fitted value of the Higgs mass is restricted to 32 lie within a stated range. This has been done for the Higgs searches using the $\gamma\gamma$ [4] and W⁺W⁻ [5] 33 channels, with the aim of extending this method to a combination of all channels. 34

35 3 Combination of Higgs search channels

In this section a brief description of each of the four search channels is given. For each channel, the method used to obtain the likelihood ratio is described, and values of the test variable q_{μ} as defined by Eq. (10) for discovery and by Eq. (16) for limits are tabulated for several values of the integrated luminosity *L* and Higgs mass $m_{\rm H}$. For the discovery sensitivity where one tests $\mu = 0$, the median value of q_0 is given under the assumption of $\mu = 1$; for exclusion sensitivity, the median of q_1 is given under the assumption of $\mu = 0$.

In addition, for each channel we show distributions of q_{μ} under the assumption of μ for the two cases $\mu = 0$ and $\mu = 1$. For the approximations used in this note to be valid, these should be close to the asymptotic form described in Section 2.4. This limiting form for the distribution $f(q_{\mu}|\mu)$ is a mixture of HIGGS - STATISTICAL COMBINATION OF SEVERAL IMPORTANT STANDARD MODEL HIGGS ...

a delta function at $q_{\mu} = 0$ and a chi-square distribution for one degree of freedom, where each component

² has equal weight. We refer to this as a $\frac{1}{2}\chi_1^2$ distribution. For the case of $\mu = 0$ we compare directly the

³ Monte Carlo distribution of q_0 with the $\frac{1}{2}\chi_1^2$ distribution and the delta-function term at $q_0 = 0$ is clearly

4 visible. For $\mu = 1$ we show the equivalent comparison but for reasons of convenience only the events

⁵ with $\hat{\mu} \leq \mu$ are shown, i.e., the events with $\hat{\mu} > \mu$ that contribute to the delta function at $q_1 = 0$ are left ⁶ off. That is, for exclusion it is the conditional pdf $f(q_1|\mu = 1, \hat{\mu} \leq 1)$ that is compared to a chi-square

⁷ distribution for one degree of freedom; there is no delta function term.

All channels use data driven background estimation methods. This way, the uncertainties in the back-8 ground shape and normalization are treated within the framework of the profile likelihood as nuisance g parameters. Using control samples the effect of many uncertainties like energy scales and fake rates on 10 the background estimate can be constrained by the control samples. Uncertainties on the signal efficiency 11 do not affect the discovery sensitivity which is testing the presence or absence of a signal; however, this 12 is not the case for exclusion sensitivity. As one would expect, uncertainty in the signal efficiency does 13 reduce the exclusion sensitivity. This uncertainty was incorporated into the profile likelihood calculation 14 by adding an extra term to the likelihood function for every channel as described in Section 2.1: a Gaus-15 sian relating the nominal efficiency estimated in an auxiliary measurement, the true efficiency, and the 16 uncertainty of that auxiliary measurement. 17

18 3.1 $H \rightarrow \gamma \gamma$

Details on the $H \rightarrow \gamma \gamma$ channel are given in Ref. [4]. We perform an unbinned maximum-likelihood fit to extract the signal and background event yields by using the diphoton invariant mass, $m_{\gamma\gamma}$, as a discriminating variable. The $H \rightarrow \gamma \gamma$ distribution of $m_{\gamma\gamma}$ forms a Gaussian peak with tails to lower values from photon energy losses before the calorimeter. It is well modelled by a *Crystal Ball* function. The signal probability density function, $p_H(m_{\gamma\gamma})$, is given by

$$p_{H}(m_{\gamma\gamma}) = N \cdot \begin{cases} \exp\left(-t^{2}/2\right), & \text{for } t > -\alpha, \\ (n/|\alpha|)^{n} \cdot \exp\left(-|\alpha|^{2}/2\right) \cdot (n/|\alpha| - |\alpha| - t)^{-n}, & \text{otherwise}, \end{cases}$$
(37)

where $t = (m_{\gamma\gamma} - m_H - \delta m_H) / \sigma(m_{\gamma\gamma})$, *N* is a normalisation parameter, m_H is the Higgs boson mass, δm_H is an offset and σ represents the diphoton invariant mass resolution. The non-Gaussian tail is parametrised by *n* and α . We include an additional, broader Gaussian term in Eq. 37 to improve the description of the tails of the distribution. Within a sufficiently narrow mass window, the background of $m_{\gamma\gamma}$ is modeled by an exponential distribution with a single slope parameter ξ .

The resulting median profile likelihood ratios for discovery, $\lambda(\mu = 0)$ (using toy s + b Monte Carlo experiments and taking the median of the $\lambda(\mu = 0)$ distribution) are given in Table 1 for a few Higgs masses at some given luminosities.

The distribution of the test statistic q_0 under the null background only hypothesis, for $m_H = 120 \text{ GeV}$ with an integrated luminosity of 2 and 10 fb⁻¹, is shown in Fig. 6. A $\frac{1}{2}\chi_1^2$ distribution is superimposed, showing the validity of the asymptotic approximation.

The median profile likelihood ratio for exclusion, $-2\ln\lambda(\mu)$ (using toy background-only Monte Carlo experiments and taking the median of the $\lambda(\mu)$ distribution) is given in Table 2 for a few Higgs masses at several integrated luminosities and for a signal strength $\mu = 1$, corresponding to a Standard Model Higgs Boson.

The distribution of the test statistic q_1 for $\hat{\mu} \le 1$ under the s + b hypotheses for $m_H = 150$ GeV with an integrated luminosity of 2 and 10 fb⁻¹ is shown in Figures 7. A χ_1^2 distribution is superimposed, showing the validity of the asymptotic approximation.

L	$m_H (\text{GeV})$								
$(\mathbf{f}\mathbf{b}^{-1})$	115	120	130	140					
1	0.35	0.55	0.75	0.67					
2	0.75	1.07	1.45	0.95					
5	1.95	2.95	3.65	2.55					
10	3.95	5.86	7.35	5.05					
30	11.85	17.72	21.99	15.05					

Table 1: Median values of $-2\ln\lambda(\mu)$ (evaluated at $\mu = 0$) obtained from fits to simulated data generated with $H \rightarrow \gamma\gamma$ signal plus background ($\mu = 1$) for several values of the Higgs mass and integrated luminosity.



Figure 6: The distribution of the test statistic q_0 (for $H \to \gamma \gamma$), under the null background only hypothesis, for $m_H = 120 \text{ GeV}$ with an integrated luminosity of 2 (a) and 10 (b) fb⁻¹. A $\frac{1}{2}\chi_1^2$ distribution is superimposed.

Table 2: Median values of $-2\ln\lambda(\mu)$ (evaluated at $\mu = 1$) obtained from $H \to \gamma\gamma$ background-only ($\mu = 0$) simulated data for several values of the Higgs mass and integrated luminosities.

L	m_H (GeV)								
$(\mathbf{f}\mathbf{b}^{-1})$	115	120	130	140					
1	0.72	0.73	0.91	0.67					
2	0.97	1.21	1.39	0.97					
5	2.11	2.59	3.13	2.23					
10	3.55	4.87	5.71	4.04					
30	8.47	10.50	11.63	9.00					



Figure 7: The distribution of the test statistic q_1 for $\hat{\mu} \le 1$ under the s + b hypothesis (for $H \to \gamma\gamma$), for $m_H = 120$ GeV with an integrated luminosity of (a) 2 fb⁻¹ and (b) 10 fb⁻¹. A χ_1^2 distribution is superimposed.

1 3.2 $H \rightarrow W^+W^-$

² The $H \rightarrow W^+W^-$ search is divided into two topologies, production of a Higgs with no jets (H + 0j) and ³ with two additional jets (H + 2j), using in both cases the decay mode $H \rightarrow WW \rightarrow ev\mu v$. The present ⁴ study does not yet consider the final states *evev* or $\mu v \mu v$, nor those with hadronic W decays. Future ⁵ inclusion of these channels is expected to improve the search sensitivity particularly for the high Higgs ⁶ mass region. The search is described in detail in Ref. [5].

7 **3.2.1** H + 0j

⁸ The analysis of the H + 0j channel uses a two dimensional maximum-likelihood fit of the transverse ⁹ mass and the transverse momentum of the WW system in two bins of the dilepton opening angle in the ¹⁰ transverse plane. The fit includes control samples to measure the backgrounds from $t\bar{t}$ and $Z \rightarrow \tau\tau$.

The QCD WW background requires particular attention. Its distributions of Higgs-candidate transverse mass and $p_{\rm T}$ are described with functions containing several adjustable (nuisance) parameters, and several others whose values are determined from a full Monte Carlo simulation and thereafter treated as fixed. The distribution of the test statistic q_0 under the background-only ($\mu = 0$) hypothesis is shown in Fig. 8(a) for $m_H = 150$ GeV for an integrated luminosity of 10 fb⁻¹. The same fixed QCD WW shape parameters are used both to generate the data and for calculating the likelihood ratio. A $\frac{1}{2}\chi_1^2$ distribution is superimposed, showing the level of agreement of the asymptotic approximation.

For this channel, further investigation of the systematic uncertainties was carried out. For the fixed shape parameters related to p_T and transverse mass distributions for the QCD WW background, the values used to generate the data were varied relative to what was used when determining the likelihood ratio. This was done in a manner that minimized the sensitivity of the resulting q_0 distribution to variations in other fixed parameters such as the QCD Q^2 scale. The resulting distributions of q_0 are thus no longer expected to follow the $\frac{1}{2}\chi_1^2$ form, as can be seen in Fig. 8(b).

Because the chi-square approximation is not valid in this case, the *p*-values are calculated using the q_0 distribution obtained directly from the Monte Carlo. An exponential is fitted to the tail region in order to extrapolate to large q_0 values, and the median value of q_0 under the hypothesis of signal plus background is determined using the same variation of the background parameters. It was found that the median *p*-value of the background-only hypothesis, with the median computed under assumption of the s + b hypothesis, is very similar to the original case where the QCD shape parameters are not varied and the $\frac{1}{2}\chi_1^2$ distribution is used.



Figure 8: The distribution of the test statistic q_0 for $H + 0j \rightarrow WW + 0j$, under the background-only hypothesis, with the same fixed QCD WW shape parameters used at both the generator and the fit level, for $m_H = 150$ GeV and for an integrated luminosity of 10 fb⁻¹ (a) with the same shape parameters for event generation and fitting; (b) with altered shape parameters. A $\frac{1}{2}\chi_1^2$ distribution is superimposed.

For the combination of results for discovery (i.e., testing $\mu = 0$), we have used the *p*-values as 1 described above for the case of the H + 0i channel. The same variation of QCD WW parameters was 2 also investigated for the case of exclusion (i.e., testing $\mu > 0$, and in particular $\mu = 1$), and it was done 3 as well for the H + 2j channel. In those studies, however, the distributions, under the assumption of 4 μ , of the test statistic q_{μ} were found to agree quite well with the expected $\frac{1}{2}\chi_1^2$ distribution, even after 5 the parameter variation. Therefore for these cases we have based the combination of results on the 6 asymptotic approximations for the q_{μ} distributions (as done in this paper for the other Higgs channels). 7 To simplify the comparison with the other channels, the median *p*-values of the background-only 8 $(\mu = 0)$ hypothesis for the H + 0j channel were converted into effective values of the variable $q_0 =$ 9

 $(\mu = 0)$ hypothesis for the H + 0j channel were converted into effective values of the valuable $q_0 = -2\ln\lambda(0)$ according to $q_0 = Z^2 = (\Phi^{-1}(1-p))^2$. These are given in Table 3 for several Higgs masses and integrated luminosities.

Table 3: Median values of $-2\ln\lambda(\mu = 0)$ obtained from H + 0j fits using data simulated under the assumption of signal plus background ($\mu = 1$) for several values of the Higgs mass and integrated luminosity.

L	m_H (GeV)										
(fb^{-1})	130	140	150	160	170	180	190				
1	1.64	4.30	9.56	16.52	15.39	7.37	3.17				
2	2.87	8.60	19.36	34.67	30.58	13.73	5.63				
5	6.55	20.15	42.77	74.39	63.58	31.54	13.54				
10	11.52	33.27	70.67	113.33	103.44	51.06	22.78				

The median profile likelihood ratio for exclusion, $\lambda(\mu)$ (using background-only MC experiments and taking the median of the $\lambda(\mu)$ distribution), is given in Table 4 for several Higgs masses and integrated luminosities for the signal strength $\mu = 1.0$, corresponding to a SM Higgs Boson.

The distribution of the statistic q_1 for $\hat{\mu} \le 1$ under the s+b hypothesis is shown in Fig. 9 for $m_H = 150$ GeV and for an integrated luminosity of 2 and 10 fb⁻¹. A χ_1^2 distribution is superimposed, showing the validity of the asymptotic approximation.

L				m_H (GeV)			
(fb^{-1})	130	140	150	160	170	180	190
1	1.37	3.93	8.69	14.83	14.23	7.26	3.26
2	2.57	7.47	15.59	25.20	23.65	12.91	5.84
5	5.85	15.26	30.05	45.60	41.13	25.41	12.02
10	10.01	24.69	45.24	62.13	57.69	37.42	20.03

Table 4: Median values of $-2\ln\lambda(\mu = 1)$ obtained from H + 0j fits using simulated background-only ($\mu = 0$) data for several values of the Higgs mass and integrated luminosity.



Figure 9: The distribution of the test statistic q_1 for $\hat{\mu} \le 1$ under the s + b hypothesis for $H + 0j \rightarrow WW + 0j$, for $m_H = 150$ GeV with an integrated luminosity of (a) 2 fb^{-1} and (b) 10 fb^{-1} . A χ_1^2 distribution is superimposed.

3.2.2 H + 2i1

The H + 2j analysis uses a two-dimensional fit based on the transverse mass and the output of a Neural 2

Network, which takes as input several kinematic variables related to the jet activity in the event. The fit is 3

performed simultaneously in signal-enriched and background-enriched regions distinguished by lepton 4

angular variables, which are nearly uncorrelated to the jet variables used in the Neural Network. 5

The median profile likelihood ratios for discovery, $-2\ln\lambda(\mu=0)$ (using toy s+b MC experiments 6 and taking the median of the $\lambda(\mu = 0)$ distribution), are given in Table 5 for several Higgs masses and 7

integrated luminosities. 8

Table 5: Median values of $-2\ln\lambda(\mu)$ for $\mu = 0$ obtained from H + 2i fits to simulated data with signal plus background ($\mu = 1$) for several values of the Higgs mass and integrated luminosity.

L				m_H (GeV)		
(fb^{-1})	130	140	150	160	170	180	190
1	0.39	0.89	1.61	2.73	2.89	1.85	1.23
2	0.70	1.75	3.22	5.34	5.67	3.66	2.07
5	2.01	5.29	8.87	14.20	14.58	9.22	5.71
10	3.82	9.14	16.56	26.35	26.05	16.68	9.93

The distribution of the statistic q_0 under the background-only hypothesis is shown in Fig. 10 for 9 $m_H = 150$ GeV and for an integrated luminosity of 2 and 10 fb⁻¹. A $\frac{1}{2}\chi_1^2$ distribution is superimposed, 10 showing the validity of the asymptotic approximation. 11



Figure 10: The distribution of the test statistic q_0 (for $H + 2j \rightarrow WW + 2j$), under the null background only hypothesis, for $m_H = 150$ GeV and for an integrated luminosity of 2 (a) and 10 (b) fb^{-1} . A $\frac{1}{2}\chi_1^2$ distribution is superimposed.

12 The median profile likelihood ratio for exclusion, $-2\ln\lambda(\mu)$ with $\mu = 1$, where the median is computed using background-only MC data, is given in Table 6 for several Higgs masses and integrated 13 luminosities. 14

The distribution of the test statistic q_1 for $\hat{\mu} \leq 1$ under the s + b hypothesis is shown in Figures 11 15

for $m_H = 150$ GeV with an integrated luminosity of 2 and 10 fb⁻¹. A χ_1^2 distribution is superimposed, 16

showing the validity of the asymptotic approximation. 17

Table 6: Median values of $-2\ln\lambda(\mu)$ for $\mu = 1$ obtained from H + 2j fits to simulated background-only ($\mu = 0$) data for several values of the Higgs mass and integrated luminosity.

L				m_H (GeV)		
(fb^{-1})	130	140	150	160	170	180	190
1	0.37	0.75	1.31	2.29	2.44	1.99	0.95
2	0.87	2.49	5.06	8.18	7.74	3.85	1.90
5	1.40	3.13	5.86	9.58	9.90	7.89	4.59
10	2.80	6.56	10.72	16.14	16.62	13.94	8.74



Figure 11: The distribution of the test statistic q_1 for $\hat{\mu} \leq 1$ under the s + b ($\mu = 1$) hypothesis (for $H + 2j \rightarrow WW + 2j$), for $m_H = 150$ GeV with an integrated luminosity of (a) 2 fb^{-1} and (b) 10 fb^{-1} . A χ_1^2 distribution is superimposed.

1 3.3 $H \rightarrow \tau^+ \tau^-$

The sensitivity of the ATLAS detector to a Higgs boson produced via Vector Boson Fusion and decaying
 to tau leptons has been investigated [6].

Two tau decay channels are considered: *ll* and *lh*. Although there are several neutrinos in the event, it is possible to reconstruct the $\tau^+\tau^-$ invariant mass, $m_{\tau\tau}$, by making the collinear approximation, in which the decay products of the τ are assumed to be collinear with the τ direction in the laboratory frame. After other event selection criteria are imposed, the spectrum of $m_{\tau\tau}$ is used to extract the signal. Particular care has been given to the incorporation of uncertainty in both the rate and shape for the signal

9 and backgrounds.

A data-driven background estimation technique has been established for the major backgrounds. Each technique has been developed to address the aspects of the background estimation which are most relevant for the analysis: the shape of the $m_{\tau\tau}$ tail from the irreducible $Z \rightarrow \tau\tau$, the fake tau contribution in the *lh*-channel, and the normalization of the QCD backgrounds.

In addition to the $m_{\tau\tau}$ spectrum from the $Z \rightarrow \tau\tau$ and QCD control samples, a track multiplicity distribution is used to constrain the fraction of QCD events in the *lh*-channel. This likelihood term is denoted $L_{track}(r_{QCD}, r_{tau})$, where r_{tau} (r_{QCD}) denotes the fraction of real taus (fakes from jets) in the sample. The track multiplicity distribution for the QCD jets is modelled from samples of QCD di-jets that produce tau candidates.

¹⁹ The shape of the $m_{\tau\tau}$ distribution for signal and $Z \rightarrow \tau\tau$ events is dictated by the resolution of \not{E}_T and ²⁰ the kinematics of the collinear approximation. The parameterization of $m_{\tau\tau}$ for the signal and $Z \rightarrow \tau\tau$ ²¹ background are based on the kinematics of the collinear approximation and reproduces an asymmetric ²² distribution with non-Gaussian tails. This distribution is dependent on the overall width width, $\sigma_{H/Z}$, and ²³ a mean, $m_{H/Z}$.

²⁴ A $Z \rightarrow \tau \tau$ control sample is used to constrain the mean, m_Z , and the overall width of the distribution, ²⁵ σ_Z , which are the only free parameters in the $Z \rightarrow \tau \tau$ background model. The error bars in the control ²⁶ sample were scaled to 10% to account for the 10% shape uncertainty in the $\mu \rightarrow \tau$ rescaling method.

The shapes for W+jets and $t\bar{t}$ are very similar and are modelled with a single distribution. A conservative 50% error is applied to each bin in the combined QCD (i.e., $t\bar{t}$ and W+jets) control sample to reflect uncertainty in how this shape changes as the remainder of the analysis cuts are applied.

The shape of the QCD background was parametrized with the following equation:

$$L_{QCD}(m_{\tau\tau}|a_1, a_2, a_3) = \mathcal{N}\left(\frac{1}{m_{\tau\tau} + a_1}\right)^{a_2} m_{\tau\tau}^{a_3} .$$
(38)

³⁰ The form is motivated by a competition between the parton distribution functions and the matrix element.

In the *lh*-channel, the normalization of the backgrounds with fake taus can be constrained by using the track multiplicity method described above. We apply a conservative 50% systematic on this fraction.

By fitting the $m_{\tau\tau}$ spectrum to a model that accurately describes the signal and various backgrounds it is possible to directly incorporate uncertainty in the background shape and take advantage of the shape of the signal within the mass window. We utilize the profile likelihood ratio as our test statistic. The likelihood function corresponding to the simultaneous fit is simply a product of the likelihoods from the individual measurements:

$$L(data|\mu, m_H, \mathbf{v}) = L_{track}(\text{track multiplicity}|r_{QCD}) \times L_Z(Z + \text{jets control}|m_Z, \sigma_Z) \\ \times L_{QCD}(\text{QCD control}|a_1, a_2, a_3) \\ \times L_{s+b}(\text{signal candidates}|\mu, m_H, \sigma_H, m_Z, \sigma_Z, r_{QCD}, a_1, a_2, a_3),$$
(39)

where the a_i are the parameters used to parametrize the QCD background and v represents all nuisance

parameters of the model: $\sigma_H, m_Z, \sigma_Z, r_{QCD}, a_1, a_2, a_3$.

1 The data-driven background estimation methods described above have been developed so that un-

² certainty in the background shape and normalization are included directly into the significance calcu-

³ lation. Because the discovery criterion is simply testing the presence or absence of the signal, it is not

sensitive to some of the sources of systematic uncertainty. In contrast, measurement and exclusion of $\sigma(pp \rightarrow qqH) \times BR(H \rightarrow \tau\tau)$ are sensitive to the uncertainty on the signal selection efficiency. Both

⁶ experimental and theoretical sources of uncertainty on the signal efficiency have been evaluated. The jet

⁷ energy scale uncertainty dominates in this channel, and a signal efficiency uncertainty of 18% was used

⁸ when estimating the exclusion sensitivity.

⁹ The median profile likelihood ratios for discovery, $\lambda(\mu = 0)$ (using the Asimov data sets with $\mu_A =$ ¹⁰ 1), are given in Table 7 for a few Higgs masses at some given luminosities. The distribution of the test

statistic q_0 under the null background only hypothesis, for $m_H = 130$ GeV with an integrated luminosity

¹² of 2 and 10 fb⁻¹, is shown in Figure 12. A $\frac{1}{2}\chi_1^2$ distribution is superimposed, showing the validity of the

13 asymptotic approximation.

Table 7: Median values of $-2\ln\lambda(\mu=0)$ obtained from $H \to \tau^+\tau^-$ simulated data generated with signal plus background ($\mu = 1$) for several values of the Higgs mass and integrated luminosity.

L	$m_H (\text{GeV})$							
(fb^{-1})	110	120	130	140				
1	0.59	0.88	0.72	0.46				
2	1.18	1.74	1.40	0.89				
5	2.91	4.43	3.34	2.08				
10	5.81	8.23	6.37	3.91				
30	17.2	23.6	17.6	10.6				

The resulting median profile likelihood ratio for exclusion, $\lambda(\mu)$ (using the Asimov data sets, $\mu_A = 0$), is given in Table 8 for a few Higgs masses at some given luminosities and signal strength $\mu = 1.0$ corresponding to a Standard Model Higgs Boson.

The distribution of the test statistic q_1 for $\hat{\mu} \le 1$ under the s + b hypotheses for $m_H = 130$ GeV with an integrated luminosity of 2 and 10 fb⁻¹ is shown in Figures 13. A χ_1^2 distribution is superimposed, showing the validity of the asymptotic approximation.

Table 8: Median values of $-2\ln\lambda(\mu=1)$ obtained from $H \to \tau^+\tau^-$ background-only ($\mu=0$) simulated data for several values of the Higgs mass and integrated luminosities.

L	$m_H({ m GeV})$							
(\mathbf{fb}^{-1})	110	120	130	140				
1	0.64	0.71	0.52	0.32				
2	1.31	1.41	1.03	0.64				
5	3.30	3.42	2.52	1.56				
10	6.93	7.79	7.18	3.48				
30	15.3	16.6	12.8	8.48				





Figure 12: The distribution of the test statistic q_0 for $H \to \tau^+ \tau^-$ under the null background-only hypothesis, for $m_H = 130 \,\text{GeV}$ with an integrated luminosity of 2 (a) and 10 (b) fb⁻¹. A $\frac{1}{2}\chi_1^2$ distribution is superimposed. Figures (c) and (d) show $1 - F(q_0)$ where $F(q_0)$ is the corresponding cumulative distribution. The small excess of events at high q_0 is statistically compatible with the expected curves, as can be seen by comparison with the dotted histograms that show the 68.3% central confidence intervals for $p = 1 - F(q_0|0)$. The lower dotted line at 2.87×10^{-7} shows the 5σ discovery threshold.



Figure 13: The distribution of the test statistic q_1 for $\hat{\mu} \leq 1$ under the s + b hypothesis for $H \to \tau^+ \tau^-$, for $m_H = 130 \,\text{GeV}$ with an integrated luminosity of (a) 2 fb⁻¹ (b) and 10 fb⁻¹. A χ_1^2 distribution is superimposed.

1 **3.4** $H \rightarrow ZZ^{(*)} \rightarrow 4l$

² Details of the $H \rightarrow ZZ^{(*)} \rightarrow 4l$ channel can be found in Ref. [7]. The main challenge of this channel ³ for what concerns statistical analysis is its relevance over a very wide mass range (from m_H around ⁴ 120 GeV up to 700 GeV), over which the shapes and cross sections of both signal and background show ⁵ considerable variations.

⁶ While in principle it would be possible to divide the mass range in different regions and use sepa-⁷ rate models for each of them, using a unique background model for the whole phase space is a better ⁸ approach, since it allows to estimate discovery and exclusion significance at any mass, without having ⁹ to worry about boundaries between different models. This will be needed of course when the analysis is ¹⁰ performed using real data, where m_H is unknown.

The main background after event filtering in this channel is the irreducible $ZZ \rightarrow 4l$ process. Reducible backgrounds such as $Zb\bar{b} \rightarrow 4l + X$ or $t\bar{t}$ give a negligible contribution to the overall shape, with the only exception of the $m_H = 120 \text{ GeV}$ case, where $Zb\bar{b} \rightarrow 4l + X$ modifies the background shape in the low mass region and it must therefore be taken into account

the low mass region, and it must therefore be taken into account.

¹⁵ The irreducible background has been modelled using a combination of Fermi functions which are suit-

¹⁶ able to describe both the plateau in the low mass region and the broad peak corresponding to the second

¹⁷ Z coming *on shell*. The chosen model is described by the following function:

$$\frac{p0}{(1+e^{\frac{p6-M_{ZZ}}{p7}})(1+e^{\frac{M_{ZZ}-p8}{p9}})} + \frac{p1}{(1+e^{\frac{p2-M_{ZZ}}{p3}})(1+e^{\frac{p4-M_{ZZ}}{p5}})}.$$
(40)

The first plateau, in the region where only one of the two Z bosons is *on shell*, is modelled by the first term, and its suppression, needed for a correct description at higher masses, is controlled by the *p*8 and *p*9 parameters. The second term in the above formula accounts for the shape of the broad peak and the tail at high masses. This function can describe with a negligible bias the ZZ background shape with good accuracy over the full mass range.

As already mentioned, the $Zb\bar{b}$ contribution is relevant only when searching for very light Higgs bosons (in this study, only $m_H = 120 \text{ GeV}$). In this case, an additional term is added to the ZZ continuum, with a functional form similar to the second part of equation 40. For what concerns signal modelling, a simple Gaussian shape has been used for $m_H \leq 300 \text{ GeV}$, while a relativistic Breit-Wigner formula was needed to properly describe the big tails arising at higher values of the Higgs mass.

In the fits to determine the profile likelihood ratio, m_H is fixed to the hypothesized value, while σ_H is allowed to float in a ±20% range around the value obtained from the signal Monte Carlo distributions. All the parameters describing the background shape are floating within sensible ranges. Given the complexity of the model involved, the fit can from time to time get trapped into local minima. While there is no easy way to avoid this problem, the fake measurements obtained in this case are easy to distinguish from the correct ones, and a repetition of the fit from a different starting point is enough to solve the problem.

The resulting median profile likelihood ratios for discovery, $-2\ln\lambda(\mu = 0)$, with the median computed using toy s + b MC data (i.e., with $\mu = 1$), are given in Table 9 for several values of the Higgs mass and integrated luminosity.

The distribution of the test statistic q_0 under the null background-only hypothesis, for $m_H = 200 \text{ GeV}$ with an integrated luminosity of 2 and 10 fb^{-1} , is shown in Fig. 14. A $\frac{1}{2}\chi_1^2$ distribution is superimposed, showing the validity of the asymptotic approximation.

The resulting median profile likelihood ratio for exclusion, $\lambda(\mu)$ (using toy background-only MC experiments and taking the median of the $\lambda(\mu)$ distribution), is given in Table 10 for several Higgs masses and luminosities using a signal strength $\mu = 1.0$ corresponding to a Standard Model Higgs boson. The distribution of the test statistic q_1 for $\hat{\mu} \leq 1$ under the s + b hypothesis for $m_H = 200$ GeV with

	-											
L		$m_H~({ m GeV})$										
(fb^{-1})	120	130	140	150	160	165	180	200	300	400	500	600
1	0.22	1.20	3.98	5.35	1.65	0.49	0.86	6.86	5.20	3.55	0.88	0.31
2	0.44	2.40	7.96	10.7	3.30	0.98	1.71	13.7	10.4	7.68	1.74	0.62
5	1.05	5.97	19.9	26.7	8.26	2.46	4.27	34.3	25.8	17.8	4.34	1.54
10	2.18	12.0	39.8	53.5	16.5	4.9	8.55	68.7	51.6	35.5	8.47	3.11
30	6.56	35.8	120	160	48.9	14.8	25.6	206	162	108	27.9	10.9
60	13.1	71.6	239	321	99.1	29.5	51.3	407	310	213	52.6	18.6

Table 9: Median values of $-2\ln\lambda(\mu = 0)$ obtained from $H \rightarrow ZZ^{(*)} \rightarrow 4l$ simulated data generated with signal plus background ($\mu = 1$) for several values of the Higgs mass and integrated luminosity.

Figure 14: The distribution of the test statistic q_0 (for $H \to 4l$), under the null background only hypothesis, for $m_H = 200$ GeV with an integrated luminosity of 2 (a) and 10 (b) fb⁻¹. A $\frac{1}{2}\chi_1^2$ distribution is superimposed. Figures (c) and (d) show $1 - F(q_0)$ where $F(q_0)$ is the corresponding cumulative distribution. The small excess of events at high q_0 is statistically compatible with the expected curves, as can be seen by comparison with the dotted histograms showing the 68.3% central confidence intervals for $p = 1 - F(q_0|0)$. The lower dotted line at 2.87×10^{-7} shows the 5σ discovery threshold.

L						m_H (GeV)					
(fb^{-1})	120	130	140	150	160	165	180	200	300	400	500	600
1	0.16	0.93	2.25	3.12	1.23	0.39	0.51	4.86	2.86	2.36	0.87	0.28
2	0.33	1.85	4.50	6.22	2.4	0.75	1.90	9.64	5.68	4.69	1.74	0.56
5	0.83	4.60	11.2	15.4	6.09	2.28	4.74	23.5	14.0	11.6	4.32	1.39
10	1.60	9.14	22.0	30.3	12.1	3.86	5.05	45.2	27.3	22.7	8.57	2.77
30	4.78	26.8	63.0	85.3	35.1	11.4	14.7	105	74.1	63.0	24.9	8.22
60	8.90	51.7	117	155	66.9	22.3	28.0	174	129	113	47.6	16.1

Table 10: Median values of $-2\ln\lambda(\mu = 1)$ obtained from $H \rightarrow ZZ^{(*)} \rightarrow 4l$ simulated data generated with background only ($\mu = 0$) for several values of the Higgs mass and integrated luminosity.

an integrated luminosity of 2 and 10 fb^{-1} is shown in Figures 15. A χ_1^2 distribution is superimposed, showing the validity of the asymptotic approximation.

Figure 15: The distribution of the test statistic q_1 for $\hat{\mu} \le 1$ under the s + b hypothesis (for $H \to 4l$), for $m_H = 200$ GeV with an integrated luminosity of (a) 2 fb⁻¹ and (b) 10 fb⁻¹. A χ_1^2 distribution is superimposed.

3.5 Limitations of the approximations used

The distributions shown in Sections 3.1 through 3.4 show varying levels of agreement between the asymptotic chi-square form and the results of Monte Carlo simulations. For the *WW* (0 jet) channel (Fig. 8), the discrepancy in the distribution of q_0 is very large, and this is understood to arise from the special manner in which the systematic uncertainties for this channel were treated. The distribution of q_0 for the *WW* (0 jet) channel therefore does not use the asymptotic formula. This is the only channel for which the approximation was not applied. For other cases such as the distribution of q_1 for the $H \rightarrow \gamma \gamma$ channel shown in Fig. 7, the Monte Carlo

distribution falls off significantly faster than the chi-square curve. This means that the significance with
 which one excludes the tested hypothesis will be less when estimated from the chi-square curve, leading

to conservative limits. As the integrated luminosity increases, one expects to the asymptotic formula to
 become more accurate.

In some of the distributions such as that of q_0 for the $H \to ZZ^{(*)} \to 4l$ channel shown in Fig. 14, the Monte Carlo simulation indicates a slight excess over the chi-square curve in the tail region. The level of

¹⁷ the excess is not statistically significant in the part of the distribution that can be meaningfully assessed

given the amount of Monte Carlo data available (out to q_0 between around 9 to 16, i.e., to the level of a 2 3 to 4σ discovery). At present it is not practical to verify directly that the chi-square formula remains

valid to the 5σ level (i.e., out to $q_0 = 25$). Thus the results on discovery significance presented here rest

4 on the assumption that the asymptotic distribution is a valid approximation to at least the 5σ level.

5 The validation exercises carried here out indicate that the methods used should be valid, or in some

⁶ cases conservative, for an integrated luminosity of at least 2 fb⁻¹. At earlier stages of the data taking, ⁷ one will be interested primarily in exclusion limits at the 95% confidence level. For this the distributions ⁸ of the test statistic q_{μ} at different values of μ can be determined with a manageably small number of ⁹ events. It is therefore anticipated that we will rely on Monte Carlo methods for the initial phase of the ¹⁰ experiment.

4 Results of the combination

12 4.1 Combined discovery sensitivity

The full discovery likelihood ratio for all channels combined, $\lambda_{s+b}(0)$, is calculated using Eq. 33. This uses the median likelihood ratio of each channel, $\lambda_{s+b,i}(0)$, found either by generating toy experiments under the s+b hypothesis and calculating the median of the $\lambda_{s+b,i}$ distribution or by approximating the median likelihood ratio using the Asimov data sets with $\mu_{A,i} = 1$. Both approaches were validated to agree with each other. The discovery significance is calculated using Eq. 36, i.e., $Z \approx \sqrt{-2 \ln \lambda(0)}$, where $\lambda(0)$ is the combined median likelihood ratio.

The resulting significances per channel and the combined one are shown in Fig. 16 for an integrated luminosity of 10 fb^{-1} .

Figure 16: The median discovery significance for the various channels and the combination with an integrated luminosity of 10 fb^{-1} for (a) the lower mass range (b) for masses up to 600 GeV.

The median discovery significance as a function of the integrated luminosity and Higgs mass is shown colour coded in Fig. 17. The full line indicates the 5σ contour. Note that the approximations used do not hold for very low luminosities (where the expected number of events is low) and therefore the results below about 2 fb^{-1} should be taken as indications only. In most cases, however, the approximations tend to underestimate the true median significance.

4.2 Combined exclusion sensitivity

The full likelihood ratio of all channels used for exclusion for a signal strength μ , $\lambda_b(\mu)$, is calculated using Eq. 34 with the median likelihood ratios of each channel, $\lambda_{b,i}(\mu)$, calculated, either by generating

Figure 17: Significance contours for different Standard Model Higgs masses and integrated luminosities. The thick curve represents the 5σ discovery contour. The median significance is shown with a colour according to the legend. The hatched area below 2 fb⁻¹ indicates the region where the approximations used in the combination are not accurate, although they are expected to be conservative.

- toy experiments under the *b*-only hypothesis and calculating the median of the $\lambda_{b,i}$ distribution or ap-
- ² proximating the median likelihood ratio using the Asimov data sets with $\mu_{A,i} = 0$. Both approaches were

³ checked to agree with each other. A signal strength $\mu = 1$ corresponds to the Standard Model Higgs ⁴ boson.

Any exclusion of $\mu(m_H)$ smaller than 1 corresponds to an exclusion of a Standard Model Higgs 5 boson with a mass m_H . To probe the median sensitivity for excluding a Standard Model Higgs boson we 6 follow Eq. 35 and calculate the corresponding p-value for $\mu = 1$, p_1 for a given luminosity at a given 7 Higgs mass. A p-value of 0.05 corresponds to a significance (Eq. 36) of 1.64. The resulting p_1 for the 8 various channels as well as for the combination, for a luminosity of 2 fb^{-1} , are shown in Fig. 18. Note 9 that any *p*-value below 0.05 indicates an exclusion. We therefore conclude that with a luminosity of 2 10 fb⁻¹ ATLAS has the median sensitivity to exclude a Standard Model Higgs boson heavier than 115 GeV 11 at the 95% Confidence Level. This can also be seen from Fig. 19, which shows the luminosity required 12 to exclude a Higgs boson with a mass m_H at a given confidence level from the combination of the four 13 channels explored in this note. 14

The sharp increase in the required luminosity for lower m_H seen in Fig. 17 reflects the decrease in sensitivity to the Higgs when using only the set of channels considered here. Further developments will increase the sensitivity in this region. For example, improved analysis methods for the $H \rightarrow \gamma\gamma$ channel are described in Ref. [4], including a separation of the events into those with zero or two accompanying jets. Additional final states such as $t\bar{t}H$ with $H \rightarrow b\bar{b}$ will help somewhat, although the contribution to the sensitivity will be small because of the large uncertainties in the background.

For the WW channel, the present study includes only the $e\nu\mu\nu$ decay mode, but it is planned to include $e\nu e\nu$, $\mu\nu\mu\nu$ and $qql\nu$ as well. The $ZZ^{(*)}$ channel here only includes Z decays to ee and $\mu\mu$, but in future analyses $qq\nu\nu$ will be included. The additional WW and $ZZ^{(*)}$ modes have been found to have sensitivity for a high-mass Higgs. Finally, combination with the results from ATLAS with those of CMS will of course result in an overall increase in sensitivity.

HIGGS – STATISTICAL COMBINATION OF SEVERAL IMPORTANT STANDARD MODEL HIGGS ...

Figure 18: The median *p*-value obtained for excluding a Standard Model Higgs Boson for the various channels as well as the combination for (a) the lower mass range (b) for masses up to 600 GeV.

Figure 19: The expected luminosity required to exclude a Higgs boson with a mass m_H at a confidence level given by the corresponding colour. The hatched area below 2 fb⁻¹ indicates the region where the approximations used in the combination are not accurate, although they are expected to be conservative.

HIGGS – STATISTICAL COMBINATION OF SEVERAL IMPORTANT STANDARD MODEL HIGGS ...

Conclusions 5 1

The procedure for combination of search results based on the profile likelihood ratio has been applied 2 to a study of the search for the Standard Model Higgs boson using four search channels: $H \to \tau^+ \tau^-$, 3 $H \to W^+W^- \to e\nu\mu\nu$, $H \to \gamma\gamma$ and $H \to ZZ^{(*)} \to 4$ leptons. The combination method is very general 4 and can be applied to essentially any search that will be carried out at the LHC. 5

The study here has not exploited all of the search channels that will be investigated and therefore the 6 current estimates of the sensitivity can be regarded as conservative. For example, using further decay 7 modes in the ZZ and WW channels will provide additional sensitivity especially for a Higgs boson in the 8 higher mass range. g

The studies have exploited a series of useful approximations that allow one to determine the median 10 discovery and exclusion sensitivities from a combined fit in a manner that only requires separate input 11 ingredients from the individual channels. The determination of the significance for a given (e.g., real) 12 data set, however, will require a simultaneous fit of all of the channels. 13

It is not practical at present to generate enough Monte Carlo data to verify directly that the tail of 14 the profile likelihood distribution is well described to the level required for discovery at the 5σ level, 15 corresponding to an upper tail area of 2.87×10^{-7} . The estimates of discovery significance presented 16 here therefore rely on the assumption that the large-sample approximation used remains valid out to this 17 level. 18

The validation studies shown in Section 3 indicate that the approximations used should be reasonably 19 accurate or lead to conservative limits for an integrated luminosity of at least 2 fb⁻¹. For the earlier 20 stages of the experiment it is expected that one will need to rely on Monte Carlo methods, which should 21 be feasible for exclusion limits at the 95% confidence level. 22

The profile likelihood ratio treats systematic errors by associating the uncertainties with adjustable 23 (nuisance) parameters. Other methods for treating systematic uncertainties can also be considered. Us-24 ing Bayesian methods, for example, one would associate a prior probability density with the nuisance 25 parameters. We plan to develop and use this and other approaches in parallel with the profile likelihood 26 method for searches at the LHC. 27

The study presented in this paper provides the discovery significance for a Higgs boson of a specific 28 mass. That is, the traditional discovery threshold p-value of 2.87×10^{-7} corresponding to a 5 σ effect 29 for a given hypothesized Higgs mass refers to the false discovery rate for a Higgs of that mass. The false 30 discovery rate for a Higgs of *any* mass is higher, and several approaches are being pursued to quantify this 31 (the so-called 'look-elsewhere effect'). The most mature of these methods involves using a simultaneous 32 fit of the Higgs mass $m_{\rm H}$ and the strength parameter μ (or equivalently the Higgs production rate), as has 33 been discussed in the studies of $H \rightarrow \gamma \gamma$ [4] and $H \rightarrow W^+W^-$ [5]. 34

To summarize, the studies based on the four channels considered in this note confirm the good dis-35 covery and exclusion sensitivities already shown in the ATLAS Technical Design Report (TDR) [10]. 36 Furthermore the results here are based on better knowledge and a more realistic simulation of the detec-37 tor than what is described in the TDR. Because of the approximations used, the present studies are valid 38 only for luminosities above 2 fb^{-1} . With a luminosity of 2 fb^{-1} the expected (median) sensitivity is at the 39 5σ level or greater for discovery of a Higgs boson in the mass range between 143 and 179 GeV, and the 40 expected upper limit at 95% confidence level on the Higgs mass is 115 GeV. 41

Comparison with procedures used at LEP Α 42

In this appendix we compare the procedures described in the present analysis with those used in searches 43 carried out at LEP. More details on these methods be found in [8]. The important differences involve 44

the definition of the test statistic used and the treatment of systematic uncertainties. In addition, the LEP 45

analyses adopted a special procedure to prevent spurious exclusion due to a downward fluctuation of the number of events (the CL_s method, see below).

In the LEP Higgs searches, a hypothesized value of the Higgs mass m_H was tested by constructing the statistic

$$Q = \frac{L_{s+b}}{L_b} = \frac{L(\mu = 1)}{L(\mu = 0)},$$
(41)

⁵ where as before b ($\mu = 0$) represents the background-only hypothesis and s + b ($\mu = 1$) refers to back-⁶ ground plus signal at the rate predicted by the Standard Model. For convenience the equivalent logarith-⁷ mic variable $q = -2 \ln Q$ was used.

The sampling distribution of *Q* was determined by Monte Carlo simulation. The Monte Carlo was also used to incorporate systematic errors by sampling values of the corresponding nuisance parameters from pdfs that reflected their uncertainties. That is, one effectively integrated the product of the likelihood and prior pdfs for the nuisance parameters.

For a give observed value $q_{obs} = -2 \ln Q_{obs}$, the *p*-values for the *s* and *s*+*b* hypotheses were determined as

$$p_{s+b} = \int_{q_{\text{obs}}}^{\infty} f(q|s+b) dq \equiv \operatorname{CL}_{s+b}, \qquad (42)$$

$$p_b = \int_{-\infty}^{q_{\text{obs}}} f(q|b) dq \equiv 1 - \text{CL}_b .$$
(43)

Having determine the *p*-values, the LEP analyses then based exclusion of the s + b hypothesis not on the

¹⁵ *p*-value of s + b but rather on the ratio CL_s, defined as

$$CL_{s} = \frac{CL_{s+b}}{CL_{b}} = \frac{p_{s+b}}{1-p_{b}}.$$
(44)

The signal-plus-background hypothesis was said to be excluded at confidence level $CL = 1 - \alpha = 0.95$ if one finds

$$CL_s < \alpha$$
 . (45)

Since $CL_b \leq 1$, one has $CL_s \geq CL_{s+b}$. Therefore the CL_s method will not exclude as large a region of parameter space as that based on the signal-plus-background *p*-value (CL_{s+b} method). As the CL_{s+b} method was designed to provide an interval that brackets the true value of the parameter with a probability of at least $1 - \alpha$, the CL_s limit must cover the true parameter with a greater probability; it is in this sense conservative. The CL_s method was devised so as to avoid the problem where a downward fluctuation in the number of background events can lead to exclusion of the Higgs mass considered, even for hypothesized mass values where one does not expect to be sensitive to Higgs production [9].

In contrast, in the present analysis we test a hypothesized value of the strength parameter μ using

$$q_{\mu} = -2\ln\frac{L(\mu,\hat{\theta})}{L(\hat{\mu},\hat{\theta})}, \qquad (46)$$

as described in Section 2. With this definition, the sampling distribution of the test statistic $f(q_{\mu}|\mu)$ approaches a well defined form related to the chi-squared distribution for a sufficiently large data sample. The ability to exploit this approximate form is very useful as the relevant *p*-value for a 5 σ discovery is 2.87 × 10⁻⁷, and therefore to determine this from Monte Carlo would require an extremely large number of simulated experiments. At LEP this was not a crucial issue as the statistical treatment focused

² primarily on exclusion limits at 95% CL, not on discovery at the 5σ level.

3 Systematic uncertainties here have been incorporated using the profile likelihood, rather than with

⁴ the integrated likelihoods used at LEP. For the uncertainties most relevant to the analyses at the LHC, the

⁵ broadening of the likelihood function obtained by both procedures is similar. It is the profile likelihood

⁶ ratio and not the ratio of integrated likelihoods, however, that approaches the chi-squared form in the

7 large sample limit in accordance with Wilks' theorem.

⁸ The CL_s method has not been applied in the present analysis but studies of its application to searches ⁹ in ATLAS are ongoing.

A final difference with the LEP procedures concerns the definition of significance Z. Here we have defined its relation to the *p*-value as the number of standard deviations of a Gaussian variable that would give a one-sided tail area of *p*, as described in Section 2.1. A significance of Z = 5 corresponds to $p = 2.87 \times 10^{-7}$. The LEP Higgs group defined this relation using a two-sided fluctuation of a Gaussian variable, i.e., a 5σ significance corresponded to $p = 5.7 \times 10^{-7}$.

15 **References**

- ¹⁶ [1] Isaac Asimov, *Franchise*, in *Isaac Asimov: The Complete Stories*, Vol. 1, Broadway Books, 1990.
- [2] S.S. Wilks, *The large-sample distribution of the likelihood ratio for testing composite hypotheses*,
 Ann. Math. Statist. 9 (1938) 60-2.
- [3] A. Stuart, J.K. Ord, and S. Arnold, *Kendall's Advanced Theory of Statistics*, Vol. 2A: *Classical In- ference and the Linear Model* 6th Ed., Oxford Univ. Press (1999), and earlier editions by Kendall
 and Stuart.
- [4] The ATLAS Collaboration, *Prospects for the Discovery of the Standard Model Higgs Boson Using* the $H \rightarrow \gamma \gamma$ Decay with the ATLAS Detector, this volume.
- [5] The ATLAS Collaboration, *Higgs Boson Searches in Gluon Fusion and Vector Boson Fusion* using the $H \rightarrow WW$ Decay Mode, this volume.
- [6] The ATLAS Collaboration, Search for the Standard Model Higgs Boson viaa Vector Boson Fusion
 production process in the di-tau channels with ATLAS, this volume.
- [7] The ATLAS Collaboration, Search for the Standard Model $H \rightarrow ZZ^* \rightarrow 4l$ with the ATLAS Detector, this volume.
- [8] ALEPH, DELPHI, L3 and OPAL Collaborations, Search for the Standard Model Higgs Boson at
 LEP, CERN-EP/2003-011, Phys. Lett. B565 (2003) 61-75.
- [9] Alex Read, *Modified frequentist analysis of search results (the CL_s method)*, proceedings of the Workshop on Confidence Limits, CERN 2000-005.
- [10] The ATLAS Collaboration, *ATLAS: Detector and physics performance technical design report*,
 Volume 2, CERN-LHCC-99-15, ATLAS-TDR-15, May 1999.