DELPHI Collaboration

DELPHI 97-158 PHYS 737
29 October, 1997

Optimal Statistical Analysis of Search Results based on the Likelihood Ratio and its Application to the Search for the MSM Higgs Boson at $\sqrt{s} = 161$ and 172 GeV

A. L. Read
Physics Department, University of Oslo, Blindern, 0316 Oslo 3, Norway

Abstract

A statistical analysis of search results based on the likelihood ratio is described. The analysis method can be used to set optimal exclusion limits for searches consisting of one or more distinct channels and is able to make optimal use of one or more discriminating variables that distinguish the signal from the background. The method was used to analyze the results of the search for the MSM Higgs Boson in the 1996 DELPHI data, taking into account the reconstructed invariant masses of the Higgs candidates, and a 95% confidence level lower bound of 65.9 GeV was obtained, to be compared to the published result of 66.2 GeV which did not take into account the mass information. Results obtained with the new method are the DELPHI input to a combined LEP limit being computed by the LEP working group for Higgs Boson searches.

1 Introduction

In this report an optimal method for setting exclusion limits for multichannel searches[1] where there is additional information about the relative compatibility of candidates with a known background compared to a hypothetical signal, in the form of the probability distribution of a discriminating variable, will be described. The method described here evolved from W. Murray's suggestion to use a maximum likelihood fit to set exclusion limits on the Higgs mass [1].

After reviewing the general concepts used in hypothesis testing in Sec. 2, the likelihood ratio test-statistic is introduced in Sec. 3. The proof that the likelihood ratio is the optimal test-statistic for the type of hypothesis testing done in new particle searches is reviewed in Appendix A. The Monte Carlo integration of the likelihood ratio probability distribution is explained, and several simple examples are presented in Secs. 3.1-3.4 to illustrate some of the properties of the method.

A likelihood ratio analysis can be performed on any search results. In this sense the search itself, i.e. the selection of candidates, may be considered decoupled from the likelihood ratio analysis. However, the properties of the likelihood ratio motivate a particular style of candidate selection which leads to further optimization of exclusion and discovery confidence levels. This is discussed in Sec. 3.5.

In Sec. 4 the analysis of the results from the search for the Minimal Standard Model (MSM) Higgs Boson in the 1996 DELPHI data using the likelihood ratio method is presented. A brief description of the technique used to combine these results with search results from the other LEP experiments is given in Sec. 5.

A description of the current implementation of the likelihood ratio method in the computer program ALRMC is given in Appendix B. The method used to interpolate the probability distributions of the signal between the discrete points where they are available, e.g. where detector simulation (DELSIM) results were obtained, is described in Appendix C.

2 Hypothesis testing of search results

The goal of a search is to either exclude as strongly as possible the existence of a signal in its absence or to confirm the discovery of a true signal as strongly as possible. This can be formulated in terms of a hypothesis test. The null hypothesis is that the signal is absent and the alternate hypothesis is that it exists. An analysis of search results is simply a formal definition of the procedure which quantifies the degree to which one of the hypotheses is favored or excluded by an experimental observation.

The first step in defining a search results analysis is to identify the observables in the experiment which comprise the search results. The simplest observable is the number of candidates satisfying a certain set of criteria. More advanced observables may be some feature of the candidates such as reconstructed invariant mass, b-quark tagging probabilities, or even composite properties such as the output of a discriminant or artifical neural-network analysis. The next step is to define a *test-statistic* or function of the observables and the population parameters (particle mass, production rate, etc.) of the known background and hypothetical signal which ranks experiments from the least

[1]When the term *multichannel* appears in this report it is implicit that the channels are independent.

to most signal-like (most to least background-like). The last step is to define *rules* for exclusion and discovery i.e. specify ranges of values of the test-statistic in which observations will lead to one conclusion or the other. In practice one wishes to specify the significance of the exclusion or discovery, and not simply give a true or false answer. In other words a *confidence level* for the exclusion will be quoted. A *confidence limit* is the value of a population parameter (such as a particle mass or a production rate) which is excluded at a specified confidence level. A confidence limit is a lower (upper) limit if the exclusion confidence is greater (less) than the specified confidence level for all values of the population parameter below the confidence limit.

Correcting the data for known backgrounds (background subtraction) may result in unphysical rates for small or absent signal rates. A standard technique for dealing with this situation is to normalize the confidence level observed for the signal+*background* hypothesis, CL_{s+b}, to the confidence level observed for the background-only hypothesis, CL_b. This also makes it possible to obtain sensible exclusion limits on the signal even when the observed rate is so low that the background hypothesis is called into question. It is commonly said that this *modified frequentist* procedure (as it will be called here) is performed in order to obtain conservative limits on the signal hypothesis, but a better interpretation is that it gives an approximation to the confidence in the signal hypothesis, CL_s, one might have obtained if the experiment had been performed in the absence of background, or in other words, if it had been possible to make a precise correction to the data for the presence of background.

The test-statistic X is constructed to increase monotonically for increasingly signal-like (decreasingly background-like) experiments so that the confidence in the signal+*background* hypothesis is given by the probability that the test-statistic is less than or equal to the value observed in the experiment, X_{obs}:

$$\mathrm{CL}_{s+b} = P_{s+b}(X \leq X_{obs}), \tag{1}$$

where

$$P_{s+b}(X \leq X_{obs}) = \int_0^{X_{obs}} \frac{dP_{s+b}}{dX} dX \tag{2}$$

and where dP_{s+b}/dX is the probability distribution function (p.d.f.) of the test-statistic for signal+*background* experiments. Similarly, the confidence in the background hypothesis is given by the probability that the test-statistic is less than or equal to the value observed in the experiment, X_{obs}:

$$\mathrm{CL}_b = P_b(X \leq X_{obs}), \tag{3}$$

where

$$P_b(X \leq X_{obs}) = \int_0^{X_{obs}} \frac{dP_b}{dX} dX \tag{4}$$

and where dP_b/dX is the p.d.f. of the test-statistic for background-only experiments.

The modified frequentist re-normalization described above is simply

$$\mathrm{CL}_s \equiv \mathrm{CL}_{s+b}/\mathrm{CL}_b. \tag{5}$$

Although CL_s is not, strictly speaking, a confidence (it is a ratio of confidences), the signal hypothesis will be considered excluded at the confidence level CL when

$$1 - \mathrm{CL}_s \leq \mathrm{CL}. \tag{6}$$

The consequence of CL_s not being a true confidence is that the hypothetical false exclusion rate is generally less than the nominal rate of $1 - CL$. The difference between CL_s and the actual false exclusion rate will increase as the p.d.f.'s of the signal+*background* and background hypotheses become more and more similar.

3 The likelihood ratio test-statistic

The likelihood ratio[2], Q, is the ratio of the probability densities for a given experimental result for two alternate hypotheses. In searches for new particles the appropriate likelihood ratio is $Q = \mathcal{L}(s+b)/\mathcal{L}(b)$, that is the ratio of probability densities for the signal+*background* hypothesis to the signal-free or background hypothesis.

The likelihood ratio is commonly used as a test-statistic in hypothesis testing since it maximizes the probability of rejecting a false hypothesis at a given confidence level [4],[5] and conversely minimizes the probability of making a false discovery at a given discovery confidence level. This is a consequence of the Neyman-Pearson theorem which is presented in Appendix A.

The likelihood ratio for an experiment with independent channels is simply a product of the likelihood ratios of the individual channels, so that the combination of additional histogram bins, independent search channels, experiments or center-of-mass energies is straightforward, unambiguous, and automatically optimal for both discovery and exclusion.

Under many circumstances it is possible to obtain the p.d.f.'s of the likelihood ratio for background and signal+*background* experiments in analytic form so that the computation of the signal confidence level in Eq. 5 is straightforward. The single-channel counting experiment is one such case; the result is the well-known PDG formula [2] which is simply a ratio of sums of Poisson probabilities (it will be derived in the present framework in Sec.3.3). For experiments with large samples, the likelihood ratio p.d.f.'s converge to χ^2-distributions so that tables or standard computer functions of χ^2 probabilities may be consulted to obtain the confidence levels.

In the context of searching for small signals in the presence of small backgrounds, in several distinct channels, and where more information than simply the number of candidates will enter the likelihood ratio, it is in general not possible to obtain analytic expressions for the likelihood ratio p.d.f.'s. The p.d.f's may instead be obtained by the Monte Carlo generation of experiments according to the relevant hypothesis and confidence levels computed as the fraction of Monte Carlo experiments satisfying $Q \leq Q_{obs}$. An advantage of this solution is that the effects of uncertainties on the population parameters (background rates, detection efficiencies, theoretical uncertainties, etc.) on the confidence integrals can be taken directly into account by folding them in with the Monte Carlo experiment generation.

This method of obtaining signal exclusion confidences by Monte Carlo integration of the probability distribution of the likelihood ratio, including the modified frequentist step that corresponds to an approximate background subtraction, will be called the MFLR method in the rest of this report [3].

[2] Although the term *likelihood* is reserved for a posteriori probabilities[5], I prefer to call the ratio of a priori probabilities that this method is based on the likelihood ratio due to it familiarity.

[3] The method is called the DELPHI method in the LEP working group for Higgs Boson searches.

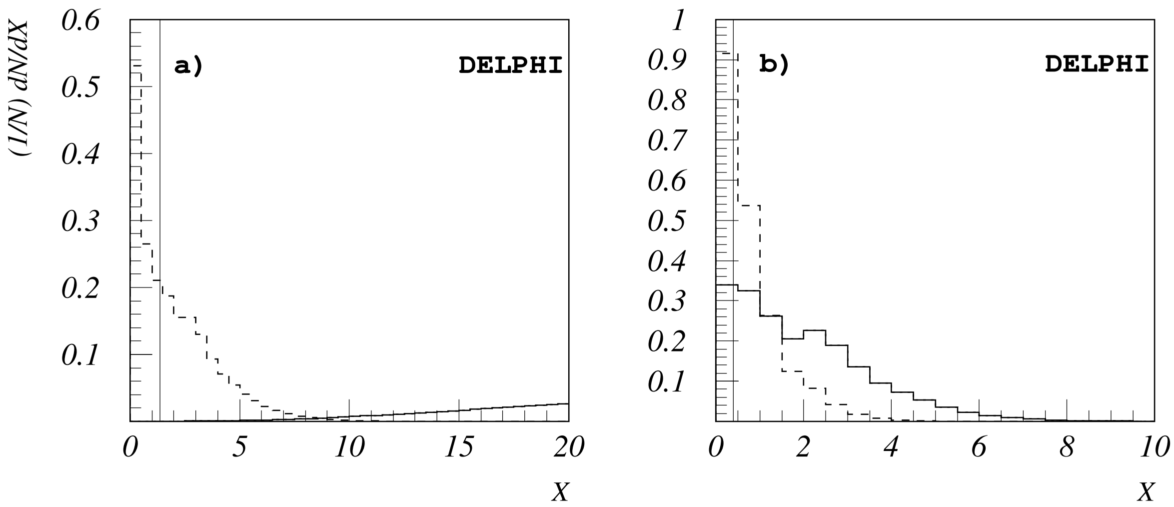

Figure 1: The probability distributions for $X = \ln(Q(m_H)) + s$ for signal+*background* (solid line) and background hypotheses (dashed line) for (a) $m_{\mathrm{H}} = 45$ GeV and (b) $m_{\mathrm{H}} = 77$ GeV, for the DELPHI Higgs boson search. The vertical lines indicate the experimental result $X_{obs}(m_{\mathrm{H}})$. The areas under the curves to the left of the experimental values are the confidence levels CL_{s+b} (solid line) and CL_b (dashed line) in Eqs. (1-4).

Figure 1 illustrates the correspondences between the p.d.f.'s of the likelihood ratio for the signal+*background* and background hypotheses and the confidence levels defined in Eqs. 1–3. From these plots, produced during the analysis described in Sec. 4, it is clear that the DELPHI search results strongly exclude the existence of a 45 GeV Higgs boson but at 77 GeV the signal+*background* and background distributions are quite similar and the exclusion confidence is severely weakened.

3.1 Generic likelihood ratio for searches

The likelihood ratio Q for experiments with N_{chan} independent search channels and measurements of a single discriminating variable x for each candidate, and where the absolute signal and background rates are known, can be written as

$$Q(m_{\mathrm{H}}) = \frac{\prod_{i=1}^{N_{chan}} \frac{e^{-(s_i+b_i)}(s_i+b_i)^{n_i}}{n_i!}}{\prod_{i=1}^{N_{nchan}} \frac{e^{-b_i}b_i^{n_i}}{n_i!}} \frac{\prod_{j=1}^{n_i} \frac{s_iS_i(x_{ij})+b_iB_i(x_{ij})}{s_i+b_i}}{\prod_{j=1}^{n_i} B_i(x_{ij})} \tag{7}$$

which can be simplified to

$$Q = e^{-s_{tot}} \prod_{i=1}^{N_{chan}} \prod_{j=1}^{n_i} \left(1 + \frac{s_iS_i(x_{ij})}{b_iB_i(x_{ij})}\right), \tag{8}$$

where n_i is the number of observed candidates in each channel, x_{ij} is the value of the discriminating variable measured for each of the candidates, s_i and b_i are the integrated signal and background rates per channel, s_{tot} is the total signal rate for all channels, and

$S_i(x)$ and $B_i(x)$ are the probability distribution functions of the discriminating variable for the signal and background respectively.

If the p.d.f.'s of the discriminating variable are identical for the signal and background or if none is measured, the likelihood ratio simplifies further to

$$Q = e^{-s_{tot}} \prod_{i=1}^{N_{chan}} \prod_{j=1}^{n_i} \left(1 + \frac{s_i}{b_i}\right). \tag{9}$$

Note that in the complete absence of background ($b = 0$) and the observation of one or more candidates, an alternate null-hypothesis must be chosen, such as that the signal is the one that maximizes the likelihood function $\mathcal{L}(s)$.

3.2 Weighted event counting

A simple derivation shows that the likelihood ratio method is effectively based on counting weighted events. Since $P(Q \leq Q_{obs}) = P(\ln(Q) \leq \ln(Q_{obs}))$ we can write

$$\ln(Q) = -s_{tot} + \sum_{k=1}^{n} n_k w_k \tag{10}$$

where n is the total number of events observed in all channels and the weight for each candidate k is given by

$$w_k = \ln\left(1 + \frac{s_k}{b_k}\frac{S_k(M_h, m_k)}{B_k(m_k)}\right), \tag{11}$$

where the k index also assigns the candidate to the search channel in which it was observed. Since the constant s_{tot} appears on both sides of the expression $\ln(Q) \leq \ln(Q_{obs})$ the method consists basically of comparing the observed number of weighted events with the distributions expected for the signal+*background* and the background hypotheses.

3.3 Single channel counting experiment

For a counting experiment with a single channel all the candidate events have the same weight, $\ln(1 + s/b)$, so that Eq. 5 takes the form

$$\frac{P(X \leq X_{obs})}{P(X_b \leq X_{obs})} = \frac{P(n \leq n_{obs})}{P(n_b \leq n_{obs})} = \mathrm{CL}_s, \tag{12}$$

where n_b and n come from the Poisson distributions of the number of events for the background and signal+*background* hypotheses respectively, and n_{obs} is the number of candidates observed in the experiment. Thus the modified frequentist signal exclusion confidence becomes

$$\mathrm{CL} = 1 - \frac{\sum_{n=0}^{n_{obs}} \frac{e^{-(b+s)}(b+s)^n}{n!}}{\sum_{n=0}^{n_{obs}} \frac{e^{-b}b^n}{n!}}, \tag{13}$$

the well-known PDG formula. An identical result is obtained by computing the Bayesian confidence interval (with uniform a priori probability for the signal s')

$$\mathrm{CL}_s = \frac{\int_s^\infty \mathcal{L}(s', b)ds'}{\int_0^\infty \mathcal{L}(s', b)ds'}. \tag{14}$$

This coincidence may well be responsible for the comment by the PDG that "This [Bayesian] is the approach favored in the older literature, and has (unfortunately and incorrectly) been referred to as the 'PDG method' in certain papers."

Channel	B_i	ϵ_i	b_i	$B_i\epsilon_i/b_i$	n_{obs}
lepton-jet	4/9	0.616	0.6	0.46	0
lepton-lepton	1/9	0.431	1.5	0.03	2

Table 1: Parameters of a hypothetical two-channel chargino search.

Channel	PDG	Bayes (± 0.1)	MFLR (± 0.1)
lepton-jet	10.93	10.9	10.9
lepton-lepton	105.9	106.0	106.0
total	14.8	11.0	10.3

Table 2: 95% confidence limits for the hypothetical chargino search computed separately for each channel and for the whole experiment with the three methods described in the text. The uncertainties in columns 3 and 4 are due to the trial statistics of the Monte Carlo integrations.

3.4 Two-channel counting experiment

A two-channel counting experiment analyzed with the likelihood ratio and the Bayes confidence interval will illustrate the power of the MFLR method. This study is inspired by an old chargino search [7] but the actual numbers used in the study bear no relation to published results. Table 1 lists the branching ratios, detection efficiencies, expected backgrounds and observed candidates for the lepton-jet and lepton-lepton decay channels for this hypothetical search. The next to last column in the table contains the product of efficiency and branching fraction divided by the expected background, a quantity which indicates the relative sensitivity of the channel.

Table 2 shows the 95% confidence level upper limits on the signal rate computed with three different techniques. Column 2 (PDG) shows naive results based on Eq. 13 where sums over candidates, expected backgrounds and efficiency-branching fraction products are made and the results are considered a single measurement. Column 3 (Bayes) shows the results when the two channels are treated separately and the Bayesian confidence interval with uniform a priori probability for the signal is computed, and column 4 (MFLR) shows the results of the Monte Carlo based likelihood ratio technique. For each technique the confidence limits are computed for each channel separately and for the combined measurement. The confidence limits computed for the single channels are the same for the three techniques, as expected. The observation of a slight excess in the least sensitive channel leads to a large increase in the confidence limit of the two-channel search using the naive (PDG) technique. This is because the information that the candidates were observed in the channel where a relatively larger background was expected has been ignored.

The superiority of the likelihood ratio method advocated here is in fact not demonstrated by the results for a "typical" experimental result but by the comparison of the mean of the exclusion confidence level for a large sample of (simulated) background experiments and the probability to exclude the false hypothesis at the 95% confidence level.

method	$<1-\mathrm{CL_b}>$ (%)	$P(1-\mathrm{CL_b}>95\%)$ (%)
Bayes	94.7	63
MFLR	96.0	75

Table 3: Mean exclusion confidence levels for background experiments and the probability to exclude the false signal hypothesis for the example two-channel chargino search for an expected signal rate of 14 events.

To illustrate this the performance figures for a hypothetical signal rate of 14 events (observable signal of 4.5 events due to branching ratios and efficiencies) are shown in Tab. 3. The difference between the methods is not enormous but the MFLR method is clearly more effective than the Bayesian confidence integral.

3.5 Analysis optimization

Many searches are simplified to single channel counting experiments. The consequence of this is that all candidates have equal weight and the observation of a single candidate, even if it is believed to be highly inconsistent with the signal hypothesis due to the observed value of some property, has a large impact on the exclusion limit set.

It is almost always possible to maximize the probability of excluding the false hypothesis by varying the selection criteria. However the false exclusion probability may vary relatively slowly over reasonable ranges of the selection criteria while the observed candidates are fixed. This often leads to the unpleasant situation that the observed exclusion limits may depend strongly on the fine-tuning of the selection criteria while the expected exclusion limits do not.

The same information which is used to optimize the counting experiment may be used to construct a likelihood ratio-based analysis. The counting experiment optimization consists of varying the selection criteria, computing the integrated background rate and detection efficiency for the signal, and computing the mean confidence expected for background experiments. The optimized selection criteria give the minimum expected confidence. The normalized differential background rates and efficiencies used in the optimization are precisely the probability distributions of the expected background and signal in histogram form which play the role of $S_i(x)$ and $B_i(x)$ in Eq.8 where x is the parameter which controls the selection criteria.

Using the likelihood ratio method instead of a counting method has several, related, advantages: The most important is that the expected confidence for the false hypothesis is minimized. The second is that even if the selection criteria are a bit sloppy, candidates from background processes appearing in signal-free regions enter with weight 0 and do not affect the exclusion limit. Another is that if a selection criterion in a signal-poor region is required to reject a background that is poorly understood, the candidates just inside or outside the criterion contribute approximately the same small weight and the analysis is therefor relatively immune to fine-tuning of the selection criterion.

When the background is poorly understood the selection criteria against the background must be optimized even when using the likelihood ratio method since a large background rate with a large uncertainty can increase the expected confidence if it is not sufficiently suppressed.

Another issue is global optimization for searches with independent channels. The brute force approach is to vary all the (unavoidable) selection criteria in all channels to obtain the globally minimized expected confidence, but in many cases this will be impractical. Optimizing each channel individually may not result in the ideal global optimization but it is reasonable to assume it is a good approximation. Another possibility, guided by the natural optimization of the likelihood ratio and the corresponding event weights in Eq. 11, is to place the selection criteria at a common signal to background ratio in all the individual channels. It is reasonable to expect this to be an even better approximation to the global optimization.

4 The 1996 MSM Higgs search

In the minimal Standard Model the Higgs boson cross-sections are functions of its mass so that the likelihood ratio $Q(m_{\rm h})$ for experiments with N_{chan} distinct search channels and measurements of a single discriminating variable for each candidate (for this search DELPHI has measured the invariant masses of the Higgs candidates) can be written as

$$Q(m_{\rm H}) = \frac{\prod_{i=1}^{N_{chan}} \frac{e^{-(s_i(m_{\rm H})+b_i)}(s_i(m_{\rm H})+b_i)^{n_i}}{n_i!}}{\prod_{i=1}^{N_{nchan}} \frac{e^{-b_i} b_i^{n_i}}{n_i!}} \frac{\prod_{j=1}^{n_i} \frac{s_i(m_{\rm H})S_i(m_{\rm H},m_{ij})+b_iB_i(m_{ij})}{s_i(m_{\rm H})+b_i}}{\prod_{j=1}^{n_i} B_i(m_{ij})} \tag{15}$$

which can be simplified to

$$Q(m_{\rm H}) = e^{-s_{tot}(m_{\rm H})} \prod_{i=1}^{N_{chan}} \prod_{j=1}^{n_i} \left(1 + \frac{s_i(m_{\rm H})S_i(m_{\rm H}, m_{ij})}{b_iB_i(m_{ij})}\right) \tag{16}$$

where n_i is the number of observed candidates in each channel, m_{ij} is the invariant mass of the candidates, $s_i(m_{\rm H})$ and b_i are the integrated signal and background rates per channel, s_{tot} is the total signal rate for all channels, and $S_i(m_{\rm H}, m_{ij})$ and $B_i(m_{ij})$ are the probability distribution functions (p.d.f.'s) of the invariant mass for the signal and background respectively evaluated at the observed masses.

The p.d.f.'s S_i and B_i can be modified or extended to include other information about the candidates which help distinguish signal from the background (such as b-tagging probabilities, production angles, etc.) which will result in lower expected confidences in the absence of a Higgs signal and more significant discoveries in the presence of an observable signal, at the cost of increased complexity of the analysis.

The present analysis is based on results obtained for the published [3] DELPHI search in the 1996 data. The 12 channels for this search consisted of the 6 decay topologies listed in Tab. 4 at the two center-of-mass energies[4] 161.3 and 172.0 GeV.

The search results required to compute the exclusion confidence level are provided by the DELPHI Higgs search team to the LEP working group on the World Wide Web [8]

[4]The data at 172.0 GeV consist of 9 pb^{-1} at 172.3 and 1 pb^{-1} at 170.0 GeV. These two energy points were combined because the efficiencies and backgrounds at the two points were quite similar and the total signal rate corresponded to putting all the luminosity at 172.0 GeV. This reduced the number of search channels from 18 to 12, simplifying the analysis.

and consisted of the integrated background rates in Tab. 4, the integrated signal rates[5] shown in Fig. 4 and for $m_{\mathrm{H}} = 65\,\mathrm{GeV}/c^2$ in Tab. 4, the invariant mass distributions of the backgrounds shown in Figs. 2 and 3, the invariant mass distributions of the signals, and the candidates listed in Tab. 4 and their invariant masses. Two events were selected from the data: one in the $\mathrm{H}^0\nu\bar{\nu}$ channel at 161 GeV, with a mass of 64.6 GeV, and one in the $(\mathrm{H}^0\!\to\!\mathrm{b\bar{b}})\mathrm{q\bar{q}}$ channel at 172 GeV, with a mass of 58.7 GeV/c^2.

Final State	Background		Signal		Candidates.	
	161	172	161	172	161	172
$\mathrm{hq\bar{q}}$	0.30	0.50	1.10	1.65	0	1
$\mathrm{h}\nu\bar{\nu}$	0.65	0.61	0.57	0.89	1	0
$\mathrm{he^+e^-}$	0.13	0.20	0.09	0.12	0	0
$\mathrm{h}\mu^+\mu^-$	0.04	0.13	0.16	0.22	0	0
$(\mathrm{h} \to \mathrm{q\bar{q}})\tau^+\tau^-$	0.31	0.22	0.05	0.06	0	0
$(\mathrm{h} \to \tau^+\tau^-)\mathrm{q\bar{q}}$	0.32	0.91	0.08	0.12	0	0
Total	1.75	2.57	2.05	3.06	1	1

Table 4: Background and signal rates (expected events) for a Higgs boson mass of $m_{\mathrm{H}} = 65\,\mathrm{GeV/c^2}$ and the number of observed candidates. The left-hand entries are for 161 GeV and the right-hand entries are for 170–172 GeV. The expected numbers of signal events do NOT take into account small contributions from $\mathrm{W^+W^-}$and $\mathrm{Z^0Z^0}$fusion.

The invariant mass distributions of the signals obviously depend on the Higgs mass hypothesis. Histograms of the signal distributions were provided for Higgs masses between 45 and 80 GeV in steps of 5 GeV. In order to compute the exclusion confidence level at intermediate masses an interpolation procedure (described in Appendix C) for probability distribution functions and histograms was developed. This procedure was used also to artificially create reference histograms at the few points where they were missing. Although not developed for extrapolation purposes, the interpolation procedure was also used to create full sets of artificial histograms at high masses (75, 80 GeV) for the 161 GeV channels in order to include these channels in the combined LEP limit computation where the expected limit is well over 70 GeV. The invariant mass distributions of the signal for $m_{\mathrm{h}} = 70$ GeV at $\sqrt{s} = 172$ GeV are shown in Fig. 5.

The expected and observed signal confidence levels computed with the modified frequentist likelihood ratio technique (MFLR) are shown in Fig. 6. The intersections of the curves with the horizontal line at 5% render the 95% confidence level lower limits on the mass of the MSM Higgs boson:

$$m_{\mathrm{H}} > 65.9\,\mathrm{GeV/c^2} \text{ (observed 95\% CL limit)}, \tag{17}$$
$$m_{\mathrm{h}} > 65.4\,\mathrm{GeV/c^2} \text{ (expected 95\% CL limit)}. \tag{18}$$

The observed and expected signal confidence levels versus Higgs mass hypothesis computed as the Bayesian confidence integrals for multichannel event counting (the method

[5] An alternative set of signal inputs would consist of the integrated luminosities, the total cross-sections, the branching ratios and the efficiencies. The computation of the signal rates was performed separately for the sake of simplicity.

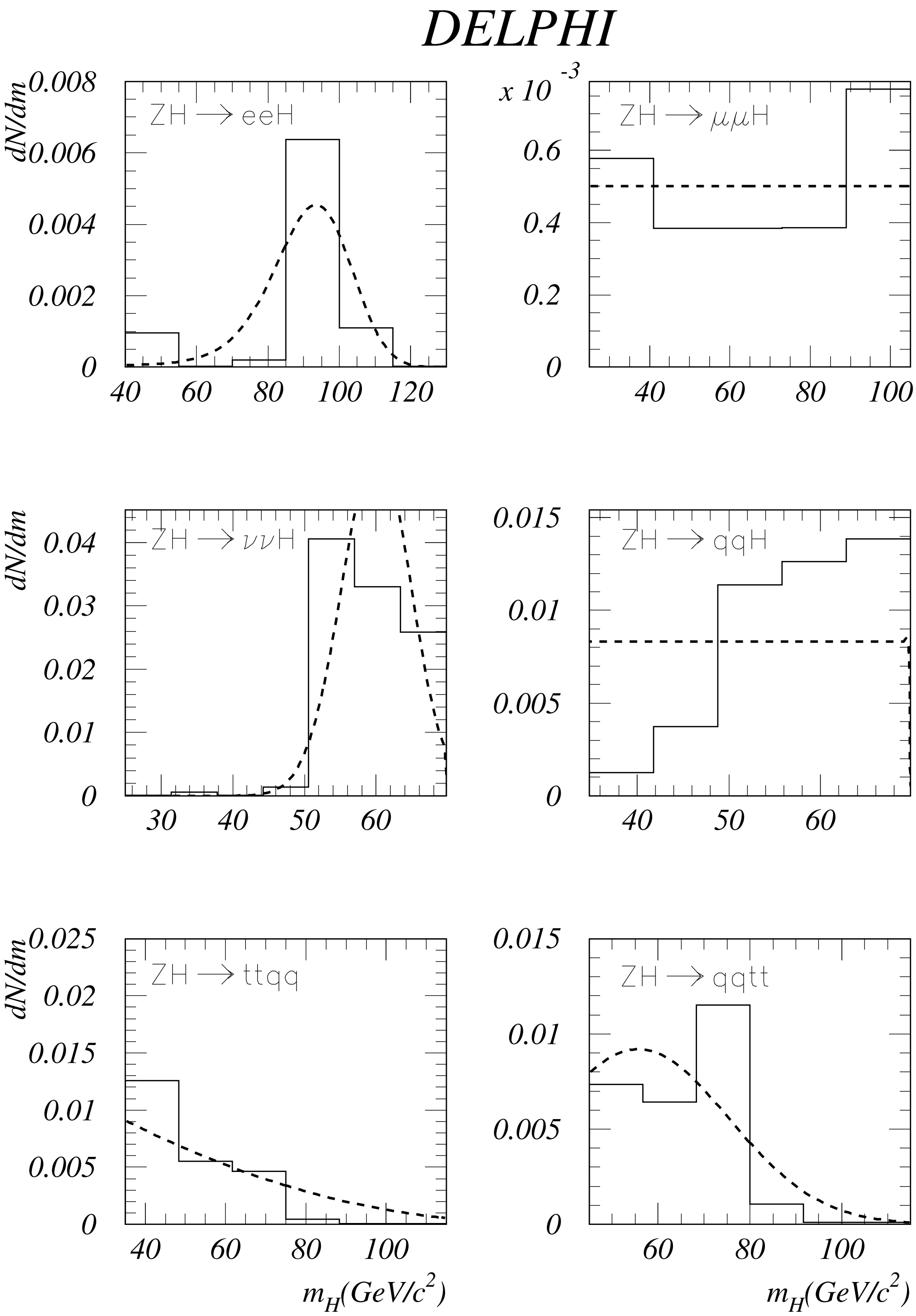

Figure 2: The invariant mass distributions (normalized to integrated luminosity) for the backgrounds at $\sqrt{s} = 161$ (solid histograms) and the corresponding smoothed and fitted functions used in the likelihood ratio (dashed curves).

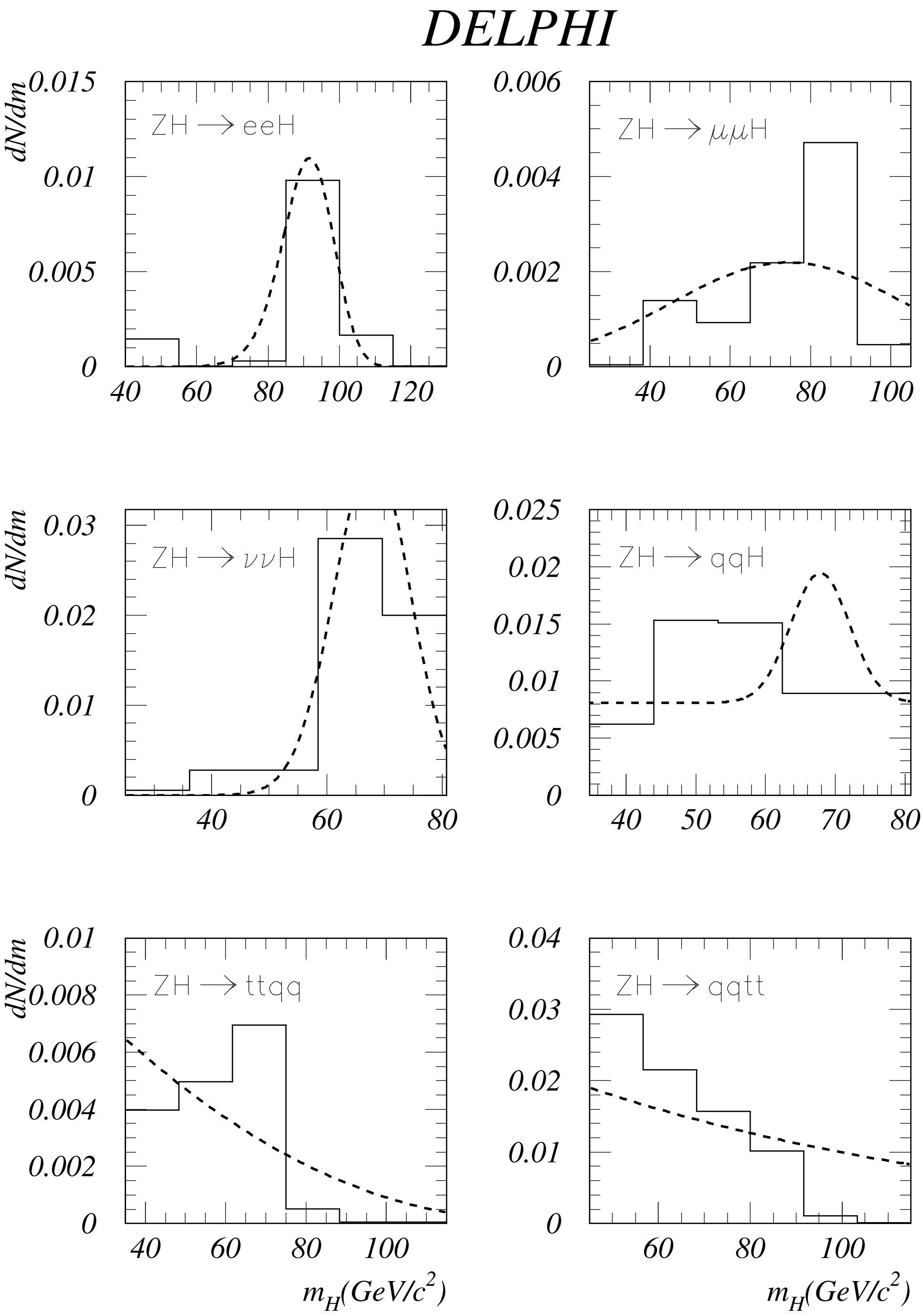

Figure 3: The invariant mass distributions (normalized to integrated luminosity) for the backgrounds at $\sqrt{s} = 172$ (solid histograms) and the corresponding smoothed and fitted functions used in the likelihood ratio (dashed curves).

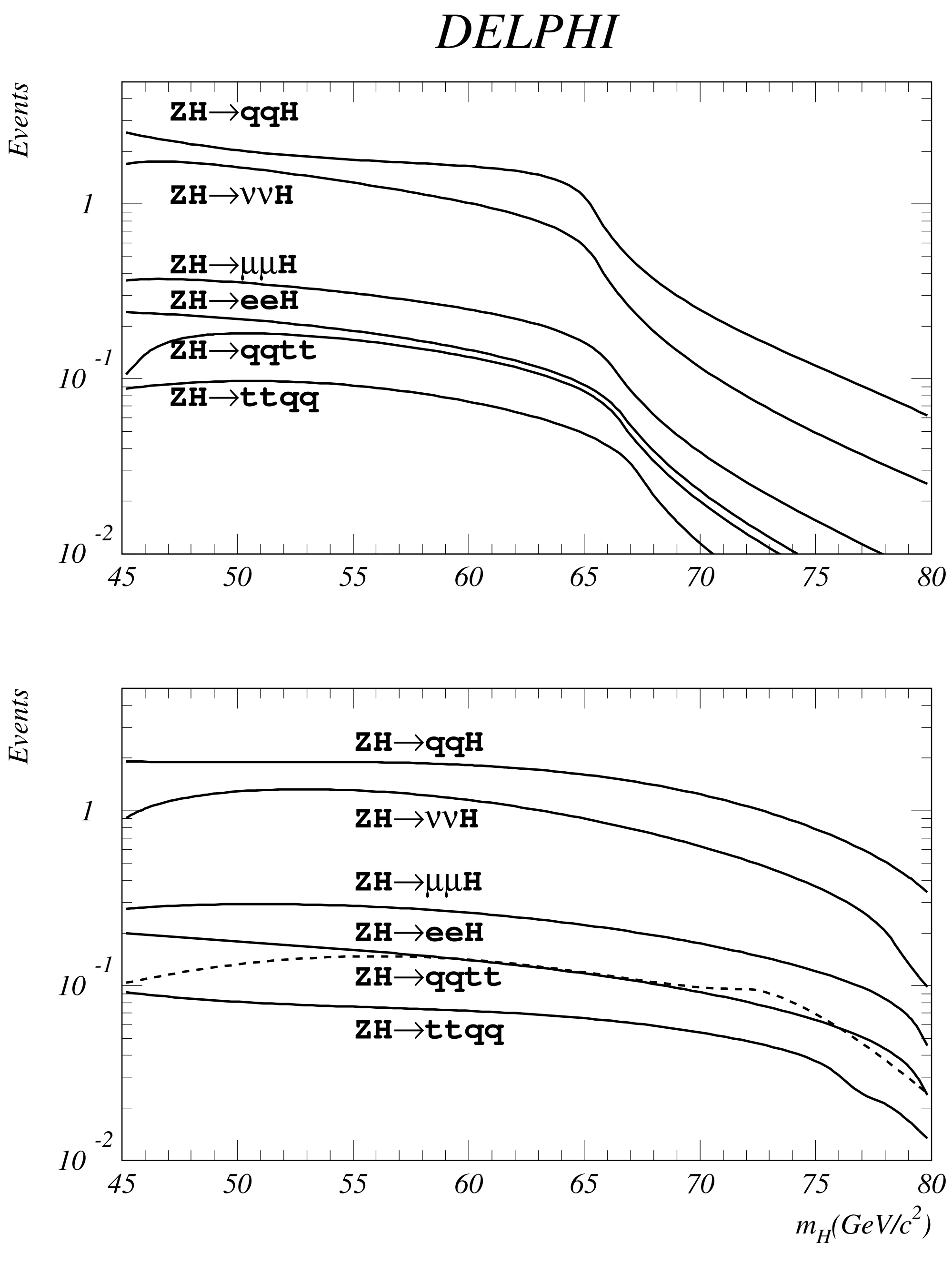

Figure 4: The smoothed and fitted expected Higgs signal rates versus mass hypothesis at $\sqrt{s}=161$ (upper plot) and 172 GeV (lower plot).

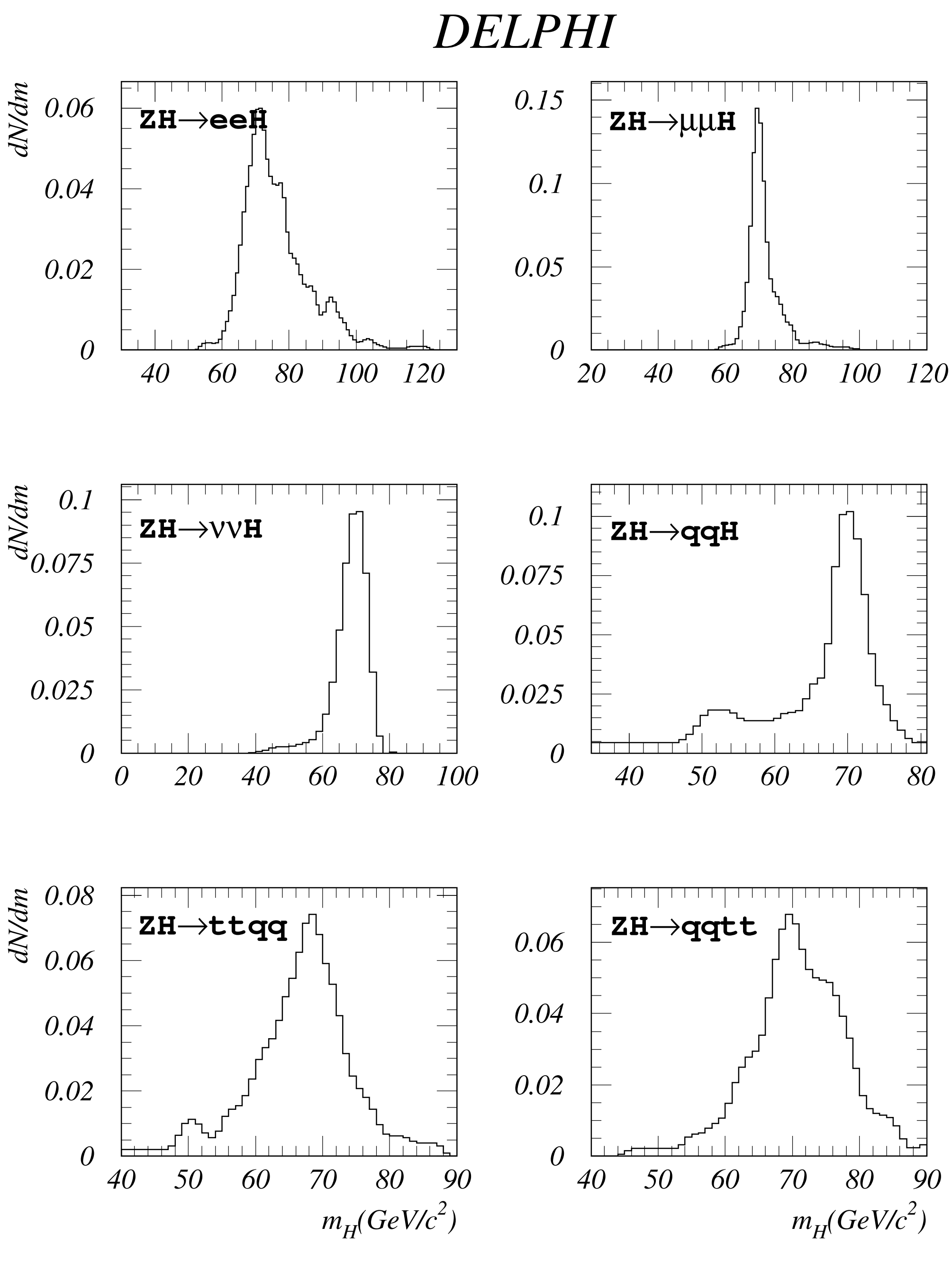

Figure 5: The invariant mass distributions of the Higgs signal for $m_{\mathrm{h}} = 70$ GeV/c^2 at $\sqrt{s} = 172$ GeV.

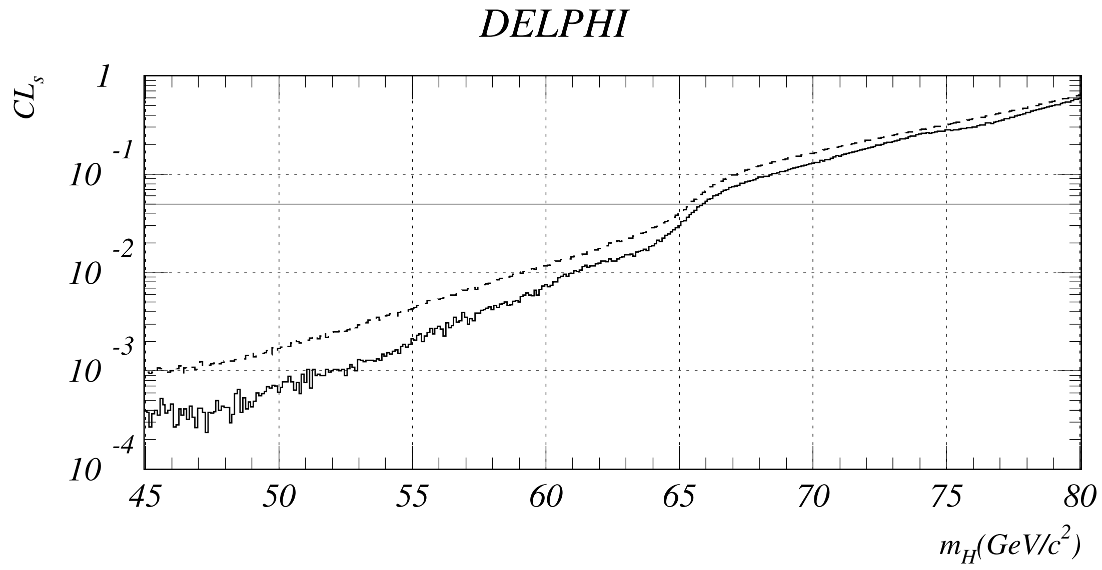

Figure 6: The expected (dashed line) and observed (solid line) confidence levels for the 1996 DELPHI MSM Higgs boson search computed with the MFLR technique.

used for the published DELPHI results [3]) are shown in Tab.5 together with the MFLR results. Except for a small mass region around the candidates, the observed confidences computed with the MFLR method are lower, and far from the candidates substantially lower, than for the published method. The expected confidences are always lower and the probabilities to exclude the false hypothesis at the 95% confidence level are always larger for the MFLR method, as expected.

The expected and observed confidence levels were computed with the background functions fitted to the background histograms shown in Figs. 2 and 3. The mass limits obtained with the direct use of the histograms in the likelihood ratio instead of the functions increased the expected limit by roughly 150 MeV and the observed limit by roughly 300 MeV, both similar to the difference between the observed MFLR mass limit and the published DELPHI result. This gives a rough estimate of the systematic uncertainty on the limit due to uncertainty in the shapes of the background distributions. The less aggressive limits based on the background functions are the ones propagated to the LEP working group for Higgs boson searches.

5 The combined LEP mass limit

The MFLR technique is one of four methods being proposed to combine the search results of the four LEP experiments in order to determine the combined 95% confidence level lower limit on the Higgs boson mass. Preliminary results are presented in [6] and a report with final results, including a detailed comparison of the candidate methods, is in preparation.

As pointed out in Sec. 3 the MFLR technique is perfectly suited to combining results

	Published			MFLR		
$m_{\rm H}$ (GeV/c^2)	$<\mathrm{CL}_s>_b$	CL_s	P_{95}	$<\mathrm{CL}_s>_b$	CL_s	P_{95}
45	1.0	0.2	96	0.1	0.0	100
50	1.5	0.2	95	0.2	0.1	100
55	1.6	0.4	90	0.4	0.2	99
60	3.1	0.7	83	1.2	0.8	96
65	7.1	3.0	53	4.0	3.0	77
70	23.9	17.3	0	16.3	12.9	0
75	42.1	40.7	0	31.8	28.1	0
80	74.0	76.7	0	63.9	56.3	0

Table 5: The observed (CL_s) and expected ($< \mathrm{CL}_s >_b$) signal confidence levels and the probability for excluding the false hypothesis (P_{95}) at the 95% confidence level for the 1996 DELPHI Higgs boson search computed with the published Bayesian method and the MFLR technique (all results in %).

from different experiments, the main disadvantage being that all the input data from the experiments must be available simultaneously. One of the strong candidates to become THE combination method of the working group takes as input the confidences, expected confidences, and expected background and signal rates (all versus Higgs mass hypothesis) computed for each experiment, in other words a substantially reduced amount of information that is easier in practical terms to exchange in a standard format. The general features of this confidence-combining technique are sketched below and compared to the MFLR technique.

5.1 Combining limits

Janot and Le Diberder are publishing an elegant, analytic method (JL) to combine confidence levels measured in two or more different experiments or channels [9] that they claim is optimal. It is interesting to compare the performance of their method to the likelihood ratio technique which is proven to be optimal on general grounds. The comparison is clearest for the case of large expected signals, i.e. when the confidence distribution for signal+*background* experiments is uniform from ~ 0 to 1.

They propose a test-statistic f which is a weighted product of the measured confidences c_i,

$$f = \prod_{i=1}^{n} c_i^{a_i}, \tag{19}$$

and compute confidence levels as in Eq. 1,

$$1 - \mathrm{CL} = \mathrm{P}(\mathrm{f} \leq \mathrm{f_{obs}}). \tag{20}$$

They assume a specific form for the background confidence distributions,

$$\frac{1}{N}\frac{dN}{dc_i} = \beta_i c_i^{-a_i}, \tag{21}$$

Background	LR test-statistic
$\beta_i e^{-a_i c_i}$	$\sum_i a_i c_i$
$\beta_i e^{-a_i c_i^2}$	$\sum_i a_i c_i^2$
$\beta_i c_i^{-a_i}$	$\prod_i c_i^{a_i}$

Table 6: Three possible p.d.f.'s for the signal+*background* confidence of background experiments and the test-statistics which are derived from the likelihood ratio.

	$<$combinedconfidence$>_{\mathrm{b}}$		
Distribution	$\sum_i a_i c_i$	$\sum_i a_i c_i^2$	$\prod_i c_i^{a_i}$
$b_i \exp^{-a_i c_i}$	**0.114**	0.138	0.119
$b_i \exp^{-a_i c_i^2}$	0.146	**0.108**	0.176
$b_i c_i^{a_i}$	0.104	0.148	**0.099**

Table 7: The mean confidence for background experiments when combining two experiments with various individual confidence distributions (first column) and evaluated with various test-statistics. The mean confidence evaluated with the test-statistic that are expected to give optimal results are in bold face. The parameters of all three individual background distributions were tuned to give individual mean background confidences of 0.25.

and prove that the test-statistic they have chosen produces optimal results, in other words that the mean confidence for combined background experiments is minimized.

The likelihood ratio Q is the ratio of the signal+*background* probability density to that of the background. After transforming the p.d.f. of Q for signal+*background* experiments to a uniform confidence distribution from ~ 0 to 1 the background is transformed into another p.d.f. Since the signal+*background* probability densities are now uniform the likelihood ratio is given simply by the inverse product of the background confidence densities and the signal+*background* confidence can now be visualized as the confidence hyper-volume delineated by contours of constant background density. If the background density in each channel is of the form $\beta_i c_i^{-a_i}$ the confidence level is identical to that proposed by Janot and Le Diberder in Eqs. 19 and 20. In other words, they have confirmed that the likelihood ratio test-statistic is optimal for the specific choice of background distribution they chose to study.

The correspondence between a number of possible background confidence distributions and the optimized test-statistic obtained from the likelihood ratio are given in Tab. 6. The test-statistic proposed in [9] for combining confidences is optimal for a specific case, and may be a reasonable approximation in many others, but it is by no means guaranteed to be optimal for all combined measurements. This is illustrated in Tab. 7. Signal confidence distributions for background experiments of the various types in Tab. 6, with their parameters tuned to give the same mean of 0.25, were generated by Monte Carlo methods and the mean confidence for combining two measurements was evaluated with the various test-statistics. For each distribution the test-statistic derived from the likelihood ratio gave the best performance, as expected.

The LEP Working Group for Higgs boson searches is evaluating whether the JL confidence-combination method is a sufficiently good approximation to the best of the direct-combination methods being considered that it can be used by the working group by default in the future to set combined exclusion limits.

6 Conclusions

The MFLR method for determining exclusion limits in searches for new or rare processes has been presented and some of its properties illustrated. The method has been used to analyze the data taken by DELPHI during 1996 at LEP.

6.1 Search method

The MFLR method is based on the Monte Carlo integration of the probability distributions of the likelihood ratio for the signal+*background* and background hypotheses. The difficulty of subtracting the background in the presence of small or absent signals is dealt with by the standard technique of normalizing the confidence in the signal+*background* hypothesis to that of the background hypothesis.

Some of the properties of the MFLR method introduced here are:

- The procedure simplifies directly to the PDG formula for counting experiments with only one search channel.
- For simple multichannel counting experiments the method simplifies to a calculation of a sum of Poisson probabilities over a finite set of accessible experimental configurations of candidates.
- Channels with different background rates and signal efficiencies are automatically combined in an optimal way, and without the need to assume a specific parameterization of the confidence distribution for background experiments.
- The channel combination part of the method is identical to the *elitist* prescription of the combination method described in [9] when the p.d.f.'s of $\mathrm{CL_{s+b}}$ for background experiments are perfectly described by the parameterization of the background used in the other method.
- For large samples $-2\ln(Q)$ is well approximated by changes in χ^2 so that confidences can be easily extracted from standard tables or computer functions of χ^2 probabilities.
- The Monte Carlo integration technique allows the effects of uncertainties on the various inputs to be taken into account in a simple and direct manner.

The formal definitions of the search method and the exclusion confidence level reveal the simple criteria which can be used to rank the discovery and exclusion potential of alternate analyses of the same search results. These are the average exclusion level for background-only experiments $1- <\mathrm{CL_{s+b}}>_\mathrm{b}$ and the average discovery significance for signal+*background* experiments $1- <\mathrm{CL_b}>_\mathrm{s+b}$. The inclusion of the p.d.f.'s of one or more discriminating variables in the likelihood ratio will not only reduce $<\mathrm{CL_{s+b}}>_\mathrm{b}$ but will reduce the sensitivity of the result to the placement of cuts on the discriminating variables when selecting candidates.

This note has dealt mainly with the issue of (false) signal exclusion, but only slight changes to the method and its implementation are required in order to carry out a maximum likelihood analysis of the Higgs boson mass and cross-section in the presence of a true signal and to compute the confidence in the discovery hypothesis, i.e. the probability that the background can account for the observation. A brute-force extension of the present implementation of the MFLR method for discovery consists of computing the Higgs mass where the likelihood ratio is maximized (or simply maximizing the likelihood, both give the same result) for a large set of background experiments and counting the fraction $1 - \mathrm{CL}$ of these which give a likelihood ratio equal to or larger than the maximum likelihood ratio observed for the experiment. The value of CL is the confidence level of the discovery hypothesis.

6.2 1996 DELPHI Higgs Search

The MFLR method been used to obtain from the 1996 data a 95% confidence level lower limit on the MSM Higgs boson mass of 65.9 GeV, similar to the published limit of 66.2 GeV which was obtained with a method which did not take into account the reconstructed invariant masses of the candidates. The MFLR results for DELPHI (also for the other three LEP experiments and the combined LEP "experiment") have been submitted to the LEP Working Group for Higgs boson searches [6] whose mandate includes the combination of the individual Standard Model Higgs searches of the 4 experiments into a single, combined result.

7 Acknowledgements

I thank O. M. Røhne, G. McPherson, W. Murray, and L. Pape for many helpful discussions, L. Bugge for bringing this interesting problem to my attention, and V. Ruhlmann-Kleider for proofreading the draft. I am grateful to the DELPHI Higgs search team for the opportunity to re-analyze their 1996 results just before they were published and to the LEP working group for Higgs boson searches for allowing me to join them and to submit this new method to them at a rather late stage of their work.

A Neyman-Pearson theorem and the likelihood ratio

In this appendix it will be shown that the likelihood ratio test-statistic maximizes the probability to discover a true signal (the discovery potential) at any specified background confidence level and simultaneously maximizes the probability to exclude the false signal hypothesis (the exclusion potential) at any specified signal confidence level. In addition, it will be shown that in the absence of systematic uncertainties in the background hypothesis, the optimal search region includes 100% of the signal, in other words, that the discovery and exclusion potentials are never improved by excluding measurements in regions where the signal to noise ratio is poor.

The proof of the discovery optimization follows closely the presentation of the Neyman-Pearson theorem in [5], although the notation has been adapted here to the analysis of search results. The proofs of the exclusion optimization and the search region optimization are also straightforward applications of the Neyman-Pearson theorem.

The term likelihood ($\mathcal{L}$) will be used here to represent the probability distributions of the signal and background hypotheses due to its familiarity, even though it is normally reserved for a posteriori probabilities. The probability density for the signal+*background* hypothesis is denoted $\mathcal{L}(\vec{X}, s+b)$ and for the pure background hypothesis $\mathcal{L}(\vec{X}, b)$, where $\vec{X}$ represents the set of measured observables for an arbitrary experimental result.

A.1 Discovery potential optimization

The discovery potential optimization begins by the choice of the discovery region R_d motivated by the likelihood ratio:

$$\mathcal{L}(\vec{X}, s+b)/\mathcal{L}(\vec{X}, b) \geq Q_0 \quad \text{for} \quad \vec{X} \in R_d \text{ and} \tag{22}$$

$$\mathcal{L}(\vec{X}, s+b)/\mathcal{L}(\vec{X}, b) \leq Q_0 \quad \text{for} \quad \vec{X} \notin R_d, \tag{23}$$

where $\vec{X}_0$ is the set of measured observables for a particular experiment and

$$Q_o = \mathcal{L}(\vec{X}_o, s+b)/\mathcal{L}(\vec{X}_o, b). \tag{24}$$

An arbitrary alternate discovery region R'_d is chosen to give the same probability as R_d to exclude the background hypothesis at a specified confidence level α:

$$\alpha = \int_{R_d} \mathcal{L}(\vec{X}, b) d\vec{X} = \int_{R'_d} \mathcal{L}(\vec{X}, b) d\vec{X}. \tag{25}$$

The relationship between the two discovery regions is illustrated in Fig. 7. Regions A and B represent the sets of observations in regions R_d and R'_d which are included or not included in region R_d and region I is common to both discovery regions. Comparing Eq. 25 to Fig. 7 we see that

$$\int_A \mathcal{L}(\vec{X}, b) d\vec{X} = \int_B \mathcal{L}(\vec{X}, b) d\vec{X}. \tag{26}$$

Since $A \subset R_d$ and B is completely outside R_d, Eqs. 22,23 imply

$$\int_A \mathcal{L}(\vec{X}, s+b) d\vec{X} \quad \geq \quad Q_0 \int_A \mathcal{L}(\vec{X}, b) d\vec{X} \text{ and} \tag{27}$$

$$\int_B \mathcal{L}(\vec{X}, s+b) d\vec{X} \quad < \quad Q_0 \int_B \mathcal{L}(\vec{X}, b) d\vec{X}. \tag{28}$$

The probability of making an observation in the discovery region R_d if the signal hypothesis is true is

$$P_{s+b}(\vec{X} \in R_d) \quad = \quad \int_{R_d} \mathcal{L}(\vec{X}, s+b) d\vec{X} \tag{29}$$

$$= \quad \int_A \mathcal{L}(\vec{X}, s+b) d\vec{X} + \int_I \mathcal{L}(\vec{X}, s+b) d\vec{X} \tag{30}$$

By substituting the results from Eqs. 27, 25 and 28, in this order, we find

$$P_{s+b}(\vec{X} \in R_d) \quad \geq \quad Q_0 \int_A \mathcal{L}(\vec{X}, b) d\vec{X} + \int_I \mathcal{L}(\vec{X}, s+b) d\vec{X} \tag{31}$$

$$\geq \quad Q_0 \int_B \mathcal{L}(\vec{X}, b) d\vec{X} + \int_I \mathcal{L}(\vec{X}, s+b) d\vec{X} \tag{32}$$

$$> \quad \int_B \mathcal{L}(\vec{X}, s+b) d\vec{X} + \int_I \mathcal{L}(\vec{X}, s+b) d\vec{X}. \tag{33}$$

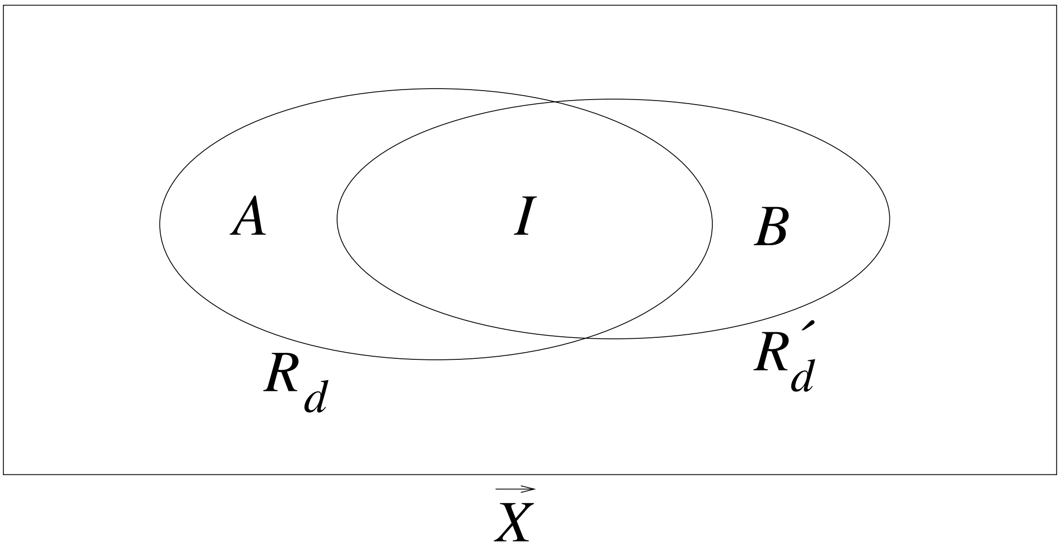

Figure 7: The set of all possible observations $\vec{X}$ (the large box), the discovery region motivated by the likelihood ratio (R_d), the alternate discovery region (R'_d) and the non-overlapping subsets of the discovery regions $A \subset R_d$, $B \subset R'_d$, and $I = R_d \cap R'_d$.

Since $B \cup I = R'_d$ and

$$P_{s+b}(\vec{X} \in R'_d) = \int_{R'_d} \mathcal{L}(\vec{X}, s+b) d\vec{X}, \tag{34}$$

we conclude that $P_{s+b}(\vec{X} \in R_d) > P_{s+b}(\vec{X} \in R'_d)$, in other words that the likelihood ratio test-statistic maximizes the discovery potential for any specified background confidence level compared to any other method of specifying a discovery region.

A.2 Exclusion optimization

The proof that the likelihood ratio maximizes the probability of excluding a false hypothesis follows very closely the arguments in the previous section. We start by defining the exclusion region R_{ex} motivated by the likelihood ratio:

$$\mathcal{L}(\vec{X}, s+b)/\mathcal{L}(\vec{X}, b) \leq Q_0 \quad \text{for} \quad \vec{X} \in R_{ex} \text{ and} \tag{35}$$

$$\mathcal{L}(\vec{X}, s+b)/\mathcal{L}(\vec{X}, b) > Q_0 \quad \text{for} \quad \vec{X} \notin R_{ex}. \tag{36}$$

An arbitrary alternate exclusion region R'_{ex} is chosen to give the same probability β as R_{ex} to exclude the signal hypothesis at a specified confidence level:

$$\beta = \int_{R_{ex}} \mathcal{L}(\vec{X}, s+b) d\vec{X} = \int_{R'_{ex}} \mathcal{L}(\vec{X}, s+b) d\vec{X}. \tag{37}$$

The relationship between the two exclusion regions is illustrated in Fig. 7 (by replacing R_d and R'_d with R_{ex} and R'_{ex}, respectively) as well as the non-overlapping subsets $A \subset R_{ex}$, $B \subset R'_{ex}$, and $I = R_{ex} \cap R'_{ex}$. Comparing Eq. 37 to Fig. 7 we see that

$$\int_A \mathcal{L}(\vec{X}, s+b) d\vec{X} = \int_B \mathcal{L}(\vec{X}, s+b) d\vec{X}. \tag{38}$$

Since $A \subset R_{ex}$ and B is completely outside R_{ex}, Eqs. 35,36 imply

$$\int_A \mathcal{L}(\vec{X}, s+b) d\vec{X} \leq Q_0 \int_A \mathcal{L}(\vec{X}, b) d\vec{X} \text{ and} \tag{39}$$

$$\int_B \mathcal{L}(\vec{X}, s+b) d\vec{X} > Q_0 \int_B \mathcal{L}(\vec{X}, b) d\vec{X}. \tag{40}$$

The probability of making an observation in the exclusion region R_{ex} if the signal hypothesis is false is

$$P_b(\vec{X} \in R_{ex}) = \int_{R_{ex}} \mathcal{L}(\vec{X}, b) d\vec{X} \tag{41}$$

$$= \int_A \mathcal{L}(\vec{X}, b) d\vec{X} + \int_I \mathcal{L}(\vec{X}, b) d\vec{X} \tag{42}$$

By substituting the results from Eqs. 39, 38 and 40, in this order, we find

$$P_b(\vec{X} \in R_{ex}) \geq \frac{1}{Q_0} \int_A \mathcal{L}(\vec{X}, s+b) d\vec{X} + \int_I \mathcal{L}(\vec{X}, b) d\vec{X} \tag{43}$$

$$\geq \frac{1}{Q_0} \int_B \mathcal{L}(\vec{X}, s+b) d\vec{X} + \int_I \mathcal{L}(\vec{X}, b) d\vec{X} \tag{44}$$

$$> \int_B \mathcal{L}(\vec{X}, b) d\vec{X} + \int_I \mathcal{L}(\vec{X}, b) d\vec{X}. \tag{45}$$

Since $B \cup I = R'_{ex}$ and

$$P_b(\vec{X} \in R'_{ex}) = \int_{R'_{ex}} \mathcal{L}(\vec{X}, b) d\vec{X}, \tag{46}$$

we conclude that $P_b(\vec{X} \in R_{ex}) > P_b(\vec{X} \in R'_{ex})$, in other words that the likelihood ratio test-statistic maximizes the exclusion potential for any specified signal confidence level compared to any other method of selecting an exclusion region.

A.3 Search region optimization

The object of this section is to prove that if a likelihood ratio analysis has been restricted to a part of the observation space $\vec{X}$ where the signal-to-noise ratio is "favorable", the effect of adding an element of the observation space where the signal-to-noise ratio is "poor" is always favorable, in other words both the discovery and exclusion potentials are improved. The expected improvement due to accepting more signal and background obviously must be weighed against practical considerations and the effects of uncertainties in the signal and background hypotheses.

The proof is based entirely on the results in the previous two sections. We start with the optimal discovery and exclusion regions R_d and R_{ex} and imagine that we decide to ignore a small, specified part of the observations, e.g. we "cut" the candidates appearing in the last bin of a 1-dimensional histogram where the signal rate is small and the background large. By ignoring some measurements (the candidates removed by a cut effectively enter the likelihood ratio with a factor 1) we have in fact selected alternate critical regions R'_d or R'_{ex}, in the notation of the previous sections, and we have already proved that these critical regions are not optimal. Note that the critical region is a range of values of the test-statistic for the experiment and NOT the range(s) of values of the discriminating variable or variables which enter the test-statistic. Since the discovery and exclusion potentials

$P_{s+b}(\vec{X} \in R_d)$ and $P_b(\vec{X} \in R_{ex})$ are always reduced by selecting alternate critical regions, corresponding to using a test-statistic other than the likelihood ratio, we conclude that increasing the acceptance for the signal (in the discriminating variable or variables that enter the likelihood ratio) always increases the discovery and exclusion potentials.

A.4 Optimization of mean confidence levels

The performance of a hypothesis test may also be measured by computing, in the case of discovery optimization, the mean background confidence level expected for signal+*background* experiments and, in the case of exclusion optimization, the mean signal+*background* confidence level expected for background experiments. In the following derivation R denotes the set of all possible experimental outcomes.

The mean signal+*background* confidence level expected for background experiments is

$$
\begin{aligned}
<CL_{s+b}>_b &= \int_R CL_{s+b}(\vec{X'})\mathcal{L}(\vec{X'};b)d\vec{X'} && (47)\\
&= \int_R P_{s+b}(\vec{X} \in R_{ex}(\vec{X'}))\mathcal{L}(\vec{X'},b)d\vec{X'}, && (48)
\end{aligned}
$$

where $R_{ex}(\vec{X'})$ is the exclusion region motivated by the likelihood ratio for a given observation $\vec{X'}$. We integrate by parts to obtain

$$
\begin{aligned}
<CL_{s+b}>_b &= P_{s+b}(\vec{X} \in R)P_b(\vec{X} \in R) - \int_R \mathcal{L}(\vec{X'};s+b)P(\vec{X} \in R_{ex}(\vec{X'}))d\vec{X'} && (49)\\
&= 1 - \int_R \mathcal{L}(\vec{X'};s+b)P_b(\vec{X} \in R_{ex}(\vec{X'}))d\vec{X'} && (50)
\end{aligned}
$$

In section A.2 it was shown that $P_b(\vec{X} \in R_{ex}) > P_b(\vec{X} \in R'_{ex})$. Working backwards from Eq. 50 to Eq. 47, this implies that $<CL_{s+b}>_b$ is automatically minimized by the likelihood ratio test-statistic.

Using a similar integration by parts and the result $P_{s+b}(\vec{X} \in R_d) > P_{s+b}(\vec{X} \in R'_d)$ from Sec. A.1, it is straightforward to show that the mean background confidence level expected for signal+*background* experiments is automatically maximized by the likelihood ratio test-statistic.

A similar derivation shows that the the mean of the ratio CL_{s+b}/CL_b, the modified frequentist approximation for CL_s (which is unavailable unless the experiment is performed in a background-free environment), expected for background experiments is also minimized by the likelihood ratio test-statistic.

A.5 Summary

The Neyman-Pearson theorem is the proof that for simple hypothesis testing the likelihood ratio test-statistic simultaneously maximizes the probabilities for excluding a false hypothesis and for discovering a true signal for arbitrary discovery and exclusion confidence levels. The same arguments show that, in the absence of systematic uncertainties and practical considerations, the discovery and exclusion potentials are always improved by considering the complete set of experimental measurements, i.e., there is no reason to suppress search channels or histogram bins from a likelihood analysis simply because the signal to background ratio is poor.

It was also shown that alternate measures of the performance of a test-statistic, the mean signal+*background* confidence level expected for background experiments and the mean background confidence level expected for signal+*background* experiments, will automatically and simultaneously be optimized by the likelihood ratio test-statistic.

It is important to note that these optimizations refer to the expected results. Alternate discovery and exclusion regions specified by another test-statistic can result in stronger discovery or exclusion confidence levels for individual experiments.

B ALRMC – A Likelihood Ratio Monte Carlo

In this appendix the implementation of the likelihood ratio search method in the computer program ALRMC will be described.

The program is written in FORTRAN-77 and makes extensive use of CERNLIB. It was developed on a Linux platform where the g77 compiler was used and can behave less reliably on other operating systems due to the wide range of floating point numbers appearing in the Monte Carlo integration of the likelihood ratio distributions. The frequency of floating point exceptions may be drastically reduced, at the cost of increased computing time, by changing the test-statistic from the likelihood ratio to the natural logarithm of the likelihood ratio.

A **running** version (on DXPLUS), not necessarily the latest, is installed under AFS at CERN. A brief description of the installation is given in Tab. 8. The file `000000.readme` contains a summary of updates to the installation.

The program is controlled by FFREAD title cards that are read from logical unit 19 and described in the next section. After this a few simple examples will be presented.

Two versions of the program may be compiled with the `Makefile` which accompanies the installation. The default result of executing the `make` command is the creation of the executable `alrmc` which analyses the 1996 DELPHI Higgs search results, but by executing "`make alrmc_gen`" instead an executable is produced which does not know about the DELPHI Higgs search but rather an idealized or generic Higgs search (or similar search) with Gaussian mass resolutions, uniform background distributions and completely user-specified branching ratios and expected signal rates. The user should examine `generic.f` to understand in more detail the type of search result it represents.

In principle a user with a new search analysis has only to provide the equivalent of the input initialization module of which `generic.f` and `delphi.f` are two examples. Users performing counting experiments should be able to use either of the existing executable versions and a new or modified FFREAD title cards file to obtain results. If an analysis includes a discriminating variable and the total number of histograms times histogram bins per histogram is small, each bin may be considered a separate search channel. If there are many search channels and many bins per channel it is more convenient to follow the example of the '96 Higgs search.

It should be noted that the current version of the program is not a finished product and has several known deficiencies. For instance, an iterative procedure to solve the non-linear equation $\epsilon(s) = 1 - \mathrm{CL}$ for s was developed and heavily optimized to reduce the number of iterations needed to obtain a satisfactorily precise result. However, the optimization is very delicate and does not work under all conditions. Also, some of the outputs produced by the analyses are poorly formatted and some contain large amounts

File	Use	Description
Documents	000000.readme	Log of changes
Compilation	Makefile	Read by the "make" command
Executables	alrmc	DELPHI Higgs search
	alrmc_gen	Generic studies
FORTRAN	alrmc.f	Main program and kernel routines
	delphi.f	1996 DELPHI Higgs search
	generic.f	Generic studies
	lep_util.f	Higgs search and combined LEP utilities
Include files	background.h	COMMON /background/
	confidences.h	COMMON /confidences/
	control.h	COMMON /control/
	minuit.h	Minuit interface
	montecarlo.h	COMMON /montecarlo/
	parameters.h	Constants and compile-time parameters
	qstorage.h	COMMON /qstorage/
	signal.h	COMMON /signal/
	statistics.h	COMMON /statistics/
FFREAD cards	count.tit	Example counting experiments
	sample.tit	Example DELPHI Higgs search
Results	count.res	Output of alrmc with fort.19 -> count.tit
	sample.res	Output of alrmc with fort.19 -> sample.tit
Directory	inputs/	Experimental inputs to DELPHI's '96 Higgs search including FORTRAN code for backgrounds and signal rates.

Table 8: Contents of the directory /afs/cern.ch/delphi/tasks/lep200/searches-/limits/alrmc where the ALRMC program is installed.

of extraneous information. A more serious deficiency is that the program was designed with a model for the signal rate and p.d.f.'s depending on a single parameter. Searches in multi-parameter models will require the code to be more flexible than it is today. I hope to fix the known problems and deficiencies in the near future. I advise early users to start from the working examples. The source code is heavily commented and users are invited to read it.

B.1 The title cards

The title cards consist of parameter cards and action cards (subroutine calls). The action cards are preceded by an asterisk and trigger a specific piece of analysis. The parameter cards are used to configure the search results prior to the next action. The following action cards have been defined:

***COUNT** Analyze a multichannel counting experiment.

***MINUIT** Starts interactive MINUIT for the configured likelihood function.

***PDGLIMIT** Compute a naive upper limit based on the PDG formula by letting the observation be the sum of the observations and the background the sum of the backgrounds - this is of course correct for a single channel.

***TEST** Performed tests of various standard components of the analysis during development.

The following action cards are specific to the analysis of the Higgs search:

***HIGGS** $<$ T/F $>$ Computes confidence levels. Specifying "T" as the argument triggers the results format used by the LEP working group. The default format includes (not completely reliably implemented) results on the upper limits of the cross-section. Since the input is signal rates instead of cross-section, branching ratios and integrated luminosities, the upper limit produced in the output table is the fraction of the expected rate and equivalently the fraction of the Standard Model cross-section.

***BHIGGS** An experimental analysis based on the Bayesian confidence interval.

***FALSE** Computes the false exclusion rate.

***CLINFTY** Computes the mean confidence level of background experiments.

***QSCAN** Crude listing of the likelihood ratio versus Higgs mass hypothesis.

***PERFORMANCE** Computes the expected confidences and probability for excluding the false signal hypothesis for Bayesian and MFLR methods.

The following parameter cards have been defined:

CONFIDENCE <cl> The confidence level used to compute exclusion limits.

argument type REAL, default 0.95

IDEBUG <idebug> The level of debugging output.

- **-1** Debugging output is off.
- **0,1** Increasingly detailed debugging information which may be useful to an ordinary user.
- **2,3** Increasingly detailed debugging information for experts.

MNPRINT <mnprint> Sets the level of MINUIT printout ('set print <mnprint>').

argument type INTEGER, default -1

WARNINGS <T/F> Enables/disables warning messages from MINUIT.

argument type LOGICAL, default .FALSE.

NEXPS Number of experiments to generate in experiment simulations.

argument type INTEGER, default 1000

NMCEXPS Number of Monte Carlo trials in numerical integrations.

argument type INTEGER, default 1000

CHANNELS Number of search channels.

argument type INTEGER, default 0, maximum 50 (set at compile time)

SCAN <m1> <m2> <istep> Determines a Higgs mass (or other model parameter) range and step-size in units of MeV ($\times 10^{-3}$) to analyze.

argument type INTEGER vector, default (10000,70000,5000)

BRANCHING The branching ratio per channel.

argument type REAL vector, default 1.0

EFFICIENCY The nominal signal detection efficiency per channel.

argument type REAL vector, default 1.0

BACKGROUND The nominal integrated background rate per channel.

argument type REAL vector, default 0.0

LUMINOSITY The integrated luminosity per channel.

argument type REAL vector, default 1.0

OBSERVED The number of candidates observed per channel.

argument type INTEGER vector, default 0

EETYPE The efficiency error type per channel.

argument type INTEGER vector, default 0

0 No error.

1 Normal distribution.

2 Poisson distribution.

3 Binomial distribution.

EEP1 Parameter for generating efficiency errors per channel. The interpretation depends on the error type.

argument type REAL, default 0

Normal distribution Standard deviation.

Binomial distribution Number of events in parent sample.

Poisson distribution Number of events selected from parent sample.

BETYPE The background error type.

argument type INTEGER vector, default 0

0 No error.

1 Normal distribution.

2 Poisson distribution.

3 Binomial distribution.

BEP1 Parameter for generating background errors per channel. The interpretation depends on the error type.

argument type REAL vector, default 0

Normal distribution Standard deviation.

Binomial distribution Number of events selected from parent sample.

Poisson distribution Number of events selected from parent sample.

BEP2 Second parameter for generating background errors per channel. The interpretation depends on the error type.

argument type REAL vector, default 0

Binomial distribution Number of events in parent sample.

MAXBMASS The end of the mass range considered for generic studies.

argument type REAL, default 0.

MINBMASS The beginning of the mass range considered for generic studies.

argument type REAL, default 90.

EXPERIMENT Selects which type of Gaussian mass resolution for generic studies. Model 1 is taken from studies performed by W. Murray and is a mass-hypothesis dependent resolution of $11.303 - 0.10347 \times m_{\mathrm{H}}$ GeV (obviously not relevant above 109 GeV). Model 2 is a constant resolution of 0.5 GeV and model 3 is a constant resolution of 4 GeV. Users wishing to define other resolutions or distributions should provide their own version of `FUNCTION sigmah` which has the channel number as the sole argument and obtains the current mass hypothesis from `COMMON /statistics/` in `statistics.h`.

argument type INTEGER, default 1

USECHAN $<$T/F$>$ per channel to enable/disable it in the analysis. In this way it is possible to study individual channels or subsets of channels once a multichannel search has been configured.

argument type LOGICAL vector, default `.TRUE.`

USEMASS $<$T/F$>$ per channel to enable/disable the use of the mass information for enabled channels in the analysis. In this way it is possible to compute combined limits for channels with and without mass measurements.

argument type LOGICAL vector, default `.FALSE.`

MINSIG The lower end of the mass spectrum for the signal per channel. Used for the Higgs search.

argument type REAL vector, default 0.0

MAXSIG The upper end of the mass spectrum for the signal per channel. Used for the Higgs search.

argument type REAL vector, default 100.0

In addition a series of cards called **MAS1**, **MAS2**, ..., **MA10**, **MA11**, ... are declared (according to the number of channels specified by `nchan_max` in `parameters.h`) to accept the observed invariant masses (or other discriminating variable) for each of the candidates in each channel.

B.2 Example: Chargino search in Section 3.4

Figure 8 shows the FFREAD title cards used to compute some of the results in Tab. 2. The most relevant part of the output is shown in Fig. 9.

The "upper limit" is computed by generation of Monte Carlo experiments and iterative solution of $1 - \mathrm{CL} = P(Q(s) \leq Q_{obs}(s))$ for s and is quoted together with the statistical error due to the finite size of the Monte Carlo sample. The next line shows the result of a more direct computation where instead of Monte Carlo experiments the Poisson probabilities of the hypothetical experimental configurations with $Q \leq Q_{obs}$ are summed [10]. The "PDG" limit is computed by reducing the 2-channel search into a single channel search where the background is the sum of the individual backgrounds and the hypothetical signal rate is the sum of the individual signal rates. The final result is the limit computed with the Bayes confidence interval treating the channels individually.

```
LIST
CHANNELS 2
USECHAN T T
USEMASS F F
BACKGROUND 0.6    1.5
EFFICIENCY 0.616  0.431
BRANCH     0.4444 0.1111
NMCEXPS 250000
OBSERVED   0      2
*COUNT
END
```

Figure 8: FFREAD title cards for the example chargino search.

```
++++++++++++++++++++++++++++++++++++++++++++++++
Number of channels      :      2
Observations            :      0        1
Expected background     :   0.60     1.50
Branching ratios        :   0.44     0.11
Efficiencies            :   0.62     0.43
Confidence level        :  95.0%
Number of MC experiments: 250000
Upper limit             :  10.35 +-  0.00
Upper limit, direct comp:  10.31 , CL=  0.95
PDG (naive) upper limit :  14.86
Bayes combined limit    :  11.03 +-  0.04
+++++++++++++++++++++++++++++++++++++++++++++++
```

Figure 9: Part of the ALRMC output for the example chargino search.

```
LIST
MINBMASS  0.0
MAXBMASS 90.0
NMCEXPS 25000
NEXPS   25000
IDEBUG -1
MNPRINT -1
EXPERIMENT 3
WARN FALSE
CONFIDENCE 0.95
CHANNELS 1
BRANCHING_FRACTION 1.0
EFFICIENCY 1.00
SCAN 30000 50000 2000
OBSERVED 1
USEMASS T
MAS1 40.0
BACKGROUND 1.0
*HIGGS F
END
```

Figure 10: FFREAD title cards for the generic Higgs search.

B.3 Example: Generic Higgs search in single channel

Figure 10 shows the FFREAD title cards used to compute the upper limit on the signal rate for a simple, generic Higgs search. The search is carried out for reconstructed masses between 0 and 90 GeV (`MINBMASS` and `MAXBMASS`) with a mass-dependent Gaussian mass resolution (`EXPERIMENT 3`). There is only one channel which has 100% detection efficiency and branching ratio and there is one candidate with a reconstructed mass of 40 GeV. Results are computed between 30 and 50 GeV with 2 GeV steps (`SCAN 30000 50000 2000`).

The most relevant part of the output is shown in Fig. 11. The output columns are described in Tab. 9. Note the bell-shape of the upper limit on the signal rate around the candidate. The upper limit for the mass hypothesis at the observed mass never exceeds 4.74, the 95% confidence level upper limit for 1 observed event and no background in a single channel counting experiment, and far away from the candidate the upper limit converges to 3.0 (the 95% confidence level upper limit for no candidate observed, independent of the background hypothesis).

C Histogram interpolation

In order to compute exclusion confidences for an arbitrary Higgs boson mass it was necessary to do some sort of interpolation between the mass points (from typically 45 to 80 GeV/c^2 in steps of 5 GeV) where Monte Carlo simulation results were available. The signal rates did not, in general, evolve linearly between the mass points but the theo-

`MH`	Higgs mass hypothesis (GeV)
`s_lim`	Upper limit on signal rate (events)
`s_hat`	Maximum likelihood estimate of signal rate at this mass hypothesis
`s_lim1`	Quick guess at upper limit on signal rate (events)
`xs_lim`	Upper limit on signal cross-section
`s_sm`	Standard Model signal rate
`s_exp`	Observable signal rate of Standard Model
`xs_sm`	Standard Model cross-section
`-2lnL`	-2 times the log-likelihood at this mass hypothesis
`CL_s(MH) %`	Modified frequentist confidence in the signal hypothesis
`CL_b(MH) %`	Confidence in the background hypothesis
`CL_sb(MH) %`	Confidence in the signal+*background* hypothesis

Table 9: Definitions of label headings in Fig. 11.

```
  MH       s_lim       s_hat  s_lim1      xs_lim       s_sm   s_exp    xs_sm  -2lnL CL_S(MH)% CL_b(MH) %  CL_sb(MH) %
 ----- ------------- ----- ------    ------------- ------ -----    ----- ------ -------- ----------  -----------
30.000 3.056+- 0.040  0.000  3.000   3.056+- 0.040  1.000  1.000   1.000  12.33   37.91    79.74       30.23
32.000 3.119+- 0.021  0.176  3.308   3.119+- 0.021  1.000  1.000   1.000  11.41   38.28    84.13       32.20
34.000 3.319+- 0.042  0.657  4.160   3.319+- 0.042  1.000  1.000   1.000  10.27   41.16    86.82       35.74
36.000 3.704+- 0.045  0.816  4.459   3.704+- 0.045  1.000  1.000   1.000   9.27   47.44    90.74       43.05
38.000 4.201+- 0.031  0.874  4.555   4.201+- 0.031  1.000  1.000   1.000   8.62   56.65    94.53       53.55
40.000 4.491+- 0.023  0.889  4.580   4.491+- 0.023  1.000  1.000   1.000   8.40   66.71    97.94       65.33
42.000 4.151+- 0.029  0.874  4.556   4.151+- 0.029  1.000  1.000   1.000   8.62   56.03    94.82       53.13
44.000 3.692+- 0.037  0.816  4.459   3.692+- 0.037  1.000  1.000   1.000   9.27   47.71    90.58       43.22
46.000 3.314+- 0.040  0.658  4.161   3.314+- 0.040  1.000  1.000   1.000  10.27   40.81    87.12       35.55
48.000 3.095+- 0.020  0.177  3.311   3.095+- 0.020  1.000  1.000   1.000  11.41   38.54    83.33       32.12
50.000 3.001+- 0.015  0.000  3.000   3.001+- 0.015  1.000  1.000   1.000  12.33   37.75    79.84       30.14
```

Figure 11: The main results of the example generic Higgs search.

retical cross-sections and branching ratios are easy to compute for arbitrary Higgs boson mass and the mass dependence of the detection efficiency was quite well described by smooth parameterizations so that the signal rate could be computed for arbitrary mass with relatively minor uncertainty.

The invariant mass distributions of the signal also evolve and this must be taken into account as well since the significance of a candidate most likely does **not** evolve linearly as the Higgs mass hypothesis changes by 5 GeV. The question is how to take into account kinematic suppressions, experimental cutoffs, changes in the mass resolution, i.e., all the features of an invariant mass distribution in a complicated situation.

The method proposed here and implemented in the ALRMC computer program is based on the linear interpolation of the inverse cumulative probability distribution function. Suppose we start with the two probability distribution functions (p.d.f.'s) $df_1(x)/dx$ and $df_2(x)/dx$ (defined in the physical region $x \geq 0$) and define the cumulative p.d.f.'s

$$f_1(x_1) = \int_0^{x_1} \frac{df_1(x)}{dx} dx \text{ and} \tag{51}$$

$$f_2(x_2) = \int_0^{x_2} \frac{df_2(x)}{dx} dx. \tag{52}$$

We want to obtain a new p.d.f. $d\bar{f}(x)/dx$ with its corresponding cumulative p.d.f.

$$\bar{f}(\bar{x}) = \int_0^{\bar{x}} \frac{d\bar{f}(x)}{dx} dx. \tag{53}$$

We propose to define a linear interpolation of the form

$$\bar{x} = ax_1 + bx_2, \tag{54}$$

where $a + b \equiv 1$, which expresses the interpolation distance between the extreme values of the relevant parameter for the two existing distributions. For the Higgs search, for instance, this would be the relative position of the Higgs mass hypothesis between two mass points where the invariant mass distributions have been simulated. The linear interpolation between the p.d.f.'s $df_1(x)/dx$ and $df_2(x)/dx$ is then defined by solving

$$y = \bar{f}(\bar{x}) = f_1(x_1) = f_2(x_2) \tag{55}$$

for all possible values of $\bar{x}$, solving for the inverse of the new cumulative p.d.f.,

$$\bar{f}^{-1}(y) = af_1^{-1}(y) + bf_2^{-1}(y), \tag{56}$$

deriving this with respect to y and solving for the interpolated p.d.f.,

$$\frac{d\bar{f}(\bar{x})}{dx} = \frac{\frac{df_1(x_1)}{dx}\frac{df_2(x_2)}{dx}}{a\frac{df_2(x_2)}{dx} + b\frac{df_1(x_1)}{dx}}. \tag{57}$$

The prescription for linearly interpolating p.d.f.'s in Eqs. 54 and 57 takes a particularly elegant form when the initial p.d.f.'s can both be described by the same function with different arguments of a particular form. Suppose that

$$\frac{df_1(x_1)}{dx} = G\left(\frac{x_1 - x_{1o}}{k_1}\right) \text{ and} \tag{58}$$

$$\frac{df_2(x_2)}{dx} = G\left(\frac{x_2 - x_{2o}}{k_2}\right), \tag{59}$$

where x_{io} and k_i regulate the offset and width features of p.d.f. i. At the point in the derivation of the interpolated p.d.f. where the cumulative p.d.f.'s are inverted we obtain

$$y = G\left(\frac{x_1 - x_{1o}}{k_1}\right) = G\left(\frac{x_2 - x_{2o}}{k_2}\right) \equiv G(\lambda). \tag{60}$$

From this and the definition of $\bar{x}$ we find that

$$\begin{aligned} \lambda &= \frac{x - \bar{x}}{\bar{k}} \text{ and} && (61) \\ \frac{d\bar{f}(\bar{x})}{dx} &= G\left(\frac{x - \bar{x}}{\bar{k}}\right), \text{ where} && (62) \\ \bar{x} &= a x_{1o} + b x_{2o} \text{ and} && (63) \\ \bar{k} &= a k_1 + b k_2. && (64) \end{aligned}$$

This means, for instance, that if both initial p.d.f.'s are normal distributions, the interpolated distribution is also a normal distribution with both the standard deviation and offset linearly interpolated between those of the initial distributions. For exponential decay distributions the decay constant of the interpolated distribution is linearly interpolated between the decay constants of the initial distributions.

The same elegant result is obtained whenever both initial distributions can be described by the same power series in $(x_i - x_{io})/k_i$. However, for a pair of nontrivial initial distributions obtained by histogramming the output of a complex physics and detector simulation, it may not be practical to identify a particular function which can describe both initial distributions, but it is always possible to carry out the computation in Eqs. 54 and 57 in order to obtain an interpolated histogram. Figure 12 shows how well the interpolation procedure manages to describe a 60 GeV Higgs decaying in the ZH$\to$ $e^+e^-\mathrm{H}^0$ channel at $\sqrt{s} = 161$ GeV based on simulations at $m_{\mathrm{H}} = 50$ and $70\,\mathrm{GeV/c^2}$, despite the large difference between the 3 mass distributions (these distributions were chosen for the comparison precisely because they have several, rapidly evolving features which might be expected to ruin the interpolation). Since the interpolation is able to predict so well the main features of the 60 GeV simulation results, it is plausible that the smaller interpolation distances used in the Higgs search should lead to reliable results, far superior to simply using the nearest available distribution.

In conclusion, the derivation of a method for linearly interpolating probability distribution functions or histograms has been presented here. The method gives particularly intuitive results for certain well-known p.d.f.'s, such as the normal distribution, where one simply linearly interpolates the offset and standard deviation parameters between those of the initial distributions. The method can also be applied to histogrammed simulation data where the distribution functions are not known and very different from each other, and surprisingly reasonable results may be obtained.

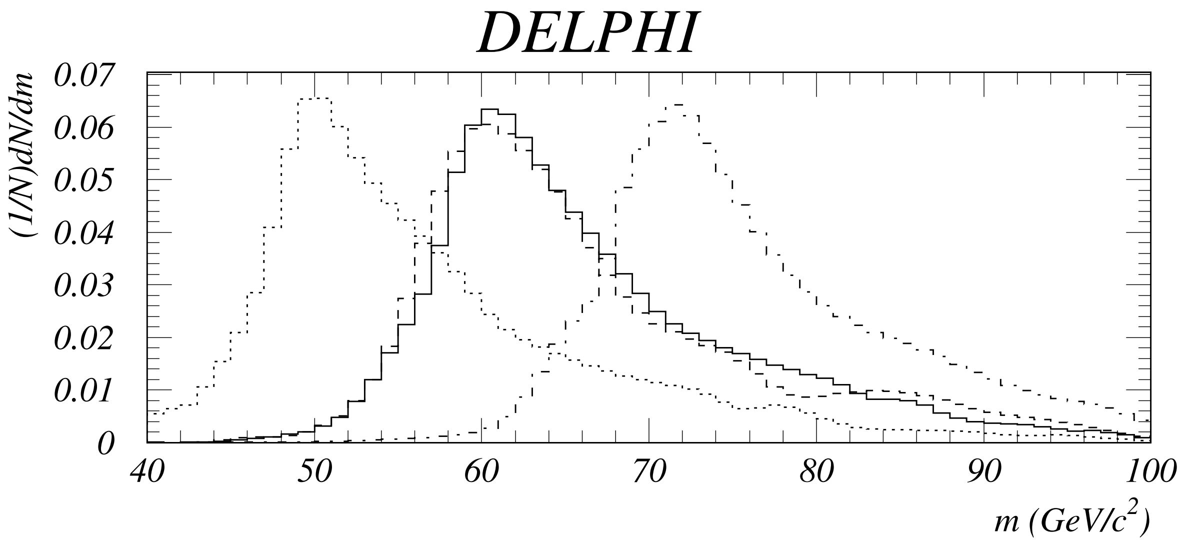

Figure 12: Example of histogram interpolation. The reconstructed invariant mass distribution for a 60 GeV/c^2 Higgs (solid line) has been obtained by interpolation between the simulation results for 50 (dotted line) and 70 (dash-dotted line) GeV/c^2 Higgs masses and overlaid with the simulation results of a 60 GeV/c^2 Higgs (dashed curve).

References

[1] W.Murray, *Use of a Maximum Likelihood fit to set limits on the Higgs mass*, DELPHI Note DELPHI 97-18 PHYS 674, 4 March, 1997.

[2] Particle Data Group (PDG), *Review of Particle Properties*, Physical Review **D**50 (1994) 1272-1282.

[3] DELPHI Collaboration, *Search for neutral and charged Higgs bosons in e^+e^- collisions at $\sqrt{s}=161$ GeV and 172 GeV*, CERN preprint PPE 97–85 (17 July 1997), (to appear in Zeit. Phys. C).

[4] A. G. Frodesen, O. Skjeggestad, and H. Tøfte, *Probability and Statistics in Particle Physics*, Universitetsforlaget, 1979, ISBN 82-00-01906-3.

[5] S. L. Meyer, *Data Analysis for Scientists and Engineers*, John Wiley and Sons, 1975, ISBN 0-471-59995-6.

[6] The LEP working group for Higgs boson searches, *Lower bound for the SM Higgs boson mass: combined results from the four LEP experiments* (short version), August 13, 1997, DELPHI note 97-150 PHYS 729.
The LEP working group for Higgs boson searches, *Lower bound for the SM Higgs boson mass: combined results from the four LEP experiments* (full version), under preparation.

[7] L. Pape, DELPHI Collaboration, *private communication.*

[8] http://www.cern.ch/LEPHIGGS/protected/DELPHI/SM/delphi96final.html

[9] P. Janot and F. Le Diberder, *"Combining 'Limits' "*, CERN PPE 97-053 and LPNHE 97-01 (submitted to Nucl. Instrum. Methods).

[10] FORTRAN code by L. Pape, 23.07.97.