EUROPEAN LABORATORY FOR PARTICLE PHYSICS

CERN-EP/98-046
April 1, 1998

Lower bound for the Standard Model Higgs boson mass from combining the results of the four LEP experiments

ALEPH, DELPHI, L3 and OPAL Collaborations

The LEP working group for Higgs boson searches [1]

Abstract

Various statistical methods are considered to combine the searches for the Standard Model Higgs boson performed by the four LEP collaborations. These methods are applied to the published data collected in 1996 at centre-of-mass energies close to 161 GeV and 172 GeV. A 95% confidence level lower bound of 77.5 GeV/c^2 is obtained for the mass of the Standard Model Higgs boson.

[1]Contributions from P. Bock, J. Carr, S. De Jong, F. Di Lodovico, E. Gross, P. Igo-Kemenes, P. Janot, W. Murray, M. Pieri, A. Read, V. Ruhlmann-Kleider, A. Sopczak.

1 Introduction

The statistical combination of the searches for the Standard Model Higgs boson performed by the four LEP collaborations is a rather complex task due to the numerous search channels, the different centre-of-mass energies and the various analysis procedures adopted. There is no unique, exact, method to deal with this problem. The LEP working group for Higgs boson searches has considered several statistical procedures. The proposed schemes have been submitted to comparative tests based on "gedanken experiments" before being applied to combine the currently available, published, search results based on data accumulated in 1996 at $\sqrt{s}$ close to 161 GeV and 172 GeV [1]-[4]. The data collected before 1995 at the Z resonance and those taken in 1995 at 130 and 136 GeV have not been included since their expected contributions to a possible signal with mass close to the combined mass limit were found to be negligibly small.

2 Statistical procedures

The statistical procedures considered are in most cases extensions of the methods already used by the LEP collaborations to combine their own results in different channels and at different centre-of-mass energies. The procedures start with the construction of a test-statistic X based on the number of selected events and on their distribution in a variable which discriminates signal from background events. The reconstructed Higgs boson mass is a convenient discriminating variable; it is used, for example in the individual publications [1, 2, 4]. However, any other observable with discriminating power (like the weight, W_h, in [3] which combines several discriminating variables into a single one) could also be taken.

The value of the test-statistic measured in the data, X_{obs}, is compared to distributions of the same test-statistic obtained on the basis of a large number of simulated gedanken experiments in which the presence of a signal with mass m_{H} is assumed in addition to the background (*signal+ background*, index "$s+b$" hereafter). Using the test-statistics, the probability $CL_{s+b}(m_{\mathrm{H}})$ to obtain $X_{s+b}(m_{\mathrm{H}}) \leq X_{obs}$ from such gedanken experiments is calculated:

$$CL_{s+b}(m_{\mathrm{H}}) = P(X_{s+b}(m_{\mathrm{H}}) \leq X_{obs}). \tag{1}$$

According to the classical definition, a mass hypothesis m_{H} is excluded at the confidence level $1 - CL_{s+b}(m_{\mathrm{H}})$. However, Eq. 1 is not complete since, in the present context, one is interested in the exclusion of a signal, assuming that the background behaves as expected for its rate and distributions. This shortcoming is corrected by defining, in analogy to Eq. 1, the probability

$$CL_b(m_{\mathrm{H}}) = P(X_b(m_{\mathrm{H}}) \leq X_{obs}), \tag{2}$$

based on gedanken experiments which assume the presence of *background only* (index "b" hereafter), and by using the normalised, approximate, signal probability (index "s"):

$$CL_s(m_{\mathrm{H}}) = \frac{CL_{s+b}(m_{\mathrm{H}})}{CL_b(m_{\mathrm{H}})}. \tag{3}$$

The 95% confidence level lower limit for the Higgs boson mass is thus the lowest value of $m_{\rm H}$ which yields $CL_s(m_{\rm H}) = 0.05$.

The expectation value of the normalised signal probability, $\langle CL_s \rangle$, can be obtained by averaging over a large number of gedanken experiments where the observed number of candidate events is varied according to the *background only* hypothesis. The value of $1 - \langle CL_s \rangle$ is a measure of the sensitivity of a given search, and the search procedure can be optimised by maximising that value. Instead of using simulated gedanken experiments, the above probabilities and $\langle CL_s \rangle$ can also be obtained analytically.

The statistical procedures proposed below differ in their definitions of the test-statistics. Other, practical, differences concern the level and detail of the information that needs to be provided by the four collaborations.

Methods A and B: These two methods make use of the probability density or likelihood function for the *signal + background* hypothesis:

$$\mathcal{L}(x) = \prod_{i=1}^{N} \frac{exp[-(x\frac{s_i(m_{\rm H})}{s(m_{\rm H})} + b_i]\ (x\frac{s_i(m_{\rm H})}{s(m_{\rm H})} + b_i)^{n_i}}{n_i!} \times \prod_{j=1}^{n_i} \frac{x\frac{s_i(m_{\rm H})}{s(m_{\rm H})} S_i(m_{\rm H}, m_{ij}) + b_i B_i(m_{ij})}{x\frac{s_i(m_{\rm H})}{s(m_{\rm H})} + b_i}, \qquad (4)$$

where N is the number of channels [2] to be combined, n_i is the number of observed candidates in channel i and m_{ij} is the value of m, the reconstructed Higgs boson mass (or any other discriminating variable) in the case of candidate j in channel i. The quantities $s_i(m_{\rm H})$ and b_i are the integrated signal and background rates in channel i with $s(m_{\rm H}) = \sum_{i=1}^{N} s_i(m_{\rm H})$ and $b = \sum_{i=1}^{N} b_i$ as the total expected signal and background in all channels. The functions $S_i(m_{\rm H}, m)$ and $B_i(m)$ are the probability distributions for the signal and background, respectively. The argument x is zero for the *background only* hypothesis and $s(m_{\rm H})$ for the *signal + background* hypothesis. The above notation assumes that the background related quantities b_i and $B_i(m)$ do not depend on $m_{\rm H}$. If the selection criteria in any one channel are $m_{\rm H}$ dependent, b_i and $B_i(m)$ have to be replaced by $b_i(m_{\rm H})$ and $B_i(m_{\rm H}, m)$.

The two methods differ in the construction of the test-statistic:

- Method A [5] uses the ratio of the likelihood for *signal + background* ($x = s(m_{\rm H})$) to the likelihood for *background only*($x = 0$):

$$X(m_{\rm H}) = \frac{\mathcal{L}(s(m_{\rm H}))}{\mathcal{L}(0)}. \qquad (5)$$

- Method B [6] uses a Bayesian probability with a flat prior probability distribution:

$$X(m_{\rm H}) = \frac{\int_{s(m_{\rm H})}^{\infty} \mathcal{L}(x)dx}{\int_0^{\infty} \mathcal{L}(x)dx}. \qquad (6)$$

[2] Identical search channels at different centre-of-mass energies or treated by different LEP collaborations are considered as independent channels.

Method C: This method is based on fractional event counting [7]. Using the previous notations, a weight $w_{ij}(m_{\rm H})$ is assigned to each selected candidate j in channel i:

$$w_{ij}(m_{\rm H}) = K(m_{\rm H}) \cdot \frac{1}{C + \frac{b_i B_i(m_{ij}) \cdot s(m_{\rm H})}{S_i^{max}(m_{\rm H}) \cdot s_i(m_{\rm H})}} \cdot \frac{S_i(m_{\rm H}, m_{ij})}{S_i^{max}(m_{\rm H})} \tag{7}$$

where $S_i^{max}(m_{\rm H})$ is the largest value taken by S_i. The factor $K(m_{\rm H})$ is chosen such as to fix the largest value of $w_{ij}(m_{\rm H})$ to unity for a given $m_{\rm H}$, arbitrary i and m_{ij}.

The sum of the weights of all candidates defines the test-statistic:

$$X(m_{\rm H}) = \sum_{i=1}^{N} \sum_{j=1}^{n_i} w_{ij}(m_{\rm H}). \tag{8}$$

The constant C is chosen to maximise $1 - \langle CL_s \rangle$.

Summarising for Methods A, B, and C, the following input is required from each of the collaborations, for each channel i to be combined.

- The integrated signal and background rates, $s_i(m_{\rm H})$ and b_i;
- The probability distributions for the signal and for the background, $S_i(m_{\rm H}, m)$ and $B_i(m)$;
- The numbers n_i of observed candidates, with the reconstructed Higgs boson mass (or the chosen discriminating variable) m_{ij} for each candidate j in channel i.

Method D: This method assumes that each collaboration has already combined its own search channels using their preferred statistical procedures. The collaborations provide the *result* of their searches expressed in the form of confidence levels as functions of $m_{\rm H}$ which are then combined analytically [8] to obtain a combined $CL_s(m_{\rm H})$ according to Eq. 3.

The input required from each collaboration i is the following.

- The measured function $CL^i_{s+b}(m_{\rm H})$;
- The average expected function $\langle CL^i_{s+b}(m_{\rm H}) \rangle$;
- The measured function $CL^i_b(m_{\rm H})$.

Furthermore, the numbers of expected signal and background events, $s_i(m_{\rm H})$ and b_i, have to be provided as well.

The test-statistic is defined as the product

$$X(m_{\rm H}) = \prod_{i=1}^{n} [CL^i_{s+b}(m_{\rm H})]^{a_i}. \tag{9}$$

The "weights" a_i, between zero and one, are tuned to minimise the expected, combined $\langle CL_{s+b}\rangle$; they depend on the intrinsic exclusion power of the experiments.

To make the method totally analytical, it is assumed that the knowledge of the expected $\langle CL_{s+b}\rangle$ and of the total expected background are sufficient to describe entirely the a priori statistical power of each experiment, while the exact shape of the background confidence level distribution in the *signal + background* hypothesis would in principle be needed. Indeed, it has been shown [8] that the weight optimisation procedure is insensitive to the details of these distributions. The observed combined confidence level is by construction independent of these considerations. In addition, the expected combined confidence level obtained this way was shown to be a good approximation of the actual value.

2.1 Comparative tests

Since each of the four methods uses different test-statistics, the resulting exclusion limits are expected to show differences even for identical experimental inputs. The methods are compared below using simulated experiments with a single channel (tests (I) and (II)) and with two channels to be combined (test (III)). Test (I) addresses the question of how the various test-statistics treat candidate events with well defined reconstructed masses; test (II) compares the probabilities for falsely excluding a signal. In test (III), the 95% confidence level exclusion limits obtained by combining two channels with differing sensitivities (signal to background ratios), are compared. Since Method D only deals with several channels to be combined, that method is not considered in the single channel tests (I) and (II).

Test (I): A ficticious experiment has been considered in which three events are observed, with reconstructed Higgs boson masses of 34, 35 and 55 GeV/c^2, while the background, four events in total, is assumed to be flat in mass between zero and 100 GeV/c^2. The mass resolution is assumed to decrease linearly from 10.5 GeV/c^2 at 10 GeV/c^2 mass to 3.3 GeV/c^2 at 80 GeV/c^2 mass. This situation is close to what is experienced at the current LEP energies and integrated luminosities.

Figure 1 shows the 95% confidence level "experimental" upper bound for the number of signal events, $\mathrm{N}_{\mathrm{sig}}^{\mathrm{lim}}$, as a function of m_{H}, obtained by Methods A, B, and C. The three statistical methods show differences in the vicinity of candidate events. For Method C, the "peaks" at the candidate masses are sharper than for Methods A and B. Further simulations show that these local differences, of the order of 15% in the present example, tend to decrease when the background is increased.

In order to quantify the difference not for a particular case but for an average experiment, the above experiment was repeated 600 times using Poisson statistics to simulate the background conditions. In each case the 95% confidence level upper bound for the signal was calculated at m_{H}=40 GeV/c^2 for Methods A, B, and C. The distribution of the limits obtained is shown in Figure 2. (The smallest limit is at $\mathrm{N}_{\mathrm{sig}}^{\mathrm{lim}} = 3$ events, from cases with no candidate with mass close to 40 GeV/c^2.) On average, the test-statistics of Methods A, B and C produce limits which agree within 3%.

Test (II): Methods A, B, and C have also been compared with regards to the probability for falsely excluding an existing signal, at the 95% confidence level, using the same ficticious experiment as in the previous test. Figure 3 shows this probability for $m_{\rm H}$=77 GeV/c^2 as a function of the number of signal events expected. The curves obtained using the three methods have very similar behaviours and converge at the level of 5% for large signal rates. (The deviation from 5% is the consequence of the division by CL_b in Eq. 3.)

Test (III): All four methods have been submitted to a comparative test where *two channels* with differing experimental characteristics were combined. For each method, the average combined signal probability $\langle CL_s \rangle$ was determined in 600 simulated *background only* experiments. The two channels were assumed to have, respectively, 2.9 and 1.6 expected background events, and 1.1 and 2.7 signal events. The mass resolution functions had tails comparable to those typical in LEP experiments for $m_{\rm H}$ close to the current mass limits.

The four methods A, B, C and D gave the following respective values of $\langle CL_s \rangle$: 0.103, 0.109, 0.117 and 0.109, with statistical errors of $\pm$0.004. The differences, not exceeding $\pm$7%, would translate into a shift of $\pm$0.15 GeV/c^2 at most in the expected mass limit at the current values of $\sqrt{s}$ and integrated luminosities. For Method D this test has been repeated using in turn the test statistics of Methods A, B and C. The respective $\langle CL_s \rangle$ values, 0.104, 0.111 and 0.117, were found to be fully consistent with those obtained using Methods A, B and C.

3 Experimental data: summary

The data used in this note have been collected at $\sqrt{s}$ close to 161 GeV and 172 GeV. The integrated luminosities of the samples are: 10.9 and 10.6 pb^{-1} for ALEPH, 10.0 and 10.0 pb^{-1} for DELPHI, 10.8 and 10.2 pb^{-1} for L3, and 10.0 and 10.4 pb^{-1} for OPAL where, in each case, the first number refers to $\sqrt{s} \approx$ 161 GeV and the second to $\sqrt{s} \approx$ 172 GeV.

Each collaboration has performed searches for the process $\rm e^+e^- \to ZH$ in the following final states:

- $(\rm Z \to q\bar{q})(H \to q\bar{q})$ (q = any quark flavour), with branching ratio $Br \approx$ 64%;
- $(\rm Z \to \nu\bar{\nu})(H \to q\bar{q})$, $Br \approx$ 18%;
- $(\rm Z \to e^+e^-, \mu^+\mu^-)(H \to q\bar{q})$, $Br \approx$ 6.2%;
- $(\rm Z \to \tau^+\tau^-)(H \to q\bar{q})$, $Br \approx$ 3.1%;
- $(\rm Z \to q\bar{q})(H \to \tau^+\tau^-)$, $Br \approx$ 5.4%.

The methods employed by the four collaborations to select the possible Higgs boson signal are described in the individual publications [1]-[4]. Since the dominant decay channel of the

Higgs boson is H$\to$b$\bar{\text{b}}$, the presence of b flavour is required explicitly in most of the searches for processes with hadronic decays of the Higgs boson; however, these searches also keep some sensitivity to other quark flavours, or to a Higgs boson decaying into a pair of gluons. The searches for (Z$\to\nu\bar{\nu}$)H and (Z$\to$e$^+$e$^-$)H may also have sensitivity to small contributions to the production cross-section from WW and ZZ fusion diagrams; the ALEPH, L3 and OPAL analyses take these into account in their calculations of the expected signal event rates.

In the cases of ALEPH, DELPHI and OPAL, the selection criteria are independent of m_{H}; in the case of L3 they do depend on m_{H}, and therefore the residual background and the number of selected candidate events are also functions of m_{H}.

Table 1 summarises the main results from the selections of the four collaborations, channel by channel, for m_{H}=70 GeV/c^2, taking into account the small differences in the exact definitions of the search channels and related detection efficiencies. The precise final states to which the numbers apply are listed in the first column. In the columns listing the residual backgrounds, observed candidates, signal efficiencies and expected signal events, the numbers to the left (right) refer to the data collected and selections employed at $\sqrt{s}$ of 161 GeV (172 GeV).

Summed over all search channels of the four collaborations and over the two energy settings, the expected background and the observed number of events are 16.4 and 10, respectively. The expected Higgs boson signal for m_{H}=70 GeV/c^2 is of 12.5 events. The Higgs boson mass is reconstructed with resolutions typically between 2 and 4 GeV/c^2.

The selected candidate events are listed in Table 2. These events enter as Higgs boson candidates into the statistical procedures yielding the combined mass limits. In the case of L3 where the selection depends on m_{H}, the events listed are those selected assuming m_{H}=70 GeV.

Figure 4 shows for the four collaborations the expected numbers of signal events and residual backgrounds, as functions of m_{H}. To produce this figure, the numbers from all search channels and the two energy settings have been added.

Figure 5 shows the observed and expected confidence levels versus m_{H} for the two hypotheses of *signal + background* and *background only.* These are obtained by each of the four collaborations using their preferred statistical procedures, as described in the individual publications. The quantities displayed in Figures 4 and 5 are the input information required by Method D.

4 Combined mass limits

The input information required by Methods A, B, C and D have been provided by each of the four collaborations. The combined results obtained using each of the four methods are summarised in Figure 6. The curves display the normalised signal probabilities $CL_s(m_{\mathrm{H}})$ (see Eq. 3). The full-line curves are the observed CL_s and the dashed-line curves the expected average $\langle CL_s \rangle$. The intersections with the horizontal lines at $CL_s = 0.05$ define the 95% confidence level lower bounds on the Higgs boson mass, observed and expected on average;

these are listed in Table 3.

One is led to the following observations:

- The observed, experimental, limits are in each case higher than those expected. This reflects the fact that for each collaboration the numbers of observed candidates, although compatible with the Standard Model predictions within the errors, are smaller than expected (see Table 1, total background vs. data). This may be a statistical fluctuation (since there are only four observations), but it may also indicate a correlated bias, for example in the background simulations. The discrepancy between the observation and the SM background prediction is not dramatic, as can be seen from Figure 7, showing the observed combined confidence level for the *background only* hypothesis. This confidence level (i.e. the probability for finding less events in the data than expected) is larger than 5% for all masses, and about 40% in the region of the combined mass limit.
- The spread of ± 0.2 GeV/c^2 in the observed limits can be attributed to a large extent to intrinsic differences in the test-statistics, particularly to their ways of treating candidate events close to the mass limits. The expected limits show a similar spread; it is a measure for residual differences in treating the same experimental input (e.g. interpolation between resolution functions provided at fixed values of m_{H}).

As a test of consistency, the three methods A, B and C have been applied separately to each of the inputs from DELPHI, L3 and OPAL. The results were again all within about ± 0.2 GeV/c^2. (This test cannot be applied to Method D which only deals with the *results* from the searches of the four collaborations.) Furthermore, all four methods have been applied to the ALEPH input alone. Since in this case no candidate is observed, all methods reduce to the same method of simple "event counting" which only uses the expected signal and background rates and should yield identical results. Indeed, identical results were obtained. The outcome of these tests is summarised in Table 4.

The treatment of experimental uncertainties affecting the input from the four collaborations has not been incorporated so far in any of the methods described here. Since in each scheme the expected background is subtracted from the observed data, there is an uncertainty arising from the estimation of the background rates, typically less than 25%. The two-channel test described in Section 2 (III) has been used to investigate the influence of such an uncertainty on the expected limit. If the background is overestimated by 25% coherently in both channels, the expected limit changes by ≈ 0.25 GeV/c^2 in each of the four methods. Since the methods for estimating the background differ from one experiment to the other, it is unlikely that there are strong correlations between them, and the quoted change in mass limit is likely to be an overestimate of the uncertainty from this source.

On the basis of the remaining uncertainties, the lowest of the four combined limits, 77.5 GeV/c^2, is taken as the 95% confidence level lower bound for the Standard Model Higgs boson mass.

5 Summary

Four statistical methods were investigated to combine the searches of the ALEPH, DELPHI, L3 and OPAL Collaborations for the Standard Model Higgs boson. The comparison of the methods shows similar performances. Each of the methods is applied to combine the published data from the four LEP collaborations collected in 1996 at energies close to 161 GeV and 172 GeV. A 95% confidence level lower bound of 77.5 GeV/c^2 is obtained for the mass of the Standard Model Higgs boson.

References

[1] ALEPH Collaboration, R. Barate et al., Phys. Lett. **B412** (1997) 155.

[2] DELPHI Collaboration, P. Abreu et al., Eur. Phys. Journ. **C2** (1998) 1.

[3] L3 Collaboration, M. Acciarri et al., Phys. Lett. **B411** (1997) 373.

[4] OPAL Collaboration, K. Ackerstaff et al., Eur. Phys. Journ. **C1** (1998) 425.

[5] A. L. Read, *Optimal statistical analysis of search results based on the likelihood ratio and its application to the search for the MSM Higgs boson at 161 and 172 GeV*, DELPHI note 97-158 PHYS 737.

[6] A. Favara and M. Pieri, *Confidence level estimation and analysis optimisation*, Firenze University preprint DFF-278/4/1997, E-preprint hep-ex 9706016.

[7] P. Bock, *Determination of exclusion limits for particle production using different decay channels with different efficiencies, mass resolutions and backgrounds*, Heidelberg University preprint HD-PY-96/05 (1996), to be submitted to Nucl. Instr. Meth.

[8] P. Janot and F. Le Diberder, *Optimally Combined Confidence Limits*, CERN PPE 97-053 and LPNHE 97-01, to be published in Nucl. Instr. Meth.

$\sqrt{s}$ (GeV) = CHANNEL	Background 161	172	Data 161	172	Signal eff. (%) 161	172	Expected signal 161	172
ALEPH								
$Hq\bar{q}$	0.17	0.23	-	-	21.1	21.9	0.24	1.12
$H\nu\bar{\nu}$	0.06	0.09	-	-	26.3	42.9	0.11	0.70
$H(e^+e^-\ ,\ \mu^+\mu^-)$	0.06	0.11	-	-	64.2	74.8	0.08	0.40
$H\tau^+\tau^-$	0.02	0.02	-	-	18.8	20.4	0.01	0.05
$(H\to\tau^+\tau^-)q\bar{q}$	0.05	0.03	-	-	17.4	17.4	0.02	0.07
Total	0.36	0.48	-	-			0.46	2.39
DELPHI								
$Hq\bar{q}$	0.30	0.50	-	1	21.6	23.6	0.25	1.21
$H\nu\bar{\nu}$	0.65	0.61	1	-	36.3	42.8	0.12	0.63
He^+e^-	0.13	0.20	-	-	41.7	37.2	0.02	0.09
$H\mu^+\mu^-$	0.04	0.13	-	-	69.0	69.8	0.04	0.17
$(H\to q\bar{q})\tau^+\tau^-$	0.31	0.22	-	-	22.9	24.4	0.01	0.06
$(H\to\tau^+\tau^-)q\bar{q}$	0.32	0.91	-	-	22.1	24.4	0.02	0.10
Total	1.74	2.50	1	1			0.46	2.26
L3								
$(H\to q\bar{q})q\bar{q}$	0.77	3.68	2	4	28.1	38.5	0.37	1.87
$(H\to q\bar{q})\nu\bar{\nu}$	0.40	1.46	-	-	46.0	69.4	0.17	0.97
$(H\to q\bar{q})e^+e^-$	0.03	0.18	-	-	45.5	65.8	0.03	0.15
$(H\to q\bar{q})\mu^+\mu^-$	0.04	0.15	-	-	34.4	48.3	0.02	0.11
$(H\to q\bar{q})\tau^+\tau^-$	0.008	0.23	-	-	13.5	34.9	0.01	0.08
$(H\to\tau^+\tau^-)q\bar{q}$	(*)	0.25	-	-	(*)	17.7	(*)	0.07
Total	1.25	5.96	2	4			0.60	3.26
OPAL								
$(H\to bb)q\bar{q}$	0.75	0.88	-	1	30.8	28.2	0.35	1.29
$(H\to q\bar{q})\nu\bar{\nu}$	0.90	0.55	1	-	38.1	41.3	0.16	0.66
He^+e^-	0.06	0.08	-	-	52.6	65.3	0.04	0.17
$H\mu^+\mu^-$	0.04	0.06	-	-	67.8	70.6	0.04	0.18
$(H\to q\bar{q})\tau^+\tau^-$	0.10	0.41	-	-	17.3	21.7	0.01	0.05
$(H\to\tau^+\tau^-)q\bar{q}$	0.06	0.18	-	-	17.0	19.3	0.02	0.08
Total	1.91	2.16	1	1			0.62	2.43

Table 1: *Summary of the event selections: expected residual backgrounds, selected data events, signal detection efficiencies, and the expected numbers of signal events for* m_H=*70* GeV/c^2. *In each column, the entries to the left (right) are for* $\sqrt{s}\approx$*161 GeV (172 GeV). The signal detection efficiencies refer to the final states listed in the first column. The searches for H*$\to q\bar{q}$ *are also sensitive to Higgs decays to gluon pairs. Wherever the Higgs decay is not specified, the symbol H stands for H*$\to$ *all channels. In the part pertaining to the L3 experiment, the symbol (*) designates cases where the particular channel has been completely eliminated by a cut which limited the signal to background ratio to values larger than 0.15.*

Experiment	$\sqrt{s}$ (GeV)	Channel	$m_{\mathrm{H}}{}^{\mathrm{REC}}$ (GeV/c^2)
DELPHI	161	$\mathrm{H}\nu\bar{\nu}$	64.6
	172	$\mathrm{Hq\bar{q}}$	58.7
L3	161	$(\mathrm{H}\to\mathrm{q\bar{q}})\mathrm{q\bar{q}}$	73.9
	161	$(\mathrm{H}\to\mathrm{q\bar{q}})\mathrm{q\bar{q}}$	73.7
	172	$(\mathrm{H}\to\mathrm{q\bar{q}})\mathrm{q\bar{q}}$	73.3
	172	$(\mathrm{H}\to\mathrm{q\bar{q}})\mathrm{q\bar{q}}$	57.0
	172	$(\mathrm{H}\to\mathrm{q\bar{q}})\mathrm{q\bar{q}}$	70.9
	172	$(\mathrm{H}\to\mathrm{q\bar{q}})\mathrm{q\bar{q}}$	68.2
OPAL	161	$(\mathrm{H}\to\mathrm{q\bar{q}})\nu\bar{\nu}$	39.3
	172	$(\mathrm{H}\to\mathrm{b\bar{b}})\mathrm{q\bar{q}}$	75.6

Table 2: *Data events selected by the four collaborations. No candidates are selected by ALEPH, two by DELPHI, six by L3 (assuming m_{H}=70 GeV/c^2) and two by OPAL. In the channel designations, wherever the Higgs decay is not specified, the symbol H stands for H$\to$ all channels. The last column indicates the reconstructed Higgs boson mass, $m_{\mathrm{H}}^{\mathrm{REC}}$.*

Mass limit (GeV/c^2) (95% c.l.)	Statistical method A	B	C	D	Spread (GeV/c^2)
Observed	77.5	77.8	77.7	77.9	$\pm$0.2
Expected	75.8	76.0	75.6	75.7	$\pm$0.2

Table 3: *The 95% confidence level combined mass limits, observed and expected on average, using the statistical methods A, B, C and D to combine the results of the four collaborations. The last column shows the spread of the limits obtained, i.e. $\pm$ one-half of the largest difference.*

Experiment	Statistical method				Spread (GeV/c^2)
	A	B	C	D	
Observed 95% c.l. mass limits (GeV/c^2)					
DELPHI	65.9	65.9	65.5	-	$\pm$0.20
L3	69.4	69.3	69.2	-	$\pm$0.10
OPAL	69.0	68.6	68.9	-	$\pm$0.20
ALEPH	69.6	69.6	69.6	69.6	-
Expected 95% c.l. mass limits (GeV/c^2)					
DELPHI	65.4	65.3	65.1	-	$\pm$0.15
L3	66.1	65.7	65.0	-	$\pm$0.55
OPAL	65.9	65.6	65.3	-	$\pm$0.30
ALEPH	68.5	68.8	68.6	68.5	$\pm$0.15

Table 4: *The observed and expected 95% confidence level mass limits for the individual collaborations using, in each case, various statistical methods. The last column indicates the spread of the limits obtained, i.e. $\pm$ one-half the largest difference. The entries for Method D pertaining to the DELPHI, L3 and OPAL Collaborations are empty since this method was only used to combine the results and not to recalculate the individual limits.*

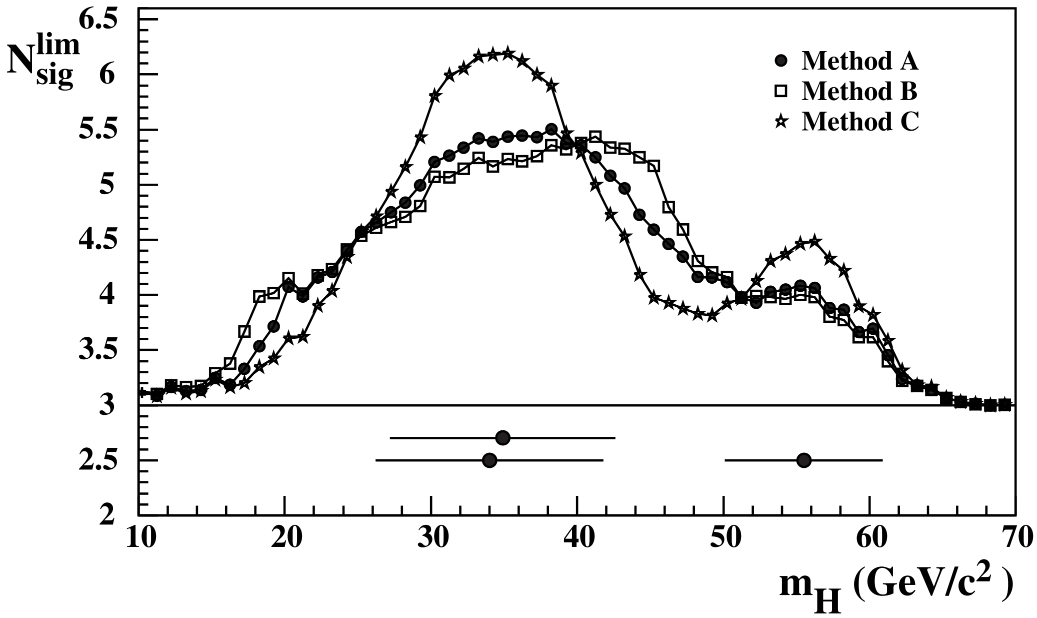

Figure 1: *Test (I); the 95% confidence level "experimental" upper bound for the number of signal events, N_{sig}^{lim}, as a function of m_H, obtained by Methods A, B and C. The smallest value, three, obtained far from the candidate events, is indicated by the horizontal line. The three dots with error bars show the masses and assumed resolutions of the candidate events.*

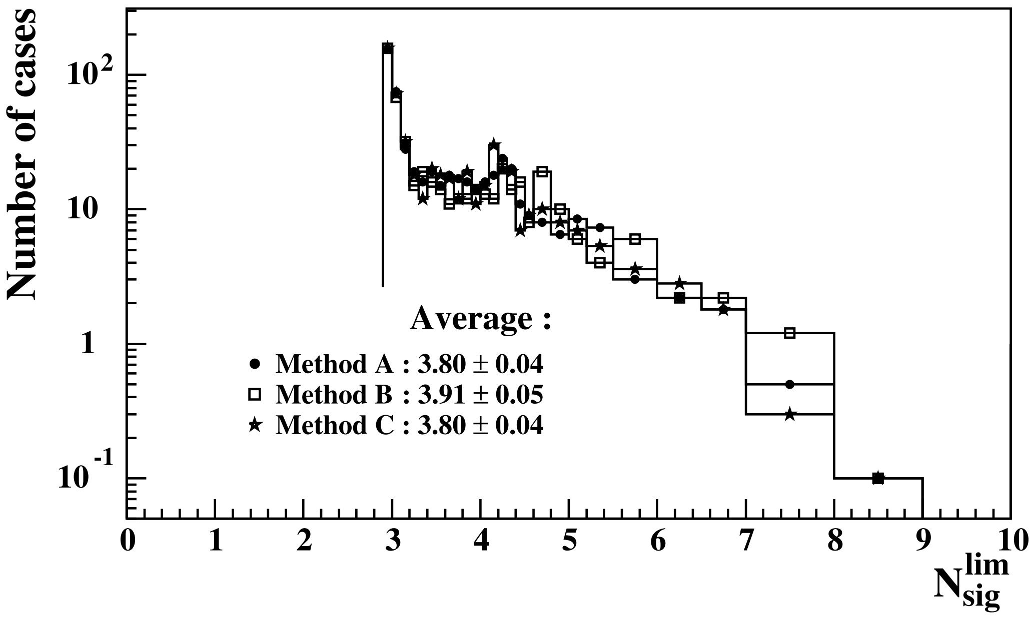

Figure 2: *Test (I); the distribution of the 95% confidence level "experimental" upper limit for the number of signal events,* $\mathrm{N}_{\mathrm{sig}}^{\mathrm{lim}}$*, assuming* $m_{\mathrm{H}}=40\ \mathrm{GeV/c^2}$*, in 600 gedanken experiments with background only, as obtained by Methods A, B and C. The average values of the limits are also given in the figure.*

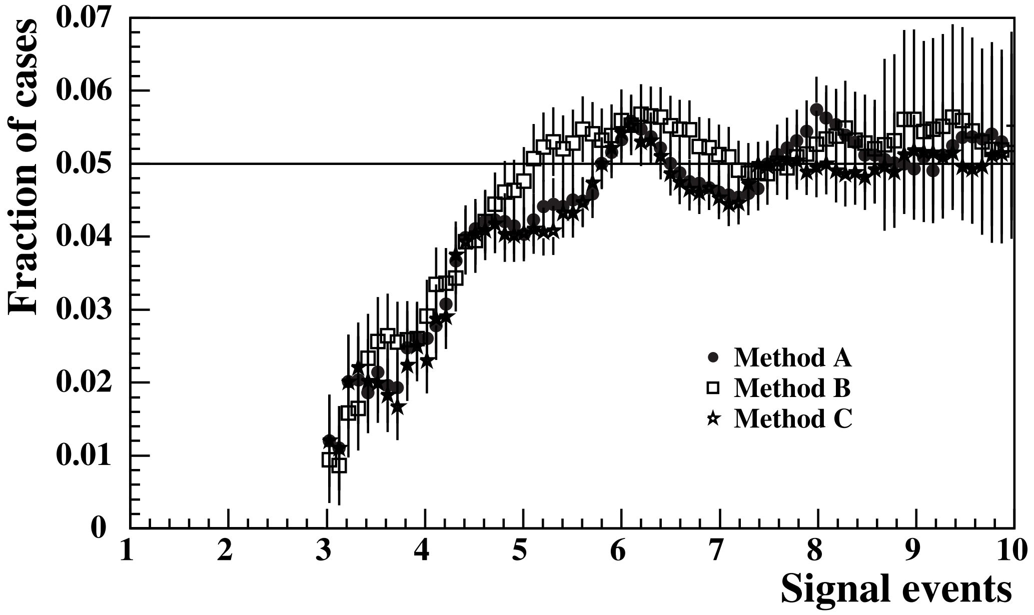

Figure 3: *Test (II); the fraction of gedanken experiments with signal+background in which an existing Higgs boson with 77* $\mathrm{GeV/c^2}$ *mass is actually excluded using Methods A, B and C, as a function of the signal event rate.*

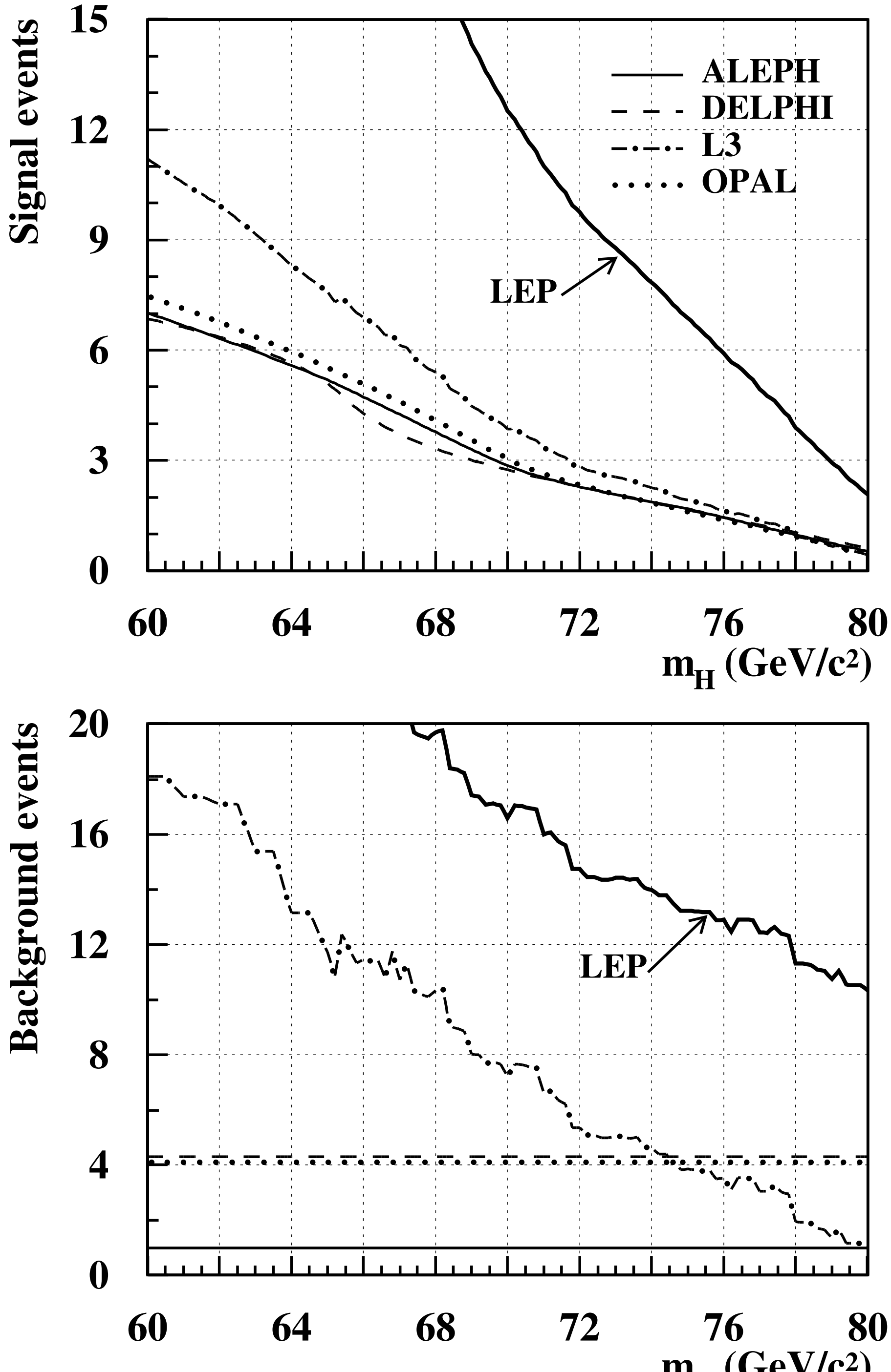

Figure 4: *Total expected numbers of signal and background events for ALEPH (full lines), DELPHI (dashed lines), L3 (dash-dotted lines) and OPAL (dotted lines), as functions of m_{H}. Data from $\sqrt{s}\approx$ 161 GeV and 172 GeV in all search channels have been added. The curves labeled "LEP" are the sums over the expected signal and background events of the four experiments.*

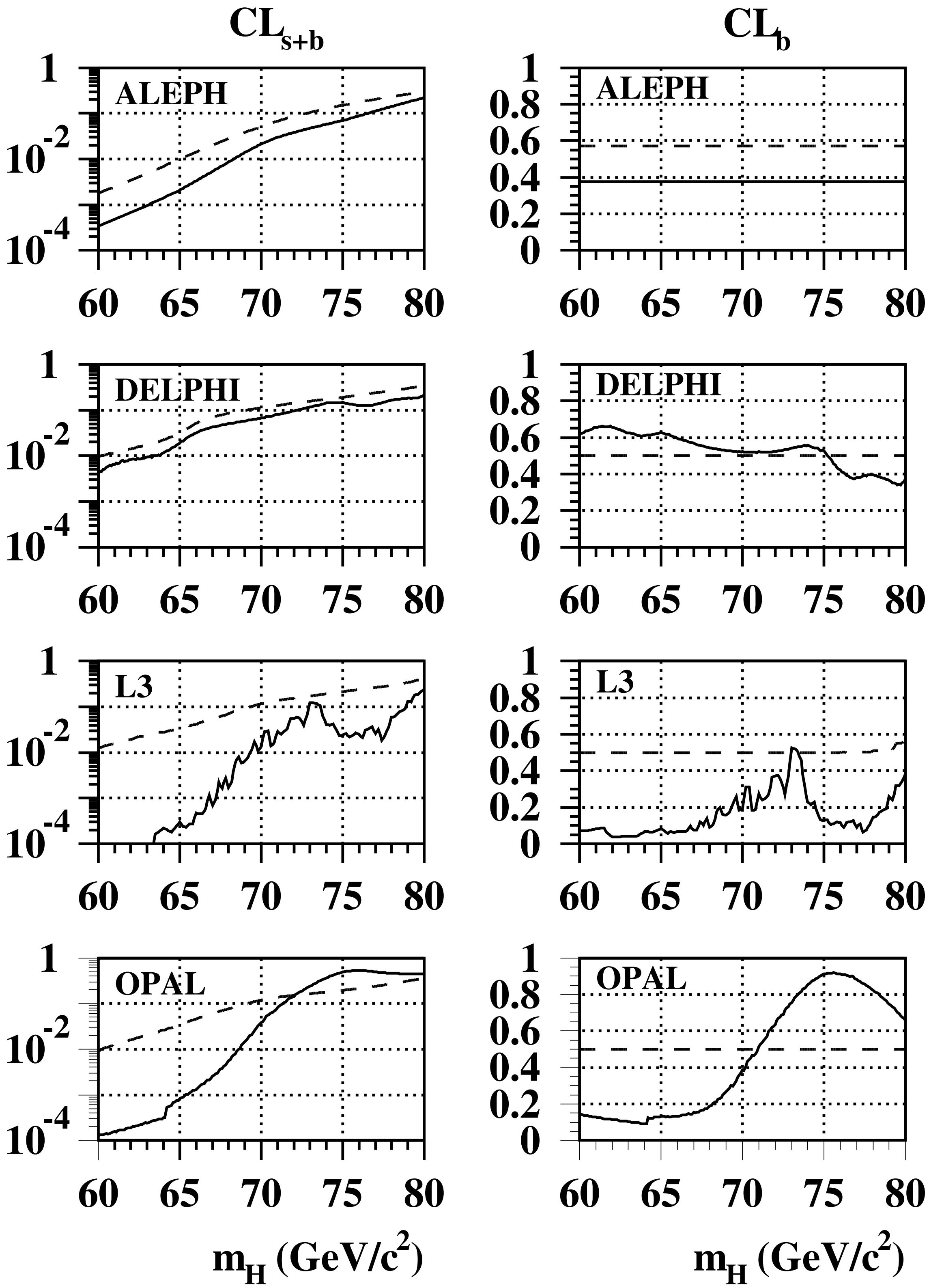

Figure 5: *Observed (full line) and expected (dashed line) confidence levels from the individual collaborations for the signal + background (left) and background only (right) hypotheses, as functions of* m_{H}.

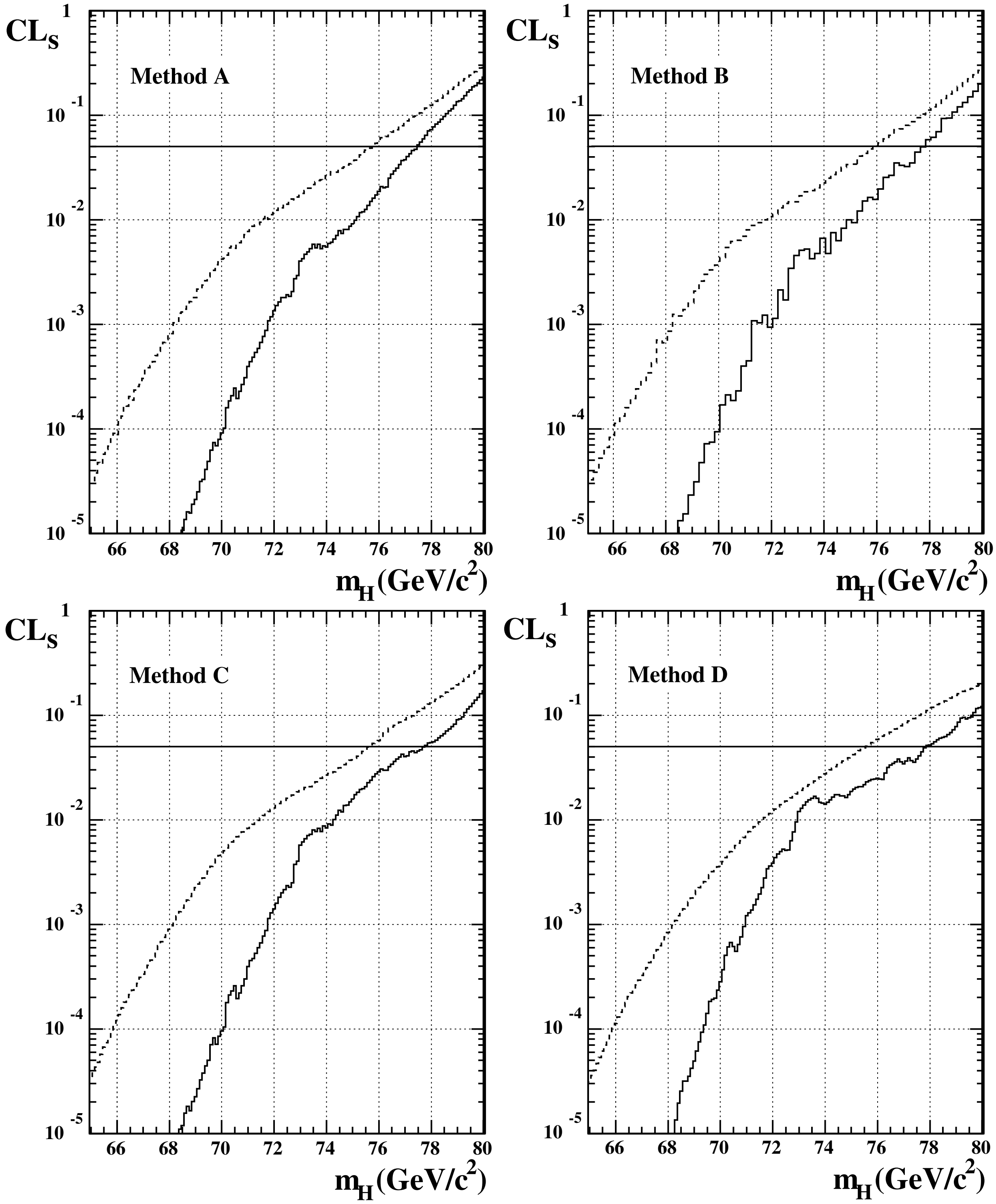

Figure 6: *Average expected (dashed lines) and observed (solid lines) confidence levels, $CL_s(m_{\mathrm{H}})$, obtained from combining the results of the four LEP collaborations using the four statistical methods investigated. The intersections of the curves with the 5% horizontal line define the 95% confidence level lower bounds, expected and observed, for the mass of the SM Higgs boson.*

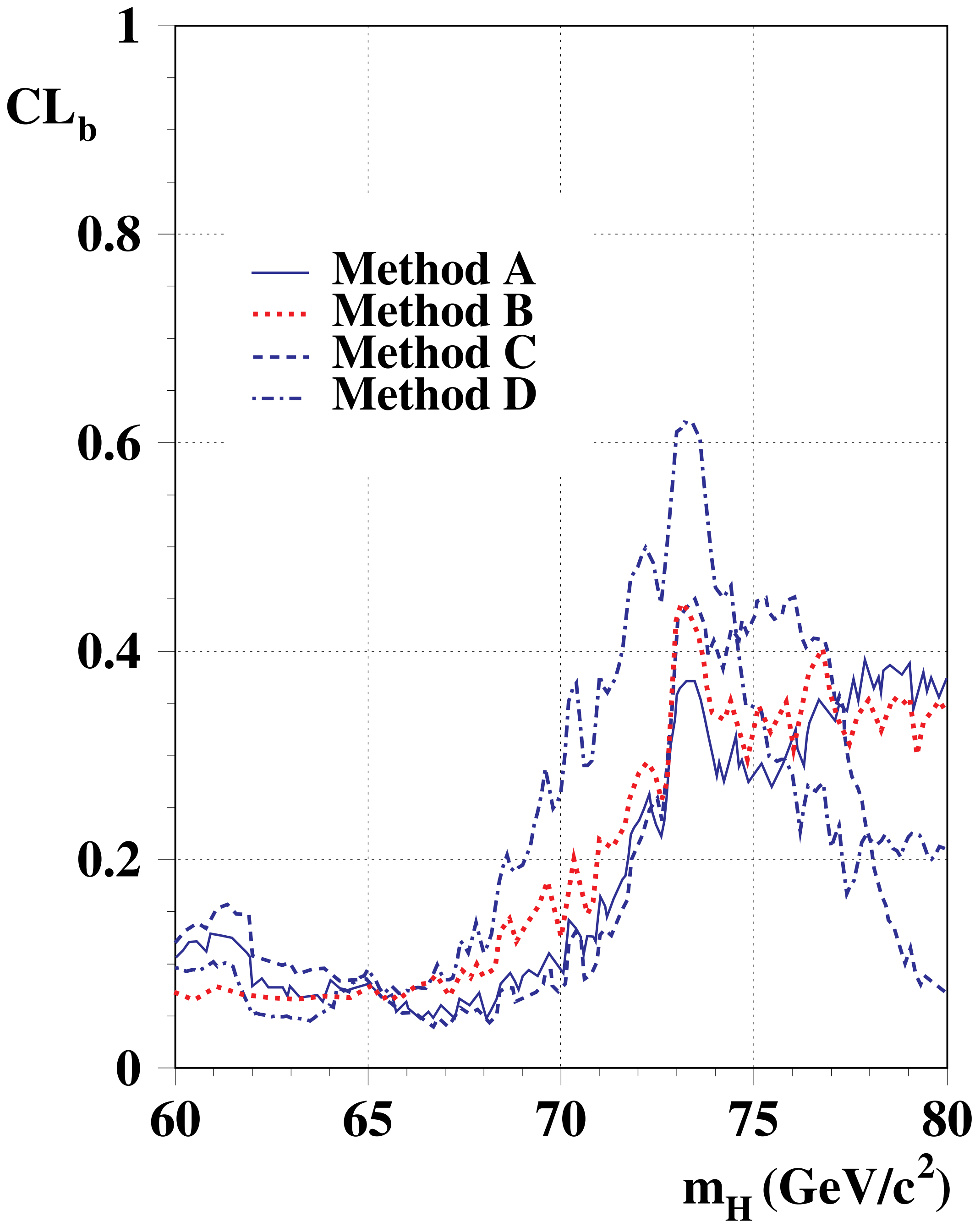

Figure 7: *Combined confidence levels for the background only hypothesis, $CL_b(m_{\rm H})$, obtained by each of the four statistical methods investigated.*