Use of discriminant analysis in search of a neutral Higgs boson

T.G.M.Malmgren and K.E.Johansson

Fysikum, Stockholm University, Box 6730, S-113 85 Stockholm, Sweden

Abstract

When searching for the Higgs boson produced in e^+e^- collisions at energies close to the Z^0 mass, a method to reject a very large background is needed. A method to pick up the signal of a Higgs boson produced in association with a pair of neutrinos has been investigated. After a preselection, which eliminates the majority of the background events, a discriminant analysis was used to achieve a background rejection factor of 168 while keeping an efficiency for Higgs boson events of 68% for a Higgs boson mass of $55 \text{ GeV}/c^2$. For the combined preselection and discriminant analysis, the rejection factor is 50 000 and the efficiency is 48%.

Submitted to *Nuclear Instruments and Methods A*

1 Introduction

At LEP, the Standard Model Higgs boson (H^0) is expected to be produced through the Bjorken process[1]. At center of mass energies close to the Z^0 mass, the Z^0 radiates the H^0 and decays into a pair of fermions. We focus on the largest decay channel with a distinct topology, namely the one for which the H^0 decays into a pair of heavy quarks and the Z^0 decays to a pair of neutrinos. This channel is different from the massive background consisting of two collinear jets. The H^0 predominantly decays into a b quark pair which is boosted due to the H^0 movement.

$$e^+ + e^- \rightarrow Z^0 \rightarrow \begin{array}{c} Z^{0*} + H^0 \\ \searrow \quad \swarrow \\ \nu \bar{\nu} \quad b \bar{b} \end{array}$$

This gives the signature of isolated missing energy and acollinear jets of particles from the b quarks. There are three main backgrounds. The ordinary Z^0 decay to a quark pair, the $q\bar{q}$ background, resembles H^0 events when particles escape detection due to weaknesses in the detector and when it has a physical H^0 -like signature, such as semi-leptonic decays of heavy quarks where isolated neutrinos are present. Events with both lepton- and quark pairs (four fermion processes) can also resemble H^0 events when the leptons are taus and decay into neutrinos. This process results in missing energy. The two photon processes can give rise to events that might look like H^0 events when particles escape detection and the scattered e^+e^- escape along the beam direction.

The cross-section for the H^0 decreases with an increasing H^0 mass. In addition, the signature becomes more like the $q\bar{q}$ background. For a luminosity of one event per pb, about one-fifth of an event is expected at an H^0 mass of 55 GeV/ c^2 , while roughly 31 000 events are expected from the $q\bar{q}$ background.

The problem of picking up a clear signal from the more than five orders of magnitude larger background has been solved by using discriminant analysis.

1.1 Discriminant analysis

In order to obtain a good separation between signal and background, a separating surface is searched for in the feature space. The latter is defined as the multi-dimensional space spanned by the event variables. Sequential cuts could be visualized as a rectangular box in this space. Linear discriminant analysis (DA)[2] may be visualized as a multi-dimensional hyperplane. However, the optimal separative surface is, in general, non-linear. Hence, we expand the linear DA to a non-linear by going from the first degree $D_1 = \mathbf{x}^T \mathbf{a}$ to the second degree $D_2 = \mathbf{x}^T (\mathbf{a} + \mathbf{B}\mathbf{x})$, where the scalar D is the output, the transpose vector $\mathbf{x}^T$ is the input feature variables while the vector $\mathbf{a}$ and the matrix $\mathbf{B}$ are the coefficients to be determined in the optimization. By expanding $\mathbf{a}$ to contain the elements of $\mathbf{B}$ and by expanding $\mathbf{x}$ to contain the elements of $\mathbf{x}\mathbf{x}^T$, the linear deduction of the coefficients may be used. The two criteria, that the local variance should be as small as possible while the separation should be as large as possible, could be expressed such that the ratio

$$R = \frac{(\mathbf{a}^T \Delta \mu)^2}{\mathbf{a}^T \mathbf{V} \mathbf{a}} \quad (1)$$

should be maximized where $\mathbf{V}$ is the variance matrix and $\Delta \mu$ the mean differences between the signal and the background. This leads to $\mathbf{a} = \mathbf{V}^{-1} \Delta \mu$.

2 Analysis

2.1 Event samples and preselection

The data used in the analysis were collected with the DELPHI detector[3] during 1994. A total of about 1.41 million hadronic Z^0 decays were collected, which corresponds to a total integrated luminosity of 45 events per pb. Events generated with the DELPHI Monte Carlo simulation routines[4, 5, 6, 7] were used. Jets were reconstructed using LUCIUS[8]. A total of 2.2 million $q\bar{q}$ and 2 500 four fermion events were generated. The background from two photon processes was ignored in this analysis as we have earlier shown this background to be negligible[9].

A total of 27 000 H^0 events were simulated[10] with H^0 masses ranging from 50 GeV/c² to 75 GeV/c² in steps of 5 GeV/c² with large samples of 6 000 events at 55-65 GeV/c² for optimization and testing the analysis method.

To reduce the samples to a workable size, a preselection was done. It also served to reject irrelevant components present in data but not in simulation like beam-gas interactions. The selection was based on feature variables such as acollinearity, invariant mass, transverse momenta and charged multiplicity[11]. The preselection retained 70% (4 246 events) of the H^0 events for an H^0 mass of 55 GeV/c² while reducing the $q\bar{q}$ background with a factor of 303 to 7 281 events.

2.2 Selection of variables

Formula 1 in the linear case was used to pick out the most discriminating variables. First, one variable was selected with the highest ratio R . Then, assuming there are m selected variables, variable x_{m+1} was picked if, combined with variables $x_1 \dots x_m$, the highest ratio R was achieved. The first fifteen variables were selected to be consistent with previous years

- the sum of the opening angles when forcing a three jet event, S_θ ,
- the invariant mass,
- the b-tag value,
- the transverse momentum,
- the balance between the two hemispheres,
- the number of jets,
- the energy within 20° of missing momentum,
- the b-tag value for hemisphere one,
- the b-tag value for hemisphere two,
- the minimum angle, in the x-y plane, between any jet and missing momentum,
- the energy within 40° of missing momentum,

- the minimum opening angle between jets when forcing a three jet event,
- the maximum opening angle between jets when forcing a three jet event,
- the absolute value of the cosine of the thrust axis,
- the ratio of the number of tracks used in the b-tagging and the total number of tracks.

2.3 Discriminant analysis in the H^0 search

To demonstrate the importance of having enough events to generalize the two event samples and to optimize the separation and yet to avoid overoptimizing, an optimization was done with various number of events. This is shown in figure 1 where a hypersurface in two dimensions can be seen. The two variables used were S_θ and the invariant mass.

Figure 1: Hypersurfaces for various number of events. The diamonds represent the H^0 events while the crosses mark $q\bar{q}$ events.

Figure 1a is made with six events and shows the risk of overoptimizing when using too few events. In figure 1b, the number of events used (50) is ten times the degrees of freedom

as suggested for neural networks[12]. The separating surface is beginning to converge. In figure 1c, the number of events used was set so that the number of events per degree of freedom equaled that of the analysis of 1994 data[11], i.e. 79 H^0 events and 135 $q\bar{q}$ events. Finally, in figure 1d, the full statistics is used (2 123 H^0 events and 3 640 $q\bar{q}$ events). It is clear that, while ten times the degrees of freedom might be enough for a good generalization, a larger event sample is desirable for a good convergence of the separating surface and a safe estimate of the efficiency and background.

An example of overoptimizing can be seen in figure 2 for a second degree DA with fifteen variables, where the coefficients are retrieved using five times the degrees of freedom. The figure shows the H^0 efficiency versus the $q\bar{q}$ rejection factor for an optimization

Figure 2: H^0 efficiency and $q\bar{q}$ rejection factor for the optimization- and test sample.

sample and a test sample. It is clear that while the optimization sample shows very good results (i.e. high H^0 efficiency and high rejection factor), a test sample shows a worse performance. It shows the importance to determine the real performance with an independent test sample. In this plot, the test samples consisted of independent samples equal the size of the full samples.

To explore this effect further, a second degree DA was optimized using 15 variables and event samples ranging from 135 events each to the full sample in ten steps. Figure 3a shows the difference in H^0 efficiency between the optimization sample and the test sample. With a small number of events per degree of freedom, the difference is large. This means that the optimization sample gives an overestimate of the efficiency. Figure 3b shows the difference between the number of $q\bar{q}$ events from the optimization sample and the test sample. At a low number of events per degree of freedom, the difference is large yielding an underestimate in the optimization sample. The arrow marks the number of events per degree of freedom for the analysis of 1994 data. It is evident that with an event sample of the order of or larger than ten times the degrees of freedom, the performance estimated with the test and optimization samples is very similar.

To study whether a higher degree DA gives a better performance than more variables, two optimizations were made by using, first, four variables and a second degree DA and, second, two variables and a fourth degree DA. The results can be seen in figure 4. It shows

Figure 3: Differences between optimization- and test samples. a) H^0 efficiencies, b) number of $q\bar{q}$ events remaining. (RF= $q\bar{q}$ Rejection Factor).

the rejection factor of the $q\bar{q}$ background and the H^0 efficiency. It can be concluded that, since both curves have the same degrees of freedom, adding more variables yields a higher gain in performance than going to a higher degree DA.

In figure 5, the H^0 efficiency as a function of the $q\bar{q}$ rejection factor for a first degree DA and a second degree DA is shown for different number of variables used. After having selected about ten variables, there is no large gain in the performance for both DAs.

2.4 Final results

In the final analysis, all 7 281 $q\bar{q}$ events and half of the 4 246 H^0 events remaining after the preselection were used to optimize the DA. This corresponds to event samples of 54 and 17 times the number of degrees of freedom, respectively. A second degree DA with 15 input variables was used to retrieve the 135 coefficients. All $q\bar{q}$ and four fermion events were passed through the analysis. An overall good separation between signal and background was achieved as can be seen in figure 6. The cut in the DA output was selected so as to optimize the selection efficiency while rejecting all background events when combined with an additional post-selection. After the DA cut, 33 $q\bar{q}$ events were left corresponding to 21 expected $q\bar{q}$ events in the data sample. This corresponds to a rejection factor of 168. The total rejection factor of the preselection and the DA stage is about 50 000. From the four fermion events, 0.09 were selected. An H^0 efficiency of 68% was achieved for an H^0 mass of 55 GeV/c^2 . The selection efficiencies can be seen in figure 7. It shows the preselection-, DA- and combined efficiencies. The decrease in efficiency at higher H^0 masses is due to the increasing similarity to the $q\bar{q}$ background making it harder to separate the two samples. In order to remove a few remaining $q\bar{q}$ events, a post-selection is applied. This led to a total H^0 efficiency of 32.1% for an H^0 mass of 55 GeV/c^2 .

Figure 4: Efficiency/rejection for the same degrees of freedom.

3 Conclusions

Nine Higgs particle events with a mass of $55 \text{ GeV}/c^2$ were expected to be produced in 1994 in the neutrino channel in the DELPHI detector at LEP. A total number of around 1.41 million Z^0 decays were accumulated, representing the potential background to the Higgs particle events. To search for the Higgs particle, a very good background rejection and signal efficiency is needed. A preselection is applied to remove the bulk of background letting a discriminant analysis reduce the more Higgs particle like background. A rejection factor of about 168 for $q\bar{q}$ events is achieved while keeping a Higgs particle efficiency of about 68% for $55 \text{ GeV}/c^2$ Higgs particle signal. Adding the preselection, the total $q\bar{q}$ rejection factor is about 50 000 and the $55 \text{ GeV}/c^2$ Higgs particle signal efficiency is 48%.

An additional post-selection was done in order to remove remaining background $q\bar{q}$ events leaving a total efficiency of 32% for a $55 \text{ GeV}/c^2$ Higgs particle. No candidates were found in the 1994 data.

The discriminant analysis proves to be a very effective tool in the Standard Model Higgs boson search in the neutrino channel.

Acknowledgments

We are greatly indebted to our DELPHI collaborators and to the funding agencies for their support in building and operating the DELPHI detector, and to the members of the CERN-SL Division for the excellent performance of the LEP collider.

Thanks to Dr.K.Hultqvist and Dr.R.Jacobsson for their assistance with the physics.

References

- [1] P.J.Franzini et al., *Z Physics at LEP*, CERN 89-08 Vol.2.

Figure 5: The H^0 efficiency as a function of the $q\bar{q}$ rejection factor for discriminant analyses with 1 to 15 variables. Left: first degree DA, right: second degree DA.

- [2] M.G.Kendall and A.Stuart, *The Advanced Theory of Statistics*, Volume 3.
- [3] DELPHI Collaboration, *The DELPHI detector at LEP*, Nucl. Instr. and Meth. **A303** (1991) 233.
- [4] DELPHI Collaboration, *DELSIM User's Guide*, DELPHI 89-67 PROG 142.
- [5] DELPHI Collaboration, *DELSIM Reference Manual*, DELPHI 89-68 PROG 143.
- [6] T.Sjöstrand, *High-energy-physics event generation with PYTHIA 5.7 and JETSET 7.4*, Comp. Phys. Comm. **82**(1994)74.
- [7] DELPHI Collaboration, *DELANA User's Guide*, DELPHI 89-44 PROG 137.
- [8] T.Sjöstrand, *The Lund Monte Carlo for Jet Fragmentation and e^+e^- Jet Physics: JETSET Version 6.2*, Comp. Phys. Comm. **39** (1986) 347.
- [9] DELPHI Collaboration: P. Abreu et al, *Search for the Standard Model Higgs Boson in Z^0 Decays*, Nucl. Phys. **B421** (1994) 3.
- [10] F.A.Berends and R.Kleiss, *Initial-state radiation at LEP energies and the corrections to Higgs boson production*, Nucl. Phys. **B260** (1985) 32.
- [11] DELPHI Collaboration, *Update of the Standard Model Higgs Boson Search*, DELPHI 95-69 PHYS-504, Contribution to the International Europhysics Conference on High Energy Physics, Brussels, Belgium.
- [12] L.Lönnblad, C.Peterson and T.Rögnvaldsson, *Pattern recognition in high energy physics with artificial neural networks – JETNET 2.0*, Comp. Phys. Comm. **70** (1992) 167.

DELPHI

Figure 6: Discriminant analysis output variable from a second degree DA.

Figure 7: Efficiencies after different selections.