NEURAL NETWORK APPLICATIONS FOR THE ANALYSIS OF LEP DATA*

Borut Paul Kersevan

Jozef Stefan Institute, Jamova 39, SI-1000 Ljubljana, Slovenia
and
Faculty of Mathematics and Physics, University of Ljubljana
Jadranska 19, SI-1000 Ljubljana, Slovenia

(Received November 12, 2003)

The general framework of the data analyses at LEP is presented, with the emphasis on the application of Artificial Neural Networks in event filtering and parameter estimation.

PACS numbers: 12.15.Ji, 13.66.Jn, 07.05.Kf, 07.05.Mh

1. The LEP legacy

The Large Electron Positron collider (LEP) was in operation at the European Laboratory for Particle Physics (CERN) at Geneva, Switzerland between the years 1989 and 2000 (see figure 1). During this period four multi-purpose detectors were collecting data of the reactions taking place in electron-positron collisions: `ALEPH`, `DELPHI`, `L3` and `OPAL`. The operation of LEP can roughly be divided into two eras:

- **LEP I** period, during which the e^+e^- beams were colliding at the centre-of-mass energy of about $\sqrt{s} = 91$ GeV, which corresponds to the mass peak of the Z^0 weak gauge boson. During this period more than 4 million Z^0 production events were recorded by each of the four experiments.
- **LEP II** era when the e^+e^- beams were accelerated to the range of energies from the production threshold of W^+W^- weak boson pairs (about $\sqrt{s} = 161$ GeV up to the highest energies of $\sqrt{s} = 210$ GeV which were reached in the year 2000.

* Presented at the Workshop on Applications of Neural Networks, Zakopane, Poland, May 30–June 8, 2003.

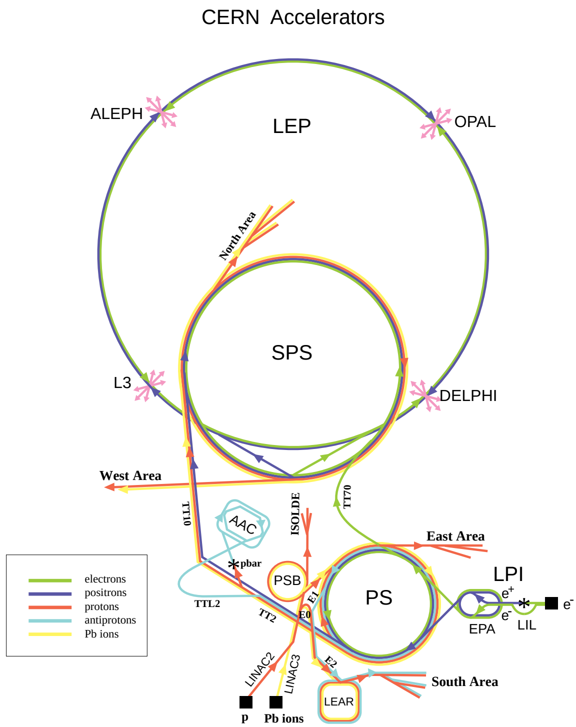

Fig. 1. The Large Electron Positron collider acceleration chain.

The topics of research, which were conducted by the four collaborations at LEP, can be divided into the following areas:

- **Precision measurements:** The field denotes very accurate measurements of the properties of already discovered particles which fit into the predictions of the Standard Model, *e.g.* the production cross-sections, masses, decay branching ratios *etc.* Figure 2 (left) shows an example of the $e^+e^- \to W^+W^-$ cross-section measurement and its agreement with the Standard Model predictions.

- **Tests of the Standard Model:** The high quality and quantity of the data collected at LEP made it possible to probe the Standard Model predictions to a very high precision and also consequently enabled one to set more precise limits on the missing parameters of the Standard Model (*e.g.* the Higgs boson mass) using indirect measurements. In figure 2 (right) an example on the Higgs boson mass constraints using the W boson and top quark mass measurements is presented.

- **New Physics Searches:** This area of research was concentrated on looking for processes and/or particles which would indicate disagreement with- or possible extensions of the Standard Model (*e.g.* Supersymmetric partners of the existing particles, Technicolor effects *etc.*), as well as the search for the not yet discovered Higgs boson particle, the mass of which remains the missing parameter of the Standard Model; a plot of the LEP estimates of the Standard Model Higgs boson mass is shown in figure 2 (bottom).

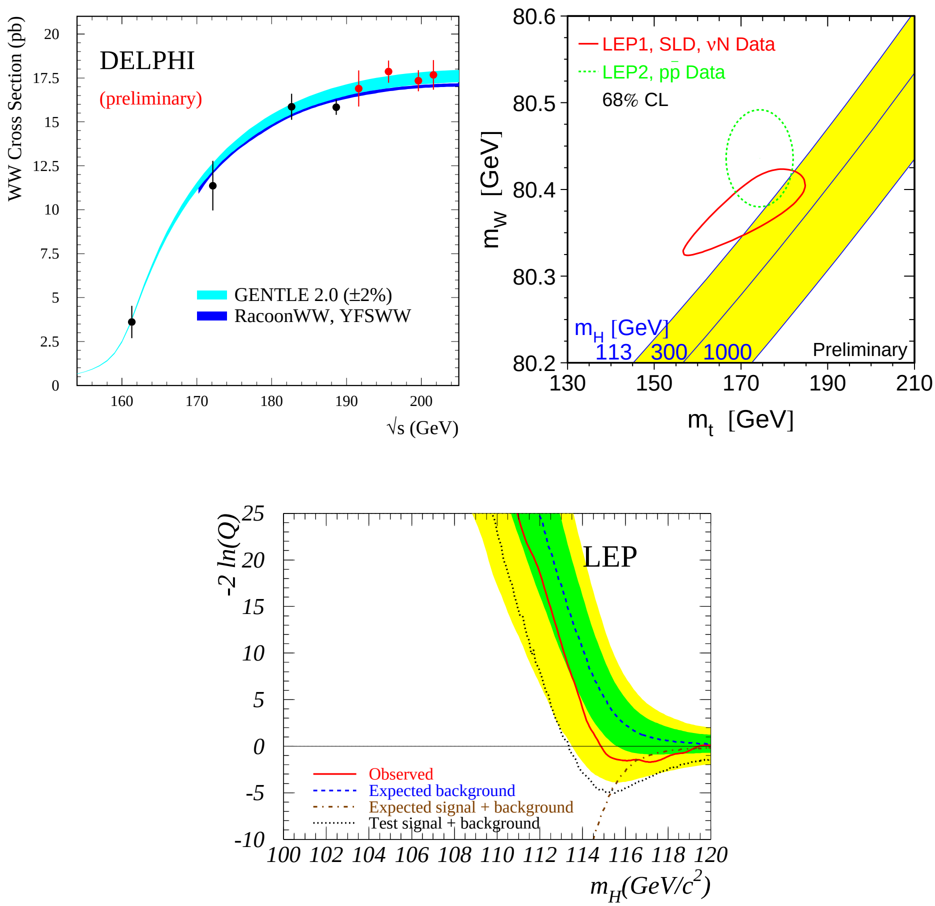

Fig. 2. Examples of fields of research at LEP: $e^+e^- \to W^+W^-$ cross-section measurement [4] (left), Higgs boson mass constraints using the W boson and top quark mass measurements (right) and estimation of the Standard Model Higgs boson mass probability [5] (bottom) .

2. The analysis chain

Let us outline a typical situation encountered when trying to estimate a certain parameter (or sets thereof) from the data collected by a LEP detector: As an example we can take one of the most significant processes discovered at LEP II, the W^+W^- production. In figure 3 a diagrammatic scheme of the process:

$$e^+e^- \to Z^0/\gamma \to W^+W^- \to \bar{s}c\tau^-\bar{\nu}_\tau \to \ldots$$

is presented. To describe it briefly, the electron and positron annihilate via the intermediate Z^0/γ to a pair of W bosons produced in a trilinear vertex. The two bosons then decay further into quark or lepton pairs which then hadronise and/or decay into final state (stable) particles, which can be observed inside a detector.

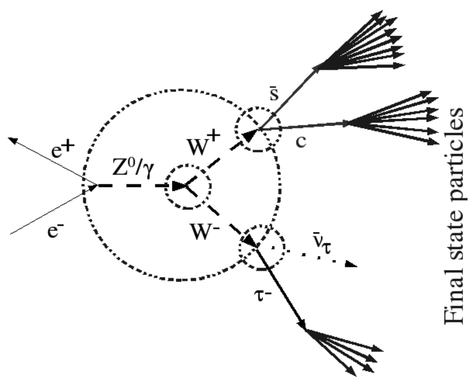

Fig. 3. A diagrammatic scheme of the process $e^+e^- \to Z^0/\gamma \to W^+W^- \to \bar{s}c\tau^-\bar{\nu}_\tau \to \ldots$. Possible regions/parameters of interest (resonances, couplings *etc.*) are marked with dashed circles.

The reaction itself has many theoretical parameters of interest, *e.g.* the value of its production cross-section, the $W^\pm$ resonance masses and couplings, $W^\pm$ couplings (branching ratios) to their decay products *etc.* One however, has to keep in mind that only the *final state* of the process is observable and that the process itself is *stochastic* which means that only the distributions of the kinematic quantities of the final state particles can be observed.

Due to the fact that only the final state particles are measured inside a detector, it might well happen that there exist other quantum processes which result in a similar final state configuration/topology. The 'signal' process we attempt to study can thus be obscured by various 'background' processes we are typically not interested in (*i.e.* we often know them better than the process we are analysing) but can severely interfere with our measurements. An example of a background process, the $e^+e^- \to Z^0/\gamma \to s\bar{s}g$, to the above signal is shown in figure 4. In practice many background processes are involved in a signal process analysis at LEP.

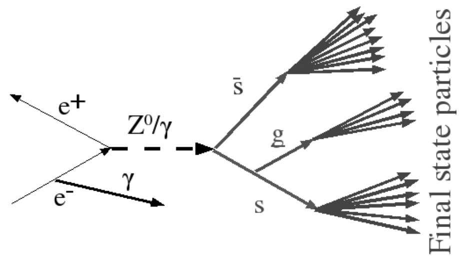

Fig. 4. A diagrammatic representation of a process $e^+e^- \to Z^0/\gamma \to s\bar{s}g$ resulting in a similar final state topology as the above 'signal process, thus contributing as potential 'contamination' (background) source inside a selected data sample.

A further fact which is often overseen by theoretical physicists is that the process information is collected by a detector, which as any machine has its implicit and explicit imperfections:

- Some particles like *e.g.* neutrinos are not detected.
- There is a finite measurement precision of any quantity, from particle energies and momenta to their production and decay vertex positions.
- Detectors are often at least partially blind to some particle properties, *e.g.* flavour or polarisation.

The four LEP detectors were excellent machines but were nevertheless still burdened by the limitations listed above. In addition, they were built as multi-purpose detectors to be able to detect as wide range of reactions as possible, thus almost no *a priori* background rejection was performed. The reader can get an impression of the quantities measured inside a detector in figure 5, which shows a computer reconstruction of a four jet event inside the `DELPHI` detector.

To summarise the above discussion into a few points to remember:

- In a data sample there is (potentially) a set of signal events along with a quantity of background events with similar kinematic signature.
- The event topologies are smeared and/or only partially reconstructed due to the detector precision and limitations.

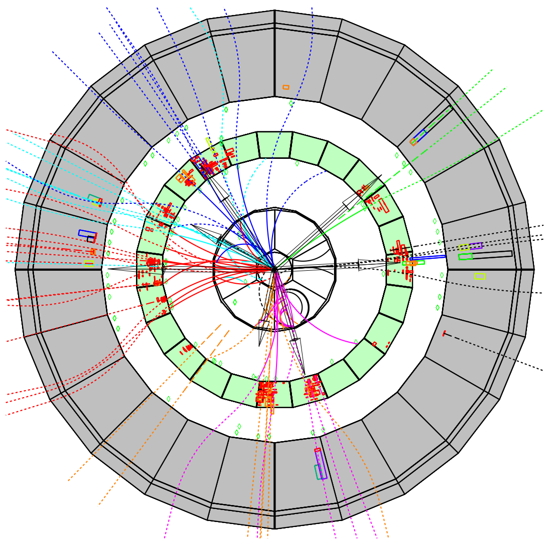

Fig. 5. A computer reconstruction of a four jet event inside the DELPHI detector. Only the final state particles are detected, so the original process resulting in the four jet topology cannot be estimated directly.

In order to obtain an estimate of the quantity (parameter) one wants to measure, the data sample has to be processed to:

- *Filter out* the signal candidate events.
- *Extract/reconstruct* the quantity of interest.

In the LEP environment the task was very demanding since one was dealing with precision measurements in a very complex environment.

3. Event filtering

In modern high energy physics experiments it would be quite inadequate to expect that the parameters of interest can be extracted by looking at the data alone (although it remains very desirable); in an advanced physics analysis at LEP extensive prior knowledge and assumptions about the detector performance and the underlying physics processes was employed to gain sufficient information for effective filtering algorithms. Indeed, one can often in advance assume some kinematic/topological quantities which would differ in signal and background events, however quantifying them is often a very tedious task.

To this end, the Monte Carlo simulation of processes and detector performance is predominantly used, which should, if properly tuned, give accurate predictions of kinematic distributions and event rates for different processes one expects to observe inside a detector[1].

Once an adequate Monte Carlo simulation of physics processes and detector response is set up one can effectively employ it in the process of event filtering by constructing (multi-dimensional) probability density functions for signal and background processes and look for optimal separation (filtering) procedure on this basis.

The subsequent procedures which all stem from the above basis are varied; the perhaps easiest and most 'robust' approach is the so called 'cut-based' analysis, which is based on a simple iterative procedure:

- Find a signal-sensitive variable, *i.e.* a kinematic quantity which is supposed to exhibit at least somewhat different behaviour in signal and background processes.
- By using the Monte Carlo simulation one can determine the expected kinematic distributions of reconstructed signal and background events (from the statistical point of view the distributions are actually projections of a multi-dimensional PDF with respect to the selected variable).
- Make a selection *cut* on this parameter, which is estimated to give an optimal background rejection while still retaining as many signal events as possible (the criterion is thus mostly the maximal value of the product of the expected sample 'purity' times the expected selection 'efficiency').
- Repeat the procedure until all the (reasonable) options are exhausted.

An example of two sequential cuts while filtering the $e^+e^- \to W^+W^- \to 4q$ signal in presence of various backgrounds at LEP II is shown in figure 6.

Although the above cut-based procedure is expected to be the least problematic due to its simplicity there are still certain disadvantages in using it:

- At LEP, cuts on *many* parameters are often required to achieve an efficient filtering procedure.
- The kinematic variables used are often correlated (*i.e.* non-orthogonal) which reduces the filtering efficiency.

In principle one could imagine that the cuts procedure in effect isolates a N-dimensional hypercube around a complex (non-trivial) signal PDF region. An example of such a case is given in figure 7.

[1] The Cracow group is indeed very well known inside the LEP community since they produced excellent Monte Carlo tools which were widely used in the LEP analyses, *e.g.* KoralZ, KoralWW, YfsWW, Tauola, Bhlumi ... to name just a few.

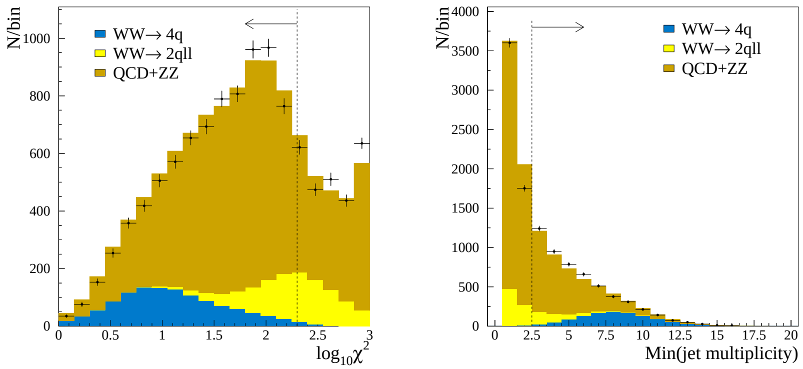

Fig. 6. Examples of two cuts applied on the data sample when trying to filter out the (four-jet) signal events $e^+e^- \to W^+W^- \to 4q$ in presence of several background processes. The data is represented with error bars and the Monte Carlo predictions with histograms.

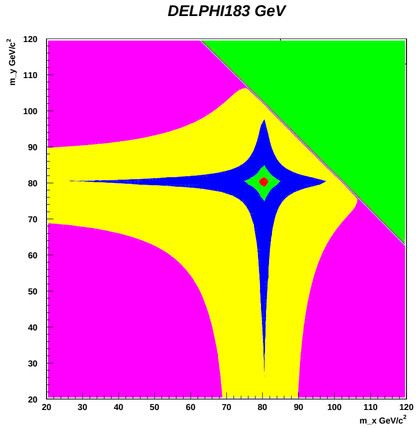

Fig. 7. Example of the expected PDF distribution of the masses of two W bosons produced in a $e^+e^- \to W^+W^- \to \ldots$ process. The shape is a star-like distribution originating in a product of two Breit–Wigner functions; the green area represents the kinematically inaccessible region. Clearly, cuts in terms of $|m_1 - m_W| < \varepsilon \times |m_2 - m_W| < \varepsilon$, which would result in a rectangular selected PDF region, would be far from optimal.

In the LEP environment one can clearly due better due to the high quality of the data collected by the LEP experiments, good understanding of detector performances and the availability of very accurate simulations of physics processes. Indeed, many advanced techniques were developed, among which the use of *Artificial Neural Networks* (ANNs) was established as the most effective one.

4. Use of Artificial Neural Networks in the LEP analyses

4.1. Event selection

At LEP various Artificial Neural Network tools (programs) were used, among which the SNNS package (IPVR Stuttgart) [1] and JETNET program (Lund University) [2] were the most widely used ones. The general strategy when using the ANNs in the event filtering procedure at LEP can in brief be outlined as follows:

- First, some loose (preliminary) kinematic cuts were made to reduce the size of the data sample.
- A *feed-forward* ANN is constructed, with a selection of kinematic variables fed in at the input neurons; a few hidden layers and one output neuron, the value (a real-valued parameter) is interpreted as the signal probability.
- The ANN is then trained by using the Monte Carlo simulated events which are accurately tagged on the output neuron; after a sufficient amount of training cycles the ANN should a fair reproduction of the input value on its output neuron/layer. An example of a typical ANN structure is given in figure 8.
- A final selection cut is made on the neural network output to separate the signal event candidates from the predicted background events.

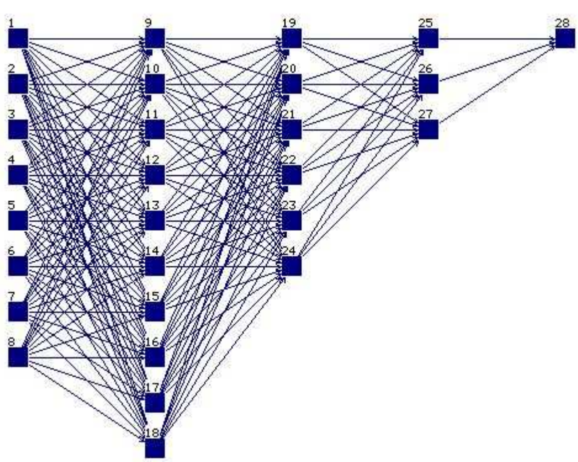

Fig. 8. Example of a Artificial Neural Network layout presented by the SNNS package. One input layer on the left with three hidden layers and one neuron in the output layer is chosen, the neurons are chosen to be fully connected to the neighbouring layers.

Two examples of ANN-based selections at LEP are listed below:

- `DELPHI` measurement of the production rate $\sigma(e^+e^- \to W^+W^- \to q_1\bar{q}_2q_3\bar{q}_4)$ [4], *i.e.* a four quark jet topology is expected in the final state recorded by the detector:
  - The JETNET package was used.
  - 10 input variables (neurons) were selected.
  - One hidden layer with eight neurons was found to be sufficient.
  - The *standard back-propagation* algorithm was used in the training phase.
  - The resulting (estimated) selection efficiency was on the order of 90% with the expected event purity of about 75%; comparatively, an advanced cut analysis managed to achieve a 85% selection efficiency and 75% sample purity. The five percent improvement might not seem much but in precision analyses at LEP it was certainly an significant achievement.
  - An example of the signal probability distribution for Monte Carlo simulation and LEP data recorded at $\sqrt{s} = 207$ GeV is given in figure 9.

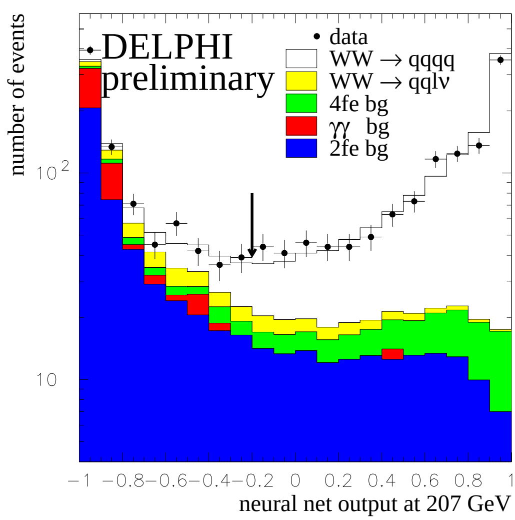

Fig. 9. An example of the signal probability distribution as given by an ANN for Monte Carlo simulated signal and background (histograms) and LEP data (error bars) as recorded at $\sqrt{s} = 207$ GeV by the `DELPHI` detector [4].

- ALEPH search for the Higgs boson [6]: Artificial Neural Networks were indeed used in Higgs boson searches by all four LEP experiments due to their versatility. The signal processes with the highest expected event rate were: $e^+e^- \to HZ^0 \to b\bar{b}q_3\bar{q}_4$ and $e^+e^- \to HZ^0 \to b\bar{b}b\bar{b}$, *i.e.* processes with hadronic Z^0 decays and Higgs boson decaying into a b-quark pair.
  - Both SNNS and JETNET packages were used.
  - 14–22 input variables (neurons) were chosen in different analyses.
  - One or two hidden layers with from 8 to 10 neurons were employed.
  - From one to three output neurons were used to give the signal classification and probability.
  - The standard back-propagation algorithm was used in the ANN training procedure.
  - Additional ANNs were used for *b-tagging*, *i.e.* determining the provenance of hadronic jets in the event.
  - An example of a ANN output for the Higgs searches at ALEPH in the year 2000 is shown in figure 10.

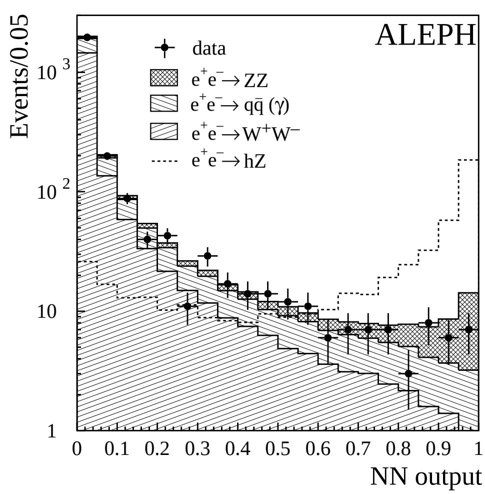

Fig. 10. An example of the signal $e^+e^- \to HZ^0 \to \ldots$ probability distribution as given by an ANN for Monte Carlo simulated signal and background (histograms) and LEP data (error bars) in the Higgs boson searches at ALEPH.

4.2. Kinematic reconstruction

In addition to event filtering the Artificial Neural Networks were at LEP also extensively used to extract (reconstruct) a parameter of interest inside the recorded (and filtered) events. The procedure of parameter extraction is often by an order of magnitude more taxing than event filtering since in the LEP environment:

- One attempts to reconstruct a quantity inside a very complex event topology.
- One has to disentangle often very *minuscule* signatures to tag the quantity of interest.

The use of ANNs for such task was found to be a very efficient method. Two examples of the most widely used applications of ANNs at LEP are presented below.

- *b-tagging* procedures: The task at hand is to differentiate and tag the hadronic jets which originate in b-quark fragmentation. The efficient b-tagging was of utmost importance in *e.g.* Higgs boson searches. For example the `ALEPH` b-tagging procedure was set up as follows:
  - The JETNET package was used.
  - Six input variables (neurons) were employed.
  - Two hidden layers were constructed.
  - One output neuron value served as the b-tag probability.
  - The standard back-propagation algorithm was used in ANN training phase.
  - The achieved *non-* b-jet rejection was about 85% with the b-jet tagging efficiency of 85%. A sample output of the ANN b-tagging probability is presented in figure 11.
- Inclusive reconstruction of *B hadrons* in $e^+e^- \to Z^0 \to b\bar{b}$ reactions at `DELPHI`: The BSAURUS package [3], possibly the most advanced ANN setup at LEP, was used and consisted of several ANNs in various processing stages: vertex estimation, B-meson flavour tag, b-quark flavour tag *etc.*
  - The JETNET package was used throughout.
  - At least ten input variables (neurons) were used at each stage.
  - One hidden layer with $N = N_{\mathrm{input}} + 1$ nodes was found to be sufficient.
  - One output neuron produced the tagging probability.

– The standard back-propagation algorithm was used in ANN training phase.
– The achieved resolution was rather staggering, an example of quark flavour tag output is shown in figure 12.

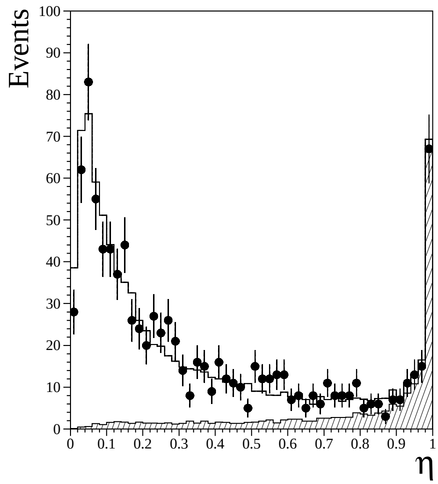

Fig. 11. An example of the ANN output when used in the b-tagging procedure; the simulated b-jet sample is shown with the full and non b-jet sample as the empty histogram; the LEP data is drawn with error bars.

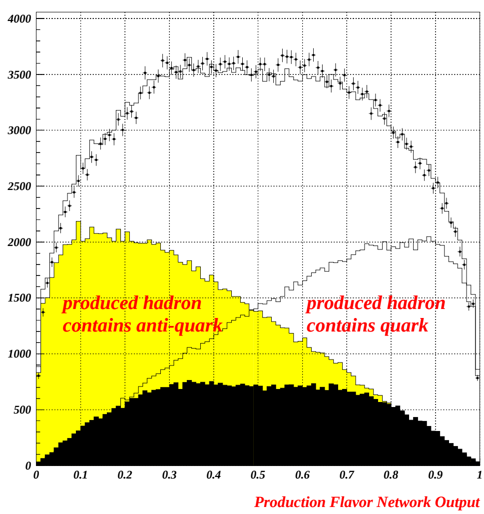

Fig. 12. An example of quark flavour tag output as given by the ANNs in the BSAURUS package is shown, presenting the achieved separation between the hadrons containing quarks and anti-quarks. The predictions are drawn with histograms and the LEP data with error bars.

5. LEP experience with ANNs

The LEP analyses using ANNs in event filtering or reconstruction procedures often encountered a certain degree of skepticism when presented in the scientific community, the main objection to the use of ANNs being that they are sort of a *black box* tool, which does not give a physicist sufficient control over the procedure and does not enable one to make a simple (proper) statistical (and systematic) uncertainty estimation because of *non-trivial* error propagation. The latter is actually with a closer statistical inspection somewhat of a *dogma*, since it can be shown, that the same problems in uncertainty estimation are encountered in any analysis using a complicated (non-Gaussian) probability density function and/or non-linear transformations of observed quantities. As it turns out, a variance estimate can still be obtained by randomly smearing the input quantities within the expected uncertainty limits as done in any advanced statistical analysis using non-Gaussian PDF-s.

The true problem of the application of ANNs are on the other hand possible *biases* in estimators due to the non-linear response of the ANNs, *e.g.* the danger of over-emphasising small fluctuations in the observed event distributions. The origins of such (possible) biases in ANN-based procedures have been found to be threefold:

- One has to rely heavily on the simulation predictions which, as good as they might be, could still exhibit some imperfections in certain regions of phase-space.
- Extremely large Monte Carlo samples are needed to adequately populate the whole phase space of interest; the sample sizes are often going into millions of simulated events. If the sample is too small certain phase space regions might remain empty, resulting in ambiguous ANN output since it was not trained to respond to occurrences in the given phase space region.
- Danger of *overtraining* the ANNs, *i.e.* the ANN has become too adapted to the simulated sample used in training, giving erroneous values to any events not belonging to this sample. This issue is the least problematic since adequate tests for over-training have been developed; there are however specific cases (*e.g.* exotic particle searches) where even slight over-training might result in potentially disastrous effects.

The possibilities of such occurrences in LEP analyses have certainly been carefully studied; on the whole the ANN performance turned out to be quite robust with respect to such fluctuations. An additional widely applied approach was to use a parallel cut-based analysis as a cross-check, as done *e.g.* in ALEPH Higgs boson searches [6], the example being given in figure 13.

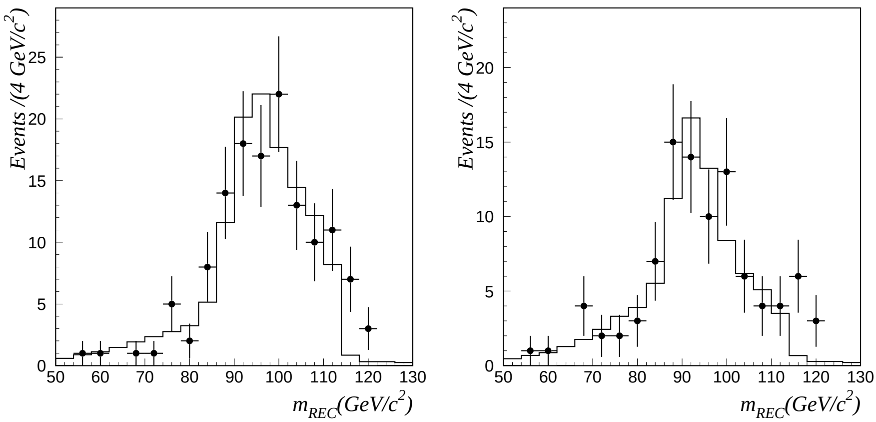

Fig. 13. The (reconstructed) mass distributions of the events passing the event filter in the ANN-based analysis (left) and the cut-based analysis (right) in the ALEPH Higgs boson searches in the year 2000 [6]. The background predictions are drawn with histograms and the data with error bars. A compatible number of excess events (Higgs event candidates) was found, observable as the excess of events in the high mass region.

The ANN-based procedures in LEP experience are still not *automatised* tools which can be applied with little human intervention. To list a few (minor) grudges related to the use of ANNs:

- The optimal ANN topology is not a-priori obvious (number and choice of input parameters, number of hidden layers) and often requires a lot of guesswork and iterations.

- The ANN analyses are still quite time consuming, both from the view of personal and above all computing time.

Nevertheless, during the LEP era the Artificial Neural Networks have become widely used in the high energy physics community and have clearly shown to be one of superior tools in many fields of scientific application.

REFERENCES

[1] Stuttgart Neural Network Simulator SNNS, Institute for Parallel and Distributed High Performance Systems, University of Stuttgart, 1993.

[2] L. Lonnblad, C. Peterson, H. Pi, T. Rognvaldsson, JETNET — A Neural Network program for jet discrimination and other High Energy Physics triggering situations, Department of Theoretical Physics, Univ. of Lund, Sweden, 1994.

[3] T. Allmendinger, G. Barker, M. Feindt, C. Haag, M. Moch, BSAURUS: A Package for Inclusive B Reconstruction in DELPHI, DELPHI-2000-069-PHYS-868, IEKP-KA-2001-3, `hep-ex/0102001`.

[4] P. Buschmann *et. al.*, DELPHI Collaboration, Measurement of the W-pair Production Cross-section and Branching Ratios at $\sqrt{s} = 205$ and 207 GeV, Preliminary results for the 2001 Summer Conferences, DEPHI 2001-104-CONF-532, 2001.

[5] P. Lutz, Higgs Working Group Report Towards the Final LEP Results, talk at the LEP Jamboree, 2002.

[6] R. Barate *et al.* [ALEPH Collaboration], *Phys. Lett.* **B495**, 1 (2000).