The Results of a Neural Network Statistical Event Class Analysis

SNO-STR-96-001

S.J. Brice, Oxford University

January 26, 1996

Abstract

Using a neural network technique, this paper shows that hit pattern recognition can identify electron scattering, charged, and neutral current event rates to a statistical accuracy of between 2 and 7% with only one year of salt fill data. Similar performance is also achieved with one year of pure D_2O data. A study of the suspected dominant systematic is made and indicates that this uncertainty will be less than 2%.

1 Introduction

A proper description of the feedforward neural network technique has been given in a previous STR [Bri95a], but a brief summary is appropriate. The neural network itself is no more than a functional mapping from some N dimensional space to a space of 1,2 or 3 dimensions (the mapping can be more general than this but the work of this paper only requires this low dimensional output space). That is, it takes a set of N numbers and calculates a function of them to produce 1,2, or 3 numbers depending upon the form of the network. The power of the technique arises from the ability to optimise the network function to carry out a desired mapping.

To carry out the analysis of this paper the first task is to decide upon a set of parameters to extract from each hit pattern that will serve to describe the pattern. For each event these parameters are the N numbers that are input into the network function. The desired functional mapping is one which tends to cluster events belonging to a particular event class (NC, CC, etc.) in a particular part of the output space with maximum separation between the clusters. The optimisation process that produces this mapping requires a labelled training set of events, that is, events whose class is known. These can come from Monte Carlo, or preferably from real/calibration data. The optimisation process is described in [Bri95a] and results in a network that carries out a functional mapping as close as possible to that desired. Once the network has been trained to produce its optimal mapping, a set of hit patterns for each event class is fed through the function and a histogram of the network output values for each event class is formed. These histograms can be viewed as probability distribution functions, one for each class of event. An example of these PDF histograms for a network with one output, distinguishing between two event classes is given in Fig. 1. Given a set of real data events of mixed and unknown classes, the task is to determine the fraction of the data events belonging to each event class. This is done by extracting parameters from each event and feeding the real data set through the trained network to form a real data histogram from

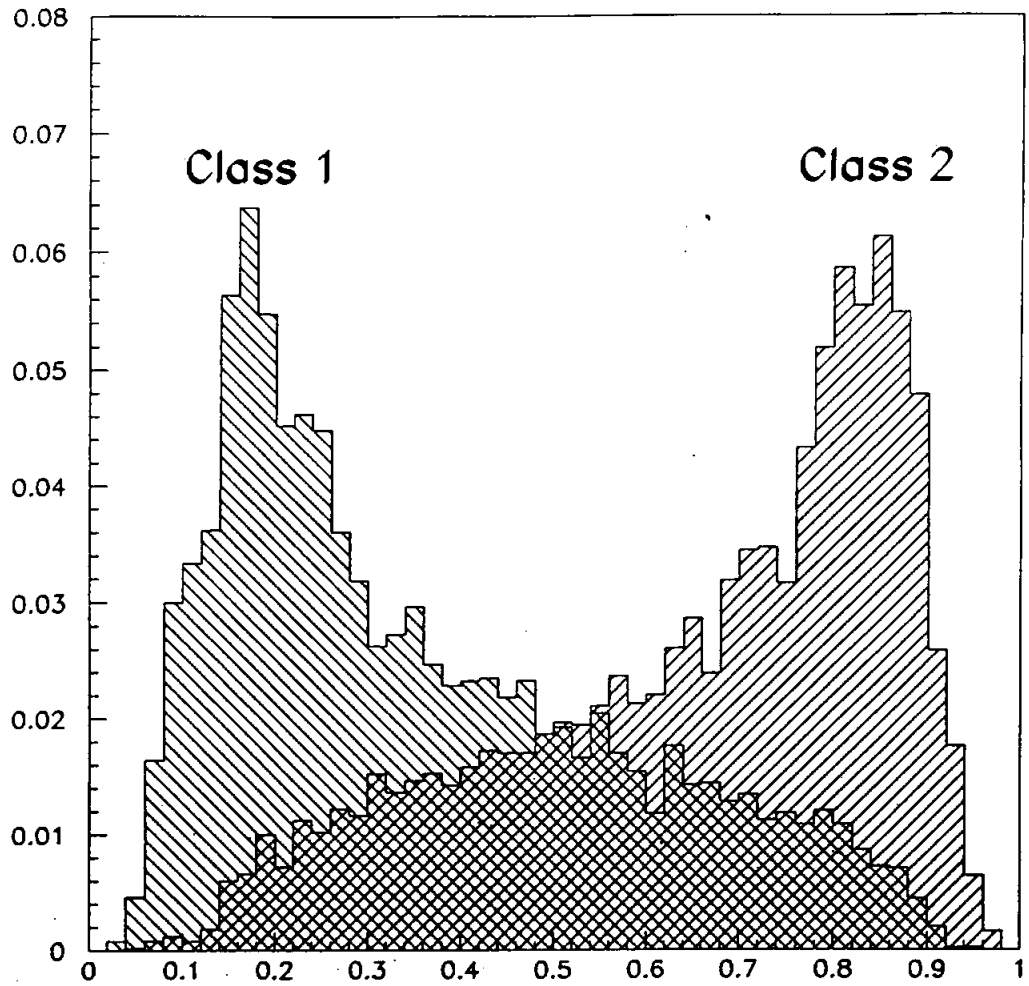


Figure 1: Normalised PDF histograms for a network with one output attempting to distinguish between two event classes

the network outputs. This real data histogram is then fitted to a linear combination of the PDF histograms thus giving the required event class fractions. This procedure is described as statistical hit pattern recognition in [Bri95a].

This paper is laid out as follows: Sections 2 and 3 describe the logic and mathematics of the parameters extracted from each hit pattern. Section 4 covers the various detector run scenarios assumed in this analysis and then Section 5 covers some aspects of training and fitting. Section 6 presents the results of the event class analysis with some further discussion taking place in Section 7. Section 8 makes a first pass at assessing systematic uncertainties. Finally Section 9 outlines some further applications of neural networks that may be possible.

2 Hit Pattern Features

The first step in defining a set of parameters which will describe each hit pattern is to delineate the hit pattern features which are relevant. One can then go on to find parameters which cover each feature. These relevant features are:-

- **Spatial Hit Distribution**:- This feature covers the positions of the hits that comprise an event. If the Čerenkov light is considered to arise from a point (an approximation that is reasonably valid even for gamma cascades and $\beta\gamma$ events, given the resolution of SNO) then the entire spatial hit distribution is held in the set of θ and ϕ values, giving the direction of each hit from the event position.

- **Temporal Hit Distribution**:- If all the Čerenkov light genuinely arises from a point then the information held in the PMT hit times is redundant given that held in the distances of each hit from the event position (this excludes scattered and reflected light). However, this pointlike approximation is known to break down for some event classes and in such cases there will be additional information in the temporal hit distribution.

- **Spatio-Temporal Hit Distribution**:- There will also be useful information in the correlations between hit positions and hit times and there should be parameters which mix spatial and temporal information.

- **Solar Angle**:- The charged current and electron scattering cross-sections have a distinctive angular dependence. This means that there is useful information in the angle between the fitted direction of an event and the direction of the sun.

- **Number of Hits**:- Different event classes have very different NHIT spectra, but great care must be taken with this parameter, as described below.

- **Radial Position of Event**:- The distribution of event radius for neutral current events shows a characteristic drop off as the acrylic vessel is approached. This is behaviour not seen in the other signal classes

In the analysis described in this paper all of the above event features have been used in characterising hit patterns with the exception of the radial position of the event. This potentially valuable feature has been omitted because it is distorted, to an *a priori* unknown degree, by the neutral current background originating in the acrylic vessel.

The NHIT spectra can be predicted for all event classes with the exception of charged current and electron scattering events. For these two the MSW effect may distort the ν_e spectrum. This means that unless the network can be trained on real data sets the NHIT parameter should probably not be used. It also means that great care must be taken to ensure that none of the other parameters have a strong NHIT dependence.

To a good approximation the detector can be taken to be spherically symmetric and, with the exception of solar correlations, the distribution of event directions is isotropic. With the solar directional correlations being taken care of by the solar angle parameter, these facts strongly suggest that the parameters describing the spatial hit distribution should be rotational invariants. This means that a spatial hit pattern occuring on one side of the detector will have exactly the same extracted parameters as the same hit pattern transformed by a rotation about the event position. This rotational invariance gives a fixed number of spatial parameters far greater descriptive ability, as well as removing from them the errors inherent in a fitted direction.

3 Extracted Parameters

This section takes each feature outlined in Sec. 2 and describes the parameter or parameter set which encapsulates the information. The first step in extracting parameters from a hit pattern is to fit a position for the event (a direction is also needed for the solar angle parameter). This position is then used as the coordinate origin.

3.1 Spatial Hit Distribution — The Harmonic Parameters

First define the spatial hit pattern function:-

$$f(\theta,\phi) = \frac{1}{N}\sum_{i=1}^{N}\delta(\cos\theta - \cos\theta_i)\,\delta(\phi-\phi_i) \tag{1}$$

which is a function of the polar and azimuthal angles of each hit with respect to the event position with a delta spike centered on each hit. With N as the number of hits this function is the spatial hit pattern. Next decompose this function into its spherical harmonic components

$$f(\theta,\phi) = \sum_{l,m}\alpha_{lm}Y^*_{lm}(\theta,\phi) \tag{2}$$

The α_{lm} can be found in the usual way

$$\begin{aligned}\alpha_{lm} &= \int f(\theta,\phi)Y_{lm}(\theta,\phi)\,d\Omega \\ &= \tfrac{1}{N}\textstyle\sum_{i=1}^{N} Y_{lm}(\theta_i,\phi_i)\end{aligned}$$

The next step is to define a set of rotationally invariant quantities β_l

$$\begin{aligned}\beta_l &= \textstyle\sum_m|\alpha_{lm}|^2 \\ &= \tfrac{1}{N^2}\textstyle\sum_{i,j}\sum_m Y_{lm}(\theta_i,\phi_i)Y^*_{lm}(\theta_j,\phi_j) \\ &= \tfrac{2l+1}{4\pi N^2}\textstyle\sum_{i,j}P_l(\cos\omega_{ij})\end{aligned}$$

where ω_{ij} is the angle between the hits labelled i and j. This last line makes explicit the fact that the β_l quantities are rotationally invariant. However, the double summation includes the combination of each hit with itself, and writing this part explicitly yields

$$\beta_l = \frac{2l+1}{4\pi N^2}\left[N + 2\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}P_l(\cos\omega_{ij})\right] \tag{3}$$

It is clear from this last equation that the β_l quantities are explicitly dependent upon the number of hits. This dependence can simply be subtracted away to give

$$\begin{aligned}\beta'_l &= \tfrac{N}{N-1}\left(\beta_l - \tfrac{2l+1}{4\pi N}\right) \\ &= \tfrac{2l+1}{2\pi N(N-1)}\textstyle\sum_{i=1}^{N-1}\sum_{j=i+1}^{N}P_l(\cos\omega_{ij})\end{aligned}$$

These β'_l quantities are the harmonic parameters which encode a spatial hit pattern in a rotationally invariant way.

3.2 Temporal Hit Distribution — The Residual Parameters

First define the residual R_i for the i^{th} hit

$$R_i = |\mathbf{r}_{\text{fit}} - \mathbf{r}_i| - v(t_{\text{fit}} - t_i) \tag{4}$$

where $\mathbf{r}_{\text{fit}}$ and t_{fit} are the fitted event position and time, $\mathbf{r}_i$ and t_i are position and time of the i^{th} hit, and v is the speed of light in water. This residual quantity is the difference between the vertex to hit distance as measured by the vertex and hit positions and the vertex to hit distance as measured by the vertex and hit times.

The mean of the residual distribution of an event can be defined in the usual way.

$$\overline{R} = \frac{1}{N}\sum_{i=1}^{N} R_i \tag{5}$$

as can the central moments of the distribution

$$\mu_l = \frac{1}{N-1}\sum_{i=1}^{N}\left[R_i - \overline{R}\right]^l \tag{6}$$

where l is integer.

With μ_0 defined as $\overline{R}$, the μ_l are the residual parameters encoding the timing information of the hit pattern.

3.3 Spatio-Temporal Hit Distribution — The Residual Harmonic Parameters

By incorporating the residual for each hit within the spatial hit pattern function defined earlier the full hit pattern function can be defined:-

$$F(\theta,\phi) = \frac{1}{N}\sum_{i=1}^{N} R_i\,\delta(\cos\theta - \cos\theta_i)\,\delta(\phi - \phi_i) \tag{7}$$

This particular form for incorporating timing information is somewhat arbitrary and job could be done in a number of other ways. The form chosen here is merely the simplest. Just as with the harmonic parameters decompose this function into spherical harmonic components.

$$F(\theta,\phi) = \sum_{l,m} A_{lm} Y_{lm}^{*}(\theta,\phi) \tag{8}$$

and define the rotationally invariant quantity γ_l

$$\gamma_l = \sum_{m} |A_{lm}|^2 \tag{9}$$

Manipulate and remove the NHIT dependence as before to give

$$\gamma_l' = \frac{2l+1}{2\pi N(N-1)}\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} R_i R_j P_l(\cos\omega_{ij}) \tag{10}$$

These γ_l' quantities are the residual harmonic parameters which encode the hit pattern along with its timing information in a rotationally invariant way.

4 Analysis Scenarios

4.1 Relevant Event Classes

Having selected the parameters to extract from each hit pattern, the next task is to decide which event classes should be considered in the analysis. Monte Carlo studies [SR94] indicate that with an NHIT threshold at 60 all backgrounds (except those producing neutrons) are negligible. The analysis of this paper is conducted throughout with this 60 hit threshold and means that only 4 classes of event need to be considered. These are precisely defined to be:-

(CC) Charge Current $\nu_e + {}^2\mathrm{H} \rightarrow p + p + e^-$

(ES) Electron Scattering $\nu_x + e^- \rightarrow \nu_x' + e^{-'}$

(NC) Chlorine Neutral Current $n + {}^{35}\mathrm{Cl} \rightarrow {}^{36}\mathrm{Cl} + \gamma\mathrm{s}$

(ND) Deuteron Neutral Current $n + {}^2\mathrm{H} \rightarrow {}^3\mathrm{H} + \gamma$

The approximate number of events per year which satisfy the NHIT cut and reconstruct within a radius of 7m can be obtained using SNOMAN and is shown in Table 1. Standard Solar Model and best fit MSW rates are given for both a pure D_2O and a chlorine fill. These numbers do not include background neutrons and are the rates used throughout this paper.

		CC	ES	NC	ND
1 Year Cl	SSM	8832	1208	2585	62
1 Year Cl	MSW	3397	536	2585	62
1 Year D_2O	SSM	8832	1208	0	634
1 Year D_2O	MSW	3397	536	0	634

Table 1: The number of events with NHIT > 60 which reconstruct within 7m

4.2 Detector Run Scenarios

To optimise the network mapping, labelled training sets of data are needed. It is clearly preferable that these should come from the detector as opposed to Monte Carlo, and for the ND and NC event classes this can easily be achieved with a neutron source during pure D_2O and chlorine runs respectively. For the CC and ES classes the situation is a little more tricky. It is probably the case that a β decay source will not be an adequate mimic of these classes because of the light scattering effects of the source itself and the sensitivity of the network algorithm to such differences (this sensitivity is, after all, the origin of the algorithm's power). It is, however, possible to obtain a pure data set of mixed CC and ES events by adding a neutron poison to the D_2O. This would enable training of the network on data entirely derived from the detector itself, with all the credibility and avoidance of systematics that that entails. It is hoped that this paper will be seen as another good argument for using a neutron poison. However, powerful reasons exist for not adding a poison and for this reason CC and ES training sets may have to come from Monte Carlo. It should also be noted that whilst the NC and ND events may show sufficient differences to be separated by a neural algorithm there appears to be no point in attempting this as both classes indicate the same thing - the presence of a free neutron in the detector.

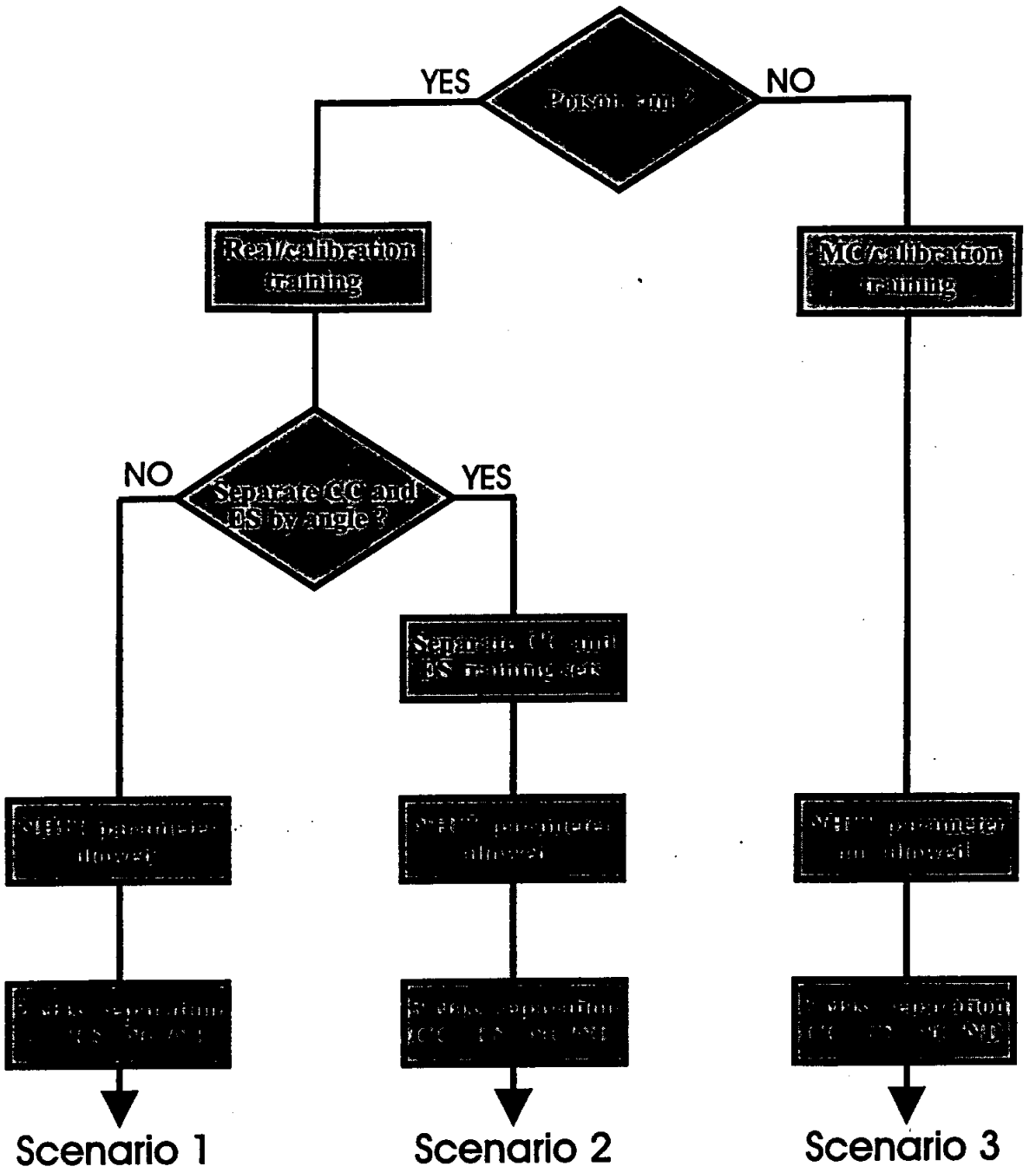


Figure 2: A flow chart defining the three operating scenarios assumed in the analysis of this paper

With the preceeding paragraph in mind the analysis has been conducted under the three different scenarios defined in the flow chart of Fig. 2.

- **Scenario 1** With a poison run and a neutron source calibration then a training set of mixed CC/ES and a set of mixed NC/ND can be obtained. A network can then be trained to separate these two classes and NHIT information can be used since real CC/ES data is available.

- **Scenario 2** Since the angular cross-section for ES events is so sharply forward peaked it should be possible to take the mixed CC/ES data set from a poison run and separate it on an event by event basis into a CC and an ES data set by simply using the angle between fitted direction and the sun. A network can now be trained to distinguish between the three classes CC, ES, NC/ND with NHIT information also allowed.

- **Scenario 3** If a poison run is not carried out then CC and ES training sets must be obtained from Monte Carlo meaning that NHIT information should not be used as the true spectrum is uncertain. A network is trained in this case to distinguish the three classes CC, ES, NC/ND.

Scenario 1 is clearly the most preferable as it uses data derived entirely from the real detector and for the same reason Scenario 3 is to be least favoured. Scenario 2 has been included for completeness only and is seriously flawed on two counts: firstly it is not clear that a good enough direction fit can be made to reliably separate CC and ES on a event by event basis, and secondly there will probably be insufficient ES events to properly train a network to pick out this class alone.

5 Training and Fitting

5.1 The Training Sets

When optimising the network function, training data sets are used in the analysis which contain 5000 events for each class. So a network for two class separation is trained with 10000 events and one for three classes uses 15000 events. Such numbers of events are easily obtained from Monte Carlo or calibration, but may not be possible with a year of poison data. The figure of 5000 is chosen so as to completely avoid overtraining [Bri95a], but this number could be taken as low as 2000 and still produce the same network mapping function. Consequently the use of more CC/ES training events in this paper than will probably be available from a poison run is not deemed to be a problem.

5.2 Handling 3 Class Separation

When two classes of events are separated by a network only one output is needed and events from each class tend to cluster near 0 or 1 in the output space as indicated in Fig. 1. To separate three classes of events it is conventional (though not strictly necessary) to use a network with three outputs. The trained network then tends to cluster events from a particular class in such a way that a particular individual output tends to be high, i.e. each individual output deals with a particular class. If an event is deemed by the network to belong to a certain class then a certain output will be high, otherwise it will be low. With such a network configuration it can be shown using Bayes theorem [Bri95a] that the actual numerical value of an output is the posterior probability of the event belonging to that class.

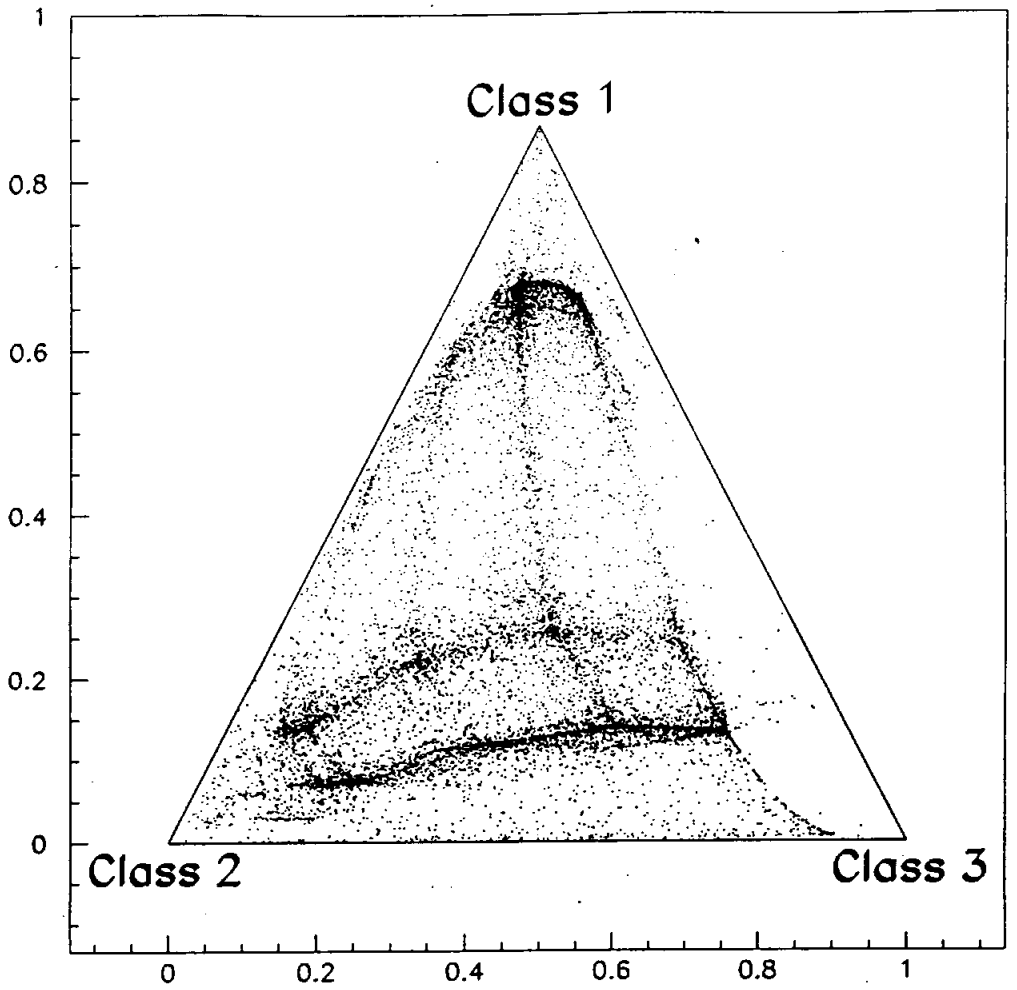


Figure 3: A scatter plot of the output of a network designed to distinguish between three classes. Note the clustering of events near each corner.

Normalisation of the probability then means that for any event fed into the network the sum of the output values must equal unity. Individual output values are also constrained to lie between 0 and 1. For a network with three outputs labelled X, Y, and Z these constraints can be written

$$X + Y + Z = 1$$

$$0 < X, Y, Z < 1$$

In the three dimensional output space of a three output network events are therefore constrained to lie on the plane $X + Y + Z = 1$ in the positive octant i.e. within a two dimensional equilateral triangle. Each of the three classes will tend to cluster near its own

corner of this triangle. By taking 3D output vectors (X,Y,Z) and applying the following transformation

$$\frac{1}{X+Y+Z}\begin{pmatrix} 1 & \frac{1}{2} & 0 \\ 0 & \frac{\sqrt{3}}{2} & 0 \end{pmatrix}\begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

then a 2D vector is produced that lies within the equilateral triangle as show in Fig. 3. This scatter plot also shows the clustering of each class near its own corner. Fig. 3 is the three class analogue of the two class PDFs of Fig. 1. There are now three 2D PDFs and exactly the same fitting procedures can be applied to extract the fraction of events belonging to each of three classes in a mixed real data set.

5.3 Fitting Procedures

The procedure for extracting the fraction of events belonging to each class in a set of real data simply consists of fitting the real data network output histogram to a linear combination of the PDF histograms. This can be done most easily with a χ^2 fit. With the following assignments:-

N	=	Number of real data events
G_i	=	Network output histogram of real data (unnormalised)
g_i	=	Network output histogram of real data (normalised to unity)
N^k	=	Number of PDF events belonging to class k
F_i^k	=	PDF histogram for class k (unnormalised)
f_i^k	=	PDF histogram for class k (normalised to unity)
A^k	=	Number of events in the real data set belonging to class k
α^k	=	Fraction of events in the real data set belonging to class k

then the χ^2 function can be written as

$$\chi^2 = \sum_i \left[\frac{G_i - \sum_k A^k f_i^k}{\sigma_i}\right]^2$$

where

$$\begin{aligned} \sigma_i^2 &= \sigma_{G_i}^2 + \textstyle\sum_k (A^k \sigma_{f_i^k})^2 \\ &= G_i + \textstyle\sum_k \frac{A^k}{N^k} A^k f_i^k \\ &= N\left[g_i + \textstyle\sum_k (\frac{N}{N^k}\alpha^k)\alpha^k f_i^k\right] \end{aligned}$$

Entirely using normalised histograms the χ^2 can therefore be written as

$$\chi^2 = N \sum_i \frac{\left[g_i - \sum_k \alpha^k f_i^k\right]^2}{\left[g_i + \sum_k (\frac{N}{N^k}\alpha^k)\alpha^k f_i^k\right]}$$

This function is then minimised with the constraint that the α^k sum to unity to find the fraction of events in each class. Statistical uncertainties can be obtained from the error

matrix in the usual way. It should be noted here that the fitting procedure outlined in Sec. 3.2 of [Bri95a] is wrong in a number of places and the χ^2 described above is the one actually used in this paper.

In addition to using χ^2 each fit has also been carried out using a maximum likelihood technique specifically designed for fits of one histogram to a linear combination of others where there are statistical uncertainties on all histograms [ELL93]. As well as allowing the α^k to vary this technique regards the f_i^k as estimators for the true PDF histograms and allows their values to be tweaked according to Poisson statistics. For a full description of the technique see [ELL93]. In all cases both fitting methods gave the same answers to within 1σ uncertainties.

6 Results

Throughout the analysis the data has been produced using SNOMAN 2_08 with the vertex fit being made with the Grid Fitter [Moo95] and the direction fit using the Elastic Fitter [Bri95b]. The extracted parameters used are:-

	9	Harmonic parameters
	9	Residual parameters
	9	Residual harmonic parameters
	1	Solar angle parameter
and, where allowed,	1	NHIT parameter

This amounts to 28 or 29 parameters. The network architectures have 20 hidden nodes and either 1 or 3 output nodes depending upon the number of classes that are separated. The three different analysis scenarios therefore use network architectures defined by Table 2.

	No. input nodes	No. hidden nodes	No. output nodes
Scenario 1	29	20	1
Scenario 2	29	20	3
Scenario 3	28	20	3

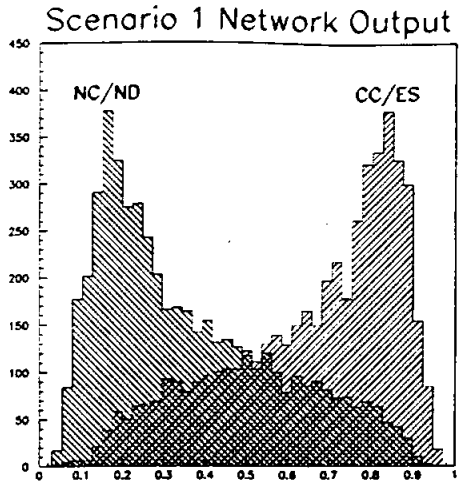
Table 2: Network architectures for the three analysis scenarios

The neural network analysis is conducted for two detector configurations (Cl and pure D_2O fill) in each case all three analysis scenarios are investigated (Scenarios 1, 2, and 3 as defined by Fig. 2) for both types of neutrino flux assumptions (SSM and MSW as defined by Table 1). This gives a total of 12 fits.

6.1 1 Year Cl Fill

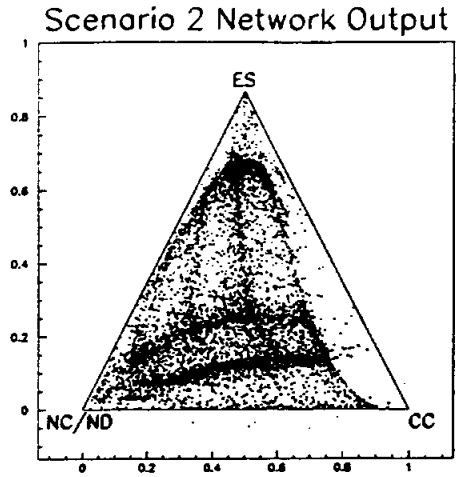
The results of the fits to 1 year of Cl fill data and the corresponding PDF histograms and scatter plots are shown in Fig. 4.

It should be noted that with the fit class fractions constrained to sum to unity the uncertainties for each class are NOT independent. In all the results presented the ES and NC/ND uncertainties are independent (subject to off diagonal error matrix elements) whereas the CC or CC/ES uncertainty quoted is the RMS of the ES and NC/ND errors.



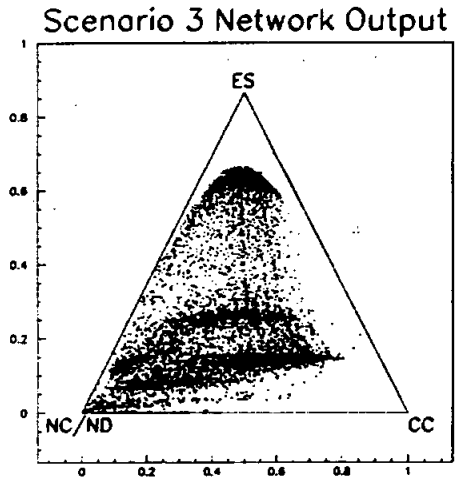
Scenario 1 Fit

	CC/ES	NC/ND
SSM fit	10119(116)	2568(116)
SSM true	10040	2647
MSW fit	3963(90)	2617(90)
MSW true	3933	2647



Scenario 2 Fit

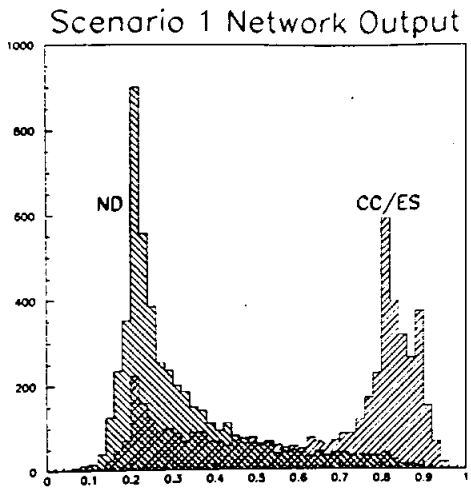
	CC	ES	NC/ND
SSM fit	8871(132)	1220(69)	2596(113)
SSM true	8832	1208	2647
MSW fit	3415(99)	525(47)	2640(87)
MSW true	3397	536	2647



Scenario 3 Fit

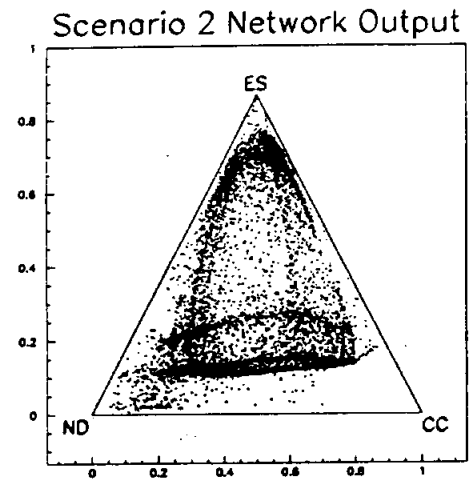
	CC	ES	NC/ND
SSM fit	8837(143)	1236(70)	2614(125)
SSM true	8832	1208	2647
MSW fit	3437(108)	549(48)	2594(97)
MSW true	3397	536	2647

Figure 4: The fit results and PDF histograms and scatter plots for each of the three analysis scenarios applied to Cl fill data



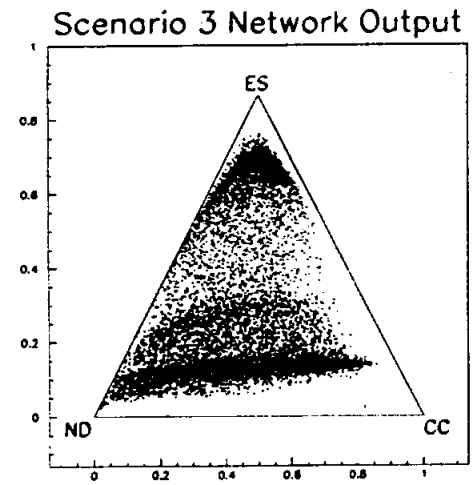
Scenario 1 Fit

	CC/ES	ND
SSM fit	10157(105)	517(105)
SSM true	10040	634
MSW fit	3838(73)	729(73)
MSW true	3933	634



Scenario 2 Fit

	CC	ES	ND
SSM fit	8893(124)	1220(67)	561(104)
SSM true	8832	1208	634
MSW fit	3334(85)	541(46)	692(72)
MSW true	3397	536	634



Scenario 3 Fit

	CC	ES	ND
SSM fit	8818(141)	1211(67)	645(125)
SSM true	8832	1208	634
MSW fit	3306(96)	536(45)	725(85)
MSW true	3397	536	634

Figure 5: The fit results and PDF histograms and scatter plots for each of the three analysis scenarios applied to pure D_2O data

The results show that, given a year of Cl fill data, the CC rate can be determined to ~2% statistical accuracy, the ES rate to ~7%, and the NC/ND rate to ~3%. It is interesting to note the increased uncertainties when going from scenario 2 to scenario 3, which differ only in that NHIT is allowed in 2, but not in 3. This indication that NHIT is an important separating parameter is born out by a comparison of the PDF scatter plots for scenarios 2 and 3. The separation in the plot for scenario 2 is clearly superior to the rather compressed distribution of points in the scenario 3 plot. That being said, the statistical uncertainties of scenarios 2 and 3 are pretty close and show that NHIT, although important, is not crucial. The clear structure in the scatter plots is discussed in Sec. 7.

6.2 1 Year Pure D_2O Fill

The results of the fits to 1 year of pure D_2O data and the corresponding PDF histograms and scatter plots are shown in Fig. 5.

As with the results from the Cl fill, with a year of pure D_2O data, the CC and ES rates can be determined to ~2% and ~7% respectively. The smaller number of ND events means, however, that the uncertainty in their rate increases to ~15%.

On the face of it the D_2O results are surprisingly good, comparing very favourably with the Cl results. This surprise is based on the assumption that what enables CC and NC events to be separated is the difference between an event that produces a single particle and one that produces more than one, be they electrons or gammas. One would therefore expect the separation between CC and ND, events that both produce only one particle, to be much worse. A closer look at the gamma cascade of the NC event class demonstrates the error in this argument. The Cl gamma cascade has a total energy of 8.6 MeV, but this is released principally in the form of a single gamma of about 6 MeV and a couple of low energy gammas. These low energy gammas will have a low efficiency for producing Čerenkov light. Now the Cl gamma cascade starts to look more like the single 6.25 MeV gamma that comes from the deuteron neutron capture of the ND events and the similar performance given a year of CL or a year of pure D_2O data is perhaps something that should be expected. This leads on to the conclusion that what enables the CC separation from NC is the multiple Compton scattering of the major gamma of the Cl cascade rather than the production of more than one gamma. This hypothesis is currently being investigated.

7 Discussion

7.1 Picking Apart the Network Mapping

The most distinct feature of the scatter plots of Figs 4 and 5 is the amount of detailed structure within them. Shown in Fig. 6 are a smoothed histogram version of the scenario 2 scatter plot of Fig. 4, the scatter plot itself, and a simple diagram picking out the line structures within the plot. The histogram is composed of seven ridges with peaks occurring where the ridges intersect. By isolating the events that form each ridge and looking at their extracted parameters (i.e. their network inputs) it is possible to roughly identify the ridges with particular properties of hit patterns. These are outlined below:-

- **Ridge 1:** Events pointing directly away from the sun.
- **Ridge 2:** Events pointing in the hemisphere away from the sun.
- **Ridge 3:** Events pointing in the hemisphere towards the sun.

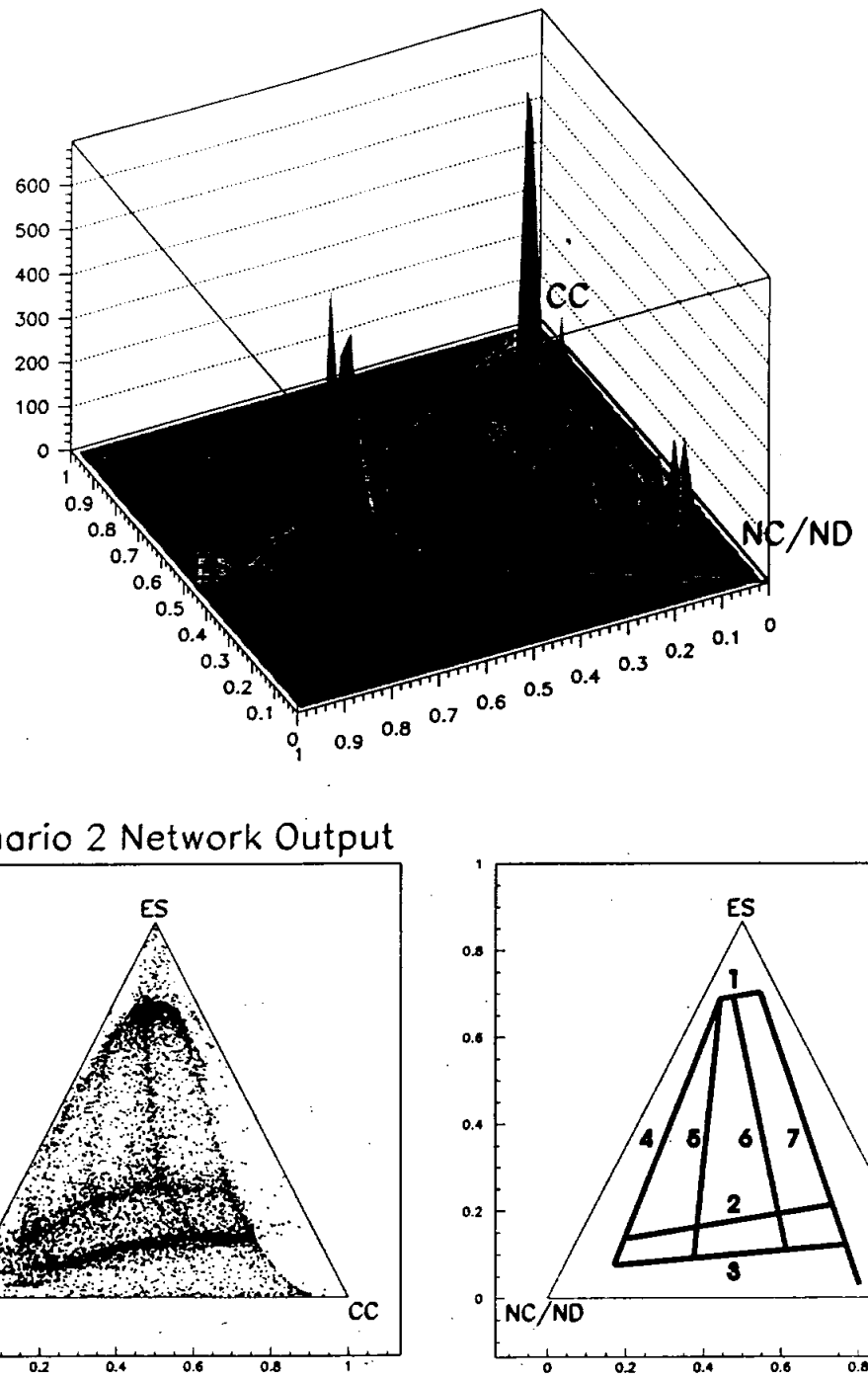


Figure 6: On top is a combined and smoothed histogram of the 3 PDFs of scenario 2 applied to Cl fill data. Below and on the left is a reproduction of the corresponding scatter plot with the diagram on the right picking out the key features.

- **Ridge 4:** Events which are abnormally isotropic.
- **Ridge 5:** No clear characteristics besides the events being somewhat more isotropic.
- **Ridge 6:** No clear characteristics besides NHIT being quite high.
- **Ridge 7:** Events with a large number of hits.

These gross properties of the ridges are exactly what would be expected. The ES events are picked out entirely on the basis of their sharp forward peaking in a direction away from the sun, with a weaker distinction between forward and backward hemispheres for the CC and NC/ND classes. Differences between isotropy and NHIT then serve to distinguish between CC and NC/ND events.

The preceeding paragraph begs the question:- If such simple characteristics are all the network is using then why not avoid the neural network technique entirely and construct some cuts based on solar angle, NHIT and isotropy? This could certainly be done and may well prove sucessful, but there seems to be little point. The network technique is relatively simple and straightforward and, as Ridges 5 and 6 above demonstrate, does use more than the three simplest parameters. Such cuts, once optimised, may be just as mathematically complex as the function carried out by the network and there would still be the necessity of obtaining pure data sets from the detector or Monte Carlo.

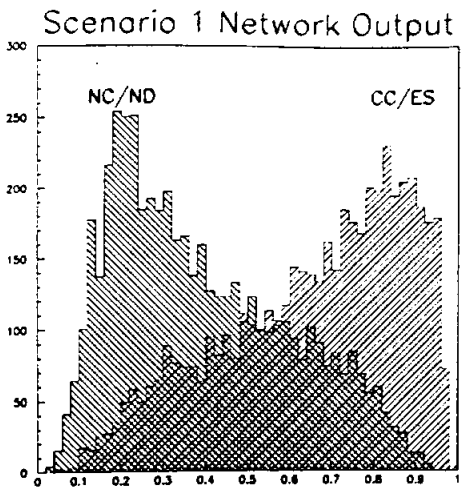
7.2 High but Known Scattering

The analyses of Figs. 4 and 5 are carried out with the known or best guess values for the light scattering and absorption of the various detector components included in the Monte Carlo. These may or may not prove to be accurate and so a check has been carried out where the analysis of Fig. 4 is repeated but with the amount of Rayleigh scattering in the Monte Carlo increased by an order of magnitude for all the training, fitting and 'real' data sets. Rayleigh scattering is chosen not because it is uncertain to an order of magnitude, but because it is a simple mimic of other forms of scattering. Fig. 7 shows the results of this high scattering analysis.

It is clear from each of the three plots in Fig. 7 that the degree of separation is poorer than that of their counterparts in Fig. 4. This is only to be expected as the more scattering there is in the detector the more isotropic the hit patterns of every event class become. That being said the separation is still pretty good and this is born out by the fit results which are accurate and have uncertainties only slightly higher than those of Fig. 4. This is therefore an impressive demonstration of the fact that the network technique should work in any reasonable detector conditions so long as those conditions are known.

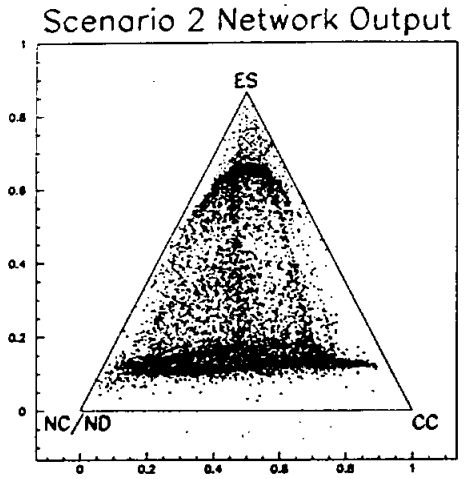
8 Systematics

The neural technique has shown good statistical uncertainties and a resilience to detector conditions, but nothing has yet been said about its susceptibility to systematic errors. Such errors can be defined as systematic differences between the data sets used to form the PDF histograms and the data set to which these PDFs are fitted. It is here that the true benefit of using CC and ES data from a poison run rather than Monte Carlo comes in. With data from a poison run and NC/ND data from calibration the only systematics that can occur will be due to the detector conditions changing between when the PDF data was taken and when the real data run occured. Although this can happen the leeway for temporal changes should be smaller than that for differences between Monte Carlo and reality.



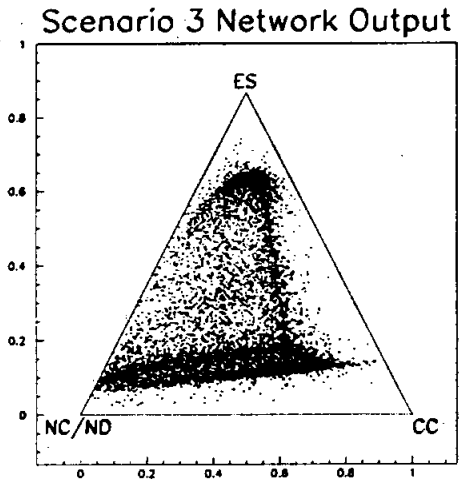
Scenario 1 Fit

	CC/ES	NC/ND
SSM fit	10062(134)	2625(134)
SSM true	10040	2647
MSW fit	3791(98)	2789(98)
MSW true	3933	2647



Scenario 2 Fit

	CC	ES	NC/ND
SSM fit	8782(153)	1223(71)	2682(136)
SSM true	8832	1208	2647
MSW fit	3313(112)	530(49)	2737(100)
MSW true	3397	536	2647



Scenario 3 Fit

	CC	ES	NC/ND
SSM fit	8865(166)	1213(69)	2609(151)
SSM true	8832	1208	2647
MSW fit	3330(125)	556(49)	2694(115)
MSW true	3397	536	2647

Figure 7: A repeat of the analysis of Fig. 4, but with Rayleigh scattering increased by an order of magnitude for all training, fitting, and 'real' data sets.

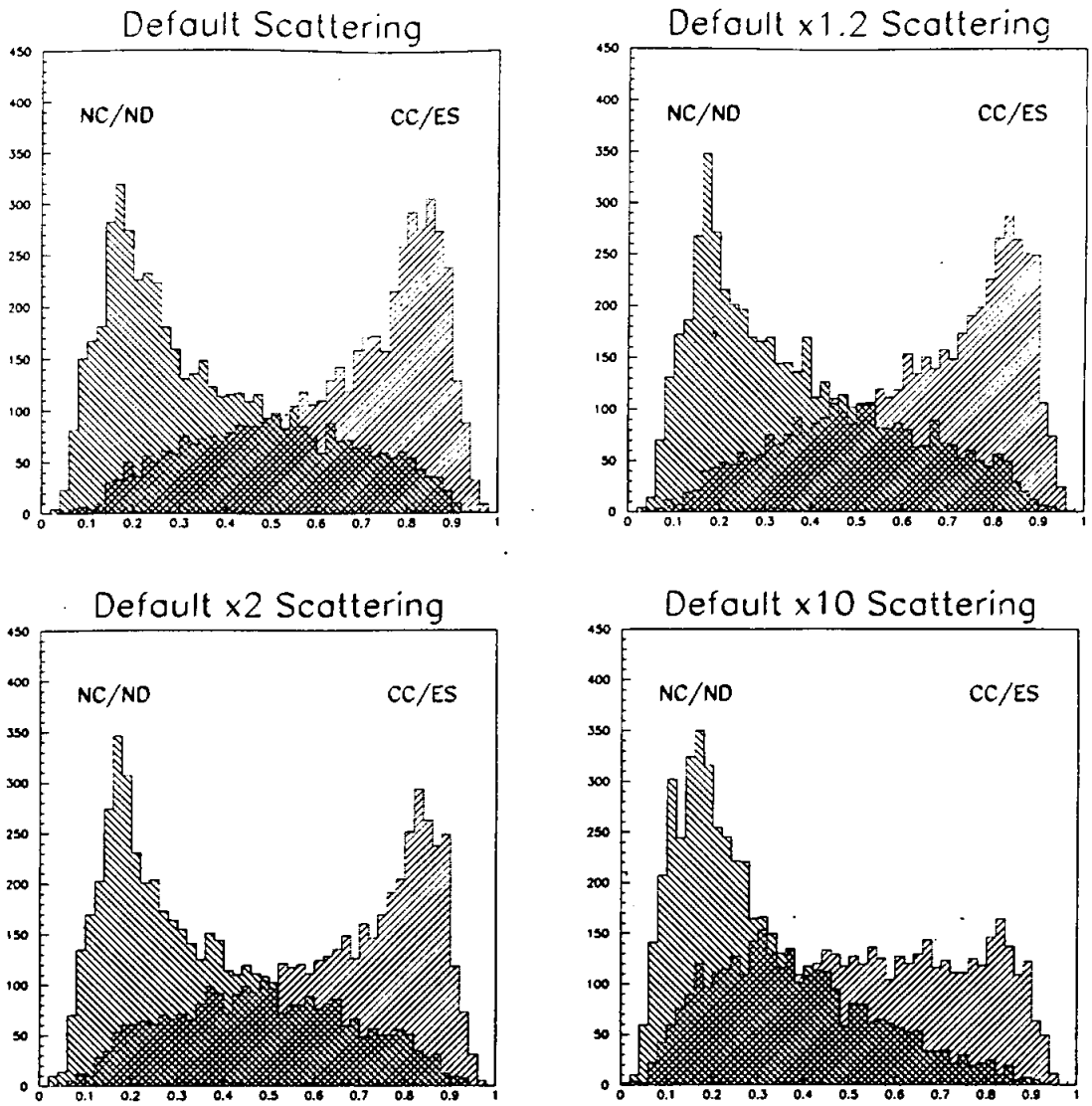


Figure 8: Four plots showing the network output of data sets of increasing levels of scattering for a network trained on the default scattering level.

As a first pass at assessing the level of systematic error a network is trained and PDF histograms formed using data produced with the default level of scattering. However, these PDFs are then fitted to 'real' data sets produced with increased Rayleigh scattering. The discrepancy between true and fitted numbers then gives the systematic error. Increases in Rayleigh scattering are chosen as a mimic for more general scattering increases. Fig. 8 shows histograms of the 'real' data using scenario 1 at four different levels of scattering. It has been fed through the same network trained on the default scattering level. The degradation with increased scattering is entirely predictable as more scattering produces more isotropic hit patterns. Thus both event classes NC/ND and CC/ES get pulled to the left, with NC/ND, the class that shows more isotropy being little affected, but CC/ES being gradually smeared flat. The results of the fits are shown in Table 3. Default scattering PDFs are fitted to 'real' data with 1.2 and 2 times the default scattering level. The systematic errors made are ~2%

Scenario 1 Fit

	CC/ES	NC/ND
×1.2 scattering SSM fit	10042(119)	2645(119)
×2.0 scattering SSM fit	9583(124)	3104(124)
SSM true	10040	2647
×1.2 scattering MSW fit	3840(91)	2740(91)
×2.0 scattering MSW fit	3544(92)	3036(92)
MSW true	3933	2647

Scenario 2 Fit

	CC	ES	NC/ND
×1.2 scattering SSM fit	8754(134)	1204(69)	2729(115)
×2.0 scattering SSM fit	8401(138)	1151(68)	3135(120)
SSM true	8832	1208	2647
×1.2 scattering MSW fit	3295(98)	491(46)	2794(87)
×2.0 scattering MSW fit	3031(100)	482(45)	3067(189)
MSW true	3397	536	2647

Scenario 3 Fit

	CC	ES	NC/ND
×1.2 scattering SSM fit	8834(144)	1209(70)	2644(126)
×2.0 scattering SSM fit	8105(153)	1125(68)	3457(137)
SSM true	8832	1208	2647
×1.2 scattering MSW fit	3380(108)	482(46)	2718(98)
×2.0 scattering MSW fit	2917(111)	466(45)	3197(101)
MSW true	3397	536	2647

Table 3: Fits of PDF histograms formed with the default level of scattering to 'real' data with increased scattering.

and ~5% respectively for CC events, ~6% and ~7% for ES events, and ~4% and ~15% for NC/ND events. Two features should be noted: The NC/ND numbers are consistently high and the CC numbers consistently low. This follows naturally from the slewing observed in Fig. 8. Secondly the ES numbers are only slightly affected by the systematic (the high percentage error arises from the low number of ES events). This is also no surprise as the ES separation is almost entirely based upon the fitted solar angle and the levels of scattering are not yet high enough to have a serious effect on the quality of direction fit. It is claimed that the overall level of scattering in the detector will be known to within 10% with a lightball calibration. A conservative estimate of the systematic error due to detector scattering would therefore be about 2% and it is likely that this will form the largest contribution to the systematic error budget.

9 Further Applications

The success of the neural technique in separating CC, ES, and NC/ND events has led to thoughts of applying it to other hit pattern identification problems in SNO. Work is currently proceeding on the separation of hit patterns from ^{208}Tl and ^{214}Bi $\beta\gamma$ events that fall within a window from 30 to 40 hits [Che95]. The idea, suggested by Martin Moorhead [Moo95], is that the Tl events can easily get above 30 hits with a β that doesn't necessarily have to be close to its 4.99 MeV endpoint, a few low energy γs and the 2.6 MeV γ. In order for the Bi event to get above 30 hits the β has to be pretty close to its 3.27 MeV endpoint and there can't be any accompanying γs. Consequently these two types of events produce Čerenkov light from more than one electron and from a single electron respectively and so ought to show characteristic differences in their hit patterns. This has proven to be the case and reasonable separation has been achieved using the same neural technique as outlined in this paper.

The more refined the SNO position/direction fitting algorithms and hit pattern recognition techniques get, the more entwined the two tasks become. To improve position fits beyond those of the Time and Quad fitters it is necessary to bring in information about the type of event being fitted. This is the case with fitters that use some kind of PDF as the Grid Fitter and, to a lesser extent, the Elastic Fitter do. On the other hand any hit pattern recognition technique will greatly benefit from position and direction fit information. In the analysis of this paper such information enables individual PMT hits to be specified by three numbers (θ, ϕ, and t) rather than four (x, y, z, t). This reduction in the number of degrees of freedom produces a better separation of event classes. It seems that one is being powerfully drawn towards a technique that will both fit an event and perform hit pattern recognition as one 'organic' process. Such a technique may not be realisable, but certainly warrants investigation. The best place to start is with the receptive fields neural network technique [HKP91] which has been used by the U.S. Postal Service to develop a neural network for zip code recognition on envelopes [C$^+$89b, C$^+$89a, KRL90] a problem whose translational invariance is analogous to the rotational invariance of the SNO detector. A receptive fields technique has also been used on SNO data [DC93], but applied to the somewhat less ambitious task of distinguishing Čerenkov hit patterns from random coincidences.

10 Conclusion

It has been shown that neural network hit pattern recognition can separate the electron scattering, charged and neutral current rates with a statistical accuracy of between 2 and 7% with one year of salt fill data and has similar, but slightly worse performance, with one year of pure D_2O data. A check on what is likely to be the worst systematic has yielded a conservative uncertainty of 2% and the robustness of the technique to varying but known detector conditions has been demonstrated.

References

[Bri95a] S.J. Brice. An Overview of the Feedforward Neural Network Technique and its Application to SNO Event Classification, 1995. SNO-STR-95-037.

[Bri95b] S.J. Brice. The Elastic Fitter, 1995. SNO-STR-95-040.

[C$^+$89a] Y. Le Cun et al. Backpropagation Applied to Hand-Written Zip Code Recognition. *Neural Computation*, 1:541–551, 1989.

[C+89b] Y. Le Cun et al. Handwritten Digit Recognition with a Back-Propagation Network. In D.S. Touretzky, editor, *Advances in Neural Information Processing Systems II*, pages 396–404, 1989.

[Che95] X. Chen. A Monte Carlo Study of $\beta\gamma$ Backgrounds from ^{208}Tl and ^{214}Bi, 1995.

[DC93] D.W. Dong and Y. Chan. Neural network for recognizing Ĉerenkov radiation patterns, 1993.

[ELL93] P. Eberhard, G. Lynch, and D. Lambert. Fits of Monte Carlo Distributions to Data. *Nuclear Instruments and Methods in Physics Research A*, 326:574–580, 1993.

[HKP91] John Hertz, Anders Krogh, and Richard G. Palmer. *Introduction to the Theory of Neural Computation.* Addison Wesley, 1991.

[KRL90] J.D. Keeler, D.E. Rumelhart, and W. Leow. Integrated Segmentation and Recognition of Hand-Printed Numerals. In *Advances in Neural Information Processing Systems III*, pages 557–563, 1990.

[Moo95] M. Moorhead. Grid Fitter for Reducing Tails in Spatial Distributions, 1995.

[SR94] P. Skensved and B.C. Robertson. Summary of Backgrounds in SNO, 1994. SNO-STR-94-013.