SAND99-0150C

Real-Time Analysis Of Individual Airborne Microparticles Using Laser Ablation Mass Spectroscopy And Genetically Trained Neural Networks

Eric P. Parker, Stephen E. Rosenthal, Michael W. Trahan, and John S. Wagner

Sandia National Laboratory

Department 9512: Pulsed Power and Laser Initiatives

William B. Whitten, Rainer A. Gieray, Peter T. A. Reilly, Alexandru Lazar, and J. Michael Ramsey

Oak Ridge National Laboratory

Laser Spectroscopy and Microinstrumentation Group

DISCLAIMER

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

Abstract

We are developing a method for analysis of airborne microparticles based on laser ablation of individual molecules in an ion trap mass spectrometer. Airborne particles enter the spectrometer through a differentially-pumped inlet, are detected by light scattered from two CW laser beams, and sampled by a pulsed excimer laser as they pass through the center of the ion trap electrodes. After the laser pulse, the stored ions are separated by conventional ion trap methods.

The mass spectra are then analyzed using genetically-trained neural networks (NNs). A number of mass spectra are averaged to obtain training cases which contain a recognizable spectral signature. Averaged spectra for a bacteria and a non-bacteria are shown to the NNs, the response evaluated, and the weights of the connections between neurodes adjusted by a Genetic Algorithm (GA) such that the output from the NN ranges from 0 for non-bacteria to 1 for bacteria. This process is iterated until the population of the GA converges or satisfies predetermined stopping criteria. Using this type of bipolar training we have obtained generalizing NNs able to distinguish five new bacteria from five new non-bacteria, none of which were used in training the NN.

Introduction

Recently there has been an increase in interest in the detection and identification of biological samples, especially those which are or can be made airborne [1-19]. There are a number of motivating factors which make timely detection and identification of airborne microorganisms desirable, not the least of which is the prospect of hostile third world powers either acquiring or developing weapons of mass destruction based on biological agents. Traditional methods used to identify bioaerosols are time consuming, as they involve collecting samples on culture media and identifying them via colony growth and identifying the way the samples respond to various agents [11]. Given the present world climate it is not surprising that faster methods of identification are being developed. These techniques tend to fall into two categories, the first of which involves using highly specific reactions such as antigen-antibody reactions or the hybridization of gene probes to DNA or RNA. The second category involves the use of conventional methods of chemical analysis. Reviews of conventional analyses [10] further divide analytical methods into bulk and single particle techniques. Single particle techniques are primarily based on mass spectrometry [14-15,17] and optical emission spectrometry [17-19]. Spectroscopic methods rely on optical resolution to detect single cells, while mass spectrometric methods rely on particle beam technology to make measurement times shorter than the time between particle events.

In this paper we describe experiments aimed at distinguishing bacteria from non-bacteria using laser ablation mass spectrometry and genetically trained neural networks. Laser ablation mass spectrometry involves using particle beam techniques to introduce a single particle into a quadrupole ion trap where it is ionized by a pulsed excimer laser beam [20]. The resulting ions can then either be resolved into a mass spectrum or further analyzed by MS/MS techniques [21]. Recent developments in on-line analysis of airborne particles using laser ablation mass spectrometry are reviewed in reference [8].

Intelligent Algorithms

Before proceeding it is useful to describe the components of the intelligent algorithms used in data analysis. Genetic algorithms and neural networks are both based on biological models and have been used successfully in a variety of data analysis applications[22-24]. Neural Networks are based on a simplified model of the mammalian brain [25]. In this model, neurons, which consist of a stoma (base) and numerous synapses (connections to other stoma), are represented by neurodes and their interconnections. Each neurode receives multiple inputs and gives a single output. The individual input signals are each assigned a relative weight and then summed to give the total weighted input. Individual inputs can either excite the neurode to fire or inhibit the neurode from firing. Excitatory inputs are positive, while inhibitory inputs are negative. The output signal from the neurode is computed as the result of a transfer function applied to the weighted sum of the inputs. If the output is above a certain threshold the

neurode fires, sending an input to each neurode in the following layer. The transfer function can take many forms: step functions, ramp functions, sigmoid functions, etc.

The structure of a NN is defined by the way in which neurodes and connections are arranged. The neural networks used in this work are comprised of three layers: an input layer containing a neurode for each piece of information passed into the NN, a middle layer, and an output layer. The middle layer is called a "hidden layer" since it has no direct connections to the outside world. We use a fully-connected feed-forward architecture, meaning that each neurode in the input layer is connected to each neurode in the hidden layer, and each neurode in the hidden layer is connected to each neurode in the output layer. No neurodes are allowed to send signals to neurodes in preceding layers or in the same layer.

Traditionally, neural networks learn to recognize and interpret patterns through a process of supervised training. The connections between neurodes in a NN are initialized with random weights. The NN is then shown many training sets, each consisting of a set of inputs and expected outputs. After a training set is shown to the NN the actual and expected outputs are compared. The difference between the expected and actual output is calculated for each output neurode, and the weights of the connections to that neurode are adjusted to give the correct output. This training method is called back-propagation of errors. In this work we do not use back-propagation to train NNs. Instead we use a genetic algorithm (GA) to train a population of NNs.

Genetic algorithms are global optimization algorithms based on mathematical models of the genetic process [26]. GA's use the principles of natural selection and "survival of the fittest" to drive the evolution of a population of NNs. During an iteration of the GA (a generation) each NN is shown a set of training cases and then ranked according to the accuracy of its responses. Individuals with higher rank are deemed to be more fit than individuals of lower rank, and thus more likely to reproduce.

Each individual NN contains a set of chromosomes describing its features. In this case, the chromosomes describe the neurodes in each layer and the weights of the connections between them. Reproduction occurs through a process of selection and recombination of chromosomes. Selection favors highly fit individuals, while the least fit are rarely chosen and thus tend to die out. Recombination or "crossover " takes information from each of two parents and places the mixture into a member of the next generation. Offspring may or may not have higher fitness values than their parents due to the random nature of the crossover processes. As this process is repeated over many generations better chromosomes tend to dominate. As a result, the overall fitness of the population increases and the most promising areas of the search space are explored.

To ensure that the search space is thoroughly explored the GA must maintain genetic diversity. This happens both through the selection/recombination process and through

mutation. Mutation introduces random changes into the population during reproduction, reducing the chance that the population will converge to a local optima.

Methods

Mass Spectroscopy

The methods used to collect the data have been described in depth previously [20]. In brief, airborne particles pass through a collimating inlet system which consists of a capillary and two skimmers separating differentially-pumped chambers before entering the vacuum chamber that houses the ion trap electrodes. Incoming particles are sensed as they pass through two continuous wavelength laser beams prior to entering the vacuum chamber. A trigger pulse is generated to fire an excimer laser at 308 nm to strike the detected particles as they pass through the center of the ion trap electrodes, after which a mass spectrum of either positive or negative ions is acquired for each particle.

Data were collected in this fashion for two different types of airborne microparticles: bacteria and pollen. Six different examples were studied in each category, making a total of twelve species. The six bacteria studied were *Azotobacter vinelandii*, *Bacillus subtilis*, *Escherichia coli*, *Enterobacter aerogenes*, *Micrococcus lysodeikticus*, and *Bacillus subtilis* spores, while the six pollen species were *Ambrosia trifida*, *Artemisia tridentata*, *Betula alba*, *Dactylis glomerata*, *Juglans nigra*, and *Agrostis alba*. Both positive and negative ion data were collected for each species. The dataset for each sample studied (positive or negative) consists of hundreds of mass spectra, each spectrum being generated from a single microparticle. Only a few (one or two) of each type of sample is used in training the neural networks (NNs).

Due to the highly random manner in which a microparticle fragments inside the mass spectrometer, a number of spectra must be averaged together to get training cases with a recognizable fingerprint. As a result, the data set used to train the NNs consists of multiple averaged spectra (training cases) for each of the selected bacteria and non-bacteria. The number of spectra averaged to yield the training spectra is determined empirically.

Neural Network Training

As mentioned above, the neural networks used in this work consist of three layers: an input layer, one hidden layer, and an output layer. The number of neurodes in the input layer is determined by (M/Z range)/(# points averaged to yield a single input to the NN), where M/Z is the mass/charge ratio. The optimal number of neurodes in the hidden layer

can be determined empirically (see Figure 1), but using a prime number of neurodes greater than the number of neurodes in the input layer generally yields NNs which work well. The output layer consists of a single neurode whose output varies between 0 (for non-bacteria) and 1 (for bacteria).

Rather than using the standard technique of back-propagation of errors to train a single NN, a GA is used to train a population of fully-connected, feed-forward NNs to distinguish between laser ablation mass spectral data from bacteria and non-bacteria. During training, all of the training cases for each of the chosen species are shown to the population of NNs and the response of each NN is evaluated. The most fit individuals -- those with the most accurate response -- have a higher probability of reproducing.

During reproduction the GA adjusts the weights of the connections between neurodes in each member of the population such that the output from that NN is 1 for bacteria and 0 for non-bacteria. This is done by selecting chromosomes (sets of weights for specific connections between neurodes) from each parent and combining them to create a new NN or offspring. Over the course of many generations (hundreds or thousands) the average fitness of the population increases as NNs which do not perform well are bred out of the population. This process is iterated until the population of the GA converges or reaches a predetermined fitness level.

When training NNs using negative ion data we found that studying the range M/Z=80-300 was sufficient. Ten points were binned into each input for a total of 22 neurodes in the input layer. Using 23 neurodes in the hidden layer gave satisfactory results after 10000 generations. Fifty spectra were averaged for each training case, and a logistic sigmoid was chosen as the neurode activation function.

Training NNs using the positive ion data proved to be somewhat more difficult. In this case the best activation function was determined to be a simple ramp function, as the NNs were not able to generalize when trained using the logistic sigmoid. It was also necessary to study M/Z = 40-370 and average 75 spectra per training case. Five points were binned into each input, and the range M/Z = 56-57 was set to 0. The input layer contained 66 neurodes, while the hidden layer contained 41 neurodes.

Results and Discussion

After the population of NNs had been trained, the sets of averaged spectra for the species not used in training were submitted to the NNs and their response evaluated. In order to be considered successful the average response for a bacterial sample had to be greater than 0.7 (the ideal is 1.0) and the average response for a non-bacterial sample had to be less than 0.3 (the ideal is 0.0). We have used this type of bipolar training on both positive and negative ion data for each species studied.

When working with negative ion data we have obtained NNs which, having been trained on a single bacterial sample and a single pollen sample, are able to distinguish five new bacteria from five new non-bacteria (pollens), none of which were used in training the NN (see Figure 2). Working with positive ion data gives equally good results; yielding NNs which can distinguish between all of the bacteria and all of the non-bacteria. The numerical results plotted in Figure 2 are given in Table I for NN training using negative ion data and in Table II for NN training using positive ion data. The species marked with an asterisk in tables I&II are the species used to train the NNs.

The fact that NNs can generalize is notable for several reasons. First, it shows that the NNs are not merely memorizing the patterns of the species used in training and thus learning to recognize only them. Second, it means that there are enough common features in the mass spectral data that different types of species can be identified. Third, once learned, recall of information is highly resistant to noise. Finally, being able to recognize bacteria which were not included in the training set decreases the likelihood that a new or mutated biological agent could slip by undetected.

This technique has multiple possible uses other than acting as a sentry against biological attack. It could be used to monitor air in hospitals to detect/track the spread of disease. It could be used to detect/identify bacteria in indoor ventilation systems – looking for "sick" buildings. It could be used in the food industry to detect possible airborne contaminants in processing plants. It can also be used to distinguish between bacteria and other potentially dangerous airborne contaminants.

Future Work

The next step in this project is to expand the types of airborne particles studied and to include the new particulate data in the NN training. This will demonstrate whether NNs can differentiate between multiple types of microparticles, or whether it will be necessary to train a series of NNs to recognize different types of particles. This effort is currently underway.

Another obvious next step is to try evaluating individual mass spectra for both bacteria and non-bacteria to see if the NNs can recognize them as either bacteria or non-bacteria. If so, the method can be used to identify bacteria on a real-time basis. This also means that the method could serve as a trigger for other, more time-consuming traditional methods of identification to determine exactly what bacteria are being detected.

Finally, the feasibility of building another arm on the mass spectrometer is being explored. This would allow simultaneous detection of both positive and negative ion data. Presently, either positive or negative ion data can be collected. It is possible that having both ionic polarities collected simultaneously would provide additional useful

data, as there may be correlations between the patterns of positive and negative ions coming from different types of microparticles.

Acknowledgments

This work was sponsored by the U.S. Department of Energy, Office of Research and Development.

Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

References

1) R.G. Pinnick, S.C. Hill, P. Nachman, G. Videen, G. Chen, R.K. Chang, "Aerosol Fluorescence-Spectrum Analyzer For Rapid Measurement Of Single Micrometer-Sized Airborne Biological Particles", Aerosol Science And Technology **28**, 95-104 (1998).

2) Y. Suyama, J. Muramatsu, Y. Takahashi, A. Noro, K. Takahashi, "Evaluation Of Airborne Microorganisms And Bioaerosols In Dental Clinic Environment", Journal Of Dental Research **77**, 1771-1771 (1998).

3) D.S. Gross, K.A. Prather, "Real-Time Analysis Of Individual Ambient Aerosol-Particles", Abstracts Of Papers Of The American Chemical Society **214**, pp.7 (1997).

4) J. Lacey, "Sampling And Rapid Assay Of Bioaerosols: Proceeding Of A Joint Meeting Of The British-Aerobiology-Federation And The Aerosol Society", Journal of Aerosol Science **28**, 345-348 (1997).

5) P. Nachman, G. Chen, R. Pinnick, S. Hill, K. Chang, M. Mayo, and G. Fernandez, "Conditional-sampling spectrograph detection system for fluorescence measurements of individual airborne biological particles", Applied Optics **35**, 1069-1076 (1996).

6) C.A. Noble, K.A. Prather, "Real-time measurement of correlated size and composition profiles of individual atmospheric aerosol particles", Environmental Science And Technology **30**, 2667-2680 (1996).

7) G. Chen, P. Nachman, R.G. Pinnick, S.C. Hill, R.K. Chang, "Conditional-firing aerosol-fluorescence spectrum analyzer for individual airborne particles with pulsed 266-nm laser excitation", Optics Letters **21**, 1307-1309 (1996).

8) M.V. Johnson, A.S. Wexler, "Analysis of individual aerosol particles", Analytical Chemistry **67**, 721A-726A (1995).

9) S.C. Hill, R.G. Pinnick, G. Chen, R.K. Chang, M.W. Mayo, and G.L. Fernandez, "Aerosol-fluorescence spectrum analyzer: real-time measurement of emission spectra of airborne biological particles", Applied Optics **34**, 7149-7155 (1995).

10) K.R. Spurny, "On the chemical detection of bioaerosols", Journal of Aerosol Science **25**, 1533-1547 (1994).

11) W.D. Griffiths, G.A.L. DeCosemo, "The assessment of bioaerosols: a critical review", Journal of Aerosol Science **25**, 1425-1458 (1994).

12) D.C. White, D.B. Ringelberg, D.B. Hendrick, D.E. Nivens, "Rapid identification of microbes from clinical and environmental matrices", Mass spectrometry for the Characterization of Microorganisms, C. Fenselau Ed., ACS Symposium Series 541, American Chemical Society, Washington DC, 1994.

13) O. Kievit, J. Marijnissen, P. Verheijen, and B. Scarlett, "Online measurement of particle size and composition", Journal Of Aerosol Science **23**, S301-S304 (1992).

14) M.P. Sinha, "Analysis of individual biological particles in air", Rapid Detection and Identification of Microorganisms, W. H. Nelson Ed., VCH Publishers, Deerfield beach, FL, 1985, pp. 165-192.

15) M.P. Sinha, R.M. Platz, V.L. Viker, S.K. Friedlander, "Analysis of individual biological particles by mass spectrometry", International Journal Of Mass Spectrometry Ion Processes **57**, 125-133 (1984).

16) M.P. Sinha, "Laser-induced volatilization and ionization of microparticles", Review of Scientific Instruments **55**, 886-891 (1984).

17) S.H. Wood, K.A. Prather, "Time-Of-Flight Mass-Spectrometry Methods For Real-Time Analysis Of Individual Aerosol Particles", Trends in Analytical Chemistry **17**, 346-356 (1998).

18) R.W. Hardin, "Lasers Help Identify Airborne Particles In Real-Time", Photonics Spectra **31**, pp.42 (1997).

19) B. Spengler, K.P. Hinz, R. Kaufmann, "Airborne Particle Analysis", Science **274**, pp.1996 (1996).

20) R.A. Gieray, P.T.A. Reilly, M. Yang, W.B. Whitten, J.M. Ramsey, "Real-time detection of individual airborne bacteria", Journal of Microbiological Methods **29**, 191-199 (1997).

21) P.T.A. Reilly, R.A. Gieray, M. Yang , W.B. Whitten, J.M. Ramsey, "Tandem mass spectrometry of individual airborne microparticles", Analytical Chemistry **69**, 36-39 (1997).

22) P.C. Gray, I.R. Shokair, S.E. Rosenthal, G.C. Tisone, J.S. Wagner, L.D. Rigdon, G.R. Siraguse, R.J. Heinen, "Distinguishability of biological material by use of ultraviolet multispectral fluorescence", Applied Optics **37**, 6037-6041 (1998).

23) M.W. Trahan, J.S. Wagner, I.R. Shokair, G.C. Tisone, P.C. Gray, "The Use of Intelligent Algorithms in Multispectral UV Analysis", CALIOPE Program 4th Interim Technical Review Proceedings, Vol. 1, 1997.

24) J.S. Wagner, M.W. Trahan, W.E. Nelson, G.C. Tisone, B.L. Preppernau, "How intelligent chemical recognition benefits from multivariate analysis and genetic optimization", Computers in Physics **10**, 114-118 (1995).

25) J. Hertz, A. Krogh, R. Palmer, Introduction To The Theory Of Neural Computation, Addison-Wesley Publishing Company, Inc., Reading Massachusetts, 1991.

26) D.E. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning, Addison-Wesley Publishing Company, Inc., Reading Massachusetts, 1989.

Figure Captions

Figure 1: Training results as a function of the number of Hidden Layer Neurodes.

a) The average fitness for a population of NNs trained on negative ion data is plotted as a function of the number of hidden layer neurodes (HLN's). As the number of neurodes increases from 1 to approximately 20 the fitness improves markedly, indicating that at a low number of HLN's the net is not large enough to handle the amount of data it has to process. After about 20 HLN's the fitness still increases, but the effect is not as dramatic. As stated in the text, using a prime number of neurodes greater than the number of neurodes in the input layer yields acceptable results. In this case, there are 22 neurodes in the input layer, making a guess of 23 HLN's a reasonable place to begin training.

b) The NN results for the two training cases are plotted as a function of the number of HLN's. After the NN training was complete all of the training cases were shown to the population of NNs and the response calculated. The results for *Bacillus subtilis* spores (top curve) and *Agrostis alba* pollen (bottom curve) both cross into the acceptable region somewhere between 15 and 19 HLN's, again indicating that 23 is a reasonable number of HLN's to work with.

c) The NN results for all species not used in NN training. The point where the results for each of these species becomes acceptable is lower that that where the results become acceptable for the training cases, indicating that the training cases were the most difficult species to recognize. This shows that a reasonable training strategy involves training the population of NNs on the species which are hardest to recognize, thereby forcing the NNs to generalize.

Figure 2: Neural Net Training Results for Negative Ion Data.

a) The architecture of the NNs trained on Negative Ion Data is illustrated in both the NN connection map and the weights histogram. The weights are (roughly) symmetrically distributed, indicating that the NNs have been through enough generations to become well-organized. This is also indicated by the fact that most weights are either inhibitory (negative) or excitatory (positive), while the number of weights in the vicinity of zero is small.

b) The results tabulated in Table I are plotted here. The training poles (*Bacillus subtilis* spores and *Agrostis alba* pollen) are shown in blue, while the other species are shown in red. The dotted lines represent the threshold criteria: values greater than 0.7 are classified by the NNs as bacteria while values less than 0.3 are classified by the NNs as non-bacteria.

Figure 3: Neural Net Training Results for Positive Ion Data.

a) The architecture of the NNs trained on Positive Ion Data. The number of connections with inhibitory and excitatory weights are approximately identical, indicating that the NNs are not as well-organized as those trained on the negative ion data

b) b) The results tabulated in Table II are plotted here. The training poles (*Bacillus subtilis*, *Artemisia tridentata* pollen, and *Juglans nigra* pollen) are shown in blue, while the other species are shown in red. As in Fugure 2, values greater than 0.7 are classified by the NNs as bacteria while values less than 0.3 are classified by the NNs as non-bacteria.

Table I: Neural Net Training Results for Negative Ion Data

Species (abbreviation)	NN Estimate
Azotobacter vinelandii (avin)	0.7811
Bacillus subtilis (bsubt)	0.8008
Escherichia coli (ecoli)	0.7992
Enterobacter aerogenes (enter)	0.7743
Micrococcus lysodeikticus (micro)	0.8229
*Bacillus subtilis spores (spore)	0.7311
Ambrosia trifida (ambtri)	0.1811
Artemisia tridentata (arttri)	0.1261
Betula alba (betalb)	0.1642
Dactylis glomerata (dacglo)	0.2167
Juglans nigra (juglan)	0.0733
*Agrostis alba (grass)	0.2869

* Species used in NN training

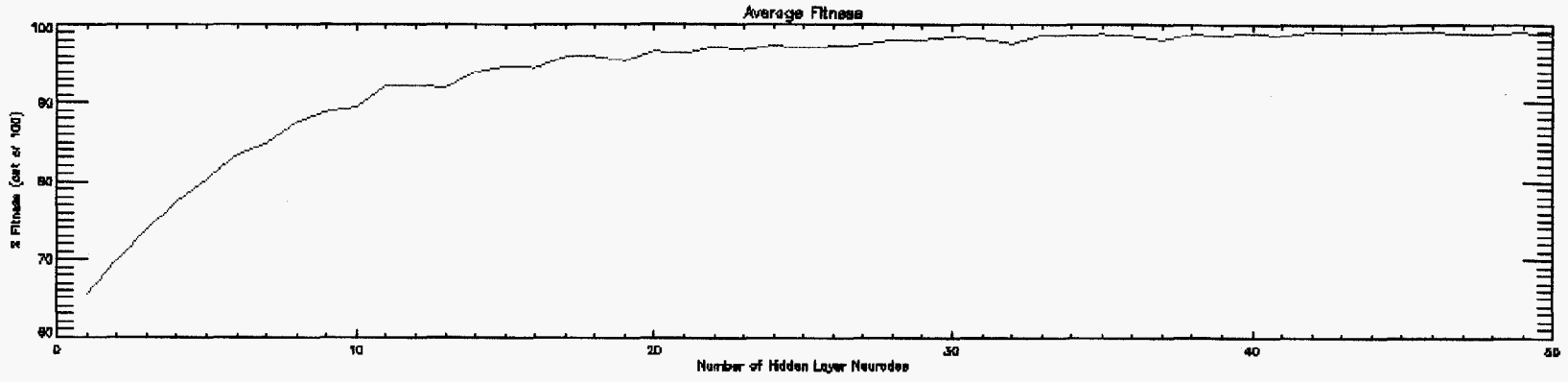
Table II: Neural Net Training Results for Positive Ion Data

Species (abbreviation)	NN Estimate
Azotobacter vinelandii (avin)	1.0000
*Bacillus subtilis (bsubt)	0.9821
Escherichia coli (ecoli)	1.0000
Enterobacter aerogenes (enter)	0.9086
Micrococcus lysodeikticus (micro)	0.7137
Bacillus subtilis spores (spore)	0.8747
Ambrosia trifida (ambtri)	0.0638
*Artemisia tridentata (arttri)	0.0000
Betula alba (betalb)	0.0477
Dactylis glomerata (dacglo)	0.0000
*Juglans nigra (juglan)	0.0000
Agrostis alba (grass)	0.1058

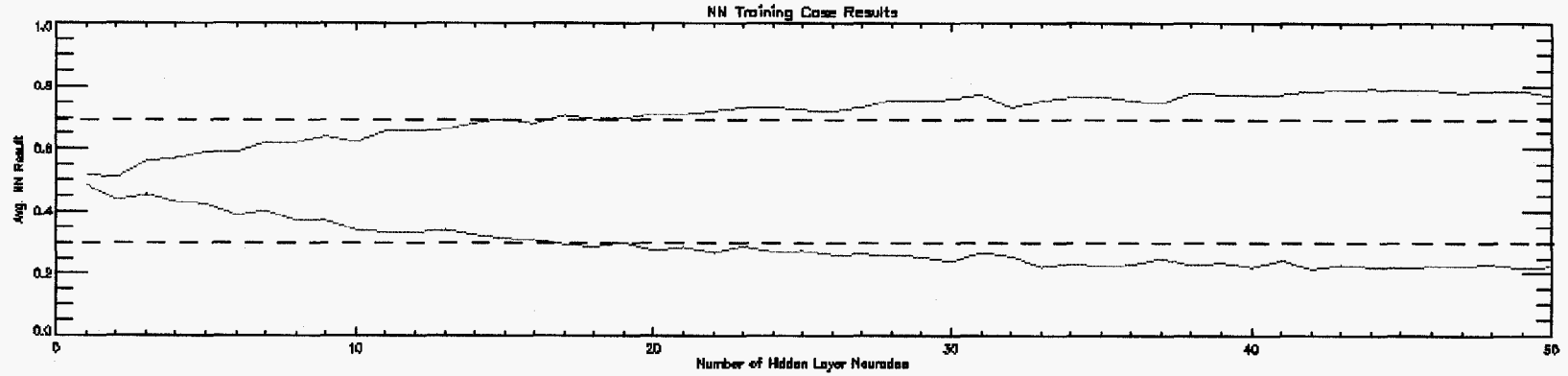
* Species used in NN training

Figure 1: Training Results as a function of the Number of Hidden Layer Neurodes

a)



b)



c)

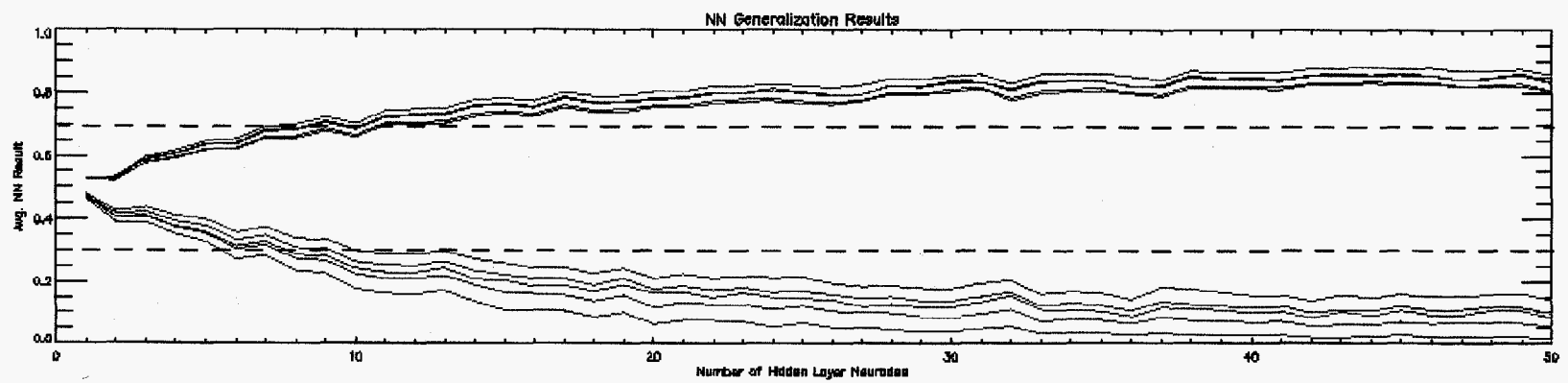
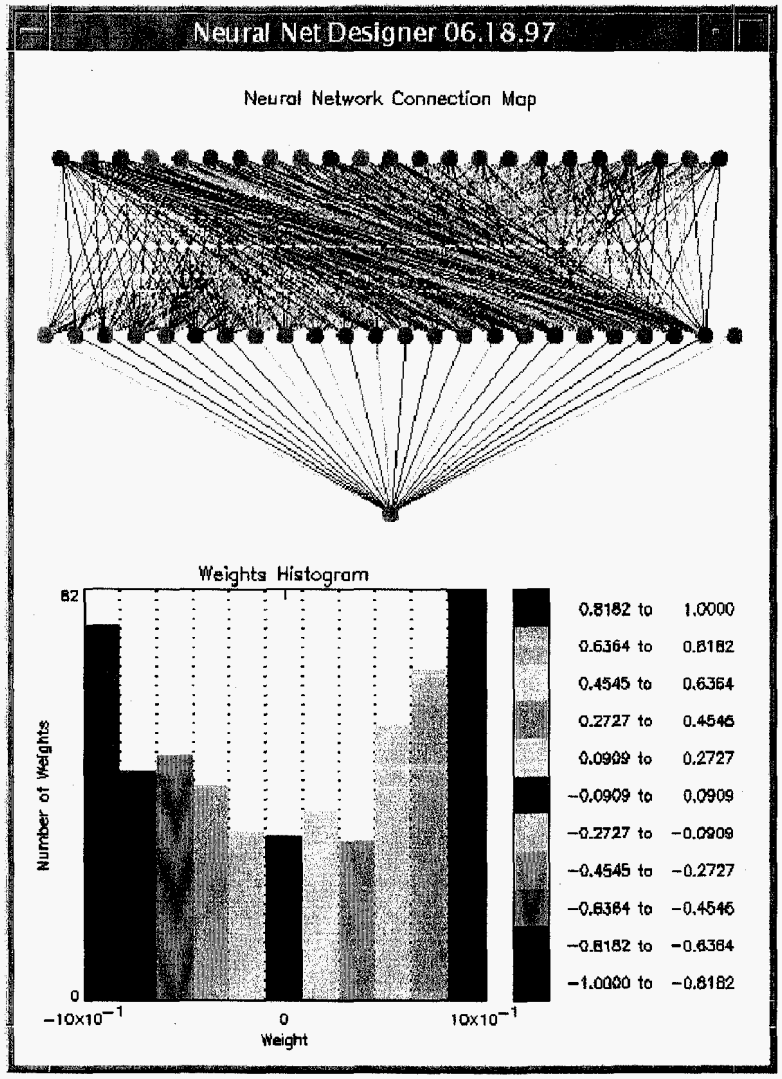


Figure 2: Neural Net Training Results for Negative Ion Data

a) Negative Ion NN Map and Weights Histogram



b) Negative Ion Data NN training results

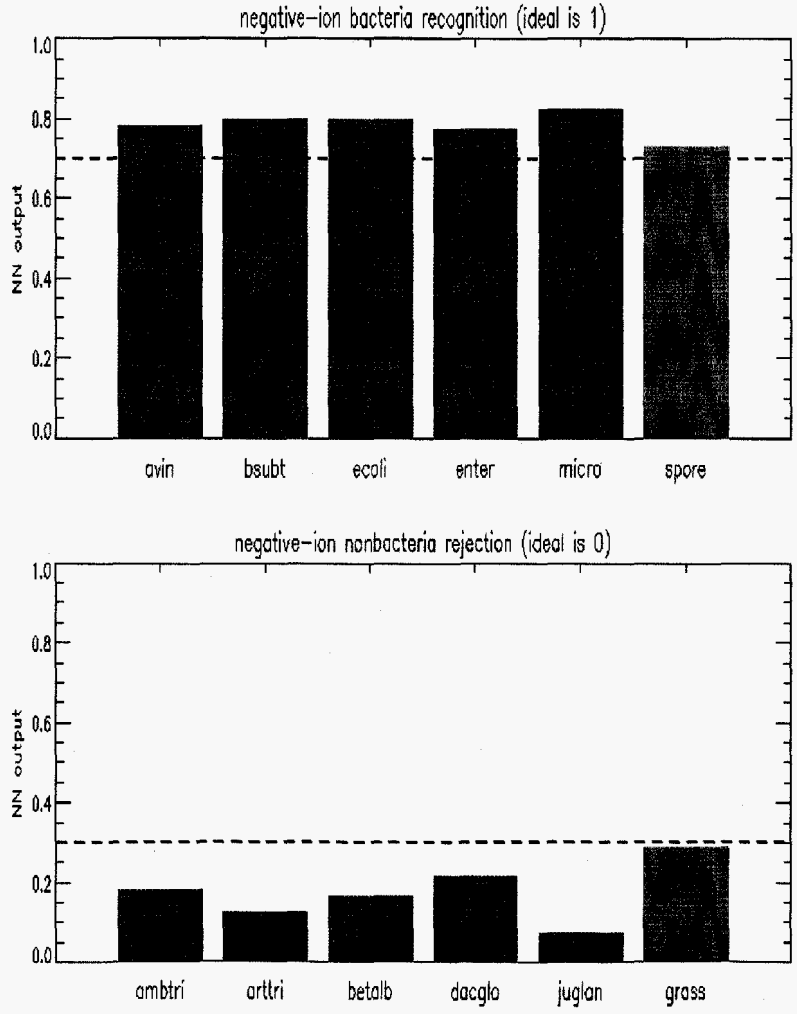
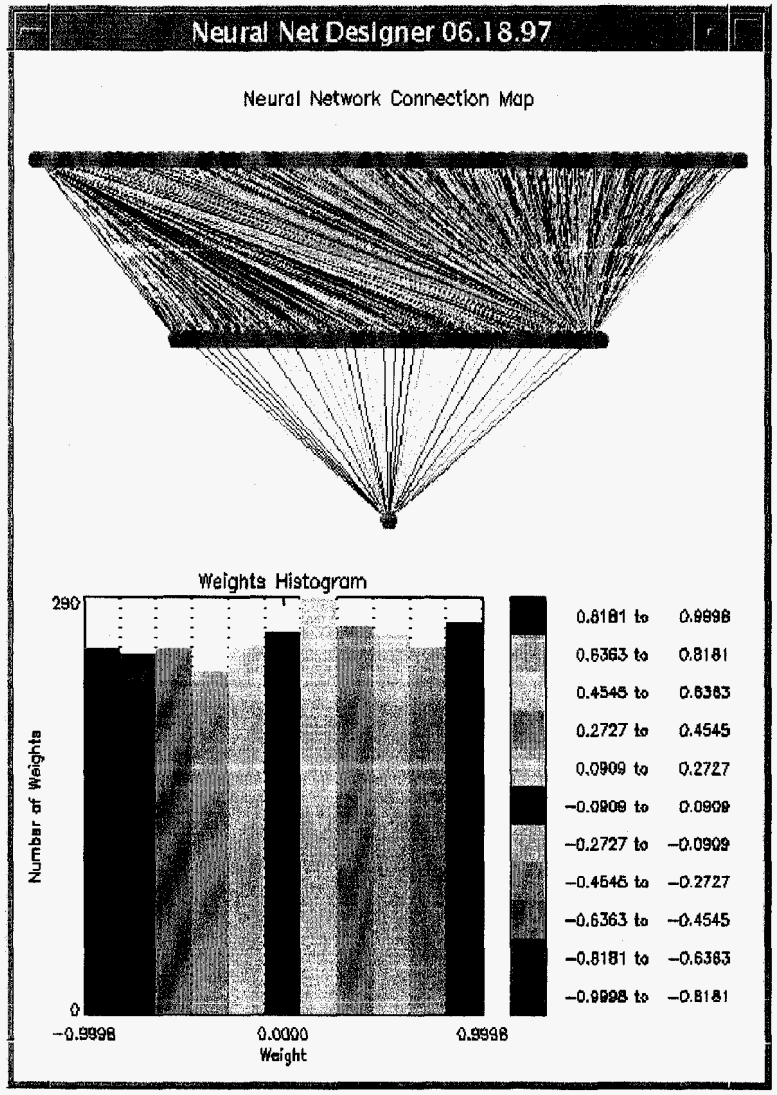


Figure 3: Neural Net Training Results for Positive Ion Data

a) Positive Ion NN Map and Weights Histogram



b) Positive Ion Data NN training results