Available online at www.sciencedirect.com
SCIENCE DIRECT®
Nuclear Instruments and Methods in Physics Research B 246 (2006) 471–478
NIM B
Beam Interactions with Materials & Atoms
www.elsevier.com/locate/nimb

Artificial neural network analysis of RBS and ERDA spectra of multilayered multielemental samples

N.R. Nené [a], A. Vieira [b], N.P. Barradas [a,c,*]

[a] Instituto Tecnológico e Nuclear, Apartado 21, E.N. 10, 2686-953 Sacavém, Portugal
[b] Instituto Superior de Engenharia do Porto, R. António Bernardino de Almeida 431, 4200 Porto, Portugal
[c] Centro de Física Nuclear da Universidade de Lisboa, Av. Prof. Gama Pinto 2, 1649-003 Lisboa, Portugal

Received 12 July 2005; received in revised form 28 November 2005
Available online 28 February 2006

Abstract

We have developed artificial neural networks (ANNs) for simultaneous analysis of Rutherford backscattering spectrometry and elastic recoil detection analysis data. The ANNs developed were applied to a highly complex problem, namely the analysis of multilayered silica–titania films doped with Ag and Er, where 11 parameters are required to describe the samples. Extensive optimization of network architecture, connectivity and pre-processing is presented. The optimized ANN was applied to experimental data leading to accurate results.

PACS: 02.60.−x; 82.80.Yc; 07.05.Mh; 07.05.Kf

Keywords: Rutherford backscattering; Neural networks; Data analysis; Ion beam analysis

1. Introduction

Rutherford backscattering spectrometry (RBS) is a nuclear analytical technique that allows one to determine elemental concentration profiles [1]. Depending on the experimental details, it can probe layers from 1 nm to tens of μm thick. It is fully quantitative and does not require standards. As such, it is widely used in the analysis of materials for applications in the most diverse fields. It is not sensitive to hydrogen, a common contaminant in samples. A complementary ion beam analysis technique, elastic recoil detection analysis (ERDA), is usually employed to quantify the amount of hydrogen.

Given a sample structure and the measurement setup, it is fairly easy to calculate the expected theoretical spectrum for both techniques. The inverse problem is however ill-posed, leading to different approaches to data analysis. The most common is still interactive and iterative comparison of the data with a simulation, which is refined until a good visual agreement is obtained [2,3]. Systematic treatments based on Bayesian inference with the Markov chain Monte Carlo method have been successfully utilized [4–7], but are difficult to use, computationally expensive (data analysis can take orders of magnitude longer than data collection), and are thus still inadequate to handle large quantities of data.

We have previously used artificial neural networks (ANNs) for automated analysis of Rutherford backscattering (RBS) data [8,9]. ANNs can be used as regression machines, designed for a specific problem containing similar data [10]. In the ANNs first developed by us, only a single spectrum could be analyzed from each sample. This is a limitation since in cases where more than one spectrum had been collected, they had to be analyzed separately, leading to loss of correlated information and hence to a reduction

* Corresponding author. Address: Instituto Tecnológico e Nuclear, Apartado 21, E.N. 10, 2686-953 Sacavém, Portugal. Tel.: +351 219946150; fax: +351 219941039.
E-mail address: nunoni@itn.pt (N.P. Barradas).

0168-583X/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.nimb.2006.01.016

in accuracy. Furthermore, when complementary RBS and ERDA data are collected, simultaneous and self-consistent analysis of all data is essential.

A second limitation is that the ANNs previously developed were dedicated to relatively simple systems, involving very few parameters to be determined. This situation is common in some cases, for instance where the required parameters are only the implanted dose and depth. In these cases, the information from which those parameters are retrieved is contained entirely in the implant peak [8], and the rest of the spectrum can be essentially ignored. In complex multilayered multielemental samples, the information is distributed in the whole spectrum (or spectra, if more than one is collected), and a large number of outputs may be required.

In this paper, we present an ANN for the analysis of RBS and ERDA spectra collected from the same sample. We apply the ANN to a highly complex case, namely the analysis of silica–titania-based sol–gel films doped with erbium and silver. Different annealing treatments lead to extensive diffusion and redistribution of the Ag to the surface and bottom of the films, and also to hydrogen loss in a surface region. Extensive work on optimization of network architecture and pre-processing of the input data is done. The best ANNs developed perform practically as well, and in some cases better, than standard analytic data analysis codes.

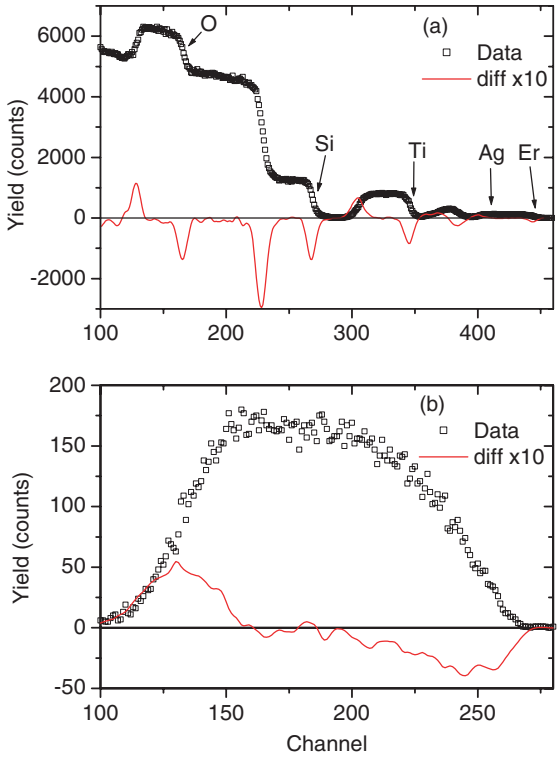

Fig. 1. (a) RBS and (b) ERDA data collected for sample 1. The differentiated data are also shown, scaled by a factor of 10.

2. Experimental data

Silica–titania films doped with erbium and silver were produced by spin-coating on Si substrates. The films have an H contamination, which is not homogeneous in depth: a surface layer has a reduced or enhanced H concentration, due to diffusion. The samples were subjected to different annealing procedures. The experimental details are given elsewhere [11].

RBS and ERDA experiments were done using a 1.925 MeV $^4He^+$ beam. The details of the measurements have been given elsewhere [12]. A total of 17 different samples was measured. Most of them had both Ag and Er dopants. A few had no Ag, a few had no Er, and one had neither Ag nor Er. Due to the different annealing procedures, some of the samples developed a thin Ag layer between the film and the Si substrate. In a few of those samples, a thin Ag film also appeared on the surface.

The RBS and ERDA data of sample 1 are shown in Fig. 1. The RBS spectrum carries all the information about the Si, Ti, O, Ag and Er. Furthermore, the Ag and Er signals are extensively superimposed on each other, and exist in the channel range 350–460 only. The signal of the lighter elements O, Si and Ti are at channels smaller than 350. The ERDA spectrum carries all the experimental information on the H.

The data were analyzed with a general-purpose data analysis code, NDF [13], that calculates analytic spectra based on standard algorithms [14]. NDF uses tabulated values for the energy loss of ions in the samples [15]. Energy straggling due to the statistical nature of the energy loss process, to the energy and angular spread of the analyzing beam, geometric straggling cause by finite size of the beam spot and detector and multiple scattering is calculated with the code DEPTH [16,17]. The background due to plural scattering [18] is expected to be small in these samples and was disregarded. The background due to pulse pileup in the MCA was calculated with the algorithm given in [19].

3. ANN architecture and connectivity

ANNs can approximate any arbitrary unknown function [10]. They are particularly well suited to high-dimensional non-linear regression analysis with noisy signals and incomplete data [20]. In the case of RBS and ERDA data, the signal (position and height) of each element depends on all other elements in a non-linear way. Experimental data are by nature noisy. Also, the problem of finding a compositional depth profile from the data is ill-posed: for each element present in the sample, information on a different depth range is available; the sensitivity for some elements can be either zero or very small; in the general case, the yield in one case cannot be assigned a priori to a given element, with extensive superposition of signals. Therefore regularisation constraints must be imposed (experience of a well-trained physicist being the method most often used). One should also notice that it is not

possible to obtain a closed-form formula to calculate an RBS or ERDA spectrum with a degree of accuracy remotely suitable for data analysis.

A feedforward neural network consists of an array of input nodes connected to an array of output nodes through the nodes of successive intermediate layers. Each connection between nodes has a weight, initially random, which is adjusted during a training process. The output of each node of a specific layer is a function (in this work a sigmoid was used [8]) of the sum of the weighted signals coming from the previous layer. The crucial points in the construction of an ANN are the selection of inputs and outputs, the architecture of the ANN, that is, the number of layers and of nodes in each layer, the connectivity of the ANN, that is, the way the nodes are connected to each other and finally, the training algorithm.

The ANN architecture can be represented by $(N, I_1, \ldots, I_n, M)$, where N is the number of inputs and M the number of outputs and I_i the number of nodes in each intermediate hidden layer. In this case, the inputs can be, for instance, the yield in each channel (i.e. in each multichannel bin). The outputs are: the thickness t_{film} of the silica–titania film; the Si, Ti, O, H, Ag and Er concentrations in the film; the thickness $t_{\text{Ag\_top}}$ and $t_{\text{Ag\_bottom}}$ of the top and bottom Ag layers; the thickness of the surface layer with changed H concentration, which we call the H depletion thickness $t_{\text{depletion}}$; and a H depletion factor $f_{\text{depletion}}$, which is the ratio between the H concentration in the H depleted layer and in the rest of the film. This makes a total of 11 outputs, which is unusually large in ANNs, and is a consequence of the complexity of the problem. Note that for the silica–titania film, the concentration parameters are not independent, since they must add up to 100 at.%. Nevertheless, they were treated by the ANN as independent, and then normalised to 100 at.%. The layer areal density is independent of the elemental concentrations, because relative concentrations, not absolute, were used.

We used supervised learning with the backpropagation algorithm [8,10]. The training is done by presenting a large set of examples, the training set, for which the inputs and corresponding outputs are known. The difference between the ANN prediction and the known output values is minimized by adjusting the connection weights. The quality of the prediction is verified by testing the ANN with an independent set of data, the test set. Root mean square errors e_{rms} are calculated for both sets. This is also calculated for the real data, taking as reference the results given by NDF. Training is repeated until no improvement in the test set can be found.

The training and test sets were generated by constructing theoretical spectra corresponding to a very broad range of sample parameters (outputs), to which realistic Poisson noise was added. The parameter range in the training and test sets was $t_{\text{film}} \in [1300, 2000]$ (10^{15} at/cm^2), Si $\in [6, 45]$ (at.%), Ti $\in [0, 50]$ (at.%), O $\in [33, 85]$ (at.%), H $\in [0, 42]$ (at.%), Ag $\in [0, 1]$ (at.%), Er $\in [0, 8]$ (at.%), $t_{\text{Ag\_top}} \in [0, 10]$ (10^{15} at/cm^2), $t_{\text{Ag\_bottom}} \in [0, 10]$ (10^{15} at/cm^2), $t_{\text{depletion}} \in [200, 800]$ (10^{15} at/cm^2), $f_{\text{depletion}} \in [0.25, 5]$ (for this parameter only, the distribution was not uniform, but instead a larger density of examples in the $[0.25, 2]$ range, where most experimental cases were expected to be, was used). These ranges are much broader than expected in real samples. The inputs were each normalised to lie in the $[0, 1]$ interval, which facilitates the training.

We developed ANNs with different types of inputs. In one case, the inputs were simply the yield values without any pre-processing. In another case, pre-processing of the spectra was done by differentiating them using cubic fitting splines and adaptive smoothing weights with prescribed third end point derivatives $f'''(x_0) = f'''(x_n) = 0$ [21]. This procedure leads to differentiated data with minimized fluctuations due to statistical uncertainties. The objective of using the differentiated data as inputs is to enhance relevant small signals and eliminate slowly changing backgrounds, which can lead to better performance of the ANN [9]. We show in Fig. 1 the differentiated spectra of sample 1.

Furthermore, the relevant region of the raw data has 360 channels for the RBS spectra and 180 for the ERDA spectra, which leads to 540 inputs. As each channel has an energy width smaller than the energy resolution of the experiment, compression of the data can lead to smaller, and thus more efficient, ANNs, without loss of information. We trained ANNs with the inputs uncompressed, and compressed by adding 4, 8 or 16 channels together, and otherwise the same architecture.

In ion beam analysis, the signal of each element cannot, in principle, be analyzed independently from the other ones, because, for instance, relevant quantities such as stopping powers require knowledge of the signal from all the elements. However, if each given signal was analyzed independently, assuming for instance equal concentrations for all other elements, a first approximation to the real values would be obtained. Those first approximations could be used all together to obtain a more refined result.

The function of the first hidden layers is to perform a non-linear transformation of the inputs into a form more easily treated by the following layers. We developed two different types of network connectivity. In the first case, full connectivity from all nodes in one layer to all nodes in the next layer was used. This is the traditional perceptron multilayer, used in the vast majority of ANNs. In this case, it corresponds to analyzing all spectra simultaneously.

In the second case, we consider different clusters of nodes in the input data. Each cluster is connected only to a given set of nodes in the first hidden layer. Other clusters are not connected to that set of nodes. In this way, a first treatment of each group of signals is made, with the results being integrated in a second step. Pre-processing occurs in the first hidden layer, with the advantage of substantially reducing the total number of connections, which reduces the complexity of the problem and facilitates efficient training of the network [22].

Although the signal of the different elements depends on the other elements as well, some properties of the samples

and of the experimental data can be used: first, the ERDA spectrum carries all the information on the H; second, the Ag and Er are impurities with small concentration, so they have a very small influence on the signal from the other elements; third, the major elements in the silica–titania film are the Si, Ti and O, and the signals of these elements are superimposed. We thus chose three clusters: the ERDA spectrum, corresponding to the H signal; the low energy range of the RBS spectrum (channels 100–350), corresponding to the O, Si and Ti signals; and the high energy range of the RBS spectrum (channels 350–460), corresponding to the Ag and Er signals.

Finally, we would like to note that the basic algorithms behind feedforward ANNs trained with backpropagation are simple to implement and extremely well documented in the vast literature on the subject. The difficulty of developing a ANN for a given problem lies on developing a ANN architecture and connectivity that leads to efficient training and high accuracy of the results. This optimisation must be made for each new problem to be solved. Once that knowledge is obtained, to write the corresponding ANN computer code is a trivial task.

4. Results and discussion

4.1. Network optimization

Results for cluster-linked ANNs without compression and with different levels of compression are shown in Tables 1–4. The training times indicated were all obtained with the same computer and are thus directly comparable. There is an optimal network size, with two hidden layers with around 50 and 30 nodes, respectively. Smaller ANNs lead to higher errors as they are too small to handle this complex problem. Large ANNs have larger errors because they are harder to train, and may have higher generalization errors.

For a given architecture, training is faster with a smaller number of inputs, as expected for a smaller network with fewer nodes and fewer connections. More interestingly, the errors e_{rms} obtained for the train and test sets are consistently smaller with raw data (i.e. no compression) and

Table 1
Results for cluster-linked networks with uncompressed yield values as inputs

Architecture	e_{rms}			Training time (s)
	Train set	Test set	Experimental set	
542, 50, 11	0.0549	0.0498	0.2768	32,333
542, 70, 11	0.0490	0.0472	0.3223	68,442
542, 70, 50, 11	0.0295	0.0288	0.1934	72,192
542, 70, 30, 11	0.0313	0.0308	0.2520	70,152
542, 50, 30, 11	0.0282	0.0272	0.1501	56,604
542, 50, 20, 11	0.0345	0.0333	0.2671	33,917
542, 50, 15, 11	0.0370	0.0358	0.4254	31,237
542, 40, 30, 11	0.0372	0.0352	0.2919	21,623
542, 40, 20, 11	0.0374	0.0355	0.4651	21,338
542, 40, 15, 11	0.0382	0.0370	0.3278	20,941
542, 30, 20, 11	0.0383	0.0365	0.4108	15,700
542, 30, 15, 11	0.0429	0.0406	0.4309	15,254

Twenty thousand examples were used in the training.

Table 2
Results for cluster-linked networks with 4-channel compression yield values as inputs

Architecture	e_{rms}			Training time (s)
	Train set	Test set	Experimental set	
136, 50, 11	0.0631	0.0608	0.3296	2909
136, 70, 11	0.0656	0.0641	0.3408	5968
136, 70, 50, 11	0.0338	0.0330	0.2004	7007
136, 70, 30, 11	0.0360	0.0343	0.4068	6922
136, 50, 30, 11	0.0341	0.0328	0.1148	6822
136, 50, 20, 11	0.0412	0.0391	0.1566	4882
136, 50, 15, 11	0.0393	0.0375	0.4476	3800
136, 40, 30, 11	0.0372	0.0358	0.1712	3208
136, 40, 20, 11	0.0420	0.0401	0.1556	2960
136, 40, 15, 11	0.0420	0.0400	0.3400	2837
136, 30, 20, 11	0.0429	0.0405	0.4651	2274
136, 30, 15, 11	0.0445	0.0420	0.3418	2188

Twenty thousand examples were used in the training.

Table 3
Results for cluster-linked networks with 8-channel compression yield values as inputs

Architecture	e_{rms}			Training time (s)
	Train set	Test set	Experimental set	
68, 50, 11	0.0755	0.0732	0.3947	2067
68, 70, 11	0.0715	0.0694	0.4303	2916
68, 70, 50, 11	0.0359	0.0346	0.2431	5245
68, 70, 30, 11	0.0379	0.0367	0.3647	4545
68, 50, 30, 11	0.0341	0.0325	0.1672	3806
68, 50, 20, 11	0.0393	0.0377	0.4517	3638
68, 50, 15, 11	0.0409	0.0389	0.3313	3404
68, 40, 30, 11	0.0357	0.0342	0.2373	3389
68, 40, 20, 11	0.0434	0.0414	0.2807	2795
68, 40, 15, 11	0.0472	0.0450	0.3433	2100
68, 30, 20, 11	0.0453	0.0434	0.4646	1754
68, 30, 15, 11	0.0551	0.0527	0.3731	1573

Twenty thousand examples were used in the training.

Table 4
Results for cluster-linked networks with 16-channel compression yield values as inputs

Architecture	e_{rms}			Training time (s)
	Train set	Test set	Experimental set	
34, 50, 11	0.1123	0.1089	0.4717	1267
34, 70, 11	0.1093	0.1053	0.4531	1589
34, 70, 50, 11	0.0561	0.0540	0.3281	4283
34, 70, 30, 11	0.0632	0.0601	0.3712	3823
34, 50, 30, 11	0.0626	0.0601	0.3531	2313
34, 50, 20, 11	0.0624	0.0606	0.2164	1452
34, 50, 15, 11	0.0691	0.0659	0.4540	1309
34, 40, 30, 11	0.0620	0.0601	0.1998	1401
34, 40, 20, 11	0.0687	0.0655	0.4448	1244
34, 40, 15, 11	0.0701	0.0682	0.2781	1066
34, 30, 20, 11	0.0735	0.0714	0.3747	940
34, 30, 15, 11	0.0813	0.0781	0.2289	841

Twenty thousand examples were used in the training.

increase only slightly after using 4- and 8-channel compression. However, for a 16-channel compression the errors increase markedly. The reason is that the energy resolution in the RBS spectra corresponds to roughly 13 channels. In a compression smaller than 13 channels, each point corresponds to more than one energy resolution, and there is at worst a small loss of information. On the other hand, a compression larger than 13 channels added together leads to a severe loss of information and thus to worse errors by the ANN.

For the experimental data the best error values were obtained using a 4-channel compression, which is the best compromise between smaller ANNs and information loss due to the compression. In particular, the best results were obtained for the architecture (136, 50, 30, 11), and thus it will be used in further tests.

First, we tested the influence of the size of the training and test sets in the results obtained. We show in Tables 5 and 6 the errors obtained with a (136, 50, 30, 11) ANN trained with data sets with different sizes (the test set was 10% of the data), for cluster-linked and fully linked networks, respectively. Better results are obtained when larger data sets are used in the training. The reason is that the ANN considered is still fairly large, with 4270 connections, and requires a large number of examples to achieve good generalization capabilities. On the other hand, this leads to a large increase in the time required for the training. This increase is more than linear, which is due to the highly non-linear properties of ANNs. Note that using fewer examples than connections leads to errors that would be acceptable if the ANN goal was to provide good initial guesses for use in some local minimisation scheme, or if high numerical accuracy was not required. Furthermore, the test set was generated randomly, and in principle better training could be achieved with a small test set if carefully chosen examples were used. Finally, the problem is highly non-linear and the parameters are correlated. For instance, the position of the back Ag signal depends on the thickness of the silica–titania film, and the height of the signal of each element depends on all other elements.

Table 5
Results for cluster-linked (136, 50, 30, 11) networks with 4-channel compression yield values as inputs

# Data	e_{rms}			Training time (s)
	Train set	Test set	Experimental set	
1000	0.1189	0.1076	0.5214	202
2000	0.0863	0.0954	0.4434	404
5000	0.0559	0.0551	0.3091	1016
10,000	0.0477	0.0477	0.2513	2199
20,000	0.0341	0.0328	0.1148	6822

Different number of examples were used in the training of each ANN.

Table 6
Results for fully linked (136, 50, 30, 11) networks with 4-channel compression yield values as inputs

# Data	e_{rms}			Training time (s)
	Train set	Test set	Experimental set	
2000	0.0956	0.1059	0.4530	404
20,000	0.0396	0.0388	0.3723	19,348

Different number of examples were used in the training of each ANN.

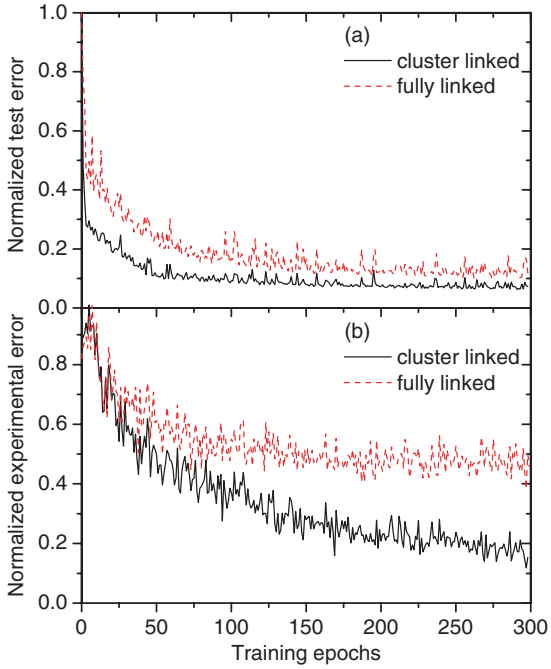

Fig. 2. Normalized (a) test set error and (b) experimental set error as a function of training iteration (in one iteration all training examples are presented to the ANN), for fully linked and cluster-linked (136, 50, 30, 11) ANNs.

We show in Fig. 2 the evolution of the test and experimental errors as a function of the training epochs, for (136, 50, 30, 11) cluster-linked and fully linked ANNs. At each epoch the entire training set is presented to the network. The cluster-linked network not only achieved a lower error, but it also has a faster rate of improvement. One should also note that for the same number of training iterations, the computer time required is much larger for fully linked ANNs. This clearly shows that an effective pre-processing is occurring in the first hidden layer in the cluster-linked network, leading to more efficient training and an increased efficiency of the network.

Finally, we note that using differentiated data, for the RBS data, for the ERDA data, or for both, did not lead to any improvement in the results. The raw data have a clear structure that the ANN is capable of utilizing to achieve the final results.

4.2. Network performance

We show in Figs. 3–7 the results obtained for all the experimental spectra available. As seen in Fig. 3, the ANN performed excellently in determining the thickness of the sol–gel film, being capable, in most cases, of

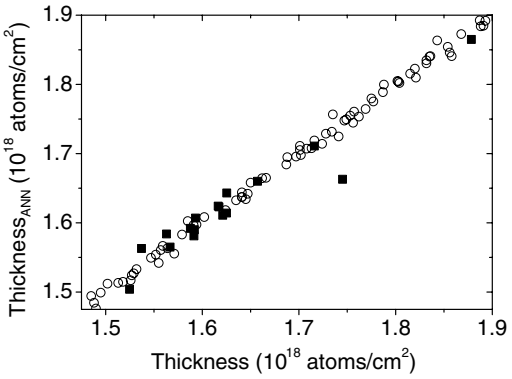

Fig. 3. Results for the sol–gel film thickness of real samples (solid squares), obtained with a cluster-linked (136, 50, 30, 11) ANN. Results for some test set examples are also shown (open circles).

recognizing differences in thickness larger than 2×10^{16} at/cm^2. Indeed, the low energy part of the RBS data carry redundant information, since the width of the O, Si and Ti signals can all be used to determine this parameter.

The results for the concentration of all elements in the sol–gel film are shown in Fig. 4. In general the ANN results match closely the reference values obtained with NDF. In particular, the results for hydrogen are excellent. This is explained by the fact that the hydrogen concentration is almost directly obtained from the yield of the ERDA spectrum, which carries information exclusively about the hydrogen. On the other hand, the ANN overestimates the Er concentration by a fairly constant amount. This may be explained by the fact that the Er signal is the smallest present in the data, and even a small background, due for instance to pileup [19], can lead to a significant error. This is confirmed by the results obtained for the test set, where the ANN performs well. Also, if a manual background subtraction procedure is applied to the experimental spectra before being input to the ANN, the ANN results improve. This shows that the bias observed for the Er is not intrinsic to the network, but indeed due to the background present in the experimental data.

The results obtained for the top and bottom silver layers are shown in Fig. 5. For the bottom layer, the ANN results

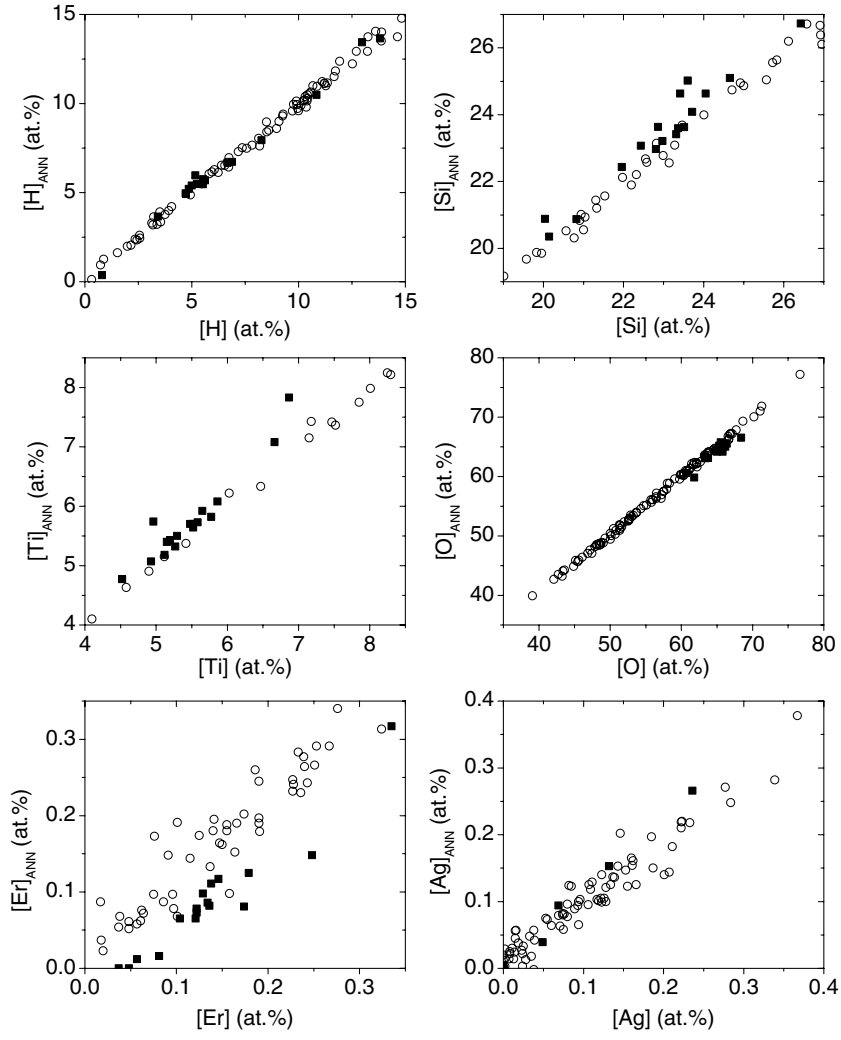

Fig. 4. Results for the sol–gel film composition of real samples (solid squares), obtained with a cluster-linked (136, 50, 30, 11) ANN. Results for some test set examples are also shown (open circles).

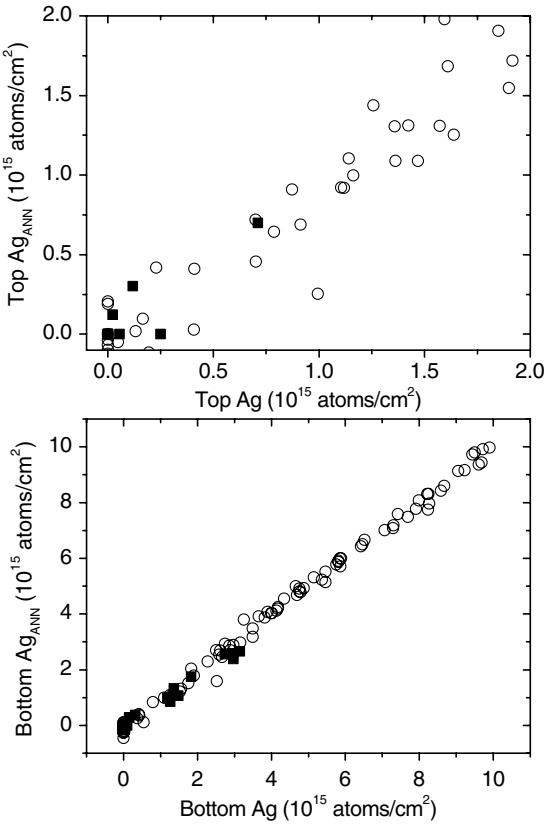

Fig. 5. Results for the thickness of the top and bottom Ag layers of real samples (solid squares), obtained with a cluster-linked (136, 50, 30, 11) ANN. Results for some test set examples are also shown (open circles).

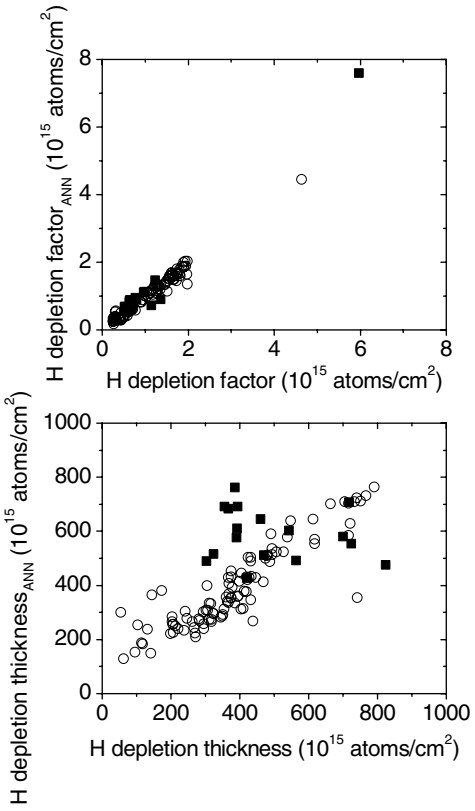

Fig. 6. Results for the H depletion factor and H depletion thickness of real samples (solid squares), obtained with a cluster-linked (136, 50, 30, 11) ANN. Results for some test set examples are also shown (open circles).

are quite good considering the very small thickness of this layer. For the top Ag layer, the ANN performance is worse. This is probably due to superposition of the signal of this layer and the background due to the Er signal. From the five samples where this top layer exists, the ANN only gives a non-zero output for three. On the other hand, it correctly gives a zero output for the 12 samples where this layer does not exist. All in all, the ANN discrimination capability is around 0.2×10^{15} at/cm^2 for this parameter.

The results obtained for the H depletion layer are shown in Fig. 6. While the trend is fairly well reproduced for the depletion factor, the results for the depletion thickness seem to be very bad: in fact, the ANN results seem to be uncorrelated with the reference values given by NDF. Several considerations are required to understand this. First of all, the resolution in the ERDA experiment is around 200×10^{15} at/cm^2, and any difference smaller than this value cannot be distinguished. Second, an hydrogen depletion factor equal to one means that there is no depletion layer, and thus there is no sensitivity to its thickness; for depletion factors close to one, the sensitivity is very small, and the NDF reference results, which come from a fit to the data, do not mean much. Only seven samples had a depletion factor outside the 0.8–1.2 range. The results obtained for the test set also show a high dispersion, although slightly smaller than for the experimental data. Finally,

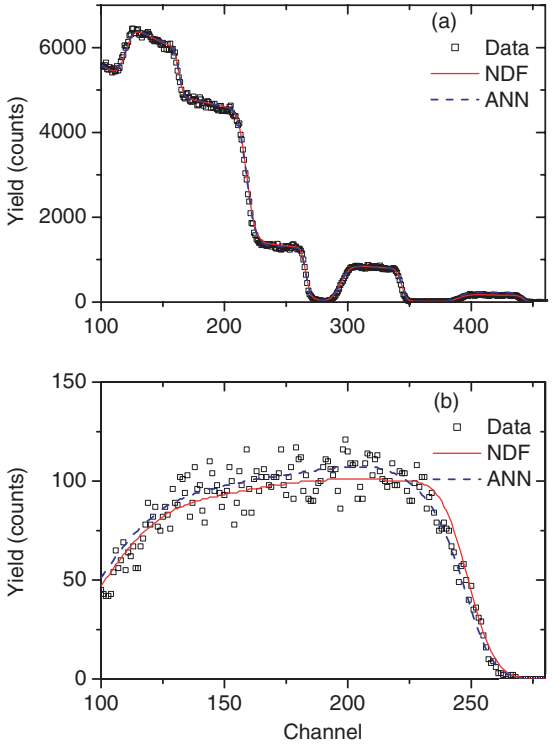

Fig. 7. (a) RBS and (b) ERDA data collected for one sample 5, together with simulations done with NDF and for the ANN output.

the NDF analysis considered simultaneously the RBS and ERDA spectra of the same sample. This means that a good overall fit may be achieved with a sub-optimal fit to the ERDA data. On the contrary, in cluster-linked ANNs the ERDA data is first considered independently of the RBS data. This may, in some cases, lead to a better fit to the ERDA data than obtained with NDF. This is illustrated in Fig. 7, where we show the data of sample 5, together with simulations considering the NDF reference values and the ANN outputs. The NDF results indicate a depletion factor 0.95 and a depletion thickness 762×10^{15} at/cm^2, while the ANN outputs are 0.78 and 386×10^{15} at/cm^2. It is clear that, while the two simulations of the RBS data are not distinguishable at first sight, the ANN simulation of the ERDA spectrum is superior. On the contrary and as mentioned above, the ANN overestimates the Er concentration.

5. Conclusions

We developed artificial neural networks capable of analyzing multiple ion beam analysis spectra collected from the same sample. The ANNs developed are capable of using the information present in all the data to produce accurate final results, in some cases better than those obtained with standard data analysis codes.

The ANNs were successfully applied to a highly complex problem: determination of the thickness and composition of silica–titania sol–gel thin films doped with Ag and Er subject to different annealing treatments, which lead to extensive Ag diffusion towards the bottom and surface of the film, and also to a surface layer of the film with changed H concentration. A total of 11 outputs were required to represent the sample structure, which is an unusually high number in ANN analysis.

A cluster-linked neural network architecture was proposed. In this architecture, different sections of the data are connected to a subset of nodes in the first hidden layer that are exclusively dedicated to that data section. A thorough study of network architecture, connectivity and effectiveness of pre-processing was made. Effective automatic pre-processing (as opposed to a priori pre-processing) is achieved with cluster linking, leading to very efficient and easy to train networks. We found that cluster-linked networks performed better than classic fully linked ANNs.

Acknowledgments

The authors thank the financial support of FCT under grant POCTI/CTM/40059/2001. The algorithms and codes used in this work can be obtained from the authors.

References

[1] J.R. Tesmer, M. Nastasi (Eds.), Handbook of Modern IBA, MRS, Pittsburgh, 1995.
[2] L.R. Doolittle, Nucl. Instr. and Meth. B 9 (1985) 344.
[3] M. Mayer, Nucl. Instr. and Meth. B 194 (2002) 177.
[4] R. Fischer, M. Mayer, W. von der Linden, V. Dose, Phys. Rev. E 55 (1997) 1.
[5] V.M. Prozesky, J. Padayachee, R. Fischer, W. von der Linden, V. Dose, R.A. Weller, Nucl. Instr. and Meth. B 136–138 (1998) 1146.
[6] N.P. Barradas, C. Jeynes, M. Jenkin, P.K. Marriott, Thin Solid Films 343–344 (1999) 31.
[7] P. Neumaier, G. Dollinger, A. Bergmeier, I. Genchev, L. Görgens, R. Fischer, C. Ronning, H. Hofsäss, Nucl. Instr. and Meth. B 183 (2001) 48.
[8] N.P. Barradas, A. Vieira, Phys. Rev. E 62 (2000) 5818.
[9] V. Matias, G. Öhl, J.C. Soares, N.P. Barradas, A. Vieira, P.P. Freitas, S. Cardoso, Phys. Rev. E 67 (2003) 046705.
[10] C.M. Bishop, Neural Networks for Pattern Recognition, Oxford University Press, Oxford, 1995.
[11] A.C. Marques, R.M. Almeida, A.R. Ramos, E. Alves, J. Sol–Gel Sci. Tech. 31 (2004) 287.
[12] A.C. Marques, A. Ramos, E. Alves, R.M. Almeida, Nucl. Instr. and Meth. B 219–220 (2004) 923.
[13] N.P. Barradas, C. Jeynes, R.P. Webb, Appl. Phys. Lett. 71 (1997) 291.
[14] E. Kótai, Nucl. Instr. and Meth. B 85 (1994) 588.
[15] J.F. Ziegler, J.P. Biersack, U. Littmark, Stopping and Ranges of Ions in Solids, Pergamon, New York, 1985.
[16] E. Szilágyi, F. Pászti, G. Amsel, Nucl. Instr. and Meth. B 100 (1995) 103.
[17] E. Szilágyi, Nucl. Instr. and Meth. B 161–163 (2000) 37.
[18] A. Weber, H. Mommsen, W. Sarter, A. Weller, Nucl. Instr. and Meth. 198 (1982) 527.
[19] C. Jeynes, Z.H. Jafri, R.P. Webb, M.J. Ashwin, A.C. Kimber, Surf. Interface Anal. 25 (1997) 254.
[20] B. Cheng, D.M. Titterington, Statist. Sci. 9 (1994) 2.
[21] G. Engeln-Müllges, F. Uhlig, Numerical Algorithms with Fortran, Springer-Verlag, Berlin, Heidelberg, 1996.
[22] H.F.R. Pinho, A. Vieira, N.R. Nené, N.P. Barradas, Nucl. Instr. and Meth. B 228 (2005) 383.