Nuclear Instruments and Methods in Physics Research A 369 (1996) 642–647

NUCLEAR INSTRUMENTS & METHODS IN PHYSICS RESEARCH Section A

ELSEVIER

Statistical modelling of neural networks in γ-spectrometry

V. Vigneron[a,c,*], J. Morel[b], M.C. Lépy[b], J.M. Martinez[a]

[a] *CEA/Saclay, DRN/DMT/SERMA, 91191 Gif-sur-Yvette, France*
[b] *CREL, 161, Rue de Versailles, 78180 Le Chesnay, France*
[c] *CEA/Saclay, LPRI, BP 52, 91193 Gif-sur-Yvette cedex, France*

Abstract

Layered Neural Networks are a class of models based on neural computation and have been applied to the measurement of uranium enrichment. The usual methods consider a limited number of X- and γ-ray peaks, and require calibrated instrumentation for each sample. Since the source-detector ensemble geometry conditions critically differ between such measurements, the spectral region of interest is normally reduced to improve the accuracy of such conventional methods by focusing on the $K_\alpha X$ region where the three elementary components are present. Such measurements lead to the desired ratio. Experimental data have been used to study the performance of neural networks involving a Maximum Likelihood Method. The encoding of the data by a Neural Network approach is a promising method for the measurement of uranium ^{235}U and ^{238}U in infinitely thick samples.

1. Introduction

Neural computing has generated widespread interest and popularity in recent years. The popularity of this technique is due in part to the analogy between Artificial Neural Networks (ANNs) and biological neural systems. Many applications have been investigated using ANNs, and we demonstrate their use below in the analysis of photon spectra, from uranium-enrichment measurements to determine the $^{245}U/U_{total}$ isotope ratio. With modern detector systems, complex and precise spectral data can be collected that impose a demanding need for efficient interpretational methods.

Traditional non-destructive analysis for uranium-enrichment measurement involves the use of several X- and γ-ray peaks, mainly in the 60 to 200 keV region. Most of these methods were developed more than 20 years ago, and are based on measurements of the full energy peak at 185.7 keV [1–4]). This approach requires calibration of the system and the measurement conditions to remain constant. Other methods have been developed using several γ-ray peaks and calibration with a limited number of peaks [5,6].

Calibration procedures and matrix effects can be avoided by focusing the spectra analysis on the $K_\alpha X$ region (which contains the main uranium components) and by using infinitely thick samples. Such samples sufficiently thick that any further increase does not affect the γ-ray emissions.

The spectral processing of the $K_\alpha X$ region involves quantification of the photon emission identified with ^{235}U, ^{238}U and X-ray fluorescence. This approach requires well-defined data for the photons emissions, together with he detector characteristics and geometry.

Under such circumstances a Neural Network would be a useful tool in developing a search procedure for an "optimum" regression function among a set of acceptable functions. ANNs belong to evaluation techniques for non-parametric models called *tabula rasa*. Like most statistical methods, ANNs are able to process vast amounts of data and to make predictions that can be surprisingly accurate. Compared with usual automatic spectra analysis methods, ANNs use full-parallel computing, are simple to implement, are insensitive to outliers and contain nonlinearities. We describe below the most suitable method based on neural networks to quantify the uranium content.

2. Experimental studies

The efficiency response for the quantification of uranium is difficult to establish due to the dearth of peaks that can be used. This problem can be minimised by reducing the region of spectral interest to the relatively complex so that the $K_\alpha X$ energy range from 83 to 103 keV (Fig. 1).

This region contains enough information to allow the determination of ^{235}U and ^{238}U and is sufficiently small for the efficiency to be defined as constant. It is however very complex to analyze, due to several interfering X- and γ-rays that can be grouped as follows:

– ^{235}U and daughters: 84.21 keV (γ ^{231}Th), 89.95 keV (γ

* Corresponding author. Tel. +33 1 69 08 94 01, fax +33 1 69 08 98 81, e-mail vvigne@soleil.serma.cea.fr.

SSDI 0168-9002(95)00964-7

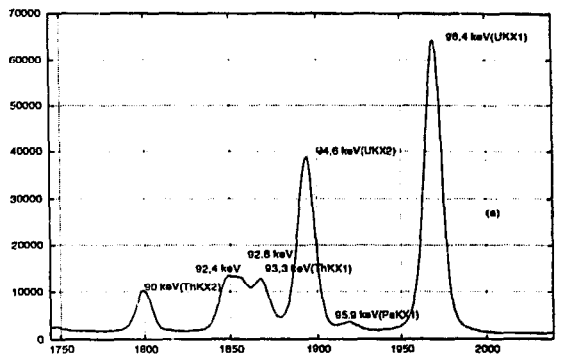

Fig. 1. X- and γ-rays in the spectral analysis of the $K_{\alpha}X$ region.

^{231}Th, Th $K_{\alpha_2}X$), 92.28 keV (Pa $K_{\alpha_2}X$), 93.35 keV (Th $K_{\alpha_1}X$), 95.86 keV (Pa $K_{\alpha_1}X$).

- ^{238}U and daughters: 83.30 keV (γ ^{234}Th), 92.28 keV (Pa $K_{\alpha_2}X$), 92.38 keV (γ ^{234}Th), 92.79 keV (γ ^{234}Th), 94.65 keV (U $K_{\alpha_2}X$), 95.86 keV (Pa $K_{\alpha_1}X$), 98.43 keV (U $K_{\alpha_1}X$), 99.85 keV (γ ^{234}Pa).
- Uranium X-ray fluorescence: 94.65 keV ($K_{\alpha_2}X$), 98.43 keV ($K_{\alpha_1}X$).

The spectral processing of this region by the standard approach takes into account three groupings: to ^{234}U and his daughters, ^{238}U and its daughters and the uranium X-ray fluoresence spectrum. These spectral emissions are represented by mathematical expressions taking into account the shapes of the X- (Voigt profile) and γ-ray (Gaussian) peaks, their energies, and emission probabilities. A conventional least squares method is used such as the MGA-U code [7]. The enrichment is fully determined by correcting for the presence of ^{234}U using the 120.9 keV peak.

Six infinitely-thick uranium oxide standards with different enrichments (from 0.7 to 9.6%) were counted several times by γ-ray spectrometry to test the neural procedure. These samples were bare cylindrical pellets with certified enrichments, and their main characteristics are listed in Table 1.

The Ge(HP) planar detector used in the measurement system had the following specifications: surface area of 2.00 cm^2, thickness of 1.00 cm, and FWHM of 190 eV at 6 keV and 480 eV at 122 keV. All the measurements were made under the same conditions, i.e. reduction of 0.05 keV per channel and a distance between the source and detector-window of 1.1 cm. Ten 20 000 s spectra for each standard pellet were analysed. The ^{234}U concentration was relatively low, although a ^{233}U/^{235}U mass ratio varying from 0.5 to 1.1% (depending on the pellet) was determined by γ-ray spectrometry using both the 53.2 and 120.9 keV peaks for ^{234}U and the 185.7 keV peak for ^{235}U.

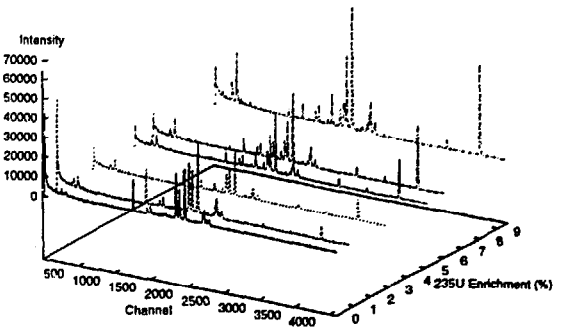

Fig. 2. 3D-Representation of the UO_2 spectra.

65 sets of experimental data were obtained using the concentrations given in Table 1, and shown in Fig. 2.

3. Layered neural network and training method

3.1. Neural networks

Neural Networks are nonlinear black-box model structures that can be used with conventional parameter estimation methods. Their details and basic concepts are clearly described in a paper to be published [8]. ANN consists of a large number of neurons, i.e. simple linear or nonlinear computing elements interconnected in complex ways and often organized into layers [9]. The collective or parallel behaviour of the network is determined by the way in which the nodes are connected and the relative type and

Table 1
Characteristics of UO_2 standards

Diameter [cm] × Height [cm]	$\frac{U}{O}$ ratio [g g^{-1}%]	Stated enrichment [g g^{-1}%]	$\frac{^{235}U}{^{235}U + ^{238}U}$ ratio [g g^{-1}%]
1.30 × 2.00	88.00	0.7112±0.004	0.7112
1.30 × 1.90	88.00	1.416±0.001	1.416
0.80 × 1.10	88.00	2.785±0.004	2.786
0.80 × 1.02	87.96	5.111±0.015	5.112
0.80 × 1.00	87.98	6.222±0.018	6.225
0.92 × 1.35	87.90	9.548±0.04	9.558

strength (excitory or inhibitory) of the interactions among them [10].

The objective of ANNs is to construct a suitable model which, when applied to a ^{235}U enrichment spectrum, produces an output (y) which approximates the exact uranium enrichment ratio. A connectionist approach is adopted to substitute a neural model and the learning procedure of the network for classic mathematical algorithms, that are based on the separation of a given curve for each individual peak and the background.

An example of a multi-layer network is given in Fig. 3a. The notation convention is such that the square represents a computational unit into which the input variables (x_j) are fed and multiplied by a respective weights (ω_j). The fundamental processing element of an ANN is a node (Fig. 3b), which is analogous to neurons in biological systems. Each node has a series of weighted inputs, ω_i, which may be either an external signal or the output from other nodes. The sum of the weighted inputs is transformed with a linear or a nonlinear transformation function (often the logistic function $f(x) = 1/(1 + e^{-x})$. This standard Neural Network is called Multi-Layered Perceptron (MLP, an dis analogous to the Multivariate Nonlinear Regression.

Transmission of information between units of two neighboring layers is performed through oriented links involving connection weights. The construction is as follows:

- input layer: each input unit receives input-variables, selected through a free parameters reduction procedure.
- hidden layer: acts as an array of feature detectors that pick up features without regard to position. Information is fed to the input units is coded on the hidden layer into an internal representation so that these units contribute to the input of second-layer unit. The hidden-layer is fully-connected to the output.
- output layer: applies a sigmoïd activation function to the weighted sum of the hidden outputs.

The role of the hidden layer is fundamental: a network without hidden units will be unable to perform the necessary multi-input/multi-output mappings, particularly with non-linear problems. An input pattern can always be encoded if there are enough hidden units, so that the appropriate output pattern.

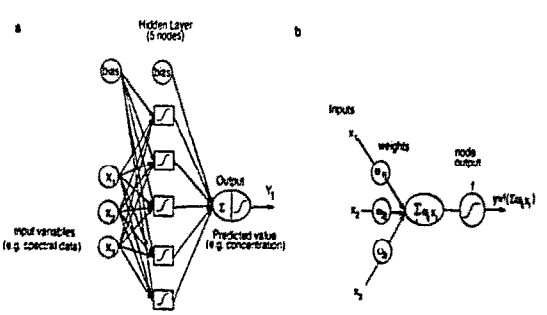

Fig. 3. (a) MLP 3-5-1 with nonlinear threshold and (b) node in an ANN.

The training data are denoted by $\chi = (x, y^d)_{t=1}^N$ where N is the number observations and x is the feature vector corresponding to the tth observation. The expected response $y^d = (y_1, \ldots, y_M)$ is related to the inputs $x = (x_1, x_2, \ldots, x_N)$ according to

$$y = \phi(x, \omega), \tag{1}$$

where ω are the *connection weights*.

The approximation results are non-constructive, and the weights have to be chosen to minimize some fitting criterion, e.g. least squares:

$$J(\omega) = \frac{1}{2}\sum_{p}^{N} (y_p^d - \phi(x_p, \omega))^2, \tag{2}$$

with respect to all the parameters, where y_p^d is the target for the pth example pattern. The minimization has to be done by a numerical search procedure called *nonlinear optimization* in which the parameter estimate is defined as the minimizing argument:

$$\hat{\omega} = \text{argmin}_{\omega}\, J(\omega). \tag{3}$$

Most efficient search routines are based on local iteration along a "downhill" direction from the current point. This method involves an iterative scheme defined by:

$$\hat{\omega}^{(i+1)} \leftarrow \hat{\omega}^{(i)} - \eta \frac{\partial J}{\partial \omega^{(i)}} \tag{4}$$

where $\hat{\omega}^{(i)}$ is the parameter estimate after iteration i, $\eta(>0)$ is the step size, and $\partial J/\partial \omega^{(i)}$ is an estimate of the gradient of $j(\omega^i)$. The practical difference between this procedure and the statistical approach lies in the way the training data are used to dictate the values of ω. There are 2 main features to the processing: (1) specifying the architecture of a suitable network, (2) training the network to perform well with reference to a training set.

3.2. Application of ANN

The ANN method has been applied to 65 sets of experimental data: five pure ^{235}U spectra, and ten of each standard (see Table 1). Each spectrum contains 4096 data points. Computations of the spectra are compared using two regression models: MLP MODEL (Fig. 3) in which the inputs are spectra data, and MIXTURES OF EXPERTS MODEL (Fig. 4) [11] where the inputs are the enrichment values. The network specifications for the networks created for the calibration of the simulated data are listed in Table 2 and they were found to be optimal for low prediction bias [8]. While the choice of the right architecture is mainly intuitive and implies arbitrary decisions, an attempt to apply ANN directly fails due to dimensionality. Therefore, the dimension of the input vector has been reduced dramatically by Principle Components Analysis (PCA),

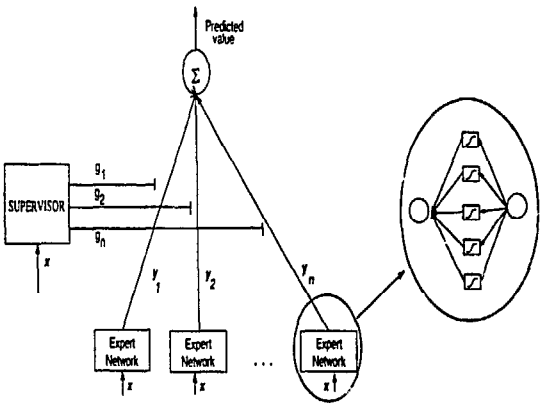

Fig. 4. Mixtures of Experts model.

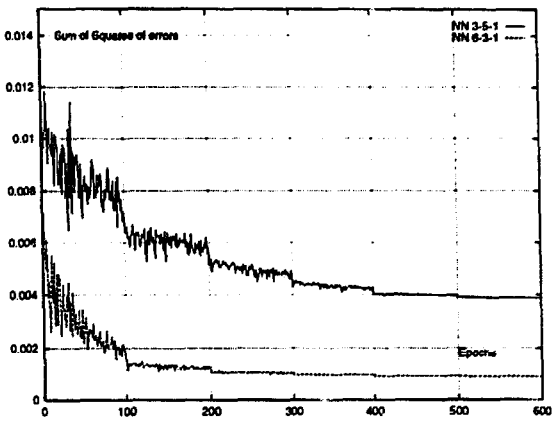

Fig. 5. Sum of squares of bias on the training set for MLP architectures.

leading to an adequate reduction of weights from the first layer of the ANN.

The MLP MODEL (Fig. 3), consists of an input layer of 6 or 3 units leading up through one layer of hidden units to an output layer of a single unit that corresponds to the desired enrichment. This network represents a poor parametrized model, but the training dataset $(x, y^{(d)})_{i=1}^{65}$ is small. The network is initialized with random weights and trained, and the bias is evaluated for each pattern (Eq. (2)). This quantity decreases rapidly at the beginning (Fig. 5), and the training is stopped to avoid overfitting when the network reaches a minimum error. After 32000 successful training passes, the bias rate ranged form −0.05 to 0.04% for the 6-3-1 net (from −0.031 to 0.061% for the 3-5-1 net).

Each item in the MIXTURES OF EXPERTS MODEL (MEX) is associated with a vector of measurable features, and a target y^d which represents the enrichment. The network receives the input x and creates the output vector y as a predicted value of the unknown y^d. The model consists of 210 independent, fully-connected networks (Fig. 4): one expert is assigned to a channel of the $K_\alpha X$ region, each expert being an observer trying to identity a "signal" due to radioactive decay in a large amount of noise. The variance of each count is proportional to the enrichment of a particular sample and the background level of the particular observation. A cooperation–competition procedure driven by a supervisor between the expert outputs leads to the choice of the most appropriate concentration.

Let $y_1, y_2, \ldots$ denote the output vectors of the experts, and $g_1, g_2, \ldots$ represent the supervisor output units. The output of the entire architecture is $y = \Sigma_{i=1}^{210} g_i y_i$ and the supervisor decides whether expert i is applicable or not; the winning expert is the network with the smallest bias $(y^d - y_i)$.

4. Discussion

The initial base included only 65 data sets, and we wished to keep a maximum of these example data for the training base. Redundancies in the data-set enrichments present one major advantage: since we measure more than one response for each case, information from all of the responses can be combined to provide more precise parameter estimations and determine a more realistic model.

The measure of system performance in MEX-simulations is the *cross-entropy* error (according to the Poissonnian error model) and the mean square error with MLP.

Table 2
ANNs specification and parameters

Parameter	MLP 6-3-1	MLP 3-5-1	Mixtures of experts
Type of input	Spectral data	Spectral data	Enrichment value
Input nodes	6	3	1
Hidden node	3	5	1050
Output node	1	1	210
Learning rule	BP	BP	Maximum Likelihood
Input layer transfer function	Linear	Linear	Linear
Hidden layer transfer function	Sigmoïdal	Sigmoïdal	Sigmoïdal
Output layer transfer function	Linear	Linear	Exponential

Table 3
Ranges of calculated Enrichments with MLP and MEX

Declared enrichment [5]	MLP 3-5-1	MLP 6-3-1	MEXs
0.711	0.691–0.723	0.700–0.720	0.702–0.710
1.416	1.394–1.426	1.406–1.435	1.406–1.416
2.785	2.732–2.822	2.762–2.799	2.784–2.790
5.111	5.066–5.148	5.089–5.132	5.112–5.136
6.122	6.105–6.162	6.117–6.133	6.088–6.112
9.548	9.531–9.570	9.541–9.550	9.542–9.552

Bias rates obtained by MEX are benchmarked against the results obtained by MLP in Table 3 and Figs. 5 and 7. Fig. 5 shows the learning curves (i.e. learning performances) for the two MLP networks using a random training procedure. The horizontal axis gives the number of iterations and the vertical axis represents the mean square errors value (MSE). Clearly, the 6-3-1 network learned significantly faster than the 3-5-1 network: this difference can be explained by the information gain of the 6-input compared with the 3-input network.

Fig. 6 shows the predicted enrichment values (one of each of the 210 experts) when a 5.111%–^{235}U spectrum is analysed by the MEX model. The final predicted values of the simulations are listed in the right-hand column of Table 3. Compared with the MLP, the MEX method is shown to be extremely reliable: for example, the bias between the predicted and calculated 2.785% enrichments ranges from 2.784 to 2.790%. As noted above, after 32 000 successful training passes, the greatest bias occurs for 5.111 and 6.122% enrichments; This relative lack of precision can be ascribed to the small size of the training dataset.

Fig. 7 compares the results of the three models, in which the bias between the predicted and desired enrichments is plotted for each of the 65 samples. The results suggest that the strong dispersion of the bias when using MLP is significantly attenuated when MEX is applied, although this observation must be moderated for the samples with the 6.122-enrichment ratio. A comparison of the absolute bias curves suggest that, the Mixtures of Experts gives the most robust performance.

The modular approach has three main advantages over the MLP models: able to model the observed behaviour, learns faster than a global model and the representation is easier to interpret. Modular architecture takes advantage of task decomposition, but the learner must decide which variables to allocate to the networks. No hypothesis is made with respect to any aspect of the spectra, including the extent of spectra resolution, nature of the features being analysed, and whether you select the most significant areas of the spectrum only (MLP models) or a significant fraction of the full spectrum (MEX model). At the same time, the method is highly specific because the ANN must learn to recognise representative spectra of the radionuclides to be identified. Furthermore, other tests have revealed to us that ANNs are resistant to noise, although this observation may arise from the extremely small size of the training dataset.

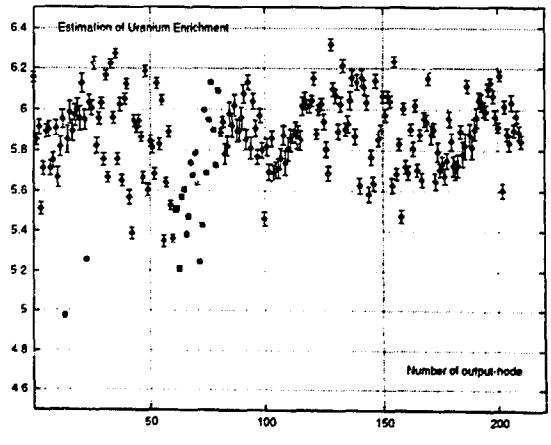

Fig. 6. Example of enrichment value (at 5.785%) predicted by the Mixtures of Experts.

5. Conclusions

The simulation studies of real UO_2 photon spectra have shown that Neural Networks can be very effective in predicting ^{235}U enrichment. This approach appears to be useful when a fast response is required with reasonable accuracy, no spectral hypothesis are made, and no defini-

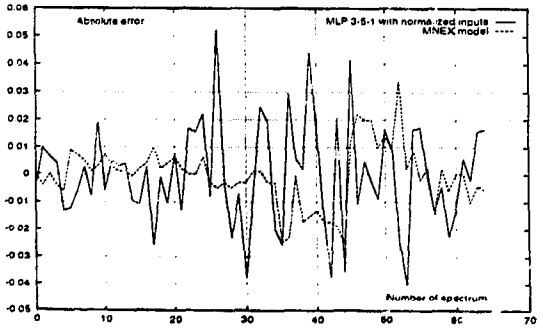

Fig. 7. Absolute bias in the enrichmen: estimation.

tive mathematical model can be assigned a priori. The resistance to noise is certainly one of the most powerful characteristics of this type of analysis. A suitable network with connections and weighting functions could be easily implemented using commercial data processing hardware. The good results show that this type of analysis can be considered the most appropriate method for the production of quantitative estimates of the concentration of radionucleidic components in mixtures under well-defined experimental conditions: the resulting data may be better than those obtained when using standard methods. The Neural Network method of analysis has also been successfully used in X-ray fluorescence studies [12].

There is no single learning procedure which is appropriate for all tasks. It is of fundamental importance the the special requirements of each task are analyzed an that appropriate training algorithms are developed for families of tasks. The efficient use of neural networks requires extremely careful analysis of the problem, an analysis which can often be neglected by impatient users.

Acknowledgements

The authors are extremely grateful to the staff of SAR and LPRI, and in particular J.L. Boutaine, A.C. Simon and R. Junca for their support and useful discussions during this work. V. Vigneron wishes to thank C. Fuche from CREL for her support during this research.

References

[1] T.D. Reilly, R.B. Walton and J.L. Parker, Report LA-4065-MS, Los Alamos (1970).
[2] L.A. Kull, R.O. Gonaven and G.E. Glancy, Atomic Energy Rev. 144 (1976) 681.
[3] J. Morel, H. Goenvec, J. Dalmazzone and G. Malet, IAEA Nuclear Safeguard Technology (IAEA, 1978).
[4] P. Matussek, Report KfK-3752, Karlsruhe (1985).
[5] R.J.S. Harry, J.K. Aaldijk and J.P. Brook, Report SM201/6 (IAEA, Viena, 1980).
[6] R. Hagenauer, Nucl. Mater. Manag. Proc., 1991.
[7] R. Gunning, W.D. Ruther, P. Miller, J. Goerten, M. Swinhoe, H. Wagner, J. Verplancke, M. Bickel and S. Abousahl, Report UCRL-JC-11114713 (Lawrence Livermore National Laboratory, 1994).
[8] V. Vigneron and J.M. Martinez, submitted to Nucl. Instr. and Meth. (1995).
[9] Y. LeCun, Technical report, ATT & Bell Laboratories, 1993.
[10] J.M. Martinez, C. Parey and M. Houkari, Proc. 5th Int. Neuro-Nimes, Paris, 1992, CEA (Saclay), EC2, pp. 431–442.
[11] M.I. Jordan, Summer school, CEA-INRIA-EDF, MIT, 1994.
[12] A.C. Simon and R. Junca, in: Note technique DTA/DAMRI/SAR/S/94-433, Saclay, 1994.