Reconstruction problem by maximum entropy method applied on a mixture of experts

Vincent Vigneron, Jean-Marc Martinez, Marie-Christine Lepy, Jean Morel

Reconstruction problem by maximum entropy method

V. Vigneron[*,1,2], **J. Morel**[3], **M.C. Lépy**[3], **J.M. Martinez**[1]

[1]CEA Saclay
DRN/DMT/SERMA
91191 Gif-sur-Yvette
FRANCE

[3]CREL
161, rue de versailles
78180 Le Chesnay
FRANCE

[2]CEA Saclay
DAMRI/LPRI, BP 52
91193 Gif-sur-Yvette cedex
FRANCE

Abstract

Layered Neural Networks, which are a class of models based on neural computation , are applied to the measurement of uranium enrichment. The usual methods consider a limited number of γ-ray and X-ray peaks, and require previously calibrated instrumentation for each sample. But since, in practice, the source-detector ensemble geometry conditions are critically different, a mean of improving the above conventional methods is to reduce the region of interest ; this is possible by focusing on the $K_\alpha X$ region where the three elementary components are present. The measurement of these components in mixtures leads to the desired ratio. Real data are used to study the performance of neural networks and training is done with a Maximum Likelihood Method. We show that the encoding of data by Neural Networks is a promising method to measure uranium ^{235}U and ^{238}U quantities in infinitely thick samples.

1 Introduction

In the past few years, the topic of neural computing has generated widespread interest and popularity. The popularity of this technique is due in part to the analogy between Artificial Neural Networks (ANNs) and biological neural networks. Many applications have been investigated using ANNs. We

[0] *Corresponding author*: e-mail `vvigne@soleil.serma.cea.fr`.

demonstrate here how they can be used, with photons spectra, in the particular case of uranium enrichment measurements, to determine the $\frac{^{235}U}{U_{total}}$ isotope ratio. Indeed, with modern detectors and high technology, spectral data are collected with even finer sampling and with even large precision, what imposes a perpetual need for efficient interpretation methods.

Traditional non-destructive methods for uranium enrichment use several X- and γ-ray peaks, mainly in the 60 to 200 keV region. Most of these methods, which were developed more than 20 years ago, are based on measurements of the full energy peak at 185,7 keV ([10],[5],[9],[8]). They require a prior calibration of the system and the measurement conditions to be constant. Other methods have been developed using several γ-ray peaks [3],[2]. In fact, these latter methods require a self-calibration with a limited number of peaks, making them difficult to implement.

Calibration procedures and matrix effects can be avoided by first, focusing the spectra analysis on a limited region, called $K_\alpha X$ region , containing the main uranium components, second, by using so-called infinitely thick samples. These samples are such that any further thickness increase does not affect the γ emission.

The processing of the $K_\alpha X$ region requires taking into account 3 elemental images corresponding to ^{235}U, ^{238}U and X-ray fluorescence. This approach requires that all the parameters for constituting each elemental image are well-known and is based on the use of external data characterising the photon spectral emission together with the detector characteristics and geometry.

It is precisely in this context that a Neural Network appears to be a useful tool. In fact, the training by ANNs can be considered as a search procedure for an "optimum" regression function among a set of acceptable functions using a set of training examples. From the statictical point of view, ANNs belong to the evaluation techniques for non-parametric models, still called *tabula rasa*. ANNs, like most statistical methods, are able to process vast amounts of data and to make predictions that are sometimes surprisingly accurate. This does not make them intelligent in the usual sense of the word. ANNs learn in much the same way that many statistical algorithms do estimation. But in contrast to usual automatic spectra analysis methods, ANNs use full-parallel computing, are simple to implement, not very sensitive to outliers and contain nonlinearities.

In the following, we describe the identification method based on neural networks to quantify uranium quantities. Section II covers the experimental procedure and the neural networks technique is explained in section III. Finally, Section IV gives the outlook and conclusion.

2 Experimental Aspect

2.1 Preliminaries

In the case of uranium spectra, the efficiency response is difficult to establish due to insufficient number of peaks that can be used. This can be overcome by reducing the region of interest of the spectrum so that the variation in the detector efficiency is limited. This is possible by considering only the relatively complex $K_\alpha X$ region, which extends from 83 to 103 keV, where many peaks are superimposed (Fig. 1).

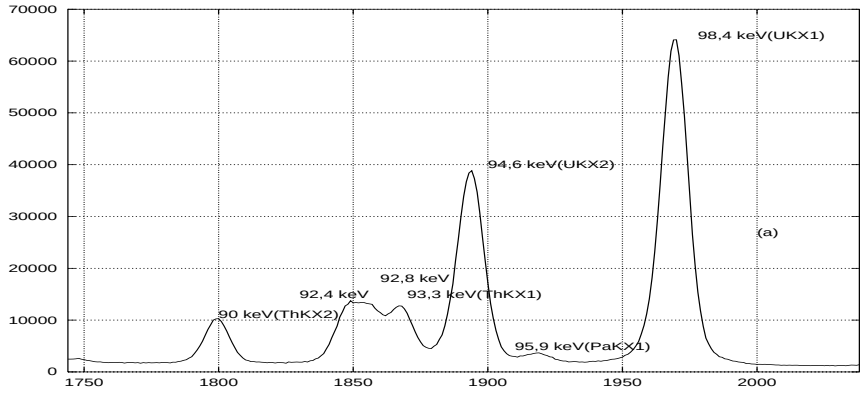

Figure 1: Principal useful $X-$ and $\gamma-$rays in the spectral analysis of the $K_\alpha X$ region.

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

- ^{235}U and daughters : 84.21 keV (γ^{231}Th), 89.95 keV (γ^{231}Th, $ThK_{\alpha_2}X$), 92.28 keV ($PaK_{\alpha_2}X$), 93.35 keV ($ThK_{\alpha_1}X$), 95.86 keV ($PaK_{\alpha_1}X$)
- ^{238}U and daughters : 83.30 keV (γ^{234}Th), 92.28 keV ($PaK_{\alpha_2}X$), 92.38 keV (γ^{234}Th), 92.79 keV (γ^{234}Th), 94.65 keV ($UK_{\alpha_2}X$), 95.86 keV ($PaK_{\alpha_1}X$), 98.43 keV ($UK_{\alpha_1}X$), 99.85 keV (γ^{234}Pa)
- Uranium X-ray fluorescence : 94.65 keV ($K_{\alpha_2}X$), 98.43 keV($K_{\alpha_1}X$).

In the standard approach, the processing of the considered region takes into account the 3 elemental images, the first corresponding to ^{235}U and his daughters, the second to ^{238}U and its daughters and the third to the uranium X-ray fluoresence spectrum. These images are then represented by mathematical expressions taking into account the shapes of the X-ray (Voigt profile) and γ-ray

(Gaussian) peaks, their energies, and intensities. The determination is then carried out conventionally with a least squares method, like the MGA-U code [1]. The final enrichment is obtained by correcting for the presence of ^{234}U with the 120.9 keV peak.

2.2 Experimental protocol

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

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

The Ge(HP) planar detector used in the measurement system had the following specification : surface, 2.00 cm^2 ; thickness, 1.00 cm ; FWHM, 190 eV at 6 keV and 480 eV at 122 keV. All the measurements were made under the same conditions, i.e. with 0.05 keV per channel and a distance between source and detector-window of 1.1 cm. Ten 20000-s. spectra for each standard pellet were analysed by our procedure. The ^{234}U concentration is relatively low, although a $\frac{^{234}U}{^{235}U}$ mass ratio varying from 0,5 to 1,1%, depending on the pellet, was determined by γ-ray spectrometry by using both the 53.2 and the 120.9 keV peaks for ^{234}U and the 185.7 keV peak for ^{235}U.

In short, 65 sets of experimental data from real-life experiments were prepared using the concentrations given in Table 1, and are illustrated in Fig. 2.

3 Layered Neural Network and Training method

3.1 Using Neural Networks

The purpose of this section, rather than the presentation of the neural network theory, is to present the place of the connectionnist approach in γ-spectrometry

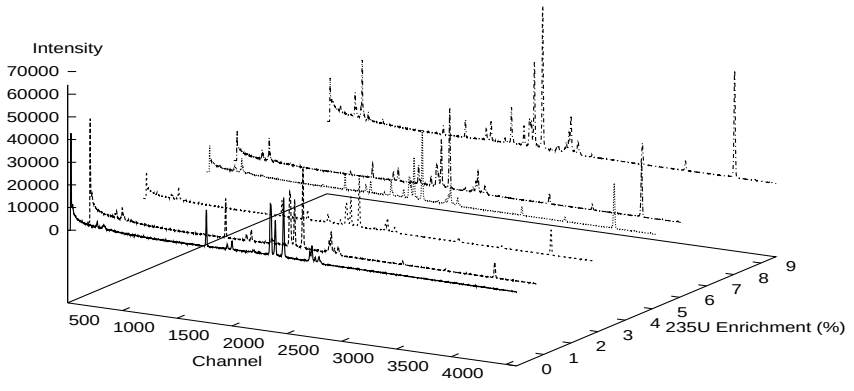

Figure 2: 3D-Representation of the UO_2 spectra set.

problems. Neural Networks are non-linear black-box model structures, to be used with conventional parameter estimation methods. Most details and basic concepts are clearly described in a paper to be published [12]. 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 [6]. The collective or parallel behaviour of the network is determined by the way in which the nodes are connected and the relative type and strengh (excitory or inhibitory) of the interactions among them [7].

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. The principal idea of the connectionist approach is to substitute a neural model and the learning procedure of the network for classical fitting algorithms, which make use of complex mathematical algorithms, generally based on the separation of a given curve, associated to each individual peak, plus a background.

An exemple of multi-layer network is given in Fig. 3.a. The notation convention is such that the square represents a computational unit into which the input variables x_j's are fed and multiplied by the respective weights ω_j's. The fundamental processing element of an ANN is a node (Fig. 3.b). Nodes are 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 non-linear transformation function (often the logistic function $f(x) = \frac{1}{1+e^{-x}}$) . In the statistical context, this standard Neural Network called *M*ulti-*L*ayered *P*erceptron (*MLP*), is analogous to the Multivariate Nonlinear Regression.

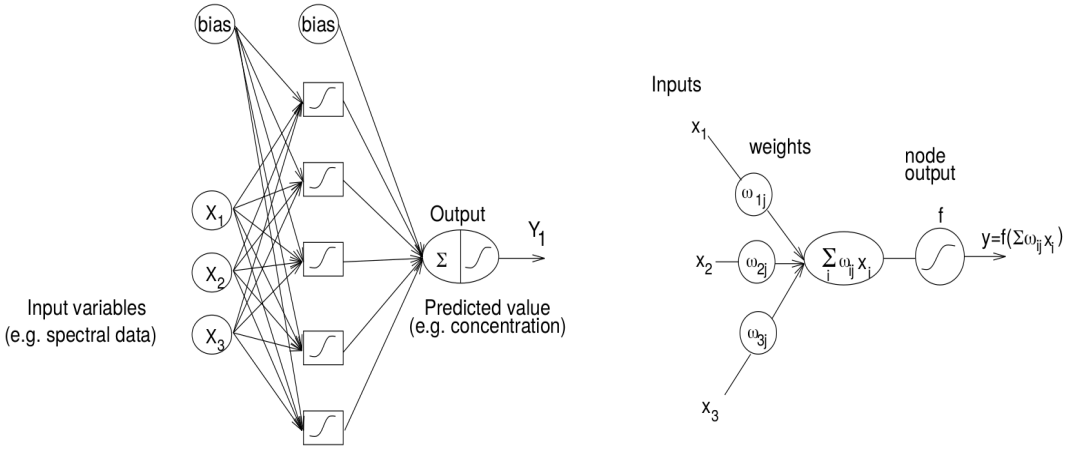

Figure 3: (a) MLP 3-5-1 with nonlinear threshold and (b) schematic representation of a node in an ANN.

Transmission of information between units of two neighboring layers is performed through oriented links. These links are level-headed by connection weights. The essential of the construction is as follows :

- input layer : this layer contains input units. Each unit receives input-variables, selected through a free parameters reduction procedure.
- hidden layer : this layer acts as an array of feature detectors picking up features without regard to position. The information coming to the input units is coded on the hidden layer into an internal representation Thus, the input-layer units contribute to the input of each second-layer unit. It is fully-connected to the output.
- output layer : it 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, in particular with non-linear problems. Input pattern can always be encoded, if there are enough hidden units, in a form so that the appropriate output pattern can be generated from the corresponding input pattern.

The training data are denoted by $\chi = (\mathbf{x}, \mathbf{y^d})_{t=1}^{N}$ where N is the number of observations and $\mathbf{x}$ is the feature vector corresponding to the t^{th} observation. The expected response $\mathbf{y^d} = (y_1, y_2, \ldots y_M)$ is related to the inputs $\mathbf{x} = (x_1, x_2, \ldots x_N)$ according to

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

where ω are the *connection weights*.

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

$$J(\omega) = \frac{1}{2}\sum_{p}^{N}(\mathbf{y}_p^{\mathbf{d}} - \phi(\mathbf{x}_p, \omega))^2, \tag{2}$$

with respect to all the parameters, where $\mathbf{y}_p^{\mathbf{d}}$ is the target for the pth example pattern. The minimization has to be done by some numerical search procedure. This is called *nonlinear optimization*. The parameter estimate is defined as the minimizing argument :

$$\hat{\omega} = argmin_{\omega} J(\omega) \tag{3}$$

Most efficient search routines are based on local iteration along a "downhill" direction from the current point. We then have an iterative scheme of the following kind :

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

where $\hat{\omega}^{(i)}$ is the parameter estimate after iteration number i, $\eta(> 0)$ is the step size and $\frac{\partial J}{\partial \omega^{(i)}}$ an estimate of the gradient of $J(\omega^i)$. The practical difference between this device and the statistical version lies in the way the training data are used to dictate the values for ω. It turns out that there are 2 main aspects 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 the ANN

To check that this method was general and reliable, we have applied it to 65 sets of experimental data from real-life experiment : five ^{235}U-pure idealized spectra, and ten of each precited standard (see Table 1). Each spectrum contains 4096 points. The computations of the spectra are compared on two regression models: the MLP MODEL (Fig. 3), where the inputs are spectral data, and the MIXTURES OF EXPERTS MODEL (Fig. 4) [4] where the inputs are the enrichment values.

The specifications for the networks created for the calibration of the simulated data are listed in Table 2. They were found to be optimal according to the rigourous methodology described in [12], for low prediction bias. The choice of the right architecture is mainly intuitive and implies arbitrary decisions. But an attempt to apply ANN directly fails due to dimensionability. In acccordance with this, the dimension of the input vector has been reduced dramatically by Principle Components Analysis (PCA), leading to the adequate

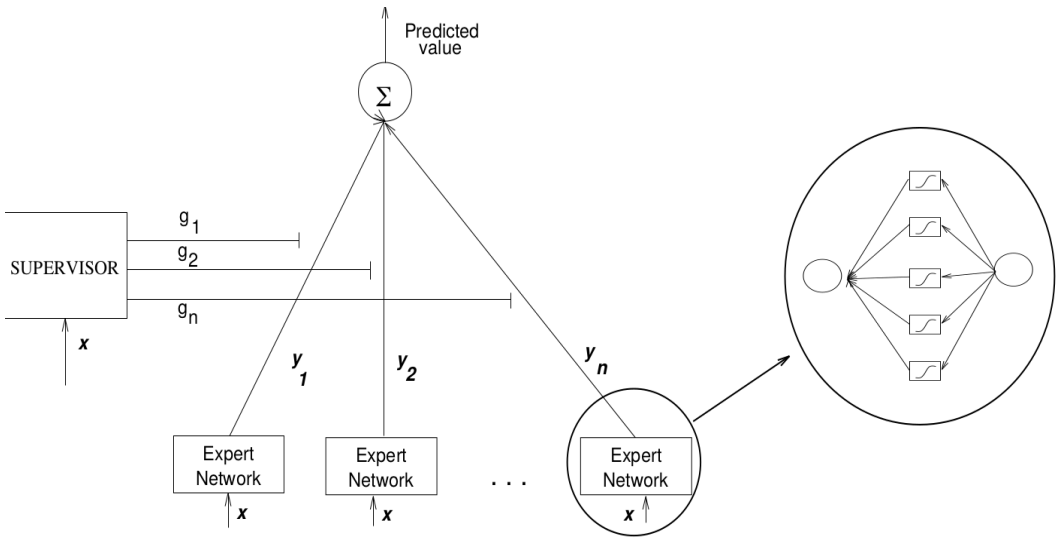

Figure 4: Mixtures of Experts model.

reduction of weights emerging from the first layer of the ANN.

Table 2: ANNs specifications 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

The MLP MODEL, depicted in 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 $(\mathbf{x}; \mathbf{y}^{(\mathbf{d})})_{t=1}^{65}$ was small. The network is initialized with random weights and trained. For each pattern, the bias, Eq. 2, is evaluated. This quantity decreases rapidly (Fig. 5) in the beginning, and the training is stopped when the network reaches a minimum error on the training set, because this is an efficient way to avoid overfitting. After 32 000 successful training passes, the bias rate range from -0.05 to 0.04% for the 6-3-1 net (from -0.031 to 0.061% for the 3-5-1 net).

In the case of mixtures of experts (MEX), each item is associated with a vector of measurable features, and a target $\mathbf{y}^{\mathbf{d}}$ which represents the enrichment. The network receives the input $\mathbf{x}$ and creates the output vector $\mathbf{y}$ as

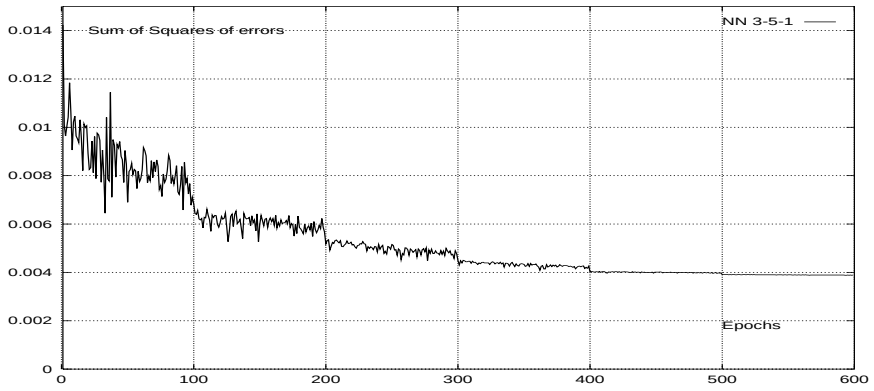

Figure 5: Sum of squares of bias on the training set for MLP architectures

a predicted value of the unknown $\mathbf{y^d}$. This model consists of 210 independant fully-connected networks (Fig. 4) : One expert is put for one channel of the $K_\alpha X$ region, each expert being an observer, trying to find a "signal" due to radioactive decay in a large amount of noise, the variance of each count being proportional to the level and thus depending on the enrichment of a particular sample and on the background level of the particular observation. A cooperation-competition procedure driven by a supervisor between the expert's outputs leads to the choice of the most appropriate concentration.

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

4 Discussion of the Results using ANN

As the initial base included only 65 examples, we wanted to keep a maximum of examples for the training base. Redundances in the data-set enrichments present one main advantage : as we measure more than one response for each case, information from all the measured responses can be combined to provide more precise parameter estimation and to determine a more realistic model.

In all simulations, the measure of the system's performance is the Mean Square Error. The bias rates obtained by using MEX are benchmarked against the results obtained by using MLPs in Table 3 and on Fig. 5 and Fig. 7. The Fig. 5 shows the learning curves (i.e. the learning performances) for the two MLP networks using a random training procedure. The horizontal axis gives the number of epochs ; the vertical axis gives the Mean Square Errors value (MSE). Clearly, the 6-3-1 network learned significantly faster than the 3-5-

1. This difference can be explained by the information gain of the 6-inputs network vs the 3-inputs network.

Table 3: Ranges of calculated Enrichments with MLP and MEX

Declared enrichment	**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

The Fig. 6 concerns the Multi-Expert model. The plotted points are predicted enrichment value (one for each of the 210 experts) when a $5.111\% - {}^{235}U$ spectrum is presented to the MEX model. The credit assignement procedure on these 210 contributions is supervised to produce a final estimation. In the right most column of Table 3, the final predicted values of the simulations with MEX can be seen. Compared with the MLPs, this shows that MEX method is really reliable ; for example, the bias between the predicted and the calculated 2.785% enrichments range from 2.784 to 2.790%. As noted above, after 32 000 successful training passes, the larger bias happens for 5.111 and 6.122% enrichments. This relative lack of precision can be ascribed to the small size of the training dataset.

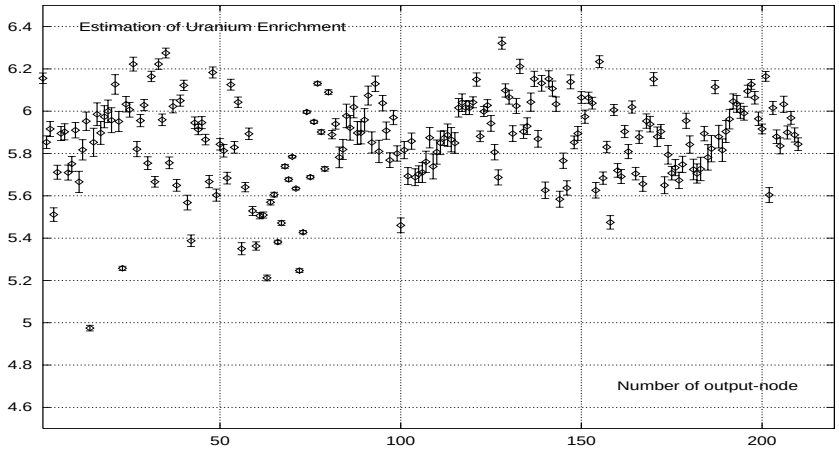

Figure 6: Example of enrichment value (at 5,785 %) predicted by the Mixtures of Experts

Fig. 7 compares the results of the three models. The bias between the predicted and the desired enrichments is plotted for each of the 65 samples.

The darkest line is put for the MEX. The results suggest that the strong dispersion of the bias with MLP is significantly attenuated when MEX is applied. This judgement must be moderated for the 6.122-enrichment-ratio samples. A comparison of the absolute bias curves suggest that, among of the three systems studied, the Mixtures of Experts is capable of showing the most robust performance.

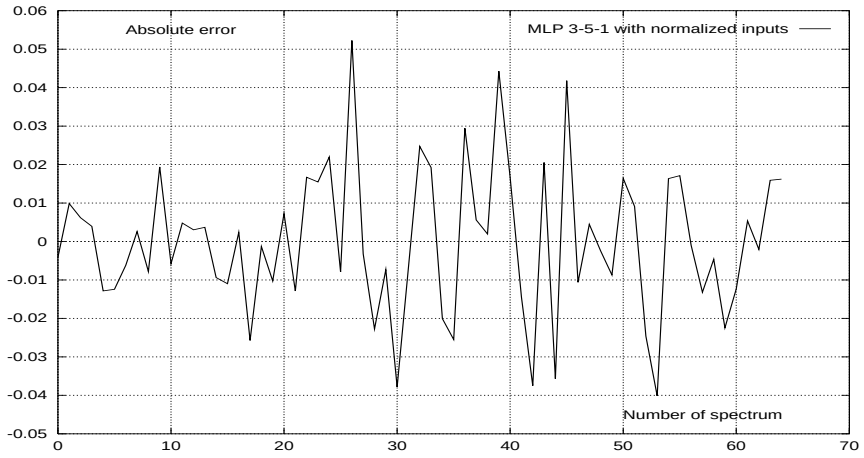

Figure 7: Absolute bias in the enrichment estimation

In fact, the modular approach presents three main advantages on the MLP models: it is able to model behavior, it learns faster than a global model and the representation is easier to interpret. The modular architecture takes advantage of task decomposition, but the learner must decide which variables to allocate to the networks. This method is, at the same time, very general and very specific. It is very general in the sense that no hypothesis is made on the aspect of the spectra : it does not depend on whether the spectra are well resolved or not, whether they are very likely or not, whether you select most significative areas of spectrum only (MLP models) or a global part of the spectrum (MEX model). But, at the same time, the method is very specific because the ANN must learn representative spectra of the family spectra to identify. Furthermore, other tests proved to us that ANNs are resistant to noise. Presently, we must put the blame on the excessively short size of the training dataset.

5 Conclusion

The simulation studies on UO_2 real spectra have shown that Neural Networks can be very effective to predict ^{235}U enrichment. They appear to be useful when a fast response is needed with a reasonnable accuracy, when no hypoth-

esis is made on the aspect of the spectra or when no definite mathematical model can be assigned *a priori*. The resistance to noise is certainly one of the most powerful characteristics of this method. Final network with connections and weighting functions could be easily implemented using commercial digital processing hardware. The good results obtained show that this method can be considered at the state of art to produce quantitative estimates of the concentrations of isotopic components in mixtures with fixed experimental conditions : they may be better than those obtained with standard methods in similar cases. This method has also been already successfully used in an X-ray fluorescence application [11].

There is no single learning procedure which is appropriate for all tasks. It is of fundamental importance that special requirements of each task are analyzed and that appropriate training algorithms are developed for families of tasks. However, an efficient use of the networks requires as careful as possible analysis of the problem, an analysis that is often ignored by impatient users.

Acknowlegments

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 in this research.

References

[1] Gunning, R., Ruther, W. D., Miller, P., Goerten, J., Swinhoe, M., Wagner, H., Verplancke, J., Bickel, M., and Abousahl, S. MGAU: A new analysis code for measuring ^{235}U enrichments in arbitrary samples. Report UCRL-JC-114713, Lawrence Livermore National Laboratory, (Livermore), 1994.

[2] Hagenauer, R. Non destructive uranium enrichment determination in process holdup deposits. In *Nucl. Mater. Manag. Proc* (1991).

[3] Harry, R. J. S., Aaldijk, J. K., and Brook, J. P. Gamma spectrometric determination of isotopic composition without use of standards. In *Report SM201/6 IAEA, Viena* (1980).

[4] Jordan, M. I. *Lectures on neural networks.* Summer school, CEA-INRIA-EDF, MIT, 1994. Private Communication.

[5] Kull, L. A., Gonaven, R. O., and Glancy, G. E. A simple γ spectrometric technique for measuring isotopic abundances in nuclear materials. *Atomic Energy Review 144* (1976), 681–689.

[6] LeCun, Y. Efficient learning and second-order methods. Tech. rep., ATT & Bell Laboratories, 1993. Private Communication.

[7] Martinez, J. M., Parey, C., and Houkari, M. Learning optimal control using neural networks. In *Proc. Fifth International Neuro-Nimes* (Paris, 1992), CEA (Saclay), EC2, pp. 431–442.

[8] Matussek, P. Accurate determination of the ^{235}U isotope abundance by γ spectrometry, a user's manual for the certified reference material EC-NMR-171/NBS-SRM-969. In *Report KfK-3752* (Karlsruhe, 1985).

[9] Morel, J., Goenvec, H., Dalmazzone, J., and Malet, G. Références pour la détermination de l'uranium 235 dans les combustibles nucléaires. *IAEA Nuclear Safeguard Technology* (1978).

[10] Reilly, T. D., Walton, R. B., and Parker, J. L. The enrichment metter. In *Report LA-4065-MS* (Los Alamos, 1970).

[11] Simon, A. C., and Junca, R. *Dosage de l'uranium et du plutonium par l'appareil de fluorescence X SAPRA γX/Ir: simplification des codes de calcul dans le cas de solutions pures.* In *Note technique DTA/DAMRI/SAR/S/94-433* (Saclay, 1994).

[12] Vigneron, V., and Martinez, J. M. A looking at Neural Network methodology. *Submitted to Nuclear Instruments and Methods in Physics Research* (1995).

List of Figures

1 Principal useful $X-$ and $\gamma-$rays in the spectral analysis of the $K_\alpha X$ region. 3
2 3D-Representation of the UO_2 spectra set. 5
3 (a) MLP 3-5-1 with nonlinear threshold and (b) schematic representation of a node in an ANN. 6
4 Mixtures of Experts model. 8
5 Sum of squares of bias on the training set for MLP architectures 9
6 Example of enrichment value (at 5,785 %) predicted by the Mixtures of Experts 10
7 Absolute bias in the enrichment estimation 11