



Measurement of uranium quantities by fluorescence-X using neuronal techniques

Vincent Vigneron, Jean-Marc Martinez, Anne-Catherine Simon, Robert Junca

► To cite this version:

Vincent Vigneron, Jean-Marc Martinez, Anne-Catherine Simon, Robert Junca. Measurement of uranium quantities by fluorescence-X using neuronal techniques. ERSADA 17th symposium on safeguards and nuclear material management, May 1995, Aachen, Germany. cea-02524489

HAL Id: cea-02524489

<https://cea.hal.science/cea-02524489>

Submitted on 24 Jan 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Measurement of Uranium quantities by fluorescence-X using Neuronal techniques

V. Vigneron, J.M. Martinez* A.C. Simon, R. Junca†

Abstract

Layered neural networks are a class of models based on neural computation in biological systems. Connexionists models are made of a large number of simple computing structures, highly interconnected. The weights assigned to the connections enable the encoding of the knowledge required for a task. They can be trained to learn any input-output relation after selecting a suitable architecture. This method appears useful in those cases in which a simple operation and a fast response are needed, together with a reasonable accuracy. They are applied here to the automatic analysis of X-ray fluorescence spectra, obtained with plastic bottles irradiated by collimated beam of photons, emitted by a sealed source of iridium. The method could allow the surveillance of the measurement of uranium quantities in nuclear fuel cycles in processing plants.

Key words *Pattern recognition, data analysis, Neural networks, supervised learning, regression, Poisson distribution.*

1 Introduction

Neural networks consist of a large number of neurons, i.e. simple linear or nonlinear computing elements, interconnected in often complex ways and organized into layers [7]. The collective or parallel behaviour of the network is determined by the way in which the nodes are physically connected and the relative strength and type (excitatory or inhibitory) of the interaction amongst them. Artificial Neural Networks (ANNs) are used by engineers, physicists, neurophysiologists, or computer scientists in three main ways:

- as models of biological nervous systems and intelligence
- as real-time adaptative signal processors or controllers for applications such as robots or nuclear plants
- as data analysis methods.

The ANNs, like many statistical methods, are capable of processing vast amounts of data and making 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. This paper is concerned with ANNs for data analysis and the application to parameters estimation of a chemical calibration process. Selection and validation of model structures are discussed. The reasons for using ANNs are that:

- they can be trained on-line
- they can adapt to slow variations of system and will be able to diagnose abnormal conditions
- they are fault tolerant and hence, reliability is significantly improved.

There exist some cases where a situation is so precise and clear that a purely mathematical model can be devised that is accepted as true within clearly defined conditions. However, in most circumstances we lack clear knowledge of the underlying mechanisms and relationships in the situation, our data is limited in coverage and subject to error, namely in our fluorescence problem. In this context, ANNs are a useful *black-box* alternative for the no apriori information modelling.

*CE de Saclay, DRN/DMT/SERMA (France)
email: vvigne@soleil.serma cea.fr

This research was supported in part by a grant from the french government.

†CEA, CE de Saclay, DAMRI/SAR (France)

2 Physical system

3 Statistical perspectives

3.1 Pattern recognition

The human observer is very good at recognizing patterns. He is able to associate a character such as "a" or "b" with a particular sound and to distinguish a person from other people. It is clear the combination of the human brains and eyes are very superior a computer algorithm. Note that, while making the figure for two variables is not at all difficult, it becomes much more difficult to draw it for three variables or dimensions. Clearly, three variables constitute the limit of our visual ability. This explains why much of pattern recognition is concerned with the question: how does one condense m -dimensional information in two or three dimensions? We have used the word *pattern* in an intuitive way. It seems our eyes and brains use features to describe a sample and distinguish it from others. This list of features constitutes a *pattern*. In the same way, we note that the fluorescence X spectra are all geometrically similar (the distance between objects is therefore a measure of similarity).

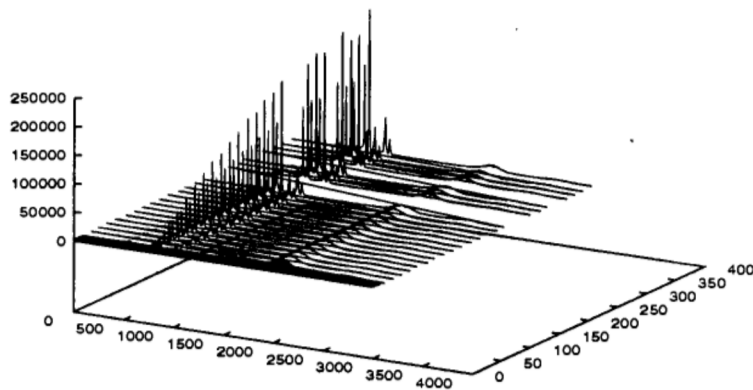


Figure 1: 3D multi-spectral representation

The main sources of constant error for the determination of an analyte are:

- matrix effects due to the presence of components which inhibit the measurements
- random interferences influencing the sensitivity of the sensor
- saturation phenomenon in the limits of detection.

But these errors cannot be easily quantified, hence the model may be finally wrong in form, inaccurate in the numerical values we use in it, inappropriate in some new situation to which we apply it. It should be clear that we cannot expect our models to be true descriptions of reality. We have to judge it not in terms of right or wrong but in terms of insight and use.

3.2 Reduction of the data space

From a statistical point of view, training a feedforward neural network with a fixed topology can be viewed as estimating the parameters of a nonlinear regression function. If the topology is allowed to vary, then network training can be viewed as an exercise in nonparametric regression [1]. The basic design principle is to minimize the number of free parameters that must be determined by the learning algorithm i.e. the "Back-Propagation", without overly reducing the computational power of the network: it is obtained by introducing a certain amount of *a priori* knowledge about the problem, clearly in focusing our attention on wellknown energies. [9]. data in a small of a data set.

In Principal Component Analysis (PCA), the new variables are considered as linear combinations of the old ones (*feature reduction*). Alternatively, one may consider the m variables among n the most *informative* (*feature selection*). In this discrimination view, the suppressed variables reduce the signal-to-noise ratio. Our method mixed the two approaches intuitively and use knowledge of fluorescence-X

phenomenon ¹.

A much used method is to select the first p principal components (PC) in such a way that they account for at least 80 or 90% of the total variation $\sum_{i=1}^m \lambda_i$ for example the following condition:

$$\frac{\sum_{i=1}^p \lambda_i}{\sum_{i=1}^m \lambda_i} \geq 0,8 \quad (1)$$

Another criterion used to select PC is to keep eigenvalues of the correlation matrix which exceed 1. Kaiser[6] introduced a selectivity parameter based on the elements of the data matrix, i.e. the sensitivity coefficients of the calibration process. The smallest value ξ is the weakest part of the procedure and according Kaiser, this minimum value determines the selectivity of the whole procedure. The knowledge of correlations between analytical results may lead to important decision about the design of the analytical procedure too (see Fahlman). Herault[4] propose new perspectives of data analysis with Independent Component Analysis with MLPs.

7 variables per spectrum has been retained by using our mixed method. Figure 2 gives an impression of the degree of correlation between the set of

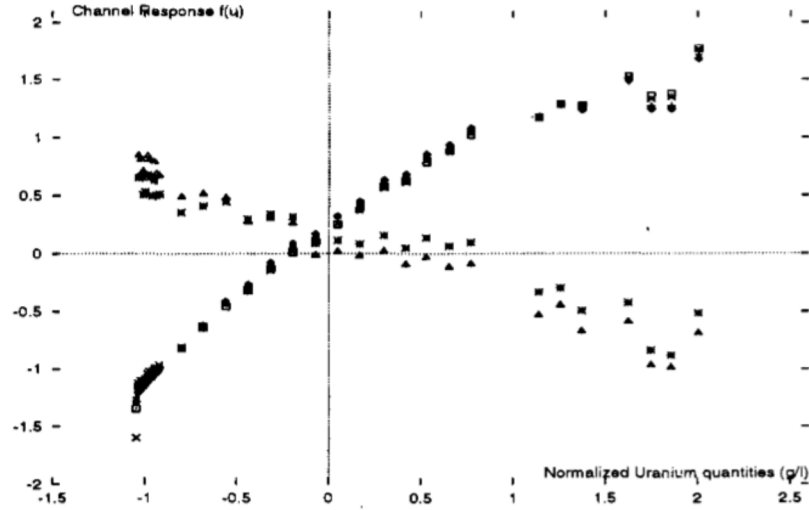


Figure 2: Normalized data Space

7 features (consisting in the main $K\alpha$ peaks of ^{92}U and ^{192}Ir) fixed to forecast the calibration process. These 7 measurements define the 7-variate pattern or input space. This figure show clearly the isolation of two clusters, the 7 objects are then characterized by two variables.

4 Neural architecture

"Classical" solutions [8] to the automatic spectrum analysis problem make use of complex mathematical algorithm, generally based on the separation of a given curve (the spectrum) into a combination of individual curves, associated to each individual peak, plus a background. The latter is considered as a negative element that detracts from the quality of analysis. It is not essential with ANNs to have *a priori* knowledge of the real physical model. The basic idea is to construct a function to summarize the data:

$$Cu = F(Y) \text{ with } Y \in \mathcal{R}^n \subset \mathcal{R}^{4096}, Cu \in \mathcal{R} \quad (2)$$

if Cu quantify the uranium quantities and Y is the input vector of the counts and with the assumption $n \ll 4096$. The estimate of the parameter values of the matrix F is called *direct approach*. The function F compile each channel response of the spectra, as the one shown on the diagramm 3; it is clear that the relation is far from the linear case.

4.1 Preprocessing the inputs

The number of examples needed to train a network depends on the nature of the calibration function to be approximated and the network architecture. The learning network is directly fed with the previous 31

¹PCA plays a key role in filtering out the systematic error. In practice, PCA has important economic consequences.

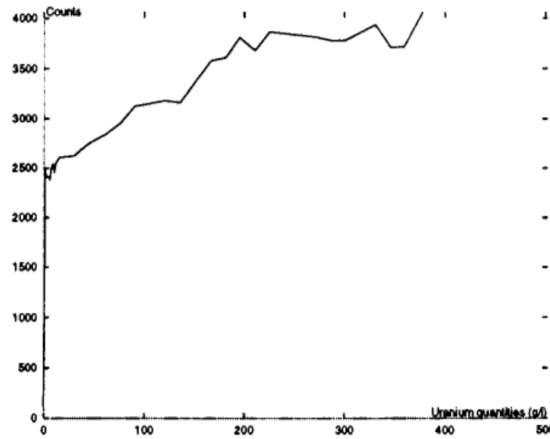


Figure 3: One channel Response

feature vectors (see figure 4). Former authors (LeCun and Boser, 1990) have shown that the architecture of the network strongly influences the network's generalization ability.

Each input variable should be sifted so that its mean (averaged over the training set) is close to 0 because *correlations* between input variables introduce "preferred directions for weight changes" [11]. Decorrelation has been performed by a PCA (Karhunen-Loeve expansion). This transformation preserves the aspect of the data base. The output is composed of the concentration in ^{92}U .

The connectionist must now deal with overfitting effects: indeed, when the net is overparametrized, convergence is a nonsense: the net has learned the function and the interferences, he will tend to generalize poorly. To help the net to fluctuate outside the training set values, a way consist of adding to the 20×31 "simulated" spectra obtained from real data by introducing ourselves uncertainties equal to $\sqrt{y_i}$ from a Poisson distribution ² on each channel. The data collection was such that it covered a number of possible configurations of the fluorescence system.

4.2 The model

The method of solution reported here relies more heavily on automatic learning, and much less on signal processing. MLPs are general purpose, flexible, non-linear models that, given enough hidden neurons and enough data, can approximate virtually any function to any desired degree of accuracy. In others words, MLPs are *universal approximators* (Hornik, 1989). They can be used when you have little knowledge about the form of the relationship between variables. You can vary the complexity of the MLP model by varying the number of hidden layers and the number of neurons in each hidden layer. A fully connected network with enough discriminative power for the task (one input neuron per channel for a 4096 channel spectrum) would have far too many parameters to be capable of generalizing correctly and cause overfitting. With a small number of hidden neurons, an MLP is a parametric model that provides a useful alternative to polynomial regression. The activity function in a MLP is analogous to the link function in a *Generalized Linear Model* (GLIM) [10] [5]. GLIMs are fitted by Maximum Likelihood for a variety of distributions in the exponential family. Maximum Likelihood for Poisson events is also used for MLPs in which case the criterion is called *cross entropy*. Thus, in theory, GLIMs and MLP are almost the same thing. In this simulation of the real process with a backpropagation approach we have difficulties with:

- the number of data required to train a net
- the structure of the ANN (neurons on hidden layers for a given number on the input and output layer)
- is there a stabilized statistical structure of the data ?

A fully connected network with enough discriminative power for the task (one input neuron per channel for a 4096 channel spectrum) would have far too many parameters to be capable to generalize correctly and cause overfitting [3]. If f is the unknown input-output relation $y = f(x)$ and w the weights-parameters vector, then our aim is to train the network, i.e. to select w such that the cost function

²The probability of obtaining i counts for a real channel counting λ counts can then be calculated as $P(i) = \frac{e^{-\lambda} \lambda^i}{i!}$.

$E(y, f(w, x))$ usually measured by the squared error between the target output y and that predicted by the network f_N is minimised over all possible training examples in the weight space:

$$E(f) = \sum_{p=1}^P \sum_{i=1}^I (y_i^{(p)} - f_N(w, x_i^{(p)}))^2 \quad (3)$$

This equation can be used for the online training of the network. Iteratively, the weights are updated by a stochastic least-mean-square error gradient descent rule. The training process is assumed to have reached a stochastic equilibrium centered around a global minimum [1].

The choice of the right architecture is mainly intuitive and imply arbitrary decisions. The network allows the fusion of information. Another complexity reduction method for neural networks consists of training an oversized network and to reduce its size and to remove elements of the structure until the cost function be minimized. To validate our model [2], we have used a popular heuristic for the prevention of overfitting of the training data called *cross validation*: it consists of observing the generalization performance of the network with a test dataset (that is not used in the training process), then, as soon as its performance begin to deteriorate, training is stopped.

4.3 Results

Our fully connected architecture detailed in figure 4 is composed of an active layer of 5 inputs, one hidden layer and one output. After 100.000 training passes the Mean Square Error (MSE) was $0,43 \times 10^{-3}$. On the whole test set the MSE was $0,33 \times 10^{-3}$. We think this can be attributed to the large amount of correlation present in real data. The error rates are compared on figure 5: the curve with rises and drops is for the conventional regression and the curve that rises slowly corresponds to neural method. After successful training, the larger errors occurred on the shorter concentrations. The adapted net was able to predict the larger uranium quantities better than the usual technics developped by Martinelli & al. We must put the blame on the excessively short size of the training dataset. The mere fact of getting a reliable identification in such adverse conditions can be considered as a success of the method.

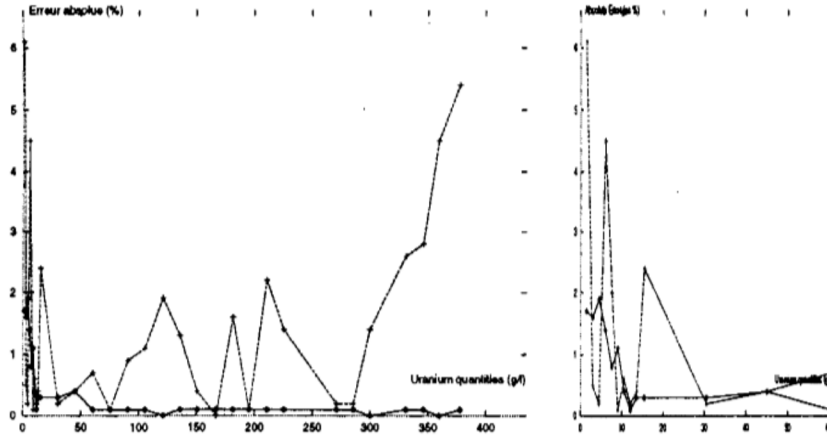


Figure 4: Comparison of quality of concentration estimation obtained with neural and classical regression models

5 Conclusions

We have studied the feasibility of applying ANNs for x-ray fluorescence spectra calibration. The network is shown to be able to calculate uranium quantities. Our results appear to be at the state of art in automatic quantifying methods of isotopes in a mixture of components. Based on a very simple application of the Back-Propagation learning, this method has been demonstrated as a reliable tool for dealing with real world task with low resolution detectors working even in adverse situations. Results may be better than those obtained with standards methods in similar cases. Final network of connections and weights could be easily implementable on commercial digital processing hardware.

Because of its architecture and the short size of real dataset, the network must be trained on a representation of data that had minimal preprocessing to avoid fitting. The net codes its own representation of

the data in the hidden layer whose nonlinear activation function allows the reduction of the complexity size problem. Even the problem has been readily simplified, preliminary results show that the method can be directly extended to larger tasks. A limitation of the method when applied to complex cases is the training the need of training the net with numerous samples covering all the data space. The extension to multiple outputs should be obvious.

Acknowledgments

The authors wish to thank the staff of SAR and LPRI, and in particular J.L. Boutaine, J. Morel and M.C. Lepy for their support and useful discussions during these works.

References

- [1] FINNOFF, W., AND ZIMMERMANN, H. Detecting structure in small datasets by network fitting under complexity constraints. *Preprint* (1991).
- [2] GANS, D. *Data fitting in the chemical sciences by the method of Least Squares*. John Wiley & sons, Leeds, 1993.
- [3] GILCHRIST, W. *Statistical Modelling*, second ed. John Wiley & sons, Leeds, 1984.
- [4] HERAULT, J., AND JUTTEN, C. *Réseaux neuronaux et traitement du signal*, vol. 1. Traité des Nouvelles Technologies- Hermes, 1984.
- [5] JORDAN, M. *Lectures on neural networks*. Ecole d'été du bréau-sans-nappe, MIT, 1994.
- [6] KAISER, H. On the definition of selectivity, specificity and sensitivity of analytical methods. *Z. Analytical Chemistry* 260 (1972).
- [7] LECUN, Y. Efficient learning and second-order methods. Tech. rep., ATT & Bell Laboratories, 1993. Preprint.
- [8] LEPY, M., AND MOREL, J. *Traitement de l'information en spectrométrie γ ou X*. Dta/damri/lpri 17-52, CEA, CE Saclay, France, 1985.
- [9] MARTINELLI, AND BOUTAINE, J. Détermination of uranium and/ or plutonium using x-ray fluorescence analysis excited by ^{192}Ir sealed sources. *Nuclear Instruments and Methods in Physics Research* 242 (1986), 569-573.
- [10] MCCULLAGH, AND NELDER, J. *Generalized Linear Models*, second ed. Chapman & al., Amsterdam, 1988. Monographs on Statistics and Applied Probability 37.
- [11] SAPORTA, G. *Probabilités, analyse de données et statistiques*. Technip, 1992.