X-RAY SPECTROMETRY
X-Ray Spectrom. **28**, 173–177 (1999)

Simultaneous Determination of Lead and Sulfur by Energy-Dispersive X-Ray Spectrometry. Comparison between Artificial Neural Networks and Other Multivariate Calibration Methods

I. Facchin, C. Mello, M. I. M. S. Bueno and R. J. Poppi*
Universidade Estadual de Campinas, Instituto de Química, C.P. 6154, 13083-970, Campinas, SP, Brazil

The need for mathematical methods to model data in energy-dispersive x-ray fluorescence (EDXRF) spectrometry is common owing to the overlapping of intense spectral lines in complex samples. This overlapping generally produces a large amount of scatter in the analytical curve, preventing simultaneous direct determinations of some elements without data treatment. This work demonstrates the performance of artificial neural networks (ANN) and other methods of multivariate calibration (linear or not) for the simultaneous determination of sulfur and lead, when overlapping of the sulfur Kα spectral line (2.308 keV) and the lead Mα line (2.346 keV) is observed. The performance of neural networks was compared by the *f*-test with five other data treatment methods: PLS (partial least squares), POLYPLS (polynomial partial least squares), NNPLS (partial least square neural networks), LR (linear regression) and CI (corrected intensity). It was verified that the ANN produces better predictions than the other methods, for both sulfur and lead, allowing their simultaneous determination in solid samples with good accuracy.

INTRODUCTION

Nowadays the use of instrumental techniques that allow simultaneous determinations have aroused the attention of the scientific community owing to the growing need for chemical analysis in a great variety of matrices. In this sense, energy-dispersive x-ray fluorescence (EDXRF) spectrometry is a very useful spectroscopic technique, since it can be applied to several different types of samples, including solid (powdered or not), liquid (of high or low viscosity) and even those of a pasty consistency, without extensive sample treatment. Nevertheless, a drawback of EDXRF is the difficulty of quantifying species when they suffer spectral interference from another species present in the sample. There is commonly a loss of the calibration linearity[1–3] and to model the data the use of mathematical methods is essential.

An example of spectral interference in EDXRF is treated in this paper, viz. the determination of sulfur in the presence of lead. In this case, there will be strong overlapping between the S Kα line and the Pb Mα line. A possible practical application for this particular case of signal overlapping will be the simultaneous monitoring of S and Pb in fuels, to which are added organic compounds of lead. Generally fossil fuels contain appreciable amounts of sulfur and its control is important for the environment.

A mathematical tool frequently used to model cases of spectral interference is the correction intensity (CI) method.[4,5] In this case, correction coefficients are calculated from data obtained with standards to model the inter-element effects. The most popular algorithm is the Lucas and Price method, in which the analysis of an element I with n possible interfering elements is expressed in accordance with the equation

$$C_I = B_I + K_I R_I + \sum_{J=1}^{N} A_{IJ} R_I R_J \qquad (1)$$

where C is the concentration of element I, R is the relative intensity of the element, A is a constant that represents the effect of element J on the intensity of element I and B and K are calibration coefficients of element I as a function of its intensity.

Another commonly used technique in spectroscopic data modeling is the partial least squares (PLS) method.[6–8] The basis of this method is to associate mathematically the signals from several channels with the concentrations of the species involved,[9] as is done in a typical calibration procedure. Nevertheless, instead of using signal intensities, linear combinations of them are used. This procedure reduces the signal dimensions to a much smaller parameter collection that is still very representative.

Some PLS alternatives can be used in cases where a non-linear relationship is present.[10,11] Among them, polynomial PLS (POLYPLS) and PLS with neural networks (NNPLS) will be considered. POLYPLS is basically identical with linear PLS, the only difference being that relationships are expressed by a variable degree polynomial.

Another very useful alternative for the treatment of non-linear data is artificial neural networks (ANN), which can

* Correspondence to: R. J. Poppi, Universidade Estadual de Campinas, Instituto de Química, C.P. 6154, 13083-970, Campinas, SP, Brazil.

CCC 0049–8246/99/030173–05 $17.50

Received 30 March 1998
Accepted 12 January 1999

be defined as a set of mathematical models and computational algorithms specially designed to imitate human brain information processing and knowledge acquisition. So designed, the ANN are able to acquire information and provide models, even when the information and data are complex, noise contaminated, non-linear or incomplete.[12–14]

The neural network is now considered a mathematical tool of widespread application, especially in analytical chemistry. Excellent models with high generalization capacities are being designed for simultaneous analyses of multicomponent systems using spectroscopic,[15] chromatographic[16] and electrochemical[17] techniques. Besides modeling linear or non-linear data, ANN find applications in pattern recognition,[18] signal processing and the analytical chemistry of processes.[19]

The ANN used in this work is composed of three neuron layers, which are represented by black circles in Fig. 1. The input signals are the x-ray fluorescence intensities at several energies of the EDXRF spectra and the output signals are the sulfur and lead concentrations in the standard samples used. After entering the input layers, the signals will be sent to the hidden layer neurons where they will be processed and sent to the output neurons. There they will be processed again before exiting the neural network. In the processing steps, the signals in the hidden layer or in the output layer are multiplied by each neuron weight and summed over all the layer. This procedure gives rise to the entrance net signal, or simply Net, represented by

$$\text{Net} = \sum w_i I_i \tag{2}$$

where w_i represents the weight of the ith neuron and I_i is the ith X-ray fluorescence intensity. Since Net is calculated, a sigmoidal function (known as the transfer function) is applied to it. The value of this function is the signal sent to the output layer neurons where the Net will again be calculated. Finally, a linear transfer function is applied to the output layer Net and the resulting signal is sent out of the ANN. In Fig. 1, a special type of neuron called bias can be observed, and it is utilized to add a proportional constant to the Net.[20]

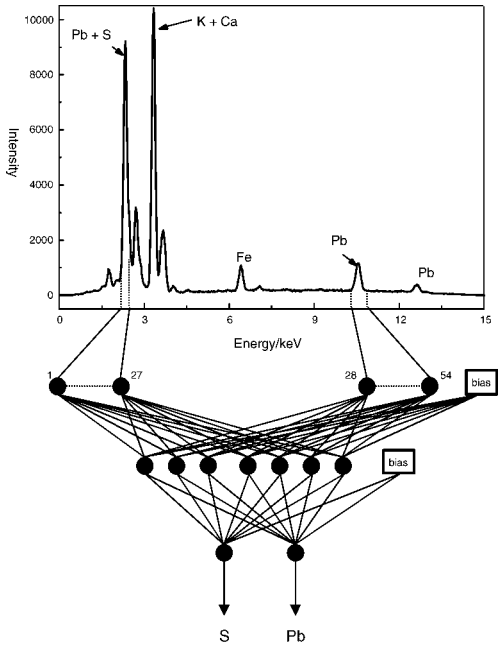

Figure 1. Typical EDXRF spectrum of solid samples in a charcoal matrix, showing the regions of interest and the associated neural network used. The dotted lines limit the energy range selected as input in the neural network, the black circles are the neurons and the continuous lines are the connections between them.

The weights and biases, in hidden and output layers, are adjusted in the training step in such a way that the ANN outputs are equal to the sulfur and lead concentrations used in the calibration samples. This is done in accordance with a convergence criterion, previously determined, as is done in a typical calibration procedure. The back-propagation algorithm[20] is normally used to weight corrections, mainly owing to the facility with which it can be implemented. However, this algorithm has problems in convergence properties. In this work, the weight corrections were made via a variant of the Gauss–Newton method, known as the Levenberg–Marquardt method,[21–23] that is faster in convergence and robust. The corrections are calculated as

$$\Delta W = (\boldsymbol{J}^{\mathrm{T}}\boldsymbol{J} + \lambda \boldsymbol{I})^{-1}\boldsymbol{J}^{\mathrm{T}}\boldsymbol{g} \tag{3}$$

where $\boldsymbol{J}$ is the Jacobian error matrix for each weight, λ is a positive scalar, $\boldsymbol{I}$ is the identity matrix and $\boldsymbol{g}$ is an error vector. The sum of the squared errors is usually used as the vector $\boldsymbol{g}$:

$$\boldsymbol{g} = \sum_{i=1}^{m} (y_{\text{real},i} - y_{\text{predicted},i})^2 \tag{4}$$

where m is the total number of samples used in the training step of the ANN.

The training step ends when the difference between the real lead and sulfur concentrations and those evaluated by the ANN ($y_{\text{predicted}}$) reach the convergence criterion. At this point, the network is trained and a set of distinct data (EDXRF spectra and sulfur and lead concentrations) can be used to evaluate the generalization properties of the ANN. These properties are usually good, especially when the data show nonlinearity.[24,25]

In addition to the preceding methods, a combination of ANN with PLS can be applied, known as NNPLS. This procedure facilitates the modeling both of linear cases and of those with weak non-linearities. The other steps are identical with those of linear PLS. It must be kept in mind that the difference between the models is that NNPLS is based on scores, whereas the conventional ANN is not. However, it is possible to elaborate models with ANN using scores as inputs into the ANN, instead of pure spectra.

Linear regression was also used in this work for comparison with the other methods of multivariate calibration listed above. The relationship taken was the signal maximum of each element with its concentration.

EXPERIMENTAL

Samples

A set of 38 samples was prepared from mixtures made with K_2SO_4 (Synth) and $Pb(NO_3)_2$ (ECIBRA) in charcoal (NORIT), in such a way that the sulfur and lead

Table 1. Compositions of synthetic mixtures

Sample No.	Composition (%, m/m) Pb	S	Sample No.	Composition (% m/m) Pb	S
1	1.960	5.004	20	2.014	15.996
2	2.009	8.000	21	2.997	1.997
3	1.995	10.990	22	2.996	15.998
4	1.978	14.010	23	7.998	7.002
5	2.001	16.993	24[a]	7.982	10.994
6	5.017	2.006	25	7.987	14.005
7	7.990	2.000	26	7.987	15.999
8	11.004	1.999	27	7.005	7.998
9	14.006	1.998	28	10.995	8.003
10[a]	16.991	2.000	29	2.989	12.007
11	1.994	1.997	30	3.993	9.002
12	4.970	4.999	31	6.014	10.996
13	4.997	7.996	32	8.997	4.001
14	5.013	10.998	33[a]	11.994	2.998
15	4.996	13.996	34[a]	14.997	2.000
16[a]	4.008	16.993	35	2.009	15.001
17[a]	7.972	5.000	36	3.986	6.000
18[a]	11.001	4.994	37	3.992	11.000
19	1.996	3.001	38[a]	10.997	4.000

[a] Samples used for prediction.

concentrations were between 2.00 and 17.00% (m/m) and the total mass was equivalent to 1.0000 g. Before weighing, the solids were sieved to homogenize their grain size (100–130 mesh).

Instrumentation and measurements

Spectra were taken using a Spectrace 5000 x-ray fluorescence spectrometer with an irradiation time of 50 s under vacuum and without a radiation filter. The rhodium x-ray tube was operated with a voltage of 15 kV and a current of 0.02 mA. As indicated in Table 1, 30 samples were selected to elaborate the calibration models and the remainder were used for prediction purposes.

Pretreatment of the EDXRF spectra

The dotted lines in Fig. 1 show the spectral regions used to create the calibration models. In ANN, the spectra were normalized between 0 and 1 and the concentrations between 0.2 and 0.8. In using PLS and its non-linear variants, the data were centered on the average and scaled for unity variance.

Data modeling

MATLAB[26] programs were developed to model the data, using routines of PLS Toolbox version 1.3[27] and of Neural Network Toolbox version 2.0.[28] The calculations were made with a Pentium microcomputer operating at 150 MHz with 32 Mbyte of RAM.

RESULTS AND DISCUSSION

The standard errors of calibration (SEC) and prediction (SEP) were used to analyze the performance of the six models, as shown in Table 2. The SEC values were obtained using the equation

Table 2. Standard error of the calibration (SEC) and prediction (SEP) results using the six calibration models

Mathematical model	SEC (%) S	SEC (%) Pb	SEP (%) S	SEP (%) Pb
Artificial neural networks	5.0	6.3	4.0	9.2
PLS neural networks	7.3	14.2	8.4	12.2
PLS polynomial	7.3	14.3	8.6	13.3
PLS linear	12.3	16.8	7.8	17.2
Correction intensity method	11.1	12.0	8.3	10.2
Linear regression	17.5	21.5	20.3	22.5

$$\mathrm{SEC}(\%) = \frac{100}{c_\mathrm{m}} \left[\frac{\sum_{i=1}^{n} (C_{ij} - \hat{C}_{ij})^2}{n - k - 1} \right]^{1/2} \quad (5)$$

where C_{ij} is the actual concentration, $\hat{C}_{ij}$ is the estimated concentration of the jth component of the ith standard, c_m is the average concentration of the standards and k is the number of degrees of freedom. For PLS, k is the number of latent variables. In ANN, the number of degrees of freedom is unknown,[29] so n, the number of calibration samples, is used as an approximation for the denominator in Eqn (5).

The SEP columns in Table 2 represent the percentage relative standard deviation of the prediction for the jth component of the validation set. In the equation

$$\mathrm{SEP}(\%) = \frac{100}{c_\mathrm{m}} \left[\frac{\sum_{i=1}^{n} (C_{ij} - \hat{C}_{ij})^2}{p} \right]^{1/2} \quad (6)$$

C_{ij}, $\hat{C}_{ij}$ and c_m are the same as in Eqn (5) and p is the number of samples used in the validation set.

From Table 2, it can be seen that the LR method was not efficient, since SEC and SEP were high for both lead and sulfur. The weak performance of this method for sulfur is basically due to the non-linearities and for lead to the high noise level around the Pb Lα line used in the calibration step. Figure 2(a) and (b) illustrate this fact, since they are the analytical curves for these two elements evaluated using the maximum at each peak (Pb Lα and S Kα).

The CI model was chosen to evaluate the effects of potassium, present in the matrices (K Kα = 3.313 keV), to verify the possibility of interference of this element in the quantification of the other species. The values of the A_{ij} coefficients [Eqn (1)] for the interference between the elements S–K and Pb–K were -1.50×10^{-4} and -1.58×10^{-3}, respectively, so this interference was not considered relevant. When compared with LR, this method allowed modeling of the system with a significant decrease in the SEP values, equal to 8.3% for sulfur and 10.2% for lead (Table 2).

Using linear PLS, four latent variables were retained, based on cross-validation,[6] two more than theoretically necessary, since the system has two components. The need to include more components indicates again the

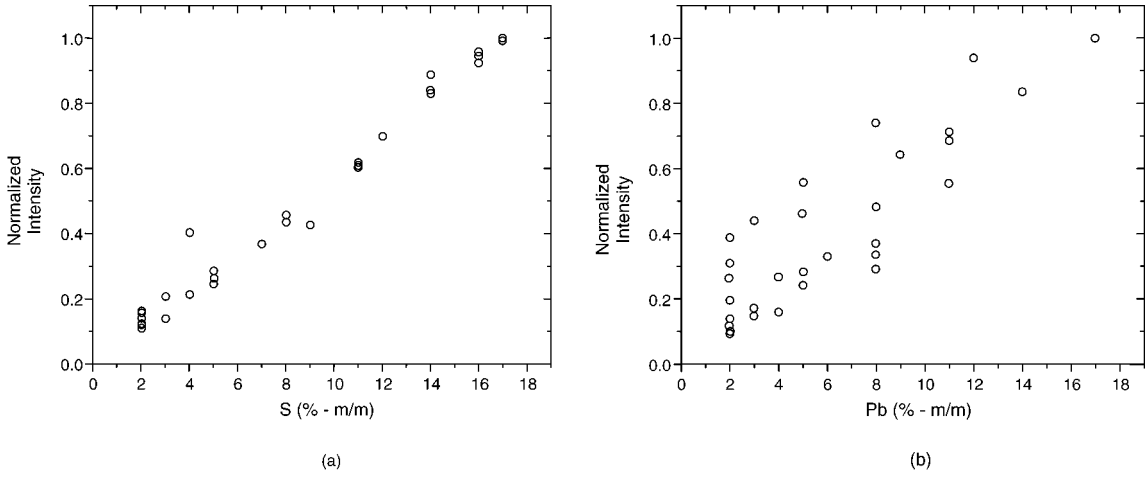

Figure 2. Linear regression models using (a) the maximum values of the S Kα line x-ray fluorescence intensity (normalized between 0 and 1) versus sulfur concentration and (b) the maximum values of the Pb Lα line x-ray fluorescence intensity (normalized between 0 and 1) versus lead concentration.

non-linearity, being difficult to model by linear PLS. Here the SEP were 7.8% for sulfur and 17.2% for lead.

In POLYPLS, four latent variables and a second-degree polynomial provided the best model. Table 2 shows a lower SEP for sulfur compared with the linear PLS values. This is an indication that part of the non-linearities were modeled in this case.

In NNPLS, four latent variables were again taken and the ANN that provided the best model had a hidden layer with three neurons. The transfer function between the input and the hidden layer was sigmoidal, and that between the hidden layer and the output layer was linear. In this modeling, the network was trained with the Levenberg–Marquardt method. The values of SEC and SEP were very close to those obtained by POLYPLS. Hence the performances of NNPLS and POLYPLS in modeling the system can be considered equivalent.

The ANN architecture was organized in three layers: the input layer with 54 neurons (one for each selected energy, as represented by the dotted lines in the EDXRF spectrum, Fig. 1), the hidden layer with seven neurons and the output layer with two neurons, one for sulfur and the other for lead. The number of neurons used in the hidden layer was chosen by continuously adding neurons to it and then evaluating the average SEP for lead and sulfur. Using this procedure, seven neurons provided the lowest value for this average SEP, as shown in Fig. 3.

The inputs were the normalized spectra between 0 and 1 and the outputs were the sulfur and lead concentrations. A sigmoidal function was used as a transfer function between the hidden and the output layers, whereas a linear function was used at the output layer. Finally, the ANN was trained by the Levenberg–Marquardt algorithm to have the weights that provide the lowest SEP and SEC. ANN was the best model, as can be seen in Table 2.

An F-test was used to compare the relative performances of all six models, at the 95% confidence level:[29]

$$F(p_i, p_j) = \left(\frac{\mathrm{SEP}_i}{\mathrm{SEP}_j}\right)^2 \qquad (7)$$

where p_j is the number of samples used in ANN and p_i is the number of samples used in PLS, POLYPLS, NNPLS, CI and LR.

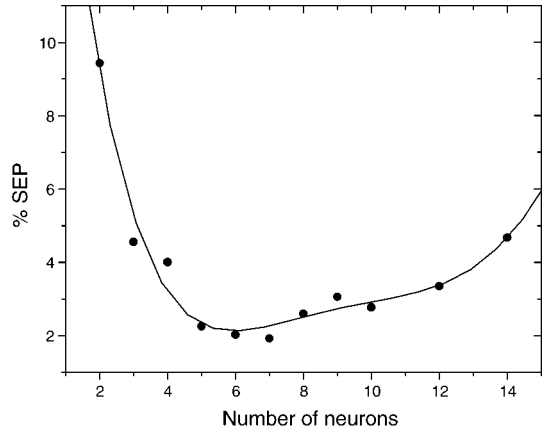

Figure 3. Plot of the average SEP versus the number of neurons added to the hidden layer.

By using the F-test in the prediction data set as a figure of merit, one can see in Table 3 that for sulfur, the ANN is better than all the others, at the 95% confidence level ($F_{\mathrm{critical}} = 3.44$). This shows the complex non-linear behavior of this element, which cannot be properly modeled for non-linear variations of the PLS, such as the corrected intensity method. The scatter (or deviation from the straight line) of the S Kα line can be attributed to two factors: Pb Mα line interference and noise.

In the case of lead, the application of NNPLS, POLYPLS and CI showed similar results to those of ANN, since the F-values were less than F_{critical} (Table 3). In principle, this result can be attributed to the fact that lead

Table 3. Comparison of neural networks with five other data treatment methods using the F-test

	F-test	
Mathematical model	S	Pb
PLS neural networks	4.4	1.6
PLS polynomial	4.6	2.0
PLS linear	3.8	3.5
Correction intensity method	4.3	1.2
Linear regression	25.7	6.0

has a line free of interference, Lα, and a linear behavior is expected. Nevertheless, Table 3 illustrates that even here the non-linear methods provide better results than the linear methods. In addition to the instrumental noise, the differential x-ray absorption coefficients of the standards as a function of the relative matrix constitution can also explain this effect.

CONCLUSION

This work demonstrated that the use of artificial neural networks in modeling non-linear data in EDXRF is a very useful alternative. The relative standard errors of prediction (SEP), when compared with univariate linear methods, was lowered by 35% for lead and 100% for sulfur. In addition, the computational time spent to train the network, owing to the use of the Levenberg–Marquardt algorithm for weight correction (about 2 min), was of the same order of magnitude as those with the other methods.

Acknowledgments

The authors are grateful to Dr Carol H. Collins for helpful revision of the manuscript.

REFERENCES

1. D. E. Leyden, *Fundamentals of X-Ray Spectrometry as Applied to Energy Dispersive Techniques*. Tracor X-Ray, Fort Collins, CO (1984).
2. T. Ellis, D. E. Leyden, W. Wegscheider, B. B. Jablonsky and W. B. Bodnar, *Anal. Chim. Acta* **142**, 73 (1982).
3. R. T. Mainardi and R. A. Barrea, *X-Ray Spectrom.* **23**, 36 (1994).
4. N. Broll, *X-Ray Spectrom.* **15**, 271 (1986).
5. G. R. Lachance, *X-Ray Spectrom.* **8**, 190 (1979).
6. P. Geladi and B. R. Kowalski, *Anal. Chim. Acta* **185**, 1 (1986).
7. K. R. Beebe and B. R. Kowalski, *Anal. Chem.* **59**, 1007A (1987).
8. E. V. Thomas, *Anal. Chem.* **66**, 795A (1994).
9. Y. Wang, X. Zhao and B. R. Kowalski, *Appl. Spectrosc.* **44**, 998 (1990).
10. S. Sekulic, M. B. Seasholtz, Z. Wang, S. E. Lee, B. R. Holtz and B. R. Kowalski, *Anal. Chem.* **14**, 835A (1993).
11. S. J. Qin and T. J. MacAvoy, *Comput. Chem. Eng.* **16**, 379 (1992).
12. J. Zupan and J. Gasteiger, *Anal. Chim. Acta* **248**, 1 (1991).
13. G. Kateman and J. R. M. Smits, *Anal. Chim. Acta* **277**, 179 (1993).
14. J. R. Long, V. G. Gregoriou and P. J. Gemperline, *Anal. Chem.* **62**, 1791 (1990).
15. M. Catasus and E. D. Salin, *Appl. Spectrosc.* **49**, (1995) 798.
16. B. Valejo-Cordoba, G. E. Artega and S. J. Nakai, *J. Food Sci.* **60**, 885 (1995).
17. Z. Wang, J. N. Hwang and B. Kowalski, *Anal. Chem.* **67**, 1497 (1995).
18. D. Wienke and L. Buydens, *Trends Anal. Chem.* **14**, 8 (1995).
19. P. Bhagat, *Chem. Eng. Prog.* **86**, 55 (1990).
20. J. Zupan and J. Gasteiger, *Neural Networks for Chemists: An Introduction*. VCH, Weinheim (1992).
21. K. Levenberg, *Q. Appl. Math.* **2**, 164 (1944).
22. D. W. Marquardt, *J. Soc. Ind. Appl. Math.* **11**, 431 (1963).
23. J. R. S. Jang, C. T. Sun and E. Mizutani, *Neuro Fuzzy and Soft Computing: a Computational Approach to Learning and Machine Intelligence*. Prentice Hall, Upper Saddle River, NY (1997).
24. T. Poggio and F. Girosi, *Proc. IEEE* **78**, 1481 (1990).
25. R. Hecht-Nielsen, *Neurocomputing*. Addison Wesley, New York (1990).
26. *Matlab*. Mathworks, Natick , MA (1995).
27. B. M. Wise and N. B. Gallagher, *PLS Toolbox for Use with Matlab*. Eigenvector Research, Inc., Manson, WA (1996).
28. H. Demuth and M. Beale, *Neural Network Toolbox for Use with Matlab*. Mathworks, Natick, MA (1995).
29. P. J. Gemperline, J. R. Long and V. G. Gregoriou, *Anal. Chem.* **63**, 2313 (1991).

