COURSE OF OPTIMIZATION

Figure 1. Flow chart for using the "moving least-squares" algorithm [8].

References

[1] Frank, I. E., and Kowalski, B. R., Chemometrics, Anal. Chem. **54**, 232R (1982).

[2] Eckschlager, K., and Stepanek, V., Information Theory as Applied to Chemical Analysis, Wiley, New York (1979).

[3] Eckschlager, K., and Stepanek, V., Anal. Chem. **54**, 1115A (1982).

[4] Eckschlager, K., and Stepanek, V., Analytical measurements and Information, Research Studies Press, Letchworth (1985).

[5] Eckschlager, K., and Stepanek, V., Chemom. Intell. Lab. Syst. **1**, 273 (1987).

[6] Martin, M., Herman, D. P., and Guiochon, G., Anal. Chem. **58**, 2200 (1986).

[7] Schoenmakers, P. J., Chromat. Lib. **35**, Elsevier Science Publishers, Amsterdam (1986).

[8] Lankmayr, E. P., and Wegscheider, W., Automated Optimization of HPLC Separation Methods, Pittsburgh Conference and Exposition, New Orleans, 1985.

[9] Criss, J. W., NRLXRF, A Fortran Program for X-ray Fluorescence Analysis, distrib. through COSMIC, Suite 112 Barrow Hall, Athens, GA (1977).

[10] Otto, M., Wegscheider, W., and Lankmayr, E. P., A Fuzzy Approach to Peak Tracking in Chromatographic Separations, submitted (1987).

[11] Otto, M., and Wegscheider, W., Anal. Chim. Acta **180**, 445 (1986).

[12] Jochum, C., Jochum, P., and Kowalski, B. R., Anal. Chem. **53**, 85 (1981).

[13] Lohninger, H., and Varmuza, K., Anal. Chem. **59**, 236 (1987).

[14] Wold, S., Wold, H., Dunn, W. J., III, and Ruhe, A., Umea University, Report UMINF-83.80, 1982.

[15] Martens, H., Multivariate Calibration: Combining Harmonies from an Orchestra of Instruments into Reliable Predictions of Chemical Composition, 46th Session of the Intern. Statistical Institute, Tokyo (1987).

[16] Hellberg, S., A Multivariate Approach to QSAR, Research Group for Chemometrics, Umea University (1986).

[17] Wold, S., Sjoestroem, M., Carlson, R., Lundstedt, T., Hellberg, S., Skagerberg, B., Wikstroem, C., and Oehman, J., Anal. Chim. Acta **191**, 17 (1986).

[18] Bandemer, H., and Otto, M., Mikrochim. Acta 1986 II, 93.

[19] Blaffert, T., Anal. Chim. Acta **161**, 135 (1984).

[20] Otto, M., and Bandemer, H., Anal. Chim. Acta **184**, 21 (1986).

[21] Otto, M., and Bandemer, H., Chemom. Intell. Lab. Syst. **1**, 71 (1986).

[22] Kateman, G., J. Res. Natl. Bur. Stand. (U.S.) **93**, 217 (1988).

[23] Schostack, K., Parekh, P., Patel, S., and Malinowski, E. R., J. Res. Natl. Bur. Stand. (U.S.) **93**, 256 (1988).

[24] Janssens, K., Van Borm, W., and Van Espen, P., J. Res. Natl. Bur. Stand. (U.S.) **93**, 260 (1988).

[25] Derde, M. P., and Massart, D. L., Anal. Chim Acta **191**, 1 (1986).

Increased Accuracy in the Automated Interpretation of Large EPMA Data Sets by the Use of an Expert System

K. Janssens, W. Van Borm, and P. Van Espen

Department of Chemistry
University of Antwerp (UIA)
Universiteitsplein 1
B-2610 Wilrijk/Antwerp, Belgium

1. Introduction

Characterization of particulate material is one of the major applications of Electron Probe Micro Analysis (EPMA). This involves the collection of an energy dispersive x-ray spectrum for each particle to determine its chemical composition. Since for each aerosol sample typically 1000 particles are measured, very large data sets are obtained. Because of limitations in computer time and mass storage capacity, these spectra are not stored but are processed on-line, i.e., they are converted into tables of peak energies and intensities, permitting

Accuracy in Trace Analysis

the characterization of 1000 particles in 4 hours of instrument time without operator intervention. The off-line data processing consists of the interpretation of the peak table associated with each particle in terms of their chemical constituents. By using the $K\alpha$ or $L\alpha$ lines of each element, a particle vs elemental x-ray intensity matrix is built which is issued as input for multivariate classification techniques.

Since almost all detailed spectral information is lost in the initial data reduction process, qualitative interpretation by a conventional computer program produces erroneous results when peak overlap occurs. However, as human interpreters can obtain better results based on the same limited information, it was decided to capture the additional interpretation knowledge used by the chemist into an expert system, implemented in the OPS5-language [1]. Before the expert system starts an interpretation session, the peak tables generated during the on-line data reduction phase (see table 1a) are converted into a representation which is more suitable for the expert system. For each peak, a library of principal x-ray lines is searched. Since each peak can be associated with several types of x-ray lines of different elements (e.g., a peak at 1.479 keV corresponds to Al-K or Br-$L\alpha$), a (sparse) matrix of possible identifications is obtained (see table 1b). These matrices are read directly by the expert system.

2. The Expert System

Inside the expert system's data base, the data present in each row of an identification matrix are stored into an OPS5 working memory element (WME) of type "PEAK." WME's are complex data structures having several distinct fields. The structure of a peak-WME is represented in figure 1a. As in the identification matrices, each peak can be associated with seven elements. A probability value (e.g., p_{Ka}, p_{Kb}, ...) corresponds with every association. X-ray data pertaining specifically to a chemical element is stored into WME's of type "ELEMENT" (see fig. 1b).

Schematically, the functioning of the expert system is represented in figure 2. The systems production rules are organized in several modules (e.g., CLEAN, ANALYZE, OVERLAP, ...) each dealing with a particular phase of the interpretation. Interaction between the modules is handled by meta-rules. Table 2a lists a meta-rule and its OPS5-

equivalent. At present, the knowledge base contains about 80 chemical knowledge rules. Table 2b lists a rule from the CLEAN module. By using these rules, the system decreases/increases the probability values of a peak as more evidence is found that the associated chemical element is absent/present. The functioning of the expert system is described in more detail elsewhere [2].

Table 1a. Peak table

Peak Number	Intensity (counts)	x-ray energy (keV)
1	167	1.244
2	212	1.479
3	1147	1.742
4	379	3.682
5	1724	4.511
6	301	4.928
7	289	5.894
8	12873	6.392
9	1666	7.043

Table 1b. Identification matrix obtained after data reduction of a spectrum collected from a Mg, Al, Si, Al, Ti, Mn and Fe containing particle

		Possible Identifications				
$K\alpha$	$K\beta$	$L\alpha$	$L\beta1$	$L\beta2$	$L\gamma$	$M\alpha$
Mg	**	As	**	**	**	**
Al	**	Br	**	**	**	**
Si	**	**	**	**	**	**
Ca	**	**	Sn	**	**	**
Ti	**	Ba	**	**	**	**
V	Ti	**	**	**	**	**
Mn	Cr	**	**	**	**	**
Fe	**	**	**	**	**	**
**	Fe	**	**	**	**	**

3. Results and Discussion

The performance of the expert system (method A) was evaluated by comparing the expert system results with those obtained by manual interpretation (method B) and by a conventional FORTRAN interpretation program (method C), using aerosol samples collected in a suburban area [3]. The conventional program operates by summing all peak tables of a data set, yielding a summary spectrum. A set of windows is constructed in which the peak intensities are accumulated while the window positions and widths are continuously adjusted during the summing process. After the summation, each

window is associated with a chemical element by comparing its mean energy with energies of principal x-ray line energies after which the elements associated with each individual particle are determined. Thus, a particle vs elemental x-ray intensity matrix is obtained.

Table 2a. Example of one of the expert systems meta-rules

Rule	**Go back to "CLEAN" phase after "INTENS"**
If	the current interpretation phase is "INTENS" one or more "INTENS"-rules has been used
Then	modify the interpretation phase to "CLEAN" *set the number of used "CLEAN"-rules to zero.*

OPS5-form:
(p intens::control:back-to-clean
{<PHASE> (phase ^name intens ^intenscount > 0)}
→ (modify <PHASE> ^name clean ^cleancount 0))

Table 2b. Example of one of the rules from the CLEAN module, which removes $K\beta$-entries for a given element if no $K\alpha$-entry is found in the data base

Rule	**Remove isolated $K\beta$-entries from the database**
If	the current interpretation session is "CLEAN" a peak is found corresponding to the $K\beta$-line of an element *no peak is found which corresponds to the $K\alpha$-line* of this element
Then	remove the $K\beta$-entry from the data base.

OPS5-form:
(p clean::k:kb-no-ka
{<PHASE> (phase ^name clean ^cleancount <n>)}
{<KB-PEAK> (peak ^kb {<el> <> **})}
--(peak ^ka <el>)
→ (modify <KB-PEAK> ^kb **)
(modify <PHASE> ^cleancount (compute <n> + 1)))

As an illustration, the fine fraction of a data set of 1000 particles is considered below. After performing the qualitative interpretation of the data set in three different ways, the composition of each particle was calculated using a standardless ZAF correction procedure [4]. The resulting three data matrices were subsequently used as input to hierarchical cluster analysis (using Ward's errors sum strategy) to extract information on the different types of particles present in the sample. The resulting dendrograms are shown in figure 3. Although at first glance dendrogram C differs greatly from dendrograms A and B, roughly the same groups of particles can be distinguished. When the soil dust (Si,Al) or gypsum (Ca,S) groups are compared among the three dendrograms, approximately the same mean composition is obtained. However,

when the particles containing heavy metals are considered, significant differences appear:

• In all dendrograms the Pb group consists of two subgroups. In A and B, the first subgroup mainly contains Pb (70%) and Br (24%), the second almost pure Pb (93%). In dendrogram C, however, both groups also contain S, with mean compositions of 60%Pb, 20%Br, 8%S and 82%Pb, 11%S respectively. Clearly, method C could not distinguish between S-$K\alpha$ and Pb-$M\alpha$ while method A could.

• Similarly, in the Zn containing particles, Na is found in all cases by method C since Na-K overlaps with Zn-$L\alpha$ while no Na is found by either the manual or expert system interpretation methods.

• In the V,Ni group, two subgroups also appear, one contains soil dust elements and the other does not (methods A and B). In dendrogram C however, because the interpretation program found Cr in some of these particles, two other groups (containing Cr and not containing Cr) were formed. In this case, the interpretation errors not only yield incorrect particle compositions but have the more important effect of influencing the way particles are clustered together by introducing non-existent correlations.

• In dendrograms A and B, a number of particles of miscellaneous composition is present which do not belong to any of the larger groups. Among them is a group of five Ti particles and one Ba particle. Also pure As and Se particles are present and were identified by both the expert system and the human interpreter. In dendrogram C, however, the As particle belongs to the Pb group while a cluster of six Ba particles is present. Although the significance of these few particles in the entire data set is very small, this shows that the expert system is also capable of handling exotic particles correctly while the conventional program is clearly not.

A significant drawback in the use of the expert system is the considerable amount of computer memory and time it requires. While the conventional program takes about 5 min CPU-time to interpret a data set of 1000 particles. the expert system takes approximately 30 min, depending on the number of peaks in each spectrum.

Conclusion

In this work, an expert system, implemented in the language OPS5, for the automated interpretation of large EPMA data sets is discussed. The

interpretation results were evaluated by comparing them with those obtained from a manual interpretation method and from a conventional interpretation program using a windowing technique. This comparison shows that the results produced by the expert system are in good accordance with the results obtained by manual interpretation since the expert system is able to deal with the frequently occurring case of spectral overlap and retains only physically realistic identification possibilities.
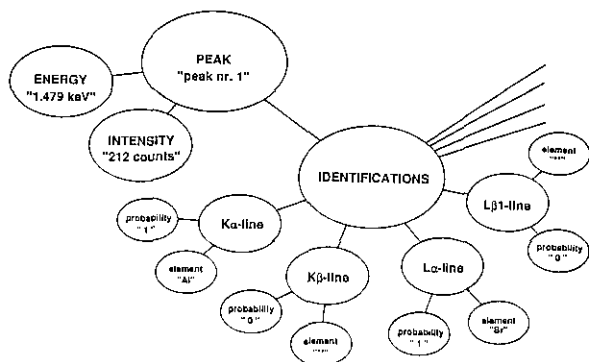
A
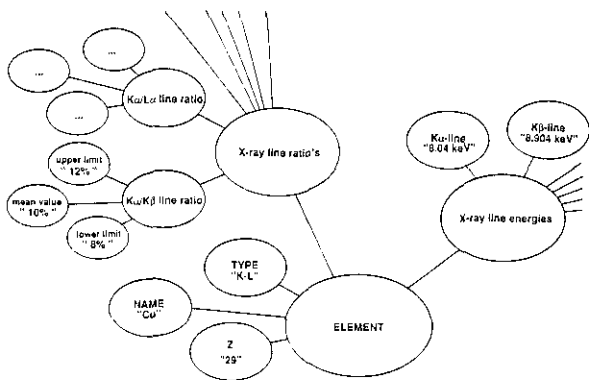
B

Figure 1. Structure of the PEAK and ELEMENT-type of data objects used by the expert system to store peak- and element-specific information.
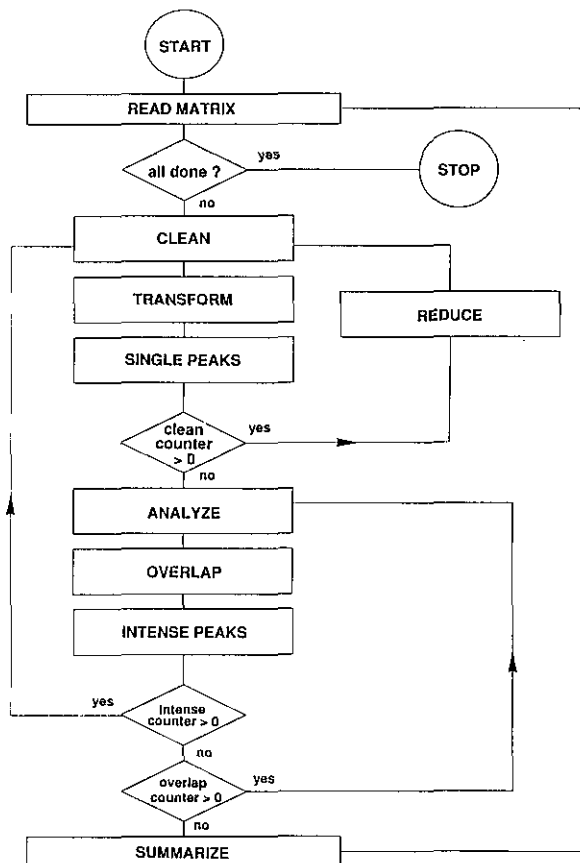

Figure 2. Schematic overview of the interaction between the rule modules present in the expert system's knowledge base.
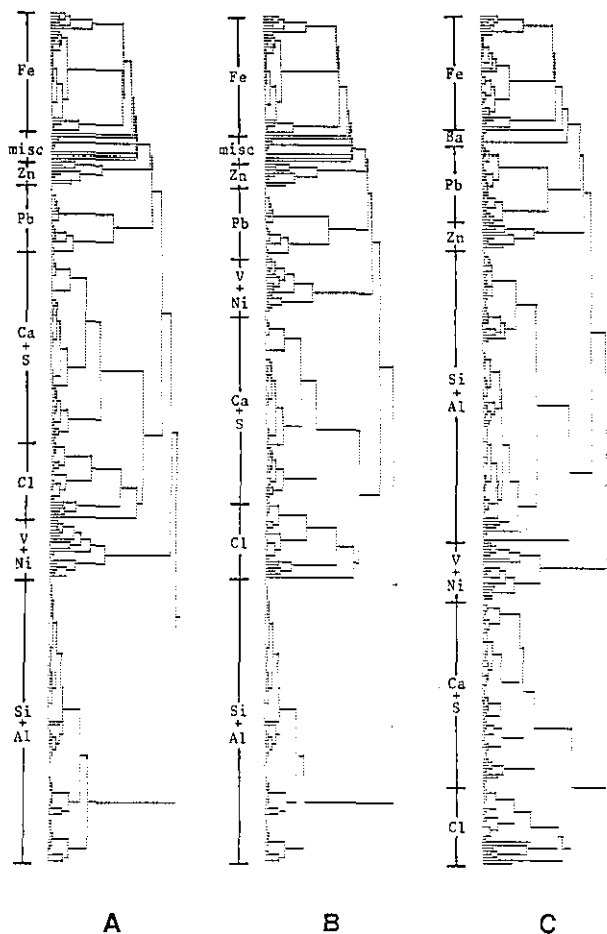
Accuracy in Trace Analysis

Figure 3. Dendrograms obtained by applying hierarchical cluster analysis to the same data set using different interpretation methods: (a) expert system, (b) manual interpretation, (c) conventional program.

References

[1] Forgy, C. L., OPS5 User's Manual, Technical Report CMU-CS-81-135, Dept. of Computer Science, Carnegie Mellon University (July 1981).

[2] Janssens, K., Dorinné, W., Van Espen, P., Chemo. Lab, The development process of an expert system for the automated interpretation of large EPMA data sets, submitted.

[3] Van Borm, W., Adams, F., Maenhaut, W., Environ. Sci. Technol., Source apportionment of air particulate matter in Antwerp, Belgium, submitted.

[4] Raeymaekers, B., Van Espen, P., Adams, F., Anal. Chem., A fast standardless ZAF correction for Electron Probe Micro Analysis, submitted.

Heteroscedastic Calibration Using Analyzed Reference Materials as Calibration Standards

Robert L. Watters, Jr.

National Bureau of Standards
Gaithersburg, MD 20899

and

Raymond J. Carroll and Clifford H. Spiegelman

Department of Statistics
Texas A&M University
College Station, TX 77843

Many instrumental analytical techniques exhibit a definable relationship between instrument response and analyte concentration over wide concentration ranges. This response is usually fit to an accepted model during the calibration phase of the measurement process. Often the calibrated concentration range (x values) is such that the measured response (y values) exhibits non-constant variance. The use of weighted regression techniques to properly estimate model parameters for this case has been described for a number of analytical applications. An inherent, if not stated, assumption in these treatments is that negligible error resides in the concentrations of the calibration standards.

A separate issue regarding calibration is the desire to minimize bias in the analysis by using calibration standards that are matched to the sample to be analyzed. It has been suggested that analyzed reference materials (ARMs) of a chemical matrix similar to that of the sample be used as calibration standards. Since the concentrations of analytes in these materials are estimates from measurements with error, using ARMs as calibration standards leads to errors in both x and y values for fitting the model. Therefore, the standard regression assumptions are not valid. A number of schemes have been developed for treating the calibration problem where both x and y have errors. However, when this problem is combined with heteroscedastic calibration, appropriate procedures are more complex.

We have recently reported an approach to heteroscedastic calibration that yields multiple-use calibration estimates and confidence intervals [1]. The first step is to obtain calibration data from