

MULTIVARIATE DATA ANALYSIS FOR X-RAY FLUORESCENCE SPECTROMETRY

M. BOS

*Technical University Twente, Department of Chemical Technology, P.O. Box 217,
7500 AE Enschede (The Netherlands)*

(Received 9th May 1984)

SUMMARY

A multivariate data analysis procedure that uses singular value decomposition and the Ho-Kashyap algorithm is proposed to obtain calibration constants for x-ray fluorescence spectrometry. These calibration constants can be used to obtain results from experimental data by means of a simple dot product calculation. The method was tested on experimental data from the literature. Comparison of results showed that the method performs at least as well or better than the Raspberry—Heinrich method or its modifications. The method can be used to express calibration results obtained with a theoretically based program in such a way that they can be used conveniently in routine applications.

In multicomponent determinations, the problem is to extract information about the separate components of a mixture from a set of measurements on the sample. For most instrumental methods, this is possible only after calibration against a number of samples of known composition. Depending on the complexity of the sample and the selectivity of the measurements, the complete procedure of calibration and evaluation can vary from the straightforward to the very complicated.

Especially in the field of ultraviolet-visible spectrophotometry and fluorescence measurements, many methods that deal with this problem have been published, e.g., least squares [1, 2], generalized standard addition [3–5], multiple regression [6], Kalman filtering [7, 8], factor analysis followed by multiple regression [9], partial least-square models in latent variables [10, 11], and matrix rank annihilation [12]. The newer methods take care of the problems of highly correlated measurements which give rise to mathematical difficulties in the straightforward least-squares and multiple regression techniques.

Presumably because of the greater inherent selectivity of x-ray fluorescence (x.r.f.), the approach to data analysis in this field has been either of theoretical [13, 14] or empirical [15] nature, in which interactions between the components in the sample during spectrometry are taken into account by means of influence coefficients or correction constants.

In this work, the experimental data given by Budesinsky [16] in a paper on the comparison of the various mathematical methods and the data of Rasberry and Heinrich [15] were treated by a general procedure for multivariate data analysis developed earlier [17], in order to assess the ability of the procedure to correct automatically for interelement effects. The advantage over the method of Rasberry and Heinrich is that the results of a calibration are expressed as a set of calibration constants that can be used in a simple non-iterative procedure (the calculation of a dot product) which allows very fast computation of the results.

The method was also applied to calibration data obtained with a fundamental method by means of the computer program NRLXRF [13, 18], to test its functioning as a fast and convenient method of data evaluation that can be done by hand on a small computer after the NRLXRF calculations have been completed on a mainframe computer.

THEORY

The general theory of multivariate quantitative data analysis with the use of singular value decomposition and the Ho—Kashyap algorithm (SVDHK method) was given earlier [17] but will be restated here in terms of x.r.f. parameters and equations.

Mathematically, the problem is to find from a set of n intensity measurements I_1-I_n on each of m standard samples with known concentrations of the component under study

$$\begin{array}{ll}
 I_{11}, I_{12}, \dots, I_{1n} & C_1 \\
 I_{21}, I_{22}, \dots, I_{2n} & C_2 \\
 \dots & \cdot \\
 \dots & \cdot \\
 I_{m1}, I_{m2}, \dots, I_{mn} & C_m
 \end{array} \tag{1}$$

the evaluation function that relates the measured intensities of an unknown sample to the concentration of the component under study. The models in use for this evaluation function have been surveyed [15]. A general model that includes most types of interactions can be written as [19]

$$\text{Conc.} = \sum_{j=1}^n b_j I_j + \sum_{j=1}^n \sum_{l=1}^n d_{jl} I_j I_l + a \tag{2}$$

where conc. denotes the concentration of the component under study, I is the measured fluorescence intensity at the lines of the various components and b , d and a are the calibration constants.

To calculate the calibration constants, the data matrix has to be extended to include the quadratic and cross-intensity terms:

$$\begin{aligned}
& I_{11}, I_{12}, \dots, I_{1n}, I_{11}^2, I_{11}I_{12}, \dots, I_{11}I_{1n}, I_{12}^2, I_{12}I_{13}, \dots, I_{12}I_{1n}, \dots, I_{1n}^2, 1 \\
& I_{21}, I_{22}, \dots, I_{2n}, I_{21}^2, I_{21}I_{23}, \dots, I_{21}I_{2n}, I_{22}^2, I_{22}I_{23}, \dots, I_{22}I_{2n}, \dots, I_{2n}^2, 1 \\
& \dots\dots\dots \\
& I_{m1}, I_{m2}, \dots, I_{mn}, I_{m1}^2, I_{m1}I_{m2}, \dots, I_{m1}I_{mn}, I_{m2}^2, I_{m2}I_{m3}, \dots, \\
& \quad I_{m2}I_{mn}, \dots, I_{mn}^2, 1 \tag{3}
\end{aligned}$$

If this data matrix is denoted by A , the calibration constants b , d and a by the vector w and the concentrations of the compound under study in the set of standard samples (the training set) by the vector c , the problem is to find the vector w from the matrix equation

$$Aw = c \tag{4}$$

Generally the set of Eqns. 4 is overdetermined and has no exact solution. Moreover, it is very likely that the data matrix will contain a number of dependent or nearly dependent columns. By singular value decomposition and use of the Ho–Kashyap algorithm [20], Eqn. 4 can be solved under these conditions, as described in detail earlier [17].

Solving the set of Eqns. 4 in this way constitutes the training or calibration phase and produces the vector w . This vector can then be used in recognition (i.e., to check how well it reproduces the calibration results) or in prediction of analytical results for unknown samples with the following equation

$$\begin{aligned}
\text{Conc.} = & w_1 I_1 + w_2 I_2 + \dots + w_n I_n + w_{n+1} I_1^2 + w_{n+2} I_1 I_2 \\
& + \dots + w_{n(n+1)/2} I_n^2 + w_{[n(n+1)/2 + 1]} \tag{5}
\end{aligned}$$

RESULTS AND DISCUSSION

The SVDHK procedure applied directly to experimental data

The Fe/Ni/Cr system. The data at 45 kV on the Fe/Ni/Cr system from the paper by Raspberry and Heinrich [15] were used to test the performance of the method applied directly to experimental data. The training set consisted only of the data concerning the ternary mixtures and included the samples treated as unknowns in Table 3 of the paper, except for specimen number 3987. The data on two-component mixtures could not be used, because no intensities were given at the line position of the absent element. For this ternary system the calibration results are expressed as

$$\begin{aligned}
\text{Conc.} = & w_1 I_{\text{Fe}} + w_2 I_{\text{Ni}} + w_3 I_{\text{Cr}} + w_4 I_{\text{Fe}}^2 + w_5 I_{\text{Fe}} I_{\text{Ni}} + w_6 I_{\text{Fe}} I_{\text{Cr}} \\
& + w_7 I_{\text{Ni}}^2 + w_8 I_{\text{Ni}} I_{\text{Cr}} + w_9 I_{\text{Cr}}^2 + w_{10} \tag{6}
\end{aligned}$$

Table 1 shows the values of the vectors w for Fe, Ni and Cr calculated with seven non-zero singular values and the resulting root mean square (RMS) error in the recognition process. This RMS error is calculated from $\sqrt{\Sigma(\text{conc}_{\text{calc}} - \text{conc}_{\text{chem}})^2/10}$.

TABLE 1

SVDHK calibration results for Fe/Ni/Cr data set [15]

Intensity term	Values of $w_1 - w_{10}$		
	Fe	Ni	Cr
I_{Fe}	1.27250	-0.67920	-0.54945
I_{Ni}	-0.58703	0.62360	-0.27812
I_{Cr}	0.17149	-0.59771	0.33422
$I_{Fe}I_{Fe}$	-0.80201	0.17902	0.38542
$I_{Fe}I_{Ni}$	-0.90793	0.70134	0.00483
$I_{Fe}I_{Cr}$	1.61510	-1.29050	-0.54450
$I_{Ni}I_{Ni}$	0.58780	-0.55446	-0.06831
$I_{Ni}I_{Cr}$	0.39872	0.09506	-0.38565
$I_{Cr}I_{Cr}$	-1.01780	0.05392	0.68526
1.0000	0.10682	0.67828	0.32797
RMS error in recognition (% absolute)	0.002	0.002	0.0005

To test the predictive power of the method, the leave-one-out technique was followed, i.e., 10 members of the data set were used as the training set and the concentrations of the eleventh member were calculated with Eqn. 6. The results are given in Table 2. As can be seen, the SVDHK method performs as well as the Raspberry—Heinrich method in all but the very low concentration ranges.

The Cu/Fe/S system. A second ternary test system was found in the data of Budesinsky [16] on copper smelter mattes. Table 3 shows the SVDHK calibration results together with the RMS errors in the recognition process in comparison with the standard deviations obtained by Budesinsky who used the expanded square-intensity method. The results were calculated with six non-zero singular values. Prediction performance for a number of samples taken from various parts of the concentration range, obtained with the leave-one-out technique, is summarized in Table 4. The RMS error is increased somewhat compared to the recognition, but is still very close to the errors obtained by the Raspberry—Heinrich method or its modifications, even though the latter are not given for the leave-one-out situation.

The Cu/SiO₂/Fe/S system. To test the method on a more complicated system, the data of Budesinsky [16] on copper smelter converter slags were used. The calibration results then comprise four 15-element w vectors. With the use of six non-zero singular values, the recognition results listed in Table 5 were obtained. It is clear that when this way of representing the calibration results is used, the data are followed more closely by the six latent variables of the SVDHK method than by the six theoretical interactive constants plus two empirical constants used in the best performing method from the comparisons made by Budesinsky.

TABLE 2

Prediction of the composition of Fe/Ni/Cr samples with the leave-one-out method

Sample ^a no.	Fe (%) SVDHK	Fe (%) ^a Chem.	Absolute error (%)	Fe (%) ^a R + H	Ni (%) SVDHK	Ni (%) ^a Chem.	Absolute error (%)	Ni (%) ^a R + H	Cr (%) SVDHK	Cr (%) ^a Chem.	Absolute error (%)	Cr (%) ^a R + H
5074	68.65	68.38	0.27		4.87	4.98	-0.11		25.16	25.25	-0.09	
5181	68.86	69.45	-0.59		10.11	9.96	1.49		19.94	19.88	-0.06	
5324	53.19	52.80	0.39		18.52	19.27	-0.75		27.13	26.96	0.17	
5321	59.11	59.19	-0.08		20.57	20.02	0.55		19.86	19.88	-0.02	
7271	73.08	71.59	1.49		7.45	8.29	-0.84		18.75	18.79	-0.04	
161	15.09	15.01	0.08		62.58	64.29	-1.71		17.66	16.88	0.78	
1189	4.02	1.40	2.62		75.96	72.60	3.36		20.73	20.30	0.43	
5054	72.12	72.50	-0.38	72.29	0.54	0.15	0.39	0.15	25.86	25.77	0.09	25.22
5202	63.28	63.03	0.25	62.27	14.56	14.80	-0.24	15.07	21.23	21.30	-0.07	20.90
5364	46.62	47.21	-0.59	47.50	24.50	23.57	0.93	24.07	27.67	27.84	-0.17	26.75
1188	6.18	6.60	-0.42	6.55	74.21	72.65	1.56	72.36	14.34	15.40	-1.06	14.21
RMS absolute error (%)			1.0	0.5			1.4	0.4			0.4	1.0

^aFrom Rasberry and Heinrich [15].

TABLE 3

SVDHK calibration results for the Cu/Fe/S system for experimental data from Budesinsky [16]

Intensity term	Values of $w_1 - w_{10}$		
	Cu	Fe	S
I_{Cu}	2.43060	-0.26105	1.71770
I_{Fe}	0.63527	0.41521	0.59317
I_S	1.59250	0.22511	1.42390
$I_{Cu}I_{Cu}$	0.11686	-0.01742	-0.06147
$I_{Cu}I_{Fe}$	1.26370	0.20399	-0.10877
$I_{Cu}I_S$	0.19492	0.18370	-0.01923
$I_{Fe}I_{Fe}$	-0.37549	1.78100	0.17309
$I_{Fe}I_S$	-0.36715	1.59710	-0.04623
$I_S I_S$	0.00324	-0.28706	0.33389
1.0000	0.67779	0.04782	0.56552
RMS error	0.088	0.107	0.092
(% absolute)			
Std. Dev. ^a	0.245	0.211	0.136
(% absolute)			

^aFrom Budesinsky [16].

TABLE 4

Prediction of Cu/Fe/S samples in the leave-one-out procedure

Sample no.	Cu (%)		Fe (%)		S (%)	
	Found	Lit. ^a	Found	Lit. ^a	Found	Lit. ^a
1	32.12	31.81	33.72	33.97	24.95	24.81
2	34.33	33.36	34.17	34.62	26.69	26.98
3	35.44	35.46	32.34	32.07	25.29	25.28
4	36.97	37.06	31.50	31.63	25.35	25.44
10	45.08	45.10	25.75	25.63	25.03	25.10
15	52.05	51.94	20.20	19.97	24.51	24.38
19	58.81	59.34	14.61	14.46	23.45	23.54
RMS error	0.47		0.27		0.15	
(% absolute)						

^aFrom Budesinsky [16].

Table 6 shows the prediction performance of the SVDHK method in the leave-one-out procedure for a number of samples chosen in different regions of the data set. The data on reveratory slags with six components [16] could not be evaluated with the SVDHK method because the w vector then has 28 elements which cannot be calculated from a data set with 20 members.

TABLE 5

SVDHK calibration results for the Cu/SiO₂/Fe/S system for experimental data from Budesinsky [16]

Intensity term	Values of $w_1 - w_{1s}$			
	Cu	SiO ₂	Fe	S
I_{Cu}	0.37521	0.83572	0.97300	0.03764
I_{Si}	0.09310	2.79550	2.10110	-0.07505
I_{Fe}	0.45545	11.07800	10.66700	-0.43682
I_S	0.00758	1.01480	0.78676	0.01660
$I_{Cu}I_{Cu}$	0.25059	-0.66370	-1.50160	-0.10125
$I_{Cu}I_{Si}$	0.37466	-3.12210	-2.62140	0.28917
$I_{Cu}I_{Fe}$	1.57400	-0.41092	0.92427	0.41387
$I_{Cu}I_S$	0.06218	0.49322	-0.09122	-0.02707
$I_{Si}I_{Si}$	-0.00098	3.49980	1.86640	-0.08174
$I_{Si}I_{Fe}$	-0.11268	0.49476	-1.94530	0.12763
$I_{Si}I_S$	-0.01531	-1.62030	-1.73430	0.16992
$I_{Fe}I_{Fe}$	-0.07578	-1.88390	0.00934	0.07295
$I_{Fe}I_S$	-0.14930	0.07106	-0.47556	0.28607
$I_S I_S$	-0.12587	0.99937	0.40766	-0.00982
1.0000	0.19634	4.76170	4.49810	-0.18865
RMS error (% absolute)	0.045	0.285	0.242	0.076
Std. dev. ^a (% absolute)	0.079	0.751	0.325	0.128

^aFrom Budesinsky [16].

TABLE 6

Prediction of Cu/SiO₂/Fe/S samples in the leave-one-out procedure

Sample no. ^a	Cu (%)		SiO ₂ (%)		Fe (%)		S (%)	
	Found	Lit. ^a	Found	Lit. ^a	Found	Lit. ^a	Found	Lit. ^a
1	1.36	1.32	28.55	28.65	47.43	47.47	0.84	0.92
2	1.40	1.45	25.95	26.31	48.97	48.94	0.96	1.11
3	1.82	1.82	25.58	25.18	48.99	49.35	1.01	1.03
4	2.22	2.19	24.80	24.87	49.48	49.16	1.18	1.12
10	4.79	4.74	22.66	22.43	48.61	48.63	1.96	1.90
19	16.22	15.90	18.66	16.78	41.20	41.91	5.45	4.60
RMS error (% absolute)	0.15		0.88		0.43		0.39	

^aFrom Budesinsky [16].

The SVDHK procedure applied to data from a theory-based computer program

The Cu/Fe/S system. The NRLXRF computer program of Criss [18] can generate relative intensity data for samples of known composition and can calculate results from measured intensity data. It is a large program running

on a mainframe computer and is therefore not well suited for routine application on large series of samples or for on-line use. The use of the SVDHK method to represent calibration data generated off-line by the NRLXRF program should make the results of the latter program applicable for routine analysis.

To test this suggestion, relative intensity data for copper, iron and sulphur were generated for samples containing the following elements in the concentration ranges 25–65% Cu, 10–40% Fe, 20–30% S and 10–50% O. Altogether, 300 samples covering these ranges evenly were simulated. Pseudo-random, normally distributed noise with zero mean and a standard deviation of 10^{-5} was added to the intensity data before the SVDHK procedure was entered. The calibration results of the SVDHK procedure with seven non-zero singular values are given in Table 7. It should be noted that these calibration vectors are valid only for the relative intensities generated by the NRLXRF program, which are relative to the pure elements, and for concentrations expressed as weight fractions not as percentages. To use the results on experimental data, the experimentally found intensities have to be converted to this type of relative intensity and the results will be found as weight fractions.

The experimental data of Budesinsky on the reverberatory mattes containing Cu, Fe and S, used above in the direct approach, were converted with the use of the NRLXRF prediction of relative intensities (RI) for the first sample: RI (Cu) = 0.224489, RI (Fe) = 0.413568, RI (S) = 0.107115. Table 8 shows the recalculated relative intensities and the results found with the calibration constants from the SVDHK procedure, compared with the concentrations given by Budesinsky. The RMS errors between the calculated results and the wet chemical results given by Budesinsky [16] are around 1% absolute and are probably mainly due to the deviation between the NRLXRF-generated and experimentally measured data. For instance, sample number 10 is predicted by the NRLXRF program to have relative intensities of 0.336949

TABLE 7

Calibration results for the Cu/Fe/S system for data generated by the NRLXRF program

Intensity term	Values of $w_1 - w_{10}$		
	Cu	Fe	S
I_{Cu}	-1.4682	-0.33434	0.15855
I_{Fe}	-1.1980	-0.29151	0.10975
I_S	2.8775	1.39480	2.35970
$I_{Cu}I_{Cu}$	3.0724	0.49056	0.21352
$I_{Cu}I_{Fe}$	6.9195	2.16510	1.02760
$I_{Cu}I_S$	-2.1887	-0.33772	0.48990
$I_{Fe}I_{Fe}$	1.6765	1.89250	0.27390
$I_{Fe}I_S$	-1.9160	-0.79475	0.37874
$I_S I_S$	2.5858	1.00470	0.78684
1.0000	-0.1662	-0.16219	-0.30525

TABLE 8

Prediction results for the composition of Cu/Fe/S samples based on calibration data from the SVDHK/NRLXRF programs

Sample no.	Cu			Fe			S		
	RI	Found (%)	Lit. ^a (%)	RI	Found (%)	Lit. ^a (%)	RI	Found (%)	Lit. ^a (%)
1	0.224489	30.94	31.81	0.413568	32.17	33.97	0.114600	24.00	24.81
2	0.231637	34.31	33.36	0.411805	33.35	34.62	0.122975	26.78	26.98
3	0.252205	35.23	35.46	0.395410	31.02	32.07	0.115183	25.07	25.28
4	0.264721	37.03	37.06	0.387277	30.45	31.63	0.114934	25.39	25.44
5	0.275469	38.03	38.20	0.377784	29.54	30.45	0.114314	25.48	25.30
6	0.292490	40.14	40.19	0.366548	28.61	29.52	0.112864	25.48	25.21
7	0.310959	41.58	42.21	0.349074	26.96	27.73	0.112321	25.48	25.21
8	0.317179	42.14	42.99	0.343624	26.49	27.23	0.112246	25.54	25.16
9	0.327709	43.81	43.93	0.338154	26.24	26.67	0.112120	25.87	25.14
10	0.338649	44.57	45.10	0.329308	25.33	25.63	0.110409	25.55	25.10
11	0.345435	45.03	45.84	0.322008	24.73	24.99	0.111199	25.75	25.16
12	0.355299	45.39	46.70	0.313039	23.74	24.04	0.109354	25.29	25.15
13	0.368413	46.97	47.92	0.305163	23.18	23.42	0.108275	25.34	24.99
14	0.389709	48.18	49.99	0.286150	21.44	21.70	0.106953	25.07	24.74
15	0.413213	50.51	51.94	0.270591	20.25	19.97	0.104943	24.97	24.38
16	0.429041	51.85	53.32	0.259310	19.35	19.29	0.103641	24.83	24.31
17	0.449609	53.70	54.92	0.245896	18.27	17.96	0.101216	24.49	24.00
18	0.433937	50.25	52.78	0.253691	17.82	18.08	0.095651	22.49	22.67
19	0.503986	57.65	59.34	0.206564	15.21	14.46	0.096872	23.76	23.54
RMS error (% absolute)			1.1			0.9			0.4

^aFrom Budesinsky [16].

for copper, 0.322943 for iron and 0.107115 for sulphur, whereas the experimental data recalculated with the predictions for the first sample are 0.345435 for copper, 0.329308 for iron, and 0.110409 for sulphur. Another indication for this explanation can be found from a comparison between Tables 8 and 4; the RMS error of the results from the direct application of the SVDHK procedure is much lower. Nevertheless, it will be clear that the SVDHK way of summarizing data from NRLXRF simulations can be used successfully if the required accuracy is not below 1% absolute.

REFERENCES

- 1 I. M. Warner, E. R. Davidson and G. D. Christian, *Anal. Chem.*, 49 (1977) 2155.
- 2 H. Gampp, M. Maeder and A. D. Zuberbuehler, *Talanta*, 27 (1980) 1037.
- 3 B. E. H. Saxberg and B. R. Kowalski, *Anal. Chem.*, 51 (1979) 1031.
- 4 C. Jochum, P. Jochum and B. R. Kowalski, *Anal. Chem.*, 53 (1981) 85.
- 5 J. H. Kalivas, B. R. Kowalski, *Anal. Chem.*, 54 (1982) 560.
- 6 D. L. Massart, A. Dijkstra and L. Kaufman, *Evaluation and Optimization of Laboratory Methods and Analytical Procedures*, Elsevier, Amsterdam, 1978, p. 296.
- 7 H. N. J. Poullisse, *Anal. Chim. Acta*, 112 (1979) 361.
- 8 C. B. M. Didden and H. N. J. Poullisse, *Anal. Lett.*, 13 (1980) 921.
- 9 H. Martens, *Anal. Chim. Acta*, 112 (1979) 423.
- 10 R. W. Gerlach, B. R. Kowalski and H. O. A. Wold, *Anal. Chim. Acta*, 112 (1979) 417.

- 11 M. Sjostrom, S. Wold, W. Lindberg and J. Persson, *Anal. Chim. Acta*, 150 (1983) 61.
- 12 C. N. Ho, G. D. Christian and E. R. Davidson, *Anal. Chem.*, 53 (1981) 92.
- 13 J. W. Criss, L. S. Birks and J. V. Gilfrich, *Anal. Chem.*, 50 (1978) 33.
- 14 N. Broll and R. Tertian, *X-Ray Spectrom.*, 12 (1983) 30.
- 15 S. D. Rasberry and K. F. J. Heinrich, *Anal. Chem.*, 46 (1974) 81.
- 16 B. W. Budesinsky, *X-Ray Spectrom.*, 12 (1983) 121.
- 17 M. Bos and G. Jasink, *Anal. Chim. Acta*, 103 (1978) 151.
- 18 J. W. Criss, *NRLXRF, A Fortran Program for X-Ray Fluorescence Analysis*, Naval Research Laboratory, Washington, DC, 1977.
- 19 H. J. Lucas-Tooth and C. Pyne, *Adv. X-Ray Anal.*, 7 (1964) 523.
- 20 Y. C. Ho and R. L. Kashyap, *IEEE Trans. Electron. Comput.*, EC-14, No. 5, (1965) 683.