JAAS
Journal of Analytical Atomic Spectrometry

Variable selection and multivariate calibration models for X-ray fluorescence spectrometry

M. J. Adams* and J. R. Allen

School of Applied Sciences, University of Wolverhampton, Wulfruna Street, Wolverhampton WV1 1SB, UK

Application of partial least squares models to quantitative XRF analysis is shown to be dependent on the efficiency of the variable selection routine used to choose the spectroscopic variables employed in the calibration model. Three variable selection techniques have been examined, based on cumulative covariance of emission intensity with analyte concentration, signal-to-noise ratio of linear sensitivity of individual variables to error variance and signal-to-noise ratio of logarithmic sensitivity of individual variables to error variance. Applying the models to a wide variety of analytes in a range of sample types, logarithmic scaling of the sensitivity term in the signal-to-noise ratio is demonstrated as providing sensible and efficient selection of appropriate variables. Results obtained are comparable to those using element selective correction models such as Lucas–Tooth and Price, which require user knowledge of the sample and analytical conditions.

Keywords: *X-ray fluorescence analysis; partial least squares regression; simple partial least squares; variable selection; modelling*

X-ray fluorescence spectrometry is a widely used technique for both qualitative and quantitative analysis, particularly for solid samples, and is a popular choice often because of the limited sample preparation required and the rapid rate of analysis. However, even when using high resolution wavelength dispersive spectrometers, the accuracy and precision of quantitative analyses may need to be supported by other techniques, such as plasma emission spectrometry or atomic absorption spectrometry, both of which usually require dissolution of the sample prior to analysis.

The two main types of interference in quantitative XRF analysis are spectral overlap and inter-elemental matrix effects. Because of the relative simplicity of XRF spectra, compared with, for example, ICP-OES, spectral interference due to emission band overlap is rarely severe, and the more common cases are well documented. Matrix effects, on the other hand, can lead to severe problems in quantitative analysis, and can arise from the chemical composition and physical nature of the sample. The recorded emission intensity from a constituent element can be influenced by absorption effects from other species within the sample, and enhanced by stimulated emission from other elements. The combination of excitation and absorption mechanisms in multi-component systems gives rise to matrix effects that complicate quantitative analysis, and a recorded XRF spectrum can rarely be described as a simple sum of constituent elemental spectra.[1]

In the most simple case, quantitative analysis follows the classic procedure of producing a working curve from the measured intensity of a characteristic line, usually the most intense present, as a linear function of analyte concentration in a range of standard samples. Owing to interference effects, however, the use of a simple univariate linear function is often not viable and alternative methods have to be employed, ranging from the use of empirical coefficients with complex polynomial regression equations to fundamental parameter models based on a detailed knowledge of the sample and instrument characteristics.[1–4]

Interference effects can be minimised by appropriate selection and/or treatment of the samples and standards. Thus, selection of suitable standards to achieve matrix matching with the unknown sample is common, but does rely on the sample being well characterised prior to analysis and suitable standards being available. Where suitable matched standards are not available, or are difficult to prepare, then global correction methods such as borate fusion techniques can be employed. By diluting the sample in a low density flux, inter-element effects are reduced and standard additions methods can be used to achieve a suitable range of reference concentrations.[4] Obviously such techniques are time consuming and not always applicable.

The current situation, therefore, is that the XRF user has an apparent wide range of sampling techniques and models available for quantitative analysis. In most cases, however, the analysis requires considerable expertise from the user and extensive knowledge of the sample composition. The aim of the present study is to evaluate partial least squares regression (PLSR) multivariate calibration with prior, automatic variable selection for quantitative XRF analysis. PLSR has been reported extensively in the literature for application with, for example, molecular, broad-band spectra. Its application to-date in analysing atomic line spectra has been more limited, although as demonstrated here, its use can provide considerable benefit. Wang *et al.* have discussed the use of PLSR with WDXRF data and demonstrated its application, but commented on the need for selection of appropriate spectral regions used for calibration.[5] Urbanski and Kowalska, and Swerts *et al.*, have reported the application of PLSR with low resolution EDXRF spectra.[6,7]

VARIABLE SELECTION AND CALIBRATION MODELS

Least squares linear regressions models are widely available and commonly employed for XRF analysis. They relate analyte concentration, y, to measured emission intensity, x, according to one of two forms

$$y = a_0 + \sum a_i x^i \quad (1)$$

in which i denotes the order of the polynomial, usually no more than cubic, used to fit intensity to concentration, or

$$y = a_0 + \sum a_i x_i \quad (2)$$

where i denotes the number and identity of the emission lines considered to be appropriate in predicting concentration; a_0 and a_i are the regression coefficients corresponding to intercept and local gradient terms, respectively.

The univariate model, eqn. (1), may be acceptable with simple matrices with little or no inter-element effects. Even in these cases a straight line, first-order, fit may not be sufficient to describe the data adequately and, commonly, the user may select a quadratic or cubic equation. Although higher order polynomials than first-order may fit the calibration data better, the efficacy of the models for prediction with unknown samples could be poorer.

The use of univariate, polynomial calibration models is

rarely sufficient to overcome interference effects and non-linear relationships between analyte concentration and emission intensity. Where such effects are present, correction terms based on empirical influence coefficients can be developed and applied in order to alleviate these problems. This technique is typified by such models as that associated with Lucas–Tooth and Price (see ref. 1), which is a form of eqn. (2). For a sample containing two elements which can interfere with the analyte emission intensity, then the model is

$$y_i = a_0 + x_i[a_i + a_{ii}x_i + a_{ij}x_j + a_{ik}x_k] \quad (3)$$

where y and x are concentration and intensity terms, respectively, the subscripts i, j and k refer to the analyte and the two known interfering species. The model is comprised of linear, quadratic and cross-product terms.

Lucas–Tooth and Price models, or similar functions, are commonly available with commercial spectrometer software and, not surprisingly, generally give rise to better predictive models than simple polynomial functions. However, as can be readily appreciated, the use of inter-element correction functions, such as implied in eqn. (3), rely on the user selecting the number of potentially interfering species and the most appropriate emission lines to use in the correction model, *i.e.* the user is required to have extensive knowledge of the sample and awareness of the mechanisms leading to interference.

Multivariate calibration models potentially requiring little user input or decisions include multiple linear regression, principal components regression and PLSR. These methods have been widely exploited in molecular spectroscopic studies for overcoming spectral band overlap and background interference. Their efficient application to XRF analysis would be enhanced by a suitable automatic variable (wavelength) selection algorithm prior to data analysis. A digitised XRF spectrum consists of over 1000 datum values, and contains a high degree of colinearity and considerable redundancy with 'white-space' between spectral peaks. It is to be expected, therefore, that an efficient variable selection algorithm could reduce considerably the dimensionality of the data, with minimal loss of analytical information, prior to calibration modelling. In a set of sample spectra, those spectral regions containing maximum analytical information might be expected to exhibit greatest variance, and simply ranking variables according to variance can provide a rapid means of ordering the variables, with some arbitrary number of the top variables selected for modelling. Since this scheme can lead to variables being selected which are not characteristic of the analyte, it is better to rank the variables by covariance with the analyte and selecting a subset for further analysis. Since covariance is proportional to variable variance and correlation, then this procedure results in a relatively small set of variables exhibiting maximum variance and greatest correlation with the analyte.

An alternative and more sophisticated variable selection technique has been proposed by Brown *et al.* for use in near-IR spectroscopy.[8,9] At all wavelengths in the XRF spectrum a simple linear model relating analyte concentration, y, to fluorescence emission, x is assumed,

$$y = b_{0,k} + b_{1,k} + \varepsilon_k \quad (4)$$

where $b_{1,k}$ represents the slope, or sensitivity, for wavelength k ($k = 1 \ldots K$) for the analyte. The error associated with using this model, ε_k, has a variance of ${\sigma_k}^2$. The method of variable selection adopted by Brown *et al.* chooses K' wavelengths ($K' \ll K$) to correspond to the K' largest values of signal-to-noise ratio, s, given by

$$s = \frac{{b_{1,k}}^2}{{\sigma_k}^2} \quad (5)$$

The selection process is achieved using eqn. (6)

$$Z = \frac{C_{(\eta)}}{\sqrt{\sum_{k=1}^{K'} \left(\frac{{b_{1,k}}^2}{{\sigma_k}^2}\right)}} \quad (6)$$

where $(1 - \eta)$ is the confidence level and

$$C_{(\eta)} = \sqrt{({X_{1-\eta}}^2[k])} \quad (7)$$

is the square root of the tabulated chi-squared values on k degrees of freedom.

Since both $C_{(\eta)}$ and the denominator of eqn. (6) are monotone-increasing in k, a compromise is reached in practice, and the number of wavelengths is a monotonically decreasing function of the significance level η. To choose the value for K', and the corresponding spectral wavelengths, the calculated s values are ordered from largest to smallest and the function Z [eqn. (6)] is minimised. The typical form for eqn. (6) is illustrated in Fig. 5. Brown commented on the intuitively appealing ordering and selection based on the s ratio. This ratio will only be large at wavelengths linearly sensitive to analyte concentration. Departures from the model of eqn. (4), resulting from non-linearity or interaction effects will inflate ${\sigma_k}^2$, deflate the ratio and, hence, reduce the likelihood of selecting that wavelength. However, as demonstrated below, because of the wide dynamic range of the analytical signal compared with absorption techniques, the values for s with line emission spectra are dominated by the most intense lines, and the selection procedure is improved by logarithmic scaling of the sensitivity coefficient, *i.e.,*

$$s' = \frac{\log(b_{1,k})^2}{{\sigma_k}^2} \quad (8)$$

Following data reduction by variable selection, multivariate calibration models can be applied directly. PLSR has received considerable attention in the chemometrics literature and detailed discussions and algorithms are provided by, amongst others, Martens and Naes.[10] The most widely used implementation is the non-linear iterative partial least squares (NIPALS), PLS1 algorithm, which proceeds by iterative decomposition of the data matrix as each relevant factor is extracted.[10] More recently alternative algorithms have been published, including a computationally efficient algorithm from de Jong, referred to as simple partial least squares (SIMPLS).[11] This algorithm avoids deflection of the full data matrix, and for unicomponent analyses SIMPLS has been shown to yield equivalent results to PLS1.[11]

EXPERIMENTAL

Materials and methods

Spectra were recorded using an ARL Model 8410 wavelength-dispersive, scanning spectrometer (Thermo-Fi, Crawley, Sussex, UK) measuring emission intensity (counts s^{-1}) as a function of wavelength, or scanning angle 2θ. Digitisation was conducted at a 0.1° interval in the range 17–154.5° using LiF(200) and poly(ethylene terephthalate) (PET) crystals. Different dispersing crystals were used depending on the elements studied: aluminium and silicon with PET, and elements with atomic number above that of potassium with LiF(200). Each LiF and PET spectrum consisted of 1375 datum values.

A wide range of solid sample, standard materials were examined, including iron-based alloys, copper-based alloys and geological samples obtained from the Bureau of Analysed Standards (Middlesborough, Cleveland, UK) and MBH Analytical (Barnet, Hertfordshire, UK). Prior to analysis the brass and steel samples were cleaned with ethanol to remove surface grease and dirt. Geological standards were homog-

enised with bakelite powder (85+15% by mass, respectively), pressed into aluminium cups at 10 tonnes pressure and allowed to cure in a warm oven (60 °C) overnight to yield a durable sample disk. This range of samples provides a variety of material types, both low and high density matrices, of varying elemental composition.

All software, data analysis and calibration models were coded in-house using Mathematica (V3.0, Wolfram Research, Champaign, IL, USA).

RESULTS AND DISCUSSION

Fifteen certified iron alloy standards were analysed for manganese (0.12–1.5%), nickel (0.069–2.2%), chromium (0.13–24%), iron (70.05–98.6%) and silicon (0.033–1.39%).

Twenty certified copper alloy samples were analysed for copper (58.5–99.16%), zinc (0.01–39.5%), lead (0.012–5.2%) and nickel (0.01–3.33%).

Twenty five geological certified samples, comprising soils and igneous and metamorphic rocks, were analysed for aluminium (3.06–31.25%), silicon (13.20–34.34%), titanium (0.0024–0.970%), manganese (0.0007–0.596%), iron (0.04–10.03%), potassium (0.099–10.88%) and calcium (0.029–14.64%).

Each group of samples was analysed using a miss-one-out strategy and using calibration models for univariate ordinary least squares regression (OLSR), Lucas–Tooth and Price (LTP) and PLSR. For each model evaluated all samples less one were used to derive the model coefficients, the odd sample used as a prediction or validation sample and the process repeated with each sample in turn being used for prediction purposes. This strategy ensures that all samples are used in both model development, *i.e.*, calibration, and model validation, *i.e.* prediction. To compare the relative merits of the calibration models the accumulated root-mean-square error of prediction (RMSEP) was employed:

$$RMSEP = \sqrt{\frac{\sum_{i=1}^{n} (\hat{y}_i - y_i)^2}{n}} \qquad (9)$$

where $(\hat{y}_i - y_i)$ is the residual between predicted analyte value and actual, or reference, analyte value, and n is the total number of samples examined for each analyte. By using RMSEP values then the models can be compared according to their prediction performance rather than solely on goodness-of-fit to the calibration data.

Detailed consideration of the analysis for copper in brass will serve to provide an example of the data processing and manipulation strategies employed. A plot of predicted *versus* actual copper concentration (by OLSR) is presented in Fig. 1(*a*), and it is evident that the simple, univariate calibration model is inadequate for quantitative analysis. Potential sources of error in analysing for copper by XRF are inter-element effects due to zinc and tin in the sample.[1,4] Selective correction models such as that of Lucas–Tooth and Price aim to compensate for such interferents by including specific terms in the regression model. In the case of copper in brass alloys, and correcting for the presence of zinc and tin, the calibration model may be expressed as

$$[Cu] = a_0 + x_{Cu}[a_1 + a_2 x_{Cu} + a_3 x_{Zn} + a_4 x_{Sn}] \qquad (10)$$

where $[Cu]$ represents the predicted concentration of copper, x_{Cu} the intensity of the copper Kα emission line, x_{Zn} the intensity of the zinc Kα emission line and x_{Sn} the intensity of the tin Lα emission line. Application of eqn. (10) leads to a marked improvement in the predictive quality for copper, as illustrated in Fig. 1(*b*).

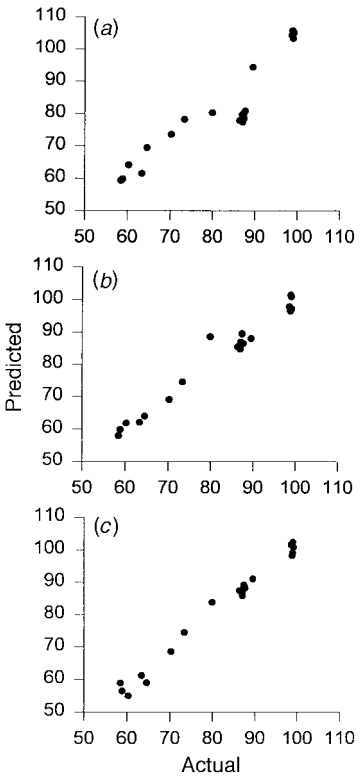

Fig. 1 Predicted *versus* actual certified values for the concentration of copper in a series of copper-based alloys determined by: (*a*) univariate, OLSR; (*b*) Lucas–Tooth and Price inter-element correction model; and (*c*) application of variable selection by minimisation of the Z function [eqn. (6)].

Remaining errors in the residuals from Fig. 1(*b*) could be due, in part, to physical effects relating to the surface finish on the analysed samples. Visual examination of the surface of the copper alloy samples indicated that several had scratches and score marks that could shield secondary copper fluorescence emissions. To compensate for such effects, and other inter-element interferences, further terms can be added to the calibration model, but the range of potential additional correction factors soon becomes increasingly complex and the model unwieldy. Instead, and in common with other spectroscopic techniques, so-called 'full-spectrum' models can be applied, such as PLSR; the efficacy of which is enhanced by suitable variable selection pre-processing.[5]

The average log(intensity) *versus* 2θ, LiF(200) spectrum from 20 copper-based alloys is shown in Fig. 2, and is characterised by a broad-band emission envelope superimposed with sharp-band atomic emission lines. The absolute covariance values of emission intensity with copper concentration is illustrated graphically in Fig. 3(*a*) with the highest values centred about identifiable 2θ values. A plot of cumulative covariance from a ranked list of all covariance values is presented in Fig. 3(*b*) and it can be observed that 95% of the summed covariance is contained in just 53 variables. This list of 53 variables groups into nine distinct spectral regions which are shown and identified in Fig. 2(*b*). In addition to the first-order Cu Kα line, the 95% summed covariance limit has also selected first-order Cu Kβ, Zn Kβ, Sn Lα, Fe Kα, Mn Kα and second-order Cu Kα and Zn Kα transitions.

PLSR using the SIMPLS algorithm and miss-one-out strategy was applied to the 53 variables contributing to the top 95% of the cumulative covariance. Computed RMSEP values for copper as a function of the number of factors extracted are displayed in Fig. 4 and a minimum RMSEP value is evident with a four-factor model. This result (RMSEP = 1.13) is

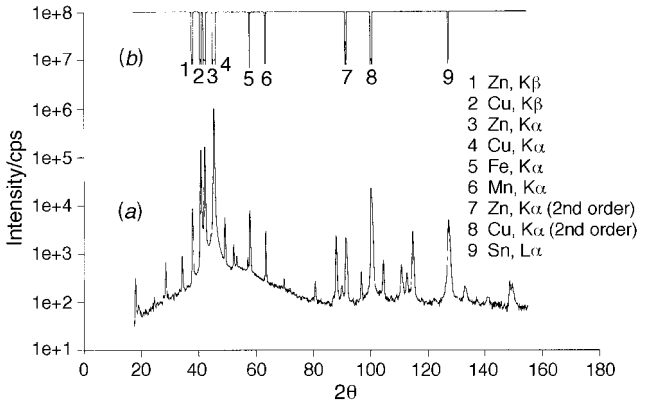

Fig. 2 Average spectrum [log(intensity) *versus* 2θ angle] for the copper-based alloys recorded using an LiF(200) crystal (*a*). The spectral regions selected as contributing to the 95% of the cumulative, ranked covariance values are shown in (*b*).

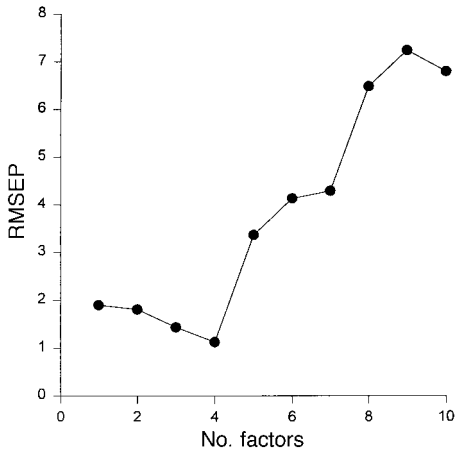

Fig. 4 RMSEP values using the miss-one-out strategy for the determination of copper in copper-based alloys and the SIMPLS calibration model as a function of the number of factors extracted by the model.

superior to that obtained using the inter-element correction model of eqn. (10) (RMSEP = 2.47) and a considerable improvement on that using a single variable, OLSR model (RMSEP = 6.10).

The second variable selection technique examined is minimisation of the function Z, eqns. (5) and (6). For copper in the 20 copper-based alloys, the form of Z, with a 1% significance level ($\eta = 0.01$), is illustrated in Fig. 5, with the minimum indicated by five variables, identified as centred on the Cu Kα and Zn Kα emission lines. Using these five variables to describe the spectra and the PLSR model, then a minimum RMSEP value of 1.79 is obtained with two factors. Although this value is greater than that obtained using selected maximum covariance and PLSR, it indicates a better model than both univariate OLSR and the empirical polynomial model; the results are illustrated in Fig. 1(*c*).

The situation is far from satisfactory, however, if a minor constituent is examined, *e.g.*, lead in copper alloys, rather than the major components. In this case the results using the two variable selection algorithms and PLSR are inferior to applying the simple OLSR model, Table 1A. The reason for this is clear when the identities of the variables selected are examined. By applying the 95% limit to the cumulative ranked covariance values, then 66 variables are selected centred about not only the Pb Lα line, but also emission lines from copper, zinc, iron, nickel and tin. Although most of these are only weakly correlated to lead concentration, the high intensity of these lines dominates the covariance structure. The effect of false correlation is more strikingly illustrated using variable selection from the signal-to-noise ratio and eqn. (6). In this case, Z is a minimum with only five variables which centre about the Cu Kα and Zn Kα lines; the lead emission lines are not even selected and used in the model.

In order to overcome this problem and reduce the effect of major components in samples dominating the variable selection procedure, eqn. (5) was modified as described above [eqn. (8)] to scale logarithmically the linear sensitivity coefficient. By this technique the minimum Z indicated two variables for lead and these were both indicative of the Pb Lα line.

The three variable selection methods, with PLSR calibration, were applied to the range of analytes in copper alloys, iron alloys and geological samples. The results are summarised in Table 1 along with the results from OLSR models and, where

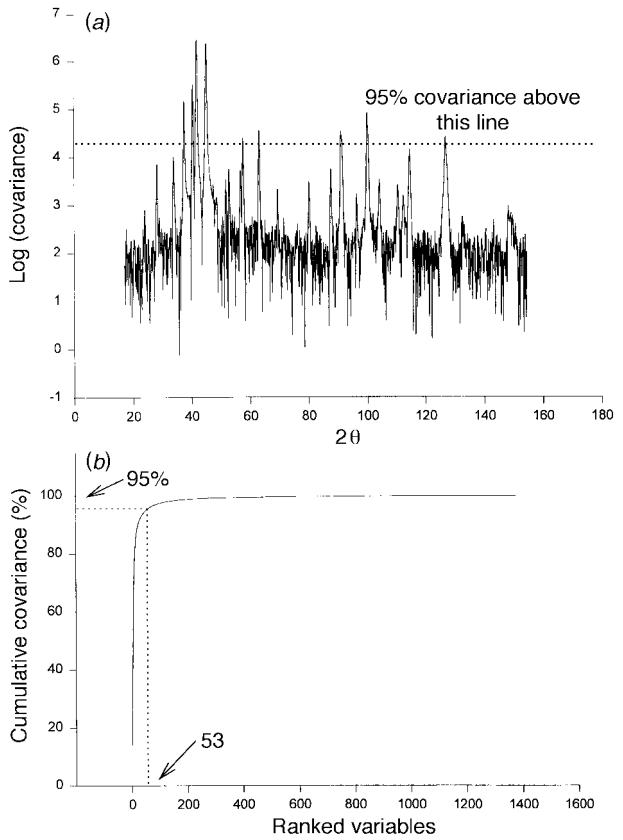

Fig. 3 (*a*) Absolute covariance values of intensity with copper concentration as a function of spectral angle. (*b*) Cumulative sum of ranked covariance values for copper in copper-based alloys.

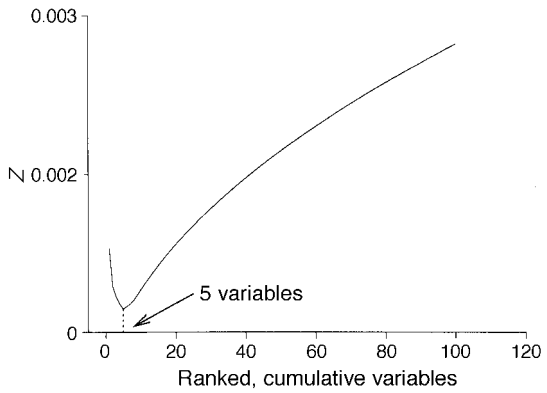

Fig. 5 The form of Z [eqn. (6)] for the first 100 ranked variables for copper in copper-based alloys, illustrating the minimum value employed to select the variables for subsequent modelling.

Table 1 RMSEP values for the analytes examined in a range of sample types. The calibration models employed are univariate OLSR, LTP using emission lines from the elements indicated for inter-element correction and PLSR. For PLSR with variable selection using the logarithmically scaled sensitivity coefficient [eqn. (8)] the number of variables selected is given along with identification of the emission lines associated with these variables as well as the number of factors extracted by the SIMPLS algorithm giving the minimum RMSEP value

	Analyte and range (%)						
A. *Copper alloys* —							
	Cu	Zn	Ni	Pb			
	58.50–99.16	0.01–39.5	0.01–3.33	0.012–5.20			
OLSR	6.10	0.79	0.10	0.24			
LTP	2.47	0.45	0.03	N/D*			
Interfering elements	Zn, Sn	Cu, Sn	Fe	—			
PLSR†	1.13	0.55	0.39	0.96			
PLSR‡	1.79	0.59	0.05	1.16			
PLSR§	1.80	0.57	0.06	0.08			
Number of factors	2	1	2	1			
Number of variables	17	5	5	2			
Lines selected	Zn Kα, Cu Kα, Zn Kβ, Cu Kβ	Zn Kα	Ni Kα	Pb Lα			
B. *Iron alloys* —							
	Fe	Cr	Mn	Ni	Si		
	70.05–98.6	0.13–24.0	0.12–1.50	0.07–2.20	0.03–1.39		
OLSR	1.87	0.73	0.04	0.16	0.16		
LTP	2.91	0.09	0.05	0.13	N/D		
Interfering elements	Ni, Cu	Mn, Fe	Co, Ni	Fe	—		
PLSR†	1.47	0.68	0.17	0.50	0.03		
PLSR‡	1.78	0.68	0.03	0.50	0.12		
PLSR§	1.79	0.61	0.04	0.10	0.09		
Number of factors	1	2	1	1	2		
Number of variables	12	9	4	5	8		
Lines selected	Fe Kα, Fe Kβ, Fe Kα (2nd order)	Cr Kα, Cr Kβ	Mn Kα	Ni Kα	Si Kα		
C. *Geological samples* —							
	Al	Si	Ti	Mn	Fe	K	Ca
	3.06–31.25	13.2–34.34	0.0024–0.97	0.0007–0.60	0.04–10.03	0.10–10.9	0.03–14.64
OLSR	0.84	1.43	0.14	0.01	0.80	0.17	1.28
LTP	N/D	N/D	0.05	0.01	0.45	0.13	1.51
Interfering elements	—	—	Cr	Co, Ni	Ni, Cu	Ca	Ti, Sc
PLSR†	0.62	1.10	0.23	0.10	0.71	0.23	0.58
PLSR‡	0.47	1.42	0.23	0.02	0.76	0.17	0.53
PLSR§	0.52	1.42	0.04	0.01	0.76	0.17	0.53
Number of factors	1	1	1	3	1	2	1
Number of variables	14	15	6	4	14	6	4
Lines selected	Al Kα	Si Kα	Ti Kα, Ti Kβ	Mn Kα	Fe Kα, Fe Kβ, Fe Kα (2nd order)	K Kα, K Kβ	Ca Kα

* N/D, no LTP model developed.
† PLSR using ranked cumulative covariance for prior variable selection.
‡ PLSR using ranked signal-to-noise ratio for prior variable selection (ref. 8).
§ PLSR using ranked modified signal-to-noise ratio for prior variable selection (see text).

appropriate, application of a Lucas–Tooth and Price inter-element correction polynomial model. Modifying the variable selection technique of Brown[8] [eqn. (8)] can be seen to provide a simple, spectroscopically interpretable and efficient choice of variables for subsequent modelling by PLSR. The procedure has provided results consistently better than application of the univariate OLSR model, and which are generally comparable to those obtained using more user demanding schemes such as Lucas–Tooth and Price models.

CONCLUSION

From the results obtained and summarised in Table 1, it can be seen that the efficient and effective automated application of PLSR models to XRF data for calibration and quantitative analysis relies on suitable variable selection strategies. For the range of analytes and sample types examined here, covering both light and heavy elements in light and dense sample matrices, the best overall technique of those studied and developed is based on the application of the signal-to-noise ratio proposed by Brown[8] and modified by scaling the sensitivity coefficients to prevent major components and false correlation from dominating the variance structure of the data. Although sometimes giving slightly inferior results to those obtained from application of user-selected correction terms in polynomial models such as that of Lucas–Tooth and Price, the latter techniques rely heavily on user expertise and prior knowledge of the analytical technique and the sample under study. It is anticipated that the results demonstrated here could lead to the development of a variable selection and calibration scheme that can be incorporated into an automatic software analysis program.

REFERENCES

1 Lachance, G. R., and Claisse, F., *Quantitative X-Ray Fluorescence Analysis*, John Wiley, Chichester, 1995.
2 Heinrich, K. F. J., and Rasberry, S. D., *Anal. Chem.*, 1974, **46**, 81.
3 Leyden, D. E., and Gilfrich, N. L., *Trends Anal. Chem.*, 1988, **7**, 327.
4 Jenkins, R., and De Vries, J. L., *Practical X-Ray Spectrometry*, Macmillan, London, 1970.

5 Wang, Y., Zhao, X., and Kowalski, B. R., *Appl. Spectrosc*., 1990, **44**, 998.
6 Urbanski, P., and Kowalska, E., *X-Ray Spectrom*., 1995, **24**, 70.
7 Swerts, J., Van Espen, P., and Geladi, P., *Anal. Chem*., 1993, **65**, 1181.
8 Brown, P. J., *J. Chemom*., 1992, **6**, 151.
9 Brown, P. J., Spiegelman, C. H., and Denham, M. C., *Philos. Trans. R. Soc. London, A*, 1991, **537**, 311.
10 Martens, H., and Naes, T., *Multivariate Calibration*, John Wiley, Chichester, 1996.
11 de Jong, S., *Chemom. Intell. Lab. Syst*., 1993, **18**, 251.

Paper 7/06212E
Received August 26, 1997
Accepted October 2, 1997