MULTIVARIATE CALIBRATION IN SPECTROSCOPY: APPROACHING THE LIMITS OF DETECTION

WOLFHARD WEGSCHEIDER¹ GOTTFRIED BAUER¹ PETER KARPATI¹ and HUGO M. ORTNER²¹Institute of Analytical Chemistry, Micro- and Radiochemistry, Graz University of Technology, A-8010 Graz, Austria²Materials Sciences, Technical University Darmstadt, Petersenstraße 21, D-6100 Darmstadt, Federal Republic of Germany

Abstract - A complete error model for (locally) linear systems is outlined that was recently introduced[7]. It takes into account errors of concentration values in standards, errors in measurements of standards and errors in measurements of unknowns. It is applied to the analysis of bauxites and of high purity tungsten. This error model can also be used to estimate realistic detection limits in presence of interferences that depend on the selectivity of the sensors and on the concentration levels of interferences.

Key words multivariate calibration, error, detection limit, Kálmán filter, tungsten, bauxite, inductively coupled plasma, optical emission spectrometry

Multivariate calibration utilizing the information of several analytical signals simultaneously can serve two purposes: to correct for interferences and to accurately determine several analytes at a time[1]. This is done in a variety of ways and applies to very diverse techniques like spectrophotometric determinations of metal ions[2], ion selective electrode measurements[3], chemical sensing[4], infrared spectrometry[5] and many others.

Although the exact scheme of performing multivariate measurements depends strongly on the analytical technique, certain of steps are common to all. In short, these are: (a) identify at least as many sensors (wavelengths) as the number of interferences and analytes; (b) select or prepare a sufficient number of samples to account for the anticipated variability of interferences and analytes (jointly these two will be called 'components'); (c) select the particular (mathematical) model useful for correction; (d) determine sensitivities of all components at all sensors (wavelengths); (e) apply this model to analyze the unknowns. Multivariate calibration is inherently more relevant for methods with low selectivity, such as NIR or spectrophotometry, but also methods considered highly selective, such as optical emission spectrometry, can benefit from it: particularly in ultratrace analysis when extreme differences in signal intensities of matrix constituents and analytes are encountered there is hardly an analytical method that can be regarded to be of sufficient selectivity. Generally, in optical emission this problem is treated by overlap correction techniques, most of which give inadequate results for structured background. Such a case was recently encountered in the determination of traces of boron in high-purity tungsten[6] and then approached by a digital filtering technique for background correction.

Recently, overlap and background correction was developed on the basis of a general linear model. In these studies it was necessary not only to determine trace elements with good accuracy, but also to estimate reliably the random errors made in this novel procedure. This led to the development of an error estimation procedure for multivariate calibration[7] that can be easily extended to estimate also multivariate detection limits[8]. Here, a brief account of this work is given along with some new results on the selection of multiple wavelengths in inductively coupled plasma optical emission spectrometry. It is stressed that a reliable estimation of detection limits in presence of interferences is only possible by establishing a proper error models for ALL components and ALL sensors (wavelengths) at low concentrations, thus close to the limit of detection.

THEORY

In the framework of a linear model multivariate calibration can be expressed as a two-stage process: calibration to find the coefficients of the system and estimation to compute the concentration of the unknown. We use here the following notation:

$$r_j = \sum_{k=1}^K S_{jk} * c_k + b_j \quad /1/$$

where the signal r (response) at sensor (wavelength) j depends on the sensitivity S_{jk} of this sensor to analyte (element) k and the concentration c_k of this element and b is an additive background term for each sensor. For more than one wavelength eqn 1 can be conveniently expressed in vector/matrix notation as:

$$r = S c + b \quad /2/$$

In calibration all elements of S are determined from several samples of known composition C giving a matrix of signals R . The background B can in principle be different for every sample used in calibration; in practice one will try to find a matrix-matched sample where all analytes are well below the detection limit. The first step to estimate the sensitivity matrix is given by:

$$S = (R - B) C^+ \quad /3/$$

where C^+ is the pseudoinverse of C and it is assumed that for this $K \times I$ matrix $I \geq K$ with at least K linearly independent rows and columns. The pseudoinverse of S then serves to estimate the concentrations in the unknowns from the measurements r :

$$c = S^+ * (r - b) = (S^T S)^{-1} S^T (r - b) \quad /4/$$

Eqn 4 also gives the definition of the pseudoinverse.

The straightforward nature of this derivation should not lead one to forget that there are errors involved when applying these equations: no signal can be measured free of errors and no standard can be prepared without a certain error in concentrations. So R , B and C (therefore also C^+) in calibration (eqn 3), as well as r and b in eqn 4 are subject to errors. From this it follows that S in eqn 3 carries over its errors into the estimation of unknowns in eqn 4:

$$dc = -S^+ (dR - dB) C^+ c + dC C^+ + S^+ (dr - db) \quad /5/$$

This equation is derived in an earlier paper[7] and contains only directly accessible error terms of the signals and the standards. Instead of using deterministic errors this expression can be recast in terms of variance and gives in component notation:

$$\sigma(c_k) = \left(\sum_{m=1}^J \sum_{l=1}^I \sum_{n=1}^K (S_{km}^+ * C_{ln}^+)^2 * [\sigma^2(R_{mi}) + \sigma^2(B_{mi})] + \right. \\ \left. + \sum_{i=1}^I \sum_{n=1}^K [\sigma(C_{ki}) * C_{in}^+ * c_n]^2 + \sum_{j=1}^J (S_{kj}^+)^2 * [\sigma^2(r_j) + \sigma^2(b_j)] \right)^{\frac{1}{2}} \quad /5a/$$

It can clearly be seen that the total error is composed of three additive terms: the errors of calibration measurements, the errors from calibration standards and the errors from measurements of the unknown sample.

The expression in eqn 5 is also suitable to estimate realistic detection limits for multiple sensor, multiple analyte calibrations. The generally accepted definition of the univariate detection limit

$$r_{A,L} = z(\alpha) * \sigma(r_{A,0}) \quad /6/$$

is expressed in the signal domain where z is a statistical factor and σ is the standard deviation of the zero net analyte signal that can be converted to the concentration domain by dividing by the sensitivity. For $z = 3$ the rearrangement of eqn 6 leads to a relative error of 33 % as long as the sensitivity is assumed to be known free of errors.

This approach in multivariate calibration is meaningless as serious random errors are introduced in the computation of the net signal from the raw data that need to be accounted for. Alternatively, eqn 5 gives the relative error of analyte concentrations that only needs to be solved for a zero net signal. It also presumes, of course, that the signal errors close to the detection limits are either known or that they can be measured. Strictly speaking this requires an iterative solution for $\sigma(c_k)$ if standard deviation changes with concentration (heteroscedastic case), but the iteration quickly converges to the final solution.

EXPERIMENTAL

Instrumentation and samples. For an experimental test of the theory inductively-coupled plasma optical emission spectrometry (ICP-OES) was chosen. A sequential spectrometer Perkin-Elmer Plasma II was operated at 1.2 kW power at 27.12 MHz and the signal was observed 12 mm above the load coil. The resolution was 9 pm in first order. Phosphorus from various samples from refractory metals industry were preconcentrated according to a literature method[9]. Alternatively, bauxite samples of Hungarian origin were decomposed by a mixture of mineral acids according to a recently proposed variant[10].

RESULTS AND DISCUSSION

Error models and overlap correction. In principle all terms in eqn 5 can be made available from measurements. While this is relatively straightforward for the signals (R, r, B, b) it is more cumbersome for the errors of these signals. In ICP-OES it can be assumed that one deals with proportional errors with some minor constant error that is unavoidable even for signals close to zero. This and alternatively a constant error model are given in Fig. 1 and one clearly sees how the latter would spoil any attempt to estimate the error: At low concentrations this would lead to a very high

detection limit, while at higher concentrations the rel. error would be judged unrealistically low. While all of these error models are debatable it has been shown that the (const. + proport. error) model can be confirmed for our experimental conditions [7,8]. For a scanning monochromator instrument it does, however, break down at the edges of the peak if the mechanical stability does not permit to reposition the wavelength with sufficient accuracy. Such a case is shown in Fig. 2 for the measurement

Figure 1 (left): Two error models for ICP-OES

Figure 2 (right): Repeated ($N=10$) scans of the Al line at 396.150 nm: mean spectrum, standard deviation of the spectrum and relative standard deviation. The scan width is 80 pm, observation height is 12 mm above coil, forward power 1200 W.

of Al in bauxite where the error at the background and the peak top is indeed close to 1 % for repeated measurements of this line. In the edges of the peak it rises, however, close to ten times this value. This has serious consequences for the error of a "full-spectrum" method such as the Kálmán filter technique that was recently introduced by Van Veen and de Loos-Vollebregt in ICP-OES [11,12]: Without further stabilization of the wavelength drive the $\sqrt{I}$ advantage (I .. number of wavelengths) will be far smaller than expected due to the relations described in eqn 5. Although multivariate calibration methods can be used whenever more than one wavelength is chosen for analytical measurements, they are most relevant if interference corrections are necessary. This is the case for the determination of Cu in bauxites, as the sensitive 224.700 nm - line of Cu is overlapped by an iron line (Fig. 3). The particular variant of the linear model is really of minor importance: Without further discussion of the Kálmán method we show in Fig. 4 the good agreement of the classical least-squares solution for multivariate calibration and the results of the Kálmán filter for the determination of Cu in a bauxite sample, while a three-point background correction can obviously not correct the effect of the iron overlap. For both methods 25 wavelengths 1 pm apart from each other around the Cu-line were selected. The copper content was determined to be $88 \pm 13 \mu\text{g/g}$.

Error estimation and detection limits [7,8]. The prediction of errors from eqn 5 was tested experimentally for a system containing Mo, W and P by running different calibration sets of three samples each. In this

Figure 3 (left): Iron interference with copper line at 224.700 nm: standards and bauxite sample

Figure 4 (right): Determination of Cu using the 224.700 nm-line summarized as box plots (N=10). Left: Automatic three point correction. Middle: Results from Kálmán filter. Right: Results from ordinary least squares.

experiment the composition of these calibration samples was changed. Also repeated scans were run to check the validity of the proportional error model. The agreement was very satisfactory and the prediction of error by eqn 5 agreed satisfactorily with the actually observed error from replicated measurements [7].

Also, a study of the effect of interferents and wavelength selection was performed [8]. In this particular case an ultra-pure tungsten sample was available as "true blank" whose phosphorus content was ascertained by SIMS to be well below an anticipated detection limit of ICP-OES. A solution of this sample was spiked with increasing amounts of P and the errors were determined from this experiment. This result again was checked against the prediction of multivariate detection limits. Table 1 shows a short account of the results of this work.

TABLE 1: Relative errors [%] and detection limits [ng/ml] for the determination of P at the 213.618 line in aqueous matrix and in a tungsten/molybdenum matrix (adapted from [7,8]).

wavelengths [nm]					rel. errors		detection limits [ng/ml]			
213.xxx exact position:					at 1 μ g/ml P		aqueous			
					[%]					
					1%excl.	1%incl.	1%excl.	1%incl.	1%excl.	1%incl.
Mo	BG1	P	W	BG2	stds.	stds.	stds.	stds.	stds.	stds.
x	x	x			35	52	100	100	920	1300
x	x		x		21	31	290	290	690	980
x		x	x		8	12	110	110	270	390
	x	x	x		9	13	100	100	290	420
x	x	x	x		8	12	100	100	260	380
x	x	x	x	x	8	12	100	100	260	380

It presents the error estimates and estimated detection limits at 5 selected wavelengths around the phosphorus line at 213.618. The molybdenum line at 213.606 nm and the tungsten line at 213.633 nm are partially overlapping. Both elements vary in concentration and additionally the background point at 213.615 (BG1) is affected both by Mo and P. In accordance with the aforementioned studies on error models, 1 % error was assumed: If this is also assumed for the concentration error in the standards the prediction error for P rises from 35 to 52 % in the worst case, from 8 to 12 % in the best case. In an aqueous solution the detection limit is hardly deteriorated by the presence of a 1 % error in standard concentrations. The situation is, however, different when overlapping signals are resolved. Then, the detection limit is seriously degraded as the concentration of interferences increase (Table 1, last two columns): For every compositions of sample a separate detection limit applies[13].

The limitations of the described approach concern non-linear signal/concentration dependences that are observed for very wide concentration ranges or for other analytical techniques. This is more a problem for general error estimation than for the estimation of detection limits: Around the blank it is almost always sensible to entertain a linear model. Locally, it will always be possible to estimate a linear model unless serious lack-of-fit is involved: For many analytical techniques lack-of-fit usually points to unmodelled interferences that cannot be accommodated by the current approach.

Conclusion. The development outlined in this paper enables one to deal with any type of error model, as for each signal the associated error is explicitly accounted for. Additionally, a different error might be involved for every component in each standard. Thus, the limitations imposed by the assumption of homoscedasticity in earlier work are overcome[14-16].

This work was supported in part by the "Forschungsförderungsfonds der gewerblichen Wirtschaft" (Vienna) within the frame of the A-3 of the European COST 503/II-action. P.K. thanks the Hungarian Government for a scholarship.

REFERENCES

1. H. Martens and T. Naes, *Multivariate Calibration*, Wiley, New York, 1989.
2. M. Otto and W. Wegscheider, *Anal. Chem.* **57**, 63 (1985).
3. M. Otto, J.D.R. Thomas, K. Murray and P.M. May, *Anal. Chem.*, **57**, 1511 (1985)
4. W.P. Carey, K.R. Beebe and B.R. Kowalski, *Anal. Chem.*, **59**, 1529 (1987)
5. D.M. Haaland, *Anal. Chem.* **60**, 1508 (1988)
6. G. Bauer, I. Rehana, H.M. Ortner and W. Wegscheider, *Spectrochim. Acta*, **43B**, 971 (1988).
7. G. Bauer, W. Wegscheider and H.M. Ortner, *Spectrochim. Acta*, **46B**, 1185 (1991).
8. G. Bauer, W. Wegscheider and H.M. Ortner, *Fresenius' J. Anal. Chem.*, **340**, 135 (1991).
9. G. Bauer, W. Wegscheider and H.M. Ortner, *Mikrochim. Acta (Wien)* **III**, 399 (1989)

10. L. Siróki, Doctoral Dissertation, University of Veszprém, Hungary 1990.
11. E.H. Van Veen and M.T.C. De Loos-Vollebregt, *Spectrochim. Acta*, 45B, 313 (1990).
12. E.H. Van Veen and M.T.C. De Loos-Vollebregt, *Spectrochim. Acta*, 45B, 1109 (1990).
13. G. Bauer, W. Wegscheider and H.M. Ortner, Proceedings European Winter Conference on Plasmaspectrochemistry, Dortmund 1991, *Spectrochim. Acta B*, in press.
14. C. Jochum, P. Jochum and B.R. Kowalski, *Anal. Chem.*, 53, 85 (1981)
15. A. Lorber, A. Harel, Z. Goldbart and I.B. Brenner, *Anal. Chem.*, 59, 1260 (1987).
16. G. Bergmann, B. von Öpen and P. Zinn, *Anal. Chem.*, 59, 2522 (1987)