Chemometrics and Intelligent Laboratory Systems, 7 (1989) 39–51
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

Interpretation of Latent-Variable Regression Models

OLAV M. KVALHEIM * and TERJE V. KARSTANG

Department of Chemistry, University of Bergen, N-5007 Bergen (Norway)

(Received 2 November 1988; accepted 1 June 1989)

CONTENTS

ABSTRACT

Kvalheim, O.M. and Karstang, T.V., 1989. Interpretation of latent-variable regression models. *Chemometrics and Intelligent Laboratory Systems,* 7: 39–51.

In this work, we show that the projections of the predictors on the normalized regression vectors represent a target rotation with the responses (concentration vectors) as targets. By means of this operation the predictive ability of a latent-variable (LV) regression model and the importance of each predictor for all the responses is obtained. The two features can be portrayed simultaneously and quantitatively in an LV regression BIPLOT display. This graph shows how modelled interferents influence prediction, information as important as the detection of and correction for unmodelled interferents when using a regression model for prediction.

For samples characterized by whole digital profiles rather than a collection of peaks, graphs showing the covariances between the responses and the original or the reproduced predictor space appear to provide the most useful information for interpreting an LV regression model.

0169-7439/89/$03.50

1 INTRODUCTION

Many procedures have been developed for the establishment of a predictive relationship between two sets of variables [1]. Among these methods, latent-variable (LV) regression techniques, such as principal component (PC) [2] and partial-least-squares (PLS) [3] regression, are increasingly used in chemistry, geochemistry and related disciplines. The instrumental techniques used in these fields typically produce data tables with a low sample-to-variable ratio, implying strong collinearity in data and, thus, problems with matrix inversion using any of the standard regression techniques.

Validation is prerequisite for any model obtained by LV regression methods (Fig. 1). Statistical validation is necessary to ensure models with good predictive abilities. The validation concerns: (i) the number of LV components to be included in a model; and (ii) detection and rejection of outlying samples and variables in the data. In addition to the statistical validation of the regression model, a chemical validation is necessary, in order to check that the model makes sense and to be able to detect possible shortcomings or limitations of the model. In order to accomplish this task, an interpretation of the model in chemical terms is mandatory.

There is some confusion about what principles to use in order to interpret an LV regression model. Martens and coworkers, in their interpretations of single response PLS regression models, tend to look at the regression vector (ref. 3 and references therein). Thus, the regions with large amplitudes on the regression coefficients are assumed to be the most important for the chemical interpretation. However, high amplitudes on the regression coefficients can occur for predictors with almost no correlation to the response [4,5]. Haaland and Thomas [5], in their elaboration of the PLS algorithm, claim that the vector of the first variable weightings, which for a single response variable is proportional to the covariances between response and predictors, is more useful than the regression vector for the qualitative interpretation of a PLS regression model. However, for a model of a single response, the vector of the first PLS variable weightings represents the variable loadings obtained by projecting the sample vectors directly onto the normalized regression vector for the full least-squares solution. Thus, for a full least-squares decomposition, the regression vector and its corresponding loading vector are picturing the same latent variable, although in two different representations.

The outline of this work is as follows: first, the

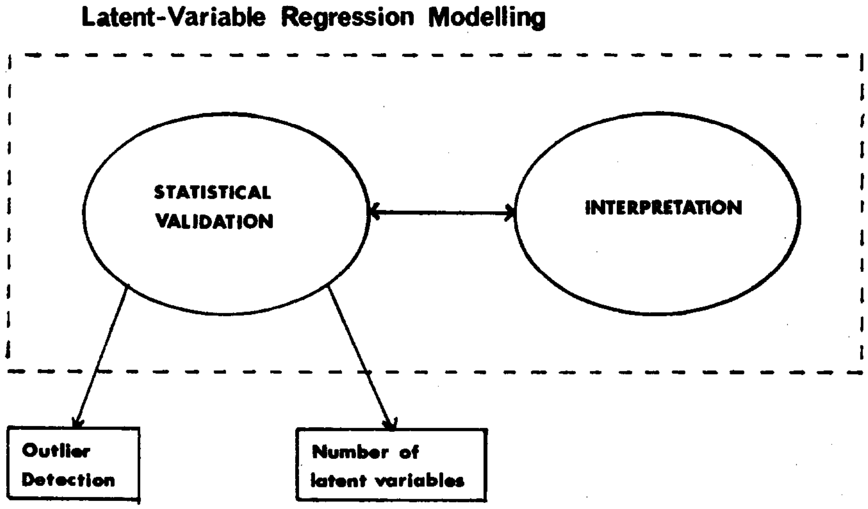
Latent-Variable Regression Modelling
STATISTICAL VALIDATION
INTERPRETATION
Outlier Detection
Number of latent variables

Fig. 1. Statistical and chemical validation of regression models.

representations in variable and object space of the matrix of predictors are used to interpret the regression vector. The proportionality between the sample scores on the regression vector and the projection of the vector of predicted responses onto the reproduced predictor space is used to show that the regression vector defines a target rotation in predictor space with the vector of responses being the target. This observation reveals that the variable loadings, corresponding to

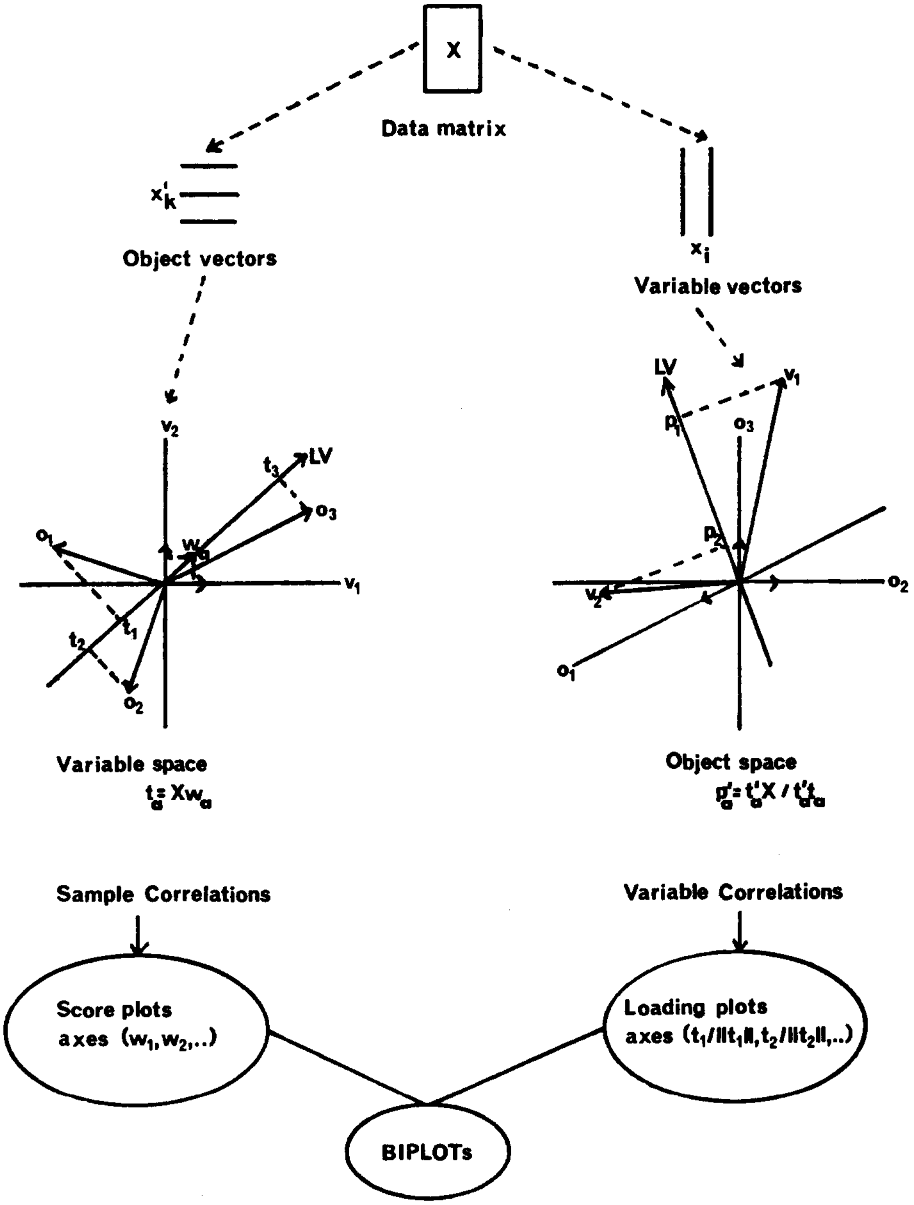

Fig. 2. The two ways of looking at a data matrix **X**, their connection to the point representations in variable and object space, and their combination into the BIPLOT representation. Three vectors, w_a, t_a, and p_a, are necessary to define the latent variable (LV) in the two spaces. The labels o and v are used to denote axes or vectors related to objects and variables, respectively.

the covariances between the reproduced predictor space and the response vector, contain the most useful information for an interpretation of an LV regression model. By combining variable and object space into BIPLOTs, the most significant information can be displayed in a quantitative and comprehensible form as regression BIPLOTs. For interpretation of data of continuous type, i.e. in the form of spectral profiles, we suggest the use of covariance graphs obtained for the vectors of rotated variable loadings and the raw data. The covariance graph for the raw data is calculated between each response and the profiles. Thus, for a single response, it is proportional to the PLS variable weightings. Both the BIPLOT regression display and the various displays for interpretation of profiles are discussed with reference to LV models obtained for real data.

2 DECOMPOSITION OF PREDICTOR SPACE

The establishment of an LV regression models consist of two parts: (i) decomposition of the predictor space into uncorrelated latent variables; and (ii) the calculation of a regression vector for each response variable. The first part of this procedure is expressed by the following decomposition formula:

$$\mathbf{X} = \mathbf{U}\mathbf{G}^{1/2}\mathbf{P}' + \mathbf{E} \qquad (1)$$

The matrix $\mathbf{X}$ contains the values of the M predictors arranged in N rows, one row for each object (sample). In order to account for the possibility of interaction between variables, the predictor space may be extended by including products of single variables [4,6]. The data matrix $\mathbf{X}$ is assumed to be column-centered. The score matrix $\mathbf{U}$ and the loading matrix $\mathbf{P}$ of dimension $N \times A$ and $M \times A$, respectively, contain the information about sample correlations ad variable correlations in the predictor space reproduced by the A latent variables extracted. The matrices $\mathbf{U}$ and $\mathbf{P}$ are composed of column vectors, each corresponding pair of vectors containing the projections of the object vectors (defined by the rows of $\mathbf{X}$) and variable vectors (defined by the columns of $\mathbf{X}$), respectively, onto a latent variable. For singular-value decompositions both $\mathbf{U}$ and $\mathbf{P}$ are orthonormal matrices, while all other decompositions are only orthonormal in $\mathbf{U}$ [7].

The diagonal matrix $\mathbf{G}$ provides the information about the importance of each latent variable with respect to the variance accounted for in predictor space. Thus, the diagonal elements of $\mathbf{G}$ are the squared norms of the latent-variable score vectors prior to normalization. Each norm is obtained by using the normalization condition on the vector of non-normalized scores $\boldsymbol{t}_a$ obtained by projecting the samples on the corresponding coordinate vector $\boldsymbol{w}_a$, called weightings by Martens [3], normalized to unit length. The decomposition procedure can be visualized with reference to variable and object space defined by $\mathbf{X}$ (see Fig. 2 and ref. 7). Note that $\boldsymbol{u}_a = \boldsymbol{t}_a / \| \boldsymbol{t}_a \| = \sqrt{g_a}\, \boldsymbol{t}_{\mathrm{a}}$. Note also that the coordinate vectors, $\{ \boldsymbol{w}_a \}$, are orthonormal [8].

The matrix $\mathbf{E}$ of dimension $N \times M$ contains the residuals of the matrix $\mathbf{X}$ after extraction of A components.

Eq. 1 contains all the information necessary for constructing a BIPLOT representation (Fig. 2) of the reproduced predictor space under the assumption of orthonormal score vectors [7]. With the additional constraint of orthonormal loading vectors $\{ \boldsymbol{p}_a \}$, eq. 1 reduces to the well-known singular-value decomposition (SVD) predictor space.

3 THE REGRESSION VECTOR

3.1 Calculating the regression vector

In order to establish the regression model we need to find the relation between the reproduced predictor space and the $N \times 1$ column vector $\boldsymbol{y}$ of responses. Minimizing, in the least-squares sense, difference between predicted and measured responses, the $N \times 1$ regression vector $\boldsymbol{b}$ connecting the response and the predictors is given by [8]:

$$\boldsymbol{b} = \mathbf{X}^{+}\boldsymbol{y} \qquad (2)$$

The matrix $\mathbf{X}^{+}$ is the (Moore–Penrose [8]) generalized inverse of the reproduced $\mathbf{X}$ matrix. Thus, in order to obtain the regression coefficients we need to calculate a generalized inverse. For singular-value decomposition both $\mathbf{U}$ and $\mathbf{P}$ are orthonormal matrices, and from eq. 1 the generalized inverse is obtained simply as $\mathbf{X}^+ = \mathbf{P}\mathbf{G}^{-1/2}\mathbf{U}'$. For the general case $\mathbf{P}$ is neither orthonormal, nor orthogonal. Starting instead from an alternative expression for the decomposition of predictor space [8];

$$\mathbf{X} = \mathbf{U}\mathbf{R}\mathbf{W}' + \mathbf{E} \qquad (3)$$

we obtain the generalized inverse as

$$\mathbf{X}^+ = \mathbf{W}\mathbf{R}^{-1}\mathbf{U}' \qquad (4)$$

$\mathbf{R}$ is a triangular matrix of dimension $A \times A$. Comparison of eqs. 1 and 3 shows that $\mathbf{R}^{-1} = (\mathbf{P}'\mathbf{W})^{-1}\mathbf{G}^{-1/2}$. By using this expression for $\mathbf{R}^{-1}$ in eq. 4, then inserting eq. 4 into eq. 2, and using the equivalence $\mathbf{U} = \mathbf{T}\mathbf{G}^{-1/2}$, the regression vector is obtained as:

$$\boldsymbol{b} = \mathbf{W}(\mathbf{P}'\mathbf{W})^{-1}\mathbf{G}^{-1}\mathbf{T}'\boldsymbol{y} \qquad (5)$$

Note that eq. 5 holds for any LV regression model as long as the score vectors are orthogonal. A special case is PLS regression with one or several response variables. Another special case of particular interest is PC regression. We have previously shown [9] that for a PC decomposition $\mathbf{P} = \mathbf{W}$. Thus, the matrix $(\mathbf{P}'\mathbf{W})$ is the $A \times A$ identity matrix, reducing the problem of obtaining the regression vector from eq. 5 to a trivial task. In cases where not only one, but a set of responses is to be predicted from one LV decomposition of predictor space, the regression vector for each response is obtained by repeating the calculation defined by eq. 5 for each response vector $\boldsymbol{y}$.

3.2 Interpretation of the regression vector

Eq. 5 shows that the regression vector is a linear combination of the coordinate vectors $\{\boldsymbol{w}_a\}$. Indeed, it follows from the the least-squares procedure used to arrive at the regression vector $\boldsymbol{b}$, that $\boldsymbol{b}$ defines the line of best fit to the response vector in the reproduced variable predictor space. Thus, $\boldsymbol{b}$ defines the target rotation of the coordinate vectors $\{\boldsymbol{w}_a\}$ with the vector of measured response as target.

For each response vector, scores and loadings can then be obtained by the following procedure [4,9]:

1. Normalize the regression vector. This gives the new coordinate vector $\boldsymbol{w}^* = \boldsymbol{b}/\|\boldsymbol{b}\|$.
2. Project the variable space of the predictors onto $\boldsymbol{b}/\|\boldsymbol{b}\|$. This gives the target-rotated scores, i.e. $\boldsymbol{t}^* = \boldsymbol{X}\boldsymbol{b}/\|\boldsymbol{b}\|$.
3. Normalize the score vector obtained in the previous step, i.e. $\boldsymbol{t}^*/\|\boldsymbol{t}\|$.
4. Project the object space of the predictors onto the normalized score vector. This gives the target-rotated loadings, i.e. $\boldsymbol{p}^{*\prime} = \boldsymbol{t}^{*\prime}\mathbf{X}/\|\boldsymbol{t}\|$.

Note that due to the least-squares criterion applied for the calculation of $\boldsymbol{b}$, the same results are obtained for $\boldsymbol{t}^*$ and $\boldsymbol{p}^*$, respectively, whether the reproduced variable and object space or the full spaces of the predictors are projected upon in the algorithm above. Comparison with the results of an ordinary target rotation of an LV decomposition with orthogonally-constrained scores of the predictor space [10] proves the equivalence between target rotation and the above projection procedure.

The above procedure defines the regression line in the variable space of the predictors in the basis of $\{\boldsymbol{w}_a\}$ (through eq. 5), and the line in the object space of the predictors for each response by calculating the correlations between the rotated $\boldsymbol{t}^*$ and the original score vectors $\{\boldsymbol{t}_a\}$. Thus, we can construct quantitative BIPLOT displays showing simultaneously the predictive ability of the regression model (because the rotated scores are proportional to the predicted responses, see step 2 in the algorithm above), and the covariance between each response vector and the predictors. We will call such a display a regression BIPLOT.

We further note that for $\mathbf{E} = \mathbf{0}$, i.e. a full least-squares decomposition of predictor space, the normalized target-rotated loading vector $\boldsymbol{p}^*$ is identical to the first PLS variable weightings $\boldsymbol{w}_1$ for a single response. In bypassing, we observe that the rotated vector in this case is exactly the latent variable provided by the decomposition technique known as redundancy analysis (RA) [11].

4 BIPLOT INTERPRETATION OF AN LV REGRESSION MODEL

The usefulness of the BIPLOT technique for the interpretation and evaluation of LV regression models is proved by an example from organic

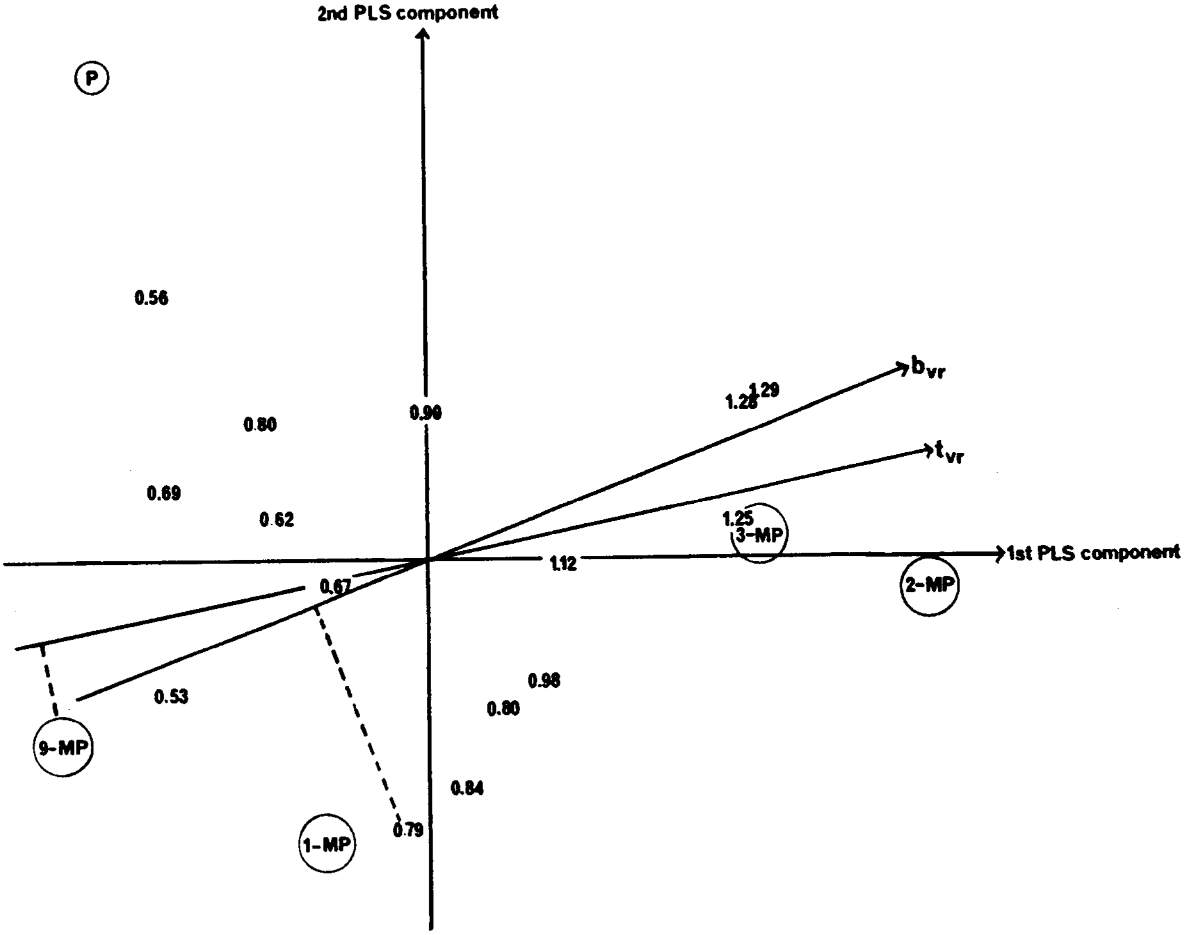

Fig. 3. BIPLOT display of a two-component PLS regression model. Loadings and scores are given on the same scale. Thus, the sum of squared loadings is equal to the sum of squared scores. The samples are labeled with their measured response (vitrinite reflectance). For the variables the abbreviations P (phenanthrene) and MP (monomethylphenanthrene) are used. The different monomethylphenanthrenes are distinguished by giving the position of the methyl group. The projection of the maturity factor in variable and object space, b_{vr} and t_{vr}, respectively, is included to quantitatively display its correlation pattern with samples and variables.

geochemistry. The relative abundances of phenanthrene and four monomethylphenanthrenes in 15 coal extracts were used as predictors for an LV regression model with vitrinite reflectance in crushed coal as response. The regression was performed to check if and how the distribution of phenanthrene and monomethylphenanthrenes in organic matter was related to maturation ('geologic temperature') as assessed by vitrinite reflectance measurements. (Experimental details and data are published in ref. 12.)

PC and PLS regressions of the tabulated data gave almost identical results. Fig. 3 shows the BIPLOT representation of a two-component PLS regression model accounting for 92.3% of the variance in the predictors. By plotting the samples with labels corresponding to their measured responses, the overall predictive ability of the model is easily comprehended just by looking at the sample pattern on the regression vector. Similarly, Fig. 3 provides significant information about the sensitivity of each predictor to variation in the response by looking at the reproduced variable correlation pattern on the rotated score vector, which, as shown above, is proportional to the vector of predicted responses. Thus, from Fig. 3 it is obvious that maturity in coal, as expressed by vitrinite reflectance, is related to the distribution of the monomethylphenanthrenes: increased relative abundance of 2- and 3-monomethylphenanthrenes compared to 9- and 1-monomethylphenanthrenes implies increased maturity. We note that this interpretation is only possible after including the line corresponding to the projection of the response in the reproduced predictor space.

More information is available from the regression BIPLOT display. Thus, in Fig. 3 we observe that the variation in relative abundance of phenanthrene is almost orthogonal to the score vector corresponding to the 'maturity factor'. Indeed, the angle between the maturity factor and the line connecting phenanthrene and origin in the regression BIPLOT is found to be 114° (Fig. 4). In other words, phenanthrene represents and interference which barely influences the prediction of maturity. Such information can be of crucial importance for assessing the reliability of a prediction. For instance, in the present case, the sample of lowest maturity classified as a strong outlier using the approximate *F*-test devised by Wold [13]. Despite this fact the prediction of vitrinite reflectance for this sample was extremely close to the measured value: 0.52 compared to a measured

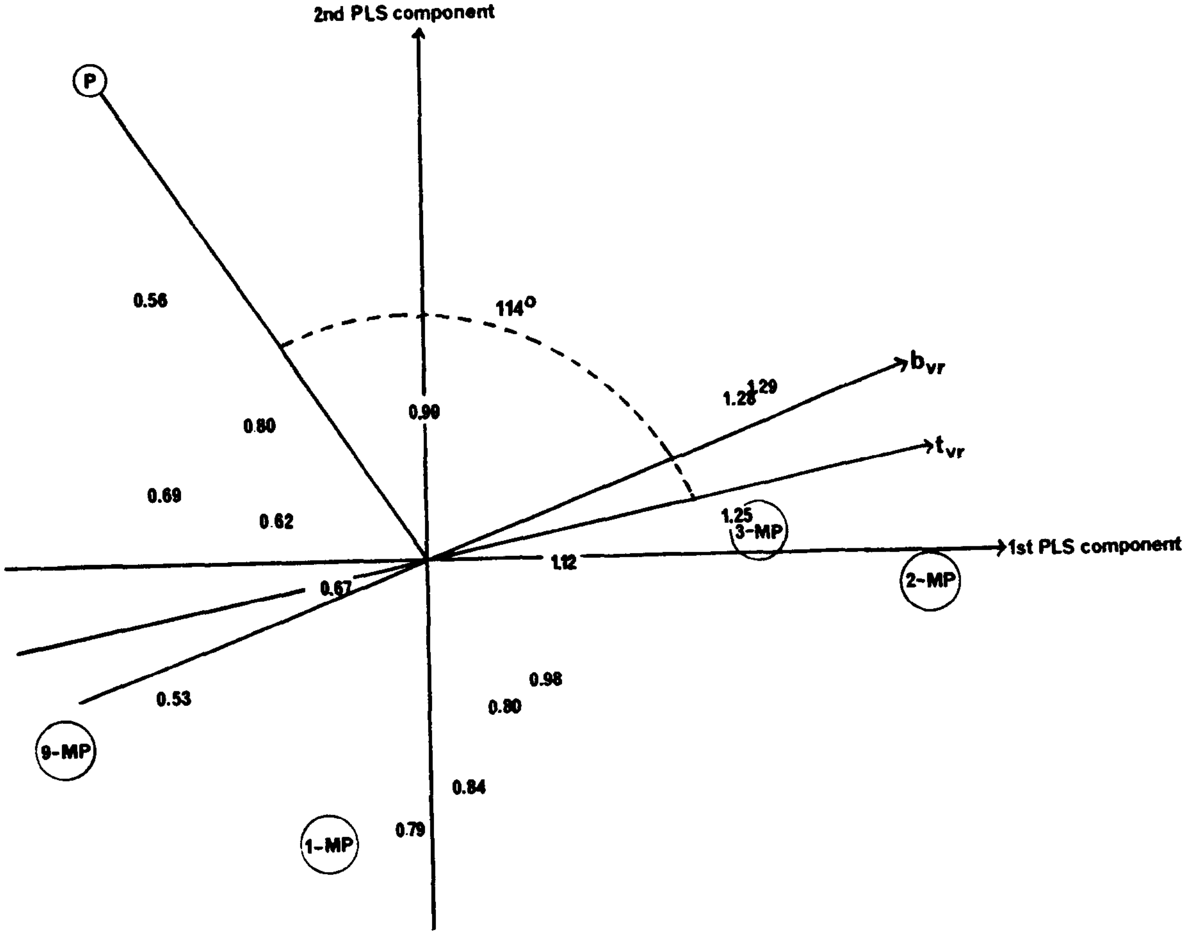

Fig. 4. Same as Fig. 3, but with the line corresponding to the interference due to phenanthrene drawn to show its almost orthogonal projection to the maturity factor.

of 0.53. Inspection of Figs. 3 or 4 reveals the reason for this apparently contradictory result. Compared to the other samples in the low-maturity region of the regression BIPLOT, the sample is very close to the regression line, thus implying an abnormal low content of phenanthrene. This observation is confirmed by looking directly at the measured data for the samples [12]. However, since the variation in relative abundance of phenanthrene does not influence maturity significantly, the prediction is still reliable. In other words, the prediction is more robust towards this interference than implied by the criterion used for the detection of outlying samples.

The outlier testing is based on the construction of a class box [13] around the two PLS components. Thus, we ought to modify the class box in accordance with our present understanding of the model. Two simple solutions come to mind: (i) remove the almost orthogonal interference and recalculate the model; or, better, (ii) reduce the weight of the interference. The latter can be obtained by either weighting each PLS component with its correlation with the target-rotated scores (i.e. the predicted responses) prior to outlier testing, or, by constructing a class box with a better orientation, i.e. around the regression line. Both procedures for reducing the weight of destructive interferences can be generalized to cases with many response variables. In such cases, we can use techniques similar to Procrustes rotation [14] to obtain a class box with better orientation.

5 INTERPRETATION OF LV REGRESSION MODELS OF PROFILES

When whole spectral of chromatographic profiles are used as predictors, the regression BIPLOT display is no longer a convenient tool for interpretation. Although, in principle, it is possible to reduce the number of variables to a manageable size by using a selection of representative points from the continuous variable loading profiles, it is in general better to retain the whole variable loading pattern for interpretation. For instance, if the spectral profiles of a set of mixtures are used as predictors for the measured concentrations of the constituents in the mixtures, the spectral loading patterns may be recognized as corresponding to combinations of spectra of pure constituents. This information is easily lost if only selected wave lengths are plotted. Thus, the regression BIPLOT display must be replaced by other graphic representations.

5.1 Correlation and covariance graphs

In most cases, an LV decomposition does not provide components that relate directly to the responses. Only in case of an orthogonal design in the response variables and with no unmodelled colinear interferences present is it possible to obtain orthogonal latent variables, for instance, as spectra of pure constituents corresponding to concentration vectors. This assumption is not fulfilled in most situations, either because the samples are obtained from natural systems or because the data for each sample have been normalized to constant sum. As for the regression BIPLOT display, alternative displays must necessarily lean themselves to the target-rotated vectors in order to be useful for a chemical interpretation of the model. Cowe et al. [15] start from the correlation graph calculated between each response vector and the profiles. The correlation graph gives a continuous profile with values ranging from -1 to $+1$. Cowe et al. [15,16] showed how correlation graphs could be utilized for a preliminary interpretation prior to selecting the regression method. For instance, if the pure constituent spectra of a set of mixtures are known, the correlation graphs can directly detect regions of interfering constituents or effects which has to be accounted for by the regression model. The approach used by Cowe et al. [15] can easily be extended to obtain correlation graphs between the responses and the reproduced predictor space. Thus, an interpretation of the sensitivity of each response to the variation in the profiles is possible by using the reproduced correlation graph.

Although the correlation graph is a very useful tool for spectral interpretation, a similar display which we shall call the covariance graph, seems to be even simpler to use. For a single response, the covariance graph normalized to unit area is obtained automatically in PLS regression as the first

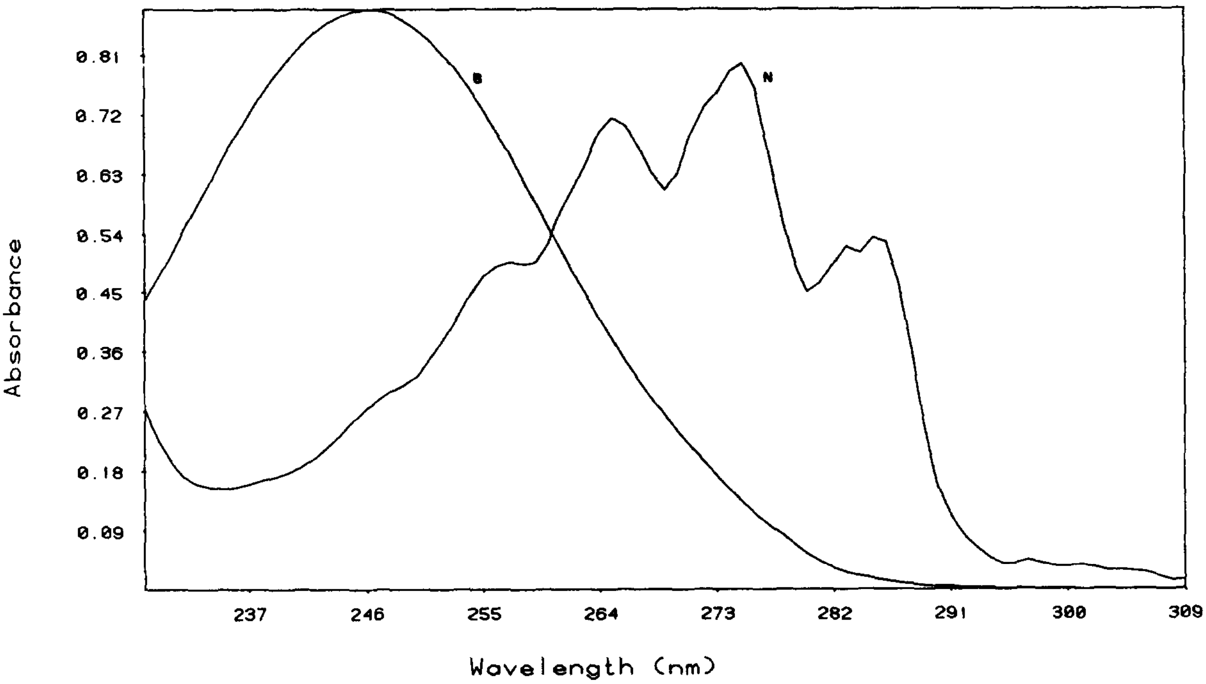

Fig. 5. Spectra of pure biphenyl (B) and naphthalene (N).

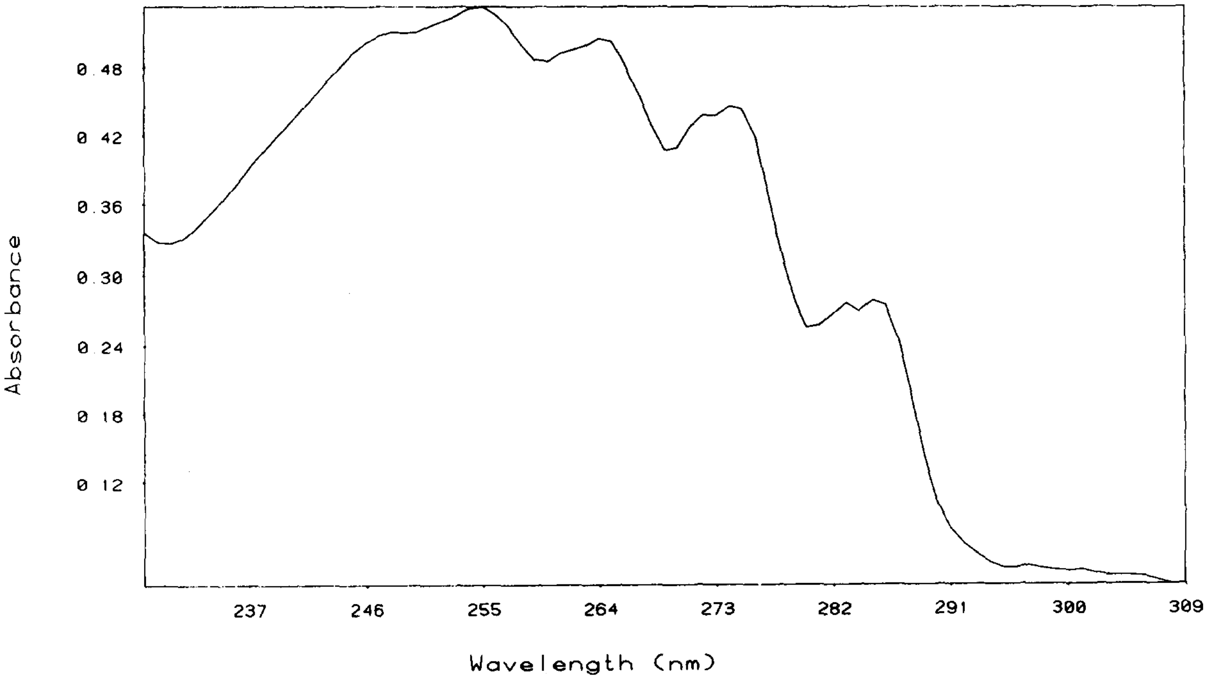

Fig. 6. Average spectrum for the mixtures.

variable weighting vector. In regions of the profiles free from interfering effects, the covariance graph shows exactly the same intensity pattern as the spectrum of the pure constituent. Covariance graphs can be calculated for multiple responses independently of what regression method is actually used. Comparison of each covariance graph with its corresponding target-rotated variable loading pattern can then be used to check if the LV regression model is adequately describing all the responses and to reveal colinear interferences.

5.2 Display of profile data

The following simple example illustrates the procedure for interpretation advocated in the above section. Biphenyl and naphthalene and eight mixtures of varying concentrations of these two constituents were characterized by UV spectroscopy. Fig. 5 shows the UV absorbance spectra of the pure constituents. The average spectrum of the eight mixtures is shown in Fig. 6.

Fig. 7 shows the covariance graphs calculated between the mixture spectra and the two concentration vectors. The similarity between these two graphs and the constituent spectra is striking. Even in the region with strong overlap between the absorbance bands of the two constituents are the covariance graphs almost identical to the spectra. This nice resolution is a result of the design of the mixtures, the concentrations being chosen so that the two concentration vectors were almost orthogonal.

PLS regression with the concentrations as responses and the profiles as predictors gave a two-component model accounting for more than 99% of the total variation. The regression vectors and their corresponding variable loadings were calculated and compared. Fig. 8a and 8b show a close similarity between the profiles corresponding to the regression vectors and the variable loadings. Thus, from this particular example it appears that the same qualitative information is obtained whether we look at the regression vectors or the target-rotated loadings vectors. However, this is not generally the case as has been shown in some recent applications [4,5,17].

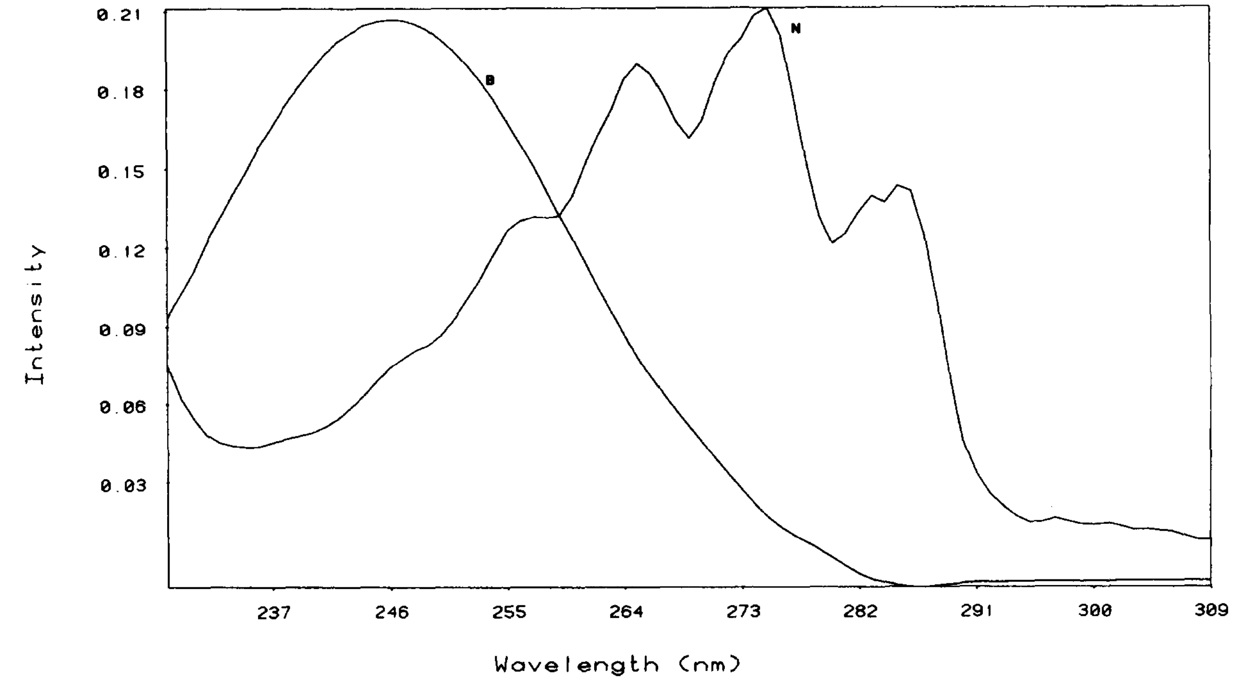

Fig. 7. Covariance graphs obtained between the spectra of mixture and the concentration vector of biphenyl (B) and naphthalene (N), respectively.

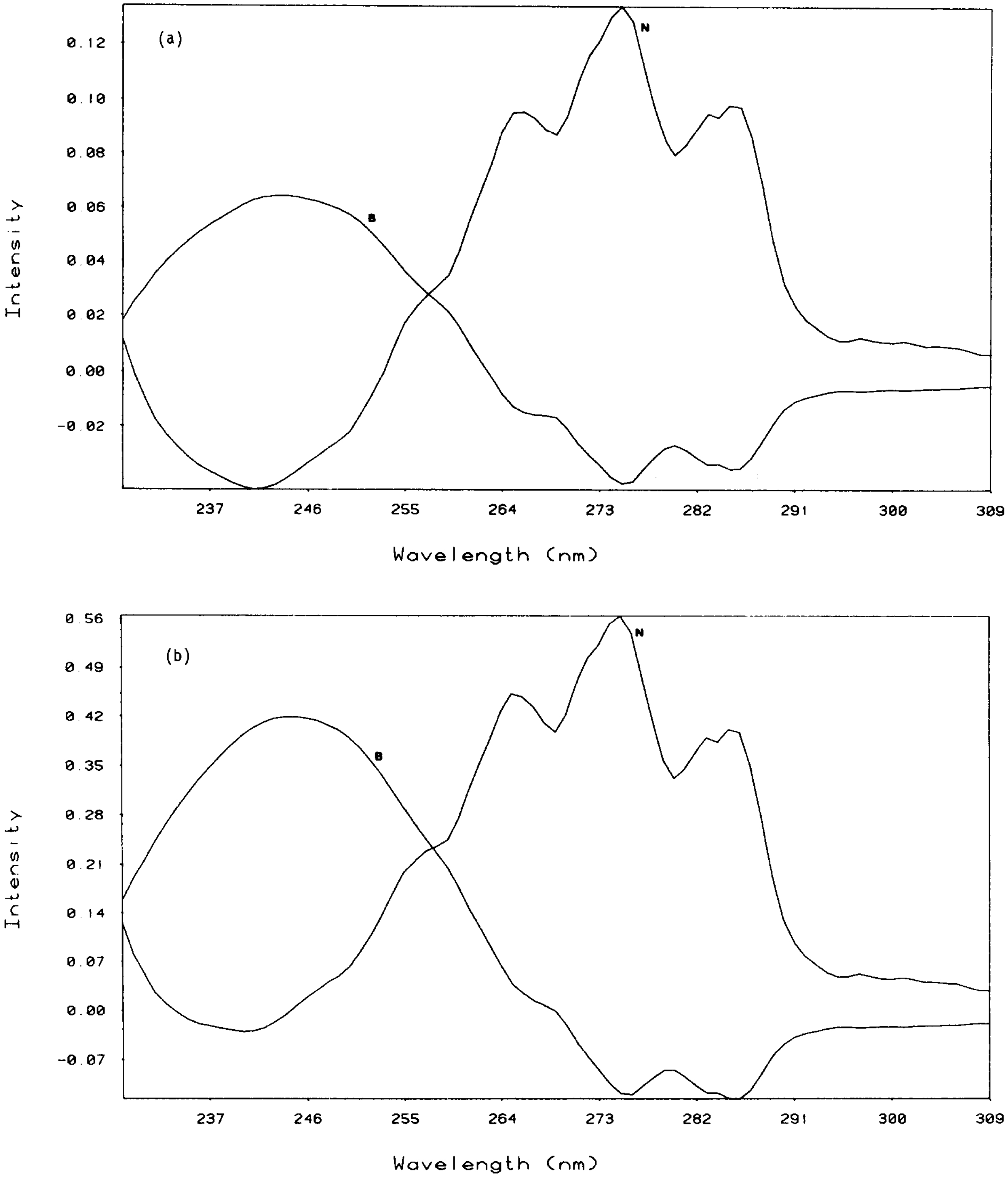

Fig. 8. (a) Regression vectors, and (b) corresponding loading vectors for biphenyl (B) and naphthalene (N) calculated from a two-component PLS2 regression model.

The target-rotated variable loadings (Fig. 8b) are seen to be very similar to the covariance graphs (Fig. 7). This latter observation is not surprising since we know that each vector of target-rotated variable loadings converges towards the corresponding vector of covariance between the response and the profiles. Thus, a model accounting for more than 99% of the variation in profiles should show this similarity. In cases where larger parts of the profile variation is left unmodelled, such a similarity is no longer mandatory and the covariance graphs alone are not sufficient for interpretation. If only one display is to be used for interpretation, the target-rotated variable loadings represent the best choice since they are directly related to the predicted responses. However, a comparison between the target-rotated variable loadings and the covariance graphs is always profitable since it tells how well the LV regression model accounts for the profile variation of significance for the responses.

A final suggestion for obtaining useful displays for interpretation of profiles has to be made. In order to enhance the most important regions of the regression model, a procedure of successively squaring the covariances between the responses and the original or reproduced profiles may be useful. Such graphs have proved to be especially useful when profiles of very complex samples are used as input to LV regression modelling by effectively reducing the amplitudes of the profile regions of little importance for the model [17].

6 CONCLUSION

The vector of regression coefficients for a response variable represents the line of best fit in the reproduced variable space of the predictors. Thus, sample scores and variable loadings corresponding to the rotation using the vector of measured responses as target, can be obtained by projecting the reproduced predictor space onto the normalized regression vector. The target-rotated scores have been shown to be proportional to the predicted responses, while the explained variance of each predictor by a response is proportional to its squared target-rotated loading. Thus, the scores and loadings obtained by projecting on the normalized regression vector show the predictive ability of a latent-variable (LV) regression model and the importance of each predictor for a response, respectively. These two features can be portrayed simultaneously and quantitatively in a regression BIPLOT display.

Application of the regression BIPLOT technique to a two-component partial-least-squares (PLS) regression model showed that the BIPLOT presentation can give important clues about how interferences influence prediction and if and how such interferences can be accounted for in prediction. This information is as important as the detection and correction of unmodelled interferences. By presenting only the regression vector (the line labelled $\boldsymbol{b}_{\mathrm{vr}}$ in Figs. 3 and 4) and residual diagnostics as customary in standard packages for regression analysis, crucial information about the model is lost.

For samples characterized by whole profiles rather than a collection of peaks, covariance graphs of the original and reproduced predictor space is the most useful displays for interpretation and chemical validation of an LV regression model.

7 ACKNOWLEDGEMENTS

This paper is based on a lecture given at the SSC1 conference — First Scandinavian Symposium on Chemometrics, Lappeenranta, Finland, 1988. We would like to thank the organizers, and, in particular, Pentti Minkkinen, for the invitation to present this work. Grants from the Norwegian Research Council for Science and Humanities (NAVF) and from Norsk Hydro Research Center are gratefully acknowledged.

REFERENCES

1 N.R. Draper and H. Smith, *Applied Regression Analysis*, Wiley, New York, Chichester, Brisbane, Toronto, Singapore, 2nd ed., 1981.

2 I.T. Jolliffe, *Principal Component Analysis*, Springer Verlag, Berlin, 1986.

3 H. Martens, *Multivariate calibration — Quantitative interpretation of non-selective chemical data*, Dr. techn. thesis, Technical University of Norway, Trondheim, 1985.

4 O.M. Kvalheim, Latent-variable regression models with higher-order terms: An extension of response modelling by orthogonal design and multiple linear regression, *Chemometrics and Intelligent Laboratory Systems,* 8 (1990) in press.

5 D.M. Haaland and E.V. Thomas, Partial least-squares methods for spectral analyses. 1. Relation to other quantitative calibration methods and the extraction of qualitative information, *Analytical Chemistry*, 60 (1988) 1193–1202.

6 O.M. Kvalheim, Model building in chemistry, a unified approach, *Analytica Chimica Acta,* 223 (1989) 53–73.

7 O.M. Kvalheim, Interpretation of direct latent-variable projection methods and their aims and use in the analysis of multicomponent spectroscopic and chromatographic data, *Chemometric and Intelligent Laboratory Systems*, 4 (1988) 11–25.

8 R. Manne, Analysis of two partial-least-squares algorithms for multivariate calibration, *Chemometrics and Intelligent Laboratory Systems,* 2 (1987) 187–197.

9 O.M. Kvalheim, Latent-structure decompositions (projections) of multivariate data, *Chemometrics and Intelligent Laboratory Systems*, 2 (1987) 283–290.

10 O.M. Kvalheim, A partial-least-squares approach to interpretative analysis of multivariate data, *Chemometrics and Intelligent Laboratory Systems*, 3 (1988) 189–197.

11 A.L. van den Wollenberg, Redundancy analysis — an alternative for canonical correlation analysis, *Psychometrika*, 42 (1977) 207–219.

12 O.M. Kvalheim, A.A. Christy, N. Telnæs and A. Bjørseth, Maturity determination of organic matter in coals using the methylphenanthrene distribution, *Geochimica et Cosmochimica Acta*, 51 (1987) 1883–1888.

13 S. Wold, Pattern recognition by means of disjoint principal components models, *Pattern Recognition*, 8 (1976) 127–139.

14 K. Mardia, J. Kent and J. Bibby, *Multivariate Analysis,* Academic Press, London, 1979.

15 I.A. Cowe, J.W. McNicol and D.C. Cuthbertson, A designed experiment for the examination of techniques used in the analysis of near infrared spectra. Part 1. Analysis of spectral structure, *The Analyst,* 110 (1985) 1227–1232.

16 I.A. Cowe, J.W. McNicol and D.C. Cuthbertson, A designed experiment for the examination of techniques used in the analysis of near infrared spectra. Part 2. Derivation and testing of regression models, *The Analyst*, 110 (1985) 1233–1240.

17 A.A. Christy, O.M. Kvalheim, T.V. Karstang, K. Øygard and B. Dahl, Maturity of kerogen and asphaltenes determined by partial-least-squares (PLS) calibration and target-rotation of diffuse reflectance Fourier-transformed infrared spectra, in preparation.