Analytica Chimica Acta, 185 (1986) 19—32
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

AN EXAMPLE OF 2-BLOCK PREDICTIVE PARTIAL LEAST-SQUARES REGRESSION WITH SIMULATED DATA

PAUL GELADI

Chemometrics Group, Department of Organic Chemistry, Umeå University, S 901 87 Umeå (Sweden)

BRUCE R. KOWALSKI

Laboratory for Chemometrics, Department of Chemistry, University of Washington, Seattle, WA 98195 (U.S.A.)

(Received 29th October 1985)

SUMMARY

Practical examples with simulated data are described to illustrate the material described in the preceding tutorial on partial least-squares regression. Starting from the perfect model for two dimensions, noise, nonlinearities and interference are added gradually in order to study their influence. The examples are related to calibration in chemical analysis.

An algorithm for partial least-squares (PLS) regression was described in the preceding paper [1] in which it was explained how PLS emerged from studies of the flaws in multiple linear regression (MLR) and in principal component regression (PCR). The advantages of the PLS method were outlined. Examples of how the PLS method can be used are given here. Simulated data with known properties are used because such data are more suitable for illustrative purposes than practical laboratory data with their unknown sources and extents of vitiating effects. Extensions to real situations can readily be visualized. The focus here is on calibration in chemical analyses. This paper has two purposes: to provide a worked example with all the numerical data for testing computer programs and to study the influence of random noise, nonlinearities and interfering extra components. This last topic is considered in an empirical way with the hope that a more theoretical approach can be developed later. The example is a regression for two chemical components (i.e., two dimensions, tank 2), into which complications including noise, nonlinearities and interferences are gradually introduced.

The PLS method

In the PLS method, an X block of independent variables is related to a Y block of dependent variables by describing outer relations [1]:

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E} = \sum \mathbf{t}_h \mathbf{p}_h' + \mathbf{E} \qquad (1)$$

0003-2670/86/$03.50

$$\mathbf{Y} = \mathbf{UQ}' + \mathbf{F}^* = \sum \mathbf{u}_h \mathbf{q}_h' + \mathbf{F}^* \quad (2)$$

(Symbols are defined in Table 1 [1]). The sums are over the number of dimensions necessary or desired. The regression part of PLS is an inner relation:

$$\hat{\mathbf{u}}_h = b_h \mathbf{t}_h \quad (3)$$

where $b_h = \mathbf{u}_h' \mathbf{t}_h / \mathbf{t}_h' \mathbf{t}_h$; the "hat" indicates that the vector is an estimated one. This leads to the mixed relation [1]

$$\mathbf{Y} = \mathbf{TBQ}' + \mathbf{F} \quad (4)$$

where $\|\mathbf{F}\|$ has to be minimized with the condition that $\|\mathbf{E}\|$ in Eqn. 1 is reduced.

In order to obtain orthogonal scores as in PCR, it is necessary to introduce weighting (**w**); PLS is achieved by using the residuals after each dimension:

$$\mathbf{E}_h = \mathbf{E}_{h-1} - \mathbf{t}_h \mathbf{p}_h'; \ \mathbf{X} = \mathbf{E}_0 \quad (5)$$

$$\mathbf{F}_h = \mathbf{F}_{h-1} - b_h \mathbf{t}_h \mathbf{q}_h'; \ \mathbf{Y} = \mathbf{F}_0 \quad (6)$$

Graphic representations of Eqns. 1 and 2 and of the matrices and vectors used in the PLS algorithm for each dimension are in the preceding paper [1]. The PLS dimensions can also be called components or factors. In this paper, the term dimension is used in order to avoid confusion; the symbol **B** indicates the diagonal matrix of the b_h.

It is important to recognize some of the properties of the PLS model: SS_Y, F and $XVAL$. SS_Y is the residual for the Y block after n dimensions. This is usually expressed as the sum of squares for the residual matrix $\mathbf{F}_h$ (Eqn. 4) divided by the number of members of the matrix. For the Y block, this is termed SS_Y. Sometimes, a similar statistic (SS_X) can be useful for the X block.

One way of evaluating the PLS model is to check the inner relation (Eqn. 3). The F-test can be applied to the linear regression in this case (see Draper and Smith [2]). High F values indicate a good inner relation and low F values a poor inner relation; very low F values show that the inner relation is meaningless. The $XVAL$ method [3, 4] describes the predictive quality of a certain PLS dimension by the use of cross-validation and is calculated on the "leave-some-out" basis. The method is described in Appendix A. A value of $XVAL > 1$ means a nonsignificant dimension; $XVAL < 1$ means a significant dimension. The cross-validation for the calculation of $XVAL$ means extra calculations; **F** on the inner relation and SS_Y do not need these.

DATA MATRICES IN CALIBRATION FOR CHEMICAL ANALYSIS

Multivariate calibration means measuring a vector of properties (variables) for calibration standards of known content. This vector may be spectral intensities, current measurements or any relevant collection of data. In this

example, two analytes A and B are used and the responses of the sensors are shown in Fig. 1. Table 1 gives the concentrations of each of these analytes in nine different samples. It is often advantageous to use a factorial design for such an example. Here, a 3^2 design was chosen because of the nonlinearities that will be introduced later. However, it may not always be possible to mix suitable standards (e.g., in food research, geology, environmental samples). In such cases, a collection of samples spanning the maximal expected range in each variable dimension must be chosen. Suitable choice of a design can lead to good results (see below).

In the simple case considered, the data matrix will be the sum of two planes, each consisting of the outer product of the concentration vector and the response vector as in Fig. 2(a). The X-matrix is built from

$$\mathbf{X} = \mathbf{c}_1\mathbf{s}_1' + \mathbf{c}_2\mathbf{s}_2' \tag{7}$$

where $\mathbf{c}_i$ are the concentration vectors from Table 1 and $\mathbf{s}_i'$ are the responses from Fig. 1. In reality, this situation is complicated by the existence of random noise, nonlinearities and interferences, as indicated in Fig. 2(b). The sum matrix in Fig. 2(b) is the independent (X) matrix for the regression of the two vectors of concentrations from Table 1. The following eight cases are possible: linearity with and without random noise for only two components; nonlinearity with and without random noise for two components; linearity with and without random noise for two components and one interfering (background) component, i.e., an extra component not recognized during calibration; and nonlinearity with and without random noise for two components and one interfering component. It is clear here that the last case with random noise, and even more complicated situations, will be encountered in practice, but for illustrative purposes four of the above models will be considered in the following paragraphs. The data will always be used in mean-centered form.

Case 1: Linearity without random noise for two components

In this case (cf. Fig. 9 [1]), it is known in advance that there will only be two components. The criteria for testing the model dimensionality, SS_Y, F and *XVAL*, indicate this. Indeed, for the second PLS dimension, *XVAL*

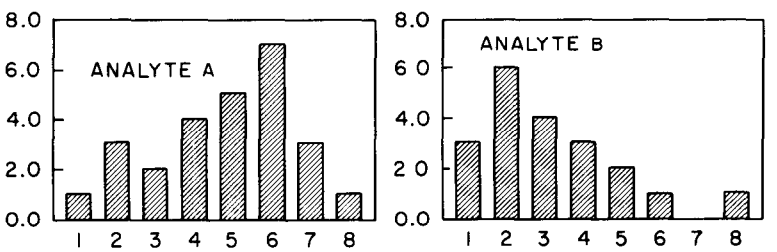


Fig. 1. The responses ("spectra") of an array of eight sensors for analytes A and B. The numbers are arbitrary.

TABLE 1

Concentrations of the analytes A and B used for the PLS calibration

Sample	1	2	3	4	5	6	7	8	9
Concn. (analyte A)	1	3	5	1	3	5	1	3	5
Concn. (analyte B)	1	1	1	3	3	3	5	5	5

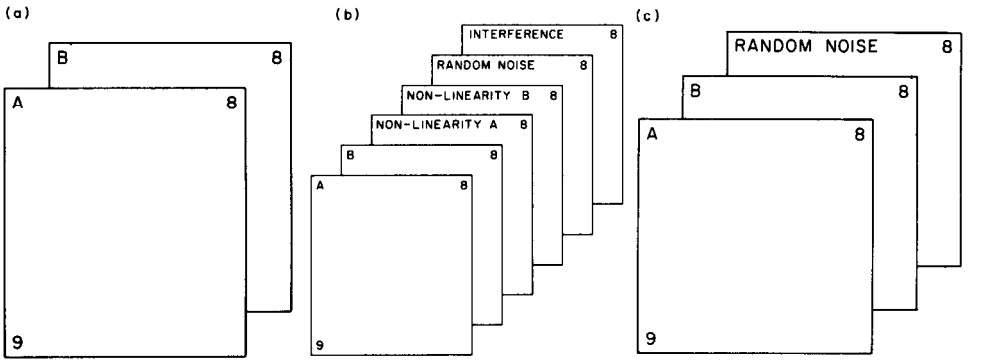


Fig. 2. Composition of the data matrices X: (a) for case 1 with two components (analytes A and B) and linearity but without noise and interfering or background components; (b) the most complicated case in which both components have nonlinearities, and both random noise and an interfering (or background) component are present; (c) random noise only added to (a).

becomes zero (the rounding error in most computers) and F becomes infinity (the inverse of the rounding error in most computers). Both SS_X and SS_Y become zero after two components.

It can be shown that the X scores obtained are exactly the same (sometimes with different sign) as those obtained by PCR, i.e., PLS and PCR behave identically in the noise-free case. For PLS, the loadings and weights for the X block are identical. The inner relationships are straight lines, as expected, and so are not shown here. It is of interest to look at the scores plot in Fig. 3. Because of the choice of the concentrations in the experimental design, the plot has the shape of a grid. It should be noted that each dimension describes exactly 50% of the variance in the Y block. This is true both for variance-scaled and non-variance-scaled data. In real cases, noise-free measurements will be very rare. The data for this case are given in Appendix B.

Conclusion. For case 1, SS_X, SS_Y, F and $XVAL$ clearly show that there are two dimensions. The PLS and PCR methods are equivalent in this case, and the PLS loadings and weights are the same. These results are independent of linear scaling of the variables.

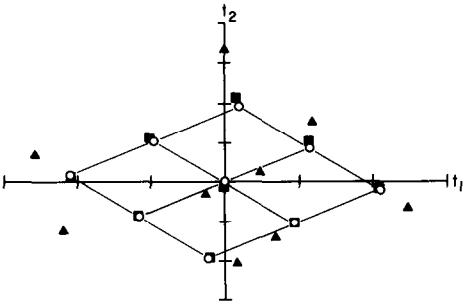


Fig. 3. The scores plot for the data matrix: (○) noise-free data; (■) 10% noise added; (▲) 100% noise added. Because of the choice of the concentrations as a 3^2 design, the points representing the samples fall on a grid for the noise-free data. Large amounts of noise completely destroy the grid structure.

Case 2: Linearity with random noise for two components

The X matrix in this case is shown schematically in Fig. 2(c). The best way to describe the influence of noise is to add increasing amounts of noise to the noise-free data and to see how the PLS models behave. In order to do this, mean-centered noise distributed as $N(0, p)$ (with p = 0.01, 0.02, 0.05, 0.1, 0.2, 0.5 and 1.0) was added to every column of the X matrix after variance-scaling. This way of adding noise ensures that each variable is affected equally; there are many other ways of adding noise to data, so this method simply illustrates a general trend that can be applied to other cases. No noise was added to the Y block; it is assumed that the concentrations of the standards are accurate.

The addition of random noise increases the rank of the X matrix from 2 to the maximum (8 in this case), because the noise is distributed equally in all the dimensions of the variable space. Figure 4 provides a graphic explanation of this phenomenon. This can also be checked by doing a principal component analysis on the noise alone.

As long as the components have a larger variance than the noise in their own dimension, there will be an adequate 2-dimensional model, but the noise may cause slight rotations of the scores. For larger amounts of noise, total confusion of noise and data can be expected. Because the noise will be of different character in real samples, increasing amounts of noise are expected to give increasing prediction errors. It should be noted that random noise is hypothetical. In practice, application of this random noise behaves as a bias. The difference from a real bias is that noise can be averaged out by repeating measurements.

Results are given in Table 2. They are based on only one application of the noise, and so the picture given is only rough. For up to 5% noise added, the first dimension describes almost 50% of the Y block and two dimensions describe almost all the rest. The F-test on the inner relation shows decreasing values with increasing noise, meaning that the inner relation becomes worse.

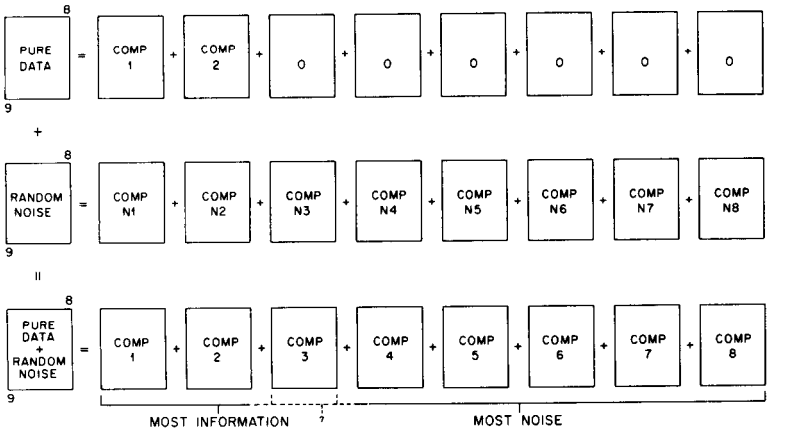


Fig. 4. The influence of noise on the PLS model. For pure analyte data (A and B) the rank of the X block is 2. For only random noise, the rank of the X block is maximal (here 8). For random noise added to the pure analyte data, there will be some disorder: most useful information will end up in the first dimensions and most noise will end up in higher dimensions. PLS can handle small amounts of random noise but higher amounts of noise can prevent the construction of useful models.

TABLE 2

The influence of noise added to the X block on the PLS model[a]

After	Noise added, $N(0, p)$, p =						
	0.01	0.02	0.05	0.1	0.2	0.5	1.0
A: F-test on the inner relation							
Dim. 1	2000000	11800	66000	23000	7300	400	32
Dim. 2	116000	69000	26000	1460	524	59	14
Dim. 3	—	23	5	69	36	3	2.2
Dim. 4	—	—	—	57	21	16	15
B: SS_Y *(%, 100% for no. dimensions calculated)*							
Dim. 1	50	50	50	50.02	50.05	50.9	59
Dim. 2	0.001	0.008	0.019	0.26	0.71	6.2	26
Dim. 3	—	0.005	0.013	0.036	0.15	4.8	22
C. XVAL							
Dim. 1	0.93	0.93	0.93	0.92	0.92	0.91	1.04
Dim. 2	0.009	0.015	0.028	0.076	0.22	0.47	0.87
Dim. 3	1.04	0.99	0.93	0.48	0.57	1.43	1.11
Dim. 4	—	1.08	1.01	0.85	0.75	0.93	0.92

[a]Blank spaces mean that the PLS stopped because residuals were too low.

There is also a steep fall in F after two dimensions. For very noisy data, less significant inner relations are found or the more significant relations appear in later dimensions. *XVAL* shows a 2-dimensional model for up to 5% noise. With 10% or more noise, the models are of higher rank, which confirms the failure to distinguish between data and noise. Figure 3 shows the influence of noise on the scores plot.

Prediction was done by using the 2-dimensional model for noise-free data with test sets with different noise contents. The results are given in Table 3 as the square root of the PRESS (see Appendix A). This statistic behaves like a standard deviation. The results in the table indicate that $(\text{PRESS})^{1/2}$ increases almost linearly with the amount of noise added. The very low result for $p = 0.05$ is due to the nature of the noise applied, as was shown by repeating some calculations.

Conclusion. For case 2, random noise in the X matrix gives it maximum rank. Increasing amounts of random noise shift information towards higher dimensions and make it difficult to decide on the correct number of dimensions. Up to 5% noise gives a precise, quantitative model, but over 5% and up to 100% noise provides a model that only allows qualitative conclusions. Prediction shows that the square root of PRESS increases almost linearly with the random error added.

Case 3: Nonlinearity without random noise for two components

Nonlinearities can be introduced by adding terms that are quadratic in the concentrations:

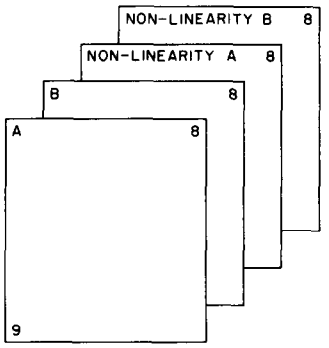
$$X = \mathbf{c}_1\mathbf{s}_1' + \mathbf{c}_2\mathbf{s}_2' + b_1\mathbf{c}_1^*\mathbf{c}_1\mathbf{s}_1' + b_2\mathbf{c}_2^*\mathbf{c}_2\mathbf{s}_2' \qquad (8)$$

where * indicates the direct vector product and the b_i are constants determining how much nonlinearity is present. A representation with matrices as planes is given in Fig. 5. For linear responses, $b_1 = b_2 = 0$. The Y block contains the linear concentrations so that the nonlinearities will obviously cause

TABLE 3

The result of random error added to a test set on the prediction error, $(\text{PRESS})^{1/2}$ for analytes A and B

Random error $N(0, p)$ p =	$(\text{PRESS})^{1/2}$		$(\text{PRESS})^{1/2}/p$	
	Anal. A	Anal. B	Anal. A	Anal. B
0.01	0.011	0.013	1.1	1.3
0.02	0.016	0.020	0.80	1.0
0.05	0.027	0.024	0.54	0.48
0.1	0.091	0.149	0.91	1.49
0.2	0.194	0.235	0.97	1.18
0.5	0.517	0.530	1.03	1.06
1.0	1.19	1.56	1.19	1.56



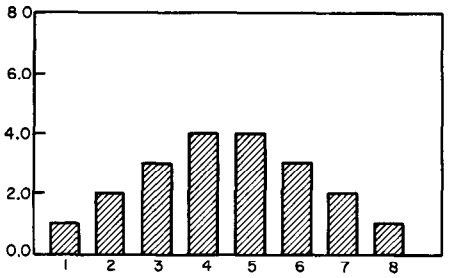


Fig. 5. The data matrix X for two components with nonlinearities. Because of the way that the nonlinearities are defined in Eqns. 8 and 9, each analyte has a set of them.

Fig. 6. The responses (spectrum) of an array of eight sensors for the interfering or background component. The numbers are arbitrary units.

problems. These nonlinearities are not random, but act as a concentration-dependent bias.

The nonlinearities do not increase the rank of the X matrix. This can be seen by writing Eqn. 8 in another way:

$$X = (\mathbf{c}_1 + b_1\mathbf{c}_1^*\mathbf{c}_1)\mathbf{s}_1' + (\mathbf{c}_2 + b_2\mathbf{c}_2^*\mathbf{c}_2)\mathbf{s}_2' \qquad (9)$$

The results for some model calculations with different b_1 and b_2 are listed in Table 4. The data were used in mean-centered and variance-scaled form. These results show that increasing amounts of nonlinearity in a certain component introduce increasing amounts of residual sum of squares for the respective column vector in the Y block. Each analyte behaves independently with respect to the nonlinearity of the other. The F values for the inner relations show that these also become rather curved. The fact that the rank of the X block does not increase, as shown in Eqn. 9, is typical of this mode of adding nonlinearities. The situation changes when cross-terms are added, e.g., if there is mutual interference between the spectra of two analytes as happens frequently in real situations.

Prediction with nonlinear data was tested by using the model parameters for the linear data for prediction with test sets containing nonlinear data. Because of the independent behavior of the two analytes, simplified results are given in Table 5. These results show that higher b_i values cause higher errors and that the higher the concentration the higher the error. Expressing the inaccuracies as percentage shows that they increase linearly with b_i and quadratically with concentration. This is what would be expected from Eqns. 8 and 9.

Conclusion. For case 3, nonlinearities of the type used here do not increase the rank of the X block. Increasing amounts of nonlinearity give larger residuals for the Y block. The results for each analyte are not influenced by the

TABLE 4

Some properties of PLS models for nonlinear data[a]

b_1	b_2	SS_{y1}	SS_{y2}	F_1	F_2
0.01	0	0.006	0	96100	153000
0.1	0	0.26	0	1790	5400
1.0	0	1.32	0	264	10900
0	0.01	0	0.006	137000	103000
0	0.1	0	0.26	2290	3300
0	1.0	0	1.32	276	4100
0.01	0.01	0.006	0.006	59000	59000
0.1	0.1	0.26	0.26	1340	1340
1.0	1.0	1.32	1.32	257	257

[a] b_i and SS_{yi} are for the first and second analyte, respectively, for the complete 2-dimensional model. F_i are for the first and second inner relations.

TABLE 5

Prediction results for nonlinear data

True concn.	Concn. found for b_i =		
	0.01	0.1	1.0
1	1.01 (1%)	1.1 (10%)	2.0 (100%)
3	3.09 (3%)	3.9 (30%)	12 (300%)
5	5.25 (5%)	7.5 (50%)	30 (500%)

nonlinearities for the other analyte. For prediction, the inaccuracies increase linearly with b_i and quadratically with concentration.

Case 4: Linearity without random noise for two components plus one interfering component

There are many possibilities for simulating an interfering or background component. The uncommon case in which the interfering component gives exactly the same analytical response (e.g., spectrum) as one of the required components will not be discussed here, nor will the case in which the concentrations of the interferent are exactly correlated with those of one of the analytes. A trivial situation occurs if the interfering component has the same concentration in all samples; the constant background so obtained is simply removed by mean-centering. The spectrum of the interfering product is given in Fig. 6. A simple pattern of concentrations of interferent is given in Table 6. The amount d can be varied to study its influence on the PLS model. A schematic representation of the X block for this example is the same as that shown in Fig. 2(c); interference simply replaces random noise in the figure. With the interfering component present, the X matrix will have rank 3. The

TABLE 6

Concentrations of analytes A and B and interferent I. The amount d is varied to study its influence on the PLS model

Sample number	Concentrations		
	A	B	I
1	1	1	1d
2	3	1	2d
3	5	1	1d
4	1	3	2d
5	3	3	1d
6	5	3	2d
7	1	5	1d
8	3	5	2d
9	5	5	1d

amount of variation of the Y block explained by the first two components will depend on d and on the correlation of the interferent spectrum with the analyte spectra.

Some results for the PLS model for different d values are shown in Table 7. The F-test clearly shows rapid deterioration of the inner relation as the amount of interferent increases. The SS_Y shows a more than linear increase in the amount of the Y block, which remains unexplained by a 2-dimensional model until a plateau is reached. The $XVAL$ indicates a 2-dimensional model, but it is poor for high values of d. A complicating problem with interferences is that there can be a large variety of spectra to choose from and that the influence on the model depends greatly on the correlation between the analytical responses of the interferents and those of the analytes.

Conclusion. For case 4, the prediction properties were checked for data containing interferent with the model for interference-free data. There is a positive bias which is linearly related to the concentration of the interfering component. Every interfering component increases the rank of the X block by 1. The influence of the interferent on the model is nonlinearly related to the amount of interferent present.

General conclusion

The use of simulated data illustrates that many interesting properties of PLS regression can be found by testing models and prediction. The criteria of dimensionality, SS_Y, F and $XVAL$, each contribute particularly in different situations. In the general case, calculation of all three criteria of dimensionality (and maybe even others) is necessary to obtain a fairly realistic picture of the behavior of a PLS model. Only four of the eight cases mentioned initially were actually tested but it is hoped that these papers will provide a better understanding of the PLS method.

TABLE 7

Some properties of the PLS model as a function of d

	d =						
	0.1	0.2	0.5	1.0	2.0	5.0	10.0
A: F-test on the inner relation							
Dim. 1	18400	4600	740	186	47	8.9	2.7
Dim. 2	172000	19000	7000	1860	570	239	384
B: SS_Y *(%, 100% start situation)*							
Dim. 1	50.02	50.08	50.5	51.8	56.4	72.0	86.0
Dim. 2	0.02	0.08	0.52	0.2	7.1	25.0	40.0
SS_Y/d	0.2	0.4	1.0	2.0	3.5	5.0	4.0
C: XVAL							
Dim. 1	0.93	0.93	0.93	0.94	0.96	1.03	1.07
Dim. 2	0.02	0.05	0.11	0.22	0.40	0.75	0.90

The colleagues in Seattle and Umeå are thanked for their remarks and stimulating discussions. The Science Department of the Belgian Ministry of Foreign Affairs provided P. Geladi with a NATO grant 10/B/84/BE, which is gratefully acknowledged.

Appendix A Cross-validation in PLS

In building and using a regression model, robust models are needed which will allow prediction with minimal error. The prediction quality can be tested by predicting the independent variables from the dependent ones for some standards of known composition and comparing the results predicted from the model with the known composition. But in many cases, such absolute standards are not available and when they are available it is best to include them in the model. Cross-validation is a method for testing the internal consistency of a regression model while it is being constructed.

Cross-validation can be done for every dimension or for a fixed number of dimensions. The important thing is that a fraction of the calibration standards are omitted in the model building. This fraction can vary from a single object (bootstrapping) to half the objects. This can be written as

$X_m \langle \longrightarrow Y_m$

$X_t \qquad Y_t$

The successive steps are: (1) the model is built between X_m and Y_m; (2) from the model parameters, $\hat{Y}_t$ is predicted from X_t; (3) the sum of squares ss of $(Y_t - \hat{Y}_t)$ is calculated; (4) this sum of squares is added to SS; (5) the fraction that was omitted is reinstated and another fraction of approximately the same size is omitted. Calculations then revert to step 1.

This procedure is repeated until all the objects have been in X_t once. Then SS is the accumulated prediction error sum of squares for all the objects. Finally SS is divided by the size of the Y matrix to give PRESS. (The size of the Y matrix can be replaced by a more fancy number of degrees of freedom.) When this is tested for every dimension, the square root of PRESS is compared with the standard deviation of the Y block of the previous dimension as follows: $XVAL = (\text{PRESS})^{1/2}/s_Y$. If this $XVAL > 1$, the current

TABLE A1

The raw data matrices X and Y for case 1 with mean and standard deviations (SD) and the final data matrices

Object	Variable									
	$x1$	$x2$	$x3$	$x4$	$x5$	$x6$	$x7$	$x8$	$y9$	$y10$
Raw data										
1	4	9	6	7	7	8	3	2	1	1
2	6	15	10	15	17	22	9	4	3	1
3	8	21	14	23	27	36	15	6	5	1
4	10	21	14	13	11	10	3	4	1	3
5	12	27	18	21	21	24	9	6	3	3
6	14	33	22	29	31	38	15	8	5	3
7	16	33	22	19	15	12	3	6	1	5
8	18	39	26	27	25	26	9	8	3	5
9	20	45	30	35	35	40	15	10	5	5
Mean	12	27	18	21	21	24	9	6	3	3
SD	5.4772	11.619	7.7459	8.6603	9.3274	12.247	5.1962	2.4495	1.7321	1.7321
Mean-centered and variance-scaled data matrices										
1	−1.4606	−1.5492	−1.5492	−1.6166	−1.5010	−1.3064	−1.1547	−1.6330	−1.155	−1.155
2	−1.0954	−1.0328	−1.0328	−0.69282	−0.42885	−0.16330	0	−0.81650	0	−1.155
3	−0.73030	−0.51640	−0.51640	0.23094	0.64327	0.97980	1.1547	0	1.155	−1.155
4	−0.36515	−0.51640	−0.51640	−0.92376	−1.0721	−1.1431	−1.1547	−0.81650	−1.155	0
5	0	0	0	0	0	0	0	0	0	0
6	0.36515	0.51640	0.51640	0.92376	1.0721	1.1431	1.1547	0.81650	1.155	0
7	0.73030	0.51640	0.51640	−0.2309	−0.6433	−0.9798	−1.1547	0	−1.155	1.155
8	1.0955	1.0328	1.0328	0.69282	0.42885	0.16330	0	0.81650	0	1.155
9	1.4606	1.5492	1.5492	1.6166	1.5010	1.3064	1.1547	1.6330	1.155	1.155

dimension has no predictive properties. If $XVAL < 1$, the current dimension is useful and the next one should be tested.

An alternative is to collect *SS* for every variable in the Y block instead for over the whole block. This then gives an *XVAL* for each Y variable. Checking these *XVAL*'s is considered a fine-tuning of the cross-validation process and is certainly worthwhile when many Y variables are considered.

Cross-validation can also be done for many dimensions simultaneously. In this case, PRESS is again calculated but for whole models of increasing dimensionality. If this PRESS is plotted against the model dimensionality, it gives a minimum for the model with the best predictive properties. An example is shown in Fig. 6 of the earlier paper.

Appendix B: Example for case 1

This appendix presents a worked example for case 1. The raw data, mean and standard deviations, and the mean-centered, variance-scaled data matrices are shown in Table A1.

Because of the choice of a 3^2 design, interesting patterns show up in both blocks. The center element of the design has zeroes for every variable. The seventh X variable is perfectly correlated to the first Y variable because the second sensor sensitivity is zero here. The columns for the second and third X variables are identical because the ratios for the sensor sensitivities for both are the same. The loadings and scores for two dimensions are shown in Tables A2 and A3. It is very common that the first X-block loading has all-positive elements. Because of the orthogonality requirement, the second X-block loading must have some negative elements. The special property of orthogonality for the Y-block loadings is only valid for this example. The construction of the data matrix as a 3^2 design together with mean-centering and variance-scaling gives symmetrical properties which makes it easier to understand what is happening and to detect erratic behavior.

TABLE A2

Loadings for two dimensions[a]

p	Variable									
	*x*1	*x*2	*x*3	*x*4	*x*5	*x*6	*x*7	*x*8	*y*9	*y*10
1	0.33026	0.35600	0.35600	0.38850	0.37018	0.33148	0.29936	0.38729	0.76965	0.6384
2	−0.4480	−0.3417	−0.3417	0.04154	0.26034	0.44373	0.54149	−0.0787	0.63847	−0.7696

[a]In this case, the loadings and weights for the X block are identical. The Y block has no weights.

TABLE A3

Scores for two dimensions

	Object									
	1	2	3	4	5	6	7	8	9	b[a]
*t*1	−4.180	−1.895	0.3894	−2.285	0	2.285	−0.3894	1.895	4.180	0.3889
*t*2	0.1786	1.048	1.917	−0.8693	0	0.8693	−1.917	−1.048	−0.1786	0.8481
*u*1	−1.626	−0.7372	0.1515	−0.8887	0	0.8887	−0.1515	0.7372	1.626	
*u*2	0.1515	0.8887	1.626	−0.7372	0	0.7372	−1.626	−0.8887	−0.1514	

[a]Regression coefficients for the inner relation.

REFERENCES

1 P. Geladi and B. Kowalski, Anal. Chim. Acta, 185 (1986) 1.
2 N. Draper and H. Smith, Applied Regression Analysis, Wiley, New York, 1981, pp. 33—40.
3 S. Wold, H. Martens and H. Wold, in Ruhe and Kågström (Eds.), Matrix Pencils, Springer, Heidelberg, pp. 286—293.
4 S. Wold, in H. Martens and H. Russwurm (Eds.), Food Research and Data Analysis, Applied Science Publishers, London, 1983, pp. 147—188.