JOURNAL OF CHEMOMETRICS
J. Chemometrics **13,** 461–471 (1999)

A SERIAL EXTENSION OF MULTIBLOCK PLS

ANDERS BERGLUND* AND SVANTE WOLD
Research Group for Chemometrics, Department of Organic Chemistry, Umeå University, S-901 87 Umeå, Sweden

SUMMARY

A novel multiblock PLS algorithm called S-PLS (serial PLS) is presented. S-PLS models the separate predictor blocks serially, making it a supplement to hierarchical PLS. In the S-PLS algorithm the predictor blocks are connected only via the response **Y**. The block models are calculated using the **Y** residuals from the previous block model. This allows for an independent interpretation of the separate block models. In each block model the classical PLS algorithm is used. The principles of S-PLS are demonstrated on two chemical applications. In the first example, which is non-linear, S-PLS makes it possible to separate the linear and non-linear parts in the model. The second example illustrates how a model with two predictor blocks can be analysed with S-PLS.

KEY WORDS: multiblock PLS; non-linear regression; latent variables

INTRODUCTION

In the fields of chemometrics and statistics the common problem is to analyse measured or calculated variables for a set of observations collected in a data table, the predictor block. This predictor block or **X** block can come from many different sources. In chemistry it may contain different properties for a set of molecules or the state of some chemical process. The predictor block can then be analysed with several techniques, e.g. principal component analysis (PCA),[1] in order to find classes, outliers, etc. The predictor block may also be related to a response, the **Y** block, which may be the biological activity for molecules or a yield for the process. Partial least squares (PLS)[2] has proven to be valuable in these applications.

Today, as data get more and more complex, several types of predictor blocks may be collected in a specific problem, e.g. industrial processes.[3–6] The easiest way to handle several predictor blocks is to merge the different blocks into a single predictor block and then use PCA or PLS. This is simple and straightforward and works well in many situations. The problem is that the result can be hard to interpret owing to the large number of variables. Furthermore, the different blocks may have different numbers of variables with different variances, making a block scaling necessary. Which type of block scaling to use is a decision that is sometimes hard to make. In order to improve these models with several predictor blocks, there have been recent papers published on the use of multiblock models.[3,4,6,7] The algorithms in these papers are all based on an algorithm by H. Wold,[8] who outlined both a PLS and a PCA algorithm for several predictor blocks. This extension of the classical PCA and PLS algorithm was also presented in a paper by S. Wold *et al.*[9]

A later paper by S. Wold *et al.*[4] presents a modified PLS algorithm called hierarchical PLS (Hi-

* Correspondence to: A. Berglund, Research Group for Chemometrics, Department of Organic Chemistry, Umeå University, S-901 87 Umeå, Sweden. E-mail: anders.berglund@chem.umu.se
Contract/grant sponsor: Swedish Science Research Council (NFR).

CCC 0886–9383/99/040461–11 $17.50
Received 15 June 1998
Accepted 17 March 1999

PLS) together with a process example. In all these algorithms the different predictor blocks are treated in a parallel mode, i.e. all predictor blocks are related to the same response variables. The algorithms try to find an underlying structure that is present in all the predictor blocks.

Another important question is whether there is unique information in the separate blocks. If, for instance, two blocks of data are collected for the same set of objects from NIR (near-infrared reflectance) and NMR (nuclear magnetic resonance), will one block be sufficient or are both needed? If the latter, which information is not unique for the blocks?

Another example where there is a clear division between different predictor blocks is in the INLR (implicit non-linear latent variable regression) approach.[10] This method has proven to work well in several non-linear examples. Here the first block contains the original variables and the second block contains the squares of the original variables. To improve the predictions and to make it possible for the model to converge to a linear PLS model, the method was modified with a pre-weighting of the expanded blocks.[11] In the INLR model it is interesting to divide the model into a linear part and a non-linear part. Also in this situation it is interesting to see if the second block is necessary for a good model.

In this paper we present an alternative multiblock PLS algorithm called serial PLS (S-PLS). In this method the predictor blocks are treated in a serial mode where they are only connected via the response **Y**. There is also only one level of information, not two as in Hi-PLS. With S-PLS it is easier to see the differences between the predictor blocks. S-PLS also divides the models for the separate predictor blocks so that the results can be examined separately.

BACKGROUND

PLS

In PLS, both the predictor block and the response block are decomposed with latent variables **T** according to

$$\boldsymbol{X} = \boldsymbol{T}\boldsymbol{P}^{\mathrm{T}} + \boldsymbol{E}, \qquad \boldsymbol{Y} = \boldsymbol{T}\boldsymbol{C}^{T} + \boldsymbol{F} \tag{1}$$

where **P** is the loading matrix for the **X** matrix, **C** is the weight matrix for **Y**, and **E** and **F** are the residuals for the **X** and **Y** matrices respectively. This is done so that **T** describes both **X** and **Y** as well as possible. In this way, PLS finds the underlying structure, the latent variables that are related to the response. These latent variables are linear combinations of the measured, observable, variables. It is also possible to calculate the regression coefficients for the regression model with PLS:

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{B} + \boldsymbol{F} \tag{2}$$

For PLS the regression coefficient matrix **B** is estimated from a subspace of **X**. This subspace is defined by the weight vectors **W**. There have recently been several published papers on the theory of PLS. Phatak and De Jong[12] explain PLS from a geometrical point of view. Burnham *et al.*[13] compare PLS and related techniques using a common framework. This is the first time an objective function has been developed for the PLS algorithm where there are several responses and for more than one PLS dimension. Similar results have also been obtained using the continuum regression method.[14] There also exist several other publications about the theory of PLS.[15–18]

Non-linear PLS

A fairly common problem in chemical applications is that the relationship between **X** and **Y** is non-

linear. Since PLS is a linear method, this may cause problems in the modelling. There are several suggestions on how to make PLS suitable for non-linear problems.[19–29]

One easy way to make PLS deal with non-linear examples is to use the INLR approach. We showed in Reference 10 that in the expansion of the predictor block the cross-product terms $\mathbf{X}_k\mathbf{X}_{k+1}$ can be excluded if there is a latent structure present in the data, i.e. the rank of the matrix is much smaller than the number of variables. This was done by showing that the block with only the squared **x** variables includes both the squares and the cross-product terms of the latent variables:

$$\boldsymbol{x}_k^2 = \sum_{a=1}^{A}\sum_{b=1}^{A} \boldsymbol{t}_a\boldsymbol{t}_b\boldsymbol{p}_{ka}^T\boldsymbol{p}_{kb}^T + \boldsymbol{e} \tag{3}$$

This is in good agreement with the PLS philosophy, since it is the latent variables that are related to the response and not the directly observable variables. The pre-weighting of the expanded blocks presented in Reference 11 improved the predictions. In this modification the expanded block is pre-weighted with a scalar ranging from zero up to some predefined number. Cross-validation[30,31] was used to estimate the best value of the scalar and the optimum number of PLS components. This also makes it possible to obtain a linear PLS model if the scalar value equals zero.

SERIAL PLS (S-PLS)

In the following, only two predictor blocks will be considered in the explanation of S-PLS, denoted $\mathbf{X}_1$ and $\mathbf{X}_2$. The algorithm is easily extended to the case with three or more predictor blocks.

The philosophy behind S-PLS differs from the philosophy of other multiblock PLS algorithms. Instead of treating all the blocks in a parallel mode, the blocks are treated in a serial mode. The variation in **Y** that is not modelled by the first block will be left to be explained by the second block. In terms of the latent variable model, S-PLS can be written as

$$\begin{aligned} \boldsymbol{X}_1 &= \boldsymbol{T}_1\boldsymbol{P}_1^{\mathrm{T}} + \boldsymbol{E}_1 \\ \boldsymbol{X}_2 &= \boldsymbol{T}_2\boldsymbol{P}_2^{\mathrm{T}} + \boldsymbol{E}_2 \\ \boldsymbol{Y} &= \boldsymbol{T}_1\boldsymbol{C}_1^{\mathrm{T}} + \boldsymbol{T}_2\boldsymbol{C}_2^{\mathrm{T}} + \boldsymbol{F} \end{aligned} \tag{4}$$

where each block results in a PLS model connected only via the response **Y**. The regression model can analogously be divided into two parts:

$$\boldsymbol{Y} = \boldsymbol{X}_1\boldsymbol{B}_1 + \boldsymbol{X}_2\boldsymbol{B}_2 + \boldsymbol{F} \tag{5}$$

where $\mathbf{B}_1$ and $\mathbf{B}_2$ are the regression coefficient matrices for the respective predictor blocks and **F** contains the residuals. Figure 1 shows how the two blocks are connected via the response **Y**. The first model is calculated using the residuals from the second model, and the second model is calculated using the residuals from the first model (Figure 1). This is done in an iterative way, where the residuals are updated after each iteration. The algorithm can be divided into the following steps.

1. Set $\mathbf{F}_2 = \mathbf{Y}$.
2. Calculate the first PLS model with $\mathbf{X}_1$ and $\mathbf{F}_2$.
3. Calculate the residuals $\mathbf{F}_1 = \mathbf{Y} - \mathbf{T}_1\mathbf{C}_1^{\mathrm{T}}$.
4. Calculate the second PLS model with $\mathbf{X}_2$ and $\mathbf{F}_1$.
5. Calculate the residuals $\mathbf{F}_2 = \mathbf{Y} - \mathbf{T}_2\mathbf{C}_2^{\mathrm{T}}$.
6. Check for convergence; if not converged, return to step 2.

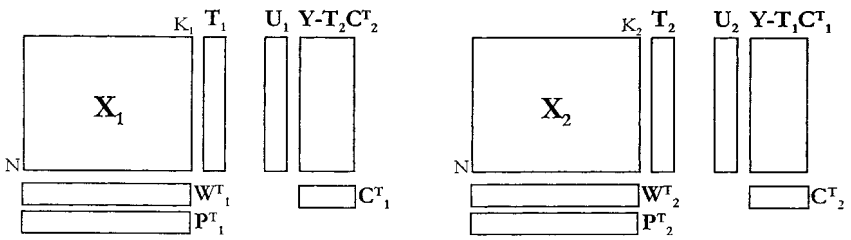

Figure 1. A schematic diagram of S-PLS. $\mathbf{X}_1$ and $\mathbf{X}_2$ are the separate predictor blocks, with $\mathbf{T}$ as the score matrix, $\mathbf{P}$ as the loading matrix and $\mathbf{W}$ as the matrix with the weight vectors. $\mathbf{C}$ is the matrix of weight vectors for $\mathbf{Y}$, and $\mathbf{U}$ is the score matrix for $\mathbf{Y}$. The first model is calculated with $\mathbf{X}_1$ and $\mathbf{Y}\text{-}\mathbf{T}_2\mathbf{C}_2^{\mathrm{T}}$ and the second model with $\mathbf{X}_2$ and $\mathbf{Y}\text{-}\mathbf{T}_1\mathbf{C}_1^{\mathrm{T}}$. Thus each separate model explains the variation in $\mathbf{Y}$ that is unexplained from the other model

As seen in the scheme above, the algorithm is simple and straightforward. The iteration may seem unnecessary but has been shown to give a better model, i.e. better prediction on the data sets on which we have tested the algorithm. With the iteration, each PLS model is only calculated on the residuals from the other PLS model. Without any iteration, only the second PLS model would be calculated on the residuals and not the first PLS model.

In the algorithm given above, the number of components in each part has to be decided before the calculations start; the residuals from the different models are dependent on how many components are used. To estimate the number of components for each block, full cross-validation is used. In full cross-validation, all components in the separate models for each cross-validation round are calculated at once and the deflation is done with the objects to be predicted removed; compared with partial cross-validation where the deflation is done with all objects present, i.e. the cross-validation is done for one component at a time. The number of components has to be estimated for both the first and the second block independently. In order to estimate the best number of components for each model, all combinations are generated and the solution with the best cross-validation value is used. For a two-block model this will result in a matrix with cross-validation values where the rows correspond to the number of components for the first model and the columns to the number of components for the second model. It is of course possible to have separate numbers of components in the different models and also to have the ordinary PLS model (zero components for the second model) as a solution.

All the figures, interpretation and statistics that are possible with the classical PLS algorithm are also valid for the S-PLS model. The only difference is that there are two separate models, one for each block. The regression coefficients for the two separate models can simply be calculated as

$$\boldsymbol{B}_1 = \boldsymbol{W}_1(\boldsymbol{P}_1^T\boldsymbol{W}_1)^{-1}\boldsymbol{C}_1^T, \qquad \boldsymbol{B}_2 = \boldsymbol{W}_2(\boldsymbol{P}_2^T\boldsymbol{W}_2)^{-1}\boldsymbol{C}_2^T \tag{6}$$

The responses for a new sample can then easily be predicted. Since the two PLS models are calculated with different responses, the residuals from the other model, the variance in $\mathbf{Y}$ for the two models will differ. This will also be reflected in the coefficients; they will usually be smaller for the second block, since the first block most likely explains most of the variance in $\mathbf{Y}$. Even so, investigating and comparing the coefficients can still give valuable information; large coefficients for the first block and small coefficients for the second block indicate that most of the variance is modelled by the first block and that the second block is not that important. The coefficient for the second block will also show if there is a single important variable for the second block.

It is important to remember that the first predictor block will explain as much as possible of the

response, leaving only a small part for the second block. In other words, the results are dependent on the order of the different predictor blocks. Because of this, the method is more suitable for cases where the order of the blocks is clear. This is also why examples with many blocks work less well with this method. With more than three or four blocks, if there are such examples, the analysis should probably be done with only two of the blocks at a time.

S-PLS for evaluating predictor blocks

In QSARs (quantitative structure–activity relationships), molecules are characterized by a set of predictor variables that are related to the properties of the molecules. With the increase in different computational techniques, the number of possible predictors has increased rapidly. Usually, a calculation generates a whole block of new predictors describing everything from atom-specific parameters to parameters related to the whole molecule or to separate fragments of the molecule. It is important to remember that the true underlying dimensionality of the problem does not increase with the number of variables. An increase in the number of variables can only help to describe the underlying function in a better way. However, it can also make the model worse if the new variables do not contribute any new information but only add noise to the model.

Hence the question is whether it is possible to investigate the effect a new predictor block has on a model.

To illustrate this, consider the following example. To a fairly good model with one predictor block and one response block, a second predictor block is added. The second predictor block may have some information that not is present in the first block.

In this example the question is whether the second block is needed or whether the first block will suffice. This question could also of course be solved by simply merging the two blocks and using classical PLS on the whole block. However, this is not totally satisfactory, since an independent interpretation of the results is difficult, as discussed below.

One possibility is that the classical PLS model with both predictor blocks is clearly better than the one with only the first block. In this case it is quite obvious that the second block is needed for the model. However, it can be difficult to determine which variables in the second block have unique information. A variable with a large weight in the second block can have this weight either owing to its correlation with some important variables in the first block or because it has unique information which is not present in the first block. Both these reasons will generate a high weight value for a variable.

Another possibility is that the classical PLS model based on both blocks gives almost the same result as the PLS model based on the first block. This may come from the fact that the two separate blocks contain the same information and nothing is gained by including the second block. Another explanation can be that the second block contributes both information and noise, but the latter cancels the positive effect of more information. Also here it is difficult to investigate the contribution of separate variables to the model.

The third and last possibility is that the classical PLS model with both blocks gives a clearly worse result than the single-block PLS model. Here the most likely explanation is that there is no information in the second block that contributes to the explanation of the variation in $\mathbf{Y}$; alternatively, it is possible that any unique information in the second block is drowned by noise.

S-PLS makes it possible to facilitate the interpretation of these different cases, since the different blocks are treated separately and only variables with unique information in the second block will show up as important.

S-PLS using the INLR approach

The INLR solution does not give information on how non-linear the latent variables are in relation to **Y**. Our first method for solving this problem was to try to rotate the solution into the 'real' latent variables and their squares and cross-products. This is necessary since the latent variables from the INLR model are a combination of the real underlying latent variables and their squares and cross-products. This can be exemplified with a simple example.

Assume that **X** contains two true latent variables, here called $\mathbf{s}_1$ and $\mathbf{s}_2$, and that **Y** is a quadratic function of these, $\mathbf{Y} = f(\mathbf{s}_1,\mathbf{s}_2)$. If linear PLS is used to model the relationship between **X** and **Y**, the first latent variable $\mathbf{t}_1$ will probably be very correlated with $\mathbf{s}_1$, but it may also be tilted by the squares $\mathbf{s}_1^2$ and $\mathbf{s}_2^2$ and the cross-term $\mathbf{s}_1\mathbf{s}_2$, The latent variable $\mathbf{t}_1$ can be said to be confounded by the squares and cross-term. This is why an expansion of the latent variables in the inner relation does not solve the problem of confounding of the latent variables. This is also true for the INLR model; the extracted latent variables are confounded by the squares and cross-terms of the latent variables. This result affects only the interpretation of the results and not the modelling power of the model. However, even though the non-linearity in the latent variables cannot be explicitly analysed, the interpretation of the separate variables and their squares is still possible.

S-PLS solves that problem, since the linear and non-linear parts are divided in the S-PLS model. The latent variables extracted from the linear model are related to the linear part, while the non-linear model gives the non-linear part. The quadratic PLS method of Höskuldsson[22] also derives the 'real' latent variables and their squares and cross-terms, but in a completely different way. In that algorithm the H-principle[32] is used for estimating the significance of linear and non-linear latent variables.

EXAMPLES

The first example, from our laboratory, 'Pulp', is a multivariate characterization of pulp using NIR. The predictor block **X** consists of 233 absorbances between 400 and 2800 nm. The single response **y** was the beating grade. There were 39 samples available and 28 of them were selected as a training set. The remaining eleven samples formed the test set. The **X** matrix was centred and scaled to unit variance before the analysis. The Pulp example was also used in Reference 10 and the result from INLR will be compared with the S-PLS results.

The second example, 'Bitter', is a comparison between different types of predictor blocks. The data set is a series of 48 bitter tasting dipeptides. Two predictor blocks describe the peptides. In the first predictor block, five *z*-scales are used to describe each amino acid.[33–35] Hence the first predictor block $\mathbf{X}_1$ will have ten variables since both of the residues in the dipeptides are varied. In the second predictor block $\mathbf{X}_2$ the amino acids are described with 22 different calculated parameters from the program Molsurf,[36] giving a total of 44 variables for the second block. The bitter threshold is used as the response. The 48 different peptides were divided into a training set with twelve peptides and a test set with 36 peptides. The training set was selected according to a D-optimal design.[37] The two predictor blocks were centred and scaled to unit variance before the analysis.

To compare the results of the different models, the explained variance in **Y**, R^2, and the cross-validated explained variance in **Y**, Q^2, are used. RMSEP (root mean square error of prediction) is calculated on the test set and gives an estimation of the predictive power of the model.

RESULTS

The Pulp example

A classical PLS model was first calculated for these data, resulting in five significant components. The statistics for this model are given in Table 1, together with the results from the other models. The plot

Table 1. Results from Pulp example

Model	A[a]	R^2	Q^2	RMSEP
PLS	5	0·93	0·86	6·23
INLR	3	0·94	0·90	3·80
W-INLR	2	0·93	0·91	5·23
S-PLS	2 ($\mathbf{X}_l$) 2 ($\mathbf{X}_n$)	0·94	0·93	3·67

[a] Number of components; for the S-PLS model the number of components in each block is given.

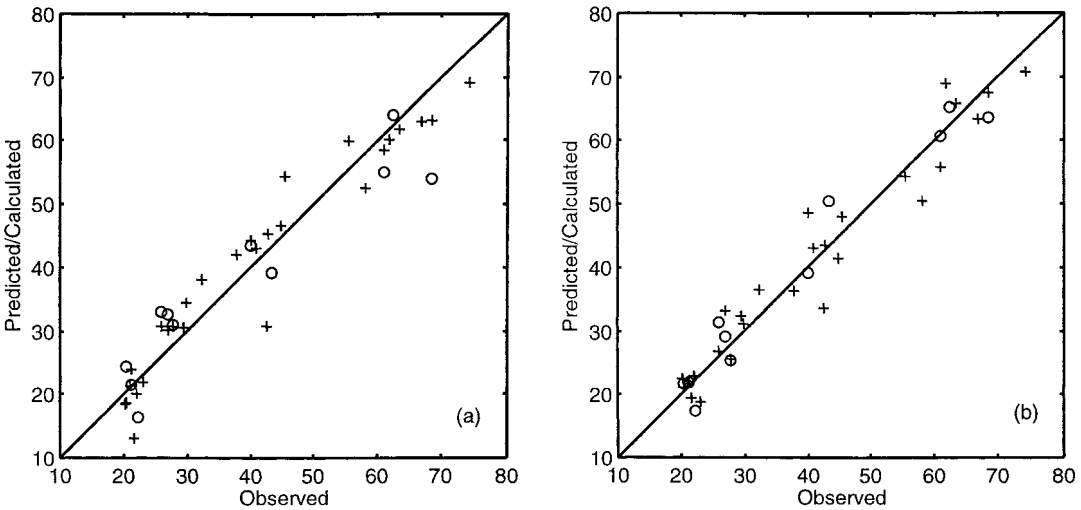

Figure 2. Observed versus predicted/calculated values for Pulp example: ○, test set; +, training set; (a) PLS model; (b) INLR model

of observed versus predicted/calculated values, Figure 2(a), shows a non-linear relationship. It is clear from this picture that there is a non-linear relationship between **X** and **Y** that classical PLS was not able to model. In Reference 10 it was shown that INLR was able to model this relationship and gave much better predictions. There was a 39% improvement in the predictions with the INLR model compared with classical PLS. Also the weighted INLR approach was tested on the Pulp example. This model suggested that the square block should be weighted down with a factor of 0·2. Unfortunately, this made the prediction worse compared with the INLR model.

The block with the original variables, the linear block, formed the first block $\mathbf{X}_l$. The squares of the variables in the first block, here called the non-linear block, formed the second block $\mathbf{X}_n$. For this model the regression equation can be written as

$$\boldsymbol{Y} = \boldsymbol{X}_l\boldsymbol{B}_l + \boldsymbol{X}_n\boldsymbol{B}_n + \boldsymbol{F}$$

The results from the S-PLS model are also presented in Table 1. This model is slightly better than the INLR model. It is also notable that the S-PLS model uses two linear components and two non-linear components, while the INLR model used three PLS components. The plot of observed versus predicted/calculated values for the S-PLS model is presented in Figure 2(b). This plot reveals a much

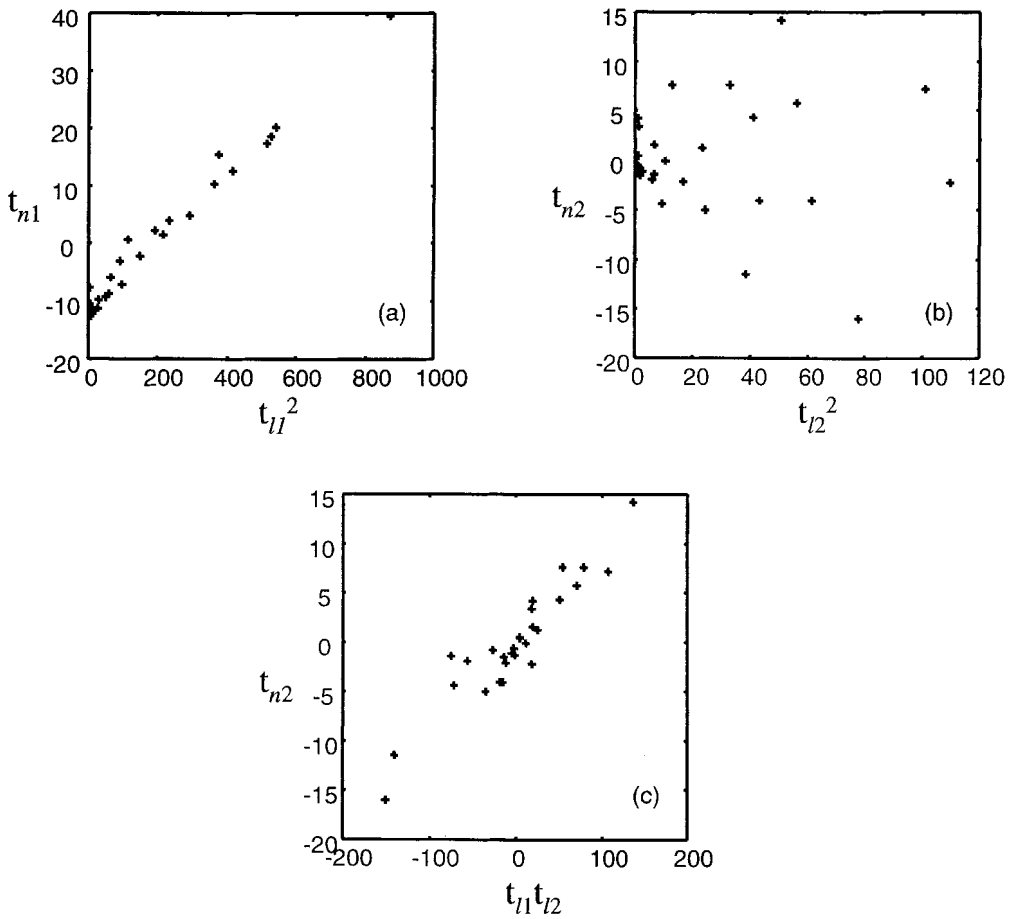

Figure 3. Scores from non-linear block plotted versus squares and cross-term of scores from linear block

more linear behaviour compared with Figure 2(a), classical PLS.

The idea behind S-PLS was that it should be possible to distinguish between the linear and the non-linear model. In the INLR model this is difficult, since the two blocks are treated as one. In Figure 3 the scores of the non-linear part are plotted versus the squares and the cross-term of the linear scores. The first score vector of the non-linear part, $\mathbf{t}_{n1}$, is strongly correlated with the square of the first linear score vector, $\mathbf{t}_{l1}^2$, as seen in Figure 3(a). A comparison of the corresponding vectors for the second component, $\mathbf{t}_{n2}$ and $\mathbf{t}_{l2}^2$, reveals that they are not correlated at all (Figure 3(b)). It is instead the score vector of the non-linear part for the second component, $\mathbf{t}_{n2}$, and the cross-product of the first- and second-component linear score vectors, $\mathbf{t}_{l1}\,\mathbf{t}_{l2}$, that are correlated (Figure 3(c)). Thus we can say that the relationship between the original $\mathbf{X}$ and $\mathbf{Y}$ is described by $\mathbf{t}_1$, $\mathbf{t}_2$, $\mathbf{t}_1^2$ and $\mathbf{t}_1\,\mathbf{t}_2$. That the type of non-linearity can be found and analysed is a clear improvement compared with the INLR method.

The Bitter example

Here the question addressed is whether the second block with Molsurf predictors, $\mathbf{X}_2$, is necessary for the modelling of $\mathbf{Y}$. The PLS models for the *z*-block $\mathbf{X}_1$ and for the amalgamated *z*- and Molsurf block

Table 2. Results from Bitter example. Predictor block used in model: Z, *z*-scales; M, Molsurf

Model	Block	A[a]	R^2	Q^2	RMSEP
PLS	Z	3	0·98	0·85	0·30
PLS	Z + M	3	0·98	0·68	0·34
S-PLS	Z + M	3 ($\mathbf{X}_1$) 1 ($\mathbf{X}_2$)	0·97	0·87	0·28

[a] Number of components; for the S-PLS model the number of components in each block is given.

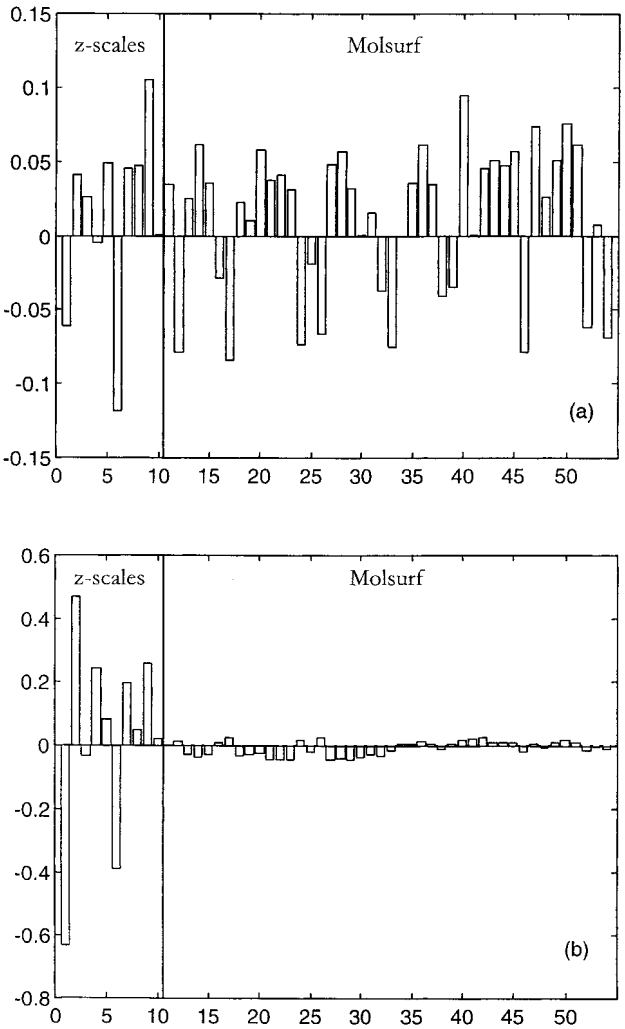

Figure 4. Coefficients for (a) PLS model and (b) S-PLS model

are given in Table 2. As can be seen, the addition of the Molsurf block did not improve the model; both Q^2 and RMSEP got worse. As seen in Figure 4(a), where the coefficients for the PLS model for both blocks are plotted, the coefficients of the Molsurf block are equal in size to the coefficients of the *z*-block. This makes it impossible to see if there are variables in the Molsurf block that might contribute new information to the model.

The statistics for the S-PLS model, where the first block is the *z*-scales and the second block is the Molsurf variables, are given in Table 2. The highest Q^2 value is obtained with three PLS components for the *z*-scales and with only one for the Molsurf block. The improvement is not large, but it is worth the effort to investigate it further. The result from the S-PLS model with three components in the *z*-scale block and one component in the Molsurf block is presented in Table 2 and the coefficients for the S-PLS model are shown in Figure 4(b). It is possible to say that this model is slightly better than the PLS model with only the *z*-scales, since it has slightly better Q^2 and RMSEP. This would indicate that there is some new information in the Molsurf block that is not captured in the *z*-scales. However, the differences are small and this would suggest that for this example the *z*-scales are sufficient for the modelling of the response. That most of the response is modelled by the *z*-scales can also be seen in the coefficient plot, Figure 4(b). The figure clearly shows that the *z*-scale coefficients are larger than the coefficients of the Molsurf block. This is also in agreement with the fact that the Molsurf block only slightly improved the model. It is important to remember that this result does not mean that the Molsurf parameters do not have any predictive power; the results only indicate that the information in the *z*-scales, for the example above, is sufficient for modelling the response.

CONCLUSION

In this paper we have presented a new multiblock PLS algorithm, S-PLS. In S-PLS the different predictor blocks are treated in a serial mode, in contrast with other multiblock PLS methods where the blocks are treated in a parallel mode. This enables a separate analysis of the contribution of each block to the model. The classical PLS algorithm is used in each block model, where by the S-PLS model retains all the features that an ordinary PLS model has.

With S-PLS it is possible to see if a new set of variables, a new predictor block, contributes any new or more accurate information to a model where there are already one or several predictor blocks. This is demonstrated on a QSAR example where the molecules are described with two different predictor blocks, *z*-scales and Molsurf. The result indicates that the *z*-scales have enough information for modelling the response. Only small improvements were achieved when both blocks were used. This example demonstrated also the difference between S-PLS with clearly separated blocks and PLS with the two blocks amalgamated, as seen in Figures 4(a) and 4(b). The sizes of the coefficients are in better agreement in the S-PLS model, since the sizes of the Molsurf coefficients are much smaller for the S-PLS model than for the PLS model on the amalgamated block. With the PLS model a separate analysis of the second block was not possible, since the weights for the variables are dependent on the first block.

S-PLS was also demonstrated on an INLR model, where a clear division can be made between the linear and non-linear predictor blocks and a separate interpretation of these is warranted. With the S-PLS model it was possible to separate the linear and non-linear parts in a simple way. For the example shown, the response was related to the latent variables $\mathbf{t}_1$ and $\mathbf{t}_2$, the square of the first of these, $\mathbf{t}_1^2$, and the cross-term $\mathbf{t}_1\mathbf{t}_2$, This separation of the results is not possible if classical PLS is used. In the classical PLS model the resulting score vectors will be a combination of the linear and non-linear latent variables. S-PLS also improved the predictions for this example.

S-PLS is a promising new technique for applications where more than one predictor block is available.

ACKNOWLEDGEMENTS

Financial support from the Swedish Science Research Council (NFR) is gratefully acknowledged.

REFERENCES

1. J. E. Jackson, *A Users Guide to Principal Components*, Wiley New York (1991).
2. S. Wold, A. Ruhe, H. Wold and W. J. Dunn III, *SIAM J. Sci. Statist. Comput.* **5**, 735 (1984).
3. J. A. Westerhuis and P. M. J. Coenegracht, *J. Chemometrics*, **11**, 379 (1997).
4. S. Wold, N. Kettaneh and K. Tjessem, *J. Chemometrics*, **10**, 463 (1996).
5. T. Lundstedt and B. Thelin, *Chemometrics Intell. Lab. Syst.* **29**, 255 (1995).
6. J. F. MacGregor, C. Jaeckle, C. Kiparissides and M. Koutoudi, *AIChE J.* **40**, 826 (1994).
7. A. Berglund, M. C. De Rosa and S. Wold, *J. Comput. Aided Mol. Des.* **11**, 601 (1997).
8. H. Wold, 'Soft modeling. The basic design and some extensions', in *Systems under Indirect Observation. 2*, ed. by K.-G Jöreskog and H. Wold, pp. 1–53, North-Holland, Amsterdam (1982).
9. S. Wold, S. Hellberg, T. Lundstedt, M. Sjöström and H. Wold, PLS model building: theory and application. PLS modeling with latent variables in two or more dimensions, presented at Symposium, Frankfurt am Main, September 1987.
10. A. Berglund and S. Wold, *J. Chemometrics*, **11**, 141 (1997).
11. A. Berglund and S. Wold, 'INLR (implicit non-linear latent variable regression). II. Blockscaling of expanded terms with QSAR examples', in *Computer-assisted Lead Finding and Optimisation, Current tools for Medicinal Chemistry*, ed. H. van de Waterbeemd, B. Testa and G. Folkers, pp. 65–79, Wiley–VCH, Basel (1997).
12. A. Phatak and S. De Jong, *J. Chemometrics*, **11**, 311 (1997).
13. A. J. Burnham, R. Viveros and J. F. MacGregor, *J. Chemometrics*, **10**, 31 (1996).
14. M. Stone and R. J. Brooks, *J. R. Statist. Soc. B*, **52**, 237 (1990).
15. A. Höskuldsson, *J. Chemometrics*, **2**, 211 (1988).
16. P. H. Garthwaite, *J. Am. Statist. Assoc.* **89**, 122 (1994).
17. R. Manne, *Chemometrics Intell. Lab. Syst.* **2**, 187 (1987).
18. I. Helland, *Commun. Statist. Simul. Comput.* **17**, 581 (1988).
19. S. Wold, N. Kettaneh-Wold and B. Skagerberg, *Chemometrics Intell. Lab. Syst.* **7**, 53 (1989).
20. S. Wold, *Chemometrics Intell. Lab. Syst.* **14**, 71 (1992).
21. I. Frank, *Chemometrics Intell. Lab. Syst.* **8**, 109 (1990).
22. A. Höskuldsson, *J. Chemometrics*, **6**, 307 (1992).
23. V.-M. Taavitsainen and P. Korhonen, *Chemometrics Intell. Lab. Syst.* **14**, 185 (1992).
24. V.-M. Taavitsainen and H. Haario, *Chemometrics Intell. Lab. Syst.* **23**, 51 (1994).
25. T. Næs and T. Isaksson, *Appl. Spectrosc.* **46**, 34 (1992).
26. T. J. McAvoy and S. J. Qin, *Comput. Chem. Engng.* **16**, 379 (1992).
27. T. R. Holcomb and M. Morari, *Comput. Chem. Engng.* **16**, 393 (1992).
28. H. Yoshida and K. Funatsu, *J. Chem. Info. Comput. Sci.* **37**, 1115 (1997).
29. G. Andersson, P. Kaufmann and L. Renberg, *J. Chemometrics*, **10**, 605 (1996).
30. M. Stone, *J. R. Statist. Soc. B*, **36**, 55 (1974).
31. S. Wold, *Technometrics*, **20**, 397 (1978).
32. A. Höskuldsson, *Chemometrics Intell. Lab. Syst.* **14**, 139 (1992).
33. M. Sjöström and S. Wold, *J. Mol. Evol.* **22**, 272 (1985).
34. J. Jonsson, L. Eriksson, S. Hellberg, M. Sjöström and S. Wold, *Quant. Struct.–Act. Relat.* **8**, 204 (1989).
35. S. Hellberg, M. Sjöström, B. Skagerberg and S. Wold, *J. Med. Chem.* **30**, 1126 (1987).
36. P. Sjöberg, *Molsurf Version 2·1*, Qemist AB, Karlskoga (1997).
37. T. J. Mitchell, *Technometrics*, **16**, 203 (1974).

