JOURNAL OF CHEMOMETRICS
J. Chemometrics 2002; **16**: 119–128

Orthogonal projections to latent structures (O-PLS)

Johan Trygg* and Svante Wold
Research Group for Chemometrics, Institute of Chemistry, Umeå University, SE-901 87 Umeå, Sweden

Received 20 June 2000; Revised 15 April 2001; Accepted 20 April 2001

A generic preprocessing method for multivariate data, called orthogonal projections to latent structures (O-PLS), is described. O-PLS removes variation from X (descriptor variables) that is not correlated to Y (property variables, e.g. yield, cost or toxicity). In mathematical terms this is equivalent to removing systematic variation in X that is orthogonal to Y. In an earlier paper, Wold *et al.* (*Chemometrics Intell. Lab. Syst.* 1998; 44: 175–185) described orthogonal signal correction (OSC). In this paper a method with the same objective but with different means is described. The proposed O-PLS method analyzes the variation explained in each PLS component. The non-correlated systematic variation in X is removed, making interpretation of the resulting PLS model easier and with the additional benefit that the non-correlated variation itself can be analyzed further. As an example, near-infrared (NIR) reflectance spectra of wood chips were analyzed. Applying O-PLS resulted in reduced model complexity with preserved prediction ability, effective removal of non-correlated variation in X and, not least, improved interpretational ability of both correlated and non-correlated variation in the NIR spectra. Copyright © 2002 John Wiley & Sons, Ltd.

KEYWORDS: orthogonal projections to latent structures (O-PLS); orthogonal signal correction (OSC); NIPALS PLS; multivariate data analysis; calibration; preprocessing

1. INTRODUCTION

Multiple-measurement vectors and arrays are increasingly being used for the characterization of solid, semisolid, fluid and vapor samples. Examples of methods giving such multiple measurements are near-infrared (NIR) spectroscopy and nuclear magnetic resonance (NMR) spectroscopy. Frequently the objective with this characterization is to determine the value of one or several concentrations of analytes in the samples. Multivariate calibration is then used to develop a quantitative relation between the digitized spectra, the matrix **X**, and the concentrations, the matrix **Y**, as reviewed in Reference [1]. NIR and other spectroscopy methods are also increasingly used to infer properties of samples other than concentrations, e.g. the strength and viscosity of polymers, the thickness of a tablet coating or the octane number of gasoline.

The first step of a multivariate calibration is often to preprocess input data. The reason is that spectra, as well as other multiple-measurement arrays, often contain systematic variation that is unrelated to the responses **Y**. For solid samples this systematic variation is due to, among others, light scattering and differences in spectroscopic path length and may often constitute the major part of the variation of the sample spectra. Another reason for systematic but non-correlated variation in the sample spectra may be that the analyte of interest absorbs only in small parts of the spectral region. The variation in **X** that is unrelated to **Y** may disturb the multivariate modeling, cause imprecise predictions for new samples and also affect the robustness of the model over time.

For the removal of undesirable systematic variation in the data, two types of preprocessing methods are commonly reported in the analytical chemistry literature, namely differentiation and signal correction. Popular approaches to signal correction include Savitzky–Golay smoothing [2], multiplicative signal correction (MSC) [3], Fourier transformation [4], principal component analysis (PCA) [5], variable selection [6] and baseline correction [7]. These signal corrections are different cases of filtering, where a signal (e.g. an NIR spectrum) is made to have 'better properties' by passing it through a filter. The objectives of filtering are often rather vague; it is not always easy to specify what is meant by 'better properties'. Even in the case of calibration, where it is possible to specify this objective in terms of lowered prediction errors or simpler calibration models, it has been difficult to construct general filters that indeed improve these properties of the data.

Projections to latent structures by means of partial least squares (PLS) is one of the main methods for analyzing multivariate data where a quantitative relationship between a descriptor matrix **X** and a response matrix **Y** is sought. Multivariate calibration, classification, discriminant analysis

* Correspondence to: J. Trygg, Research Group for Chemometrics, Institute of Chemistry, Umeå University, SE-901 87 Umeå, Sweden
E-mail: johan.trygg@chem.umu.se
Contract/grant sponsor: National Graduate School in Scientific Computing (NGSSC).
Contract/grant sponsor: Centre for Forest Biotechnology and Chemistry.
Contract/grant sponsor: Swedish Natural Science Research Council (NFR).

DOI:10.1002/cem.695

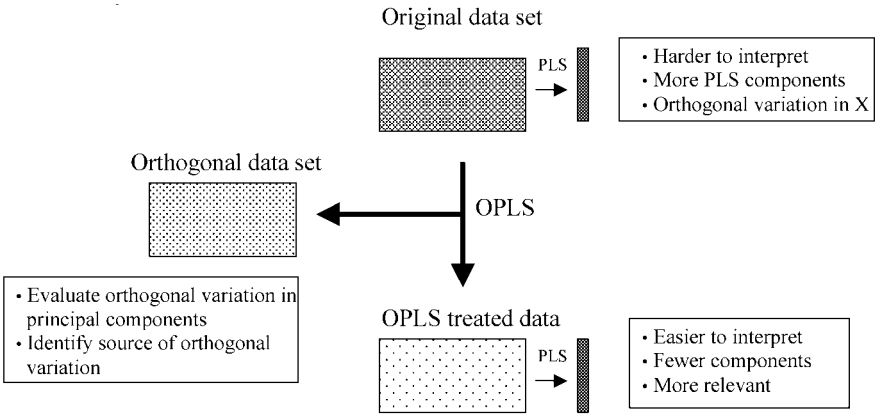

Figure 1. Overview of orthogonal projections to latent structures (O-PLS).

and pattern recognition are a few areas where PLS has shown to be a useful tool [1]. PLS can cope with collinearity among variables, noise in both **X** and **Y** and moderate amounts of missing data in both **X** and **Y**; these are the main reasons for its success. These types of complicated data are common owing to the frequent use of analytical instruments such as HPLC, LC/UV, LC/MS and spectroscopy instruments.

Improved and modified PLS methods, using the nonlinear iterative partial least squares (NIPALS) method [8], have been proposed since the birth of PLS in 1977. Recently, the orthogonal signal correction (OSC) method was proposed by Wold *et al.* [9]. The idea was to remove systematic information in **X** not correlated to the modeling of **Y** in order to achieve better models in multivariate calibration.

In this paper an additional modification of the NIPALS PLS algorithm is described, called orthogonal projections to latent structures (O-PLS). Its objective is to improve interpretation of PLS models and reduce model complexity. O-PLS provides a way to remove systematic variation from an input data set **X** not correlated to the response set **Y**; in other words, to remove variability in **X** that is orthogonal to **Y**. The proposed O-PLS method analyzes the disturbing variation in each regular PLS component. The non-correlated variation in **X** is separated from the correlated variation, with the additional benefit that the non-correlated variation itself can be studied and analyzed. Removing non-correlated variation in data prior to data modeling is not only interesting from a predictive point of view, but the interpretational ability of resulting models also improves. Thus more information and knowledge of a system can be retrieved and analyzed, and developed further.

As an example, we use NIR reflectance spectra of wood chips, measured at ASSI Domän, Piteå, Sweden, to determine the dry content of the chips.

2. ORTHOGONAL CORRECTION METHODS

The idea with the orthogonal correction methods is to remove variation in **X** that is not correlated to **Y**. There are three criteria put on the orthogonal correction methods.

- Component should involve the large systematic variation in **X**.
- Component must be predictive by **X** (in order to apply on future data).
- Component must be orthogonal to **Y**.

The first two criteria are easily met by performing a PCA of **X**. However, the third and most important criterion is not as easily met.

2.1. Orthogonal signal correction (OSC)

The OSC method introduced by Wold *et al.* [9] contains an internal time-consuming iteration to find orthogonal components. The method often converges quickly, but still it needs 5–20 iterations. The OSC method does not give a unique solution but depends on the starting vector **t**. Therefore PCA is a good choice to produce the starting vector, because it gives the longest score vector **t** that can be predicted by **X**. Two of the criteria above are then automatically met. During the OSC iterations the norm of the OSC component will decrease some in order to converge.

The OSC method generally works very well. The main problem with the method concerns the overfitting of the estimated OSC components. Cross-validation or any other validation method is not used owing to implementational problems and the additional time needed for this. The correct number of 'internal PLS components' (step 4 in Appendix A of Reference [9]) for estimation of the orthogonal components is therefore difficult to estimate, leading to a risk of overfitting and sometimes even degradation of resulting calibration models.

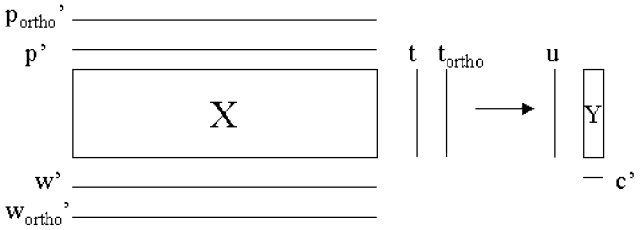

Figure 2. More schematic overview of the steps involved in O-PLS.

2.2. Orthogonal filtering method (T. Fearn)

Fearn [10] recently reported an alternative way of estimating orthogonal components than that proposed by Wold *et al.* Fearn solved the eigenvalue problem of the matrix $\mathbf{MX}^T\mathbf{X}$, where $\mathbf{M} = \mathbf{I} - \mathbf{X}^T\mathbf{Y}(\mathbf{Y}^T\mathbf{X}\mathbf{X}^T\mathbf{Y})^{-1}\mathbf{Y}^T\mathbf{X}$. This was accomplished by doing an eigenvalue decomposition and using the eigenvectors as weight vectors $\mathbf{w}$ to produce the orthogonal scores $\mathbf{t} = \mathbf{Xw}$, and thereafter finding the loading vector $\mathbf{p} = \mathbf{t}^T\mathbf{X}/(\mathbf{t}^T\mathbf{t})$ and subsequently updating the $\mathbf{X}$ matrix by removing the found orthogonal variation, $\mathbf{E} = \mathbf{X} - \mathbf{tp}^T$. For additional orthogonal components, $\mathbf{E}$ is set to $\mathbf{X}$ and the procedure is repeated. The approach suggested by Fearn works fairly well according to our limited experience. Some problems encountered have been increased prediction errors in some cases (as in an example given later in this paper) and the inability to get a resulting one-component PLS model with preserved prediction errors. A resulting PLS model with more than one component implies that latent orthogonal variation is still present in the $\mathbf{X}$ data, assuming a linear model. Also, there is an increase in calculations compared to O-PLS (below).

2.3. Orthogonal projections to latent structures (O-PLS)

The O-PLS method proposed here is a modification of the original NIPALS PLS algorithm. The development of O-PLS has, like OSC, been driven by the large amount of non-correlated variation present in data sets today, especially in a multivariate calibration situation. The O-PLS method can be seen as either a pure preprocessing method to remove systematic orthogonal variation from a given data set $\mathbf{X}$, or it can be made an integrated part of the regular PLS modeling to provide simpler models and with the additional advantage that the orthogonal variation can be analyzed separately (Figure 1). The simplicity of O-PLS and the fact that only a small modification of the regular NIPALS PLS method needs to be done should make O-PLS a generally applicable preprocessing and filtering method.

An outline of the proposed O-PLS method is shown here for a single $\mathbf{y}$. For a detailed description of the O-PLS method with a $\mathbf{Y}$ matrix, see Appendix I (Figure 2).

- Optionally transform, center and scale the raw data to give the matrices $\mathbf{X}$ and $\mathbf{y}$.

1. $\mathbf{w}^T = \mathbf{y}^T\mathbf{X} \;/\; (\mathbf{y}^T\mathbf{y})$. Steps 1–6 are the normal NIPALS PLS method for single $\mathbf{y}$.
2. $\mathbf{w} = \mathbf{w} \;/\; \|\mathbf{w}\|$.
3. $\mathbf{t} = \mathbf{Xw} \;/\; (\mathbf{w}^T\mathbf{w})$.
4. $c^T = \mathbf{t}^T\mathbf{y} \;/\; (\mathbf{t}^T\mathbf{t})$.
5. $\mathbf{u} = \mathbf{y}c \;/\; (c^Tc)$.
6. $\mathbf{p}^T = \mathbf{t}^T\mathbf{X} \;/\; (\mathbf{t}^T\mathbf{t})$.
7. $\mathbf{w}_{ortho} = \mathbf{p} - [\mathbf{w}^T\mathbf{p} \;/\; (\mathbf{w}^T\mathbf{w})]\mathbf{w}$. The $\mathbf{p}$ vector in step 7 can also be an arbitrary vector of the same length as $\mathbf{w}$ (e.g. PCA loading of $\mathbf{X}$).
8. $\mathbf{w}_{ortho} = \mathbf{w}_{ortho} \;/\; \|\mathbf{w}_{ortho}\|$.
9. $\mathbf{t}_{ortho} = \mathbf{Xw}_{ortho}/(\mathbf{w}_{ortho}^T\mathbf{w}_{ortho})$.
10. $\mathbf{p}_{ortho}^T = \mathbf{t}_{ortho}^T\mathbf{X}/(\mathbf{t}_{ortho}^T\mathbf{t}_{ortho})$.
11. $\mathbf{E}_{O-PLS} = \mathbf{X} - \mathbf{t}_{ortho}\mathbf{p}_{ortho}^T$. $\mathbf{E}_{O\text{-}PLS}$ are the filtered data.
12. Save found parameters $\mathbf{T}_{ortho} = [\mathbf{T}_{ortho} \;\; \mathbf{t}_{ortho}]$, $\mathbf{P}_{ortho} = [\mathbf{P}_{ortho} \;\; \mathbf{p}_{ortho}]$, $\mathbf{W}_{ortho} = [\mathbf{W}_{ortho} \;\; \mathbf{w}_{ortho}]$. For additional orthogonal components, return to step 3 and set $\mathbf{X} = \mathbf{E}_{O\text{-}PLS}$, otherwise continue to step 13.
13. $\mathbf{X}_{ortho} = \mathbf{T}_{ortho}\mathbf{P}_{ortho}^T$. Analyze orthogonal variation component-wise, or run PCA on $\mathbf{X}_{ortho}$ (step 14).
14. $\mathbf{X}_{ortho} = \mathbf{T}_{pca\_ortho}\mathbf{P}_{pca\_ortho}^T + \mathbf{E}_{pca\_ortho}$. PCA of $\mathbf{X}_{ortho}$ to summarize the systematic orthogonal variation. Removing all estimated orthogonal variation from $\mathbf{X}$ is one option; another option is to remove only the principal orthogonal components estimated in step 14. This corresponds to adding $\mathbf{E}_{pca\_ortho}$ back into $\mathbf{E}_{O\text{-}PLS}$.
15. New or future samples (the prediction set) are corrected using $\mathbf{W}_{ortho}$ and $\mathbf{P}_{ortho}$ of the calibration model. For each new observation vector $\mathbf{x}_{new}^T$, repeat steps 16–18 for each orthogonal component estimated in the calibration model.
16. $t_{new\_ortho} = \mathbf{x}_{new}^T\mathbf{w}_{ortho}/(\mathbf{w}_{ortho}^T\mathbf{w}_{ortho})$.
17. $\mathbf{t}_{new\_ortho}^T = [\mathbf{t}_{new\_ortho}^T \;\; t_{new\_ortho}]$. Save orthogonal scores for prediction set. The first $\mathbf{t}$ in the brackets is a vector while the second t is a scalar.
18. $\mathbf{e}_{new\_O-PLS}^T = \mathbf{x}_{new}^T - t_{new\_ortho}\mathbf{p}_{ortho}^T$. Orthogonal component in $\mathbf{x}_{new}^T$is removed. Set $\mathbf{x}_{new}^T = \mathbf{e}_{new\_O-PLS}^T$ for additional components and return to step 16, otherwise proceed to step 19.
19. $\mathbf{x}_{new\_ortho}^T = \mathbf{t}_{new\_ortho}^T\mathbf{P}_{ortho}^T$.
20. $\mathbf{t}_{new\_pca\_ortho}^T = \mathbf{x}_{new\_ortho}^T\mathbf{P}_{pca\_ortho}$. If only the orthogonal latent components from PCA on $\mathbf{X}_{ortho}$ were removed, estimate new scores from PCA loadings in step 14, $\mathbf{e}_{new\_pca\_ortho}^T = \mathbf{x}_{new\_ortho}^T - \mathbf{t}_{new\_pca\_ortho}^T\mathbf{P}_{pca\_ortho}^T$, and then add back $\mathbf{e}_{new_pca_ortho}^T$to $\mathbf{e}_{new\_O-PLS}^T$.

2.3.1. *Proof of the orthogonality between* $\mathbf{t}_{ortho}$ *and* $\mathbf{y}$

The orthogonality property $\mathbf{t}_{ortho}^T\mathbf{y} = 0$for each orthogonal component can be shown with the following proof.

$$\mathbf{t}_{ortho}^T\mathbf{y} = \mathbf{w}_{ortho}^T\mathbf{X}^T\mathbf{y} = \{\mathbf{p}^T - [\mathbf{p}^T\mathbf{w}/(\mathbf{w}^T\mathbf{w})]\mathbf{w}^T\}\mathbf{X}^T\mathbf{y}$$

Substituting and simplifying with

$$(\mathbf{w}^T\mathbf{w}) = 1, \qquad \mathbf{X}^T\mathbf{y} = \mathbf{w}\|\mathbf{X}^T\mathbf{y}\|$$

gives

$$\mathbf{t}_{ortho}^T\mathbf{y} = [\mathbf{p}^T - (\mathbf{p}^T\mathbf{w})\mathbf{w}^T]\mathbf{w}\|\mathbf{X}^T\mathbf{y}\|$$

$$= \mathbf{p}^T\mathbf{w}\|\mathbf{X}^T\mathbf{y}\| - (\mathbf{p}^T\mathbf{w})\mathbf{w}^T\mathbf{w}\|\mathbf{X}^T\mathbf{y}\|$$

Simplifying with

$$(\mathbf{w}^T\mathbf{w}) = 1$$

gives

$$\mathbf{t}_{ortho}^T\mathbf{y} = (\mathbf{p}^T\mathbf{w})\|\mathbf{X}^T\mathbf{y}\| - (\mathbf{p}^T\mathbf{w})\|\mathbf{X}^T\mathbf{y}\| = 0$$

2.3.2. *O-PLS properties*

In the area of semiempirical modeling, the obvious advantages with O-PLS are more parsimonious PLS models and easier interpretation, because the non-correlated variation and the correlated variation have been separated. O-PLS should give an improved detection limit for outliers in the scores, because the non-correlated variation in $\mathbf{X}$ could have

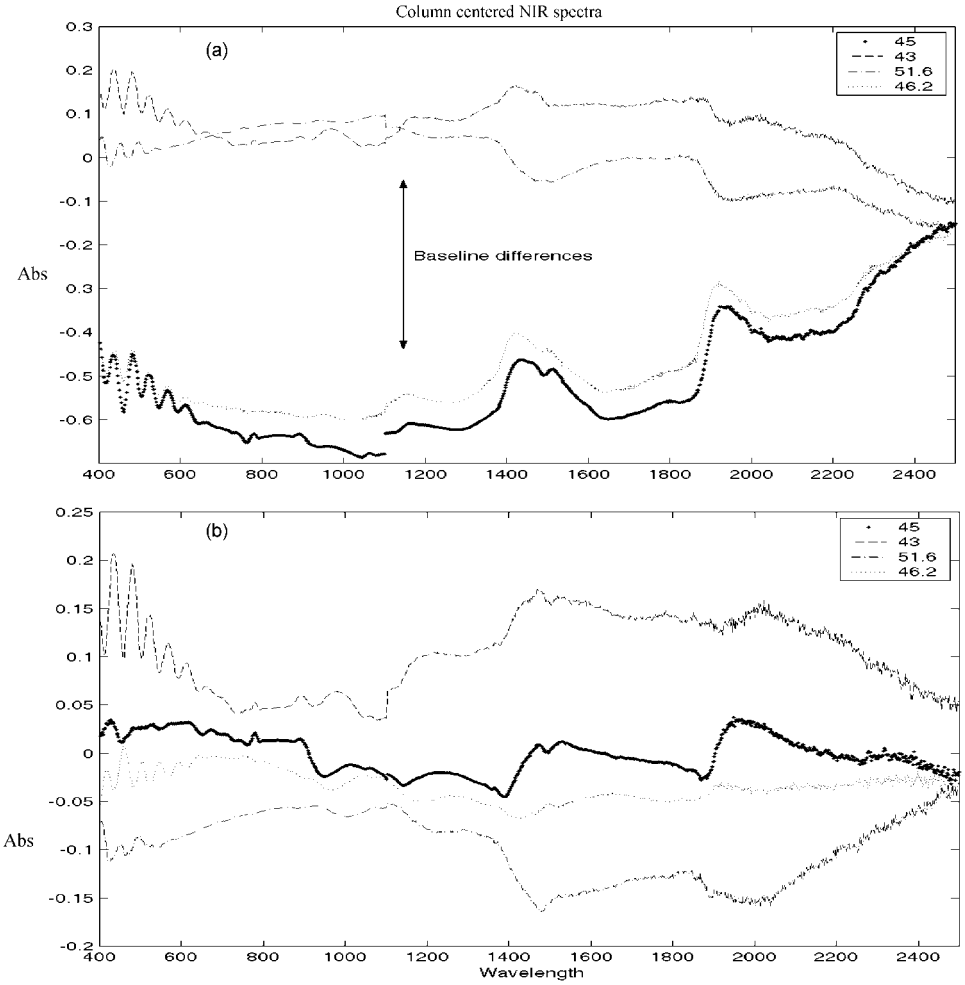

Figure 3. Example of the effect of O-PLS: (a) column-centered untreated NIR spectra; (b) O-PLS-pretreated NIR spectra. The baseline and slope problems have been reduced with the O-PLS method.

different statistical distributions than the correlated variation, producing a disturbance in the calculation of e.g. the Hotelling T^2 statistic. Another advantage with O-PLS compared to the earlier proposed OSC method is that no time-consuming internal iteration is present, making it very fast to calculate.

Analyzing the non-correlated variation is most informative; the source of disturbing variation can perhaps be identified and removed, or at least understood. PCA can be used to decompose the orthogonal matrix $\mathbf{X}_{\text{ortho}}$ into principal components (step 14 in the O-PLS method). The number of principal components can be chosen according to some significance criterion, e.g. cross-validation or large eigenvalues. It is important to realize that the scores and loading plots can be interpreted without regarding their relation to **Y**. The information from such analysis is very revealing, not least for industrial process data which contain large unknown variations due to fluctuating process environments that are hard to remove, but awareness of what they are could be vital for further process improvements. Also, only removing the orthogonal principal components from the original data has shown to decrease the total number of components used drastically.

2.3.3. *Projected orthogonal signal correction (POSC)*

There also exists another method closely related to O-PLS for estimating and removing orthogonal variation from **X**. This method is herein named projected orthogonal signal correction. Initially it requires a fully estimated PLS model, and it cannot extract an orthogonal component for each PLS component as O-PLS can. However, POSC can be seen as a special case of O-PLS. Both methods yield the same results if one adds the residual matrix of the O-PLS-treated PLS model (which can be regarded as 'systematically orthogonal'; that is, not necessarily algebraically orthogonal, but, considering noise level and so forth, it is still irrelevant for **Y**) into the orthogonal variation matrix ($\mathbf{X}_{\text{ortho}}$). The POSC algorithm is outlined in Appendix II.

2.3.4. *Number of orthogonal components*

The risk of overfitting is greatly reduced with O-PLS in comparison with OSC, because appropriate cross-validation and/or some eigenvalue criterion can be used, resulting in systematic and correlated components being calculated and extracted. Here cross-validation [11] is used to estimate the number of orthogonal components. The cross-validation performed is full cross-validation; the partial cross-valida-

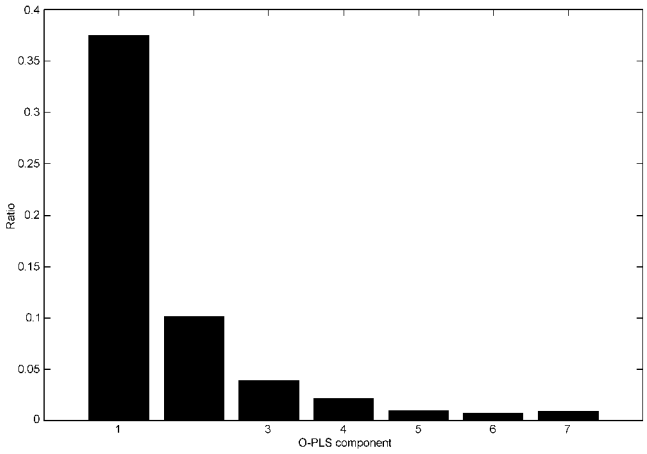

Figure 4. Plot of the ratio $\|\mathbf{w}_{ortho}\|/\|\mathbf{p}\|$ for each O-PLS component. It is a good indicator of the number of O-PLS components to remove from **X**. According to cross-validation, four O-PLS components were removed from **X**, which this plot also indicates.

tion method cannot be used because the **c** weights for all O-PLS components are always zero. By full cross-validation we mean the approach where *G* matrices are set up before the analysis and separately analyzed until the last component. In partial cross-validation a single matrix is restored between the components, and the residuals after each component are calculated just for this single matrix, using the parameters based on the previous single matrix.

The eigenvalue approach is to analyze $\|\mathbf{p} - [\mathbf{w}^T\mathbf{p}/(\mathbf{w}^T\mathbf{w})]\mathbf{w}\| \;/\; \|\mathbf{p}\|$, which becomes zero for correlated PLS components if no orthogonal variation is present in **X**. (Note that the norm of **p** will decrease for subsequent PLS components and eventually be regarded as noise.) A plot of $\|\mathbf{p} - [\mathbf{w}^T\mathbf{p}/(\mathbf{w}^T\mathbf{w})]\mathbf{w}\| \;/\; \|\mathbf{p}\|$ versus the number of orthogonal components gives a good indication of the number of orthogonal components to extract. In Figure 4 in Section 4 such a plot can be found. (The plot is interpreted as the regular scree plot [12] in principal component analysis.) A combination of the cross-validation method and the suggested eigenvalue approach is recommended.

2.3.5. *Outlier detection*

All projection methods working after some least squares methodology are sensitive to abnormal occurrences in data. PLS and PCA are no different. Detection and investigation of abnormal data or outliers represent an important part in multivariate data analysis and semiempirical modeling. In PLS and PCA the abnormal samples of data can be detected and analyzed by looking at scores and residuals. Outlier detection in O-PLS presents no additional problem, because the same principles apply.

2.3.6. *Rescaling*

Improvements in predictions in the resulting O-PLS-pretreated PLS model compared with the original PLS model can occur if the O-PLS-pretreated data matrix $\mathbf{E}_{O\text{-}PLS}$ is rescaled prior to PLS modeling. This was done in the NIR example, where PCA was used on $\mathbf{X}_{ortho}$ to find the orthogonal latent components. Scaling methods such as unit variance (UV) scaling, where each column is divided by the standard deviation of that column, or pareto scaling, where the weight factor is the square root of the standard deviation for that column, are recommended.

2.3.7. *The O-PLS method is versatile and can be designed to remove specific variation*

The suggested O-PLS method is generic. If properly used, it will improve data modeling and interpretation regardless of most types of data properties. Suppose that a given data set **X** does not contain any non-correlated systematic variation. In that case, O-PLS will not find any orthogonal components, and the resulting PLS model becomes a regular one-component PLS solution (if only single **y**). This is the case for data from designed experiments, where columns are orthogonal with respect to each other, and no orthogonal latent variables are present. Also consider the opposite case where the data set **X** only consists of non-correlated systematic variation. O-PLS then finds orthogonal components, but no PLS component can be extracted, and therefore it converges to a PCA solution.

Instead of removing systematic non-correlated variation in **X**, O-PLS can be used to remove specific types of variation in the **X** data. One option is to set the unwanted property to **Y**; another option is to set an unwanted profile to **w** and start the O-PLS algorithm with a given **w** at step 2. This could provide an opportunity to analyze the specific systematic variation in **X** originating from the unwanted property, and further analyze the O-PLS-pretreated data $\mathbf{X}_{ortho}$ with reduced influence of the disturbing variation.

2.3.8. *Noise in* **Y** *and missing data*

The O-PLS method uses **X** and **Y** to filter and remove variation in **X** not correlated to **Y**. If the given **Y** data include a great deal of noise, then there has been some concern that the orthogonal filtering methods might not perform as well as they should, although the information removed from **X** is

Table I. PLS model summary from NIR example

Original PLS model				O-PLS-preprocessed PLS model			
PLS comp.	R2Xcum	R2Ycum	Q2cum	PLS comp.	R2Xcum	R2Ycum	Q2cum
1	0.948	0.107	0.09	1	0.11	0.80	0.74
2	0.987	0.499	0.48				
3	0.995	0.629	0.58				
4	0.996	0.757	0.70				
5	0.997	0.800	0.74				

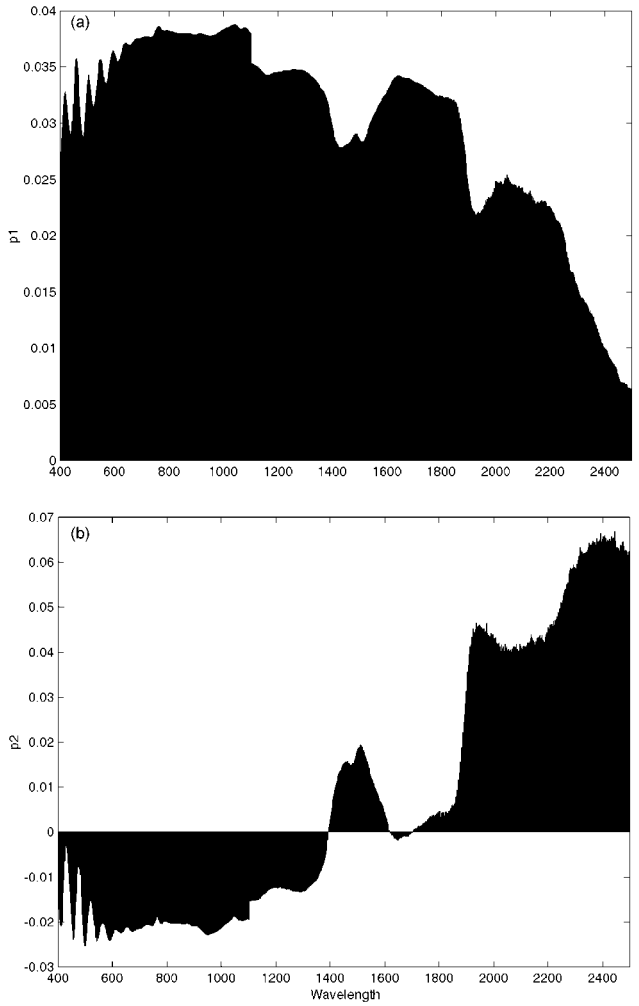

Figure 5. Plot of the (a) first and (b) second loading $\mathbf{P}_{\text{pca-ortho}}$ of a principal component analysis of the orthogonal variation in **X**. The non-correlated variation in **X** corresponds to baseline and slope variations.

indeed orthogonal to **Y**. Results from initial studies with O-PLS do not show any degradation of results compared to non-treated data. The NIPALS method can handle moderate amounts of missing data (up to around 20%) [13]. Since the O-PLS method is based on the NIPALS method for computation, it can also cope with moderate amounts of missing data.

2.3.9. *Improved interpretability with O-PLS*

Perhaps the greatest advantage with O-PLS is the simplified interpretation of the data. Imagine a 12-component PLS model raising the questions: What are the interesting variables for prediction of the response variable **y**? Which plots and parameters are interesting? Some would analyze the regression coefficient vector, because it is used to predict **y** from **X**. Others would take a combination of all loadings, scores and the coefficient vector, together with a look at the residuals. That sounds rather tough, but used in conjunction with prior knowledge of the data, this approach often works. The O-PLS method makes the interpretation easier. First, it separates the correlated variation from the non-correlated orthogonal variation. Second, it gives an opportunity to analyze the non-correlated variation in the data and understand what the sources are for that. Third, the number of PLS components in the O-PLS-pretreated PLS model is reduced to a single component (if single **y**). Interpreting and analyzing that model clearly becomes much easier.

The first loading **w** in the original PLS model is identical to the first loading **w** in the O-PLS-pretreated PLS model. This will always be the case and is easily understood when realizing that **w** is the projection of the matrix **X** onto the vector **u,** **y** if only one column in **Y**, using the NIPALS algorithm. Removing orthogonal components from **X** does not disturb the correlation between **X** and **y**, because orthogonal columns in **X** do not influence the projection $\mathbf{w}^{\mathrm{T}} = \mathbf{u}^{\mathrm{T}}\mathbf{X}/(\mathbf{u}^{\mathrm{T}}\mathbf{u})$. This raises an interesting point about the interpretation of PLS models.

2.3.10. *Simulated example*

Consider the following simulated PLS example with two data matrices $\mathbf{X}_0$ (with no orthogonal variation) and $\mathbf{X}_1$ (where orthogonal variation has been added to $\mathbf{X}_0$) and **y**:

$$\mathbf{X}_0 = \begin{bmatrix} -1 & -1 \\ 1 & -1 \\ -1 & 1 \\ 1 & 1 \end{bmatrix}, \qquad \mathbf{X}_1 = \begin{bmatrix} -2.18 & -2.18 \\ 1.84 & -0.16 \\ -0.48 & 1.52 \\ 0.83 & 0.83 \end{bmatrix},$$

$$\mathbf{y} = \begin{bmatrix} 2 \\ 2 \\ 0 \\ -4 \end{bmatrix}$$

1. PLS model with no latent orthogonal variation present, $\mathbf{X}_0$, **y**:

$$\mathbf{w}1(X_0)^{\mathrm{T}} = [-0.45 \quad -0.89],$$

$$\mathbf{p}1(X_0)^{\mathrm{T}} = [-0.45 \quad -0.89],$$

$$\mathbf{b}_{\mathrm{PLS}}(X_0)^{\mathrm{T}} = [-1 \quad -2]$$

Theresulting one-component PLS model parameters are shown above. Only one PLS component is necessary. Note that the loading vectors **p**1 and **w**1 are identical. The regression coefficients $\mathbf{b}_{\mathrm{PLS}}$ display negative correlations for both variables and with twice the magnitude for the second variable.

2. PLS model with latent orthogonal variation present, $\mathbf{X}_1$, **y**:

$$\mathbf{w}1(X_1)^{\mathrm{T}} = [-0.45 \quad -0.89],$$

$$\mathbf{p}1(X_1)^{\mathrm{T}} = [-0.69 \quad -0.77],$$

$$\mathbf{b}_{\mathrm{PLS}}(X_1)^{\mathrm{T}} = [0.08 \quad -1.08]$$

Above,**p**1 and **w**1 vectors from the first PLS component are shown together with the regression coefficients for the resulting two-component PLS model. Two PLS components were necessary to deal with the added orthogonal variation. Notice that the loading vectors **p**1 and **w**1 are not identical; this implies that latent orthogonal variation is present in **X**,

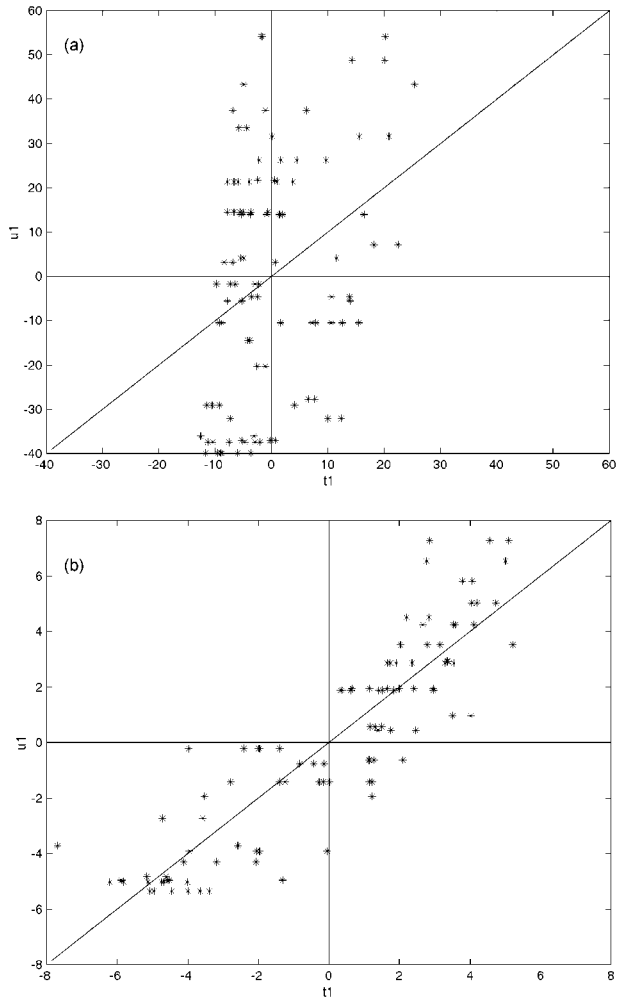

Figure 6. First-PLS-component **t–u** score plot for (a) the original PLS model and (b) the O-PLS-pretreated PLS model. A more distinct linear relation is present for the O-PLS-pretreated PLS model.

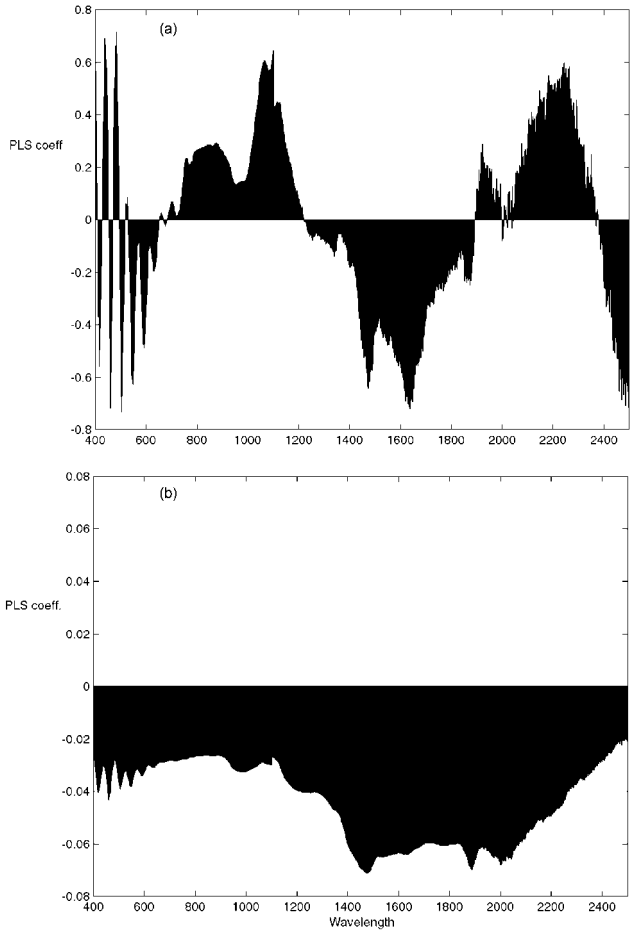

Figure 7. Regression coefficients for (a) the original PLS model and (b) the O-PLS-pretreated PLS model. The regression coefficients are clearly different, including the overall interpretation, which is a result of the non-correlated variation present in **X** for the original PLS model.

and hence O-PLS should be used. The regression coefficents are also different for the estimated two-component PLS model. The regression coefficient for the first variable has switched sign (positive), and the ratio between the two regression coefficients has also changed. This means that the interpretation of the regular PLS model will be different when latent orthogonal variation is present in **X** compared to the case with no orthogonal variation present.

3. O-PLS-pretreated PLS model with latent orthogonal variation present, $\mathbf{X}_1$, **Y**:

$$\mathbf{w}1_{\text{ortho}}(X_1)^{\text{T}} = [-0.89 \quad 0.45],$$

$$\mathbf{p}1_{\text{ortho}}(X_1)^{\text{T}} = [-1.16 \quad -0.09],$$

$$\mathbf{t}1_{\text{ortho}}(X_1)^{\text{T}} = [0.97 \quad -1.71 \quad 1.11 \quad -0.37]$$

Theone-component O-PLS model parameters are shown above. The amount of orthogonal variation removed ($\mathbf{t}1_{\text{ortho}}\mathbf{p}1_{\text{ortho}}^{\text{T}}$) corresponds to 43% of $\mathbf{X}_1$. A regular one-component PLS model was estimated on the O-PLS-pretreated matrix:

$$\mathbf{X}_{\text{O-PLS}} = \mathbf{X}_1 - \mathbf{t}1_{\text{ortho}}\mathbf{p}1_{\text{ortho}}^{\text{T}}$$

$$\mathbf{w}1(X_{\text{OPLS}})^{\text{T}} = [-0.45 \quad -0.89],$$

$$\mathbf{p}1(X_{\text{OPLS}})^{\text{T}} = [-0.45 \quad -0.89],$$

$$\mathbf{b}(X_{\text{OPLS}})^{\text{T}} = [-0.41 \quad -0.82]$$

Theone-component PLS model parameters are shown above. Note that the loading vectors $\mathbf{p}1$ and $\mathbf{w}1$ are identical. The regression coefficients display negative correlations for both variables and with twice the magnitude for the second variable.

The interpretation of the regression coefficients for the original PLS model where no latent orthogonal variation is present (1) is equivalent to that for the O-PLS-pretreated PLS model where orthogonal variation is present (3). This is not the case for the regular PLS model where orthogonal variation exists (2). It gives a different interpretation (switched sign) of the variables and their relationship with **Y**, which is important to note.

Table II. PLS modeling results from NIR example (data set ASSI NIR)

Method	Orthogonal comp.	PLS comp.	R2Y (cum)	Q2Y (cum)	RMSEP
PLS	–	5	0.80	0.74	2.95
O-PLS/PLS	4	1	0.80	0.74	2.95
O-PLS(PCA)/PLS	2	1	0.80	0.74	2.94
POSC/PLS	1	1	0.80	0.73	2.94
OSC/PLS	2	1	0.81	0.72	3.06
Fearn/PLS	4	1	0.74	0.68	3.19

3. EXPERIMENTAL

NIR-VIS spectra were collected from the wavelength region 400–2500 nm. An NIR-Systems 6500 spectrometer was installed on top of a conveyer belt, and 151 baskets filled with different wood chip compositions were measured next to the conveyer belt at ASSI Domän pulp plant in Piteå, Sweden. The dry content was measured using a reference method. The wood chip dry content varied from 39% to 58% (w/w). From the data set $N \times K$, where $N = 151$ samples and $K = 1050$ digitized wavelengths, 51 spectra were randomly removed as a prediction set, leaving 100 spectra for use as a calibration set.

4. RESULTS

For comparison, a regular PLS model and the O-PLS-pretreated PLS model will be discussed in some detail. The unfiltered data display a clear baseline variation with little relevance for moisture content. Figure 3(a) shows the raw column-centered NIR spectra. The irrelevant baseline and slope variations have been greatly reduced for the O-PLS-filtered spectra, as depicted in Figure 3(b). The number of PLS components was calculated according to cross-validation [11]. However, another indicative measure of the correct number of orthogonal components is to plot the ratio $\|\mathbf{w}_{\text{ortho}}\|/\|\mathbf{p}\|$ (from steps 6 and 7 in the O-PLS algorithm) for each O-PLS component. In Figure 4 this ratio for each O-PLS component is shown, and four orthogonal components were removed from **X** according to cross-validation, which the plot also indicates, because after four components the amplitude is down at noise level.

It is very informative to analyze the non-correlated variation in **X** as principal components. Here the O-PLS components removed 88% of the total variation in **X**. It should be clear that these plots, scores and loadings can be interpreted without regarding their relation to **Y**, dry content. The first two orthogonal loadings are plotted in Figure 5. Figure 5(a) illustrates the first principal orthogonal loading of the disturbing variation in **X**, and Figure 5(b) the second component. The orthogonal components represent offset shift (baseline) and slope differences. Those non-correlated variations were detected and suppressed with the O-PLS method. The corresponding score plot for the orthogonal components is not shown.

All single-**y** PLS models with more than one PLS component would benefit from using O-PLS. Consider designed data with orthogonal variables; only one PLS component is needed, because no latent orthogonal variation exists. The NIR spectra **X** contain large baseline variations not correlated to the dry content. This creates problems for the PLS algorithm. PLS is forced to include **X–Y** covariance in each PLS component even though a great deal of the **X** variation is orthogonal to **Y**. PLS solves this by peeling off variation from the **X** matrix in a couple of components, leading to a more complex PLS model. A strong indication of when to use O-PLS can be seen in Table I. The amount of variance explained in **Y**, R2Y cum (or cross-validated Q2cum), is relatively small in the first component. The amount of explained variation in **X**, R2Xcum, is rather large. This implies that large orthogonal variation with regard to **Y** exists in **X**. The O-PLS-pretreated data required only one PLS component, because all orthogonal latent variation had been removed.

In Figure 6(a) the original PLS model does not show much **t–u** correlation in the first PLS component. The **t–u** correlation in the O-PLS-pretreated PLS model is much more distinct and clear, as can be seen in Figure 6(b).

The original PLS model and the O-PLS-pretreated PLS model have the same first loading weight vector **w**, but different scores **t** are produced. This implies that the PLS weight vector **w** is not very useful when orthogonal variation is present. The loading vector **p** is more relevant to analyze with respect to the scores. However, the loadings **P** are influenced by both the correlated and the non-correlated variation in **X** mixed together. This makes the interpretation of the original PLS model ambiguous. Preprocessing the data with O-PLS strongly helps the interpretation of the PLS model.

The regression coefficients of the original PLS model and the O-PLS-pretreated PLS model are very different as a result of the orthogonal variation present in the NIR spectra; see Figure 7. The regression coefficients for the O-PLS-pretreated PLS model show negative correlation to dry content for all wavelengths and resemble an NIR spectrum, contrary to the original PLS model coefficient plot which gives a quite different interpretation. The difference in interpretation originates from the amount of orthogonal variation in **X**.

Table II shows that O-PLS gave a distinct reduction in the total number of components needed when PCA was used on $\mathbf{X}_{\text{ortho}}$; instead, only the principal component found was removed from **X**, and not the whole $\mathbf{X}_{\text{ortho}}$ matrix. Scaling to unit variance was thereafter applied for the PCA case. The number of resulting PLS components is always reduced to one (if single **y**), and as a rule the total number of components should never exceed the number of components for the original PLS model. The OSC method and the orthogonal filtering method proposed by Fearn increased the prediction residuals slightly, probably owing to a minor overfit.

5. CONCLUDING REMARKS

The proposed O-PLS method has been shown to be generic and versatile. The clearest advantage of using O-PLS is the improvement in interpretation of PLS models and their parameters scores, loadings and residuals. The interpretation of a regular PLS model that contains systematic orthogonal variation is ambiguous, and it has been demonstrated with both a synthetic and a real data set that O-PLS effectively helped to remove the non-correlated variation from **X** and thereby improved the interpretation of the resulting one-component PLS model. In the NIR example the orthogonal variation in the data was further analyzed with principal component analysis, which revealed that the non-correlated variation arose from disturbing baseline and slope variations.

O-PLS can be embedded as an integrated part of the regular PLS modeling. O-PLS does not only apply for multivariate calibration purposes, but for all types of filtering where non-correlated systematic variation in the data **X** needs to be removed or analyzed. In industrial processes it is not always possible to remove disturbing variation. The O-PLS method offers the advantage to at least know what type of disturbing variation exists and perhaps find methods to reduce it. Calibration transfer is another area where orthogonal correction methods have proven useful [14,15].

It is often possible to construct a response matrix **Y** if it does not exist. In classification problems, **Y** is constructed using dummy variables (1/0). O-PLS would then find the variation in **X** that separates the groups (correlated variation) and the variation in **X** that combines the groups (non-correlated variation). Using O-PLS on time series and setting time as **Y** would separate variations in **X** related to time. Another area where the use of O-PLS will prove warranted is in quantitative structure–activity relationships (QSARs), where the **X** data matrix often contains many thousands of variables but the majority have little relevance to e.g. biological activity. O-PLS helps to clean up the data prior to PLS analysis. Combinations of O-PLS and other pretreatment methods such as wavelets [16,17] will be looked into in the near future.

Acknowledgements

The authors are indebted to Dr Henrik Antti, Umea University and Dr Lars Wallbäcks, ASSI Domän, Piteå, Sweden for supplying the NIR data set. Financial support from the National Graduate School in Scientific Computing (NGSSC), the Centre for Forest Biotechnology and Chemistry and the Swedish Natural Science Research Council (NFR) is gratefully acknowledged.

REFERENCES

1. Martens H, Naes T. *Multivariate Calibration* (2nd edn), vol. **1**. Wiley: Chichester, 1989.
2. Savitsky A, Golay MJE. Smoothing and differentiation of data by simplified least squares procedures. *Anal. Chem.* 1964; **36**: 1627–1639.
3. Geladi P, MacDougall D, Martens H. Linearization and scatter-correction for near-infrared reflectance spectra of meat. *Appl. Spectrosc.* 1985; **3**: 491–500.
4. Williams PC, Norris K. *Near-infrared Technology in Agricultural and Food Industries*. American Cereal Association: St Paul, MN, 1987.
5. Sun J. Statistical analysis of NIR data: data pretreatment. *J. Chemometrics* 1997; **11**: 525–532.
6. Baroni M, Clementi S, Cruciani G, Constantino G, Riganelli D. Predictive ability of regression models. Part 2: Selection of the best predictive PLS model. *J. Chemometrics* 1992; **6**: 347–356.
7. Barnes RJ, Dhanoa MS, Lister SJ. Standard normal variate transformation and detrending of near-infrared diffuse reflectance spectra. *Appl. Spectrosc.* 1989; **43**: 772–777.
8. Wold H. Nonlinear estimation by iterative least squares procedures. In *Research Papers in Statistics*, David F (ed.). Wiley: New York, 1996; 411–444.
9. Wold S, Antti H, Lindgren F, Öhman J. Orthogonal signal correction of near-infrared spectra. *Chemometrics Intell. Lab. Syst.* 1998; **44**: 175–185.
10. Fearn T. On orthogonal signal correction. *Chemometrics Intell. Lab. Syst.* 2000; **50**: 47–52.
11. Wold S. Cross-validatory estimation of the number of components in factor and principal components models. *Technometrics* 1978; **20**: 397–405.
12. Malinowski ER. *Factor Analysis in Chemistry* (2nd edn). Wiley: New York, 1991.
13. Rannar S, Geladi P, Lindgren F, Wold S. A PLS kernel algorithm for data sets with many variables and few objects. Part II: Cross validation, missing data and examples. *J. Chemometrics* 1995; **9**: 459–470.
14. Sjöblom J, Svensson O, Josefsson M, Kullberg H, Wold S. An evaluation of orthogonal signal correction applied to calibration transfer of near infrared spectra. *Chemometrics Intell. Lab. Syst.* 1998; **44**: 229–244.
15. Geladi P, Bärring H, Dåbakk E, Trygg J, Antti H, Wold S, Karlberg B. Calibration transfer for predicting lake-water pH from near infrared spectra of lake sediments. *J. Near Infrared Spectrosc.* 1999; **7**: 251–264.
16. Eriksson L, Trygg J, Johansson E, Bro R, Wold S. Orthogonal signal correction, wavelet analysis, and multivariate calibration of complicated process fluorescence data. *Anal. Chim. Acta* 2000; **420**: 181–195.
17. Trygg J, Wold S. PLS regression on wavelet compressed NIR spectra. *Chemometrics Intell. Lab. Syst.* 1998; **42**: 209–220.
18. Kvalheim OM, Karstang TV. Interpretation of latent-variable regression models. *Chemometrics Intell. Lab. Syst.* 1989; **7**: 39–51.
19. Christie OHJ. Data laundering by target rotation in chemistry-based oil exploration. *J. Chemometrics* 1996; **10**: 453–461.

APPENDIX I

An outline of the proposed O-PLS method is shown here for a matrix **Y**.

- Optionally transform, center and scale the raw data to give the matrices **X** and **Y**.

1. $\mathbf{w}^T = \mathbf{y}^T\mathbf{X} / (\mathbf{y}^T\mathbf{y})$. For each column in **Y**, estimate the corresponding **w** and create a matrix $\mathbf{W} = [\mathbf{W}\ \mathbf{w}]$.
2. $\mathbf{W} = \mathbf{T}_w\mathbf{P}_w^T + \mathbf{E}_w$. Estimate with PCA the principal components of **W** as long as the ratio of the sum of squares of the current score vector $\mathbf{t}_w$ divided by the sum of squares of **W** is larger than a given threshold, typically 10^{-10}.
3. Estimate a regular multi-*Y* PLS component with given **X** and **Y** (steps 4–9).

4. Initialize multi-Y PLS calculation by setting a column in $\mathbf{Y}$ to $\mathbf{u}$.
5. $\mathbf{w}^T = \mathbf{u}^T\mathbf{X} / (\mathbf{u}^T\mathbf{u})$. Repeat steps 5–9 until convergence.
6. $\mathbf{w} = \mathbf{w} / \|\mathbf{w}\|$.
7. $\mathbf{t} = \mathbf{X}\mathbf{w} / (\mathbf{w}^T\mathbf{w})$.
8. $\mathbf{c}^T = \mathbf{t}^T\mathbf{Y} / (\mathbf{t}^T\mathbf{t})$.
9. $\mathbf{u} = \mathbf{Y}\mathbf{c} / (\mathbf{c}^T\mathbf{c})$. Check convergence; if $\|\mathbf{u}_{new} - \mathbf{u}_{old}\| / \|\mathbf{u}_{new}\| < 10^{-10}$, continue to step 10, otherwise return to step 5.
10. $\mathbf{p}^T = \mathbf{t}^T\mathbf{X} / (\mathbf{t}^T\mathbf{t})$. To estimate an orthogonal component, go to step 11, otherwise go to step 17.
11. $\mathbf{p} = \mathbf{p} - [\mathbf{t}_w^T\mathbf{p}/(\mathbf{t}_w^T\mathbf{t}_w)]\mathbf{t}_w$. Orthogonalize $\mathbf{p}$ to each column in $\mathbf{T}_w$, then set $\mathbf{w}_{ortho} = \mathbf{p}$. In this way, orthogonality to all $\mathbf{Y}$ variables is ensured for resulting orthogonal score vector in step 13. The $\mathbf{p}$ vector in this step can also be an arbitrary vector (e.g. PCA loading of $\mathbf{X}$).
12. $\mathbf{w}_{ortho} = \mathbf{w}_{ortho} / \|\mathbf{w}_{ortho}\|$.
13. $\mathbf{t}_{ortho} = \mathbf{X}\mathbf{w}_{ortho}/(\mathbf{w}_{ortho}^T\mathbf{w}_{ortho})$.
14. $\mathbf{p}_{ortho}^T = \mathbf{t}_{ortho}^T\mathbf{X}/(\mathbf{t}_{ortho}^T\mathbf{t}_{ortho})$.
15. $\mathbf{E}_{O-PLS} = \mathbf{X} - \mathbf{t}_{ortho}\mathbf{p}_{ortho}^T$. $\mathbf{E}_{O\text{-}PLS}$ are the filtered data.
16. Save found parameters $\mathbf{T}_{ortho} = [\mathbf{T}_{ortho}\ \mathbf{t}_{ortho}]$, $\mathbf{P}_{ortho} = [\mathbf{P}_{ortho}\ \mathbf{p}_{ortho}]$, $\mathbf{W}_{ortho} = [\mathbf{W}_{ortho}\ \mathbf{w}_{ortho}]$. Return to step 4 and set $\mathbf{X} = \mathbf{E}_{O\text{-}PLS}$.
17. To find orthogonal variation for the next PLS component, remove current PLS component from $\mathbf{X}$ and $\mathbf{Y}$ and save the parameters for this PLS component for future samples, $\mathbf{E} = \mathbf{X} - \mathbf{t}\mathbf{p}^T$, $\mathbf{F} = \mathbf{Y} - \mathbf{t}\mathbf{c}^T$, $\mathbf{T}_{pls} = [\mathbf{T}_{pls}\ \mathbf{t}]$, $\mathbf{W}_{pls} = [\mathbf{W}_{pls}\ \mathbf{w}]$, $\mathbf{P}_{pls} = [\mathbf{P}_{pls}\ \mathbf{p}]$, and return to step 1 and set $\mathbf{X} = \mathbf{E}$ and $\mathbf{Y} = \mathbf{F}$. Otherwise to stop, go to step 18.
18. $\mathbf{X}_{ortho} = \mathbf{T}_{ortho}\mathbf{P}_{ortho}^T$. Analyze orthogonal variation component-wise, or run PCA on $\mathbf{X}_{ortho}$ (step 19).
19. $\mathbf{X}_{ortho} = \mathbf{T}_{pca\_ortho}\mathbf{P}_{pca\_ortho}^T + \mathbf{E}_{pca\_ortho}$. PCA of $\mathbf{X}_{ortho}$ to summarize the systematic orthogonal variation. Removing all estimated orthogonal variation from $\mathbf{X}$ is one option; another option is to remove only the principal orthogonal components estimated in step 19. This corresponds to adding $\mathbf{E}_{pca\_ortho}$ back into $\mathbf{E}_{O\text{-}PLS}$.
20. $\mathbf{E}_{O-PLS} = \mathbf{E}_{O-PLS} + \mathbf{T}_{pls}\mathbf{P}_{pls}^T$. Add the PLS components removed back into $\mathbf{E}_{O\text{-}PLS}$, which now contains the filtered data.
21. New or future samples (the prediction set) are corrected using $\mathbf{W}_{ortho}$, $\mathbf{P}_{ortho}$, $\mathbf{W}_{pls}$ and $\mathbf{P}_{pls}$ from the calibration model. For each new observation vector $\mathbf{x}_{new}^T$, repeat steps 22–26 for each component (O-PLS or PLS) in the order they were calculated in the calibration model.
22. If component = OPLS: $t_{new\_ortho} = \mathbf{x}_{new}^T\mathbf{w}_{ortho}/(\mathbf{w}_{ortho}^T\mathbf{w}_{ortho})$.
23. If component = OPLS: $\mathbf{t}_{new\_ortho}^T = [\mathbf{t}_{new\_ortho}^T t_{new\_ortho}]$. Save orthogonal scores for prediction set. The first $\mathbf{t}$ in the brackets is a vector while the second t is a scalar.
24. If component = OPLS: $\mathbf{e}_{new\_O-PLS}^T = \mathbf{x}_{new}^T - t_{new\_ortho}\mathbf{p}_{ortho}^T$. Orthogonal component in $\mathbf{x}_{new}^T$ is removed. Set $\mathbf{x}_{new}^T = \mathbf{e}_{new\_O-PLS}^T$ for additional components and return to step 21, otherwise proceed to step 27.
25. If component = PLS: $t_{new\_pls} = \mathbf{x}_{new}^T\mathbf{w}_{pls}/(\mathbf{w}_{pls}^T\mathbf{w}_{pls})$, $\mathbf{t}_{new\_pls}^T = [\mathbf{t}_{new\_pls}^T\ \ t_{new\_pls}]$. The first $\mathbf{t}$ in the brackets is a vector while the second t is a scalar.
26. If component = PLS: $\mathbf{e}_{new\_pls}^T = \mathbf{x}_{new}^T - t_{new\_pls}\mathbf{p}_{pls}^T$. PLS component in $\mathbf{x}_{new}^T$ is removed. Set $\mathbf{x}_{new}^T = \mathbf{e}_{new\_pls}^T$ and return to step 22.
27. $\mathbf{x}_{new\_ortho}^T = \mathbf{t}_{new\_ortho}^T\mathbf{P}_{ortho}^T$.
28. $\mathbf{t}_{new\_pca\_ortho}^T = \mathbf{x}_{new\_ortho}^T\mathbf{P}_{pca\_ortho}$. Estimate new scores from PCA loadings in step 19, $\mathbf{x}_{new\_ortho}^T = \mathbf{t}_{new\_pca\_ortho}^T\mathbf{P}_{pca\_ortho}^T + \mathbf{e}_{new\_pca\_ortho}^T$. If only the orthogonal latent components from PCA on $\mathbf{X}_{ortho}$ were removed, then $\mathbf{e}_{new\_pca\_ortho}^T$ should be added back to $\mathbf{e}_{new\_O-PLS}^T$.
29. $\mathbf{e}_{new\_O-PLS}^T = \mathbf{e}_{new\_O-PLS}^T + \mathbf{t}_{new\_pls}^T\mathbf{P}_{pls}^T$. Add the PLS components back into $\mathbf{e}_{new\_O-PLS}^T$, which now contains the filtered data.

APPENDIX II

An outline of the proposed POSC method is shown here for the matrices $\mathbf{X}$ and $\mathbf{Y}$.

- Optionally transform, center and scale the raw data to give the matrices $\mathbf{X}$ and $\mathbf{Y}$.

1. $\mathbf{T} = \mathbf{X}\mathbf{W}$. Calculate the normalized regression coefficients $\mathbf{W}$ from some regression method (e.g. PLS) to estimate $\mathbf{T}$, representing the best systematic correlation to $\mathbf{Y}$.
2. $\mathbf{T} = \mathbf{T}_{pca}\mathbf{P}_{pca}^T + \mathbf{E}_{pca}$. Estimate with PCA the principal components of $\mathbf{T}$ as long as the ratio of the sum of squares of the current score vector $\mathbf{t}_{pca}$ divided by the sum of squares of $\mathbf{T}$ is larger than a given threshold, typically 10^{-10}.
3. $\mathbf{p}^T = \mathbf{t}_{pca}^T\mathbf{X}/(\mathbf{t}_{pca}^T\mathbf{t}_{pca})$. Estimate $\mathbf{p}$ for each column in $\mathbf{T}_{pca}$, resulting in matrix $\mathbf{P}$.
4. $\mathbf{X}_{ortho} = \mathbf{X} - \mathbf{T}_{pca}\mathbf{P}^T$.
5. $\mathbf{X}_{ortho} = \mathbf{T}_{ortho}\mathbf{P}_{ortho}^T + \mathbf{E}_{ortho}$. Calculate a PCA model.
6. $\mathbf{X}_{posc} = \mathbf{X} - \mathbf{T}_{ortho}\mathbf{P}_{ortho}^T$. Remove the systematically irrelevant variation.
7. New or future data (the prediction set) are corrected using $\mathbf{W}$, $\mathbf{P}_{pca}$, $\mathbf{P}$ and $\mathbf{P}_{ortho}$ from the calibration model. For each new observation vector $\mathbf{x}_{test}^T$, repeat the steps below.
8. $\mathbf{t}_{test}^T = \mathbf{x}_{test}^T\mathbf{W}$.
9. $\mathbf{t}_{testpca}^T = \mathbf{t}_{test}^T\mathbf{P}_{pca}$.
10. $\mathbf{x}_{test\_ortho}^T = \mathbf{x}_{test}^T - \mathbf{t}_{testpca}^T\mathbf{P}^T$. Orthogonal variation in new data.
11. Repeat steps 12–14 for each orthogonal principal component removed in step 6.
12. $\mathbf{t}_{test\_ortho} = \mathbf{x}_{test\_ortho}^T\mathbf{p}_{\_ortho}$.
13. $\mathbf{t}_{test\_ortho}^T = [\mathbf{t}_{test\_ortho}^T\ \ t_{test\_ortho}]$. Save t_{test_ortho}. The first $\mathbf{t}$ in the brackets is a vector while the second t is a scalar.
14. $\mathbf{e}_{test\_ortho}^T = \mathbf{x}_{test\_ortho}^T - t_{test\_ortho}^T\mathbf{p}_{\_ortho}^T$. For each remaining orthogonal component, set $\mathbf{x}_{test\_ortho}^T = \mathbf{e}_{test\_ortho}^T$ and return to step 12 to remove any additional orthogonal component estimated earlier, else proceed to step 15.
15. $\mathbf{x}_{posc\_test}^T = \mathbf{x}_{test}^T - \mathbf{t}_{test\_ortho}^T\mathbf{P}_{ortho}^T$. Filtered new data $\mathbf{x}_{posc\_test}^T$.