Research Article

Journal of CHEMOMETRICS

Received: 2 February 2010, Revised: 9 September 2010, Accepted: 16 September 2010, Published online in Wiley Online Library: 29 December 2010

(wileyonlinelibrary.com) DOI: 10.1002/cem.1357

Path modelling by sequential PLS regression

T. Næs[a]*, O. Tomic[a], B.-H. Mevik[b] and H. Martens[a]

This paper presents a new approach to path modelling, based on a sequential multi-block modelling in latent variables. The approach is explorative and focused on interpretation. The method breaks with standard traditions of estimating all paths using one single modelling. Instead, one separate model is estimated for each endogenous block. Each separate model is constructed by stepwise use of the standard PLS regression on matrices that are orthogonalised with respect to each other. The advantages of the approach are that it can allow for different dimensionality within each block, it is invariant to relative weighting of the blocks and it is based on simple and standard methodology allowing for simple outlier detection, validation and interpretation. No convergence problems are involved and the method can be used for situations with many more variables than samples. An application based on sensory analysis of wines will be used to illustrate the method. Copyright © 2010 John Wiley & Sons, Ltd.

Keywords: path models; LS-PLS; PLS; regression; multi-block; interpretation

1. INTRODUCTION

In modern science there is often a strong need for understanding the relationship between a number of different sets/blocks of variables, for instance, between different measurement instruments, between measurements taken at different places in a process (be it an industrial or a biological process) or between different characteristics related to individuals, e.g. medical patients or consumers in a buying situation. The most basic methodology for this type of problem is regression analysis, where the focus initially was on quantitatively relating one variable to another. Such traditional one-to-one relationships can be easily displayed in a simple *x/y* plot and the linear model parameters estimated on a pocket calculator. When relating one data table to another one, good methods also exist that do require statistical parameter estimation in a computer, but that also provide both overview and detailed insight into the relationships within and between the tables (see, e.g. Reference [1]).

In many cases, however, there is a need for understanding the relations between *several* data sets. This is a much more complex task, for several reasons. The structural form of each of the different relationships needs to be defined in such a way that it represents the underlying unknown reality, yet being practical and easy to parameterise. The individual model parameters then need to be established—defined by the scientist from prior studies or estimated from data. Collinearity and noise of various types usually force the scientist to make simplifying shortcuts, creating a compromise between all the detailed links that one would like to know and the few coarse links that the data allow to be determined. Finally, it is a challenge to represent and display the results so that the scientists can understand the finally obtained relations, both with respect to the overall picture, the individual estimated values and the various types of uncertainties involved.

An important contribution to methodology for handling several 'blocks' of data was made by Wright [2–4]. He developed basic concepts, concrete models and methodology that have later found important applications in a number of disciplines. The methodology is often today referred to as *path modelling*. In path modelling, a complicated system of information flows, consisting of parallel and serial model elements, are linked according to prior knowledge or according to what one is interested in. Usually, the links are implemented as linear approximation models. The information flow usually has a one-way direction, thus representing so-called *directed graphs* (in contrast to, e.g. cyclic graphs). Wright's original concept was to link individual manifest (observed) variables to each other through such a directed graph built on a set of regressions whose parameters are estimated from data. The directions of the 'arrows' of the graph may be defined so that they sort the blocks of variables in some expected natural order—the flow of information points forward in time or outward in space, or one may impose some other causality or interpretability structure. This approach thus represents one 'extreme' variant of multi-block modelling. The advantage is simplicity in developing and interpreting the model. Its main drawback is that someone must select the salient set of informative variables beforehand.

The other 'extreme' approach is to lump all the blocks together into one and search for clear co-variation patterns among the full set of variables by, e.g. Principal Component Analysis (PCA, see e.g. References [5,6]). This is a simple approach, but lacks interpretation tools for more detailed analysis of differences and similarities between the blocks. In order to remedy this problem a number of multi-block versions of the PCA approach have been developed (see, e.g. References [7–9]). These give insight into the general consensus structure of the blocks as well as how the individual blocks relate to the overall PCA patterns. The advantage of these methods is the overview that they give; their

* *Correspondence to: T. Næs, Nofima Mat, Osloveien 1, 1430 Ås, Norway. E-mail: tormod.naes@nofima.no*

a *T. Næs, O. Tomic, H. Martens*
Nofima Mat, Osloveien 1, 1430 Ås, Norway

b *B.-H. Mevik*
USIT, University of Oslo, Norway

drawback is that it is difficult to impose an explicit flow of information between the blocks. Other related approaches can be found in References [10–13] where other criteria for relating several blocks are proposed and analysed (see also Reference [14]). These criteria range from optimising correlations between linear combinations of the blocks to various compromises between correlation and explained variance. In Reference [8] a number of methods are discussed which impose a predictive relation between several input blocks on one side and one or several output blocks on the other. The methods are based on partial least squares (PLS) regression and bear some relationship to the path modelling method proposed in this paper. Tenenhaus and Hanafi [15] present an important review of some of the methods discussed above and also show how they relate to the PLS path modelling approach to be discussed below.

The original path modelling methodology proposed by Wright ([2–4]) was extended and fine-tuned in the seventies (see, e.g. References [16] and [17]) and developed into a unified framework called structural equations modelling (SEM). This framework is based on combined modelling at two levels: a set of equations linking so-called latent variables in the different blocks to each other, combined with a measurement model that defines the relation between the measured (manifest) variables and the latent variables, thus representing a kind of compromise between the two 'extremes' mentioned above, the original manifest variables version and the class of multi-block methods. Lots of theoretical contributions describing the basic properties of the SEM methods have also been made (see, e.g. Reference [18]).

The most used criteria for estimation of parameters in path models (SEM models) are based on maximum likelihood (ML) and partial least squares (PLS, see References [19,20]). Both approaches are based on joint fitting of all data blocks, allowing for all data to influence all structural elements in the system. The use of the two different criteria ML and PLS for estimation has, for the last couple of decades, developed into two different modelling schools, with quite different philosophies and ideas and only moderate communication between them. In our point of view the difference between the two traditions highlights a number of important methodological aspects which are useful for obtaining a deeper understanding of path modelling and also for further development. We refer to the Section 2 below and to Reference [21] for further discussion of some of these issues: we also refer to References [22] and [23] for alternative PLS-based approaches to path modelling. These are different from the standard PLS path modelling approach, but are based on similar path model diagrams and use the original PLS principles of covariance optimisation. So far they have obtained less attention than the original version by Wold [24].

In the present paper we will start by discussing path modelling more generally, with the distinction between the two philosophies mentioned above (PLS and ML) as the point of departure. This discussion will lead to the development of a new, sequential approach to path modelling which addressed some of the problems associated with the established methods (as discussed in Section 2). From one perspective, the new method takes one step back (towards the original approach) and breaks the paradigm that all data in the system should influence everything. Instead, the method will be based on splitting the modelling process into a consciously chosen sequence of PLS modelling steps for each endogenous (dependent variable) block versus the blocks which have an input relationship to it. For estimation we use the sequential multi-block LS-PLS method [25], which is simple to use and compute and which has been shown to have good prediction and interpretation properties for multi-block data with a predictive structure [26,27]. The new method is interpretation oriented, but interpretation is done in models which are developed with as good prediction ability as possible using sequential PLS methodology. It should be emphasised that the proposed method is particularly useful for explorative analysis and should therefore be considered as a method of special interest for an early-stage investigation of a block diagram. This fits well with the fact that only well established and easy-to-use and easy-to-compute methodology is involved. The method will be tested on a previously published data set concerning sensory descriptive testing of wine [28,29].

2. CULTURAL ASPECTS OF AND EXISTING APPROACHES TO PATH MODELLING

Path models are models that link a number of data blocks to each other [18,30], thereby establishing a more or less complicated set of paths along which information may be considered to flow between the blocks, representing a known time sequence, an assumed causality order, or some other chosen organising principle. Usually, a visual diagram is used for highlighting how the blocks are related. A data block used for prediction only is called *exogenous*, while all other blocks are called *endogenous* (see, e.g. Reference [18]). An illustration of such a path or arrow diagram is depicted in Figure 1. This is an example obtained from Martens *et al*. [29] and is related to descriptive sensory analysis and sensory quality of wine (see also Reference [28]). Different characteristics of wine, representing the different blocks, are assessed in sequence, and one is interested in how the different sensory property types are linked. For instance, how much is gained for predicting (either mathematically or cognitively, depending on perspective) the hedonic assessment of overall wine quality from the tasting attributes when the odour and visual attributes have already been taken into account?

In many cases, the input variables (manifest variables) in the individual blocks can be very numerous and inter-correlated. Thus, direct fitting of data blocks to each other by, for instance, least squares (LS) becomes impossible. In structural equation modelling this is handled by introducing a small number of so-called latent variables which describe the main variability in the manifest variables. The model for each block resembles a

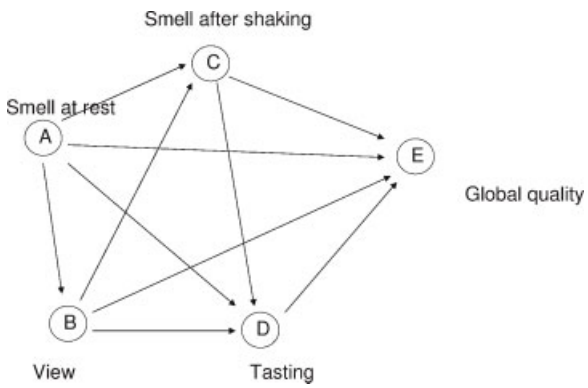

Figure 1. Path diagram for wine data set. The different blocks A–E represent different aspect of the sensory profile (measured in alphabetical order). The arrows indicate the relations of interest to study.

regular factor analysis model and is called the measurement model. The structural relations are then modelled using the latent variables instead of the manifest ones.

In our approach discussed below, the rigorous model structure underlying the general SEM modelling is of less relevance because of the explorative nature and will therefore not be discussed further here. The interested reader is referred to References [18,30].

2.1. General aspects

There are a number of aspects within path modelling that seem to be traditionally accepted and that are underlying most method developments in the area. We will now first discuss some of these aspects on a general basis and afterwards relate them to the two basic principles of ML and PLS before using this discussion as a basis for proposing an alternative methodology.

2.1.1. *Causal modelling versus regression fitting*

It seems to be a quite strong tradition within the area of SEM modelling to consider the different blocks of measurements as linked through so-called *causal* relations. This is particularly true for the ML approach discussed below. Causal modelling is, however, always extremely difficult and is strongly related to the concept of controlled experiments or detailed mechanistic insight. In observational studies there is always a chance that there are other unmeasured variables that are responsible for the estimated structure and which thus influence the parameter estimates, unless there is a clear separation in e.g. time or space. Moreover, living systems—including people—invariably experience mutual causality through various feedback mechanisms. This can sometimes make causal inference of even temporally well-separated phenomena difficult. Pearl [31] showed that pointing out unidirectional causalities in such webs of reciprocal influences is rather risky.

The *direction* of the SEM paths may be causal, but the estimates may be highly misleading if interpreted as the causal *strength* (effect size) that one set of variables has on another. In our point of view, however, this potential misuse of the causality concept does not reduce the importance of the path modelling methodology as such, and a more pragmatic and useful way of looking at the SEM models is to consider them simply as networks of regression models that can be used to estimate the apparent, predictive ability that the blocks may have on each other. This is a point of view which is in line with standard ways of looking at regression models, namely as methods which are useful for estimating correlations and relations. It is usually only together with additional mechanistic knowledge that presumed causal relations can be imposed and tested. And given the natural human propensity for ascribing causality to observed correlations, claims of causality need to be viewed critically unless substantiated by independent empirical evidence. Such a viewpoint will be underlying the discussion in the rest of this paper. Note that this type of reasoning is also quite common within the area of PLS path modelling which will be discussed below (see also Reference [32]).

2.1.2. *One-dimensional versus multi-dimensional block structure*

Many applications of path modelling have traditionally focused on one-dimensional block models: This seems to have its background in the social sciences where a given block is often supposed to represent a battery of measurements for the *same* underlying phenomenon. A device called Cronbach's alpha has been developed in order to evaluate the realism of this assumption. Presumably, the difficulty in measuring this single latent variable precisely and accurately is supposed to be remedied by getting at it from several different angles. But this is clearly an assumption which is questionable in many important applications both in the social sciences and in the natural sciences. If, for instance, sensory analysis or a chemical instrument is involved as one of the blocks, there is every reason to expect collinearity and thus low rank, but no reason at all to assume a rank of one, i.e. uni-dimensionality. Different sensory or instrumental descriptors will usually reflect a set of several underlying phenomena, although in different ways. Only in ill-designed studies will *all* the salient underlying phenomena co-vary completely and thus collapse to uni-dimensionality, which can make interpretation really difficult. In other words, there is a need for an easy-to-use approach that is suitable for well-designed studies, which opens up for different dimensionality within each of the blocks (see the discussion on hypothetical constructs and indicators in Reference [33]).

For the classical PLS path modelling method to be discussed below, this rank-one assumption is inherent in one of the modes of the methodology (mode A), although there is ongoing research attempting at relaxing this assumption (References [34–36], XLSTAT-PLSM software (www.xlstat.com), see also below). Further discussion on this issue is given below. For the ML-based methods there is an opening for having several dimensions for each block, but as far as we are aware of, this possibility is mainly used in ML factor analysis and not fully explored for more general path models. Allowing several dimensions in each block clearly complicates both software convergence and model identification [18], as well as the graphical display of the solution and its interpretation.

2.1.3. *The same input and output latent variable in each block*

All the methods discussed above assume implicitly that for a block that is both endogenous ('dependent') with respect to another block and exogenous ('independent') with respect to a third, the *same* latent variable is used for prediction and to be predicted. There is, however, no obvious reason to believe that the same aspect of **X** that can be predicted from **Z** is useful for prediction of **Y**. It may, for instance, be that a chemical instrument is excellent for predicting sensory variable A and not at all able to predict sensory variable B, but it is a combination of both A and B which is used to predict consumer acceptance. In such cases one common latent variable is not very useful.

2.1.4. *Direct and additional effect*

Interpreting regression models is always difficult in situations with many explanatory variables, in particular if the latter are inter-correlated. The theory of SEM models introduces the concepts of total, direct and indirect effects which are very useful for interpreting networks of models [18,37]. There is, however, another aspect that seldom seems to be taken into account in this type of modelling, namely the concept of 'additional effect'. This has to do with the additional effect of adding a variable or a block of variables to the model. If variable **X** is already in the model, what is the additional effect of adding variable **Z** for predicting **Y**? In regression theory this type of information is often

obtained by the use of orthogonalisation, which is the basis for all significance testing in regression and analysis of variance (ANOVA). As will be shown below, this concept links well with the proposed approach, since it is based on orthogonalising blocks with respect to each other. In this paper the size of the additional effect as measured by, for instance, R^2 or another measure of explained variance will be in focus.

2.2. Two established traditions

2.2.1. ML-based SEM modelling

This is a method which is well established in the statistical literature (References [16] and [38]) and is often referred to as LISREL. It is based on modelling all relevant relations by assuming a structural equation and fitting the corresponding covariance structure to the data by ML optimisation assuming normality. The method is quite well understood theoretically, but also has problems associated with it.

For instance, for technical reasons (at least for the standard approach) the number of observations N must be larger than the number of variables p in each of the manifest blocks. In situations with, for instance, several hundred or thousand manifest variables in one or several of the blocks, which may often be the case in modern science, this approach becomes impossible to use. In some cases, there are also convergence problems associated with the ML algorithm and there may be complex identification problems related to parameter uniqueness.

A third weakness, which is not so much method oriented, but more related to user culture, is that the classical LISREL approach focuses on the structural information only (estimates intended to describe the population), not on the scores or residuals related to the objects (e.g. individual subjects). It is known from many types of applications in regression that the latter types of information can be just as useful as the structural information, both for interpretation and for diagnostic purposes, and particularly so when displayed graphically. Methods for calculating scores are available for special situations (FA), but usually this aspect is not focused [18,37]. Model adequacy is usually tested using a likelihood ratio test that follows an approximate chi square distribution. If only a moderate or small data set is available, rather serious discrepancies can, however, pass unnoticed, which again points to the estimation and graphical display of score and residuals to provide the user with opportunities to employ his or her more or less tacit background knowledge as additional information in the data modelling process.

2.2.2. PLS-based path modelling

This technique is an important alternative to the ML estimation and solves the problem of more variables than samples and has few or no convergence problems [19]. It is designed to extract and display both structural (population) and individual (scores) information. Two modes are available, one (mode A, reflective mode) where the manifest variables can be considered as linear function of an underlying latent variable plus noise and the other mode (mode B, formative mode) in which it is more natural to consider the latent variable simply as linear functions of the manifest ones, often identified by LS-like regressions.

One of the possible problems with this method is that it is algorithmically oriented, making it less transparent what it actually optimises and then also somewhat more difficult to study theoretically. The method is based on an iterative algorithmic multi-step estimation process iterating between finding relations between manifest and latent variables and finding relations between the latent variables themselves. Each of the local steps in the multi-step algorithm is based on least squares minimisations. But until recently global optimisation criteria and their convergence properties have not been available. Hanafi [39] has shown that for mode B, the solution can be characterised by a sum of a number of absolute values of correlations between latent variables. No similar results are known to the authors of this paper regarding mode A. A new criterion-based method for mode A estimation was developed by Tenenhaus and Tenenhaus [40] which does not have this problem.

As also discussed above, another possible problem with one of the modes of the original PLS path modelling method (mode A) is that only one latent variable is allowed for each block. This limits the complexity of each block to be basically one-dimensional, and thus also its realism and predictive power. Generalisations to several latent variables can, however, be envisioned. One possibility is to deflate each block separately [34], as in multi-block PCA [7]. But this may lead to correlations among the first and second components among the different blocks that may need a re-estimation of inner relations, as an extension of Martens' version of the two-block PLS Regression [1]. Another approach is to use PLS regression for estimating relations both between and within the blocks (References [32]; pp. 54–55 and [35]). This approach is simple and represents an elegant way of solving some of the problems mentioned. It also builds a bridge between the two modes in the sense that it can be used for blocks that are neither uni-dimensional (mode A) nor full-dimensional (mode B).

An advantage of the PLS method is that no identification problems exist for the method, since the solution is defined through the algorithm.

2.3. A new proposal—basic ideas

The features that we deem desired for multi-block path modelling are the following: natural inter-correlations between variables should not be treated as a 'collinearity problem', but as a stabilising advantage. The method should be easy to use also when $N < p$ and unwarranted and over-simplified ideas about causality and one-dimensionality should not be imposed on the data. The criterion for optimisation should be as clear as possible and the algorithm for estimating the model parameters should not contain large, complex iteration cycles whose stability and numerical precision are unknown. Equal attention should be given to structural and individual aspects. And finally, the results should be easy to validate and display graphically. As is clear from the above, our point of view is that the Wright manifest path modelling, the multi-block PCA, the LISREL and the PLS path modelling fail with respect to one or more of these criteria. This is the background for our new proposal, which in some sense both extends perspective and simplifies methodology. It will be discussed in Sections 6 how these different requirements are met for the new method.

The methodology developed here has an explorative character which makes it particularly useful in an early-phase investigation of the data. In some cases, one may follow up with more detailed and confirmative studies using, for instance, one of the other approaches mentioned. The method models each endogenous block separately as a function of all blocks that are input to it. A

major reason for this is that we want a method which can be used also for situations where the predictable part of an endogenous block is different from the part of the same block that is useful for prediction. In addition, it provides models which are simple to estimate and which also have some other advantages as discussed below. The estimation method will be the LS-PLS (least squares—PLS) method which has a number of advantages [25] and which also has been shown to have good predictive properties, also as compared to regular multi-block PLS (concatenated PLS, Reference [25]). Subsequent interpretation, based on either LS-PLS directly or the simplified PCP [41] method for post modelling interpretation (as will be in focus here), is then based on as stable and predictive models as possible within the PLS paradigm. The focus of the method is interpretation of relationships between blocks that enter into the same predictive relation, but only interpretation of models with good predictive power.

The first step of our new proposal is based on sequential use of regular two-block reduced-rank PLS regression [42], as proposed in References [25] and [27]. This method is called LS-PLS and has the advantages that it allows for different complexity within each block and it is invariant to the relative scales/units of the blocks. Both these aspects represent an advantage as compared to standard multi-block PLS where all input blocks are concatenated, since no more components than needed are extracted from each of the blocks (parsimony) and since one does not need to do any relative weighting of the blocks to compensate for different scales/units used. A possible challenge may be to choose the sequence of the blocks to incorporate, but in many relevant cases, the sequence is quite obvious as will be discussed below. The sequence is also controlled by what type of information is wanted, and in most cases, the predictive ability will probably not be influenced too much, since each step is based on predicting residuals from fitting of previous blocks. If in doubt, one may try alternative sequences. For the purpose of comparing alternatives, the CV-ANOVA method proposed in Reference [43] for comparison of prediction abilities may be useful. The method is simply based on two-way ANOVA for the prediction errors in a cross-validation context. One will then normally select one which is significantly better than the others tested. If none appears better than the others, interpretation criteria as just mentioned should be used for selecting the most natural solution (see also below).

In this paper, there is generally less need for the LS (least squares) step of the LS-PLS method and it is therefore natural to rename the method 'sequential and orthogonalised PLS regression '(SO-PLS) comprising LS-PLS as a special case (with the maximum number of PLS components). Note that this method has nothing to do with the O-PLS method [44] which is also based on some type of orthogonalisation. The SO-PLS method as discussed here is not based on one single optimisation criterion, as is the case, for instance for LS regression but is based on sequential use of PLS regression, with PLS in each step predicting the residuals after fitting of the previous blocks in the sequence. In other words, for each new block in the sequence the best prediction (in the PLS sense) of what is left to explain is obtained. The method thus relies at each step on the well-established properties of PLS regression to give good prediction results, also for collinear data.

The second step in the approach is interpretation of the models. The different SO-PLS models (for each endogenous block) may be interpreted directly using PLS scores and loadings along the same lines as followed in References [25] and [26]. This may be useful for both direct interpretation and outlier detection. In this paper, however, we will put main emphasis on another way of interpreting the results, namely by the use of the simpler PCP method (principal components of prediction, Langsrud and Næs [41]) for post modelling interpretation of established prediction models (in a similar way as done in Reference [45] for regular PLS models). The method takes the predicted **Y**-values as point of departure and uses them as the input to a PCA. The predicted principal components are known to be linear functions of the input data (i.e. linear functions of linear functions) and these functions can easily be found and plotted as discussed below. The advantage of the PCP method is that it reduces the number of components, since it focus directly on the main variation in **Y** that can be explained (see Reference [41] for further arguments and examples). Another advantage in this context is that interpretation can be focused on one plot only for each endogenous block.

A possible drawback with the new approach as compared to one-component traditional PLS path models is that one loses the very nice graphical illustration of results with paths, regression coefficients and explained variances presented in the same plot. In the new approach, the plots to look at will be several (one for each endogenous block), but on the other hand, they will be of the same type as regularly used in PLS regression, with scores, **X**-loadings and **Y**-loadings. Another possible drawback is that certain standard SEM concepts such as 'measurements model' are not explicitly defined. In other words, there is detailed information that is highlighted in other approaches that are not covered here, since we focus on overall tendencies and structures. It may therefore be necessary at a later stage to investigate some of the blocks or some of the relations in more detail.

3. BRIEF DESCRIPTION OF SO-PLS AND PCP

In this section we briefly describe the two main building blocks of the present path modelling approach, SO-PLS [25] and PCP [41], but we refer to the original publications for an even more detailed explanation. The method will here be discussed for two input blocks, but the extension to several blocks follows exactly the same structure. In both cases we will use general regression symbols, **X**, **Y** and **Z**, where **Y** denotes a block of response variables and **X** and **Z** denote two blocks of input variables. The dimensionality of **Y** is N^*p, the dimensionality of **X** is N^*q and the dimensionality of **Z** is N^*r. The below procedure can be used for situations with full column rank in **X** and **Y** as well as for situations which call for data compression by, for instance, PLS. The method works for all constellations of N, p and q, although in situations where the space spanned by the columns of **Z** is already incorporated in the space spanned by the columns of **X,** only **X** will play a part in the modelling (see comment 2 below).

3.1. Basic model and estimation

The SO-PLS method was developed for estimating regression equations with two or several blocks of independent variables, i.e.

$$\mathbf{Y} = \mathbf{XB} + \mathbf{ZC} + \mathbf{E} \quad (1)$$

for the purpose of both prediction and interpretation. The original applications of the method were developed for process modelling where **X** corresponds to processing settings, **Z** to raw

material properties and **Y** to end product properties. In this case, the dimensionality of the matrices **X** and **Z** is typically very different: **X** corresponds to a few (often designed) variables and **Z** will typically represent spectra with several hundred spectroscopic readings. It has been shown [25,26] that for interpretation purposes it is advantageous to treat the two input blocks separately and not jointly by, for instance, a concatenated PLS regression. Other advantages as compared to multi-block PLS of **Y** versus (**X**,**Z**) were mentioned above and will also be discussed below.

The SO-PLS method is based on first fitting **Y** to **X**, and then on fitting the estimated residuals to **Z** after orthogonalisation with respect to **X**, or more precisely orthogonalisation with respect to the extracted PLS components of **X** for the first model. This means that one first 'empties' **X** for information about **Y** and then extracts the predictive information from **Z** orthogonalised with respect to **X**. Note that this can safely be done, since the part of **Z** that can be projected onto **X** has already implicitly been involved in the first model. It will only be the orthogonalised part that will contribute to better explained variance in **Y**.

The main reasons for the orthogonalisation process are that we obtain scale invariance, we can allow for different dimensionality in each block and the procedure is non-iterative. In addition, the space spanned by (**X**,**Z**) and (**X**,$\mathbf{Z}^{\text{orth}}$), where $\mathbf{Z}^{\text{orth}}$ is **Z** orthogonalised with respect to **X**, is the same, so there is no loss of information in the procedure. Therefore, the procedure gives a model which in addition to being efficient (see references on LS-PLS) also has a number of advantages as compared to regular concatenated PLS.

In other words, the idea is to first identify the column space of **X**, i.e. the space spanned by the columns of **X**, that best fits the **Y** by the use of PLS regression. The scores of this regression define the relevant column space of **X**. The next step is to do the same for the **Z** after orthogonalisation with respect to **X** (or more precisely after orthogonalisation with respect to the scores of the first PLS regression). Note that the sum of the rank of the two column spaces of **X** and the orthogonalised **Z** has nothing to do with the column space of **Y**. It may, for instance, often happen when using PLS that several components are extracted from **X** in order to explain one single **Y**.

It should be mentioned that this splitting in two steps based on orthogonalisation is inspired by regular LS fitting where the orthogonalisation ensures that the fitting of **Y** to **X** and **Z** can be done independently for **X** and the orthogonalised **Z** without changing fit and model residuals.

In more detail, the SO-PLS procedure goes as follows (we will use the same nomenclature as used in Reference [1]):

(1) Fit **Y** to **X** by PLS regression, compute the **X**-scores $\mathbf{T}_X$ and **X**- and **Y**-loadings $\mathbf{P}_X$, $\mathbf{Q}_X$ and compute the predictive model $\mathbf{T}_X\mathbf{Q}_X^t$ and residuals $\mathbf{E} = \mathbf{Y} - \mathbf{T}_X\mathbf{Q}_X^t$.
(2) Orthogonalise **Z** with respect to the scores $\mathbf{T}_X$, i.e. compute the orthogonalised **Z** as $\mathbf{Z}^{\text{orth}} = \mathbf{Z} - \mathbf{T}_X(\mathbf{T}_X{}^t\mathbf{T}_X)^{-1}\mathbf{T}_X^t\mathbf{Z}$
(3) Then fit the residuals $\mathbf{E} = \mathbf{Y} - \mathbf{T}_X\mathbf{Q}_X^t$ to $\mathbf{Z}^{\text{orth}}$ using PLS regression and compute scores $\mathbf{T}_Z^{\text{orth}}$and **X**- and **Y**-loadings $\mathbf{P}_Z^{\text{orth}}$, $\mathbf{Q}_Z^{\text{orth}}$ and predictive model $\mathbf{T}_Z^{\text{orth}}\left(\mathbf{Q}_Z^{\text{orth}}\right)^t$.
(4) Compute full predictive model as $\hat{\mathbf{Y}} = \mathbf{T}_X\mathbf{Q}_X{}^t + \mathbf{T}_Z^{\text{orth}}\left(\mathbf{Q}_Z^{\text{orth}}\right)^t$

Comments:

(1) At this step, the regular PLS regression is computed for **Y** versus **X** with the additional calculation of scores, loadings, predicted values and residuals. If the **X** has full column rank (q), and q PLS components are extracted, the PLS solution for **X** reduces to the LS solution for that block. This was the situation that was considered in Reference [25], and was the reason for the original name of LS-PLS.
(2) This is the step where **Z** is orthogonalised with respect to the information in **X** (scores) that is used for prediction of **Y**. It is worth mentioning that the orthogonalisation essentially splits the contribution of **Z** into two, the part that is orthogonal to $\mathbf{T}_X$ (i.e. $\mathbf{Z}^{\text{orth}} = \mathbf{Z} - \mathbf{T}_X(\mathbf{T}_X{}^t\mathbf{T}_X)^{-1}\mathbf{T}_X^t\mathbf{Z}$) and the part that is contained in the column space of $\mathbf{T}_X$ (i.e. $\mathbf{T}_X(\mathbf{T}_X{}^t\mathbf{T}_X)^{-1}\mathbf{T}_X^t\mathbf{Z}$). The latter has here no additional contribution to the fit, since it is already implicitly incorporated in the first regression. Therefore, step 3 in the algorithm focuses only on the part of **Z** that contributes in addition to **X** for explaining variability in **Y**. In this sense it is also meaningful to talk about incremental or additional contribution of the latter. Note that when the maximum number of components (p) can be extracted from **X**, one could in principle orthogonalise with respect to **X** directly, but this is essentially the same as using the procedure described using p PLS scores. When the two column spaces of **X** and **Z** are already orthogonal, the orthogonalisation is unnecessary. If the column space of **Z** is already incorporated in the column space of $\mathbf{T}_X$, the $\mathbf{Z}^{\text{orth}}$ will be identical to 0. This will, for instance, happen when N is smaller or equal to p and the column space of **X** has dimension N. Note that the orthogonalised $\mathbf{Z}^{\text{orth}}$ can also be formulated as the residuals from fitting **Z** to $\mathbf{T}_X$.
(3) Then a new PLS is conducted for the residuals versus the orthogonalised **Z**, with calculation of scores, loadings and predicted values. Note that at this point one could also have fitted **Y** directly to the $\mathbf{Z}^{\text{orth}}$ without changing the model (since $\left(\mathbf{I} - \mathbf{T}_X\left(\mathbf{T}_X^t\mathbf{T}_X\right)^{-1}\mathbf{T}_X^t\right)$ used to obtained the residuals and the orthogonalised **Z** is idempotent).
(4) Here the predicted values from the steps 1 and 3 are added to each other. Note that because of the orthogonalisation in step 2, the scores $\mathbf{T}_X$ and $\mathbf{T}_Z^{\text{orth}}$ are automatically orthogonal, and consequently, the predictive model can automatically be obtained by joining the two equations $\mathbf{T}_X\mathbf{Q}_X^t$ and $\mathbf{T}_Z^{\text{orth}}\left(\mathbf{Q}_Z^{\text{orth}}\right)^t$ as done. Alternative one could regress **Y** onto $\mathbf{T}_X$ and $\mathbf{T}_Z^{\text{orth}}$ simultaneously.

As can be seen, the SO-PLS orthogonalises the second block to the relevant components of the first and then does a separate PLS fitting which can then be done independently because of orthogonality. The only differences between this and regular LS splitting and fitting used as inspiration is that we here orthogonalise based on the relevant information in the first block only and then use PLS as fitting criterion. Note that if it is possible to fit the maximum number of PLS components, the result is the LS solution, showing that our approach is a generalisation of the LS orthogonalisation approach. Note also the similarity between the SO-PLS idea and the Type I sums of squares strategy for analysis of variance.

It should be mentioned that the first step, i.e. fitting of **Y** to **X**, identifies the important dimensions of **X** and leaves most of the noise as residuals. Thus the projection of **Z** is done onto the most stable and reliable part of **X**. The rest of **Z** used for regression will then be filtered the same way leaving the noise primarily in the residuals.

It should also be stressed that the main reasons for orthogonalisation are that it allows for different dimensionality for each block and it provides invariance with respect to the scale

of the different blocks. Another aspect is that it gives rise to a method which is easy to compute. As a by-product it also explicitly provides information about incremental contributions of the blocks which can be important for interpretation (see also the PCP section and the example). The PCP itself can be used for any orthogonalisation and also without as discussed at the end of this section, but it should be remembered when interpreting the model which ortogonalisation is used. In the orthogonalised model, for instance, the interpretation for the block **X** represents the effect of that block only while the interpretation of **Z** represents what is obtained in addition to **X**.

If **Z** is not orthogonalised to **X**, the two become dependent on each other in a regression context. An iterative LS-PLS version without orthogonalisation was proposed in the study by Jørgensen *et al*. [25]. The present approach is simpler, since it does not involve iterations, and is therefore chosen here. For more information about properties and empirical results of SO-PLS, we refer to Jørgensen *et al*. [25,26] and Jørgensen and Næs [46].

For more than two blocks the algorithm continues by adding new points 2, 3 and 4, with orthogonalisation always being done with respect to scores of all preceding blocks. Note that the number of components can be chosen independently for each block leading to optimal parsimony within each block.

If wanted, one can replace the scores by the original measurements in the predictive model:

$$\hat{\mathbf{Y}} = \mathbf{X}\mathbf{V}_X\mathbf{Q}_X{}^t + \mathbf{Z}^{orth}\mathbf{V}_Z^{orth}(\mathbf{Q}_Z^{orth})^t \quad (2)$$

where the **V**s define how the original measurements are used to form the scores (composed of the loadings and the so-called loading weights, see, e.g. Reference [1]).

If one does not like interpretation of the orthogonalised **Z**, which is clearly not in the space spanned by **Z**, it will be shown below that the SO-PLS estimation procedure leads to a model that can easily be formulated in terms of the original variables **X** and **Z** (see equation 3). This model can also be interpreted for each of the blocks, but as also proposed below we focus on one overall interpretation based on the PCP method. As also described this can be done for both the original and the orthogonalised **Z** data. In other words, the results can be presented and interpreted in terms of the original variables and the SO-PLS procedure can be looked upon simply as a method of estimation with the above-mentioned properties.

The process of presenting the model in terms of the original variables can be formulated in the following way:

$$\begin{aligned}\hat{\mathbf{Y}} &= \mathbf{X}\mathbf{V}_X\mathbf{Q}_X{}^t + \mathbf{Z}^{orth}\mathbf{V}_Z^{orth}(\mathbf{Q}_Z^{orth})^t \\ &= \mathbf{X}\mathbf{V}_X\mathbf{Q}_X^t + \left(\mathbf{I}-\mathbf{T}_X(\mathbf{T}_X^t\mathbf{T}_X)^{-1}\mathbf{T}_X^t\right)\mathbf{Z}\mathbf{V}_Z^{orth}(\mathbf{Q}_Z^{orth})^t \\ &= \mathbf{X}\mathbf{V}_X\left(\mathbf{Q}_X^t-(\mathbf{T}_X^t\mathbf{T}_X)^{-1}\mathbf{T}_X^t\mathbf{Z}\mathbf{V}_Z^{orth}(\mathbf{Q}_Z^{orth})^t\right) + \mathbf{Z}\mathbf{V}_Z^{orth}(\mathbf{Q}_Z^{orth})^t \\ &= \mathbf{X}\mathbf{V}_X\mathbf{Q}_X^{*t} + \mathbf{Z}\mathbf{V}_Z^{orth}(\mathbf{Q}_Z^{orth})^t\end{aligned} \quad (3)$$

where the $\mathbf{Q}_X^*$ is the symbol used for the whole parenthesis following $\mathbf{V}_X$. The predicted values as well as the coefficients for **Z** and $\mathbf{Z}^{orth}$ are identical for the two. This means that if wanted, one can interpret the model using PCP in the original measurements in the same way as it is used for interpretation in terms of **X** and $\mathbf{Z}^{orth}$. For the purpose of this paper we focus on one of the possibilities, namely the non-orthogonalised version, i.e. the version based on model (3).

3.2. Validation and determining the number of components

Validation is important both for determining the number of components (A_x and A_z) and for obtaining information about the size of the prediction errors. As usual, cross-validation (CV) or prediction testing are the two simplest and most natural tools to be used for validation of the models developed. Also in this case, either the prediction errors or the predicted R^2s can be used as criterion for the quality of the model. In this paper some attention will also be given to incremental contribution of the blocks, i.e. how much the prediction ability improves when adding a new block.

The cross-validation, as done and recommended in this paper, is based on leaving out one (or, alternatively, blocks of samples) doing the model fitting based on the rest of the data and testing it out on the sample(s) kept out. The only possible complication in this context is the orthogonalisation. It can, however, be observed that $\mathbf{Z}^{orth}$ in the description of the algorithm is the residual matrix obtained from a linear regression of **Z** onto $\mathbf{T}_X$, i.e. $\mathbf{Z}^{orth} = \mathbf{Z}-\hat{\mathbf{Z}}$ where $\hat{\mathbf{Z}} = \mathbf{T}_X\mathrm{D}$ with $\mathrm{D} = (\mathbf{T}_X{}^t\mathbf{T}_X)^{-1}\mathbf{T}_X{}^t\mathbf{Z}$ being the regression coefficients. The $\mathbf{Z}^{orth}$ for a new object can then be obtained as prediction residuals from the same model, i.e. the same **D**.

With two or several input blocks in a model one has basically two different choices when choosing the number of components. The first and probably most natural one is to do a full optimisation of all components simultaneously, i.e. try to find the combination of factors that gives the best overall prediction results (global optimisation). The other possibility is to optimise the number of components in the first block first, then in the second one keeping the first fixed etc. (sequential optimisation). This puts more emphasis on the incremental aspect mentioned above. In the example below, both strategies will be tested. As usual, one should of course favour solutions with as few components as possible.

In some cases, one may end up with a regression solution where the optimal number of components is 0 for one or a number of the blocks. This essentially means that information in this block does not add to the predictive ability of the block of response values (i.e. in the endogenous block considered). The information is already taken care of either by earlier blocks in the sequence or by blocks incorporated later (in the global optimisation). In such cases, the actual block is simply eliminated from the sequence of input blocks. It is worth mentioning that these blocks can be useful for other regression models in the system (i.e. for other endogenous blocks) of regression models, so they are not eliminated from the diagram.

For determining the number of components to use, we chose the same method as proposed in Reference [27], where the prediction ability is given on the vertical axis and the total number of components for the whole model across all **X**-blocks on the other. For each number of components, the different combinations of components for the individual blocks are indicated with numbers. The method is illustrated in Figure 3.

3.3. The PCP method

In the SO-PLS procedure, a number of PLS regressions are performed, each of which produces both regular PLS scores and PLS loadings. The scores and loadings can be used for interpretation and outlier detection. Generally, we recommend such an exercise to be conducted, but here we give main attention to the simpler PCP (principal components of prediction) method for post modelling interpretation of regression models (see References [41] and

[45]). Using this method we will end up with one scores and loadings plot for each endogenous block of data.

The PCP method was developed for improved interpretation of complex multivariate regression equations. It was based on the fact that in many applications of the PLS-2 method rather many components are needed to obtain the best possible prediction results. From an interpretation point of view it is difficult to interpret all these components. It is also questionable whether all these components really represent the relevant dimensionality of the problem or whether they are only needed for obtaining good prediction ability. As an alternative the PCP takes the predicted values of **Y** as the point of departure without any concrete reference of how they are calculated. The method can therefore in principle be used for any type of regression equation.

The first step of the PCP is to use PCA on the predicted values of **Y**. This gives **Y**-scores and **Y**-loadings. Then, since it is known that the predicted values of **Y** are linear functions of the input data and principal components scores are linear functions of the raw data, the principal components scores of the $\widehat{\mathbf{Y}}$ are also linear functions of the input data. Therefore, one can directly express the relation between the independent variables and the principal components of **Y**. These values are as stable as they can be within the PLS paradigm, since the original model is parsimonious and well validated. The method is tested and compared to regular PLS interpretations in Reference [41]. The method is focused on explaining the part of **Y** that can be predicted and how this relates to both **Y** and the input data.

In more detail, $\widehat{\mathbf{Y}}$ can be formulated (see equation (2)) as a linear function of **X** and $\mathbf{Z}^{\text{orth}}$. Since the principal components of $\widehat{\mathbf{Y}}$ can be written as $\mathbf{T}_{\hat{\mathbf{Y}}}=\hat{\mathbf{Y}}\mathbf{P}$ with **P** being the **Y**-loadings, the principal components of prediction can be written as

$$\mathbf{T}_{\hat{\mathbf{Y}}} = \left(\mathbf{X}\mathbf{V}_{\mathbf{X}}\mathbf{Q}_{\mathbf{X}}{}^{t} + \mathbf{Z}^{\text{orth}}\mathbf{V}_{\mathbf{Z}}^{\text{orth}}\left(\mathbf{Q}_{\mathbf{Z}}^{\text{orth}}\right)^{t}\right)\mathbf{P} \tag{4}$$

i.e. as linear functions of the original measurements **X** and $\mathbf{Z}^{\text{orth}}$. The coefficient values are considered as '**X**-loadings' (i.e. input loadings) for the model.

Usually, one will concentrate on the first two components of $\widehat{\mathbf{Y}}$, but as usual, more components are possible if two and two are considered simultaneously. For both the input and output data, the loadings are interpreted in the usual way. The loadings that span the first dimension contribute or are related to the variability in this direction in the score plot. This will be described and commented on in the example. For one-dimensional solutions, there will be only one set of loadings for the input and one for the output data. Note that this situation resembles the original PLS-path modelling situation and that the coefficients can also be plotted for that situation in the geometrical path model as will be shown in the example below (Figure 7).

Note that PCP is independent of the number of components. It only requires that a good model is established. The number of components issue is only relevant for obtaining stable and parsimonious models. This is here obtained by PLS with possibly different components for each block. Interpretation can be done for both the original and the orthogonalised variables (see Section 3.1). In this paper we focus on the former.

4. HOW TO USE THE NEW METHOD

The first step in the procedure proposed here is to establish the 'dependence' diagram. This is a diagram conveying more or less the same information as the path diagram, but in a way that makes it more explicit and clear which blocks that are predicted from which. All the blocks are lined up adjacent to each other and for each of the endogenous blocks, a relation is established (using arrows) with all the blocks that have an effect on it. In addition to representing all the relations, the dependence diagram also shows the order in which the blocks are incorporated in the orthogonalisation process. The diagram is established such that orthogonalisation always takes place from left to right in the diagram.

For orthogonalisation the general rule is that if two input (exogenous) blocks have the same relation to the output (endogenous) block, one can just choose which one to start out with. But if one of the input blocks only has a direct relation to the output and another one both a direct and an indirect effect (through the first), which is the case in, for instance, the second model below, the rule is that the block with only a direct relation is orthogonalised with respect to the other one. The reason for this is that this is the most natural block to consider as additional to the other, i.e. this is the block for which one is usually most interested in considering the additional effect, i.e. how much extra variance that can be explained.

If one is interested in other aspects of additional or incremental variance one can, however, choose otherwise. In particular when there are several equivalent exogenous blocks with the same relation to the output, it may be relevant to try several alternatives. An alternative is to use the so-called parallel LS-PLS modelling suggested by Måge *et al*. [27] for situations where several blocks are introduced in parallel. In most cases, the order of the orthogonalisation will not matter too much when prediction power is concerned, since one always tries to model the residuals the best possible way by using PLS regression. Note that the orthogonalisation is mainly imposed for the purpose of creating invariance and for allowing for different dimensionality, and that the PCP method works equally for all possible ways of orthogonalising the data.

Note that the method as proposed here estimates a separate model for each endogenous block. This means that each model depends only on those blocks that have a path (either direct or indirect) onto the actual endogenous block. An endogenous block is only dependent on and modelled as response for the blocks that are exogenous to it.

A possible dependence diagram for the path model in Figure 1 is presented in Figure 2. As can be seen there are four endogenous blocks, which means that there will be 4 independent models to be estimated. The choice of the block order, following the general rule above, gives rise to the following four models:

Model 1: Block B predicted from block A,
Model 2: Block C predicted from the blocks A and B, with B orthogonalised to A,
Model 3: Block D predicted from the blocks A, B and C, with B orthogonalised to A and C orthogonalised to both A and B,

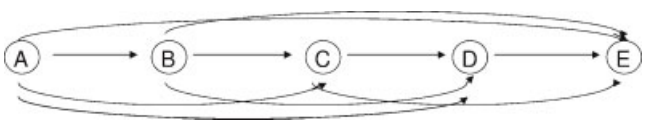

Figure 2. Dependence diagram. This highlights the predicted relations between the different blocks and also illustrates the sequence in which the different blocks are orthogonalised and incorporated.

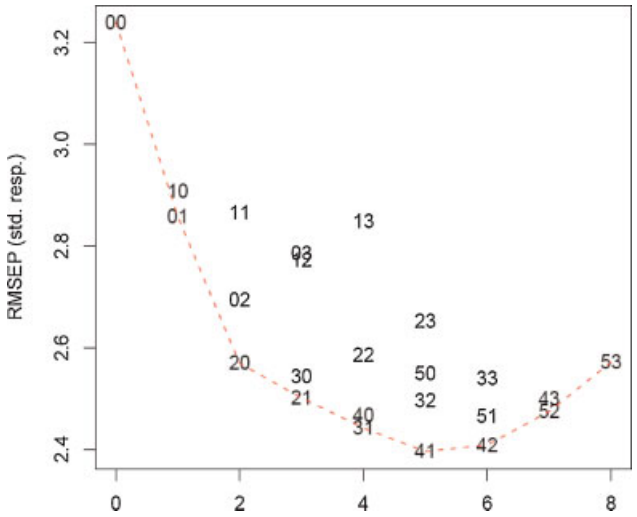

Figure 3. Plot of the root mean squares error of prediction (RMSEP) based on cross-validation (CV). Along the horizontal axis, the total number of components is presented and along the vertical axis the prediction ability is presented. The numbers indicate the number of components in the first block (A) and the second block (B), respectively. As can be seen, the solution with 4 components in the first and 1 components in the second block gives the best results, but the solution 31 is not much poorer.

Model 4: Block E predicted from blocks A, B, C and D also here with orthogonalisation in alphabetic order.

Note that each of the models can be considered a multi-block model and regular multi-block PLS methods [8] could therefore in principle also be used. As stated above, these models do not allow for different dimensionality of the blocks and are dependent on the relative scales of the measurements. Therefore, the SO-PLS is a more natural choice.

In the Section 2 of the paper it was discussed that different dimensions of a block can be used for prediction and to be predicted. In a concrete situation information about this can be obtained by regressing the predicted values when the block is predicted onto the scores for the same endogenous block when it is used for predicting another endogenous block. This will be illustrated for the example below.

5. EXAMPLE

5.1. Data set used

The data used for illustration here is the wine data also used for illustration of regular PLS path modelling in Reference [29]. The data set is based on sensory analysis of 21 French wines (see Reference [28] for a detailed description). A number of sensory attributes are measured in sequence starting with attributes related to 'smell at rest', going via visual appearance attributes, attributes related to odour after shaking, then tasting attributes and ending up with an overall assessment of quality. The average over the panel is used for the study (see also Figure 1).

In more detail, the sensory attributes measured are:

Block A—smell at rest: Rest1, smell intensity at rest; Rest2, aromatic quality at rest; Rest 3, fruity note at rest; Rest 4, floral note at rest; Rest 5, spicy note at rest.

Block B—view: View 1, visual intensity; View 2, shading (from orange to purple); View 3, surface expression.

Block C—smell after shaking: Shaking 1, smell intensity; Shaking 2, smell quality; Shaking 3, fruity note; Shaking 4, floral note; Shaking 5, spicy note; Shaking 6, vegetable note; Shaking 7, phenolic note; Shaking 8, aromatic intensity in mouth; Shaking 9, aromatic persistence in mouth; Shaking 10, aromatic quality in mouth.

Block D—tasting: Tasting 1, intensity of attack; Tasting 2, acidity; Tasting 3, astringency; Tasting 4, alcohol; Tasting 5, balance (acidity, astringency, alcohol); Tasting 6, mellowness; Tasting 7, bitterness; Tasting 8, ending intensity in mouth; Tasting 9, harmony.

Block E—overall quality.

5.2. Results

The main focus of the study (as it also was in Reference [29]) is to investigate how well it is possible to predict attributes measured later based on previous measurements, in particular in which way attributes such as taste and overall quality can be predicted by odour and visual attributes.

The prediction results with the use of cross-validation were evaluated using the same plot as proposed by Måge *et al.* [27]. This plot shows each possible combination of factors for the X-blocks and the corresponding prediction error. An example of such a plot is given in Figure 3 (for model 2). The first digit in the numbers corresponds to the number of components in the A block, and the second digit corresponds to the number of components in the B block. As can be seen from Figure 3, the best results, i.e. the results with the best prediction ability, are obtained by using 4 components in the first block (smell at rest) and 1 component in the second (view). But the result obtained with 3 and 1 components respectively is not much poorer.

The optimal prediction results (global optimisation) together with the corresponding number of components are given in Table I. Table II presents the results obtained by optimising the number of components in the first block first, then fixing this number and optimising the second block etc. (sequential optimisation). As can be seen, the results in Table I are better in terms of prediction ability, except for the two first models where they are identical. The second strategy can, as mentioned above, possibly be somewhat more relevant when there is strong focus on incremental contributions of the blocks. As will be shown below, they may also be useful to look at together.

The difference between the two approaches is very visible for model 3 where the first block vanishes from the first approach

Table I. Explained variances (cross-validation) for the different input matrices in all the four models. Global optimisation of the number of components. Incremental values can be obtained as differences between the percentages given. The values in parentheses are the number of components needed

	Model 1	Model 2	Model 3	Model 4
Block A	37 (1)	42.5 (4)	0.0 (0)	0.0 (0)
Block B		45.3 (1)	41.1 (2)	0.0 (0)
Block C			50.9 (2)	78.4 (2)
Block D				96.5 (3)

Table II. Explained variances (cross-validation) for the different input matrices in all the four models. Sequential optimisation of the number of components. Incremental values can be obtained as differences. The values in parentheses are the number of components needed

	Model 1	Model 2	Model 3	Model 4
Block A	37 (1)	42.5 (4)	22.6 (3)	42.8 (4)
Block B		45.3 (1)	44.4 (2)	55.6 (1)
Block C			44.9 (1)	76.7 (1)
Block D				92.1 (1)

while for the other the optimal number of components is 3. This means that the first block (A-'smell at rest') is important for predicting the taste, but that the information from block A is contained in the other two blocks 'view, block B' and 'small after shaking, block C' when predicting block D. This means that for predicting 'taste', 'smell at rest' has no effect if the other two blocks are measured (when global optimisation of components is used). If sequential optimisation is used, however, we see that both 'smell at rest' and 'view' are important, but that 'smell after shaking' adds little. This means that one can, in addition to 'view', decide to use either 'smell at rest' or 'smell after shaking' with a preference for the latter, since this gives the best overall prediction ability of 'taste'.

As can also be noted, the total number of components can become quite large, which again is an argument for the use of the simpler strategy PCP, which focuses on understanding the dimensionality of the predicted Y.

In model 2 with block C as output and block A and B as input it is primarily the first block (A) which contains the important information while the second one adds little. This means that viewing the wines adds very little to estimating 'smell after shaking' if 'smell at rest' is already measured.

For model 3 with D as output and A, B and C as input, both the 'view' and the 'smell after shaking' are important for predicting taste. The view explains 41.1% while the 'smell after shaking' adds about 10% to this. As can also be seen, the 'smell at rest' adds nothing if 'smell after shaking' is already there (since adding components beyond 0 gives no improvement). Note that 0 components in a matrix means that this matrix is essentially eliminated from the calculations of this particular model.

For model 4 with block E as output and blocks A, B, C and D as input, both the 'smell after shaking' and 'tasting' are important. Note that 'tasting' adds a substantial amount of information (18%). Again 'smell at rest' and also 'view' have no effect if 'smell after shaking' and 'tasting' are present for the globally optimised model.

By considering the amount of information accounted for in the predicted **Y**-values (cross-validated, see Table III), it is obvious that at least the block C (model 2)) is clearly two-dimensional. The two components describe 61% and 20% of the information, respectively.

In order to illustrate the interpretation by the use of PCP, we use model 2 as an example (Figures 4–6, note that the number of components here is the same for both optimisation approaches). In this illustration we use the original input variables (see equation (3)). Along the first axis, Shaking 6 (vegetable note) is seen as opposite to the others. Shaking 2, 3, 8, 9 and 10 have the largest negative values. This corresponds well to the first axis in the PLS path modelling approach (see Reference [29]). The second axis is dominated by the samples 20 and 21 which seem to have more intensity and spiciness for both the 'tasting' attributes and the 'smell after shaking' attributes. This information is not visible in the PLS path modelling approach, since the solution in that case is one-dimensional. No outliers can be spotted in the scores plot.

Table III. Explained variances (in %) of the **Y** for the four models after 1, 2 and 3 principal components (CV). Note that since the predicted **Y** is one-dimensional (see Table Ia), the explained variance of **Y** is 100% after just one component

	Model 1	Model 2	Model 3	Model 4
1 Component	100	61	85	100
2 Component		81	96	
3 Component		92	97	

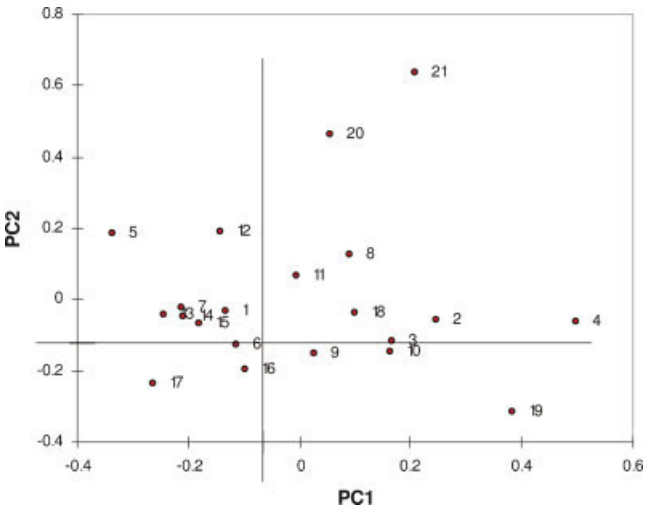

Figure 4. The principal components scores plot of the predicted values for block C (model 2).

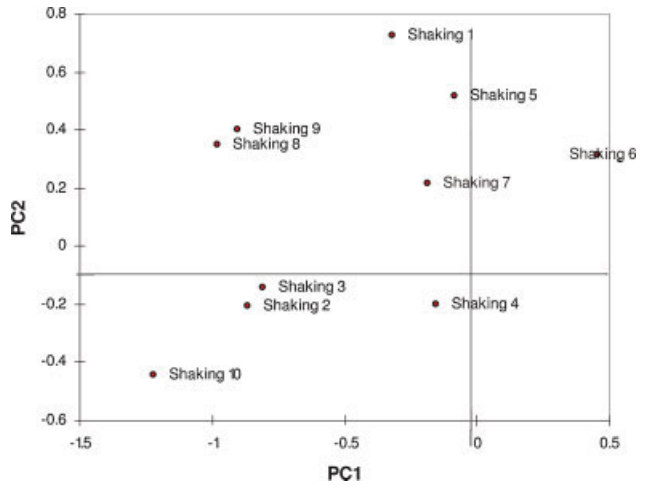

Figure 5. The PCA **Y**-loadings plot for block C (model 2).

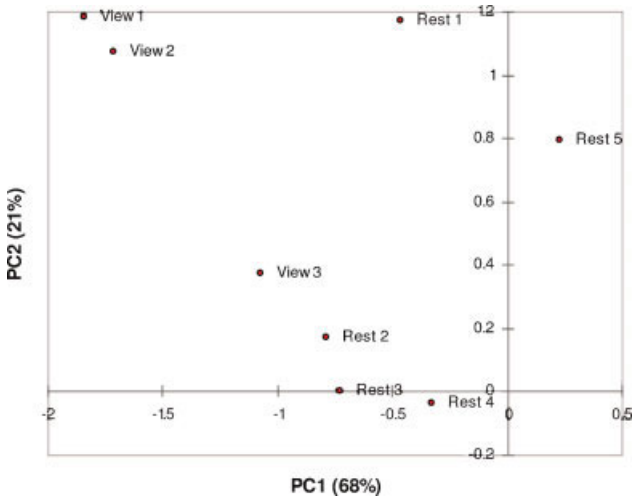

Figure 6. The PCP **X**-loadings plot for the input variables in model 2. (original variables).

Note that the prediction results obtained by this approach cannot be directly related to the PLS path modelling approach, since in that case all predictions are for the latent variables. Our approach is more directly related to the measured variables, since prediction ability is measured with respect to these. Explained variances in Table III are, however, more easily comparable to the prediction abilities given in the PLS path modelling approach.

For one-dimensional $\hat{\mathbf{Y}}$, the loadings can be plotted in the same way as done for standard path modelling PLS. Figure 7 shows the loadings for the 'view' block (Block B) with three manifest variables. As can be seen, the **Y**-loadings are all positive, but of somewhat different size. The latent variables in the standard PLS path modelling approach are scaled to unit variance so the scale of the two approaches cannot be directly related. But as can be seen in Reference [29] also for the PLS path modelling all coefficient values are positive with the first two being the largest. The **X**-loadings have the same sign, with Rest 1 having the largest value, number 2, 3 and 4 having intermediate values and Rest 5 having very small loading value. For the standard PLS path

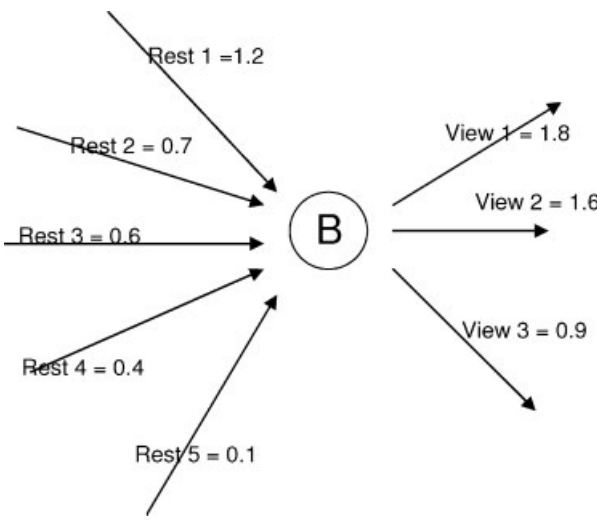

Figure 7. The output and input PCP coefficients for Block B (model 1, one-dimensional PCP).

modelling, the final one was eliminated as non-significant. The first variable is, however, found more important in the PLS-path modelling than here. This may be related to the fact that in standard PLS path modelling, the loadings describe the local variability in the actual block while for the present PCP approach they are only related to prediction. It seems that apart from this, the present approach and the PLS-path modelling approach give similar results for the one-dimensional model for block B.

For illustration we evaluated the overlap between the predicted values of block B (model 1) and the scores for block B when used in the model for predicting block C (model 2). This was done by regular cross-validated PLS regression and no prediction ability was found. Tables I and II show that block 2 contributes in addition to A for predicting C, but this contribution has no relation to the predicted values of B from block A. This shows that the part of block B that can be predicted from A has no overlap with the part of B that adds to predicting C.

6. PROPERTIES OF THE METHOD AND GENERAL DISCUSSION

This section presents a discussion of the main properties of the method with a special focus on relations to established methodologies. It should be mentioned that many of the properties are shared also by established methods (for instance, standard PLS path modelling and the method of Vinzi [35]), but some of the properties are unique for the present proposal.

The new method represents an explorative approach to path modelling based on sequential use of PLS regression for each endogenous block followed by interpretation using the post modelling interpretation tool PCP. The focus is interpretation, but only for models with good prediction ability. The method is explorative, is data driven, requires few assumptions and is as such particularly suitable for an early-stage investigation of a complex system. It may be possible or useful afterwards to follow up by more detailed investigations using, for instance, one of the other methods mentioned above.

The proposed approach is simple to compute and both steps, SO-PLS and PCP, are based on well-established and well-understood methodologies. The three principles used are orthogonalisation, PLS and PCA, which are easily available in most statistical software packages. No convergence problems exist, since only these simple principles are involved. The method of estimation is a sequential PLS approach. This does not have a global optimisation criterion, but is based on sequential PLS prediction of **Y**-residuals from data blocks orthogonalised with respect to those previously used. In this sense, the method is based on sequential PLS optimisation of prediction ability of variability not already accounted for in other blocks. General experience shows that this gives stable and reliable prediction models.

Both scores and loadings are computed for each of the blocks in each SO-PLS model. These can be used for interpretation of the different blocks, how the latent variables relate to the manifest ones and also for detection of outliers. Outlier detection can be done for each of the sequential models following the same ideas as for regular PLS (see, e.g. Reference [1]). The interpretation method focuses here is based on the PCP method which provides one score plot and one loadings plot (**X-** and **Y-**loadings) for each endogenous block. These plots are interpreted in the same way as regular PLS plots. In addition, using the incremental R^2s

one gets information about how much the different blocks contribute.

The new method allows for different dimensionality (different number of latent variables) of the blocks. This means that it can also combine categorical design variables (full rank) in one block with collinear variables in another. This makes the method as parsimonious as possible for each block within the PLS paradigm. This aspect is difficult or impossible to handle by LISREL and also standard PLS path modelling. The method also allows for situations where the variability that can be predicted in a block is not necessarily the same as the part of the same block used for prediction of another. No attempt at finding common dimensions that carry through the system is made (see Section 2). This property is not shared by any of the established methods. The property can be checked in practice as described above. The method is invariant to the relative scaling of the blocks. This property is not shared by the other PLS-based approaches.

The method can be used for many variables and few observations and has no problems with collinearity among the variables within a block. This property is shared by all PLS-based approaches, but not by ML. Variable selection can, if wanted (see Martens [47]), be done for each of the steps in the sequence, but this is a complex point, since orthogonalisation is involved. Therefore, the effect of this is unclear at the present stage of development.

Since PLS can be used for missing value situations, the same is true for the present path modelling approach. This is also the case for the other PLS-based approaches to path modelling. No identification problems exist, since the solution is defined by the algorithm. This is the same for the PLS path modelling approaches, but not for LISREL.

The present approach represents a splitting of the path diagram into separate multi-block regression models with one block of response variables and several blocks of input variables. In this sense it belongs to a certain extent more to the class of multi-block methods than to the class of classical path modelling methods. The class of multi-block methods is large. Most of the methods relate blocks of data to each other without any specific predictive direction (see the introduction for references). These methods could have been used in this context, but we think it is more useful to take the prediction direction into account if it exists. As such, the present method is more similar to the class of multi-block methods which takes the prediction direction into account. A good review of the PLS-based methods in this class can be found in Reference [8].

As already mentioned in the introduction, there are other PLS-based approaches to path modelling; first of all there is the classical method and modifications of it suggested by Vinzi [35] and Tenenhaus and Tenenhaus [40]. These methods are important, but have somewhat different properties as also discussed in Section 2. Other classes of PLS-based methods were suggested by Høskuldsson [22] and Wangen and Kowalski [23]. These are also interesting, but are based on several steps of regression and are therefore less transparent as compared to the methods proposed here. They are all based on modelling all the data in one analysis.

The SO-PLS method used as a basis for the present approach also has some similarities with the ASCA method (Jansen *et al*. [48,49]), where a matrix is fitted to one or several input matrices. Each of the matrices is, after estimation, decomposed by the use of principal component analysis. The main differences lie in the fact that for ASCA, the PCA is used instead of PLS and also that fitting is done simultaneously for all input blocks.

Note that the present approach does not formally distinguish between formative and reflective blocks. All blocks can from one perspective be considered as reflective, since PLS is used for estimation. The number of components can, however, be extended to the maximum, which is natural when a block consists of only a few (for instance, design) variables, and in such cases, the block will essentially be fitted by LS regression. In this case it resembles a formative approach.

Interactions between blocks can, if wanted, be incorporated by adding the products of two blocks as a new block. For instance, all variables in block A can be multiplied by all variables in block B and incorporated as a new block (AB). The new block will most naturally be incorporated after the main blocks and then orthogonalised with respect to them in the analysis. This is an area which requires more research before a clear recommendation can be made. Similar techniques can probably be used also for some of the other approaches.

The method was tested and illustrated on a data set from sensory analysis of wines. One of the blocks was clearly two-dimensional and it has been shown how this can be determined and interpreted. It was also shown that the information that can be predicted within a block is not necessarily identical to the one used for prediction of another block. As could be seen, both the incremental R^2s and the PCP plots were important for interpreting the result. The method worked well both for the block with more than one dimension and for the block with one dimension investigated in detail.

Acknowledgements

We would like to thank NFR (Norwegian Research Council) for financial support through the project 'ConsumerCheck'. Thanks to Jerome Pages and Michel Tenenhaus for providing the data and allowing us to use them. We would also thank the referees for important comments leading to an improved version of the paper.

REFERENCES

1. Martens H, Næs T. *Multivariate Calibration*. John Wiley and Sons: Chichester, UK, 1989.
2. Wright S. On the nature of size factors. *Genetics* 1918; **3**: 174–367.
3. Wright S. The method of path coefficients. *Ann. Math. Stat.* 1934; **5**: 161–215.
4. Wright S. Path coefficients and path regressions: alternative or complementary concepts. *Biometrics* 1960; **16**: 189–202.
5. Mardia KV, Kent JT, Bibby JM. *Multivariate Analysis*. Academic Press: UK, 1979.
6. Tucker LR. Some mathematical notes on three-mode factor analysis. *Psykometrica* 1966; **(31)**: 279–311.
7. Kohler A, Hanafi M, Bertrand D, Oust Janbu A, Møretrø T, Naderstad K, Qannari M, Martens H, New concepts for investigating several sets of data in modern bioscience. In *Modern Concepts in Biomedical Vibrational Spectroscopy*, Lasch P, Kneipp J (eds). Blackwell Publishing: USA, 2009; Chapter 15.
8. Westerhuis JA, Kourti T, MacGregor JF. Analysis of multiblock and hierarchical PCA and PLS models. *J. Chemometr.* 1998; **12**: 301–321.
9. Wold S, Kettaneh N, Tjessem K. Hierarchical multi-block PLS and PC models for easier model interpretation and as alternative to variable selection. *J. Chemometr.* 1996; **10**: 463–482.
10. Carroll JD. Generalisation of canonical analysis to three or more sets of variables *Proceedings of the 76th convention of the American Psychological Association, vol. 3*. 1968; 227–228.
11. Dahl T, Næs T. A bridge between Tucker-1 and Carroll's generalised canonical analysis. *Comput. Stat. Data Anal.* 2006; **50**(11): 3086–3098.

12. Hanafi M, Kiers H. Analysis of K sets of data, with differential emphasis on agreement between and within sets. *Comput. Stat. Data Anal.* 2006; **51**: 1491–1508.
13. Kettenring JR. Canonical analysis of several sets of variables. *Biometrika* 1973; **56**: 433–451.
14. van de Geer JP. Linear relations between k sets of variables. *Psychometrika* 1984; **49**: 79–94.
15. Tenenhaus M, Hanafi M. A bridge between PLS path modelling and multi-block data analysis. In *Handbook of Partial Least Squares Concepts, Methods Applications*, V, Esposito Vinzi WW, Chin J, Henseler H Wang (eds). Springer: Berlin, 2010.
16. Jøreskog KG. Structural analysis of covariance and correlation matrices. *Psychometrika* 1978; **43**: 443–477.
17. Wiley DE. The identification problem in structural equations models with unmeasured variables. In *Structural Equation Models in the Social Sciences* **(Vols 69–83)**, AS, Goldberger OD Duncan (eds). Academic Press: New York, 1973.
18. Bollen KA. *Structural Equations with Latent Variables*. Wiley: NY, 1989.
19. Jøreskog KG, Wold H. *Systems Under Indirect Observations. Part I and Part II*. North Holland Publishing Company: North Holland, Amsterdam, 1982.
20. Tenenhaus M, Vinzi VE, Chatelin Y-M, Lauro C. PLS path modelling. *Comput. Stat. Data Anal.* 2005; **48**: 159–205.
21. Tenenhaus M, Vinzi VE. PLS regression, PLS path modelling and generalised Procrustean analysis: a combined approach for multiblock analysis. *J. of Chemometr.* 2005; **19**: 145–153.
22. Høskuldsson A. Multi-block and path modelling procedures. *J. Chemometr.* 2008; **222**: 571–579.
23. Wangen LE, Kowalski BR. A multiblock partial least squares algorithm for investigating complex chemical systems. *J. Chemometr.* 1989; **3**: 3–20.
24. Wold H. Soft modelling. The basic design and some extensions. In *Systems Under Indirect Observation–Part II*. Jøreskog and Wold (eds). North Holland Publishing Company: North Holland, Amsterdam, 1982; 1–54.
25. Jørgensen K, Segtnan V, Thyholt K, Næs T. A comparison of methods for analysing regression models with both spectral and designed variables. *J. Chemometr.* 2004; **18**(10): 451–464.
26. Jørgensen K, Mevik B-H, Næs T. Combining designed experiments with several blocks of spectroscopic data. *Chemolab* 2007; **88**(2): 143–212.
27. Måge I, Mevik B-H, Næs T. Regression models with process variables and parallel blocks of raw material measurements. *J. Chemometr.* **22**: 2008; 443–456.
28. Escofier B, Pages J. *Analyses factorielles simples et multiples*. Dunod: Paris, 1988.
29. Martens M, Tenenhaus M, Vinzi VE, Martens H. The use of partial least squares methods in new food product development. In *2007 Consumer-led Food Products Development*, MacFie H (ed.). Woodhead Publishing Lmt: Cambridge, England, 2007; 492–523.
30. Loehlin JC. *Latent Variable Models*. Laurance Erlbaum Associates: New Jersey, 2004.
31. Pearl J. *Causality. Models, Reasoning and Inference*. Cambridge University Press: UK, 2000.
32. Vinzi VE, Trinchera L, Amato S. PLS path modeling: from foundations to recent developments and open issues for model assessment and improvement. In *Handbook of Partial Least Squares: Concepts, Methods and Applications*, V, Esposito Vinzi WW, Chin J, Henseler H, Wang (eds). Springer: Berlin, 2010.
33. Skrondahl A, Rabe-Hesketh S. *Generalised Latent Variables Modelling*. Chapman and Hall: UK, 2004.
34. Tenenhaus M. Personal communication about research on PLS path modelling. 2008.
35. Vinzi VE. 2009; A Joint Partial Least Squares Component-based Approach to Structural Equation Modeling and Multi-block Data Analysis. *PLS'09-6th International Conference on Partial Least Squares and Related Methods, BeiHang University*, Beijing, China, 06 Sept. 2009. (Plenary Invited Speaker).
36. Vinzi VE, Russolillo G, Trinchera L. 2009; A joint use of PLS Regression and PLS Path Modelling for a data analysis approach to latent variable modelling *57th Session of the International Statistical Institute (ISI)*, International Statistical Institute (ISI), Durban, South Africa, 18 August 2009 (Invited Paper Session).
37. Kaplan D. *Structural Equation Modelling*. Sage Publications: UK, 2000.
38. Jøreskog KG, Sørbom D. *LISREL 7. A Guide to the Program and Applications*. SPSS, Inc.: Chicago, 1988.
39. Hanafi M. PLS path modelling: computations of latent variables with the estimation model B. *Comput. Stat.* 2007; **22**: 276–292.
40. Tenenhaus A, Tenenhaus M. A PLS approach to regularised generalised canonical correlation analysis. Psykometrika (in press). 2010.
41. Langsrud Ø, Næs T. Optimised score plot by principal components of prediction. *Chemolab* 2003; **68**: 61–74.
42. Wold S, Martens H, Wold H. The multivariate calibration problem in chemistry solved by the PLS method. Proc. Conf. matrix Pencils. (A. Ruhe, B. Kågström ets.). March 1982. *Lecture Notes in Mathematics*, Springer Verlag: Heidelberg, 1983; 286–293.
43. Indahl UG, Næs T. Evaluation of alternative spectral feature extraction methods of textural images for multivariate modelling. *J. Chemometr.* 1998; **12**(4): 261–278.
44. Trygg J, Wold S. Orthogonal projections to latent structures (O-PLS). *J. Chemometr.* 2002; **16**: 119–128.
45. Ergon R. Compression into two-component PLS factorizations. *J. Chemometr.* 2003; **17**(6): 303–312.
46. Jørgensen K, Næs T. The use of LS-PLS for improved understanding, monitoring and prediction of cheese. *Chemolab* 2008; **93**: 11–19.
47. Martens H, Martens M. *Multivariate Analysis of Quality: An Introduction*. Wiley: Chichester, UK, 2001.
48. Jansen J, Hoefsloot J, van der Greef M, Timmerman E, Westerhuis J, Smilde AK. ASCA: analysis of multivaraite data obtained from an experimemtal design. *J. Chemometr.* 2005; **19**(9): 469–481.
49. Jansen J, Bro R, Huub C, Hoefsloot J, van den Berg FWJ, Westerhuis J, Smilde AK. PARAFASCA: ASCA combined with PARAFAC for the analysis of metabolic fingerprinting data. *J. Chemometr.* 2008; **22**: 114–121.