Chemometrics and Intelligent Laboratory Systems, 7 (1989) 53–65
Elsevier Science Publishers B.V., Amsterdam — Printed in The Netherlands

Nonlinear PLS Modeling

SVANTE WOLD *, NOUNA KETTANEH-WOLD and BERT SKAGERBERG
Research Group for Chemometrics, Umeå University, S-901 87 Umeå (Sweden)

(Received 6 October 1988; accepted 1 May 1989)

ABSTRACT

Wold, S., Kettaneh-Wold, N. and Skagerberg, B., 1989. Nonlinear PLS modeling. *Chemometrics and Intelligent Laboratory Systems*, 7: 53–65.

The linear two block predictive PLS model (PPLS2) is often used to model the relation between two data matrices, **X** and **Y**. Applications include multivariate calibration, quantitative structure–activity relationships (QSAR), and process optimization.

In each PPLS2 model dimension the matrices **X** and **Y** are decomposed as bilinear products plus residual matrices:

$$\mathbf{X} = \boldsymbol{tp}' + \mathbf{E}$$

$$\mathbf{Y} = \boldsymbol{uq}' + \mathbf{F}$$

In addition, a linear model is assumed to relate the score vectors $\boldsymbol{t}$ and $\boldsymbol{u}$ ($\boldsymbol{h}$ denotes residuals):

$$\boldsymbol{u} = b\boldsymbol{t} + \boldsymbol{h}$$

This allows **Y** to be modeled by $\boldsymbol{t}$ and $\boldsymbol{q}$ as:

$$\mathbf{Y} = \boldsymbol{tq}'b + \boldsymbol{f}^*$$

In the present work the linear PPLS2 model is extended to the case when the inner model relating the block scores $\boldsymbol{u}$ and $\boldsymbol{t}$ is nonlinear ($\boldsymbol{h}$ is a vector of residuals):

$$\boldsymbol{u} = f(\boldsymbol{t}) + \boldsymbol{h}$$

An algorithm is outlined for the model where the inner relation is a quadratic polynomial:

$$\boldsymbol{u} = c_0 + c_1\boldsymbol{t} + c_2\boldsymbol{t}^2 + \boldsymbol{h}$$

This will be referred to as the QPLS2 model (standing for quadratic PLS with two blocks).

Applications to cosmetics qualimetrics and a drug structure–activity relationship are used as illustrations.

0169-7439/89/$03.50

INTRODUCTION

A common data analytic problem arising in chemistry and biology is that of modeling the relation between two data matrices (data tables), **X** and **Y**, as outlined in Fig. 1. Often the objective is to extract the information in **X** that has predictive power for **Y** [1–4].

Typical examples are found in:

(i) multivariate calibration in analytical chemistry [5–7] and process analytical chemistry [8], where **X** contains signals measured on samples (objects) with concentrations **Y**;

(ii) structure–activity relationships in bioorganic chemistry [9,10], where the structure description matrix **X** of chemical compounds (objects) is used to predict the biological activity matrix **Y**; and

(iii) the optimization of complex processes [11,12], where **X** contains the values of predicting factors of experimental runs (objects). Often the squares and cross terms are included in the **X** matrix. The **Y** matrix contains the values of the dependent variables, responses, of each experimental run.

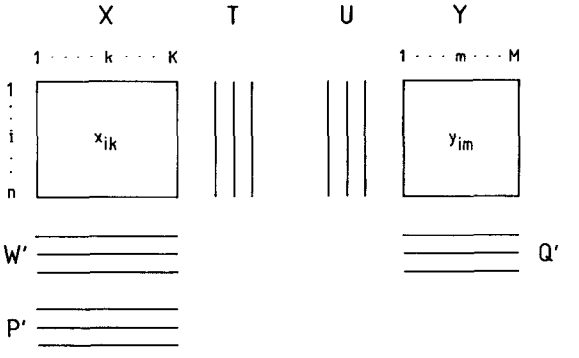

Fig. 1. The matrices and vectors involved in the PPLS2 model. Here the columns and rows correspond to variables (indices k and m) and objects (index i), respectively. Basically, the matrices **X** and **Y** are decomposed as:

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E}$$

$$\mathbf{Y} = \mathbf{UQ}' + \mathbf{F}$$

Here, **E** and **F** are residual matrices. Simultaneously, **U** is related to **T** by the diagonal matrix **B**, giving (**H** is a residual matrix):

$$\mathbf{U} = \mathbf{TB} + \mathbf{H}$$

This gives, in turn, for **Y**:

$$\mathbf{Y} = \mathbf{TQ}'\mathbf{B} + \mathbf{F}^* = \mathbf{TQ}^{*\prime} + \mathbf{F}^*$$

For these problems, the two block predictive PLS model (PPLS2) developed by Wold et al. [2], has been found suitable. PLS is an acronym for partial least squares projection to latent structures.

With few **X**- and **Y**-variables, PPLS2 gives results equivalent to multiple regression. However, PPLS2 is uniquely suitable in cases when collinearities are present in the matrices **X** and **Y**, i.e. the inherent dimensionality of the system studied (here denoted by A), is smaller than the number of measured **X**-variables, K, and/or **Y**-variables, M.

The PPLS2 model is, however, linear in the relation between the two blocks **X** and **Y**. In many applications it is desirable to assume a second order model, i.e. a curved relationship between the blocks. The analytical calibration curve may depart from linearity, and the structure–activity relationship and response surface used for the optimization may have a maximum or a minimum in the domain investigated.

In the present paper we indicate how PPLS2 can be extended to assume a quadratic or nonlinear inner relations between the **X**- and **Y**-blocks. The two block predictive model with a quadratic inner relationship will be referred to as QPLS2. This work represents a preliminary effort to explore the possibilities of nonlinear PLS modeling, and is more a presentation of ideas than a final methodology.

As illustrations we will use a case study from the cosmetic industry and a structure–activity relationship for beta-receptor adrenergic drug candidate compounds. In both cases we will compare the results of the linear PPLS2 with that of PLS using a quadratic inner relation, QPLS2.

PLS MODELS, LINEAR AND NONLINEAR

We first briefly review various two block PLS models.

Linear with linear projections (PPLS2)

The two block, predictive, PLS model (PPLS2) is shown in Fig. 1. The scope of the model is to

find decompositions (projections) of **X** and **Y** as $\mathbf{X} = \mathbf{TP}'$, and $\mathbf{Y} = \mathbf{UQ}'$, respectively, such that:

(a) **X** is well modeled by $\mathbf{TP}'$, and

(b) the matrix **T** predicts **U** well by a linear model, $\mathbf{U} = \mathbf{TB}$ where **B** is a diagonal matrix. Thereby **Y** is linearly modeled by $\mathbf{Y} = \mathbf{TQ}'\mathbf{B}$.

Thus PPLS2 decomposes the **X**- and **Y**-blocks into bilinear models similar to principal components (PC). We note, however, that the PPLS2 model differs from the PC models both statistically and numerically. PPLS2 addresses a different problem than PC analysis, and hence gives different numerical values for the elements of the parameter matrices **T**, **P**, and **Q**.

The dimensionality of the model, i.e. the number of columns in **T** and rows in $\mathbf{P}'$, A, is usually determined by cross-validation.

As mentioned above, the PPLS2 model has been extensively used in various branches of chemistry to model complicated relationships between sets of variables.

Theoretically, the PPLS2 model has been shown to converge towards the multiple regression model when the PLS model dimensionality, A, tends towards the number of X-variables, K [2,16], and to have a number of other interesting statistical properties [16–20].

Linear with quadratic projection of X

Gnanadesikan [21] showed that by extending the **X** matrix with squares (x_k^2) and cross-terms ($x_j * x_k$), the PC projection of the resulting matrix onto a plane in this extended space corresponds to the projection of the original matrix **X** onto a quadratic surface. This result can be extended to PPLS2, and has indeed been used for PLS response surface modeling [3,11,12].

For a limited number of independent **X**-variables (K) this approach is practical for modeling curved relationships. In the real multivariate situation with many and collinear **X**-variables, the approach of Gnanadesikan becomes unwieldy since the number of cross-terms becomes large, the computations become cumbersome, and the results become difficult to interpret.

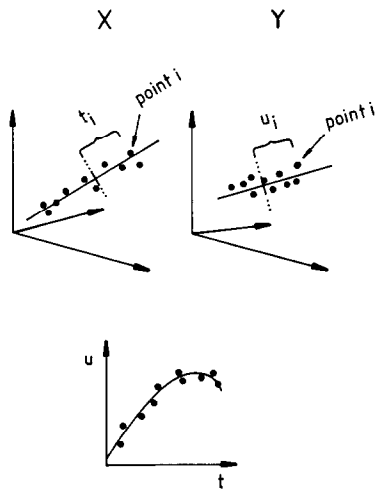

Fig. 2. In the quadratic PLS model, the X-data and Y-data are separately projected first on a line in the **X**- and **Y**-spaces, respectively, by means of the projection vectors $\boldsymbol{p}$ and $\boldsymbol{q}$. The coordinates of each point along the two lines are $\boldsymbol{t}$ and $\boldsymbol{u}$ in **X**- and **Y**-space. These coordinates are related to each other by a quadratic model. Since the score vectors $\boldsymbol{t}_a$ are orthogonal to each other, the first dimension can be peeled off from **X** and **Y** analogously to PPLS2, letting the set of vectors of the next dimension be estimated from the residual matrices $\mathbf{E} = \mathbf{X} - \boldsymbol{tp}'$ and $\mathbf{F} = \mathbf{Y} - (c_0 + c_1 t + c_2 t^2)\,\boldsymbol{q}'$, instead of **X** and **Y**, respectively.

Quadratic inner relation (QPLS2)

To model curved relationships between **X** and **Y**, the projections of **X** and **Y** should be polynomially or nonlinearly related to each other.

The polynomial model is the simplest extension of the PPLS2 model, which in the simplest quadratic form becomes (the third equation relates to each model dimension, a):

$$\mathbf{X} = \mathbf{TP}' + \mathbf{E}$$

$$\mathbf{Y} = \mathbf{UQ}' + \mathbf{F}$$

$$\boldsymbol{u}_a = c_{0a} + c_{1a}\boldsymbol{t}_a + c_{2a}\boldsymbol{t}_a^2 + \boldsymbol{h}_a$$

Geometrically, this corresponds to **X** and **Y** being represented as point swarms in two multidimensional spaces with K and M axes, respectively, and then approximated by lines, planes or hyperplanes so that the projections down on each model dimension (line) in **X** and **Y** space are quadratically related to each other (Fig. 2).

We shall use this simplest quadratic extension of PPLS2, briefly called QPLS2, to illustrate the methodology, and then outline the general case with $\boldsymbol{u}$ as a nonlinear function (continuous and differentiable) in $\boldsymbol{t}$ instead of the third equation above:

$$\boldsymbol{u}_a = f(\boldsymbol{t}_a) + \boldsymbol{h}_a$$

As with PPLS2, the QPLS2 model is 'based' on **X**; the vectors $\boldsymbol{t}_a$ form an orthogonal expansion of **X**. This gives the model desirable properties; one model dimension after another can be estimated and 'peeled off', the model dimensions can be interpreted independently from each other (apart from a rotation), and so on.

We can, of course, use a quadratic projection of **X** if desired, by extending **X** with squares and cross-terms, as outlined above.

We note that the QPLS2 model is inherently different from the linear PLS model with a quadratic projection of the **X**-matrix, and represents a different approximation of the real system.

ESTIMATION ALGORITHM

What follows is an overview of the QPLS2 algorithm, together with the corresponding steps of PPLS2. Details are given in the appendix.

After preprocessing of the data (usually centering and scaling), the algorithm computes the parameters of one model dimension (index a) at a time in an identical way. This rests on the fact that the $\boldsymbol{t}_a$ vectors of different model dimensions are orthogonal.

The QPLS2 algorithm also works when some data are missing, analogously to PPLS2. This is why, below, the division by $\boldsymbol{q}'\boldsymbol{q}$ and $\boldsymbol{w}'\boldsymbol{w}$ is written out explicitly, although it is unnecessary when the data matrices are complete. For matrices with missing data, a sum such as the numerator in $t_i = \{\mathbf{X}\boldsymbol{w}\}_i / \boldsymbol{w}'\boldsymbol{w}$ is calculated only over the existing elements in the ith row of **X**, and the denominator, i.e. $\boldsymbol{w}'\boldsymbol{w}$, is calculated only over the elements corresponding to the ones present in this row of **X**.

Basically, the idea of QPLS2 is to project **X** and **Y** on $\boldsymbol{t}$ and $\hat{\boldsymbol{u}}$ so as to (1) approximate **X** and **Y** as $\boldsymbol{t}\boldsymbol{p}'$ and $\hat{\boldsymbol{u}}\boldsymbol{q}$, respectively; and (2) at the same time fulfil a quadratic inner relation between $\hat{\boldsymbol{u}}$ and $\boldsymbol{t}$.

This is accomplished by starting with a linearly derived PLS weight vector $\boldsymbol{w}$, which is then updated by a Newton–Raphson-like linearization of the quadratic inner relation estimated by linear PLS. At each iteration, ordinary PLS steps are used to update the vectors $\boldsymbol{t}$, $\boldsymbol{q}$, and $\boldsymbol{u}$, and ordinary least squares to estimate the coefficients $\boldsymbol{c}$ of the inner relation.

We see that the **X**-scores $\boldsymbol{t}$ still have the character of a projection of **X**, but the 'projection coefficients' $\boldsymbol{w}$ are derived from the correlation of $\boldsymbol{u}$ with a combination of $\boldsymbol{x}_k$ and the 'quadratic term' $\boldsymbol{x}_k\boldsymbol{t}$.

Thus, for model dimension a (starting with $a = 1$) we have the steps in Table 1.

Cross-validation

Cross-validation can be used in the same way as in ordinary PPLS2 to check the statistical significance of the ath model dimension. Some objects are kept out, the coefficients $\boldsymbol{t}$, $\boldsymbol{w}$, $\boldsymbol{q}$, and $\boldsymbol{c}$ are calculated from the remaining data as indicated above, and finally the y-values for the excluded objects are predicted (see Appendix) from the corresponding x-vectors and the developed model, and the squared differences between predicted and actual y-values are added to the predictive sum of squares, PRESS. This is repeated a number of times until each object has been kept out once and once only.

Cross-validation is even more important with nonlinear PLS than with linear, because of the greater flexibility of nonlinear models and the consequently increased risk of overfitting.

Predictions and confidence intervals

Once the nonlinear PLS model has been computed from a 'training set', predictions can be obtained for new 'test set' objects by inserting their x-values into the model to give t-values, which then, via the inner nonlinear relation, give predicted y-values (see Appendix for details).

Jackknife estimates of the uncertainty of the predicted y-values (confidence intervals) may be computed analogously to the linear model.

TABLE 1

Steps in the QPLS2 and PPLS2 algorithms

QPLS2	PPLS2
0. Start. Use one column in **Y** as a starting vector for u. Get starting vectors for w and t by a linear PPLS2:	*0. Start.* Use one column in **X** as a starting vector for t.
$w' = u'\mathbf{X}/u'u$	
norm w: $\|w\| = 1$	
$t = \mathbf{X}w/w'w$	
1. $X \rightarrow Y$	*1.* $X \rightarrow Y$
Estimate coefficients c of the inner relation by least squares:	In the linear inner relation we have $u = bt + h$ which makes $r = \hat{u} = bt$. Since the constant b is immaterial because of later normalizations, we use directly $r = \hat{u} = t$.
$u = c_0 + c_1 t + c_2 t^2 + h$ $r = \hat{u} = c_0 + c_1 t + c_2 t^2$	
Calculate q from r and **Y**:	Calculate q from r and Y:
$q' = r'\mathbf{Y}/r'r$	$q' = r'\mathbf{Y}/r'r = t'\mathbf{Y}/t't$
norm q: $\|q\| = 1$	norm q: $\|q\| = 1$
2. $Y \rightarrow Y$	*2.* $Y \rightarrow Y$
$u = \mathbf{Y}q/q'q$	Same as QPLS2 step 2.
3. $Y \rightarrow X$ (see appendix)	*3.* $Y \rightarrow X$
Nonlinear regression gives an update of w improving the fit of the quadratic inner relation	$w' = u'\mathbf{X}/u'u$
$u = c_0 + c_1(\mathbf{X}w) + c_2(\mathbf{X}w)^2 + h$	
norm w: $\|w\| = 1$	norm w: $\|w\| = 1$
4. $X \rightarrow X$	*4.* $X \rightarrow X$
$t = \mathbf{X}w/w'w$	Same as QPLS2 step 4.
Check convergence on t or u, e.g.	
$\|t - t_{\text{old}}\|/\|t\| < e$	
5. If no convergence, and if number of iterations < maxiter (here 50), return to step 1, Else, step 6.	5. Same as QPLS2 step 5.
6. Calculate new $r = \hat{u}$ (step 1), q (step 2), u (step 3), and again c (first part of step 1) with the latest t-values	6. Calculate new u (step 2), and the slope, b, of the inner relation ($u = bt + h$).
7. Calculate **X**-loadings:	7. Same as QPLS2 step 7.
$p' = t'\mathbf{X}/t't$	
8. Residuals:	8. Same as QPLS2 step 8.
$\mathbf{E} = \mathbf{X} - tp'$ $\mathbf{F} = \mathbf{Y} - \hat{u}q'$	
Use **E** as **X** and **F** as **Y** in the next model dimension calculations.	

A 'linear' algorithm

As long as the curvature in the inner relation is not too large, i.e. the relation between $\boldsymbol{u}$ and $\boldsymbol{t}$ is monotonic, the PLS-weights $\boldsymbol{w}$ will be very similar for the linear PPLS2 algorithm and the 'nonlinear' QPLS2 algorithm outlined above. This may be utilized to devise a 'quick and dirty' QPLS2 algorithm for such cases, which consists of the standard PPLS2 algorithm until its convergence (step 5), followed by the fitting of a quadratic inner relation in step 6. We call this algorithm the 'linear quadratic' one or L-QPLS2. The same procedure is used by Frank [32].

Other forms of inner relations

We realize that the inner relation can take any functional form as long as the function $f(\boldsymbol{t}) = f(\mathbf{X}\boldsymbol{w})$ is continuous and differentiable with respect to the elements w_k. The only thing that needs to be changed is the details of the nonlinear regression in step 3, and, if $f(\boldsymbol{t})$ is nonlinear in its parameters, also employ nonlinear regression for the fitting of the inner relation (steps 1 and 6).

$$\boldsymbol{u} = f(\boldsymbol{t}) + \boldsymbol{h}$$

Cubic polynomials, exponential and logarithmic functions are commonly employed in scientific and technical models, and in this way 'nonlinear' PLS provides a bridge between the empirical models usually employed in data analysis and 'fundamental' models derived from basic theory.

CHOICE OF MODEL

PLS models represent approximations of an underlying complicated reality, with the purpose of explaining the dependent variables (**Y**) by means of the predictor variables (**X**). A useful model (i) approximates data well in the region of interest, and is (ii) parsimonious and consistent with fundamental knowledge. There is usually a tradeoff between (i) and (ii).

Given the possibility of chosing between a linear and nonlinear PLS model, we see this tradeoff in data sets with mild nonlinearities. As seen in the two present examples, the same degree of fit may be obtained by a linear model with several factors (model dimensions), or by a nonlinear model with fewer dimensions. The nonlinear is more parsimonious, and thereby gives more stable predictions, and is also simpler to interpret; in short, it is preferable.

In cases where the system displays strong nonlinearities with maxima or minima within the experimental domain, the linear model cannot provide a good approximation even with many factors, and therefore only the nonlinear model is appropriate.

In cases where the system investigated is basically linear in its relation between **X** and **Y**, the linear PLS model is adequate and the nonlinear one presents unwarranted complications. In such cases the QPLS2 algorithm gives the coefficient c_2 a value close to zero. This has been confirmed on runs with several different data sets such as the Ketone data [28].

In practice one should always start with the simpler model, here the linear one. If plots of the Y-scores, $\boldsymbol{u}$, against the X-scores, $\boldsymbol{t}$, of the first few model dimensions indicate curvature, and if several model dimensions (three or more) are significant, then a PLS model with quadratic or nonlinear inner relation may be warranted.

In cases where knowledge about the underlying mechanism of the investigated system motivates a nonlinear model, this should of course be tried immediately.

ILLUSTRATIONS

The cosmetics example

The data of this example have been slightly altered so as not to reveal the actual source and type of cosmetic formulations used as illustrations. There are $n = 17$ different formulations of, say, face creams, which are composed of $K = 8$ chemical constituents such as glycerine, water, emulsifier, vaseline, etc., comprising the (17×8) 'composition' matrix, **X**.

These 17 creams have been subjected to a test in which each cream has been applied to one half

TABLE 2

Raw data of the cosmetics example, centered and scaled to unit variance, and rounded to four decimals

Each 'object' occupies three lines. The first eight entries are the x-values and the following eleven are the y-values.

1.7301	−1.0629	0.7277	−0.6592	0.7324	−0.6409	−0.5196	1.4716	−0.319
1.246	1.9574	1.7677	1.7788	0.2649	−0.449	0.1288	0.1475	1.81
−0.8346								
−0.859	0.8351	−0.8685	−0.6592	−1.1656	−0.6409	−0.5196	−0.879	1.7056
−1.5052	−2.3426	−1.9721	0.2086	0.2649	−1.6328	−1.475	0.1475	−1.0344
−0.8346								
0.2506	0.0759	−0.8685	−0.6592	0.7324	1.9533	1.6887	−0.879	−0.319
−0.6044	−1.4142	−0.3173	0.5125	−1.719	0.1225	1.4773	−0.9038	−1.0344
1.3362								
−0.859	1.2147	−0.8685	−0.6592	0.7324	−0.6409	−0.5196	−0.879	0.1148
−1.0426	−1.5608	−1.7735	−1.387	0.2649	−0.5715	−0.5638	0.1475	0.0756
−0.602								
−0.859	1.5943	1.5258	−0.6592	−0.4064	−0.6409	−0.5196	1.2198	1.4887
0.4669	−0.1926	0.1129	0.5125	0.2649	−1.0613	−0.5638	1.1988	−0.8263
−1.3385								
0.9904	−1.0629	0.7277	−0.6592	1.8712	1.9533	−0.5196	0.8	−1.0059
1.5382	0.2716	1.3374	0.0566	0.6403	−1.6328	0.8213	0.1475	0.0756
−0.098								
0.8794	−1.0629	0.3286	0.1272	−0.1786	1.3048	−0.5196	0.3802	−2.5967
−2.1139	0.125	0.7417	0.3605	−1.0756	0.449	1.04	0.6327	0.4919
1.1036								
1.2123	−1.0629	0.7277	−0.6592	0.7324	−0.6409	0.9526	−0.879	0.1148
0.4669	0.5648	−0.5159	−1.0577	0.6403	0.6531	−0.0899	1.1988	−0.5835
0.6385								
−0.859	1.2147	1.5258	−0.6592	−1.9248	−0.6409	−0.5196	0.3802	0.8018
−0.2879	−0.046	−0.7145	0.3605	0.2649	−0.7348	0.8213	0.1475	0.9428
−0.602								
0.9904	−1.0629	0.7277	−0.6592	0.7324	1.3048	−0.5196	−0.879	−0.5721
−0.434	0.7358	0.3115	−1.387	−2.0943	0.3674	−0.7824	1.1988	−1.4506
0.6385								
−0.859	−1.0629	−0.8685	2.29	−1.5452	−0.6409	−0.5196	−0.879	−0.319
−0.434	0.125	−0.3173	−0.9058	−1.0756	0.9797	−1.0376	0.6327	0.2837
0.1345								
0.2506	0.0759	1.5258	−0.6592	0.7324	−0.6409	−0.5196	−0.879	0.5849
1.0756	0.7358	−0.3173	−1.5389	−0.3785	−0.8981	−1.0376	0.1475	−0.3753
−1.3385								
−0.859	0.911	−0.8685	1.3069	0.7324	−0.6409	−0.5196	1.2198	−0.1021
−0.2879	0.4182	0.7417	0.3605	0.6403	0.8981	−0.3451	0.1475	0.4919
0.6385								
−0.859	0.911	−0.8685	1.3069	−0.786	−0.1221	2.4248	1.2198	0.5849
0.6373	0.2716	0.7417	0.6898	1.9808	−0.1225	0.5661	−2.3595	−1.2425
−1.5711								
1.4268	−1.0629	−0.8685	−0.6592	−0.0268	−0.6409	1.6887	−0.879	0.8018
0.9295	−0.3392	−0.5159	−0.2726	−0.0568	1.8369	−0.5638	−0.9038	0.9428
1.6075								
−0.859	0.1518	−0.8685	0.9137	−0.4064	−0.6409	−0.5196	−0.879	−0.3913
−0.2879	0.5648	−0.2511	−0.07	0.533	0.6531	−0.5658	0.1475	−0.3772
0.9873								
−0.859	0.4555	−0.8685	1.3069	−0.5582	0.6562	−0.5196	1.2198	−0.5721
0.6373	0.125	0.9403	1.7788	0.6403	1.143	2.1699	−1.8743	1.81
0.1345								

of the face of ten women models, while at the same time a 'standard cream' has been applied to the other half of the face. Trained evaluators, together with the women models, have given their opinions about $M = 11$ different 'quality indicators' such as 'ease of application', 'greasiness', 'skin smoothness', 'skin shininess', 'overall appeal', etc., all relative to the 'standard cream'. The results for each cream are averaged over the ten women models to give a (17×11) 'quality' matrix, **Y**.

The raw data in centered and scaled form are given in Table 2.

This example is typical of the 'qualimetrics' of complicated products, where a single response variable expressing the quality of the product is difficult to measure, but a battery of 'quality indicators' together may contain the information regarding quality.

The data analytical problem is to develop a model relating the cream composition (**X**) to the 'quality indicators' (**Y**), in order to be able to predict 'quality' (**Y**) from cream composition (**X**). This may allow the formulation of an 'optimal' cream by choosing the appropriate composition.

Results

Linear PLS

Applying linear PLS resulted in a four component model (significant according to cross-validation), describing about 54% of the variance of **Y**. The two first components are fairly strong, each explaining about 17%, while the last two components were weaker, each explaining about 11% of the **Y** variance. The explained variance for each of the 11 y-variables is shown in Table 4.

Plotting the Y-scores ($\boldsymbol{u}$) against the X-scores ($\boldsymbol{t}$) for the first two components reveal curvature in the first dimension (Fig. 3) but none in the second.

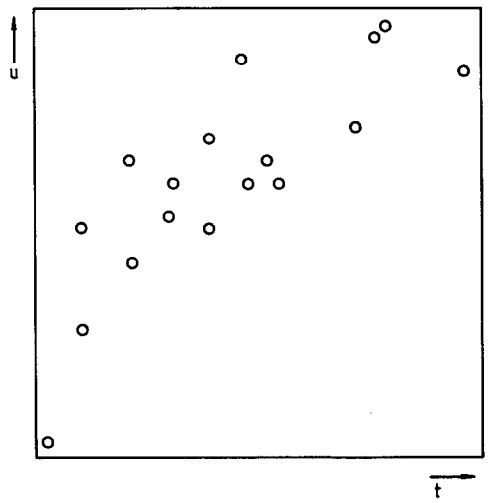

Fig. 3. First dimension Y-scores ($\boldsymbol{u}_1$) plotted against first dimension X-scores ($\boldsymbol{t}_1$) for the PPLS2 model of the cosmetic data.

QPLS2

The quadratic PLS model gave two strong components only, which together explain 52% of the **Y** variance in a way very similar to the four component PPLS2 model (see e.g. Table 3). This shows that the last two components of the linear PPLS2 model are merely compensating for nonlinearity.

This is also seen in that the first score vector, $\boldsymbol{t}_1$, is almost identical for the QPLS2 and PPLS2 models, while there are somewhat larger differences in the second dimension. The $\boldsymbol{u}$-vectors for the first two dimensions are virtually identical for the two models. Also the loadings and PLS-weights ($\boldsymbol{w}$) are very similar in the first two dimensions for the two models (Table 4).

TABLE 3

Explained variance of the eleven y-variables for the linear PLS model with four components (PPLS2) and the quadratic PLS model with two components (QPLS2). The last row shows the results of the 'linear' algorithm for QPLS2

	1	2	3	4	5	6	7	8	9	10	11	total
PPLS2	0.45	0.01	0.56	0.91	0.44	0.18	0.31	0.42	0.49	0.11	0.35	0.54
QPLS2	0.45	0.16	0.39	0.89	0.52	0.32	0.26	0.32	0.12	0.16	0.43	0.52
LIN-Q	0.45	0.17	0.43	0.91	0.54	0.39	0.11	0.37	0.20	0.15	0.30	0.52

TABLE 4

Resulting PLS-weights (w) and inner relation coefficients (c) for, respectively, the QPLS2-algorithm (nl), the L-QPLS2 (lnl), and the PPLS2 algorithm (l) for the first two dimensions for the cosmetic example

	1	2	3	4	5	6	7	8
w1–nl	−0.404	0.558	0.061	−0.419	−0.294	−0.376	−0.205	−0.277
w1–lnl	−0.459	0.527	−0.128	−0.199	−0.321	−0.437	0.052	−0.399
w1–l	−0.505	0.580	−0.141	−0.219	−0.353	−0.480	0.057	−0.439
w2–nl	−0.149	0.306	0.169	0.186	−0.126	−0.092	−0.056	0.891
lnl	−0.335	0.313	−0.079	0.470	−0.206	−0.008	−0.029	0.760
l	−0.234	0.304	−0.146	0.397	−0.222	−0.008	0.163	0.807
c1–nl	−0.325	0.969	0.209					
c1–lnl	−0.39	1.113	0.219					
c2–nl	0.052	1.255	−0.041					
c2–lnl	0.027	0.990	−0.020					

L-QPLS2

The 'linear quadratic' algorithm in this case gives very similar results to the 'nonlinear quadratic' algorithm both for parameters and explained variances (see Table 3). This can be understood in the light of the limited curvature in the first dimension which makes the PLS-weights, w, very similar in the two algorithms (see Table 4).

The beta-receptor agonist example

The data of this example are taken from Mukherjee et al. [13] who made 37 substituted phenylethylamines, of which 15 turned out to be agonists to the beta-adrenergic receptor.

Dunn et al. [14], and later Hellberg et al. [15] formulated structure–activity relationships for these compounds. In this, they described the structural variation among the $N = 15$ agonists by $K = 6$ descriptor variables such as the partial lipophilicities and Hammett sigmas of the varying substituents. This means that **X** has the dimensions (15×6).

The biological activity of the agonist compounds was reported in ref. 13 as: (i) the concentration needed to cause half-maximal adenyl cyclase stimulation; (ii) the competitive binding to the beta-adrenergic receptor relative to Halprenolol; and (iii) the intrinsic activity measured as the maximum enzyme stimulation relative to isoproterenol. Hence the **Y** matrix is (15×3).

By developing a model relating **X** to **Y**, a so-called quantitative structure–activity relationship (QSAR), one may understand which structural factors influence the biological activity **Y**, and how to change them to move **Y** in a desired direction (drug optimization).

Results

PPLS2

The earlier analysis by linear PLS, i.e. PPLS2, resulted in three model dimensions significant according to cross-validation [15]. The variance of the three y-variables explained by this model is 87, 85, and 72%, respectively. A plot of $\boldsymbol{u}_1$ against $\boldsymbol{t}_1$ indicates a mild positive curvature [14,15].

QPLS2

The PLS model with quadratic inner relation has two strong components, which together explain 87, 88, and 72% of the three y-variables, respectively. Hence, QPLS2 gives a more parsimonious model.

There is nonlinearity in both components as shown by the coefficients (c_0–c_2) of the inner quadratic relation. These are for component one

and two: (−0.263, 0.821, 0.127) and (−0.263, 0.955, 0.330), respectively.

L-QPLS2

With the 'linear-quadratic' algorithm, two components explain 87, 88, and 72% of y_1 to y_3, respectively, which is the same as the quadratic algorithm. The c-values become a little different, (−0.588, 1.244, 0.231) and (−0.156, 0.955, 0.209), showing that there is some flexibility in a two-dimensional quadratic model in approximating a given data set.

MISCELLANEOUS

Design

Statistical designs in latent variables are essential to get stable PPLS2 models with good predictive power in the domain spanned by the design [10,23–25]. We realize that statistical designs are essential for quadratic and nonlinear PLS models, and that designs such as composite designs [26,27] in the latent variables are appropriate to support quadratic inner relations.

Missing data

Since all steps of the QPLS2 algorithm involve projections, i.e. partial least squares parameter estimations [22], they work in the ordinary way also when **X** and **Y** have missing observations, as long as the pattern of 'holes' is not strongly nonrandom.

DISCUSSION

In the present paper, we have indicated the possibility of extending the linear two block predictive PLS model to a model with quadratic or nonlinear inner relations. The resulting estimation problem becomes nonlinear in one step of the problem (No. 3 above). Therefore we call this mode nonlinear PLS, although we have so far only made use of models with quadratic (i.e. still statistically linear) inner relationships.

For data where the connection between **X** and **Y** is nonlinear the proposed PLS models show some advantages. They give more parsimonious models, which are easier to interpret, and which are more stable in their predictions. This is seen in the two examples, where two curved components explain as much **Y**-variance as four and three linear components, respectively.

In multivariate analysis, linear and bilinear models have so far been the rule. These models can approximate a system locally, as long as higher terms in the Taylor expansion of the underlying function are negligible. With the availability of nonlinear models, the domain where a single model is applicable increases in size, but we must remember that any model remains an approximation in a limited domain of a more complicated, and usually highly nonlinear, reality.

Typical areas where nonlinear PLS will apply are process optimization and quantitative structure–activity relationships. In such applications there is often a maximum or a minimum in the domain of interest, and a quadratic or higher polynomial model is necessary.

With the present interest in nonlinear dynamic systems, often showing 'chaotic' and fractal behaviour [30], we note that PLS models have been shown to apply in time series analysis [31]. Thus, nonlinear PLS models may be applicable to the modeling of chaotic behaviour, or at least to more well-behaved nonlinear dynamic systems.

Much work remains to make nonlinear PLS a routine method. Our experience and understanding of the properties of nonlinear models must be much increased.

In the linear PLS model (PPLS2), the first score vector, t_1, is an eigenvector of the matrix $\mathbf{XX'YY'}$, the second, t_2, to $\mathbf{EE'YY'}$ (**E** is the first residual **X**-matrix), and so on [16,17,29]. It would be interesting to find out whether there is any eigenvalue problem that also underlies the nonlinear PLS models.

The present algorithm is fairly complicated, and converges slowly when the data lack structure. Hopefully this situation may be improved by better algorithms.

For nonlinear PLS models much research is needed on modeling diagnostics and statistically

important problems connected with inference, such as confidence intervals of parameters, and prediction errors.

ACKNOWLEDGEMENTS

We are grateful for financial support from the Swedish Natural Science Research Council (NFR), and the National Swedish Board for Technical Development (STU).

APPENDIX

Here we give the QPLS2 algorithm in more detail. We note that this is the first version of the algorithm. It works fairly well, but certainly can and will be improved.

As in all multivariate data analysis by means of projections, the results depend on the scaling of the data. Hence, the first step of the analysis is the appropriate transformation and scaling of the data. This is followed by the estimation of one model dimension (index a) at a time. This is possible because of the orthogonality of the subsequent t-vectors. This orthogonality follows from the form of steps C7 and C8 below as shown by Höskuldsson [17].

Thus the algorithm goes as follows:

(A) Transform and scale the matrices **X** and **Y** according to the prior knowledge of the situation. If no such prior knowledge is available, center **X** and **Y** by subtracting column means, and scale the columns in **X** and **Y** to unit variance by dividing by their standard deviation (autoscaling).

(B) Set the dimension index, a, to zero.

(C0) Increment a: $a = a + 1$.
Find a set of starting vectors $\boldsymbol{w}$, $\boldsymbol{t}$, and $\boldsymbol{u}$, by means of one interation in the linear PPLS2 model:
(i) take $\boldsymbol{u}$ as the Y-column with max variance.
(ii) $\boldsymbol{w}' = \boldsymbol{u}'\mathbf{X}/\boldsymbol{u}'\boldsymbol{u}$ normed to length 1.0
(iii) $\boldsymbol{t} = \mathbf{X}\boldsymbol{w}/\boldsymbol{w}'\boldsymbol{w}$

(C1) Estimate the coefficients c by least squares:

$$\boldsymbol{u} = c_0 + c_1\boldsymbol{t} + c_2\boldsymbol{t}^2 + \boldsymbol{e}$$

Denote the estimated $\boldsymbol{u}$ ('$\boldsymbol{u}$-hat') by $\boldsymbol{r}$:

$$\boldsymbol{r} = c_0 + c_1\boldsymbol{t} + c_2\boldsymbol{t}^2$$

(C1-B) Y-loadings, $\boldsymbol{q}$:

$$\boldsymbol{q}' = \boldsymbol{r}'\mathbf{Y}/\boldsymbol{r}'\boldsymbol{r}$$

Norm $\boldsymbol{q}$ to length one.

(C2) New $\boldsymbol{u}$-values:

$$\boldsymbol{u} = \mathbf{Y}\boldsymbol{q}/\boldsymbol{q}'\boldsymbol{q}$$

With the new $\boldsymbol{u}$-values, recalculate the coefficients in the inner relation, c_0, c_1, and c_2 by least squares.

(C3) Calculate corrections to $\boldsymbol{w}$.
We have:

$$\begin{aligned}\boldsymbol{u} &= c_0 + c_1\boldsymbol{t} + c_2\boldsymbol{t}^2 + \boldsymbol{e}\\ &= c_0 + c_1\mathbf{X}\boldsymbol{w} + c_2(\mathbf{X}\boldsymbol{w})^2 + \boldsymbol{e}\end{aligned}$$

Rewriting this as $\boldsymbol{u} = F(\mathbf{X}, \boldsymbol{w}, \boldsymbol{c})$ and linearizing with respect to all the unknown parameters ($\boldsymbol{c}$ and $\boldsymbol{w}$), we get around the point F_{00}:

$$\begin{aligned}u_i &= F_{00} + \mathrm{Sum}_j \Delta c_j \, \mathrm{d}F/\mathrm{d}c_j \\ &\quad + \mathrm{Sum}_k \Delta w_k \, \mathrm{d}F/\mathrm{d}w_k \\ &= c_{00} + c_{01}t_i + c_{02}t_i^2 \\ &\quad + \Delta c_0 + \Delta c_1 t_i + \mathrm{Sum}_k[c_1 x_{ik} \, \mathrm{d}w_k] \\ &\quad + \Delta c_2 t_i^2 + \mathrm{Sum}_k[2c_2 t_i x_{ik} \, \mathrm{d}w_k]\end{aligned}$$

Combining c_{00} with Δc_0 to give a new c_0, etc., we have:

$$\begin{aligned}u_i &= c_0 + c_1 t_i + c_2 t_i^2 \\ &\quad + \mathrm{Sum}_k[(c_1 + 2c_2 t_i)x_{ik} \, \mathrm{d}w_i]\end{aligned}$$

This equation is used for updating $\boldsymbol{w}$ as follows:

(a) calculate the corrections $\mathrm{d}w_k$ by one dimension of linear PLS (here also the cs are treated as unknown parameters, but they are not used thereafter):
(i) Construct the matrix **Z** with its first K columns equalling $(c_1 + 2c_2 t_i)x_{ik}$ and its last three

columns 1, t_i, and t_i^2, respectively.

(ii) $\boldsymbol{v}' = \boldsymbol{u}'\mathbf{Z}/\boldsymbol{u}'\boldsymbol{u}$; normed to length 1.0

(iii) $\boldsymbol{s} = \mathbf{Z}\boldsymbol{v}/\boldsymbol{v}'\boldsymbol{v}$

(iv) $b = \boldsymbol{s}'\boldsymbol{u}/\boldsymbol{s}'\boldsymbol{s}$

(v) $\mathrm{d}\boldsymbol{w} = b\boldsymbol{v}$ (first K elements)

(b) Update $\boldsymbol{w}$ as: $\boldsymbol{w} = \boldsymbol{w} + \mathrm{d}\boldsymbol{w}$ (normed)

(C4) Calculate $\boldsymbol{t}$

$$\boldsymbol{t} = \mathbf{X}\boldsymbol{w}/\boldsymbol{w}'\boldsymbol{w}$$

(C5) Check convergence on

$$\| \boldsymbol{u}_{\text{new}} - \boldsymbol{u}_{\text{old}} \| / \| \boldsymbol{u}_{\text{old}} \| < 10^{-10}$$

If convergence, continue with (C6)
If no. of iterations > maxiter (50 or 100), go to (C6)
Else return to (C1)

(C6) Calculate final values of $\boldsymbol{r} = \hat{\boldsymbol{u}}$ (step C1), $\boldsymbol{q}$ (step C2), $\boldsymbol{u}$ (step C3), and $\boldsymbol{c}$ (first part of step C1) from the latest t-values.

(C7) Loadings $\boldsymbol{p}$

$$\boldsymbol{p}' = \boldsymbol{t}'\mathbf{X}/\boldsymbol{t}'\boldsymbol{t}$$

(C8) Residuals:

$$\mathbf{E} = \mathbf{X} - \boldsymbol{t}\boldsymbol{p}'$$

$$\mathbf{F} = \mathbf{Y} - \boldsymbol{r}\boldsymbol{q}'$$

Use residuals $\mathbf{E}$ and $\mathbf{F}$ as $\mathbf{X}$ and $\mathbf{Y}$ in next dimension. Return to step (C0).

Predictions of y for a new vector x

This is made analogously to PPLS2 as follows:

(1) Preprocess the vector $\boldsymbol{x}$ in the same way as the training set (e.g. subtract previously calculated means and multiply by scaling weights).
Set the predicted y-values equal to 0.0

(2) For each model dimension (a):
Calculate $\boldsymbol{t} = \boldsymbol{x}\boldsymbol{w}_a$ If $\boldsymbol{x}$ has missing data, use instead:

$$\boldsymbol{t} = \boldsymbol{x}\boldsymbol{p}_a/\boldsymbol{p}_a'\boldsymbol{p}_a$$

Calculate $\hat{\boldsymbol{u}} = \boldsymbol{r} = c_0 + c_1\boldsymbol{t} + c_2\boldsymbol{t}^2$
Update the predicted y-values by adding $\boldsymbol{r}\boldsymbol{q}_a'$
Form x-residuals as $\boldsymbol{x} - \boldsymbol{t}\boldsymbol{p}_a'$ to use as $\boldsymbol{x}$ in next dimension

Comments: Step C3 corresponds to the ordinary PPLS2 step $\boldsymbol{w}' = \boldsymbol{u}'\mathbf{X}/\boldsymbol{u}'\boldsymbol{u}$ ($\boldsymbol{w}$ normed to length one).

By the NIPALS interpretation [22] this PPLS2 step can be seen as regression each X-column on $\boldsymbol{u}$. In QPLS2 this is not always possible, since $\boldsymbol{u}$ is a function of $\boldsymbol{t}$, but $\boldsymbol{t}$ is not necessarily a function of $\boldsymbol{u}$. Hence, instead the elements in $\boldsymbol{w}$ are updated to make $\boldsymbol{t}$ a better fit to $\boldsymbol{u}$ in the quadratic relation. This leads to a nonlinear problem, which is solved by linearization and a PLS1 estimation.

We realize that instead of the quadratic relation between $\boldsymbol{u}$ and $\boldsymbol{t}$ used above, we can have any functional relation $\boldsymbol{u} = f(\boldsymbol{t})$ as the inner relation in nonlinear PLS, as long as it can be differentiated with respect to $\boldsymbol{t}$, and thereby with respect to w_k.

REFERENCES

1 S. Wold, C. Albano, W.J. Dunn III, K. Esbensen, S. Hellberg, E. Johansson and M. Sjöström, Pattern recognition: finding and using patterns in multivariate data, in H. Martens and H. Russwurm Jr. (Editors), *Food Research and Data Analysis*, Applied Science Publishers, London, 1983, p. 147.

2 S. Wold, A. Ruhe, H. Wold and W.J. Dunn III, The collinearity problem in linear regression. The partial least squares approach to generalized inverses, *SIAM Journal of the Science of Statistics and Computation*, 5 (1984) 735–743.

3 S. Wold, C. Albano, W.J. Dunn III, U. Edlund, K. Esbensen, P. Geladi, S. Hellberg, E. Johansson, W. Lindberg and M. Sjöström, Multivariate data analysis in chemistry, in B.R. Kowalski (Editor), *Chemometrics. Mathematics and Statistics in Chemistry*, Reidel, Dordrecht, 1984, p. 17.

4 P. Geladi and B.R. Kowalski, Partial least squares regression (PLS): a tutorial, *Analytica Chimica Acta*, 185 (1986) 1–17.

5 W. Lindberg, J.-Å. Persson and S. Wold, Partial least-squares method for spectrofluorimetric analysis of mixtures of humic acid and ligninsulfonate, *Analytical Chemistry*, 55 (1983) 643–648.

6 H. Martens, *Multivariate Calibration*, Thesis, Technical University of Norway, Trondheim, 1985.

7 H. Martens and T. Naes, Multivariate calibration, I and II, *Trends in Analytical Chemistry*, 3 (1984) 204 and 266.

8 A. Lorber and B.R. Kowalski, The effect of interferences and calibration design on accuracy: implications for sensor and sample selection, *Journal of Chemometrics*, 2 (1988) 67–79.

9 W.J. Dunn III, S. Wold, U. Edlund, S. Hellberg and J. Gasteiger, Multivariate structure–activity relationships between data from a battery of biological tests and en ensem-

ble of structure descriptors: the PLS method, *Quantitative Structure–Activity Relationships*, 3 (1984) 131–137.
10 S. Hellberg, *A. Multivariate Approach to QSAR*, Thesis, Umeå University, 1986.
11 S. Clementi, G. Cruciani, G. Curti and B. Skagerberg, PLS response surface optimization: The Carso procedure, *Journal of Chemometrics*, (1989) in press.
12 O.M. Kvalheim, Latent-structure decompositions (projections) of multivariate data, *Chemometrics and Intelligent Laboratory Systems*, 2 (1987) 283–290.
13 C. Mukherjee, M.C. Caron, D. Mulliken and R.J. Lefkowitz, *Molecular Pharmacology*, 12 (1976) 16.
14 W.J. Dunn, S. Wold and Y. Martin, Structure–activity study of beta-adrenergic agents using the SIMCA method of pattern recognition, *Journal Medicinal Chemistry*, 21 (1978) 922–930.
15 S. Hellberg, S. Wold and W.J. Dunn, Multivariate quantitative structure–activity relationships of beta-adrenergic agents using the MACUP method, in A. Höskuldson and K. Esbensen (Editors), *Proceedings of a Symposium on Applied Statistics held in Copenhagen 1982*, NEUCC and RECKU, Copenhagen, 1982, pp. 55–85.
16 R. Manne, Analysis of two partial least squares algorithms for multivariate calibration, *Chemometrics and Intelligent Laboratory Systems*, 2 (1987) 187–197.
17 A. Höskuldsson, PLS regression methods, *Journal of Chemometrics*, 2 (1988) 211–228.
18 T. Naes and H. Martens, Comparison of prediction methods for multicollinear data, *Communications on Statistics, Simulation and Computing*, 14 (1985) 545–576.
19 I.S. Helland, The structure of partial least squares regression, *Communications on Statistics, Simulation and Computing*, 17 (1988) 581–607.
20 A. Lorber, L.E. Wangen and B.R. Kowalski, A theoretical foundation for the PLS algorithm, *Journal of Chemometrics*, 1 (1987) 19–31.
21 R. Gnanadesikan, *Methods for Statistical Data Analysis of Multivariate Observations*, Wiley, New York, 1977.
22 H. Wold, Nonlinear estimation by iterative least squares procedures, in F. David (Editor), *Research Papers in Statistics. Festschrift for Jerzy Neuman*, Wiley, New York, 1966, p. 411.
23 S. Wold, M. Sjöström, R. Carlson, T. Lundstedt, S. Hellberg, B. Skagerberg, C. Wikström and J. Öhman, Multivariate design, *Analytica Chimica Acta*, 191 (1986) 17–32.
24 R. Carlson, T. Lundstedt and C. Albano, Screening of suitable solvents in organic synthesis. Strategies for solvent selection, *Acta Chemica Scandinavica*, B39 (1985) 79.
25 J. Jonsson, L. Eriksson, M. Sjöström, S. Wold and M.L. Tosato, A strategy for ranking environmentally occurring chemicals, *Chemometrics and Intelligent Laboratory Systems*, 5 (1989) 169–186.
26 G.E.P. Box, W.G. Hunter and J.S. Hunter, *Statistics for Experimenters*, Wiley, New York, 1978.
27 G.E.P. Box and N.R. Draper, *Empirical Model-Building and Response Surfaces*, Wiley, New York, 1986.
28 *Simca-3B. Program Manual*, Institute of Chemistry, Umeå University, 1985.
29 S. Wold, M. Sjöström and S. Hellberg, Chemometrics: multivariate analysis and design, *International Statistical Institute Bulletin*, 46 (1988) Vol. 4, 477–495.
30 J.M.T. Thompson and H.B. Stewart, *Nonlinear Dynamics and Chaos*, Wiley, New York, 1986.
31 N.L. Ricker, The use of biased least squares estimators for parameters in discrete-time pulse-response models, *Industrial and Engineering Chemistry Research*, 27 (1988) 343–350.
32 I. Frank, Nonlinear PLS model, Talk given at the IX ICCCRE, Lago del Garda, Italy, June 1989.