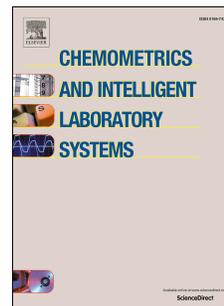


Journal Pre-proof

Residual spaces in latent variables model inversion and their impact in the design space for given quality characteristics

S. Ruiz, L.A. Sarabia, M.C. Ortiz, M.S. Sánchez



PII: S0169-7439(19)30469-1

DOI: <https://doi.org/10.1016/j.chemolab.2020.104040>

Reference: CHEMOM 104040

To appear in: *Chemometrics and Intelligent Laboratory Systems*

Received Date: 25 July 2019

Revised Date: 30 April 2020

Accepted Date: 3 May 2020

Please cite this article as: S. Ruiz, L.A. Sarabia, M.C. Ortiz, M.S. Sánchez, Residual spaces in latent variables model inversion and their impact in the design space for given quality characteristics, *Chemometrics and Intelligent Laboratory Systems* (2020), doi: <https://doi.org/10.1016/j.chemolab.2020.104040>.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2020 Published by Elsevier B.V.

Credit Author Statement:

S. Ruiz: Conceptualization, Methodology, Software, Formal analysis, Supervision, Writing, Review & Editing

L.A. Sarabia: Conceptualization, Methodology, Formal analysis, Supervision,

M.C. Ortiz: Conceptualization, Formal analysis, Supervision, Writing - Review & Editing, Funding acquisition

M.S. Sánchez: Conceptualization, Methodology, Formal analysis, Software, Writing - Original Draft, Supervision, Review & Editing

Residual spaces in latent variables model inversion and their impact in the design space for given quality characteristics

S. Ruiz^a, L.A. Sarabia^a, M.C. Ortiz^{b*}, M.S. Sánchez^a

^aDpt. Mathematics and Computation, ^bDpt. Chemistry

Facultad de Ciencias, UNIVERSIDAD DE BURGOS, Pza. Misael Bañuelos s/n, 09001 Burgos, Spain

Abstract

The paper contains a discussion about the null spaces associated to linear prediction models for the particular case of Partial Least Squares regression models. The discussion separately considers the two existing null spaces: the one in the input space related to the projection onto the latent space and the null space, coming from the projection space, corresponding to the mapping of the scores onto the predicted responses.

The paper also explores the impact of such null spaces in the definition of the design space around some feasible solutions obtained by inverting the prediction model, via several cases with simulated and real data from the literature. The case-studies serve to illustrate the discussion and the need of considering points in the two null spaces, rather than just take into account the null space within the latent space.

They also serve to show how to address the use of the resulting vectors in the design space to maintain the desired quality by modifying the tunable (maneuverable) process variables to compensate for variations due to some other feature variables not so easy to control.

Keywords

Partial Least Squares; Process Analytical Technology; Quality by Design; linear application; null space; model inversion.

1. Introduction

In the context of process control, prediction models are commonly used to estimate the expected quality of a product as a function of the characteristics of the process and/or some feature variables, such as raw material properties or environmental factors. Given

* Corresponding author, mcortiz@ubu.es

the frequently large number of process or feature variables and quality characteristics and their correlation (in both spaces), these prediction models are usually Latent Variables Models (LVM).

Besides their standard use to predict the expected quality characteristics (the response when fitting the model), these models are also useful when the focus is on maintaining a given quality, like in the framework of Quality by Design (QbD) established for pharmaceutical processes. In the latter situation, the prediction model must be inverted to find the settings of the input variables necessary to guarantee this given or desired quality.

The linear LVM selected is a Partial Least Squares (PLS) regression model. Briefly, a PLS model can be seen as the mathematical composition of two linear maps: first a reduction of dimensionality by projecting the points onto a space of lower dimension (the latent or projection space spanned by the selected latent variables) and then regressing the scores onto the response space.

The present work focuses on the variability in both the input and latent spaces due to the linear structure of the PLS model being used, rather than on the variability of the predicted response itself. The reader interested in the statistical uncertainty of the predictions can consult Refs. [1, 2] where some confidence intervals are computed for the response predicted with PLS models.

The use of subspaces to solve linear problems is usual in engineering and signal processing. In control system, which is area directly related to PAT (Process Analytical Technology) and QbD, first principles models (i.e. based on physical or chemical laws) have been traditionally of use to describe the behavior of a system to subsequently control it. However, in the last years, the idea of using data driven models even if they have no direct physical meaning gained interest among engineers, ending up in so called 'subspace' methods. Their name reflects the fact that linear models can be computed from row and column (sub)spaces of certain matrices, which are obtained from input-output data of the system under observation. Typically, the column space of such data matrices contains information about the model, while the row spaces are somehow representative of the state sequence, and they are estimated directly from input-output data without 'a priori' knowledge of any first principles models. The decomposition based on canonical classical models has some alternatives. For example, in the book in ref. [3], the use of QR decomposition is proposed for the projection (orthogonal or oblique) and generalized singular value decomposition to determine finite dimensional subspace and, then, least squares to obtain the linear relations. Least Squares Regression (LSR) method is introduced for subspace segmentation in ref. [4], but multivariate analysis with principal component analysis (PCA) and partial least squares (PLS) are somehow representative methods [5] among several data-driven techniques for process monitoring, thanks to their simplicity and efficiency in processing huge amount of process data. As stated in ref. [6], a more important objective of process monitoring is to provide assurance of good product quality that is impacted by the process conditions, and quality variables are better taken into account with PLS than with PCA.

Techniques of searching for subspaces appropriate to model the system response are extended to non-linear models by using KPCA (Kernel Principal Component Analysis) to model nonlinear dependences between variables to reproduce nonlinear dynamic behavior [7]. Regarding decompositions, ref. [8] is an introduction that shows the utility of Krylov subspaces to solve linear systems $\mathbf{Ax} = \mathbf{b}$, where \mathbf{A} is a square matrix and \mathbf{b} is a vector. When \mathbf{A} is invertible, the solution is just $\mathbf{A}^{-1}\mathbf{b}$ but the calculation of the inverse of \mathbf{A} for large matrices can be computationally inefficient and space consuming. On this subject, the book by Liesen and Strakos [9] contains a very complete mathematical and computational foundation, along with some interesting historical notes. The development of efficient algorithms to solve linear system of equations with sparse matrices is a matter of constant research, see for example the recent paper in [10].

Since the Krylov subspace methods are projection methods onto a subspace, its use as the projection techniques to solve large-scale control problems is a natural ‘extension’ of its use in numerical linear algebra for matrix problems. Basically, it consists of projecting the original large control problem onto a m -dimensional Krylov subspace (m ‘small’) by constructing a basis of the subspace, and then use a standard well-established technique to solve the projected smaller problem. In that way, an approximate solution of the original large problem is obtained from the solution of the projected problem [11,12].

PLS was initially developed as a heuristic technique [13] to solve least squares problems in multilinear regression, with no optimality properties and criticized for a long time as a somewhat ‘ad-hoc’ solution. Later, it was shown that it is equivalent to some of the most sophisticated numerical algorithms to date for solving systems of linear equations, such as the Lanczos bidiagonalization or the conjugate gradient methods [14]. The property that the vector $\hat{\mathbf{y}}$ consisting of the values predicted by a PLS model with a latent variables is the projection of the data vector \mathbf{y} onto a subspace spanned by the predictor variables in \mathbf{X} (which is a subspace that can be described as the span of a certain Krylov sequence), is used in ref. [15] to characterize several regression methods (ridge regression, continuum power regression, PLS and principal components regression, PCR) that attempt to reduce the variance of the OLS (ordinary least squares) regressor $\mathbf{b}_{OLS} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{y}$ by replacing the ill-conditioned matrix $(\mathbf{X}^T\mathbf{X})^{-1}$ by a more stable alternative.

The interpretation of PLS in terms of Krylov subspaces is also used in ref. [16] to show that various latent variables regression methods share a common structure, called Multilayer Basic Sequence (MBS), which is a particular case of a Krylov sequence. It allows unifying several algorithms used to estimate the PLS coefficients and studying the relation between Martens’ PLS [17] (orthogonal MBS) and Wold’s PLS [18,19] (oblique MBS). For any orthogonal-MBS model, there is an equivalent oblique-MBS model so that they share the same regression space (or Krylov spanned space), leading to equivalent predictive performance. That means that similar regression results can be obtained via multiple linear regressions onto the response variable. Martens’ PLS and Wold’s PLS are a special case of this situation. This explains the identical regression

coefficients and prediction results. The mapping nature of orthogonal-MBS and oblique-MBS from latent score \mathbf{T} -space to the original \mathbf{X} -space reflects the corresponding deflation techniques.

Although, conceptually, the interpretation of the latent variables is irrelevant in the methods of least squares based on subspace decomposition, it becomes ‘relevant’ during the stage of using the model, where the practitioner looks for this interpretation. This is especially true in Chemometrics where PLS used to be a method to predict concentrations of samples with known multivariate signals, that is, as a calibration method. It is worth mentioning that, in this context with, usually, more variables than samples, the solution by using standard linear regression models does not exist.

When exploring the possible interpretation of the PLS calibration models, it turned out that the subspace spanned by the a selected latent variables included information not directly related to the response \mathbf{y} . This issue was addressed with two methodologies: i) pretreating the signals to eliminate ‘artifacts’ of the signal (usually physical) not related to the response or ii) modifying the search for the solution limiting it to subspaces correlated with the response. The first approach has led to several pretreatment methods of the predictor variables, which greatly impact the subsequent PLS calibration model [20,21]. On the second methodology mentioned, the most generalizable contribution is possibly the Orthogonal Signal Correction (OSC) by Wold et al. [22], an appropriate modification of the PLS algorithm to eliminate systematic \mathbf{y} -orthogonal parts of the matrix \mathbf{X} . The Net Analyte Preprocessing (NAP) approach [23] and its exactly equivalent Direct Orthogonalization (DO)[24] propose some loading weights (not necessarily contained in the row space of \mathbf{X}) representing phenomena considered to be irrelevant to the modelling of \mathbf{y} . Other proposals, as Orthogonal-PLS (O-PLS)[25], the alternative way of computing O-PLS [26] and the regression procedure based on the Orthogonal Signal Correction by Fearn [27], are discussed by Indhal [28], clarifying the way these OSC methods work and giving a rigorous explanation of the reason why the entire OSC concept may be both confusing and superfluous. Yu and MacGregor [29] point out that ‘the subspace of \mathbf{X} uncorrelated to \mathbf{Y} is often high dimensional and it is not at all clear what part of this subspace is removed by different OSC algorithms with different number of OSC components. As a result, any interpretation of the OSC components may be different’ and propose the application of a canonical correlation analysis as post-processing of a PLS model (PLS-CCA). This approach to define the subspace most related to the quality variables is used in ref. [30] to model a batch process by looking for the subspace common to the different slices defined with the processing time.

Therefore, activity in the field of least-squares regression on subspaces is focused on modeling the subspace of the predictors S_m most related to the response. In general, the orthogonal subspace S_m^\perp is only considered for defining a limit, usually known as the squared prediction error (SPE), to decide if a new object is compatible with the fitted PLS (or PCR) model. This index is essential when building multivariate calibration models or multivariate control charts based on latent variables [31]. Its use has even

been generalized to the non-linear case of multivariate control in the industry using KPCA [7]. Besides the definition of SPE, the null space of the S_m^\perp regression model has been used, for example, for early diagnosis of structural damages or machinery malfunctions resulting in reduction of the maintenance cost and increasing reliability and safety [32,33,34].

This subsidiary use of the null space has been surpassed in the field of QbD and PAT, because the null space is part of the design space. This is consequence of the own definition of null space: when adding any vector of the null space (of the regression between \mathbf{X} -containing process variables, raw material or environmental properties, and \mathbf{Y} -with the quality characteristics) to any vector of the space spanned by \mathbf{X} , the predicted quality characteristics do not change. Nonetheless, the analyses on the null space are limited to the one related to regressing \mathbf{Y} on the latent variables. The QbD context requires to define the values of process variables, raw material or environmental properties that result in the same (or similar enough) value of the measured quality characteristics and, if only the referred null space is used, the resulting estimates are ‘indirect’ because the range of admissible values of the original variables is computed from their range on the latent variables.

Therefore, in what follows, the predicted value remains invariant. In fact, we set a desired value of the response \mathbf{y}_{des} (the target quality) and look for the corresponding settings of the input variables whose predicted value is exactly \mathbf{y}_{des} . Additionally, around these settings, when they exist, it is advisable to determine the corresponding design space, that is, the region in the input space where the predicted response is the same and thus the expected target quality is maintained.

When working with linear models, this design space includes the kernel of the mapping, also known as the null space, which we have already said that is the vector subspace that contains the points mapped onto zero, i.e. the vectors that do not modify the predictions when added to any other.

To our knowledge, Jaeckle and MacGregor [35] introduced for the first time the idea of a null space associated to the inversion of a PLS model, but its effect is only considered in the latent space. This null space is also taken into account in ref. [36] when backpropagating the confidence interval in ref. [1] with different procedures to calculate a kind of confidence regions to bracket the design space.

However, focusing on the fitted PLS model, part of the variability reflected in the design space comes from what we can call the ‘residual’ space related to the input variables. The term ‘residual’ refers to the fact that it is discarded when selecting the number of latent variables because it is supposed to be the part that is not related to the predicted quality characteristics. In particular, the variability in the residual space related to the input variables (the part of the process or feature variables space that is discarded after the decomposition) does not have any impact on the predicted quality. Therefore, it reflects the part of the space where the input variables can be modified

without altering the quality of the resulting product, that is, it is necessarily part of the design space.

The present paper intends to describe the ‘shape’ of this part of the design space throughout some production processes with a certain predefined quality. Technically, this requires the inversion of the regression model because the quality characteristics of the product are the ones kept fixed. Given that PLS is a linear model, under certain conditions (relation among the dimensions of the spaces –input variables, latent, and quality characteristics), this inversion can be done algebraically by solving matrix systems [37]. There are also some computational alternatives [38] that do not depend on the mentioned dimensions. Regardless of the picked solution, the null spaces related to the linear models being used play an important role in the subsequent design space.

Certainly, this ‘division into two parts’ of the null space can be ignored and use directly its algebraic characterization by using the regression coefficients of the fitted PLS model. However, not all points in this null space belong to the design space, because they also must be adequate for the process being modelled, i.e., feasible settings for the process/feature variables that, in addition, are projected onto the latent space defined by the PLS model.

The basic notation and some theoretical background are presented in section 2 and different small sample sized case-studies are analyzed in section 3 to illustrate the properties and practical applicability. The paper ends with some conclusions.

2. Theoretical properties and definitions

Let \mathbf{X} ($n \times p$) and \mathbf{Y} ($n \times q$) be the matrices of p (predictor) input variables and the corresponding q (response) quality characteristics, respectively. Frequently these are historical data of a process, arranged in two matrices used to fit, in our case, a PLS model to predict quality characteristics given some settings of the process/feature variables. These data are assumed to be representative of the process because the information they contain is the only one used to fit the model and to make decisions, in a completely data-driven approach.

In what follows, we are assuming that p and q are less than n , that up to p new orthogonal variables can be obtained in the input space, and that both \mathbf{X} and \mathbf{Y} are autoscaled (that is, each column is centered by subtracting the mean, and has variance one after dividing by the standard deviation).

In addition, in any process under control, the process/feature variables are bounded. To emulate this situation, the admissible domain of the predictor variables will be defined by the range of the observed values in \mathbf{X} . The design space we are looking for must be inside this domain and with the settings of the input variables that define feasible solutions.

A PLS prediction model is fitted with \mathbf{X} and \mathbf{Y} by computing a orthogonal new variables (the latent variables) with a common scores matrix \mathbf{T}_a ($n \times a$) such that

$$\mathbf{X} = \mathbf{TP}^T = \mathbf{T}_a \mathbf{P}_a^T + \mathbf{R}_X \quad (1)$$

$$\mathbf{Y} = \mathbf{TQ}^T + \mathbf{R}_Y \quad (2)$$

Matrices \mathbf{P} ($p \times p$) and \mathbf{Q} ($q \times a$) are the loadings, and \mathbf{R}_X and \mathbf{R}_Y the matrices containing the residuals of the decomposition. As usual, upper T means transposing the corresponding matrix or vector. Finally, sub-index a denotes that we are only considering the first a columns of the corresponding matrix, so for instance, \mathbf{T} is the $n \times p$ matrix of \mathbf{X} scores, whose first a columns form the common scores matrix \mathbf{T}_a . The subspace spanned by the columns of \mathbf{P}_a is the latent variables space or projection subspace (for now on, latent space).

Once a PLS model in equations (1) and (2) is fitted, there are also some constraints to adequately apply the model. We will use the 95 % confidence limits for the usual Q and T^2 statistics to define the proper region where the feasible predictor variables are projected.

The value of the T^2 statistic for a given sample is the Hotelling's distance from the projection to the origin of the latent space (i.e., all scores equal to zero) and indicates the position of the projected sample. On the other hand, Q -residuals are calculated as the sum of squares of each row in \mathbf{R}_X , eq. (1), the residuals after projecting the sample through the model. It is then a measure of the distance from the sample to the orthogonal projection onto the latent space. Imposing limits in both statistics is to impose a threshold value for the variation of points both within and outside the latent space.

Accordingly, the fitted model must incorporate both constraints, so that the PLS model is properly applied only to those \mathbf{x} with values of Q and T^2 below the threshold values imposed in both statistics. In other words, we are defining a bounded domain for the application of the PLS model. Geometrically, this domain can be viewed as a "PLS-box" that includes the part of the latent space and the part of the residual space bounded with, in our case, the 95 % confidence level for Q and T^2 statistics.

In that sense, the feasible solutions are those that belong simultaneously to the domain of the input variables and to what we have named the PLS-box.

For vector calculations, we will respect their structure in matrix form. For example, a set of p predictor variables is written as a vector \mathbf{x}^T because these variables are rows in matrix \mathbf{X} . Similarly, \mathbf{y}^T is a row vector containing quality characteristics.

With this notation, vector \mathbf{x} containing some settings for the input variables is projected onto the PLS-latent space by using some weights in \mathbf{W} ($p \times a$ matrix such that

$\mathbf{P}_a^T \mathbf{W} = \mathbf{I}_a$), so that the a -dimensional vector of scores is $\mathbf{t}_a^T = \mathbf{x}^T \mathbf{W}$. Then, to obtain the expected (predicted) characteristics, $\hat{\mathbf{y}}$, it suffices to compute $\hat{\mathbf{y}}^T = \mathbf{t}_a^T \mathbf{Q}^T$. In short, given values of p input variables in \mathbf{x} , the vector of q predicted quality characteristics $\hat{\mathbf{y}}$ is computed as:

$$\hat{\mathbf{y}}^T = \mathbf{x}^T \mathbf{W} \mathbf{Q}^T = \mathbf{x}^T \mathbf{b} \quad (3)$$

It is clear from equation (3) that the fitted PLS model is a linear mapping, L , between the p -dimensional input space to the q -dimensional quality characteristics space defined by the matrix of estimated coefficients \mathbf{b}^T ($q \times p$).

As any other linear map, it has a null space associated to it, that is, the subspace that contains all the vectors that are mapped onto the null vector. Because $a < p$, the dimension of such subspace (the kernel of L , $\ker(L)$) is $d = p - \text{rank}(\mathbf{b}) = \max\{p - a, p - q\}$ and it is easy to compute a basis \mathbf{B}_{null} ($p \times d$ matrix) to characterize $\ker(L)$. That means that for any d -dimensional vector \mathbf{v} , the corresponding vector $\mathbf{x}_0^T = \mathbf{v}^T \mathbf{B}_{\text{null}}^T$ belongs to $\ker(L)$. In other words, if the null space does not reduce to the null vector, it contains the set of vectors \mathbf{x}_0 that can be added to any other \mathbf{x} in that space without modifying the predicted response, that is,

$$\hat{\mathbf{y}}^T = \mathbf{x}^T \mathbf{b} = L(\mathbf{x}) = L(\mathbf{x} + \mathbf{x}_0) = (\mathbf{x}^T + \mathbf{x}_0^T) \mathbf{b} \quad \forall \mathbf{x}_0 \in \ker(L) \quad (4)$$

This null space is relevant when defining the design space for a given set of quality characteristics, because it is part of the allowed variation in the process/feature variables that does not modify the predicted response. Consequently, the set of points $\{\mathbf{x} + \mathbf{x}_0, \mathbf{x}_0 \in \ker(L)\}$ is contained in the design space around \mathbf{x} .

On the other hand, the interest now is focused on the inversion of the PLS prediction model for a desired or given \mathbf{y}_{des} (values of the target quality characteristics). This issue is usually tackled in the latent space by determining the a scores $\hat{\mathbf{t}}_a$, equation (2), such that

$$\mathbf{y}_{des}^T = \hat{\mathbf{t}}_a^T \mathbf{Q}^T \quad (5)$$

Then, to obtain the corresponding predictor variables, it suffices to use the loadings \mathbf{P}_a in equation (1) to ‘transform’ $\hat{\mathbf{t}}_a$ to a p -dimensional \mathbf{x} irrespective of the values that the discarded dimensions could have, that is, $\hat{\mathbf{x}}_{des}^T = \hat{\mathbf{t}}_a^T \mathbf{P}_a^T$.

However, the solution $\hat{\mathbf{t}}_a$ of equation (5) may not be unique. When $a > q$, there is a null space as defined in [35], which is a subspace inside the latent space that contains the vectors \mathbf{t}_0 for which the predicted response is zero, that is,

$$\mathbf{t}_0^T \mathbf{Q} = \mathbf{0}_q \quad (6)$$

Equation (6) shows that \mathbf{t}_0 belongs to the null space related to the linear map defined by matrix \mathbf{Q} in equation (2), provided $a > q$ (otherwise, the kernel is reduced to the null vector). To avoid confusion, we will denote this space as \mathbf{Q} -null space.

Therefore, for any \mathbf{t}_0 in \mathbf{Q} -null space, $\hat{\mathbf{t}}_a + \mathbf{t}_0$ is also solution of equation (5) and is mapped to \mathbf{y}_{des} , and thus, the inversion problem has infinitely many solutions in the latent space. When multiplying by the loadings \mathbf{P}_a , we have infinitely many settings for the input variables with the same expected response, i.e., that predict the same quality characteristics, namely

$$\hat{\mathbf{X}}_{des} = (\hat{\mathbf{t}}_a^T + \mathbf{t}_0^T) \mathbf{P}_a^T = \hat{\mathbf{t}}_a^T \mathbf{P}_a^T + \mathbf{t}_0^T \mathbf{P}_a^T = \hat{\mathbf{x}}_{des}^T + \mathbf{x}_0^T \quad (7)$$

for any $\hat{\mathbf{t}}_a$ solution of equation (5) and \mathbf{t}_0 in \mathbf{Q} -null space, hence $\mathbf{x}_0 \in \ker(L)$ and all the points in matrix $\hat{\mathbf{X}}_{des}$ has the same predicted response, \mathbf{y}_{des} .

Nevertheless, this procedure for computing different settings of input variables to obtain the desired quality characteristics does not cover the whole set of possible points in $\ker(L)$ because there is another null subspace, orthogonal to the latent space, which is ignored when re-constructing the predictor variables as in equation (7).

This subspace is precisely the one spanned by the loadings of the discarded latent variables. For the training set in equation (1), we can write

$$\mathbf{X} = \mathbf{T}_a \mathbf{P}_a^T + \mathbf{T}_r \mathbf{P}_r^T \quad (8)$$

where sub-index r means the last r columns of the corresponding scores or loadings matrices ($r = p - a$). For any r -dimensional point \mathbf{s} ,

$$\mathbf{x}_r^T = \mathbf{s}^T \mathbf{P}_r^T \quad (9)$$

is a p -dimensional vector in the input space such that $\mathbf{x}_r^T \mathbf{b} = \mathbf{0}_q$, i.e. it belongs to $\ker(L)$.

In fact, is

$$\mathbf{x}_r^T \mathbf{W} = \mathbf{0}_a \quad (10)$$

meaning that any of the \mathbf{x}_r is in fact projected onto the origin of the latent space (null scores).

Therefore, if p latent variables can be computed for the p -dimensional \mathbf{X} , the loadings of the discarded latent variables are a basis of the kernel of the projection from the input space onto the latent space. Now, we will denote this space as \mathbf{W} -null space, related to the residual space, to distinguish it from the previous \mathbf{Q} -null space inside the latent space.

With this notation, and always in the global $\ker(L)$, we will differentiate between points \mathbf{x}_r in \mathbf{W} -null space, solution of equation (10) computed as in equation (9), and points \mathbf{x}_0 that will correspond to $\mathbf{x}_0^T = \mathbf{t}_0^T \mathbf{P}_a^T$, that is, points reconstructed from any \mathbf{t}_0 in \mathbf{Q} -null space defined in equation (6). These two ‘types’ of vectors have additional distinguishing properties: any \mathbf{x}_r will have null value for the T^2 statistic since it is always projected onto the origin of the latent space, and any \mathbf{x}_0 will have Q -residual equal to 0 because it is computed directly from the latent space.

In any case, as we are trying to estimate the design space around a given $\hat{\mathbf{x}}_{des}$ to achieve \mathbf{y}_{des} , the variability due to the \mathbf{W} -null space must be taken into account together with the one due to the \mathbf{Q} -null space.

Aside from the residual variability in \mathbf{R}_Y in equation (2) (which will be reflected in the statistical uncertainty when estimating a given response \hat{y}), we are focusing on the one in \mathbf{R}_X , equation (1), which will exert an impact on the input variables but not on the predicted response nor even in the latent space. In other words, this variability will be reflected only in the design space.

According to equations (6) and (10), the first property to notice is that, in all the discussed cases, the variation around any point in the input space due to the null space does not depend on the actual values, neither the variables \mathbf{x} nor the predicted (or desired) quality characteristics in the corresponding \mathbf{y} . However, this property, in practice, can be useless.

In fact, all the properties in the previous paragraphs are theoretical properties. They help in understanding the spaces we are exploring but fail in the precise description of the design space. This is so because the points that we can obtain with the bases computed for the different null spaces, indeed, do not modify the predicted response since they always add zero, but we have no guarantee that the PLS model is properly applied to them, nor even that the calculated input variables are inside the allowed domain. Put differently, any point in $\ker(L)$ is not necessarily a feasible solution, inside both the input domain and the PLS-box.

Therefore, besides the theoretical property of being in $\ker(L)$, we also need to characterize which points are not only within the domain where the predictor variables can vary, but also have projections inside the latent space delimited by the fitted PLS model. Both Q and T^2 statistics are then the parameters to be used to decide that a process is in-control when this control is based on LVM.

3. Results and discussion

Different case-studies are discussed, illustrating different properties or approaches. If the data correspond to a process, then \mathbf{X} will contain some values of process variables or features with some other characteristics (raw material, environmental factors, etc.) and \mathbf{Y} will be then some quantitative measure(s) of a quality characteristic of the product.

In a general context, the data sets consist of a matrix \mathbf{X} with predictor or input variables and, a vector \mathbf{y} (in all the section is $q = 1$) with the response to be modelled. Both \mathbf{X} and \mathbf{y} are supposed to be autoscaled before fitting the model.

Irrespective of the type of data at hand, the discussion is related to the properties of the PLS model fitted, together with some suggestions as to use these properties in practice, for example, in process control.

For each data \mathbf{X} - \mathbf{y} , the procedure has the following steps:

- Step 1.** Set the domain with the allowed range of variation of the input variables. We have defined it as the rectangular parallelepiped (rectangular box) formed with the variation range of \mathbf{X} .
- Step 2.** Set the domain for the PLS model (the PLS-box), with coefficients vector \mathbf{b} . To do it,
- fit a PLS model with autoscaled \mathbf{X} - \mathbf{y}
 - remove any point with values of both Q and T^2 statistics above the threshold (critical values at 95 % confidence level)
 - remove any point with standardized residual (difference between y and \hat{y}) outside the interval $[-2.5, 2.5]$
 - re-scale the remaining points and repeat until no sample is outside the box
- Step 3.** Define a target or desired value of y_{des} (target response or desired quality characteristic).
- Step 4.** Look for score(s) $\hat{\mathbf{t}}_a$, solution of eq. (5), and inside the latent space. If $a > q$, generate m points \mathbf{t}_0 in \mathbf{Q} -null space, eq. (6), and add them to $\hat{\mathbf{t}}_a$.
- Step 5.** By using the loadings, as in eq. (7), compute $\hat{\mathbf{X}}_{des}$ (m settings of predictor variables, all of them with the property of $\hat{\mathbf{X}}_{des} \mathbf{b} = y_{des}$).

Step 6. Generate points belonging to \mathbf{W} -null space, following eq. (10), and add them to all the computed $\hat{\mathbf{X}}_{des}$ (the matrix is, in fact, a single point when $a = q$).

Step 7. Remove those settings outside the PLS-box or outside the input domain defined in step 1, that is, retain only the feasible solutions.

The last steps of the procedure are summarized in fig. 1. They correspond to the algebraic inversion of the PLS model plus the added solutions taking into account the two null spaces we have introduced in section 2.

Figure 1 around here

After finishing the steps explained, all the remaining points and their convex combinations (line segments between points in the set) are part of the design space corresponding to the inversion of the PLS fitted to predict y_{des} due to the linear nature of the PLS model. Remember that a set of points is said to be convex [39] if the line segment between any two of its points is entirely contained in the set, and the convex hull of a set of points is the smallest convex polygon that encloses all the points in the set.

3.1 Case 1. Simulated linear relation between \mathbf{X} and \mathbf{y}

The first case study consists of a simulated linear relation with data in low dimension to allow the usual Cartesian representations to illustrate more easily the different spaces explained in section 2 and their effect on the design space, or the part of the domain of the input variables where they can be modified without altering the predicted response.

Therefore, let us consider three input variables (in this case, three predictor variables or regressors) X_1, X_2, X_3 and a single response Y . The linear relation is defined in eq. (11), just the sum of the variables plus an offset term.

$$Y = X_1 + X_2 + X_3 - 1.5 \quad (11)$$

To fit a PLS model in order to recover this linear relation, we need some data. The following subsections show the effect of the available data and the selected number of latent variables on the PLS model fitted and, thus, on the part of the design space estimated with the corresponding model.

In all three cases, we simulate a predictor matrix \mathbf{X} with 50 samples and 3 variables and compute response vector \mathbf{y} according to eq. (11) with the actual values in \mathbf{X} . A comparative summary of the three situations is at the end of the present section.

3.1.1 PLS model with a single latent variable. Uncorrelated predictor variables

In this case, the three variables are generated as values independently chosen with a uniform distribution in $[0, 1]$. That means that there is neither correlation among the predictor variables nor experimental error, only the relation between predictor and response, which is a linear relation that we want to model via PLS.

Table 1 contains the characteristics of the decomposition when fitting the PLS models. A single latent variable (LV) suffices to explain 98.96 % of the variance of the response \mathbf{y} with 36.84 % of the variance in \mathbf{X} .

Table 1. Characteristics of the PLS model fitted to the simulated data with uncorrelated predictors.

Number of LV	Variance explained in \mathbf{X} (%)	Variance explained in \mathbf{y} (%)	RMSECV
1	36.84	98.96	0.104
2	71.96	99.99	0.015

RMSECV: Root Mean Squared Error in Cross-Validation

With this single latent variable for the PLS model, in terms of the notation established in section 2, we have $p = 3$, $a = 1$, $q = 1$. Since $a = q$, if we need to find values for the regressors to predict, say $y_{\text{des}} = 0.31$ (an intermediate quantity among the autoscaled values of response), there is a unique solution for determining the corresponding score in eq. (5), step 4 in fig. 1, namely \hat{t}_a such that $\hat{t}_a \mathbf{Q} = 0.31$.

Applying the loadings in \mathbf{P}_a in eq. (1), step 5 in fig. 1, there is then a unique $\hat{\mathbf{x}}_{\text{des}} = (0.12, 0.22, 0.21)$ that will predict the desired characteristic y_{des} .

It is necessary to check that this solution belongs to the PLS-box as well as to the domain of the input variables. The latter condition is directly examined with the coordinates. For the former step, in fact, as its Q -residual is always null, we only have to check if its T^2 value is less than the threshold limit, which is 4.04 for the fitted PLS model with one latent variable.

In this case, $\hat{\mathbf{x}}_{\text{des}}$ meets all the constraints, defining thus a feasible solution. Fig. 2 is the Cartesian representation of the three-dimensional input space. The range of \mathbf{X} , green diamonds, defines the domain. We can see that $\hat{\mathbf{x}}_{\text{des}}$, the red filled circle, is well inside it.

Figure 2 around here

However, there are two dimensions discarded when fitting the PLS model that impact the predictor variables space but not the latent space. Because of the projection, there is a plane in the three-dimensional input space where we can move without modifying the predicted response. This plane is consequence of the space that we have called the **W**-null space, related to the residual space, which in turn is orthogonal to the latent space.

To explore the effect of the **W**-null space associated to y_{des} , we generate $m = 1000$ points uniformly distributed inside the range defined by the discarded scores \mathbf{T}_r , equation (8), and multiply them by \mathbf{P}_r , the discarded loadings that are a basis of the **W**-null space. As we have already indicated, all the points thus generated have null value of T^2 .

Next, we add all of them to $\hat{\mathbf{x}}_{des}$, step 6 in fig. 1, which does not modify the predicted response according to eq. (10) although it does modify the values of Q and T^2 . Therefore, the final step (step 7 in fig. 1) is to remove those resulting points outside either the domain of the input variables or the latent space (in this case, only if their Q -residuals are larger than the critical value 5.68).

The remaining 726 settings are plotted as small black dots in fig. 2. To facilitate visualization of the plane they all lie on, the convex hull of the points is depicted in light gray and it is clear that it bounds the whole plane (containing $\hat{\mathbf{x}}_{des}$) to the points inside the defined domain for the input variables and the PLS-box.

3.1.2 PLS model with two latent variables. Uncorrelated predictor variables

For the sake of illustration, for the same data sets \mathbf{X} - \mathbf{y} , let us suppose that we had chosen two LV for the PLS prediction model. As we can see in table 1, this new model explains 99.99 % of the variance in \mathbf{y} and 71.96 % of the variance of \mathbf{X} . Due to the uncorrelated predictor variables, this second latent variable only explains variance in \mathbf{X} , in a percentage similar to the first one.

Figure 3 shows the scores (green diamonds) of \mathbf{X} in the two-dimensional latent space. Now $a = 2$, so there is a one-dimensional null space in the latent space, what we have called the **Q**-null space, resulting from the infinitely many solutions of equation (5). One of its valid solution has score zero on the second LV and, thus, with the same \hat{t}_1 on the first LV as in 3.1.1. This solution is in fact the solution in the least squares sense to the system in equation (5). It is represented in figure 3 by a red filled circle.

The remaining scores, solution of equation (5), theoretically follow a straight line, the one depicted in fig. 3 with blue filled squares, obtained by taking points equally spaced along it, step 4 in fig. 1 with $a > 1 = q$.

This line is in fact limited to the segment made up with the points inside the PLS-box in the latent space, graphically, the segment bounded by the ellipse depicted in fig. 3 with black dashed line. The points enclosed by the ellipse have T^2 values less than the critical value established when fitting the model, namely 6.51 at 95 % confidence level. Notice that the ellipse is almost a circumference due to the lack of correlation among the predictor variables.

Figure 3 around here

By using the loadings of \mathbf{P}_a in equation (1), this \mathbf{Q} -null space is expressed (step 5 in fig. 1) in the input space also as a line in the three-dimensional space. This line is the one defined by the 18 blue filled squares in fig. 4 that correspond to the blue filled squares in fig. 3 whose re-constructed predictor variables are feasible solutions (i.e., belong to both the PLS-box and the input domain). Since they are computed directly from the latent space by using the loadings, their Q -residuals are always null. Notice that the central position is occupied by the red filled circle which, also in this space, is exactly $\hat{\mathbf{x}}_{des}$, the same as in section 3.1.1.

As in the previous section, there is some additional variation allowed in the input space which is not being considered by only taking into account the \mathbf{Q} -null space defined in the latent space. The possibilities that are missing come from the \mathbf{W} -null space, which in this case is the one-dimensional null space related to the projection onto the latent space, and orthogonal to it.

With eq. (10) we generate 50 points in \mathbf{W} -null space and, following step 6 in fig. 1, add them to each of the 19 solutions already found (the 18 blue squares plus the red circle in fig. 3). Finally, step 7 in fig. 1, we keep the 760 feasible solutions.

Figure 4 around here

These feasible solutions are the small black dots defining parallel lines in figure 4 where they clearly show that the direction determined by the black points in \mathbf{W} -null space is orthogonal to the aforementioned line defined by the solutions coming from the \mathbf{Q} -null space, blue squares.

These two directions, together, define a plane, but only a part of it contains the settings of the predictor variables for which the problem makes sense, i.e., only the convex hull depicted in fig. 4 in light gray is part of the design space around $\hat{\mathbf{x}}_{des}$ for predicting \mathbf{y}_{des} .

In any case, the bounded regions in gray in fig. 2 and 4, inside the theoretical planes with equal predicted response, illustrate that if only the part of $\ker(L)$ coming from \mathbf{Q} -null space is considered, the design space would be underestimated, it is necessary to add points in \mathbf{W} -null space as well.

3.1.3 Correlated predictor variables

It might appear that the constraints imposed for defining the PLS-box have a certain degree of arbitrariness, but in fact they are the key of the control of a process by using the fitted model.

To illustrate this, in the same established situation, let us collect new data for fitting the model, this time with variables highly correlated. Then, new 50 samples with three correlated variables are simulated and, again, the true value of the response in \mathbf{y} is computed as in eq. (11).

Table 2 contains the new structure of the PLS models fitted, with 1 and 2 latent variables. One of the points was well beyond the 95 % limits of Q -residuals and T^2 statistics and it was removed.

Table 2. Characteristics of the PLS model fitted to the simulated data with correlated predictor variables.

Number of LV	Variance explained in \mathbf{X} (%)	Variance explained in \mathbf{y} (%)	RMSECV
1	95.08	81.85	0.429
2	98.80	98.64	0.127

RMSECV: Root Mean Squared Error in Cross-Validation

Compared to the models in table 1 for uncorrelated variables in \mathbf{X} , in the present case, the first latent variable is mostly related to the correlation among variables in \mathbf{X} (95.08 % of variance in \mathbf{X}) and the second latent variable is needed to better predict the response in \mathbf{y} , up to 98.64 % of explained variance.

Figure 5 shows the new latent space, with the same coding as figure 3: Green diamonds are the projection of the data points in \mathbf{X} , the red circle is the least squares solution of eq. (5) for $y_{\text{des}}=0.28$ (an intermediate non-null value of autoscaled \mathbf{y}), and the blue squares are along the segment consequence of the \mathbf{Q} -null space, inside the latent space, related to the red point. The subspace is again a line segment but with a very different slope, compared to the one in figure 3. Finally, the ellipse in fig. 5 is the boundary of the region of the latent space belonging to the PLS-box. We see that, now, is far from circular.

Fig.5 around here

The region of feasible settings of the predictor variables within the design space related to y_{des} is depicted in fig. 6 as the convex hull of the solutions computed following the procedure of fig. 1.

Figure 6 around here

The orthogonal position of the lines defined with the blue points and the ones defined with the black points is still evident, but the yellow region inside the theoretical plane is now very different from those in fig. 2 or fig. 4. In this case it is a kind of narrow band that extends in the direction defined by the \mathbf{Q} -null space (blue squares), with width due to the \mathbf{W} -null space (black dots) and controlled, precisely, by the threshold on the Q statistic. The small variation allowed is not surprising taking into account that selecting two latent variables, there is only near 1 % of the variance in \mathbf{X} left in the residual space.

Summing up, even though the functional relation between \mathbf{X} and \mathbf{y} is the same in all the cases, the estimated regions inside the design space, consequence of the null space, are not. In fact, the whole kernel subspaces are different. To prove it analytically, we use directly the corresponding regression coefficients to obtain the two-dimensional null space ‘at once’ without making distinction for the source of such variability (\mathbf{Q} - or \mathbf{W} -null spaces).

With a single response, the kernel of the linear map defined by the regression coefficients \mathbf{b} can be interpreted as the set of points orthogonal to \mathbf{b} (see for instance page 137 in [39]), so that the regression vector \mathbf{b} is the normal vector of the plane we are looking for. With this property, the analytical equations of the planes in figures 2, 4 and 6 are, respectively:

$$\begin{aligned}\pi_1 : 0.45X_1 + 0.63X_2 + 0.63X_3 - 0.33 &= 0 \\ \pi_2 : 0.54X_1 + 0.58X_2 + 0.61X_3 - 0.32 &= 0 \\ \pi_3 : 0.68X_1 + 0.47X_2 + 1.13X_3 - 0.28 &= 0\end{aligned}\tag{12}$$

The first two planes in eq. (12), π_1 and π_2 , are very similar to each other, and correspond to the situation with uncorrelated predictor variables. The reason for the difference is the addition of the second latent variable, which hardly modifies the predictions computed with the PLS model with a single latent variable (due to the independent variables in \mathbf{X}), but that modifies the residual and latent spaces. These spaces are completely different in the case of the third plane π_3 obtained with a PLS model with 2 LV fitted with correlated input variables.

Therefore, it is clear that the model decomposition is driven by the available information, that is, depends on the data at hand. The correlation of the input variables in the available data forces the shape and orientation of the PLS-box and, thus, also the ‘inherited’ structure imposed in the input space (where the design space lies).

In any case, we have already pointed out that the expressions in eq. (12) alone are useless unless we impose the constraints of the model. However, they can be useful to make decisions in some practical situations. For example, suppose that there was a limitation in the availability of -say, X_3 (a very expensive reactive or shortage in some material). The corresponding equation gives the conditions to be imposed in the other two variables to maintain the predicted response, providing alternatives to find feasible solutions after applying the constraints.

The main conclusion is that the design space for a given \mathbf{y}_{des} should be defined taking into account all the possible variations around a computed solution of the inversion. The situation in figures 2, 4 and 6 illustrates that if only the part of $\ker(L)$ coming from \mathbf{Q} -null space is considered, the design space would be underestimated, it is necessary to add points in \mathbf{W} -null space as well.

Besides, as expected, the design space is dependent not only on the value of the desired response but also on the structure of \mathbf{X} and \mathbf{y} , provided they are properly ‘collected’ via the prediction model. The limits in Q and T^2 statistics when using a PLS model help defining the domain of applicability and controlling the process.

In particular, ‘how much’ of the residual space is, in fact, affecting the model (i.e., it is inside the PLS-box) is mostly controlled by the threshold value imposed in the Q statistic. In the cases discussed in this section, if only the prediction ability of the PLS model is taken into account, tables 1 and 2 show that we should select one latent variable for the case of uncorrelated predictor variables (that explains 98.96 % of the variance in \mathbf{y} with less than 37 % in \mathbf{X}), and two for the case of correlated variables (explaining 98.62 % of \mathbf{y} and 98.8 % of \mathbf{X}). In that case, there is almost a 65 % of the variance of \mathbf{X} unaccounted for in the first case, which is ‘reflected’ in the large region in figure 2, whereas the residual space will account for less than 2 % of the variance in \mathbf{X} in the second case, with a narrow region related to the inversion in figure 6.

3.2 Byproduct in Alumina production

Table 2 in [40] contain data from a study to optimize the technological conditions in a process, data that we will use as representative of the process. According to the authors, the process of an alumina production line consists of separating out the aluminum hydroxide $\text{Al}(\text{OH})_3$ from a saturated solution of sodium aluminate NaAlO_2 as a result of reaction with carbon dioxide, CO_2 . Unfortunately, the silicon dioxide SiO_2 present in the sodium aluminate also separates out and mixes with the $\text{Al}(\text{OH})_3$ degrading the quality of the extracted aluminum hydroxide.

The SiO_2 content in $\text{Al}(\text{OH})_3$ is the characteristic recorded in the dataset. It consists of 31 samples of the production process characterized by seven feature variables, those

listed in table 3, where we can see that there are proper process variables as well as some features measured in the raw materials.

Table 3. Process and feature variables for the Alumina production process

Notation	Description
X_1	the total alkaline concentration of the original solution
X_2	the aluminum concentration of the original solution
X_3	the ratio of SiO_2 and Al_2O_3 concentrations
X_4	the total alkaline concentration of the solution after discontinuing the connection of CO_2
X_5	the decomposition rate
X_6	the rest time of the solution before connecting the CO_2
X_7	the time connecting the CO_2

The seven feature variables work as the predictor variables to model the quantity of the byproduct SiO_2 that should be as low as possible. In the experimental data this quantity varies in [0.0302, 0.0653].

After autoscaling data to obtain matrices \mathbf{X} (31×7) and \mathbf{y} (31×1), a PLS model is fitted with the characteristics in table 4. Looking at the percentage of variance explained by adding latent variables, it is observed that, although the fifth latent variable captures more than 10 % of variance in \mathbf{X} , this is not related to the SiO_2 content we are fitting (which is not surprising since the process is for alumina production, not SiO_2). Besides, cross validation with venetian blinds shows a sustained decrease of RMSECV (Root Mean Squared Error in Cross Validation) up until the fourth latent variable, then no improvement is observed by adding more latent variables.

Table 4. Variance captured by PLS when adding latent variables and the corresponding Root Mean Squared Error in Cross-Validation (RMSECV).

Number of latent variables	Variance captured in \mathbf{X} (%)	Cumulative variance captured in \mathbf{X} (%)	Variance captured in \mathbf{Y} (%)	Cumulative variance captured in \mathbf{Y} (%)	RMSECV
1	36.77	36.77	74.20	74.20	0.548
2	17.16	53.93	14.18	88.38	0.408
3	13.14	67.07	5.36	93.74	0.295
4	5.72	72.79	3.64	97.38	0.223
5	11.43	84.23	0.23	97.61	0.200
6	8.59	92.81	0.05	97.66	0.200
7	7.19	100.00	0.00	97.66	0.201

Therefore, the PLS predicting model has 4 latent variables explaining 72.79 % of the variance in the predictor variables related to 97.38 % of the variance in the response \mathbf{y} . The critical values at 95 % confidence level for Q and T^2 statistics are 5.26 and 12.12 respectively.

With the notation established in section 2, we have $p = 7$, $a = 4$, $q = 1$, so both the \mathbf{W} -null space related to the ‘residual’ space discarded with the projection and the \mathbf{Q} -null space inside the latent space are three-dimensional subspaces.

In this case, we explore deeper the scope of knowing the ‘shape’ of the allowed variations around any given point, as consequence of the null space, and whether it can be useful for a general description of the design space. We have already stated that the variability around a given point due to the different null spaces does not depend on the values of \mathbf{y}_{des} or the computed $\hat{\mathbf{t}}_a$ or $\hat{\mathbf{x}}_{des}$. That means that we can explore the points that belong to the different null spaces without the need of defining a target quality, or equivalently, considering that the target value is zero (the mean value of the response when it is seen in the raw scales of the feature variables).

With this aim and starting with \mathbf{W} -null space (step 6 in fig. 1), we take 100 points, uniformly distributed, within the hypercube defined by the scores on the three discarded latent variables, \mathbf{T}_r in eq. (8), and then reconstruct the corresponding predictor variables, \mathbf{X}_r in eq. (9), by multiplying by the loadings \mathbf{P}_r . All the rows in \mathbf{X}_r are points of $\ker(L)$, in other words, represent variations that added to any setting of input variables do not modify the response predicted with those settings.

On the other hand, for \mathbf{Q} -null space, we can apply the steps in fig. 1 for $y_{des} = 0$, and compute scores to predict y_{des} by using the loadings \mathbf{Q} of eq. (2). Irrespective of the value of y_{des} , when mathematically inverting the model, there is not unique solution but infinitely many vectors that belong to the three-dimensional \mathbf{Q} -null space inside the four-dimensional latent space. To explore this space, $m=100$ points are also generated in it, limiting their range to the range of the actual scores \mathbf{T}_a in eq. (8). The predictor variables are then computed by using the loadings \mathbf{P}_a to have \mathbf{X}_0 .

For both \mathbf{X}_0 and \mathbf{X}_r , the prevention of using the range of values within the corresponding scores does not guarantee that the solutions obtained will be directly valid. So, by discarding those objects outside the input domain or outside the PLS-box, \mathbf{X}_0 still contain 88 vectors (points) coming from \mathbf{Q} -null space, and \mathbf{X}_r another 71 belonging to \mathbf{W} -null space.

Figure 7 is the parallel coordinates plot of the computed points, in the scale used with the PLS model (i.e., autoscaled values). In the plot, each point is represented by one broken line that intersects each coordinate at the height corresponding to the value of the point in this coordinate. The red dashed lines in fig. 7 correspond to samples in \mathbf{X} as a reference of the domain of the process/feature variables. The continuous blue lines are for \mathbf{X}_0 , computed from \mathbf{Q} -null space, in fig. 7a); and for \mathbf{X}_r , the settings in \mathbf{W} -null space, in fig. 7b).

Figure 7 around here

Figure 7 shows that, in both cases, the points are around the null vector (that always belongs to the null space) but affect differently depending on the variable we look at, and to different variables depending on the null space in question.

The lines in fig. 7a), related to \mathbf{Q} -null space inside the latent space, show that there is a shorter range of allowed variation for X_3 and X_4 than for the rest, especially X_5 .

However, in fig. 7b), related to \mathbf{W} -null space, we see less variation permissible for X_2 , X_3 , practically no variations around X_5 , but a larger variation in X_4 than in fig. 7a).

Comparing fig. 7a) and 7b) it is clear that the two types of solutions with null prediction are different and, thus, will exert a different impact on the design space.

In addition to the points represented, any linear combination of them also belongs to the kernel, that is, any of them can be added to any settings of variables without affecting the intended response, though the addition does modify the values of Q and T^2 statistics and even the new points may have moved outside the domain. Of course, we can reduce somehow the magnitude of the values in \mathbf{X}_0 and \mathbf{X}_r so that the region covered by the points with null prediction would be smaller, i.e., a better estimation of the part of $\ker(L)$ related to the definition of the corresponding design space.

Regardless, once the solutions in \mathbf{X}_0 and \mathbf{X}_r are available, and inside both the domain and the PLS-box, it is important to remember that, although any linear combination of points in $\ker(L)$ also belongs to $\ker(L)$, only the convex combinations of points in \mathbf{X}_0 or \mathbf{X}_r are part of the valid variations that can be added to a feasible solution when defining the design space.

To illustrate the practical meaning of having such a description, we will define a target response, say 0.0302 (the actual minimum in the domain) and look for the settings of feature variables to obtain this small quantity of the subproduct of the reaction, as predicted with the PLS model.

For the algebraic inversion, fig. 1, we have to solve eq. (5) looking for scores $\hat{\mathbf{t}}_a$ such that $y_{des} = \hat{\mathbf{t}}_a \mathbf{Q}^T$, where y_{des} is the autoscaled value (using the mean and standard deviation of the original responses) of the desired response 0.0302. Solving the equation by least squares provides a unique solution $(t, 0, 0, 0)$, the one with only one score non-null. The reconstructed settings $\hat{\mathbf{x}}_t$ define an unfeasible solution because it is outside the domain defined by the range of \mathbf{X} , although it has null Q -residual and the value for the T^2 statistic is less than the threshold value. Therefore, another solution of eq. (5) is needed related to a feasible solution when reconstructing the corresponding predictor variables. We denote this solution as $\hat{\mathbf{X}}_{des}$.

Undoing the scaling, their corresponding process/feature variables are written in the second row in table 5. For comparative purposes, row 1 contains the actual values of the

variables with which 0.0302 of SiO₂ content in AL(OH)₃ was obtained. We see that the settings in row 1 (experimental) and row 2 (calculated) are not very different.

In addition, the last two rows in table 5 describe the domain established for the problem, that is, the range of each process variable in the training set of historical data.

Provided that we follow the rows of the matrices depicted in figure 7, we can add any of them to $\hat{\mathbf{x}}_{des}$ without modifying the predicted y_{des} . However, again, even in this case that we made sure that both matrices \mathbf{X}_0 and \mathbf{X}_r are in the domain and in the PLS-box, we still do not have any guarantee that the addition of these points to an actual $\hat{\mathbf{x}}_{des}$ (no null) will provide feasible solutions. It only means that the addition of such points does not alter the predicted response.

In other words, the theoretical property of being in $\ker(L)$, by itself, does not give any 'hint' about the feasibility of the solutions obtained when adding $\hat{\mathbf{x}}_{des}$ to one of these points, irrespective of whether it belongs to \mathbf{W} -null space or comes from \mathbf{Q} -null space.

Table 5. Settings for the seven features to obtain SiO₂ content of 0.0302. The first row contains, in fact, the experimental conditions in the training set.

#	X_1	X_2	X_3	X_4	X_5	X_6	X_7
1	108.8	89.6	627	117.6	77.6	2.00	1.60
2	110.1	85.3	626	117.1	81.8	2.37	2.33
3	107.0	83.8	577	118.5	81.7	1.60	1.52
4	109.2	87.9	623	118.2	79.3	2.06	1.62
5	106.7	81.0	570	116.8	83.9	1.77	2.38
6	111.5	86.8	636	119.9	81.2	2.32	1.34
7	110.7	83.4	623	117.8	83.7	2.44	2.32
8	113.7	90.8	645	120.6	77.4	1.48	2.27
9	107.8	89.8	640	117.4	77.9	2.04	1.54
10	103.4	89.2	611	118.3	77.4	2.45	1.34
11	117.1	90.2	635	117.8	77.4	1.32	2.12
12	115.2	90.0	645	116.3	77.8	1.47	1.85
13	113.8	88.6	611	113.3	77.5	1.77	1.36
14	114.6	90.0	632	117.7	77.4	1.53	1.96
Minimum	103.4	81.0	570	113.3	77.4	1.32	1.34
Maximum	117.1	90.8	645	120.6	83.9	2.45	2.38
Minimum in domain	94.5	78.8	193	104.6	77.3	0.20	1.31
Maximum in domain	119.4	101.4	647	123.1	91.8	2.77	4.14

In fact, after imposing the constrains for the vectors obtained by adding elements of \mathbf{X}_0 and \mathbf{X}_r and sums of them to $\hat{\mathbf{x}}_{des}$, rows 3 to 14 in table 5 show some of the resulting feasible settings for the variables that are expected to give SiO₂ content of 0.0302. Rows 15 and 16 summarize the range of these solutions.

Comparing the range in rows 15 and 16 with the allowed range (last two rows), we can see that the design space related to the minimum content of byproduct is a subregion of the hypercube we have defined as the domain to work inside. In fact, this region is also a hypercube because any convex combination of rows 2-14 in table 5 also provides a feasible solution of the inversion problem we are tackling.

In terms of the process analytical technology (PAT), table 5 can be seen as the possibilities to use the process itself to correct deviations in some process variables (or in some feature variables) above all those which are not directly controllable.

For example, say that for some reason, the ratio of SiO_2 and Al_2O_3 concentrations (X_3) decreases from 623 to 570, as from row 4 to row 5. Then, following the mentioned rows, we can compensate this decrease by reducing both the total alkaline concentration of the original solution (X_1) until 106.7 and the aluminum concentration of the original solution, X_2 , to 81; together with a slight reduction of X_4 , the total alkaline concentration of the solution after discontinuing the connection of CO_2 , and less waiting time before connecting the CO_2 , X_6 . At the same time X_5 , the decomposition rate, should increase up to 83.9 and longer time connecting the CO_2 , X_7 , up to 2.38.

Besides, not necessarily the selected values of the variables must be exactly one of those written in table 5. We have already said that any convex combination also provides a feasible solution. In the process, that means that we can use the information in table 5 for any value for X_3 between 570 (minimum) and 645 (maximum), provided we maintain the relation among rows. For example, $X_3 = 615$ is not in table 5 but it is the result of the convex combination $570 \times 0.4 + 645 \times 0.6$ (values in rows 5 and 8). Then, with the same convex combination, the remaining process variables should be $X_1 = 110.9$, $X_2 = 86.9$, $X_4 = 119.1$, $X_5 = 80.0$, $X_6 = 1.60$ and $X_7 = 2.31$, and we use the controllable process variables to compensate the variability in the raw material.

3.3 Supercritical Carbon Dioxide Extraction

Zhang et al. [41] describe the optimization of a process of oil extraction using supercritical carbon dioxide. Table 1 in [41] shows the settings of three process variables and the corresponding responses (experimental results) obtained when conducting a Box-Behnken design in a cubic domain to fit a second order polynomial model, in the context of the well-known experimental strategies inside the so-called Response Surface Methodology (for details about RSM, see for example, [42]).

The process depends on three factors (process variables) that are described in the first three rows of table 6 along with the minimum and maximum values of each process variable, which define the experimental domain.

The last row in table 6 corresponds to the response, namely the percentage of oil yield of *N. glandulifera* seed. Table 6 also shows the range of the percentage of oil yield obtained with the experiments, as reported in the original paper.

Table 6. Experimental domain, response and optimal conditions determined in ref. [41] for the process to obtain oil yield of *N. glandulifera*

Variable	Factor/Response	Minimum	Maximum	Conditions and expected maximum of Y
X_1	Pressure (MPa)	22	35	30.84
X_2	Temperature (°C)	35	45	40.57
X_3	CO ₂ flow rate (L h ⁻¹)	18	22	22.00
Y	Oil yield (%)	31.9	37.1	38.19

The maximum oil yield obtained experimentally is 37.1 %. However, with the aid of the fitted model, the authors report a maximum yield attainable of 38.19 % with the settings shown in the last column of table 6. We see that the CO₂ flow rate is 22 L h⁻¹, the maximum allowed value, thus, the settings computed are at the boundary of the experimental domain.

For comparative purposes but to avoid working at the boundary of the domain, we will set our target value, y_t , equal to 38 %, which still would be an increase compared to the best oil yield obtained experimentally.

To look for the corresponding process variables, we will invert a PLS model fitted with the same data (available in the paper itself). The model will be of degree 2, which means that there is not a linear relationship between the three process variables and the oil yield. It also means that the algebraic inversion is not available. We will develop this idea in the following paragraphs.

To fit a second order model by using a linear method like PLS, it is not enough to consider just the three process variables. Predictor matrix \mathbf{X} is in fact made by 9 columns corresponding to the settings of the process variables and the additional columns with cross products and squares, according to the second addend that appears in the model in eq. (13). Both \mathbf{X} and \mathbf{y} are autoscaled so no need to consider an independent term for the prediction model.

$$y = \sum_{i=1}^3 \beta_i X_i + \sum_{i=1}^3 \sum_{j=i}^3 \beta_{ij} X_{ij} \quad (13)$$

With these matrices \mathbf{X} - \mathbf{y} , a PLS model was fitted with 3 latent variables that explain 33.95 % of the variance of the predictor variables and 99.04 % of the variance of the response. Figure 8 is the three-dimensional latent space with the scores of \mathbf{X} depicted, like in the remaining case-studies, by green filled diamonds.

Figure 8 around here

In terms of the PLS model, $p = 9$, $a = 3$ and $q = 1$, so the \mathbf{Q} -null space inside the latent space is two dimensional (i.e., a plane in the three-dimensional latent space).

3.3.1 Algebraic inversion

We start by algebraically inverting the model following the procedure detailed at the beginning of section 3 and summarized in fig. 1, with the loadings of \mathbf{y} and the desired value \mathbf{y}_t . Then, adding vectors of the \mathbf{Q} -null space (step 4 when $a > 1$) we obtain the scores depicted also in figure 8 as red points, where only the bigger ones belong to the PLS-box.

In any case, all the red points share the property of having \mathbf{y}_t as their predicted response, thus, the plane they all lie on is the 2-dimensional \mathbf{Q} -null space already mentioned corresponding to \mathbf{y}_t .

We also see that there are no scores of \mathbf{X} lying on this plane. This is so because the selected desired value \mathbf{y}_t is not among the values already in \mathbf{y} . In fact, we are primarily trying to determine whether there are values of the process variables to obtain \mathbf{y}_t and, in this case, to describe the design space around them.

This design space is inside the three-dimensional experimental domain depicted in figure 9, where, again, the green filled diamonds are the settings of the experimental factors in the design conducted.

With eq. (7), step 5 in fig. 1, the settings of the experimental factors corresponding to the computed scores in fig. 8 that belong both to the experimental domain and the PLS-box are depicted in fig. 9 as red filled circles. The already mentioned extreme position of scores in the latent space places their corresponding process variables near the boundary of the domain, which explains the few values obtained.

The next step is to explore the variability due to the residuals of the projection onto the latent space, which was not taken into account during the inversion. This is described by considering points in the \mathbf{W} null space defined by the projection, for instance, by using the basis of the discarded loadings. Then, step 6 in fig. 1, adding up these points to the computed values during the inversion does not change the predicted response, but they add spreading in the \mathbf{X} -space, therefore, better describing the design space associated to the yield of $\mathbf{y}_t = 38\%$. These new conditions are depicted with small black dots in figure 9, after removing the solutions outside the experimental domain or the PLS-box, step 7 in the procedure.

Figure 9 around here

As we can see, the points in \mathbf{W} -null space (black dots) extend in lines parallel to the one defined by the points coming from \mathbf{Q} -null space (red filled circles) in a wider region of

the experimental domain. In any case, the most distinguishable common property is that the third factor (CO₂ flow rate is at high level).

3.3.2 Computational inversion for the target 38 %

The calculations made in section 3.3.1 are not completely right because the algebraic inversion of the PLS model provides in fact values for nine coordinates, computed without taking into account the precise relation that exists among the nine columns that made up the predictor matrix \mathbf{X} , relation defined from the first three columns following the model in eq. (13). Therefore, this precise relation among the last six coordinates of rows in \mathbf{X} with the first three is not maintained in any way when inverting the linear application.

Actually, the points depicted in fig. 9 are the first three coordinates of the computed solutions, dismissing the remaining six. To get an idea about the meaning of the comments in the previous paragraph, let us take the black and red points in fig. 9 and apply the PLS model to these values as if there were settings of the process variables. The predicted oil yield varies from 33.6 to 38.1 %, some of them far from the expected 38 % and, besides, with almost the same range of variation as the experimental values themselves, from 31.9 to 37.1 % (table 6).

The conclusion is that, if we should preserve the relation among the final columns (which is mandatory), then the mathematical inversion in the way described in fig. 1 is not possible.

Nevertheless, the inversion as proposed in [38] is still viable, because it works maintaining the search space defined with the problem. In this case, this domain is the three dimensional experimental domain where the factors (the process variables) vary. Once the domain is set, the procedure consists of an elitist genetic algorithm that always search for experimental conditions inside the domain defined by further imposing the restrictions about the Q and T^2 statistics. In that way, the algorithm evolves always using feasible settings for the process at hand.

Briefly (for details consult the referenced paper), the algorithm starts with an initial population of several different settings randomly selected in the domain of the three process variables and inside the PLS-box (i.e., values of Q -residuals less than 12.45 and values of T^2 less than 11.47, which are the corresponding limits at 95 % confidence level). Then, it evolves for a determined number of generations by creating new potential solutions, always in both the domain and the PLS-box, and with increasing percentages of the predicted oil yield.

Additionally, the initial population can be seeded to include any potential solution that we want (solutions already known due to previous runs, or the conditions of the design conducted, or some points generated from the algebraic solution of eq. (5), etc.). At the end, the better ones will survive the evolution process.

For the process at hand, and the scaled value corresponding to 38 % as target value, we run the algorithm with a population size of 100. The initial population is randomly chosen inside the experimental domain. In each subsequent generation, new 100 offsprings are computed by sequentially selecting (at random with uniform probability) pairs of solutions in the actual population and single crossover between each pair, forcing the offsprings to be different from the parents. Ten percent of the resulting offsprings are mutated inside the domain.

The new solutions are then evaluated in terms of the PLS model. From those solutions inside the domain and the PLS-box, the next population is made up by the 100 solutions with predicted values closest to the desired value. The evolution stops after 300 generations, with all the solutions in the final population predicting the target value.

The different settings for the process variables of this final population are depicted as magenta filled squares in fig. 9, and now they are discrete estimates of solutions in the design space for the targeted response.

Fig. 9 shows that these settings spread not far from the red circles (coming from the theoretical Q-null space) and near the region of possible spread of solutions due to the **W**-null space related to the residual space of the PLS model (black dots). However, besides predicting the same yield, the magenta points (true solutions of the inversion) do not define any recognizable mathematical relation, they do not follow the direction marked by the red points, neither the direction defined by the black points.

3.3.3 Computational inversion for maximizing oil yield

The computational procedure just explained evolves by searching settings inside the experimental domain with the goal of improving a given property. In section 3.3.2 the criterion was to approach 38 % of yield. However, the goal was to find the best possible percentage of yield inside the experimental domain.

Therefore, by defining a new criterion to drive the evolution, we can just invert the same PLS model but not necessarily to predict a given value but directly looking for the maximum oil yield achievable in the domain. The modification of the criterion implies a slight modification with respect to the evolution process already explained: the updating of the population in each generation is made by selecting the points with largest predictions to survive to the next generation.

When using the criterion of maximizing oil yield in this domain, the population converges to a single solution with an estimated oil yield of 38.19 % with the same process variables as the one in the last column of table 6. These experimental conditions are depicted in fig. 9 by a blue filled triangle, which is around the magenta squares, farther from the filled circles.

3.4 Vapor pressure of a distillation tower (larger data set)

The data are available at <https://openmv.net/info/distillation-tower>, where it is stated that they come from an unspecified industrial source. They correspond to measurements on 26 variables from a distillation column, measured over two and a half years. It is also explained that variable names are coded so that details of the process are not disclosed. The names provided are in table 7 where we can only identify the type of variables: there are temperatures, flow, different pressures, etc., the process conditions when each of the 253 samples were taken. The range of these variables is also in table 7 to see the defined domain for the process variables. The last row, Vapour Pressure, is a quality variable measured in the laboratory. Therefore, $p = 26$, $q = 1$.

Table 7. Process variables as well as their range for the distillation tower case-study. The last row contains the quality (response) variable Y .

Number	Variable	Minimum	Maximum
X_1	Temp1	111.03	168.04
X_2	FlowC1	174.74	505.67
X_3	Temp2	318.45	406.74
X_4	TempC1	75.59	132.20
X_5	Temp3	420.78	528.08
X_6	TempC2	419.02	526.39
X_7	TempC3	119.21	229.41
X_8	Temp4	124.60	235.62
X_9	PressureC1	199.82	254.59
X_{10}	Temp5	418.62	524.79
X_{11}	Temp6	419.30	525.31
X_{12}	OC1	1.86	8.50
X_{13}	Temp7	890.11	1179.20
X_{14}	Temp8	421.45	557.36
X_{15}	TempC9	34.51	99.98
X_{16}	FlowC2	61.14	83.59
X_{17}	Temp9	400.32	507.31
X_{18}	Temp10	432.45	547.13
X_{19}	FlowC3	4.62	8.85
X_{20}	FlowC4	5.85	10.76
X_{21}	Temp11	12.89	35.53
X_{22}	Temp12	419.43	526.14
X_{23}	InvTemp1	1.90	2.38
X_{24}	InvTemp2	2.46	3.14
X_{25}	InvTemp3	1.97	2.50
X_{26}	InvPressure1	3.71	4.66
Y	Vapour Pressure	25.16	64.31

Process variables are correlated, a few of them highly, approximately 13 % of the correlation coefficients are, in absolute value, greater than 0.9. However, 61 % are less

than 0.5 (again, in absolute value). When taking into account the response quality variable to be predicted with PLS, RMSECV (Root Mean Square Error in Cross Validation) maintains a slow decrease up until seven latent variables. From the third latent variable the variance explained by the latent variables is mostly from \mathbf{X} . In any case, the selected PLS model requires seven latent variables ($a = 7$), after removing 18 samples because of their standardized residuals (outside the interval $[-3, 3]$) or because of their high values of Q -residual or/and T^2 statistics. The resulting model explains 95.7 % of the variance of \mathbf{X} with 97.3 % of the variance of \mathbf{Y} , with crossvalidated coefficient of determination of 96.9 %, estimated extracting random samples with 15 splits and 5 iterations.

The \mathbf{Q} -null space is thus a six-dimensional space in the latent space, whereas the \mathbf{W} -null space is in the 19-dimensional space. By directly estimating points from the null spaces, without a target quality value, figure 10 is the parallel coordinates plot for the points already inside the input space (26 dimensions) for matrices \mathbf{X}_0 in blue continuous line and \mathbf{X}_r in dotted red lines. \mathbf{X}_0 comes from the \mathbf{Q} -null space following equation (7) with null target scores, while \mathbf{X}_r contains points in \mathbf{W} -null space, computed with equation (9). In both cases, the points outside the domain or outside the PLS-box were removed.

Figure 10 around here

It is clear that there are much more variation allowed around the null value of any variable when taking points of the null space coming from the latent space, blue lines, than when the points correspond to the null space in the residual space, in red, which is consistent with the few variance of \mathbf{X} in the residual space, despite the fact of being a subspace of dimension 19.

Still, there is some variables around which the variation allowed without modifying the predicted response is larger due to \mathbf{W} -null space in red, such as variables $X_5, X_6, X_{10}, X_{11}, X_{17}, X_{22}, X_{23}$ and X_{25} .

4. Conclusions

When working with linear models, such as PLS, part of the estimated design space is related to the kernel defined by the linear mapping. Through some case studies, the present work shows the need to consider the part of the kernel due to the orthogonal residual space, along with the one resulting from the projection (latent variables) space.

Case 3.1 highlights the need to take into account what we have called the \mathbf{W} -null space, and the close relationship between the training data and the fitted PLS model.

The a-priori description of the kernel associated to the PLS model, directly with the regression coefficients or in two steps exploring both \mathbf{Q} - and \mathbf{W} -null spaces, does not help in constructing feasible solutions. This is explored in case 3.2 that focuses on the fact that, although the kernel is a characteristic of the linear mapping, its impact on the

design space depends on the desired quality characteristic. Besides, PAT framework includes the use of the process to compensate for deviations in some values of the process or feature variables (above all for the variables that cannot be directly modified). Case 3.2 exemplifies how to do it with the information obtained from the points in the kernel of the corresponding PLS model to fit the content of a byproduct in a step of a production process of $\text{Al}(\text{OH})_3$.

In Case 3.3 with a process for extracting oil yield of *N. glandulifera*, PLS is used to describe a nonlinear relation between process or features variables and quality characteristics. Therefore, there is not algebraic solution for the inversion, and the exploration of the design space is based on a computational inversion.

Case 3.4 illustrates the situation for larger dimensions, the estimation procedure is the same, but the visualization of the convex space defined inside the design space where the predicted response does not change is not viable in Cartesian coordinates.

Acknowledgements

This work was supported by the European Regional Development Fund (FEDER) through Spanish Agencia Estatal de Investigación (project CTQ2017-88894-R).

References

- [1] K. Faber, B.R. Kowalski, Propagation of measurement errors for the validation of predictions obtained by principal component regression and partial least squares, *J. Chemometrics* 11 (1997) 181-238.
- [2] P. Facco, F. Dal Pasto, N. Meneghetti, F. Bezzo, M. Barolo, Bracketing the Design Space within the Knowledge Space in Pharmaceutical Product Development, *Ind. Eng. Chem. Res.* 54 (2015) 5128–5138.
- [3] P. Van Overschee, B. De Moor, *Subspace Identification for Linear Systems Theory Implementation Applications* (1996) Kluwer Academic Publishers.
- [4] Lu CY., Min H., Zhao ZQ., Zhu L., Huang DS., Yan S., Robust and Efficient Subspace Segmentation via Least Squares Regression. In: Fitzgibbon A., Lazebnik S., Perona P., Sato Y., Schmid C. (eds) *Computer Vision – ECCV 2012*. ECCV 2012. Lecture Notes in Computer Science, vol 7578. Springer, Berlin, Heidelberg.
- [5] S.X. Ding, Data-driven design of monitoring and diagnosis systems for dynamic processes: A review of subspace technique based schemes and some recent results, *Journal of Process Control* 24 (2014) 431–449.
- [6] Gang Li, S. Joe Qin, Donghua Zhou, Geometric properties of partial least squares for process monitoring, *Automatica* 46 (2010) 204-210.
- [7] V.-H. Nguyen, C. Rutten, J.-C. Golinval, Comparison of subspace analysis methods for fault detection in industrial systems, *Proceedings of ISMA2010 including USD2010*, (2010) 985-998.

- [8] S. Fan, An Introduction to Krylov Subspace Methods A Less Mathematical Way to Understand, arXiv:1811.09025v1 [math.OC] (2018).
- [9] J. Liesen, Z. Strakos, Krylov Subspace Methods. Principles and Analysis, (2013) Oxford University Press, Oxford.
- [10] I. Fernández de Bustos, V. García-Marina, G. Urkullu, M. Abasolo, An efficient LDU algorithm for the minimal least squares solution of linear systems, *Journal of Computational and Applied Mathematics* 344 (2018) 346–355.
- [11] B.N. Datta, Krylov subspace methods for large-scale matrix problems in control, *Future Generation Computer Systems* 19 (2003) 1253–1263.
- [12] B.N. Datta, Numerical methods for linear control systems design and analysis, (2004) Elsevier Academic Press, San Diego.
- [13] H. Wold, Estimation of principal components and related models by iterative least squares. In: Krishnaiah, P.R. (ed.) *Multivariate Analysis*, pp. 391–420. (1966) Academic, New York.
- [14] Y Takane, S. Loisel On the PLS Algorithm for Multiple Regression (PLS1), in H. Abdi, V. Esposito Vinzi, G. Russolillo, G. Saporta, L. Trinchera (eds) *The Multiple Facets of Partial Least Squares and Related Methods* (2016) Springer.
- [15] A. Björkström Krylov Sequences as a Tool for Analysing Iterated Regression Algorithms, *Scandinavian Journal of Statistics* 37 (2010) 166–175.
- [16] T. N. Tran, L. Blanchet, N.L. Afanador, L. M.C. Buydens, Novel unified framework for latent modeling and its interpretation, *Chemometrics and Intelligent Laboratory Systems* 149 (2015) 127–139.
- [17] H. Martens, T. Naes, *Multivariate Calibration*, (1989) John Wiley & Sons, Chichester.
- [18] S. Wold, E. Johansson, M. Cocchi, PLS-partial least-squares projections to latent structures, pp. 523–550 in H. Kubinyi (ed.) *3D QSAR in Drug Design*, (1994) Kluwer/Escom Springer.
- [19] S. Wold, L. Eriksson, J. Trygg, N. Kettaneh, The PLS method -partial least squares projections to latent structures- and its applications in industrial RDP (research, development, and production PLS). *Industrial RPD for Prague 1* (2004), 1-44.
- [20] J. Engel, J. Gerretzen, E. Szymanska, J. J. Jansen, G. Downey, L. Blanchet, L. M. C. Buydens, Breaking with trends in pre-processing? *Trends Anal. Chem.* 50 (2013) 96–106.
- [21] C.Mas, L.Rubio, L.Valverde-Som, L.A.Sarabia, M.C.Ortiz, Impact of the pretreatment of ATR-FTIR signals on the figures of merit when PLS is used, *Chemometrics and Intelligent Laboratory Systems* 201 (2020) 104006.
- [22] S. Wold, H. Antti, F. Lindgren, J. Ohman, Orthogonal signal correction of near-infrared spectra, *Chemometrics and Intelligent Laboratory Systems* 44 (1998) 175– 185.
- [23] H.C. Goicoechea, A.C.Olivieri, A comparison of orthogonal signal correction and net analyte preprocessing methods. Theoretical and experimental study. *Chemom Intell Lab Syst.* 56 (2001) 73–78.
- [24] C.A. Andersson, Direct orthogonalization, *Chemometrics and Intelligent Laboratory Systems* 47 (1999) 51– 63.
- [25] J. Trygg, S. Wold, Orthogonal projections to latent structures (OPLS), *Journal of Chemometrics* 16 (2002) 119–128.
- [26] R. Ergon, PLS post-processing by similarity transformation (PLS+ST): A simple alternative to OPLS. *J. Chemom.* 96 (2005) 1–4.

- [27] T. Fearn, On orthogonal signal correction, *Chemometrics and Intelligent Laboratory Systems* 50 (2000) 47– 52.
- [28] U.G Indhal, The O-PLS methodology for orthogonal signal correction -is it correcting or confusing? *Journal of Chemometrics* 34 (2020) e2884.
- [29] H. Yu, J. F. MacGregor, Post processing methods (PLS–CCA): simple alternatives to preprocessing methods (OSC–PLS) *Chemometrics and Intelligent Laboratory Systems* 73 (2004) 199– 205.
- [30] Y. Qin, C. Zhao, X. Wang, F. Gao, Subspace decomposition and critical phase selection based cumulative quality analysis for multiphase batch processes. *Chemical Engineering Science* 166 (2017) 130–143.
- [31] T. Kourti, *Multivariate Statistical Process Control and Process Control, Using Latent Variables in S.D. Brown, R. Tauler, B. Walczak (eds.) Comprehensive Chemometrics Vol. 4, pp. 21-54.*
- [32] A.M. Yan, F. Marin, J.C. Golival, Null-Subspace Analysis for Structural Damage Monitoring in Fatigue Testing of Luminaires, proceedings of ISMA2004, 557-587.
- [33] A.M. Yan, J.C. Golival, Null subspace-based damage detection of structures using vibration measurements, *Mechanical Systems and Signal Processing* 20 (2006) 611–626.
- [34] A. R. M. Rao, V. Kasireddy, N. Gopalakrishnan, K. Lakshmi, Sensor fault detection in structural health monitoring using null subspace–based approach, *Journal of Intelligent Material Systems and Structures* 26 (2015) 172–185.
- [35] C.M. Jaeckle, J.F. MacGregor, Industrial applications of product design through the inversion of latent variable models, *Chemometrics and Intelligent Laboratory Systems* 50 (2000) 199–210.
- [36] G. Bano, P. Facco, N. Meneghetti, F. Bezzo, M. Barolo, Uncertainty back-propagation in PLS model inversion for design space determination in pharmaceutical product development, *Computers & Chemical Engineering* 101 (2017) 110-112.
- [37] M. Ottavian, E. Tomba, M. Barolo, Advanced process decision making using multivariate latent variable methods, pp. 159-189 in M. G. Ierapetritou, R. Ramachandran (Eds.), *Process simulation and data modeling in solid oral drug development and manufacture*, (2016) Humana Press, Springer, New York.
- [38] S. Ruiz, M.C. Ortiz, L.A. Sarabia, M.S. Sánchez, A computational approach to partial least squares model inversion in the framework of the process analytical technology and quality by design initiatives, *Chemometrics and Intelligent Laboratory Systems* 182 (2018) 70–78.
- [39] S. Lang, *Introduction to Linear Algebra*, second Ed., 1986, Springer-Verlag New York Inc.
- [40] Z. Cheng, E. Zhu, N. Chen, Application of orthogonal expansion to mapping and modelling, *Journal of Chemometrics* 7 (1993) 243-253.
- [41] J-P. Zhang, X-L. Hou, T. Yu, Y. Li, H-Y. Dong, Response Surface Optimization of *Nigella glandulifera* Freyn Seed Oil Yield by Supercritical Carbon Dioxide Extraction, *Journal of Integrative Agriculture* 11 (2012) 151-158.
- [42] L.A. Sarabia, M.C. Ortiz, Response Surface Methodology, pp. 345-390 in Brown S, Tauler R, Walczak R (eds.) *Comprehensive Chemometrics*, 1st Edition, (2009), Elsevier, Oxford.

Figure captions

Figure 1. Diagram that summarizes the procedure of adding points in null spaces to the inversion of a PLS model to predict a given value y_{des} .

Figure 2. Three-dimensional input space in Case 3.1.1. Green diamonds represent the samples in \mathbf{X} , the red filled circle is $\hat{\mathbf{x}}_{des}$, black points belong to the \mathbf{W} -null space. The convex hull of the black points is in gray.

Figure 3. Latent variables space in Case 3.1.2. Green diamonds are the scores of \mathbf{X} , the red filled circle is the LS solution of eq. (5) and the blue filled squares follow the line along which the predicted response is the same. The ellipse marks the limit of the T^2 statistic (95 % confidence level).

Figure 4. Three-dimensional input space in Case 3.1.2. Green diamonds represent the samples in \mathbf{X} , the red filled circle is $\hat{\mathbf{x}}_{des}$, blue filled squares are the predictor variables that correspond to the \mathbf{Q} -null space, black points are in the \mathbf{W} -null space. In gray, their convex hull.

Figure 5. Latent variables space in Case 3.1.3. Green diamonds are the scores of \mathbf{X} , blue squares are points in the \mathbf{Q} -null space and the red circle is the LS solution of eq. (5). The ellipse marks the limit of the T^2 statistic at 95 % confidence level.

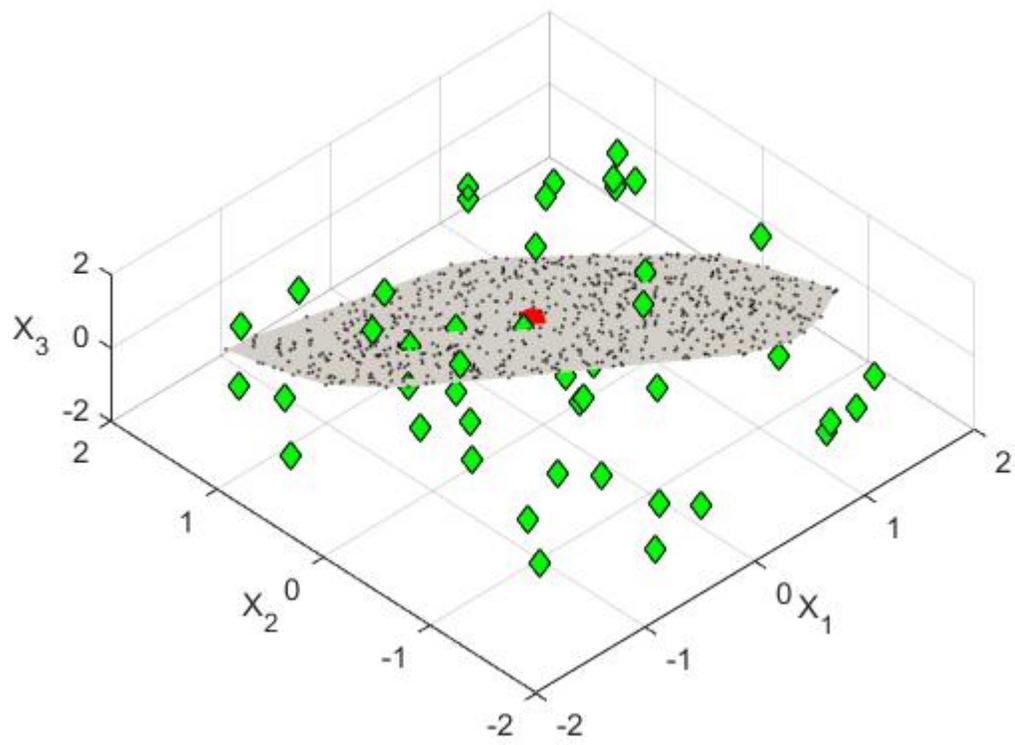
Figure 6. Three-dimensional input space in Case 3.1.3. In yellow, the region of feasible solutions inside the design space for y_{des} . Green diamonds represent the points in \mathbf{X} . Blue squares come from the \mathbf{Q} -null space. Black points come from the \mathbf{W} -null space.

Figure 7. Case 3.2. Parallel coordinates plot of \mathbf{X} in red dashed lines. In blue some settings: (a) coming from \mathbf{Q} -null space, and (b) belonging to \mathbf{W} -null space.

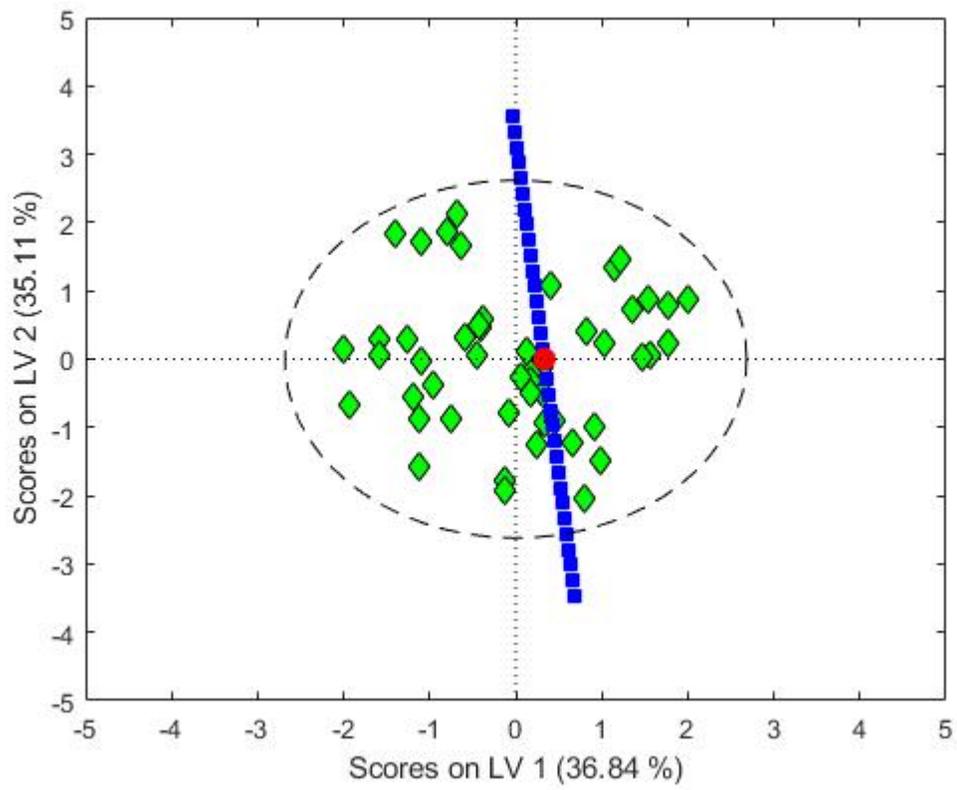
Figure 8. Three-dimensional scores space in Case 3.3. Green diamonds are the scores of \mathbf{X} , red points are inside \mathbf{Q} -null space, with those bigger (filled circles) also inside the PLS-box (with values of T^2 below the 95 % confidence limit).

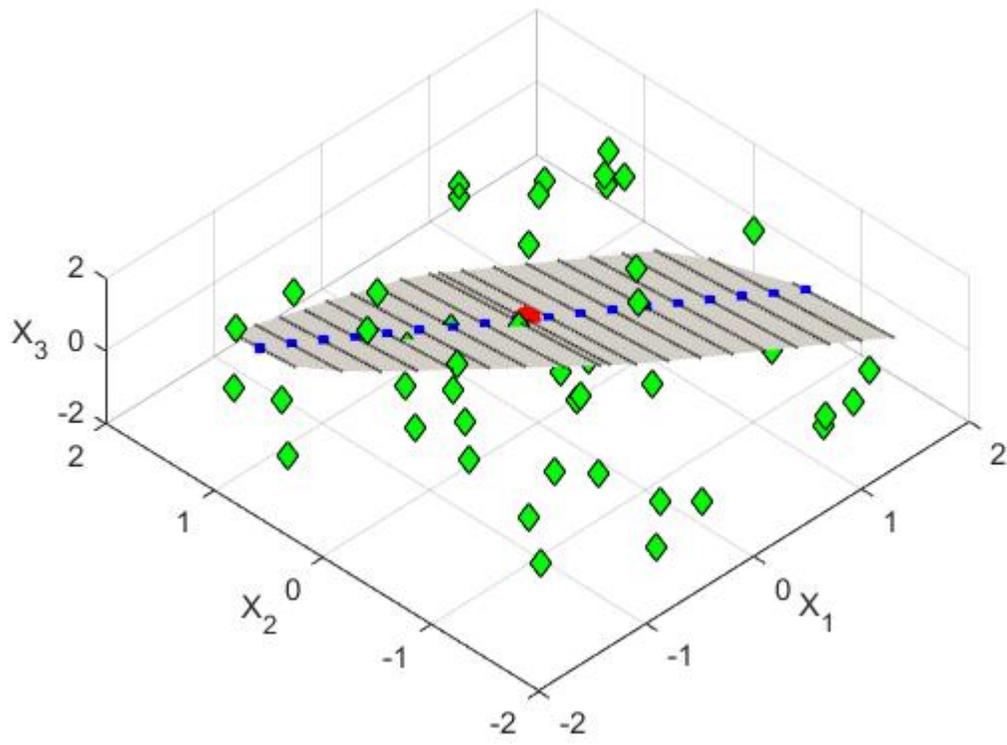
Figure 9. Experimental domain in Case 3.3. Green diamonds are the experiments conducted. Red circles correspond to the red circles in fig. 8 (\mathbf{Q} -null space), and black points are in \mathbf{W} -null space. Magenta squares are the points found with the computational inversion in section 3.2.2. The blue triangle marks the settings obtained with the computational inversion for maximizing yield in section 3.2.3.

Figure 10. Process variables space for case 3.4. Parallel coordinates plot for points in the kernel of the fitted PLS model. Blue continuous lines are for points from \mathbf{Q} -null space, dotted red lines are from points in \mathbf{W} -null space.

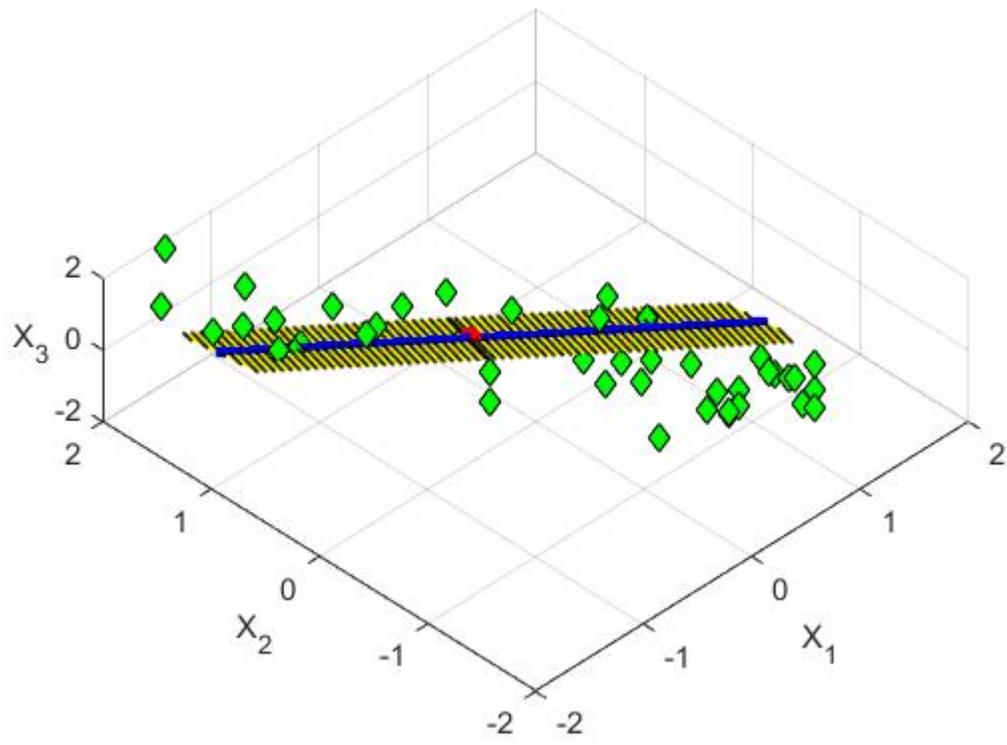


Journal

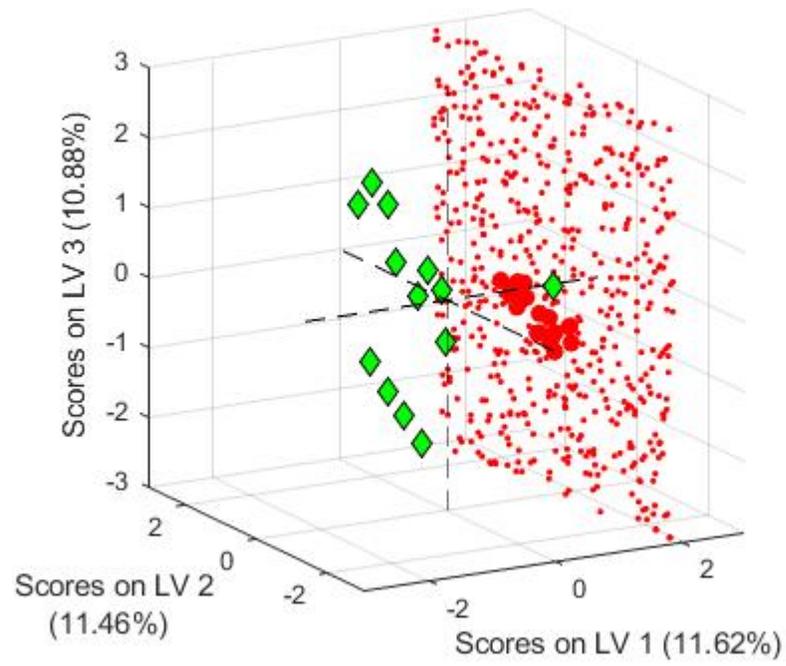




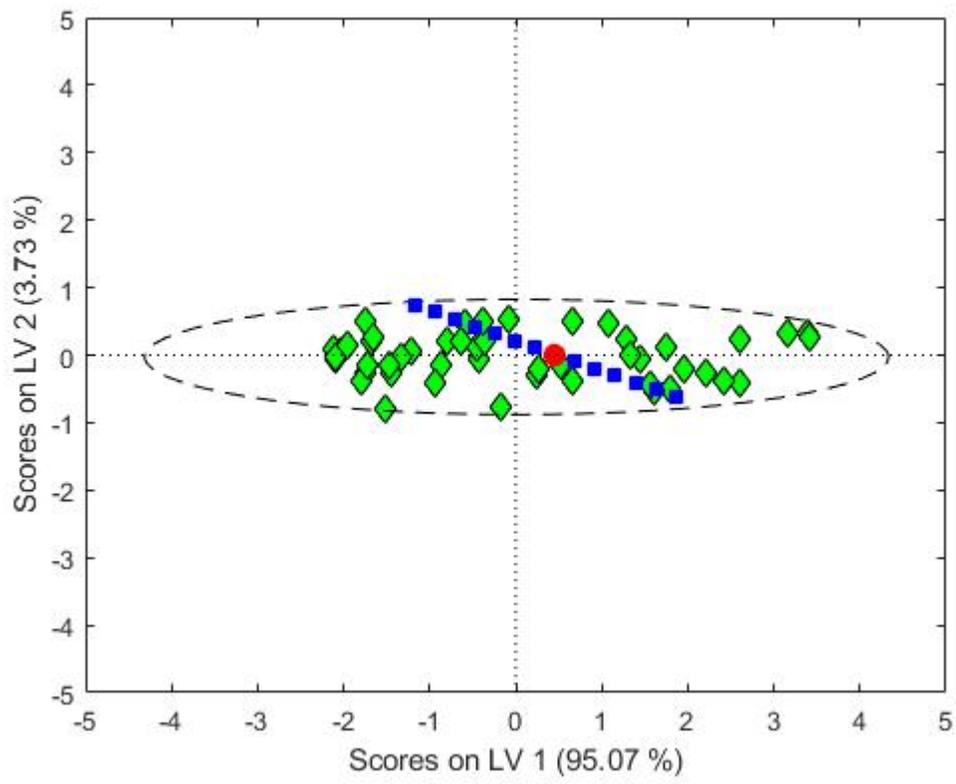
Journal



Journal



Journal



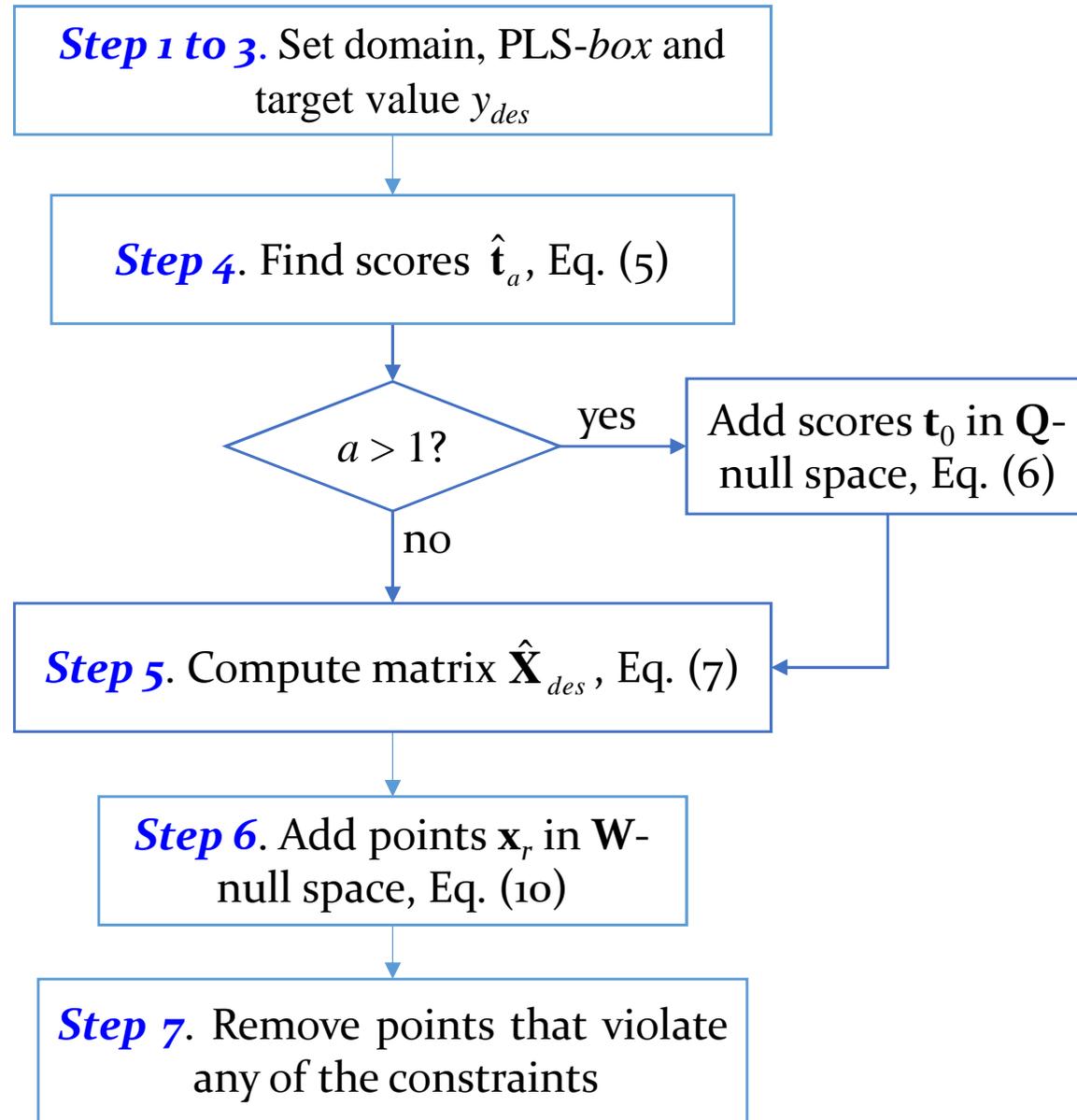
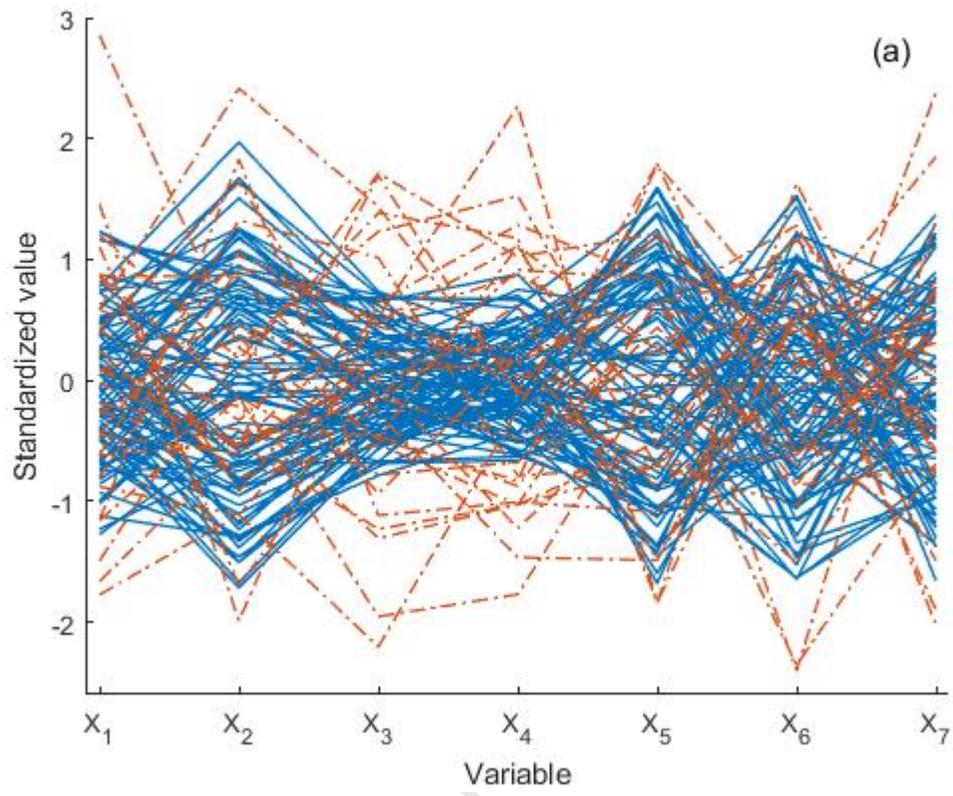
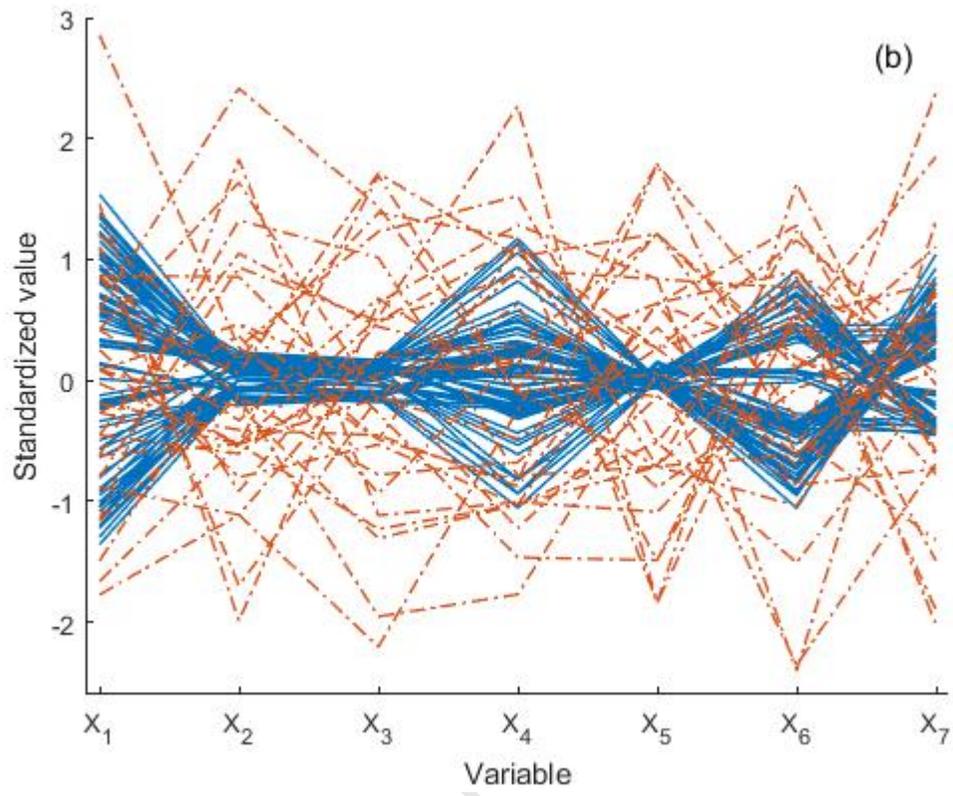
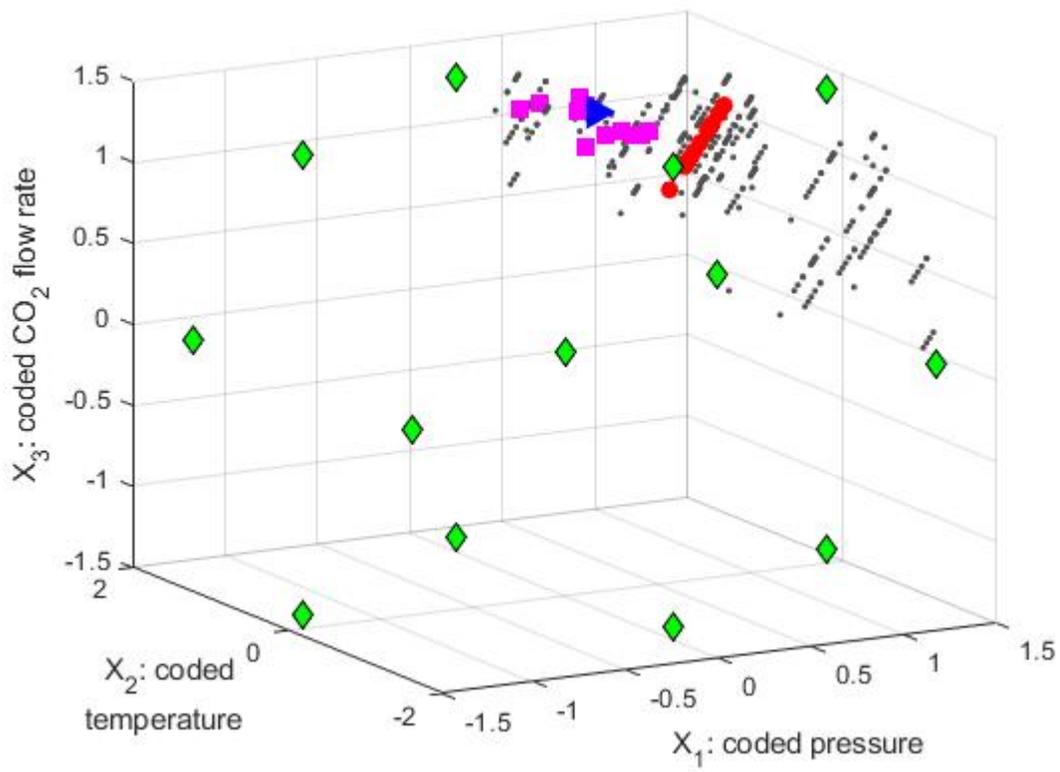
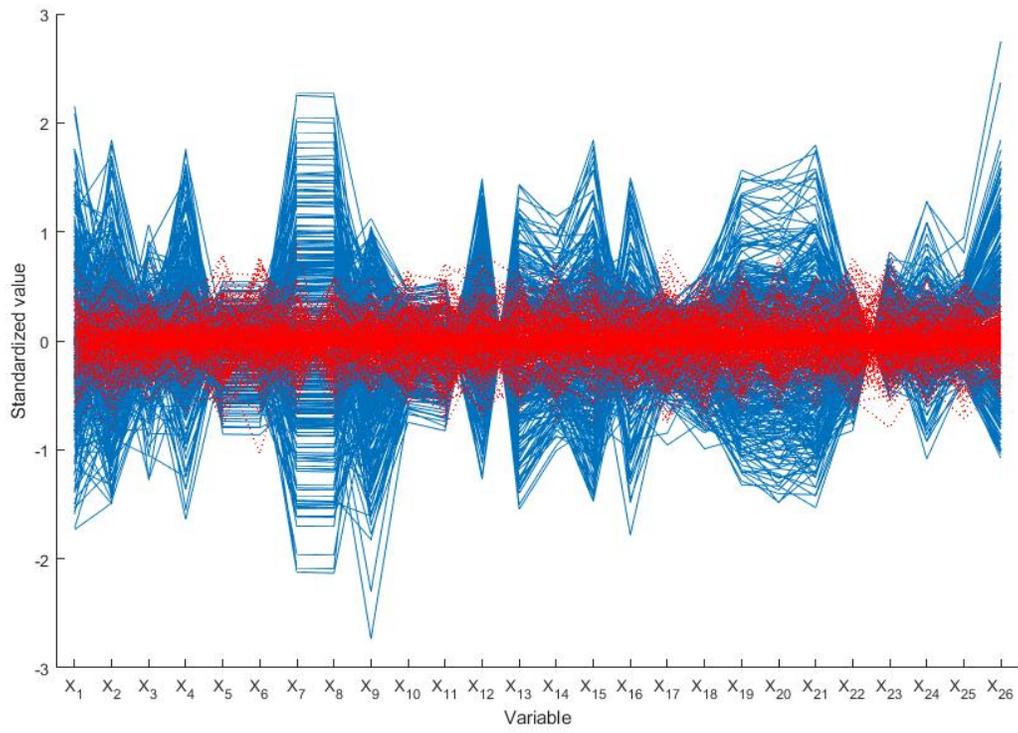


Fig.1









Inversion of PLS models to keep a target quality in process control

Definition of design space from the null spaces defined by PLS

Explicit consideration of the variability due to the discarded latent variables

Computational inversion when a second-order polynomial is fitted with PLS

The design space to compensate uncontrollable with controllable process variables

Journal Pre-proof

Declaration of interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests:

Journal Pre-proof