Available online at www.sciencedirect.com
SCIENCE DIRECT®
Chemometrics and Intelligent Laboratory Systems 81 (2006) 165–179
Chemometrics and intelligent laboratory systems
www.elsevier.com/locate/chemolab

Process control and optimization with simple interval calculation method

Alexey Pomerantsev [a,*], Oxana Rodionova [a], Agnar Höskuldsson [b]

[a] Institute of Chemical Physics, Kosygin Str. 4, Moscow, 119991, Russia
[b] Technical University of Denmark, Building 358, 2800 Lyngby, Denmark

Received 20 June 2005; received in revised form 28 November 2005; accepted 2 December 2005
Available online 14 February 2006

Abstract

Methods of process control and optimization are presented and illustrated with a real world example. The optimization methods are based on the PLS block modeling as well as on the simple interval calculation methods of interval prediction and object status classification. It is proposed to employ the series of expanding PLS/SIC models in order to support the on-line process improvements. This method helps to predict the effect of planned actions on the product quality and thus enables passive quality control. We have also considered an optimization approach that proposes the correcting actions for the quality improvement in the course of production. The latter is an active quality optimization, which takes into account the actual history of the process. The advocate approach is allied to the conventional method of multivariate statistical process control (MSPC) as it also employs the historical process data as a basis for modeling. On the other hand, the presented concept aims more at the process optimization than at the process control. Therefore, it is proposed to call such an approach as multivariate statistical process optimization (MSPO).

Keywords: MSPC; Optimization; Expanding PLS modeling; SIC approach; MSPO

1. Introduction

Multivariate statistical process control (MSPC) is nowadays a very popular approach that helps to understand and to run real-world technological processes [1–5]. In order to secure the quality of final products, authorities in different countries provide recommendations or requirements concerning the process control. A good example is the PAT initiative [6] presented by US Food and Drug Administration as a draft guidance for industry. MSPC combines the old-known statistical methods such as statistical process control (SPC, e.g., the Sheward cards [7]) with modern multivariate data analysis techniques (MDA, e.g., PLS [8,9]) in order to produce new knowledge about the process in question. This knowledge gives an easy way to monitor and control the process, but it does not offer a method to *optimize* the process performance. However, the general aim of any statistical analysis of technology is to improve or stabilize the final product quality or/and to reduce production costs.

The main MSPC concept is to apply historical data on performance attributes (**X** matrix) for construction of a *linear* (calibration) model, which explains how the final results (*y*-vector) depend on the observed *X* variables and to verify that the process is remaining in a 'state of statistical control'. Studying this model, one can suggest a program of actions that can improve performance *in general*. However, this is a *post factum* optimization, while the most important issue in production is an in situ optimization, which prescribes immediate actions in the course of production in order to correct its current state and to improve the future. In the paper, the MSPC concept is extended in order to develop an approach for the in-line process optimization. This approach may be termed as multivariate statistical process optimization (MSPO).

Two mathematical methods are implemented in this work. The first one is the PLS regression [8,9] applied to build various calibration models. The second technique is simple interval calculation (SIC). This is a method of linear modeling that gives the result of prediction directly in the interval form [10,11] and also provides wide possibilities for the leverage-type *object status classification*. In Section 2, we present a brief description of these methods. Section 3 introduces a real-world data that are collected in a batch process fit within the present framework. In

* Corresponding author.
E-mail address: forecast@chpr.ras.ru (A. Pomerantsev).

0169-7439/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.chemolab.2005.12.005

Section 4, a method of passive optimization is considered. It is performed using a series of the expanding PLS models combined with the SIC interval estimation. Section 5 proposes a method of active optimization. It is based on the block PLS modeling and the SIC object status classification. We consider different optimization strategies and illustrate the theory with a case study, based on the batch process described in Section 3.

2. Mathematical methods

In this section, we present a brief description of mathematical methods used in the paper. They are PLS regression and SIC method.

2.1. PLS method

Two versions of PLS algorithm are employed in the paper. They are: PLS1 that is used for the single-response regression and its multi-response extension, known as PLS2 method. There are numerous papers and tutorials on PLS published, e.g., see Refs. [8,9] and references cited herein.

Let us consider a linear regression model

$$\boldsymbol{y} = \mathbf{X}\boldsymbol{a} + \boldsymbol{\varepsilon} \tag{1}$$

where $\boldsymbol{y}$ is the n-dimensional response vector, $\boldsymbol{a}$ is the p-dimensional vector of unknown parameters, $\mathbf{X}$ is the $(n\times p)$ predictor matrix, $\boldsymbol{\varepsilon}$ is an unknown error vector; ordinarily rank of matrix $\mathbf{X}$ is less than p. The main PLS1 concept can be presented as the simultaneous bilinear decomposition of matrix $\mathbf{X}$ and vector $\boldsymbol{y}$

$$\mathbf{X} = \mathbf{T}\mathbf{P}^t + \mathbf{E} \qquad \boldsymbol{y} = \mathbf{T}\boldsymbol{q}^t + \boldsymbol{f}. \tag{2}$$

Here $\mathbf{T}$ is the $(n\times k)$ score matrix, $\mathbf{P}$ is the $(p\times\kappa)$ loading matrix and $\boldsymbol{q}$ is the $(k\times 1)$ loading vector; $\mathbf{E}$ and $\boldsymbol{f}$ are matrix and vector of residuals; k denotes the number of PLS components (PC), $k\le p$. In such a decomposition, the initial regression problem (1) is projected onto a low-dimensional (k) subspace, where the new problem (2) already has full rank.

In practice, the PLS principal components, i.e., score vectors $\boldsymbol{t}_1$, $\boldsymbol{t}_2$, etc., are calculated using a recurring algorithm (known as NIPALS [8]). At each step, one vector $\boldsymbol{t}_i$ is obtained, as well as the corresponding $\mathbf{Y}$-score vector $\boldsymbol{u}_i$, calculated as $\boldsymbol{u}_i=\boldsymbol{y}\boldsymbol{q}_i$. Correlation between these vectors

$$r_i = \mathrm{cor}(\boldsymbol{t}_i, \boldsymbol{u}_i) \tag{3}$$

serves as an important indicator showing that the PLS decomposition is completed and a proper number of PLS components, k, has already been achieved. This happens when the value of r_i essentially decreases in comparison with the previous step. To verify the completeness of the PLS decomposition, two additional characteristics are also used. These are the explained X and Y variances that are calculated as the mean squared residuals of models (2)

$$E_\mathrm{p} = 1-\frac{\sum E_{ij}^2}{\sum X_{ij}^2} \qquad E_\mathrm{r} = 1-\frac{\sum f_i^2}{\sum y_i^2}. \tag{4}$$

The accuracy of calibration could also be characterized by the residual response variance s^2

$$s^2 = \frac{1}{n-k}\sum_{i=1}^{n}(\boldsymbol{y}_i-\hat{\boldsymbol{y}}_i)^2$$

where $\hat{\boldsymbol{y}}$ is the vector of predicted response values.

In the conventional calibration approach, a regression model is constructed using calibration data set ($\mathbf{X}$, $\boldsymbol{y}$). Further on, it is validated using either an independent test set or by means of the cross-validation technique [8]. The square root of the residual variance calculated at the validation stage is called the root mean squared error of prediction (RMSEP), in contrast to the root mean squared error of calibration (RMSEC), which is calculated for the calibration data set.

It can be seen that initial regression problem given by Eq. (1) has no intercept term and therefore $\boldsymbol{y}=0$ at $\boldsymbol{x}=0$. To agree a raw data ($\mathbf{X}_\mathrm{raw}$, $\boldsymbol{y}_\mathrm{raw}$) with the model, a conventional centering transformation is applied

$$\boldsymbol{y} = \boldsymbol{y}_\mathrm{raw}-m_0\boldsymbol{1}, \qquad \mathbf{X} = \mathbf{X}_\mathrm{raw}-(\mathrm{m}_1\boldsymbol{1}, \mathrm{m}_2\boldsymbol{1}, \ldots, \mathrm{m}_\mathrm{p}\boldsymbol{1}).$$

Here m_0 is the mean value of response vector $\boldsymbol{y}$, m_i are the mean values calculated for the columns of matrix $\mathbf{X}_\mathrm{raw}$ and $\boldsymbol{1}$ is the vector of units. It is also important to perform an appropriate scaling of the raw data. The reason for this is that the strength of relationship is measured by covariance matrices. If data are not scaled, the results may depend on some variables that have a large variance but weak modeling power. Scaling of data can also be viewed as a way to obtain a stable PLS algorithm solution [9].

PLS2 method is a natural extension of the conventional PLS method (which will be termed below as PLS1) to the multi-response regression. In this case, a regression model is presented as

$$\mathbf{Y} = \mathbf{X}\mathbf{D} + \mathbf{E}$$

where $\mathbf{Y}$ is the $(n\times q)$-dimensional response matrix, $\mathbf{D}$ is the $(p\times q)$-dimensional matrix of unknown parameters, $\mathbf{X}$ is the $(n\times p)$ predictor matrix and $\mathbf{E}$ is an unknown error matrix. The related PLS2 decomposition may be written as

$$\mathbf{X} = \mathbf{T}\mathbf{P}^t + \mathbf{E} \qquad \mathbf{Y} = \mathbf{U}\mathbf{Q}^t + \mathbf{F}$$

Therefore, PLS2 gives one set of X and Y scores ($\mathbf{T}$ and $\mathbf{U}$) and one set of X and Y loadings ($\mathbf{P}$ and $\mathbf{Q}$), which are valid for all Y variables simultaneously.

2.2. Simple interval calculation

The SIC approach is based on a single assumption that all errors involved in calibration problem (1) are *limited* (measurement errors, modeling errors, etc.) [14]. The error finiteness means that there exists a maximum error deviation (MED) of error ε, which equals β, i.e.

$$\exists\beta{>}0 \quad \mathrm{Prob}\{|\varepsilon|{>}\beta\} = 0 \text{ and}$$
$$\text{for any } 0{<}b{<}\beta \ \mathrm{Prob}\{|\varepsilon|{>}b{>}0\} \tag{5}$$

where Prob{•} denotes probability that an event occurs. Relying on assumption in Eq. (5) and employing given calibration data set (**X**, ***y***) with n samples, it is possible to build the entire system of inequalities regarding the unknown regression parameters ***a***,

$$A = \{\boldsymbol{a}\in\mathbf{R}^{p} : \boldsymbol{y}^{-}<\mathbf{X}\boldsymbol{a}<\boldsymbol{y}^{+}\},\ \text{where } y_i^{-} = y_i-\beta,\ y_i^{+} = y_i+\beta \tag{6}$$

A is a closed convex set in the parameters' space; it is called the *region of possible* (parameter) *values* (RPV). This is a volumetric analogue of the conventional parameter point estimates vector $\hat{\boldsymbol{a}}$, which is calculated by some traditional regression method, e.g., PLS.

Using the obtained RPV, it is possible to solve a prediction problem for any given predictor vector ***x*** (e.g., a new spectrum or similar). If parameter ***a*** varies over A, it is clear that the predicted value $\boldsymbol{y}=\boldsymbol{x}^{t}\boldsymbol{a}$ belongs to the interval

$$V = [v^{-}, v^{+}],\ \text{where } v^{-} = \min_{\boldsymbol{a}\in A}(\boldsymbol{x}^{t}\boldsymbol{a}),\quad v^{+} = \max_{\boldsymbol{a}\in A}(\boldsymbol{x}^{t}\boldsymbol{a}). \tag{7}$$

The interval V is the result of SIC prediction. To find this interval, it is not necessary to present RPV explicitly, as the solutions of Eq. (7) may be obtained by linear programming methods [12], which are commonly used to find the optima of a linear function on a convex set. However, the limited solutions of a linear programming problem can be found if and only if the set A is bounded, i.e., **X** is a full-rank matrix [13]. In the opposite case, it is necessary to apply a regularization procedure, e.g., the PLS projection and further on use a score matrix **T** instead of **X** in the SIC method.

Usually, MED value is unknown and some estimate b is used instead of β. In the present work, we use two β estimates. Estimator $b_{\min}$ is defined as follows

$$b_{\min} = \min\{b,\quad A(b)\neq\varnothing\}. \tag{8}$$

This is a consistent but biased ($b_{\min}\le\beta$) estimate and $b_{\min}$ is the low limit of all possible β values. To estimate the upper limit of β, we apply a traditional statistical approach [18] to the regression residuals $\boldsymbol{e}=\hat{\boldsymbol{y}}-\boldsymbol{y}$. Therefore, it is possible to find an estimator b_{SIC} such that $\mathrm{Prob}\{b_{\mathrm{SIC}}>\beta\}>0.90$ and b_{SIC} is as close to β as possible. This enhanced estimator, b_{SIC}, can be calculated by formula [14]

$$b_{\mathrm{SIC}} = b_{\max}C(n, s^{2}). \tag{9}$$

Here $b_{\max}=\max(|e_1|,\ldots,|e_n|)$ and empirical function $C(n, s^2)$ depend on n that stands for the number of objects in the calibration set and on the residual variance s^2.

The calculation of different β estimators is rather comprehensive and is outside the scope of this paper. However, we can present a rule of thumb that helps to evaluate the estimators roughly. This could be termed the '1-2-3-4 sigma rule'. Let s_c be the root mean square error of calibration (RMSEC), then $b_{\min}\approx 2s_c$, $b_{\max}\approx 3s_c$ and $b_{\mathrm{SIC}}\approx 4s_c$. Certainly, this rule represents only a tendency, which also depends on the number of samples in the calibration set. Nevertheless, our experience in application to numerous examples shows that this rule appropriately characterizes the situation. Below, we will confirm this claim using Table 2 in Section 4.2.

To quantify a quality of SIC prediction, two measures are used [14]. The *SIC residual* is the difference between the center of the prediction interval (7) and the reference value y (scaled by β), so this is a characteristic of the bias:

$$r(\boldsymbol{x}, y) = \frac{1}{\beta}\left(y-\frac{v^{+}(\boldsymbol{x})+v^{-}(\boldsymbol{x})}{2}\right). \tag{10}$$

The *SIC leverage* is calculated as the width of the prediction interval, divided by MED β, so it has the character of β-normalized precision:

$$h(\boldsymbol{x}) = \frac{1}{\beta}\left(\frac{v^{+}(\boldsymbol{x})-v^{-}(\boldsymbol{x})}{2}\right). \tag{11}$$

In paper [10], a new object status classification (OSClas) concept was proposed. To understand this classification, it could be useful to interpret calibration/prediction as a continuous process of new sample treatment. Let us imagine a sequence of new objects $(\boldsymbol{x}_i, y_i)$, $i=n+1,\ldots$, that enter the calibration model, which was primary evaluated with n samples. The new samples can be considered in two different ways. Firstly, the model can be fixed at these n previous calibration objects so the incoming samples are considered as the unknown objects for prediction. This is a prediction point of view. Secondly, the model can be considered as an open system, which is replenished with the incoming objects and then recalibrated. This is a calibration point of view. It is evident that the addition of a new sample $(\boldsymbol{x}_{i+1}, y_{i+1})$ to the calibration set could modify RPV (6) in only one of the following ways: (i) RPV does not change, i.e., $A_{n+1}=A_n$; (ii) RPV shrinks, i.e., $A_{n+1}\subset A_n$; (iii) RPV disappears, i.e., $A_{n+1}=\varnothing$. The first case (i) corresponds to a sample, which is termed *insider*. In the second case (ii), such object is termed *outsider*. The third case (iii) corresponds to *outlier* in every sense of the term.

Looking at the objects from the prediction point of view, we can claim that the insiders are trusted absolutely as they agree completely with the model. Outsiders are less than perfect. There may be two reasons for this. Either the width of the prediction interval (that is the SIC leverage) is greater than the calibration error or there is a bias (characterized by the SIC residual). Outliers are the worst case and they cannot be used in prediction at all. From the calibration point of view, the object status interpretation is rather different. Insiders are then the worst objects, as they do not change the model but increase the experiment costs with redundant measurements. Outsiders are the most desirable objects as they improve the calibration accuracy and expand the domain of applicability. Outliers are dubious objects in the calibration. If the case the outrage mistake is excluded, such objects are very valuable in calibration as they outline the limits of prediction.

It was shown in [10] that OSClas could easily be performed without the explicit construction of the complex RPV in the parameter space. It is instead based on the following statements.

An object $(\boldsymbol{x}, y)$ is an insider iff $|r(\boldsymbol{x}, y)|\le 1-h(\boldsymbol{x})$; it is an outsider iff $|r(\boldsymbol{x}, y)|>1-h(\boldsymbol{x})$; an object is an outlier,

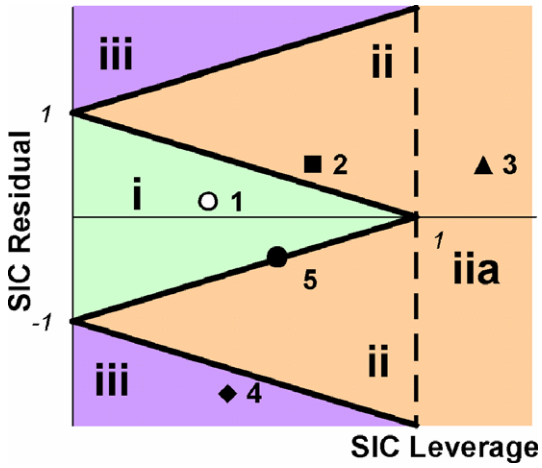


Fig. 1. Object status plot. (i) Insiders (○), (ii) outsiders (■), (iia) abs. outsiders (▲), (iii) outliers (◆).

iff $|r(\boldsymbol{x}, y)|>1+h(\boldsymbol{x})$; and an object $(\boldsymbol{x}, y)$ is an absolute outsider for any y iff $h(\boldsymbol{x})>1$. Using these statements, one can construct an *object status plot* (OSP) [10], the archetype of which is shown as Fig. 1. This OSP has the same appearance for *any* dimensionality of the initial data $(\mathbf{X}, \boldsymbol{y})$ and for *any* number of model parameters, which makes it a very powerful tool. OSP plane may be divided into three areas, each corresponding to one of the three object categories: insiders (area i in Fig. 1), outsiders (area ii) and outliers (area iii).

3. Real world data

We consider a multi-stage technological process that is represented by 25 process variables $\boldsymbol{x}$ and one output variable y, which is the final quality of the end-product. The food process under consideration is the well-known Russian strong drink manufacture. We consider the process more precisely in [15]. The production cycle (see Fig. 2) is divided into seven stages numbered by the Roman numerals. Each stage may be described by the input, current and further variables. Variables used in all previous stages are fixed input variables, current variables are the controlled ones, and the variables that characterize the following production stages are out of scope at the moment. Moving along the process, variables change their roles.

The first stage (I) is represented by six input variables (W1, W2, W3 and S1, S2, S3) that stand for the properties of the raw components S and W. At the second stage (II), the component W is refined and the process is characterized by the variables WR1 and WR2. Variables CW1, CW2 and CW3 (stage III) represent the properties of the outcome product CW. The next stage (IV) is mixing of the raw component S and the refined component CW. The result M is characterized by the variables M1, M2 and M3. Later on, the blend M is also refined (stage V) with the process characteristics MR1 and MR2, and the properties of outcome CM are presented by the variables CM1, CM2 and CM3 (stage VI). The last stage (VII) stands for the ultimate improvements, which are done with additives A1,..,A6. The output variable $(\mathbf{P}=y)$ is the final product quality.

At the end of each stage, a production engineer could analyze the intermediate results and correct the process attributes (variables) of the next stage. Both the analysis and the correction should be performed regarding the foreseeable output property y and with respect to the admissible range of correcting actions. Evidently, it is unreasonable to suggest theoretically improving actions, which, however, cannot be implemented in practice. The MSPC approach based on the real historical production records is the appropriate tool that is able to give the reasonable solution of this problem.

4. Process control

In this section, we will demonstrate how a process can be controlled without attempts to interfere into it. This may be called a *passive optimization* within the whole MSPO framework. For this purpose, we apply a method of expanding process modeling, which is based on the multi-block regression concept [16].

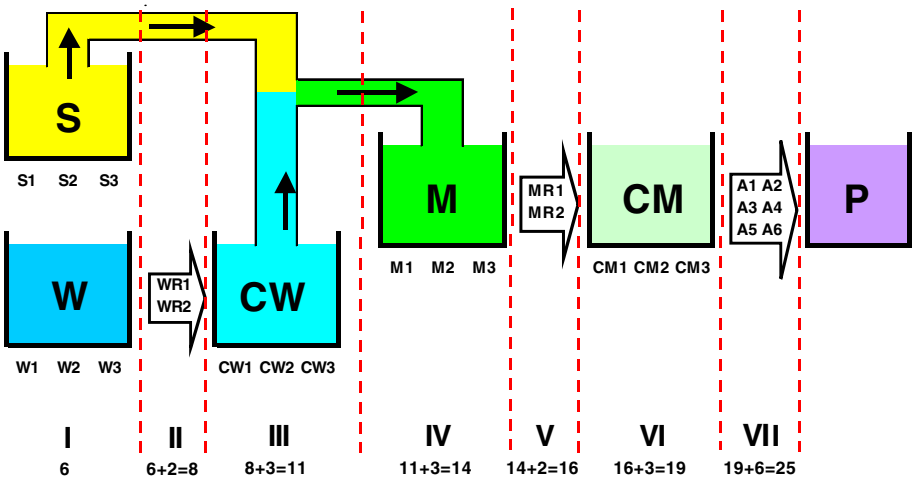


Fig. 2. Production cycle.

4.1. Theory

Let us assume that there is a collection of actual historical data measured for n samples that characterize a *proper* process performance. Each sample corresponds to the entire production cycle (batch); an example is shown in Fig. 2. The related row vector consists of the values of instrumental variables $x_1, x_2, \ldots, x_p$, and it also includes the quality variable y. The whole data set $(\mathbf{X}, \boldsymbol{y})$ is divided vertically (by variables) into L blocks in conformity with the stages:

$$\mathbf{X} = (\mathbf{X}_{\mathrm{I}}, \mathbf{X}_{\mathrm{II}}, \ldots, \mathbf{X}_L).$$

The last $L+1$ block consists of quality values $\mathbf{Y}=\boldsymbol{y}$.

In the example presented in Section 3, the data block $\mathbf{X}_{\mathrm{IV}}$ is the $(n\times 3)$ matrix, which includes three variables: M1, M2 and M3. The notation $\mathbf{X}_{(M)}$ will be used to represent a matrix that is composed with the $\mathbf{X}$ data from all previous M stages. For the process shown in Fig. 2, the block $\mathbf{X}_{(\mathrm{III})}=(\mathbf{X}_{\mathrm{I}}, \mathbf{X}_{\mathrm{II}}, \mathbf{X}_{\mathrm{III}})$ is the $(n\times 11)$ matrix. Applying this notation to the M-th stage of the process, it can be seen that matrix $\mathbf{X}_{(M-1)}$ presents the input data and matrix $\mathbf{X}_M$ stands for the controlled current data. At the first stage (I), there are no input data and, at the last stage L, $\mathbf{X}_{(L)}=\mathbf{X}$. It is assumed that the data are centered and scaled in such a way that each variable, including quality variable y, varies within the range $(-1, +1)$ and that all values outside this interval are not valid. It is also supposed that the highest product quality corresponds to $y=+1$, while the lowest one corresponds to $y=-1$.

Using the whole data set, it is possible to build an overall PLS1 regression model

$$XY : \mathbf{X}\Rightarrow\boldsymbol{y} \tag{12}$$

with k principal components. The notation XY is used here for an operator that maps block $\mathbf{X}$ to vector $\boldsymbol{y}$ with the help of PLS1 regression.

It is also assumed that *additional data scaling* has already been made with the X block. This scaling is performed by multiplication of some X columns (X variables) by factor -1, in order to make all estimates of the regression coefficients $\boldsymbol{a}$ in Eq. (1) positive. Such a scaling is useful in the optimization procedure described in Section 5. Scaling standardizes the process reply, because after it the increase of any process variable X will lead to the improvement of quality y. To make the notation simpler, the same symbols $\mathbf{X}$ and $\boldsymbol{y}$ will be used for the preprocessed data.

Using these data, we construct the following series of $L-1$ PLS1 regression models

$$XY_{\mathrm{I}} : \mathbf{X}_{(\mathrm{I})}\Rightarrow\boldsymbol{y}, \quad XY_{\mathrm{II}} : \mathbf{X}_{(\mathrm{II})}\Rightarrow\boldsymbol{y}, \quad \ldots,$$
$$XY_{L-1} : \mathbf{X}_{(L-1)}\Rightarrow\boldsymbol{y}. \tag{13}$$

Each model is denoted here by the operator XY_M, which maps the X block, $\mathbf{X}_{(M)}$, to the Y block, $\boldsymbol{y}$. In this series, the X block is expanded through the process time. Each XY model uses the same number of PLS principal components k that was chosen in the overall model given in Eq. (12).

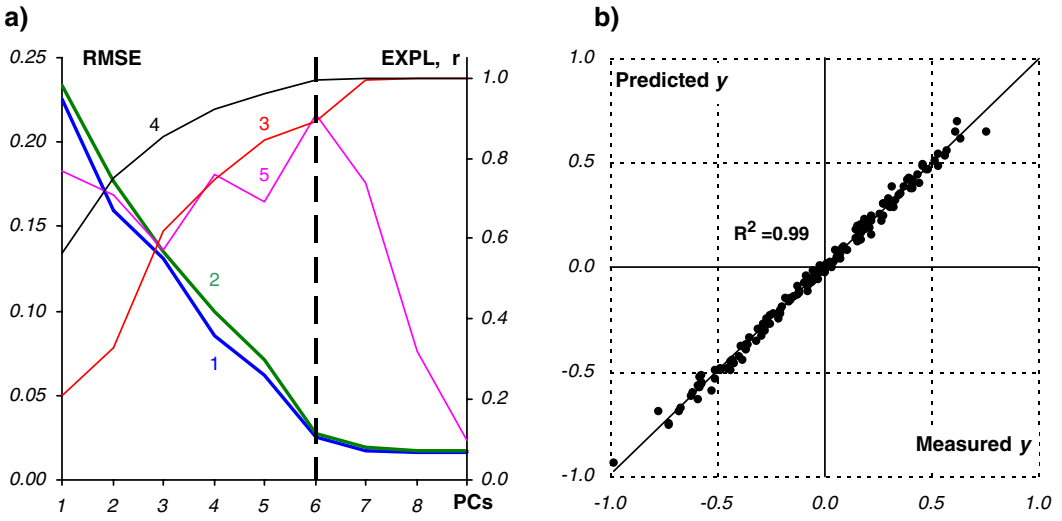
The main purpose of these models is the prediction of the output quality variable y at each (M-th) stage of production process. The predicted value could be further compared with a desired quality level. Too large difference signalizes that something is wrong and the process demands active improvements at the next $(M+1)$-th stage. To verify these corrections, a process engineer may try out various values of the variables that characterize stage $M+1$. The corresponding model XY_{M+1}: $\mathbf{X}_{(M+1)}\Rightarrow\boldsymbol{y}$ can validate the solution. Therefore, the system of models (13) serves as an "adviser" that helps the engineer to make a decision. However, this adviser cannot predict the future outcome y exactly. There is always some uncertainty. To present it, the corresponding SIC models are used. These models are built on the base of the relative PLS models with a given number of principal components, k. The maximum deviation value, β, is calculated as it is explained in Section 2.2.

4.2. Case study

Current section describes the application of the introduced theory to the real world data presented in Section 3. There is a set of historical data collected for 154 samples (batches). Seven vertical blocks $\mathbf{X}=(\mathbf{X}_{\mathrm{I}}, \mathbf{X}_{\mathrm{II}}, \ldots, \mathbf{X}_{\mathrm{VII}})$ represent the process stages with 25 instrumental variables and the very last $\boldsymbol{y}$ block relates to the final product quality, as it is shown in Fig. 2. The data are centered and scaled all in accordance with procedure described in the previous subsection.

For construction of the overall PLS regression model (12), we use six PLS components. This number is chosen with respect to the 10% out cross-validation analysis performed with the whole data set. The results are shown in Fig. 3 where the important characteristics of the model are plotted for the different number of PCs. These characteristics are: RMSEC, RMSEP; the rate of explained X and Y data (E_p, E_r, Eq. (4)); and the coefficient of correlation r (Eq. (3)). Fig. 3a shows that six principal components are enough for the PLS modeling. At this point, correlation coefficient r has the maximum $r=0.91$. At six PCs, there have been explained 89% of X-variance and 99% of Y-variance. It also can be seen that at six PCs both RMSEP and RMSEC have stabilized near value 0.026. We suppose that a higher dimension, i.e., seven PCs, may result in model over-fitting. Fig. 3b presents the plot of predicted y values versus measured y data. It demonstrates that all points are located close to the line with slope 45°. The correlation coefficient, R, between the measured and the predicted y values is $R=0.99$. The obtained RMSEP (which could also be called RMSECV due to employed cross-validation procedure) value 0.026 shows that, after the last stage (VII), the response can be estimated with an uncertainty that 95% of cases is smaller than $2\times 0.026=0.052$ and 99.99% of cases is smaller than $4\times 0.026=0.104$. Further, this uncertainty limit will be compared with the β value calculated in the SIC method.

The series of seven expanded PLS models given by Eq. (13) is built with the obtained number of principal components, i.e.,



Characteristics vs. number of PC. Left Y axis: 1- RMSEC, 2- RMSEP. Right Y axis: 3- E_p, 4- E_r, 5 – r=cor(***t***, ***u***)

Predicted vs. measured *y* values for the cross-validation analysis.

Fig. 3. Overall XY model with six PLS components.

for $k=6$. For each stage of the process, the X block consists of variables accumulated by this stage, e.g.

$$\mathbf{X}_{(\mathrm{I})} = \mathbf{X}_{\mathrm{I}} = (S1, S2, S3, W1, W2, W3)$$
$$\mathbf{X}_{(\mathrm{II})} = (\mathbf{X}_{\mathrm{I}}, \mathbf{X}_{\mathrm{II}}) = (S1, S2, S3, W1, W2, W3, WR1, WR2)$$

etc. To validate the PLS models, each data block is divided horizontally (by samples) in two parts: the calibration set (first 102 objects) and the test set (last 52 objects). The calibration set is used for the PLS modeling, while the test set is utilized for the prediction testing only. Some general characteristics of the models are presented in Table 1.

Table 2 presents the important characteristics of the related SIC models. Table 2 reads as follows. The first row, marked b, represents the upper limit for value β, calculated by Eq. (9). The next row, marked $b_{\min}$, demonstrates the β values calculated by Eq. (8). Both values decrease when the PLS model is expanded. This agrees with a general concept that modeling error should reduce when data set is enlarged. The next two rows represent ratios b/s_c and $b_{\min}/s_c$. It can be seen that the first ratio is about 4.5, while the second one is about 2.5. This confirms the rule of thumb claimed in Section 2.2. Row w contains the mean values of width of the SIC intervals (Eq. (7)) obtained for the test set. They show that uncertainty is reduced while the data set is enlarging. Row h represents the mean values of the SIC leverages (Eq. (11)) in the test set. It demonstrates a rather stable behavior along the process and does not vary too much.

Fig. 4 represents the OSP constructed for the test set that is predicted with overall PLS model given by Eq. (12). It might be useful to match this plot with the OSP archetype shown in Fig. 1. Such comparison shows that among 52 test samples there are 24 insiders, which lie within the triangle, e.g., samples no. 10 and no. 52. The 28 residuary samples are outsiders, e.g., samples no. 46, no. 24 and no. 50. Among them, there are 11 absolute outsiders, e.g., sample no. 24. No outliers can be found in the test set. The SIC leverage values are less than 1.5 with two extreme samples no. 40 and no. 46 that are situated at the right side of the plot.

For illustration purposes, we select five samples from the test set. They are nos. 10, 16, 24, 50 and 52 that are marked with the larger dots in Fig. 4. These samples represent the most typical cases with respect to the product quality as well as to the SIC status. The expanded PLS modeling (13) and the correspondent SIC modeling for these selected samples are presented in Fig. 5. Each plot demonstrates the results of prediction of future quality y that are obtained at every process stage. They are the PLS point estimates (black dots)

Table 1
General characteristics of the expanded PLS model (13)

	Stage	I	II	III	IV	V	VI	VII
Calibration	E_p	1.00	0.99	1.00	1.00	0.99	0.99	0.97
	E_r	0.78	0.81	0.79	0.84	0.95	0.98	0.99
	RMSEC	0.157	0.147	0.152	0.133	0.078	0.048	0.035
Validation	E_p	1.00	0.99	0.99	0.99	0.99	0.99	0.95
	E_r	0.85	0.87	0.87	0.88	0.96	0.98	0.99
	RMSEP	0.148	0.142	0.143	0.133	0.076	0.050	0.037

Table 2
General characteristics of the expanded SIC model (13)

Stage	I	II	III	IV	V	VI	VII
b	0.85	0.83	0.80	0.55	0.37	0.22	0.15
$b_{\min}$	0.51	0.47	0.51	0.30	0.21	0.12	0.08
b/s_c	5.41	5.65	5.26	4.14	4.72	4.62	4.24
$b_{\min}/s_c$	3.25	3.20	3.36	2.26	2.68	2.52	2.26
w	0.68	0.68	0.67	0.46	0.33	0.19	0.13
h	0.80	0.82	0.83	0.84	0.89	0.87	0.84

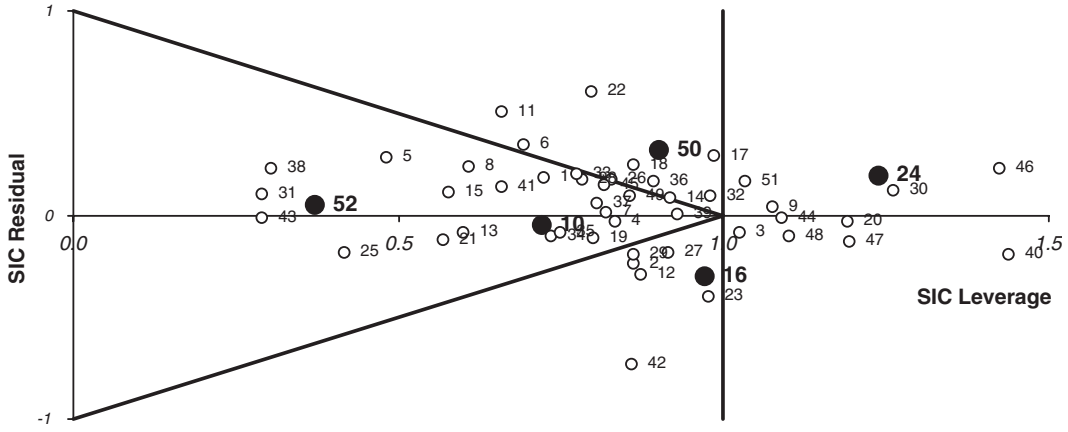


Fig. 4. Object status plot for test set. Overall PLS model after stage VII.

and the SIC intervals (gray bars). The actual y values (open rhombus) obtained at the end of production are also shown. The following results are clearly seen in Fig. 5. The SIC prediction intervals decrease through the process, but the reference value is always located inside the intervals. The width of the SIC interval (i.e., the degree of uncertainty) is smaller for the insiders (sample nos. 10 and 52) and it is the largest for the absolute outsider (sample no. 24).

Let us have a closer look at the sample no. 24 after stage IV. At stage V, the two instrumental variables MR1 and MR2 can be adjusted in order to improve the future quality value y. Four feasible solutions are shown in Fig. 6. Solution 0 is given here as a control reference and represents the usage of the actual historical MR1 and MR2 values. Solution 1 represents the application of average values MR1 and MR2 that are zeros, since X data are centered. Solution 2 corresponds to the PLS2 prediction $\mathbf{X}_{(\mathrm{IV})} \Rightarrow \mathbf{X}_{\mathrm{V}}$, which will be viewed in details in Section 5. Solutions 3 and 4 are arbitrary values that could be set by an experienced production engineer. The forecasted quality values for each solution, y, are shown in Fig. 6 with gray bars that present the SIC intervals and with closed dots that stand for the points of the PLS prediction. Horizontal line represents the control quality level that was actually obtained for sample no. 24. It can be seen that solutions 1 and 2 do not improve quality, while solutions 3 and 4 make it better. However, it may be supposed that solution 4 is too drastic and so it might be inadmissible. This problem will be considered in the next section.

5. Process optimization

In this section, we shall explain how to find the optimal correcting actions that could be performed at the end of each stage, i.e., how to perform *active optimization*. Actually, this means a proper choice of the controlled variables that become the input variables for the next coming stage. This choice should meet two crucial conditions. Firstly, it must improve the quality of the end-product, i.e., to maximize y value. Secondly, the choice must be performed within a range of acceptable bounds of the controlled variables. The underlying theory is based on the path modeling technique [17] and the SIC object status concept [10]. This is the second approach within the proposed MSPO concept.

5.1. Theory

Consider a task of modeling three data blocks. The situation is illustrated in Fig. 7. Variables in blocks **X** and **Z** are available historical process data and those in $\boldsymbol{y}$ are quality data, which are also known. The primary objective is to provide with prediction of quality for a new value of process variable ($\boldsymbol{x}$, $\boldsymbol{z}$). However, some of these variable values are not available. They are indicated by the row vector $\boldsymbol{z}$ in Fig. 7. Values in $\boldsymbol{x}$ associated with **X** block are available, and they may be used to predict both $\boldsymbol{z}$ and $\boldsymbol{y}$ values. It is necessary to find 'an optimal' $\boldsymbol{z}$ value, such that $\boldsymbol{z}$ maximizes the prediction for value of y. Thus, we are not looking for an apt prediction of $\boldsymbol{z}$ per se, but we want to find the values of $\boldsymbol{z}$ such that they are good for entering predictions of y.

In order to find an appropriate $\boldsymbol{z}$, two relevant PLS models can be used. The first one is the overall PLS1 model that predicts the response block $\boldsymbol{y}$ over the joint blocks **X** and **Z**, i.e.

$$XY: \ (\mathbf{X}, \mathbf{Z}) \Rightarrow \boldsymbol{y}. \tag{14}$$

Let $\boldsymbol{b}$ and $\boldsymbol{c}$ be the raw regression coefficients obtained at calibration of the historical data (**X**, **Z**, $\boldsymbol{y}$) by the XY regression, i.e.,

$$\hat{\boldsymbol{y}} = XY(\mathbf{X}, \mathbf{Z}) = \mathbf{X}\boldsymbol{b} + \mathbf{Z}\boldsymbol{c}.$$

It should be mentioned that the joint coefficients $\boldsymbol{b}$ and $\boldsymbol{c}$ give the coefficient vector $\boldsymbol{a}$, defined in Eq. (1). Then the optimization problem may be stated as

$$\text{maximize}(\boldsymbol{x}^t\boldsymbol{b} + \boldsymbol{z}^t\boldsymbol{c}) \quad \text{w.r.t. } z, \quad \text{subject to } \boldsymbol{z} \in L_z \tag{15}$$

Here L_z is the region of acceptable $\boldsymbol{z}$ values. This area will be defined bellow, whereas now it may be discussed in general terms.

The main problem in linear optimization is not to find a solution, but to restrict the area L_z where this solution is reached. Optimization of a linear model in unreasonable region always gives a senseless solution, where the optimized value goes out of the feasible value range. At first, the restrictions for area L_z

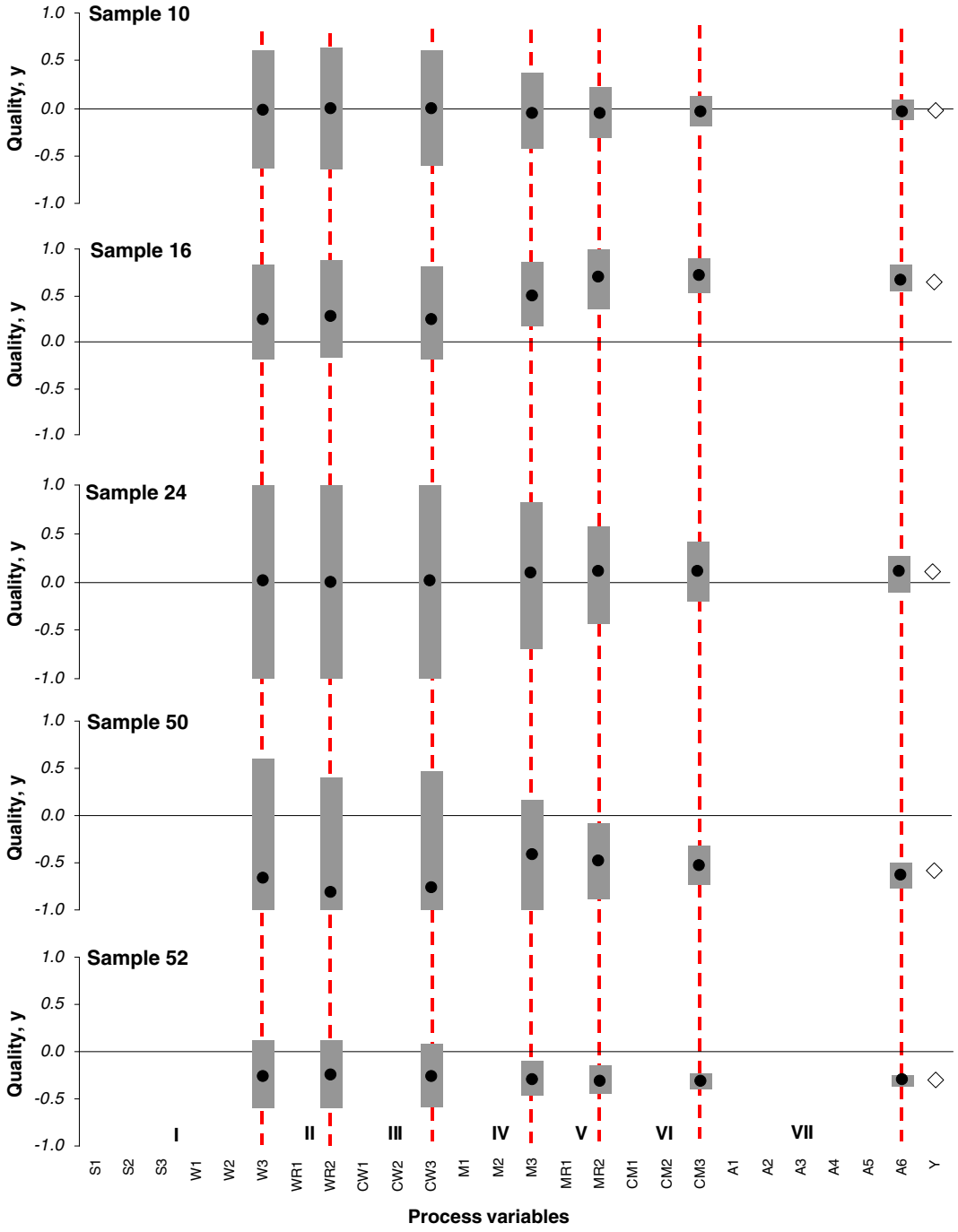


Fig. 5. Prediction for quality at each process stage. SIC intervals (gray bars) and PLS prediction (closed dots) for the selected samples. Open rhombus in the right part of the plots shows the quality value y that was actually obtained in the production.

come from physical or technological meaning of the process variables. Without loss of generality, these constraints may be presented as follows:

$$|z_i|<1 \text{ for } i = 1, 2, \ldots \text{ and } |\boldsymbol{x}^t\boldsymbol{b} + \boldsymbol{z}^t\boldsymbol{c}|<1$$

It is evident that, with respect to the *additional scaling* of X data described in Section 4, the overall XY model has a feature that an increase of any variable z_i in vector $\boldsymbol{z}$ leads to the growth of the quality value y. Therefore, the optimum is always reached at the upper boundary of the related variable region. Consequently, all that has to be done in the optimization is to define the reasonable bounds. The rough limits come from data scaling. However, bounds $(-1, +1)$ are too wide and optimization within this range gives the unrealistic result in prediction of quality y, which becomes greater than $+1$, its upper scaled limit. This happens due to the evident correlations between the process variables, which make such combination of process values as $(+1, +1, \ldots, +1)$ unavailable and impracticable. Therefore, the area L_z should contain only such $\boldsymbol{z}$ values that do not contradict the history of the process. This means that the

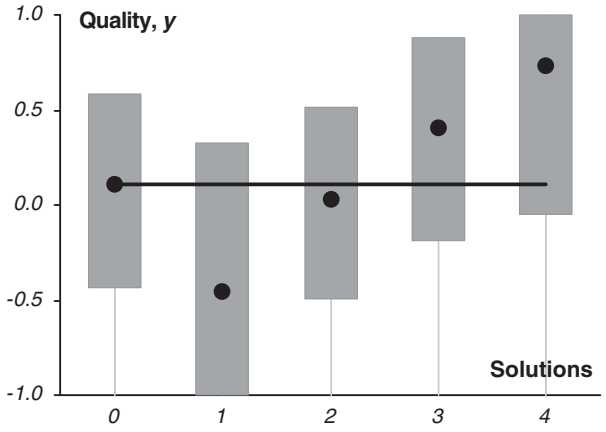


Fig. 6. Predicted quality for solutions at stage V.

allowable $\boldsymbol{z}$ values should lie within PLS model (14), which describes the process. Here we could mention essential resemblance in L_z definition to the known problem called 'outliers in prediction' [8,19,20]. Predicting a new unknown object, it is important to verify whether it could be used in the established calibration model, or whether the sample has any abnormalities that make prediction doubtful or improper. As long as the response value y is unavailable for a new object, the detection of outliers in prediction is primarily based on the X variables. Two manifest distance criteria are used for this purpose. The first one is a leverage measure, which characterizes the PLS spanned subspace distance to a new sample. The second criterion is a transversal distance that represents the unmodeled residuals in X data. Both measures are random values so some statistical approach should be applied to these criteria.

Let us give a formal definition of area L_z, using the SIC object status classification. Let $\boldsymbol{h}$ be a vector of the SIC leverages and let $\boldsymbol{d}$ be a vector

$$d_i = \sqrt{\boldsymbol{e}_i^t \boldsymbol{e}_i}, \quad \text{where } \boldsymbol{e}_i^t = \boldsymbol{x}_i^t - t_i \mathbf{P}^t \tag{16}$$

of the root mean squared X residuals that represent the object distances to the PLS hyperplane. Vectors $\boldsymbol{h}$ and $\boldsymbol{d}$ could be obtained at validation of the overall XY model (14) using an independent test set ($\mathbf{X}_{\text{test}}$, $\mathbf{Z}_{\text{test}}$) or by the cross-validation method. However, they cannot be calculated at the calibration stage, in which all SIC leverages are less than 1. Treating vectors $\boldsymbol{h}$ and $\boldsymbol{d}$ as representative samplings, it is possible to find their critical levels. For example, four limits might be set for each criterion:

$$\begin{aligned} l_0 &= m_h, \quad l_1 = m_h + s_h, \quad l_2 = m_h + 2s_h, \quad l_3 = m_h + 3s_h, \\ r_0 &= m_d, \quad r_1 = m_d + s_d, \quad r_2 = m_d + 2s_d, \quad r_3 = m_d + 3s_d, \end{aligned} \tag{17}$$

where m_h and m_d are the means, and s_h and s_d are the standard deviations calculated from vectors $\boldsymbol{h}$ and $\boldsymbol{d}$ correspondingly. These levels define the critical position of a new object regarding the overall PLS model.

Having those historically confirmed boundaries, it is possible to determine whether a new vector $\boldsymbol{z}$ is an admissible solution for the process optimization. For that, two values h_z and d_z are calculated for the joint row ($\boldsymbol{x}$, $\boldsymbol{z}$), and then they are compared with the chosen limits l_c and r_c. If $h_z < l_c$, and $d_z < r_c$, then $\boldsymbol{z} \in L_z$. Employing various limits (e.g., given by Eq. (17)), one may perform a variety of optimization strategies that may be termed as cautious, bold, etc.

Now, let us consider the second relevant PLS2 regression that predicts the response block $\mathbf{Z}$ over the predictors block $\mathbf{X}$, i.e.

$$XZ: \ \mathbf{X} \Rightarrow \mathbf{Z} \tag{18}$$

This model may be calibrated using the available historical data set ($\mathbf{X}$, $\mathbf{Z}$), i.e., $\hat{\mathbf{Z}} = XX(\mathbf{X}) = \mathbf{XD}$, where $\mathbf{D}$ is the corresponding parameter matrix. Applying the model to a new row vector $\boldsymbol{x}$, a prediction for $\boldsymbol{z}$ can be found. This vector, $\hat{\boldsymbol{z}} = XX(\boldsymbol{x})$, is not a solution for the optimization problem (15); however, $\hat{\boldsymbol{z}}$ is obviously a feasible solution, which belongs to the area L_z. During optimization, each component of $\hat{\boldsymbol{z}}$ could be changed until new row vector $\boldsymbol{z}^+$ is within area L_z. This action can be presented via operator G

$$G(\hat{\boldsymbol{z}}) = \boldsymbol{z}+ \tag{19}$$

that specifies a strategy of optimization. It is clear that, due to additional scaling of X variables, the row vector $G(\boldsymbol{z})$ has to be greater than or equal to $\boldsymbol{z}$. A unit operator

$$G(\hat{\boldsymbol{z}}) = \hat{\boldsymbol{z}} \tag{20}$$

gives a trivial solution. In this case, no enlargement is exercised. Other kinds of the enlarging operator G will be considered in the next subsection.

5.2. Case study

Now the theory of process optimization will be illustrated using the batch process presented in Section 3. Applying the theory one can act in two ways. In the first case, each component of $\boldsymbol{z}$ is arbitrary enlarged until the conditions $\boldsymbol{z} \in L_z$ are violated. This, however, is a cumbersome tactics as all Z variables are obviously correlated. The second way is more

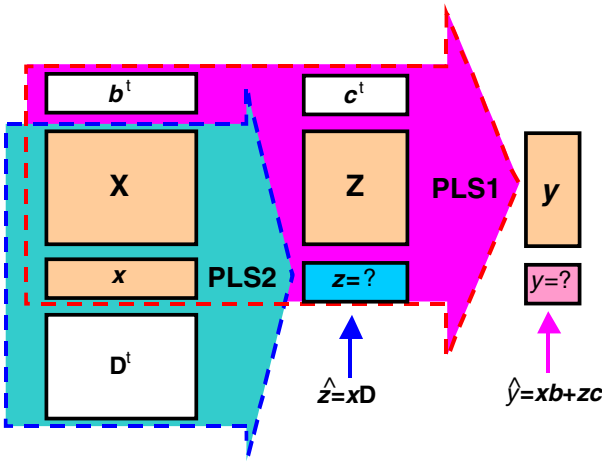


Fig. 7. The scheme of three data block modeling.

regular. It is proposed to choose some general method of $\boldsymbol{z}$ optimization, to apply it systematically to the whole production cycle, and then to verify that optimized values $\boldsymbol{z}^+$ agree with the criteria (17). The second way was chosen as more interesting for the theory exploration.

The series of the expanding PLS1 models given by Eq. (13)

$$XY_{\mathrm{I}}:\ \mathbf{X}_{(\mathrm{I})}\Rightarrow\boldsymbol{y},\ XY_{\mathrm{II}}:\ \mathbf{X}_{(\mathrm{II})}\Rightarrow\boldsymbol{y},\ \ldots,\ XY_{L-1}:\ \mathbf{X}_{(L-1)}\Rightarrow\boldsymbol{y}$$

has already been constructed in Section 4. These models correspond to the model given in Eq. (14). There is also a series of the PLS2 models corresponding to the model given in Eq. (18).

$$XX_{\mathrm{I}}:\ \mathbf{X}_{(\mathrm{I})}\Rightarrow\mathbf{X}_{\mathrm{II}},\ XX_{\mathrm{II}}:\ \mathbf{X}_{(\mathrm{II})}\Rightarrow\mathbf{X}_{\mathrm{III}},\ \ldots,$$
$$XX_{\mathrm{VI}}:\ \mathbf{X}_{(\mathrm{VI})}\Rightarrow\mathbf{X}_{\mathrm{VII}} \qquad (21)$$

These models predict a block of the future X variables ($\mathbf{Z}$) using the known block of the preceding X variables ($\mathbf{X}$) and they are constructed using the calibration set, as explained in Section 4. Each model is denoted by an operator XX_M that maps the $\mathbf{X}_{(M)}$ block to the $\mathbf{X}_{M+1}$ block. These models are of different complexity, i.e., the number of PCs, the MED β, etc. Some general characteristics are shown in Table 3. For example, a model XX_{III}: $\mathbf{X}_{(\mathrm{III})}\rightarrow\mathbf{X}_{\mathrm{IV}}$ predicts the forth block, $\mathbf{X}_{\mathrm{IV}}$, of variables (M1, M2 and M3) that is the (102×3)-matrix, over the block $\mathbf{X}_{(\mathrm{III})}$ of the first 11 variables represented with the (102×11)-matrix, and it uses six PCs. The X block was explained near 100%. Each response variable, M1, M2 and M3, has its own PLS2 characteristics. The responses, M1, M2 and M3, were explained as 99%, 100% and 98%, and the β values were estimated as $b=0.12$, 0.06 and 0.14, respectively. In comparison with RMSEC values, $s_c=0.027$, 0.015 and 0.033, they give the following boundary ratios $b/s_c=4.40$, 4.11 and 4.11.

Such approach implies the general concept of multivariate data analysis, namely, that difference between predictors and responses is very problem dependent [21]. This turns on one's choice and intention, which data block is reckoned as the predictor matrix or the response matrix in a given problem. In application to the process example, model (21) gives the most reasonable estimates of the future variables, which agree with the historical experience collected at the actual production. These models may be applied to any new process realization in order to obtain the prediction of the expected blocks of X values.

Table 3
General characteristics of PLS2 model (21)

Model	PCs	E_p	Variable	E_r	b	s_c	b/s_c
XX_{I}: $\mathbf{X}_{(\mathrm{I})}\Rightarrow\mathbf{X}_{\mathrm{II}}$	6	1.00	WR1	1.00	0.03	0.010	2.63
			WR2	1.00	0.05	0.012	3.79
XX_{II}: $\mathbf{X}_{(\mathrm{II})}\Rightarrow\mathbf{X}_{\mathrm{III}}$	6	1.00	CW1	1.00	0.03	0.008	3.70
			CW2	1.00	0.05	0.011	4.16
			CW3	1.00	0.02	0.005	4.30
XX_{III}: $\mathbf{X}_{(\mathrm{III})}\Rightarrow\mathbf{X}_{\mathrm{IV}}$	6	1.00	M1	0.99	0.12	0.027	4.40
			M2	1.00	0.06	0.015	4.11
			M3	0.98	0.14	0.033	4.11
XX_{IV}: $\mathbf{X}_{(\mathrm{IV})}\Rightarrow\mathbf{X}_{\mathrm{V}}$	6	1.00	MR1	0.99	0.08	0.029	2.90
			MR2	0.99	0.06	0.020	3.03
XX_{V}: $\mathbf{X}_{(\mathrm{V})}\Rightarrow\mathbf{X}_{\mathrm{VI}}$	7	1.00	CM1	1.00	0.06	0.014	4.14
			CM2	1.00	0.03	0.006	4.28
			CM3	1.00	0.01	0.002	4.13
XX_{VI}: $\mathbf{X}_{(\mathrm{VI})}\Rightarrow\mathbf{X}_{\mathrm{VII}}$	8	1.00	A1	1.00	0.05	0.012	3.84
			A2	1.00	0.05	0.011	4.48
			A3	1.00	0.05	0.009	4.96
			A4	1.00	0.07	0.016	4.24
			A5	1.00	0.08	0.017	4.42
			A6	1.00	0.06	0.012	4.90

Let us consider different solutions for optimization problem given by Eq. (15). The test set consisting of 52 samples represents a set of new samples; therefore, this set will be termed now as the *control* set. There will be necessity to compare different solutions with each other, as well as to compare each solution with the control set results. For this purpose, we use a sample distribution of quality variable y. To construct it, the set of 52 y values was distributed over 10 bins, which uniformly cover the range $[-1.0, +1.0]$ and then the frequencies were scaled with the number of samples. The histogram obtained for the control set is shown in Fig. 8a. This plot demonstrates a rather symmetrical distribution of quality value y among the control set with the mean value, $M=-0.10$, and the standard deviation value, $S=0.38$.

In application to the example, a general algorithm of the batch process optimization can be described as follows. Let us consider the process state after stage I, when a new-coming block of X variables $\mathbf{X}_{\mathrm{II}}$ should be adjusted. The PLS2 model XX_{I}: $\mathbf{X}_{\mathrm{I}}\Rightarrow\mathbf{X}_{\mathrm{II}}$ gives a block $\hat{\mathbf{X}}_{\mathrm{II}}$ of predicted X values that could be used as a base for the adjustment. To obtain an optimal solution some enlarging operator G, $G(\hat{\mathbf{X}}_{\mathrm{II}})=\mathbf{X}^+_{\mathrm{II}}$ (see Eq. (19)) is applied to this block. The result, $\mathbf{X}^+_{\mathrm{II}}$, is then combined with block $\mathbf{X}_{\mathrm{I}}$ and the joint block, $\mathbf{X}^+_{(\mathrm{II})}=(\mathbf{X}_{\mathrm{I}}, \mathbf{X}^+_{\mathrm{II}})$, is used as the input data for stage III. Block $\mathbf{X}^+_{(\mathrm{II})}$ is further processed in a similar way, i.e.

$$XX_{\mathrm{II}}(\mathbf{X}^+_{(\mathrm{II})})=\hat{\mathbf{X}}_{\mathrm{III}}$$

$$G(\hat{\mathbf{X}}_{\mathrm{III}})=\mathbf{X}^+_{\mathrm{III}}$$

$$(\mathbf{X}^+_{(\mathrm{II})},\mathbf{X}^+_{\mathrm{III}})=\mathbf{X}^+_{(\mathrm{III})}$$

Repeating the procedure for every stage, one can calculate a set of adjusted X variables, $\mathbf{X}^+=(\mathbf{X}_{\mathrm{I}}, \mathbf{X}^+_{\mathrm{II}}, \ldots, \mathbf{X}^+_{\mathrm{VII}})$, which may be used at each stage of the process for prediction of the quality values $\boldsymbol{y}$. This can be done using the expanding PLS1 model (13)

$$\boldsymbol{y}^+_M=XY\left(\mathbf{X}^+_{(M)}\right),\ M=II,III,\ldots,VIII$$

for the point prediction and the related SIC models for the interval prediction as it is explained in Section 4.2.

This general algorithm was implemented with three enlarging operator G (see Eq. (19)) that represents the various strategies. The *first strategy* employs unit operator G. In this case, the PLS2 estimates calculated with the help of model (21) are used as the optimized solutions. The result of this optimization is shown in Fig. 8b. It may be seen that the histogram of quality value y

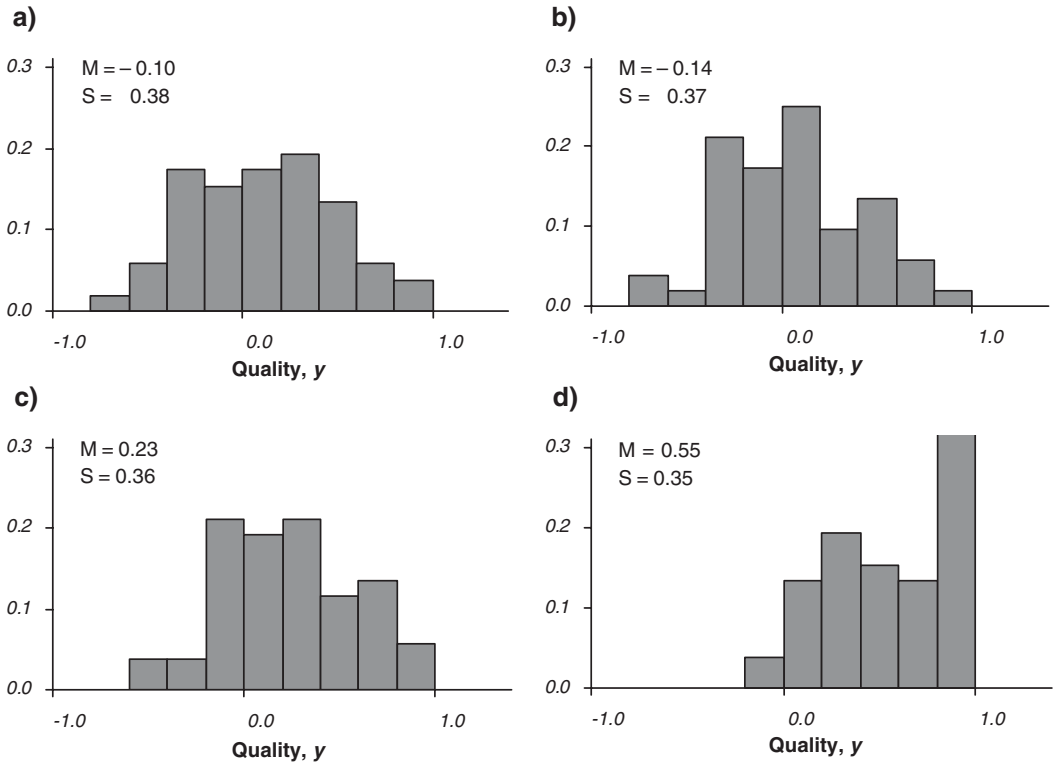


Fig. 8. Distribution of quality variable: (a) control set (no optimization), (b) optimization of the *insider* type, (c) optimization of the *outsiders* type and (d) optimization of the *outliers* type.

predicted over the whole X block, i.e., $\boldsymbol{y}^+ = XY(\mathbf{X}^+)$ is very similar to the distribution of the control values $\boldsymbol{y}$.

The *second optimization strategy* presumes a more resolute enlargement of the X variables. This is the addition of some constant values to each of the adjusted variables. Let values g_1, g_2, … be positive numbers. Then, the action of operator G on each variable z_i of the row vector $\boldsymbol{z}$ may be defined as

$$G(\hat{z}_i) = \hat{z}_i + g_i$$

We propose to chose g_1, g_2, … using the SIC concept of the MED β, namely, to select them as $g_i = b_i$, where b_i are the MED values estimated for each $\mathbf{Z}$ variable.

In our example, these values are given in Table 3, column b. For instance, a model XX_{I}: $\mathbf{X}_{(\mathrm{I})} \Rightarrow \mathbf{X}_{\mathrm{II}}$ predicts the second block $\mathbf{X}_{\mathrm{II}}$ of variables (WR1, WR2) that is the (102 × 2)-matrix, over the block $\mathbf{X}_{(\mathrm{I})}$ of the first six variables represented with the (102 × 6)-matrix, and it uses six PCs. The related β values were estimated as $b = 0.03$ and 0.05. In this strategy, the expected value $\hat{\mathrm{x}}_i$ is substituted with new value $x_i^+ = \hat{\mathrm{x}}_i + b_i$. Such an optimization is applied to each object from the control set of our example and gives an evident gain (see Fig. 8c) expressed in augmentation of the mean quality level M, namely, it is $0.23 - (-0.10) = 0.33$ regarding the control set.

Following the main concepts of the SIC object status classification, it is natural to examine the outermost type of the X variables enlargement. *This optimization* can be done by the following operator

$$G(\hat{z}_i) = \hat{z}_i + b_i(1 + h_i),$$

where z_i are the components of $\boldsymbol{z}$, h_i are the SIC leverages and b_i are the MED values estimated for each X variable related to the block $\mathbf{Z}$. The histogram of quality acquired by this optimization is shown in Fig. 8d. It can be seen that there is a great gain with respect to the control set, as the augmentation of the mean quality is $0.55 - (-0.10) = 0.65$.

Object status classification helps to select different strategies of optimization and the underlying object status plot (OSP) illustrates the idea of each strategy. It is unattainable to present all OSPs but they look very similar. Fig. 9 represents the OSP for predicted variable x_7, named as WR1 in the PLS2 model XX_{I}: $\mathbf{X}_{\mathrm{I}} \Rightarrow \mathbf{X}_{\mathrm{II}}$. For calculation of the SIC residuals, the known control values of variable WR1 are used as the reference response values. Fig. 9a shows that the SIC residuals have a rather widespread and the SIC leverages are less than 1.5. In total, there are 23 insiders and 29 outsiders among 52 samples. This is a rather typical appearance of the OSP for an ordinary test set (compare with Fig. 4).

Fig. 9b demonstrates how the layout of the 52 control samples is modified when we change the optimization strategies. The OSP allows us to understand the status of each adjusted sample variable, i.e., to understand whether it has a strange or an ordinarily value regarding the historical data. Here the SIC residuals are calculated with respect to the optimized variables x_7^+ that are regarded as the reference values (see Eq. (10)). It should be mentioned that this is more a treatment of the problem of outliers, which, however, should be considered as the special new objects with known reference values equal to the proposed optimal solutions. Applying the OSClas method,

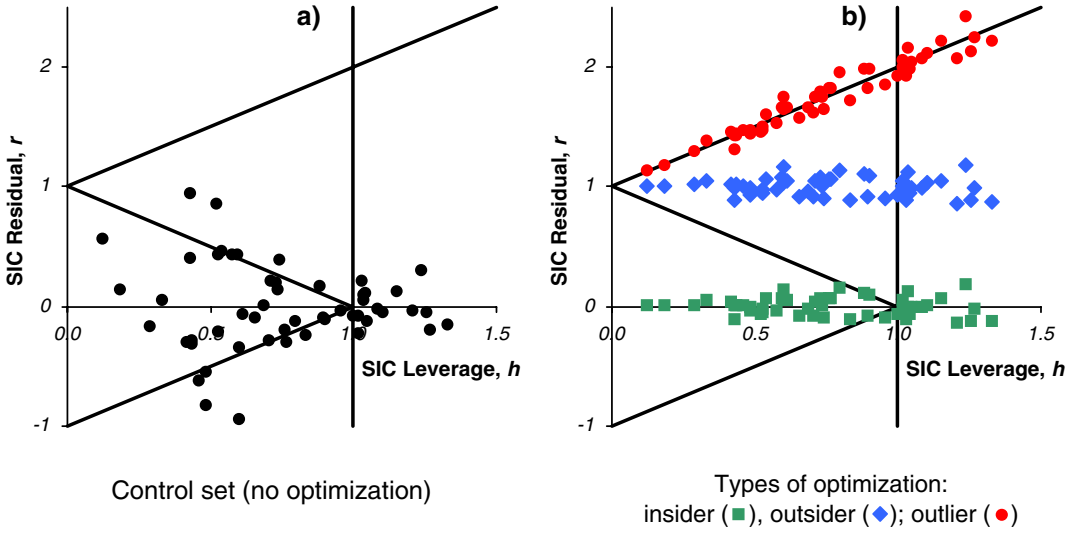


Fig. 9. OSP for model XX_I, variable WR1.

one can understand how far these solutions are from the average historically predicted values.

For the *first optimization strategy* (squares on Fig. 9b), we can indicate that the most of the adjusted variable values (35 of 52) are insiders (see Section 2.2). Therefore, the trivial optimization given by the unit operator G (Eq. (20)) may be termed as the *insider type* of optimization. This strategy, however, gives no gain in quality; as such optimization merely reproduces the historical experience with all its losses and gains.

For the *second optimization strategy* (diamonds in Fig. 9b), one can see that all samples are outsiders (compare with Fig. 1). Due to Section 2.2, such enlargements do not contradict the model. The same disposition is repeated in the OSPs for other X variables. Therefore, this approach may be termed as the *outsider type* optimization.

The idea of the *last optimization strategy* becomes clear from comparison the layout of adjusted samples (close dots in Fig. 9b) with the OSP archetype in Fig. 1. One can see that all objects are located now on the border of the outliers region and therefore this kind of optimization may be called the *outlier type*.

Applying three optimization strategies for our example, we yielded three data sets $\mathbf{X}^+$ that more or less enhance the predicted quality with respect to the control ones. These sets (strategies) were named insider, outsider and outlier types. It is worthy of mentioning that such status definitions relate to the XX models defined in Eq. (21) and they do not concern the overall XY model given in Eq. (12). This would be a great disappointment if the sets of adjusted X variables $\mathbf{X}^+$ disagree with the overall XY model calibrated over the block $\mathbf{X}=\mathbf{X}_{(VII)}$ of the historical, control X values. Such a case would mean that the crucial claim for $z \in L_z$ (see Eq. (15)) has been broken and the adjusted $\mathbf{X}^+$ values are unacceptable. Therefore, it is necessary to check through the sets $\mathbf{X}^+$ whether they agree with criteria given by Eq. (17). In that way, we verify that the optimized variable sets are situated not very far along the overall PLS model subspace and that their transversal distances to this hyperplane are not too large as well. Table 4 presents the distribution of the SIC leverages h calculated for the control set $\mathbf{X}$ and, for three sets of adjusted X variables, $\mathbf{X}^+$. Also, the mean values (m_h) and the standard deviations (s_h) of $\boldsymbol{h}$ for each optimization strategy are presented.

Table 5 shows the similar results for the root mean squared X residuals $\boldsymbol{d}$ (Eq. (16)). Studying Tables 4 and 5, one can conclude that all adjusted $\mathbf{X}^+$ sets agree with crucial claim for $\mathbf{X}^+ \in L_z$, in general. This can also be confirmed by comparisons of the mean values for each type of optimization. However, from Table 5, it could be seen that optimization of the outlier type is more drastic than the outsider one and the application of the outlier strategy should be done with caution.

Application of the OSClas theory gives us the instrument for understanding what is good or bad strategy and how to choose

Table 4
Distribution of the SIC leverages, $\boldsymbol{h}$ for the different optimization strategies. l_i values are calculated by Eq. (17)

Optimization strategy	$\boldsymbol{h} \le l_0$	$l_0 < \boldsymbol{h} \le l_1$	$l_1 < \boldsymbol{h} \le l_2$	$l_2 < \boldsymbol{h} \le l_3$	$\boldsymbol{h} > l_3$	m_h	s_h
Control	26	19	5	2	0	0.835	0.261
Insiders	36	12	3	1	0	0.723	0.277
Outsiders	28	17	5	1	1	0.809	0.305
Outliers	26	11	10	5	0	0.854	0.385

Two last columns present the mean and the standard deviation values of $\boldsymbol{h}$.

Table 5
Distribution of the root mean squared X residuals, $\boldsymbol{d}$ for the different optimization strategies. r_i values are calculated by Eq. (17)

Optimization strategy	$\boldsymbol{d} \le r_0$	$r_0 < \boldsymbol{d} \le r_1$	$r_1 < \boldsymbol{d} \le r_2$	$r_2 < \boldsymbol{d} \le r_3$	$\boldsymbol{d} > r_3$	m_d	s_d
Control	26	17	9	0	0	0.052	0.031
Insiders	35	10	3	4	0	0.047	0.030
Outsiders	19	25	5	3	0	0.068	0.024
Outliers	0	9	29	10	4	0.105	0.027

Two last columns present the mean and the standard deviation values of $\boldsymbol{d}$.

the enlargement operator G. For this purpose, we use the corresponding object status plots (OSPs) that are constructed for all PLS2 models XX_M given by Eq. (21). The position of the adjusted X variables x_i^+ in these plots explains the strategy as follows:

1. No improvement in quality will be obtained if the adjusted variables x_i^+ are located in the insiders' area (area i in Fig. 1).
2. If we want to yield a considerable improvement of quality variable $\boldsymbol{y}$, then optimized variable values x_i^+ should be located in the outsiders' area between the insiders and outliers (area ii in Fig. 1).
3. Values x_i^+ that are located in the outliers' area (area iii in Fig. 1) may contradict the historical experience and therefore such values could be applied with a great caution.
4. It is obligatory to verify that optimized values $x_i^+ \in L_z$, i.e., that they do not contradict the process history.

It is also interesting to compare the results of the expanded PLS/SIC modeling for the subset of the selected samples that is

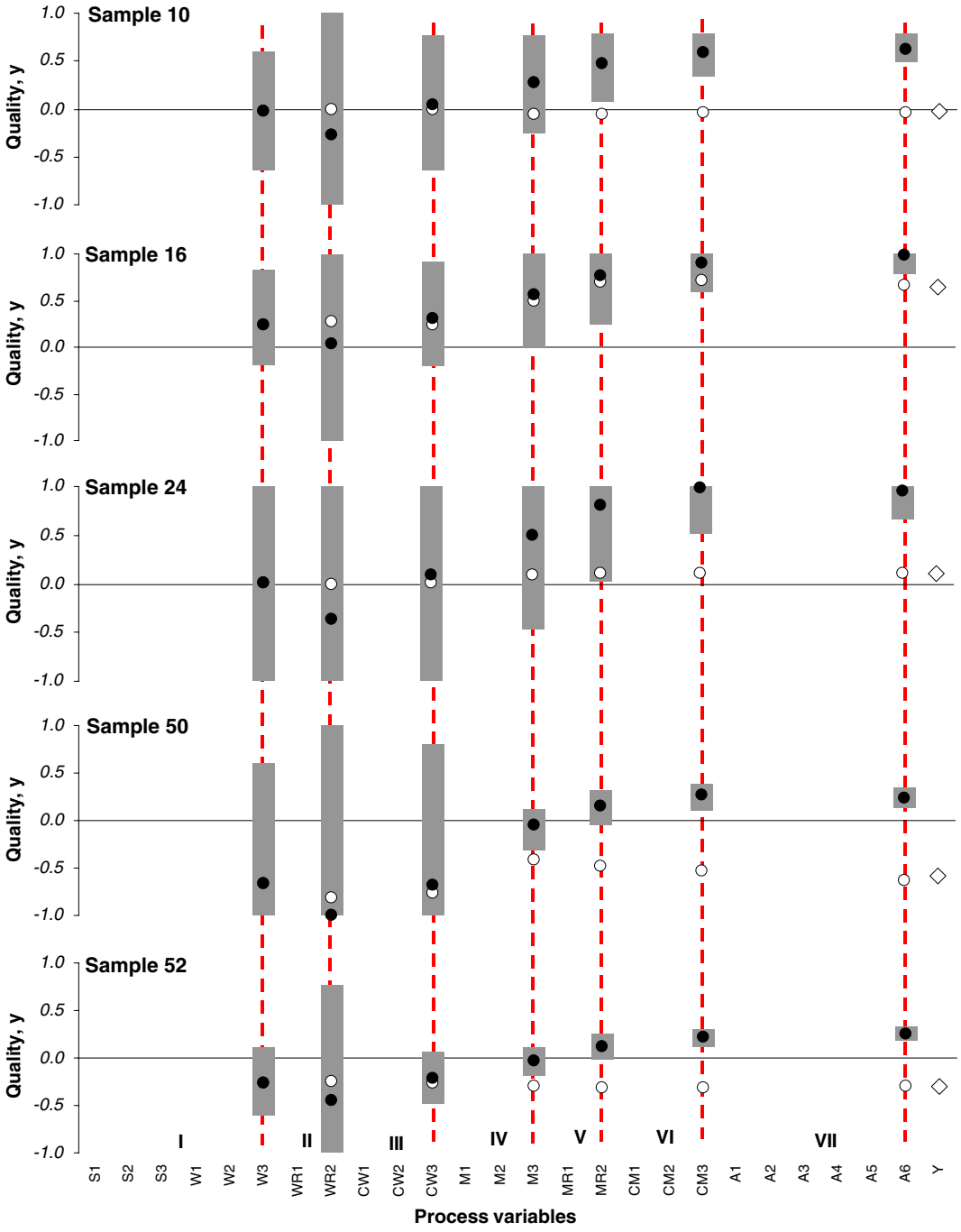


Fig. 10. Optimization by the outlier type. SIC intervals (gray bars) and PLS prediction (closed dots) for the selected samples. Open rhombus in the right part of the plots shows the historical quality value y that was actually obtained in the production. Open dots represent the PLS predictions for the control set.

presented in Fig. 5, with the corresponding plots obtained for the optimized $\mathbf{X}^{+}$ values. They are shown in Fig. 10 for the outlier type of optimization.

The layout of these plots is similar to those plots in Fig. 5. The single difference are the open dots that represent the PLS prediction points for the control set. It is clear that this optimization actually, more or less, improves the output quality. It also can be seen that the intermediate y evaluation at stage II predicts the reduction of quality for all samples. This results from the partial block model XY_{II}: $\mathbf{X}_{(\mathrm{II})} \Rightarrow \boldsymbol{y}$ that has the negative regression coefficients at variables WR1 and WR2, while in the overall model XY: $\mathbf{X} \Rightarrow \boldsymbol{y}$ all the coefficients are positive. There can be seen some differences between the optimization performed with respect to the overall regression model (12) and the optimization regarded with the partial block model (13). These may be considered as two objects of optimization that could be termed as the global and the local style. In the first case, all the variables are undoubtedly increased at every other stage, while in the second case the variables that are associated with the negative regression coefficients of the current block model (13) are conversely decreased. The global style follows from the overall PLS model (12) that has the positive regression coefficients due to additional X scaling, as presented in the paper. An alternative style is to optimize the quality variable y with respect to each partial local model (13), where some regression coefficients may be negative. The local style of optimization has also been applied to the example process and there were rather small differences in the ultimate results. At the moment, we cannot explain this effect, but it obviously demonstrates a stability of the proposed optimization procedure.

6. Conclusions

This paper presents methods of process control and optimization and duly illustrates them with a real world example. The optimization methods are based on the PLS block modeling as well as on the simple interval calculation methods of interval prediction and object status classification. It is proposed to employ the series of expanding PLS/SIC model (13) in order to support the on-line process improvements. This method helps to predict the effect of planned actions on the product quality and thus enables passive quality control. We have also considered an optimization approach that proposes the correcting actions for the quality improvement in the course of production. The latter is an active quality optimization, which takes into account the actual history of the process and finds available adjustments that can be made during the production cycle. Active optimization is based on the object status classification that is an ensuing consequence of the SIC method.

This approach is allied to the conventional method of multivariate statistical process control (MSPC) as it also employs the historical process data as a basis for modeling. On the other hand, the presented concept aims more at the process optimization than at the process control. Therefore, it was proposed to call such an approach as multivariate statistical process optimization (MSPO).

In conclusion, we would like to discuss some specific problems connected to the proposed optimization procedure. It is worthy of mentioning that it is possible to conduct optimization with respect to other aims than maximization. For example, it is possible to look for the correcting actions that tend quality measure y to its mean, zero value, instead of maximum. The exciting item is the selection of actual data records that represent the historical knowledge about the process. This data set serves as a basis for regressions, imposes restrictions on correcting action and, as a matter of fact, determines the ultimate result of optimization. In the explored process, this selection was done with respect to the following considerations. All failure realizations, as well as the records, characterized by the experienced production engineer as atypical or incidental, were declined. Generally speaking, we made there nothing except the conventional data pretreatment including outlier detection and explanatory analysis. However, this issue is not properly investigated yet and requires further revisions.

References

[1] P. Nomikos, J.F. MacGregor, Multivariate SPC charts for monitoring batch processes, Technometrics 37 (1) (1995) 41–59.
[2] T. Kourti, J.F. MacGregor, Recent developments in multivariate SPC methods for monitoring and diagnosing process and product performance, J. Qual. Technol. 28 (4) (1996) 409–428.
[3] X. Wang, U. Kruger, B. Lennox, Recursive partial least squares algorithms for monitoring complex industrial processes, Control Eng. Pract. 11 (2003) 613–632.
[4] M. Kano, S. Hasebe, I. Hashimoto, H. Ohno, Evolution of multivariate statistical process control: application of independent component analysis and external analysis, Comput. Chem. Eng. 28 (2004) 1157–1166.
[5] T. Kourti, J. MacGregor, Process analysis, monitoring and diagnosis, using multivariate projection methods. Tutorial, Chemom. Intell. Lab. Syst. 28 (1995) 3–21.
[6] Guidance for Industry PAT—A Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance, U.S. Department of Health and Human Services, Food and Drug Administration, Center for Drug Evaluation and Research (CDER), Center for Veterinary Medicine (CVM), Office of Regulatory Affairs (ORA), September 2004, Pharmaceutical CGMPs.
[7] W.A. Shewhart, Economic Control of Quality of Manufactured Product, Van Nostrand, Princeton, NJ, 1931.
[8] H. Martens, T. Naes, Multivariate Calibration, Wiley, New York, 1998.
[9] A. Höskuldsson, Prediction Methods in Science and Technology, Thor Publishing, Copenhagen, Denmark, 1996.
[10] Rodionova O. Ye, K.H. Esbensen, A.L. Pomerantsev, Application of SIC (Simple Interval Calculation) for object status classification and outlier detection—comparison with PLS/PCR, J. Chemometr. 18 (2004) 402–413.
[11] A.L. Pomerantsev, Rodionova O. Ye, Hard and soft methods for prediction of antioxidants' activity based on the DSC measurements, Chemom. Intell. Lab. Syst. 79 (2005) 73–83.
[12] G. Dantzig, Linear Programming and Extensions, Princeton University Press, Princeton, NJ, 1963.
[13] S. Gass, Linear Programming, 4-th ed.McGow-Hill, New York, 1975.
[14] Rodionova O. Ye, A.L. Pomerantsev, Principles of simple interval calculations, in: A.L. Pomerantsev (Ed.), Progress in Chemometrics Research, NovaScience Publishers, NY, 2005, pp. 43–64.
[15] A.L. Pomerantsev, Rodionova O. Ye, Multivariate statistical process control and optimisation, in: Pomerantsev (Ed.), Progress in Chemometrics Research, NovaScience Publishers, NY, 2005, pp. 209–227.
[16] L.E. Wagen, B. Kowalski, A multiblock partial least squares algorithm for investigation complex chemical systems, J. Chemometr. 3 (1998) 3–20.

[17] A. Höskuldsson, Causal and path modelling, Chemom. Intell. Lab. Syst. 58 (2001) 287–311.
[18] E. Gumbel, Statistics of Extremes, Columbia University Press, NY, 1962.
[19] J.A. Fernandez Pierna, L. Jin, M. Daszykowski, F. Wahl, D.L. Massart, A methodology to detect outliers/inliers in prediction with PLS, Chemom. Intell. Lab. Syst. 68 (2003) 17–28.
[20] M.S. Larrechi, M.P. Callao, Strategy for introducing NIR spectroscopy and multivariate calibration techniques in industry, Trends Anal. Chem. 22 (2003) 634–640.
[21] K.H. Esbensen, Multivariate Data Analysis—In Practice, 4-th ed.CAMO, 2000.