Available online at www.sciencedirect.com
SCIENCE DIRECT
Computers and Chemical Engineering 30 (2006) 521–534
Computers & Chemical Engineering
www.elsevier.com/locate/compchemeng

Optimal sensor location for distributed-sensor systems using multivariate regression

Riccardo Muradore*, Fabrizio Bezzo, Massimiliano Barolo

DIPIC, Dipartimento di Principi e Impianti di Ingegneria Chimica, Università di Padova, Via Marzolo 9, 35131 Padova PD, Italy

Received 30 September 2004; received in revised form 12 October 2005; accepted 17 October 2005
Available online 29 November 2005

Abstract

The performance of monitoring and control techniques for distributed-sensor systems is affected by the choice of the measurement sensor location. In this paper a methodology is suggested to solve the optimal sensor location problem. The suggested algorithm does not require any explicit knowledge of the plant model, and is based upon a sequential procedure selecting at each iteration the most informative measurement input, and updating the input and output spaces by subtracting the information explained by the computed regressor. The effectiveness of the proposed algorithm is assessed by means of two simulated case studies concerning the location of sensors in tubular reactors where product composition need to be estimated from temperature measurements.
© 2005 Elsevier Ltd. All rights reserved.

Keywords: Optimal sensor location; Distributed sensor systems; Multivariate regression

1. Introduction

A common problem in the monitoring of processing systems concerns the estimation of product quality variables (primary variables) using process operating variables (secondary variables). In fact, quality variables (e.g. concentration, melt index, and viscosity) are often not available on-line at a sufficiently fast sampling frequency (sometimes they are not available on-line at all), whereas process variables (e.g. temperature, flow rate, pressure, and level) can be obtained on-line very frequently, accurately, and at low cost.

Data-driven estimation techniques offer the desirable feature that, at least in principle, no a priori knowledge about the process (fundamental mechanisms or model parameters) is needed to build a quality variable estimator. In many cases several process measurement sensors are distributed within an existing system, or can be potentially installed for a system being designed anew. In such circumstances, an issue arises regarding which process measurements should be used, among all the available ones, as inputs to the primary variable estimator. The problem exists both for single-unit systems, where several measurements of the same kind (e.g. temperatures) can be possibly allocated along the unit, as well as for plantwide systems, where the above issue is further complicated by the fact that measurements of different nature (e.g. temperature, pressure, and flow rate) must be integrated together. The topic of this paper is the optimal selection of process measurement sensors to be fed as inputs to a data-driven product quality estimator (soft sensor). The sensor selection problem is addressed for single-unit systems only, although the proposed procedure can be easily extended to plantwide systems.

The inherent distributed nature of two important classes of processing units, i.e. tubular reactors and distillation columns, may make the sensor selection problem a challenging task also for a single-unit system. For example, when the temperature measurements are used to estimate on-line the composition of the product streams, several different choices for locating the temperature sensors along the units are possible, and the performance of the quality estimator is strongly related to the choice of the temperature input set.

If an existing unit is being considered, the number and locations of temperature sensors is usually assigned (i.e. the sensors are physically already installed on the unit), and the simplest approach would be to build the estimator by making use of all of the available sensors. Because the available measurements are almost invariably correlated, a suitable multivariate statistical method (e.g. partial least-squares regression (PLS); Geladi & Kowalski, 1986; Rao & Toutenburg, 1999) can then be exploited to solve the collinearity issues. Although this approach (Mejdell & Skogestad, 1991) is sound from a theoretical point

* Corresponding author. Tel.: +39 049 827 5468; fax: +39 049 827 5461.
E-mail address: muradore@dei.unipd.it (R. Muradore).

0098-1354/$ – see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compchemeng.2005.10.009

of view, it may nevertheless be impractical, because using too many sensors can make the estimator more exposed to sensor failure. A different way to tackle the problem would be to find out a subset of the available sensors that (in some sense) prove to be optimal for the estimation of the primary variables. This is the approach taken, for example, by Kano, Miyazaki, Hasebe, and Hashimoto (2000) in the estimation of the distillate and bottoms compositions from tray temperature measurements in a distillation column. However, no indication is given by the authors about how the tray temperatures subset can be identified in a systematic way.

If a new unit is being designed together with its measurement system, the selection problem can be stated in a slightly more general way. In principle, one can assume to be able to locate the measurement sensors wherever along the unit (e.g. on each tray of a distillation column; evenly spaced along the whole length of a tubular reactor), and then select which sensor (and how many sensors) among all the possible ones is the most suitable for product quality estimation purposes.[1] In this perspective, Kookos and Perkins (1999) proposed an algorithmic approach to the solution of the optimal sensor location problem in distillation columns based on a mixed-integer linear programming problem formulation. The algorithm is very effective, but explicitly requires a first-principles model of the plant. Zamprogna, Barolo, and Seborg (2005) used principal component analysis (PCA) to screen out the optimal temperatures for a PLS-based composition estimator in a batch distillation column. The proposed procedure is completely data-driven, but requires the calculation of a sensitivity matrix, which has not a strong statistical interpretation. van den Berg, Hoefsloot, Boelens, and Smilde (2000) found the best temperature sensor location for a tubular reactor by specifying different scalar measures of observability, and selecting the best sensor as the one that maximizes the system observability. However, a single sensor is selected by that procedure, and extension to multiple-sensor selection is not straightforward; moreover, the selection procedure is inherently model-based. Vande Wouwer, Point, Porteman, and Remy (2000) found the optimal sensor locations in a tubular reactor by maximizing the independence amongst the sensor responses based on the Gram determinant value. Since the selection procedure requires a hard maximum search on a discretized surface of dimension $(n + 1)$, where n is the number of sensors, the search becomes complicated when $n > 2$.

In the present paper, a data-driven solution is proposed to solve the sensor location problem; the optimal locations are chosen by using only a sequence of inputs and outputs. For ease of demonstration, we will refer to two tubular reactor case studies, where the compositions (or flows) of the desired components in the product streams must be estimated from temperature measurements. It will be supposed that the measurement system is to be designed anew, i.e. that temperature sensors can be potentially allocated in any section of the reactors. The selection procedure is asked to find out the optimal number and optimal location of sensors; subsequently, the temperature measurements that have been identified as optimal are used to design a linear static estimator via a multivariate regression technique. The proposed method has two main desirable features. First, it does not require to use a first-principles model within the selection algorithm. Second, the method relies on a linear static estimator that does not require stability assessments and expensive on-line computations. Furthermore, it has a sequential formulation that allows to select the sensor location in a step-by-step way, and it is easy to extend to systems where different types of sensors have to be integrated. In fact, the sequential formulation makes it easy to add sensors (in their optimal locations) until some specified criterion is satisfied. Quite obviously, the price to pay when working with linear static estimators based on black-box models is a possible loss of accuracy when the nonlinearity or the dynamics of the plant are high.

[1] This relies on the fact that a first-principles model (namely, the same model that was used to design the unit itself) is available, but cannot be used on-line for estimation purposes (e.g. because it is computationally burdensome).

2. Problem statement

The problem under study is the optimal temperature sensor location for tubular reactors when product composition information must be inferred from temperature measurements. The class of tubular reactors under study is described by a set of nonlinear partial differential equations with algebraic constraints of the form

$$\frac{\partial \xi}{\partial t} = -v\frac{\partial \xi}{\partial z} + D\frac{\partial^2 \xi}{\partial z^2} - g(\xi), \qquad \zeta = h(\xi, z), \tag{1}$$

where $\xi = \xi(t, z)$ is the state vector containing component concentrations (or flows) and temperatures as function of time t and space z, ζ the measurement signal related to ξ through the nonlinear function $h(\cdot, \cdot)$. The function $g(\xi)$ and the parameters v, D, are related to the particular reactor.

In the cases under study, the output signal ζ is a set of temperature measurements along the reactor length. An evenly spaced distribution of the available locations is assumed, and the goal is to select the best ones. In order to avoid the introduction of the infinite dimensional system theory to study systems like (1) (see Alonso, Kevrekidis, Banga, & Frouzakis, 2004; Curtain & Zwart, 1995; El Jaï & Pritchard, 1988; Kubrusly & Malebranche, 1985; Waldraff, Dochain, Bourrel, & Magnus, 2000), a static black-box approach is used. It is important to note that the sensor locations and the class of estimators are intrinsically related. In general, a particular sensor location could be optimal for a static linear estimator, but not for a Kalman filter, or for a nonlinear static estimator.

2.1. Mathematical framework

As mentioned above, we intend to derive a procedure to select a set of temperature measurements along the reactor to estimate the species concentrations and/or flows at the reactor outlet ($z = L$, where L is the total reactor length). In order to choose the optimal locations $z_{j_1}, \ldots, z_{j_{n_{\text{opt}}}}$ among the available locations $z_1, \ldots, z_n$ (with $n_{\text{opt}} \ll n$) and to design the estimator, some well-known facts about linear regression, which will be used throughout the paper, are here briefly recalled. The model

relating temperatures to component concentrations (or flows) is assumed to be linear around the nominal steady state and is approximated with the linear stochastic equation:

$$y(t) = Ax(t) + b + e(t) \tag{2}$$

where $y(t) \in \mathbb{R}^m$ is the vector of quality measurements (concentrations $\xi_i, i = 1, \ldots, m$ or flows $F_i, i = 1, \ldots, m$), $x(t) \in \mathbb{R}^n$ is the vector of process measurements (temperatures T_j, $j = 1, \ldots, n$), the matrix $A \in \mathbb{R}^{m\times n}$ and the vector $b \in \mathbb{R}^m$ are deterministic, but unknown, parameters, and $e(t) \in \mathbb{R}^m$ is the vector of measurement errors (random vector), which is assumed to be a white process. In the following the variables $y(t)$ and $x(t)$ will be also called the output vector and the input vector, respectively.

Starting from N observed samples of $x(\cdot)$ and $y(\cdot)$ at time $t_1, t_2, \ldots, t_N$, the following tail matrices $X \in \mathbb{R}^{n\times N}$ and $Y \in \mathbb{R}^{m\times N}$:

$$X = [\, x(t_1) \quad x(t_2) \quad \cdots \quad x(t_N)\,] \tag{3}$$

$$Y = [\, y(t_1) \quad y(t_2) \quad \cdots \quad y(t_N)\,] \tag{4}$$

are constructed. Observe that the number of rows of matrix X is equal to the number of available temperature sensor locations. By construction, also matrices X and Y satisfy Eq. (2), i.e.[2]

$$Y = AX + b + E, \tag{5}$$

where $E = [\, e(t_1) \quad e(t_2) \quad \cdots \quad e(t_N)\,]$ is the error matrix. Eq. (5) is the actual regression model, and the objective is to find an estimation $\hat{A}$ for the unknown matrix A, and an estimation $\hat{b}$ for the unknown vector b, such that the estimation $\hat{Y} = \hat{A}X + \hat{b}$ is optimal in least-squares sense. The symbol $\hat{}$ will be used to identify estimated signals or estimated matrices.

The estimation of Y given X is nothing else that the projection of the rows of Y onto the linear space $\mathcal{L}(X)$ generated by the rows of X (Anderson, 1958; Doob, 1990). Such projection will be denoted with the symbol $\hat{\mathbf{E}}[Y|\mathcal{L}(X)]$, where $\hat{\mathbf{E}}$ is the sample expectation. The estimation takes the form (Rao & Toutenburg, 1999):

$$\hat{Y} = \hat{\mathbf{E}}[Y|\mathcal{L}(X)] = \underbrace{YX^{\mathrm{T}}(XX^{\mathrm{T}})^{-1}}_{\hat{A}} X + \underbrace{m_Y - YX^{\mathrm{T}}(XX^{\mathrm{T}})^{-1}m_X}_{\hat{b}} \tag{6}$$

where

$$m_X = \hat{\mathbf{E}}[X] \triangleq \frac{1}{N}\sum_{k=1}^{N} x(t_k), \quad m_Y = \hat{\mathbf{E}}[Y] \triangleq \frac{1}{N}\sum_{k=1}^{N} y(t_k)$$

are the sample means of the inputs X and the outputs Y, respectively. Starting from the matrix $\hat{A}$ and $\hat{b}$ defined in Eq. (6), it is easy to compute the estimate $\hat{y}(t)$ of a new primary variable $y(t)$ given the corresponding value of the secondary variable $x(t)$. The ordinary least squares estimator (Rao & Toutenburg, 1999) is

$$\hat{y}(t) = \hat{A}x(t) + \hat{b}. \tag{7}$$

[2] With a slight misuse of notation, we write $Y = AX + b + E$ instead of $Y = AX + B + E$ with $B = [\, b \quad b \quad \cdots \quad b\,]$ to increase readability.

Note that the product matrices XX^{T} and XY^{T} in (6) are related to important statistical quantities, i.e.

$$\Sigma_X = \hat{\mathbf{E}}[XX^{\mathrm{T}}] \triangleq \frac{1}{N}XX^{\mathrm{T}}, \quad \Sigma_{XY} = \hat{\mathbf{E}}[XY^{\mathrm{T}}] \triangleq \frac{1}{N}XY^{\mathrm{T}}$$

where Σ_X is the sample variance of x and Σ_{XY} is the sample covariance of x and y (Anderson, 1958).

A difficulty in the design of the estimator (6) arises when the rows of the input matrix X are highly collinear. In this case the matrix XX^{T} is near to singularity and its inverse is difficult to compute. A solution to circumvent such problem is to select a smaller number of rows of X, $n_{\mathrm{opt}} \ll n$, in order to have a new input matrix X_{opt} retaining all the useful information contained in the original one, but farther from singularity. This means

$$n_{\mathrm{opt}} = \arg\min_{n^*}\{\mathsf{rowspan}\,(X) \simeq \mathsf{rowspan}\,(X_{\mathrm{opt}}), \quad X_{\mathrm{opt}} \in \mathbb{R}^{n^*\times N} \text{submatrix of } X \in \mathbb{R}^{n\times N}\} \tag{8}$$

In this manner the singularity problem is solved using a smaller number of sensors (rows of X) and disregarding the redundant information.

It will be shown that the methodology defined in this paper is such that it also quantifies the approximation introduced in the simplification of the input matrix. In other words, the proposed procedure quantifies the symbol "$\simeq$" in (8).

2.2. Motivation for the proposed approach

Several methodologies exist which try to solve the optimal sensor selection problem. The most widespread data-driven procedures are based on a modification of the PLS regression approach (see Geladi & Kowalski, 1986; Rao & Toutenburg, 1999; Kresta, Marlin, & MacGregor, 1994, and the Appendix A for the algorithm and the nomenclature). The PLS approach comprises the following steps:

(1) perform a PLS regression using all the available sensors;
(2) choose the optimal number n_{opt} of sensors and select their optimal locations according to one of the criteria (a), …, (d) reported below;
(3) extract the rows of the X matrix corresponding to the n_{opt} sensors;
(4) compute a regression using only this subset of X;
(5) check if the estimation is good enough.

However, a major weakness in the above procedure is that a general criterion to carry out the optimal sensor selection (step 2) cannot be properly defined. Some proposed criteria are:

(a) in any of the first n loadings, extract the component that has the greatest absolute value. By construction, the i-th sensor location corresponds to the i-th component in the w or p loading vectors (Geladi & Kowalski, 1986);
(b) in the first loading, extract the n components that have the greatest absolute values (Zamprogna et al., 2005);

(c) select the variable that has the greatest absolute value in the first loading, then remove it and perform another PLS regression concerning the remaining rows of X. Repeat such procedure until n sensors have been selected (Kaspar & Ray, 1992);
(d) detect whether clusters among the components of the first two or three loadings exist and (if so) select a variable for each clusters (Kaspar & Ray, 1992).

Note that stopping criterion (b) privileges the most significant direction, since the sensors corresponding to the larger components in the first loading vector are selected. Therefore, data with a score space of dimension greater than one do not have an accurate estimation. Stopping criterion (c) suffers from the same problem. Stopping criterion (a) appears in principle to be more effective since it selects a sensor from each of the first n_{opt} loading vectors. Stopping criterion criterion (d) is quite subjective and difficult to automate.

The approach presented in this paper is intrinsically sequential and avoids the ambiguity inside the PLS-based procedures. The underlying idea is to add sensors (i.e. rows of X) until the prediction of Y is "good enough", i.e. until some accuracy bound is satisfied.

With respect to the PLS method, the proposed approach has the following desirable characteristics:

(1) it is sequential in its formulation (sensor-by-sensor approach);
(2) its stopping rule has a strong statistical interpretation;
(3) it provides an indication of the correlation between the selected sensor and the other ones at each iteration of the sequential procedure;
(4) no difficulties arise when different types of sensors have to be integrated. Conversely, when the assigned measurements are different in nature (pressure measurements and/or concentration measurements, for example), it is not easy to select the optimal sensor with the PLS method only by looking at the value of the loading components, because information is displayed in an aggregated form.

The proposed stepwise regression methodology to address the sensor selection problem retains some similarities with the SROV algorithm of Shacham and Brauner (2003), and of Brauner and Shacham (2000). For example, also in the SROV algorithm the candidate input variables are selected to enter the regression model according to their level of correlation with the variables to be estimated, and an updating of the input and output spaces is carried out by subtraction from those spaces of the information already explained. However, it should be noted that Shacham's and Brauner's approach holds true for a single dependent variable only, while the proposed approach is explicitly designed for multi-outputs systems. Shacham and Brauner suggest as a stopping criterion that an input variable is removed from consideration of inclusion in the soft sensor when its residual information is at the noise level. A different stopping criterion will be considered herein, as it will be illustrated in the next section.

3. Proposed algorithm

The two underlying features of the proposed selection procedure concern the stopping criterion and the selection criterion. The stopping criterion is related to the Y-block explained variance (Geladi and Kowalski, 1986), which gives a measure on how the secondary variables are able to estimate the primary variables. When the explained variance increases, the prediction power of the estimator raises and the estimation error decreases. The selection criterion is based on the correlation between the primary and the secondary variables: choosing at any step the sensor having the larger correlation with the outputs, but taking also into account the previous selected sensors, those sensors that are more related to the outputs and have less common information are singled out. This is useful in order to reduce the numerical problems related to the calculation of the estimator matrices $\hat{A}$ and $\hat{b}$.

The data needed to the selection algorithm are the measurements of the secondary variables $X \in \mathbb{R}^{n\times N}$ and of the primary variables $Y \in \mathbb{R}^{m\times N}$; these values can be obtained (for example) by past measurements. The measurement selection algorithm proceeds as a series of iterations (starting from iteration no. 1), each one characterized by a sequence of steps; one measurement sensor (input) is selected at each iteration. Subscript i will be used to identify an iteration. The subscript i on a matrix (such as X_i and Y_i) is used to indicate that the matrix has been updated after iteration $(i-1)$; therefore, X_i and Y_i represent the working matrices at the beginning step of the i-th iteration. Before starting the first iteration, the sample means $m_X \in \mathbb{R}^n$ and $m_Y \in \mathbb{R}^m$ are removed from the input and output matrices, and all rows are scaled to unit variance obtaining the normalized matrices X_1 and Y_1.

The j index is used to identify an input measurement. Therefore, at the start of the i-th iteration, the X_i matrix can be thought as a stack of $(n-i+1)$ row vectors $X_i(j_i)$, each one having dimension N. The objective of iteration i is to identify the i-th most explanatory input measurement among the ones available at that iteration; this measurement will be denoted by j_i^*. The following steps are carried out in sequence at the i-th iteration.

The algorithm's steps (illustrated in Fig. 1) are next commented on.

(1) Compute the sample correlation matrix between X_i and Y_i

In order to select the i-th sensor (row of X_i) that is the most predictive with respect to the outputs Y_i, the sample correlation matrix[3] $\Gamma_{X_iY_i} \in \mathbb{R}^{(n-i+1)\times m}$ is first computed as

[3] When the data are only mean centered, the sample correlation matrix $\Gamma_{X_iY_i} \in \mathbb{R}^{(n-i+1)\times m}$ is computed as

$$\Gamma_{X_iY_i} = \Lambda_{X_i}^{-1/2}\Sigma_{X_iY_i}\Lambda_{Y_i}^{-1/2} \tag{9}$$

where $\Lambda_{X_i} = \mathsf{diag}\{\Sigma_{X_i}\}$ and $\Lambda_{Y_i} = \mathsf{diag}\{\Sigma_{Y_i}\}$ are the diagonal elements of the sample variances Σ_{X_i} and Σ_{Y_i}, and $\Sigma_{X_iY_i}$ is the sample covariance between X_i and Y_i.

(Anderson, 1958):

$$\Gamma_{X_iY_i} = \hat{\mathbf{E}}[X_i Y_i^{\mathrm{T}}] = \frac{1}{N} X_i Y_i^{\mathrm{T}}. \tag{10}$$

Note that the (j, h) element of $\Gamma_{X_iY_i}$, $\gamma_{j,h} = \Gamma_{X_iY_i}(j, h)$, explains the correlation between the j-th input measurement and the h-th quality output.

It is important to highlight that the row-dimension of matrix $\Gamma_{X_iY_i}$ decreases after an iteration has been completed. In fact, after an iteration has been completed, an input is selected and then removed from the set of input sensors available at the next iteration ($\Gamma_{X_1Y_1} \in \mathbb{R}^{n\times m}$ at the start of the first iteration, $\Gamma_{X_2Y_2} \in \mathbb{R}^{(n-1)\times m}$ at the start of the second iteration, $\Gamma_{X_iY_i} \in \mathbb{R}^{(n-i+1)\times m}$ at the start of the i-th iteration).

(2) Select the input $X_i(j_i^*)$ most correlated with the output Y_i

The goal is to choose the input index j_i^* representing the location of the sensor that is more predictive with respect to all components of the output vector. To this purpose, the sum of the absolute values of all correlation coefficients between each component (input) in X_i and all components (outputs) in Y_i is analyzed ($\sum_{h=1}^{m} |\Gamma_{X_iY_i}(j, h)|, \forall j = 1, \ldots, n - i + 1$). Then, the index j_i^* is selected by solving the following problem

$$j_i^* = \arg \max_{1\le j\le n-i+1} \sum_{h=1}^{m} |\Gamma_{X_iY_i}(j, h)|. \tag{11}$$

If some components in Y_i (i.e. some quality variables) are more important than others, a weighted sum can be considered. The selection criterion (11) guarantees that at any iteration the sensor having the largest cumulative correlation with all the outputs is selected.

(3) Compute the i-th regression

Using the principle of ordinary least squares (OLS), the OLS estimator $\hat{Y}_i = \hat{\mathbf{E}}[Y_i|\mathcal{L}(X_i)]$ is given by $\hat{a}_{j_i^*} X_i(j_i^*)$, where the i-th regression coefficient $\hat{a}_{j_i^*}$ is defined as follows (Rao & Toutenburg, 1999):

$$\begin{aligned}\hat{a}_{j_i^*} &= Y_i X_i^{\mathrm{T}}(j_i^*)(X_i(j_i^*)X_i^{\mathrm{T}}(j_i^*))^{-1} \\ &= \arg \min_{a\in\mathbb{R}^m} \|Y_i - aX_i(j_i^*)\|.\end{aligned} \tag{12}$$

Note that this regression coefficient is calculated in a different way with respect to the general form shown in Eq. (6). In fact, the data matrices are mean centered, which implies that the $\hat{b}$-matrix vanishes, and the soft sensor input is a scalar (only one measurement sensor), i.e. the $\hat{A}$-matrix is a column vector. This eliminates the numerical issues typically associated with the calculation of $(XX^{\mathrm{T}})^{-1}$ in Eq. (6).

(4) Update the input and output matrix data

When the i-th estimator is computed, data need updating. This implies the projection of the data Y_i onto the space orthogonal to $\mathcal{L}(X_i(j_i^*))$, or, in other words, to subtract $\hat{Y}_i$ from Y_i according to:

$$Y_{i+1} \triangleq Y_i - \hat{\mathbf{E}}[Y_i|\mathcal{L}(X_i(j_i^*))] = Y_i - \hat{Y}_i = Y_i - \hat{a}_{j_i^*} X_i(j_i^*) \tag{13}$$

The input matrix must be updated as well. After defining $X_i^c(j_i^*)$ as the matrix made of the rows of X_i minus the selected j_i^* row, i.e.

$$X_i^c(j_i^*) = \begin{bmatrix} X_i(1) \\ \vdots \\ X_i(j_i^* - 1) \\ X_i(j_i^* + 1) \\ \vdots \\ X_i(n - i + 1) \end{bmatrix}, \tag{14}$$

the updated input data matrix X_{i+1} is therefore given by

$$\begin{aligned} X_{i+1} &\triangleq X_i^c(j_i^*) - \hat{\mathbf{E}}[X_i^c(j_i^*)|\mathcal{L}(X_i(j_i^*))] \\ &= X_i^c(j_i^*) - \sum_{X_i^c(j_i^*)X_i(j_i^*)}^{-1} \sum_{X_i(j_i^*)} X_i(j_i^*). \end{aligned} \tag{15}$$

The above equations state that, in the next iteration, only the information contained in the output data and not already explained by the previous regressors $j_1^*, j_2^*, \ldots, j_i^*$ need explaining.

(5) Stopping rule

The above steps are repeated until a stopping rule is satisfied. This rule should establish whether the accuracy reached by the estimator is sufficient or not. One possible approach is to assume that the accuracy is measured by the explained variance σ_i of the Y-block at the i-th step, which is estimated from sensors $j_1^*, j_2^*, \ldots, j_i^*$:

$$\begin{aligned}\sigma_i &= \left\{1 - \frac{\sum_{k=1}^{N} \|Y_i(k,:) - \hat{Y}_i(k,:)\|^2}{\sum_{k=1}^{N} \|Y(k,:)\|^2}\right\} \times 100 \\ &= \left\{1 - \frac{\sum_{k=1}^{N} \|Y_{i+1}(k,:)\|^2}{\sum_{k=1}^{N} \|Y(k,:)\|^2}\right\} \times 100\end{aligned} \tag{16}$$

where $\|\cdot\|$ is the Euclidean norm and $Y(k, :)$ is the k-th row of Y. The algorithm stops when either the explained variance σ_i for the Y-block at the i-th iteration is large enough, or the incremental variation $\sigma_i - \sigma_{i-1}$ is not sufficient to motivate a further iteration since adding additional sensors would simply make the estimator less robust.

The algorithm ends by giving back the n_{opt} optimal locations. The estimator is then built using these measurements only.

4. Results and discussion

In this section, the optimal sensor location problem for the selected case studies is solved by applying the proposed algorithm. Two performance indices are introduced to evaluate and compare the estimators. The first performance index is the explained variance of the input-output signals (which appears also in the stopping rule of the selection algorithm, see

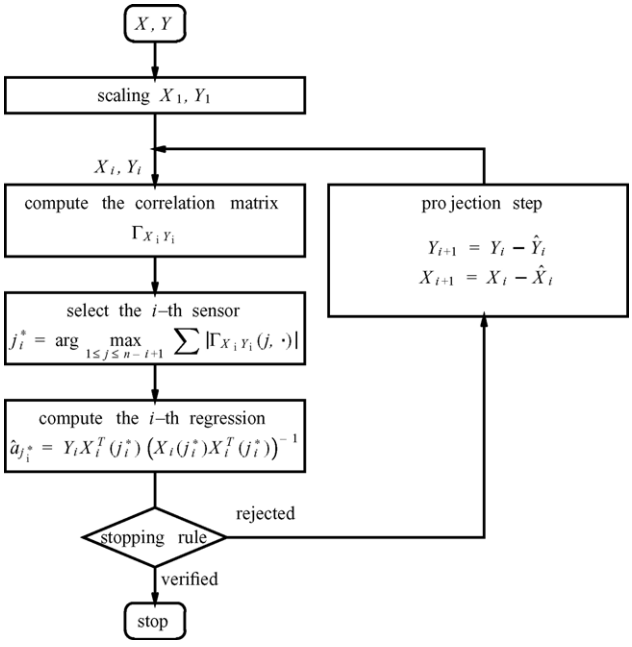

Fig. 1. Flowchart of the sequential method.

Eq. (16)) and the second one is the root mean square error (RMS):

$$\mathrm{RMS} = \left\{ \frac{1}{N} \sum_{k=1}^{N} \| y_i(k) - \hat{y}_i(k) \|^2 \right\}^{1/2} \tag{17}$$

where $y_i(\cdot)$ and $\hat{y}_i(\cdot)$ are the i-th output component and its estimation, respectively. Trajectories of measured and estimated signals are also compared graphically to visualize the estimation accuracy and the effect of the temperature measurement noise variance.

Two case studies are taken into account.

4.1. Hydro-dealkylation of toluene

The hydro-dealkylation (HDA) of toluene is a familiar process in chemical engineering. The adiabatic tubular reactor (HDA reactor in the following), where two vapour-phase reactions generate methane (M), diphenyl (D) and benzene (B), from reactants toluene (T) and hydrogen (H), is considered. The model equations and parameters are taken from Douglas (1988) and Luyben, Tyréus, and Luyben (1998) and are reported in Appendix B. Fig. 2a shows a typical steady-state temperature profile along the reactor, and Fig. 2c shows the evenly spaced distribution of the available temperature sensors along the reactor. Assuming to have 12 available sensors $T_1, \ldots, T_{12}$ numbered from 1 (at the reactor inlet) to 12 (at the reactor outlet), the objective is to find the most favorable locations for a subset of the available sensors in order to estimate the concentrations of benzene ξ_B and diphenyl ξ_D at the reactor outlet. The subset dimension should also be determined.

After the addition of white noise to corrupt the input and output data, the sequential algorithm was applied to the data time–series. Two plots are displayed in Fig. 3 for each one of the first three iterations of the selection algorithm. The plots on the left show the correlations, $\Gamma_{T_i\xi_B}$, and $\Gamma_{T_i\xi_D}$, between the temperature measurements along the reactor (T_i, $i = 1, \ldots, n$) and the concentrations of benzene $\xi_B(t, L)$ and diphenyl $\xi_D(t, L)$ at the reactor outlet, and the cumulative correlation obtained as the componentwise sum of the absolute values of such correlations $|\Gamma_{T_i\xi_B}| + |\Gamma_{T_i\xi_D}|$. The location corresponding to the largest cumulative correlation is selected at each iteration. Therefore, temperature sensor T_{11} is selected at the first iteration. This sensor is then removed, and sensor number 12 (T_{12}), which displays the greatest cumulative correlation, is selected at the second iteration. The third iteration removes both the previous locations (T_{11}, T_{12}), and selects temperature sensor T_{10}. It is important to observe that, by increasing the number of selected sensors, the magnitude of the correlation coefficients decreases. In fact, after each iteration, part of the information contained in the

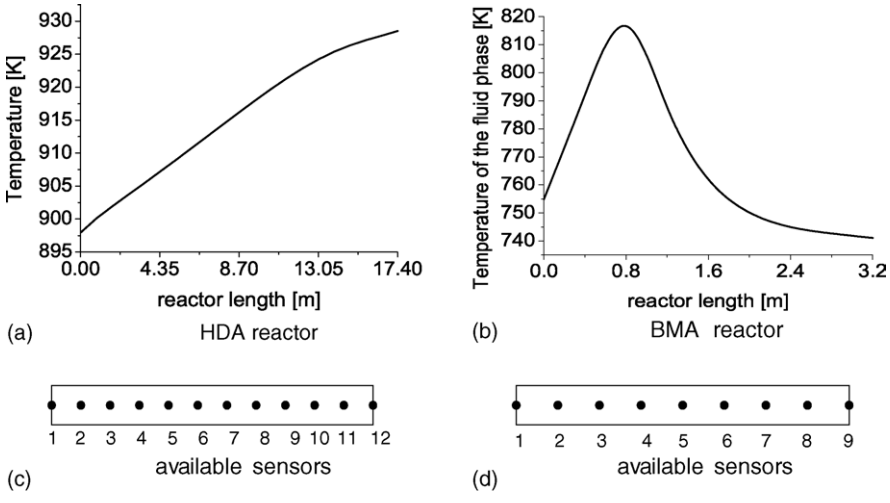

Fig. 2. (a) Steady-state temperature profile along the HDA reactor; (b) steady-state temperature profile along the BMA reactor; (c) available sensors for the HDA reactor; (d) available sensors for the BMA reactor.

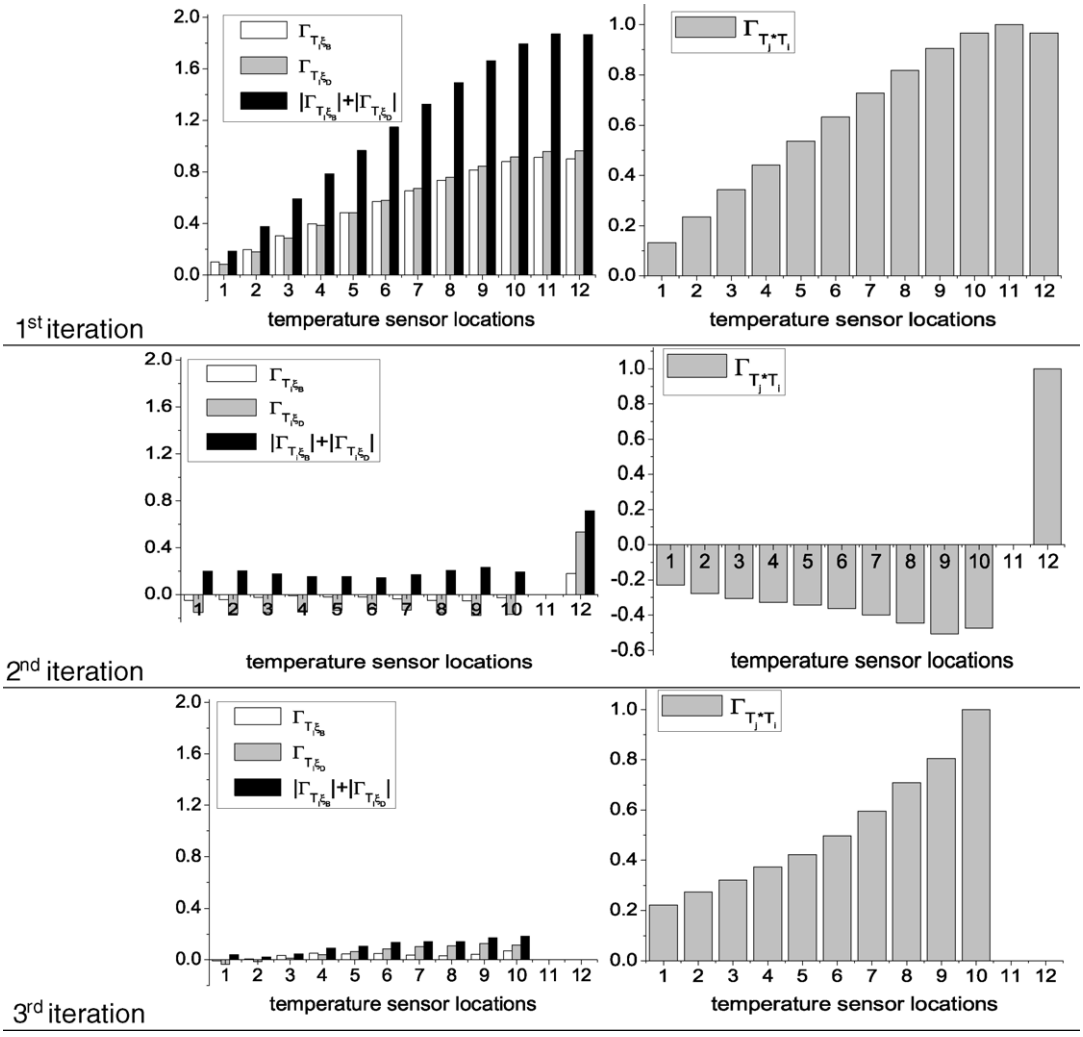

Fig. 3. HDA reactor. Left: correlations $\Gamma_{T_i\xi_B}$, $\Gamma_{T_i\xi_D}$ of the sensors T_i with the outputs ξ_B, ξ_D and the sum of the absolute value of such correlations, at each iteration of the algorithm. Right: correlation of the optimal sensor with the others sensors (and also with itself) at each iteration.

already selected sensor measurements is removed by projection (step (4) of the selection algorithm). At the end (third iteration) only "small" correlation coefficients with approximatively equal values remain. This is precisely the effect of the noise in the measurements. The plots on the right of Fig. 3 display the correlations between the optimal temperature measurement and the other temperature measurements at each iteration. It can be observed that, quite reasonably, the sensors having the largest correlation with the selected sensor are the ones located in its neighborhood.

Table 1 reports the explained variance of the output data (*Y*-block) for different sets of input measurements. To provide additional information, the explained variance is broken up in terms of latent variables (LVs) so as to display the most significant directions of the data. The first LV is the linear combination of the original variables that describes the direction of greatest variability, the second LV is the linear combination of the original variables that describes the second direction of greatest variability orthogonal to the first, and so on (Geladi & Kowalski, 1986). The explained variance of the *X* block is not reported since in the proposed approach the number of latent variables is equal to the dimension of the reduced *X* space and, accordingly, the variance explained for the *X*-block is always 100%. Note that the variance explained for the calibration *Y*-block, when three sensors are used, is only slightly larger than the one explained when using two sensors. Therefore, a soft sensor with only the first two temperature sensors (T_{11}, T_{12}) could be built; adding one more sensor (T_{10}) would mostly result in explaining measurement noise, which can be detrimental for estimation purposes. In fact, Table 2 shows that the RMS error on the calibration data decreases when three sensors are used, but that

Table 1
Sequential algorithm applied to the HDA reactor: how the *Y*-block explained variance increases by adding sensors

	Explained variances on calibration data (*Y*-block)			
Location(s)	First LV	Second LV	Third LV	Total
11	93.41	–	–	93.41
11,12	94.07	0.72	–	94.79
11,12,10	94.03	0.85	0.08	94.96

Table 2
Sequential algorithm applied to the HDA reactor: RMS for the estimation of benzene and dyphenil concentrations on calibration and validation data

Optimal location(s)	RMS calibration data			RMS validation data		
	ξ_B ($\times 10^3$)	ξ_D ($\times 10^3$)	Total ($\times 10^3$)	ξ_B ($\times 10^3$)	ξ_D ($\times 10^3$)	Total ($\times 10^3$)
11	2.0780	0.2593	2.0800	2.5449	0.3414	2.5500
11, 12	2.0737	0.2050	2.0762	2.5007	0.3194	2.5103
11, 12, 10	2.0276	0.2038	2.0304	2.5468	0.3160	2.5564

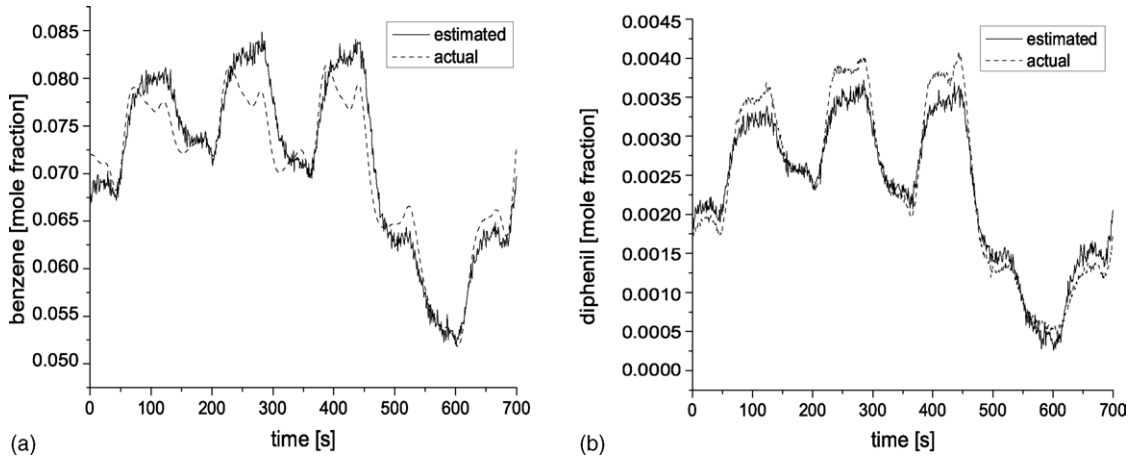

Fig. 4. Sequential algorithm applied to the HDA reactor (validation data): (a) benzene concentration, (b) diphenyl concentration (reactor exit).

adding the third sensor worsens the RMS error on the validation data.

Fig. 4a and b show the trajectories of the actual concentrations $\xi_B(t, L)$ and $\xi_D(t, L)$, together with the estimated ones $\hat{\xi}_B(t, L)$ and $\hat{\xi}_D(t, L)$. Note that the variance of $\hat{\xi}_B(t, L)$ (and $\hat{\xi}_D(t, L)$) is bigger than the variance of $\xi_B(t, L)$ (and $\xi_D(t, L)$). This depends on the fact that the estimator processes the temperature measurements, which are corrupted by noise (see Table B.3 in Appendix B). This noise propagates to the composition estimations, giving rise to an estimation noise larger than the measurement noise. The trajectories also highlight a small bias in the 'quasi steady-state' conditions. Since the estimator is static, no feedback is introduced to compensate for steady-state errors as, for example, in a Kalman filtering approach.

To check if the selected locations are indeed "optimal" for estimation purposes, the explained variance for the calibration *Y*-block data was calculated for all possible measurement pairs. The results for the most significant pairs are reported in Table 3. It can be seen that the (T_{11}, T_{12}) pair is actually the optimal one.

Table 3
HDA reactor: calibration *Y*-block explained variance for the most significant pairs of temperature sensors

(T_i, T_j)	8	9	10	11	12
6	62.95	75.72	87.21	93.61	92.83
7	63.48	76.83	87.67	93.58	93.04
8	–	77.73	88.37	93.62	93.41
9	–	–	88.47	93.64	93.89
10	–	–	–	93.67	94.54
11	–	–	–	–	94.79

4.2. Partial oxidation of benzene to maleic anhydride

The second case study considers the partial oxidation of benzene (B) to maleic anhydride (A). This case study is the same as the one considered by van den Berg et al. (2000). The model equations and parameters are taken from that paper and are reported in Appendix B. Fig. 2b and d show a typical steady-state temperature profile along the reactor (BMA reactor in the following) and the nine available temperature sensors, respectively. The objective is to select a suitable number of temperature sensors to obtain the best estimation for molar flow rates F_A and F_B at the reactor outlet.

Fig. 5 shows the correlations $\Gamma_{T_i F_A}$, $\Gamma_{T_i F_B}$ between temperature sensors and the component molar flow rates F_A and F_B when three sensors are selected, whereas Tables 4 and 5 report the RMS and the explained variances in this case study. They show that the optimal locations are more spatially distributed than in the previous case study (see Fig. 2b). This is related to the particular temperature profile in the BMA reactor. Note that a sensor is located near the temperature peak (hot spot). This

Table 4
Sequential algorithm applied to the BMA reactor: how the *Y*-block explained variance increases by adding sensors

Location(s)	Explained variances on calibration data (*Y*-block)			
	First LV	Second LV	Third LV	Total
7	89.91	–	–	89.91
7,9	87.90	3.09	–	90.99
7,9,4	85.63	8.66	2.25	96.54

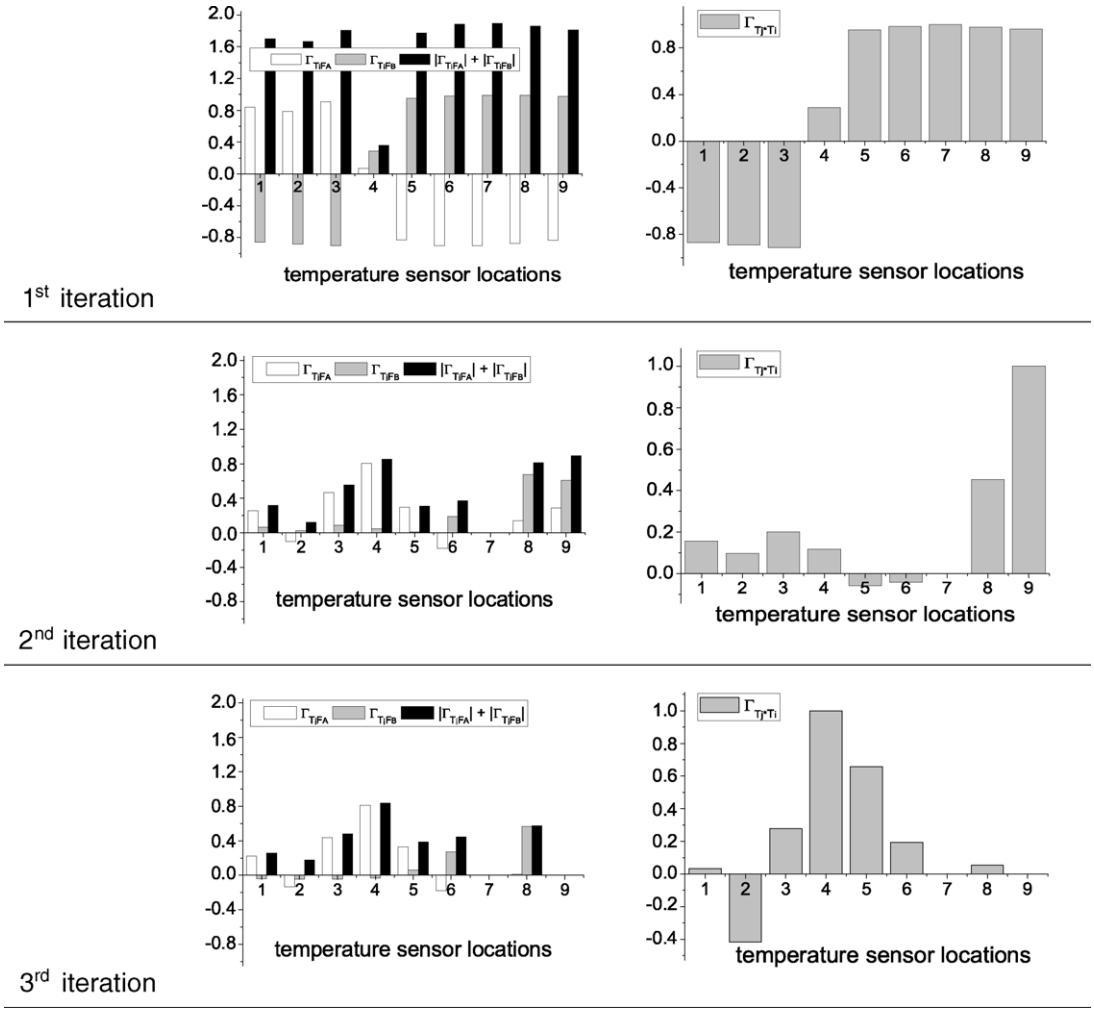

Fig. 5. BMA reactor. Left: correlations $\Gamma_{T_i F_A}$, $\Gamma_{T_i F_B}$ of the sensors T_i with the outputs F_A, F_B and the sum of the absolute value of the correlations, at each iteration of the algorithm. Right: correlation of the optimal sensor with the others sensors (and also with itself) at each iteration.

result agrees with what was obtained by van den Berg et al. (2000).

Fig. 6a and b show the measured and estimated flow rates, F_A, $\hat{F}_A$ and F_B, $\hat{F}_B$, at the reactor outlet. The same observations reported for the HDA reactor also hold here, although the estimation accuracy is much better in this case. Table 6 compares the Y-block explained variance for the best 10 triples of temperature sensors (out of the 1320 possible ones). In this case, the set of locations selected by the algorithm (T_4, T_7, T_9) is not the true "optimal"; however, the variance explained by this set is lower by less than 1% than the maximum achievable. Note however that techniques have been developed (e.g. Shacham & Brauner, 2003) to improve the selection sequence by changing the order in which the input variables enter the sequence itself.

Now let us suppose that only one product flow needs to be estimated, e.g. the benzene flow rate $F_B(t, L)$. Tables 7 and 8 show the explained variances and the RMS obtained in this subcase. Note that the optimal locations change with respect to the previous situation. This implies that the sensors which are more correlated with the benzene flow rate are located by the reactor outlet, whereas the sensor located near the hot spot are more important to estimate the maleic anhydride flow. Moreover, two sensors are sufficient and the estimation is improved with respect to the one obtained in the general case (see Tables 7 and 8).

Table 5
Sequential algorithm applied to the BMA reactor: RMS for the estimation of maleic anhydride and benzene molar flow rates on the calibration data

Optimal location(s)	RMS calibration data ($\times 10^5$ mol/s)			RMS validation data ($\times 10^5$ mol/s)		
	F_A	F_B	Total	F_A	F_B	Total
7	5.0588	1.8535	5.0601	6.8326	1.9769	6.8335
7, 9	4.8464	1.4720	4.8522	6.6048	1.6102	6.6189
7, 9, 4	2.8431	1.4713	2.8678	2.8327	1.6021	2.8940

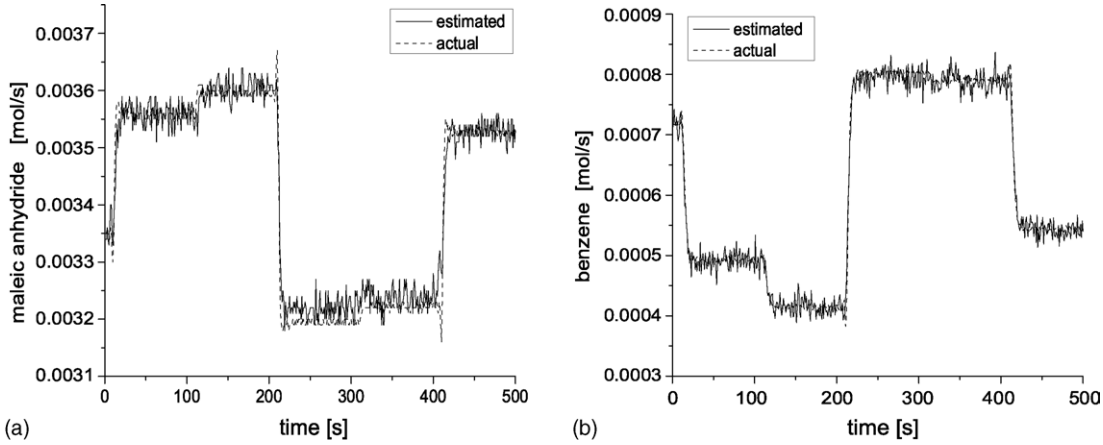

Fig. 6. Sequential algorithm applied to the BMA reactor (validation data): (a) trajectory of $F_A(t, L)$ and $\hat{F}_A(t, L)$, (b) trajectory of $F_B(t, L)$ and $\hat{F}_B(t, L)$ (reactor outlet).

Table 6
BMA reactor: calibration Y-block explained variance for the most significant triples of temperature sensors

T_i	T_j	T_h	Explained variance
4	6	8	97.13
4	6	9	97.12
4	6	7	96.97
4	5	7	96.66
4	7	9	96.54
2	4	7	96.43
3	4	7	96.42
4	5	8	96.31
4	7	8	96.31
1	4	7	96.31

Table 7
Sequential algorithm applied to the BMA reactor to estimate the benzene molar flow rate only: how the Y-block explained variances increase by adding sensors

	Explained variances on calibration data (Y-block)			
Location(s)	First LV	Second LV	Third LV	Total
7	98.27	–	–	98.27
7,8	99.04	0.02	–	99.06
7,8,9	99.08	0.17	0.004	99.26

Table 8
Sequential algorithm applied to the BMA reactor to estimate the benzene molar flow rate only: RMS for the estimation of benzene molar flow rate on the calibration and validation data

Optimal location(s)	RMS calibration data $F_B(\times 10^3)$	RMS validation data $F_B(\times 10^3)$
7	1.8535	1.9769
7,8	1.3679	1.4450
7,8,9	1.2155	1.3338

5. Conclusions

In this paper a multivariate statistical method has been presented to determine the optimal sensor measurement location in distributed sensor systems. The suggested algorithm is based upon a sequential procedure selecting at each iteration the most informative measurement input, and updating the input and output spaces by subtracting the information explained by the computed regressor. A static linear composition estimator is developed using the selected optimal measurement set. Neither the measurement selection algorithm nor the estimator require any explicit knowledge about the plant model. Although optimality of the input sequence cannot be formally guaranteed, extensive simulations showed that the proposed selection method is able to identify the sequence that, for practical purposes, is indeed the optimal one. Note that, even if two single-unit case studies have been considered to highlight the features of the proposed selection algorithm, the method can be easily extended to plantwide systems where several measurements of different nature must be integrated together to built a data-based property estimator.

Acknowledgement

This research was carried out in the framework of the MIUR-PRIN 2005 project “Integrated techniques for performance assessment and statistical process control in the process industry".

Appendix A. PLS algorithm

Let X and Y be the mean-centered and scaled to unit variance inputs and outputs data. The nonlinear iterative partial least squares (NIPALS) algorithm is (Geladi & Kowalski, 1986):

(1) Start: set u equal to a column of Y,
(2) Regress columns of X on u to get loadings: $w^T = \frac{u^T X}{u^T u}$,
(3) Normalize w to unit length: $w \leftarrow \frac{w}{w^T w}$,
(4) Calculate the scores: $t = \frac{Xw}{w^T w}$,
(5) Regress columns of Y on t : $q^T = \frac{t^T Y}{t^T t}$,
(6) Calculate new score vector for Y: $u = \frac{Yq}{q^T q}$,
(7) Check convergence of u: if YES go to 8, if NO go to 2,

(8) Calculate X matrix loadings by regressing columns of X on t : $p^{\mathrm{T}} = \frac{t^{\mathrm{T}}X}{t^{\mathrm{T}}t}$,
(9) Calculate residual matrices: $E = X - tp^{\mathrm{T}}$ and $F = Y - tq^{\mathrm{T}}$,
(10) To calculate the next set of latent vectors, replace X and Y by E and F and repeat.

Remark: the scores t_k are orthogonal by construction.

Appendix B. Model parameters

This Appendix reports the characteristic parameters of the examples under study and describes how the simulations were performed. The numerical values for the parameters (Table B.2) and the nominal conditions (Table B.1) are the same reported in Douglas (1988) and van den Berg et al. (2000) for the HDA and BMA reactors, respectively (Table B.2).

B.1. HDA reactor

The following two reactions are considered:

$$C_7H_8 + H_2 \rightarrow C_6H_6 + CH_4 \tag{B.1}$$

$$2C_6H_6 \rightleftharpoons C_{12}H_{10} + H_2. \tag{B.2}$$

The mass and energy balances are described by the following equations:

$$\frac{\partial \xi_{\mathrm{B}}}{\partial t} = -v\left(\frac{\partial \xi_{\mathrm{B}}}{\partial z} + \frac{\xi_{\mathrm{B}}}{T}\frac{\partial T}{\partial z}\right) + \frac{RT}{P}k_1\,\mathrm{e}^{-A_1/RT}(\xi_{\mathrm{T}}P)(\xi_{\mathrm{H}}P)^{0.5} - 2\frac{RT}{P}k_2\,\mathrm{e}^{-A_2/RT}(\xi_{\mathrm{B}}P)^2 + 2\frac{RT}{P}k_3\,\mathrm{e}^{-A_3/RT}(\xi_{\mathrm{D}}P)(\xi_{\mathrm{H}}P) \tag{B.3}$$

$$\frac{\partial \xi_{\mathrm{D}}}{\partial t} = -v\left(\frac{\partial x_{\mathrm{D}}}{\partial z} + \frac{x_{\mathrm{D}}}{T}\frac{\partial T}{\partial z}\right) + \frac{RT}{P}k_2\,\mathrm{e}^{-A_2/RT}(\xi_{\mathrm{B}}P)^2 - \frac{RT}{P}k_3\,\mathrm{e}^{-A_3/RT}(\xi_{\mathrm{D}}P)(\xi_{\mathrm{H}}P) \tag{B.4}$$

$$\frac{\partial \xi_{H}}{\partial t} = -v\left(\frac{\partial \xi_{\mathrm{H}}}{\partial z} + \frac{\xi_{\mathrm{H}}}{T}\frac{\partial T}{\partial z}\right) - \frac{RT}{P}k_1\,\mathrm{e}^{-A_1/RT}(\xi_{\mathrm{T}}P)(\xi_{\mathrm{H}}P)^{0.5} + \frac{RT}{P}k_2\mathrm{e}^{-A_2/RT}(\xi_{\mathrm{B}}P)^2 - \frac{RT}{P}k_3\,\mathrm{e}^{-A_3/RT}(\xi_{\mathrm{D}}P)(\xi_{\mathrm{H}}P) \tag{B.5}$$

$$\frac{\partial \xi_{\mathrm{M}}}{\partial t} = -v\left(\frac{\partial \xi_{\mathrm{M}}}{\partial z} + \frac{\xi_{\mathrm{M}}}{T}\frac{\partial T}{\partial z}\right) + \frac{RT}{P}k_1\,\mathrm{e}^{-A_1/RT}(\xi_{\mathrm{T}}P)(\xi_{\mathrm{H}}P)^{0.5} \tag{B.6}$$

$$\frac{\partial \xi_{\mathrm{T}}}{\partial t} = -v\left(\frac{\partial \xi_{\mathrm{T}}}{\partial z} + \frac{\xi_{\mathrm{T}}}{T}\frac{\partial T}{\partial z}\right) - \frac{RT}{P}k_1\,\mathrm{e}^{-A_1/RT}(\xi_{\mathrm{T}}P)(\xi_{\mathrm{H}}P)^{0.5} \tag{B.7}$$

Table B.1
Nominal operating conditions for the reactors

(a) HDA reactor	
Gas velocity v (m/s)	0.1977
Feed temperature $T(t, 0)$ (K)	904
Pressure P (kPa)	3468
Feed composition (mole fraction)	
$\xi_{\mathrm{B}}(t, 0)$	0.0053
$\xi_{\mathrm{D}}(t, 0)$	0
$\xi_{\mathrm{H}}(t, 0)$	0.4291
$\xi_{\mathrm{M}}(t, 0)$	0.4800
$\xi_{\mathrm{T}}(t, 0)$	0.0856
(b) BMA reactor	
Gas velocity v (m/s)	2.48
Temperature of the fluid phase $T_{\mathrm{f}}(t, 0)$ (K)	733
Temperature of the solid phase $T_{\mathrm{s}}(t, 0)$ (K)	733
Flow rates (mol/s)	
$F_{\mathrm{B}}(t, 0)$	0.009
$F_{\mathrm{A}}(t, 0)$	0

$$c_{\mathrm{p}}\frac{\partial T}{\partial t} = -vc_{\mathrm{p}}\frac{\partial T}{\partial z} - \Delta H_1\frac{RT}{P}k_1\,\mathrm{e}^{-A_1/RT}(\xi_{\mathrm{T}}P)(\xi_{\mathrm{H}}P)^{0.5} - \Delta H_2\frac{RT}{P}k_2\,\mathrm{e}^{-A_2/RT}(\xi_{\mathrm{B}}P)^2 - \Delta H_3\frac{RT}{P}k_3\,\mathrm{e}^{-A_3/RT}(\xi_{\mathrm{D}}P)(\xi_{\mathrm{H}}P) \tag{B.8}$$

Table B.2
Parameters of the reactor models

(a) HDA reactor	
k_1 (kmol/(s m^3 kPa$^{1.5}$))	54352.8
k_2 (kmol/(s m^3 kPa^2))	336.22
k_3 (kmol/(s m^3 kPa^2))	1433.79
A_1/R (K)	25616
A_2/R (K)	25616
A_3/R (K)	25616
R ((m^3 kPa)/(kmol K))	8.314
P (kPa)	3468
ρ (kg/m^3)	7.87
C_p (kJ/(kmol K))	68.80
ΔH_1 (kJ/kg)	50009
ΔH_2	0
ΔH_3	0
L (m)	17.4
(b) BMA reactor	
A_1 (s^{-1})	86760
A_2 (s^{-1})	37260
A_3 (s^{-1})	149.4
ΔH_1 (kJ mol^{-1})	-1490
ΔH_2 (kJ mol^{-1})	-2322
ΔH_3 (kJ mol^{-1})	-832
E_1 (J mol^{-1})	71711.7
E_2 (J mol^{-1})	71711.7
E_3 (J mol^{-1})	36026.3
U_{fw} (s^{-1})	10.6
U_{sf} (s^{-1})	84.0
T_{w} (K)	733
c_{s} (s K J^{-1})	0.729
D_{eff} ($\times 10^{-3}$ m^2 s^{-1})	3.17
k_{eff} ($\times 10^{-3}$ m^2 s^{-1})	3.17
L (m)	3.2

where $\xi_B, \xi_D, \xi_H, \xi_M, \xi_T$ are the mole fractions of benzene, diphenil, hydrogen, methane, toluene, respectively, P the pressure, and v is the velocity. It is assumed that pressure P and velocity v are constant through the reactor. The meaning and the value of the other parameters and the nominal conditions are reported in Tables B.2(a) and B.1(a).

B.2. BMA reactor

The feed stream is air mixed with benzene. Three exothermic, irreversible gas-phase reactions take place within a tubular packed reactor:

$$C_6H_6 + 4O_2 \rightarrow C_4H_2O_3 + CO + CO_2 + 2H_2O \quad \text{(B.9)}$$

$$C_6H_6 + 6O_2 \rightarrow 3CO + 3CO_2 + 3H_2O \quad \text{(B.10)}$$

$$C_4H_2O_3 + 2O_2 \rightarrow 2CO + 2CO_2 + H_2O. \quad \text{(B.11)}$$

The first reaction represents the desired outcome, while the other two the undesired burning of both reactant and product. The heat balance is defined for both the fluid phase and the stagnant solid-phase catalyst. The mass and heat balances are as follows:

$$\frac{\partial F_B}{\partial t} = -v\frac{\partial F_B}{\partial z} + D_{eff}\frac{\partial^2 F_B}{\partial z^2} - A_1\, e^{E_1/RT_s}F_B - A_2\, e^{E_2/RT_s}F_B \quad \text{(B.12)}$$

$$\frac{\partial F_A}{\partial t} = -v\frac{\partial F_A}{\partial z} + D_{eff}\frac{\partial^2 F_A}{\partial z^2} + A_1\, e^{E_1/RT_s}F_B - A_3\, e^{E_3/RT_s}F_A \quad \text{(B.13)}$$

$$\frac{\partial T_f}{\partial t} = -v\frac{\partial T_f}{\partial z} + k_{eff}\frac{\partial^2 T_f}{\partial z^2} - U_{fw}(T_f - T_w) - U_{sf}(T_s - T_f) \quad \text{(B.14)}$$

$$\frac{\partial T_s}{\partial t} = -U_{sf}(T_s - T_f) + c_s\Delta H_1 A_1\, e^{E_1/RT_s}F_B + c_s\Delta H_2 A_2\, e^{E_2/RT_s}F_B + c_s\Delta H_3 A_3\, e^{E_3/RT_s}F_A \quad \text{(B.15)}$$

where F_B, F_A are benzene and maleic anhydride molar flow rates, respectively, T_f the temperature of the fluid phase, and T_s is the temperature of the stagnant solid phase. The other parameters and the nominal conditions can be found in Tables B.2(b) and B.1(b). Velocity v is assumed to be constant along the reactor.

B.3. Simulated data

The simulated data were obtained by perturbing the inputs of the models reported in the above subsections: concentrations/temperature for the HDA reactor ($\xi_B(t, 0), \xi_D(t, 0), \xi_H(t, 0), \xi_M(t, 0), \xi_T(t, 0), T(t, 0)$) and flow rate/temperature for the BMA reactor ($F_B(t, 0)$, $T_f(t, 0)$). These perturbations were such that the reactors never attained a steady state condition. The inputs are perturbed around their nominal operating conditions (Table B.1) adding square signals having different amplitude and length. Figs. B.1 and B.2 show the obtained spatial and temporal behavior of all the interesting variables. The input–output data were partitioned in two groups: the calibration data

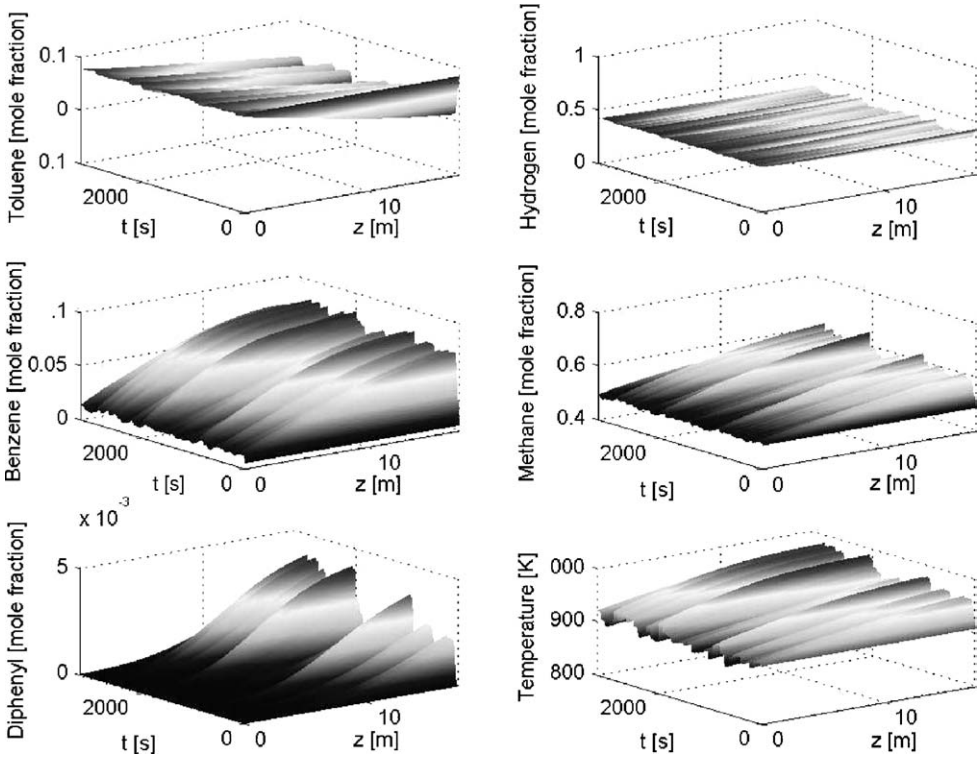

Fig. B.1. HDA reactor: concentrations and temperatures (calibration and validation data).

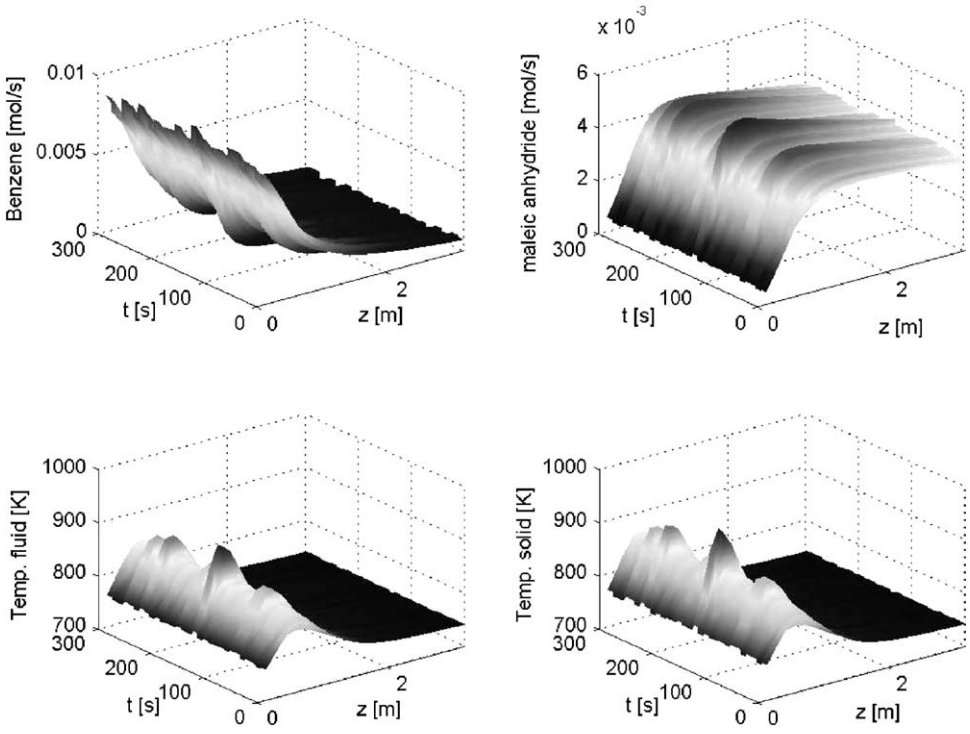

Fig. B.2. BMA reactor: molar flow rates and temperatures (calibration and validation data).

Table B.3
Standard deviations for (a) concentrations and temperature measurements for the HDA reactor; (b) flow rates and temperature measurements for the BMA reactor

(a) HDA reactor	
σ_{ξ_H} ($\times 10^{-3}$ mole fraction)	3.33
σ_{ξ_B} ($\times 10^{-5}$ mole fraction)	3.33
σ_{ξ_T} ($\times 10^{-4}$ mole fraction)	3.33
σ_{ξ_D} ($\times 10^{-5}$ mole fraction)	3.33
σ_{ξ_M} ($\times 10^{-3}$ mole fraction)	3.33
σ_T (K)	1.66
(b) BMA reactor	
σ_{F_A} ($\times 10^{-6}$ mol/s)	5
σ_{F_B} ($\times 10^{-6}$ mol/s)	3
σ_{T_f} (K)	1.66

(which are used to select the optimal sensor locations and to design the estimators) and the test data (which are used to validate the estimators). Both data are afterward corrupted by white Gaussian noises with standard deviations reported in Table B.3.

References

Alonso, A. A., Kevrekidis, I. G., Banga, J. R., & Frouzakis, C. E. (2004). Optimal sensor location and reduced order observer design for distributed process systems. *Computers & Chemical Engineering*, *28*, 27–35.

Anderson, T. W. (1958). *An Introduction to Multivariate Statistical Analysis* (3rd ed.). New York: John Wiley & Sons.

Brauner, N., & Shacham, M. (2000). Considering precision of data in reduction of dimensionality and PCA. *Computers & Chemical Engineering*, *24*, 2603–2611.

Curtain, R. F., & Zwart, H. (1995). *An Introduction to Infinite–Dimensional Linear Systems Theory*. New York: Springer-Verlag.

Doob, J. L. (1990). *Stochastic Processes*. New York: A Wiley/Interscience Publication. John Wiley & Sons, Inc.

Douglas, J. M. (1988). *Conceptual Design of Chemical Processes*. Singapore: McGraw-Hill International.

El Jaï, A., & Pritchard, A. J. (1988). *Sensors and Controls in the Analysis of Distributed Systems*. Chichester: Ellis Horwood Ltd.

Geladi, P., & Kowalski, B. R. (1986). Partial least-squares regression: A tutorial. *Analytica Chimica Acta*, *185*, 1–17.

Kano, M., Miyazaki, K., Hasebe, S., & Hashimoto, I. (2000). Inferential control system of distillation compositions using dynamic partial least squares regression. *Journal of Process Control*, *10*, 157–166.

Kaspar, M. H., & Ray, W. H. (1992). Chemometric method for process monitoring and high-performance controller design. *AIChE Journal*, *38*, 1593–1608.

Kookos, I. K., & Perkins, J. D. (1999). A systematic method for optimum sensor selection in inferential control systems. *Industrial and Engineering Chemistry Research*, *38*, 4299–4308.

Kresta, J. V., Marlin, T. E., & MacGregor, J. F. (1994). Development of inferential process models using PLS. *Computers & Chemical Engineering*, *18*, 597–611.

Kubrusly, C. S., & Malebranche, H. (1985). Sensors and controllers location in distributed systems—A survey. *Automatica Journal IFAC*, *21*, 117–128.

Luyben, W. L., Tyréus, B. D., & Luyben, M. L. (1998). *Plantwide Process Control*. New York, USA: McGraw-Hill.

Mejdell, T., & Skogestad, S. (1991). Estimation of distillation compositions from multiple temperature measurements using partial-least-squares regression. *Industrial and Engineering Chemistry Research*, *30*, 2543–2555.

Rao, C. R., & Toutenburg, H. (1999). *Linear Models. Least Squares and Alternatives* (2nd ed.). New York: Springer-Verlag.

Shacham, M., & Brauner, N. (2003). The SROV program for data analysis and regression model identification. *Computers & Chemical Engineering*, *27*, 701–714.

van den Berg, F. W. J., Hoefsloot, H. C. J., Boelens, H. F. M., & Smilde, A. K. (2000). Selection of optimal sensor position in a tubular reactor using robust degree of observality criteria. *Chemical Engineering Science*, *55*, 827–837.

Vande Wouwer, A., Point, N., Porteman, S., & Remy, M. (2000). An approach to the selection of optimal sensor locations in distributed parameter systems. *Journal of Process Control*, *10*, 291–300.

Waldraff, W., Dochain, D., Bourrel, S., & Magnus, A. (2000). On the use of observability measures for sensor location in tubular reactor. *Journal of Process Control*, *8*, 497–505.

Zamprogna, E., Barolo, M., & Seborg, D. (2005). Optimal selection of soft sensor inputs for batch distillation columns using principal component analysis. *Journal of Process Control*, *15*, 39–52.