

Virtual Sensing Technology in Process Industries: Trends and Challenges Revealed by Recent Industrial Applications

Manabu KANO and Koichi FUJIWARA

Department of Systems Science, Kyoto University, Yoshida-Honmachi, Sakyo-ku, Kyoto-shi, Kyoto 606-8501, Japan

Keywords: Softsensor, Input Variable Selection, Just-in-Time Modeling, Model Maintenance, Partial Least Squares, Industrial Application

Virtual sensing technology is crucial for high product quality and productivity in any industry. This review aims to clarify the trend of research and application of virtual sensing technology in process industries. After a brief survey, practical issues are clarified by introducing recent questionnaire survey results: 1) changes in process characteristics and operating conditions, 2) individual difference of equipment, and 3) reliability of soft-sensors. Since input variable selection is crucial for high estimation performance, conventional methods and new group-wise variable selection methods are introduced, and the usefulness of the group-wise variable selection methods is demonstrated through industrial case studies. Just-in-time (JIT) modeling is dealt with as a promising virtual sensing technology that can cope with changes in process characteristics as well as nonlinearity. Recent developments leading to successful industrial applications are introduced: correlation-based JIT (CoJIT) modeling and locally weighted regression (LWR), especially LW-PLS, with modified similarity measures. Manufacturing processes in different industries are quite different in appearance, but they have very similar problems from the viewpoint of quality issue. There remain practical issues requiring further research efforts to realize high-performance, maintenance-free virtual sensing technology.

Introduction

Virtual sensing technology, known as soft-sensors in the area of chemical engineering, is a key to success in estimating product quality or other important variables when online analyzers are not available. Even when hardware sensors can be used, operators and engineers have found the problems listed in **Table 1**. These problems with hardware sensors were identified as the result of a questionnaire to 26 companies in Japan (JSPS PSE 143 Committee, 2004). Soft-sensors are judged to be useful for addressing these problems.

Ookita (2006) investigated the present state of soft-sensor technologies in Mitsubishi Chemical Corporation. A part of the investigation results, i.e. the purpose of using soft-sensors, is shown in **Table 2**. The major purposes of using soft-sensors are: to stabilize product quality via online estimation, to reduce energy and material consumption via effective operation close to specifications/constraints, and to validate online analyzers by comparison with the soft-sensors.

Virtual sensing technology has been successfully applied to various processes in various industries. A great number of researches have been conducted to improve the estimation performance and to solve practical problems. This review aims to clarify the trend of research and application of

virtual sensing technology in the process industries. First, we look back over the research and development in this field by focusing on partial least squares (PLS). Then, practical issues concerning virtual sensing technology are pointed out by introducing the recent questionnaire survey results. The following issues are emphasized: 1) changes in process characteristics and operating conditions, 2) individual differences in equipment, and 3) reliability of soft-sensors. In addition, input variable selection is discussed. Besides well-known conventional methods, new group-wise variable selection methods are introduced. Those methods are compared through two case studies: an estimation of ethane concentration in an ethylene fractionator from online measured process variables and an estimation of the freezing temperature of diesel fuel from near-infrared (NIR) spectra. Furthermore, just-in-time (JIT) modeling is dealt with as a promising virtual sensing technology that can cope with changes in process characteristics as well as nonlinearity.

Table 1 Problems with hardware sensors (from the survey JSPS PSE143 2004)

Recognized problem	Percentage
Time-consuming maintenance	27%
Need for calibration	21%
Aged deterioration	15%
Insufficient accuracy	13%
Long dead-time, slow dynamics	10%
Large noise	8%
Low reproducibility	2%
Others	4%

Received on June 20, 2012; accepted on September 4, 2012

DOI: 10.1252/jcej.12we167

Correspondence concerning this article should be addressed to M. Kano (E-mail address: manabu@human.sys.i.kyoto-u.ac.jp).

Table 2 Purposes in using soft-sensors in Mitsubishi Chemical Corporation (Ookita, 2006)

Purpose	Percentage
Stabilize quality	37%
Reduce feed and utility	28%
Improve reliability with online analyzer	18%
Stabilize operation	9%
Reduce manual analysis	6%
Avoid installing special analyzer	1%
Analyze key factors	1%

Recent developments leading to successful industrial applications are introduced: correlation-based JIT (CoJIT) modeling and locally weighted regression (LWR), especially LW-PLS, with modified similarity measures. Finally, the review is concluded by discussing practical issues that we are facing.

1. Virtual Sensing Technology in the Past

Let us look back over the research and development of virtual sensing technology in the process industries. Here we focus on distillation processes as an example because distillation is a dominant unit operation for soft-sensor application (Kano and Ogawa, 2010). In addition, the history of soft-sensor development is divided into two eras by focusing on PLS because PLS changed approaches to soft-sensor design considerably.

1.1 Before the PLS era

In the petroleum refining and chemical industries, distillation is the most major target for soft-sensor application (Kano and Ogawa, 2010). In distillation columns, it is rarely the case that measurements of product compositions can be directly used as controlled variables, because online accurate measurement of compositions is difficult. Most analyzers such as gas chromatographs and near-infrared (NIR) spectrometric analyzers suffer from large measurement delays and high investment and maintenance costs.

In place of composition control using a product analyzer, tray temperature control has been widely used. The temperature control is based on the assumption that the product composition can satisfy its specification when an appropriate tray temperature is kept constant at its set-point. Of course, it is quite difficult to keep a product composition at its set-point by using temperature control when a feed composition or a feed flow rate changes in multicomponent distillation. Pressure changes also cause temperature variations. To solve these problems, many approaches have been proposed. Rademaker *et al.* (1975) suggested that the influence of nonkey components could be reduced by locating a temperature measurement in the region of the column where their compositions were nearly constant. Yu and Luyben (1984) used the other differential temperature for nonkey-component compensation. Whitehead and Parnis (1987) used a weighted average of many differential temperatures

for disturbance compensation and also used the temperature difference for pressure compensation. Bozenhardt (1988) used multiple temperatures to track the maximum temperature difference between two trays in an alcohol–water–ether column. They found that the position of this maximum difference was strongly correlated to the product composition. Although the performance of tray temperature control might not be sufficient, process knowledge is fully utilized to realize better composition control.

To realize online composition control without a product analyzer, inferential control can be used; a product composition is estimated from other measured process variables and the estimates are used for control. It is crucial to build a highly accurate inferential model to enhance the performance of inferential control. A first-principle model is preferred as far as it is available and provides sufficient accuracy with reasonable modeling effort and computational load. However, if it is not the case, an empirical (statistical) model determined from process data are an inevitable choice.

Although multiple regression analysis (MRA) has been widely used to build soft-sensors (Kano and Ogawa, 2010), this method does not always function well due to the highly correlated nature of process data, i.e. collinearity. For example, in distillation, tray temperatures close to each other change in nearly the same way. Applying MRA to such highly correlated data leads to numerical errors and singularity. In general, the results are not useful for prediction due to the problem called overfitting or overparametrization.

The simplest solution to the problem of collinearity and overfitting is to select a few input variables, which are uncorrelated, from all measured variables. Many articles have been published on this matter (Weber and Brosilow, 1972; Joseph and Brosilow, 1978; Morari and Stephanopoulos, 1980; Moore *et al.*, 1987). However, this simple approach is not optimal, because additional measurements may improve the performance of an estimator.

On the other hand, Brosilow and co-workers (Weber and Brosilow, 1972; Joseph and Brosilow, 1978) proposed a composition estimator called the Brosilow estimator, in which temperatures and flow rates were used for estimating unmeasured disturbances and then the derived disturbance values were used to estimate product compositions. This estimator is based on a linearized process model.

These researches were conducted in the 1970s and 1980s before PLS was applied to soft-sensor design.

1.2 In the PLS era

In the 1990s, composition estimators using PLS were proposed (Mejdell and Skogestad, 1991; Kresta *et al.*, 1994). In their work, inferential models of product compositions were built using steady-state data. However, it is difficult to obtain ideal steady-state data since industrial distillation columns are perturbed by disturbances. Even if steady-state data can be obtained, a steady-state estimator may not work well for dynamic operating conditions. Mejdell and Skogestad (1993) compared three different estimators using a linear model of a binary distillation column, and they concluded

that good control performance could be achieved with the steady-state principal component regression (PCR) estimator, which was almost as good as the dynamic Kalman filter, because the steady-state estimator had an inherent feedforward effect. The inherent feedforward effect was investigated in more detail by Showchaiya *et al.* (2001) and Kano *et al.* (2003a). They suggested using predictive inferential control with a dynamic inferential model within the cascade control configuration to achieve good performance without demanding iterative modeling. Kano *et al.* (2000) further investigated PLS-based inferential models, which can estimate the product compositions of the multicomponent distillation column. They compared steady-state, static, and dynamic inferential models and found that the estimation accuracy could be greatly improved by using dynamic models. In other words, it is beneficial to take process dynamics into account to improve the performance of soft-sensors.

PLS is popular in industry; nevertheless artificial neural networks (ANN) have been dominant in the literature since the middle 1990s (Kano and Nakagawa, 2008). ANN is a useful tool for building nonlinear models and is supposed to be suitable for industrial processes. However, linear models have produced satisfactory results in many cases because industrial processes are operated within certain range to produce the required products and linear approximation functions well. Thus, PLS has been popular as a practical tool for soft-sensor design. In recent years, support vector machine (SVM), support vector regression (SVR), and other kernel-based methods have emerged (Boser *et al.*, 1992; Cortes and Vapnik, 1995). These methods have attracted researchers' and engineers' attention and have been used for soft-sensor design (Yan *et al.*, 2004; Desai *et al.*, 2006).

Another method for developing soft-sensors is subspace identification (SSID), which can build a state space model from input and output data (Verhaegen and Dewilde, 1992; Overschee and Moor, 1994). SSID is a useful tool to build a dynamic inferential model of a multivariable process, and it is suitable for soft-sensor design because the performance of soft-sensors can be improved by taking process dynamics into account (Kano *et al.*, 2000). Amirthalingam and Lee (1999) used SSID for inferential control of a continuous pulp digester. Amirthalingam *et al.* (2000) developed a two-step procedure to build SSID-based inferential control models, in which the stochastic part was identified from historical data and the deterministic part was identified from plant test data. In addition, Kano *et al.* (2009) proposed two-stage SSID to develop soft-sensors that can estimate unmeasured disturbances without assumptions that the conventional Kalman filtering technique must make. Thus, it can outperform the Kalman filtering technique when innovations are not Gaussian white noises or the properties of disturbances do not stay constant with time. The superiority of the two-stage SSID over conventional methods was demonstrated through their application to an industrial ethylene ractionator.

2. Virtual Sensing Technology in the Present

In this section, the recent questionnaire survey result related to industrial soft-sensor application is introduced, and then practical issues concerning virtual sensing technology are described. Another literature survey provided by Kadlec *et al.* (2009) is also helpful to understand the present state of data-driven soft-sensors.

2.1 Questionnaire survey

In Japan, the 143rd committee on process systems engineering (PSE143) in the Japan Society for the Promotion of Science (JSPS) has provided sites where engineers in industry and researchers in academia can meet and discuss particular fascinating topics. For example, a task force named "Workshop No. 25 Control Performance Monitoring (WS25-CPM)" had been active since 2002 for about two years, then another task force named "Workshop No. 27 Process Control Technology (WS27-PCT)" since 2007 for about two years. These two task forces, headed by the first author, were financially supported by PSE143, which is operated with membership fees paid by about 30 companies. At present in 2012, PSE143 consists of about 70 members from academia and 60 members from industry. The participating companies cover petrochemical, chemical, refining, engineering, food, instrumentation, software, and other industries.

WS27-PCT, consisting of 32 engineers from industry and 12 researchers from universities, conducted a questionnaire survey of process control applications in 2009 (Kano and Ogawa, 2010). The respondents are 21 engineers of 15 companies, which include 7 petrochemical, 6 chemical, 3 engineering, 2 petroleum refineries, and 3 other companies.

A part of the questionnaire survey results related to soft-sensors is introduced here. This questionnaire asked control engineers the number of soft-sensor applications, targeted processes, methods for designing soft-sensors, and problems to be solved. The survey result, summarized in **Table 3**, clarifies the state of the art of soft-sensor application in Japan. The total number of soft-sensor applications answered was 439. A major targeted process is distillation (331/439), followed by reaction (86/439) and polymerization (20/439). A major modeling method is MRA (293/439), followed by PLS (93/439). Old soft-sensors tend to be built with MRA, and more recent ones are built with PLS. Nonlinear modeling methods such as ANN are rarely used. It is confirmed that linear regression such as MRA and PLS can achieve sufficient estimation accuracy for most distillation and reaction processes. Although the number of soft-sensor applications has increased rapidly, that of ANN has decreased for more than a decade. This interesting contrast reveals the gap between theory and practice (Kano and Nakagawa, 2008).

On the other hand, Ookita (2006) reported other statistics of soft-sensor application in Mitsubishi Chemical Corporation as summarized in **Table 4**. Most soft-sensor applications were found in distillation processes. In addition, it was reported that most soft-sensors were developed through

Table 3 Statistics of soft-sensor applications (from the survey JSPS PSE143 WS27-PCT 2009)

Process	Methodology							Total
	Phys	MRA	PLS	O.L.	ANN	JIT	Gray	
Distillation	20	256	41	6	0	5	3	331
Reaction	5	32	43	0	0	5	1	86
Polymerization	0	4	8	0	3	0	5	20
Others	0	1	1	0	0	0	0	2
Total	25	293	93	6	3	10	9	439

Phys: physical model, MRA: multiple regression analysis, PLS: partial least squares regression, O.L.: other linear regression, ANN: artificial neural network, JIT: just-in-time model, Gray: gray-box model: hybrid model between physical model and statistical model.

PLS.

The questionnaire of WS27-PCT in 2009 asked engineers what are problems related with applications of soft-sensors; the answers are summarized in **Table 5**. This result confirms that the maintenance of models is the most critical issue concerning soft-sensors. In fact, a total of 36% of the engineers pointed out the necessity to cope with changes in process characteristics and operating conditions in order to keep the performance of soft-sensors. It is also very important to lessen the burdens of modeling including data acquisition and preprocessing. Other issues are to attain sufficient accuracy, to evaluate reliability, and to justify cost performance.

The authors believe it is crucial for engineers to use a modeling method suitable for the target. The method and the derived model should be as simple as possible in order to achieve the desired performance including robustness by spending minimum time and cost. Engineers do not have to use a complex technique unless necessary.

Hereafter in this section, practical issues concerning virtual sensing technology are explained in more detail.

2.2 Changes in process characteristics and operating conditions

Generally, building a high performance soft-sensor is very laborious, since input variables and samples for model construction have to be selected carefully and parameters have to be tuned appropriately. Even if a good soft-sensor is developed successfully, its estimation performance deteriorates when process characteristics or operating conditions change. In chemical processes, for example, equipment characteristics are changed by catalyst deactivation, scale adhesion, and others. In semi-conductor processes, equipment characteristics and operating conditions change drastically after preventive maintenance of the equipment. Another common cause is the variation of feed properties; this problem is observed in various industries. In particular, the variation of feed properties is recognized as a critical problem regarding quality assurance in the pharmaceutical industry. Such changes and variations may lead to a deterioration of estimation performance and consequently incur a quality risk. In practice, the maintenance of soft-sensors is crucial in keeping their estimation performance as revealed from the questionnaire survey results. Soft-sensors should be updated as the process characteristics change, but manual

Table 4 Processes which soft-sensors are applied to in Mitsubishi Chemical Corporation (Ookita, 2006)

Process	Percentage
Distillation	73%
Reaction	25%
Evaporation	1%
Others	1%

Table 5 Problems of soft-sensor applications (from the survey JSPS PSE143 WS27-PCT 2009)

Accuracy deterioration due to changes in process characteristics	29%
Burden (time/cost) of data acquisition	22%
Burden of modeling itself	14%
Burden of data preprocessing	7%
Inadequate accuracy since installation	7%
Inadequate accuracy due to changes in operating conditions	7%
Difficulty in evaluating reliability	7%
Unjustifiable cost performance	7%

and repeated construction of them should be avoided due to the heavy workload.

To cope with changes in process characteristics and operating conditions, recursive methods such as recursive PLS were developed. The basic idea is updating covariance matrices when new data become available in order to update statistical models automatically. Helland *et al.* (1992) proposed an original recursive PLS algorithm, which updates the model based on new data without increasing the size of data matrices. This algorithm is recursive but not exponentially weighted, thus its adaptation is very slow. To solve this problem, Dayal and MacGregor (1997) proposed exponentially weighted PLS and applied it to a flotation circuit. Moreover, to solve the numerical stability problems that the original recursive algorithm suffered from, Qin (1998) proposed a block-wise recursive PLS algorithm with a moving window and forgetting factor adaptation schemes. This recursive PLS was also extended to dynamic modeling and nonlinear modeling.

A number of papers regarding recursive techniques have been published, and industrial applications and validation results with real plant data have been provided in the literature. Several recent examples are as follows. Mu *et al.*

(2006) proposed a dual updating strategy by integrating recursive PLS and the model output offset update. The high performance of this strategy was demonstrated through its application to an industrial purified terephthalic acid (PTA) purification process in which the average crystal particle size was predicted. Similarly, Ahmed *et al.* (2009) used a recursive PLS as well as a model bias update to predict the melt index (MI) in grade-changing polymerization of high density polyethylene (HDPE) processes. The prediction of MI was also investigated by Liu (2007). In the proposed recursive algorithm, principal component analysis (PCA) was used to eliminate collinearity, fuzzy c-means (FCM) clustering was used to decompose the operating space into several local regions, and fuzzy Takagi–Sugeno (FTS) modeling was used to build a local linear model for each region. The developed algorithm was validated through its application to MI prediction in a polyethylene process producing multiple grade products. Sbarbaro *et al.* (2008) proposed the use of constrained parameter estimation algorithms to take account of some process prior knowledge and enhance the performance of the soft-sensor based on the recursive least squares. The proposed method was applied to an industrial grinding circuit. Liu *et al.* (2010) proposed fast moving window PLS (FMWPLS) based on the dissimilarities between the new and oldest data that were incorporated into the kernel algorithm for PLS, FMW-PLS was applied to oxygen concentration estimation in an air separation process. Kadlec and Gabrys (2011) proposed the incremental local learning soft sensing algorithm (ILLSA) by exploiting the local learning framework and applied it to a polymerization reactor. In this method, locally valid models were built and maintained to model the nonlinear relationship between the input and output with a combination of linear models. Ni *et al.* (2012) proposed a localized and adaptive recursive PLS (LARPLS). In this algorithm, both local model adaptation and local time regions adaptation were conducted, and also three adaptive strategies, i.e. means and variances adaptation, adaptive forgetting factor, and adaptive extraction of local time regions, were provided. LARPLS was applied to three types of polymerization processes. Galicia *et al.* (2012) extended the reduced-order dynamic PLS (RO-DPLS) to its adaptive version to address frequent process changes in a pulp digester. The result of applying recursive RO-DPLS to the digester under feedback control have shown that the control performance was improved by feeding estimates back to a PID controller. In the semiconductor industry, Pan *et al.* (2011) proposed an adaptive virtual metrology (VM) based on recursive canonical variate analysis (RCVA) and applied it to an industrial sputtering process.

These recursive methods can adapt models to new operating conditions gradually. However, the model may adapt excessively and not function in a sufficiently wide range of operating conditions when a process is operated within a narrow range for a certain period of time. To prevent excessive recursive PLS update, Ahmed *et al.* (2009) proposed a scheme that minimizes the number of recursive PLS update runs while maintaining the model.

In addition, recursive methods cannot cope with abrupt changes in process characteristics. Such abrupt changes are very common in the semiconductor industry, where process characteristics change drastically after preventive maintenance of equipment. In such a situation, JIT modeling is desirable.

A review of adaptation techniques was given by Kadlec *et al.* (2011). The concept drift theory was exploited to classify the algorithms into three different types: 1) moving windows techniques, 2) recursive adaptation techniques, and 3) ensemble-based methods. They explained significant algorithms and also provided a list of publications, in which adaptive soft-sensors were proposed and applied to practical problems. This review has mainly focused on recursive algorithms and has taken little account of JIT modeling. Hence, JIT modeling and its industrial application are reviewed in detail here in Section 4.

2.3 Individual differences in equipment

In addition to changes in process characteristics and operating conditions, the individual differences in production devices should be considered in some industrial applications. In semiconductor processes, for example, tens of parallelized production devices are used in each of hundreds of processes, and they have different characteristics even if their catalog specifications are the same. Thus, a soft-sensor developed for one device is not always applicable to another device, and it is very laborious to customize soft-sensors according to their individual difference.

To cope with such individual differences in equipment as well as changes in process characteristics, Fujiwara *et al.* (2012a) proposed the nearest correlation (NC) method and integrated it with CoJIT modeling. The NC method is a correlation-based pattern recognition method that can detect samples whose correlation is similar to the query without any teacher signals on the basis of sample geometry. On the other hand, CoJIT modeling is a kind of JIT modeling, but it selects the samples used for local modeling on the basis of correlation together with distance (Fujiwara *et al.*, 2009). CoJIT and its industrial application are described in more detail in Section 4.1. NC-CoJIT is based on the idea that the individual differences in equipment are expressed as the difference of the correlation among variables. Fujiwara *et al.* (2012a) demonstrated the advantage of NC-CoJIT over conventional JIT modeling through a case study of product composition estimation for a parallelized chemical reaction process.

Fujiwara *et al.* (2010) extended the NC method by integrating it with spectral clustering. Spectral clustering is a clustering method based on the graph theory (Ding *et al.*, 2001; Ng *et al.*, 2001). It can partition a weighted graph, whose weights express affinities between nodes, into subgraphs through cutting some of their arcs. In the new clustering method, referred to as NC-spectral clustering (NCSC), the correlation-based affinities between samples are calculated by using the NC method to construct a weighted graph, and the constructed weighted graph is

partitioned by spectral clustering. On the basis of NCSC, a correlation-based soft-sensor design method was proposed, and its usefulness was demonstrated through a case study of parallelized batch processes (Fujiwara *et al.*, 2010). The performance of NCSC-based soft-sensor has been shown to be better than that of the conventional distance-based methods.

NCSC can be used not only for clustering samples and soft-sensor design but also for clustering variables and selecting variables for modeling. Variable selection based on NCSC is explained in Section 3.8.

2.4 Reliability of soft-sensors

A great deal of research has been conducted to develop data-driven soft-sensors for various processes. A data-driven soft-sensor, however, does not always function well. The estimation performance deteriorates when one or more sensors fail and measurements are not true, since the validity of soft-sensors relies on the validity of measurements of input variables. In addition, a soft-sensor is not valid when a process is operated outside certain conditions where operation data used for modeling are obtained. In other words, statistical models can be used for interpolation but should not be used for extrapolation. The product quality and process performance will deteriorate if estimates of the soft-sensor are blindly believed by operators and used in a control system. Online monitoring of the validity of soft-sensors will avoid such a dangerous situation. The simplest approach is to check whether an estimation error exceeds its control limit when a measurement becomes available. This approach can detect the inconsistency between the analyzer and the soft-sensor, but the validity of the soft-sensor can be checked only when analyzed data are available. Thus, another approach to online validation needs to be developed.

To address such practical problems and to monitor the validity of soft-sensors online, Qin *et al.* (1997) proposed a self-validating inferential sensor approach based on multivariate statistical process control (MSPC) (Jackson and Mudholkar, 1979; Kresta *et al.*, 1991). First, input sensors are validated before making an estimation. If a faulty sensor is detected and identified, the corresponding measurement is reconstructed with its estimate based on PCA. Then, it is checked whether the inferential model is extrapolating through a χ^2 test for the PCA scores. Finally, output variables are estimated by using principal components as input variables, i.e. principal component regression (PCR). At this step, reconstructed principal components are used for estimation if one sensor is faulty. This approach was applied to air emission monitoring for a boiler process, where NO_x emissions were monitored.

Later, Liu *et al.* (2010) extended this reconstruction-based approach to the case of multiple sensor failures. Since a moving window approach including their FMW-PLS is sensitive to outliers, the confidence intervals of predictions, which were derived from the uncertainties of the output and input variables, were created to avoid inferential models being misled by abrupt noise from the online analyzers. This approach was applied to oxygen concentration estimation in

an air separation process.

In industry practice, it is usually assumed that estimation error is caused by inaccurate estimation; however, this assumption is not always true because analyzers are not always reliable. For example, a hardware sensor cannot provide accurate measurements when blockage occurs within a sampling line. To address such a practical problem and to monitor the validity of soft-sensors online, Kamohara *et al.* (2004) proposed a PLS-based framework, in which a dynamic PLS-based soft-sensor was integrated with an online monitoring system to judge whether the soft-sensor was reliable or not. The monitoring system was based on MSPC in which the dynamic PLS model was used. In addition, simple rules were established for checking the performance of a process gas chromatograph by combining the soft-sensor and the monitoring system. The developed system was applied to an industrial ethylene plant. Since the system was able to check the validity of the soft-sensors online, only reliable estimates of product quality were fed to a multivariable model predictive control (MPC) system.

Another approach was proposed by Kaneko *et al.* (2009), who adopted independent component analysis (ICA) to detect abnormal situations and improve the prediction accuracy. By using an ICA-based fault detection and classification model, a PLS-based inferential model can be updated with only normal samples. In the last decade, ICA has been recognized as a useful technique for fault detection and diagnosis (Kano *et al.*, 2003b, 2004; Lee *et al.*, 2004). The combination of virtual sensing technology and statistical process monitoring is effective in practice.

To check the validity of estimates, it seems useful to integrate statistical process monitoring with virtual sensing technology as several researches have suggested. However, the online implementation of this approach confronts a difficult problem: how to distinguish between normal changes and faults. As mentioned in Section 2.2, soft-sensors are required to cope with changes in process characteristics, operating conditions, and feed properties. To adapt the soft-sensor to a new situation, it needs to be updated by using new data that look different from old data. However, such new data will be recognized as faulty data by statistical process monitoring, which is based on historical normal data. This problem is critical and needs to be solved when data-driven techniques including virtual sensing technology and statistical process monitoring should be adaptive.

3. Input Variable Selection

In general, a soft-sensor can be fitted to samples used for modeling when the number of input variables is large. However, its estimation performance may deteriorate due to overfitting when input variables that do not physically relate to the output are used. Thus, appropriate input variables have to be selected when a soft-sensor is constructed. Although physical knowledge of a process is useful for selecting input variables, it is difficult to understand the effect of each input variable on output variables when the number

of variables becomes large. In addition, physical phenomena have not been perfectly clarified in many processes. In such cases, it is difficult to select input variables of a soft-sensor on the basis of physical knowledge. If all possible combinations of candidate variables are tested, the computational load increases remarkably. To improve estimation performance as well as efficiency of soft-sensor design, a systematic methodology for selecting appropriate input variables needs to be developed.

This section briefly explains conventional variable selection methods such as the stepwise method and the least absolute shrinkage and selection operator (Lasso). PLS can cope with collinearity, i.e. correlated input variables, to which MRA cannot be directly applied; nevertheless variable selection is important for improving the estimation performance of PLS. To select appropriate input variables for PLS, several measures have been proposed such as PLS-Beta, variable influence on projection (VIP), and the selectivity ratio (SR). These PLS-based variable selection measures are also described here.

In addition, a new methodology for selecting variables in PLS modeling, referred to as NCSC-based variable selection (NCSC-VS), is introduced. In NCSC-VS proposed by Fujiwara *et al.* (2012b), variables are clustered into some variable classes based on the correlation between variables by using NCSC (Fujiwara *et al.*, 2010, 2011), and each variable class is examined as to whether it should be used as input variables. This method does not evaluate individual variables but evaluates groups of variables similarly to group Lasso proposed by Yuan and Lin (2006). Since variable groups need to be generated in advance for group Lasso, NCSC can be used to generate variable groups. This idea derives NCSC-based group Lasso (NCSC-GL).

The above-mentioned variable selection methods are compared through their applications to soft-sensor design for an industrial chemical process and the estimation of the freezing temperature of diesel fuel from near-infrared (NIR) spectra.

3.1 Partial least squares (PLS)

PLS is a linear regression method that can cope with the collinearity problem; thus it has been used as a modeling tool in various industries where process variables are highly correlated. Here PLS is briefly explained to define symbols.

In PLS with one output variable, input data $\mathbf{X} \in \mathcal{R}^{N \times M}$ and output data $\mathbf{y} \in \mathcal{R}^N$ are decomposed as follows.

$$\mathbf{X} = \mathbf{T}\mathbf{P}^T + \mathbf{E} \quad (1)$$

$$\mathbf{y} = \mathbf{T}\mathbf{b} + \mathbf{f} \quad (2)$$

where $\mathbf{T} \in \mathcal{R}^{N \times R}$ is a latent variable matrix whose columns are latent variables $\mathbf{t}_r \in \mathcal{R}^N$ ($r = 1, 2, \dots, R$), $\mathbf{P} \in \mathcal{R}^{M \times R}$ is a loading matrix of \mathbf{X} and its columns are loading vectors \mathbf{p}_r , $\mathbf{b} = [b_1, b_2, \dots, b_R]^T$ is a loading vector of \mathbf{y} , and \mathbf{E} and \mathbf{f} are errors. N , M , and R denote the number of samples, that of input variables, and that of adopted latent variables, respectively.

The nonlinear iterative partial least squares (NIPALS) algorithm can be used to construct the PLS model (Wold *et al.*, 2001). Suppose that the first to $(r-1)$ th latent variables $\mathbf{t}_1, \mathbf{t}_2, \dots, \mathbf{t}_{r-1}$, the loading vectors $\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_{r-1}$ and b_1, b_2, \dots, b_{r-1} are given. The r -th residual input and output can be written as follows.

$$\mathbf{X}_r = \mathbf{X}_{r-1} - \mathbf{t}_{r-1}\mathbf{p}_{r-1}^T \quad (3)$$

$$\mathbf{y}_r = \mathbf{y}_{r-1} - b_{r-1}\mathbf{t}_{r-1} \quad (4)$$

where $\mathbf{X}_1 = \mathbf{X}$ and $\mathbf{y}_1 = \mathbf{y}$. The latent variable \mathbf{t}_r is a linear combination of the columns of \mathbf{X}_r , that is, $\mathbf{t}_r = \mathbf{X}_r\mathbf{w}_r$, where $\mathbf{w}_r \in \mathcal{R}^M$ is the r -th weighting vector. PLS aims to maximize the covariance between \mathbf{y}_r and \mathbf{t}_r under the constraint $\|\mathbf{w}_r\| = 1$. By using the Lagrange multipliers method, \mathbf{w}_r is derived as

$$\mathbf{w}_r = \frac{\mathbf{X}_r^T \mathbf{y}_r}{\|\mathbf{X}_r^T \mathbf{y}_r\|} \quad (5)$$

The r -th loading \mathbf{p}_r and b_r are as follows.

$$\mathbf{p}_r = \frac{\mathbf{X}_r^T \mathbf{t}_r}{\mathbf{t}_r^T \mathbf{t}_r} \quad (6)$$

$$b_r = \frac{\mathbf{y}_r^T \mathbf{t}_r}{\mathbf{t}_r^T \mathbf{t}_r} \quad (7)$$

Finally, the above procedure is repeated until the number of adopted latent variables R is achieved; R can be determined by cross validation.

The estimation performance of PLS models can be improved by taking process dynamics into account. For this purpose, the past information is used as inputs in addition to the present information. In such a case, the input variable vector \mathbf{z}_i is written as

$$\mathbf{z}_i = [\mathbf{x}_i^T, \mathbf{x}_{i-k_1}^T, \mathbf{x}_{i-k_2}^T, \dots]^T \quad (8)$$

where k_1, k_2, \dots are i or less arbitrary natural numbers. This method is referred to as dynamic PLS (Ricker, 1998; Kano *et al.*, 2000). Galicia *et al.* (2011) have shown that dynamic PLS is theoretically a dynamic estimator that can adequately capture the process dynamics if the past horizon is long enough. In addition, they proposed reduced-order dynamic PLS (RO-DPLS) to reduce the number of lagged variables that are required in dynamic PLS, and applied it to a Kamyr digester.

3.2 Stepwise

The stepwise method selects input variables of an MRA model on the basis of a hypothesis test. It repeats model construction by adding and eliminating a variable step by step, and tests whether the true value of the regression coefficient of the added or eliminated variable is zero by using the F -test (Hocking, 1976).

In the stepwise method, the threshold of the p -value is used for variable selection. In addition, the use of Akaike's information criterion (AIC) and Bayesian information criterion (BIC) has been proposed (Ward, 2008). However, it

is not possible to determine the best variable selection index for the stepwise method since almost the same result can be achieved through tuning their threshold regardless of the index adopted (Murtaugh, 2009).

3.3 PLS-Beta

The PLS-Beta method translates a PLS model into an MRA model and selects input variables based on the magnitude of its regression coefficients. The estimate \hat{y} is expressed as follows (Chong and Jun, 2005).

$$\hat{y} = T(T^T T)^{-1} T^T y = X\beta_{pls} \quad (9)$$

$$\beta_{pls} = W(P^T W)^{-1} (T^T T)^{-1} T^T y \quad (10)$$

where $W = [w_1, w_2, \dots, w_R]$ and β_{pls} denotes the regression coefficients.

3.4 Variable influence on projection (VIP)

The VIP score of an input variable is a summary of the importance for the projection to find latent variables, expresses the contribution of the input variable to the output variable, and thus can be used for variable selection. Hence, VIP is also known as variable importance in the projection. VIP is a weighted sum of squares of the PLS weights with the weights calculated from the amount of Y -variance of each PLS component (Wold *et al.*, 2001). The VIP score of the m -th variable is defined as

$$V_m = \sqrt{\frac{M \sum_{r=1}^R (b_r^2 t_r^T t_r (w_{mr} / \|w_r\|)^2)}{\sum_{r=1}^R b_r^2 t_r^T t_r}} \quad (11)$$

where w_{mr} is the m -th element of the weighting vector of the r -th latent variable t_r . The variables satisfying $V_m > \mu$ should be selected as input variables. Since the average of squared VIP scores equals 1, $\mu = 1$ is suggested as a criterion for variable selection (Chong and Jun, 2005). However, from the authors' experience, μ needs to be adjusted in order to improve the estimation performance.

3.5 Selectivity ratio (SR)

The SR score is a variable selection index based on PLS (Rajalahti *et al.*, 2009). SR is defined as the ratio of the variance of the input variables explained by the latent variables v_j^{expl} to the variance of the errors v_j^{err}

$$s_j = \frac{v_j^{\text{expl}}}{v_j^{\text{err}}} \quad (12)$$

because the input variables are expressed as the linear combination of the latent variables and the errors as written in Eq. (1). The input variables can be individually selected in descending order of the SR scores s_j .

3.6 Least absolute shrinkage and selection operator (Lasso)

Lasso is a constrained version of MRA and tends to

shrink its regression coefficients toward zero by constraining the sum of the absolute regression coefficients (Tibshirani, 1996). Lasso minimizes the sum of the squared errors subject to the L_1 norm of the regression coefficients $\|\beta\|_1 = \sum_i |\beta_i|$ less than a constant η .

$$\beta_{\text{lasso}} = \arg \min_{\beta} \|y - X\beta\|^2 \quad (13)$$

$$\text{s.t. } \|\beta\|_1 \leq \eta \quad (14)$$

or equivalently

$$\beta_{\text{lasso}} = \arg \min_{\beta} (\|y - X\beta\|^2 + \lambda \|\beta\|_1) \quad (15)$$

In Lasso, some regression coefficients tend to be zero and this characteristic can be used for variable selection. That is, variables corresponding to regression coefficients that are close to zero are eliminated from the model.

Although Eqs. (13) and (14) are formulated as a quadratic programming (QP) problem, this problem can be solved by an efficient algorithm called the least angle regressions (LARS) (Efron *et al.*, 2004).

3.7 Group Lasso

Group Lasso is an extension of Lasso for selecting groups consisting of multiple variables (Yuan and Lin, 2006). Suppose that M variables are divided into J groups; X_j and β_j denote the input matrix and the regression coefficient vector corresponding to the j -th group, respectively. The number of variables in the j -th group is M_j , i.e. $M = \sum_{j=1}^J M_j$. The optimization problem of group Lasso is formulated as

$$\beta_{\text{gl}} = \arg \min_{\beta} \left(\left\| y - \sum_{j=1}^J X_j \beta_j \right\|^2 + \lambda \sum_{j=1}^J \sqrt{M_j} \|\beta_j\| \right) \quad (16)$$

where $\beta = [\beta_1^T, \beta_2^T, \dots, \beta_J^T]^T$, and λ is a tuning parameter. In group Lasso, variable groups have to be determined in advance.

3.8 NCSC-based variable selection

In the NCSC-based method, variables are clustered into some variable classes based on the correlation between variables, and each variable class is examined as to whether it should be used as input variables. Since NCSC is used for clustering variables, the method is referred to as NCSC-based variable selection (NCSC-VS). Although NCSC was originally proposed for sample clustering (Fujiwara *et al.*, 2010, 2011), it can be used for variable clustering.

First, NCSC-VS clusters variables into J variable classes $v_j = \{x_m | m \in \mathcal{V}_j\}$ ($j = 1, 2, \dots, J$) based on their correlation, where \mathcal{V}_j is the j th subset of indexes of variables and $\mathcal{V} = \bigcup_j \mathcal{V}_j$. To classify the measured variables by NCSC, an affinity matrix is constructed from the transposed input data matrix X^T by using the NC method, a weighted graph is constructed on the basis of the affinity matrix, and the weighted graph is partitioned by spectral clustering. Next, the j th PLS model f_j is constructed from the j th variable class matrix X_j , whose columns consist of variables belonging to v_j , and the

validity of f_j is examined on the basis of the contribution ratio to the estimates. The contribution ratio for evaluating f_j with r latent variables is defined as

$$C_{j,r} = 1 - \frac{\|\hat{\mathbf{y}}_{j,r}\|^2}{\|\mathbf{y}\|^2} \quad (17)$$

where $\hat{\mathbf{y}}_{j,r}$ denotes the estimates by f_j with r latent variables.

Finally, D ($\leq J$) variable classes are selected in descending order of $C_{j,r}$ and the final PLS model is constructed from the variables belonging to the selected variable classes. In NCSC-VS, the number of latent variables r for evaluating the contribution ratio $C_{j,r}$ and that of adopted variable classes D used in the final PLS model are tuning parameters. In addition, the number of latent variables in the final PLS model can be determined by cross validation.

3.9 NCSC-based group Lasso

As mentioned, variable groups need to be generated in advance for group Lasso; therefore, NCSC can be used to generate variable groups for group Lasso. This idea derives NCSC-based group Lasso (NCSC-GL), in which J variable groups are constructed in the same manner as NCSC-VS, and variable groups are selected by using Eq. (16).

3.10 Comparison 1: Ethylene fractionator

Fujiwara *et al.* (2012b) compared several variable selection methods through their application to soft-sensor design for an industrial chemical process. A soft-sensor for estimating ethane concentration was constructed to realize the highly efficient operation of the ethylene fractionator at the Showa Denko K.K. (SDK) Oita plant in Japan (Kamohara *et al.*, 2004).

In the ethylene fractionator, 33 variables are measured. To account for the dynamics of the ethylene fractionator, the candidates of input variables consist of the present sample and the past samples measured 5, 10, 15, 20, 25, 30, 35, 40 and 45 min earlier, since engineers and operators know that the ethylene fractionator achieves a new steady state in less than an hour. As a result, the total number of candidate variables is 330 (33×10) and appropriate input variables have to be selected to design a precise soft-sensor.

The soft-sensors for estimating the ethane concentration were designed on the basis of variable selection results

obtained with the above-mentioned methods. A dynamic PLS model was constructed using all of the candidate variables, and the number of its adopted latent variables was determined by cross validation. Input variables were selected by PLS-Beta, VIP, and SR based on the constructed dynamic PLS model. Parameters were tuned through cross validation. In addition, the SDK engineers selected input variables on the basis of their physical process knowledge (Kamohara *et al.*, 2004). Since the number of SDK selection variables is 18, the total number of input variables of the soft-sensor is 180 ($= 18 \times 10$).

The estimation results are summarized in **Table 6**. In this table, #Var is the number of selected variables, #LV is the number of adopted latent variables, R denotes the correlation coefficient between estimates and measurements, and RMSE is the root-mean-square error. The SDK selection, PLS-Beta, SR, Lasso, and the stepwise method achieved almost the same performance as dynamic PLS, while the number of input variables of the soft-sensors was reduced. However, the performance of VIP deteriorated and could not be improved, whichever threshold μ was selected. On the other hand, NCSC-VS achieved the highest performance of all the methods and RMSE was improved by about 36% in comparison with dynamic PLS.

3.11 Comparison 2: NIR spectra

Another case study is introduced here. Lasso, group Lasso, NCSC-VS, and NCSC-GL are compared in the estimation performance through a case study of diesel fuel property data provided by Eigenvector Research Inc. (<http://www.eigenvector.com>). The objective property is the freezing temperature of diesel fuel, which is estimated from near-infrared (NIR) spectrum data. Candidate input variables are the intensity measured on a total of 401 wavelengths between 750–1500 nm. Appropriate wavelengths need to be selected to build a model that can estimate the freezing temperature precisely. Since the intensity changes very similarly on neighbor wavelengths, it seems beneficial to adopt group-wise variable selection methods. There are 115 samples for model construction and 136 samples for model validation. Before modeling, a Savitzky–Golay smoothing filter of the first-order derivation (Savitzky and Golay, 1964) was applied to the spectrum data.

A PLS model employing all the wavelengths (PLS-All)

Table 6 Estimation results of ethane concentration in product ethylene of the ethylene fractionator through PLS models based on different variable selection methods

	#Var	#LV	RMSE	R
Dynamic PLS	330	47	28.7	0.88
SDK selection	180	37	23.8	0.88
Stepwise	151	40	26.6	0.89
PLS-Beta	117	25	25.7	0.87
VIP	293	26	31.5	0.87
SR	271	45	26.0	0.85
Lasso	217	30	28.1	0.88
NCSC-VS	200	35	18.0	0.92

was constructed. The wavelengths were selected by Lasso and group Lasso. Before group Lasso was used, the full-spectrum region was divided into several wavelength groups so that they consisted of the same number of successive wavelengths. Such preprocessing is commonly used in spectrum analysis (Norgaard *et al.*, 2000). The parameters were selected through cross validation.

The estimation results are summarized in Table 7. The estimation performance of Lasso was worse than PLS-All. On the other hand, group Lasso improved the estimation performance while the number of input wavelengths was reduced. In addition, NCSC-VS and NCSC-GL achieved higher performance than group Lasso; their RMSEs were improved by about 9% and 7% in comparison with PLS-All, respectively. These results demonstrate the usefulness of NCSC-based group-wise variable selection.

4. Just-in-Time Modeling

In the recent review of adaptation techniques for data-driven soft-sensors (Kadlec *et al.*, 2011), a list of publications is provided. However, this review has mainly focused on recursive algorithms and has taken little account of JIT modeling. Hence, JIT modeling and its industrial applications are reviewed in detail in this section.

Just-in-time (JIT) modeling, or JIT learning, was proposed to cope with changes in process characteristics as well as nonlinearity, and it has been used not only for virtual sensing but also for process monitoring. JIT modeling is also known as lazy learning (Atkeson *et al.*, 1997; Maron and Moore, 1997; Bontempi *et al.*, 1999; Birattari *et al.*, 1999). In JIT modeling, a local model is built from past data around a query, i.e. a new sample, only when an estimate is required. This approach is useful when a global model does not function well.

In comparison with the conventional modeling, JIT modeling has the following features: 1) when new input and output data become available, they are stored into a database; 2) only when estimation is required, a local model is constructed from samples located in the neighborhood around the query, and output variables are estimated; 3) the constructed local model is discarded after its use for estimation. In JIT modeling, samples for local modeling should be selected appropriately each time, and online computational load tends to become large.

A simple approach to selecting neighbors of a query is k nearest neighbors (k -NN). Although k -NN works in

practice, the estimation performance of JIT modeling can be improved by choosing neighbors in a more sophisticated way. Also, a simple approach to building a local model is a weighted average (Kibler *et al.*, 1989), which is preferable to other local modeling methods when the number of neighbors is limited. However, the estimation performance can be improved by adopting local linear regression (LLR), e.g., MRA and PLS. Another approach is to build a local model as a Taylor series of the global model expanded in the vicinity of the query (Su *et al.*, 2012).

Gupta *et al.* (2008) proved that the variance of estimates of LLR is bounded by the variance of measurement noise if a query is in the convex hull enclosing its neighborhood. This fact stresses the importance of defining the neighborhood so that it encloses the query. Consequently, the authors suggested three enclosing neighborhood strategies: enclosing k -nearest neighbors (ek -NN), natural neighbors, and natural neighbors inclusive. Another convex neighborhood based on the Delaunay tessellation was proposed by Jin *et al.* (2003). Zhu *et al.* (2011) compared these enclosing neighborhood strategies and non-enclosing neighborhood strategies such as k -NN and correlation-based neighbors adopted in CoJIT through their applications to soft-sensor design for estimating the ethane concentration in a deethanizer. The results confirmed that an appropriate selection of the neighborhood was critical for learning a local regression model and enclosing neighborhoods were preferable. The advantages of enclosing neighborhoods over non-enclosing neighborhoods are not only their stability but also their automatic adaptivity to the topology of the measurements, which eliminates the necessity of user-defined parameters. In the steel industry, Zheng and Kimura (2001) proposed k bipartite neighbors 2 (k -BN2) to select k neighbors and combined local linear models with a global nonlinear model to improve estimation performance. The proposed method was validated through its applications to a hot strip rolling mill and another plate rolling mill.

Instead of deciding whether or not to use each sample stored in a database, one can use all samples by prioritizing them according to their importance. The importance is evaluated by the similarity between the query and samples, e.g., the distance between the query and samples. The prioritizing strategy is more general because it is the same as deciding to use or not to use samples when the similarity is binary $\{0, 1\}$. Figure 1 illustrates the different prioritizing strategies: an exponential type and a binary type. Thus, JIT modeling is also known as locally weighted regression

Table 7 Estimation results of freezing temperature of diesel fuel from near-infrared (NIR) spectrum data through PLS models based on different variable selection methods

	#Var	#LV	RMSE	R
PLS-All	401	4	2.69	0.71
Lasso	33	4	2.80	0.74
Group Lasso	165	5	2.54	0.78
NCSC-VS	42	10	2.45	0.80
NCSC-GL	186	6	2.49	0.80

(LWR) (Cleveland, 1979).

LWR constructs a local model by prioritizing samples in a database according to the similarity between a query and the samples. Hence, LWR can cope with abrupt changes as well as gradual ones in contrast to the adaptive methods reviewed by Kadlec *et al.* (2011). In addition, it can cope with nonlinearity since it builds a piecewise model. LWR usually outperforms a typical weighted average method. To build an accurate model with LWR, the similarity needs to be properly defined. In general, similarity is defined on the basis of the Euclidean distance or the Mahalanobis distance (Walden and Prescott, 1983; Cleveland and Devlin, 1988; Naes *et al.*, 1990; Centner and Massart, 1998; Leung *et al.*, 2004; Lee *et al.*, 2005). Other similarity measures proposed so far are based on the distance and the angle (Chang *et al.*, 2001; Cheng and Chiu, 2004; Ge and Song, 2008), the distance and the output estimate derived by a global model (Wang *et al.*, 1994), the correlation among input variables (Fujiwara *et al.*, 2009, 2010), and the weighted Euclidean distance (Kim *et al.*, 2011; Shigemori *et al.*, 2011; Nakagawa *et al.*, 2012). For example, Shigemori *et al.* (2011) developed a quality design system for steel products that can determine the optimal manufacturing conditions for securing the target quality based on LWR, in which weights are defined as regression coefficients of a global MRA model. The developed system has been operated at JFE Steel's West Japan Works for more than eight years.

Since it is crucial to define the similarity appropriately in order to achieve high estimation performance, recent researches are introduced in this section with emphasis on this issue. The first one is correlation-based JIT (CoJIT) modeling, and the second one is LWR, especially locally weighted PLS (LW-PLS), with the weighted Euclidean distance.

4.1 Correlation-based JIT (CoJIT) modeling

The estimation performance of distance-based JIT modeling is not always high because samples used for local modeling are selected on the basis of the distance from the query and the correlation among variables is not taken into account. A good model cannot be developed when the correlation among input and output variables is weak even when the distance between samples is small. Conversely, a very accurate model can be developed when the correlation is strong even if the distance is large. On the basis of this idea, Fujiwara *et al.* (2009) proposed correlation-based JIT (CoJIT) modeling. In this technique, the samples used for local modeling are selected on the basis of correlation together with distance, and the Q statistic is used as an index of the correlation dissimilarity. The Q statistic, which is derived from PCA, is a measure of dissimilarity between the sample and the modeling data from the viewpoint of the correlation among variables (Jackson and Mudholkar, 1979).

In PCA, the loading matrix $V_R \in \mathcal{R}^{M \times R}$ is derived as the right singular matrix of a data matrix $X \in \mathcal{R}^{N \times M}$ whose n -th row is \mathbf{x}_n^T , and the column space of V_R is the subspace spanned by principal components. Here, M , N , and R denote the numbers of variables, samples, and principal

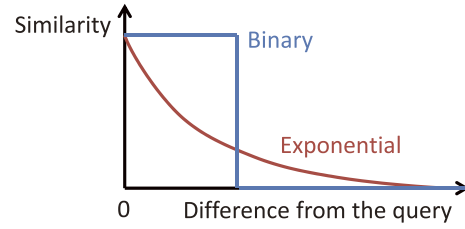


Fig. 1 Difference of prioritizing strategies: a binary type and an exponential type

components retained in the PCA model, respectively. All variables are mean-centered and appropriately scaled. The score is a projection of X onto the subspace spanned by principal components. The score matrix $T_R \in \mathcal{R}^{N \times R}$ is given by

$$T_R = X V_R \quad (18)$$

The data matrix X can be reconstructed from T_R with linear transformation V_R .

$$\hat{X} = T_R V_R^T = X V_R V_R^T \quad (19)$$

The Q statistic is defined as

$$Q = \sum_{m=1}^M (x_m - \hat{x}_m)^2 \quad (20)$$

where x_m is the m -th variable of \mathbf{x} and \hat{x}_m is its estimate. The Q statistic is the distance between the sample and the subspace spanned by principal components. In other words, Q is the correlation-based dissimilarity between the sample and the modeling data. In addition, to evaluate the distance-based dissimilarity between the sample and the modeling data, Hotelling's T^2 statistic can be used.

$$T^2 = \sum_{r=1}^R \frac{t_r^2}{\sigma_{t_r}^2} \quad (21)$$

where σ_{t_r} denotes the standard deviation of the r -th score t_r . The T^2 statistic expresses the normalized distance from the origin in the subspace spanned by principal components. The Q and T^2 statistics can be integrated into a single index (Raich and Cinar, 1996).

$$J = \lambda T^2 + (1 - \lambda) Q \quad (22)$$

where λ is a tuning parameter. In CoJIT, the index J is calculated for each dataset, and the dataset that minimizes J is selected and used for local modeling.

Figure 2 shows the difference of sample selection for local modeling between distance-based JIT modeling and CoJIT modeling. The samples are classified into two groups that have different correlations. Conventional distance-based JIT modeling selects samples regardless of the difference of correlation, while CoJIT modeling can select samples whose correlation is best fit for the query.

The procedure of CoJIT is as follows: 1) several datasets are generated from data stored in a database; 2) the index J is calculated from a query and each dataset; 3) the dataset whose J is the smallest is selected; and 4) a local model is

constructed from the selected dataset.

Not only can CoJIT cope with abrupt changes of process characteristics and also achieve high estimation performance, but can also cope with process nonlinearity. Fujiwara *et al.* (2009) reported an application of CoJIT to estimating the aroma concentration of the cracked gasoline (CGL) fractionator of the ethylene production process at the Showa Denko K.K. (SDK) Oita plant in Japan. Aroma denotes the generic name for benzene, toluene, xylene, styrene, etc. The CGL fractionator is controlled by multivariable MPC with an optimizer, and the aroma concentration in CGL is used as one of the constraints in the optimizer. Since the aroma concentration is analyzed in a laboratory usually once a day because of its long analysis time, the process must be operated in a condition that has a wide margin and is far from constraints for safety. Therefore, a soft-sensor that can estimate the aroma concentration accurately in real time needs to be developed for realizing efficient operation.

First, the aroma concentration was estimated with recursive PLS. The model was updated every 24h when the aroma concentration was analyzed in the laboratory, but a bias between the measurements and the estimates was unavoidable after the operating condition was changed. Then, the aroma concentration was estimated with CoJIT. In the initial state, the operation data obtained from about 10 months of operation were stored in the database. Then, the soft-sensor was updated and the aroma concentration was estimated by using CoJIT. The estimation of CoJIT was successful; RMSE was improved by about 28% in comparison with recursive PLS.

In the original CoJIT, candidate datasets for local modeling were constructed by using a moving time-window. Hence, it was applicable only to time-series data and was not necessarily able to generate a dataset consisting of samples that represent characteristics of a query. To solve this problem, Fujiwara *et al.* (2012a) proposed an unsupervised pattern recognition method, referred to as the nearest correlation (NC) method, and integrated it with CoJIT. The NC method can detect samples that are similar to the query from the viewpoint of the correlation without a teacher signal. The NC-CoJIT modeling can cope with not only changes in process characteristics but also the individual difference of production devices and improve the estimation performance since it can select samples for local modeling by appropriately accounting for the correlation among variables.

Liu *et al.* (2012) proposed another JIT modeling called JIT-ENS by integrating robust nearest correlation (RNC) and multi-model ensemble learning, which is an extension of NC-CoJIT. They also proposed an interval soft-sensor by integrating the JIT modeling with the inductive confidence predictor (ICP). The interval soft-sensor generates not only estimates but also their confidence intervals. This approach was applied to a wastewater treatment process. The problem of JIT-ENS and other multi-model ensemble learning methods is their heavy computational load.

CoJIT also has been used for process monitoring and

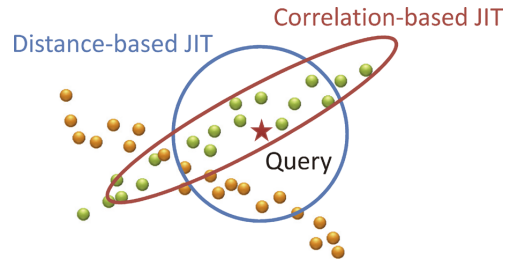


Fig. 2 Difference between distance-based JIT modeling and CoJIT modeling in sample selection

control. Ino *et al.* (2011) proposed MSPC based on CoJIT (CoJIT-MSPC) and applied it to lithography equipment. The application results have demonstrated that CoJIT-MSPC can detect defects with less false alarms and without model maintenance for half a year. Yang *et al.* (2012) integrated CoJIT with the adaptive decentralized PID controller design for multivariable systems.

CoJIT would have a potential for realizing efficient maintenance of soft-sensors in the real world. However, to popularize the use of CoJIT, it is necessary to improve its stability and simplify its parameter tuning as suggested by Zhu *et al.* (2011).

4.2 Locally weighted PLS

Locally weighted PLS (LW-PLS) is one type of LWR, in which PLS is used for local modeling. LW-PLS is preferable to conventional LWR since robust local models can be constructed even when neighbors are less than input variables. Here, LW-PLS is explained and demonstrated through its industrial applications, which clarify that an appropriate definition of similarity between the query and samples is crucial to improve the estimation performance.

4.2.1 Basic algorithm The n -th sample ($n = 1, 2, \dots, N$) of input and output variables is denoted by

$$\mathbf{x}_n = [\mathbf{x}_{n1}, \mathbf{x}_{n2}, \dots, \mathbf{x}_{nM}]^T \quad (23)$$

$$\mathbf{y}_n = [\mathbf{y}_{n1}, \mathbf{y}_{n2}, \dots, \mathbf{y}_{nL}]^T \quad (24)$$

where M and L are the numbers of input and output variables, respectively, and $\mathbf{X} \in \mathcal{R}^{N \times M}$ and $\mathbf{Y} \in \mathcal{R}^{N \times L}$ are the input and output variable matrices whose n -th rows are \mathbf{x}_n^T and \mathbf{y}_n^T , respectively.

In LW-PLS, \mathbf{X} and \mathbf{Y} are stored in a database. When an output estimation is required for a query \mathbf{x}_q , the similarity ω_n between \mathbf{x}_q and \mathbf{x}_n is calculated, and a local PLS model is constructed by weighting samples with a similarity matrix $\mathbf{\Omega} \in \mathcal{R}^{N \times N}$.

$$\mathbf{\Omega} = \text{diag}(\omega_1, \omega_2, \dots, \omega_N) \quad (25)$$

The output estimate $\hat{\mathbf{y}}_q$ corresponding to the query \mathbf{x}_q is calculated as follows.

1. Determine the number of latent variables R and set $r = 1$.
2. Calculate the similarity matrix $\mathbf{\Omega}$.
3. Calculate \mathbf{X}_r , \mathbf{Y}_r and \mathbf{x}_{qr} :

$$\mathbf{X}_r = \mathbf{X} - \mathbf{1}_N [\bar{x}_1, \bar{x}_2, \dots, \bar{x}_M] \quad (26)$$

$$\mathbf{Y}_r = \mathbf{Y} - \mathbf{1}_N [\bar{y}_1, \bar{y}_2, \dots, \bar{y}_L] \quad (27)$$

$$\mathbf{x}_{q,r} = \mathbf{x}_q - [\bar{x}_1, \bar{x}_2, \dots, \bar{x}_M]^T \quad (28)$$

$$\bar{x}_m = \frac{\sum_{n=1}^N \omega_n x_{nm}}{\sum_{n=1}^N \omega_n} \quad (29)$$

$$\bar{y}_l = \frac{\sum_{n=1}^N \omega_n y_{nl}}{\sum_{n=1}^N \omega_n} \quad (30)$$

where $\mathbf{1}_N \in \mathcal{R}^N$ is a vector of ones.

4. Set $\hat{\mathbf{y}}_q = [\bar{y}_1, \bar{y}_2, \dots, \bar{y}_L]^T$.

5. Derive the r -th latent variable of \mathbf{X}

$$\mathbf{t}_r = \mathbf{X}_r \mathbf{w}_r \quad (31)$$

where \mathbf{w}_r is the eigenvector of $\mathbf{X}_r^T \mathbf{\Omega} \mathbf{Y}_r \mathbf{Y}_r^T \mathbf{\Omega} \mathbf{X}_r$, which corresponds to the maximum eigen value.

6. Derive the r -th loading vector of \mathbf{X} and the regression coefficient vector.

$$\mathbf{p}_r = \frac{\mathbf{X}_r^T \mathbf{\Omega} \mathbf{t}_r}{\mathbf{t}_r^T \mathbf{\Omega} \mathbf{t}_r} \quad (32)$$

$$\mathbf{q}_r = \frac{\mathbf{Y}_r^T \mathbf{\Omega} \mathbf{t}_r}{\mathbf{t}_r^T \mathbf{\Omega} \mathbf{t}_r} \quad (33)$$

7. Derive the r -th latent variable of \mathbf{x}_q .

$$\mathbf{t}_{q,r} = \mathbf{x}_{q,r}^T \mathbf{w}_r \quad (34)$$

8. Update $\hat{\mathbf{y}}_q = \hat{\mathbf{y}}_q + t_{q,r} \mathbf{q}_r$.

9. If $r = R$, end. Otherwise, set

$$\mathbf{X}_{r+1} = \mathbf{X}_r - \mathbf{t}_r \mathbf{p}_r^T \quad (35)$$

$$\mathbf{Y}_{r+1} = \mathbf{Y}_r - \mathbf{t}_r \mathbf{q}_r^T \quad (36)$$

$$\mathbf{x}_{q,r+1} = \mathbf{x}_{q,r} - t_{q,r} \mathbf{p}_r \quad (37)$$

10. Set $r = r + 1$ and go to step 5.

In step 3, the weighted mean of each variable is subtracted from each column of \mathbf{X} , \mathbf{Y} , and \mathbf{x}_q^T to make the query near to the origin of the multidimensional space.

4.2.2 Similarity The estimation performance of LW-PLS depends on the definition of similarity ω_n between the query \mathbf{x}_q and the samples \mathbf{x}_n . Here, the similarity is defined on the basis of the weighted Euclidean distance d_n .

$$\omega_n = \exp\left(-\frac{d_n}{\sigma_d \varphi}\right) \quad (38)$$

$$d_n = \sqrt{(\mathbf{x}_n - \mathbf{x}_q)^T \mathbf{\Theta} (\mathbf{x}_n - \mathbf{x}_q)} \quad (39)$$

$$\mathbf{\Theta} = \text{diag}(\theta_1, \theta_2, \dots, \theta_M) \quad (40)$$

where σ_d is a standard deviation of $\{d_n\}$, φ is a localization parameter, $\mathbf{\Theta} \in \mathcal{R}^{M \times M}$ is a weighting matrix, and θ_m is a weight for the m -th input variable. The exponential-type

similarity in Figure 1 is an example in a case where difference is evaluated by the weighted Euclidean distance. The similarity decreases steeply when φ is small and gradually when φ is large. The estimation performance of LW-PLS is higher than, or at least the same as, that of PLS as long as the localization parameter φ is tuned properly, because LW-PLS becomes equivalent to PLS when $\varphi = \infty$. In other words, LW-PLS includes PLS as a special case.

4.2.3 Various similarity measures and applications The definition of similarity or weight significantly affects the estimation performance of LW-PLS. The simplest one is $\theta_m = 1$, i.e. no weighting, but the performance can be improved by adopting an appropriate similarity measure. Thus, a variety of similarities have been investigated not only for LW-PLS but also for various JIT modeling (Fujiwara *et al.*, 2010; Kim *et al.*, 2011; Shigemori *et al.*, 2011; Nakagawa *et al.*, 2012).

For example, Shigemori *et al.* (2011) proposed to define θ_m as the absolute value of the m -th input variable's regression coefficient of a global MRA model. A quality design system for steel products was developed to determine the optimal manufacturing conditions for securing the target quality by using the proposed LWR model. The developed system has been applied practically at JFE Steel's West Japan Works for more than eight years and has contributed to a reduction in quality deviations and production cost. In addition, the developed system can support designers in searching effectively for the manufacturing condition at which manufacturing cost is minimized while satisfying quality requirements. As a result, design time, which required 30–60 min with the conventional method, can be significantly shortened to 10–15 min.

To further improve the estimation accuracy, Kim *et al.* (2011) proposed to define θ_m as the absolute value of the m -th input variable's regression coefficient of an LW-PLS model constructed by defining $\theta_m = 1$. This weighting strategy can determine the weight θ_m adaptively according to the query; therefore, it can outperform the strategy based on a global MRA model. The proposed LW-PLS was used at Daiichi Sankyo Co., Ltd. to estimate active pharmaceutical ingredient (API) content in granules for tableting from spectral data of near infrared spectroscopy (NIRS). The application results have demonstrated the potential of the proposed LW-PLS as a process analytical technology (PAT) tool in the pharmaceutical industry.

Another application of LW-PLS to a pharmaceutical process was reported by Nakagawa *et al.* (2012). To develop a rapid measurement method of the drug substances remaining in manufacturing equipment after cleaning, applying LW-PLS to spectral data of infrared-reflection absorption spectroscopy (IR-RAS) spectra was investigated. The results confirmed that the residual drug substances and the excipient were detected simultaneously and quickly with IR-RAS. Also, the estimation accuracy of residual drug substances with IR-RAS was improved by using LW-PLS with a new weighting technique, in which second differential spectra of IR attenuated total reflection (IR-ATR) were used as

weights. This result suggests that the estimation performance of LW-PLS can be improved by defining the weight or the similarity on the basis of available knowledge.

The above-mentioned methods succeeded in improving estimation accuracy in particular applications, but the derived regression coefficients of LW-PLS models do not necessarily represent the strength of nonlinearity, which might change depending on the values of input variables. In such a case, the estimation performance might become worse. In general, θ should be small when an input variable has linear effect on an output variable, because large θ can lead to prioritizing local samples excessively and to cause overfitting. On the contrary, θ should be large to cope with nonlinear effect of the input variable on the output variable. Hence, the weight might need to be determined adaptively. Okajima *et al.* (2011) proposed to determine θ by using the variance of regression coefficients of multiple LW-PLS models and showed its usefulness through numerical examples.

It has been confirmed that appropriate definition of similarity between the query and samples is crucial in improving the estimation performance of JIT modeling. Although several weighting techniques have been proposed and have succeeded in particular applications, a systematic approach to defining the weight and the similarity needs to be developed.

Conclusion

The authors believe it is crucial for engineers to use a modeling method suitable for the target. The method and the derived model should be as simple as possible in order to achieve the desired performance including robustness and maintain the model by spending minimum time and cost. Engineers do not have to adopt complex techniques unless necessary. In fact, the recent questionnaire survey has revealed that linear regression such as PLS is dominant and model maintenance is the most critical issue in industrial application of virtual sensing technology. Accordingly, this review has focused on PLS-based variable selection and JIT modeling.

It is beneficial for us to recognize that manufacturing processes in different industries are quite different in appearance but they have very similar problems from the viewpoint of quality issue.

In the petrochemical industry, for example, product quality is not usually measured in real time due to the high investment and maintenance cost of online analyzers. To meet the customers' specifications, inferential control has been widely used. The key idea is to develop a soft-sensor that can estimate product quality from measured process variables and to use the estimates for feedback control. However, the estimation performance of soft-sensors deteriorates due to changes in process characteristics, operating conditions, and feed properties. Hence, model maintenance is a critical issue.

In the steel industry, it is crucial to minimize defects in steel products and improve product yield. However, the

problem is not easy. Various products are manufactured through various pieces of equipment such as blast furnace, steel converter, continuous-casting machine, and hot and cold rolling mills. Engineers have great difficulty in finding the key operating conditions to reduce defects. A key issue is again maintenance of models that are used for estimating product quality and optimizing operating conditions.

The situation is the same in the semiconductor industry, where products are manufactured through hundreds of processes. Virtual metrology (VM) plays an important role in monitoring and controlling product quality, but its maintenance is a critical problem.

In the pharmaceutical industry, a new paradigm of quality by design (QbD) for pharmaceutical development and manufacturing has emerged. In the traditional quality by testing (QbT) paradigm, product quality is assured by testing and rejecting lots that fail to meet the specification, which makes this approach wasteful and inefficient. In the QbD approach, on the other hand, quality should be built into a product with understanding of the product and process. QbD is defined as a systematic approach to development that begins with predefined objectives and emphasizes product and process understanding and process control, based on sound science and quality risk management (International Conference on Harmonization (ICH), 2005). Process analytical technology (PAT) is an important tool to enhance process understanding and to realize QbD. PAT is a system for designing, analyzing, and controlling manufacturing through timely measurements (i.e. during processing) of critical quality and performance attributes of raw and in-process materials and processes with the goal of ensuring final product quality (Food and Drug Administration (FDA), 2004). The goal of QbD and PAT is process understanding to build a model that improves process control and leads to continuous improvement and real-time release (RTR) of final products without final quality testing. RTR testing is defined as the ability to evaluate and ensure the quality of in-process and/or final product based on process data, which typically include a valid combination of measured material attributes and process controls (International Conference on Harmonization (ICH), 2005). Hence, modeling and model maintenance is a key to success.

To adapt a model to new process characteristics and operating conditions, modeling data need to be updated. In JIT modeling, a local model is built on demand each time estimation is required. This approach works only when a database is appropriately maintained. In recursive modeling, samples included in a moving-window need to be appropriately maintained. Thus, far from avoiding data preprocessing, these approaches require online efficient data preprocessing that can cope with outliers, missing values, and measurement noise. Although data preprocessing issues are not discussed in this review, further research is imperative to realize high-performance maintenance-free virtual sensing technology. In addition to the research and the development of new modeling techniques, further efforts to solve practical problems are encouraged. Virtual sensing technology

such as soft-sensors has become more important than ever in various industries. Now we are given a great opportunity to play an active part.

Literature Cited

- Ahmed, F., S. Nazir and Y. Yeo; "A New Soft Sensor Based on Recursive Partial Least Squares for Online Melt Index Predictions in Grade-Changing HDPE Operations," *Chem. Prod. Process Model.*, **4**, art. no. 33 (2009)
- Amirthalingam, R. and J. Lee; "Subspace Identification Based Inferential Control Applied to a Continuous Pulp Digester," *J. Process Contr.*, **9**, 397–406 (1999)
- Amirthalingam, R., S. Sung and J. Lee; "Two-Step Procedure for Data-Based Modeling for Inferential Control Applications," *AIChE J.*, **46**, 1974–1988 (2000)
- Atkeson, C., A. Moore and S. Schaaf; "Locally Weighted Learning," *Artif. Intell. Rev.*, **11**, 11–73 (1997)
- Birattari, M., G. Bontempi and H. Bersini; "Lazy Learning Meets the Recursive Least Squares Algorithm," *Adv. Neural Inf. Process. Syst.*, **12**, 375–381 (1999)
- Bontempi, G., M. Birattari and H. Bersini; "Lazy Learning for Local Modelling and Control Design," *Int. J. Control*, **72**, 643–658 (1999)
- Boser, B., I. Guyon and V. Vapnik; "A Training Algorithm for Optimal Margin Classifiers," Annual Workshop on Computational Learning Theory, pp. 144–152, Pittsburgh, USA (1992)
- Bozenhardt, H.; "Modern Control Tricks Solve Distillation Problems," *Hydrocarbon Process.*, **67**, 47–50 (1988)
- Centner, V. and D. L. Massart; "Optimization in Locally Weighted Regression," *Anal. Chem.*, **70**, 4206–4211 (1998)
- Chang, S., E. Baughman and B. McIntosh; "Implementation of Locally Weighted Regression to Maintain Calibrations on FT-NIR Analyzers for Industrial Processes," *Appl. Spectrosc.*, **55**, 1199–1206 (2001)
- Cheng, C. and M. Chiu; "A New Data-Based Methodology for Nonlinear Process Modeling," *Chem. Eng. Sci.*, **59**, 2801–2810 (2004)
- Chong, I. and C. Jun; "Performance of Some Variable Selection Methods When Multicollinearity is Present," *Chemom. Intell. Lab. Syst.*, **78**, 103–112 (2005)
- Cleveland, W.; "Robust Locally Weighted Regression and Smoothing Scatterplots," *J. Am. Stat. Assoc.*, **74**, 829–836 (1979)
- Cleveland, W. and S. Devlin; "Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting," *J. Am. Stat. Assoc.*, **83**, 596–610 (1988)
- Cortes, C. and V. Vapnik; "Support Vector Networks," *Mach. Learn.*, **20**, 273–297 (1995)
- Dayal, B. and J. MacGregor; "Recursive Exponentially Weighted PLS and Its Applications to Adaptive Control and Prediction," *J. Process Contr.*, **7**, 169–179 (1997)
- Desai, K., Y. Badhe, S. Tambe and B. Kulkarni; "Soft-Sensor Development for Fed-Batch Bioreactors Using Support Vector Regression," *Biochem. Eng. J.*, **27**, 225–239 (2006)
- Ding, C. H. Q., X. He, H. Zha, M. Gu and H. D. Simon; "A Min-Max Cut Algorithm for Graph Partitioning and Data Clustering," Proceedings of the 2001 IEEE International Conference on Data Mining (ICDM'01), 107–114, IEEE Computer Society, Washington, DC, USA (2001)
- Efron, B., T. Hastie, I. Johnstone and R. Tibshirani; "Least Angle Regression," *Ann. Stat.*, **32**, 407–499 (2004)
- Food and Drug Administration (FDA); Guidance for Industry: PAT, a Framework for Innovative Pharmaceutical Development, Manufacturing, and Quality Assurance, U.S. Dept. of Health and Human Services, Food and Drug Administration (FDA), MD, USA (2004)
- Fujiwara, K., M. Kano, S. Hasebe and A. Takinami; "Soft-Sensor Development Using Correlation-Based Just-in-Time Modeling," *AIChE J.*, **55**, 1754–1765 (2009)
- Fujiwara, K., M. Kano and S. Hasebe; "Development of Correlation-Based Clustering Method and Its Application to Software Sensing," *Chemom. Intell. Lab. Syst.*, **44**, 130–138 (2010)
- Fujiwara, K., M. Kano and S. Hasebe; "Correlation-Based Spectral Clustering for Flexible Process Monitoring," *J. Process Contr.*, **21**, 1438–1448 (2011)
- Fujiwara, K., M. Kano and S. Hasebe; "Development of Correlation-Based Pattern Recognition Algorithm and Adaptive Soft-Sensor Design," *Control Eng. Pract.*, **20**, 371–378 (2012a)
- Fujiwara, K., H. Sawada and M. Kano; "Input Variable Selection for PLS Modeling Using Nearest Correlation Spectral Clustering," *Chemom. Intell. Lab. Syst.*, **118**, 109–119 (2012b)
- Galicia, H., Q. He and J. Wang; "A Reduced Order Soft Sensor Approach and Its Application to a Continuous Digester," *J. Process Contr.*, **21**, 489–500 (2011)
- Galicia, H., Q. He and J. Wang; "Comparison of the Performance of a Reduced-Order Dynamic PLS Soft Sensor with Different Updating Schemes for Digester Control," *Control Eng. Pract.*, **20**, 747–760 (2012)
- Ge, Z. and Z. Song; "Online Monitoring of Nonlinear Multiple Mode Processes Based on Adaptive Local Model Approach," *Control Eng. Pract.*, **16**, 1427–1437 (2008)
- Gupta, M. R., E. K. Garcia and E. Chin; "Adaptive Local Linear Regression with Application to Printer Color Management," *IEEE Trans. Image Process.*, **17**, 936–945 (2008)
- Helland, K., H. Berntsen, O. Borgen and H. Martens; "Recursive Algorithm for Partial Least Squares Regression," *Chemom. Intell. Lab. Syst.*, **14**, 129–137 (1992)
- Hocking, R.; "The Analysis and Selection of Variables in Linear Regression," *Biometrics*, **32**, 1–49 (1976)
- Ino, T., L. Huang and M. Kano; "Maintenance-Free Multivariate SPC by Using Correlation-Based Just-In-Time Modeling," AEC/APC Symposium Asia, MT-O-023, Tokyo, Japan (2011)
- International Conference on Harmonization (ICH); ICH Harmonised Tripartite Guideline, Pharmaceutical Development Q8, U.S. Dept. of Health and Human Services, Food and Drug Administration, Rockville, MD, USA (2005)
- Jackson, J. and G. Mudholkar; "Control Procedures for Residuals Associated with Principal Component Analysis," *Technometrics*, **21**, 341–349 (1979)
- Jin, L., J. Fernandez Pierna, Q. Xu, F. Wahl, O. de Noord, C. Saby and D. Massart; "Delaunay Calibration Method for Multivariate Calibration," *Anal. Chim. Acta*, **488**, 1–14 (2003)
- Joseph, B. and C. Brosilow; "Inferential Control of Processes," *AIChE J.*, **24**, 485–509 (1978)
- JSPS PSE 143 committee; Annual Report 2003, **28**, Tokyo, Japan (2004)
- Kadlec, P. and B. Gabrys; "Local Learning-Based Adaptive Soft Sensor for Catalyst Activation Prediction," *AIChE J.*, **57**, 1288–1301 (2011)
- Kadlec, P., B. Gabrys and S. Strandt; "Data-Driven Soft Sensors in the Process Industry," *Comput. Chem. Eng.*, **33**, 795–814 (2009)
- Kadlec, P., R. Grbic and B. Gabrys; "Review of Adaptation Mechanisms for Data-Driven Soft Sensors," *Comput. Chem. Eng.*, **35**, 1–24 (2011)
- Kamohara, H., A. Takinami, M. Takeda, M. Kano, S. Hasebe and I. Hashimoto; "Product Quality Estimation and Operating Condition Monitoring for Industrial Ethylene Fractionator," *J. Chem. Eng. Japan*, **37**, 422–428 (2004)
- Kaneko, H., M. Arakawa and K. Funatsu; "Development of a New Soft Sensor Method Using Independent Component Analysis and

- Partial Least Squares," *AIChE J.*, **55**, 87–98 (2009)
- Kano, M. and Y. Nakagawa; "Data-Based Process Monitoring, Process Control, and Quality Improvement: Recent Developments and Applications in Steel Industry," *Comput. Chem. Eng.*, **32**, 12–24 (2008)
- Kano, M. and M. Ogawa; "The State of the Art in Chemical Process Control in Japan: Good Practice and Questionnaire Survey," *J. Process Contr.*, **20**, 969–982 (2010)
- Kano, M., K. Miyazaki, S. Hasebe and I. Hashimoto; "Inferential Control System of Distillation Compositions Using Dynamic Partial Least Squares Regression," *J. Process Contr.*, **10**, 157–166 (2000)
- Kano, M., N. Showchayai, S. Hasebe and I. Hashimoto; "Inferential Control of Distillation Compositions: Selection of Model and Control Configuration," *Control Eng. Pract.*, **11**, 927–933 (2003a)
- Kano, M., S. Tanaka, S. Hasebe, I. Hashimoto and H. Ohno; "Monitoring Independent Components for Fault Detection," *AIChE J.*, **49**, 969–976 (2003b)
- Kano, M., S. Hasebe, I. Hashimoto and H. Ohno; "Evolution of Multivariate Statistical Process Control: Application of Independent Component Analysis and External Analysis," *Comput. Chem. Eng.*, **28**, 1157–1166 (2004)
- Kano, M., S. Lee and S. Hasebe; "Two-Stage Subspace Identification for Softsensor Design and Disturbance Estimation," *J. Process Contr.*, **19**, 179–186 (2009)
- Kibler, D., D. Aha and M. Albert; "Instance-Based Prediction of Real-Valued Attributes," *Comput. Intell.*, **5**, 51–57 (1989)
- Kim, S., M. Kano, H. Nakagawa and S. Hasebe; "Estimation of Active Pharmaceutical Ingredients Content Using Locally Weighted Partial Least Squares and Statistical Wavelength Selection," *Int. J. Pharm.*, **421**, 269–274 (2011)
- Kresta, J., J. MacGregor and T. Marlin; "Multivariate Statistical Monitoring of Process Operating Performance," *Can. J. Chem. Eng.*, **69**, 35–47 (1991)
- Kresta, J., T. Marlin and J. MacGregor; "Development of Inferential Process Models Using PLS," *Comput. Chem. Eng.*, **18**, 597–611 (1994)
- Lee, J., C. Yoo and I. Lee; "Statistical Process Monitoring with Independent Component Analysis," *J. Process Contr.*, **14**, 467–485 (2004)
- Lee, D., J. Song, S. Song and E. Yoon; "Weighted Support Vector Machine for Quality Estimation in the Polymerization Process," *Ind. Eng. Chem. Res.*, **44**, 2101–2105 (2005)
- Leung, H., Y. Huang and C. Cao; "Locally Weighted Regression for Desulphurisation Intelligent Decision System Modeling," *Simul. Model. Pract. Theory*, **12**, 413–423 (2004)
- Liu, J.; "On-Line Soft Sensor for Polyethylene Process with Multiple Production Grades," *Control Eng. Pract.*, **15**, 769–778 (2007)
- Liu, J., D. Chen and J. Shen; "Development of Self-Validating Soft Sensors Using Fast Moving Window Partial Least Squares," *Ind. Eng. Chem. Res.*, **49**, 11530–11546 (2010)
- Liu, Y., D. Huang and Y. Li; "Development of Interval Soft Sensors Using Enhanced Just-in-Time Learning and Inductive Confidence Predictor," *Ind. Eng. Chem. Res.*, **51**, 3356–3367 (2012)
- Maron, D. and A. Moore; "The Racing Algorithm: Model Selection for Lazy Learners," *Artif. Intell. Rev.*, **11**, 193–225 (1997)
- Mejdell, T. and S. Skogestad; "Estimation of Distillation Compositions from Multiple Temperature Measurements Using Partial-Least-Squares Regression," *Ind. Eng. Chem. Res.*, **30**, 2543–2555 (1991)
- Mejdell, T. and S. Skogestad; "Output Estimation Using Multiple Secondary Measurements: High-Purity Distillation," *AIChE J.*, **39**, 1641–1653 (1993)
- Moore, C., J. Hackney and D. Canter; "Selecting Sensor Location and Type for Multivariable Processes," *Shell Process Control Workshop*, 291–308, Butterworth, Boston, USA (1987)
- Morari, M. and G. Stephanopoulos; "Optimal Selection of Secondary Measurements within the Framework of State Estimation in the Presence of Persistent Unknown Disturbances," *AIChE J.*, **26**, 247–259 (1980)
- Mu, S., Y. Zeng, R. Liu, P. Wu, H. Su and J. Chu; "Online Dual Updating with Recursive PLS Model and Its Application in Predicting Crystal Size of Purified Terephthalic Acid (PTA) Process," *J. Process Contr.*, **16**, 557–566 (2006)
- Murtaugh, P. A.; "Performance of Several Variable-Selection Methods Applied to Real Ecological Data," *Ecol. Lett.*, **12**, 1061–1068 (2009)
- Naes, T., T. Isaksson and B. Kowalski; "Locally Weighted Regression and Scatter Correction for Near-Infrared Reflectance Data," *Anal. Chem.*, **62**, 664–673 (1990)
- Nakagawa, H., T. Tajima, M. Kano, S. Kim, S. Hasebe, T. Suzuki and H. Nakagami; "Evaluation of Infrared-Reflection Absorption Spectroscopy Measurement and Locally Weighted Partial Least-Squares for Rapid Analysis of Residual Drug Substances in Cleaning Processes," *Anal. Chem.*, **84**, 3820–3826 (2012)
- Ng, A., M. Jordan and Y. Weiss; "On Spectral Clustering: Analysis and an Algorithm," *Neural Inf. Process. Syst.*, Vancouver, Canada (2001)
- Ni, W., S. Tan, W. Ng and S. Brown; "Localized, Adaptive Recursive Partial Least Squares Regression for Dynamic System Modeling," *Ind. Eng. Chem. Res.*, **51**, 8025–8039 (2012)
- Norgaard, L., A. Saudland, J. Wagner, J. Nielsen, L. Munck and S. Engelsen; "Interval Partial Least-Squares Regression (iPLS): A Comparative Chemometric Study with an Example from Near-Infrared Spectroscopy," *Appl. Spectrosc.*, **54**, 413–419 (2000)
- Okajima, R., S. Kim, M. Kano and S. Hasebe; "Selection of Similarity Measure for Locally Weighted Partial Least Squares Regression," *AIChE Annual Meeting*, 669e, Minneapolis, USA (2011)
- Ookita, K.; "Operation and Quality Control for Chemical Plants by Soft-sensors," *CICSJ Bulletin*, **24**, 31–33 (2006)
- Overschee, P. and B. Moor; "N4SID: Subspace Algorithms for the Identification of Combined Deterministic-Stochastic Systems," *Automatica*, **30**, 75–93 (1994)
- Pan, T., B. Sheng, D. Wong and S. Jang; "A Virtual Metrology Model Based on Recursive Canonical Variate Analysis with Applications to Sputtering Process," *J. Process Contr.*, **21**, 830–839 (2011)
- Qin, S.; "Recursive PLS Algorithms for Adaptive Data Modeling," *Comput. Chem. Eng.*, **22**, 503–514 (1998)
- Qin, S., H. Yue and R. Dunia; "Self-Validating Inferential Sensors with Application to Air Emission Monitoring," *Ind. Eng. Chem. Res.*, **36**, 1675–1685 (1997)
- Rademaker, O., J. Rijnsdorp and A. Maarleveld; *Dynamics and Control of Continuous Distillation Units*, Elsevier, Amsterdam, the Netherlands (1975)
- Raich, A. and A. Cinar; "Statistical Process Monitoring and Disturbance Diagnosis in Multivariable Continuous Processes," *AIChE J.*, **42**, 995–1009 (1996)
- Rajalahti, T., R. Arneberg, F. Bervend, K. Myhra, R. Ulvik and O. Kvalheim; "Biomarker Discovery in Mass Spectral Profiles by Means of Selectivity Ratio Plot," *Chemom. Intell. Lab. Syst.*, **95**, 35–48 (2009)
- Ricker, N.; "Use of Biased Least-Squares Estimators for Parameters in Discrete-Time Pulse Response Models," *Ind. Eng. Chem. Res.*, **27**, 343–350 (1998)
- Savitzky, A. and M. Golay; "Smoothing and Differentiation of Data by Simplified Least Squares Procedures," *Anal. Chem.*, **36**, 1627–1639 (1964)
- Sbarbaro, D., P. Ascencio, P. Espinoza, F. Mujica and G. Cortes; "Adaptive Soft-Sensors for On-Line Particle Size Estimation in Wet Grinding Circuits," *Control Eng. Pract.*, **16**, 171–178 (2008)

- Shigemori, H., M. Kano and S. Hasebe; "Optimum Quality Design System for Steel Products through Locally Weighted Regression Model," *J. Process Contr.*, **21**, 293–301 (2011)
- Showchaiya, N., M. Kano, S. Hasebe and I. Hashimoto; "Improvement of Distillation Composition Control by Using Predictive Inferential Control Technique," *J. Chem. Eng. Japan*, **34**, 1026–1032 (2001)
- Su, Q. L., M. Kano and M. S. Chiu; "A New Strategy of Locality Enhancement for Just-in-Time Learning Method," Proceedings of the 11th International Symposium on Process Systems Engineering (PSE), 1662–1666, Singapore (2012)
- Tibshirani, R.; "Regression Shrinkage and Selection via the Lasso," *J. R. Stat. Soc. Series B Stat. Methodol.*, **58**, 267–288 (1996)
- Verhaegen, M. and P. Dewilde; "Subspace Model Identification. I—The Output-Error State-Space Model Identification Class of Algorithms," *Int. J. Control*, **56**, 1187–1210 (1992)
- Walden, A. and P. Prescott; "Identification of Trends in Annual Maximum Sea Levels Using Robust Locally Weighted Regression," *Estuar. Coast. Shelf Sci.*, **16**, 17–26 (1983)
- Wang, Z., T. Isaksson and B. Kowalski; "New Approach for Distance Measurement in Locally Weighted Regression," *Anal. Chem.*, **66**, 249–260 (1994)
- Ward, E.; "A Review and Comparison of Four Commonly Used Bayesian and Maximum Likelihood Model Selection Tools," *Ecol. Modell.*, **211**, 1–10 (2008)
- Weber, R. and C. Brosilow; "The Use of Secondary Measurements to Improve Control," *AIChE J.*, **18**, 614–623 (1972)
- Whitehead, D. and M. Parnis; "Computer Control Improves Ethylene Plant Operation," *Hydrocarbon Process.*, **66**, 105–108 (1987)
- Wold, S., M. Sjostrom and L. Eriksson; "PLS-Regression: a Basic Tool of Chemometrics," *Chemom. Intell. Lab. Syst.*, **58**, 109–130 (2001)
- Yan, W., H. Shao and X. Wang; "Soft Sensing Modeling Based on Support Vector Machine and Bayesian Model Selection," *Comput. Chem. Eng.*, **28**, 1489–1498 (2004)
- Yang, X., L. Jia and M. S. Chiu; "Adaptive Decentralized PID Controllers Design Based on the Lyapunov Approach," *J. Process Contr.*, **22**, 1531–1542 (2012)
- Yu, C. and W. Luyben; "Use of Multiple Temperatures for the Control of Multicomponent Distillation Columns," *Ind. Eng. Chem. Res.*, **23**, 590–597 (1984)
- Yuan, M. and Y. Lin; "Model Selection and Estimation in Regression with Grouped Variables," *J. R. Stat. Soc. Series B Stat. Methodol.*, **68**, 49–67 (2006)
- Zheng, Q. and H. Kimura; "Just-in-Time Modeling for Function Prediction and Its Applications," *Asian J. Control*, **3**, 35–44 (2001)
- Zhu, Z., F. Corona, A. Lendasse, R. Baratti and J. Romagnoli; "Local Linear Regression for Soft-Sensor Design with Application to an Industrial Deethanizer," The 18th IFAC World Congress, 2839–2844, Milano, Italy (2011)