

Title	Data-based process monitoring, process control, and quality improvement: Recent developments and applications in steel industry
Author(s)	Kano, Manabu; Nakagawa, Yoshiaki
Citation	Computers & Chemical Engineering (2008), 32(1-2): 12-24
Issue Date	2008-01
URL	<a href="http://hdl.handle.net/2433/85262">http://hdl.handle.net/2433/85262</a>
Right	Copyright © 2007 Elsevier Ltd
Type	Journal Article
Textversion	author

# Data-Based Process Monitoring, Process Control, and Quality Improvement: Recent Developments and Applications in Steel Industry

Manabu Kano <sup>a,\*</sup> Yoshiaki Nakagawa <sup>b</sup>

<sup>a</sup>*Department of Chemical Engineering, Kyoto University, Nishikyo-ku, Kyoto 615-8510, Japan*

<sup>b</sup>*Sumitomo Metals (Kokura), Ltd., Kokurakita-ku, Kitakyushu 802-8686, Japan*

---

## Abstract

The issue of how to improve product quality and product yield in a brief period of time becomes more critical in many industries. Even though industrial processes are totally different in appearance, the problems to solve are highly similar: how to build a reliable model from a limited data, how to analyze the model and relate it to first principles, how to optimize operating condition, and how to realize an on-line monitoring and control system and maintain it. In this paper, statistical process monitoring and control methodologies are briefly surveyed, and our application results in steel facilities are presented. The achievements of the present work are as follows: 1) the development of a new method that can cope with qualitative quality information and relate operating conditions to product quality or product yield, 2) the simultaneous analysis of multiple processing units including a converter, a continuous caster, a blooming process, and rolling processes, and 3) the successful

application results in the steel industry.

*Key words:* Soft-sensor, Multivariate Statistical Process Control, Multivariate Analysis, Iron and Steel Process, Quality Improvement, Quantification

---

## 1 Introduction

How can we improve product quality and yield? More than ever, the answer to this question is vital as product life cycles are getting shorter and international competition is getting keener. Since this question arises repeatedly when a new product is developed, quality improvement should be achieved faster and in a more systematic way.

Although the most reliable approach to quality improvement will be the use of precise first-principle models, such models are not available in most newly developed processes and modeling of a complex industrial process is very difficult and time-consuming. In particular, it is difficult to build precise first-principle models that can explain why defects appear in products. In the steel industry, for example, the relationship of operating conditions to product defects such as surface flaws and internal defects is not clear. The product qualities have been usually maintained by skilled operators on the basis of their experience and intuition. Although much effort has been devoted to clarify the relationship between operating conditions and product quality, the problem remains unsolved. No first-principle model is available for product quality improvement.

---

\* Corresponding author.

*Email address:* manabu@cheme.kyoto-u.ac.jp (Manabu Kano).

Another option to solve this difficult situation is the use of operation data. In the last decade or so, data-based approaches have been widely accepted for process control and monitoring in various industries. In the steel industry, for example, many companies have built integrated data-bases to store many operation data from all factories and product properties data from inspection. Thus, data-based approaches are expected to become solutions.

To achieve product quality improvement, we need to develop a system having at least the following functions: 1) to predict product quality from operating conditions, 2) to derive better operating conditions that can improve the product quality, and 3) to detect faults or malfunctions for preventing undesirable operation. The first function is realized by developing a soft-sensor, which is a mathematical model to relate operating conditions to product quality. On the basis of the model, the second function is realized by formulating and solving an optimization problem. The third function is realized by multivariate statistical process control (MSPC).

In this paper, selected application results of statistical methods to iron and steel processes are presented. The objective is to improve product quality and achieve higher product yield – the percentage of good products to all products – of various products. Although many researches have been reported in this area, the achievements of the present work are 1) the development of a new method that can cope with qualitative quality information and relate operating conditions to product yield and 2) the simultaneous analysis of multiple processing units including a converter, a continuous caster, a blooming process, and rolling processes. The proposed method is referred to as Data-Driven Quality Improvement (DDQI), which was originally reported by Kano et al. (2004b). Since 2004, DDQI has been applied to many processes including iron

making, steel making, blooming, hot rolling, cold rolling, and continuous galvanic lines for improving product quality and realizing stable operation. This paper aims to formulate DDQI and demonstrate its usefulness via industrial applications.

This paper is organized in the following manner. First, data-based process monitoring and control techniques are briefly surveyed in section 2. Section 3 aims to describe past results and recent developments of DDQI, including a new quantification method of product quality. Then, application results of DDQI to iron and steel processes are described in section 4. Finally, the summary of this paper is given in conclusions.

## **2 Statistical Methods for Process Monitoring and Control: Trends and Problems**

To achieve product quality improvement, we need to predict product quality from operating conditions, to derive better operating conditions that can improve the product quality, and to detect faults or malfunctions for preventing undesirable operation. These requirements are common among various industries, although processes are totally different in appearance. The authors have focused on statistical methods, especially soft-sensors and MSPC. They are expected to be key tools for realizing significant improvement of steel product quality under the situation where a first-principle model is not available.

In this section, data-based process monitoring and control techniques are briefly surveyed. The focal points are recent trends and the remaining problems in the fields of soft-sensor design and MSPC. Both techniques are highly

Table 1

Problems with hardware sensors. The results of a questionnaire to 26 companies in Japan.

percentage	recognized problem
27%	time-consuming maintenance
21%	need for calibration
15%	aged deterioration
13%	insufficient accuracy
10%	long dead-time, slow dynamics
8%	large noise
2%	low reproducibility
4%	others

relevant to our industrial applications.

### 2.1 *Soft-sensor Design*

When hardware sensors are not available, soft-sensors are key technologies for producing high quality products. 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 results of a questionnaire to 26 companies in Japan (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

Table 2

Purpose in using soft-sensors in Mitsubishi Chemical Corporation.

percentage	purpose
37%	stabilize quality
28%	reduce feed and utility
18%	improve reliability with online analyzer
9%	stabilize operation
6%	reduce manual analysis
1%	avoid installing special analyzer
1%	analyze key factors

Table 3

Processes which soft-sensors are applied to in Mitsubishi Chemical Corporation.

percentage	process
37%	distillation
28%	reaction
18%	evaporation
9%	others

of the investigation results are shown in Tables 2 and 3. The major purposes of using soft-sensors are to 1) stabilize product quality via its online estimation, 2) reduce energy and material consumption via effective operation close to specifications/constraints, and 3) validate online analyzers by comparison with the soft-sensors. Most soft-sensor applications were found in distillation

processes. In addition, it was reported that most soft-sensors were developed through partial least squares (PLS).

Chemometric techniques such as PLS have been widely applied for process modeling, monitoring, and control. The main advantage of those methods is that they can generate latent variables, which are uncorrelated with each other, and cope with correlated input variables. This characteristic is suitable for analyzing data from industrial processes, because industrial processes are multivariable systems and a great number of variables are mutually correlated.

To develop soft-sensors for distillation processes, composition estimators using PLS have been widely used in the literature (Mejdell and Skogestad, 1991a,b; Kresta et al., 1994). In the work of these researchers, steady-state inferential models of product compositions were built. Mejdell and Skogestad (1993) compared three different estimators using a linear model of a binary distillation column. They concluded that good control performance could be achieved with the steady-state PCR (Principal Component Regression) estimator, which was almost as good as the dynamic Kalman filter, because the steady-state estimator has an inherent feedforward effect. The inherent feedforward effect was investigated in more detail by Kano et al. (2003a) and Showchaiya et al. (2001). 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 from on-line measured process variables. 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, the performance of soft-sensors can be greatly



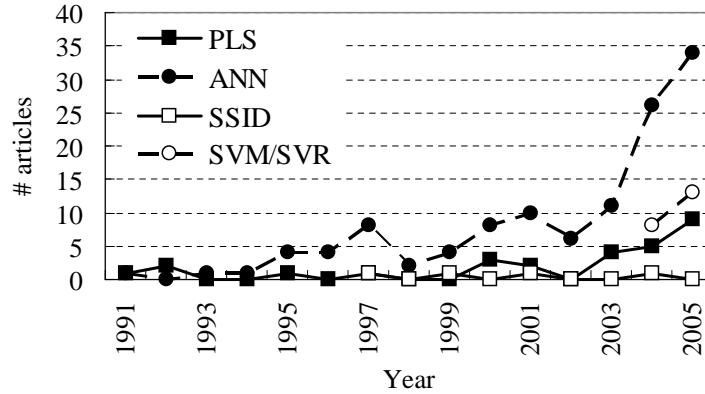


Fig. 1. The number of research articles related to soft-sensor.

improved by taking process dynamics into account.

Soft-sensors are usually used in inferential control systems. Any control algorithm is applicable to inferential control because estimates are used instead of measurements. However, statistical feedback controllers based on multivariate analysis have been proposed. Piovoso and Kosanovich (1994) developed a feedback controller based on a static PCA/PCR model and applied it to a binary distillation column. Chen et al. (1998) designed a controller in the model predictive control (MPC) framework, and it was used to control the equivalent score space representation of the process. The score predictive model for the MPC algorithm was built using PLS.

Figure 1, which is the keyword search results using Scopus, shows the research trend in the field of soft-sensor design. Artificial neural network (ANN) has been dominant in the literature since the middle 90's, while PLS is popular in industry as shown by Ookita (2006). ANN is a useful tool for building non-linear models and 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 models function well. In the last few years, support vector machine (SVM) and support vector regression (SVR) have been emerging. These methods have attracted researchers' and engineers' attention. Another method for developing soft sensors is subspace identification (SSID), which can build a state space model from input-output data (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 greatly improved by taking process dynamics into account. Amirthalingam and Lee (1999) used SSID for inferential control of a continuous pulp digester. Lee et al. (2006) proposed two-stage SSID to develop highly accurate softsensors that can take into account the influence of unmeasured disturbances on estimated key variables, and they demonstrated the superiority of the proposed method over conventional methods through their application to an industrial ethylene fractionator.

A great deal of research has been conducted to develop data-based soft-sensors for various processes. A data-based soft-sensor, however, does not always function well, because a black-box model is not valid when a process is operated outside certain conditions where operation data used for modeling were obtained. 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. On-line monitoring of the validity of the soft-sensor 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 enables us to detect the inconsistency between the analyzer and the soft-sensor, but the cause of the inconsistency cannot be identified. In industry practice, it is assumed that an estimation error is caused by inaccurate esti-

mation; however, this assumption is not always true because analyzers are not always reliable. For example, when blockage occurs within a sampling line, a hardware sensor cannot provide accurate measurements. To address such practical problems, Kamohara et al. (2004) proposed a PLS-based framework for developing a soft-sensor and monitoring its validity on-line. The on-line monitoring system was based on the dynamic PLS model designed for estimating the product quality. In addition, simple rules were established for checking the performance of a process gas chromatograph by combining the soft-sensor and the statistical monitoring system. The effectiveness of the developed system was demonstrated through its application to an ethylene production plant.

## *2.2 Statistical Process Monitoring*

For the successful operation of any process, it is important to detect process upsets, equipment malfunctions, or other special events as early as possible and then to find and remove the factors causing those events. In industrial processes, data-based process monitoring methods have been widely used (Montgomery, 1997). Conventional SPC charts such as Shewhart control charts, cumulative sum (CUSUM) control charts, and exponentially weighted moving average (EWMA) control charts are well established for monitoring univariate processes. Industrial processes are, however, multivariable systems consisting of a large number of mutually correlated variables. Unfortunately, univariate SPC charts do not function well for such multivariable processes because they cannot cope with correlation among variables. To monitor such multivariable processes, MSPC has been developed.

The original Shewhart-type control chart for correlated variables is the Hotelling

$T^2$  control chart. Jackson (1959) used principal component analysis (PCA) and proposed a  $T^2$  control chart for principal components. Later, Jackson and Mudholkar (1979) investigated PCA as a tool of MSPC and introduced a residual analysis. The control chart was introduced for the sum of squared errors ( $SPE$ ) as well as  $T^2$  of principal components retained in a PCA model.

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

$$SPE = \sum_{p=1}^P (x_p - \hat{x}_p)^2 \quad (2)$$

where  $t_r$  is the  $r$ th principal component score and  $\sigma_{t_r}^2$  is the variance of  $t_r$ .  $x_p$  and  $\hat{x}_p$  are a measurement of the  $p$ th variable and its predicted (reconstructed) value, respectively.  $R$  and  $P$  denote the number of principal components retained in the PCA model and the number of process variables, respectively. The  $T^2$  statistic is a measure of the variation within the PCA model, and the  $SPE$  statistic is a measure of the amount of variation not captured by the PCA model. Kresta et al. (1991) demonstrated the usefulness of MSPC with applications to simulated data from a fluidized bed reactor and an extractive distillation column. In the last decade or so, various extensions of MSPC have been proposed. For example, multiway PCA/PLS for monitoring batch processes (Wold et al., 1987), multiblock PCA/PLS for monitoring large processes (MacGregor et al., 1994), dynamic PCA for including process dynamics in a PCA model (Ku et al., 1995), multiscale PCA based on wavelet analysis for monitoring signals at several different frequency ranges (Bakshi, 1998), model-based PCA for integrating a first-principle model with PCA (Rotem et al., 2000), moving PCA that monitors changes in directions of principal components (Kano et al., 2001), a method called DISSIM that is based on the dissimilarity of process data sets (Kano et al., 2002), a method based on

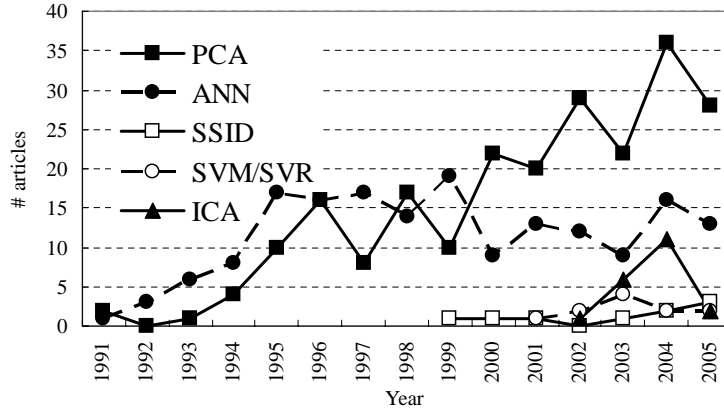


Fig. 2. The number of research articles related to multivariate statistical process control.

constrained PCA for incorporating external information into a PCA model (Yoon and MacGregor, 2001), and a method based on independent component analysis (ICA) to further improve the fault detection performance (Kano et al., 2003b).

These are only a small part of the research efforts dedicated to this field. Figure 2, which is the keyword search results using Scopus, shows the research trend in the field of MSPC. The research trend of MSPC is similar to that of soft-sensors. PCA and ANN are dominant in the literature, and in the last few years, ICA, SVM, and SVR have been emerging. These methods have attracted researchers' and engineers' attention. Another method for statistical process monitoring is SSID. The total number of articles is steadily increasing.

When an out-of-control signal is detected by  $T^2$  or  $SPE$ , it is necessary to identify the process variables that cause the out-of-control signal. This information helps operators to further diagnose the actual cause of a fault. For this purpose, contribution plots can be used. A contribution of the  $p$ th variable to the  $SPE$  statistic is defined as

$$C_p^{[SPE]} = (x_p - \hat{x}_p)^2 \quad (3)$$

On the other hand, a contribution of each variable to the  $T^2$  statistic cannot be easily defined. If only one principal component is related to the detected fault, a contribution of the  $p$ th variable to the important score can be used. However, when several scores are related to the fault, this contribution is not sufficient. Nomikos (1996) derived the contribution of the  $p$ th variable to the  $T^2$  statistic:

$$C_p^{[T^2]} = \mathbf{t}^T \mathbf{S}_T^{-1} x_p \mathbf{p}_p \quad (4)$$

where  $\mathbf{t}$  and  $\mathbf{p}_p$  denote a score vector and the transposed  $p$ th row vector of the loading matrix  $\mathbf{P}$ , respectively.  $\mathbf{S}_T$  is the covariance matrix of scores. This contribution can be positive or negative, although the  $T^2$  statistic is always positive. In practice, however, only a few variables whose contributions are positive large are important (Westerhuis et al., 2000).

Another approach to fault identification is the use of process data obtained from several past faulty operating conditions (Raich and Cinar, 1996). When such process data are available, a fault can be identified by comparing the similarity between data sets representing faulty operating conditions in the past and the data set representing an operating condition when the fault is detected. For applying MSPC-based fault identification, the following procedure is adopted.

- (1) Apply PCA to data sets representing past faulty situations and determine reference PCA models.
- (2) Calculate  $T^2$  and  $SPE$  of the current data using the reference PCA models when a faulty situation is detected.

- (3) Identify the past faulty situation whose data set gives the smallest  $T^2$  and  $SPE$  of all. Identification with  $T^2$  and  $SPE$  is performed separately.

For fault identification, the data matrix representing the current faulty situation is updated step by step, and it is scaled with the means and the standard deviations obtained from the data set representing the past faulty situations.

In process industries, PCA and PLS have been dominant tools for statistical process monitoring. However, linear discriminant analysis (LDA) or Fisher discriminant analysis is also useful for fault diagnosis (Chiang et al., 2000). LDA provides an optimal lower dimensional representation in terms of discriminating among classes of data, it has been extensively studied in the pattern classification literature, and it has advantages over PCA when the objective is to discriminate among faults. Another technique suitable for fault diagnosis is discriminant PLS (Kemsley, 1996; Chiang et al., 2000).

MSPC has been widely accepted in process industries. However, their practicability and achievable performance are limited due to the assumption that a continuous process is operated in a particular steady state. Operating conditions cannot be constant in many processes because of load changes, product grade transitions, or other causes. In addition, steady-state set-points of not only controlled variables but also manipulated variables are frequently changed by a steady-state optimizer, which gives the optimal steady-state values to multivariable control systems. Such a hierarchical control system, consisting of a steady-state optimizer at a higher level and multivariable control systems at a lower level, has been widely accepted, especially in petrochemical process industries to achieve higher product qualities and lower operating costs at the same time. Kano et al. (2004a) developed a monitoring method based on

external analysis to distinguish faults from normal changes in operating conditions such as load changes and grade transitions. Yamamoto et al. (2004) reported an application of the monitoring method based on external analysis to an industrial monomer plant.

Recently, Dudzic and Zhang (2004) reported successful application results of MSPC to steel processes. They developed an integrated on-line monitoring system for a continuous caster. The monitoring system consists of four functions: start-up monitoring, process transition monitoring, run-time operation monitoring, and process state determination. The latter is able to determine the process state based on the current process conditions, and automatically selects the appropriate model and monitoring function. Multiway PCA is used for start-up monitoring and process transition monitoring while PCA is used for run-time operation monitoring.

### *2.3 Recognized Problems*

Soft-sensors play an important role to predict product quality from operating conditions and to derive better operating conditions that can improve the product quality. Conventional techniques are useful in most cases. However, they are not directly applicable to our problems in the steel industry because the product quality is given as qualitative information and it cannot be specified quantitatively. Conventional quantification methods are useful for building a quality model, but not for solving its inverse problem. A serious problem is that the physical meaning of the quantified variable is not clear at all. To address this issue, a novel quantification method that can give physical meaning to quantified variables is described in section 3.



As introduced in this section, extensive research on soft-sensor design and MSPC has been conducted. Many research works have focused on the performance of the developed systems, i.e., the estimation accuracy and the fault detection ability. However, from a practical viewpoint, there are other important issues to address. One of such important issues is model adaptation or model maintenance. The process characteristics will change with time. For example, reaction velocity changes due to inactivation of catalyst, and overall coefficient of heat transfer changes due to adhesion of dirt. To keep the performance of process models, for example, the estimation accuracy, the models must be updated. A simple approach to adapt an inferential model to a time-variant process is bias update based on measurements. The bias update is widely used in industries but it is not always sufficient. Several more advanced model adaptation methods such as recursive PCA/PLS have been proposed (Li et al., 2000; Qin , 1998). To make the recursive algorithms useful in real time, the issues of missing values and outliers were discussed. Although industrial application results have been reported, such recursive methods are not widely used in real plants. A difficulty of applying recursive modeling methods to real plants is anxiety that a model is updated in an undesirable way. For example, the model will cover narrower operating conditions when the process continues to be operated successfully around a specific operating point. As a result, the updated model will not be applicable in wide operating conditions, in which the model is expected to function well. Engineers are hesitant, fearing that automatic model adaptation might cause serious damage to the process. In fact, the key issue of soft-sensor design and MSPC is not only selection or usage of modeling methods but selection of process data suitable for modeling and selection of input variables necessary for estimation and monitoring. In practice, most efforts are devoted to such preparation. Therefore, an efficient

and robust procedure needs to be developed for adequate data preprocessing and model adaptation.

Another obstacle standing in our way to put statistical methods in practice is engineers' and operators' refusal. The validation of a statistical model is a difficult task. To win operators' trust, models and systems need to be understandable and consistent with process knowledge. In addition, models and systems need to be easily maintained.

### **3 Data-Driven Quality Improvement**

In this section, DDQI is formulated. DDQI constructs a statistical model from operation data, analyzes the cause of inferior quality and low yield, selects manipulated variables, and optimizes the operating conditions that can achieve the desired quality. Jaeckle and MacGregor (1998) proposed a product design method based on linear/nonlinear multivariate analysis. DDQI uses such conventional methods for modeling processes and optimizing operating conditions, but DDQI has several additional important functions. The advanced features of DDQI include: 1) handling qualitative variables as well as quantitative variables in a unified framework and 2) modeling batch processes and optimizing operation profiles through wavelet analysis and multivariate analysis. These features are explained in this section.

#### *3.1 Preprocessing Data*

A quality data matrix  $\mathbf{Y} \in \mathfrak{R}^{N \times Q}$  and an operating condition data matrix  $\mathbf{X} \in \mathfrak{R}^{N \times P}$  are observed.  $N$ ,  $Q$ , and  $P$  are the numbers of samples, quality variables,

and operating condition variables, respectively. The  $n$ th measurements of the  $q$ th quality variable and the  $p$ th operating condition variable are denoted by  $y_{nq}$  and  $x_{np}$ , respectively. For simplicity, it is assumed that quality data  $\mathbf{y}_q \in \mathfrak{R}^{N \times 1}$  and operating condition data  $\mathbf{x}_p \in \mathfrak{R}^{N \times 1}$  are mean-centered and scaled if necessary. In general, their standard deviations are scaled to be unity.

$$x_p = \frac{x_p^* - \bar{x}_p}{\sigma_{x_p}} \quad (5)$$

where  $x_p^*$  and  $x_p$  are measurements before and after scaling.  $\bar{x}_p$  and  $\sigma_{x_p}$  denote the mean and the standard deviation of  $x_p^*$ , respectively. However, they can be scaled independently on the basis of their characteristics. For example, signal-to-noise ratio can be reflected in scaling as follows:

$$x_p = \frac{x_p^* - \bar{x}_p}{\sigma_{x_p}} \frac{\sigma_{x_p} - \sigma_{x_{pe}}}{\sigma_{x_p}} \quad (6)$$

where  $\sigma_{x_{pe}}$  denotes the standard deviation of noise included in  $x_p^*$ . This scaling technique is useful when  $\sigma_{x_{pe}}$  is given or can be estimated.

### 3.2 Modeling Quality and Operating Conditions

DDQI is based on a statistical quality model that relates operating conditions with product quality. Multiple regression analysis (MRA) is the simplest method for building a quality model, but it cannot be used if a colinearity problem occurs. To cope with this problem, PCR and PLS can be used. The following formulation of DDQI is based on PCR for simplicity. DDQI can be formulated on the basis of PLS in the same way.

The singular value decomposition of an operating condition data matrix  $\mathbf{X}$  is written as

$$\begin{aligned}
\mathbf{X} &= \mathbf{U}\mathbf{S}\mathbf{V}^T \\
&= \begin{bmatrix} \mathbf{U}_R & \mathbf{U}_0 \end{bmatrix} \begin{bmatrix} \mathbf{S}_R & \mathbf{0} \\ \mathbf{0} & \mathbf{S}_0 \end{bmatrix} \begin{bmatrix} \mathbf{V}_R & \mathbf{V}_0 \end{bmatrix}^T
\end{aligned} \tag{7}$$

where  $\mathbf{U}$  and  $\mathbf{V}$  are orthogonal matrices.  $R$  denotes the number of principal components to be retained in a PCA model. The diagonal matrix  $\mathbf{S}$  has singular values  $s_r$  in its diagonal elements in decreasing order. The  $r$ th principal component is given as the  $r$ th column  $\mathbf{v}_r$  of the loading matrix  $\mathbf{V}_R$ , and the  $r$ th score  $\mathbf{t}_r$  is given by

$$\mathbf{t}_r = \mathbf{X}\mathbf{v}_r = s_r\mathbf{u}_r. \tag{8}$$

The score matrix is given by

$$\mathbf{T}_R = \mathbf{X}\mathbf{V}_R = \mathbf{U}_R\mathbf{S}_R. \tag{9}$$

Since the scores  $\mathbf{t}_r$  are uncorrelated with each other, the covariance matrix of  $\mathbf{T}_R$  is diagonal.

$$\frac{1}{N-1}\mathbf{T}_R^T\mathbf{T}_R = \frac{1}{N-1}\mathbf{S}_R^2 \tag{10}$$

The dimensionality can be reduced from  $P$  to  $R$  via PCA, and  $\mathbf{T}_R$  is obtained. Then,  $\mathbf{X}$  can be reconstructed as  $\hat{\mathbf{X}}$  by projecting  $\mathbf{T}_R$  onto the original  $P$  dimensional space.

$$\hat{\mathbf{X}} = \mathbf{T}_R\mathbf{V}_R^T = \mathbf{X}\mathbf{V}_R\mathbf{V}_R^T \tag{11}$$

The size of  $\hat{\mathbf{X}}$  is the same as that of  $\mathbf{X}$ , but the rank of  $\hat{\mathbf{X}}$  is only  $R$ .

In PCR, the scores are used as input variables of MRA. The collinearity problem does not occur any more, because the scores are uncorrelated with each

other. The PCR model can be written as

$$\mathbf{Y} = \mathbf{T}_R \mathbf{K} + \mathbf{E} \quad (12)$$

where  $\mathbf{K}$  is a regression coefficient matrix and  $\mathbf{E}$  is an error matrix. By least squares,  $\mathbf{K}$  is determined.

$$\mathbf{K} = (\mathbf{T}_R^T \mathbf{T}_R)^{-1} \mathbf{T}_R^T \mathbf{Y} \quad (13)$$

The prediction of product quality can be given by

$$\hat{\mathbf{Y}} = \mathbf{T}_R \mathbf{K} = \mathbf{X} \mathbf{V}_R \mathbf{K} = \mathbf{X} \mathbf{K}_{PCR} \quad (14)$$

$$\mathbf{K}_{PCR} \equiv \mathbf{V}_R \mathbf{K} . \quad (15)$$

The coefficient matrix  $\mathbf{K}_{PCR}$  shows the influence of operating conditions on quality. Basically, the operating condition variable with the larger coefficient has greater influence on the quality. However, the estimated coefficients are biased due to correlation among input variables.

Although PCR and PLS are useful for building a quality model, they cannot cope with process nonlinearity. On the other hand, nonlinear modeling methods such as artificial neural networks are not always desirable because limited samples make it difficult to build a reliable nonlinear model; moreover, its interpretation is difficult. Therefore, in real applications of DDQI, the first step is always to build a linear regression model by using PCR or PLS. To build a reliable model, input variables should be carefully selected on the basis of process knowledge as well as statistical tests. Then, a minimum number of nonlinear terms, in particular second order terms, are introduced into a model if the linear model is not good enough. The analysis of variance (ANOVA) is useful to clarify whether significant interaction exists between specific input variables. If it exists, then an interaction term,  $x_i x_j$ , is introduced into

the quality model as an additional input variable. In addition, nonlinearity between prediction error and each input variable is analyzed, and a second order term,  $x_i^2$ , is introduced into the quality model if necessary. As a result, the derived model is easy to analyze and interpret. This approach can generate a simple quality model with minimum nonlinear terms. Another effective approach to treat process nonlinearity is the use of nonlinear transformation. The nonlinear transformation is useful when suitable transformation is given or can be estimated. Furthermore, nonlinearity analysis is useful to select manipulated variables (Fujiwara et al., 2006). Once a model is developed and the relation between product quality and operating conditions is clarified, it is possible to select a limited number of manipulated variables, which are suitable for the use in a linear control system including run-to-run (R2R) control. The basic idea is to avoid using input variables that nonlinearly affect product quality or have significant interaction with other input variables.

### *3.3 Optimizing Operating Condition*

DDQI is an off-line analysis system that provides an optimal operating condition under a given objective function and constraints. To determine the optimal operating condition, an inverse problem of the quality model needs to be solved. The method adopted in DDQI is similar to the product design method based on multivariate analysis proposed by Jaeckle and MacGregor (1998).

In general, the number of quality variables  $Q$  is less than that of principal components  $R$ , and thus, the operating condition cannot be determined uniquely. Based on Eq. (14), the operating conditions  $\tilde{\mathbf{x}} \in \Re^{P \times 1}$  that can achieve the

desired quality  $\tilde{\mathbf{y}} \in \mathfrak{R}^{Q \times 1}$  are determined. The operating conditions must exist in the space spanned by  $R$  principal components, because correlation among operating condition variables, which is theoretically defined by physical or chemical laws, is extracted from operation data by using PCA. Therefore, at the first step to solving the inverse problem, the scores  $\tilde{\mathbf{t}} \in \mathfrak{R}^{R \times 1}$  that can achieve the desired quality  $\tilde{\mathbf{y}}$  are determined. The scores  $\tilde{\mathbf{t}}$  are related to the desired quality  $\tilde{\mathbf{y}}$  via

$$\tilde{\mathbf{y}} = \mathbf{K}^T \tilde{\mathbf{t}}. \quad (16)$$

If  $\tilde{\mathbf{t}}$  is found, then the operating conditions  $\tilde{\mathbf{x}}$  can be determined.

$$\tilde{\mathbf{x}} = \mathbf{V}_R \tilde{\mathbf{t}} \quad (17)$$

When the number of quality variables  $Q$  is larger than that of principal components  $R$ , there is no score  $\tilde{\mathbf{t}}$  that can achieve the desired quality. In such a case, the score that achieves as desired quality as possible should be determined.

$$\tilde{\mathbf{t}} = (\mathbf{K}\mathbf{K}^T)^{-1} \mathbf{K} \tilde{\mathbf{y}} \quad (18)$$

From Eq. (17), the operating condition recommended for quality improvement is

$$\tilde{\mathbf{x}} = \mathbf{V}_R (\mathbf{K}\mathbf{K}^T)^{-1} \mathbf{K} \tilde{\mathbf{y}}. \quad (19)$$

On the other hand, there is a unique score vector  $\tilde{\mathbf{t}}$  when  $Q = R$ . In this case, the operating condition recommended for quality improvement is also determined by Eq. (19) or

$$\tilde{\mathbf{x}} = \mathbf{V}_R (\mathbf{K}^T)^{-1} \tilde{\mathbf{y}}. \quad (20)$$

In many cases, however,  $Q$  is less than  $R$ . Therefore, the scores  $\tilde{\mathbf{t}}$  cannot be determined uniquely. In such a case, a solution set  $\mathcal{S}_{\tilde{\mathbf{t}}}$  where any score can achieve the desired quality is searched. By using a matrix  $\tilde{\mathbf{T}}_0 \in \mathfrak{R}^{R \times (R-Q)}$  whose columns are bases of the null space (kernel) of  $\mathbf{K}^T$ ,  $\tilde{\mathbf{t}}$  can be described as

$$\tilde{\mathbf{t}} = (\mathbf{K}^T)^+ \tilde{\mathbf{y}} + \tilde{\mathbf{T}}_0 \phi. \quad (21)$$

where  $(\mathbf{K}^T)^+$  denotes the generalized inverse of  $\mathbf{K}^T$  and  $\phi \in \mathfrak{R}^{(R-Q) \times 1}$  is any coefficient vector. Finally the operating conditions that can achieve the desired quality  $\tilde{\mathbf{y}}$  are given by

$$\tilde{\mathbf{x}} = \mathbf{V}_R \left\{ (\mathbf{K}^T)^+ \tilde{\mathbf{y}} + \tilde{\mathbf{T}}_0 \phi \right\}. \quad (22)$$

Any general solution  $\tilde{\mathbf{x}}$  is an element of the column space of the loading matrix  $\mathbf{V}_R$ , and thus it is clear that  $\tilde{\mathbf{x}}$  exists in the space spanned by  $R$  principal components.

Even though the operating condition cannot be determined uniquely, it can be optimized when an objective function is provided. The objective function is optimized under the following constraints: 1) the desired product quality is achieved, 2) all operating condition variables exist within their upper and lower bounds, and 3)  $T^2$  statistic of scores is below its upper control limit or approximately 3') all scores exist within their upper and lower bounds. The last constraint is necessary for finding a new optimal operating condition within the region where the statistical model is valid. In other words, extrapolation should be avoided by using the last constraint. If there is no solution that satisfies all constraints, i.e., the imposed specifications on quality are too severe, the operating condition that achieves as desired quality as possible should be



determined.

The formulation described in this subsection is based on PCR for simplicity, but any linear/nonlinear model can be used for optimizing operating conditions.

### *3.4 Operation Profile Optimization*

In section 3.3, the method to optimize the operating conditions was explained. However, the method needs to be modified when an operation profile is crucial. A batch process is a typical example of such a situation. To apply multivariate analysis to batch process data, multiway methods such as multiway PCA (MPCA) and multiway PLS (MPLS) have been developed (Nomikos and MacGregor, 1994, 1995). In multiway methods, three-dimensional arrays are unfolded to two-dimensional arrays. For example, a three-dimensional array (batches  $\times$  variables  $\times$  time) can be unfolded to (batches  $\times$  (variables  $\times$  time)) when variables at different sampling points are regarded as different variables. In this type of approach, the number of input variables after unfolding becomes significantly larger than that of original input variables. The increase of input variables generally causes deterioration in the reliability or the estimation accuracy of the statistical model. Although one practical approach for solving this problem is to thin out data at a part of sampling points and to make the ratio of the number of samples, i.e., batches, to that of input variables higher, process information corresponding to the removed data is lost.

To relate product quality with operation profiles efficiently, the use of wavelet

analysis has been investigated by many researchers. The key idea is to realize feature extraction and data compression through wavelet analysis. However, to optimize operation profiles that can achieve the desired product quality, not only the estimation performance of product quality but also the reconstruction performance of operation profiles is important, and wavelet coefficients useful for reconstructing original signals and estimating outputs must be selected appropriately. Kano et al. (2006) proposed such a method, in which important wavelet coefficients of operation profiles are selected as input variables of a statistical model, and then further dimensionality reduction is achieved through multivariate analysis. In addition, the indicator variable technique is used to cope with the problem of unequal duration of each batch and to align batch trajectories (Nomikos and MacGregor, 1994). DDQI has this function of optimizing operation profiles through wavelet analysis and multivariate analysis.

### *3.5 Handling Qualitative Variables*

In previous subsections, the method to optimize the operating conditions was explained. However, it is applicable only to cases where all quality variables are quantitative. When the quality variables are qualitative, e.g., good and bad, the product quality cannot be specified quantitatively.

To build a quality model, qualitative variables should be quantified. As is well-known, for example, qualitative information such as good and bad can be quantified and denoted by 1 and 0, respectively.

$$y_{nq} = \begin{cases} 1, & \text{for good product} \\ 0, & \text{for bad product} \end{cases} \quad (23)$$

The quantified variable is then mean-centered, and the quality model can be built via PCR.

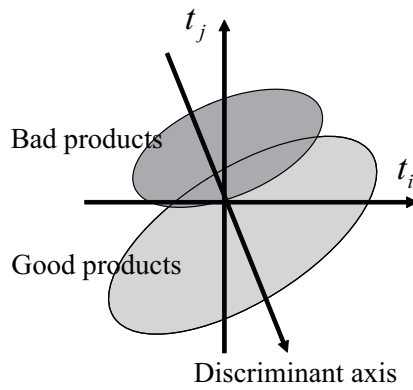
$$\hat{\mathbf{y}}_q = \mathbf{X}\mathbf{k}_{PCR} \quad (24)$$

A quality variable can be quantified in a similar way by using discriminant PLS.

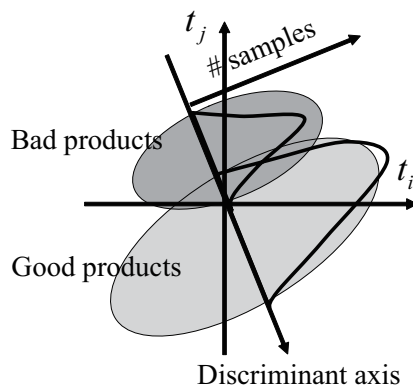
This conventional quantification method is useful for building a quality model, but not for solving its inverse problem. A serious problem is that the physical meaning of the quantified variable is not clear at all. For example, what does 0.6 mean when a quality variable is quantified through Eq. (23)? Is it acceptable or not? Nobody can answer this question. To address this problem, Kano et al. (2004b) proposed a novel quantification method that was able to give physical meaning to quantified variables.

The proposed quantification method consists of the following five steps:

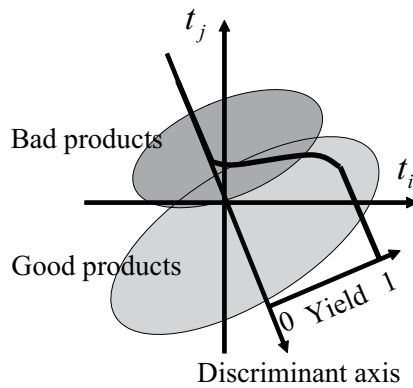
- (1) All operation data are classified into two classes, good products and bad products, on the basis of available qualitative information on product quality.
- (2) PCA is applied to the data for dimensionality reduction.
- (3) LDA is applied to the principal component scores to discriminate between good products and bad products. (Fig. 3(a))
- (4) All data are projected onto the discriminant axis, and histograms of good products and bad products are drawn against the discriminant axis.



- (a) Discrimination between good products and bad products on the subspace spanned by principal components by using LDA.



- (b) Histograms of good products and bad products against the discriminant axis.



- (c) Product yield against the discriminant axis.

Fig. 3. Quantification method based on PCA-LDA.

(Fig. 3(b))

(5) Product yield is calculated from the histograms. (Fig. 3(c))

This method is referred to as PCA-LDA. The first three steps of PCA-LDA are the same as the conventional quantification method. It should be noticed that the derived coefficient vector  $\mathbf{k}_{PCR}$  in Eq. (24) is proportional to the vector describing the discriminant function of PCA-LDA.

In PCA-LDA, the product yield can be specified instead of the quality itself on the basis of the histogram for each category. Each histogram can be obtained from operation data, and it can be drawn as the frequency distribution of good or bad samples against the axis defined by  $\mathbf{k}_{PCR}$ . Then, the product yield against  $\mathbf{k}_{PCR}$  can be defined by

$$\text{yield}(\Delta_i) = \frac{N_{good}(\Delta_i)}{N_{good}(\Delta_i) + N_{bad}(\Delta_i)} \quad (25)$$

where  $\Delta_i$  is the  $i$ th bin of the histogram.  $N_{good}(\Delta_i)$  and  $N_{bad}(\Delta_i)$  are the numbers of good and bad samples contained in  $\Delta_i$ , respectively. To calculate the product yield through Eq. (25), the position of each bin must be the same between the histogram for good products and that for bad products. Once the desired yield is specified, operating conditions that can achieve the desired yield can be found by following the above-mentioned approach.

#### 4 Applications to Industrial Iron and Steel Processes

Sumitomo Metals Kokura manufactures steel bars and wire rods. These products are widely used for automobiles, vessels, industrial machinery, electrical machinery, buildings, civil engineering, and so on. The products are produced through various processes. Iron ore and coke are fed to the blast furnace to make iron. The blast furnace is a huge chemical reactor where reduction re-

actions take place. The iron is sent to the steel making process where bloom is produced. The steel making process consists of converters for removing carbon, refiners for adjusting elements, and continuous casters. Then, the blooming process resizes bloom to billet for the next rolling process. The bar rolling process or the wire rolling process rolls the billet to the desired size. Finally, surface flaws, internal defects, hardness, and other properties are inspected.

In the iron and steel industry, process control systems have been designed on the assumption that the relationship between controlled product quality variables and manipulated variables is known for the most part. For example, a setup (feedforward) control system and a real-time (feedback) control system of strip thickness in a rolling process can be developed by using a mathematical model based on the theory of rolling (Bland and Ford, 1948). In this control application, the influence of tension between rolling stands and roll gap, both of which are manipulated variables, on strip thickness can be calculated through the mathematical model. Therefore, both manipulated variables are adjusted in advance to realize desired strip thickness. This type of control is referred to as setup control. In addition, a thickness sensor is installed at least after the last stand of a rolling mill, and strip thickness can be measured on-line. Therefore, real-time feedback control is applied to remove deviation of strip thickness during rolling.

On the other hand, the relationship of operating conditions to product defects such as surface flaws and internal defects is not clear. In general, the product quality has been maintained by skilled operators on the basis of their experience and intuition. It is empirically known that the occurrence of surface flaws and internal defects is affected by operating conditions of a furnace in a rolling process and a continuous casting equipment in a steel making process.

However, it is not clear which operating condition has an influence on product qualities and to what extent. Although much effort has been devoted to clarify the relationship between operating conditions and product quality, the problem remains unsolved because there are too many operating conditions to investigate. Therefore, effective setup control is difficult to apply. In addition, since internal defects are checked by using ultrasonic testing after a rolling process, and it may take a few days to get control results in a steel making process, real-time control cannot be applied to this type of quality control problem. To improve product yield, it is important to predict final product qualities on-line.

Due to these characteristics, surface flaws and internal defects have not been the target of process control for many years. However, the business situation is changing. To meet customers' requirements for higher product quality, to realize higher product yield, and to cope with a decrease in skilled operators (Year 2007 problem in Japan), most iron and steel companies have started to cope with product quality issues within a framework of process control. For this purpose, modeling is the key to success.

Recently, many companies build integrated data-bases. At Sumitomo Metals Kokura, huge amounts of data of operating conditions and inspection results are stored in the integrated data-base, and any technical staff in any factory can access the integrated data-base easily.

Under the above-mentioned situation, the authors have investigated a statistical approach to improve the product quality in iron and steel processes. In this section, two application results are presented.

In general, PLS outperforms PCR in the estimation performance. However,

PCA-LDA were used in both applications and others, because PCA had been used for process analysis and monitoring rather than PLS in the processes that DDQI has been applied to. In other words, PCA-LDA is more consistent with existing tools and systems, and engineers and operators are more familiar with PCA than PLS.

#### *4.1 Reduction of Surface Flaws*

Reducing surface flaws is crucial for improving product yield. Surface flaws are checked after cold rolling as one of the key product qualities, and their shape and size are varied depending on their factors and steel grades produced. In this application, surface flaws frequently occurring in a specific alloy steel product are investigated. It is empirically known that the occurrence of surface flaws is affected by operating conditions of a rolling process and a steel making process. A large amount of defective steel is produced if steel making causes surface flaws, because surface flaws cannot be detected before the final inspection process. Therefore, it is important to clarify the cause of surface flaws, to find the operating condition that can minimize them, and to realize setup control.

First, a model relating operating conditions in steel making and hot rolling to surface flaws inspection results was developed. Input variables include contents of various additive elements in a steel making process, temperature and residence time in each heating zone in a hot rolling process, and temperature at the exit of each stand in a hot rolling process. The total number of input variables selected was 55. The sample number was 138 consisting of 122 samples with surface flaws, classified into bad, and only 16 samples without



surface flaws, classified into good.

In general, the use of process knowledge for selecting input variables is key to success in building statistical models. In this application, the number of process variables that can be used as input variables is several hundreds or more. Therefore, input variable selection is crucial for effective modeling. Input variables were selected by engineers and operators who have profound knowledge of the processes. In addition, operation data were carefully examined, unreliable data were excluded, and linear interpolation was used for coping with missing data or outliers by taking the process characteristics into account. Furthermore, the developed model was carefully examined by the engineers and operators.

The dimensionality was reduced from 55 to 6 via PCA after all variables were normalized. Then, LDA was used for discriminating between two classes, good and bad. The developed PCA-LDA model was able to successfully discriminate between good products and bad products along the discriminant axis as shown in Fig. 4. This PCA-LDA model can relate operating conditions with the product yield through the proposed quantification method. Six variables having the largest influence on the product yield are listed in Fig. 5 with their regression coefficients. On the basis of this result, process engineers selected to manipulate the content of alloy element e1 considering both operation cost and operability. Figure 5 suggests that surface flaws can be reduced by increasing the content of alloy element e1. The results of verification experiments are shown in Fig. 6. The verification experiments confirmed that surface flaws can be significantly reduced by increasing the content of alloy element e1.

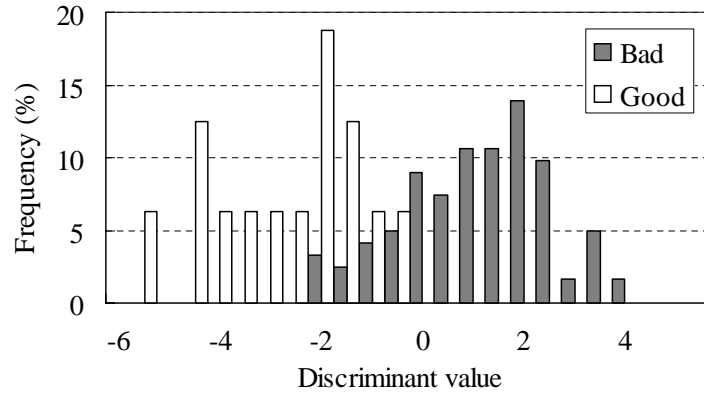


Fig. 4. Analysis of surface flaws via PCA-LDA (1). Histograms of good products and bad products are drawn against the discriminant axis. Product yield are calculated from the histograms.

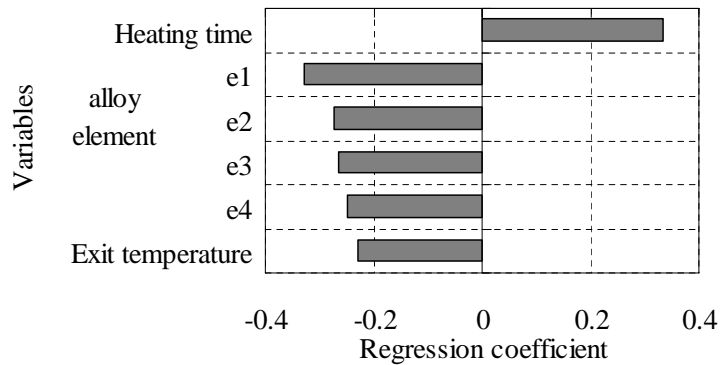


Fig. 5. Analysis of surface flaws via PCA-LDA (2). Top six influential factors.

#### 4.2 Reduction of Internal Defects

The objective in this application is to minimize internal defects. The target process consists of a steel making process, a blooming process, and a bar rolling process. Internal defects are checked by using ultrasonic testing (UST) after the bar rolling process. In this application, 40 operating condition variables are selected as input variables. The sample number is 740 consisting of 208

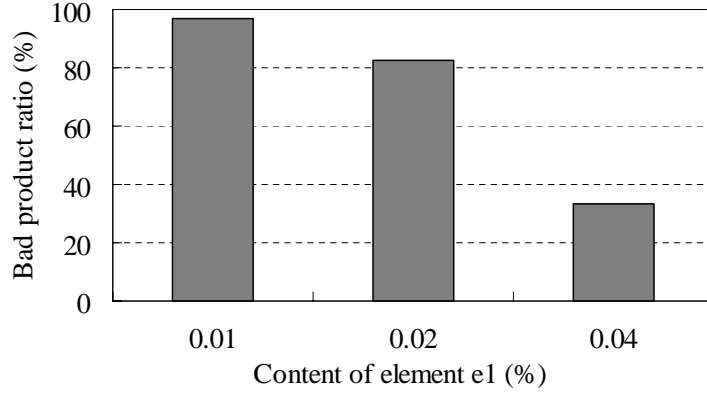


Fig. 6. The results of verification experiments. The effect of alloy element e1 content on bad product ratio.

samples with internal defects and 532 samples without internal defects.

The developed PCA-LDA model with five principal components was able to discriminate between good products and bad products along the discriminant axis. On the basis of the PCA-LDA model, process engineers selected to manipulate two factors in the steel making process and one factor in the blooming process considering both operation cost and operability. The next step is to optimize operating conditions. The optimal operating condition that can improve the product yield by 20% is searched. To verify the results, test experiments were performed at the operating condition close to the optimal point, and it was confirmed that the percentage of defective products was reduced by half.

#### 4.3 Effect of Dimensionality Reduction

In most applications, PCA-LDA was used to build a model relating operating conditions to product quality. Although other methods such as discriminant

PLS can be used for influential factor analysis, LDA should not be used directly. To clarify the advantage of PCA-LDA over LDA, these two methods were compared by analyzing influential factors on steel sheet surface flaws. A total of 73 operating condition variables with 105 samples classified into "no flaws" and 47 samples classified into "flaws" were used for modeling. In PCA-LDA, 10 principal components were retained in the PCA model. The classification power of LDA is significantly higher than PCA-LDA when they are applied to data used for modeling. However, the LDA model provides misleading results because operating condition variables are highly correlated. For example, temperatures at two different measurement points had significant coefficients with different sign in the LDA model, although these two temperatures have very strong positive correlation. It is impossible to change these two temperatures in different directions. The result suggests that PCA or another method must be used together with LDA to get a reliable result. In fact, from the PCA-LDA model with 10 principal components, it was concluded that temperatures identified by the LDA model were not key variables to reduce defects, and it was suggested that sheet steel should be cooled down rapidly after the hot rolling process. The results of the PCA-LDA model were confirmed to be correct by changing the operating condition.

#### *4.4 Integration of DDQI with MSPC*

In the process industry, a hierarchical control system has been widely accepted. The most famous one would be a model predictive control system which is integrated with a steady-state optimizer and local controllers. At the top (first) layer is the steady-state optimizer that determines the optimal steady state.

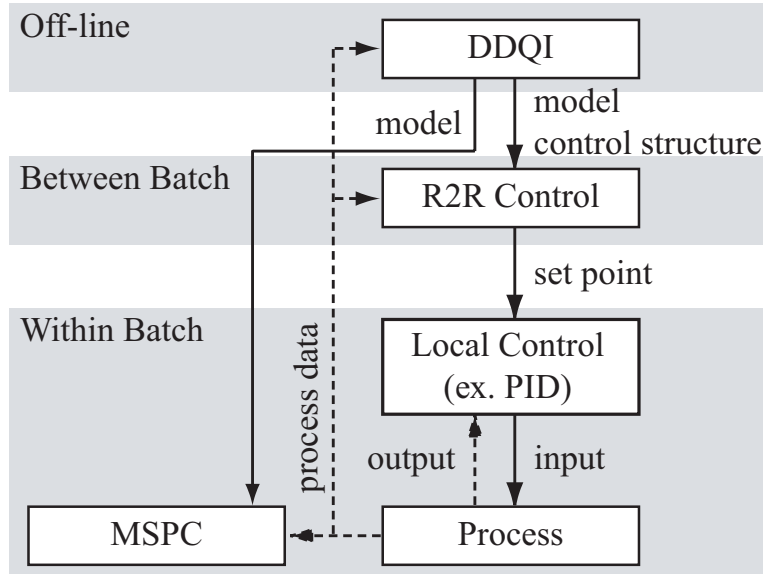


Fig. 7. A schematic diagram of hierarchical quality improvement system (HiQIS). At the second layer, model predictive control aims to bring a process to its optimal steady state. The model predictive controller optimizes future manipulated variables under a given objective function and constraints. Usually, the objective function is determined according to the trade-off between the control performance and the stability. The optimal manipulated variables are given to local controllers at the third layer as their set-points.

Qin et al. (2004) proposed a hierarchical fab-wide control framework to maintain highly efficient operations, minimize downtime of equipment, and optimize the yield of high quality products in the semiconductor industry. The fab-wide control system is analogous to the model predictive control system. The hierarchical quality improvement system (HiQIS) proposed by Fujiwara et al. (2006) is also an analogue to them. A schematic diagram of HiQIS is shown in Fig. 7. DDQI is located at the top of the hierarchy. R2R control updates operating conditions or operation profiles for the next batch (Castillo and Huriwitz, 1997; Moyne et al., 2001), and it gives set-points to local controllers on the basis of information provided by DDQI, i.e., the quality model,

the manipulated variables or control structure, and the optimal operating conditions. In addition, MSPC detects and diagnoses faults on the basis of the statistical model built in DDQI.

The authors have succeeded in applying DDQI to the iron and steel processes and reducing surface flaws and internal defects significantly. However, to keep stable operation and to realize continuous improvement, DDQI needs to be integrated with MSPC. PCA-based MSPC is used to be consistent with DDQI, in which PCA-LDA models are used. Through the applications of DDQI, process characteristics, in particular the influence of operating conditions on product quality, became more clear than ever before. As a result, good models are available not only for DDQI but also for MSPC. DDQI and MSPC are currently being integrated in the plants.

## **5 Conclusions**

How can we improve product quality and yield? This question becomes more critical in various industries. Even though the processes are totally different in appearance, the problems to solve are highly similar. Therefore, it is expected that a unified framework can be developed to answer the above question.

The model-based approach is preferable to the data-based approach if first-principle models with sufficient accuracy are available. However, in the steel industry, for example, any quantitative first-principle model to relate surface flaws or internal defects to operating conditions is not available. Therefore, data-based approaches were adopted.

In this paper, first, data-based process monitoring and control techniques were

briefly surveyed. The focal points were recent trends and remaining problems in the fields of soft-sensor design and statistical process monitoring. Then, Data-Driven Quality Improvement (DDQI) was explained. DDQI constructs a statistical model from operation data, analyzes the cause of inferior quality and low yield, selects manipulated variables, and optimizes the operating conditions that can achieve the desired quality. Since 2004, DDQI has been applied to many processes including iron making, steel making, blooming, hot rolling, cold rolling, and continuous galvanic lines for improving product quality and realizing stable operation. In this paper, the selected application results of DDQI to the iron and steel processes were presented. The achievements of the present work are as follows: 1) the development of a new method that can cope with qualitative quality information and relate operating conditions to product yield, 2) the simultaneous analysis of multiple processing units including a converter, a continuous caster, a blooming process, and rolling processes, and 3) the successful application results. From their experience in applying DDQI to several industrial processes, the authors hope to develop a unified framework that can answer to the question "How can we improve product quality and yield?" on the basis of data-based methodologies.

## **Acknowledgements**

This research was partially supported by JFE 21st Century Foundation and Innovation Plaza Kyoto, Japan Science and Technology Agency (JST).

## References

- Amirthalingam, R. & Lee, J.H. (1999). Subspace Identification Based Inferential Control Applied to a Continuous Pulp Digester. *J. Proc. Cont.*, **9**, 397–406.
- Bakshi, B.R. (1998). Multiscale PCA with application to multivariate statistical process monitoring. *AIChE J.*, **44**, 1596–1610.
- Bland, D.A., & Ford, H. (1948). The Calculation of Roll Force and Torque in Cold Strip Rolling with Tensions. *Proc. Instn. Mech. Eng.*, **159**, 144–153.
- Castillo, E.D. & Huriwitz, A.M. (1997). Run-to-Run Process Control: Literature Review and Extensions. *J. Qual. Technol.*, **29**, 184–196.
- Chen, G., McAvoy, T.J., Piovoso, M.J. (1998). A multivariate statistical controller for on-line quality improvement. *J. Proc. Cont.*, **8**, 139–149.
- Chiang, L.H., Russell, E.L., & Braatz, R.D. (2000). Fault Diagnosis in Chemical Processes Using Fisher Discriminant Analysis, Discriminant Partial Least Squares, and Principal Component Analysis. *Chemometrics and Intelligent Laboratory Systems*, **50**, 243–252.
- Dudzic, M. & Zhang, Y. (2004). On-line Industrial Implementation of Process Monitoring/Control Applications Using Multivariate Statistical Technologies: Challenges and Opportunities. *IFAC Symp. on Dynamics and Control of Process Systems (DYCOPS)*, CD-ROM, Cambridge, July 5–7.
- Fujiwara, K., Kano, M., Hasebe, S. & Ohno, H. (2006). Development of Data-Based Hierarchical Quality Improvement System (HiQIS) -Manipulated Variables Selection for Quality Control-. *Transactions of the Society of Instrument and Control Engineers*, **42**, 909–915.
- Jackson, J.E. (1959). Quality Control Methods for Several Related Variables. *Technometrics*, **1**, 359–377.



- Jackson, J.E. & Mudholkar, G.S. (1979). Control Procedures for Residuals Associated with Principal Component Analysis. *Technometrics*, **21**, 341–349.
- Jaeckle, C.M. & MacGregor, J.F. (1998). Product Design through Multivariate Statistical Analysis of Process Data. *AIChE J.*, **44**, 1105–1118.
- JSPS PSE 143 committee. (2004). Annual Rreport 2003, **28**. [in Japanese]
- Kamohara, H., Takinami, A., Takeda, M., Kano, M., Hasebe, S., & Hashimoto, I. (2004). Product Quality Estimation and Operating Condition Monitoring for Industrial Ethylene Fractionator. *J. Chem. Eng. Japan*, **37**, 422–428.
- Kano, M., Miyazaki, K., Hasebe, S. & Hashimoto, I. (2000). Inferential Control System of Distillation Compositions Using Dynamic Partial Least Squares Regression. *J. Proc. Cont.*, **10**, 157–166.
- Kano, M., Ohno, H., Hasebe, S., & Hashimoto, I. (2001). A New Multivariate Statistical Process Monitoring Method Using Principal Component Analysis. *Computers chem. Engng*, **25**, 1103–1113.
- Kano, M., Hasebe, S., Hashimoto, I., & Ohno, H. (2002). Statistical Process Monitoring Based on Dissimilarity of Process Data. *AIChE J.*, **48**, 1231–1240.
- Kano, M., Showchaiya, N., Hasebe, S. & Hashimoto, I. (2003a). Inferential Control of Distillation Compositions: Selection of Model and Control Configuration. *Control Engineering Practice*, **11**, 927–933.
- Kano, M., Tanaka, S., Hasebe, S., Hashimoto, I., & Ohno, H. (2003b). Monitoring Independent Components for Fault Detection. *AIChE J.*, **49**, 969–976.
- Kano, M., Hasebe, S., Hashimoto, I., & Ohno, H. (2004a). Evolution of Multivariate Statistical Process Control: Application of Independent Component Analysis and External Analysis. *Comput. chem. Engng*, **28**, 1157–1166.
- Kano, M., Fujiwara, K., Hasebe, S., & Ohno, H. (2004b). Data-Driven Quality

- Improvement: Handling Qualitative Variables. *IFAC DYCOPS*, CD-ROM, Cambridge, July 5–7.
- Kano, M., Fujiwara, K., Hasebe, S., & Ohno, H. (2006). Operation Profile Optimization for Batch Process through Wavelet Analysis and Multivariate Analysis. *SICE-ICASE International Joint Conference 2006*, CD-ROM, 5749-5753, Busan, Korea, Oct. 18–21.
- Kemsley, E.K. (1996). Discriminant analysis of high-dimensional data: a comparison of principal components analysis and partial least squares data reduction methods *Chemometrics and Intelligent Laboratory Systems*, **33**, 47–61.
- Kresta, J.V., MacGregor, J.F., & Marlin, T.E. (1991). Multivariate Statistical Monitoring of Process Operating Performance. *Can. J. Chem. Eng.*, **69**, 35–47.
- Kresta, J.V., Marlin, T.E. & MacGregor, J.F. (1994). Development of inferential process models using PLS. *Comp. chem. Engng*, **18**, 597–611.
- Ku, W., Storer, R.H., & Georgakis, C. (1995). Disturbance Detection and Isolation by Dynamic Principal Component Analysis. *Chemometrics and intelligent laboratory systems*, **30**, 179–196.
- Lee, S., Kano, M., Ando, K., & Hasebe, S. (2006). Softsensor Development through Two-Stage Subspace Identification. *SICE-ICASE International Joint Conference 2006*, CD-ROM, 5731–5736, Busan, Korea, Oct. 18–21.
- Li, W., Yue, H.H., Valle-Cervantes, S. & Qin, S.J. (2000). Recursive PCA for Adaptive Process Monitoring. *J. Proc. Cont.*, **10**, 471–486.
- MacGregor, J.F., Jaeckle, C., Kiparissides, C. & Koutoudi, M. (1994). Process Monitoring and Diagnosis by Multiblock Methods. *AIChE J.*, **40**, 826–838.
- Mejdell, T. & Skogestad, S. (1991a). Estimation of Distillation Compositions

- from Multiple Temperature Measurements Using Partial-Least-Squares Regression. *Ind. Eng. Chem. Res.*, **30**, 2543–2555.
- Mejdell, T. & Skogestad, S. (1991b). Composition Estimator in a Pilot-plant Distillation Column Using Multiple Temperatures. *Ind. Eng. Chem. Res.*, **30**, 2555–2564.
- Mejdell, T. & Skogestad, S. (1993). Output Estimation Using Multiple Secondary Measurements: High-purity Distillation. *AIChE J.*, **39**, 1641–1653.
- Montgomery, D.C. (1997). Introduction to statistical quality control. 132–150, John Wiley, New York.
- Moyne, J., Castillo, E.D. & Hurwitz, A.M. (2001). Run-to-Run Control in Semiconductor Manufacturing. CRC Press.
- Nomikos, P. & MacGregor, J.F. (1994). Monitoring Batch Processes Using Multiway Principal Component Analysis. *AIChE J.*, **40**, 1361–1375.
- Nomikos, P. & MacGregor, J.F. (1995). Multiway Partial Least Squares in Monitoring Batch Processes. *Chemometrics and Intelligent Laboratory Systems*, **30**, 97–109.
- Nomikos, P. (1996). Detection and diagnosis of abnormal batch operations based on multi-way principal component analysis. *ISA Trans.*, **35**, 259–266.
- Ookita, K. (2006). Operation and Quality Control for Chemical Plants By Soft-sensors. *CICSJ Bulletin*, **24**, 31–33. [in Japanese]
- Overschee, P.V. & Moor, B.D. (1994). N4SID: Subspace Algorithms for the Identification of Combined Deterministic-Stochastic Systems. *Automatica*, **30**, 75–93.
- Piovoso, M.J. & Kosanovich, K.A. (1994). Applications of Multivariate Statistical Methods to Process Monitoring and Controller Design. *International Journal of Control*, **59**, 743–765.
- Qin, S.J. (1998). Recursive PLS algorithms for adaptive data modeling. *Comp.*

- chem. Engng*, **22**, 503–514.
- Qin, S.J., Cherry, G., Good, R., Wang, J., & Harrison, C.A. (2004). Control and Monitoring of Semiconductor Manufacturing Processes: Challenges and Opportunities. *IFAC Symposium on Dynamics and Control of Process Systems (DYCOPS)*, CD-ROM, Cambridge, July 5–7.
- Raich, A. & Cinar, A. (1996). Statistical Process Monitoring and Disturbance Diagnosis in Multivariable Continuous Processes. *AIChE J.*, **42**, 995–1009.
- Rotem, Y., Wachs, A., & Lewin, D.R. (2000). Ethylene Compressor Monitoring Using Model-Based PCA. *AIChE J.*, **46**, 1825–1836.
- Showchaiya, N., Kano, M., Hasebe, S. & Hashimoto, I. (2001). Improvement of Distillation Composition Control by Using Predictive Inferential Control Technique. *J. Chem. Eng. Japan*, **34**, 1026–1032.
- Westerhuis, J.A., Gurden, S.P., & Smilde, A.K. (2000). Generalized Contribution Plots in Multivariate Statistical Process Monitoring. *Chemometrics and Intelligent Laboratory Systems*, **51**, 95–114.
- Wold, S., Geladi, P., Esbesen, K., & Ohman, J. (1987). Multi-way Principal Components- and PLS-Analysis. *J. Chemometrics*, **1**, 41–56.
- Yamamoto, T., Shimameguri, A., Ogawa, M., Kano, M., & Hashimoto, I. (2004). Application of Statistical Process Monitoring with External Analysis to an Industrial Monomer Plant. *Preprints of IFAC Symposium on Advanced Control of Chemical Processes (ADCHEM)*, 405–410, Hong Kong, Jan. 11–14.
- Yoon, S. & MacGregor, J.F. (2001). Incorporation of External Information into Multivariate PCA/PLS models. *Prep. of 4th IFAC Workshop on On-line Fault Detection and Supervision in the Chemical Industries*, 121–126, Jeju Island, Korea, June 7–8.