Available online at www.sciencedirect.com

SCIENCE DIRECT®

Control Engineering Practice 12 (2004) 945–961

CONTROL ENGINEERING PRACTICE

www.elsevier.com/locate/conengprac

Process improvement methodology based on multivariate statistical analysis methods

Young-Hak Lee[a], Kwang Gi Min[a], Chonghun Han[a,*], Kun Soo Chang[a], Tae Hwa Choi[b]

[a] Department of Chemical Engineering and Automation Research Center, Pohang University of Science and Technology, San 31 Hyojadong Namgu, Pohang, Kyungbuk 790-784, South Korea

[b] Pohang Iron and Steel Company, Pohang, Kyungbuk 790-785, South Korea

Received 14 June 2001; accepted 26 November 2003

Abstract

A systematic procedure for process improvement methodology is proposed based on multivariate statistical process control methods. To take advantage of a large amount of historical data, the procedure employs a combination of hierarchical clustering method and statistical process control methods to detect and analyze the key factors that significantly affect the performance of processes. This methodology consists of four sequential steps: (1) Data collection and multivariate statistical analysis; (2) hierarchical clustering and operation mode detection; (3) selection of dominant variables; (4) a new operational guideline and its validation. The proposed procedure was applied to improve the heat efficiency of an industrial hot stove system located at Pohang Iron & Steel Co. (POSCO) in Korea. The implementation results show that the proposed methodology helps us systematically improve the operating conditions of the hot stove system.

Keywords: Process improvement; Multivariate statistical process control; Hierarchical clustering; Hot stove system

1. Introduction

Process improvement is an essential activity in most of the industries in order to survive in the increasingly competitive global market. Therefore, there has been increasing attention on the methodologies for process improvement.

Basically, process improvement involves two complementary steps: the reduction of product variations within pre-specified limits and continuous process improvement through finding the effects of *common* (*or chance*) *causes* upon the performance of processes.

The key idea of the first step is to achieve process improvement through reducing the variations of product to a pre-specified limit based on Statistical Process Control (SPC) methods. Achieving this goal requires three sub-steps: (1) detection of the occurrence of abnormal situations based on control charts such as Shewart, CUSUM, and EWMA charts, which are extensively used for process monitoring; (2) diagnosis for the *assignable* (*special*) *causes* that are generating the abnormal situations; and (3) elimination of those causes.

In the second major step, the goal is to provide an opportunity to improve process performance through finding the relationship between product quality and common causes, presented within the process itself. In order to analyze this relationship, there are typically two methods according to the methods of data gathering: designed experimental methods and non-experimental methods.

Designed experimental methods have been developed in several related ways such as a fractional factorial design, an orthogonal array, and evolutionary operation (Taguchi, 1986; Box & Draper, 1998). These methods have been developed for providing the optimum operating conditions. However, these methods have limits in the scope, the frequency, and the number of situations in which experiments have been conducted. Furthermore, these methods have possibilities of unexpected risk and require high experimental cost for implementation.

*Corresponding author. Tel.: +82-542-795-397; fax: +82-542-798-478.

E-mail address: chan@postech.ac.kr (C. Han).

0967-0661/$ - see front matter © 2003 Published by Elsevier Ltd.
doi:10.1016/j.conengprac.2003.11.004

Nomenclature

C	criterion for sample variance
C_{pm1}, C_{pm2}	cold and hot air specific heat, respectively (kcal/m^3 °C)
CV_m	index of dominant variable
$\mathbf{E}$	residual matrix
G	combustion gas flow rate (m^3/h)
He	gas heat (kcal/m^3)
J	criterion function for clustering
m_i	mean vector of ith cluster
N_c	number of clusters
$\mathbf{P}, \mathbf{p}_r$	loading matrix and its column vector, respectively
Q_1, Q_2	cold and hot air temperature, respectively (°C)
Q_i	sum of squared prediction error for the ith batch
R	number of principal components, correlation matrix
$\mathbf{s}^2$	sample variance
$\mathbf{S}^{-1}$	sample covariance matrix
t	heating period (h/cycle)
$\mathbf{t}_r, \mathbf{T}$	score vector and matrix, respectively
T^2	Hotelling's T^2 statistic
$\Delta\mathbf{t}_{ij}$	distance between cluster i and cluster j
T	blowing time (h/60 cycle), score matrix
V	blowing flow rate (60 m^3/h)
$\Delta\mathbf{x}_{ij}$	difference of data between group i and group j
$\mathbf{v}m_{ij}$	variations of mth variable between group i and j
$\mathbf{X}$	data matrix
$\mathbf{Y}$	quality matrix
$\bar{y}$	mean of quality variable
z	scores in the reduced dimension
Z_i	ith subset of cluster

Greeks letters

ζ	mapping function
ω_i	ith cluster labels
Ω	interpretation space
Ξ	set of all possible cluster labels

On the other hand, non-experimental methods aim at extracting new and important information to improve process performance based on the process data obtained directly from process information system (PIS). Although these non-experimental methods require no cost and risk in gathering process data, there are limits in the range of possible process improvement. However, at most of modern industrial sites, it is believed that the collected historical data would be a gold mine of information, if only the "important" and relevant information could be extracted painlessly and quickly (Piovoso, Kosanovich, & James, 1992). Therefore, there has been increasing attention on the methods to collect historical data for process improvement such as higher productivity, better product quality, increased safety, and lower environmental pollution.

Thanks to the deployment of PIS in the industrial world, engineers can get a huge amount of process data representing operating conditions. The industrial data shows that variables are highly correlated and often collinear, signal-to-noise ratio (SNR) is low, and even there are missing data. These complex characteristics of the data make it difficult to analyze the data and extract the information using conventional univariate SPC charts.

Recently multivariate statistical techniques such as the principal component analysis (PCA) and the partial least squares or projection latent structures (PLS) have received increasing attention as alternative data analysis tools (Jackson, 1991; Geladi & Kowalski, 1986). This is due to their ability to efficiently handle a large number of highly correlated variables, measurement errors, and missing data. More important is their ability to provide an operator with useful information on process improvement by projecting the high-dimensional process data into the low-dimensional space defined by a few latent variables.

Based on these multivariate statistical techniques, several researchers have proposed various methodologies for monitoring the performance of a continuous process over time, which aims at detecting any abnormal situations due to the assignable (special) causes (Kresta, MacGregor, & Marlin, 1991; MacGregor, Jaeckle, Kiparissides, & Koutoudi, 1994; Wise & Gallagher, 1996). Then, after identifying assignable (special) causes for abnormalities, process improvements can be achieved through eliminating the causes by changing the process operating conditions. Nomikos and MacGregor (1994) used the multi-way principal component analysis (MPCA) as an extension of PCA, for treating batch or semi-batch processes. Monitoring charts based on MPCA are capable of tracking the progress of new batch runs and detecting the occurrence of observable upsets. Kourti, Nomikos, and MacGregor (1995) discussed the methodologies based on PCA and PLS for diagnosing the assignable (special) causes. The methods mentioned above have mainly focused on the detection and diagnosis of the assignable (special) causes through comparison with the recipe data, which represent normal operation trajectories. However, there are limitations for process improvement because such methods do not provide any information about the common causes, which can provide significant opportunities for process improvement.