*Korean J. Chem. Eng.*, **22**(1), 26-31 (2005)

# Real-Time Risk Monitoring System for Chemical Plants

**Ky-Soo Kim and Jae-Wook Ko†**

Kwangwoon University, 447-1, Wolgye-Dong, Nowon-Ku, Seoul 139-701, Korea
(*Received 12 November 2004 • accepted 11 January 2005*)

**Abstract**−This study was performed to develop a Real-Time Risk Monitoring System which helps to do fault detection using the information from plant information systems in a chemical process. In this study, to do fault detection, principal component analysis (PCA) methods of multivariate statistical analysis were used. The fundamental notions are a set of variable combinations, that is, detection of principal components which indicate the tendency of variables and operating data. Besides classical statistic process control, PCA can reduce the dimension of variables with monitoring process. Therefore, they are known as suitable methods to treat enormous data composed of many dimensions. The developed Real-Time Risk Monitoring System can analyze and manage the plant information on-line, diagnose causes of abnormality and so prevent major accidents. It's useful for operators to treat numerous process faults efficiently.

Key words: Real-Time Risk Monitoring System, Consequence Analysis, PCA, Fault Detection, Diagnosis

## INTRODUCTION

Chemical plants are composed of many units and various kinds of complex processes in order to produce a large amount of product. Because of recycle flow, reaction and vapor-liquid equilibrium, chemical processes show nonlinearity and complexity. The necessity of saving energy and materials is forcing the operating conditions to severe limits like high pressure, high temperature or extremely low pressure, temperature. In addition, usually the materials in chemical plants are toxic, flammable, explosive and dangerous, so operating chemical plant safely and keeping it under control is one of the most important issues in the chemical industry.

Two main approaches to keep chemical processes within the safe and economic conditions are offline methods and online monitoring. Offline methods include process modification, changing operating conditions etc. Real time process monitoring, control and variables estimation play an important role also. But it's not easy to meet these goals because of many obstacles in chemical processes including correlation between variables, nonlinearity, uncertainty and measurement delay.

Real-time risk monitoring system and hazard analysis have been developed separately, because their application levels are different, in the aspect of theoretical research and actual application. The reason why theyre different is that the former has been developed by design engineers and safety engineers focused on experience, while the latter has been developed by experience and technology of operators at the factory in operation. Currently, these two systems are implemented as automated systems, displaying the information of their target process with a model. Since these two systems use different methods, they use different process models to be implemented as the automated system; thus it's very hard to maintain compatibility between systems, implement user-friendly systems, create process models for system implementation, create/maintain databases, and maintain consistency of each model. However, in fact, they're based on common, basic knowledge - they share the information of behavior, structure, and material for process equipment. Therefore, we can get the benefits of eliminating unnecessary duplicated work, minimizing maintenance responsibility, effective information management and exchange, and user-friendliness through an integrated system. This integration has the same tendency as recent integration of the chemical factory across every field such as design, construction, and operation.

Correlation between variables is one of the most common problems in a chemical process. Sometimes different variables have the same meaning, because variables in a chemical process have strong correlation. For example, distillation column temperature at different trays shows the same behavior when there is a feed change or steam change. This variable correlation causes big problems during system identification or control model buildup.

To perform model identification or controller buildup, we usually solve a linear least square problem.

$$Y=XB+E$$
$$B=(X^TX)^{-1}X^TY \tag{1}$$

When there is strong correlation between variables, $X^TX$ doesn't have full rank. Namely, $X^TX$ becomes singular and the model parameter matrix, B, does not exist because there is no inverse matrix of $X^TX$. If the variables do not have so strong a correlation, $X^TX$ becomes a bad condition and the model based on B loses robustness, because a small change of X like noise of measurement error causes big change of Y. To solve this problem, variable selection or other methods are needed.

In this paper, PCA (principal component analysis) is used to overcome variable correlation and to build a robust model against noise or measurement error that is very popular in chemical processes.

## THEORY

### 1. Principal Component Analysis (PCA)

PCA was developed by Pearson in 1901 and used to analyze the relationship between variables by Hotelling. PCA is one of the multi-

†To whom correspondence should be addressed.
E-mail: jwko@daisy.kw.ac.kr

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1m} \\ a_{21} & a_{22} & \cdots & a_{2m} \\ \vdots & & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nm} \end{bmatrix}$$

**Fig. 1. Data matrix.**

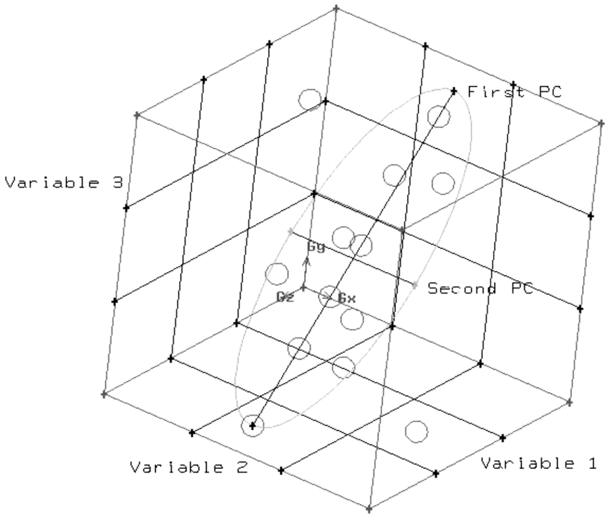



**Fig. 2. Algebraic interpretation of PCA.**

variate statistical analysis methods that can be used for feature extraction. All multivariate statistical analysis results are based on data, so the quality of data is very important. Fig. 1 shows a data matrix, where n rows mean observations or samples and the data matrix has m variables.

PCA can be understood at two different aspects, geometrical and algebraic.

Geometrically, PCA defines a new axis in data space and newly defined axis is called the principal component. PCA produces a mapping of the data set onto a new axis, defined by the span of a chosen subset of eigenvectors or loading vector, of the variance-covariance matrix of the data. Each new eigenvector captures the maximum amount of variability in the data in an ordered fashion. The first principal component explains the greatest amount of variation, the second, the next largest amount after removal of the first, and so on. PC's are orthogonal each other. The corresponding eigenvalue is associated with each eigenvector. Eigenvector $\lambda$ relates the amount of variance explained by that eigenvector. PCA is a special case of singular value decomposition. The projections onto the loading vector generate a set of scores that are linearly independent.

Algebraically, PCA represents matrix X with m rank as a linear combination of m matrix that has only one rank.

$$X=t_1p_1^T+t_2p_2^T+\cdots\cdots+t_kp_k^T+E \tag{2}$$

$t_k$=score vectors (m×1)
$p_k$=loading vectors (n×1)
E=residual matrix (m×n)

A subset of the first few scores provides information in a lower dimensional space, the score space, of the behavior of the process

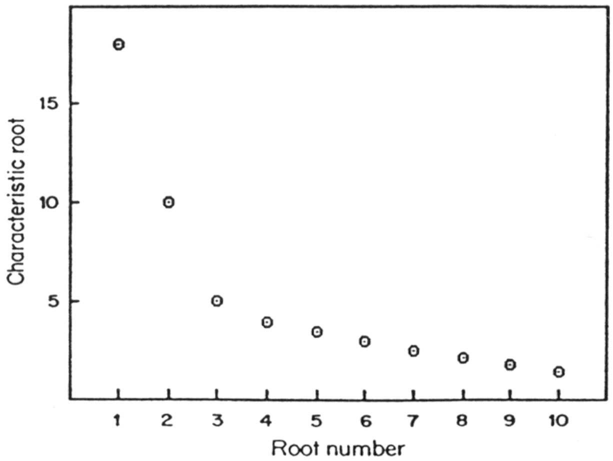



**Fig. 3. SCREE plot.**

during the period in which the measurements were made. This set of scores and the PCA loadings can be used to determine if the present process operation has changed its behavior relative to the data that were used to define the scores and loadings.

The application of PCA usually involves a prior step, which means centering scaling of the data. Mean centering implies that the average value for each variable is subtracted from the corresponding measurement. Scaling or normalization is necessary to avoid problems associated with some measurements having large values and others with small ones.

**Number of PC;** A scree plot is a method to identify how many PCs to be computed. It is useful in determining the appropriate number of components to interpret. The scree plot is a graph of the eigenvalues (representing the amount of variance explained) versus the number of factors. So as more factors are added, more of the variance is accounted for. After a certain number of factors, though, the curve flattens, which indicates that it is pointless to add more factors since not much more of the variance is being explained (see Fig. 3).

**2. Statistics**

There are several ways of interpreting the PCA results: the Q-statistic, a measure of the model mismatch; the Hotelling $T^2$-statistic, a measure of the fit of new observations to the model space; variance plots, a measure of the samples' variability; and score plots, a qualitative representation of the process performance, relative to the calibration model in the model space defined by the calibration model. The one used here is $T^2$ statistic.

**$T^2$ statistic;** The $T^2$-statistic measures unusual variability within the calibration model space. That is, if the calibration model data represent a process operation at one operating condition, and the process has shifted to a different one, then the $T^2$-statistic will show that data at this operating condition cannot be classified with the calibration data. The $T^2$-statistic is proportional to the sum of the squares of the scores on each of the principal components.

Using Eq. (3), the covariance matrix is calculated, where n is the number of observations and $\bar{x}$ is mean vector.

$$S=(n-1)^{-1}\sum_{i=1}^{n}(x_i-\bar{x})(x_i-\bar{x})^T \tag{3}$$

$T^2$ statistic for new observation x is given by Eq. (4)

$$T^2 = (x-\tau)^T S^{-1}(x-\tau) \tag{4}$$

where, $\tau$ is target value.

The confidence limits for the $T^2$-statistic can be calculated by means of the *F*-distribution

$$T^2 UCL = \frac{(n-1)(n+1)a}{n(n-a)} F_a(a, n-a) \tag{5}$$

**Contribution plot;** $T^2$ statistic is used for fault detection and contribution plot can be used to identify the variables that cause a process fault. Contribution plots decompose the scores into their summation operands and graph them versus the contributing variable. The summation operands are the products of the loadings of variable j and the corresponding value of variable j. A large product associated with a particular variable implies a correspondingly large contribution.

By comparing the contribution plot of a sample taken from the calibration set with one that is outside the confidence limits, differences in the expected variables' magnitude may provide an indication of which variables have exceeded their expected limits.

### 3. Quantitative Risk Assessment

Through quantitative risk assessment, a potential process hazard is identified and evaluated in quantitative value.

Fig. 4 shows the procedure of quantitative risk assessment. First, fault detection is performed to identify hazards in a certain process. This is followed by consequence analysis. In the consequence analysis step, a discharge model, dispersion model, fire & explosion model and effect model are used.

In this study, PCA is used to detect fault in process and risk assessment follows.

## RISK ASSESSMENT SYSTEM

### 1. System Overview

The system developed in this study is composed of plant information system (PIS), data preprocessing system, process monitoring & diagnosis system and quantitative risk assessment system. The whole structure of the developed system is shown in Fig. 5.

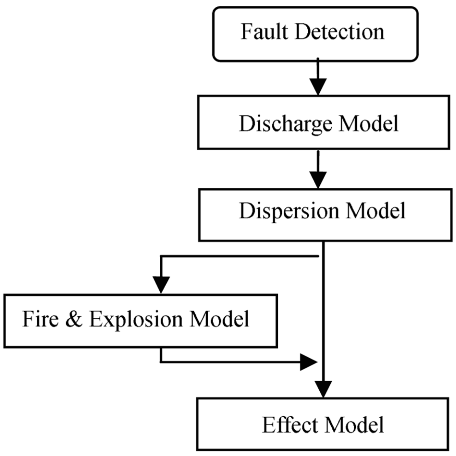



**Fig. 4. Logic diagram for consequence analysis of release of hazard materials.**

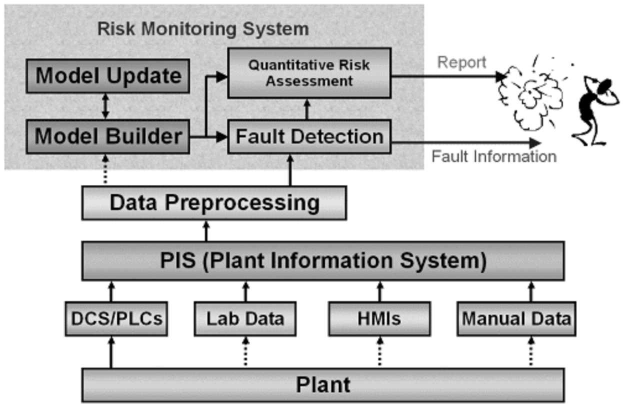



**Fig. 5. Whole system structure.**

Process data is gathered from PI$^{TM}$ of OSI. PI$^{TM}$ and linked with developed system by using API technique and ActiveX.

### 2. System Building

The following is the system development platform.

Visual Basic (Version 6.0)- Microsoft
NAG Library (Mark 18) - NAG
First Impression (Version 5.0) - Visual Components
Formula One (Version 5.0) - Visual Components

### 3. Principal Component Analysis (PCA)

Fig. 6. shows the PCA screen and Fig. 7 is the PCA algorithm. Process risk assessment is performed by using quantitative methods. The proposed system in this study takes process and fault related information from the database. Quantitative risk assessment then follows.

## CASE STUDY

Methylamine is one of the most important materials that is used as a raw material of an organic compound. Methylamine is pro-

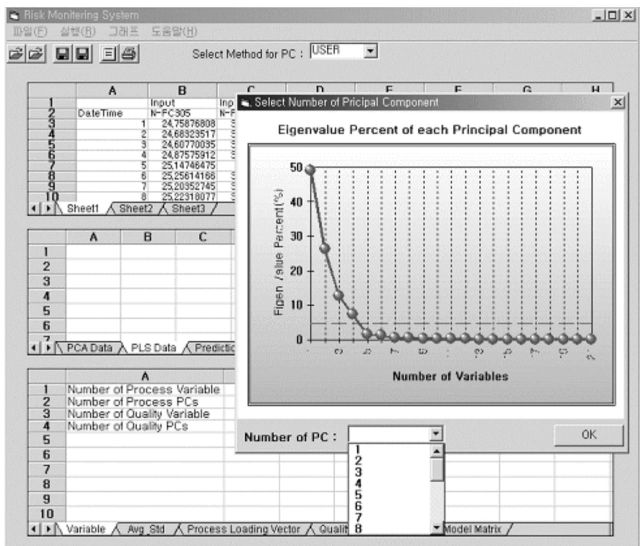



**Fig. 6. Principal component analysis screen.**

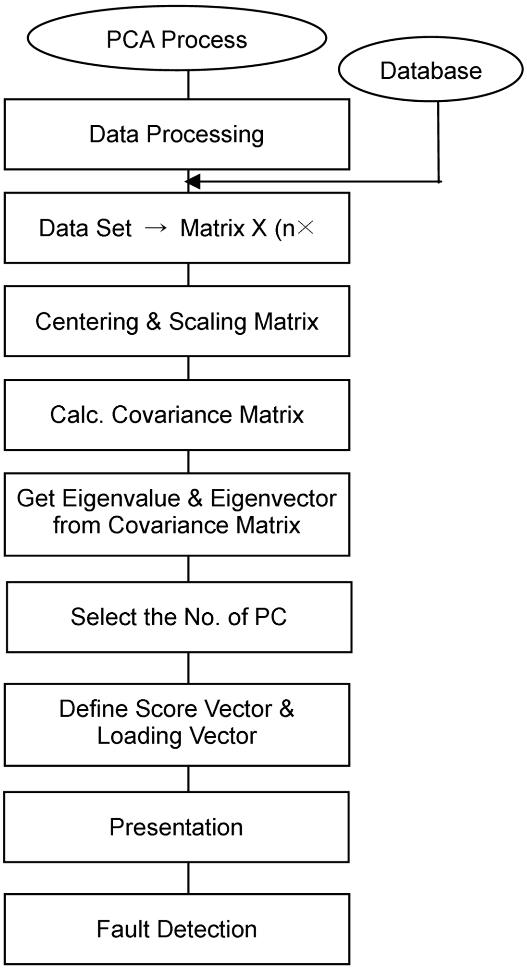



**Fig. 7. PCA algorithm.**

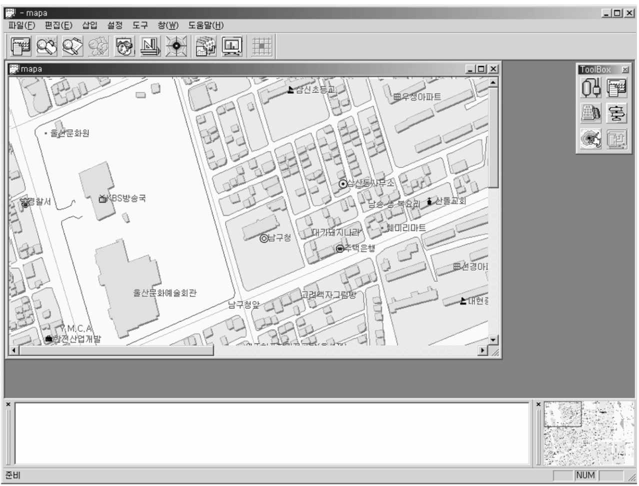


**Fig. 8. Screen capture of risk assessment tool.**

duced under 350-300 °C and 15-30 bar through the reaction of methanol and ammonia.

**1. Methylamine Process and Process Data**

Ammonia and methylamine related products separation is performed in this process. Process data is composed of 20 process variables and each variable has 575 observations. Every variable is preprocessed like outlier removal. The process variables are listed in Table 1.

**Table 1. Process input data (measurement data)**

| No | Description | No | Description |
|---|---|---|---|
| 1 | Vaporized $NH_3$ flow rate | 11 | Temperature 5 of converter |
| 2 | Level of gas separator | 12 | Reaction temperature of converter |
| 3 | $CH_3OH$ feed to converter | 13 | Reaction temperature of converter (set point) |
| 4 | Recycle liquid flow rate | 14 | Del P of $NH_3$ column |
| 5 | Pressure of gas separator | 15 | Bottom flow rate in $NH_3$ column |
| 6 | Temperature 1 of converter | 16 | Steam flow rate at heat exchanger 8 |
| 7 | Temperature 2 of converter | 17 | Water feed to $NH_3$ column |
| 8 | Temperature 3 of converter | 18 | $NH_3$ Feed to $NH_3$ column |
| 9 | Temperature 4 of converter | 19 | Pressure of $NH_3$ column |
| 10 | Temperature 5 of converter | 20 | Level of $NH_3$ column |

**Table 2. Model overview for MA plant**

| No. of PC | Eigenvalues | Percentage variation | Average value | Standard deviation |
|---|---|---|---|---|
| 1 | 6.41 | 60.66 | 5898.42 | 234.79 |
| 2 | 2.92 | 27.68 | 15979.77 | 659.32 |
| 3 | 0.90 | 8.55 | 3771.87 | 153.61 |
| 4 | 0.09 | 0.84 | 7414.05 | 307.67 |

**2. Data Analysis and Model Building**

Process data have 20 variables and 575 observations. Through principal component analysis we found that only 3 principal components among 20 variables contain almost 95% information about the MA process, so we decided to use 3 principal components. Table 2 shows the number of principal components and percentage variation of each PC. In order to do process monitoring, a time series plot, chart and contribution plot are used.

**3. Fault Detection and Risk Assessment**

After principal component analysis, we check whether the value exceeds the upper control limit using $T^2$ plot. In this study, 99% and 99.9% UCL are used.

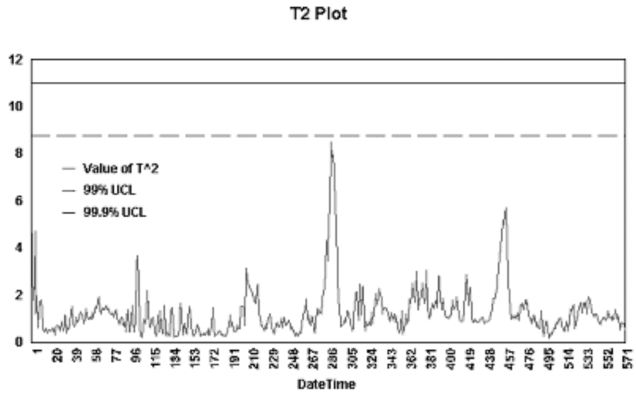
In Fig. 9(b), first and 286$^{th}$ observations show a process fault. The cause of the process fault is NH3 feed to NH3 column and water feed to NH3 column. Fig. 10 shows the contribution plot of the first and 286$^{th}$ observation.

Quantitative risk assessment is performed under the scenario that guesses ammonia discharge from an ammonia column. The temperature and pressure of the ammonia feed is 20 and 20 bar, respectively. Total discharge amount is 400 kg.

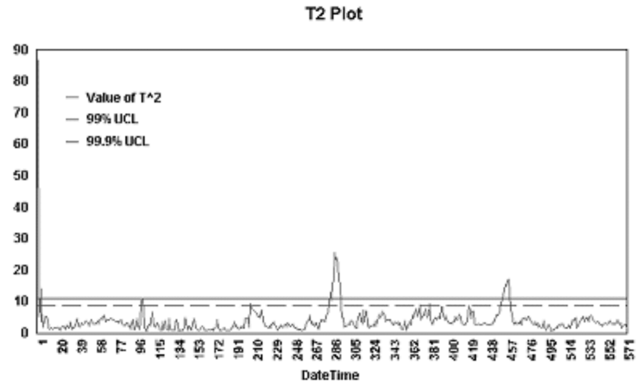
Table 3 shows the risk assessment result.

## CONCLUSIONS

In this study, fault detection is performed by using principal component analysis followed by risk assessment. Principal component

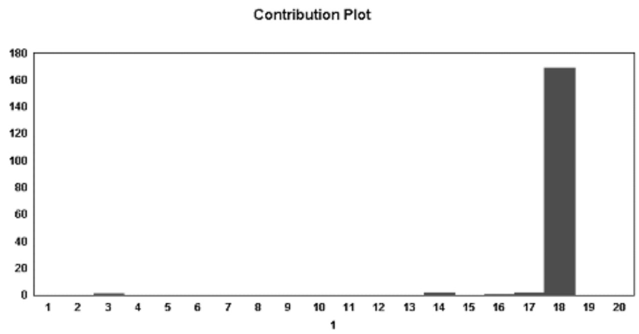



(a) Normal operation

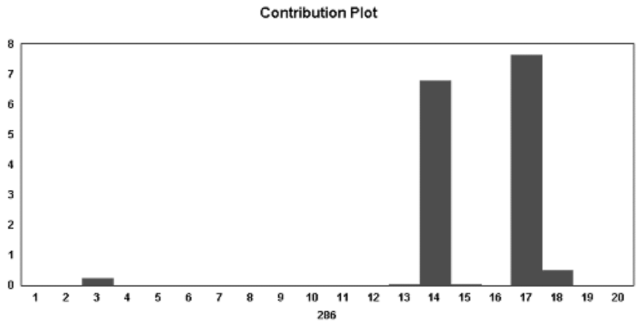



(b) Abnormal operation

**Fig. 9. $T^2$ plot of MA plant.**




(a) Contribution plot of the 1st observation




(b) Contribution plot of the 286th observation

**Fig. 10. Contribution plot of MA plant.**

**Table 3. Effect distance for wind speed and weather stability**

| Wind speed (m/sec) & stability | Effect distance for 1/2 LFL |
|---|---|
| 1.5/F | 13.3 |
| 1.5/D | 14.4 |
| 5/D | 26.0 |

analysis is a useful data analysis method when there is strong correlation between variables. The methylamine process is chosen as a case study process.

Because PCA is based on process data, the quality of process data is very important. So, first, data preprocessing, like outlier removal, is performed before principal component analysis. After data preprocessing, normal process data is used to calculate loading vector and score vector. Offline model building, monitoring and fault detection is the base of online implementation.

The number of PCs is determined based on a scree plot. $T^2$ statistic and contribution plot are used to detect process faults and to isolate fault variables. The quantitative risk assessment step is followed by fault detection. Discharge, dispersion, fire & explosion modules are included in the risk assessment system.

As the system developed through this study was applied to the methylamine plant, there was a significant decrease of ammonia and water injection volume in the ammonia column for abnormal operation, showing instability; thus $\Delta P$ of the ammonia column became abnormal. And we performed quantitative risk analysis assuming the decrease of ammonia feed flow rate as leakage, to find the length from a minimum 13 meters to maximum 26 meters is in the range of influence for 1/2 LFL. Therefore, it is necessary to consider appropriate safety actions for potential hazards.

From now on, the following research must proceed so the system developed through this research can be utilized for application to an operator training system and a means of case study for design and equipment change. As the effort to integrate and automate various fields of chemical process systems is in progress now, we also have to proceed with other research to support this effort. Especially, a close connection with safety and process design field is required.

## ACKNOWLEDGEMENT

The present research has been conducted by the Research Grant of Kwangwoon University in 2003 and supported by MOST in 2003-2004.

## NOMENCLATURE

B : model parameter matrix [m×n]
E : residual matrix [m×n]
$F_a$ : F distribution
$p_k$ : loading vectors [n×1]
S : sample covariance matrix
$t_k$ : score vectors [m×1]
$T^2$ : $T^2$-statistic
$T^2_{UCL}$ : upper control limit of the $T^2$-statistic
$x_i$ : row vector including $i^{th}$ event

$\bar{x}$ : mean vector of $x_i$'s
X : process variables matrix
Y : responses matrix

**Greek Letters**

$\lambda$ : eigenvector
$\tau$ : target value

## REFERENCES


AIChE/CCPS, *Guidelines for Use of Vapor Cloud Dispersion Models*, Center for Chemical Process Safety, AIChE, New York (1987).

AIChE/CCPS, *Plant Guidelines for Technical Management of Chemical Process Safety*, Center for Chemical Process Safety, New York (1995).

Dunia, R. S., Qin, J., Edgar, T. F. and McAvoy, T. J., "Identification of Faulty Sensors Using Principal Components Analysis," *AIChE J.*, **42**(10), 2797 (1996).

Jackson, J. E., "Principal Components Analysis: Part 1 Principal Analysis," *J. Qual. Technol.*, **12**, 201 (1980).

Kim, K. H., Shin, D. and Yoon, E. S., "Risk Analysis Using Automatically Synthesized Robust Accident Scenarios and Consequence Assessment for Chemical Processes: Process Partition and Consequence Analysis Approach," *Korean J. Chem. Eng.*, **20**, 992 (2003).

Kim, K. S., "Development of the Real-time Risk Monitoring System for Chemical Plants," PhD Dissertation, Kwangwoon University, Seoul (2000).

MacGregor, J. F. and Kourti, T., "Statistical Process Control of Multivariable Processes," *Contr. Eng. Pr.*, **3**(3), 403 (1995).

Park, M. J. and Doyle, F. J., "Nonlinear Model Order Reduction and Control of Particle Size Distribution in a Semibatch Vinyl Acetate/Butyl Acrylate Emulsion Copolymerization Reactor," *Korean J. Chem. Eng.*, **21**, 168 (2004).

Paul Geladi, Bruce R. Kowalski, "Partial Least Square Regression: A Tutorial," *Analytical Chemical Acta*, **185**, 1 (1986).

Sohn, J. H. and Jung, S. H., *Non-linear PLS Based on Non-linear Principal Component Analysis and Neural Network*, KACC (2000).