



## 주성분분석(PCA) 기법에 기반한 CNG 충전소의 이상감지 모니터링 및 진단 시스템 연구

이기준 · 이봉우 · 최동황 · 김태옥\* · †신동일\*

한국소방산업기술원, \*명지대학교 화학공학과

(2014년 5월 11일 접수, 2014년 6월 20일 수정, 2014년 6월 23일 채택)

## A Study on Fault Detection Monitoring and Diagnosis System of CNG Stations based on Principal Component Analysis(PCA)

Kijun Lee · Bong Woo Lee · Dong-Hwang Choi · Tae-Ok Kim\* · †Dongil Shin\*

Korea Institute of Fire Industry & Technology, Yongin 446-909, Korea

\*Department of Chemical Engineering, Myongji University, Yongin 449-728, Korea

(Received May 11, 2014; Revised June 20, 2014; Accepted June 23, 2014)

### 요약

본 연구에서는 비정상상태 운전을 기본으로 하는 CNG 충전소를 대상으로 다변량 통계분석방법 중의 하나인 다차원의 대용량 데이터 처리에 적합한 주성분분석(PCA) 기법을 사용하여 실시간 이상감지 및 진단이 가능한 모니터링 시스템을 제안하였다. CNG 충전소로부터 매초 간격으로 수집되는 7개의 압력센서 데이터와 5개의 온도센서 데이터의 주요 경향을 나타내는 변수들의 조합으로 주성분이라 불리는 새로운 특성 변수들을 산출하고, 분산의 분포를 통해 특성변수의 계산으로부터 모델을 구축하였다. 모니터링은 구축된 모델을 통해 운전 중의 실시간 데이터를 반영하여 진행된다. 시스템 검증 및 정확성을 개선하기 위해 모니터링 테스트를 수행한 결과, 정상상태의 모든 데이터를 정상으로 판단하였고, 이상 데이터의 성공적인 검출 시 관련 변수를 추적하여 비정상 원인을 찾아낼 수 있었다.

**Abstract** - In this study, we suggest a system to build the monitoring model for compressed natural gas (CNG) stations, operated in only non-stationary modes, and perform the real-time monitoring and the abnormality diagnosis using principal component analysis (PCA) that is suitable for processing large amounts of multi-dimensional data among multivariate statistical analysis methods. We build the model by the calculation of the new characteristic variables, called as the major components, finding the factors representing the trend of process operation, or a combination of variables among 7 pressure sensor data and 5 temperature sensor data collected from a CNG station at every second. The real-time monitoring is performed reflecting the data of process operation measured in real-time against the built model. As a result of conducting the test of monitoring in order to improve the accuracy of the system and verification, all data in the normal operation were distinguished as normal. The cause of abnormality could be refined, when abnormality was detected successfully, by tracking the variables out of the score plot.

**Key words** : fault detection, monitoring, diagnosis, principal component analysis (PCA), CNG station

†Corresponding author: dongil@mju.ac.kr

Copyright © 2014 by The Korean Institute of Gas

## 1. 서론

최근 석유공급에 대한 국제정세의 불안 및 에너지 수급의 불안정으로 국제 유가가 폭등하고 있으며, 이에 따라 친환경 및 신재생 에너지 사용에 관한 많은 연구가 이루어지고 있다. 특히, 친환경 에너지원 중에서 대표적인 압축천연가스(compressed natural gas, CNG)는 가정용, 산업용 및 발전용으로 많이 이용되고 있을 뿐만 아니라 국내 에너지 사용량의 10% 이상을 차지하는 중요한 에너지원으로 떠오르고 있다. 그러나 이와 같이 원활하게 이루어지고 있는 CNG 차량 및 CNG 충전소의 보급과 기술 발전에 비해 공정안전기술의 발전은 매우 미비한 실정이다. 더구나 최근 5년간 13건 이상의 크고 작은 사고 발생으로부터 확인할 수 있듯이 실시간 모니터링 시스템과 같은 사고예방 관점의 공정안전기술에 대한 연구가 절실히 필요한 상황이다.

CNG 충전소는 비반복적으로 CNG 차량의 충전이 이루어지고, 소비된 만큼의 CNG를 다단 압축을 통해 저장하는 형태로, 일정한 가동시간 이후에 안정화 상태를 갖게 되는 일반적인 화학공정의 운전형태와는 다르다. 따라서 CNG 충전소는 비반복적인 비정상상태의 전이구간이 계속적으로 존재하여 기존의 모니터링 기법으로는 이상감지 및 이상요인 검출이 제한되며, 설정된 임계치를 초과하여 위험정보

를 습득하더라도 그때는 이미 사고 발생을 막을 수 없는 상황이거나, 유지보수가 불가능한 경우가 대부분이다.

이상진단 모니터링 방법은 Fig. 1과 같이 공정의 정량적인 기존 모델에 기반한 기법, 정성적인 정보를 기초로 하는 기법과 통계학적 모델을 이용하는 기법으로 크게 세 가지 기법으로 구분된다[1].

정량적인 모델 기반 모니터링은 모델에 관련된 변수들이 불확실하고, 정확한 모델이 실시간으로 개선되어야 한다는 단점이 있다. 그리고 정성적인 모델 기반 모니터링은 정확한 수식 모델이 필요하지 않다는 장점이 있으나, 과거 지식 경험이나 전문가 또는 조업자의 경험적 지식이 바탕이 되어 주관적인 측면이 강하고, 다량의 데이터를 해석하기 위한 통계적 기반이 부족하다는 단점이 있다. 이에 비해 공정 데이터를 기반으로 통계적 기법을 접목시킨 통계학적 모델 기반 모니터링의 경우는 다른 방법들 보다 비교적 쉽게 모니터링이 가능하고, 공정 데이터의 분석도 가능하다는 장점을 가지고 있다. 특히, 통계학적 모델 기반 모니터링 기법은 단변량 통계와 다변량 통계 기법으로 구분할 수 있으며, 이는 해당 공정에 확률적으로 반응을 가지는 변수의 개수를 기준으로 구분하게 된다. 단변량 통계 모니터링 기법은 단순공정에 대해 비교적 간단하고 정확한 모니터링이 가능하지만, 실제 공정은 공정변수가 다양으로

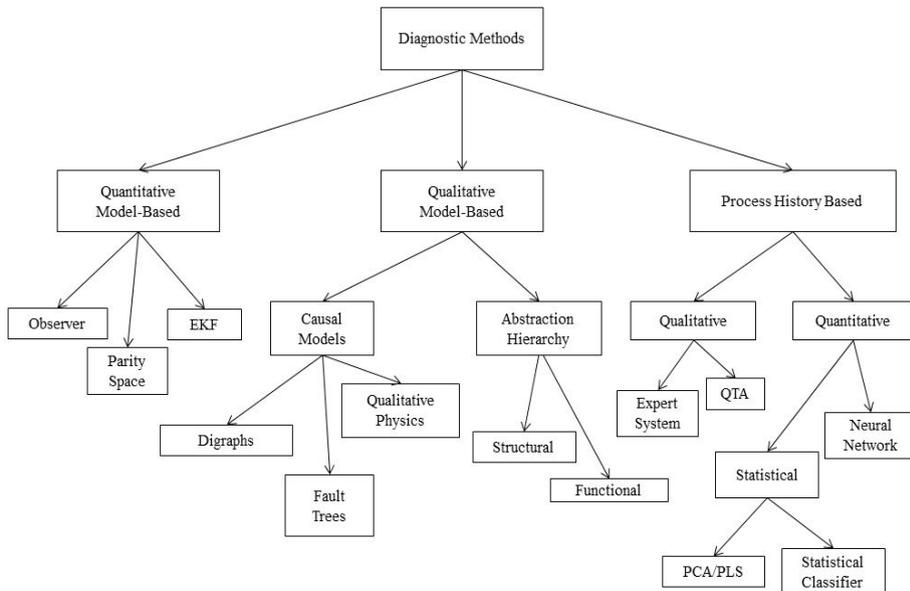


Fig. 1. Classification of diagnostic algorithms of process fault detection and diagnosis[1].

존재하는데도 불구하고, 변수간의 상관관계를 고려하지 않고 모니터링하기 때문에 복잡한 화학공정이나 본 연구의 대상인 CNG 충전소와 같이 비정상상태로 운전하는 공정에 대해서는 이상 데이터의 감지능력이 떨어지고, 적합하지 못하다. 따라서 공정변수의 상관관계를 고려하여 공정을 모니터링하는 다변량 통계분석을 통한 공정 모니터링 시스템의 연구와 개발이 지속적으로 이루어지고 있다[2-4].

본 연구에서는 비정상적인 전이구간이 존재하는 CNG 충전소의 공정운전에 대하여 다변량 통계 기법중 하나인 주성분분석(PCA)을 이용하여 모니터링을 진행하였다.

## II. 주성분분석(PCA) 이론

PCA의 기본원리는 데이터의 축소, 즉 차원의 감소에 있다. 많은 양의 데이터 중 불필요하거나, 크게 비중을 차지하지 않는 데이터를 제거함으로써 본 데이터의 성향에서 크게 벗어나지 않는 데이터의 조합을 분석하는 것이다. 즉, 원형변수들을 통해 만들어진 선형 조합에서 주요 경향이나 추세 등을 설명하는 요인들이나 변수들의 조합들을 찾아내어 주성분으로 불리는 새로운 특성변수들을 산출하고, 분산의 분포를 통해 특성변수의 계산으로 모델을 구축하는 것이 주목적이다.

Fig. 2는 차원 감소의 원리를 보여주는데, 수집된 샘플에 대해 3가지 측정된 변수들의 값들을 보여주고 있다. 3차원으로 데이터를 표현하고, 2개의 PC를 사용해서 데이터를 표현하였으며, 이때 샘플들이 타원형의 형태로 원 안에 들어가는 것을 확인할 수 있다. 이러한 과정을 거쳐 작은 분산 값을 갖는 주성분

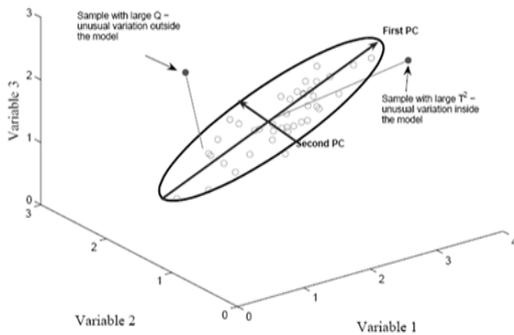


Fig. 2. Graphical representation of principal components analysis[5].

을 제거하게 되면 원래 데이터와의 상관관계도 크게 잃지 않으면서 원형변수들의 개수보다 적은 수의 특성변수들을 이용해 데이터를 얻을 수 있게 된다. 이와 같이 고차원의 데이터에서 저차원의 특성변수로 데이터의 차원을 줄이면서 고차원 문제를 해결하는 것이 PCA의 원리이다[5].

주성분은 데이터,  $X = [X_1, X_2, \dots, X_m]$ 는 스코어벡터  $t_i$ 와 로딩벡터  $P_i$ 의 곱과 잔차행렬  $E$ 의 합으로 나타낼 수 있다. 데이터,  $X = [X_1, X_2, \dots, X_m]$ 가 존재하고,  $N$ 개의 측정값  $k = 1, 2, \dots, N$ 이 새로운 행렬에 들어간다고 할 때  $T = [t_1, t_2, \dots, t_r]$  또한  $N$ 개의 측정값을 갖게 된다. 그러나 차원은  $r < m$ 이 된다. 이것을 변형행렬,  $P$ 를 이용해서 식 (1)과 같이 나타낼 수 있다.

$$T_{[N \times r]} = X_{[N \times m]} P_{[m \times r]} \quad (1)$$

여기서  $P$ 는  $P = [P_1, P_2, \dots, P_r]$ 로 나타낼 수 있으며, 이 변환형태는 직교함수이기 때문에 이것은  $P^T P = I$ 의 성질을 갖고 있으며,  $X = TP^T$ 로 나타낼 수 있다.

$$\begin{aligned} X &= t_1 P_1^T + t_2 P_2^T + \dots + t_k P_k^T + E \\ &= TP^T + E \end{aligned} \quad (2)$$

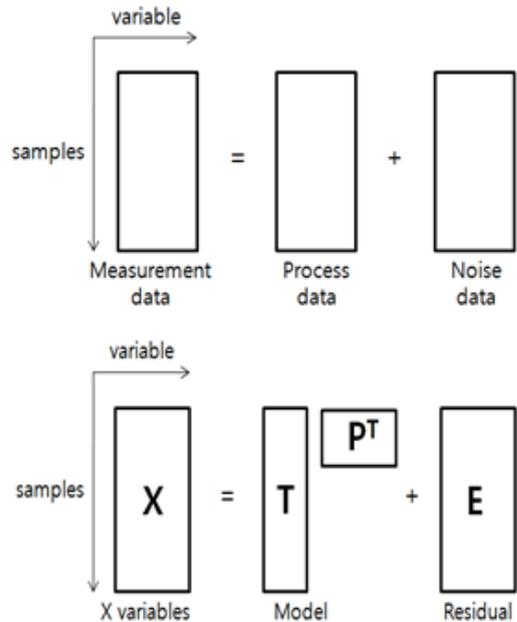


Fig. 3. Measurement data from superposition of process data and noise data[6].

즉,  $m$ 개의 변수와  $n$ 개의 샘플수로 이루어진 행렬,  $X \in R^{n \times m}$ 에 의해  $X$ 를 표현한 식은 식 (2)와 같으며, Fig. 3을 통해 원리를 알 수 있다[6].

### III. Real-Time Database (RTDB)를 이용한 예측모델 개발

#### 3.1. 데이터의 필터링 및 스케일링

공정운전의 경향이 모두 반영될 수 있도록 매 초당 7개의 압력센서 데이터와 5개의 온도센서 데이터를 바탕으로, 일주일치 데이터를 사용하여 모델 개발 및 학습시켰다.

Fig. 4는 개발된 시스템의 구동 모습이며, PI 서버를 이용하여 데이터를 수집하고 있는 상황을 보여주고 있다. 모델을 구축하기 전에 Fig. 5와 같이 Start-up 구간과 Shut-down 구간의 제거로 운전 중의 데이터만 사용할 수 있도록 필터링 과정을 거치게 되

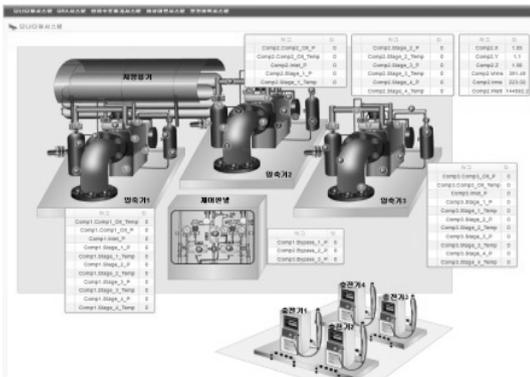


Fig. 4. Screenshot of the developed system.

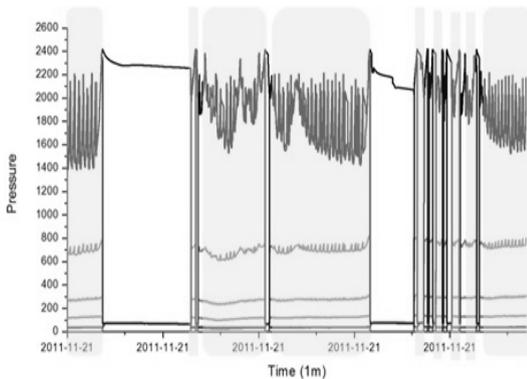


Fig. 5. Filtering of the pressure data.

며, 필터링 된 데이터를 바탕으로 스케일링을 진행한다. 변수의 단위가 달라서 정확한 모델을 구축할 수 없기 때문에 평균과 표준편차를 이용하여 스케일링하는 전처리 과정이 필요하다[7-9].

#### 3.2. PC 수의 선정

PCA의 장점인 데이터 축소를 위해서는 주성분(PC)의 수를 선정해야 한다. PC의 수를 선정하는 방법에는 Table 1에서와 같이 eigenvalue와 variance captured total을 이용하여 구하는 average root 방법과 그래픽적인 방법으로 Fig. 6과 같이 근의 값을 가로와 세로축에 위치시키면서 covariance matrix의 특성근을 모두 도면에 표시하는 SCREE test 방법이 있다. 일반적으로 variance captured total이 85%

Table 1. Percent variance captured by PCA model

Percent variance captured by PCA model			
Principal component	Eigenvalue of COV(X)	% Variance this PC	% Variance cumulative
1	3.39e+000	48.43	48.43
2	2.55e+000	36.43	84.86
3	8.98e-001	12.84	97.70
4	1.18e-001	1.69	99.39
5	2.42e-002	0.35	99.73
6	1.54e-002	0.22	99.95
7	3.45e-003	0.05	100.00

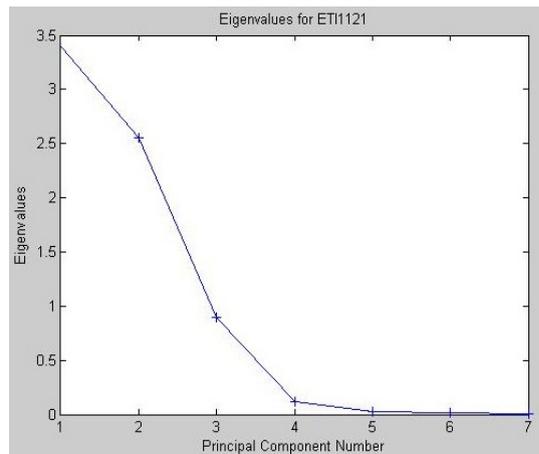


Fig. 6. SCREE plot of scaled CNG station data.

를 만족시키는 것을 PC 수로 선정하지만[6,10], 본 연구에서는 두 가지 경우를 모두 고려하여 variance captured total이 90% 이상을 만족하도록 PC의 수를 3개로 선정하였다[11].

### 3.3. 모델 구축

일주일치 데이터의 필터링 과정과 스케일링 과정을 통해 형성된 데이터 매트릭스의 주성분 분석을 통해 모델을 구축하였다. 그 결과, Fig. 7를 통해 구축된 모델을 확인할 수 있으며, 95% 신뢰한계를 보여주고 있다. 그림에서 신뢰한계를 넘어서는 값들은 Fig. 8과 같이 잔차의 분석으로 반복되어 넘지 않음을 확인할 수 있고, 이는 노이즈로 판단할 수 있다[12,13].

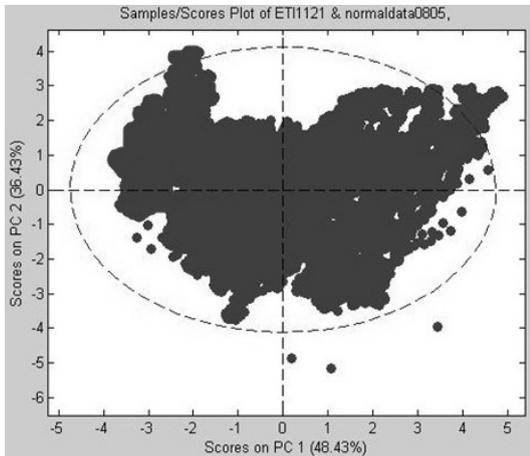


Fig. 7. Score plot of PCA model.

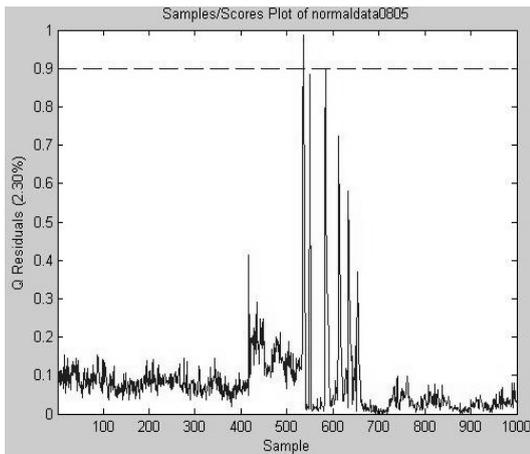


Fig. 8. Q residuals of model data.

## IV. 모니터링 진행 및 결과

구축된 모델에 실시간으로 측정되는 공정운전 중의 데이터를 투영함으로써 Fig. 9와 같이 모니터링이 진행되는데, 시스템의 검증과 정확성을 향상시키기

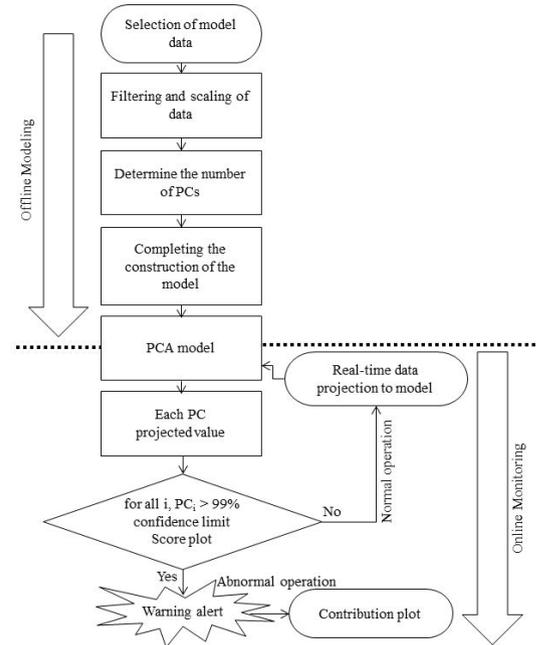


Fig. 9. Flowchart of the real-time monitoring system.

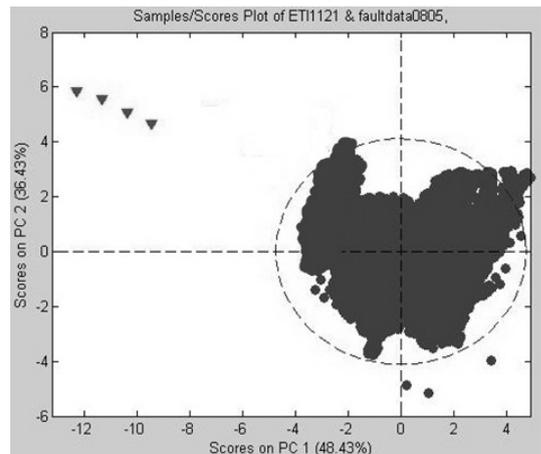


Fig. 10. Plot of detecting abnormality.

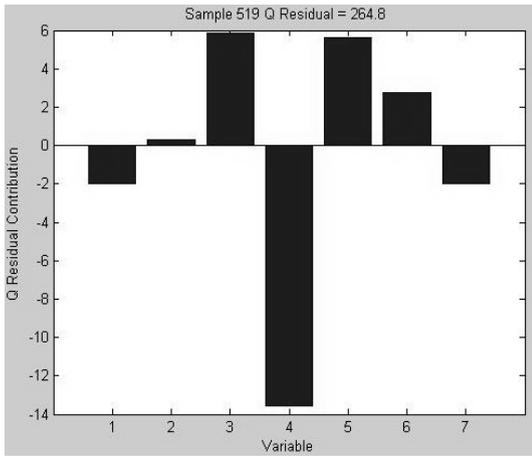


Fig. 11. Diagnosing cause of abnormality.

위해 오프라인 모니터링을 진행하였다. 모델의 학습에 이용되지 않은 과거 운전데이터를 이용해 모니터링을 진행해 본 결과, 정상운전 중의 모든 데이터를 정상으로 판별하였고, 이상 데이터의 부재로 가공된 이상 데이터를 투영하였을 때 Fig. 10과 같이 이상을 감지하고, Fig. 11과 같이 이상현상을 직접적으로 일으키는 변수의 추적으로, 이상의 원인이 되는 센서를 파악할 수 있었다. 특히, Fig. 11에 나타난 7개 변수는 압축기에 설치된 압력센서들의 이상정보를 나타내고 있으며, 그 중에서 가장 큰 값을 나타내고 있는 4번 센서는 압축기 출구 쪽에 달려있는 압력센서로, 그 부근의 이상을 감지하고 있다.

## V. 결론

비반복적인 비정상상태 전이구간이 존재하는 CNG 충전소 공정운전에 대해 일주일간의 데이터의 수집으로 모니터링 시스템을 구축하고, 모델 구축에 이용되지 않은 3개월간의 데이터를 통해 이를 검증하였다. 실질적인 이상데이터의 부재로 가공된 이상 데이터와 과거 정상 데이터의 투영을 통해 오프라인 모니터링을 진행결과, 모든 이상을 판별하고, 그 원인의 추적이 가능하였다. 이를 통해 기존의 모니터링 방법으로 한계가 있었던 비정상상태의 전이구간이 반복되는 CNG 충전소의 공정운전에 대한 모니터링의 가능 여부와 그 성능을 확인할 수 있었다.

따라서 이와 같은 결과를 바탕으로 계절변화 등의 주위 환경조건 변화를 반영하여 고정되어 있는 모델의 형태가 아닌 계속적으로 진보하는 모델을 탑재한 모니터링 시스템을 통해 CNG 충전소뿐만 아

니라 이와 유사한 형태를 지닌 비정상상태 운전을 기본으로 하는 공정의 모니터링 시스템에 적용이 가능할 것으로 판단된다.

## 감사의 글

본 연구는 소방방재청 인적재난안전기술개발사업의 지원으로 수행한 '유해화학물질 효과적 대응을 위한 방재기술연구' [NEMA-인적\_2014-41]과제의 성과입니다.

## REFERENCES

- [1] Himmelblau, D. M., *Fault Detection and Diagnosis in Chemical and Petrochemical Processes*, Elsevier, (1978)
- [2] Kourti, T. and MacGregor, J. F., "Process Analysis, Monitoring and Diagnosis Using Multivariate Projection Methods", *Chemometrics and Intelligent Laboratory System*, **28**(1), 3-21, (1995)
- [3] Eastment, H. T. and Krzanowski, W. J., "Cross-Validatory Choice of the Number of Components from a Principal Component Analysis", *Technometrics*, **24**(2), 73-77, (1982)
- [4] Wise, B. M., *PLS\_Toolbox 3.5 Manual*, Eigenvector Research, (2004)
- [5] Wise, B. M. and Gallagher, N. B., "The Process Chemometrics Approach to Process Monitoring and Fault Detection", *J. Process Control*, **6**(6), 329-348, (1996)
- [6] Yoo, C. K., Choi, S. W., and Lee, I. B., "Recent Research Trends of Process Monitoring Technology: State-of-the Art", *Korean J. Chem. Eng.*, **46**(2), 233-247, (2008)
- [7] Yang, J. M., Kim, B. S., Yong, J. W., Ko, B. S., Lee, D. H., and Ko, J. W., "A Study on Safety and Operational Management System for CNG Filling Stations", *Journal of the Korean Institute of Gas*, **15**(6), 8-13, (2011)
- [8] Lee, H. S., Lee, D. H., Yang, J. M., and Ko, J. W., "A Study on Application of USN in CNG Station", *Journal of the Korean Institute of Gas*, **15**(4), 56-61, (2011)
- [9] Hyosung Corporation, *CNG compressor package manual*, System Ver. 1.0
- [10] Manabu, K., Shinji, H., and Iori, H., "A New Multivariate Statistical Process Monitoring Me-

- thod Using Principal Component Analysis", *Computers and Chemical Engineering*, **25**, 1103-1113, (2001)
- [11] Lee, Y. T., *Monitoring and Diagnosis of Non-Steady Transient Process Operations*, M.S. Thesis, Myongji University, (2011).
- [12] Lee, K. J., *Quantitative Simulation-based Safety Analysis of Energy-filling Facilities for Improved Sustainability*, M.S. Thesis, Myongji University, (2014)
- [13] Dan, S., Moon, D. J., Yoon, E. S., and Shin, D., "Analysis of Gas Explosion Consequence Models for the Explosion Risk Control in the New Gas Energy Filling Stations", *Journal of Industrial & Engineering Chemistry Research*, **52**, 7265-7273, (2013)