KIGAS Vol. 28, No. 2, pp 38~46, 2024
(Journal of the Korean Institute of Gas)
https://doi.org/10.7842/kigas.2024.28.2.38

도시가스 배관망의 고장 탐지 및 진단을 위한 다중블록 PCA 적용 연구

백연주 · 이태룡* · 김종선* · †고홍철
(주)테크다스, *(주)해양에너지
(2024년 2월 1일 접수, 2024년 4월 25일 수정, 2024년 5월 31일 채택)

Multi-block PCA for Sensor Fault Detection and Diagnosis of City Gas Network

Yeon-ju Baek · Tae-Ryong Lee* · Jong-Seun Kim* · †Hong-Cheol Ko
Techdas Ltd., Seoul 08520, Korea
**Haeyang Energy Co. Ltd., GwangJu 62209, Korea*
(Received February 1, 2024; Revised April 25, 2024; Accepted May 31, 2024)

요 약

도시가스 배관망은 넓은 지역에 분포되며, 계층적으로 복잡하게 연결되어 있는 것을 특징으로 한다. 넓은 지역에 분포된 배관망의 압력 상태를 정밀하게 감시하기 위하여 PCA 기반의 센서 고장탐지 및 진단 기법을 적용하였다. 도시가스 배관망은 센서의 수가 많기 때문에 상호 연관성이 높은 센서들을 하나의 블록으로 묶어 블록 차원에서 전체 센서를 다루는 다중블록 PCA(MBPCA)가 추천된다. 그런데 MBPCA는 고장센서 식별 성능이 우수한 장점이 있지만, 센서의 개수가 증가할수록 고장탐지 성능이 나빠지는 경우가 많고, 지엽적인 변화에도 모델 전체를 갱신해야 하는 문제점이 있다. 이러한 이유로 모든 센서에 대해 MBPCA를 적용하는 것보다 선택적으로 MBPCA를 적용하는 것이 필요하다. 본 연구에서는 고장탐지 성능(fault detectability) 지수와 고장진단 성능(fault identifiability) 지수를 제안하였으며, 이를 통해 블록별로 MBPCA와 PCA의 성능을 비교할 수 있도록 하였다. 이 지수를 기준으로 MBPCA와 PCA 블록을 구분하여 (주)해양에너지의 도시가스 배관망의 고장 탐지 및 진단 시스템을 개발하였으며, 센서의 개수가 많은 경우에 발생하는 문제점을 해결할 수 있었다.

Abstract - The city gas pipeline network is characterized by being widely distributed and hierarchically connected in a complex manner over a wide area. In order to monitor the status of the widely distributed network pressures with high precision, Multi-block PCA(MBPCA) is recommended. However, while MBPCA has excellent performance in identifying faulty sensors as the number of sensors increases, the fault detection performance deteriorates, and also there is a problem that the model needs to be updated entirely even if minor changes occur. In this study, we developed fault detectability index and fault identifiability index to determine the effectiveness of MBPCA application block by block. Based on these indices, we distinguished MBPCA and PCA blocks and developed a fault detection and diagnostic system for the city gas pipeline network of Haean Energy Co., Ltd., and were able to solve the problems that arise when there are many sensors.

Key words : city gas network, sensor validation, multi-block PCA, fault detection and diagnosis

†Corresponding author:kohc@techdas.co.kr
Copyright ⓒ 2024 by The Korean Institute of Gas

Ⅰ. 서 론

전력, 수도, 가스, 교통, 통신 등의 주요기반시스템(Critical Infrastructure System(CIS))은 다수의 구성 요소들이 넓은 지역에 분산되어 계층적으로 복잡하게 연결되어 있는 것을 특징으로 한다. 이러한 시스템의 구성 요소들은 망의 형태로 연결되어 서로 유기적으로 영향을 주고받는다. CIS를 정교하게 운영하려면 많은 센서로부터 수집되는 다양한 형태의 데이터를 신속하게 처리하여 정확한 정보를 제공하는 것이 중요하다.

CIS 분야의 하나인 도시가스 배관망은 정부의 2030 온실가스 감축 목표 및 2050 탄소제로 정책으로 친환경 에너지의 수요 증가에 따라 그 규모가 해마다 증가하고 있다. 도시가스 시스템은 천연가스를 압축하고 저장하는 시설, 압축된 가스를 각 지역으로 운송하는 배관, 최종 수요가에게 공급하기 위해 적정 압력으로 감압하는 시설, 수요가에게 공급하는 배관으로 구성되어 있다. 전국적으로 도시가스 시스템은 2023년 현재 34개소가 운영되고 있다. 그런데 우리나라 도시가스 배관의 40%는 설치된 지 20년이 경과되어 배관 부식이 우려되고 있다. 특히 지중에 매설되어 있기 때문에 세부 관리가 어렵고 다양한 위험에 노출되어 있다[1]. 도시가스 배관망의 안전성 향상을 위해서 실시간으로 고장 발생을 감시하고, 고장의 위치를 식별하여 제공하는 고장탐지 및 진단 시스템이 필요하다[2].

본 연구에서는 도시가스 배관망의 고장탐지 및 진단 기법으로 주성분분석법(Principal Component Analysis(PCA))을 적용하였다. PCA는 고장탐지 및 진단에 가장 널리 사용되고 있는 기법이다[3]. PCA는 다수의 변수 데이터를 선형 결합하여 소수의 주성분으로 표현하는 통계적 모델링 기법으로 복잡한 시스템을 구성하는 데 장점이 있다. PCA 모델로부터 도출된 수식을 이용하여 고장탐지와 고장진단을 순차적으로 실행한다. 고장탐지는 SPE(Squared Prediction Error) 지수나 Hotelling T^2, 또는 SPE와 T^2의 융합지수가 기준치를 초과하는지 여부로 판단한다. 고장진단은 고장탐지 지수를 센서별로 분할하고, 가장 큰 값을 갖는 센서를 고장으로 식별한다. SPE 지수를 센서별로 분할한 값을 센서의 이상기여도(contribution)라 부른다.

이상기여도에는 고장센서 식별을 어렵게 하는 스며듦 현상(smearing effect)이 있다. 스며듦 현상이란 고장센서 측정값이 정상센서의 이상기여도를 왜곡시켜 정상센서가 고장으로 오인되는 현상이다. Alcala와 Qin[4]은 스며듦 현상을 개선하고자 오지시 보정을 실시한 후 이상기여도를 구하는 방법(Reconstruction Based Contribution(RBC))을 제안하였다. RBC는 스며듦 현상을 상당 부분 개선하였으나, 완전히 제거하지는 못했다[5].

한편, 센서의 개수가 증가할수록 고장센서 식별이 어려워지는 문제를 해결하기 위하여 다중블록 PCA(MBPCA)가 제안되었다[6~10]. MBPCA는 많은 개수의 센서들을 지역이나 설비 단위의 블록으로 구분하여 고장센서를 찾는 방법으로, 먼저 블록 차원에서 고장을 찾고, 이어서 블록 내의 센서에서 고장을 찾는다. 센서의 개수를 줄여서 고장센서를 식별하므로 스며듦 현상을 어느 정도 개선할 수 있다. Liu 등[8]은 MBPCA에 Alcala와 Qin[4]의 RBC를 적용하였다.

그런데 MBPCA도 센서의 개수가 증가하면 고장탐지 성능이 나빠질 개연성이 커진다. 본 연구에서 다룰 도시가스 배관망은 센서가 250개 이상으로 많고, 수요가의 변화로 인해 수시로 모델을 갱신해야 한다. 이를 고려할 때 MBPCA의 크기를 제한할 필요가 있다. 본 연구에서는 PCA와 MBPCA의 성능을 비교할 수 있는 지수를 제안하였으며, 성능이 더 우수한 경우에만 선택적으로 MBPCA를 적용할 수 있도록 하였다.

본 논문의 구성은 다음과 같다. II장에서는 (주)해양에너지 공급권역의 도시가스 배관망과 기존의 이상 감시 방법을 알아본다. III장에서는 MBPCA로서 Concensus PCA와 고장진단을 위한 RBC 알고리즘, 그리고 본 연구에서 제안하는 성능 지수를 설명한다. IV장에서는 (주)해양에너지 공급권역의 도시가스 배관망의 블록들에 대해 PCA와 MBPCA 성능을 비교하고, V장에서 결론을 기술한다.

Ⅱ. 도시가스 배관망

(주)해양에너지는 Fig. 1과 같이 광주광역시 및 전라남도 내 8개의 지역에 도시가스를 공급하고 있으며, 2022년 기준 총 연장 2,490km의 배관을 통해 836,609개의 수요가에 815백만 m^3의 가스를 공급한다.

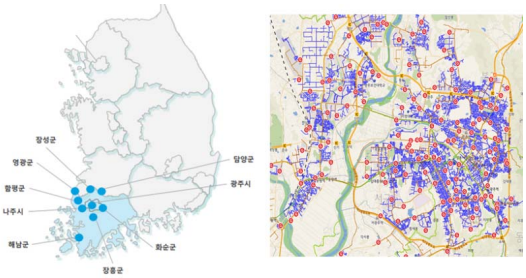


Fig. 1. City gas network managed by Haeyang Energy.

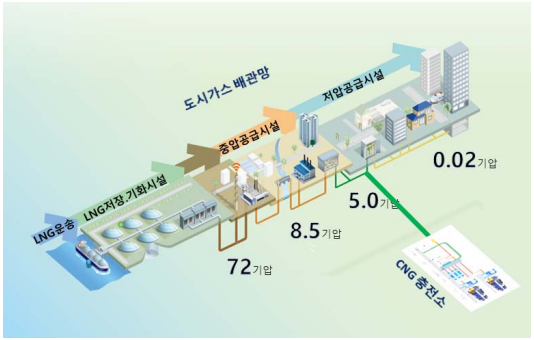


Fig. 2. Pressure regulations in city gas network.

도시가스 배관은 압력 제어를 위해 11개의 지구정압기와 241개의 지역정압기가 설치되어 있다. Fig. 2처럼 지구정압기는 8.5기압의 가스를 5.0기압으로 감압하며, 지역정압기는 0.02기압으로 재감압하여 최종 수요가에 공급한다.

(주)해양에너지는 가스누설에 의한 사고나 불안정한 압력에 의한 공급 불량을 방지하기 위하여 종합상황실에서 도시가스 배관망의 상태를 감시한다. SCADA 시스템을 통해 가스 배관의 압력 측정값을 실시간 수집하고, 압력이 기준값을 초과거나 일정치 이하로 떨어지면 정비 인력을 현장에 파견하여 조치한다. 그러나 이와 같은 기준값을 이용한 경보방식은 외부환경을 고려하지 않은 획일화된 기준을 적용하기 때문에 불필요한 이상 경보를 발행하는 문제가 있다. 그렇다고 기준값을 완화하면 이상을 놓쳐 더 심각한 문제가 발생할 수 있다. 이에 도시가스 배관망의 경보체계를 고도화하고자 빅데이터 기반의 고장탐지 및 원인 진단 시스템을 개발하고, 이를 통합관제 시스템에 구축하여 24시간 감시하고자 한다.

Ⅲ. 이론 및 기술

3.1. 고장 탐지

데이터 행렬 X를 변수를 기준으로 B개의 블록으로 나누고, 각 블록 행렬들로 분리하여 표현하면 다음과 같다.

$$X = [X_1\, X_2\, \cdots\, X_B] \tag{1}$$

여기서 b번째 블록 X_b는 평균이 0이고, 분산이 $1/m_b$로 정규화한 데이터이다. 블록들의 비중을 동일하게 하기 위하여 변수 개수 m_b로 나누어 정규화한다 [11].

PCA 모델의 로딩 $\overline{P}$를 주성분공간의 로딩 $P \in \mathbb{R}^{m \times l}$와 잔차공간의 로딩 $\tilde{P} \in \mathbb{R}^{m \times (m-l)}$로 분리하여 표현하면 $\overline{P} = [P\ \tilde{P}]$이다. 여기서 l은 모델에 사용할 주성분의 개수이다. 두 로딩을 B개의 블록으로 분리하여 표현하면, $P = [P_1^T\, P_2^T\, \cdots\, P_B^T]^T$와 $\tilde{P} = [\tilde{P}_1^T\, \tilde{P}_2^T\, \cdots\, \tilde{P}_B^T]^T$이다.

센서 고장은 잔차공간의 통계량인 SPE(Squared Prediction Error) 지수를 이용하여 탐지한다. 시료 데이터 $x \in \mathbb{R}^m$의 SPE 지수 $Q(x)$가 고장 판별 경계값 ucl보다 크면 고장으로 판단한다.

$$Q(x) = x^T \tilde{P} \tilde{P}^T x \geq ucl \tag{2}$$

본 연구에서는 고장 판별에 SPE 지수를 사용하지만, 주성분공간에서의 통계량인 Hotelling T^2나 SPE와 T^2의 융합지수도 적용할 수 있다. 고장 판별 경계값 ucl은 학습 데이터의 $Q(x)$를 오름차순으로 나열할 때 95%(또는 99%)에 해당하는 값으로 결정한다. 이는 학습 데이터의 5%(또는 1%)를 고장으로 판단하는 오류, 즉 제1종 오류를 허용한다는 의미이다.

시료 데이터를 $x = [x_1\, x_2\, \cdots\, x_B]$와 같이 블록으로 구분하면, 각 블록의 SPE 지수는 다음 식과 같다.

$$Q(x_b) = x_b^T \tilde{P}_b \tilde{P}_b^T x_b \tag{3}$$

각 블록의 $Q(x_b)$ 합이 $Q(x)$이므로, 각 블록의 $Q(x_b)$는 고장 판별 지수에 대한 블록 b의 기여 정도이다. 각 블록의 $Q(x_b)$를 비교하여, 가장 큰 값을 갖는 블록에 고장센서가 존재하는 것으로 판별한다.

3.2. 고장 식별

식 (3)은 전통적인 이상기여도를 블록 차원에서 적용한 것이고, 본 연구에서는 오지시보정 기반의 이상기여도 RBBC(Reconstruction Based Block Contribution)를 사용한다. 고장방향행렬을 Ξ_b라 표시하면 블록 b의 센서들의 고장으로 인한 오지시 보정값은 다음과 같다.

$$z_b = x - \Xi_b f_b \tag{4}$$

여기서 f_b는 블록 b의 고장 크기를 나타낸다. 오지시 보정이란 최적화를 통해 고장 판별 지수 $Q(z_b)$가 최소가 되는 f_b를 찾는 작업을 의미한다. 즉,

$$f_b = \arg\min(Q(z_b)) \tag{5}$$

이다. $Q(z_b)$를 최소로 하는 f_b는 리스트스퀘어 해를 통해 구할 수 있다.

$$f_b = (\Xi_b^T \tilde{P}\tilde{P}^T \Xi_b)^{-1} \Xi_b^T \tilde{P}\tilde{P}^T x \tag{6}$$

블록 b의 이상기여도 RBBC(Reconstruction Based Block Contribution)는 다음과 같다.

$$RBBC_b = \| \tilde{P}\tilde{P}^T \Xi_b f_b \|^2 \tag{7}$$

여기서 $(PP^T)^T = PP^T$이므로, 식 (7)에 식 (6)을 대입하여 정리하면 블록 b에 대한 이상기여도를 구할 수 있다[8].

$$\begin{aligned} RBBC_b &= f_b^T \Xi_b^T \tilde{P}\tilde{P}^T \Xi_b f_b \\ &= x^T \tilde{P}\tilde{P}^T \Xi_b (\Xi_b^T \tilde{P}\tilde{P}^T \Xi_b)^{-1} \Xi_b^T \tilde{P}\tilde{P}^T x \end{aligned} \tag{8}$$

여기서 분모 $\Xi_b^T \tilde{P}\tilde{P}^T \Xi_b$는 랭크 부족(rank-deficient)이므로 의사역행렬(pseudo inverse)을 구한다.

블록 단위의 이상기여도를 비교하여 고장블록을 식별하였으면, 이어서 고장블록에 속한 센서들의 이상기여도를 비교하여 고장센서를 식별한다[8]. 블록 b에 속한 i번째 센서의 오지시 보정식은 다음과 같다.

$$z_{b,i} = x_b - \xi_{b,i} f_{b,i} \tag{9}$$

여기서 고장방향벡터 $\xi_{b,i}$는 m x m 크기의 단위행렬에서 i번째 열벡터이고, $f_{b,i}$는 블록 b에 속한 i번째 센서의 고장 크기이다. 블록 b에 속한 i번째 변수에 대한 SPE 지수는 다음과 같다.

$$Q(z_{b,i}) = z_{b,i}^T \tilde{P}_b \tilde{P}_b^T z_{b,i} \tag{10}$$

$Q(z_{b,i})$를 최소로 하는 $f_{b,i}$는 리스트스퀘어 해로 구한다.

$$f_{b,i} = (\xi_{b,i}^T \tilde{P}_b \tilde{P}_b^T \xi_{b,i})^{-1} \xi_{b,i}^T P_b^T \tilde{P}_b^T x_b \tag{11}$$

블록 b에 속한 센서 i의 오지시보정 기반의 이상기여도 RBVC(Reconstruction Based Variable Contribution)는 다음과 같다.

$$RBVC_{b,i} = \| \tilde{P}_b \tilde{P}_b^T \xi_{b,i} f_{b,i} \|^2 \tag{12}$$

식 (11)을 식 (12)에 대입하면 $x_{b,i}$에 대한 이상기여도를 구할 수 있다[8].

$$\begin{aligned} RBVC_{b,i} &= f_{b,i}^T \xi_{b,i}^T \tilde{P}_b \tilde{P}_b^T \xi_{b,i} f_{b,i} \\ &= x_b^T \tilde{P}\tilde{P}_b^T \xi_{b,i} (\xi_{b,i}^T \tilde{P}_b \tilde{P}_b^T \xi_{b,i})^{-1} \xi_{b,i}^T \tilde{P}_b \tilde{P}_b^T x_b \end{aligned} \tag{13}$$

블록 b에 속한 센서들의 $RBVC_{b,i}$를 비교하여 가장 큰 값의 센서를 고장으로 식별한다.

3.3. 성능 지수

MBPCA 구현을 위해 다양한 방법들이 제안되었다. 가장 전통적인 방법 중의 하나가 Concensus PCA이다[6]. 최근에는 고장식별 성능을 개선하고자 LASSO (Least Absolute Shrinkage and Selection Operator)를 도입한 방법이 제안되었다[9]. 본 연구에서는 MBPCA 방법으로 Concensus PCA를, 고장식별 방법으로 Liu 등[8]의 RBBC를 사용하였다. Concensus PCA의 알고리즘을 [부록]에 소개하였다.

PCA와 MBPCA의 성능을 비교하기 위해 본 연구에서는 고장탐지 성능(fault detectability) 지수와 고장센서 식별 성능(fault identificability) 지수를 제안하였다. 전자는 고장의 발생을 탐지하는 데 사용하는 지수이고, 후자는 고장 센서를 식별하는 데 사용하는 지수이다. PCA에서 고장탐지(fault detectability) 충분조건은

$$\| \tilde{P}\tilde{P}^T \xi_i f_i \| \geq 2\sqrt{ud} \tag{14}$$

이다[7]. 여기서 f_i는 고장탐지 충분조건으로서 센서 i 고장의 크기이다. 식 (14)의 좌변은 잔차공간에서 고장의 크기이며, 우변은 정상영역의 다차원 구의 지름이다. 고장의 크기가 이 지름보다 크면 항상 고장이므로 식 (14)는 고장탐지에 충분한 조건이다. 센서 i의 고장의 크기 f_i가 고장을 판단하는 경계이므로 크기가 작을수록 고장을 쉽게 탐지할 수 있다.

$$f_i = 2\sqrt{ud} / \| \tilde{P}\tilde{P}^T \xi_i \| \tag{15}$$

MBPCA에 포함하는 센서가 많아지면 대체로 $\sqrt{ud}$과 $\| \tilde{P}\tilde{P}^T \xi_i \|$이 커진다. 만약에 $\sqrt{ud}$ 보다 $\| \tilde{P}\tilde{P}^T \xi_i \|$이 커지는 정도가 작아 f_i가 커지면, 고장탐지 성능이 나빠진다. 반면에 f_i가 작아지면 MBPCA의 고장탐지 성능이 좋아진다.

MBPCA에서 블록 단위의 ucl_b는 식 (8)의 $RBBC_b$를 이용하여 구할 수 있다. 학습 데이터에 대해 $RBBC_b$를 구하고, 크기 오름차순으로 95%(또는 99%)에 해당하는 값으로 ucl_b를 결정한다. 고장탐지를 블록 단위로 실시

하면 블록 b에 속한 i번째 센서의 고장탐지 충분조건을 만족하는 고장의 크기 $f_{b,i}$는 식 (15)로부터 구할 수 있다.

$$f_{b,i} = 2\sqrt{ucd_b} / \| \tilde{P}\tilde{P}^T \xi_i \| \quad (16)$$

다음은 고장센서 식별 성능 지수에 대해 설명하고자 한다. Alcala와 Qin은 오지시보정 기반의 이상기여도를 이용할 때 고장센서 식별을 보장하는 조건(fault identificability)으로 다음 식을 제시하였다[4].

$$\lambda_{ii} = \xi_i \tilde{P}\tilde{P}^T \xi_i \geq (\xi_j \tilde{P}\tilde{P}^T \xi_j)^{-1} (\xi_j \tilde{P}\tilde{P}^T \xi_i)^2 \quad (17)$$

$\tilde{P}\tilde{P}^T$의 대각선 요소인 λ_{ii}는 고장 벡터를 잔차공간에 투영할 때 변수 i 방향의 크기를 나타내며, 이 값이 클수록 스며듦 현상이 줄어들어 고장 식별이 쉬워진다. 따라서 어떤 블록 조합에서 PCA와 MBPCA의 λ_i 크기를 통해 성능이 더 우수한 기법을 선택할 수 있다. 여기서 $\sum \lambda_{ij} (= \sum \xi_i \tilde{P}\tilde{P}^T \xi_j)$의 크기가 1이 아니므로 정규화한 값을 사용한다.

$$\lambda_i = \lambda_{ii} / \| \tilde{P}\tilde{P}^T \xi_i \| \quad (18)$$

성능 지수를 요약하자면, PCA와 MBPCA 중에서 f_i가 더 작은 방법이 고장탐지 성능이 더 우수하고, λ_i가 더 큰 방법이 고장식별 성능이 더 우수하다.

Ⅳ. 사례연구

㈜ 해양에너지 공급권역의 도시가스 배관망에 설치되어 있는 압력 센서 총 252개에 대해 PCA 기반의 고장 탐지 및 진단 기법을 적용하였다. 2023.01.01부터 6개월 동안 5분 간격으로 수집한 데이터를 학습에 사용하였다. 실무자의 도움을 받아 센서들을 지역 기준으로 총 26개의 블록으로 분리하였다. 이 중에서 16개 블록은 상관계수가 99%이상으로 매우 컸으며, 나머지 10개 블록은 50%~70%로 상대적으로 작았다. 상관계수가 큰 16개 블록은 개별적인 PCA를 구성하였다. 나머지 10개 블록은 서로 보완적인 관계의 블록 조합을 찾아 MBPCA를 구성하였다.

본 장에서는 PCA와 MBPCA의 성능을 비교하고 본 연구에서 제안한 지수의 타당성을 분석하고자 네 가지 사례를 준비하였다. 두 블록씩 짝을 지어 구성한 세 가지 사례와 네 개의 블록으로 구성된 사례이다. 네 번째 사례는 도시가스 배관망의 고장 탐지 및 진단 시스템에 포함된 것이다.

먼저 세 가지 사례를 소개하자면, Fig. 3에 도식화한 상관계수 행렬로 이해할 수 있듯이 두 블록의 조합이 내·외적으로 상관성의 정도가 다른 경우들이다. 그림에서 셀의 색이 진할수록 종축과 횡축의 센서 간의 상관성이 높고, 옅을수록 상관성이 낮다. 먼저 Fig. 3의 (a)의 Block A와 Block B는 블록 내·외적으로 상관성이 높다. 그리고 (b)의 Block C와 Block D는 블록 내적으로 센서들의 상관성은 높지만, 두 블록 간의 상관성은 이보다 낮다. (c)의 Block E와 Block F는 블록 내·외적으로 상관성이 불균일한 경우이다.

각 블록에 속한 센서들의 24시간 동안의 시계열 추이는 Fig. 4와 같다. Fig. 4의 (a)는 Block A, (b)는 Block B(Block C와 동일함), (c)는 Block D의 센서들의 시계열 추이도이다. 센서들이 거의 동일한 패턴을 보인다. 반면에 (d)는 Block E, (e)는 Block F의 센서들의 시계열 추이도인데, 패턴이 다른 센서들이 포함되어 있다. 오전 08:00시~10:00시 사이에 오후 18:00시~22:00시 사이는 가스 사용량이 많아 압력이 낮아지는 패턴을 보인다.

세 가지 사례에 대해 PCA와 MBPCA 모델을 개발하였으며, 그 결과는 Table 1과 같다. 주성분의 개수 PC는 설명분산계수 99%로 결정하였다. 고장 판별 경계값 ucl은 SPE를 크기의 내림차순으로 나열하고 13,000회 측정 데이터의 1%인 130번째 SPE 값으로

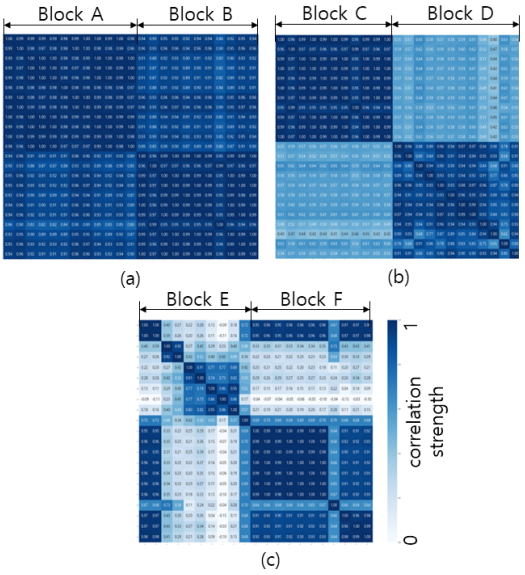


Fig. 3. Correlation matrix of the sensors in the example cases.

Table 1. Performance indices of three examples

(a) Example 1

	PCA		MBPCA	
block	A	B	A	B
PC	1	2	2	
$\sqrt{ucl_b}$	1.10	0.23	0.77	1.55
$f_{b,i}$ average	2.19	1.08	1.84	2.61
$\lambda_{b,i}$average	0.96	0.90	0.96	0.95

(b) Example 2

	PCA		MBPCA	
block	C	D	C	D
PC	2	3	5	
$\sqrt{ucl_b}$	0.23	1.50	0.79	0.95
$f_{b,i}$ average	1.08	2.88	1.92	2.37
$\lambda_{b,i}$average	0.90	0.86	0.87	0.71

(c) Example 3

	PCA		MBPCA	
block	E	F	E	F
PC	5	3	6	
$\sqrt{ucl_b}$	1.66	1.04	1.59	0.49
$f_{b,i}$ average	3.71	3.32	3.52	1.76
$\lambda_{b,i}$average	0.70	0.86	0.75	0.88

결정하였다. $f_{b,i}$average는 블록 b 센서들에 대해 식 (16)으로 계산한 $f_{b,i}$의 평균값을 의미하고, $\lambda_{b,i}$average는 식 (18)로 계산한 블록 b 센서들의 λ_i의 평균값을 의미한다. MBPCA의 $f_{b,i}$average가 PCA보다 작을수록, MBPCA의 $\lambda_{b,i}$average가 PCA보다 클수록 두 블록은 MBPCA가 더 적합하다.

Table 1의 (a)는 블록 내·외적으로 상관성이 높은 Block A와 Block B의 결과이다. $f_{b,i}$average는 Block A의 경우 PCA 2.19에서 MBPCA 1.84로 16% 개선되었으나, Block B는 1.08에서 2.61로 142% 저하되었다. $\lambda_{b,i}$average는 Block A의 경우 0.96으로 변화가 없었고, Block B는 6% 개선되었다. 이 사례는 블록 내적으로 상관성이 높아서 외부 블록의 도움이 크지 않는 경우로 Block B의 $f_{b,i}$average가 크게 저하되었기 때문에 MBPCA로 엮을 필요가 없다.

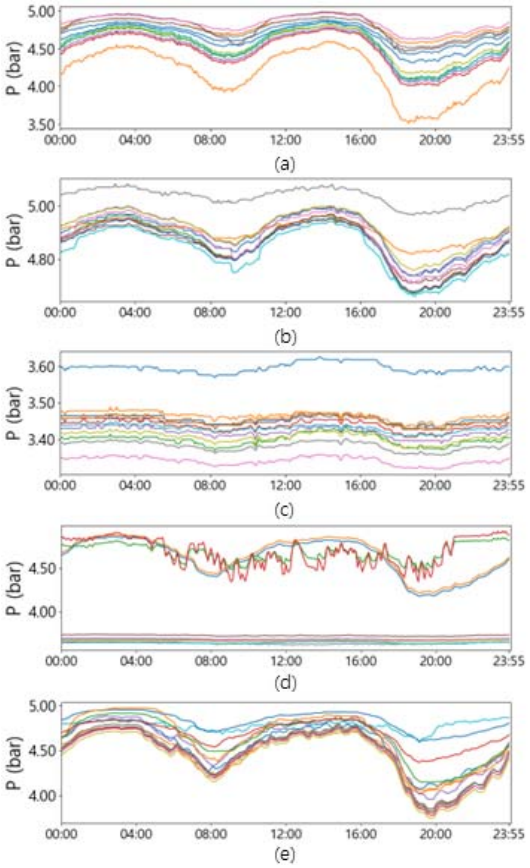


Fig. 4. Pressure trend of the sensors for 24 hours: (a) Block A, (b) Block B/C, (c) Block D, (d) Block E, and (e) Block F.

Table 1의 (b)는 블록 내적으로는 센서들의 상관성이 높지만, 두 블록 간에는 상관성이 낮은 사례의 결과이다. $f_{b,i}$average는 Block C의 경우 PCA 1.08에서 MBPCA 1.92로 77% 저하되었고, Block D는 2.88에서 2.37로 17% 개선되었다. $\lambda_{b,i}$average는 Block C의 경우 3% 저하되었고, Block D도 17% 저하되었다. $f_{b,i}$average의 개선 정도에 비해 $\lambda_{b,i}$average의 저하 정도가 커서 두 블록을 MBPCA로 엮을 필요가 없다.

Table 1의 (c)는 블록 내·외적으로 센서 간의 상관성이 일정하지 않은 Block E와 Block F의 결과이다. $f_{b,i}$average는 Block E는 5%, Block F는 47% 개선되었다. $\lambda_{b,i}$average는 Block E는 7%, Block F는 2% 개선되었다. $f_{b,i}$average와 $\lambda_{b,i}$average 모두 개선되었기 때문에 두 블록을 MBPCA로 엮을 필요가 있는 사례이다.

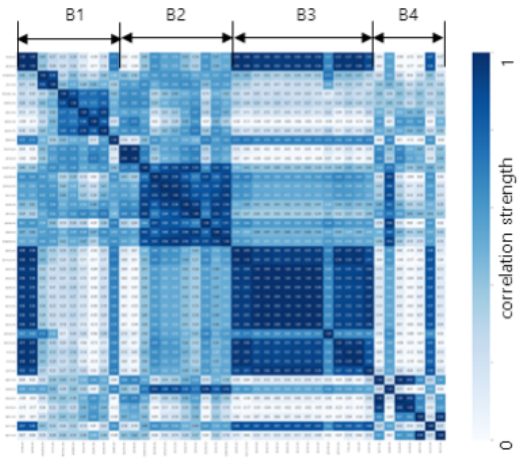


Fig. 5. Correlation matrix of the sensors of Example 4.

Table 2. Performance indices of PCA and MBPCA of Example 4

	PCA				MBPCA			
block	B1	B2	B3	B4	B1	B2	B3	B4
PC	5	4	3	4	12			
$\sqrt{ucl_b}$	1.66	0.85	1.04	1.03	1.54	0.86	0.91	0.43
$f_{b,i}$ average	3.71	2.37	3.32	7.82	3.19	2.29	2.16	2.48
$\lambda_{b,i}$average	0.70	0.79	0.86	0.58	0.78	0.91	0.87	0.72

네 번째 사례는 네 개의 블록으로 구성된 경우로 센서 간의 상관계수 정도는 Fig. 5와 같다. Table 2에 PCA와 MBPCA 적용 결과를 나타내었다. 여기서 블록 B1은 앞 예제의 Block E이고, 블록 B3는 Block F이다. 여기서 특이한 점은 블록 B4의 $f_{b,i}$average가 7.82로 매우 크다는 사실이다. PCA 대비 MBPCA의 $f_{b,i}$ average는 모든 블록이 개선되었다. 특히 블록 B4는 PCA 7.82에서 MBPCA 2.48로 68% 개선되었다. $\lambda_{b,i}$ average도 네 개 블록이 모두 개선되었으며, 블록 B4는 24% 개선되었다. 종합적으로 $f_{b,i}$average와 $\lambda_{b,i}$ average 모두 개선되어 MBPCA로 엮을 필요가 있다고 판단하였다.

본 연구에서 제안한 고장탐지 지수와 고장진단 지수가 유용한지 확인하기 위하여 측정값에 인위적으로 오차를 부여하고 시험을 실시하였다. 네 번째 사례

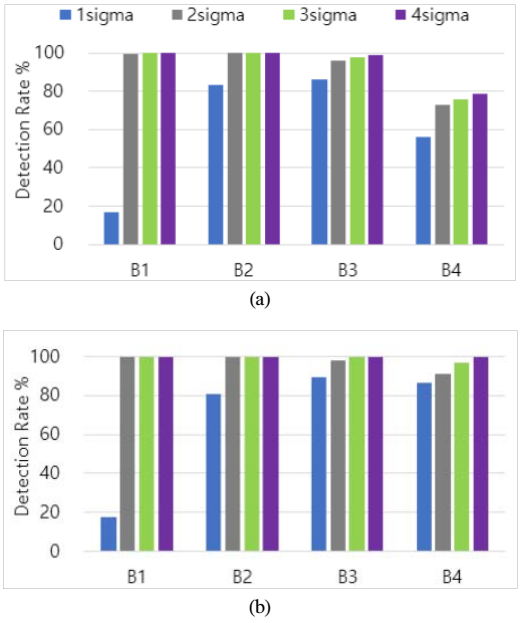
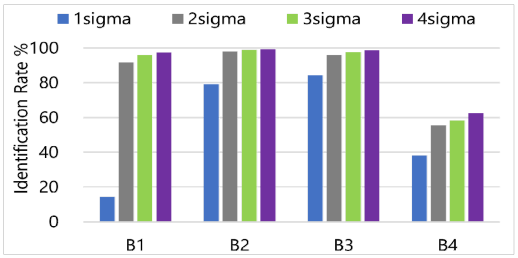


Fig. 6. Fault detection performance of Example 4: (a) PCA, (b) MBPCA.

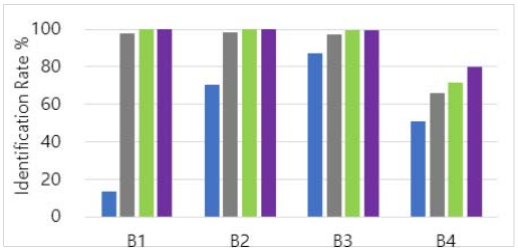
에서 정상 데이터 200세트를 추출하고, 40개 모든 센서의 측정값에 각각 1, 2, 3, 4 표준편차의 오차를 부여하였다. 이렇게 만든 오차 데이터에 대해 PCA와 MBPCA의 센서고장 탐지와 고장센서 식별 성능을 비교하였다.

Fig. 6은 센서고장 탐지 성공률을 나타낸 그래프로 (a)는 PCA, (b)는 MBPCA의 결과이다. 센서고장 탐지는 블록별로 실시하였는데, 식 (16)의 블록 b의 고장 판별 경계값 ucl_b보다 식 (8)의 $RBBC_b$가 크면 고장으로 판단하였다. 하나의 블록이라도 고장이 탐지되면 고장으로 판단하였다. 센서고장 탐지 성공률은 센서별로 고장을 부여한 데이터 200세트 중에서 고장이 탐지되는 데이터 세트 수(A)의 비율(A/200)을 각 센서에 대해 구하고, 이들의 블록 평균값이다. 전체적으로 고장의 크기가 증가할수록 고장탐지 성공률이 증가하였다. 1 표준편차에서는 블록별로 결과가 상이하였으며, 특히 B1이 낮았다. PCA 경우 3과 4 표준편차에서 B4를 제외하고 성공률이 100%에 근접하였다. MBPCA의 경우는 2 ~ 4 표준편차에서 모두 100%에 근접하였으며, B4도 80% 이상으로 높았 다. 특히 B4에서 PCA 대비 MBPCA 성공률이 20% 정도 개선되었으며, 나머지 블록은 비슷하거나 약간 개선되었다. 전체적으로 본 시험 결과는 Table 2를 따른다.

Fig. 7은 고장센서 식별 성공률을 나타낸 그래프로



(a)



(b)

Fig. 7. Fault identification performance of Example 4 (a) PCA, (b) MBPCA.

(a)는 PCA, (b)는 MBPCA의 결과이다. 고장센서 식별 성공률은 각 센서별로 고장을 부여한 데이터 200세트 중에서 고장 탐지에 성공한 데이터로 고장센서 식별을 실시하여 식별에 성공한 데이터 세트 수(B)의 비율(B/200)을 각 센서에 대해 구하고, 이들의 블록 평균값이다. 고장센서 식별 성공률도 고장탐지 성공률과 비슷한 경향을 보였다. 특히 PCA 대비 MBPCA 성공률이 B4에서 15% 정도 개선되었다. 그런데 Table 2에 의하면 B2는 PCA 대비 MBPCA 성공률이 0.79에서 0.91로 15% 개선을 예측하였는데, 시험 결과에 의하면 1 표준편차에서 오히려 나빠졌다. 저자들이 수행한 다른 분야의 문제에서도 비슷한 경우가 있었는데, 이는 오차의 크기가 작을수록 측정값 노이즈의 영향이 커지기 때문으로 판단된다. 측정값 노이즈의 영향이 작은 2 표준편차에서는 이상기여도가 100%에 더 근접하였다. 이를 제외하고는 전체적으로 Table 2를 따른다. 본 시험을 통해 본 연구에서 제안한 지수가 PCA와 MBPCA의 성능을 평가하는 데 유용함을 확인하였다.

V. 결 론

일반적으로 공정 시스템에서 압력이나 온도, 유량을 측정하는 센서 개수가 20~30개 이상이면 단일 PCA로 고장센서를 식별하는 것은 쉽지 않다. 측정값 노이즈의 모델 외란과 스며듦 현상의 이상기여도의 한계 때문이다. 이러한 이유로 MBPCA가 추천되는데, 센서의 개수가 증가할수록 고장탐지 성능이 나빠질 개연성이 커지며, 지엽적인 변화에도 모델 전체를 갱신해야 하는 유지보수의 문제도 따른다.

따라서 PCA 대비 성능이 개선되는 경우에만 MBPCA를 구성하기 위하여 두 기법의 성능을 비교할 수 있는 지수를 두 가지 제안하였다. 하나는 고장탐지 성능(fault detectability)을 평가하기 위한 것으로 고장탐지 충분조건으로서 센서의 고장 크기이며, 다른 하나는 고장진단 성능(fault identifiability)을 평가하기 위한 것으로 잔차공간에서 센서 방향의 고장 크기이다. 고장탐지 성능 지수는 작을수록, 고장센서 식별 성능을 클수록 성능이 우수하다. 측정값에 인위적으로 오차를 부여하고, 고장탐지 성능과 고장진단 성능을 비교하였으며, 본 연구에서 제안한 지수가 타당함을 확인하였다.

(주)해양에너지의 공급권역의 도시가스 배관망에 대한 센서고장 탐지 및 진단 시스템을 개발하였다. 압력 센서 252개를 26개 블록으로 나누고 고장탐지 지수와 고장진단 지수를 적용하여 블록별로 PCA와 MBPCA를 선택하여 개발하였다. 이를 통해 센서의 개수가 많은 경우에 발생하는 문제점을 해결할 수 있었다.

감사의 글

본 연구는 2022년도 산업통상자원부의 재원으로 한국산업기술진흥원의 지원을 받아 수행된 DX한걸음 프로젝트(과제번호 P0021372)로 지원에 감사를 드립니다.

부 록

Concensus PCA 알고리즘[7]

1. 데이터 행렬 X를 B개의 블록 $X = [X_1\ X_2\ \cdots\ X_B]$로 나누고, $X_{b,i} = X_b$와 $i = 1$로 설정한다. 여기서 개별 블록 X_b는 평균이 0이고, 분산이 $1/m_b$로 정규화한 것이다.

2. 임의의 $t_{T,i}$를 가정하고, 수렴할 때까지 아래의 과정을 반복한다.

$$p_{b,i} = X_{b,i}^T t_{T,i} / \| X_{b,i}^T t_{T,i} \|$$
$$t_{b,i} = X_{b,i} p_{b,i}$$
$$T_i = [t_{b,i} \cdots t_{B,i}]$$
$$P_{T,i} = T_i^T t_{T,i} / \| T_i^T t_{T,i} \|$$
$$t_{T,i} = T_i p_{T,i}$$

3. X의 잔여 분을 계산하고 주성분 개수만큼 단계 2를 반복한다.

$$X = (I - t_{t,i} t_{T,i}^T / t_{T,i}^T t_{t,i}) X_{b,i}$$

REFERENCES

[1] Chung,W.H., Park, G.J., Gu,Y.H., Lim, S.H., Yoo, S.J., Jo, Y.D., "City Gas Pipeline Pressure Prediction Model", The *Journal of Society for e-Business Studies*, 23(2), 33~47, (2018)

[2] Mujtaba, S. M., Lemma, Tamiru A. S., Vandrang, K., "Leak Diagnostics in Natural Gas Pipelines Using Fault Signatures", *International Journal of Pressure Vessels and Piping*, 199, October (2022) Volume 199, October 2022, 104698

[3] Yi, T. H., Huang, H. B., Li, H. N., "Development of Sensor Validation Methodologies for Structural Health Monitoring: a Comprehensive Review", *Measuremen,t*, 109, 200-214, October (2017)

[4] Alcala, C.F., Qin, S.J., "Reconstruction-based Contribution for Process Monitoring", *Automatica,* 45, 1593−1600, (2009)

[5] Shang, C., Huang, B., Yang, F., Huang, D., "Slow Feature Analysis for Monitoring and Diagnosis of Control Performance", *Journal of Process Control,* 39, 21-34, March (2016)

[6] Westerhuis, J. A., Kourti, T., MacGregor, J. F., "Analysis of Multiblock and Hierarchical PCA and PLS Models," *J. Chemometrics*, 12, 301−321, (1998)

[7] Qin, S., Valle, S., Piovoso, M., "On Unifying Multiblock Analysis with Application to Decentralized Process Monitoring", *Journal of Chemometrics*, 15, 715−742, (2001)

[8] Liu, Q., Chai, T., Qin, S. J., "Fault Diagnosis of Continuous Annealing Processes Using a Reconstruction-based Method," *Control Eng. Practice*, 20(5), 511−518, (2012)

[9] Shang, C., Ji, H., Huang, X., Yang, F., Huang, D., "Generalized Grouped Contributions for Hierarchical Fault Diagnosis with Group Lasso", *Control Engineering Practice*, 93, 1~12, December (2019)

[10] Mishra, P., Roger, J. M., Jouan-Rimbaud- Bouveress, D., Biancolillo, A., Marini, F., Nordon, A., Rutledge, D. N., "Recent Trends in Multi-block Data Analysis in Chemometrics for Multi-source Data Integration", *Trends in Analytical Chemistry,* 137, April (2021)

[11] Dunia, R., Qin, S. J., "A Unified Geometric Approach to Process and Sensor Fault Identification and Reconstruction: The Unidimensional Fault Case", *Computers & Chemical Engineering*, 22(7-8), 927−943, (1998)