

## 데이터 기반 이상진단법을 위한 화학공정의 조업모드 판별

이창준 · 고재욱\* · 이기백\*\*†

서울대학교 화학생명공학부  
151-742 서울시 신림동 산 56-1  
\*광운대학교 화학공학과  
139-701 서울시 노원구 월계동 447-1  
\*\*충주대학교 화학생명공학과  
380-702 충북 충주시 대학로 72  
(2007년 12월 11일 접수, 2008년 2월 5일 채택)

### Operation Modes Classification of Chemical Processes for History Data-Based Fault Diagnosis Methods

Chang Jun Lee, Jae Wook Ko\* and Gibaek Lee\*\*†

Department of Chemical and Biological Engineering, Seoul National University, San 56-1, Shilim-dong, Gwanak-gu, Seoul 151-742, Korea

\*Department of Chemical Engineering, Kwangwoon University, 447-1 Wolgye-dong, Nowon-gu, Seoul 139-701, Korea

\*\*Department of Chemical and Biological Engineering, Chungju National University, 72 Daehak-ro, Chungju, Chungbuk 380-702, Korea

(Received 11 December 2007; accepted 5 February 2008)

#### 요 약

화학공정의 안전하고 효율적인 운전에 관심이 커지면서 공정이상 원인을 조기에 진단하기 위한 다양한 이상진단 방법이 연구되어 왔다. 최근에는 통계적 모델 등 정량적 데이터에 기반한 이상진단방법이 많이 연구되고 있으나, 특정 조업영역에서 얻어진 통계적 모델을 다른 조업영역에 적용하면 오진단이 많아지게 된다. 따라서 공정특성상 다양한 조업영역이 존재하는 화학공정에 데이터기반 방법론을 적용하기에는 어려움이 있어 화학공정의 조업영역 판별법이 요구되고 있다. 이 연구에서는 유클리드 거리(Euclidean distance), FDA(Fisher's discriminant analysis), PCA(principal component analysis)의 통계모델과 이 모델들에 공정변수의 동특성을 반영한 모델을 제안하였다. 6개의 조업모드를 가진 TE(tennessee eastman) 공정에 대한 사례연구를 통해 동특성을 반영한 PCA 모델의 성능이 가장 우수함을 확인하였다.

**Abstract** – The safe and efficient operation of the chemical processes has become one of the primary concerns of chemical companies, and a variety of fault diagnosis methods have been developed to diagnose faults when abnormal situations arise. Recently, many research efforts have focused on fault diagnosis methods based on quantitative history data-based methods such as statistical models. However, when the history data-based models trained with the data obtained on an operation mode are applied to another operating condition, the models can make continuous wrong diagnosis, and have limits to be applied to real chemical processes with various operation modes. In order to classify operation modes of chemical processes, this study considers three multivariate models of Euclidean distance, FDA (Fisher's Discriminant Analysis), and PCA (principal component analysis), and integrates them with process dynamics to lead dynamic Euclidean distance, dynamic FDA, and dynamic PCA. A case study of the TE (Tennessee Eastman) process having six operation modes illustrates the conclusion that dynamic PCA model shows the best classification performance.

Key words: Fault Diagnosis, Operation Mode Classification, Principal Component Analysis, Fisher Discriminant Analysis, Euclidean Distance, Tennessee Eastman Process

#### 1. 서 론

화학산업에서의 경쟁이 더욱 격화되면서 화학공정은 한층 더 규모가 커지고 복잡해지고 있다. 이와 같은 변화는 공장에서의 사고

발생을 증가시켜 복구비용, 생산량 감소, 제품 질의 저하, 시장점유율의 하락 등 크고 작은 경제적 손실을 미치게 된다. 따라서 공정 이상의 원인을 조기에 진단함으로써 공정을 안전하게 조업할 수 있도록 도와주기 위한 다양한 이상진단 방법이 연구되어왔다[1]. 화학공정의 이상진단 방법론은 공정지식을 얻는 방법에 따라 모델기반 또는 데이터기반 방법으로 나누어지며, 각각 그 지식의 형태에 따라

†To whom correspondence should be addressed.  
E-mail: glee@cjnu.ac.kr

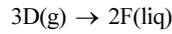
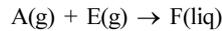
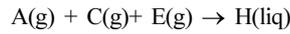
정량적, 정성적 방법으로 나뉘어진다. 최근에는 정량적 데이터 기반 방법론인 PCA(principal component analysis, 주성분분석), PLS (partial least squares) 등의 통계적방법 또는 통계적방법과 다른 방법론을 혼합한 진단방법에 대한 연구가 많이 이루어지고 있다[2, 3].

그러나 공정조건의 변화 등으로 인해 대상공정의 조업영역이 달라졌을 때 대상공정의 특정 조업영역에 대해 학습된 통계적 모델은 오진단이 많아지는 등 진단성능이 매우 나빠지게 된다. 따라서 공정 특성상 조업영역이 매우 다양한 화학공정에 통계적모델에 기반한 이상진단법을 적용하는 것은 큰 어려움이 있다[4]. 따라서 화학공정의 조업모드를 분류하여 조업모드마다 다른 진단모델을 구축하고 진단을 수행하여야 한다. 최근 Zhao 등은 화학공정의 조업영역을 판별하기 위해 각각의 조업영역 데이터에 대해 각각의 PCA 모델을 구축하는 다중 PCA 모델에 의한 조업모드 판별법을 발표하고 TE(tennessee esatman) 공정과 FCC 공정에 적용한 바 있다[4].

이 연구에서는 정량적 데이터기반 이상진단모델의 구축과 진단을 위한 조업모드 판별법으로서 유클리드 거리(Euclidean distance, ED), FDA(Fisher's discriminant analysis), PCA 등의 다양한 통계모델을 제안하고 TE 공정에 적용함으로써 제안된 방법의 판별성능을 보이고자 한다.

## 2. 대상공정

TE 공정은 공정제어기법, 최적화, 이상진단 등의 평가를 위하여 Downs와 Vogel에 의해 만들어진 벤치마크 공정으로, Eastman사의 TE 공정을 실제적으로 모사한 것이다[5]. 이 공정은 반응기, 분리기, 스트리퍼, 응축기, 압축기의 5개 주요 공정으로 이루어져 있으며(Fig. 1) 12개의 조작변수와 41개의 측정변수를 가지고 있다. 측정변수는 22개의 연속변수와 19개의 조성 측정변수로 나누어진다. 4가지의 반응물 A, C, D, E로부터 G와 H가 생산된다. 반응기에서는 다음과 같은 비가역 발열반응이 일어난다.



반응기의 생성물은 미반응물과 함께 증기상태로 반응기 상부로 빠져 나오며 응축기로 들어가 응축되어 기액 분리기로 보내진다. 기액 분리기에서는 응축되지 않은 성분을 분리하여 분리기 상부로 내보내며 압축기에 의해 반응기로 재순환된다. 비응축성분 중 일부는 외부로 방출되어 비활성물질 B와 부산물 F를 공정에서 제거한다. 기액 분리기에서 응축된 성분은 스트리퍼로 들어간다. 스트리퍼에서는 잔여미반응물을 4번 흐름으로 스트리핑하여 상부로 분리한다. 생산물은 스트리퍼의 하부로 나온다.

이 연구에서 사용한 모사기는 Ricker가 Matlab으로 개발한 것으로[6, 7] 적분간격은 0.01시간이고 모사시간은 10시간이다. TE 공정은 생성물흐름(11번 흐름)의 G/H 질량 비율과 유량에 6개 조업모드가 있으며(Table 1) 이 연구에서는 정상상태 및 이상발생시의 6개 조업모드를 판별하여 제안된 조업영역 판별법의 성능을 확인하고자 한다. Table 2의 설정치를 비교하였을 때 모드 2와 5, 모드 3과 6은 유사한 값을 보여주는데 특히 모드 2와 5는 다른 모드들과는 달리 대부분의 설정치가 매우 비슷한 값에 있어 조업모드 판별이 어려울 것으로 예측된다.

Table 1. 6 operation modes

Mode	G/H mass ratio	Production (kg/h)
1	50/50	14,076
2	10/90	14,076
3	90/10	11,111
4	50/50	Maximum
5	10/90	Maximum
6	90/10	Maximum

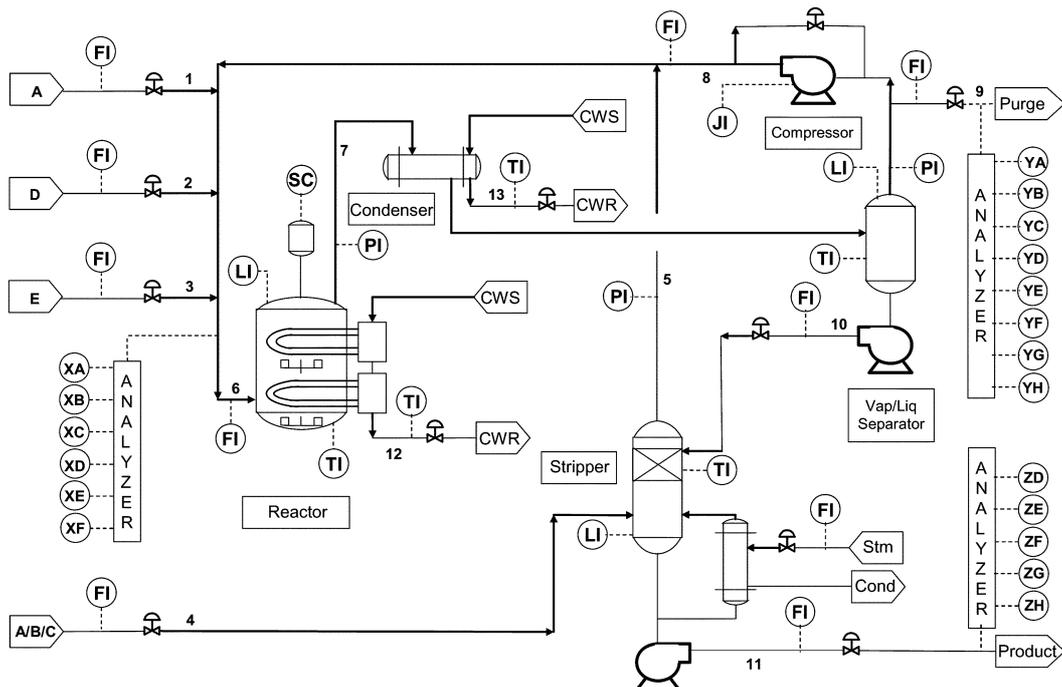


Fig. 1. The tennessee eastman challenge process.

**Table 2. Setpoints of 6 operation modes**

Setpoint	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6
1	22.89	22.73	18.04	36.04	23.55	20.2
2	50	50	50	50	50	50
3	50	50	50	50	50	50
4	65	65	65	65	65	65
5	2,800	2,800	2,800	2,800	2,800	2,800
6	53.83	11.66	90.09	53.35	11.65	90.07
7	63.21	64.18	62.11	61.95	64.03	61.47
8	50.96	54.25	47.43	58.76	54.32	48.79
9	122.9	124.2	121.9	128.2	124.6	123
10	1	1	77.621	1	1	71.166
11	1	1	1	1	1	1
12	100	100	100	100	100	100

**3. 방법론**

**3-1. 유클리드 거리에 의한 판별**

데이터 사이의 유사성을 측정하기 위해 데이터 사이의 거리가 많이 사용되고 분류를 목적으로 하는 경우에는 거리의 근접성이 이용된다. 이 연구에서는 거리를 측정하는 여러 가지 대표적인 방법들 중에서 유클리드 거리를 사용하였다. n개의 변수를 가진 2개의 데이터  $X^T = [x_1 \ x_2 \ \dots \ x_n]$ ,  $Y_2^T = [x_1 \ x_2 \ \dots \ x_n]$ 의 유클리드 거리, 즉 직선거리는 다음과 같다[8].

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \tag{1}$$

조업모드 판별을 위해 먼저 각 조업영역에서 정상상태 데이터의 산술평균을 이용하여 중심(centroid)을 찾는다. 온라인에서는 테스트 데이터와 각 조업영역의 중심 간의 유클리드 거리를 계산하고 거리가 가장 가까운 조업모드를 선택해서 그 조업모드를 테스트 데이터의 조업 모드라고 판단한다. Table 3은 TE 공정의 6개 조업모드의 중심 사이의 유클리드 거리를 보여주고 있다. 모드 2와 5사이의 거리가 다른 모드간 거리에 비해 매우 짧아 조업모드 오판의 가능성이 높음을 보여주고 있다.

**3-2. FDA에 의한 판별**

FDA는 데이터의 분류를 최대화시키기 위한 차원축소방법으로, 분류된 데이터를 각각 클래스라고 할 때 클래스 내 분산(within-class scatter)을 최소로 하면서 동시에 클래스 간 분산(between-class scatter)을 최대화시키도록 선형 변환벡터를 찾는 방법이다[9]. 클래스의 수를 p, 데이터의 수를 n이라 할 때 클래스 내 분산  $S_w$ , 클래스 간 분산  $S_b$ 에 대한 정의는 다음과 같다.

**Table 3. Euclidean distance between the centroid of each mode**

	Mode 1	Mode 2	Mode 3	Mode 4	Mode 5	Mode 6
Mode 1	0	5.5	8.8	10.6	5.8	8.6
Mode 2	5.5	0	12.8	10.9	0.8	12.7
Mode 3	8.8	12.8	0	15.1	13.0	2.8
Mode 4	10.6	10.9	15.1	0	10.3	14.1
Mode 5	5.8	0.8	13.0	10.3	0	12.9
Mode 6	8.6	12.7	2.8	14.1	12.9	0

$$S_w = \sum_{j=1}^p \sum_{x_i \in x_j} (x_i - \bar{x}_j)(x_i - \bar{x}_j)^T \tag{2}$$

$$S_b = \sum_{j=1}^p n_j(\bar{x}_j - \bar{x})(\bar{x}_j - \bar{x})^T \tag{3}$$

이 식에서,  $\bar{x}$ 는 평균벡터이며, 전체 분산(total-scatter),  $S_t$ 는 클래스 내 분산과 클래스 간 분산의 합과 같다.

$$S_t = \sum_{i=1}^n n_j(x_i - \bar{x})(x_i - \bar{x})^T = S_w + S_b \tag{4}$$

첫 번째 FDA factor의 목적은 클래스 내 분산은 최소화하고 클래스 간 분산을 최대화 하도록 계산된다. 두 번째 FDA factor는 첫 번째 FDA factor에 수직인 모든 축들 사이에서 클래스 내 분산을 최소화하고 클래스 간 분산을 최대화하도록 계산되어진다. 나머지 FDA factor도 같은 방법으로 계산되어진다. 얻어지는 FDA factor들은 다음 고유치(eigenvalue) 문제에서 고유벡터(eigenvector),  $w_k$ 의 값과 같다.

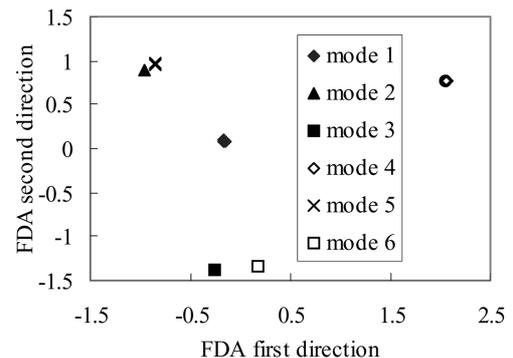
$$s_b w_k = \lambda_k S_w w_k \tag{5}$$

여기서, 고유값  $\lambda_k$ 는 클래스 사이의 전체 분산의 정도를 나타낸다. FDA는 선형 변환이 목적이기 때문에 클래스의 수가 p일 때, FDA factor의 수는 최대 p-1이 된다. 또한, FDA factor로 구성된 고유벡터행렬을  $W_p$ 라고 할 때 데이터  $x_i$ 는 다음과 같이  $z_i$ 로 선형 변환된다.

$$z_i = W_p^T x_i \tag{6}$$

FDA를 이용한 조업모드 판별을 위해 각 조업모드의 정상상태 데이터로부터 FDA 모델을 만든다. 온라인판별을 할 때는 테스트 데이터를 FDA 모델을 이용하여 선형 변환한 다음, 변환된 값과 각 조업모드의 평균값 사이의 거리를 계산한다. 거리가 가장 가까운 조업모드를 선택해서 그 조업모드를 테스트 데이터의 조업 모드라고 판단한다.

Fig. 2는 TE 공정의 6개 조업모드를 FDA 모델로 변환한 다음 2개의 FDA factor에 투영한 결과이다. 이 그림에서는 모드 2와 5, 모드 3과 6의 거리가 가까운데 특히 모드 2와 5의 거리가 지나치게 가깝다. 이로 인해 오분류의 가능성이 매우 높는데 이는 유클리드 거리의 경우(Table 3)와 비슷하다.



**Fig. 2. Projected steady state data on FDA factor.**

3-3. PCA에 의한 판별

PCA는 수백 개의 변수를 몇 개의 주성분(PC, principal component)으로 나타내는 대표적인 차원축소방법이다[10]. 공정변수의 분포에서 가장 큰 변이를 갖는 방향에 대하여 축을 정의하고 이것이 첫 번째 주성분(PC 1)이 된다. 이 축에 원래 변수를 투영시킨 값을 스코어(score) 벡터라고 한다. 다음은 PC 1과 직교하도록 다음으로 큰 변이를 갖는 방향에 대하여 두 번째 주성분(PC 2)를 정한다. 이와 같이 변이가 큰 순서로 PC를 정하면 몇 개의 PC만으로 모든 변수를 대략적으로 나타낼 수 있다. 즉, PCA란 원래 변수들이 가진 전체 변이에 대한 각 성분의 공헌도를 순차적으로 최대화하는 성분을 유도하는 방법이다.

데이터 행렬,  $X$ 는 PCA에 의해 스코어행렬  $T$ 와 가중치행렬  $P$ 의 곱으로 나타내어진다.

$$X = \bar{X} + E = \sum_{i=1}^k t_i p_i^T + \sum_{i=1}^k e_i = TP^T + E \quad (7)$$

여기서  $E$ 는 오차행렬,  $k$ 는 PC의 수이다. 가중치벡터  $p_i$ 는 PC  $i$ 에 대해 변수의 영향력을 나타내는 가중치(loading) 벡터이며,  $t_i$ 는 스코어벡터( $t_i = X p_i^T$ )이다. 즉,  $X$ 를  $k$ 개의 축에 투영시킨 값과 그 가중치의 선형적인 합으로 근사하는 것을 말한다.

PCA에서 PC의 수는 모델의 성능을 결정하는 중요한 인자로 이를 결정하는 방법에는 scree 테스트, cross-validation, parallel analysis 등이 있는데 Ku 등은 가장 좋은 방법으로 parallel analysis를 추천하였다[9, 11]. 이 연구에서는 주어진 데이터와 같은 크기인 임의의 데이터를 발생시킨 다음 주어진 데이터의 고유치가 임의의 데이터의 고유치보다 클 때까지의 PC를 선택하는 parallel analysis를 사용하였다.

PCA를 이용한 분석 도구로서  $T^2$ (Hotelling's  $T^2$  statistic)와 SPE(squared prediction error)가 많이 사용된다. Hotelling's  $T^2$ 는 새로운 데이터가 모델과 얼마나 잘 부합되는지를 보여주는 수치이며, SPE는 모델과의 오차를 측정하는 수치로 Q-statistic이라고도 불린다.

$$SPE = \sum_{i=1}^k e_i^2 \quad (8)$$

Zhao 등은 오프라인에서 각 조업모드의 데이터마다 독립적인 PCA 모델을 만들고 온라인에서는 각 PCA 모델의 SPE를 계산하여 최소 SPE가 얻어진 모델에 해당하는 조업모드를 답으로 제시하는 다중 PCA 모델을 제안하였다[4]. 이 연구에서도 같은 전략을 사용하였으며, 6개 조업모드에 대한 정상상태 데이터로부터 6개의 PCA 모델을 구축하였다.

3-4. 동특성을 반영한 판별

화학공정은 대개의 경우 시간에 따라 변화하는 동적상태에 있다. 이와 같은 특성을 고려하여 공정변수의 시간에 따른 변화를 모델에 추가함으로써 모델의 정확성을 향상시킬 수 있다. Ku 등은 ARIMA 모델과 유사한 방법으로 데이터의 동적정보, 즉 공정변수의 현재값 뿐만 아니라 과거값을 PCA 모델에 포함시킨 DPCA(dynamic PCA) 모델을 제안하였는데 이 개념은 많은 연구에서 응용되고 있다[11, 12]. 데이터 행렬,  $X$ 에  $l$ 개의 과거값을 포함시킨 DPCA 모델의 데이터 행렬,  $X_d$ 는 다음과 같다.

$$X_d = [X(t) \ X(t-1) \ \dots \ X(t-l)] \quad (9)$$

여기서  $l$ 은 과거 값, lag라고 불리며, 흔히 1이나 2인데 Ku 등은 이 값을 결정하는 체계적인 방법을 제안한 바 있다. 결정된 TE 공정의  $l$ 값은 모든 조업모드에서 2이다.

이 연구에서는 DPCA 외에도 ED(유클리드거리) 모델과 FDA 모델의 입력데이터에 공정변수의 과거값을 포함시킨 DED(dynamic ED) 모델, DFDA(dynamic FDA)의 모델도 조업모드 판별모델로서 고려하여 고려된 모델의 수는 총 6가지였다.

4. 예제 및 결과

4-1. 모드 5, IDV1의 판별

모드 5에서 이상 IDV1이 발생한 경우에 대해 조업모드를 판별하였다. 이상발생시간은 1시간이다. IDV1은 4번 흐름에서 B의 조성이 일정하면서 A/C의 비가 감소하게 된다. 이로 인해 6번 흐름에서 A가 감소하고 제어에 의해 1번 흐름의 유량이 증가하게 된다. A의 유량과 조성의 변화는 반응기의 조건을 변화시켜 온도, 압력, 수위를 변화시키게 된다.

Fig. 3(a)는 6개 모드의 centroid와 테스트 데이터 간의 유클리드 거리를 보여주고 있으나 각 모드에서 계산된 거리가 너무 가까워 그림으로는 구분하기 쉽지 않다. 처음에는 모드 2의 유클리드 거리가 가장 짧지만 7시간이 넘게 되면서 모드 5의 거리가 모드 2보다 짧아진다. 이로 인해 Fig. 4(a)에 보이는 것처럼 29.5%의 시간영역에서 모드 2로 오판별하였다. Fig. 3(b)와 Fig. 4(b)는 동특성을 고려한 유클리드 거리 계산에서도 29.2%의 시간영역에서 오판별한다는 것을 보여준다.

Fig. 3(c)는 FDA에 의해 계산된 거리를 보여주고 있는데 7시간 정도 경과하면서 모드 2와의 거리가 모드 5의 거리보다 가까워지면서 오판별하게 된다. Fig. 4(c)에 보이는 것처럼 5.5%의 오판별율을 보여주었다. 그러나, 동특성이 고려된 DFDA는 Fig. 3(d)와 Fig. 4(d)에 보이는 것처럼 오분류가 없었다.

Fig. 3(e)와 (f)는 PCA와 DPCA에 의한 SPE를 보여주고 있는데 다른 4개 모드의 SPE는 지나치게 커서 그림에서는 표시하지 않았다. PCA는 모드 2로 오분류하는 비율이 12.9%였으나(Fig. 4(e)), DPCA는 오분류하지 않았다(Fig. 4(f)). 요약하면 DFDA와 DPCA에 의한 조업모드 판별은 100% 성공하였으며 유클리드 거리를 이용한 판별은 성능이 가장 나빴다.

4-2. 판별 결과

테스트 데이터는 정상상태 데이터와 15개 이상 데이터(IDV1에서 IDV15)로 모사시간은 각각 10시간이다. Table 4는 제안된 6개 방법을 이용한 6개 조업모드의 판별결과로서 정상상태 데이터의 오판별율과 15개 이상 데이터의 평균 오판별율을 보여주고 있다.

훈련데이터인 정상상태 데이터에 대해 오판별한 모델은 유클리드 거리를 이용한 모델들이고 다른 모델들은 오판별율이 0%였다.

FDA와 DFDA 모델은 정상상태 데이터에 대해서는 오판별하지 않았으나 이상데이터에 대한 평균 오판별율은 ED와 DED에 비해 더 높게 나타났다. PCA와 DPCA에 의한 판별결과는 다른 방법들에 비해 월등히 우수하였으며 특히 DPCA는 평균 오판별율이 1% 미만이었다. 대부분의 모델들에서 6개의 조업모드 중에 모드 2와 모

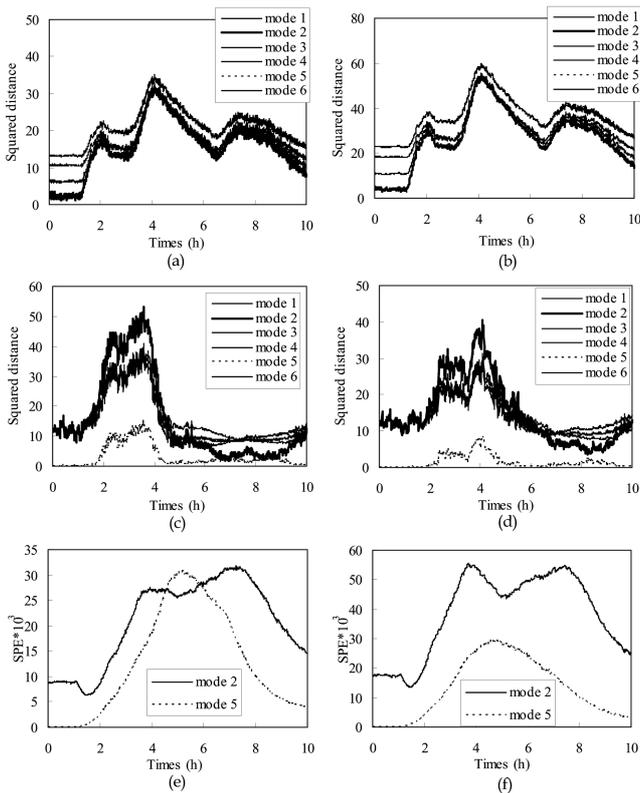


Fig. 3. Distances and SPEs for mode 5 and IDV1 by (a) ED, (b) DED, (c) FDA, (d) DFDA, (e) PCA, (g) DPCA.

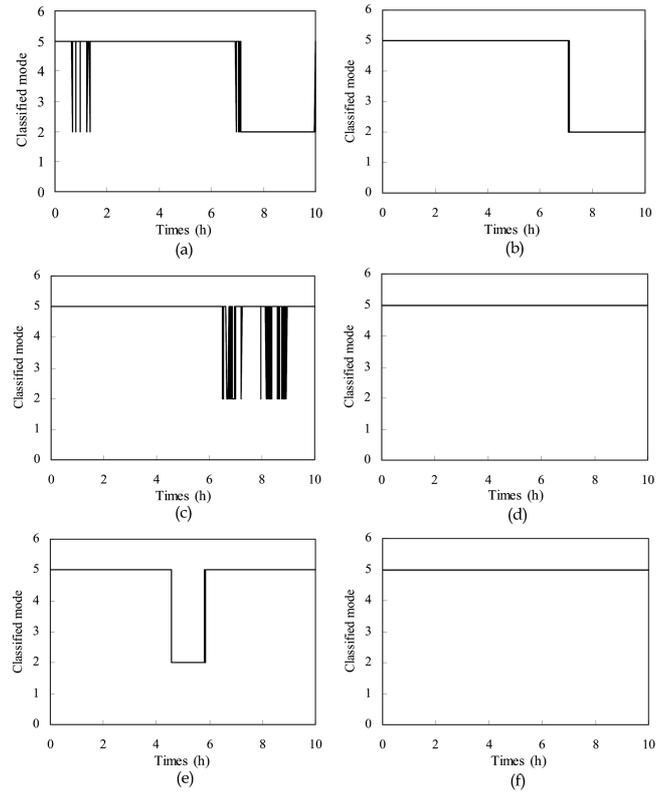


Fig. 4. Mode classification results for mode 5 and IDV1 by (a) ED, (b) DED, (c) FDA, (d) DFDA, (e) PCA, (f) DPCA.

Table 4. Miss-classification rates by 6 statistical models

Mode	Steady-state data						Faulty data					
	ED	DED	FDA	DFDA	PCA	DPCA	ED	DED	FDA	DFDA	PCA	DPCA
1	0%	0.0%	0%	0%	0%	0%	2.3%	2.3%	15.6%	23.6%	0%	0%
2	8.2%	1.1%	0%	0%	0%	0%	24.9%	20.3%	25.0%	24.3%	10.3%	3.5%
3	0%	0%	0%	0%	0%	0%	1.5%	1.2%	14.8%	23.9%	0%	0%
4	0%	0%	0%	0%	0%	0%	0%	0%	30.6%	31.5%	0.8%	0.9%
5	8.9%	1.8%	0%	0%	0%	0%	11.5%	7.1%	9.1%	11.6%	2.7%	0.9%
6	0%	0%	0%	0%	0%	0%	2.4%	2.4%	15.9%	23.3%	0%	0%
Average	2.8%	0.5%	0%	0%	0%	0%	7.1%	5.5%	18.5%	23.0%	2.3%	0.9%

Table 5. Miss-classification rates of faulty data sets by the PCA models of 3 different variables sets

Mode	22 continuous variables		19 composition variables		12 manipulated variables	
	PCA	DPCA	PCA	DPCA	PCA	DPCA
1	0%	0%	0.9%	0.9%	0%	0%
2	4.6%	1.6%	0.8%	0%	14.7%	7.8%
3	13.2%	9.2%	4.0%	0.6%	0%	0%
4	0%	0%	0%	0%	0.9%	1.0%
5	7.5%	9.1%	15.2%	20.8%	6.5%	5.5%
6	0%	0%	1.3%	4.3%	0%	0%
Average	4.2%	3.3%	3.7%	4.4%	3.7%	2.4%

드 5에 대한 오판별율이 높게 나타났는데 이것은 앞에서 예측한 것과 같이 두 조업모드의 설정치 및 변수들이 비슷한 값을 가지기 때문이다. 또한, FDA를 제외한 다른 두 모델에서는 동특성을 반영한

모델이 훨씬 우수한 결과를 보여주었다. 이것은 이상진단 모델과 같이 조업모드 판별에서도 동특성의 고려가 정확성을 향상시킨다는 것을 나타낸다[12].

이 연구에서는 TE 공정의 모든 측정 및 조절변수를 이용하였으나, Zhao 등은 TE 공정에 대한 조업모드 판별 연구에서 22개의 연속변수만을 사용하였다[4]. Table 5의 첫 번째 행은 22개 연속변수만 사용한 PCA 및 DPCA 모델을 이용하여 이상데이터의 조업모드를 판별하였을 때의 오판별율로 각각 4.2%, 3.3%를 보여주어 53개 전체 공정변수를 사용할 때보다 나쁜 결과를 나타내었다.

입력변수의 차이가 판별결과에 미치는 영향을 보기 위해 각각 19개 조성변수만 사용한 PCA 모델 및 12개 조절변수만 사용한 PCA 모델에 의한 판별 결과를 비교하였다. 19개 조성변수 또는 12개 조절변수를 사용한 PCA 모델들도 전체 공정변수를 사용한 ED, FDA 모델보다는 좋은 결과를 보였으나 전체 공정변수를 사용한 PCA 모델에 비해서는 나쁜 성능을 보였다.

## 5. 결 론

통계적 모델, 신경망 등 정량적 데이터에 기반한 이상진단방법은 다중 조업모드를 가진 화학공정에 대한 이상진단에서 오진단의 가능성이 매우 높아진다. 따라서 공정의 조업모드를 판별하고 각 조업모드에 대한 진단모델을 구축하여 진단하는 전략이 필요해진다.

이 연구에서는 조업영역 판별을 위해 통계적 모델인 유클리드 거리, FDA, PCA와 이 모델들에 동특성을 결합한 모델들을 제시하였고 TE 공정에 대한 사례연구를 통해 동특성을 반영한 PCA(DPCA) 모델이 가장 우수한 판별능력을 보임을 확인하였다.

앞으로는 이 연구에서 얻어진 결과를 바탕으로 입력변수의 선택이 조업모드 판별성능에 미치는 영향에 대한 연구가 필요할 것이다.

## 사용기호

$d$	: distance
$E$	: residual matrix for X
$e_i$	: $i$ th residual vector
$k$	: number of PCs
$l$	: number of past values
$n$	: number of variables
$P$	: loading matrix for X
$p$	: number of class
$p_i$	: $i$ th loading vector
$S_b$	: between-class scatter matrix
SPE	: squared prediction error
$S_t$	: total-scatter matrix
$S_w$	: within-class scatter matrix
$T$	: score matrix for X
$t_i$	: $i$ th score vector
$T^2$	: Hotelling's $T^2$ statistic
$w_k$	: eigenvector
$X$	: data matrix
$\bar{x}$	: total mean vector

## 그리스 문자

$\lambda_k$  : eigenvalue

## 감 사

이 논문은 충주대학교 대학구조개혁지원사업비(교육인적자원부 지원)의 지원을 받아 수행한 연구임.

## 참고문헌

1. Venkatasubramanian, V., Rengaswamy, R., Yin, K. and Kavuri, S. N., "A Review of Process Fault Detection and Diagnosis Part I: Quantitative Model-based Methods," *Comput. Chem. Eng.*, **27**(3), 293-311(2003).
2. Venkatasubramanian, V., Rengaswamy, R., Kavuri, S. N. and Yin, K., "A Review of Process Fault Detection and Diagnosis Part III: Process History Based Methods," *Comput. Chem. Eng.*, **27**(3), 327-346(2003).
3. Lee, G., Han, C. and Yoon, E. S., "Multiple-Fault Diagnosis of the Tennessee Eastman Process Based on System Decomposition and Dynamic PLS," *Ind. Eng. Chem. Res.*, **43**(25), 8037-8048 (2004).
4. Zhao, S. J., Zhang, J. and Xu, Y. M., "Monitoring of Processes with Multiple Operating Modes through Multiple Principle Component Analysis Models," *Ind. Eng. Chem. Res.*, **43**(22), 7025-7035(2004).
5. Downs, J. J. and Vogel, F. F., "A Plant-wide Industrial Process Control Problem," *Comput. Chem. Eng.*, **17**(3), 245-255(1993).
6. Ricker, N. L., "Decentralized Control of the Tennessee Eastman Challenge Process," *J. Proc. Cont.*, **6**(4), 205-221(1996).
7. <http://depts.washington.edu/control/LARRY/TE/download.html>.
8. Johnson, R. A. and Wichern, D. W., *Applied Multivariate Statistical Analysis*, 5th Ed., Prentice Hall, Upper Saddle River, NJ (2002).
9. Chiang, L. H., Russell, E. L. and Braatz, R. D., *Fault Detection and Diagnosis in Industrial Systems*, Springer, London(2001).
10. Yoon, D.-M., Lee, Y.-H., Han, C., An, H. S. and Chang, S. Y., "Fault Detection and Diagnosis in Film Processing Plants," *Korean Chem. Eng. Res.*, **41**(5), 585-591(2003).
11. Ku, W., Storer, R. H. and Georgakis, C., "Disturbance Detection and Isolation by Dynamic Principal Component Analysis," *Chemometrics Intell. Lab. Syst.*, **30**(1), 179-196(1995).
12. Lee, G., Song, S.-O. and Yoon, E. S., "Multiple-Fault Diagnosis Based on System Decomposition and Dynamic PLS," *Ind. Eng. Chem. Res.*, **42**(24), 6145-615(2003).