

# 적응형 인자 모델을 이용한 개선된 진공펌프 상태진단에 관한 연구

이규호<sup>a\*</sup> · 이수갑<sup>a</sup> · 임종연<sup>b</sup> · 정완섭<sup>b</sup>

<sup>a</sup>서울대학교 기계항공공학부, 공학연구소, 서울 151-747

<sup>b</sup>한국표준과학연구원, 대전 305-340

(2010년 12월 27일 받음, 2011년 3월 31일 수정, 2011년 4월 20일 확정)

본 논문에서는 건식 진공펌프에서 측정된 다중 변수로 구성된 배치데이터의 통계적인 특성을 소개한다. 흡입구 및 배출구 압력과 부스터/드라이 펌프의 소비전류와 같은 상태변수의 변위분포는 2개나 3개의 특징적인 구간으로 나뉘는 특성이 있다. 이런 관측을 통해 발견한 통계학적 특성을 나타내기 위해 적응형 인자 모델(APM)을 사용하였다. APM 모델기반의 배치 데이터는 건식 진공펌프의 상태를 진단하는데 적절함을 증명하였고, 이전의 동적 시간 왜곡 알고리즘과 비교하였을 때 계산시간 및 필요 메모리 면에서 효율적임을 확인하였다.

주제어 : 진공펌프, 상태진단, 주성분분석, 적응형인자모델, 상태변수

## I. 서 론

국내외 반도체 및 평판 디스플레이 제작공정에서 진공환경을 제공하는 진공 배기 시스템의 상태진단 기술에 대한 필요성은 이전부터 지적되어온 분야이다 [1]. 국내 반도체 및 평판 디스플레이 업체들의 최근 연구는 진공 배기 시스템의 핵심 부품인 건식 진공펌프들의 정밀한 상태진단에 따른 예지 보수에 초점이 맞추어져 있으며, 이는 생산되는 웨이퍼나 플랫 디스플레이의 크기가 커짐에 따른 생산 비용과 시간 단축 즉, 생산성 극대화에 반드시 필요한 기술이기 때문이다 [2].

건식 진공펌프들의 운전 상태는 기계 내부에 부착된 다중 센서들의 출력 값 즉, 상태변수들의 측정치를 이용한다. 다중 상태변수들을 동시에 고려하는 상태진단 기법들 [3]은 다양하게 개발되었지만, 특히 주성분 분석(PCA, Principal Component Analysis)을 통한 Hotelling's  $T^2$  값을 이용하는 기법의 효율성을 이전연구 [4]에서 확인한 바 있었다. 이는 시스템의 상태 변화를 지표로 나타내기 쉽고, 공정의 트렌드를 파악하는데 유용할 뿐 아니라 이상 발생에 기여한 상태변수(contributing state variables)들의 정량적 기여도에 대한 탐색도 가능하기 때문이다 [3].

PCA를 통한 Hotelling's  $T^2$ 는 일정 시간 동안 모은 다중 변수의 측정치("batch"라 호칭)를 하나의 그림으로 보고 그림 속 데이터의 통계적 거리를 합산하여 비교하는 기법이다. 따라서 진공펌프의 운전 상태를 비교하기 위해서는 매번 단위 batch(배치)를 선정해야 한다. 그러나 배치마다 행렬 데이터의 크기가 같지 않아 인위적 재구성을 통한 배치의 크기를 같게 하는 선행과정이 필요하다. 이전 연구 [4]에서는 동적 시간 왜곡(DTW, Dynamic Time Wrapping) 알고리즘 [5,6]을 사용하였다. 그러나 DTW를 이용하여 배치를 동일한 길이로 재구성하는 과정은 계산량이 많고 데이터의 저장 공간 역시 매우 크다. 이는 단일 펌프의 상태진단에는 무리가 없으나 실제 라인별로 수십 대 이상의 펌프가 사용되는 현장에 적용하기에는 개선의 여지가 많다.

본 연구에서는 배치 데이터의 통계적 특성을 몇 개의 모델 인자(model parameter)로 표현하는 적응형 인자 모델(APM, Adaptive Parameter Model)을 제안한다. APM은 배치의 길이(데이터 개수)와 무관하게 배치 데이터의 특성을 나타내는 모델 인자들로 PCA에 사용될 배치 행렬을 재구성하는 방법이다. 제2절에서는 반도체 공정에 사용되고 있는 건식 진공펌프의 다중 상태변수들과 이들의 측정 장치에 대해 소개를 한다, 제3절에서는 우선 측정된 상태변수들의

\* [전자우편] bukha16@snu.ac.kr

통계적 특징을 소개한다. 제4절에서는 다중 상태변수들이 갖는 통계적 특성치 고찰에 따른 APM 기법 개발에 대한 배경을, 제5절에서는 기존의 DTW 기법과 비교함으로써 APM 기법의 개선된 장점들을 자세히 기술한다. 마지막으로 제6절에서는 본 연구의 주요 내용을 간략히 요약한다.

## II. 실험장치 및 측정방법

반도체 공정에 사용되는 건식 진공펌프는 배기속도 증가를 위한 부스터 펌프(booster pump)와 작동 가스들이 실제 배출과 진공도를 유지하는 드라이 펌프(dry pump)로 구성되어 있다.

반도체 공정용 건식 진공펌프의 운전 상태를 측정하는데 사용되는 센서들은 일반적으로 부스터 및 드라이 펌프의 부위별 온도, 압력, 유량, 소비전류 등 십 여종 이상으로 구성 사용하고 있다. 본 연구에서 펌프의 운전 상태를 진단하고자 선정된 센서들은 부스터 펌프 흡입구 압력, 부스터 펌프와 드라이 펌프 각각의 소비전류와 몸체 진동, 그리고 드라이 펌프의 배출구 압력을 선정하였다. 흡입구 압력과 진동 센서는 실제 반도체 공정용 건식 진공펌프에는 아직 사용되지 않고 있다. 본 연구에서는 진공펌프에 인가되는 가스 부하를 측정하고자 흡입구 압력 센서를 설치하였으며 또한 부스터 및 드라이 펌프의 회전운동에서 유발되는 진동을 측정 분석하고자 가속도계를 각각 설치하였다. Fig. 1은 펌프 상태진단에 사용된 센서들의 부착 위치를 각각 보이고 있으며, Table 1에서 선정된 센서와 관련 신호 증폭기에 대한 제원들을 나타냈다.

흡입구와 배출구 압력 센서의 출력 신호 2채널, 부스터

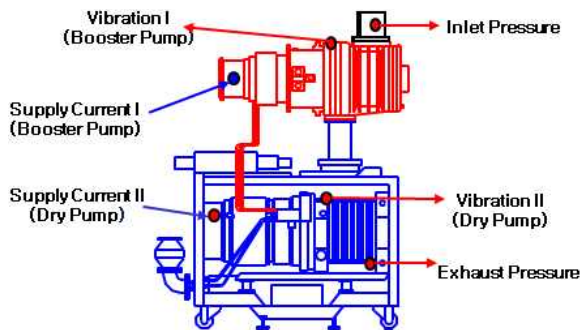


Figure 1. Experimental setup for measurement of the state variables of a dry vacuum pump.

및 드라이 펌프의 구동모터 소비전류 센서의 출력 신호 2채널, 부스터 펌프와 드라이 펌프의 몸체에 장착된 진동 가속도계(단축 모델)의 출력 신호 1채널, 그리고 드라이 펌프 몸체에 부착된 3축 가속도계 출력 3채널의 신호들은 NI사 PXI 사시(Model NI-1000B)에 장착된 실시간 신호 수집 장치(dual NI PXI-4472, 24-bit AD, 16 channels)에 의하여 모두 24-bit의 디지털 신호로 변환되어 서버 급 PC로 저장하였다. 본 연구에서 40.96kHz의 신호 수집 속도를 선정하였으며 수집된 모든 신호는 0.1초마다 두 압력신호의 평균 값과 전류 및 진동신호의 실효치 값을 각각 환산하였다. 드라이 펌프 몸체에 장착된 3축 가속도계의 신호는 3축 벡터 합성된 신호의 실효치 값을 상태변수로 선택하였다.

## III. 측정된 상태변수들의 통계적 특성 분석

Fig. 2는 6시간 동안 측정된 진공펌프의 상태변수들의 시 계열 신호들과 그리고 각 신호들의 분포 특성을 보이고 있다. 이들 그림에 보인 바와 같이 측정된 상태변수 값들이 공정에 따라 두 가지 상태가 번갈아 가며 나타나는 것을 확인할 수 있었다. 즉 펌프 시스템과 연결된 공정 챔버의 개폐밸브가 열려 배출 가스를 펌프가 흡기하는 가스 부하 상태('gas loaded' state)와 펌프 시스템은 운전 중이지만 공

Table 1. Listings of measured state variables and selected sensors.

Measured state variables		Selected sensors
Pressure (mbar)	Inlet pressure	Pfeiffer vacuum, CMR 362 (110 mbar) + MaxiGauge
	Exhaust pressure	Trafag, 8489 model (1.6 bar, absolute, 0.2% F. S.)
Supply current (A)	Booster pump	Taewa Trans., TZ84V/L (100A, 0.2%)
	Dry pump	Taewa Trans., TZ84V/L (100A, 0.2%)
Vibration acceleration (m/s <sup>2</sup> )	Accelerometer: Booster pump	Endevco, model 7210-100 (Uni-axis, 100g, 1%) + Model 136 Amplifier
	Accelerometer: Dry pump	Endevco, model 2230EM (3 axes, 500g, 1%) + Model 136 Amplifier

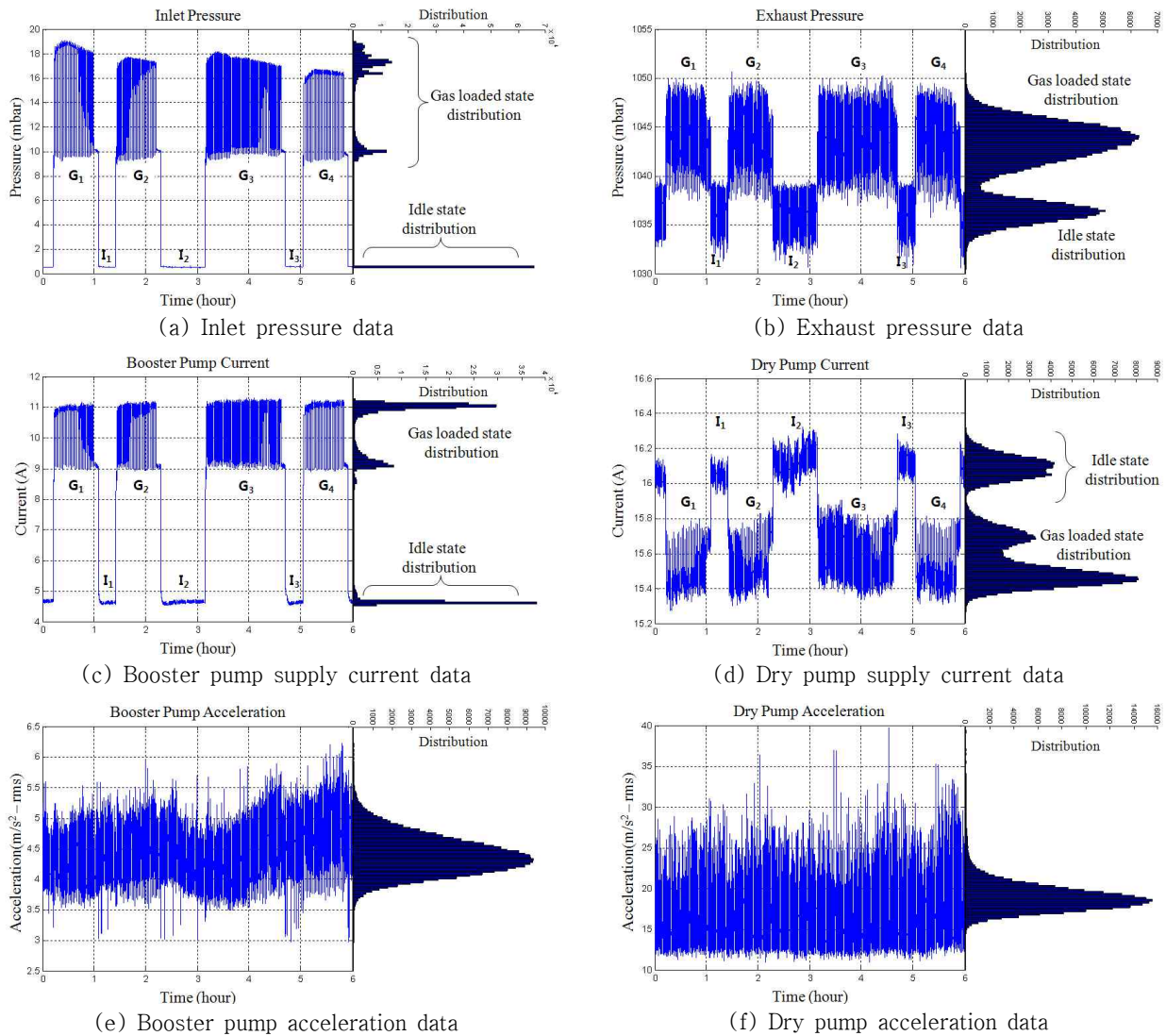


Figure 2. Measured sample data and their distribution characteristics.

정 챔버의 개폐밸브가 닫힌 상태로 배기관 내 기저압력을 유지하는 유힬 상태('idle' state)로 구분된다. 부스터 펌프의 흡입구 압력을 나타내는 Fig. 2 (a)에서 이 두 상태는 공정과정에 따른 가스 부하의 유무에 의해 신호들의 통계적 특성이 확연히 구분됨을 알 수 있었으며, 또한 부스터 펌프의 소비 전류 신호인 Fig. 2 (c)에서도 유사한 특징들이 관측됨을 알 수 있었다. 이들 두 측정 데이터는 가스 부하 상태와 유힬 상태의 측정 신호들의 통계적 분포가 서로 많은 차이점을 보였다. 유힬상태에서는 신호변화가 거의 없는 평균값에 수렴함을 볼 수 있었다. 그러나 가스 부하상태에서는 데이터가 상부 경계와 하한 경계에 둘로 분리되며 두 경계에 집중된 분포를 보였다.

Fig. 2 (b)는 배출구 압력의 변화를 보이는 데이터이며

가스 부하 상태와 유힬 상태에 따라 구분된 2종의 통계적 분포를 보였다. 그리고 Fig. 2 (d)는 드라이 펌프 구동 모터의 소비전류를 나타낸 데이터이며, 이는 유힬 상태에서는 하나의 분포를 갖는 특징을 보였지만 가스 부하 상태에서는 다시 상하 경계로 분리 중복된 2종의 분포를 보였다. 부스터 펌프의 소비전류 신호 분포와 드라이 펌프 소비전류 신호 분포는 서로 반전하는, 즉 부스터 펌프가 가스부하 상태에서 많은 전류를 소비하는 반면에 드라이 펌프는 유힬 상태에서 더 큰 전류를 소비함을 알 수 있었다. 부스터 및 드라이 펌프 몸체의 진동신호에 대한 분포 특성은 Fig 2 (e)와 (f)에 보인 바와 같았으며, 진동 가속도 데이터의 경우 평균값을 중심으로 하나의 분포를 즉, 가스부하에 따라 차이를 보이지 않는 분포 특성이 관측되었다.

이러한 측정 신호들의 분포 특성으로부터 얻어진 첫째 교혼은 반도체 공정용 진공펌프 상태진단 기법의 정확성과 신뢰도 향상을 위해서는 우선 두 상태를 분리하여 따로 진단할 필요성이 있다는 점이다. 둘째로 가스 부하 조건에서 상부 경계와 하부 경계에 둘로 분리된 분포를 보이는 상태 변수들의 특징을 고려한 상태진단 기법의 구현이 또한 필요하다. 본 논문에서는 상태변수들의 분포에 대한 형태가 다양한 가스 부하상태의 데이터들을 이용한 진공펌프의 상태탐지에 초점을 맞추었다.

#### IV. 적응형 인자 모델(APM) 기법

##### 1. 적응형 인자 모델(APM)

앞서 Fig. 2의 (a)부터 (d)에 보인 데이터 분포에서 확인하였듯이 가스 부하 상태에서 데이터는 주로 배치의 상부 경계(upper boundary)와 하부 경계(lower boundary)에 집중적으로 분포한다. 그러므로 데이터의 집중 되는 부분의 점근선, 즉 평균값의 선형적 특징을 해당 배치의 특성으로 대표하는 인자로 활용할 수 있다. 평균값들의 선형적 변화 특성 모델은 다음과 같이 나타낼 수 있다.

$$y_{k,n} = a_k \times n + b_k \tag{1}$$

$$a_k = \frac{N \cdot \sum_{n=1}^N n \cdot y_{k,n} - \sum_{n=1}^N n \cdot \sum_{n=1}^N y_{k,n}}{N \cdot \sum_{n=1}^N n^2 - \left(\sum_{n=1}^N n\right)^2} \tag{2}$$

$$b_k = \frac{\sum_{n=1}^N n^2 \cdot \sum_{n=1}^N y_{k,n} - \sum_{n=1}^N n \cdot \sum_{n=1}^N n \cdot y_{k,n}}{N \cdot \sum_{n=1}^N n^2 - \left(\sum_{n=1}^N n\right)^2} \tag{3}$$

위 Eq. 1은 점근선 y의 일차 선형모델을 나타내며 a는 점근선의 기울기이고 b는 점근선의 초기 절편 값이다. 기울기와 절편 값은 Eq. 2와 Eq. 3을 사용하여 실 측정치 {y<sub>k,n</sub>, n = 1, 2, ..., N}들로부터 환산된다. 아래첨자 k는 상부와 하부 경계선 모델의 구분을 위한 첨자이다. 이는 실제 최소 자승법에 의한 일차 함수의 최적 인자 추정법

과 동일하다. 따라서 추정된 점근선 모델과 수집 데이터와의 차이를 나타내는 표준편차 σ는 다음과 같이 계산된다.

$$\sigma_k = \sqrt{\frac{1}{N} \sum_n (y_{k,n} - a_k \cdot n - b_k)^2} \tag{4}$$

이렇게 각각의 배치에서 구한 파라미터들은 공정시간과 신호들의 특성 변화에도 적용할 수 있는 배치를 대표하는 ‘적응형 인자’이다. 이는 합리적으로 각 배치 간의 데이터 개수를 동일하게 하면서 그 수를 효과적으로 줄일 수 있다.

또한, 이 인자들은 배치의 주요 특성을 내포하고 있으며 모델 인자로부터 원래 측정 데이터들의 대표적인 형상 복원 역시 가능하다는 점이 큰 장점이다. Fig. 3은 표본데이터를 150초 단위로 최댓값과 최솟값을 각각 상부경계와 하부경계로 하여 Eq. 1의 1차 선형 모델을 이용하여 각 배치를 상태변수로 구성한 것이다. 하나의 상태변수 데이터들은 6개의 변수(상부 하부 각각의 점근선 모델 인자 2종과 표준편차 1종)로 변환되며, 총 6종류의 측정신호들로 구성된 한 배치 데이터는 결론적으로 총 36개의 모델인자로 표현된다. Fig. 3에서 나타낸 모델 인자들을 Table 2에 나타내었다. 이렇게 배치의 데이터를 일차 함수의 점근선 모델 인자들과 표준편차로 변환하는 방법은 본 연구팀에 의하여 처음으로 제안되었다 [7,8]. 본 연구는 이러한 적응형 인자 모델의 유효성 검증에 초점을 맞추어 진행되었다.

##### 2. 배치 구성 방법

앞서 측정된 6개의 상태변수로부터 모델 인자들을 얻는 과정을 설명하였다. 이에 따라 흡기 가스 데이터에서는 배치 당 상부와 하부 경계선 영역의 두 점근선으로부터 기울기와 초기 절편 값 그리고 점근선에 대한 표준편차로 각각 구성된 총 6개의 모델 인자를 얻는다. 아래 X<sub>IP</sub>는 흡입구 가스 압력신호로부터 구성된 파라미터 행렬의 예를 보이고 있으며 열의 색인 j는 배치의 순번을 나타낸다.

$$X_{IP} = \begin{bmatrix} a_{u1} & b_{u1} & \sigma_{u1} & a_{l1} & b_{l1} & \sigma_{l1} \\ a_{u2} & b_{u2} & \sigma_{u2} & a_{l2} & b_{l2} & \sigma_{l2} \\ a_{u3} & b_{u3} & \sigma_{u3} & a_{l3} & b_{l3} & \sigma_{l3} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{uj} & b_{uj} & \sigma_{uj} & a_{lj} & b_{lj} & \sigma_{lj} \end{bmatrix} \tag{5}$$

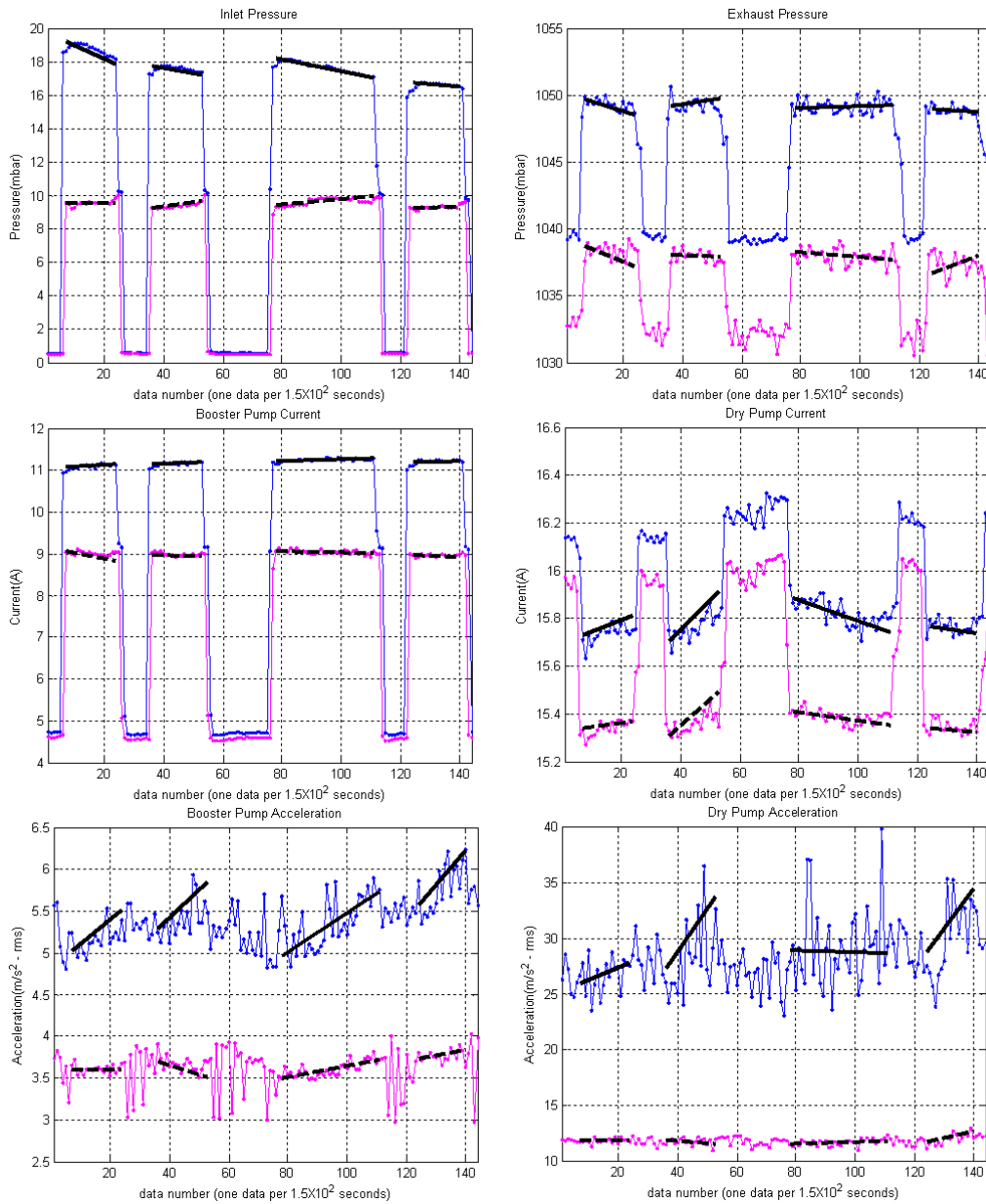


Figure 3. Gas loaded batches reconstructed by the APM algorithm.

동일한 방법으로 남은 5가지 상태변수에 대해 배기압력  $X_{EP}$ , 부스터 펌프 전류  $X_{BC}$ , 드라이 펌프 전류  $X_{DC}$ , 부스터 펌프 진동 가속도  $X_{BA}$ , 드라이 펌프 진동 가속도  $X_{DA}$  행렬을 각각 구성한 뒤 Eq. 6과 같이 상태변수들의 행렬들을 구성함으로써 상태진단에 사용하는 배치  $X$ 를 구성한다.

$$X = [X_{IP} \quad X_{EP} \quad X_{BC} \quad X_{DC} \quad X_{BA} \quad X_{DA}] \quad (6)$$

배치 행렬  $X$ 의 각 열은 동일한 물리적 단위를 갖지만 각 행은 서로 다른 물리적 단위를 가진 측정값들이다. 따라서

상태변수 모델 인자들의 기여도를 동일하게 조절해야 하므로 Eq. 7과 같은 정규화 과정(centering & scaling)을 통해  $X_N$  행렬을 만든다 [9].

$$X_N(i,j) = \frac{X(i,j) - \bar{X}(j)}{\sigma(X(j))} \quad (7)$$

위 식에서  $i$ 는 행벡터의 색인을, 위 첨자 “-”는 열벡터의 평균값을 그리고  $\sigma$ 는 열벡터의 표준편차를 각각 나타낸다.

Table 2. Model parameters evaluated from the sample data of four batches.

State variables	Pump operation		Process G1	Process G2	Process G3	Process G4
	Region	Symbol	(3109s)	(3117s)	(5605s)	(3118s)
Inlet pressure [mbar]	Upper bound	$a_U$	$-8.00 \times 10^{-2}$	$-3.13 \times 10^{-2}$	$-3.57 \times 10^{-2}$	$-1.61 \times 10^{-2}$
		$b_U$	$1.92 \times 10^1$	$1.78 \times 10^1$	$1.82 \times 10^1$	$1.67 \times 10^1$
		$\sigma_U$	$3.91 \times 10^{-2}$	$3.42 \times 10^{-2}$	$4.87 \times 10^{-2}$	$2.51 \times 10^{-2}$
	Lower bound	$a_L$	$1.97 \times 10^{-3}$	$2.41 \times 10^{-2}$	$1.57 \times 10^{-2}$	$1.18 \times 10^{-3}$
		$b_L$	$9.55 \times 10^0$	$9.24 \times 10^0$	$9.43 \times 10^0$	$9.29 \times 10^0$
		$\sigma_L$	$1.95 \times 10^{-2}$	$3.40 \times 10^{-2}$	$1.37 \times 10^{-1}$	$3.50 \times 10^{-2}$
Exhaust pressure [mbar]	Upper bound	$a_U$	$-6.98 \times 10^{-2}$	$3.39 \times 10^{-2}$	$6.79 \times 10^{-3}$	$-1.20 \times 10^{-2}$
		$b_U$	$1.05 \times 10^3$	$1.05 \times 10^3$	$1.05 \times 10^3$	$1.04 \times 10^3$
		$\sigma_U$	$3.03 \times 10^{-1}$	$3.17 \times 10^{-1}$	$4.88 \times 10^{-1}$	$2.78 \times 10^{-1}$
	Lower bound	$a_L$	$-9.26 \times 10^{-2}$	$-6.96 \times 10^{-2}$	$-1.69 \times 10^{-2}$	$7.98 \times 10^{-2}$
		$b_L$	$1.04 \times 10^3$	$1.04 \times 10^3$	$1.03 \times 10^3$	$1.03 \times 10^3$
		$\sigma_L$	$4.49 \times 10^{-1}$	$2.48 \times 10^{-1}$	$4.84 \times 10^{-1}$	$6.57 \times 10^{-1}$
Booster pump current [A]	Upper bound	$a_U$	$4.17 \times 10^{-3}$	$2.08 \times 10^{-3}$	$1.78 \times 10^{-3}$	$1.54 \times 10^{-3}$
		$b_U$	$1.11 \times 10^1$	$1.12 \times 10^1$	$1.12 \times 10^1$	$1.12 \times 10^1$
		$\sigma_U$	$3.38 \times 10^{-2}$	$2.22 \times 10^{-2}$	$2.21 \times 10^{-2}$	$1.66 \times 10^{-2}$
	Lower bound	$a_L$	$-1.38 \times 10^{-2}$	$-1.19 \times 10^{-2}$	$-1.99 \times 10^{-3}$	$-2.40 \times 10^{-3}$
		$b_L$	$9.07 \times 10^0$	$8.97 \times 10^0$	$9.07 \times 10^0$	$8.96 \times 10^0$
		$\sigma_L$	$1.69 \times 10^{-2}$	$2.09 \times 10^{-2}$	$4.05 \times 10^{-2}$	$1.82 \times 10^{-2}$
Dry pump current [A]	Upper bound	$a_U$	$4.62 \times 10^{-3}$	$1.23 \times 10^{-2}$	$-4.43 \times 10^{-3}$	$-1.94 \times 10^{-3}$
		$b_U$	$1.57 \times 10^1$	$1.57 \times 10^1$	$1.58 \times 10^1$	$1.57 \times 10^1$
		$\sigma_U$	$1.67 \times 10^{-2}$	$3.04 \times 10^{-2}$	$3.67 \times 10^{-2}$	$1.95 \times 10^{-2}$
	Lower bound	$a_L$	$1.78 \times 10^{-3}$	$1.09 \times 10^{-2}$	$-1.76 \times 10^{-3}$	$-8.90 \times 10^{-4}$
		$b_L$	$1.53 \times 10^1$	$1.53 \times 10^1$	$1.54 \times 10^1$	$1.53 \times 10^1$
		$\sigma_L$	$1.75 \times 10^{-2}$	$2.12 \times 10^{-2}$	$2.20 \times 10^{-2}$	$1.44 \times 10^{-2}$
Booster pump acceleration [m/s <sup>2</sup> ]	Upper bound	$a_U$	$2.88 \times 10^{-2}$	$3.33 \times 10^{-2}$	$2.35 \times 10^{-2}$	$4.10 \times 10^{-2}$
		$b_U$	$5.02 \times 10^0$	$5.29 \times 10^0$	$4.96 \times 10^0$	$5.57 \times 10^0$
		$\sigma_U$	$1.30 \times 10^{-1}$	$2.05 \times 10^{-1}$	$2.06 \times 10^{-1}$	$1.98 \times 10^{-1}$
	Lower bound	$a_L$	$3.47 \times 10^{-4}$	$-1.05 \times 10^{-2}$	$6.70 \times 10^{-3}$	$6.95 \times 10^{-3}$
		$b_L$	$3.60 \times 10^0$	$3.70 \times 10^0$	$3.49 \times 10^0$	$3.72 \times 10^0$
		$\sigma_L$	$3.74 \times 10^{-2}$	$5.63 \times 10^{-2}$	$5.63 \times 10^{-2}$	$5.94 \times 10^{-2}$
Dry pump acceleration [m/s <sup>2</sup> ]	Upper bound	$a_U$	$1.04 \times 10^{-1}$	$3.80 \times 10^{-1}$	$-5.59 \times 10^{-3}$	$3.57 \times 10^{-1}$
		$b_U$	$2.60 \times 10^1$	$2.73 \times 10^1$	$2.88 \times 10^1$	$2.87 \times 10^1$
		$\sigma_U$	$1.30 \times 10^0$	$3.13 \times 10^0$	$3.25 \times 10^0$	$2.69 \times 10^0$
	Lower bound	$a_L$	$-2.51 \times 10^{-3}$	$-1.89 \times 10^{-2}$	$8.85 \times 10^{-3}$	$6.27 \times 10^{-2}$
		$b_L$	$1.19 \times 10^1$	$1.19 \times 10^1$	$1.15 \times 10^1$	$1.16 \times 10^1$
		$\sigma_L$	$2.25 \times 10^{-1}$	$2.53 \times 10^{-1}$	$2.48 \times 10^{-1}$	$2.99 \times 10^{-1}$

## V. 실험결과 및 고찰

### 1. Hotelling's T<sup>2</sup>

앞 절에서 소개된 배치 재구성 방법에 따라 가스 부하조건

에서 측정된 다중 상태변수들은 최종적으로 행렬 형태의 배치로 구성된다. 공정 진행에 따라 얻어진 배치 데이터의 변화를 하나의 대푯값으로 표시하는 기법인 Hotelling's T<sup>2</sup> 환산 기법을 본 연구에서도 이용하였다. Hotelling's T<sup>2</sup> 값은 각 공정마다 측정된 배치들을 하나의 그림으로 간주하고 공정



진행에 따라 기록되는 각각의 그림들이 어느 정도 차이를 보이는가를 환산하는 일종의 다차원 거리 환산 기법이다.

$T^2$  값을 구하기 위해 전체 배치 데이터를 HDS와 ODS로 나눈다. HDS (Historical Data Set)는 초기에 정상 작동 범위 내에서 공정이 진행된 상태의 데이터이고, ODS (Observed Data Set)는 현재 펌프의 상태를 나타내는 데이터이다. HDS 구간에 대한 ODS의 변화를 통해 펌프 상태를 파악할 수 있다. 이전 연구 [4]에서 전체 데이터 중 HDS 구간은 초기 5일간의 데이터로 설정하는 것이 가장 효율적임을 밝힌 바 있다.

배치 데이터는 Eq. 7에 의해 36개의 파라미터에 대해 동일한 기여도를 부여한다. 그러나 실제로 36개의 파라미터들이 펌프상태를 반영하는 정도는 다를 수 있다. 상태 반영도가 작은 과잉 정보들을 그대로 합산하면  $T^2$  계산 시에 값이 이상 증대하는 오류가 발생하게 된다. 이를 보정하기 위해 주성분 분석 기법(PCA)을 도입하게 된다. 주성분 분석은 좌표축을 회전시켜 통계적인 기여도가 큰 데이터 성분을 추출하는 기법이다.

HDS 구간의 공정 데이터에 대한 파라미터 행렬인  $N_{ROW} \times N_{COL}$  크기인  $X_{HDS}$ 을 가정해보면,  $N_{ROW}$ 번의 공정 횟수동안  $N_{COL}$ (36)개 썩의 파라미터가 축적되었다고 볼 수 있다. 이  $X_{HDS}$ 를 이용해 주성분 분석(PCA)을 하면,  $N_{COL} \times N_{COL}$  크기의 고유벡터행렬로 이루어진 주성분 계수(Principal component coefficient) 행렬  $U$ 와  $N_{COL}$ 개의 고유근  $\lambda$ 로 이루어진 대각행렬  $S_D$ 를 얻는다. 이때 고유근이 주성분의 상태 반영도가 되므로 Eq. 8과 같이 고유근의 누적값이 총 고유근의 합의 95%가 될 때까지의 값을 유의미한 성분이라고 본다.

$$\sum_{k=1}^P \lambda_k \leq 0.95 \times \sum_{k=1}^{N_{COL}} \lambda_k \quad (8)$$

위 방식으로 주성분 계수 행렬  $U$ 는  $N_{COL} \times P$ 로 차원 축소된다. 이때  $U$ 의 열벡터인 고유벡터는 서로 직교하며 독립적이다. 여기서  $X_{HDS}$ 에 주성분 계수 행렬을 곱함으로 고유 벡터를 주좌표로 사영시킨  $N_{ROW} \times P$  크기의 주성분 (Principal Component) 행렬  $T_S$ 를 구할 수 있다 ( $T_S = X_{HDS} \times U$ ). 이때 주성분 행렬의 각각의 열벡터가 시스템의 상태를 나타내는 주성분이 되며,  $T^2$ 는 다음과 같이 구해진다.

$$S_D^{-1} = \frac{1}{\lambda_1} U_1 U_1^T + \frac{1}{\lambda_2} + \dots + \frac{1}{\lambda_p} U_p U_p^T \quad (9)$$

$$T_i^2 = (T_{Si}) S_d^{-1} (T_{Si})^T = \sum_j \frac{(T_{Si,j})^2}{\lambda_j} \quad (10)$$

이로써 정상구간인 HDS에서의  $T^2$ 값을 구하였다. 이를 기준으로 실제 펌프의 이상 유무를 판단할 구간인 ODS의  $T^2$ 값을 구하도록 한다. ODS에서는 ODS의 공분산 행렬을 이용한 고유벡터를 사용하는 것이 아니라, 정상상태인 HDS에서 구한 고유벡터  $U$ 를 사용하여 주 좌표 변환을 한다 ( $T_S = X_{ODS} \times U$ ). 이렇게 얻은 주성분 행렬  $T_S$ 로 Eq. 10와 동일한 방법으로 ODS의  $T^2$ 값을 구할 수 있다. 이와 같은 과정을 통해 ODS의 단일 변수 값의 변화뿐 아니라 변수간의 상관관계 변화 역시  $T^2$ 값에 영향을 주게 되므로 다변수간의 상태 진단이 가능하다.

## 2. 실험 결과 및 신뢰도 검증

앞 절에서의와 같은 방법으로 구한  $T^2$ 값의 변화를 Fig. 4에 도시하였다. 검증을 위해 동일한 RAW 데이터를 DTW와 APM으로 각각 재구성한 배치에 대한 결과를 비교하였

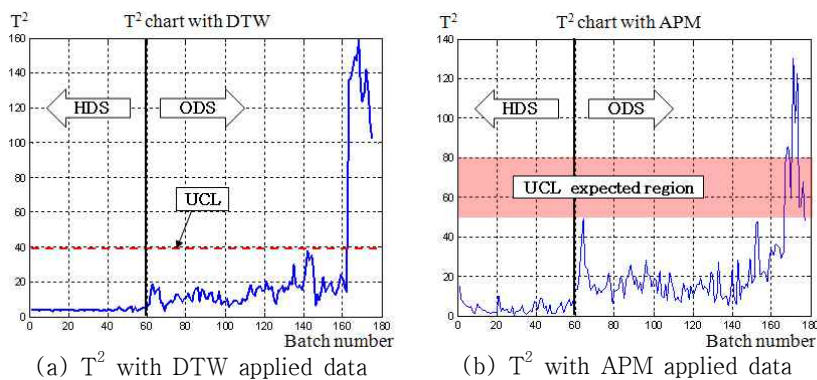


Figure 4. Hotelling's  $T^2$  statistics chart with 6 sensor signals.

다. Fig. 4 (a)는 DTW가 적용된 배치 데이터를 이용한  $T^2$  값의 변화를 나타낸 그래프이고, Fig. 4(b)는 APM이 적용된 배치 데이터를 이용한  $T^2$  값의 변화를 나타내었다. 모두 비슷한 패턴을 보이며 동일한 지점에서 이상 발생을 감지함을 확인할 수 있다. 하지만, Fig. 4와 같이 비선형적인 변화를 보이는 경우 외에도 선형적인 변화를 보이는 경우에 대해 합리적인 이상 발생 감지를 위해서는, 정량적인 판단 기준인  $T^2$  값의 정상범위의 동작 한계선인 Upper critical limit (UCL)을 설정할 필요가 있다. UCL은 주로 HDS의  $T^2$ 에 대해  $\text{mean} + \text{factor} \times \sigma$ 로 제안된다. Factor는 이전 연구 [4]에서 20으로 제안된 바 있다. 이를 DTW에 대한 결과 A에 적용한 결과  $\text{mean}(3.79) + \text{factor}(20) \times \sigma(1.79) = 39.6$ 로 Fig. 4(a)에 적용하면 잘 맞는다. 동일한 데이터에 대해 계산을 하였으므로 APM를 사용한 결과에 대해 적용을 하여도 맞아야 한다. 결과 Fig. 4(b)의 그래프를 보면 UCL이 존재해야 하는 구간은 60~40이며, 실제 UCL은  $\text{mean}(3.91) + \text{factor}(20) \times \sigma(2.95) = 62.91$ 로 UCL이 신뢰성 있는 구간에 위치한다. 그러므로 이상발생 감지를 동

일하게 했다고 볼 수 있다.

### 3. DTW 기법과 APM 기법의 비교

개선된 자기진단 평가 기법의 효율성을 평가하기 위해 DTW 기법과 APM 기법의 계산량 및 계산시간과 데이터 저장량을 정량적으로 파악하였다. 이를 위해 정규화에 필요한 주요 계산식의 비교를 통해 감소한 계산량을 평가하고 동일한 조건하에서 계산을 수행한 경우 평균적인 계산 시간의 감소량 역시 확인하였다. 또한 데이터 저장에 관하여, 계산에 사용되는 주요 변수 행렬의 크기를 비교하여 필요한 데이터 저장량 비교 역시 병행하였다. 정규화 과정에서 주요 부분을 차지하는 식과 식에 해당하는 계산량을 Table 3과 Table 4에 비교 정리하였다. DTW로 구성하는 행렬의 경우 하나의 배치에 레퍼런스 배치를 기준으로 6개의 센서에서 약 300개씩의 데이터가 입력되므로 행-300, 열-6이다. 반면 APM은 배치를 사용자 임의로 나눌 수 있으므로 행벡터의 길이 조절이 가능하며 안정적인 계산을

Table 3. Computation amount of DTW processing with one batch. Processing step equations are cited from the equation (22~24) of Ref [4] ( $N_{\text{ROW}} : \sim 300, N_{\text{COL}} : 6, \text{Iteration number} : 10$ ).

Equation	Approximated computation number
1) the local distance processing step: $d(i, j) = [B_i(i, :) - B_{\text{REF}}(j, :)] \cdot W \cdot [B_i(i, :) - B_{\text{REF}}(j, :)]^T$	$2 \times N_{\text{ROW}}^2 \times N_{\text{COL}}^3$
2) the minimum accumulated distance processing step: $D_A(i, j) = \min \begin{cases} D_A(i-1, j) + d(i, j) \\ D_A(i-1, j-1) + d(i, j) \\ D_A(i, j-1) + d(i, j) \end{cases}$	$2 \times N_{\text{ROW}}$
3) the weight matrix computation step: $W(j, j) = \left[ \sum_{i=1}^I \sum_{k=1}^{b_{\text{REF}}} [\bar{B}_i(k, j) - \bar{B}(k, j)]^2 \right]^{-1}$	$3 \times N_{\text{ROW}} \times N_{\text{COL}}$
Sum of total computation number	$\frac{N_{\text{ROW}} \times (2 \times N_{\text{ROW}} \times N_{\text{COL}}^3 + 3 \times N_{\text{COL}} + 2)}{\sim 3.8 \times 10^7}$

Table 4. Computation amount of APM processing with one batch ( $N_{\text{ROW}} : \sim 20, N_{\text{COL}} : 36$ ).

Equation	Approximated computation number
Upper and Lower bound separation	$1,500 \times N_{\text{COL}}$
1) slope coefficient (Eq. 2): $a_k$	$(N_{\text{ROW}}^2 + 2N_{\text{ROW}}) \times N_{\text{COL}}$
2) initial value of batch (Eq. 3): $b_k$	$(N_{\text{ROW}}^2 + 2N_{\text{ROW}}) \times N_{\text{COL}}$
3) standard deviation (Eq. 4): $\sigma_k$	$4 \times N_{\text{ROW}} \times N_{\text{COL}}$
Sum of total computation number	$\frac{(2 \times N_{\text{ROW}}^2 + 8 \times N_{\text{ROW}} + 1,500) \times N_{\text{COL}}}{\sim 8.9 \times 10^4}$



Table 5. Comparison of the data storage space requirement per one batch data using the size of state variable matrix and vectors.

DTW		APM	
Matrix	Size	Matrix (or vector)	Size
B	300×6	Max value	20×36
B <sub>ref</sub>	300×6	Min value	20×36
d	300×300	Mean vale	20×36
W	6×6	Coefficient of equations	2×3
Sum	93636	Sum	2166

위해 10개의 단위로 나누었다. 그리고 6개의 센서에서 6개 씩의 인자를 추출하므로 하나의 배치행렬의 크기는 행-10, 열-36이다. 하나의 배치 데이터를 정규화 하는데 필요한 계산 량은 DTW의 경우 10의 7승 그리고 APM의 경우 10의 4승으로 약 1/1000로 계산 량이 감소함을 확인할 수 있다. 또한 이는 DTW의 결과 수렴을 배제한 결과로 각 배치 데이터의 데이터양에 대한 편차가 크다면 수렴을 위해 반복 계산이 필요하므로 DTW는 수렴 횟수만큼의 배수로 계산 량이 증가하게 된다.

실제로 이 계산을 수행한 결과 DTW는 평균 43.26(초)가 소요된 반면 APM은 평균 3.38(초)가 소요됨에 따라 1/10 배 이상의 계산 시간을 단축함을 확인하였다.

그리고 데이터 저장을 위한 변수 행렬을 비교하면 DTW는 B, B<sub>ref</sub>, d, W의 4가지 행렬이 필요하고, APM은 max value vector, min value vector, mean value vector, coefficient equation parameters가 필요하다. 이 행렬의 크기를 비교하면 Table 5와 같다. 이를 통해 필요한 데이터 저장 공간이 약 100분의 1로 줄어들음을 확인할 수 있다.

## VI. 결 론

본 논문에서는 반도체 공정에서 측정된 6종류의 상태변수들의 진폭에 대한 통계적 분포함수의 특징을 상세히 소개하였다. 특히 펌프의 흡입구와 배출구 압력 그리고 부스터 펌프와 드라이 펌프의 소비전류 신호들의 진폭 신호에 대한 통계적 분포 특성이 최소 2~3개의 독립된 분포의 영역이 존재함을 확인하였다. 그리고 이들 신호의 상부와 하부의 점근적 분포특성을 분리하여 그들의 고유한 변화 특성을 모델링하는 방법을 새로이 제안하였다. 그리고 부스터 펌프와 드라이 펌프에 장착된 진동센서의 가속도 신호 등은 표준분

포와 유사한 특성을 보였지만 이들 신호 또한 상부와 하부의 점근적 모델을 각각 분리하여 이들 신호의 변화 특성을 모델링하는 방법을 제안하였다. 이러한 진공펌프의 상태변수들의 공정 가변성으로 말미암은 배치 데이터 통계적 특성들을 고려하고자 기존의 DTW 대신 적응형 인자 모델(APM) 알고리즘을 사용하였다. 반도체 공정에 사용되는 진공펌프들의 상태진단을 위해 이전연구에서 사용한 PCA를 통한 Hotelling's T<sup>2</sup>값이 매우 효과적임을 확인할 수 있다. 이전의 DTW 기법으로 얻었던 진공펌프의 이상 유무를 APM 모델을 이용한 배치결과로 부터도 동일하게 판별할 수 있었다. 이 논문에서 제안한 APM 모델기법은 DTW에 비해 계산 량의 관점에서 1/1000 수준, 계산시간에서 1/10 수준 그리고 데이터 저장량에서 1/100 수준임을 확인하였다.

## 감사의 글

본 연구는 서울대학교 공학연구소를 경유하여 협약한 지식경제부 지원 전략기술개발사업 중 “차세대 반도체용 진공공정의 실시간 측정/진단/제어 기술개발” 중 제 3 세부과제인 “스마트형 진공 배기 진단 제어 시스템 개발” 과제(과제번호: 10031858)와 지식경제부의 재원으로 한국 에너지 기술평가원(KETEP)의 지원을 받아 수행된 연구과제(과제번호: 20104010100490, 20094020100060)의 연구결과 중 일부입니다.

## 참고문헌

- [1] R. Bahren and M Kuhn, Vacuum **44**, 709 (1993).
- [2] S. Y. Lee, M. Noh, B. O. Kim, and A. S. Lee,

- J. Korean Vacuum Soc. **19**, 265 (2010).
- [3] R. L. Mason and J. C. Young, *Multivariate Statistical Process Control with Industrial Applications*, (ASA SIAM, Philadelphia, 2002). Chap.1-6.
- [4] D. Sung, J. Kim, W. Jung, S. Lee, W. Cheung, J. Lim, and K. Chung, J. Korean Vacuum Soc. **15**, 338 (2006).
- [5] A. Kassida, J. F. MacGregor, and P. A. Taylor, *Process systems Engineering*, AIChE Journal **44**, 846 (1998).
- [6] V. Pravdova, B. Walczak, and D. L. Massart, *Analytica Chimica Acta* **456**, 77 (2002).
- [7] W. S. Cheung, J. Y. Lim, K. H. Chung, and S. G. Lee, U. S. Patent No. 7,653,512 (2010).
- [8] W. S. Cheung, J. Y. Lim, K. H. Chung, and S. G. Lee, U. S. Patent No. 7,664,618 (2010).
- [9] R. Bro and A. K. Smilde, *J Chemometr* **17**, 16 (2003).
- [10] E. J. Dudewicz and S.N. Mishra, *Modern Mathematical statistics*, (Wiley, New York, 1988).
- [11] W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical recipes in Fortran the art of scientific computing second edition*, (Cambridge university press, New York, 1997).
- [12] J. E. Jackson, *A User's guide to Principal Component*, (John Wiley and Sons, 1991).

## Study on Vacuum Pump Monitoring Using Adaptive Parameter Model

Kyuho Lee<sup>a\*</sup>, Soogab Lee<sup>a</sup>, Jong-Yeon Lim<sup>b</sup>, and Wan-Sup Cheung<sup>b</sup>

<sup>a</sup>*Engineering Research Institute, Department of Mechanical and Aerospace Engineering, Seoul National University, Seoul 151-747*

<sup>b</sup>*Korea Research Institute of Standards and Science, Daejeon 305-340*

(Received December 27, 2010, Revised March 31, 2011, Accepted April 20, 2011)

This paper introduces statistical features observed from measured batch data from the multiple operation state variables of dry vacuum pumps running in the semiconductor processes. The amplitude distribution characteristics of such state variables as inlet pressures, supply currents of the booster and dry pumps, and exhaust pressures are shown to be divided into two or three distinctive regions. This observation gives an idea of using an adaptive parametric model (APM) chosen to describe their statistical features. This modelling, in comparison to the traditional dynamic time wrapping algorithm, is shown to provide superior performance in computation time and memory resources required in the preprocessing stage of sampled batch data for the diagnosis of running dry vacuum pumps. APM model-based batch data are demonstrated to be very appropriate for monitoring and diagnosing the running conditions of dry vacuum pumps.

Keywords : Vacuum pump, Diagnosis, PCA, APM, State variables

\* [E-mail] bukha16@snu.ac.kr