

MPCA 기반의 통계기법을 이용한 진공펌프 상태진단에 관한 연구

성동원^{1*} · 김재환¹ · 정원태¹ · 이수갑¹ · 정완섭² · 임종연² · 정광화²

¹서울대학교, 서울 151-747

²한국표준과학연구원, 대전 305-600

(2006년 3월 20일 받음)

반도체 공정에 사용되는 진공펌프는 가혹한 운전조건과 비선형적 특성으로 인하여 고장시점을 정확히 예측해내기가 어려운데 이로 인해 불량품이 양산되거나 불필요한 재원이 낭비되는 등의 문제가 발생하게 된다. 따라서 펌프의 운전상태를 올바르게 모니터링하고 고장 지점을 정확히 인지해 적절한 펌프 교체 시점을 알려주는 진공펌프 상태진단 모델의 개발은 매우 시급하고도 중대한 문제라 할 수 있겠다. 본 연구에서는 다변량 통계기법을 이용하여 영향력 있는 인자들을 종합적으로 고려하였으며 최종적으로 Hotelling's T² 통계량을 이용한 진공펌프 상태진단 모델을 제안하였다. 핵심적인 알고리즘으로는 Multiway Principal Component Analysis(MPCA)와 Dynamic Time Warping Algorithm(DTW Algorithm) 기법 등이 사용되었다.

주제어 : 진공펌프, 배치, 다중주성분분석, 동적 시간 왜곡 알고리즘, 통계적 거리

1. 서 론

반도체 공정에 사용되는 진공펌프는 가혹한 운전조건과 비선형적 특성으로 인하여 고장시점을 정확히 예측해내기가 어려운데 이로 인해 불량품이 양산되거나 불필요한 재원이 낭비되는 등의 문제가 발생하게 된다. 따라서 펌프의 운전상태를 올바르게 모니터링하고 고장 지점을 정확히 인지해 적절한 펌프 교체 시점을 알려주는 진공펌프 상태진단 모델의 개발은 매우 시급하고도 중대한 문제라 할 수 있다.

진공펌프의 상태진단 모델을 개발한다는 것은 곧 펌프 성능을 대변하는 주요 인자를 찾아내어 시스템의 오작동 여부를 판별하는 시스템을 구축하는 것을 의미한다. 즉, 상태진단 시스템이 완성되면 펌프의 운전 상태가 정상적인 범위를 벗어날 시에 이상징후가 자동 감지되어 경고가 발동됨으로써 사용자가 이에 반응해 펌프의 동작을 멈추거나 기타 다른 조치를 취할 수 있도록 기능하게 된다. 따라서, 이러한 시스템을 구축하는데 있어 이상징후 여부의 기준이 되는 주요 인자를 찾아내는 것이 관건이 되며 이를 통해 정상동작범위와

이상동작범위의 경계구간을 설정하게 된다.

본 연구에서는 진공펌프에 설치된 9개의 센서들로부터 신호를 얻어내 모든 분석의 기본 자료로 사용하였는데 각 신호가 상호 독립적이지 못하고 어느 정도의 상관성을 지닌다는 사실에 근거하여 각 신호를 독립적으로 분석하기보다 전반적인 추세를 반영할 수 있는 다변량 통계기법을 사용하게 되었다. 즉, MD (Mahalanobis's Distance) 또는 Hotelling's T²이라 불리는 통계적 거리(Statistical Distance) 기법을 사용하여 9개의 신호변수를 동시에 고려하여 T²이라는 하나의 통계량으로 정보를 축약하였다 [1]. T²은 개별 변수들의 이상거동뿐만 아니라 변수들간의 함수관계에 대한 고려도 포함하므로 개별신호들만의 관측으로는 얻을 수 없는 변수간 정보까지 포함하는 통계량이다.

결국, T²을 이용해 구현하게 되는 최종적인 진공펌프 상태진단 모델은 T² 값이 어느 한도를 벗어나면 시스템이 이상징후를 보이는 것으로 간주하게 되며 이때의 한계지점인 UCL(Upper Control Limit)을 어떠한 방식으로 설정할 것인가가 본 연구의 핵심사항이라고 할 수 있겠다.

II. 실험 및 분석

2.1. 실험방법

실험을 위한 진공펌프는 루츠형 부스터펌프와 스크류형 드라이펌프로 구성되었으며, 데이터를 얻기 위하여 크게 4종류의 센서를 부착하였다. 즉, 압력센서 2개, 가속도센서 4개, 음향센서 1개, 전류센서 2개를 사용하였으며, 표 1에는 센서로 측정되는 9개의 신호를 나타내었다.

진공펌프에 전류를 가하면 진공상태를 만들기 위해 압력이 형성되고 이러한 과정은 진동과 음향을 수반하므로 센서의 선정은 타당하다고 볼 수 있다. 물론 이외에도 진공펌프의 특성에 영향을 미치는 인자들이 많지만 일단 모니터링 분석 기법이 확립되면 이후에 변수를 추가하는 문제는 쉽게 구현될 수 있을 것으로 보인다.

표 1. 센서로 측정되는 9개의 신호

변수	측정 신호
1	Inlet Pressure
2	Booster Pump Acceleration
3	Dry Pump Acceleration X
4	Dry Pump Acceleration Y
5	Dry Pump Acceleration Z
6	Acoustic Pressure
7	Exhaust Pressure
8	Booster Pump Current
9	Dry Pump Current

2.2. 분석절차

분석절차는 그림 1과 같이 크게 1 단계(Phase 1)와 2 단계(Phase 2)의 두 과정으로 이루어진다. 1 단계에서는 진공펌프 교체설치 후 일정기간 동안을 정상동작 범위로 가정하고 이를 기반으로 기준자료로 사용할 T^2 의 값을 취득하게 된다. 진공펌프는 작동을 시작한 이후로 지속적으로 성능이 저하되므로 초기구간을 정상동작범위로 설정하는 것은 타당하다고 볼 수 있으며 이는 실제 많은 모니터링기법에서 사용하는 방식이다. 이 때 정상동작으로 간주된 데이터를 HDS(Historical Data Set)라 명명한다. 2 단계에서는 기준자료와의 비교를 통해 실제 모니터링을 수행하게 되며 HDS이후의 데이터를 ODS(Observed Data Set)으로 명명한다. 세부적인 사항은 이후에 자세히 논의하기로 한다.

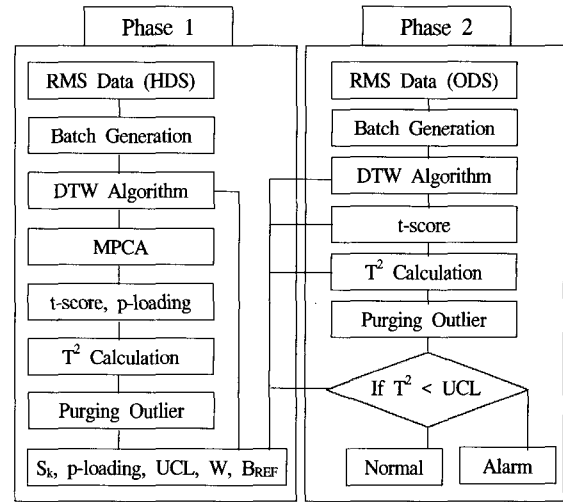


그림 1. 순서도

(1) 배치 생성(Batch Generation)

표 1의 센서로 기록된 신호들은 반도체 제작 공정이 수행되지 않는 구간들의 정보도 함께 포함하고 있으므로 실제 공정이 진행되는 구간들만을 별도로 추출하여 배치데이터로 저장한다. 즉, 배치란 하나의 공정단위를 의미한다. 공정이 진행되는 동안에는 흡입압력이 높은 값을 유지하다가 공정을 마치고 다음 공정이 시작되기까지는 흡입압력이 거의 0에 가깝게 값이 떨어지게 되므로 센서 1의 신호파형을 관측하면 공정의 진행여부를 판별할 수 있으며 나머지 변수들도 동일한 구간에서 공정과정의 데이터만을 선별하여 배치데이터로 저장한다.

(2) 동적시간 왜곡 알고리즘

(Dynamic Time Warping Algorithm)

반도체 공정의 소요시간은 상황에 따라 달라지므로 배치길이는 저마다 달라지게 된다. 이를 그대로 방치한 상태로 분석을 수행하게 되면 이후에 공분산행렬을 구하는 과정에서 역행렬이 발산할 우려가 있으므로 배치길이를 사전에 동일하게 맞춰주어야 한다.

이 때 배치길이를 조절하는 과정에서 초기입력 추세를 그대로 보존하는 것이 중요한데 이를 위해 Dynamic Time Warping(DTW) 기법을 도입해 사용한다. DTW 기법은 음성인식 분야에서 널리 사용되고 있는 기법으로서 Kassidas (1998) 등이 이를 배치공정에 적용한 바 있다[2].

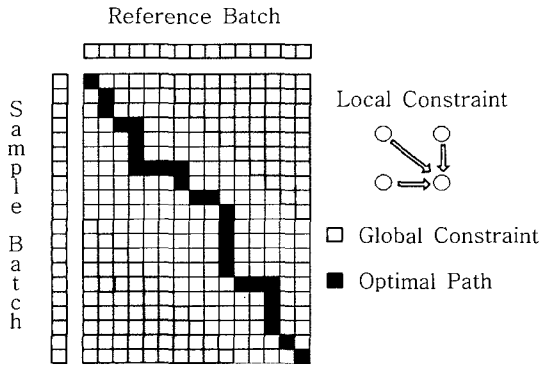


그림 2. 배치 길이 조절

표본배치(Sample Batch)의 길이를 참고배치(Reference Batch)의 길이와 동일하게 조정하기 위해 정해진 규제조건과 알고리즘에 따라 다음과 같이 최적 경로(Optimal Path)를 선정한다.(그림 2 참조)

$$d(i, j) = [B_i(i, :) - B_{REF}(j, :)] \cdot W \cdot [B_i(i, :) - B_{REF}(j, :)]^T \quad (1)$$

$$D_A(i, j) = \min \begin{cases} D_A(i-1, j) + d(i, j) \\ D_A(i-1, j-1) + d(i, j) \\ D_A(i, j-1) + d(i, j) \end{cases} \quad (2)$$

$$W(j, j) = \left[\sum_{i=1}^I \sum_{k=1}^{b_{REF}} [\tilde{B}_i(k, j) - \bar{B}(k, j)]^2 \right]^{-1} \quad (3)$$

식 (1), (2), (3)과 같이 국부거리 $d(i, j)$ 와 누적거리 $D_A(i, j)$ 를 구하는 것이 주된 알고리즘인데 이 때 가중치 W 가 수렴할 때까지 반복수행한다. 10번 정도의 반복수행으로 W 가 수렴하는 것을 확인할 수 있었다. B 는 배치데이터를 의미한다. 자세한 알고리즘은 참고문헌 [2]를 참조하였다.

$D(i, j)$ 가 구해지고 나면 Local-global Constraint에 따라 최적경로(Optimal Path)를 선정하고 이를 기반한 평균화과정을 통해 배치길이를 동일하게 맞춘다.

그림 3에 실제로 $D(i, j)$ 를 구한 후 최적경로(Optimal Path)를 산출한 결과를 나타내었다.

(3) 다중주성분분석

(Multiway Principal Component Analysis)

배치길이를 일정하게 조절함으로써 다중주성분분석(MPCA)를 수행할 수 있는 기본적인 조건이 만족되었지만 본격적인 분석에 앞서서 센터링(Centering)과 스케일링(Scaling)의 사전작업을 수행하여야 한다[3].

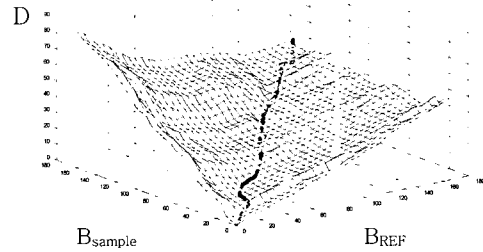


그림 3. 최적 경로 산출

MPCA의 목적은 대표적인 배치의 거동특성을 파악하는 것이기 때문에 각 배치들은 동일한 조건에서 그 특성이 비교되어야 한다. 각 배치들은 저마다 다른 평균값을 갖게 되는데 데이터에서 평균값을 빼 주는 형식으로 평균지점을 일괄적으로 0으로 맞춰주는 센터링을 수행하면 이러한 문제가 해결된다.

또한 9개의 변수들은 저마다의 단위체계를 가지고 있으므로 이에 대한 고려가 필요하다. 같은 값이라 하더라도 어떠한 단위를 쓰는가에 따라 그 비중이 달라지게 되므로 이를 보정하지 않을 경우 특정 변수의 데이터에 과도한 가중치가 부여되어 전체적인 분석에 있어 오류를 범할 수 있기 때문이다. 이는 각 데이터를 표준편차로 나눠주는 스케일링 과정을 통해 해결할 수 있다.

배치데이터는 여러 개의 변수를 가지고 있고 또한 시간의 진행에 따른 데이터를 가지고 있으므로 본질적으로 3차원의 데이터 구조가 된다. 이 상태에서 직접 센터링, 스케일링을 수행하여도 되지만 보다 시각적으로 명확하고 용이한 해석을 위하여 시간축에서 한 단 계씩 데이터를 잘라내어 횡으로 배열하면 그림 4와 같은 2차원 배열을 얻을 수 있다. 이제 종방향으로 평균과 표준편차를 구한 후 각 데이터에서 평균값을 빼주는 센터링과, 각 데이터를 표준편차로 나눠주는 스케일링을 수행하면 된다.

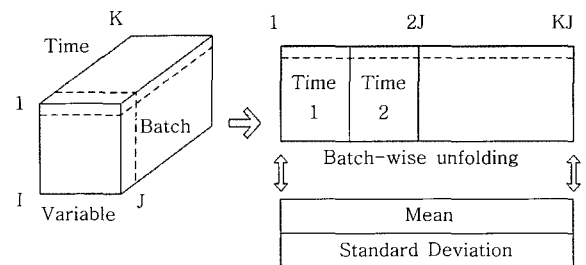


그림 4. 사전 데이터 처리

위와 같은 사전작업(Data Preprocessing)이 끝나면 비로소 다중 주성분 분석을 수행할 수 있는 모든 조건이 만족되게 된다.

다중 주성분 분석에 적합한 형태로 행렬을 재배열하면 그림 5와 같다.

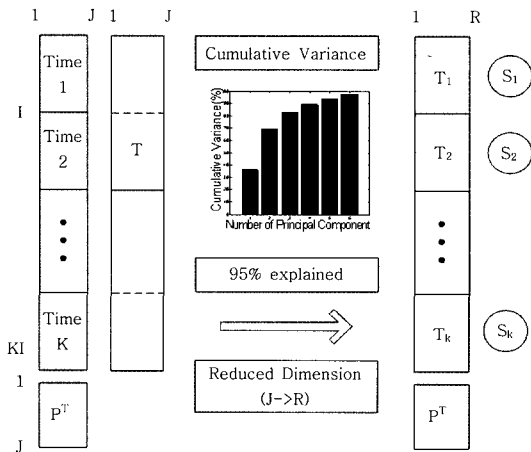


그림 5. 다중 주성분 분석

주성분분석은 좌표축을 회전시켜 기여도가 큰 데이터 성분을 추출하는 기법이다. 이 때 고유벡터가 변환된 주좌표로 사용되며 고유근은 각 주좌표 방향으로의 데이터 분산을 의미하게 되므로 결국 고유벡터와 고유근을 결정하는 것이 중요한 문제가 된다. 이는 Matlab 내부함수에서 SVD(Singular Value Decomposition) 기법을 이용해 간단히 구현된다. 이후 기존 데이터들을 각 고유벡터에 정사영시키면 주성분들을 얻을 수 있고 이 때, 각 주성분들은 고유근에 해당하는 분산을 가지게 된다. 그런데 만약 기존 데이터내에 과잉정보(redundant information)가 포함되어 있을 경우 고유근이 매우 작은 주성분들이 나타나게 되는데 이를 그대로 방치하면 이후의 T^2 계산시에 값이 이상증대하는 오류가 발생하게 된다. 주성분을 t_{score} 라 하고, 주좌표벡터(고유벡터)를 $p_{loading}$ 으로 명명하면 T^2 은 다음과 같이 계산된다.

$$T^2 = t_{score} S_D t_{score}^T = \sum_i \frac{t_{score,i}^2}{\lambda_i} \quad (4)$$

이 때 공분산행렬 S_D 는 좌표변환 후 고유근을 대각 요소로 갖는 대각행렬이 되므로 결국 T^2 은 고유근의 역수로 가중치가 부여된 t_{score} 의 제곱합이 된다. 따라

서 고유근이 0에 가까운 작은 값을 갖게 되면 T^2 값이 크게 증가하게 되는데 실제로 별 비중이 없는 성분이 T^2 값을 크게 만드는 상황이므로 이는 본래의 추세를 제대로 반영했다고 보기 어렵다.

이러한 이유로 주성분은 차원을 축소해야 하며 이는 고유근의 관찰을 통해 이루어진다.

$$C.V._m(\%) = \frac{\sum_{j=1}^m \lambda_j}{\sum_{i=1}^9 \lambda_i} \times 100 \quad (5)$$

즉, 식 (5)를 이용하여 누적분산을 구해나가다가 값이 95%를 넘어서기 전까지의 변수의 개수를 축소된 차원으로 정의한다[4].

이상의 주성분분석과정을 1 단계의 HDS에 적용하면 t_{score} , $p_{loading}$, T^2 이 구해지고 이후 2 단계의 ODS에서는 다시 $p_{loading}$ 을 구하지 않고 이미 구해진 $p_{loading}$ 에 데이터를 정사영시켜 t_{score} 를 구하고 고유근과의 조합을 통해 T^2 을 구하게 된다.

(4) 이상점 제거(Purging Outliers)

이상점(Outlier)은 데이터 중 대다수의 추세를 벗어나 특이한 경향을 보이는 몇몇 데이터를 일컫는 용어인데 이를 그대로 방치하게 되면 전체의 추세를 추정함에 있어 심한 왜곡이 발생할 수 있으므로 이를 적절히 제거하여 시스템 본래의 추세를 올바르게 보존하여야 한다.

이상점 제거는 우선 데이터의 분포곡선을 추정하여 가장 끝단의 일부 데이터를 제거한 후 재구성된 데이터의 분포곡선을 다시 추정하는 방식을 반복하면서 추정된 분포가 안정화될 때까지 반복수행을 실시하는 방식으로 이루어진다.

분포를 추정하는 방식은 식 (6)과 같이 Kernel Smoothing Technique [1]을 사용한다.

$$f(t) = \frac{1}{n} \sum_{j=1}^n \phi(t - T^2) / h \quad (6)$$

이 때 ϕ 는 평균 0, 분산 1의 정규분포함수이고 T^2 값의 분포로부터 확률분포의 합이 1이라는 제약조건을 이용해 각 정규함수의 h 값을 결정하면 정규함수의 과형이 결정되고 이들의 선형조합으로 최종 분포가

결정된다.

그림 6과 같이 일단 분포가 결정되면 오른쪽 꼬리 부분에서 α (%)에 해당하는 비율의 T^2 지점을 찾아 UCL로 선정한 후 이 값을 넘어서는 T^2 값을 데이터에서 제외시킨다. 만약 이상점의 비율이 그리 높지 않다면 이러한 작업을 한 번만 수행하여도 분포가 안정화되지만 이상점이 많을 경우 이상점 제거 작업을 반복하여 추정되는 분포가 더 이상 변화하지 않고 수렴될 때까지 작업을 계속하여야 한다.

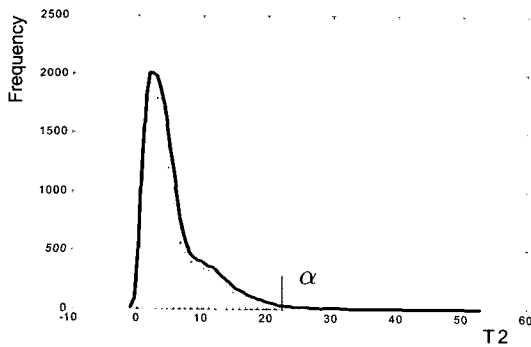


그림 6. Kernel Smoothing 기법

표 2에서는 실제 이상점 제거 과정을 보여주고 있는데 각 단계에서의 UCL과 제거되는 이상점의 개수를 나타내고 있다.

표 2. 이상점 제거 과정

	Pass1	Pass2	Pass3	Pass4	Pass5	Pass6	Pass7
UCL	14.9	12.8	11.7	10.9	10.4	10.0	9.7
No. of Outliers Removed	144	157	146	153	141	122	92

III. 결과 및 고찰

3.1. HDS 범위 선정

HDS(Historical Data Set)는 정상동작범위의 데이터로서 이후의 ODS(Observed Data Set)을 모니터링하기 위한 기준자료를 제공한다. HDS는 정상동작의 양상을 대표할 만큼의 충분한 길이를 확보해야 하지만 무작정 길게 선정하게 되면 그만큼 실제 모니터링을 수행할 수 있는 기간이 짧아지기 때문에 적절한 수준의 타협점을 찾아야 한다.

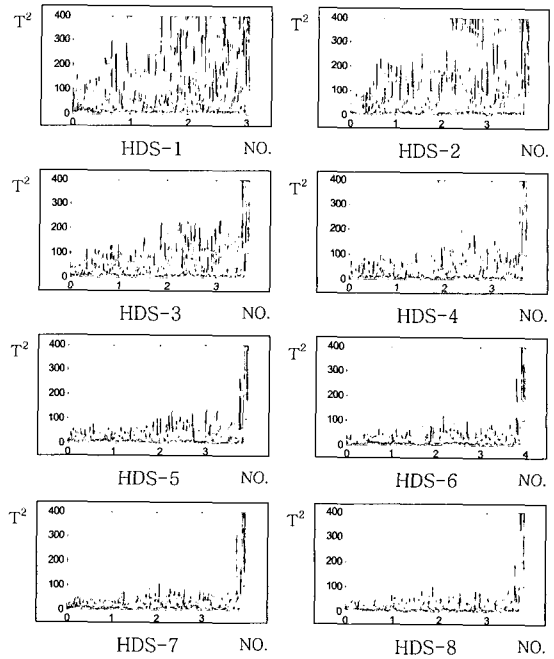


그림 7. HDS 기간 비교분석

그림 7은 진공펌프 운전 데이터를 이용해 HDS 기간을 1~8일 구간에서 바꿔가면서 데이터를 분석한 결과이다. 가로축 No.는 배치들을 일렬로 모아놓은 것으로 공정의 시간순서라고 생각하면 된다.

이 결과에서 알 수 있듯이 HDS가 너무 작으면 대표성이 결여되어 T^2 값이 쉽게 발산한다는 것을 알 수 있으며 HDS가 늘어날수록 T^2 분포는 점차 안정화되어 일정 수준에서 수렴하는 것을 관측할 수 있다. 즉, 그림 7에서 보면 HDS-5 이후로는 어느 정도 추세가 수렴하는 것을 알 수 있으며 총 6개의 사례에 대해 조사해본 결과 모두 같은 현상이 관측되었으므로 HDS는 5일 정도가 적절할 것으로 판단된다.

3.2. HDS 재설정 이유

HDS는 정상동작범위의 기준이 되는 데이터로서 진공펌프 교체 후 초기 5일간의 자료로 구성된다. 그런데 만약 모든 펌프가 정상동작시 동일한 성능을 나타낸다고 가정하면 펌프가 교체되더라도 기존의 HDS를 그대로 사용할 수 있는 상황을 생각해 볼 수 있다. HDS에서 기준자료 데이터를 추출하는 작업은 그 자체로 많은 시간이 소요되므로 만약 HDS를 한 번만 설정해 사용하는 것이 가능하다면 연산속도 향상에도 도움을 줄 수 있다. 이런 상황이 가능한지 직접 데이터 분석을 통해 확인해 보았다.

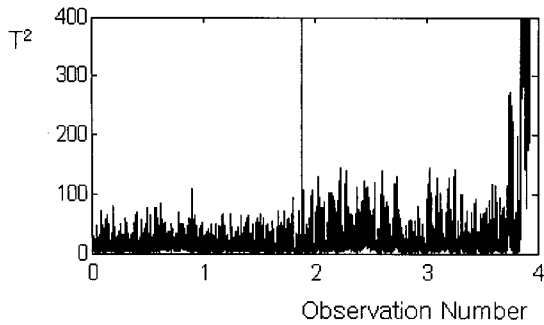


그림 8. 펌프 교체후의 T² 선도의 거동

그림 8은 펌프운전에 따른 T² 선도인데 펌프는 중반부에 교체되었다. 펌프가 교체되었으나 기존펌프와의 특성을 비교하기 위해 교체펌프의 초기자료를 HDS로 재설정하지 않고 의도적으로 ODS로 사용하였다. 그림에서 알 수 있듯이 펌프교체 직후에 T²값은 급격하게 증가하였다. 만약 교체된 펌프와 기존 펌프의 정상동작 성능이 동일하였다면 펌프 교체 후의 T²값은 그림 8의 초반부의 T² 분포와 비슷한 양상을 보였을 것이다.

현재 6가지 case의 펌프데이터를 보유하고 있는데 나머지 5개의 case에 대해서도 결과를 분석한 결과 펌프 교체 후 T²값이 증가하는 현상이 모든 case에서 항상 관찰되었으므로 펌프의 정상동작 특성은 제각기 다르다는 사실이 확인되었고 따라서 펌프가 교체되면 반드시 HDS를 재설정해 주어야 한다는 결론을 얻을 수 있다.

3.3. UCL의 선정

이제 HDS에 대한 정의가 명확해졌으므로 최종적으로 UCL을 선정하는 작업이 남아있다.

UCL은 결국 정상 동작범위의 한계선이기 때문에 그 근거를 HDS에서 찾을 수 있다. 즉, HDS의 T² 분포로부터 UCL이 결정되는 것이다. UCL을 결정짓는 방식은 기법에 따라 차이가 있는데 본 연구에서는 HDS T²의 평균과 표준편차를 이용하는 기법을 택하였다.

$$UCL = mean + factor \times \sigma \quad (7)$$

흔히 3-sigma, 6-sigma 등의 용어로 명명되어 사용되지만 이를 단순히 무조건적으로 사용할 수는 없고 실제적인 분석을 통해서 Factor를 구해내는 일이 관건이라고 하겠다.

본 연구에서 사용한 진공펌프 데이터는 총 6구간의

자료였고 그 중 하나는 정상동작만으로 이루어진 데이터였다. 즉, 진공펌프가 설치되어 작동하다가 작업을 중지하고 새로운 펌프로 대체한 경우의 센서신호데이터를 6구간에 대해 보유하고 있다. 따라서, 정상동작만으로 이루어진 자료로부터 우선적으로 대략적인 Factor를 감지해내고 이후에 나머지 5가지 자료로부터 구체적인 수치를 규정짓는 방식으로 연구를 진행하였다.

3.4. T² 분석 및 Factor 선정

우선 정상동작만으로 이루어진 펌프자료를 살펴보자. 이 때 추세의 명확성을 나타내기 위해 HDS뿐만 아니라 ODS에서도 이상점을 제거하였고 이후로 논의되는 모든 경우에 있어서도 동일한 작업을 수행하였다.

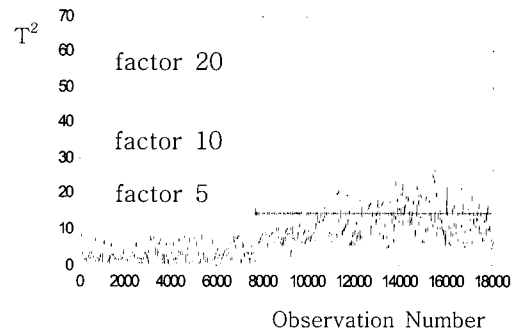


그림 9. 정상동작 펌프 데이터

그림 9에는 Factor가 각각 5, 10, 20인 UCL이 나타나 있고 HDS는 Observation Number 7691까지이다. 정상동작임에도 불구하고 잘못된 경고를 보내지 않도록 하기 위해 UCL은 T² 데이터 윗부분에 존재하여야 한다. 이로부터 UCL은 대략적으로 Factor 10 ~ 20 사이에서 결정될 것으로 생각할 수 있다.

이제 나머지 5가지 Case를 기반으로 UCL Factor를 결정해 보자.

그림 10의 각 차트에서 아래의 수평선은 Factor = 10, 위 수평선은 Factor = 20인 UCL 지점을 나타낸다. 또한 각 차트의 초반부 5일은 HDS로 사용되었으며 진공펌프는 중반부에 교체되었다.

그림 10을 관찰해 보면 T²값의 변화양상이 크게 두 가지 경우로 나뉘어짐을 알 수 있다. 즉, Case 1, 3, 5의 경우에는 펌프 교체시점 이전에 이미 T²값이 크게 상승한 상태였는데 반해 Case 2, 4의 경우에는 T²값이 상대적으로 작은 값을 유지하다가 펌프가 교체되고 나서야 큰 값으로 상승하였다.

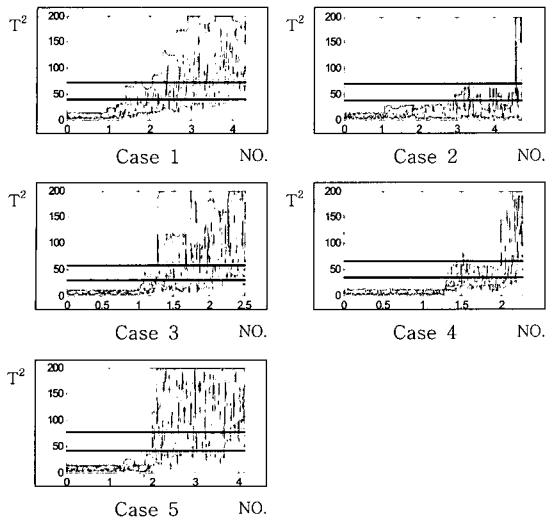


그림 10. UCL factor 선정

현재 반도체 생산 현장에서는 나름의 평가기준을 가지고 진공펌프의 교체시점을 결정하고 있는데 이는 단순히 외부로 드러나는 몇 가지 현상만을 관찰하는 방식으로 이루어지기 때문에 시스템의 전반적인 특성을 포괄적으로 고려했다고 볼 수 없으며 따라서 펌프 교체시점 또한 절대적인 것으로 간주할 수 없다.

이런 의미에서 본다면 Case 1, 3, 5의 경우에는 펌프 교체시기가 최적 시점보다 늦어졌다고 생각할 수 있으며 Case 2, 4의 경우에는 너무 성급하게 펌프를 교체한 것으로 추측해볼 수 있다. 이 경우 펌프를 교체하지 않고 계속 작동시켰다면 Case 1, 3, 5와 같이 T^2 값이 증가하는 현상을 관찰할 수 있었을 것이다.

따라서 Case 2, 4의 경우는 모두 정상동작을 보였다고는 단정할 수 없지만 적어도 T^2 값이 UCL을 크게 벗어나지는 않았을 것이라 생각할 수 있으며 이는 UCL Factor 결정에 대한 단서를 제공하게 된다. 즉, UCL Factor는 10 ~ 20 사이에서 결정되어야 한다는 이전 결과와 함께 Case 2, 4의 T^2 분포를 고려하면 UCL Factor는 20으로 선정해 볼 수 있겠다.

이제 UCL Factor가 20으로 확정되었으므로 이를 기반으로 그림 10의 각 차트를 분석해보자.

Case 2, 4의 경우에는 펌프가 교체되기 이전에 잠시 국부적으로 T^2 값이 UCL을 넘어서는 구간이 있지만 그 추세는 오래 지속되지 않고 곧 정상범위로 회복이 되므로 이러한 경우에는 정상동작으로 간주할 수 있겠다. UCL이라고 하는 것은 고장의 가능성을 제시하는 하나의 지표일 뿐 그 자체로 절대적인 수치가 아니므로 국부적인 T^2 의 변화에 주목하기보다는

전체적인 T^2 추세가 UCL을 기준으로 어떻게 변모하는가에 초점을 맞추어야 한다.

한편 Case 1, Case 3, Case 5의 경우에는 펌프 교체시기 이전부터 이미 고장의 징후가 나타났는데 특히 Case 1과 Case 5에 주목할 필요가 있다. 전자의 경우에는 T^2 값이 어느 시점부터 UCL을 넘어 점진적으로 증가하는 형태를 띠고 있고 있는 반면 후자의 경우, T^2 값이 어느 순간 급작스럽게 증가하여 계속 유지되는 현상을 보였다. 다시 말해, Case 1의 경우에는 펌프가 완전히 고장이 나서 작업에 지장을 초래하기 이전에 점진적으로 운전상태가 나빠지는 징후를 포착하기 용이한 반면, Case 5에서는 한 순간에 고장이 나기 때문에 사전에 고장징후를 파악할만한 겨를이 없는 상황이 되었다. Case 3에서는 이 두 가지의 상황이 복합적으로 나타난 것으로 보인다.

이를 진공펌프의 고장 요인과 종합하여 생각해 보면 두 가지 패턴이 극명하게 드러난다. 즉, 펌프의 성능이 점차적으로 저하되는 고장요인과 한 순간에 급격히 성능이 떨어지는 고장요인이 있을 것이다. 전자는 사전예측이 용이한 반면 후자는 사전감지가 어려운 것으로 여겨진다.

N. 결 론

다변량 통계기법인 MPCA기법을 이용하여 진공펌프 상태진단에 대한 연구를 수행하였다. 통계적인 기법을 이용함으로써 전체적인 시스템의 특성을 T^2 이라는 하나의 대표적 인자로 표현하였으며 UCL을 선정하여 정상동작 판단의 기준으로 삼았다.

이상의 연구에서 얻은 결론을 정리하면 다음과 같다.

- 1) 흡입압력 변수를 기준으로 배치데이터를 선별해 내는 알고리즘을 개발하였다.
- 2) 이상점 제거 기법을 HDS 뿐만 아니라 ODS에도 적용함으로써 T^2 의 분포 추세를 명확하게 판별해 낼 수 있었다.
- 3) HDS의 최적화된 길이를 조사하여 5일간의 데이터셋이 적절하다는 결론을 얻었다.
- 4) 펌프 교체 이후로 T^2 값이 급격히 상승하는 것을 확인함으로써 각 펌프는 서로 특성이 명확히 다르고 매 펌프교체시마다 HDS를 재설정해야 한다는 사실을 확인하였다.
- 5) T^2 이 UCL을 순간적으로 벗어나더라도 곧 정상범

위로 추세가 돌아오면 이상징후로 볼 수 없다고 간주하였다.

- 6) UCL의 기준으로는 $\text{mean} + 20 \times \text{sigma}$ 가 적절하다는 결론을 얻었다.

본 연구에서는 신호들의 상관관계에 기반을 둔 다변량 통계분석기법의 활용을 바탕으로 가혹한 운전조건하에서 작동하는 반도체 공정용 진공펌프의 상태진단 모델을 제시하였으며, 실제 펌프에서 얻어진 신호를 바탕으로 상태 이상을 예측할 수 있는 가능성을 보여주었다. 이를 바탕으로 현장에서 많은 데이터를 축적하고 실제적인 물리적 인자나 고장 요인 등과의 접목을 시도한다면 보다 실제적이며 정교한 상태진단 모델을 제시할 수 있을 것이다.

감사의 글

본 연구는 진공기술기반구축사업의 일환으로 연구되었으며 한국표준과학연구원의 재정적 지원과 과학재단 특정기초 연구 R01-2006-000-10301-0의 지원을 받았습니다. 지원에 감사드립니다.

참고 문헌

- [1] Robert L. Mason, and John C. Young, *Multivariate Statistical Process Control with Industrial Applications*, (ASA SIAM, Philadelphia, 2002), Chap. 1-6.
- [2] Athanassios Kassidas, John F. MacGregor, and Paul A. Taylor, *AIChE Journal* **44**, 864 (1998).
- [3] Rasmus Bro and Age K. Smilde, *Journal of Chemometrics* **17**, 16 (2003).
- [4] Sergio Valle, Weihua Li, and S. Joe Qin, *Ind. Eng. Chem. Res.* **38**, 4389 (1999).
- [5] Robert L. Mason, Young-min Chou, and John C. Young, *Journal of Quality Technology* **33**, 466 (2001).
- [6] Elaine Martin, Julian Morris, and Steven Lane, *IEEE Control, Systems Magazine*, 26 (2002).
- [7] Jong-Min Lee, ChangKyoo Yoo, and In-Beum Lee, *Computers & Chemical Engineering* **27**, 1903 (2003).
- [8] Jong-Min Lee, ChangKyoo Yoo, and In-Beum Lee, *Journal of Biotechnology* **110**, 119 (2004).
- [9] Henk-Jan Ramaker, Eric N.M. van Sprang, Johan A. Westerhuis, and Age K. Smilde, *Analytica Chimica Acta* **498**, 133 (2003).
- [10] V. Pravdova, B. Walczak, and D.L. Massart, *Analytica Chimica Acta* **456**, 77 (2002).
- [11] J. Edward Jackson, *A User's Guide to Principal Components* (John Wiley & Sons, Inc., Canada, 1991), Chap. 1-2.

Study on Vacuum Pump Monitoring Using MPCA Statistical Method

D. Sung^{1*}, J. Kim¹, W. Jung¹, S. Lee¹, W. Cheung², J. Lim², and K. Chung²

¹*Seoul National University, Seoul 151-747*

²*Korea Research Institute of Standards and science, Daejeon 305-600*

(Received March 20, 2006)

In semiconductor process, it is so hard to predict an exact failure point of the vacuum pump due to its harsh operation conditions and nonlinear properties, which may causes many problems, such as production of inferior goods or waste of unnecessary materials. Therefore it is very urgent and serious problem to develop diagnostic models which can monitor the operation conditions appropriately and recognize the failure point exactly, indicating when to replace the vacuum pump. In this study, many influencing factors are totally considered and eventually the monitoring model using multivariate statistical methods is suggested. The pivotal algorithms are Multiway Principal Component Analysis(MPCA), Dynamic Time Warping Algorithm(DTW Algorithm), etc.

Keywords : Vacuum pump, Batch, Multiway principal component analysis, Dynamic time warping algorithm, Hotelling's T^2

* [E-mail] true1@snu.ac.kr