

## 관리도 성능평가모형을 통한 관리한계선 개선주기 탐지기법

김종우 · 박정술 · 김준석<sup>†</sup> · 김성식 · 백준걸

고려대학교 산업경영공학과

## Update Cycle Detection Method of Control Limits using Control Chart Performance Evaluation Model

Jongwoo Kim · Cheong-Sool Park · Jun Seok Kim · Sung-Shick Kim · Jun-Geol Baek

School of Industrial Management Engineering, Korea University

Statistical process control (SPC) is an important technique for monitoring and managing the manufacturing process. In spite of its easiness and effectiveness, some problematic sides of application exist such that the SPC techniques are hardly reflect the changes of the process conditions. Especially, update of control limits at the right time plays an important role in acquiring a reasonable performance of control charts. Therefore, we propose the control chart performance evaluation index (CPEI) based on count data model to monitor and manage the performance of control charts. The CPEI could indicate the degree of control chart performance and be helpful to detect the proper update cycle of control limits in real time. Experiments using real manufacturing data show that the proper update intervals are made by proposed method.

**Keywords:** Control Chart, Count Data Model, Semiconductor Manufacturing Process, Statistical Process Control, Control Limit

### 1. 서 론

모바일 환경의 확산과 스마트폰, 태블릿을 비롯한 다양한 전자제품의 수요증가에 따라 반도체 시장은 지속해서 확대되고 있다. 또한, 반도체 업계의 치열한 경쟁과 고객의 품질 요구 수준이 높아지면서 품질에 대한 중요성은 더욱 강조되고 있다. 이에 따라 반도체 기업은 생산성 및 품질 향상을 위해 정밀한 공정제어 및 관리기술 확보에 힘쓰고 있다(Kymal and Patiyasevi, 2006). 특히 공정에서 발생하는 데이터를 분석하여 설비 및 공정 과정의 문제점을 관리하고 모니터링(Monitoring)하는 공정제어기술에 관한 연구들이 진행되고 있다.

공정제어기술에는 공정에서 발생하는 데이터를 통계적으로 관리하는 통계적 공정 관리(Statistical Process Control, SPC) 기법이 대표적이다. SPC 기법에는 생산된 제품 품질특성치의 평균

이 정규분포를 근사적으로 따른다는 가정 하에 구축되는 Shewhart 관리도, 공정 평균의 작은 변동에 민감한 CUSUM 관리도(Cumulative SUM)와 EWMA 관리도(Exponentially Weighted Moving Average) 등이 있다(Montgomery, 1997). 이러한 다양한 SPC 기법들은 제조과정에서 발생하는 품질특성치를 모니터링하기 위해 활용되며 제품의 품질 향상에 도움이 된다(Chiang et al., 2001).

제조 공정은 시대별 환경에 따라 단변량 공정에서 시작하여 빅데이터(Big Data) 공정으로 변화하였다(Kim, 2013). 반도체 제조공정은 대표적인 빅데이터 공정으로 다양하고 복잡한 공정과정을 지닌다. 제품 하나를 만들기 위해서는 수십, 수백 가지의 공정을 거쳐야 하고 단위 공정 설비에는 수백 수천 개의 센서가 있다. 각 센서는 온도, 압력 등 수많은 공정변수 데이터를 실시간으로 수집한다(Jang, 2012). 반도체 제조공정에서는

이 연구는 미래창조과학부 및 정보통신산업진흥원의 ‘정보통신연구기반구축사업’의 연구결과로 수행되었으며(NIPA-2013-(I2218-13-1004)), 2011년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 것임(2011-0025414).

† 연락처 : 김준석 연구교수, 136-701 서울시 성북구 안암동 5가 1번지 고려대학교 산업경영공학과, Tel : 02-925-5035, Fax : 02-3290-4802,

E-mail : bliths@korea.ac.kr

2013년 11월 13일 접수; 2013년 12월 24일 수정본 접수; 2014년 1월 23일 게재 확정.

관리도를 통해 수많은 공정변수를 동시에 모니터링하고 있다 (Jo, 2005). 관리도 운영 과정에서는 관리도 성능 저하를 방지하기 위해 주기적인 관리한계선 개선기법을 수행한다. 주기적인 관리한계선 개선기법은 일정 주기를 가지고 관리한계선을 동시에 개선하는 기법이다. 주기적인 관리한계선 개선은 많은 양의 데이터를 동시에 처리하기 때문에 서버 과부하가 발생한다(Chen et al., 2004). 또한, 주기적 관리한계선 개선기법의 개선주기는 주로 공정기술자들의 경험 및 단순한 직관으로 결정된다. 이는 주관적 결정일 가능성이 많으며 통계적 근거가 부족하여 추가적인 생산, 품질 비용이 발생한다(Im, 2010). 예를 들어, 공정 조건이 실제로 변하지 않았음에도 불구하고 관리한계선을 개선하였을 경우 불필요한 서버 과부하가 발생한다. 이러한 문제는 관리한계선 개선 지연 문제를 일으키며 더 나아가 공정변화로 인한 관리도의 검출 능력을 악화시킬 수 있다. 또한, 공정 조건이 실제로 변하였음에도 불구하고 관리한계선 개선을 수행하지 않았을 경우 관리도의 이상 검출 능력이 떨어지게 되어 오경보(False Alarm) 문제가 발생한다(Im, 2010). 따라서 공정 모니터링의 정확도 향상과 불필요한 과부하를 줄이기 위해서는 관리도의 성능이 일정 기준 이하가 되었을 때 관리한계선 개선을 수행해야 한다.

전통적인 관리한계선 개선기법은 공정 이상이 발생할 경우에 이상을 조치한 후 관리한계선을 개선하는 것이다(Hill, 1956; Montgomery, 1997). 그러나 반도체와 같이 여러 가지의 제품이 혼류되어 생산되는 제조공정에서는 시간의 흐름에 따라 설비 및 작업환경이 변하기 때문에 이상 발생 전까지 관리한계선을 유지하는 것은 관리도의 검출 능력 저하 문제를 발생시킬 수 있다. 이에 Im(2010)은 규칙기반으로 관리한계선을 설정하고 개선하는 기법을 제안하였다. 규칙기반 관리한계선 개선기법은 공정의 관리한계선이나 공정기술자의 기술(Know-how)을 규칙화하여 공정을 관리하는 기법이다. 규칙기반 관리한계선 개선기법은 규칙을 설정하기 위해서는 공정기술자의 경험이 필요하고 공정마다 필요한 규칙이 다르기 때문에 규칙의 추가, 삭제, 수정 등의 작업이 수동적으로 진행되는 문제가 있다(Im, 2010). Marc(2011)는 관리도 성능을 이용한 관리한계선 개선기법을 제안하였다. 관리도 성능을 이용한 관리한계선 개선기법은 관리한계선 개선 이전의 관리상한선과 관리하한선 차이와 개선한 이후의 관리상한선과 관리하한선 차이를 비율로 제시한다. 공정기술자는 제공된 정보를 이용하여 관리한계선 개선여부를 판단한다. 하지만 관리도 성능을 이용한 관리한계선 개선기법은 관리한계선 개선 시점에 관리도의 성능을 확인할 수 있을 뿐 실시간 관리한계선 성능을 확인할 수 없으며 공정기술자에게 관리한계선의 변화 정보만 제시한다. Marc(2011)가 제안한 관리도 성능을 이용한 관리한계선 개선기법은 관리한계선 개선작업이 선행되어야 관리도 성능을 판단할 수 있기 때문에 관리한계선 개선 시점을 판단하는데 많은 시간과 비용이 필요하다. 따라서 설정된 관리도 성능을 실시간으로 모니터링하고, 관리한계선 개선 시점을 탐지할 수 있는 기법에 대

한 연구가 필요하다.

본 논문에서는 관리도 성능을 평가하는 척도인 CPEI(Control chart Performance Evaluation Index)를 제시함으로써 관리도 성능을 예측하고 관리한계선 개선 시점을 모니터링할 수 있는 CPEM(Control chart Performance Evaluation Model)을 제안한다. CPEM은 <Figure 1>과 같은 형태로 CPEI를 실시간으로 모니터링하고, 성능기준치인 CP(Critical Point)에 도달할 경우에 관리한계선 개선을 수행한다.

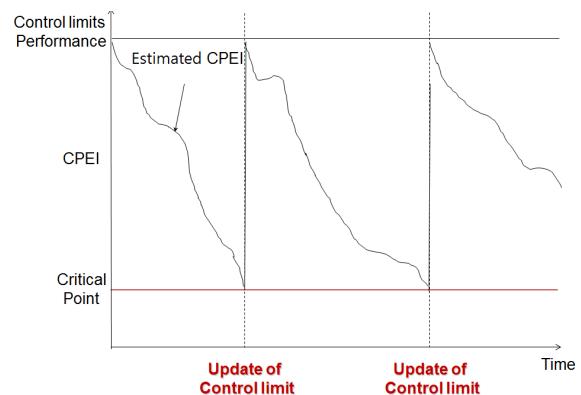
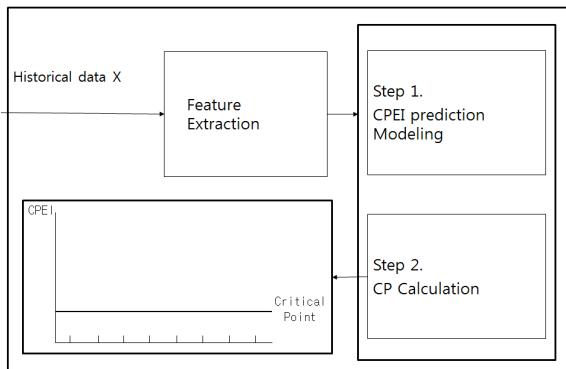


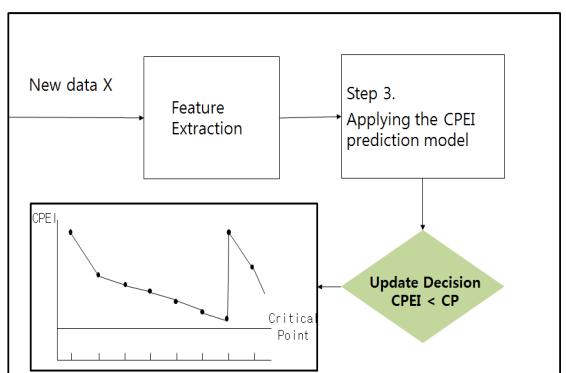
Figure 1. A basic example of the control limit update by CPEM

#### PHASE I : CPEM Construction



(a) The procedure of CPEM construction

#### PHASE II : CPEI Monitoring



(b) The procedure of CPEI monitoring

Figure 2. The procedure of CPEM

본 논문에서 제안하는 CPEM은 <Figure 2>와 같은 절차로 진행되며, CPEM 구축 단계(PHASE I)와 CPEI 모니터링 단계(PHASE II)로 나뉜다. CPEM 구축 단계에서는 CPEI 예측모형과 CP를 결정한다. CPEI 모니터링 단계에서는 CPEM 구축단계에서 설정된 모형을 바탕으로 CPEI를 추정하고 CP과의 비교를 통해 관리한계선 개선여부를 판단한다.

CPEM 구축 단계에서는 <Figure 2>(a)와 같이 수집된 과거 데이터로부터 특질변수를 추출한 후 CPEI 예측모형을 수립하고, 관리한계선 개선 기준인 CP를 결정한다.

CPEI 모니터링 단계는 <Figure 2>(b)에 나타나듯이 새롭게 생성되는 데이터로부터 특질변수를 추출한 후 수립된 CPEM을 통해 계산된 CPEI를 모니터링하고 CP와의 비교를 통해 관리한계선 개선여부를 판단한다.

본 논문에서 제안하는 기법은 CPEM을 이용해 계산된 CPEI를 통해 실시간으로 관리도 성능을 예측할 수 있다. 즉, CPEI와 성능기준치인 CP값과의 비교를 통해 적절한 시점에 관리한계선 개선을 수행할 수 있다. 따라서 주기적 관리한계선 개선기법에서 발생할 수 있는 서버 과부하를 감소시킬 수 있다. 또한 공정조건의 변화를 실시간으로 모니터링 할 수 있기 때문에 관리도 성능의 저하를 방지 할 수 있다.

본 논문의 구성은 다음과 같다. 제 2장에서는 CPEM 절차, CPEI 및 CP의 정의 그리고 CPEI 예측모형을 구축하기 위해 사용된 가산자료 분석기법에 대해 기술 한다. 또한, 제 3장에서는 반도체 제조 기업에서 사용하는 주기적 관리한계선 개선기법과 CPEM을 통한 관리한계선 개선기법의 성능 비교를 수행한다. 마지막으로 제 4장에서는 결론과 추후 연구의 내용을 제시한다.

## 2. 관리도 성능평가모델

### 2.1 CPEM 절차

본 연구에서 제안하는 기법은 CPEM 구축 단계와 CPEI 모니터링 단계로 구성되어 있다. CPEM 구축 단계는 <Figure 3>의 CPEI 예측모형 구축과 <Figure 4>의 CP 계산의 절차로 이루어지며, CPEI 모니터링 단계는 <Figure 5>와 같이 구축된 CPEM을 이용해 계산된 CPEI를 모니터링하고 CP값과의 비교를 통해 관리한계선 개선여부를 결정하는 단계이다.

<Figure 3>은 CPEI 예측모형 구축과정으로써 2가지 단계로 구성된다. 첫째, 학습데이터로부터 필요한 특질변수를 추출한다. 품질특성치의 변화를 감지하기 위해 관리도에서 발생하는 데이터의 패턴을 규칙화한 런 규칙(Run Rule)의 누적발생 횟수와 데이터의 분포 변화를 감지하기 위한 분포 통계량이 특질변수로 추출된다. 둘째, 과산포성 검정(Over-dispersion Test)을 통해 회귀모형을 선정한다. 회귀모형의 선정기준은 과산포성 검정 결과에 의해 결정된다.

```

Input = Traindata_model

Step. 1 FOR i = 1 to n DO
    x[i] = Traindata_model[i]
    feature[i] = feature_extraction(x[i])
ENDFOR

STEP. 2 result = over-dispersion_test(feature)
IF result = significant THEN
    Select Poisson regression
ELSE
    Select Negative Binomial regression
ENDIF

```

Figure 3. Pseudo-code for CPEI prediction model

<Figure 4>는 CP 설정과정이다. CP 설정과정은 다음과 같이 구성된다. 첫째, 학습데이터의 관리한계선 개선 구간정보를 이용하여 특질변수를 추출한다. 둘째, 추출된 특질변수와 구축된 CPEM을 이용해 CPEI를 예측하고, 각 구간별로 예측된 CPEI의 평균을 계산하여 CP를 결정한다.

```

Input = Traindata_CP

STEP. 1 FOR i = 1 to n DO
    x[i] = Traindata_CP[i]
    feature[i] = feature_extraction(x[i])
ENDFOR

STEP. 2 FOR i = 1 to n DO
    CPEI[i] = CPEI_prediction_model(feature[i])
ENDFOR

STEP. 3 CP = 1-mean(CPEI)

```

Figure 4. Pseudo-code for CP calculation

```

Input = Testdata

STEP 1. FOR i = 1 to n DO
    x[i] = Testdata[i]
    feature[i] = feature_extraction(x[i])
ENDFOR

STEP 2. FOR i = 1 to n DO
    predict_value[i] =
        CPEI_prediction_model(feature[i])
ENDFOR

STEP 3. FOR i = 1 to n DO
    IF 1-predict_value[i] < CP THEN
        Updating the control limit
    ENDIF
ENDFOR

```

Figure 5. Pseudo-code for control limit update

<Figure 5>는 CPEI 모니터링 과정으로써 3가지 단계로 구성된다. 첫째, 공정에서 새롭게 발생하는 데이터로부터 특질 변수를 추출한다. 둘째, 특질변수를 <Figure 3>에서 선정된 CPEI 예측모형에 적용하여 CPEI를 예측한 후 모니터링 한다. 셋째, CPEI 예측모형을 통해 예측된 CPEI는 <Figure 4>에서 결정된 CP를 기준으로 관리한계선 간선을 판단한다.

## 2.2 CPEI 정의

관리한계선은 공정이 관리 상태에 있음에도 불구하고 관리 상태에서 벗어났다고 결론을 내리는 오류인  $\alpha$ 에 의해 정의된다(Montgomery, 2007). 또한, 반도체 공정은 수백 개의 다양한 스텝으로 구성되어 있고 다양한 제품을 생산하기 때문에  $\alpha$ 로 인하여 불필요하게 공정을 중지시키는 일이 빈번하게 발생한다. 이는 곧 생산성과 직결되므로 공정에 이상이 없음에도 불구하고 공정을 중지시키는  $\alpha$ 를 줄일 필요가 있다(Ji, 2001). 또한,  $\alpha$ 는 관리도를 통해 실시간 통제가 가능하므로  $\alpha$ 를 활용하여 CPEI를 정의한다.  $\alpha$ 의 추정은 식 (1)과 같다.

$$\hat{\alpha} = \frac{\text{관측된 이상치의 수}}{\text{전체 품질특성치의 수}} \quad (1)$$

$\alpha$ 를 평가하기 위해 반응변수는 이상치의 누적 발생 횟수를 활용하였으며  $\alpha$ 을 추정하기 위해 관측된 이상치의 수를 전체 관측치의 수의 비율로 정의한다. 이상치는 관리한계선을 벗어난 품질특성치를 의미한다.

## 2.3 CP 정의

관리한계선 간선의 판단 기준인 CP는 기본적으로 기업의 생산전략에 의해 결정되며 기업 내부에서는 전략에 관련된 성능기준치를 CP로 사용할 수 있다. 하지만 반도체, 디스플레이 등의 많은 제조 기업의 품질관리 수준은 정보보안 문제로 공개하기 어려우며, 반도체 공정의 경우 복잡한 공정특성상 성능기준치가 불분명한 경우가 많다. 따라서 본 논문에서는 기존 관리한계선 간선기법과의 비교를 위해 기존 관리도의 관리한계선 간선구간별 CPEI를 예측한 값들의 평균을 CP로 정의한다. 이렇게 정의된 CP를 결정하기 위해서는 예측모형이 필요하므로 본 연구에서는 반응변수인 이상치의 누적 발생 횟수를 고려한 예측모형을 구축하기 위해 이산자료 분석에 효과적인 이산자료 분석모형을 활용한다.

## 2.4 이산자료모형

회귀분석은 일반적으로 반응변수가 연속형 변수임을 가정하는 선형회귀모형을 활용하여 설명할 수 있으나 이상치의 누적 발생 횟수와 같은 이산 자료의 경우 선형회귀모형은 적합하지

않다. 따라서 이산자료를 분석하기 위한 모형이 필요하다(Jang, 2003; Lee et al., 2002; Lord et al., 2005). 또한 이산자료 분석을 효과적으로 수행하기 위해 변수선택과정이 필요하다.

본 논문에서는 관리도로부터 특질변수를 추출하고 회귀계수의 유의성 검증을 기반으로 특질변수의 사용 여부를 판단한 후 이산자료모형을 선정한다. 이산자료를 분석하기 위한 대표적인 모형은 푸아송 회귀모형(Poisson Regression)이다. 또한, 이산자료 분석에 관한 연구는 대부분 푸아송 회귀모형을 위주로 진행되었다(Gourieroux and Visser, 1997). 하지만 현실에서는 푸아송 회귀모형의 가정인 등산포성(Equi-dispersion)을 만족하지 않는 상황이 많다(Green, 2000). 특히, 푸아송 확률분포의 분산이 평균보다 큰 과산포가 많이 발생한다. 이때는 이산자료를 모형화하기 위해 푸아송 회귀모형의 일반화된 모형인 음이항 회귀모형(Negative Binomial Regression)을 사용한다(Long, 1997; Hilbe, 2007).

### (1) 특질변수 추출

효과적인 이산자료모형을 구축하기 위해서는 변수선택과정이 필요하다. 따라서 본 논문에서는 관리도로부터 발생하는 데이터의 패턴을 규칙화한 런 규칙(Run Rule)의 누적발생 횟수와 데이터의 분포 변화를 감지하기 위한 분포 통계량을 추출한다. 특질변수는 두 가지 요소로 나누어 설명할 수 있다. 첫 번째 요소는 런 규칙이다. 런 규칙은 특정 구간에서의 품질특성치들의 연속적인 패턴을 의미한다. 이는 품질특성치의 추세, 주기 등의 패턴 감지를 목적으로 활용하며 <Table 1>과 같다.

**Table 1.** Descriptions of the variables used in the CPEI prediction model

Variables	Descriptions
Accumulated Run rule 2	Accumulated number of two of three consecutive points outside two-sigma warning limits but still inside the control limits
Accumulated Run rule 3	Accumulated number of four of five consecutive points outside one-sigma warning limits but still inside the control limits
Accumulated Run rule 4	Accumulated number of a run eight consecutive statistic on one side of the centerline
Distribution Statistics	Mean, Variance, Skewness, Kurtosis

본 논문에서 활용한 런 규칙 2는 관리도에서 연속적으로 관측된 3개의 품질 특성치 중 2시그마와 3시그마 사이에 2개 관측된 품질특성치의 누적 발생 횟수, 런 규칙 3은 연속적으로 관측된 5개의 품질 특성치 중 1시그마와 2시그마 사이에 4개 관측된 품질 특성치의 누적 발생 횟수, 런 규칙 4는 8개 연속하여 관측된 품질 특성치가 중심선의 한쪽에서 누적 발생한 횟수를 의미한다.

두 번째 요소는 분포 통계량이다. 분포 통계량은 시간의 흐름에 따른 분포변화를 반영하기 위해 활용한다. 일반적으로 정규분포가 평균과 분산으로 분포가 결정되는 반면에 다양한 형태의 분포 표현이 가능한 피어슨 분포는 평균, 분산, 왜도, 첨도, 총 네 개의 모수에 의해 결정된다. 따라서 정규분포로 설명하기 어려운 반도체 공정의 분포를 정확하게 반영하기 위해 피어슨 분포를 사용한다(Yoo et al., 2012).

### (2) 푸아송 회귀모형

푸아송 분포는 발생확률이 작은 사건을 대량으로 관찰할 때, 그 발생횟수가 만드는 분포이다. 따라서 푸아송 분포는 이상치의 누적 발생 횟수와 같은 매우 드물게 일어나는 사건의 모형에 이용하기 적합하다(Knorr and Ng, 1997). 푸아송 회귀모형을 세우기 위해서는 다음과 같은 가정이 필요하다. 첫째, 단위 시간당 일어나는 사건의 발생이 서로 독립적이다. 둘째, 단위시간 내 사건의 발생확률이 동일하며 시간에 따라 변하지 않는다. 셋째, 푸아송 확률분포의 평균과 분산이 같다는 등산포성을 만족한다. 푸아송 분포의 확률밀도식은 식 (2)와 같고 회귀모형은 식 (3)과 같다.

$$P(Y_i = y_i) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{y_i!}, \quad i = 0, 1, 2, \dots \quad (2)$$

$$\ln \lambda_i = \theta' X_i \quad (3)$$

여기서,  $Y_i$ 는 이상치의 누적 발생 횟수를 나타내는 반응변수이며  $y_i$ 는  $i$ 번째 발생하는 사건의 횟수를 의미한다.  $\lambda_i$ 는  $i$ 번째 발생하는 평균 및 분산을 의미하는 모수이다.  $X_i$ 는 설명변수 벡터,  $\theta_i$ 는 설명변수에 대한 추정계수 벡터이다. 또한, 가정에 의해 분포의 평균과 분산은 식 (4)와 같이 도출된다.

$$E[Y_i|\theta, X_i] = Var[Y_i|\theta, X_i] = \exp(\theta X_i'), \quad \exp(\theta X_i') = \lambda_i \quad (4)$$

이때 최대우도추정(Maximum Likelihood Estimation, MLE)을 활용한 로그우도함수(Log Likelihood Function)는 식 (5)와 같다.

$$\ln L_{Poi} = \sum [y_i - \exp(\theta X_i') - \ln(y_i!)] \quad (5)$$

여기서, 모수 벡터  $\theta$ 를 추정하였으며 CPEI 예측모형을 통해 예측된 이상치의 수를 활용하여 CPEI를 추정한다. CPEI 추정식은 식 (6)과 같다.

$$CPEI_i = \frac{Y_i}{n_i}, \quad n_i = \sum_{j=0}^i j, \quad i = 0, 1, 2, \dots \quad (6)$$

식 (6)에서  $n_i$ 은  $i$ 번째까지의 전체 품질특성치의 수를 의미하고 추정된 CPEI를 통해 관리도 성능을 모니터링한다.

### (3) 음이항 회귀모형

음이항 회귀모형은 가산자료를 모형화하기 위해 푸아송 회귀모형을 일반화한 모형이다. 음이항 회귀모형의 확률밀도함수는 식 (7)과 같다.

$$P(Y_i = y_i) = \frac{\Gamma(y_i + \gamma^{-1})}{\Gamma(\gamma^{-1}) \Gamma(y_i + 1)} (\gamma \lambda_i)^{y_i} (1 + \gamma \lambda_i)^{-(y_i + \gamma^{-1})} \quad i = 0, 1, 2, \dots \quad (7)$$

여기서  $\gamma$ 를 과산포 모수라 한다. 만약  $\gamma = 0$ 이란 귀무가설이 기각되지 않는다면 누적 이상치 발생 횟수는 등산포의 특성을 보이며 푸아송모형을 사용해도 무리가 없다. 하지만 귀무가설이 기각되면 푸아송 회귀모형 대신에 음이항 회귀모형을 사용해야 한다(Cameron and Trivedi, 1986). 요컨대,  $\gamma$ 값이 0이면 푸아송 회귀모형, 0보다 크면 음이항 회귀모형이 적합하다. 음이항 회귀모형에서 평균은 식 (8)과 같으며 분산은 식 (9)와 같다.

$$E(Y_i) = \lambda_i \quad (8)$$

$$Var(Y_i) = \lambda_i + \gamma \lambda_i^2 = \lambda_i (1 + \gamma \lambda_i) \quad (9)$$

따라서 분산의 구조상 어느 경우이든 간에 평균은 분산보다 작게 된다. 여기서  $\gamma$ 는 비음제약을 만족해야 하며,  $\gamma = 0$ 이면 푸아송 회귀모형과 같아진다. 이때 로그우도함수는 식 (10)과 같다.

$$\begin{aligned} \ln L_{NB} = & \sum [y_i \ln(\exp(\theta X_i') \gamma) \\ & - (y_i + \gamma) \ln(1 + (\exp(\theta X_i') \gamma)) \\ & + \ln \Gamma(y_i + \gamma^{-1}) - \ln(y_i!) - \ln \Gamma(\gamma^{-1})] \end{aligned} \quad (10)$$

식 (5)와 달리 음이항 회귀모형에서는  $\theta$ 모수벡터 뿐만 아니라  $\gamma$ 도 추정해야 한다.

### (4) 과산포성 검정

가산자료 분석을 위해 반응변수에 과산포의 존재 여부를 판단해야 한다. 이는 간단한 과산포성 검정을 통해 해결할 수 있다. 과산포의 존재 여부는 푸아송 및 음이항 분포를 기반으로 한 우도비 검정(Likelihood Ratio Test)을 이용한다. 검정가설은 식 (11)과 같으며 검정 통계량(Log Likelihood Ratio, Log LR)은 식 (12)와 같다.

$$H_0 : Var(Y_i) = \lambda_i \quad (11)$$

$$H_1 : Var(Y_i) \neq \lambda_i$$

$$Log LR = -2(\ln L_{Poi} - \ln L_{NB}) \quad (12)$$

여기서  $\ln L_{Poi}$ 와  $\ln L_{NB}$ 는 각각 푸아송 분포와 음이항 분포의 우도를 나타내며  $Log LR$ 은 자유도가 1인 카이제곱 분포를 따

른다. 또한, 식 (11)의 대립가설에서 양측을  $\lambda$ 로 나누면  $Var(Y_i)/\lambda_i \neq \lambda_i/\lambda_i$ 가 되며 양쪽 모두 과산포율로 나타난다.  $Var(Y_i) = \lambda_i$ 이라는 귀무가설이 정해진 유의수준에서 채택되면 평균과 분산이 같다는 가정이 만족하므로 푸아송 회귀모형을 사용한다. 반면에 귀무가설이 정해진 유의수준에서 기각되면 평균과 분산이 같다는 가정을 만족하지 않으므로 음이항 회귀모형을 사용하여 CPEI 예측모형을 구축한다.

### 3. 실험

#### 3.1 실험절차

본 논문에서 활용된 데이터는 반도체 공정 데이터로 3개의 챔버(Chamber)를 가진 한 설비에서 3개월간 수집되었다. 총 28 개 공정변수가 약 15,000~30,000개의 품질특성치로 구성되어 있다. 실험에 앞서 학습데이터와 검정 데이터를 분할하였다. 데이터 분할은 CPEI 예측모형을 구축하기 위한 학습데이터, CP 설정을 위한 학습데이터 및 검정데이터로 분할하였다. CPEI 예측모형을 구축하기 위한 학습데이터는 3개월간의 데이터 중 마지막 1개월의 데이터를 사용하였으며 CP 설정을 위한 학습데이터는 관리한계선 갱신구간별로 데이터를 구성하였다. 검정 데이터는 공정으로부터 새롭게 발생하는 데이터를 사용하였다. 데이터 분할을 수행한 후 각 데이터로부터 특질변수를 추출하였다.

가산자료모형의 선정은 과산포성 검정을 통해 푸아송 회귀모형과 음이항 회귀모형의 적용 여부를 판단한 후 회귀모형을 선정하였다. 다음으로 선정된 회귀모형의 적합성을 판단하기 위해 회귀계수 검정과 적합도 검정을 하였다. 마지막으로 새로운 데이터를 CPEI 예측모형에 적용하여 CPEM에 의한 관리한계선 주기를 결정하였으며 기존 관리한계선 갱신 주기와 비교하였다. 실험은 오픈 소스 통계 패키지인 R을 이용하여 실현하였다.

#### 3.2 실험결과

반도체 공정에서 추출된 데이터가 푸아송 회귀모형의 가정인 등산포성을 따른다고 전제할 수 없기 때문에 회귀모형 구축에 앞서 과산포성 검정을 진행하였다. 실험에서 반응변수로 사용된 이상치의 누적 발생횟수에 대한 과산포성 검정결과는 <Table 2>와 같다.

<Table 2>는 총 28개의 공정변수 데이터를 활용한 실험 중에 16개의 경우(PARAMETER 1-7, 12-13, 16-17, 19, 20, 24, 26, 28)은 유의수준 0.24~0.91에서 귀무가설이 채택되어 평균과 분산이 같다는 가정을 만족한다고 볼 수 있다. 즉 푸아송 회귀모형에 적합한 실험 경우라고 판단된다. 반면 나머지 12개의 경우(PARAMETER 8-11, 14-15, 18, 21-23, 25)는 귀무가설이 유의수준 0.05 이하에서 기각되어 평균과 분산이 다르다고 판

단할 수 있으며 음이항 회귀모형이 적합함을 확인할 수 있다. 또한, 과산포성 테스트 결과에 나타나듯이  $Log LR$ 이 0에 가까울수록 푸아송 회귀모형에 근접함을 확인할 수 있다.

Table 2. Results of over-dispersion test

Process parameter	$Log LR$	z-value	p-value
PARAMETER 1	$4.10 \times 10^{-5}$	$-1.01 \times 10^7$	0.73
PARAMETER 2	$-5.05 \times 10^{-13}$	$-1.11 \times 10^{16}$	0.84
PARAMETER 3	$2.00 \times 10^{-3}$	$-2.12 \times 10^4$	0.34
PARAMETER 4	0.05	$-1.71 \times 10^3$	0.25
PARAMETER 5	$3.56 \times 10^{-5}$	$-1.47 \times 10^6$	0.75
PARAMETER 6	0.03	$-1.01 \times 10^3$	0.23
PARAMETER 7	0.02	$-1.46 \times 10^3$	0.24
PARAMETER 8	2.60	10.80	$2.20 \times 10^{-16}$
PARAMETER 9	90.80	37.96	$2.20 \times 10^{-16}$
PARAMETER 10	39.46	37.77	$2.20 \times 10^{-16}$
PARAMETER 11	42.79	39.57	$2.20 \times 10^{-16}$
PARAMETER 12	-1.03	$-2.99 \times 10^{13}$	0.82
PARAMETER 13	0.01	$-1.03 \times 10^6$	0.32
PARAMETER 14	57.26	23.45	$2.20 \times 10^{-16}$
PARAMETER 15	103.44	58.48	$2.20 \times 10^{-16}$
PARAMETER 16	$6.17 \times 10^{-5}$	$1.04 \times 10^6$	0.72
PARAMETER 17	$1.14 \times 10^{13}$	2.00	0.23
PARAMETER 18	$6.57 \times 10^{15}$	16.40	$2.20 \times 10^{-16}$
PARAMETER 19	$7.47 \times 10^6$	2.00	0.23
PARAMETER 20	$1.14 \times 10^{13}$	2.00	0.23
PARAMETER 21	182.17	76.61	$2.20 \times 10^{-16}$
PARAMETER 22	8.22	96.30	$2.20 \times 10^{-16}$
PARAMETER 23	108.94	44.48	$2.20 \times 10^{-16}$
PARAMETER 24	$1.09 \times 10^{-14}$	$-3.76 \times 10^{15}$	0.65
PARAMETER 25	3,751.51	9.05	$2.20 \times 10^{-16}$
PARAMETER 26	$2.00 \times 10^{-4}$	$-4.70 \times 10^5$	0.89

과산포성 검정결과에 따라 결정된 회귀모형의 적합성을 검정하기 위해 결정계수를 사용하여 확인하였다. 가산자료모형의 결정계수는 일반 회귀모형과 다르기 때문에 가산자료모형의 결정계수 계산방식을 사용하였다(Cameron and Windmeijer, 1996). 결정계수 검정결과는 <Table 3>과 같다.

Table 3. Summary result of coefficient of determination test

Variable	Descriptive statistics		
	Max	Min	Mean
Process parameter	0.94	0.90	0.92

모든 결정계수는 90% 이상으로 모형의 설명변수들이 반응변수를 잘 설명하고 있다. 또한, 회귀계수검정 결과 대부분의 경우에 유의수준 0.05 이하에서 유의하다. 즉, 선택된 특질변

수들은 충분히 활용 가능한 것을 확인할 수 있다.

<Table 4>는 28개의 실험 중 대표적인 푸아송 회귀모형인 PARAMETER 3의 회귀모형 결과이다. 회귀계수 검정 결과 설명변수인 런 규칙 2의 누적 발생 횟수, 런 규칙 3의 누적 발생 횟수 및 런 규칙 4의 누적 발생 횟수의 회귀계수는 양측검정으로 p-value가 0.01 이하에서 유의하다는 것을 알 수 있으며 평균, 분산, 웨도, 첨도와 같은 분포 통계량은 각각 0.05~0.1사이에서 유의하다는 것을 알 수 있다. 즉, 특질변수는 충분히 활용 가능한 것을 확인할 수 있다. 또한, 설명변수가 반응변수에 미치는 영향력을 나타내는 결정계수는 92%로서 특질변수로 사용된 반응변수들이 설명변수를 매우 잘 설명하고 있음을 의미한다.

**Table 4.** Results of CPEM by poisson regression

Variable	Estimated value	Std.error	p-value
intercept	- 11.74	$4.24 \times 10^{-2}$	$2.00 \times 10^{-16}$
Run Rule 2	$-1.23 \times 10^{-6}$	- 1.26	0.02
Run Rule 3	$-3.93 \times 10^{-6}$	- 2.49	0.01
Run Rule 4	$1.64 \times 10^{-6}$	- 2.36	0.02
Mean	$1.66 \times 10^{-1}$	469.89	$2.00 \times 10^{-16}$
Variance	$7.00 \times 10^{-1}$	556.65	$2.00 \times 10^{-16}$
Skewness	$-4.07 \times 10^{-1}$	- 924.18	$2.00 \times 10^{-16}$
Kurtosis	$-1.05 \times 10^{-1}$	- 1,131.74	$2.00 \times 10^{-16}$
Poisson Regression			
Observation : 6687, R squared : 0.92			

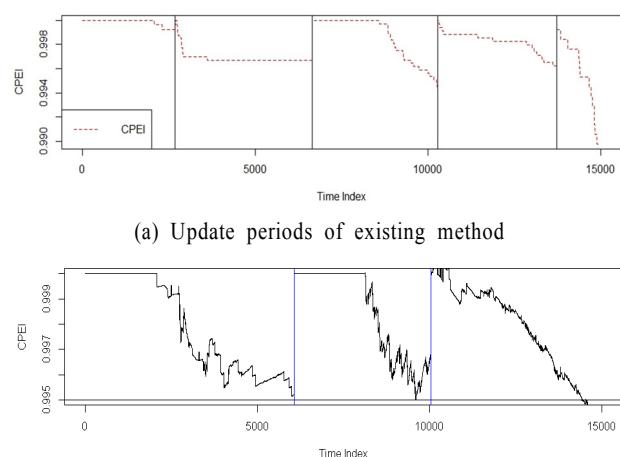
**Table 5.** Results of CPEM by negative binomial regression

Variable	Estimated value	Std.error	p-value
intercept	- 57.07	$8.53 \times 10^{-2}$	$2.00 \times 10^{-16}$
Run Rule 2	$-1.75 \times 10^{-1}$	$5.53 \times 10^{-4}$	$2.00 \times 10^{-16}$
Run Rule 3	$-1.88 \times 10^{-5}$	$2.53 \times 10^{-6}$	$1.14 \times 10^{-13}$
Run Rule 4	$3.16 \times 10^{-5}$	$2.65 \times 10^{-6}$	$2.00 \times 10^{-16}$
Mean	$4.50 \times 10^{-1}$	$6.83 \times 10^{-4}$	$2.00 \times 10^{-16}$
Variance	9.66	$1.50 \times 10^{-2}$	$2.00 \times 10^{-16}$
Skewness	- 1.67	$3.09 \times 10^{-3}$	$2.00 \times 10^{-16}$
Kurtosis	$-2.65 \times 10^{-1}$	$9.42 \times 10^{-4}$	$2.00 \times 10^{-16}$
Negative Binomial Regression			
Observation : 6687, R squared : 0.91			

<Table 5>는 28개의 실험 중 대표적인 음이항 회귀모형인 PARAMETER 10을 이용한 CPEI 예측모형 구축결과이다. 회귀계수 검정결과 설명변수인 런 규칙 2의 누적 발생 횟수, 런 규칙 3의 누적 발생 횟수 및 런 규칙 4의 누적 발생 횟수, 평균, 분산, 웨도 첨도의 회귀 계수는 양측검정으로 p-value가 0.01 이하에서 유의하다는 것을 알 수 있다. 또한, 설명변수가 반응 변수에 미치는 영향력을 나타내는 결정계수는 91%로서 특질 변수로 사용된 반응변수들이 설명변수를 매우 잘 설명하고 있

음을 확인할 수 있다.

<Figure 6>은 기존 관리도 제안기법의 관리도 성능을 나타낸 것이다. <Figure 6>(a)는 주기적 관리한계선 설정기법의 CPEI를 모니터링한 결과이다. 수직의 실선은 주기적 관리한계선 개선주기이며 수평의 점선은 CPEI를 의미한다. <Figure 6>(a)의 CPEI가 구간별로 다른 것으로 보아 주기적 관리한계선 개선기법은 관리도 성능을 고려하지 않고 있음을 확인할 수 있다. <Figure 6>(b)는 CPEM을 통한 관리한계선 개선기법이다. 수평의 점선은 CPEI를 의미하며 수평의 실선은 CP를 의미한다. <Figure 6>(b)에 나타나듯이 관리한계선 개선은 2회 되었음을 확인할 수 있다. 즉, 관리도 성능이 CP값(0.995) 이상으로 유지되는 동안 관리한계선 개선 횟수는 절반으로 줄었다.



**Figure 6.** Comparisons of CPEI monitoring

<Figure 7>은 기존 주기적 관리한계선 개선기법과 CPEM을 통한 관리한계선 개선방법을 비교한 것이다. 수평의 점선은 관리상한선을 의미하며 실선은 관리 하한선을 의미한다. 수직의 실선은 기존 주기적 관리한계선 개선기법에 따른 관리한계선 개선 시점을 나타내며 수직의 점선은 CPEM 방식의 관리한계선 개선주기이다. <Figure 6>에 나타나듯 주기적 관리한계선 개선방식은 4회, CPEM 방식은 2회 관리한계선이 개선되었다. 이는 반도체 공정의 넓은 관리한계선으로 인해 관리도의 성능이 저하되지 않았음에도 불구하고 관리한계선이 개선되고 있었음을 의미한다. 즉, 불필요한 개선이 수행되고 있었다고 볼 수 있다.

마지막으로 주기적 관리한계선 개선기법과 CPEM 개선기법의 개선주기를 비교하였다. 주기적 관리한계선 개선주기는 기업의 운영전략이므로 보안 유지 차원에서 언급하기 어렵기 때문에 기존 관리한계선 개선주기를 1로 표준화하여 관리한계선 개선주기를 상대 비교하였다. 표준화한 값 1은 기존 관리한계선의 개선주기를 의미한다. 만약 개선주기가 2라면 개선주기는 2배 늘었다는 것이므로 개선 횟수가 반으로 감소했다는 것을 의미한다.

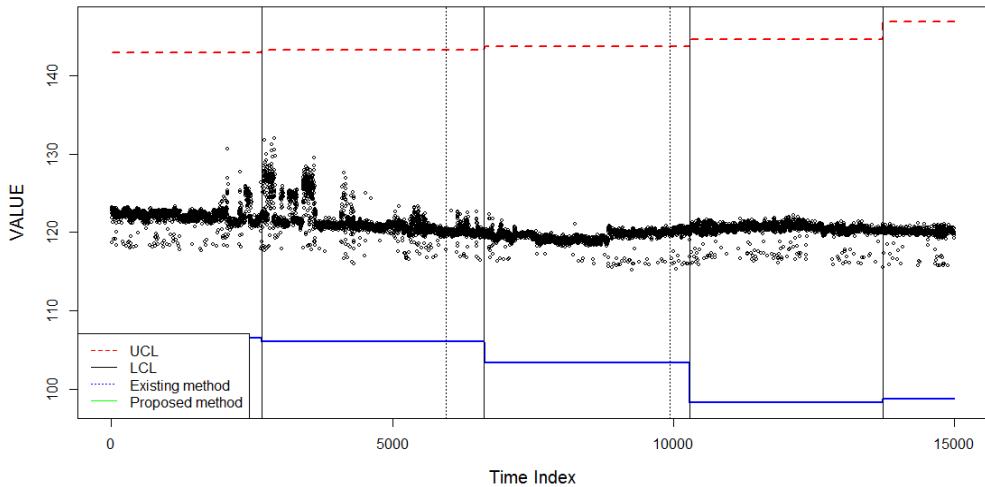


Figure 7. Comparisons of control limit update methods

즉, 관리도 성능은 유지하면서 개선 횟수의 감소를 통해 서버 과부하 발생을 줄일 수 있다. 반면 개선주기가 1/2이라면 개선 횟수가 2배 증가했다는 것을 의미한다. 즉, 관리도 성능 저하를 방지하기 위해 관리한계선 개선이 많이 필요하다는 것을 알 수 있다. 상대비교 방식은 식 (13)과 같은 관리한계선 개선 상대률도(Relative Update Rate, RUR)를 사용하였다.

$$RUR = \frac{\text{제안주기} - \text{기존주기}}{\text{제안주기}} \times 100 \quad (13)$$

RUR은 기존 관리한계선 개선 주기 대비 제안 관리한계선 주기의 개선 정도를 의미한다. RUR이 0보다 크다는 것은 기존 기법보다 관리한계선 개선주기를 길게 설정해야 함을 의미하고 0보다 작다는 것은 기존 기법보다 관리한계선 개선 주기를 짧게 설정해야 함을 의미한다.

<Table 6>은 주기적 관리한계선 개선기법과 CPEM 기법을 비교한 표이다. 모든 경우의 실험에서 기존에 비해 개선주기가 증가하였으며, 평균적으로 175%, 최대 400%까지 개선주기가 증가하였음을 확인할 수 있다. 즉, 관리한계선 개선을 기준 대비 최대 1/4정도로 감소시킬 수 있음을 알 수 있다. 따라서 본 연구에서 제안하는 관리한계선 개선기법을 통해 관리도의 성능을 유지하면서 서버 과부하 발생의 감소를 기대할 수 있다.

#### 4. 결 론

반도체 제조공정은 관리도 성능을 저하를 방지하기 위해 주기적 관리한계선 개선을 수행하고 있다. 주기적 관리한계선 개선은 공정기술자들의 경험 및 단순한 직관으로 결정된다. 이는 주관적 결정일 가능성이 많으며 통계적 근거가 부족하여 불필요한 서버 과부하나 관리도 성능 저하를 일으킨다. 따라서 본 논문에서는 CPEM을 통해 관리한계선 개선 시점을 결정하는 기법을 제안하였다. 제안기법은 각 공정변수 데이터의 특성을

Table 6. The improved RUR by CPEM

Process parameter	Updating Period		RUR(%)
	Existing Method	CPEM Method	
PARAMETER 1		5	400
PARAMETER 2		1.23	23
PARAMETER 3		2	100
PARAMETER 4		1	0
PARAMETER 5		5	400
PARAMETER 6		1.067	6.7
PARAMETER 7		5	400
PARAMETER 8		5	400
PARAMETER 9		1.57	57
PARAMETER 10		2	100
PARAMETER 11		1.21	21
PARAMETER 12		1	0
PARAMETER 13		1.5	50
PARAMETER 14		4	300
PARAMETER 15		1.09	9
PARAMETER 16		2	100
PARAMETER 17		2.8	180
PARAMETER 18		3.0	200
PARAMETER 19		5.8	480
PARAMETER 20		4.2	320
PARAMETER 21		3	200
PARAMETER 22		2.235	123.5
PARAMETER 23		4	300
PARAMETER 24		1.09	9
PARAMETER 25		4	300
PARAMETER 26		2.8	180
PARAMETER 27		3.0	200
PARAMETER 28		1.5	50
AVERAGE		2.75	175.33

반영함으로써 특성별 관리 한계선 개선 주기를 가능하게 하였다. 또한, 주기적 관리한계선 개선에 비교하여 제안기법은 공정조건의 변화에 따라 CP을 조정하여 관리한계선 개선 시점을 쉽게 결정할 수 있다. 이는 관리한계선의 불필요한 개선을 방지함으로써 서버의 과부하 방지 및 관리한계선의 예측력 저하로 인한 이상 검출 능력저하를 방지할 수 있다.

추후 연구로는 예측 변수선정을 통한 모형의 신뢰성 향상에 대한 연구와 실제 제조공정의 특성을 반영한 CP 설정에 대한 연구가 필요하다. 또한, 수백 수천의 공정변수를 효율적으로 관리하기 위한 특성별 다변량 관리기법에 대한 개선방법론 개발이 필요하다.

## 참고문헌

- Cameron, A. C. and Trivedi, P. K. (1986), Econometric models based on count data. Comparisons and applications of some estimators and tests, *Journal of applied econometrics*, 1(1), 29-53.
- Cameron, A. C. and Windmeijer, F. A. (1996), R-squared measures for count data regression models with applications to health-care utilization, *Journal of Business and Economic Statistics*, 14(2), 209-220.
- Chen, M. F. and Huang, M. C. (2004), SPC precaution system, Semiconductor Manufacturing Technology Workshop Proceedings, IEEE, 186-189.
- Chiang, L. H., Braatz, R. D., and Russell, E. L. (2001), *Fault detection and diagnosis in industrial system*, Springer, Great Britain.
- Gourieroux, C. and Visser, M. (1997), A count data model with unobserved heterogeneity, *Journal of Econometrics*, 79(2), 247-268.
- Hilbe, J. M. (2011), *Negative binomial regression*, 2, Cambridge University Press, New York, United States of America.
- Hill, D. (1956), Modified control limits, *Applied Statistics*, 5(1), 12-19.
- Im, K.-H. (2010), Rule-based Process Control System for multi-product, small-sized production, *Journal of the Korea Industrial Information System Society*, 15(1), 47-57.
- Jang, T.-Y. (2003), Modeling Traffic Accident by Taxi Drivers through Overdispersion Test, *Journal of the Korean Society of Civil Engineers*, 23(1D), 27-34.
- Jang, Y.-J. (2012), Utilizing bigdata technique in manufacturing field. *Information and Communications Magazine*, 29(11), 30-35.
- Jo, D.-H (2005), Automatic setting method of interlock range in semiconductor process FDC system, Conference of The Korean Operations Research and Management Science Society, 491-494.
- Ji, S.-S. (2001), The Design of a neural network control chart using X-R statistics in start-up process, *Journal of the Korean Industrial and System Engineering*, 24(66), 19-26.
- Kim, S.-B. (2013), [Case Material] Utilizing big data for quality improvement of manufacturing process, *Industrial Engineering Magazine*, 20(2), 42-45.
- Knorr, E. M. and Ng, R. T. (1997), A Unified Notion of Outliers : Properties and Computation, KDD-97 Proceedings, 219-222.
- Kymal, C. and Patiyasevi, P. (2006), Semiconductor quality initiatives : How to maintain quality in this fast-changing industry, *Quality Digest*, 26(4), 43-48.
- Lee, A. H., Stevenson, M. R., Wang, K., and Yau, K. K. (2002), Modeling young driver motor vehicle crashes : data with extra zeros, *Accident Analysis and Prevention*, 34(4), 515-521.
- Long, J. S. (1997), *Regression models for categorical and limited dependent variables*, Sage Publications, Inc, United States of America.
- Lord, D., Washington, S. P., and Ivan, J. N. (2005), Poisson, Poisson-gamma and zero-inflated regression models of motor vehicle crashes : balancing statistical fit and theory, *Accident Analysis and Prevention*, 37(1), 35-46.
- Marc, S. and Craig, M (2010), *DataLyzer Spectrum*, DATALYZER International, United States of America.
- Montgomery, D. C. (2007), *Introduction to statistical quality control*, 5, Wiley, United States of America.
- Park, S.-J., Park, C.-J., Kim, S.-S., Baek, J.-G. (2011), Model Parameter Based Fault Detection for Time-series Data, *Journal of The Korea Society for Simulation*, 20(4), 67-69
- Yoo, Y.-J., Park, C.-S., Kim, J.-S., Lee, Y.-H., Kim, S.-S., Baek, J.-G. (2012), Clustering Mixed Data Using Non-normal Regression Tree for Process Monitoring, ICIME 2012 International Conference on Industrial and Mechanical Engineering, Phuket.