

장치산업에서 공정개선을 위한 공정변동 탐지에 관한 연구

최형주

(주)이지나우

A study on detecting process variation for process improvement in the process industry

Hyung Ju Choi

Easynow, Inc.

Abstract

Because process variations have direct influence on yield rate in process industry, it is very important to understand process variations that occur accidentally. In process industry, quality variation due to the activities of process improvement and maintenance and chance effect such as change of work environment and difference in staffs' craftsmanship are mixed with each other, therefore it is difficult to actually detect minute process variations. In this study, objective and rational methods of detection that can detect minute process variations in process industry were designed referring to various methodologies of process management, and they were verified through similar examples.

Keywords : Change Point Model, Homogeneity Test, SNHT Test

1. 서론

공정의 변동을 최소화하여 품질을 향상시키기 위한 공정관리 방법은 크게 통계적 공정관리 (Statistical Process Control: SPC), 공학적 공정관리 (Engineering Process Control: EPC) 및 통합공정관리(Integrated Process Control: IPC)로 구분하고 있다.

SPC는 부품산업 등에서 공정변동의 근원이 되는 이상원인(Special Cause)을 탐지하고 이를 제거함으로써 공정산포를 줄이는 것이 주목적이다. EPC는 정밀한 가공을 요하는 화학산업이나 장치산업 등에서 공정을 조절하기 위해 이용된다. 대부분의 장치산업 공정에서 관측값 간에는 상관관계가 높고, 어떠한 조정도 없는 경우 공정수준이 목표값으로 부터 멀어지는 경향이 있다. 이런 공정에서는 경향의 원인을 내재하는 잡음에 의한 것으로 보고 일반적으로 피드백 콘트롤러(Feedback Controller)를 사용하여 공정수준이 목표값에 가깝게 유지하도록 하는 수정 (Adjustment)을 하게 되는데 이와 같이 공정수정을 통한 공정관리 절차를 EPC라고 한다. 이런 연구는 Kwon et al.(2005) 공정변수의 산포발생에 따른 노이즈(잡음요소)가 최종산출물의 결과와 공정변수에 영향을 미치게 되어 공정변수의 변동을 고려한 최적 조합을 찾는 방법 연구이다. Pan et al.(2003)과 Park et al.(2008))은 EPC와 SPC를 병행하여 공정 진행 중에 주기적으로 공정변수를 수정하면서 이상원인의 발생을 동시에 탐지함으로써 관리효과를 증대시키는 방법에 관한 연구가 있다. 그러나 장치산업의 경우 공정개선 및 유지보수 활동으로 인한 공정변동이 빈번하게 발생한다. 이러한 공정변동이 품질에 어떠한 영향을 미치는지에 대한 조사가 필요하나 장비의 점진적인 노화, 노이즈 등으로 인해 공정변동을 탐지하기가 어려운 실정이다. 기존 연구에서는 공정변수 변동에 따른 공정변수 조합 최적화 방법에 대한 연구가 주류를 이루고 있다. 따라서 본 연구에서는 다양한 공정관리 방법론을 참고하여 장치산업에서의 공정변동을 탐지하는 방법을 설계하고 유사 사례를 통해 검증하고자 한다.

2. 기존연구

2.1 통계적 공정관리(SPC)

SPC는 통계적 특성을 이용하여 현 공정에서 제조한 제품들이 그 분포 패턴을 벗어날 경우, 즉 이상 원인으로 인한 변동이 생길 경우 바로 확인하고 조치를 취할 수 있도록 하기 위한 수단이다. 우연 원인으로 인한 변동만 있을 경우, 그 공정은 통계적 관리 상태에 있다고 표현하며, 공정 개선은 우연 원인의 변동을 줄이기 위한 활동이라고 할 수 있다. SPC는 관리도를 이용하여 공정변동을 관리하며 슈하트관리도, EWMA관리도, 누적합(CUSUM)관리도 등이 대표적으로 사용되고 있다.

2.2 공학적 공정관리(EPC)

EPC는 조정 가능한 입력(input)변수를 최적으로 설정하여 공정의 출력값을 목표치에 가능한 한 가깝게 유지되도록 하는 방법이다. 공정조정의 방법에는 feed-back control, feed-forward control 등이 있다. feed-back system에는 입력변수의 값이 공정의 한 단위 시간에만 영향을 미치는 단반응 시스템(responsive system)과 출력변수의 여러 시간에 걸쳐 영향을 미치는 동적 시스템(dynamic system)으로 구분할 수 있다. 전통적인 EPC 이론으로 자기회귀모형에서 변수 u_t 는 이전 관측값 u_{t-1} 에 직접적으로 영향을 받아 다음과 같은 일계 이동평균 모형식으로 표현할 수 있다.

$$u_t = u_{t-1} + \xi - \theta\xi_{t-1}$$

일계 이동평균모형에서 u_t 와 u_{t-1} 간의 상관관계는 $\rho = -\theta/(1+\theta^2)$ 이고 다른 모든 lag에서 0 값을 가진다. 그러나 u_t 에서의 상관구조는 하나의 시간간격을 뒤로 늘린 것이다. 자기회귀의 통합과 이동평균기간을 포함하는 일계 혼합모형식은 다음과 같다.

$$u_t = \xi + \theta u_{t-1} + \epsilon_t - \theta\epsilon_{t-1}$$

이 모형은 화학산업에서 자주 사용되어 진다. 화학과 공정산업에서 일계자기회귀 패턴을 갖는 고정의 품질특성들은 일반적인 형태들이다. 랜덤하거나 상관관계가 없는 것으로 간주될 수 있는 오차가 있는 품질특성들은 경험에 의해 측정된다.

2.3 비동질성 검정방법

2.3.1 이중위상 회귀모형

이중위상 회귀모형은 시계열자료에서 변곡점을 탐지하는 기법으로 시계열 내 변곡점 T를 기준으로 시계열 Z_t 를 다음과 같이 두개의 단순회귀모형으로 표현한 이중위상 회귀모형을 이용하여 시계열 내 불연속의 위치를 확인하고 변화의 통계적 유의성을 F검정을 통해 확인하는 방법이다(Solow(1987)).

$$Z_t = \begin{cases} a_1 + b_1t + e_t & t = 1, \dots, T-1 \\ a_2 + b_2t + e_t & t = T, \dots, n \end{cases}$$

2.3.2 Vincent의 비동질성 검정

Vincent(2002)는 비동질성의 크기가 우연 변동과 같거나 작을 때 비동질성을 탐지하는 것은 매우 어렵다고 보았다. 그는 동질성 검증대상의 변동성을 포함하고 있으면서 인위적인 요인으로 발생한 바이어스가 시계열에 반영되어 있지 않은 동질한 시계열인 기준시계열(reference series)를 이용하여 동질성 검증 대상과 비교하는 상대적인 비동질성 검정방법을 제안하였고 상대적인 비동질성 검정방법은 기준시계열과 검증하려는 시계열의 비교를 통해 이루어진다.

- ① 동질성 검증대상 선정 및 데이터 수집: 동질성 검증대상 자료와 이와 유사한 특징을 가진 자료를 수집한다.
- ② 기준 자료 선정: 검증대상 자료에 대한 기술통계량 및 상관분석을 수행하고 상관성이 높은 5개의 기준 자료를 선정한다.
- ③ FD 시계열 생성: 선정된 기준 자료 시계열에 대해 다음 관측값(T_{t+1})에서 현재 관측값(T_t)을 뺀 시계열(FD)을 생성한다.

$$FD_i = T_{t+1} - T_t$$

- ④ 단일 시계열 생성: 5개 기준자료의 FD 시계열을 대상으로 시간대별 최소값과 최대값을 제외한 나머지 3개 값의 평균을 산정한다. 이때 시계열의 마지막 연도의 값은 산정할 수 없기 때문에 검증대상 자료의 평균값을 대입하고 다음의 식 형식으로 시계열을 변환한다.

$$T_t = T_{t+1} - FD_i$$

- ⑤ 기준시계열 산출: 시계열 평균값에 대한 아노말리로 전환하여 기준시계열을 산출한다.

Easterling 등(1994)의 방법은 검증대상 자료의 시계열과 상관관계가 높은 5개의 기준자료를 선정하여 기준시계열을 생산한다는 점에서 객관적이라 할 수 있다. 또한 단일 기준자료가 아닌 5개 기준자료의 시계열 중 비동질성을 내포할 가능성이 존재하는 양극 값을 제거하여 보다 동질한 기준시계열을 산정하는 장점이 있다.

3. 연구방법

3.1 연구목적

본 연구의 목적은 장치산업에서의 미세한 공정변동을 탐지하여 예방 활동에 활용하는 방안을 설계하는 것이다. 장치산업에서 설비효율은 부품교체, 유지보수 등의 다양한 활동을 통해 지속적으로 유지되거나 급격하게 변동하게 된다. 본 연구에서는 설비효율이 급격하게 변화

하는 시점을 탐지하고 이에 영향을 미치는 요인을 파악하기 위해 상대적 공정변동 탐지 방법을 제안한다.

3.2 연구절차

장치산업에서 공정의 이상변동을 탐지하기 위해 측정환경이나 주변 환경의 변화에 따른 자료의 변동을 제거하고 가능한 공정 자체의 변동만을 반영한 시계열을 생산하기 위해 자료수집→자료탐색→유사성평가→기준시계열생성→공정변동탐지와 같은 절차에 따라 수행한다.

3.2.1 자료수집

자료수집 단계에서는 먼저 평가대상 공정 또는 라인을 선정하고 이에 대한 자료를 수집한다. 다음으로 평가대상 자료와 유사하다고 판단되는 비교대상 공정 또는 라인에 대한 측정 자료를 수집한다.

3.2.2 자료탐색

자료탐색은 수집된 측정자료의 특성과 품질을 파악하고 분석을 위한 자료를 준비하는 단계로 데이터 탐색, 데이터 품질 및 데이터 준비 등 세가지 세부단계로 구분할 수 있다. 데이터 탐색은 수집된 데이터에 대한 간단한 기술통계, 그래프 등을 이용하여 기초적인 분포와 일반적인 특성을 파악한다. 데이터 품질은 각종 품질 점검 기법을 이용하여 통계분석에 사용될 수 있는 데이터들에 대한 품질을 파악한다. 데이터 준비는 실제 통계분석에 사용될 데이터를 선택, 정제, 파생변수 생성 및 데이터 통합 단계를 거쳐 최종적으로 통계분석을 위한 데이터셋을 생성한다.

3.2.3 유사성 평가

유사성 평가 단계에서는 평가대상자료와 비교대상자료들 간의 상관분석을 수행하여 비교대상자료를 확정하고 비교대상 자료에 대한 가중치를 결정한다.

상관분석은 피어슨의 적률상관계수(Pearson product-moment correlation coefficient)를 이용하며, 비동질성의 효과를 피하기 위해 1차 차분한 시계열 자료를 이용한다.

상관분석 결과를 토대로 비교대상자료 선정에 대한 적합성을 평가하고 비교대상 자료들 중 평가대상자료와 낮은 상관관계를 보이는 자료를 제외함으로써 최종 비교대상자료를 확정한다. 가중치 산정은 비교대상자료와 평가대상자료 간의 유사성 정도에 따라 가중치를 부여한다.

3.2.4 기준시계열 산정

기준시계열 산정 단계에서는 비교대상자료의 가중치를 이용하여 기준 시계열 자료와 아노말

리(Anomaly)를 산정한다.

평가대상자료의 기준시계열은 (1)과 같이 계산할 수 있다. (1)에서 w_j 는 비교자료 j 의 가중치이고 v_j 는 비교자료 j 의 표준화된 측정치이다.

$$\sum_{j=1}^k (w_j \times v_j) \quad (1)$$

아노말리시계열 산정은 평가대상자료를 표준화한 값에서 기준시계열자료를 표준화한 값을 차감한 값으로 계산한다(표준화 측정치-표준화 기준시계열).

3.2.5 공정변동 탐지

공정변동 탐지 단계에서는 구간별 검정통계량을 계산하고 검정통계량이 최대이면서 기준통계량보다 높은 지점(일자)을 탐지한다.

공정변동이 없으며 측정값이 x_1, \dots, x_n 인 n 개의 아노말리시계열은 F_0 분포를 따른다고 가정하자. 만약 τ 시점에서 공정변동이 발생했다면 τ 시점 이전의 분포는 F_0 분포를 따르고 이후부터는 F_0 와 다른 F_1 분포를 따르게 될 것이다($F_0 \neq F_1$). 특정시점 k 에서 공정변동이 발생되었다고 한다면 다음과 같이 가설을 세울수 있다.

$$H_0 : X_i \sim F_0(x, \theta_0), \quad i = 1, \dots, n$$

$$H_1 : X_i \sim \begin{cases} F_0(x, \theta_0), & i = 1, \dots, k \\ F_1(x, \theta_1), & i = k+1, \dots, n \end{cases}$$

여기에서 θ 는 각 분포에 대한 미지의 모수이다.

위의 가설을 검정하는 방법은 관측값의 분포에 따라 달라진다. 만약 관측값이 정규분포를 따른다면 평균변화 탐지를 위해 Student t-test를 적용하고(Hawkins 등(2003)), 변동의 변화를 탐지하기 위해 F-test를 수행한다. 정규분포를 가정할 수 없을 경우에는 비모수적 검정 방법으로 위치변화 탐지를 위해 Mann-Whitney test를 사용하며, Scale 변화를 탐지하기 위해 Mood test를 사용한다(Ross 등(2011)).

선정된 검정방법에 따라 검정통계량 $D_{k,n}$ 을 계산하며 임계값 $h_{k,n}$ 보다 크면 변동이 발생했다고 할 수 있다. 변동점을 미리 알 수 없기 때문에 가능한 모든 자료를 분리하여 $D_{k,n}$ 을 계산하고 최대값을 구한다. 이를 수식으로 표현하면 (2)와 같다.

$$D_n = \left| \max \frac{D_{k,n} \cdot -\mu D_{k,n}}{\sigma D_{k,n}} \right|, \quad 1 < k < n \quad (2)$$

따라서 변동점 추정($\hat{\tau}$)은 D_n 을 최대로하는 k 시점으로 (3)과 같다.

$$\hat{\tau} = \arg \max D_{k,n} . \tag{3}$$

두 표본에 대한 검정방법을 확장하여 새로운 측정값을 수집할 때마다 변동점을 탐지함으로써 공정변동을 최대한 빨리 탐지할 수 있다.

4. 공정자료 분석

4.1 자료 소개

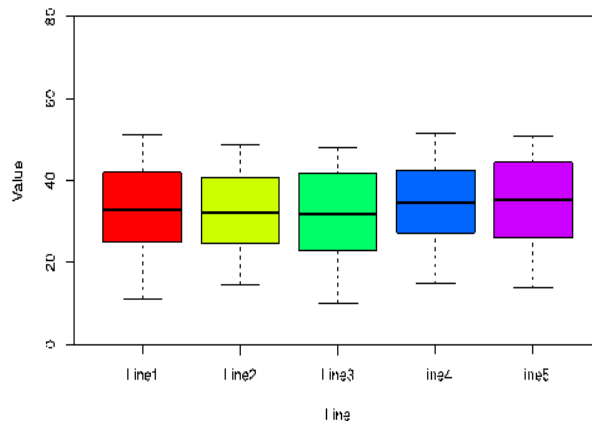
본 연구에서 설계한 공정변동 탐지 방법에 대한 정확성 및 활용성을 평가하기 위해 A기업의 생산라인1(Line1)을 평가대상으로 선정하였다. 생산라인1의 공정변동을 평가하기 위해 동일한 네 개의 생산라인을 선택하고 자료를 수집하였다. 수집된 자료는 2010년 1월 1일부터 2010년 12월 31일까지의 일생산량을 가동시간으로 나눈 시간당 생산량이다.

4.2 자료 탐색

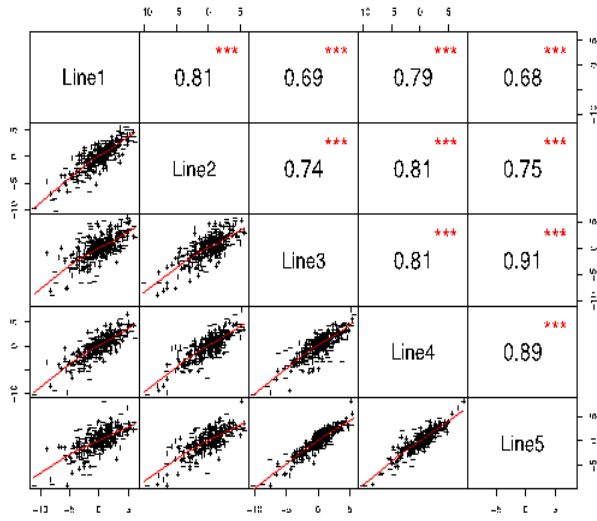
생산라인별 수집된 자료에 대한 특성을 확인하기 위해 기술통계량과 상자도표를 산출하였으며 결과는 <표 1> 및 <그림 1>과 같다. 데이터 품질검토는 이상치와 결측치를 확인하였으며 특이점을 발견하지 못하였다.

<표 1> 생산라인별 기술통계량

	N	Mean	Med	SD	CV
Line1	365	33.1	33.0	9.9	0.3
Line2	365	32.6	32.3	9.1	0.3
Line3	365	31.6	31.7	10.3	0.3
Line4	365	34.7	34.4	9.4	0.3
Line5	365	34.4	35.4	10.3	0.3



<그림 1> 생산라인별 상자도표

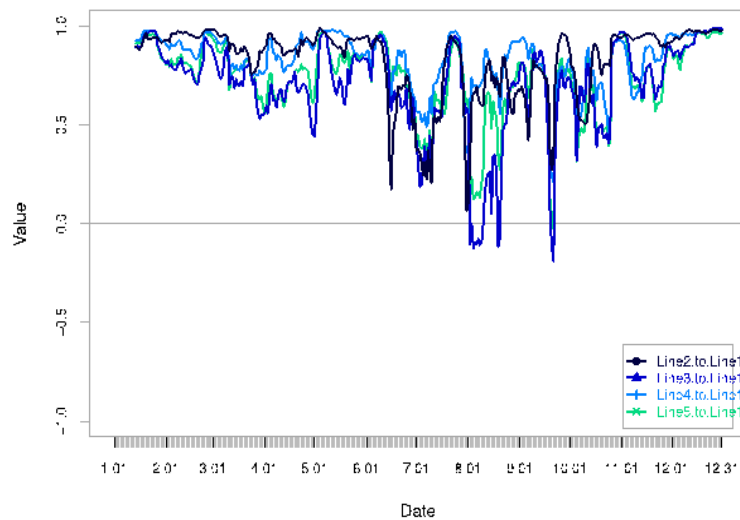


<그림 2> 생산라인간 산점도 및 상관계수

4.3 유사성 평가

평가대상인 생산라인1과 특성이 유사한 네 개의 다른 생산라인들(비교대상 생산라인)과의 유사성을 확인하기 위해 피어슨의 적률상관분석을 실시하였다. 상관분석은 자료의 비동질성 효과를 피하기 위해 1차 차분한 자료를 이용하여 전체 자료에 대한 산점도 및 상관계수와 기간별 롤링 상관계수를 산출하였다.

전체 자료에 대한 상관분석 결과 <그림2>와 같이 생산라인1과 참조 생산라인들 간의 상관계수가 0.68 이상으로 전반적으로 높은 상관관계를 보였으며, 특히 생산라인2와의 상관계수가 0.81로 가장 높게 나타났다.



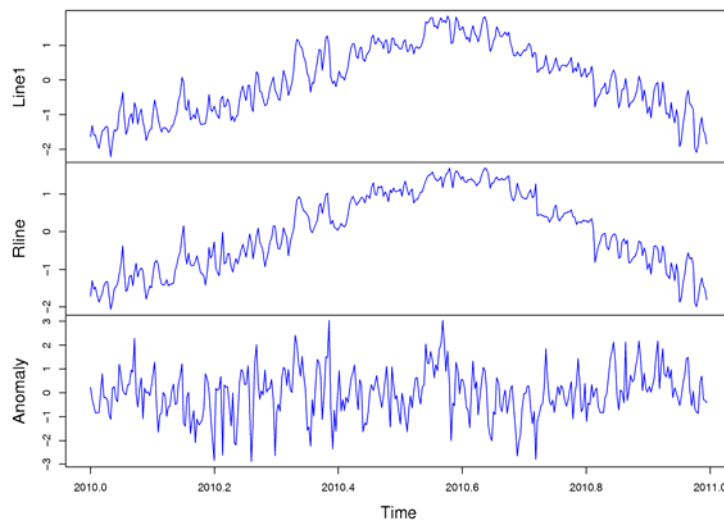
<그림 3> 생산라인1에 대한 롤링상관분석

또한 시간의 흐름에 따른 롤링 상관계수를 산정하였으며 결과는 <그림 3>과 같다. <그림 3>에서 보는 바와 같이 특정 시간대에 상관계수가 급격히 떨어짐을 확인할 수 있다. 이는 생산라인에 발생한 사건에 따라 일시적으로 상관관계가 낮아진 것으로 나타났다. 상관분석 결과를 요약하면, 비교대상 생산라인 4개 모두 생산라인1과 높은 상관관계를 보여 적합하다고 판단하였다. 기준시계열을 생산하기 위한 가중치는 시간에 따라 상관계수가 다르기 때문에 롤링상관계수를 이용하여 가중치를 산정하였다.

4.4 기준시계열 산정

생산라인1을 평가하기 위해 생산라인1을 제외한 나머지 4개 생산라인의 자료를 이용하여 기준 시계열 자료와 아노말리(Anomaly) 시계열 자료를 산정하였다. 생산라인1의 기준시계열은 (2)로 계산하였으며, 여기서 w_j 는 생산라인 j 의 가중치이고 v_j 는 생산라인 j 의 표준화된 측정치이다.

<그림 4>에서 Line1은 생산라인1의 표준화 생산량이고(상), Rline은 표준화 기준시계열 데이터를 나타내며(중), Anomaly는 표준화 아노말리(하)의 산정 결과이다.

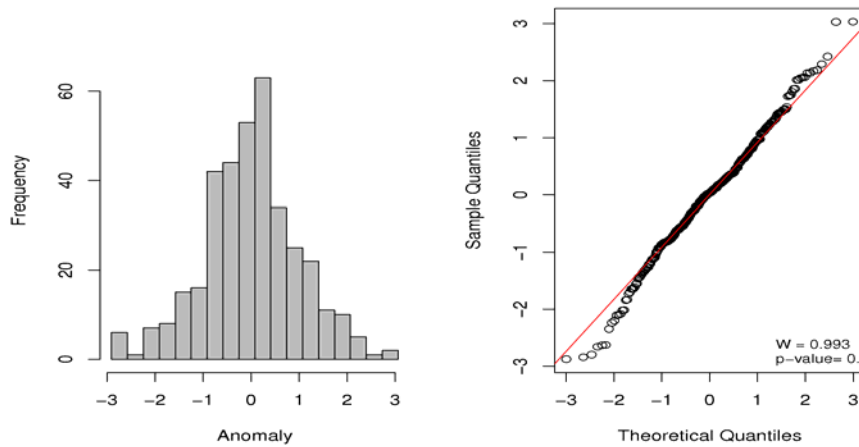


<그림 4> 생산라인1, 기준시계열 및 아노말리 시계열 그래프

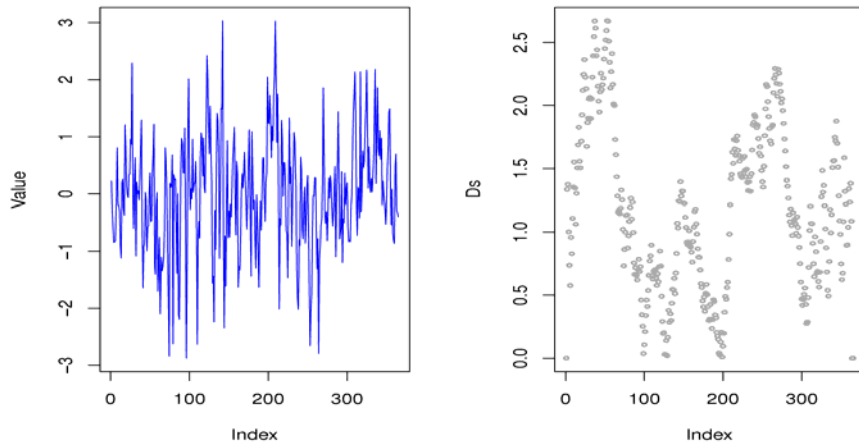
4.5 공정변동 탐지

공정변동을 탐지하기 위한 검정방법을 결정하기 위해 아노말리에 대한 정규성 검정을 실시하였다. <그림 5>은 아노말리에 대한 히스토그램과 Q-Q Plot을 나타내고 있다. 정규성 검정은 Shapiro-test를 수행하였으며 W검정통계량은 0.993으로 유의수준 0.01에서 정규분포를 따른다고 할 수 없으며, 유의수준 0.01은 미세한 공정변화를 탐지하기 위해 적용했

다.(p-value: 0.068). 그리고 정규성 검정 결과 정규분포를 가정할 수 없기 때문에 공정변동 감지를 위한 검정방법으로 위치변화 탐지를 위해 Mann-Whitney test와 Scale 변화를 탐지하기 위해 Mood test를 수행하였다.



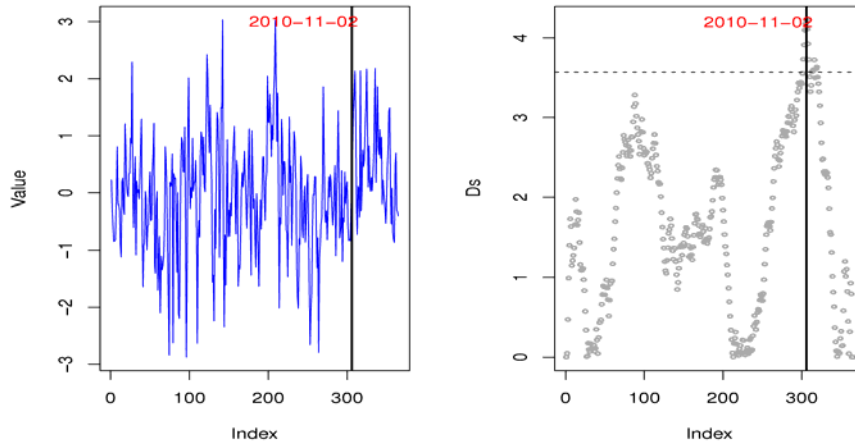
<그림 5> 아노말리에 대한 히스토그램 및 Q-Q Plot



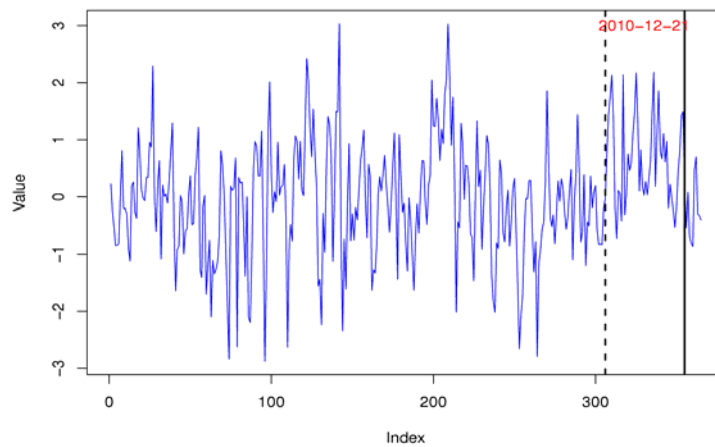
<그림 6> Mann-Whitney test 결과

공정에 대한 위치변화를 탐지하기 위한 Mann-Whitney test 결과 2010년 11월 2일에 공정변동이 탐지되었다 <그림 6>. 이 때의 검정통계량 D 는 4.23으로 유의수준 0.01에서 통계적으로 공정변동이 발생되었다고 할 수 있다. <그림 6>의 오른쪽에 있는 검정통계량 그래프에서 점선은 유의수준 0.01에 대한 임계값을 나타낸다.

공정에 대한 Scale 변화를 탐지하기 위해 Mood test를 수행한 결과는 <그림 7>과 같다. 최대 검정통계량 D 는 2.67로 유의수준 0.01에서 통계적으로 Scale 변화가 발생되었다고 할 수 없다. 즉, 생산라인1의 Scale은 다른 생산라인들과 변동에서 차이가 없는 것으로 나타났다.

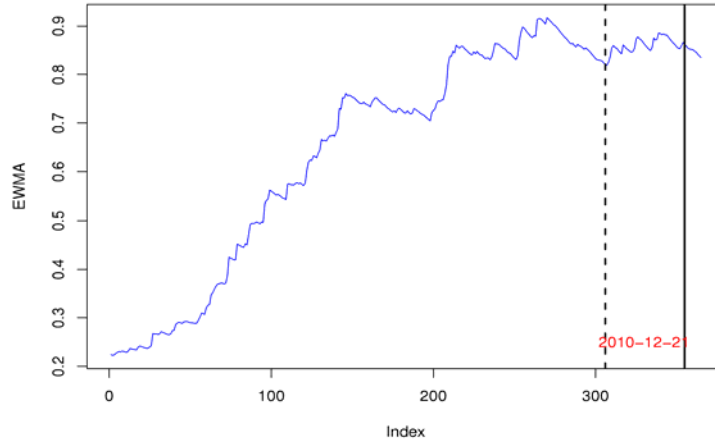


<그림 7> Mood test 결과



<그림 8> Mann-Whitney test를 통한 공정변동 탐지

생산라인의 이상 변동을 최대한 빨리 탐지하기 위해서 새로운 자료가 수집되고 있음을 가정하고 동일한 Mann-Whitney test를 수행하였으며 결과는 <그림 8>, <그림 9>와 같다. <그림 8>과 <그림 9>에서 점선은 공정변동일을 나타내며 점과선은 탐지일을 나타낸다. 즉 변동이 발생된 이후 약 49일 후에 변동을 정확하게 탐지할 수 있다.



<그림 9> Mann-Whitney test를 통한 EWMA

5. 결론

장치산업에서 공정의 변동을 파악하는 것은 생산성과 직결되기 때문에 매우 중요하며 이를 파악하기 위해서는 장기간의 정확한 데이터가 수집되어야 한다. 그러나 장치산업의 경우 공정개선 및 유지보수 활동으로 인해 정확한 측정치를 수집하기 어려운 실정이다.

본 연구에서는 다양한 공정관리 방법론을 참고하여 장치산업에서의 미세한 공정변동을 탐지할 수 있는 객관적이고 합리적인 탐지방법을 설계하고 유사 사례를 통해 검증하였다.

주요 연구결과를 요약하면 다음과 같다. 첫째, 생산라인의 변동을 탐지하기 위해 유사한 생산라인을 선정하고 기준시계열을 통한 아노말리를 분석하는 상대적 탐지방법을 사용하였다. 둘째, 기준시계열을 생성하는데 있어 시간대별 높은 상관관계를 보이는 생산라인에 높은 가중치를 부여함으로써 기준시계열 생산의 객관성을 높였다. 셋째, 아노말리에 대한 동질성 검정을 통해 공정변동 시점을 추정할 수 있었으며, 자료가 순차적으로 수집됨을 고려하여 공정변동을 탐지할 수 있는 시점을 추정하였다. 본 연구에서 수행한 공정변동 탐지 체계는 장치산업 뿐만 아니라 다양한 산업분야에서 응용될 수 있을 것이라 판단된다.

참고문헌

- [1] Easterling, D. R. and Peterson, T. C. (1994). Global Climate Laboratory, National Climatic Data Center, Asheville, NC 28803, USA, Accepted 17 November 1994.
- [2] Hawkins, D., Qiu, P., Kang, C. (2003). The Changepoint Model for Statistical Process

- Control, *Journal of Quality Technology*, 35, 355-366.
- [3] Kwon, J. B., Lee, J. S, Lee, S. H., Jun, C. H. and Kim, K. J. (2005), Multiresponse Optimization through a New Desirability Function Considering Process Parameter Fluctuation, *Journal of the Korean Operations Research and Management Science Society*, 30(1),95-104.
- [4] MacGregor, J. F. and Harris, T. J. (1990). Discussion of 'EWMA control schemes : properties and enhancement' by Lucas and Saccucci, *Technometrics*, 32, 23-26.
- [5] Pan, R. and Del Castillo, E. (2003). Integration of sequential process adjustment and process monitoring techniques. *Quality and Reliability Engineering International*, 19, 371 - 86.
- [6] Park, C. and Lee, J. (2008). An integrated process control scheme based on the future loss. *The Korean Journal of Applied Statistics*, 21, 247 - 64.
- [7] Ross, G. J., Tasoulis, D. K., Adams, N. M. (2011). A Nonparametric Change-Point Model for Streaming Data, *Technometrics*, 53(4).
- [8] Solow, A. R.(1987). Testing for climate change: An application of the two-phase regression model. *Journal of Climate and Applied Meteorology*, 26(10): 1401-1405.
- [9] Vincent, C. (2002). Influence of climate change over the 20th century on four french glacier mass balances. *Journal of Geophysical Research*, 2002, vol 107.