

이 현 철[†]

한국항공대학교 경영학과

Residual-based Robust CUSUM Control Charts for Autocorrelated Processes

Hyun Cheol Lee[†]

Department of Business Administration, Korea Aerospace University

The design method for cumulative sum (CUSUM) control charts, which can be robust to autoregressive moving average (ARMA) modeling errors, has not been frequently proposed so far. This is because the CUSUM statistic involves a maximum function, which is intractable in mathematical derivations, and thus any modification on the statistic can not be favorably made. We propose residual-based robust CUSUM control charts for monitoring autocorrelated processes. In order to incorporate the effects of ARMA modeling errors into the design method, we modify parameters (reference value and decision interval) of CUSUM control charts using the approximate expected variance of residuals generated in model uncertainty, rather than directly modify the form of the CUSUM statistic. The expected variance of residuals is derived using a second-order Taylor approximation and the general form is represented using the order of ARMA models with the sample size for ARMA modeling. Based on the Monte carlo simulation, we demonstrate that the proposed method can be effectively used for statistical process control (SPC) charts, which are robust to ARMA modeling errors.

Keywords : Residual-based Control Chart, Cumulative Sum, Robustness, Autoregressive Moving Average

1. 서 론

통계적 공정관리(statistical process control) 방법 중 관리도(control chart)는 생산 공정에 대한 적절한 모니터링을 통해 품질과 직접적으로 관련된 주요 특성의 변화를 감지하는 유용한 방법이다. 특히 다양한 통계적 공정 관리도 종류 중에서 누적합(Cumulative Sum(CUSUM)) 관리도와 Exponentially Weighted Moving Average(EWMA) 관리도는 공정 품질 특성의 미세한 변화를 감지하기 위해 널리 사용되는 방법들이다. 이 두 종류의 관리도는 Shewhart

형태의 관리도와는 달리 공정 변동의 크기와 감지 속도의 관계에서 작은 크기의 공정 변화에는 민감하게 반응하는 한편, 큰 크기의 공정 변화에 대해서는 비교적 느리게 반응한다.

누적합 관리도는 공정 변화를 감지하는 능력에 있어서 EWMA 관리도와 매우 유사한 성능을 보여준다. 그러나 이러한 성능상의 장점에도 불구하고 통계적 공정관리에서의 중요 이슈 중 하나인 자기상관 관측치(data) 적용에 관한 연구에 있어서는 EWMA 관리도만큼 연구결과가 많지 않다[1, 2]. 특히 이 주제와 관련하여, 최근의 연구결과인

논문접수일 : 2012년 05월 24일 게재확정일 : 2012년 07월 31일

[†] 교신저자 hcleee@kau.ac.kr

※ 이 논문은 2011년도 한국항공대학교 교비지원 연구비에 의하여 지원된 연구의 결과임.

Autoregressive Moving Average(ARMA) 시계열 모델을 이용한 잔차 기반 관리도의 적용이 자기상관 특성을 갖는 공정 관측치를 관리도에 직접 적용하는 방법보다 관리도의 성능이나 모델링 오차에 대한 강건성 측면에서 우수함을 증명하는 대부분의 연구에서도 EWMA 관리 통계량을 이용하여 결과를 도출하였다[7].

누적합 관리도 통계량을 이용한 ARMA 모델링 오차에 대한 강건한 방법론 개발 등의 연구에 있어 가장 큰 어려움은 누적합 통계량 형태에 의한 수학적 전개가 쉽지 않다는 점이다. 누적합 통계량은 제 2장에서 자세히 설명하겠지만 최대값 함수를 통계량에 포함하기 때문에 통계량의 형태 변화 또는 통계량의 수식 전개가 EWMA 통계량 만큼 쉽지 않다. 이러한 이유로 누적합 관리도에 대해서는 통계량에 대한 정량적인 수정을 이용한 강건 설계 방법론의 연구 결과가 거의 존재하지 않는다.

앞에서 언급한 공정 데이터의 자기상관 특성은 최근의 공정 모니터링 환경의 큰 변화에 기인한다. 최근엔 센서에 의한 실시간 측정과 저장이 가능한 계측기술이 현저하게 발전하여 양산제품에 대한 다량의 실시간 관측치 확보가 가능해졌다. 실시간으로 계측이 빈번해지면 계측 데이터의 특성은 통계적으로 독립(statistically independent)하지 않고 자기 상관관계(autocorrelation)를 갖는다. 이는 기존의 일반적인 관리도에 적용되는 품질 특성 관측치의 통계적 특성이 상호 독립적이고 동일한 분포(iid)를 따라야 한다는 가정과 배치된다.

공정 데이터가 자기상관 특성을 가질 때 관리도의 적용은 크게 세 가지 방법으로 발전되어 왔다. 첫 번째 방법은 자동 제어 이론에 기반을 둔 피드백 제어 방법을 이용하여 공정의 자기상관 특성을 직접 제거한 후 공정 데이터를 일반적인 관리도에 적용하는 방법이다 [3]. 자동 제어에 의한 자기상관 특성이 정확히 제거된다면 공정 데이터는 서로 독립적이고 동일한 분포를 갖는 일반적인 관리도 적용의 가정을 만족할 것이다. 두 번째 방법은 자기상관 특성을 갖는 공정 데이터를 ARMA 모델을 사용하여 모델링한 후, 모델의 one-step ahead prediction error, 즉 잔차를 이용하여 일반적인 관리도에 적용하는 방법이다. 이 방법에서는 만일 ARMA 모델의 모델링 오차가 존재하지 않는다면 잔차는 서로 독립이면서 동일한 분포를 갖게 되어 일반적인 관리도 적용의 가정을 만족하게 된다[4, 8, 11, 16]. 마지막으로 세 번째 방법은 공정 데이터의 자기상관 크기만큼을 관리도 통계량에 반영하여 수정된 관리도 통계량을 이용함으로써 자기상관 특성을 갖는 데이터를 직접 모니터링 하는 방법이다[12, 17, 19].

첫 번째 방법에서 공정교란(process disturbance)이 ARMA 모델을 따르고 모델링 오차가 존재할 경우, 최소평균 제

곱오차 제어 방법(minimum mean square error control)으로 자기상관을 제거한 공정 데이터를 관리도에 적용하는 경우의 강건성은 두 번째 방법에서 제공하는 ARMA 모델링 오차가 존재하는 잔차 기반 관리도의 강건성 보다 좋지 않거나 동일하다고 증명되었다[7]. 동일 논문에서 자기상관 데이터를 그대로 사용하는 방법인 세 번째 방법도 첫 번째 방법과 마찬가지로 두 번째 방법보다 일반적으로 ARMA 모델링 오차에 대한 강건성이나 성능 측면에서 우월하지 않다고 증명되었다. 이는 세 번째 방법 역시 자기상관 크기를 측정하는 상관계수가 공정 데이터로부터 추정되어야 하므로 모델링 오차 문제를 완전히 극복할 수는 없기 때문이다. 두 번째 방법 또한 현실적으로 모델링 오차를 피할 수는 없기 때문에 관리도 설정 시 의도된 관리도의 특성이 기대에 미치지 못할 수 있다. 하지만 세 가지 방법 중 두 번째 방법이 모델링 오차에 대한 강건성 측면에서 가장 우수하다고 증명되었다[14].

기존의 많은 연구가 이들 방법간 ARMA 모델링 오차에 대한 강건성 및 성능 평가를 시뮬레이션 등의 방법을 통해 비교한 반면, Lee[14]는 ARMA 모델링 오차에 대한 관리도 통계량의 분산 통계량 변화율을 모수에 대한 편미분을 이용하여 수학적식에 의한 정량화를 통해 비교결과를 제시하였다. EWMA 관리도 통계량을 이용하여 강건성과 성능을 비교한 결과, 잔차 기반 EWMA 관리도가 자기상관 데이터를 그대로 사용한 EWMA 관리도보다 두 가지 측면에서 모두 우수함을 확인하였다.

앞에서 언급한 바와 같이 지금까지 ARMA 모델링 오차에 강건한 통계적 공정관리 설계방법에 대한 연구는 주로 공정의 미세 변동을 감지하는데 효과적이면서 관리도 통계량 식의 변형이 쉬운 EWMA 관리도에 대해서 이루어졌다[5, 6, 15, 18]. 그러나 이 논문에서는 미세한 공정변화에 대한 감지 성능은 EWMA 관리도와 유사하지만 관리 통계량에 대한 수학적 변형이 쉽지 않은 특성 때문에 연구결과가 많지 않은 누적합 통계량을 이용하여 ARMA 모델링 오차에 강건한 통계적 공정관리를 제안하고자 한다. 특히, 이 논문에서 제안하는 방법은 기존의 강건 설계 방법들이 관리도 통계량의 분산 통계량에 모델링 오차를 고려하는 형태로 변형되어 적용된 점과 다르게 누적합 관리도 통계량 형태의 변화보다 잔차 자체에서 모델링 오차를 고려한 잔차의 분산 통계량을 일반식으로 유도하여 그 크기만큼을 누적합 관리도의 모수에 보정해주는 형태로 적용한다. 즉, 본 논문에서 제안하는 통계적 공정관리 설계 방법은 자기상관성을 갖는 관측치에 적용 가능한 새로운 형태의 누적합 관리도로서 ARMA 모델링 오차를 포함하는 잔차의 분산 통계량을 일반식의 형태로 유도하여 이 모델링 오차를 관리도 모수에 반영하는 것

이 주요 결과이다.

본 논문의 제 2장에서는 논문에서 가정된 모델에 대한 설명으로 ARMA 모델, 모델링 오차가 고려된 잔차와 누적합 관리도가 소개된다. 제 3장에서는 이 논문에서 제안하는 자기상관 데이터 모니터링 목적의 잔차 기반 누적합 관리도의 설계에 필요한 모델링 오차가 포함된 잔차 분산 통계량을 유도한다. 제 4장에서는 실제 ARMA 모델링 오차에 강건한 누적합 관리도의 설계방법을 예제를 통해 설명한다. 제 5장에서는 이 논문의 결과를 정리하고 결론을 맺는다.

2. 가정 모델

2.1 Autoregressive Moving Average 모델

본 논문에서 자기상관 특성을 갖는 공정 데이터를 모델링하기 위하여 시계열 모델 중 가장 널리 알려진 모델 중 하나인 ARMA 모델을 이용한다. 일반적으로 자기상관 특성을 갖는 공정 관측치는 ARMA(p, q) 모델로 충분히 정확하게 표현될 수 있고 아래와 같이 정의 된다[9].

$$\Phi(B)x_t = \Theta(B)a_t \quad (1)$$

x_t 는 자기상관 특성을 갖는 공정 관측 데이터, a_t 는 평균 0, 분산 σ_a^2 인 정규분포 가정의 random process 데이터, B 는 $Bx_t = x_{t-1}$ 로 정의된 후행 진이 연산자이다. $\Phi(B) = 1 - \phi_1 B - \phi_2 B^2 \dots - \phi_p B^p$ 는 차수 p 의 AR 다항식, $\Theta(B) = 1 - \theta_1 B - \theta_2 B^2 \dots - \theta_q B^q$ 는 차수 q 의 MA 다항식이다. 모수벡터 $\gamma = [\phi_1 \phi_2 \dots \phi_p \theta_1 \theta_2 \dots \theta_q]$ 로 나타내고, $\hat{\gamma}$ 는 γ 의 추정량이다.

2.2 ARMA 잔차

ARMA 모델에서 잔차, 즉 one-step ahead prediction error는 공정 데이터로부터 ARMA 모델의 모수 추정을 통해 아래의 식과 같이 계산된다.

$$e_t = \frac{\hat{\Phi}(B)}{\hat{\Theta}(B)} x_t = \frac{\hat{\Phi}(B)\Theta(B)}{\hat{\Theta}(B)\Phi(B)} a_t \quad (2)$$

식에서 보는 바와 같이 만일 ARMA 모델링이 오차 없이 완벽하게 수행될 수 있다면 잔차는 정확히 a_t 와 일치하게 되어 평균 0, 분산 σ_a^2 인 정규분포 가정의 random process 데이터가 된다. 즉, 모델링 오차가 없다면 잔차는 자기상관 관계가 제거되어 관리도에 적용 가능한 관측치

의 통계적 특성을 만족하게 되므로 임의의 일반적인 관리도 통계량을 적용할 수 있다. 하지만 실제 공정 데이터를 이용하여 ARMA 모델링을 할 경우, 일반적으로 모델링 오차를 피할 수 없기 때문에 모델링 오차가 포함된 실제 잔차는 ARMA($p+q, p+q$)가 된다.

2.3 누적합 관리도

누적합 관리도는 통계적 공정관리 기법 중에서 공정의 미세변동을 감지하기 위해 널리 사용되는 관리도 중 하나로 단방향 누적합 관리도와 양방향 누적합 관리도가 있다. 본 논문에서는 양방향 누적합 관리도를 이용하여 강건 설계 방법을 설명한다. 제 2.2절에서 소개된 잔차가 적용된 누적합 통계량은 아래와 같이 정의된다.

$$S_t^+ = \max\{0, S_{t-1}^+ + e_t - K\}$$

그리고

$$S_t^- = \max\{0, S_{t-1}^- - e_t - K\} \quad (3)$$

만일 S_t^+ 가 H 보다 크면 양의 방향(공정 평균의 증가)으로 공정 평균의 변화가 발생했다고 판단하고 S_t^- 가 H 보다 크면 음의 방향(공정 평균의 감소)으로 공정 평균의 변화가 발생했다고 판단한다. 여기서 K 와 H 는 사용자에게 의해 정의되는 관리도 모수로 각각 참고값(reference value), 결정 구간(decision interval)을 나타낸다. 누적합 관리도의 설정 시 K 와 H 의 결정은 공정 데이터가 통계적으로 상호 독립이고 동일한 분포(iid)라는 가정을 만족할 때 Montgomery [13]에 의해 잘 설명되어 있다.

관리도 설정 시 일반적으로 이용되는 제 1종 오류, 제 2종 오류가 누적합 관리도에서는 통계량의 특성상 사용이 어렵다. 누적합 관리도에서는 이들 오류 대신에 공정 안정 상태의 평균 런 길이(In-control average run length (ARL))와 공정 불안정 상태의 평균 런 길이(Out-of-control ARL)를 사용한다. 본 논문에서는 공정 안정 상태의 평균 런 길이가 500(제 1종 오류 = 0.002)인 경우와 370(제 1종 오류 = 0.0027)인 경우의 다양한 $K/\sigma_e, H/\sigma_e$ 조합에 대해 강건 설계 방법을 제시한다. 이 논문에서 공정 데이터가 통계적으로 상호 독립적이고 동일한 분포(iid)라는 가정을 만족할 때 사용되는 $K/\sigma_e, H/\sigma_e$ 조합은 <표 1>에 정리되어 있다. 공정 안정 상태의 평균 런 길이가 370일 때의 값은 Montgomery[13]와 동일하고, 500일 때의 값은 Monte carlo 시뮬레이션에 의해 계산되었다. 표에서 K/σ_e 가 커질수록 누적합 관리도는 Shewhart X 관리도의 특성과 비슷해진다. 제 4장에서 이 조합의 값들이 모델링 오차가 고려된 잔차 분산 통계량에 의해 적절히 보정된다.

<표 1> 데이터 iid일 때 누적합 관리도의 K, H 최적 조합
(In-control ARL = 370/500 경우)

IC ARL	K/σ _e	0.2	0.5	0.75	1.0	1.25	1.5	2.0	2.5	3.0
370	H/σ _e	9.25	4.77	3.34	2.52	1.99	1.61	·	·	·
500		9.96	5.07	3.54	2.67	2.11	1.71	1.11	0.59	0.09

3. 잔차 분산 통계량의 근사

이 장에서 ARMA 모델링 오차가 존재하는 잔차 분산 통계량의 근사식을 유도한다. 잔차 분산 통계량은 2차 Taylor 근사법에 의해 표현되고 근사식은 임의의 ARMA 모델에 대해 일반식으로 표현된다. 앞장에서 설명한 바와 같이, 모델의 잔차는 만일 모델링 오차가 존재한다면 $\Phi(B)$, $\Theta(B)$ 의 추정량을 이용하여 식 (2)로 표현 가능하다. 식 (2)는 다시 임펄스 반응 방법을 이용하여 아래 식으로 표현 가능하다.

$$e_t = \frac{\hat{\Phi}(B)}{\hat{\Theta}(B)} x_t = \frac{\hat{\Phi}(B)\Theta(B)}{\hat{\Theta}(B)\Phi(B)} a_t$$

$$= H(B)a_t = \sum_{j=0}^{\infty} h_j a_{t-j} \quad (4)$$

$H(B) = \hat{\Theta}(B)^{-1}\hat{\Phi}(B)\Theta(B) = \sum_{j=0}^{\infty} h_j B^j$ 이고 $h_j(j=0, 1, 2, 3, \dots)$ 은 ARMA($p+q, p+q$)인 전이함수(transfer function)의 임펄스 반응계수이다. 따라서 임펄스 반응 방법을 이용한 잔차의 분산 통계량은 다음의 식으로 계산할 수 있다.

$$\sigma_e^2 = \sigma_a^2 \sum_{j=0}^{\infty} h_j^2 \quad (5)$$

또한 이 잔차의 분산 통계량을 2차 Taylor 근사법을 이용해 표현한다면 이 통계량은 다시 아래와 같이 표현 가능하다. 보는 바와 같이 이 식은 $\hat{\gamma}=\gamma$ 에 대해서 Taylor 근사한 결과이다.

$$\sigma_e^2 = \sigma_e^2|_{\hat{\gamma}=\gamma} + \left[\frac{\partial \sigma_e^2}{\partial \hat{\gamma}} \Big|_{\hat{\gamma}=\gamma} \right]^T (\hat{\gamma}-\gamma)$$

$$+ \frac{1}{2} (\hat{\gamma}-\gamma)^T \left[\frac{\partial^2 \sigma_e^2}{\partial \hat{\gamma}^2} \Big|_{\hat{\gamma}=\gamma} \right] (\hat{\gamma}-\gamma) \quad (6)$$

만일 확률변수 γ 에 대하여 기대값을 취한다면 식 (6)은 아래의 식과 같이 표현된다. 여기서 우리는 N의 크기

가 커짐에 따라 $\hat{\gamma}$ 의 분포는 평균 행렬 γ , 공분산 행렬 $\Sigma_{\hat{\gamma}}$ 인 다변량 정규분포를 따른다는 사실을 이용한다[10].

$$\hat{E}[\sigma_e^2|\gamma] \cong \sigma_e^2|_{\hat{\gamma}=\gamma} + \frac{1}{2} tr \left\{ \frac{\partial^2 \sigma_e^2}{\partial \hat{\gamma}^2} \Big|_{\hat{\gamma}=\gamma} \Sigma_{\hat{\gamma}} \right\} \quad (7)$$

여기서 $\sigma_e^2|_{\hat{\gamma}=\gamma} = \sigma_a^2$ 이고 이는 ARMA 모델링 오차가 존재하지 않을 때의 a_t 의 분산을 의미한다. 즉, 근사된 잔차 분산 통계량의 기대치는 ARMA 모델링 오차가 존재하지 않을 때 a_t 의 분산 통계량(첫 번째 항)과 ARMA 모델링 오차가 존재함에 따라 생성되는 부분(두 번째 항)으로 구성된다. tr 은 행렬의 trace 연산자를 나타내고 $\Sigma_{\hat{\gamma}}$ 는 $\hat{\gamma}$ 의 공분산 행렬이다. 이제 잔차 분산 통계량의 유도에 필요한 부분은 식 (7)의 두 번째 항의 전개이다. 이를 위해 우리는 식 (5)을 이용한 $\hat{\gamma}$ 에 대한 이차 미분결과가 필요하다. 식 (5)의 양변을 두 번 미분하면 아래의 식과 같이 표현된다.

$$\frac{\partial^2 \sigma_e^2}{\partial \hat{\gamma}^2} = 2\sigma_a^2 \sum_{j=0}^{\infty} \left[h_j \frac{\partial^2 h_j}{\partial \hat{\gamma}^2} + \left[\frac{\partial h_j}{\partial \hat{\gamma}} \right] \left[\frac{\partial h_j}{\partial \hat{\gamma}} \right]^T \right] \quad (8)$$

따라서 식 (8)을 이용하여 식 (7)을 다시 표현하면

$$\hat{E}[\sigma_e^2|\gamma] \cong \sigma_a^2 + \sigma_a^2 tr \left\{ \sum_{j=0}^{\infty} h_j C_j \Sigma_{\hat{\gamma}} + \sum_{j=0}^{\infty} c_j c_j^T \Sigma_{\hat{\gamma}} \right\} \quad (9)$$

여기서 $C_j = \partial^2 h_j / \partial \hat{\gamma}^2|_{\hat{\gamma}=\gamma}$ 이고, $c_j = \partial h_j / \partial \hat{\gamma}|_{\hat{\gamma}=\gamma}$ 이다. 따라서 잔차 분산 통계량의 2차 Taylor 근사법에 의한 일반식은 아래와 같이 구할 수 있다.

$$\hat{E}[\sigma_e^2|\gamma] \cong \sigma_a^2 + \sigma_a^2 \left[\frac{p+q}{N} \right] = \sigma_a^2 \left[1 + \frac{1}{N} [p+q] \right] \quad (10)$$

여기서 $tr \sum_{j=0}^{\infty} h_j C_j \Sigma_{\hat{\gamma}}$ 과 $tr \sum_{j=0}^{\infty} c_j c_j^T \Sigma_{\hat{\gamma}}$ 는 Lee and Apley[15]의 식 (A13)과 식 (A18)로부터 각각 구할 수 있다. 두 식 모두 식 (A13)과 식 (A18)에서 $v=0$ 일 때의 경우로 각각의 식에 0을 대입하면 그 값은 0과 $(p+q)/N$ 가 된다.

즉, 잔차 분산 통계량 일반식의 최종 결과는 모델링 오차가 존재하지 않을 때의 $e_t(= a_t)$ 분산, ARMA 모델링 시 사용된 공정 데이터 개수 N개, 그리고 ARMA 모델의 차수합으로 구성되어 있음을 알 수 있다. 따라서 이 모델링 오차가 고려된 잔차 분산 통계량을 누적합 관리도의 모수에 반영하여 ARMA 모델링 오차에 강건한 잔차 기반 관리도를 다음 장에서 제안한다.

4. 강건 누적합 관리도의 설계

4.1 최적 모수의 결정

이 장에서는 제 3장에서 유도된 모델링 오차가 고려된 잔차 분산 통계량의 결과를 이용하여 최적 누적합 관리도의 설정 방법을 설명한다. ARMA 모델링 오차가 고려된 잔차 분산 통계량을 각각 K, H 와 결합하여 ARMA 모델링 오차에 강건하도록 보정하려면 식 (10)으로부터 다음의 관계를 이용한다.

$$K' = K \times \sqrt{1 + \frac{1}{N}[p+q]} \tag{11}$$

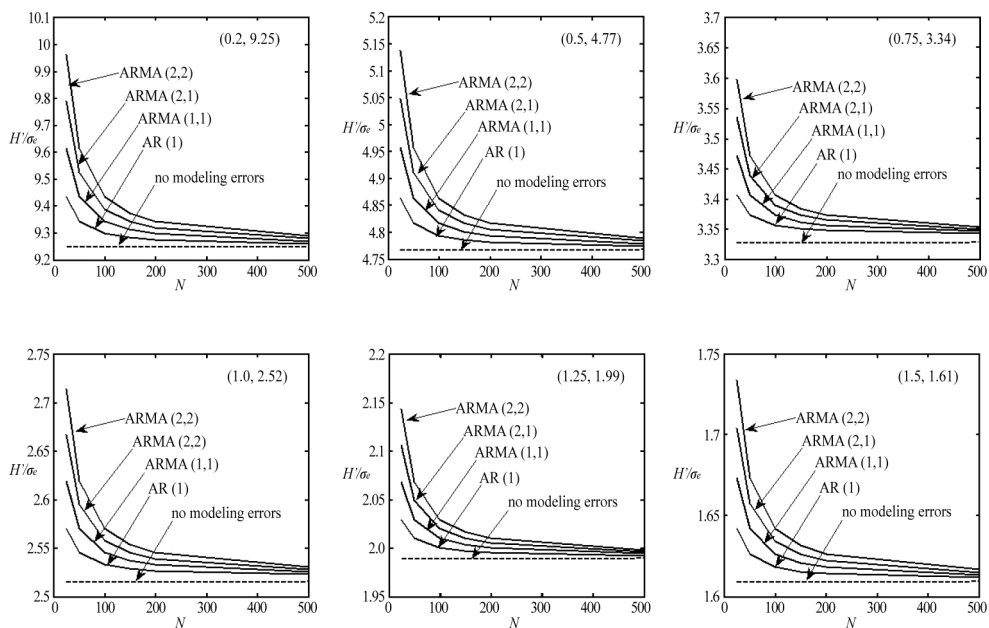
$$H' = H \times \sqrt{1 + \frac{1}{N}[p+q]} \tag{12}$$

즉, 모델링 오차 없을 때 적용하는 기존의 참고값 K , 결정구간 H 값에 모델링 오차가 고려된 잔차 분산 통계량의 양의 제곱근 값을 곱하면 ARMA 모델링 오차에 강건한 K', H' 값을 계산할 수 있다.

제 2장에서 간략히 설명한 바와 같이, 누적합 관리도는 공정 안정 상태의 평균 런 길이와 공정 불안정 상태의 평균 런 길이를 사용하여 관리도를 설정하고 성능을 평가한다. 먼저 제 1종 오차 대신 사용되는 공정 안정 상태의 평균 런 길이 중 가장 많이 사용하는 값 중 하나인 500을 이용하여 ARMA 모델링 오차에 강건한 누적합 관리도 설정 방법을 설명한다. 제안한 방법에서는 식 (11), 식 (12)를 이용하여 <표 1>의 $K/\sigma_e, H/\sigma_e$ 조합을 $K'/\sigma_e, H'/\sigma_e$ 조합으로 보정하여 사용 할 수 있다. 따라서 공정 안정 상태의 평균

런 길이가 500일 때, 제안한 방법에 의해 누적합 관리도를 설계하기 위한 다양한 조건의 $K'/\sigma_e, H'/\sigma_e$ 는 <표 2>와 같다.

ARMA 모델링 오차에 강건하도록 제안한 모수를 계산하기 위해 <표 2>에서 사용된 데이터 크기는 6가지(= 25, 50, 100, 150, 200, 500) 이고 고려된 ARMA 모델은 4가지(= AR(1), ARMA(1, 1), ARMA(2, 1), ARMA(2, 2))이다. 고려된 ARMA 모델은 자기상관 특성을 갖는 데이터가 모델링 될 때 생산 공정에서 가장 많은 빈도로 모델링 되는 모델들이다. 예를 들어, ARMA 모델링의 결과, 사용된 데이터의 크기가 25이고 ARMA(1, 1)인 경우는 <표 2>로부터 기존의 H/σ_e (= 9.96)값 보다 대략 4%(= (10.35-9.96)/9.96) 커진 H'/σ_e (= 10.35)를 사용하고 동일한 모델에서 데이터의 크기가 200으로 커지면 대략 결정구간의 크기가 0.5%(= (10.01-9.96)/9.96) 증가된 H'/σ_e (= 10.01)을 사용한다. 또한 시계열 모델링 결과가 ARMA(2, 2)일 때, 모델링에 사용된 데이터의 크기가 25이면 기존의 H/σ_e 값 보다 대략 8%(= (10.73-9.96)/9.96) 커진 H'/σ_e (= 10.73)를 사용하고 동일한 모델에서 데이터 크기가 200으로 증가하면 결정구간의 크기가 1%(= (10.06-9.96)/9.96) 증가된 H'/σ_e (= 10.06)을 사용한다. 모델링 결과에 따라 K'/σ_e 에 대해서도 <표 2>를 이용하여 K/σ_e 대신 사용할 수 있다. 물론 이장에서 사용한 ARMA 모델과 데이터 크기가 아니더라도 데이터 크기가 N 인 임의의 ARMA(p, q) 모델에 대해서 식 (11), 식 (12)를 이용하여 모든 $K/\sigma_e, H/\sigma_e$ 조합을 보정하여 사용 할 수 있다. <그림 1>은 공정 안정 상태의 평균 런 길이가 370일 때 <표 1>의 H/σ_e 에 대해서 <표 2>에서 사용한 ARMA 모델들에 대해 도식화 한 내용이다. 그림의 오른쪽 상단의 값은 모델링 오차가 존재하지 않을 경우의 $K/\sigma_e, H/\sigma_e$ 조합을 나타낸다.



<그림 1> 강건 누적합 관리도의 결정구간(In-Control ARL 370인 경우)

<표 2> 강건 누적합 관리도의 참고값과 결정구간(In-contrl ARL 500인 경우)

K/σ_e H/σ_e	N	0.2	0.5	0.75	1.0	1.25	1.5	2.0	2.5	3.0	
		9.96	5.07	3.54	2.67	2.11	1.71	1.11	0.59	0.09	
K'/σ_e H'/σ_e	AR(1)	25	0.204 10.16	0.510 5.17	0.765 3.61	1.020 2.72	1.275 2.15	1.530 1.74	2.040 1.13	2.550 0.60	3.059 0.092
		50	0.202 10.06	0.505 5.12	0.757 3.58	1.009 2.70	1.261 2.13	1.514 1.73	2.018 1.12	2.523 0.60	3.027 0.091
		100	0.201 10.01	0.503 5.10	0.754 3.56	1.005 2.68	1.256 2.12	1.508 1.72	2.010 1.12	2.513 0.59	3.015 0.09
		150	0.201 9.99	0.502 5.09	0.753 3.55	1.003 2.68	1.254 2.12	1.505 1.72	2.007 1.11	2.508 0.59	3.010 0.09
		200	0.201 9.98	0.501 5.08	0.752 3.55	1.003 2.68	1.253 2.12	1.504 1.71	2.005 1.11	2.506 0.59	3.008 0.09
		500	0.200 9.97	0.501 5.08	0.751 3.54	1.001 2.67	1.251 2.11	1.502 1.71	2.002 1.11	2.503 0.59	3.003 0.09
	ARMA (1, 1)	25	0.208 10.35	0.520 5.27	0.779 3.68	1.039 2.77	1.299 2.19	1.559 1.78	2.079 1.15	2.598 0.61	3.118 0.094
		50	0.204 10.16	0.509 5.17	0.764 3.61	1.018 2.72	1.273 2.15	1.527 1.74	2.036 1.13	2.545 0.60	3.054 0.092
		100	0.202 10.06	0.505 5.12	0.758 3.58	1.010 2.70	1.262 2.13	1.515 1.73	2.020 1.12	2.525 0.60	3.030 0.091
		150	0.201 10.03	0.503 5.10	0.755 3.56	1.007 2.69	1.258 2.12	1.510 1.72	2.013 1.12	2.517 0.59	3.020 0.091
		200	0.201 10.01	0.503 5.10	0.754 3.56	1.005 2.68	1.256 2.12	1.508 1.72	2.010 1.12	2.513 0.59	3.015 0.09
		500	0.200 9.98	0.501 5.08	0.752 3.55	1.002 2.68	1.253 2.11	1.503 1.71	2.004 1.11	2.505 0.59	3.006 0.09
	ARMA (2, 1)	25	0.212 10.54	0.529 5.37	0.794 3.75	1.058 2.83	1.323 2.23	1.588 1.81	2.117 1.17	2.646 0.62	3.175 0.10
		50	0.205 10.25	0.514 5.22	0.770 3.64	1.027 2.75	1.284 2.17	1.540 1.76	2.054 1.14	2.567 0.61	3.081 0.093
		100	0.203 10.11	0.507 5.15	0.761 3.59	1.015 2.71	1.269 2.14	1.522 1.74	2.030 1.13	2.537 0.60	3.045 0.091
		150	0.202 10.06	0.505 5.12	0.758 3.58	1.010 2.70	1.262 2.13	1.515 1.73	2.020 1.12	2.525 0.60	3.030 0.091
		200	0.202 10.03	0.504 5.11	0.756 3.57	1.008 2.69	1.259 2.13	1.511 1.72	2.015 1.12	2.519 0.59	3.022 0.091
		500	0.201 9.99	0.502 5.09	0.752 3.55	1.003 2.68	1.254 2.12	1.505 1.72	2.006 1.11	2.508 0.59	3.009 0.09
	ARMA (2, 2)	25	0.215 10.73	0.539 5.46	0.808 3.81	1.077 2.88	1.346 2.27	1.616 1.84	2.154 1.20	2.693 0.64	3.231 0.10
		50	0.207 10.35	0.518 5.27	0.777 3.68	1.036 2.77	1.295 2.19	1.554 1.78	2.072 1.15	2.589 0.61	3.107 0.094
		100	0.204 10.16	0.510 5.17	0.765 3.61	1.020 2.72	1.275 2.15	1.530 1.74	2.040 1.13	2.550 0.60	3.059 0.092
		150	0.203 10.09	0.507 5.14	0.760 3.59	1.013 2.71	1.267 2.14	1.520 1.73	2.027 1.12	2.533 0.60	3.040 0.091
		200	0.202 10.06	0.505 5.12	0.758 3.58	1.010 2.70	1.262 2.13	1.515 1.73	2.020 1.12	2.525 0.60	3.030 0.091
		500	0.201 10.00	0.502 5.09	0.753 3.55	1.004 2.68	1.255 2.12	1.506 1.72	2.008 1.11	2.510 0.59	3.012 0.09

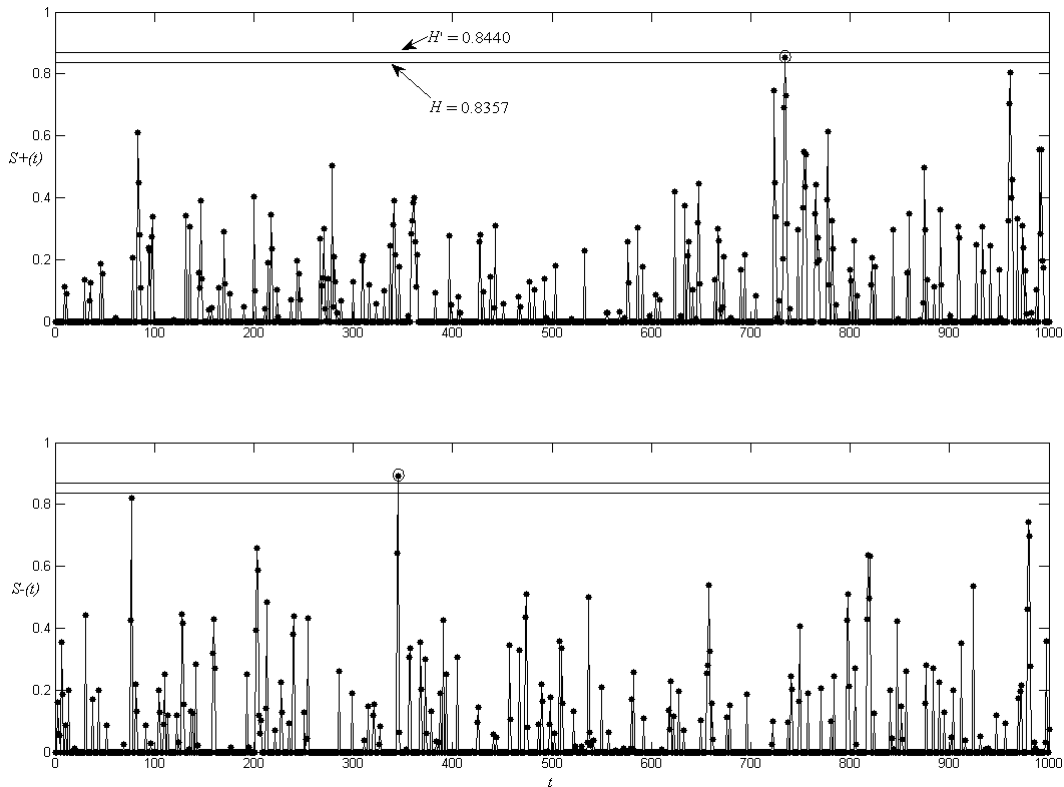
4.2 강건 누적합 관리도의 강건성 비교

이 장에서는 Box et al.[9]의 Series A 데이터를 이용하여 제안한 방법의 적용 절차를 설명한다. Box et al.[9]에 의하면 Series A 데이터는 ARMA(1, 1) 모델로 잘 표현이 되고, $\hat{\phi}_1$ 는 0.87, $\hat{\theta}_1$ 는 0.48, $\hat{\sigma}_a$ 는 0.313으로 추정되었다. 이때 사용된 데이터 개수는 100개로 가정한다. 제안한 방법에 따라 예제에 대한 누적합 관리도를 설정하면 <표 2>에서 ARMA(1, 1) 모델의 해당하는 임의의 K/α_c 크기 별 H/α_c 의 조합을 찾을 수 있다.

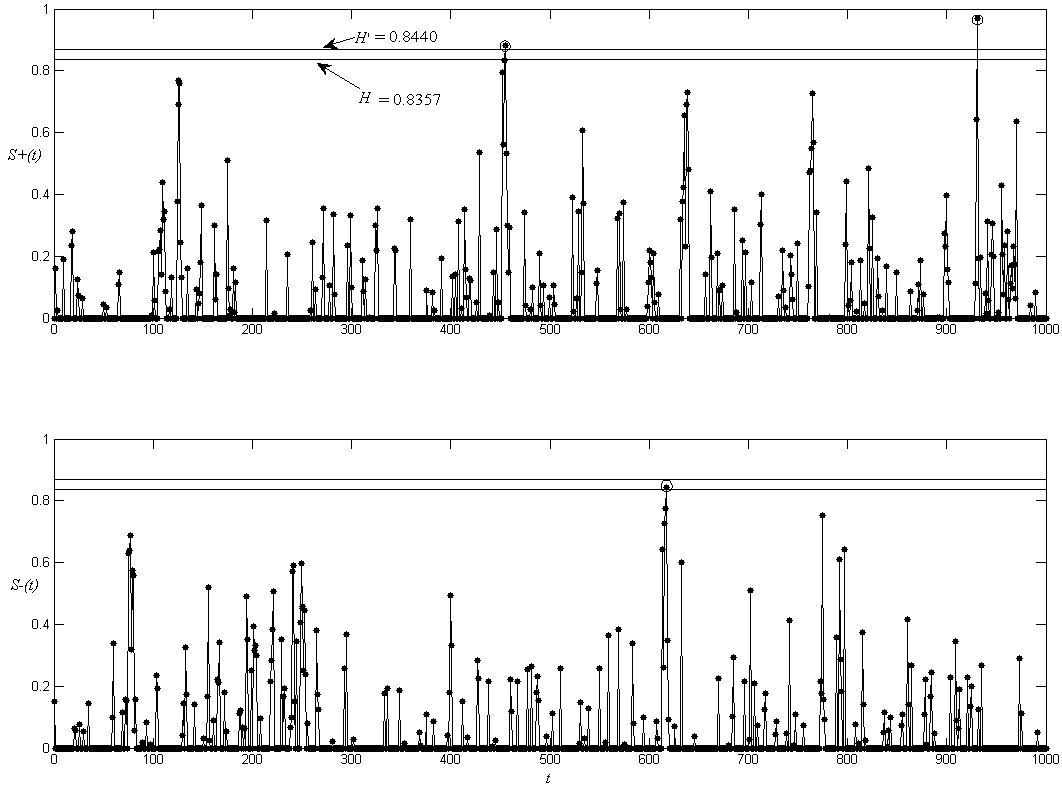
<그림 2>는 Series A 데이터가 따르는 ARMA(1, 1) 모델로부터 $\phi_1, \theta_1, \sigma_a$ 이 각각 $\hat{\phi}_1, \hat{\theta}_1, \hat{\sigma}_a$ 이 같다는 가정 하에, 즉 모델링 오차가 없다는 가정 하에 Series A로 시뮬레이션 된 데이터를 기존의 누적합 관리도와 제안한 누적합 관리도에 적용한 결과이다. 기존 누적합 관리도의 모수 조합은 (1.0, 2.67)이 이용되었고, 이 경우에 대한 강건 누적합 관리도의 모수 조합은 (1.01, 2.70)이 이용되었다. 모델링 오차가 존재하지 않기 때문에 기존의 설계방법에 의한 관리한계가 제공하는 공정 안정 상태의 평균 런 길이가 약 500에 가까울 것이라는 예측이 가능하다. 그림에서 공정변화가 존재하지 않았기 때문에 관리한계를 벗어나는 관측치는 모두 오경보를 의미한다. 기존의 누적합 관리도에 의한 오

경보는 2개(345번째, 734번째 관측치)이고, 제안한 방법의 누적합 관리도에 의한 오경보는 1개(345번째 관측치)이다.

<그림 3>은 Series A 데이터가 따르는 ARMA(1, 1) 모델로부터 기존의 관리도와 제안한 방법에 의한 관리도간의 모델링 오차에 대한 강건성을 평가하기 위한 실험결과이다. θ_1, σ_a 는 각각 $\hat{\theta}_1, \hat{\sigma}_a$ 로 같지만 ϕ_1 을 0.9로 가정하여 $\hat{\phi}_1$ 가 과소추정 되었다는 조건 하에 시뮬레이션 된 Series A 데이터를 기존의 누적합 관리도와 제안한 누적합 관리도에 적용한 결과이다. 이 그림에서도 공정변화는 존재하지 않기 때문에 관리한계를 벗어나는 관측치는 모두 오경보를 의미한다. 기존의 누적합 관리도에 의한 오경보는 3개(455번째, 617번째, 931번째 관측치)이고, 제안한 방법의 누적합 관리도에 의한 오경보는 2개(455번째, 931번째 관측치)이다. 첫 번째 실험과는 달리 모델링 오차가 존재하는 경우에는 제안한 방법에 의한 누적합 관리도가 제공하는 공정 안정상태의 평균 런 길이가 500에 가까움을 확인할 수 있다. 반면 기존의 누적합 관리가 제공하는 공정 안정 상태의 평균 런 길이는 333으로 의도된 평균 런 길이 보다 약 33%가 줄어들었음을 확인할 수 있다. 즉, 모델링 오차가 존재하는 경우에는 제안한 방법에 의한 누적합 관리도 설계방법이 의도된 크기의 공정 안정상태의 평균 런 길이를 보장할 수 있다고 결론 내릴 수 있다.



<그림 2> Series A 데이터에 대한 양방향 강건 누적합 관리도 적용(모델링 오차가 존재하지 않는 경우)



<그림 3> Series A 데이터에 대한 양방향 강건 누적합 관리도 적용(모델링 오차가 존재하는 경우)

4.3 강건 누적합 관리도의 성능 비교

ARMA 모델링 오차에 강건하기 위하여 관리한계의 폭이 늘어나면, 즉 결정구간의 값이 커지면 공정 불안정상태의 평균 런 길이에서는 단점으로 작용할 수 있다. 특히 첫 번째 모의실험 비교결과에서 보는 바와 같이 만일 모델링 오차가 무시될 만큼 실제 모수와 비슷하게 모델링이 된다면 강건 설계 방법에 의한 관리한계에서는 이런 단점이 두드러질 수 있다. 반면에 두 번째 모의실험에서 본 것처럼 어느 정도 모델링 오차가 존재하게 되면 강건 누적합 관리도는 최소한 관리도 설정 시 요구됐던 공정 안정상태의 평균 런 길이는 보장할 수 있는 장점이 있다.

<표 3>은 Monte carlo 시뮬레이션을 이용하여 기존의 누적합 관리도와 이 논문에서 제안한 강건 누적합 관리도의 다양한 공정변화의 크기에 따른 검출능력을 비교한 결과이다. 앞에서 사용된 Series A 데이터가 다시 사용되었고 실제 모수와 모델링 결과가 같다는 가정 하에 비교하였다. 모델링 결과는 ARMA(1, 1), $\hat{\phi}_1 = 0.87$, $\hat{\theta}_1 = 0.48$ 로 동일하고 실험의 편의를 위해 $\hat{\sigma}_a = 1$, $N = 100$ 으로 가정한다. 제안한 방법에 의해 강건 누적합 관리도가 설정된 내용이 굵게 표기된 부분이고 이에 대응하는 기존 누적합 관리도의 설정 내용은 바로 위 칸에 각각 표기되었다. 표의 모든 안정/

불안정 상태의 평균 런 길이(In-control/Out-of-control ARL)는 Monte carlo 시뮬레이션에 의해 계산되었다. 각각의 시뮬레이션에 대해 10,000번의 실험 회수가 반복 사용되었다.

공정 안정 상태의 평균 런 길이 관점에서 보면 이 논문에서 제안한 방법이 제공하는 누적합 관리도의 평균 런 길이가 기존의 누적합 관리도의 평균 런 길이 보다 약 10~20% 정도 크을 알 수 있다. 이는 모델링 오차가 존재할 경우에 대비한 것으로 <표 3>의 실험이 모델링 오차가 존재하지 않는다는 가정 하에 진행되었으므로 예측 가능한 결과이다. 일반적으로 공정 안정 상태의 평균 런 길이는 크면 클수록 오경보가 발생할 가능성이 줄어든다는 의미로 해석한다.

반면에 공정변동이 존재할 때 런 길이는 짧을수록 좋다. 공정 안정 상태의 평균 런 길이 관점에서 비교해 보면 전반적으로 공정변동의 크기가 비교적 작은 경우(= $1 \sim 2\sigma_a$)에는 기존 방법에 의한 누적합 관리도가 2~15% 정도 빠르게 감지하고 있음을 확인할 수 있다. 물론 공정변동의 크기가 중간 크기(= $3\sigma_a$)이거나 비교적 큰 경우(= $4 \sim 5\sigma_a$)에도 기존 방법에 의한 누적합 관리도가 모든 경우에 제안한 방법에 의한 누적합 관리도 보다 빠르게 감지하고 있지만 그 차이는 공정변동의 크기가 커짐에 따라 상당히 작아지고 있음을 볼 수 있다.

<표 3> 강건설계 방법의 공정변동에 대한 검출능력 비교(단위 : 평균 런 길이(ARL))

관리도	(참고값, 결정구간)	공정평균 변동크기(σ_a 단위)					
		0	1	2	3	4	5
CUSUM(기존 방법)	(0.2 9.96)	500	84.4	25.1	11.7	6.51	4.12
CUSUM(제안 방법)	(0.202 10.06)	534	86.9	25.5	11.9	6.64	4.20
CUSUM(기존 방법)	(0.5 5.07)	499	141	27.8	6.92	2.80	1.83
CUSUM(제안 방법)	(0.505 5.12)	549	151	29.2	7.16	2.86	1.86
CUSUM(기존 방법)	(0.75 3.54)	501	195	40.2	7.08	2.03	1.26
CUSUM(제안 방법)	(0.758 3.58)	612	230	46.4	7.91	2.14	1.29
CUSUM(기존 방법)	(1.0 2.67)	502	247	59.4	9.22	1.86	1.10
CUSUM(제안 방법)	(1.010 2.70)	549	273	65.1	9.97	1.93	1.11
CUSUM(기존 방법)	(1.25 2.11)	500	287	83.8	13.3	1.86	1.06
CUSUM(제안 방법)	(1.262 2.13)	554	317	92.4	14.7	1.96	1.07
CUSUM(기존 방법)	(1.5 1.71)	500	315	109	19.6	2.44	1.06
CUSUM(제안 방법)	(1.515 1.73)	552	350	122	22.0	2.65	1.06
CUSUM(기존 방법)	(2.0 1.11)	501	353	152	35.9	4.22	1.16
CUSUM(제안 방법)	(2.020 1.12)	550	392	168	40.5	4.85	1.18
CUSUM(기존 방법)	(2.5 0.59)	501	367	164	46.2	7.06	1.29
CUSUM(제안 방법)	(2.525 0.60)	552	405	183	51.9	7.98	1.34
CUSUM(기존 방법)	(3.0 0.09)	502	359	166	50.1	8.15	1.39
CUSUM(제안 방법)	(3.030 0.091)	552	393	184	56.5	9.17	1.46

5. 결 론

이 논문에서 공정의 미세변동에 대한 감지능력이 좋은 누적합 관리도에 대해 모델링 오차가 존재하는 자기상관 데이터를 적용하기 위한 잔차 기반의 강건 설계 방법을 제안하였다. ARMA 모델링 오차의 강건성을 확보하기 위하여 먼저 ARMA 모델링 오차가 고려된 잔차의 분산 통계량을 Taylor 2차 근사식을 이용하여 일반식의 형태로 유도하였고, 이는 임의의 ARMA(p, q) 모델에 대해 적용 가능하도록 하였다. 이 유도된 결과를 이용하여 수학적 변형이 쉽지 않은 누적합 관리도의 통계량을 직접 수정하기보다 누적합 관리도의 모수를 수정하여 설계하는 방법을 제안하였다.

제안한 강건 설계 방법은 공정 안정 상태의 평균 런 길이를 우선적으로 유지하기 위해 고안되었다. 이러한 특성으로 인해 제안한 방법에 의한 누적합 관리도가 공정변동에 대한 감지 능력이 다소 떨어질 수 있다. 하지만 제 4장의 실험 결과에서 확인한 바와 같이 제안한 방법은 작은 크기의 공정변동에 대해서는 기존 방법보다 감지가 다소 늦는 결과를 보였지만, 중간 크기나 큰 크기의 공정변동에서는 기존 방법과 성능상의 차이가 크게 존재하지 않았다. 이러한 장점은 제안한 방법이 관리도 설계 시 의도된 공정 안정 상태의 평균 런 길이를 유지하면서

도 일정 정도 감지 성능을 보장함을 보여준다.

일반적으로 실제 공정에서 공정변동을 감지하지 못하여 발생하는 피해도 중요하지만, 오경보에 의한 피해 또한 매우 중요하다. 특히 잦은 오경보로 인해 작업자들에게 통계적 공정관리에 대한 불신과 경보에 대한 불감증이 만연하게 되면 이는 공정관리와 품질 마인드 재교육 등 조직 품질문화 전반에 대한 재정립을 필요로 한다. 이는 기업 입장에서 보면 불필요한 지출이므로 손실로 작용하는 것이다. 이 논문에서 제안한 방법이 제공하는 누적합 관리도는 이러한 모델링 오차에 의한 잦은 오경보의 위험을 최소화함으로써 기업의 손실을 줄일 수 있기 때문에 매우 중요한 의미를 갖는다. 따라서 이 논문에서 제안한 방법이 위에서 논의된 적용상의 장점을 바탕으로 자기상관이 존재하는 다양한 생산 공정에서 효과적으로 사용될 수 있을 것으로 기대한다.

이 논문의 서론에서 언급한 바와 같이 EWMA 관리도는 누적합 관리도와 함께 공정의 미세 변동 감지에 있어 효과적인 방법으로 널리 사용된다. 이 논문에서는 논의되지 못하였지만 잔차 기반의 EWMA 강건 설계 방법과 이 논문의 누적합 강건 설계 방법간의 성능 비교는 후속 연구 주제로 상당한 의미가 있을 것으로 판단된다. 현재까지 두 방법론간의 비교는 모델링 오차가 고려되지 않은 상황에서 이뤄졌거나, 모델링 오차가 고려되더라도

각각의 방법론이 모델링 오차에 얼마나 강건한가에만 중점을 두어 논의되었다. 따라서 강건 방법론간의 성능비교 연구결과는 많지 않기 때문에 이에 대한 후속 연구가 필요해 보인다.

참고문헌

- [1] 이정형, 전태윤, 조신섭; “자기상관을 갖는 공정의 로버스트 누적합관리도”, 품질경영학회지, 27(4) : 123-142, 1999.
- [2] 이재준, 이종선; “자기상관 공정에 대한 누적합관리도에서 설계모수 값의 결정”, 품질경영학회지, 36(4) : 87-92, 2008.
- [3] Adams, B. M. and Tseng, I. T.; “Robustness of Forecast-Based Monitoring Schemes,” *Journal of Quality Technology*, 30(4) : 328-339, 1998.
- [4] Alwan, L. C. and Roberts, H. V.; “Time-Series Modeling for Statistical Process Control,” *Journal of Business and Economic Statistics*, 6(1) : 87-95, 1988.
- [5] Apley, D. W.; “Time Series Control Charts in the Presence of Model Uncertainty,” *Journal of Manufacturing Science and Engineering*, 124(4) : 891-898, 2002.
- [6] Apley, D. W. and Lee, H. C.; “Design of Exponentially Weighted Moving Average Control Charts for Autocorrelated Processes With Model Uncertainty,” *Technometrics*, 45(3) : 187-198, 2003.
- [7] Apley, D. W. and Lee, H. C.; “Robustness Comparison of Exponentially Weighted Moving-Average Charts on Autocorrelated Data and on Residuals,” *Journal of Quality Technology*, 40(4) : 428-447, 2008.
- [8] Apley, D. W. and Shi, J.; “The GLRT for Statistical Process Control of Autocorrelated Processes,” *IIE Transactions*, 31(12) : 1123-1134, 1999.
- [9] Box, G., Jenkins, G., and Reinsel, G.; *Time Series Analysis, Forecasting and Control*, 4th Edition, John Wiley and Sons, Inc., New Jersey, 2008.
- [10] Casella, G. and Berger, R. L.; *Statistical Inference*, 1st Edition, Duxbury, Belmont, California, 1990.
- [11] English, J. R., Lee, S., Martin, T. W., and Tilmon, C.; “Detecting changes in autoregressive processes with \bar{X} and EWMA charts,” *IIE Transactions*, 32(12) : 1103-1113, 2000.
- [12] Johnson, R. A. and Bagshaw, M.; “The Effect of Serial Correlation on the Performance of CUSUM tests,” *Technometrics*, 16(1) : 103-112, 1974.
- [13] Montgomery, D. C.; *A Modern Introduction to Statistical Quality Control*, 6th Edition, John Wiley and Sons, Inc., New Jersey, 2008.
- [14] Lee, H. C.; “Robust Design of Control Charts for Autocorrelated Processes with Model Uncertainty,” Ph.D. Dissertation, Texas A&M University, College Station, Texas, U.S.A., 2004.
- [15] Lee, H. C. and Apley, D. W.; “Improved Design of Robust Exponentially Weighted Moving Average Control Charts for Autocorrelated Processes,” *Quality and Reliability Engineering International*, 27(3) : 337-352, 2011.
- [16] Montgomery, D. C. and Mastrangelo, C. M.; “Some Statistical Process Control Methods for Autocorrelated Data,” *Journal of Quality Technology*, 23(3) : 179-193, 1991.
- [17] Schmid, W.; “On EWMA Charts for Time Series,” *Frontiers of Statistical Quality Control*, 5 : 115-137, 1997.
- [18] Testik, M. C.; “Model Inadequacy and Residuals Control Charts for Autocorrelated Processes,” *Quality and Reliability Engineering International*, 21(2) : 115-130, 2005.
- [19] Zhang, N. F.; “A Statistical Control Chart for Stationary Process Data,” *Technometrics*, 40(1) : 24-38, 1998.