

Multivariate Control Charts for Autocorrelated Process¹⁾

Gyo-Young Cho²⁾ · Mi-Ra Park³⁾

Abstract

In this paper, we propose Shewhart control chart and EWMA control chart using the autocorrelated data which are common in chemical and process industries and lead to increase the number of false alarms when conventional control charts are applied. The effect of autocorrelated data is modeled as a autoregressive process, and canonical analysis is used to reduce the dimensionality of the data set and find the canonical variables that explain as much of the data variation as possible. Charting statistics are constructed based on the residual vectors from the canonical variables which are uncorrelated over time, and the control charts for these statistics can attenuate the autocorrelation in the process data. The charting procedures are illustrated with a numerical example and simulation is conducted to investigate the performances of the proposed control charts.

1. 머리말

기술사회에서 통계적 공정관리(Statistical Process Control : SPC)는 공정의 평균이나 변동의 변화를 발견하기 위해 차트기법을 적용하여 샘플링 디자인, 파레토 분석, 실험계획법, 다변량 분석, 신뢰성 등 폭넓은 분야에서 활용되고 있다. SPC는 공학적인 고유기술과 접목되어 품질특성치의 변동 상태를 파악하여 이러한 변동을 유발하는 요인을 발견하고 이에 대한 적절한 조치를 통해 그 변동을 줄임으로서 품질특성치를 바람직한 수준으로 유지하는데 도움을 준다. 최근에는 적극적인 불량품 예방활동을 위

1) This paper was supported by Kyungpook National University Research Fund, 2001.

2) Associate Professor, Department of Statistics, Kyungpook National University, Daegu, 702-701, Korea
E-mail : gycho@knu.ac.kr

3) Part-time Instructor, Department of Statistics, Kyungpook National University, Daegu, 702-701, Korea

하여 실시간 SPC를 통해 가능한 빨리 공정의 이상원인(assignable cause)을 탐지하여 수정 조치를 취하려는 노력이 진행되고 있다. 이와 같은 목적으로 사용되는 가장 대표적인 기법이 관리도이다. 관리도는 연속생산공정에서 제품의 품질변화를 신속히 탐지하는 통계적 절차라 할 수 있다.

대표적인 관리도는 슈하르트(Shewhart) 관리도, 누적합(cumulative-sum : CUSUM) 관리도, 지수가중이동평균(exponentially weighted moving average: EWMA) 관리도 등이 있는데 이 관리도들은 일반적으로 공정으로부터 채취되는 표본들이 서로 독립이고 동일한 분포를 갖는다고 가정한다. 대부분의 관리도들은 정규성과 같은 기본 가정에 로버스트(robust)하지만 상관성이 있는 측정치에 대해서는 매우 민감하다. 특히 공정의 특성상 측정치들이 과거의 측정치에 의존하는 자기상관(autocorrelation)관계가 있는 경우 이러한 관리도에 그대로 적용하면 실제로 이상원인이 존재하지 않더라도 이상원인이 발생한 것처럼 거짓경보(false alarm)를 주는 경우가 많아 공정을 효율적으로 관리하기 어렵게 된다. 이에 대한 대안으로 시계열 모형을 이용하여 자기상관효과를 파악하고, 측정치와 시계열 모형에 의한 예측치의 차이인 잔차를 이용하여 관리도를 운영하는 기법들이 연구되었다. Alwan과 Roberts(1988)는 자기상관이 있는 연속생산공정을 Shewhart 관리도로 관리할 경우 거짓경보가 일어날 확률이 증가함을 보였고, 시계열 모형을 통하여 얻어진 예측오차를 이용하는 관리도를 운용할 것을 제안하였다.

또한 Harris와 Ross(1991)는 CUSUM 관리도와 EWMA 관리도에서의 자기상관효과를 분석한 후, Kalman filter를 이용하여 자기상관효과를 제거한 뒤에 관리도를 사용할 것을 제안하였다. Mastrangelo와 Montgomery(1995)는 시계열 예측모형으로 EWMA 통계량을 이용하는 관리도에서 공정의 이상을 빨리 탐지할 수 있도록 추적지표검정(tracking signal tests)을 보조수단으로 사용하는 EWMA 관리도에 대해 연구하였다. 그밖에 Montgomery(1992)는 자기상관자료에 대한 공정관리 기법에 대한 연구결과들을 정리하였으며, Lin과 Adams(1996)는 마코프 연쇄(Markov chain)를 이용하여 시계열 예측모형을 사용하는 기존의 관리도들의 수행도를 분석하였다.

생산공정에서 공정특성을 나타내는 품질특성치가 여러 개인 경우에는 각 품질특성치에 적절한 관리도를 독립적으로 사용하는 것이 일반적이다. 만약 품질특성치들이 서로 상관관계를 가진다면, 독립적인 관리도의 시행은 잘못된 판단을 내릴 가능성이 많아진다. 이와 같이 서로 상관관계를 가진 여러 개의 품질특성치를 관리하고자 할 때 품질특성치들간의 상관관계를 고려한 하나의 통계량을 이용하여 관리하는 다변량 관리도(multivariate control chart)를 사용한다. Hotelling(1947)에 의해 제안된 이후 최근까지 많은 연구가 진행되어 왔고 Alt(1985), Jackson(1985), Lowry와 Montgomery(1995)는 다변량 관리도에 대한 기존 연구들을 정리하였다. 그러나 다변량 관리도에 대한 기존의 연구들은 측정치들 사이의 자기상관관계를 제대로 고려하지 못하고 있어 화학공정과 같은 연속생산공정을 관리하는데 사용하기에는 적합하지 않다.

이 논문에서는 측정치들간에 자기상관이 존재하는 연속생산공정을 관리하기 위하여 시계열 모형을 세우고, 정준상관분석을 통해 자기상관효과가 제거된 잔차를 이용하는 새로운 다변량 관리도를 제안한다. 제 2절에서는 연속생산공정에서 얻어지는 측정치의 자기상관관계를 자기회귀과정(autoregressive process)으로 모형화한다. 제 3절에서는 측정치의 분산공분산행렬에서 자기상관효과를 분리해낼 수 있는 정준상관분석

(canonical analysis)에 대해서 설명한다. 제 4절에서는 정준상관분석을 이용하여 자기상관효과가 제거된 잔차를 이용하는 다변량 관리도를 제안하고 수치 예를 들어 설명한다. 제 5절에서는 모의실험을 통하여 자기상관공정을 위한 Shewhart 관리도와 EWMA 관리도를 평균런길이(average run length : ARL) 측면에서 기존의 관리도와 비교·평가한다.

2. 자기회귀모형을 이용한 연속생산공정

서로 상관관계가 있는 $k(k \geq 2)$ 개의 품질특성들을 관리해야 하는 다변량 관리도에서는 모집단이 다변량 정규분포를 따른다는 가정하에서 가설 $\mu = \mu_0$ 를 검정하는 t -통계량을 다변량으로 확장한 Hotelling의 T^2 통계량을 이용한다.

$$T^2 = (\mathbf{X}_t - \boldsymbol{\mu}_0)^T \boldsymbol{\Sigma}^{-1} (\mathbf{X}_t - \boldsymbol{\mu}_0) \quad (2.1)$$

여기서 \mathbf{X}_t : $(k \times 1)$ 확률벡터, $\boldsymbol{\mu}_0$: $(k \times 1)$ 평균벡터, $\boldsymbol{\Sigma}$: $(k \times k)$ 분산공분산행렬.

Hotelling(1931)에 의해 유도된 T^2 의 분포는 변수의 수 (k)의 함수로서 \mathbf{X}_t 가 평균이 $\boldsymbol{\mu}_0$, 분산공분산행렬이 $\boldsymbol{\Sigma}$ 인 다변량 정규분포를 따르고 표본의 크기가 충분히 크다면 T^2 통계량은 자유도가 k 인 χ^2 분포를 따르게 된다. T^2 은 제공된 양(quantity)이므로 유의수준을 α 로 하는 관리상한선 ($UCL = \chi^2_{\alpha}(k)$)만 존재한다. 여기서 $\chi^2_{\alpha}(k)$ 는 자유도가 k 인 χ^2 분포의 $(1-\alpha)$ 백분위수이다.

식 (2.1)과 같은 기존의 다변량 관리도는 시점 t 에서의 측정치 \mathbf{X}_t 가

$$\mathbf{X}_t = \boldsymbol{\mu}_0 + \mathbf{a}_t \quad (2.2)$$

와 같이 측정치간에 서로 독립이고 동일한 분포를 갖는다는 가정하에 사용된다. 여기서 $\{\mathbf{a}_t\}$ 는 백색잡음과정(white noise process)으로, \mathbf{a}_t 는 평균이 $\mathbf{0}$ 이고, $(k \times k)$ 분산공분산행렬 $\boldsymbol{\Sigma}$ 을 갖는 k 변량 정규분포를 따른다. 그러나 화학공정과 같은 연속생산공정에서는 연속된 측정치들간의 자기상관이 존재하므로, 시점 t 에서의 측정치 \mathbf{X}_t 는 시계열로 모형화 될 수 있다. 본 논문에서는 Box-Jenkins의 ARIMA(autoregressive integrated moving average) 과정 중 연속생산공정 데이터를 잘 설명한다고 알려진 p 차의 자기회귀과정 (AR(p))을 고려하여 연속생산공정의 자기상관관계를 모형화한다.

자기회귀과정은 시점 t 에서의 $(k \times 1)$ 시계열 데이터 벡터 \mathbf{X}_t 가 p 개의 과거 데이터 $\mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \dots, \mathbf{X}_{t-p}$ 와 이들로 설명되지 않는 부분인 오차항 \mathbf{a}_t 들의 가중선형결합으로 표현되는 모형으로, p 차의 자기회귀과정 AR(p)는

$$\mathbf{X}_t = \phi_1 \mathbf{X}_{t-1} + \phi_2 \mathbf{X}_{t-2} + \dots + \phi_p \mathbf{X}_{t-p} + \mathbf{a}_t \quad (2.3)$$

이다. 여기서 ϕ_i 는 $(k \times k)$ 자기회귀계수행렬로 \mathbf{X}_t 와 \mathbf{X}_{t-i} 간의 상관관계를 나타내며,

$\{\mathbf{a}_t\}$ 는 분산공분산행렬이 Σ_w 인 k 변량 정규분포를 따르는 백색잡음과정이다. 그리고 \mathbf{a}_t 와 $\mathbf{X}_{t-1}, \mathbf{X}_{t-2}, \dots$ 와는 독립이라고 가정한다. 이 모형은 시계열 데이터간의 관계가 회귀식 형태로 표현되어 직관적으로 이해하기 쉽다는 장점이 있다. 자기회귀과정 중 실제 응용성이나 수식적 편의성 때문에 차수가 1인 AR(1) 과정이 많이 이용되는데, 시계열 데이터의 평균벡터가 μ 이고 차수가 1인 AR(1) 과정은

$$\mathbf{X}_t = \mu + \phi(\mathbf{X}_{t-1} - \mu) + \mathbf{a}_t \quad (2.4)$$

으로 표현된다.

$(k \times 1)$ 확률벡터 \mathbf{x}_t 가 식(2.3)의 AR(p)를 따른다면, 시계열 데이터의 자기상관효과를 알아보기 위해 \mathbf{x}_t 의 시차(time lag)가 j 인 자기공분산행렬을 구하면

$$\begin{aligned} \Gamma_j(\mathbf{X}) &= E(\mathbf{X}_{t-j} \mathbf{X}_t^T) \\ &= E\left(\mathbf{X}_{t-j} \left[\sum_{i=1}^p \phi_i \mathbf{X}_{t-i} + \mathbf{a}_t \right]^T\right) \\ &= \sum_{i=1}^p E(\mathbf{X}_{t-j} \mathbf{X}_{t-i}^T) \phi_i^T \\ &= \sum_{i=1}^p \Gamma_{j-i}(\mathbf{X}) \phi_i^T \quad j = 1, 2, \dots, p \end{aligned} \quad (2.5)$$

가 되고, \mathbf{x}_t 의 분산공분산행렬은

$$\begin{aligned} \Gamma_0(\mathbf{X}) &= E(\mathbf{X}_t \mathbf{X}_t^T) \\ &= E\left(\left[\sum_{i=1}^p \phi_i \mathbf{X}_{t-i} + \mathbf{a}_t \right] \mathbf{X}_t^T\right) \\ &= \sum_{i=1}^p \phi_i E(\mathbf{X}_{t-i} \mathbf{X}_t^T) + E(\mathbf{a}_t \mathbf{X}_t^T) \\ &= \sum_{i=1}^p \phi_i \Gamma_i(\mathbf{X}) + \Sigma_w \\ &= \Sigma_c + \Sigma_w \end{aligned} \quad (2.6)$$

가 된다. 여기서 \mathbf{X}_t 의 분산공분산행렬은 자기상관효과에 대한 부분 Σ_c 와 백색잡음에 의한 부분 Σ_w 로 나누어진다. 따라서 추정치 \mathbf{x}_t 에서 자기상관으로 인해 나타나는 효과 Σ_c 를 제거하면 효율적인 공정관리를 이룰 수 있을 것이다.

3. 자기상관자료에 대한 정준상관분석

연속된 추정치들간에 자기상관이 있는 공정을 관리하기 위해서는 Alwan과 Roberts(1988)가 제안한 것처럼 공정을 적절히 묘사하는 시계열 모형을 세우고, 이 시계열 모형으로부터 구한 추정치와 실제 추정치간의 차이인 잔차들을 이용할 수 있다. 잔차들은 추정치에 나타나는 자기상관효과를 시계열의 추정치로 제거한 것으로, 서로 독립이 되어 기존의 관리도에 그대로 적용할 수 있을 것이다. 그러나 화학공정과 같은 연속생산공정에서는 품질특성의 개수가 매우 많으므로, 시계열 모형에 의한 잔차를 이용하여 관리도를 운영하더라도 공정의 분석이 쉽지 않다. 따라서 본 논문에서는

다수의 품질특성 중에서 공정에 특히 많은 영향을 미치는 일부의 품질특성만을 선택하여 공정을 관리하는 방법을 이용하였다. 다수의 품질특성으로부터의 자기상관효과를 제거하면서 공정의 산포에 많은 영향을 미치는 품질특성을 선택하기 위해 다음과 같은 정준상관분석을 이용한다(Box와 Tiao(1977)).

공정변수 \mathbf{x}_t 의 선형조합

$$y_t = \mathbf{m}^T \mathbf{X}_t \quad (3.1)$$

을 고려하자. 여기서 \mathbf{m} 은 선형계수를 나타내는 $(k \times 1)$ 벡터이다. \mathbf{x}_t 가 $AR(p)$ 를 따르는 경우 y_t 의 분산은 식(2.6)으로부터

$$\mathbf{m}^T \Gamma_0 \mathbf{m} = \mathbf{m}^T \Sigma_c \mathbf{m} + \mathbf{m}^T \Sigma_w \mathbf{m} \quad (3.2)$$

가 되어 자기상관효과에 대한 부분 ($\mathbf{m}^T \Sigma_c \mathbf{m}$)과 백색잡음 ($\mathbf{m}^T \Sigma_w \mathbf{m}$)으로 나누어진다. 여기서 백색잡음은 공정에 내재된 산포를 나타내므로 공정의 산포에 가장 많은 영향을 미치는 변수의 선택은 자기상관효과에 대한 부분, 즉 y_t 의 분산에 대한 자기상관효과의 비

$$\lambda = \{ \mathbf{m}^T \Sigma_c \mathbf{m} \} / \{ \mathbf{m}^T \Gamma_0 \mathbf{m} \} \quad (3.3)$$

가 큰 순서대로 선형 변환된 변수를 취하면 된다. 즉, $\lambda_1 > \lambda_2 > \dots > \lambda_k$ 에 해당하는 $\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_k$ 를 구하여 $\mathbf{M} = (\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_k)^T$ 라 하면, 선형 변환된 공정변수

$$\mathbf{Y}_t = \mathbf{M} \mathbf{X}_t \quad (3.4)$$

를 얻게 된다. 여기서 λ_i 와 \mathbf{m}_i 는 각각 행렬의 고유값과 고유벡터가 되고, \mathbf{m}_i 들은 서로 독립이다. \mathbf{Y}_t 의 분산공분산행렬을 구하면

$$\mathbf{M} \Gamma_0 \mathbf{M}^T = \mathbf{M} \Sigma_c \mathbf{M}^T + \mathbf{M} \Sigma_w \mathbf{M}^T \Gamma_0^{-1} \Sigma_c \quad (3.5)$$

가 된다. Λ 를 $(\lambda_1, \lambda_2, \dots, \lambda_k)$ 로 구성된 대각행렬이라 할 때, $\Gamma_0^{-1} \Sigma_c \mathbf{M}^T = \mathbf{M}^T \Lambda$ 와

$\mathbf{M} \mathbf{M}^T = \mathbf{M}^T \mathbf{M} = \mathbf{I}$ 에 의해

$$\mathbf{M} \Gamma_0^{-1} \Sigma_c \mathbf{M}^T = \Lambda, \quad \mathbf{M} \Gamma_0^{-1} \Sigma_w \mathbf{M}^T = \mathbf{I} - \Lambda \quad (3.6)$$

이므로, $i \neq j$ 에 대하여 $\mathbf{m}_i^T \Sigma_w \mathbf{m}_j = \mathbf{m}_i^T \Sigma_c \mathbf{m}_j = 0$ 이 된다. 따라서 $\mathbf{M} \Sigma_c \mathbf{M}^T$ 와 $\mathbf{M} \Sigma_w \mathbf{M}^T$ 는 대각행렬이 된다. 즉, \mathbf{Y}_t 는 자기상관효과를 가장 많이 설명하는 변수순으로 순서화 되어 있고, 각 공정변수들간에는 서로 독립이다.

4. 잔차를 이용한 다변량 관리도

일반적으로 연속생산공정에서 얻어진 많은 공정변수들 중 일부는 다른 공정변수들에 의해 설명되어 질 수 있고, 서로 유사한 자기상관효과를 갖는다. 예를 들어 서로 다른 위치에서 온도를 시간에 따라 측정하는 경우 측정치간에는 유사한 자기상관효과를 갖게 된다.

고유값 $(\lambda_1, \lambda_2, \dots, \lambda_k)$ 중 $(k-q)$ 개의 고유값들은 0에 가까운 값을 갖는다고 가정하면, 고유벡터 \mathbf{M} 은 처음 q 개의 열과 마지막 $k-q$ 개의 열로 구성된 행렬 \mathbf{M}_q 와 \mathbf{M}_{k-q} 로 각각 분할된다. 이때 q 개의 고유값에 해당하는 고유벡터 $\mathbf{M}_q = (\mathbf{m}_1, \mathbf{m}_2, \dots, \mathbf{m}_q)^T$ 만으로 선형 변환된 $(q \times 1)$ 벡터

$$\mathbf{Z}_t = \mathbf{M}_q \mathbf{X}_t \quad (4.1)$$

는 \mathbf{X}_t 가 갖는 대부분의 자기상관효과를 포함하며, \mathbf{X}_t 의 추정치 $\widehat{\mathbf{X}}_t$ 는 $\mathbf{M}_q^T \mathbf{M}_q \mathbf{X}_t$ 가 된다. 따라서 정준상관분석 결과 다음의 잔차를 얻을 수 있으며,

$$\mathbf{E}_t = \mathbf{X}_t - \widehat{\mathbf{X}}_t = \mathbf{X}_t - \mathbf{M}_q^T \mathbf{M}_q \mathbf{X}_t = (\mathbf{I} - \mathbf{M}_q^T \mathbf{M}_q) \mathbf{X}_t \quad (4.2)$$

이는 자기상관효과가 제거되어 측정치간에 독립성을 가정하는 기존의 다변량 관리도에 적용할 수 있게 된다.

잔차 \mathbf{E}_t 의 분산공분산행렬은 $(\mathbf{I} - \mathbf{M}_q^T \mathbf{M}_q) \Gamma_0 (\mathbf{I} - \mathbf{M}_q^T \mathbf{M}_q)^T$ 이므로 Hotelling의 T^2 통계량은

$$T_i^2 = \mathbf{E}_t^T \{ (\mathbf{I} - \mathbf{M}_q^T \mathbf{M}_q) \Gamma_0 (\mathbf{I} - \mathbf{M}_q^T \mathbf{M}_q)^T \}^{-1} \mathbf{E}_t \quad (4.3)$$

이고 T_i^2 는 자유도가 $(k-q)$ 인 χ^2 분포를 따르므로 Shewhart 관리도에서 관리상한선은 $\chi_a^2(k-q)$ 가 된다.

【 수치 예제 】

품질특성치 사이에 자기상관이 존재할 때, 구체적인 수치 데이터를 이용하여 제안된 관리도의 적용절차를 살펴본다.

평균벡터가 $\boldsymbol{\mu}^T = (0, 0, 0, 0)$ 이고, 자기회귀계수 및 백색잡음 \mathbf{a}_t 의 분산공분산행렬이 각각

$$\Phi_1 = \begin{bmatrix} 0.5 & 0.4 & 0.0 & -0.3 \\ 0.1 & 0.3 & 0.1 & 0.0 \\ 0.1 & 0.0 & -0.3 & -0.1 \\ -0.1 & 0.0 & 0.2 & 0.3 \end{bmatrix}, \quad \Sigma_w = \begin{bmatrix} 0.1 & -0.5 & -0.3 & 0.2 \\ -0.5 & 1.0 & 0.7 & 0.5 \\ -0.3 & 0.7 & 1.0 & 0.1 \\ 0.2 & 0.5 & 0.1 & 1.0 \end{bmatrix}$$

인 (4×1) 벡터 AR(1)과정에서 30개의 난수를 발생시킨 후 정리한 것이 <표 1>이다.

표 1. AR(1) 과정의 데이터

t	x_{1t}	x_{2t}	x_{3t}	x_{4t}
1	-0.44999	-0.81476	-0.33130	-0.22670
2	-0.45303	0.64183	0.71205	1.04989
3	0.99904	-0.44816	0.06310	2.03743
4	0.42551	0.41483	-1.65518	1.41046
5	-0.90606	0.47442	-0.75182	1.43031
6	-2.19327	1.14701	0.86715	-1.03465
7	0.06997	0.91412	-0.23891	1.23739
8	0.16800	0.96897	-0.40615	0.05181
9	1.54111	-1.12592	-0.12972	0.13114
10	0.56070	0.12969	2.52152	-0.23179
11	1.46976	0.48493	-1.23945	0.47649
12	-0.17676	2.02860	0.25247	0.46393
13	-0.36546	1.16941	0.85369	-0.53555
14	-0.07597	0.98883	-1.32959	0.96769
15	-0.32514	2.52024	2.26855	1.96364
16	0.04861	-1.08690	-2.83794	-0.12995
17	0.57188	-1.33782	1.03323	0.49748
18	-1.46116	1.97426	1.33037	0.11638
19	-1.82468	1.63399	0.40154	1.09421
20	0.21092	-0.63693	-2.50276	0.98468
21	-2.28138	0.87449	1.28630	-0.57097
22	0.08013	0.43164	1.90924	-0.12394
23	0.87233	1.77269	0.50623	-0.60361
24	0.02446	-0.93278	-2.33852	0.61697
25	-0.08351	-0.60959	-0.69208	-1.82022
26	0.42324	-0.22875	1.24839	0.55104
27	-1.15857	-0.64179	-0.18195	-0.89221
28	-2.15523	0.95911	0.01948	-0.66423
29	-1.94400	1.56937	-0.28485	1.70665
30	0.96819	-0.13988	1.23414	0.31003

표 2. $\Gamma_0^{-1} \Sigma_c$ 의 고유값과 고유벡터

	z_{1t}	z_{2t}	z_{3t}	z_{4t}
x_{1t}	0.6373	-0.5315	0.3740	-0.5453
x_{2t}	0.5677	0.2906	0.3776	0.6086
x_{3t}	-0.4071	-0.5264	0.3487	0.5763
x_{4t}	0.3255	0.5966	0.7720	-0.0094
λ_i	0.6338	0.3111	0.0317	0.0058

먼저 <표 1>의 데이터에 대해 T^2 통계량을 구하고 관리상태에서의 평균런길이가 200이 되도록 $\alpha=0.005$ 인 관리도에 타점한 것이 <그림 1>이다. 공정에 이상이 발생하지 않았음에도 23번째에서 관리한계선을 벗어났음을 알 수 있다.

제안된 관리도를 적용하기 위해 $\Gamma_0^{-1} \Sigma_c$ 의 고유값과 고유벡터를 구한 것이 <표 2>이다. 세 번째와 네 번째의 고유값이 각각 0.0317, 0.0058로서 0에 가까운 값을 가지므로 처음 두 개의 고유값에 대응하는 고유벡터만을 이용하여 변환한 변수

$$z_{1t} = 0.6373 x_{1t} + 0.5677 x_{2t} - 0.4071 x_{3t} + 0.3255 x_{4t}$$

$$z_{2t} = -0.5315 x_{1t} + 0.2906 x_{2t} - 0.5264 x_{3t} + 0.5966 x_{4t}$$

는 원래의 공정변수가 갖는 자기상관효과 대부분을 설명하게 된다. <그림 2>는 선형변환된 변수 \mathbf{z}_t 로부터 구한 잔차의 T^2 통계량을 관리도에 타점한 것으로 거짓경보 없이 모두 관리한계선 내에 타점됨을 알 수 있다.

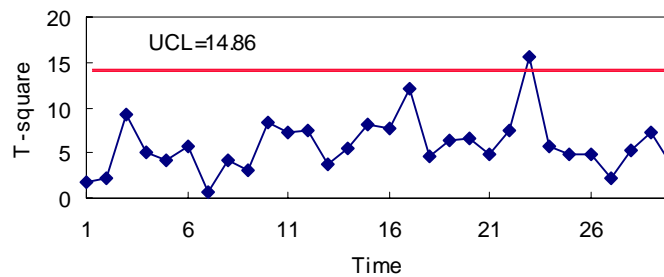


그림 1. 원자료에 대한 다변량 관리도

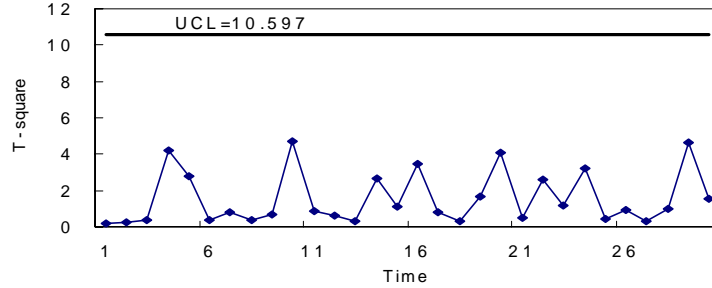


그림 2. 정준상관분석에 의한 다변량 관리도

5. 수행도분석을 위한 모의실험

측정치들 사이에 자기상관이 존재할 때, (i) T^2 관리도, (ii) 주성분분석에 의한 Runger의 다변량 관리도, (iii) 정준상관분석에 의한 Shewhart 관리도, (iv) 정준상관분석에 의한 EWMA 관리도들의 수행도를 비교·분석한다.

(i) T^2 관리도 :

측정치가 갖는 시계열 모형으로부터 구한 측정치 \mathbf{x}_t 의 분산공분산행렬 Γ_0 를 구하여 T^2 통계량

$$T_t^2 = \mathbf{x}_t^T \Gamma_0^{-1} \mathbf{x}_t$$

를 사용하는 관리도이다. 이때 Γ_0 를 시계열모형으로 구하였으므로 \mathbf{x}_t 에 나타나는 자기상관효과를 관리도에 반영하게 된다.

(ii) 주성분 분석에 의한 Runger의 다변량 관리도 :

Runger(1996)에 의해 제안된 관리도로서 \mathbf{x}_t 의 분산공분산행렬 Γ_0 에 대한 주성분분석 결과, 얻어진 잔차의 T^2 를 관리도에 타점하는 방법이다. 이 때의 잔차는 자기상관효과가 제거되어 독립성을 갖는다.

(iii) 정준상관분석에 의한 Shewhart 관리도 :

이 논문에서 제안된 방법으로 정준상관분석 결과 얻어진 잔차의 T^2 를 타점하는 방법이다.

(iv) 정준상관분석에 의한 EWMA 관리도 :

(iii)에서와 마찬가지로 정준상관분석 결과 얻어진 잔차의 T^2 를 EWMA 관리도에 적용한다. 시점 j 에서의 T_j^2 의 값을 t_j ($i = 1, 2, \dots$)이라 할 때 지수가중이동평균 Y_i 는 다음과 같다.

$$Y_i = \lambda t_i + (1-\lambda)Y_{i-1}$$

여기서 λ 는 가중치(weight)를 나타내며 $0 < \lambda \leq 1$ 이다. Y_0 는 지수가중 이동평균의 시작값으로 T^2 의 평균을 사용한다.

【 모의실험 설계 】

비교를 위해 자기상관모형으로 식 (2.4)의 AR(1)과정을 이용하였다. 분석의 편의를 위하여 백색잡음과정 $\{\mathbf{a}_t\}$ 는 평균이 $\mathbf{0}$ 이고 분산공분산행렬 $\Sigma_w = \begin{bmatrix} 1 & 0.7 \\ 0.7 & 1 \end{bmatrix}$ 인 이변량 정규분포를 따른다고 가정하였으며, 자기회귀계수행렬 $\Phi = \begin{bmatrix} \phi_{11} & \phi_{12} \\ \phi_{21} & \phi_{22} \end{bmatrix}$ 의 모수 중 교차자기회귀계수 $\phi_{12} = \phi_{21} = 0$ 으로 가정하였다. 실제로 AR(1)과정은 변수변환을 통하여 교차자기회귀계수가 0인 모형으로 변환될 수 있다.

관리가탈상태(out of control)의 관리도의 수행도 분석을 위해 측정치의 평균이 $\boldsymbol{\mu} = (0, 0)^T$ 에서 $\boldsymbol{\mu}_1 = (\mu_1, \mu_2)^T$ 으로 변환한 경우의 평균런길이를 구하였다. 이때 T^2 통계량의 확률분포는 $\boldsymbol{\mu}$ 의 원소 (μ_1, μ_2) 가 어떠한 값을 갖는지에 관계없이 비중심 모수(noncentral parameter) $\delta = \boldsymbol{\mu}^T \Gamma_0^{-1} \boldsymbol{\mu}$ 에 의존하므로 $\delta = 1, 2, 3$ 이 되도록 $\boldsymbol{\mu}_1 = (\mu_1, \mu_2)^T$ 을 구하였다.

▶ Shewhart 관리도

$\alpha = 0.005$ 인 관리한계선을 사용하여 관리상태에서의 평균런길이가 200이 되도록 하였다.

▶ EWMA 관리도

- $Y_0 = 1$: T^2 의 평균
- h : 관리한계
- $\lambda = 0.1, 0.3, 0.5$: 가중치
- h 값은 Shewhart 관리도와 비교하기 위해 공정이 관리상태에 있을때 평균런길이가 200이 되도록 관리한계선을 정하였다.

관리 통계량이 주어진 관리한계선보다 크면 우리는 난수생성을 멈추고 런길이를 계산한다. 여기서 10,000을 반복한 후 평균런길이를 계산한다.

<표 3>은 (ϕ_{11}, ϕ_{22}) 의 변화에 따른 각 관리도의 수행도를 평균런길이 측면에서 분석한 결과를 정리한 것이다.

(1) T^2 관리도는 자기회귀계수가 커짐에 따라 관리상태에서의 평균런길이의 값이 증가하는데 이는 단변량 관리도를 다룬 Reynolds(1996)등의 연구결과와 일치한다. 즉 시계열 모형에 의해 생성된 \mathbf{X}_t 의 자기상관계수가 높을수록 \mathbf{X}_t 의 분산공분산행렬은 큰 값을 갖게 되고, 이는 관리한계선의 폭을 넓히게 되어 관리상태에서의 평균런길이의 값이 증가하게 된다.

(2) Shewhart 관리도에서 주성분분석에 의한 관리도와 정준상관분석에 의한 관리도는 \mathbf{x}_t 와 이를 추정한 추정치 사이의 잔차를 이용한 통계량(T^2_t)을 타점하게 되는데, \mathbf{x}_t 에 대한 추정치는 주성분분석에 의한 관리도의 경우 공정의 산포를 많이 설명하는 변수들만의 선형결합이고 정준상관분석에 의한 관리도의 경우 자기상관효과가 큰 변수들만의 선형결합이다. 따라서 관리상한선이 $\chi^2_{0.005}(1)$ 으로 정해졌을 때, 평균런길이를 정확히 200으로 맞추는 것은 어렵다. T^2 관리도는 측정치들 사이에 자기상관효과가 존재하지만 주성분분석에 의한 관리도와 제안된 관리도는 어느 정도 자기상관효과가 제거되었으므로, 주성분분석에 의한 관리도와 제안된 관리도 모두 기존의 T^2 관리도에 비해 관리상태에서의 평균런길이가 200에 가깝게 유지됨을 알 수 있다.

(3) 관리이탈상태에서의 평균런길이를 살펴보면 제안된 관리도가 다른 두 곳의 관리도에 비해 공정의 이상상태를 빠르게 탐지함을 알 수 있다.

(4) 제안된 관리도에서 EWMA 관리도가 Shewhart 관리도보다 공정의 작은 변화를 빨리 탐지함을 알 수 있고, λ 값이 작을수록 공정의 변화에 민감함을 알 수 있다.

표 3. 자기상관자료에 대한 각 관리도의 평균런길이

(ϕ_{11}, ϕ_{22})	δ	T^2 관리도	주성분분석에 의한 관리도	제안된 Shewhart 관리도	제안된 EWMA 관리도		
					$\lambda=0.1$ h=1.843	$\lambda=0.3$ h=3.208	$\lambda=0.5$ h=4.537
(0.3, 0.1)	0	196.78	200.79	206.39	200.010	200.000	200.000
	1	43.45	205.10	38.23	28.511	32.807	37.721
	2	8.33	186.98	7.37	9.329	9.482	10.313
	3	2.79	166.20	2.60	6.363	6.172	6.361
(0.5, 0.1)	0	216.48	198.44	204.33	200.010	200.000	200.000
	1	48.68	157.10	32.69	15.944	22.272	27.436
	2	8.09	107.28	6.23	6.131	7.335	8.967
	3	2.87	52.85	2.18	4.342	4.767	5.512
(0.5, 0.3)	0	205.16	204.06	200.79	200.010	200.000	200.000
	1	46.18	190.66	40.07	20.882	21.228	23.526
	2	9.28	147.12	9.26	7.457	7.514	7.624
	3	3.37	100.48	3.42	5.129	5.203	5.247
(0.7, 0.1)	0	222.18	201.50	191.84	200.010	200.000	200.000
	1	48.32	95.80	29.71	17.749	21.736	25.659
	2	8.56	24.96	5.55	7.689	8.322	9.432
	3	2.10	5.96	2.06	5.504	5.596	6.047
(0.7, 0.3)	0	226.55	210.29	206.24	200.010	200.000	200.000
	1	49.04	114.27	33.79	14.236	16.452	19.421
	2	9.58	40.23	6.96	6.285	6.488	7.119
	3	3.15	12.41	2.65	4.515	4.588	4.806
(0.7, 0.5)	0	237.07	231.64	214.51	200.010	200.000	200.000
	1	54.84	157.62	46.06	11.006	10.763	11.610
	2	12.78	81.95	10.61	5.893	5.484	5.441
	3	4.34	40.10	4.21	4.641	4.322	4.198

표 3. (계속)

(ϕ_{11}, ϕ_{22})	δ	T^2 관리도	주성분분석에 의한 관리도	제안된 Shewhart 관리도	제안된 EWMA 관리도		
					$\lambda=0.1$ $h=1.843$	$\lambda=0.3$ $h=3.208$	$\lambda=0.5$ $h=4.537$
(0.9, 0.1)	0	285.57	193.21	202.70	200.010	200.000	200.000
	1	49.73	36.75	28.77	17.965	20.509	23.977
	2	7.30	5.82	5.22	8.591	8.780	9.644
	3	2.93	1.96	2.03	6.193	6.024	6.252
(0.9, 0.3)	0	294.49	207.67	208.67	200.010	200.000	200.000
	1	49.60	43.05	29.75	12.287	14.242	17.000
	2	9.85	7.03	6.14	6.027	6.202	6.722
	3	3.00	2.65	2.58	4.478	4.527	4.650
(0.9, 0.5)	0	304.10	223.43	215.71	200.010	200.000	200.000
	1	59.65	56.60	35.46	8.226	8.805	10.044
	2	11.19	11.59	8.31	4.843	4.884	5.023
	3	4.50	3.87	3.41	3.897	3.983	3.936
(0.9, 0.7)	0	344.57	280.98	253.35	200.010	200.000	200.000
	1	76.02	102.71	50.55	5.768	5.780	6.070
	2	16.35	27.06	12.27	4.041	4.054	4.087
	3	6.15	8.46	5.50	3.474	3.492	3.497

6. 맺음말

본 논문에서는 측정값들간의 자기상관관계를 효과적으로 제거하고 품질특성사이의 상관관계를 잘 설명할 수 있도록 하기 위해, 측정값들의 시계열 모형에 대한 정준상관분석 결과 얻어진 잔차를 이용하는 다변량 관리도를 제안하였다.

평균런길이에 대한 모의실험 결과 제안된 관리도는 관리상태에서의 평균런길이를 일정하게 유지할 수 있으며, 공정평균의 변화를 탐지하는 능력 또한 우수함을 보였다. Shewhart 관리도는 공정의 큰 변화는 빠르게 탐지하지만 공정의 작은 변화나 지속적인 변화에는 둔감하다는 단점이 있다. EWMA 관리도는 최근의 측정치에 더 큰 가중치를 주게 함으로써 공정의 작은 변화를 Shewhart 관리도보다 빨리 탐지함을 알 수 있다.

따라서 본 논문에서는 제안된 방법을 공정의 변화를 빨리 탐지할 수 있고 변화가 일어난 시간을 쉽게 파악할 수 있는 CUSUM 관리도에 응용하는 것도 의미 있는 연구가 될 것이다.

참고 문헌

1. Alt, F. B. (1985), "Multivariate Quality Control", *Encyclopedia of Statistical Sciences* (S. Kortz and Johnson, eds.), Vol. 6, John Wiley &

- Sons, New York, NY, pp. 110-122.
2. Alwan, C. A. and Roberts, H. V. (1988), "Time-Series Modeling for Statistical Process Control", *Journal of Business and Economic Statistics*, Vol. 6, pp. 87-95.
 3. Box, G. E. P. and Tiao, G. C. (1977), "A Canonical Analysis of Multiple Time Series", *Biometrika*, Vol. 64, pp. 355-365.
 4. Harris, T. J. and Ross, W. H. (1991), "Statistical Control Procedures for Correlated Observations", *The Canadian Journal for Chemical Engineering*, Vol. 69, pp. 48-57.
 5. Hotelling, H. (1931), "The generalization of Student's ratio", *Ann. Math. Statist.*, Vol 2, pp. 474-523.
 6. Hotelling, H. (1947), "Multivariate Quality Control", *Techniques of Statistical Analysis* (C. Eisenhart, M.W. Hastay, and W.A. Wallis eds.), McGraw-Hill, New York, NY, pp. 114-184.
 7. Hunter, J. S. (1986), "The Exponentially Weighted Moving Average", *Journal of Quality Technology*, Vol. 18, No.4, pp.203-210.
 8. Jackson, J. E. (1985), "Multivariate Quality Control", *Communications in Statistics-Theory and Methods*, Vol. 14, pp. 2647-2688.
 9. James, M. L and Michael, S. S (1990), "Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements", *Technometrics*, Vol. 32, No.1, pp. 1-12.
 10. Lin, W. W and Adams, B. M. (1996), "Combine Control Charts for Forecast-Based Monitoring Schemes", *Journal of Quality Technology*, Vol. 28, pp. 289-301.
 11. Lowry, C. A. and Montgomery, D. C. (1995), "A Review of Multivariate Control Charts", *IIE Transactions*, Vol. pp. 800-810.
 12. Mastrangelo, C. M. and Montgomery, D. C.(1995), "SPC with Correlated Observations for the Chemical and Process Industries", *Quality and Reliability International*, Vol. 11. pp. 79-89.
 13. Montgomery, D. C. (1992), "The Use for Statistical Process Control and Design of Experiments in Product and Process Improvement", *IIE Transactions* Vol. 24, pp. 4-17.
 14. Reynolds, M. R. and Arnold, J. C. and Baik, J. W. (1996), "Variable Sampling Interval \bar{X} Charts in the Presence of Correlation", *Journal of Quality Technology*, Vol. 28, pp. 12-30.
 15. Runger, G. C. (1996) "Multivariate Statistical Process Control for Autocorrelated Processes", *International Journal of Production Research*, Vol. 34, pp. 1715-1724.