

# 자기상관자료를 갖는 공정을 위한 다변량 관리도

남국현 · 장영순 · 배도선<sup>†</sup>

한국과학기술원 산업공학과

## Multivariate Control Chart for Autocorrelated Process

Gook Hyun Nam · Young Soon Chang · Do Sun Bai

Dept. of Industrial Engineering, Korea Advanced Institute of Science and Technology, Daejeon, 305-701

This paper proposes multivariate control chart for autocorrelated data which are common in chemical and process industries and lead to increase in the number of false alarms when conventional control charts are applied. The effect of autocorrelated data is modeled as a vector 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 therefore the control charts for these statistics can attenuate the autocorrelation in the process data. The charting procedures are illustrated with a numerical example and Monte Carlo simulation is conducted to investigate the performances of the proposed control charts.

**Keywords :** multivariate control chart, autocorrelation, canonical analysis, time series model

### 1. 서론

급변하는 경쟁사회에서 기업이 생존하기 위해서는 품질향상을 통한 제품의 원가절감 및 고급화, 다양화가 필수적이고, 이를 위해 공정을 감시하여 불량품이 발생하지 않도록 미리 예방활동을 하는 통계적 공정관리(statistical process control : SPC)의 체계적 활용이 강조되고 있다. SPC의 중요한 목적은 공정의 이상원인을 가능한 빨리 탐지하여 수정조치를 취함으로써 불량제품의 발생을 사전에 억제하고 공정으로부터 생산된 제품들의 품질특성치가 목표치로부터의 변동이 최소가 되도록 공정을 유도·조정하는 것으로, 이와 같은 목적으로 사용되는 가장 대표적인 기법이 관리도이다. 관리도는 품질특성의 변화를 그래프로 나타낸 것으로 관리상태에서의 품질특성치의 평균을 나타내는 중심선과 관리상한선 및 관리하한선으로 구성된다. 표본들을 추출하여 계산한 값들을 관리도상에 점으로 나타냈을 때 모든 점들이 관리한계선 사이에 놓이고 점들의 행태를 볼 때 공정에 이상이 있다고 판단할 만한 별다른 징후

가 없으면, 공정이 관리상태에 있다고 본다. 만약 한점이라도 관리한계선을 벗어나거나 점들이 비정상적인 행태를 보이면 공정이 관리상태를 이탈했을 가능성이 있을 것으로 판단하여 공정에 이상원인이 있는지 조사하게 된다.

관리도는 관리하고자 하는 품질특성치의 개수에 따라 단변량관리도와 다변량관리도로 나눌 수 있다. 단변량관리도는 품질특성치의 개수가 한 개인 경우에 사용되는 것으로 슈하르트 관리도, 누적합(cumulative-sum; CUSUM) 관리도, 지수가중이 동평균(exponentially weighted moving average; EWMA) 관리도 등이 있고, 다변량관리도는 상관관계가 존재하는 다수의 품질특성치들을 관리하기 위해 사용하는 것으로 Hotelling (1947)에 의해  $T^2$ 관리도가 제안된 이후 최근까지 많은 연구가 진행되어 왔으며 Alt(1985), Jackson(1985), Lowry and Montgomery (1995), Mason *et al.*(1997)은 다변량관리도에 대한 기존연구들을 정리하였다. 이러한 관리도들은 공정으로부터 채취되는 표본들이 서로 독립이고 동일한 분포를 갖는다는 가정하에서 구해진 것이다. 그러나 표본채취간격이 짧은 공정이나 장치산업과 같은 연속생산공정에는 공정의 특성상 측정치들이 과거의

이 연구는 1997년도 한국학술진흥재단의 공모과제 사업에 의해 지원되었음.

<sup>†</sup> Corresponding author: Professor Do Sun Bai, Dept. of Industrial Engineering, Korea Advanced Institute of Science and Technology, 373-1 Kusong-dong, Yusong-gu, Taejeon, 305-701 Korea; Fax +82-42-869-3110; e-mail dsbai@mail.kaist.ac.kr

2000년 3월 접수, 1회 수정 후, 2001년 7월 게재 확정.

측정치들에 의존하는 자기상관(autocorrelation)관계가 존재하기 때문에, 이러한 관리도를 이용하면 공정을 효율적으로 관리하기 어렵게 된다. 즉, 자기상관관계가 있는 측정치들을 이러한 관리도에 그대로 적용하면 실제로 이상원인이 존재하지 않더라도 이상원인이 발생한 것처럼 거짓경보(false alarm)를 주는 경우가 많아져서 관리도를 사용하는 소기의 목적을 달성할 수 없게 된다.

이러한 문제점을 해결하기 위한 단변량관리도에 대한 연구로, Alwan and Roberts (1988)는 시계열 모형을 통하여 얻어진 예측오차를 이용하는 관리도를 운용할 것을 제안하였고 Harris and Ross(1991)는 Kalman filter를 이용하여 자기상관효과를 제거한 뒤에 관리도를 사용할 것을 제안하였다. Mastrangelo and 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. 그 밖에 Montgomery (1992)는 자기상관자료에 대한 공정관리기법에 대한 연구결과들을 정리하였으며, Lin and Adams(1996)는 마코프 연쇄(Markov chain)를 이용하여 시계열 예측모형을 사용하는 기존의 관리도들의 수행도를 분석하였다. 다변량 관리도에 대한 연구로, Runger(1996)는 공정이 상태공간(steady-state)모형을 따르는 경우 주성분분석(principal component analysis)을 이용하여 자기상관을 효과적으로 제거할 수 있음을 보였고, Mason *et al.*(1996)은 몇개의 순차적인 공정 사이에 자기상관이 존재하는 경우  $T^2$  통계량을 분해하여 관리도를 구성하는 방법을 제안하였다. Bakshi (1998)는 웨이블릿 변환을 이용하여 자기상관효과를 제거한 후 주성분분석을 사용하는 방법을 제안하였다. 또한, Negiz and Çinar(1998)는 품질특성치와 상태변수(state variable) 간의 관계를 정준상관분석(canonical analysis)을 통해 모형화하는 방법을 연구하였으며 Norvilas *et al.*(2000)은 Negiz and Çinar(1998)의 방법을 적용하여 공정관리를 위한 실시간 지식기반시스템을 설계하였다.

이 논문은 정준상관 분석을 이용하여 다변량 자기상관 공정을 관리하기 위한 관리도를 설계하는 것으로, 측정치들을 자기회귀과정(autoregressive process)으로 모형화하고 정준상관 분석을 통해 분산공분산 행렬에서 자기상관효과를 분리해 낸 후 이를 이용하여 다변량  $T^2$  관리도를 구성하는 방법을 제안한다. 또한, 모의실험(simulation)을 통하여 제안된 관리도를 비교적 널리 사용되는 주성분분석을 이용한 Runger(1996)의 방법과 평균 런의 길이(average run length : ARL) 측면에서 비교·평가한다.

## 2. 자기상관 공정 모형

서로 상관관계가 있는 둘 이상의 품질특성을 관리해야 하는 다변량 관리도에서는 모집단이 다변량 정규분포를 따른다는

가정하에 가설  $\mu = \mu_0$ 를 검정하는  $t$  통계량을 다변량으로 확장한 Hotelling의  $T^2$  통계량

$$T_t^2 = (X_t - \mu_0)^T \Sigma^{-1} (X_t - \mu_0) \quad (1)$$

을 타점통계량으로 사용한다. 여기서  $(k \times 1)$  확률벡터  $X_t$ 가 평균벡터가  $\mu_0$ 이고 분산공분산 행렬이  $\Sigma$ 인  $k$ 변량 정규분포를 따른다면,  $T^2$ 통계량은 자유도가  $k$ 인  $\chi^2$ -분포를 따르게 되고, 유의수준을  $\alpha$ 로 하는 관리상한선(UCL)이  $UCL = \chi_{\alpha}^2(k)$ 인 관리도에 타점된다. 여기서  $\chi_{\alpha}^2(k)$ 는 자유도가  $k$ 인  $\chi^2$ -분포의  $(1 - \alpha) \times 100$  분위수이다.

식 (1)과 같은 기존의 다변량 관리도는 시점  $t$ 에서의 측정치  $X_t$ 가

$$X_t = \mu_0 + a_t \quad (2)$$

와 같이 서로 독립이고 동일한 분포를 갖는다는 가정하에 사용된다. 여기서  $\{a_t\}$ 는 백색잡음과정(white noise process)으로  $a_t$ 는 평균이 0이고,  $(k \times k)$  분산공분산 행렬  $\Sigma$ 를 갖는  $k$ 변량 정규분포를 따른다. 그러나 화학공정과 같은 연속생산공정에서는 연속적인 측정치들 간에 자기상관이 존재하므로,  $t$ 에서의 측정치  $X_t$ 는 시계열모형으로 모형화될 수 있다. 이 논문에서는 Box와 Jenkins의 ARIMA(autoregressive integrated moving average)과정 중 연속생산공정 데이터를 잘 설명한다고 알려진  $p$ 차의 자기회귀과정(AR( $p$ ))을 고려하여 자기상관공정을 모형화한다.

자기회귀과정은  $t$ 시점에서의  $(k \times 1)$  시계열 데이터 벡터  $X_t$ 가 과거 데이터  $X_{t-1}, X_{t-2}, \dots$ 들의 가중선형결합(weighted linear combination)으로 표현되는 모형으로,  $p$ 차의 자기회귀과정 AR( $p$ )는

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + a_t \quad (3)$$

이다. 여기서,  $\phi_i$ 는  $(k \times k)$  자기회귀계수행렬로  $X_t$ 와  $X_{t-i}$  간의 상관관계를 나타내며,  $\{a_t\}$ 는 분산공분산행렬이  $\Sigma_w$ 인  $k$ 변량 정규분포를 따르는 백색잡음과정이다. 이 모형은 시계열 데이터 간의 관계가 회귀식 형태로 표현되어 직관적으로 이해하기 쉽다는 장점이 있다. 자기회귀과정 중 실제 응용성이나 수식적 편의성 때문에 특히 차수가 1인 AR(1) 과정이 많이 이용되는데, 시계열 데이터의 평균이  $\mu$ 이고 차수가 1인 AR(1) 과정은

$$X_t = \mu + \phi_1 (X_{t-1} - \mu) + a_t \quad (4)$$

으로 표현된다.

$(k \times 1)$  확률벡터  $X_t$ 가 식 (3)의 AR( $p$ )를 따른다면, 시계열 데이터의 자기상관 효과를 알아보기 위해  $X_t$ 의  $j$ 지연(lag) 자기공분산행렬을 구하면

$$\begin{aligned} \Gamma_j(X) &= E(X_{t-j} X_t^T) \\ &= E\left(X_{t-j} \left[ \sum_{i=1}^p \phi_i X_{t-i} + a_t \right]^T\right) \\ &= \sum_{i=1}^p \Gamma_{j-i}(X) \phi_i^T, \quad j=1, 2, \dots, p \end{aligned} \quad (5)$$

가 되고,  $X_t$ 의 분산공분산행렬은

$$\Gamma_0 = \sum_{i=1}^p \phi_i \Gamma_i(X) + \Sigma_w = \Sigma_c + \Sigma_w \quad (6)$$

가 된다. 여기서  $X_t$ 의 분산공분산행렬은 자기상관효과에 대한 부분  $\Sigma_c$ 와 백색잡음에 의한 부분  $\Sigma_w$ 로 나누어진다. 따라서 추정치  $X_t$ 에서 자기상관으로 인해 나타나는 효과  $\Sigma_c$ 를 제거하면 효율적인 공정관리를 이룰 수 있을 것이다.

### 3. 정준상관 분석

연속된 측정치들 간에 자기상관이 있는 공정을 관리하기 위해서는 Alwan and Roberts(1988)가 제안한 것처럼 공정을 적절히 묘사하는 시계열 모형을 세우고, 이 시계열모형으로부터 구간 추정치와 실제 측정치 간의 차이인 잔차들을 이용할 수 있다. 잔차들은 측정치에 나타나는 자기상관 효과를 시계열의 추정치로 제거한 것으로, 서로 독립이 되어 기존의 관리도에 그대로 적용할 수 있을 것이다. 그러나 품질특성의 개수가 매우 많은 경우에는 시계열 모형에 의한 잔차를 이용하여 관리도를 운영하더라도 공정의 분석이 쉽지 않다. 따라서 이 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(Box and Tiao, 1977).

공정변수  $X_t$ 의 선형조합

$$y_t = m^T X_t \quad (7)$$

을 고려하자. 여기서  $m$ 은 선형계수를 나타내는  $(k \times 1)$  벡터이다.  $X_t$ 가 AR( $p$ )를 따르는 경우  $y_t$ 의 분산은 식 (6)으로부터

$$m^T \Gamma_0 m = m^T \Sigma_c m + m^T \Sigma_w m \quad (8)$$

가 되어 자기상관효과에 대한 부분( $m^T \Sigma_c m$ )과 백색잡음( $m^T \Sigma_w m$ )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

$$\lambda = \{m^T \Sigma_c m\} / \{m^T \Gamma_0 m\} \quad (9)$$

가 큰 순서대로 선형변환된 변수를 취하면 된다. 즉  $\lambda_1 > \lambda_2 > \dots > \lambda_k$ 에 해당하는  $m_1, m_2, \dots, m_k$ 를 구하여  $M = (m_1, m_2, \dots, m_k)^T$ 라 하면, 선형변환된 공정변수

$$Y_t = M X_t \quad (10)$$

를 얻게 되고, 식 (3)의 AR( $p$ ) 모형은

$$Y_t = \sum_{i=1}^p \tilde{\phi}_i X_{t-i} + b_t \quad (11)$$

와 같이 변환된다. 여기서  $\lambda_i$ 와  $m_i$ 는 각각 행렬  $\Gamma_0^{-1} \Sigma_c$ 의 고유값(eigenvalue)과 고유벡터(eigenvector)가 되고,  $\tilde{\phi}_i = M \phi_i$ ,  $b_t = M a_t$ 이다. 이 때,  $Y_t$ 의 분산공분산행렬을 구하면

$$M \Gamma_0 M^T = M \Sigma_c M^T + M \Sigma_w M^T \quad (12)$$

가 된다.  $\Lambda$ 를  $(\lambda_1, \lambda_2, \dots, \lambda_k)$ 로 구성된 대각행렬이라 할 때,

$$\begin{aligned} (M^T)^{-1} \Gamma_0^{-1} \Sigma_c M^T &= \Lambda, \\ (M^T)^{-1} \Gamma_0^{-1} \Sigma_w M^T &= I - \Lambda \end{aligned} \quad (13)$$

이므로,  $i \neq j$ 에 대하여  $m_i^T \Sigma_w m_j = m_i^T \Sigma_c m_j = 0$  이 된다. 즉,  $M \Sigma_c M^T$ 와  $M \Sigma_w M^T$ 는 대각행렬이 되고,  $M \Gamma_0 M^T$  역시 대각행렬이 된다. 따라서,  $Y_t$ 는 자기상관효과를 가장 많이 설명하는 변수순으로 순서화되어 있고, 각 공정변수들 간에는 서로 독립이다.

### 4. 자기상관공정에 대한 다변량관리도

일반적으로 자기상관공정에서 얻어진 많은 공정변수들 중 일부는 다른 공정변수들에 의해 설명되어 질 수 있고, 서로 유사한 자기상관효과를 갖는다. 예를 들어 연속생산공정에서 서로 다른 위치에서 온도를 시간에 따라 측정하는 경우 측정치 간에는 유사한 자기상관효과를 갖게 된다. 따라서 고유값  $(\lambda_1, \lambda_2, \dots, \lambda_k)$  중  $(k-q)$ 개의 고유값들은 0에 가까운 값을 갖게 된다. 만일  $(k-q)$ 개의 고유값들이 0이라면

$$M \Sigma_c M^T = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \quad (14)$$

와 같이 나타낼 수 있다. 여기서  $D$ 는  $(q \times q)$  대각행렬이다.

$Y_t, b_t, \tilde{\phi}_i$ 를 각각

$$Y_t = \begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix}, \quad b_t = \begin{bmatrix} b_{1t} \\ b_{2t} \end{bmatrix}, \quad \tilde{\phi}_i = \begin{bmatrix} \tilde{\phi}_i^{11} & \tilde{\phi}_i^{12} \\ \tilde{\phi}_i^{21} & \tilde{\phi}_i^{22} \end{bmatrix} \quad (15)$$

와 같이 고유값이 0이 아닌 부분과 0인 부분으로 분해하면, 고유값이 0에 해당하는 부분의  $\tilde{\Phi}_i^{21}$ 과  $\tilde{\Phi}_i^{22}$ 는 0이 되어 식(11)은

$$\begin{bmatrix} Y_{1t} \\ Y_{2t} \end{bmatrix} = \sum_{i=1}^p \begin{bmatrix} \tilde{\Phi}_i^{11} & \tilde{\Phi}_i^{12} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} Y_{1,t-i} \\ Y_{2,t-i} \end{bmatrix} + \begin{bmatrix} b_{1t} \\ b_{2t} \end{bmatrix} \quad (16)$$

이 된다. 즉,  $Y_t$ 는  $X_t$ 의 자기상관관계를 모두 포함하는  $q$ 차원 벡터  $Y_{1t}$ 와 자기상관관계가 모두 제거된  $(k-q)$ 차원 벡터  $Y_{2t}$ 로 분해된다.

따라서, 0의 값에 가까운  $(k-q)$ 개의 고유값에 해당하는 고유벡터  $M_{k-q}^T = (m_{q+1}, m_{q+2}, \dots, m_k)$ 만으로 선형변환된 벡터

$$Z_t = M_{k-q} X_t \quad (17)$$

는  $X_t$ 에서 자기상관효과가 대부분 제거된 잔차가 되어 측정치 간의 독립성을 가정하는 기존의 다변량 관리도에 적용할 수 있게 된다. 또한 변환된 벡터의 차수도  $(k-q)$  차로 줄어들고 변수 간에 독립성이 보장되어 관리도의 운영이 용이하다는 장점이 있다.  $\text{diag}\{d_{q+1}, \dots, d_k\}$ 를 대각원소가  $d_{q+1}, \dots, d_k$ 로 이루어진 대각행렬이라 하면,  $Z_t$ 의 분산공분산 행렬은

$$M_{k-q} \Gamma_0 M_{k-q} = \text{diag}\{d_{q+1}, \dots, d_k\}$$

와 같이 대각행렬이 되므로 Hotelling의  $T^2$  통계량은

$$T_t^2 = \sum_{i=q+1}^k \frac{Z_i^T Z_i}{d_i} \quad (18)$$

이고,  $T_t^2$ 는 자유도가  $(k-q)$ 인  $\chi^2$ -분포를 따르므로 UCL은  $\chi^2_{\alpha}(k-q)$ 가 된다.

[수치 예제]

여기서는 수치데이터를 이용하여 제안된 관리도의 적용절차를 살펴본다. 자기회귀계수  $\Phi_1$  및 백색 잡음  $a_t$ 의 분산공분산행렬이 각각

$$\Phi_1 = \begin{bmatrix} 0.5 & 0.4 & 0.0 & -0.3 \\ 0.1 & 0.3 & 0.1 & 0.0 \\ 0.1 & 0.0 & -0.3 & -0.1 \\ -0.1 & 0.0 & 0.2 & 0.3 \end{bmatrix}$$

와

$$\Sigma_w = \begin{bmatrix} 1.0 & -0.5 & -0.3 & 0.2 \\ -0.5 & 1.0 & 0.7 & 0.5 \\ -0.3 & 0.7 & 1.0 & 0.1 \\ 0.2 & 0.5 & 0.1 & 1.0 \end{bmatrix}$$

인  $(4 \times 1)$  벡터 AR(1) 과정에서 30개의 난수를 발생시킨 후

표 1. AR(1)과정의 데이터

t	$x_{1t}$	$x_{2t}$	$x_{3t}$	$x_{4t}$
1	0.027	0.294	0.473	1.001
2	-0.440	1.065	-0.117	0.970
3	0.591	0.747	0.907	0.688
4	1.409	-0.859	-0.188	-0.315
5	0.714	0.013	0.076	0.725
6	0.853	-0.868	-1.006	-0.299
7	0.012	-1.032	0.715	-2.093
8	-0.084	0.227	1.074	-0.405
9	1.266	-0.507	-1.290	0.191
10	-1.671	-0.031	1.275	-2.200
11	0.524	-1.257	-0.413	-1.845
12	0.592	-0.469	0.926	-0.529
13	0.414	0.813	0.684	1.102
14	-1.511	0.979	0.305	-0.490
15	0.690	-0.640	-1.231	-0.060
16	-0.096	-2.138	-0.636	-2.597
17	0.347	-0.960	-0.218	-0.246
18	0.925	-1.190	-0.775	0.632
19	1.730	-1.338	0.118	-0.391
20	0.145	0.446	0.468	0.584
21	1.162	1.379	0.311	1.171
22	1.945	1.805	2.158	1.779
23	-0.218	2.186	1.913	0.233
24	1.958	1.889	0.509	2.765
25	0.238	3.077	1.311	1.706
26	1.068	3.097	1.769	2.275
27	0.160	2.961	3.258	1.576
28	3.422	1.750	1.473	3.063
29	2.008	0.032	0.910	0.012
30	0.968	1.676	3.335	0.523

정리한 것이 <표 1>이다. 이 중 20개의 데이터는 평균벡터  $\mu$ 가  $(0, 0, 0, 0)^T$ 인 관리상태에서 발생된 것이고, 나머지 10개의 데이터는 평균이  $(1.5, 1.5, 1.5, 1.5)^T$ 로 변한 관리이탈상태로부터 얻어진 것이다.

제안된 관리도를 적용하기 위해  $X_t$ 의 분산공분산행렬  $\Gamma_0$ 와  $\Sigma_c$ 를 구하면

$$\Gamma_0 = \begin{bmatrix} 1.211 & -0.438 & -0.291 & 0.179 \\ -0.438 & 1.129 & 0.597 & 0.617 \\ -0.291 & 0.597 & 1.139 & -0.020 \\ 0.179 & 0.617 & -0.020 & 1.161 \end{bmatrix},$$

$$\Sigma_c = \begin{bmatrix} 0.211 & 0.062 & 0.009 & -0.021 \\ 0.062 & 0.129 & -0.103 & 0.117 \\ 0.009 & -0.103 & 0.139 & -0.120 \\ -0.021 & 0.117 & -0.120 & 0.161 \end{bmatrix}$$

이 되고,  $\Gamma_0^{-1} \Sigma_c$ 의 고유값과 고유벡터는 <표 2>와 같이 된다. <표 2>에서 세 번째와 네 번째의 고유값은 각각 0.032, 0.006으

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

	$z_{1t}$	$z_{2t}$	$z_{3t}$	$z_{4t}$
$x_{1t}$	0.396	0.618	0.199	-0.314
$x_{2t}$	0.736	0.023	-0.299	0.805
$x_{3t}$	-0.409	0.552	0.516	0.322
$x_{4t}$	-0.367	-0.560	0.778	-0.387
$\lambda_i$	0.634	0.311	0.032	0.006

로서 0에 가까운 값을 가지므로 처음 두 개의 고유값에 대응하는 변환은 원래의 공정변수  $X_t$ 가 갖는 자기상관효과의 대부분을 설명하게 되므로, 나머지 세 번째와 네 번째의 고유값에 대응하는 고유벡터만을 이용하여 변환한 변수

$$z_{3t} = 0.199x_{1t} - 0.299x_{2t} + 0.516x_{3t} + 0.778x_{4t}$$

$$z_{4t} = -0.314x_{1t} + 0.805x_{2t} + 0.322x_{3t} - 0.387x_{4t}$$

를 이용하여 관리도를 구성할 수 있다. 선형변환된  $Z_t$ 의 분산 공분산 행렬을 구하면

$$M_{k-q} \Gamma_0 M_{k-q}^T = \begin{bmatrix} 0.715 & 0.000 \\ 0.0006 & 1.3976 \end{bmatrix} = \begin{bmatrix} 0.715 & 0.000 \\ 0.000 & 1.397 \end{bmatrix}$$

이 되어  $T^2$  통계량은

$$T_t^2 = \frac{z_3^2}{0.715} + \frac{z_4^2}{1.397}$$

이 된다.

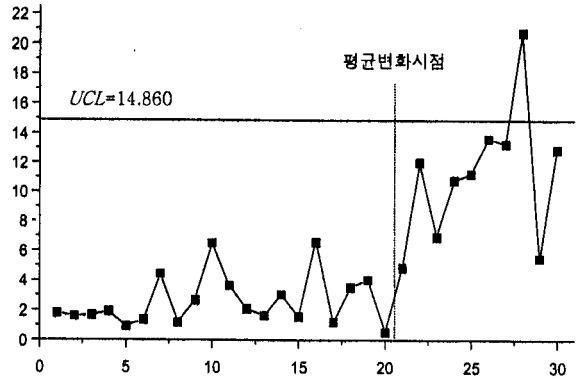
<그림 1(a)>는 <표 1>의 데이터에 대해 기존의 방법으로  $T^2$  통계량을 구하고 관리상태에서 ARL이 200이 되도록  $\alpha=0.005$ 인 관리도에 타점한 것이고, <그림 1(b)>는 제안된 방법으로  $T^2$  통계량을 구한 후 관리도에 타점한 것이다. 두 관리도 모두 관리상태에서는 모든 데이터가 관리한계선 내에 타점되었으나, 기존이 관리도에서는 공정평균이 변화한 후 8번째 데이터에서  $T_8^2 > UCL$ 이 되어 이상신호가 발생하고 제안된 관리도에서는 7번째 데이터에서 이상신호가 발생하였다.

### 5. 수행도 분석

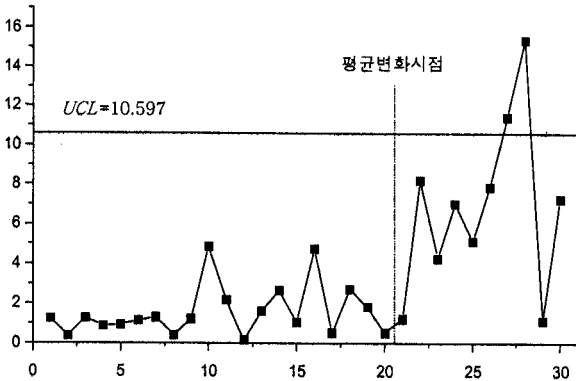
이 절에서는 추정치들에 자기상관이 존재할 때 i) 기존의  $T^2$  관리도, ii) Runger(1996)의 주성분 분석에 의한 다변량관리도, iii) 제안된 정준상관분석에 의한 다변량 관리도의 수행도를 비교·분석한다.

#### i) 기존의 $T^2$ 관리도

추정치가 갖는 시계열 모형으로부터 구한 추정치  $X_t$ 의 분산공분산행렬  $\Gamma_0$ 를 구하여  $T^2$  통계량



(a) 원 자료에 의한 다변량 관리도



(b) 정준상관분석에 의한 다변량 관리도

그림 1. <표 1>의 자료에 대한 다변량 관리도.

$$T_t^2 = X_t^T \Gamma_0^{-1} X_t$$

를 타점통계량으로 사용하는 관리도이다. 이때  $\Gamma_0$ 를 시계열 모형으로부터 구하였으므로  $X_t$ 에 나타나는 자기상관효과를 관리도에 반영하게 된다.

#### ii) 주성분 분석에 의한 다변량관리도

Runger(1996)에 의해 제안된 관리도로서  $X_t$ 의 분산공분산 행렬  $\Gamma_0$ 에 대한 주성분분석 결과 얻어진 잔차를 관리도에 타점하는 방법이다. 이때의 잔차는 자기상관 효과가 제거되어 독립성을 갖는다.

#### iii) 제안된 정준상관분석에 의한 다변량 관리도

이 논문에서 제안된 방법으로 정준상관분석 결과 얻어진 잔차를 관리도에 타점하는 방법이다.

비교를 위한 자기상관모형으로 식 (4)의 AR(1) 과정을 이용하였다. 분석의 편의를 위하여 자기회귀계수행렬  $\phi_1$ 의 모수 중 교차자기 회귀계수는 0으로 가정하였다. 실제로 AR(1) 과정은 변수변환을 통하여 교차자기 회귀계수가 0인 모형으로 변환될 수 있다.

관리가탈상태의 관리도의 수행도 분석을 위해 추정치의 평균이  $\mu=0$ 에서  $\mu_1$ 으로 변화한 경우의 ARL을 구하였다. 이

때  $T^2$  통계량의 확률분포는  $\mu$ 의 원소가 어떠한 값을 갖는지에 관계없이 비중심모수  $\delta = \mu^T \Gamma_0^{-1} \mu$ 에 의존하므로  $\delta = 1, 2, 3$ 이 되도록  $\mu_1$ 을 구하였다.

<표 3>과 <표 4>는  $\Sigma_w$ 와  $\Phi_1$ 의 변화에 따른 각 관리도의 수행도를 ARL 측면에서 분석한 결과를 정리한 것으로  $ARL(T^2)$ ,  $ARL(PC)$ ,  $ARL(CA)$ 는 각각 기존의  $T^2$  관리도, 주성분분석을 이용한 관리도, 제안된 관리도의 ARL을 의미한다. <표 3>은  $\Sigma_w = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$  이고  $\Phi_1 = \begin{bmatrix} \phi_{11} & 0 \\ 0 & \phi_{22} \end{bmatrix}$  인 2변량의 경우에 대해  $(\phi_{11}, \phi_{22}, \rho)$ 를 변화시켜 가면서 얻은 결과이고, <표 4>는 4변량 품질특성치를 갖는 경우에 대해  $(\phi_{11}, \phi_{22}, \phi_{33}, \phi_{44})$ 를 변화시켜 가면서 얻은 것이다. <표 4>의 Case 1은 다양한 양의 상관관계가 존재하는 경우로

$$\Sigma_w = \begin{bmatrix} 1.0 & 0.8 & 0.4 & 0.4 \\ 0.8 & 1.0 & 0.6 & 0.2 \\ 0.4 & 0.6 & 1.0 & 0.2 \\ 0.4 & 0.2 & 0.2 & 1.0 \end{bmatrix}$$

를 이용하였고, Case 2는 양의 상관관계와 음의 상관관계가 모두 존재하는 경우로

$$\Sigma_w = \begin{bmatrix} 1.0 & -0.5 & -0.3 & 0.2 \\ -0.5 & 1.0 & 0.7 & 0.5 \\ -0.3 & 0.7 & 1.0 & 0.1 \\ 0.2 & 0.5 & 0.1 & 1.0 \end{bmatrix}$$

를 사용하였다. 이때 <표 3>과 <표 4>의 각 관리도는  $\alpha = 0.005$ 인 관리한계선을 사용하여 관리상태에서의 ARL

표 3. 자기상관자료에 대한 각 관리도의 ARL (2변량의 경우)

$\delta$	$(\phi_{11}, \phi_{22}, \rho)$	$ARL(T^2)$	$ARL(PC)$	$ARL(CA)$	$(\phi_{11}, \phi_{22}, \rho)$	$ARL(T^2)$	$ARL(PC)$	$ARL(CA)$
0	(0.3, 0.1, -0.6)	198.52	201.54	199.46	(0.3, 0.1, 0.5)	202.85	202.27	201.71
1		43.65	29.58	74.96		43.21	196.81	43.04
2		7.62	5.33	20.37		7.62	182.06	8.47
3		2.33	1.81	7.21		2.32	167.20	2.84
0	(0.5, 0.3, -0.6)	210.65	211.87	202.78	(0.5, 0.3, 0.5)	210.27	209.99	201.29
1		48.17	33.45	66.06		48.19	189.72	47.46
2		8.92	6.11	16.58		8.78	143.26	10.04
3		2.58	1.97	5.71		2.62	98.05	3.33
0	(0.7, 0.5, -0.6)	237.39	229.86	217.27	(0.7, 0.5, 0.5)	235.96	229.06	215.15
1		58.40	41.48	66.73		57.08	162.96	50.42
2		11.54	7.80	17.00		11.05	83.20	11.07
3		3.08	2.14	5.85		3.04	42.89	3.47
0	(0.9, 0.7, -0.6)	340.22	277.75	255.54	(0.9, 0.7, 0.5)	332.68	275.46	256.22
1		89.33	58.11	72.77		79.59	100.70	55.76
2		18.08	11.28	18.25		16.45	28.50	11.79
3		4.14	2.47	5.79		4.11	9.78	3.13
0	(0.3, 0.1, 0.3)	202.86	205.99	199.97	(0.3, 0.1, 0.8)	202.12	201.63	199.04
1		43.30	194.73	47.68		44.28	197.88	35.77
2		7.48	172.37	9.93		7.67	183.84	6.45
3		2.35	147.60	3.30		2.37	164.93	2.22
0	(0.5, 0.3, 0.3)	210.02	207.75	202.22	(0.5, 0.3, 0.8)	215.51	209.24	200.97
1		47.47	173.37	50.73		48.47	188.33	39.58
2		8.74	115.69	11.09		8.56	137.65	7.70
3		2.60	70.80	3.70		2.58	96.73	2.57
0	(0.7, 0.5, 0.3)	230.42	226.96	216.41	(0.7, 0.5, 0.8)	239.04	232.97	210.45
1		56.96	137.42	54.84		55.82	161.18	41.45
2		11.37	58.91	12.11		10.81	83.84	8.30
3		3.14	26.56	3.94		3.05	41.90	2.65
0	(0.9, 0.7, 0.3)	328.75	268.42	257.40	(0.9, 0.7, 0.8)	353.37	287.81	250.75
1		82.34	87.94	60.01		77.68	112.81	46.38
2		16.82	22.89	13.25		14.88	32.60	9.31
3		4.14	7.17	3.64		3.88	10.33	2.51

표 4. 자기상관자료에 대한 각 관리도의 ARL(4변량의 경우)

Cases	$(\phi_{11}, \phi_{22}, \phi_{33}, \phi_{44})$	$\delta$	ARL( $T^2$ )	ARL(PC)	ARL(CA)
1	(0.1, 0.2, 0.2, 0.3)	0	200.18	201.98	198.67
		1	62.59	179.27	65.97
		2	11.64	134.32	14.21
		3	3.29	89.45	4.34
	(0.3, 0.5, 0.4, 0.6)	0	209.97	209.12	208.67
		1	67.98	183.07	59.80
		2	14.10	130.80	12.32
		3	3.92	84.47	3.75
	(0.7, 0.8, 0.6, 0.9)	0	290.03	247.40	232.45
		1	95.76	152.56	63.55
		2	20.62	61.22	13.10
		3	5.85	26.41	3.76
(0.2, 0.8, 0.4, 0.6)	0	239.70	209.16	204.23	
	1	73.13	137.89	53.13	
	2	14.09	60.34	10.08	
	3	3.91	27.07	3.15	
2	(0.1, 0.2, 0.2, 0.3)	0	201.60	201.63	199.15
		1	62.14	75.58	85.07
		2	11.27	18.00	22.03
		3	3.10	5.74	6.97
	(0.3, 0.5, 0.4, 0.6)	0	216.20	211.98	204.19
		1	68.40	71.26	66.30
		2	13.40	16.85	14.46
		3	3.73	5.22	4.58
	(0.7, 0.8, 0.6, 0.9)	0	295.98	255.69	235.17
		1	100.94	83.03	73.26
		2	23.12	19.29	16.39
		3	6.44	5.37	4.74
(0.2, 0.8, 0.4, 0.6)	0	235.87	216.51	204.82	
	1	75.45	72.79	51.77	
	2	14.55	17.17	9.81	
	3	3.99	5.33	2.97	

이 200이 되도록 하였다. <표 3>과 <표 4>로부터 다음의 결과를 얻을 수 있다.

- ① 세 종류의 관리도 모두 자기회귀계수가 커짐에 따라 관리상태에서의 ARL이 증가한다. 그러나 제안된 관리도는 기존의 관리도뿐 아니라 주성분분석에 의한 관리도에 비해 자기상관을 효과적으로 제거하므로 다른 두 관리도보다 관리상태에서의 ARL이 일정하게 유지된다.
- ② 제안된 관리도는 관리도를 구성하기 위해 변환된 벡터의 차수가 품질특성치의 개수보다 작게 되므로 상관계수가 음인 경우를 포함하여 자기상관계수 또는 상관계수가 작은 경우에는 기존의 관리도보다 관리이탈상태에서의 성능이 다소 떨어지게 되나, ARL면에서 큰 차이가 발생하지 않는다. 그러나 자기상관계수와 상관계수가 커짐에 따

라 제안된 관리도의 성능이 우수하게 됨을 알 수 있다.

- ③ 주성분분석에 의한 관리도와 제안된 관리도는  $X_i$ 를 공정의 산포를 많이 설명하는 변수로 선형변환한 후 나머지 잔차를 타점하게 된다. 주성분분석은 자기상관효과와 백색잡음이 혼합되어 나타나는 산포를 설명하는 변수들의 선형결합을 찾아내는 것이고, 제안된 방법은 자기상관효과가 큰 변수들만의 선형결합을 찾아내는 것이다. 따라서 제안된 관리도가 주성분분석에 의한 관리도보다 자기상관효과를 효율적으로 제거할 수 있게 되어 수행도가 우수하게 된다.

### 6. 결론

이 논문에서는 측정치들 간의 자기상관관계를 효과적으로 제거하고 품질특성 사이의 상관관계를 잘 설명할 수 있도록 하기 위해 시계열 모형에 대한 정준상관분석을 이용하는 다변량 관리도를 제안하였다. 평균런의 길이에 대한 모의실험결과 제안된 관리도는 관리상태에서의 평균런의 길이를 일정하게 유지할 수 있으며, 공정평균의 변화를 탐지하는 능력 또한 우수함을 보였다. 따라서 측정치 간에 자기상관이 있는 경우 제안된 관리도는 기존의 관리도에 비해 효과적으로 공정을 관리할 수 있을 것이다. 이 논문에서 제안된 다변량관리도는 확률한계선을 관리한계선으로 사용하는 일종의 슈하르트형 관리도이다. 슈하르트형 관리도는 공정의 큰 변화는 빠르게 탐지하지만 공정의 작은 변화나 지속적인 변화에는 둔감하다는 단점이 있다. 따라서 이 논문에서 제안된 방법을 EWMA 관리도 또는 CUSUM 관리도에 응용하는 것도 의미 있는 연구가 될 것이다.

### 참고문헌

Alt, F. B. (1985), Multivariate Quality Control, in *Encyclopedia of Statistical Sciences* 6 (S. Kotz and Johnson, eds.), John Wiley & Sons, New York, NY, 6, 110-122.

Alwan, C. A. and Roberts, H. V. (1988), Time-Series Modelling for Statistical Process Control, *Journal of Business and Economic Statistics*, 6, 87-95.

Bakshi, B. R. (1998), Multiscale PCA with Application to Multivariate Statistical Process Monitoring, *AIChE Journal*, 44, 1596-1610.

Box, G. E. P. and Tiao, G. C. (1977), A Canonical Analysis of Multiple Time Series, *Biometrika*, 64, 355-365.

Harris, T. J. and Ross, W. H. (1991), Statistical Control Procedures for Correlated Observations, *The Canadian Journal of Chemical Engineering*, 69, 48-57.

Hotelling, H. (1947), Multivariate Quality Control, in *Techniques of Statistical Analysis* (C. Eisenhart, M. W. Hastay, and W. A. Wallis eds.), McGraw-Hill, New York, NY, 111-184.

Jackson, J. E. (1985), Multivariate Quality Control, *Communications*

- in Statistics-Theory and Methods*, **14**, 2657-2688.
- Lin, W. W. and Adams, B. M. (1996), Combined Control Charts for Forecast-Based Monitoring Schemes, *Journal of Quality Technology*, **28**, 289-301.
- Lowry, C. A. and Montgomery, D. C. (1995), A Review of Multivariate Control Charts, *IIE Transactions*, **27**, 800-810.
- Mason, R. L., Champ, C. W., Tracy, N. D., Wierda, S. J. and Young, J. C. (1997), Assessment of Multivariate Process Control Techniques, *Journal of Quality Technology*, **29**, 140-143.
- Mason, R. L., Tracy, N. D. and Young, J. C. (1996), Monitoring a Multivariate Step Process, *Journal of Quality Technology*, **28**, 39-50.
- Mastrangelo, C. M. and Montgomery, D. C. (1995), SPC with Correlated Observations for the Chemical and Process Industries, *Quality and Reliability Engineering International*, **11**, 79-89.
- Montgomery, D. C. (1992), The Use of Statistical Process Control and Design of Experiments in Product and Process Improvement, *IIE Transactions*, **24**, 4-17.
- Negiz, A. and Çinar, A. (1998), Monitoring of Multivariable Dynamic Processes and Sensor Auditing, *Journal of Process Control*, **8**, 375-380.
- Norvilas, A., Negiz, A., DeCicco, J. and Çinar, A. (2000), Intelligent Process Monitoring by Interfacing Knowledge-Based Systems and Multivariate Statistical Monitoring, *Journal of Process Control*, **10**, 341-350.
- Runger, G. C. (1996), Multivariate Statistical Process Control for Autocorrelated Processes, *International Journal of Production Research*, **34**, 1715-1724.