

배치 공정의 온라인 모니터링을 위한 다변량 관리도

- Multivariate SPC Charts for On-line Monitoring the Batch Processes -

이 배 진*

Lee, Bae Jin

강 창 욱*

Kang, Chang Wook

Abstract

Batch processes are a significant class of processes in the process industry and play an important role in the production of high quality speciality materials. Examples include the production of semiconductors, chemicals, pharmaceuticals, and biochemicals. With on-line sensors connected to most batch processes, massive amounts of data are being collected routinely during the batch on easily measured process variables such as temperatures, pressures, and flowrates.

In this paper, multivariate SPC charts for on-line monitoring of the progress of new batches are developed which utilize the information in the on-line measurements in real-time. We propose the formation of statistical model which describes the normal operation of a batch at each time interval during the batch operation. An on-line monitoring scheme based on the proposed method can handle both cross-correlation among process variables at any one time and auto-correlation over time. And the control limits for the monitoring charts are established from sound statistical framework unlike previous researches which use the external reference distribution.

The proposed charts perform real-time, on-line monitoring to ensure that the batch is progressing in a manner that will lead to a high-quality product or to detect and indicate faults that can be corrected prior to completion of the batch. This approach is capable of tracking the progress of new batch runs, identifying the time periods in which the fault occurred and detecting underlying cause.

1. 서론

배치 공정(batch process)은 반도체, 화학, 제약, 식품 등의 장치산업(process industry)에서 주로 사용하는 공정으로 고부가가치 제품 생산에 중요한 역할을 수행한다. 배치 공정의 예로는 반도체의 식각 공정, 자동차의 도장 공정, 다양한 제품 생산의

* 한양대학교 대학원 산업공학과

하부공정으로 사용되는 발효 공정, 증류 공정, 결정 공정, 분류 공정 등이 있다.

배치 공정은 유한기간 공정(finite duration process)이며 일반적으로 다음과 같은 단계를 따라 대상 제품을 생산한다. 먼저 규정된 방식으로 원재료를 투입하여 배치의 생산을 시작한다. 배치의 생산이 진행됨에 따라 온도, 압력, 반응물의 이송율(feedrate) 등과 같은 공정변수들이 미리 설정되어 있는 궤적(trajjectory)을 따라가도록 제어하면서 공정을 진행시킨다. 마지막으로 생산이 끝난 최종 제품을 방출함으로써 한 배치의 생산이 완료된다. 한 배치의 생산이 끝나면 다음 배치를 투입하고 동일한 단계들을 반복한다.

배치 공정에서는 센서나 온라인 컴퓨터를 통해서 온도, 압력, 이송율 등과 같이 쉽게 측정할 수 있는 대량의 공정변수 자료들이 온라인으로 수집되며, 측정기술의 발달에 의해 한 시점에서 측정되는 공정변수의 수가 상당히 많고 측정되는 시간 간격도 짧다. 배치의 진행에 따라 동적으로 변하는 공정변수의 궤적들은 배치의 상태에 대한 정보를 담고 있으며, 배치에서 발생한 이상원인은 이러한 공정변수들의 궤적에 영향을 미친다. 따라서 배치가 완료될 때까지 기다리지 않고 공정 이상을 검출하여 필요한 조치를 취하려면 이러한 온라인 자료를 이용하여 공정을 모니터링하는 것이 바람직하다. 또한 최종 제품의 품질 특성치에 이상이 발생한 경우, 공정 진행의 어느 시점에서 이상이 발생하였는지 추적하고 그 원인을 규명하는 데 온라인 공정변수 자료들이 유용한 정보를 제공한다.

본 연구는 대규모의 공정변수들이 온라인 센서를 통해 실시간으로 수집되는 배치 공정을 대상으로 한다. 이러한 온라인 공정변수 자료들을 관리도를 이용하여 온라인으로 모니터링하기 위해서는 배치 공정의 다음과 같은 특징들을 고려하여야 한다.

- (1) 배치 공정은 유한기간 공정이다.
- (2) 공정변수의 궤적은 시간에 따라 동적으로 변화하며, 선형이 아니다.
- (3) 한 시점에서 측정되는 공정변수는 여러 개이며, 이 변수들은 동일 배치의 동일 시점에서 측정한 자료이므로 일반적으로 대단히 상관관계가 크다.
- (4) 한 시점에서의 공정변수들은 이전 시점의 공정변수들과 독립이 아니다.

본 연구의 목적은 대량의 공정변수 자료들이 온라인으로 수집되는 경우에 배치 공정을 온라인 모니터링할 수 있는 다변량 관리도를 제안하는 것이다. 배치 공정의 온라인 모니터링을 위해 배치 공정의 특성을 고려하여 배치의 진행 시점별로 통계적 모형을 수립할 수 있는 시점별 모형화 방법을 제안한다.

1.1 배치 공정 자료의 구조

배치 공정의 온라인 모니터링을 위해서는 관리상태에 있는 공정을 나타내는 통계적 모형이 필요하다. 이것을 참조분포(reference distribution)라 하며, 과거의 성공적으로 진행된 배치들의 데이터베이스를 이용하여 수립한다. 배치 공정의 온라인 모니터링은 참조분포와 현재 생산중인 배치에서 얻어진 공정변수들을 각 시점별로 비교함으로써 공정을 모니터링하는 것이다.

참조분포를 수립하기 위해서 사용되는 과거의 정상적인 배치의 수를 I 라 하면, 전

체 자료는 그림 1과 같은 3차원 배열로 표현할 수 있다. 그림 1의 3차원 배열 \mathbf{X} ($I \times J \times K$)는 배치(batch), 공정변수(variable), 시간(time)의 3개의 축으로 I 개의 배치들의 K 개의 시점에서 측정된 J 개의 공정변수들의 측정치를 표현하고 있다. 3차원 배열 \mathbf{X} ($I \times J \times K$)를 수평으로 절단한 ($J \times K$) 행렬은 각 배치의 공정변수들의 전체 궤적을 나타낸다. 또한 \mathbf{X} ($I \times J \times K$)를 수직으로 절단한 ($I \times J$) 행렬은 동일 시점에서 측정된 J 개의 공정변수들의 측정치를 나타낸다.

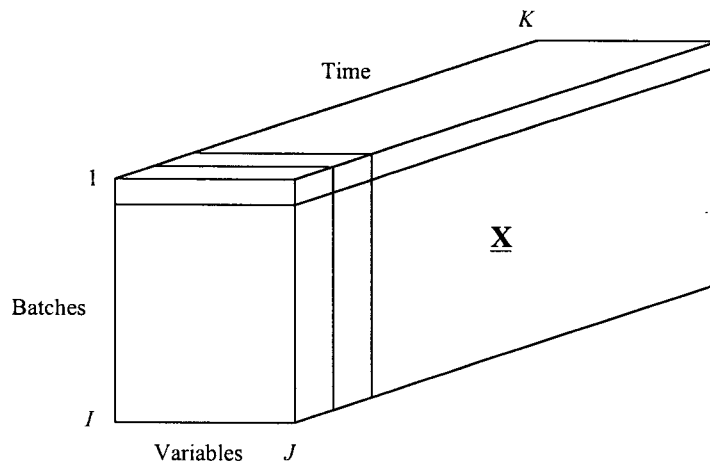


그림 1. 배치 공정 자료의 구조

1.2 기존 연구 고찰

다변량 SPC는 상관관계가 존재하는 다수의 공정변수들을 동시에 관리하는 방법이다. 대부분의 다변량 관리도 들은 Hotelling T^2 통계량을 기초로 하고 있으며, T^2 통계량은 p 개의 변수들로 이루어지는 관측 벡터 \mathbf{x} 의 평균 벡터 $\bar{\mathbf{x}}$ 로부터의 일반화된 거리(generalized distance)이다.

$$T^2 = (\mathbf{x} - \bar{\mathbf{x}})^T \mathbf{S}^{-1} (\mathbf{x} - \bar{\mathbf{x}}) \quad (1)$$

여기서 \mathbf{S} 는 표본 공분산 행렬(covariance matrix)이다.

Hotelling T^2 통계량의 계산을 위해서는 공분산 행렬의 역행렬을 저장하고 있어야 하며, 변수의 수(p)가 많은 경우에는 공분산 행렬의 크기($p \times p$)가 대단히 커진다. 또한 변수들이 상호 상관관계가 높은 경우(ill-conditioned data)에는 공분산 행렬의 역행렬을 구하는 데 문제가 발생한다. 이와 같은 대규모 공정의 경우 서로 직교하는 새로운 잠재변수(latent variable)들을 정의하고, 이러한 잠재변수들이 구성하는 저차원의 잠재변수 공간(low-dimensional latent variable space)으로 투영(projection)하여 관리도를 작성하는 방법이 제안되었다[2, 4]. 이러한 다변량 투영방법은 잠재변수를 정의하는 차원축소 기법에 의해 여러 가지로 분류할 수 있으며, 대표적인 방법으로 PCA (principal component analysis)[13]와 PLS(partial least squares)[1]가 있다. Geladi와 Kowalski[1]

는 다변량 투영방법의 잠재변수를 구하는 알고리즘을 제시하였다. PCA를 대규모 연속 공정의 공정 모니터링에 적용한 사례로는 Kresta et al.[6], MagGregor와 Kourti[8], Kourti와 MacGregor[5]의 연구가 있으며, Ku et al.[7]과 Russell et al.[11]은 자기상관 관계가 있는 연속 공정에 PCA를 적용하여 공정을 모니터링하는 방법을 제안하였다.

배치 공정의 진행 중에 측정되는 모든 공정변수들을 이용하여 배치 공정을 관리하는 다변량 관리도는 Nomikos와 MacGregor[9, 10]가 처음으로 연구하였다. 이들은 Multiway PCA(MPCA)를 이용하여 그림 1과 같은 3차원 배열을 모형화하였다. MPCA는 3차원 배열을 임의의 축 방향으로 절단하고 이 절단한 것을 전개하여 2차원 행렬로 만든 후 PCA를 수행하는 방법이다. MPCA는 전체 공정변수 궤적이 존재하는 경우에 적용하는 모형화 방법이며, 배치가 진행 중인 상황에서 주기적으로 측정되는 공정변수들을 이용한 온라인 모니터링의 목적으로 사용하기는 어렵다.

2. 배치 공정의 온라인 모니터링

본 연구에서는 3차원 배열 $\mathbf{X}(I \times J \times K)$ 를 그림 2와 같이 MPCA와 동일한 방법으로 전개한 행렬 $\mathbf{X}(I \times JK)$ 를 온라인 모니터링의 참조 데이터베이스로 사용한다. 시점 k 에서 측정된 J 개의 공정변수들의 자료들로 구성된 행렬을 \mathbf{X}_k 라 하면, 참조 데이터베이스 \mathbf{X} 는 그림 2와 같은 구조를 갖는다.

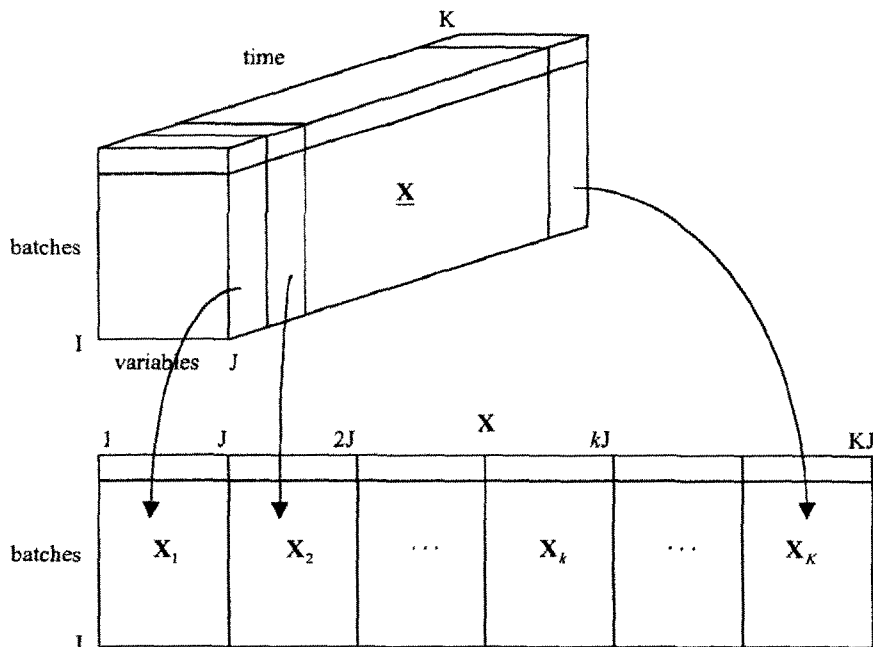


그림 2. 3차원 배열의 전개방법

2.1 시점별 참조분포

배치 공정의 온라인 모니터링은 투입된 배치의 생산이 완료되기 전에 시점 k 에서 다음의 가설을 검정하는 것이다.

H_0 : 현 시점에서 측정된 공정변수들의 관측치가 참조분포에 의해 정의되는 정상적인 배치 진행과 일치한다.

이러한 가설검정을 시점 k 에서 얻어진 측정치 \mathbf{x}_k 를 사용하여 순차적으로 수행하려면 배치의 시작 시점에서부터 각 시점별로 정상적인 공정상태를 나타내는 참조분포를 결정하여야 한다. 따라서 온라인 모니터링을 위해서는 시점별로 모두 K 개의 시점별 참조분포를 수립하여야 한다.

현재 시점 k 에 있을 때 온라인 모니터링에 사용할 수 있는 자료는 시점 k 까지의 자료뿐이므로, 시점 k 에서의 참조분포는 시점 1에서부터 시점 k 까지의 참조 데이터베이스 부분행렬 $\mathbf{X}_{1,k}$ 만을 사용하여 수립하여야 한다. 본 연구에서는 시점 k 에서의 참조분포를 수립하기 위해 사용하는 참조 데이터베이스의 범위에 따라 다음의 3가지 방법을 제안한다. 모든 방법에서 시점별 참조분포를 수립하기 위해 PCA를 사용하며, 사용하는 참조 데이터베이스의 범위만 차이가 있다. 각 방법에서 사용하는 참조 데이터베이스의 범위는 그림 3과 같다.

(1) 시점 k 의 자료만 사용하는 방법

이 방법은 k 시점의 참조 데이터베이스 행렬 \mathbf{X}_k 만을 사용하여 시점 k 에서의 참조분포를 수립하는 방법이다. 이것은 시점 k 에서의 공정변수 \mathbf{x}_k 가 이전 공정변수 $\mathbf{x}_{k-1}, \mathbf{x}_{k-2}, \dots, \mathbf{x}_1$ 과 독립이라고 가정하고 K 개의 PCA 모형을 수립하는 것으로 생각할 수 있으며, 이 방법을 시점별 PCA라 한다.

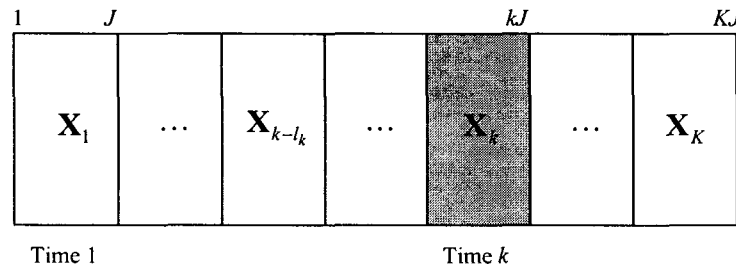
시점별 PCA는 개념적으로 간단하며, 쉽게 시점별 참조분포를 수립할 수 있는 장점이 있다. 그러나, 배치 공정에서는 현재 시점에서의 관측치가 이전 시점의 관측치와 자기상관관계를 갖는 경우가 대부분이므로, 측정시간간격이 충분히 길어서 이전 시점에 독립이라고 가정할 수 있는 경우가 아니면 배치 공정의 시점별 참조분포를 수립하는데 사용할 수 없다.

(2) 시점 1부터 시점 k 까지의 자료를 모두 사용하는 방법

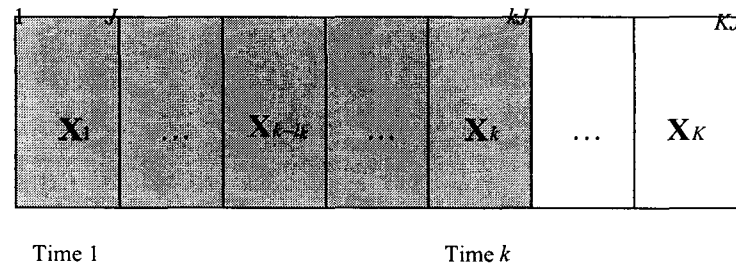
이 방법은 시작 시점에서부터 현 시점 k 까지의 참조 데이터베이스 행렬 $\mathbf{X}_{1,k}$ 를 사용하여 시점 k 에서의 참조분포를 수립하는 방법이다. 이것은 현 시점 k 를 배치의 종료시점으로 생각하고 K 개의 MPCA 모형을 수립하는 것으로 생각할 수 있으며, 이 방법을 시점별 MPCA라 한다.

시점별 MPCA는 시점간의 자기상관관계를 고려하여 시점 k 의 참조분포를 수립하므로, 공정 변화에 민감하게 반응하는 장점이 있다. 시점별 MPCA는 시점 k 에서의 공

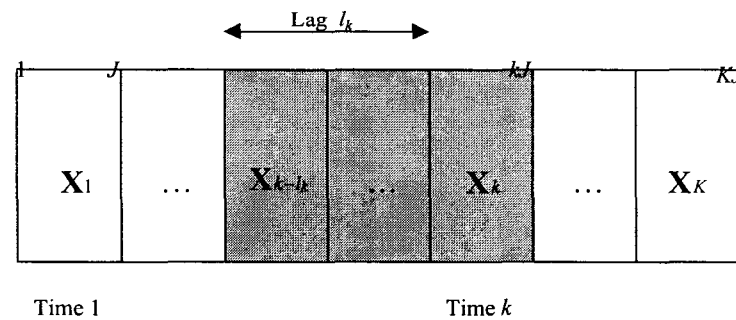
정 상태를 판정하기 위해 배치의 시작에서부터 현 시점까지 공정변수들의 모든 궤적을 사용한다. 따라서 시점 k 에서 이상신호가 발생하면, 배치 시작 시점에서부터 시점 k 사이에서 이상이 발생했음을 의미하며, 이상발생시점을 정확하게 알려줄 수 없는 단점이 있다. 또한 온라인 모니터링 단계에서 사용하기 위해 저장해야 하는 정보가 전체 측정시점의 수 K 에 따라 급속히 증가하므로 K 가 작은 경우에만 사용할 수 있다.



(a) 시점별 PCA



(b) 시점별 MPCA



(c) Lagged PCA

그림 3. 시점 k 의 참조 데이터베이스

(3) 시점 $(k-l_k)$ 부터 시점 k 까지의 자료를 사용하는 방법

이 방법은 시점 $(k-l_k)$ 에서부터 현 시점 k 까지의 참조 데이터베이스 행렬 $\mathbf{X}_{k-l_k, k}$ 를 사용하여 시점 k 에서의 참조분포를 수립하는 방법이다. Lag l_k 는 참조 데이터베이스에 포함되는 이전 시점의 수로 정의하며, 시점 k 에 따라 달라질 수 있다.

이것은 본 연구에서 제안하는 방법으로 lagged PCA라 한다.

Lagged PCA는 시점 k 의 참조분포를 수립할 때 시점 k 와 자기상관관계를 갖는 l_k 개의 이전 시점을 포함해서 PCA를 수행하는 방법이다. Lagged PCA는 시점간의 자기상관관계를 고려하므로 시점별 PCA보다 공정 변화에 더 민감하게 반응하는 장점이 있다. 또한 lag l_k 가 작은 값을 갖는 경우에는 시점별 MPCA보다 수행시간과 기억장소가 더 적게 소요되는 장점이 있다.

시점별 PCA와 시점별 MPCA는 lagged PCA의 특수한 경우로 생각할 수 있다. Lagged PCA는 $l_k = 0$ 인 경우에는 시점별 PCA, $l_k = k-1$ 인 경우에는 시점별 MPCA와 동일하다. Lag l_k 는 \mathbf{x}_k 와 이전 시점의 공정변수 $\mathbf{x}_{k-1}, \mathbf{x}_{k-2}, \dots, \mathbf{x}_1$ 의 자기상관관계를 고려하여 결정한다.

2.2 관리한계선

배치 공정의 온라인 모니터링을 위해서 배치 진행 시점별로 lagged PCA를 이용하여 참조분포를 결정하고 새로운 배치들이 진행됨에 따라 시점 k 에서 측정된 관측치들을 이용하여 T_k^2 통계량과 Q_k 통계량을 계산한다. 이러한 통계량을 각각 T_A^2 관리도와 Q 관리도에 배치의 완료시점까지 연속적으로 타점하면서 배치가 정상적으로 진행되고 있는지를 온라인으로 모니터링한다.

(1) T_A^2 관리도의 관리한계선

시점 k 에서 계산된 주성분 점수 \mathbf{t}_k 는 공정변수들의 선형결합으로 표현되며, 중심극한정리(central limit theorem)에 의해 정규분포에 근사한다. 정규분포를 따르는 주성분 점수들을 이용하여 계산된 통계량 T_k^2 은 다음과 같은 분포를 따르는 것으로 알려져 있다[12].

$$T_k^2 \frac{I(I-A_k)}{A_k(I^2-1)} \sim F_{A_k, I-A_k} \quad (2)$$

여기서 A_k 는 시점 k 에서 lagged PCA를 수행한 결과 보유하게 된 주성분의 수이다. 따라서 시점 k 에서의 T_A^2 관리도의 관리한계선은 다음의 식 3과 같다.

$$UCL_{T_k^2} = \frac{A_k(I^2-1)}{I(I-A_k)} F_{A_k, I-A_k} \quad (3)$$

(2) Q 관리도의 관리한계선

Q 관리도에 사용되는 Q_k 통계량은 예측오차들의 이차형(quadratic form)으로 표현된다. 이러한 예측오차 \mathbf{e}_k 는 참조분포가 정확하게 결정된 경우에 근사적으로 다변량 정규분포를 따르는 것으로 생각할 수 있다. Jackson과 Mudholkar[3]는 다변량 정규분포를 따르는 변수들의 이차형으로 표현되는 통계량의 근사 분포를 유도하였다.

시점 k 에서 Q 관리도의 관리한계선은 다음의 식 4와 같다.

$$UCL_{Q_k} = \theta_1 [1 - \theta_2 h_0 (1 - h_0) / \theta_1^2 + z_\alpha (2 \theta_2 h_0^2)^{1/2} / \theta_1]^{1/h_0} \quad (4)$$

여기서 $\theta_1 = \sum \lambda_j$, $\theta_2 = \sum \lambda_j^2$, $\theta_3 = \sum \lambda_j^3$, $h_0 = 1 - 2\theta_1\theta_3/3\theta_2^2$ 이며, z_α 는 h_0 와 동일한 부호를 갖는 표준정규분포의 $(1-\alpha)$ 분위수이다. 식 4에서 $\theta_i, i=1,2,3$ 는 예측오차 e_k 의 고유값 $\lambda_j, j=1, \dots, (l_k+1)J$ 의 함수로 표현되어 있다.

2.3 배치 공정의 온라인 모니터링 절차

Lagged PCA 관리도는 다변량 관리도를 설정하는 모형화 단계와 실제 공정을 모니터링하는 온라인 모니터링 단계를 통해 배치 공정을 온라인 모니터링한다. Lagged PCA를 이용한 배치 공정의 온라인 모니터링 절차를 정리하면 다음과 같다.

(1) 모형화 단계

- ① 공정이 관리상태에 있을 때 I 개의 공정변수 자료를 수집하여 참조 데이터베이스 행렬 \mathbf{X} 를 구성한다.
- ② 공정평균 벡터 $\bar{\mathbf{x}}$ 와 표준편차 벡터 \mathbf{s} 를 계산하고, 참조 데이터베이스 행렬 \mathbf{X} 를 표준화한다.
- ③ 각 시점 k 의 lag l_k 를 결정한다.
- ④ 시점별로 lagged PCA를 수행하여 보유하는 주성분의 수 A_k 를 결정한다.
- ⑤ 각 시점별로 loading $\mathbf{p}_{k,a}$ 와 고유값 $\lambda_{k,a}, k=1, \dots, K, a=1, \dots, A_k$ 를 구한다.
- ⑥ 각 시점별로 T_A^2 관리도와 Q 관리도의 관리한계선을 설정한다.

(2) 온라인 모니터링 단계

새로 투입된 배치가 진행함에 따라 시점 k 의 관측벡터 \mathbf{x}_k 가 수집되면, 시점 k 의 평균 벡터 $\bar{\mathbf{x}}_k$ 와 표준편차 벡터 \mathbf{s}_k 를 이용하여 \mathbf{x}_k 를 표준화한다. 표준화된 관측 벡터 $\mathbf{x}_{k-l_k, k}$ 를 사용하여 다음과 같은 순서로 시점 k 에서의 T_k^2 통계량과 Q_k 통계량을 계산한다.

$$t_{k,a} = \mathbf{p}_{k,a}^T \mathbf{x}_{k-l_k, k} \quad a = 1, 2, \dots, A_k \quad (5)$$

$$\hat{\mathbf{x}}_k = \sum_{a=1}^{A_k} t_{k,a} \mathbf{p}_{k,a} \quad (6)$$

$$\mathbf{e}_k = \mathbf{x}_{k-l_k, k} - \hat{\mathbf{x}}_k \quad (7)$$

$$T_k^2 = \sum_{a=1}^{A_k} \frac{t_{k,a}^2}{\lambda_{k,a}} \quad (8)$$

$$Q_k = \mathbf{e}_k^T \mathbf{e}_k \quad (9)$$

식 8의 T_k^2 통계량과 식 9의 Q_k 통계량을 각각 T_A^2 관리도와 Q 관리도에 타점하고, 관리한계선을 벗어나면 시점 k 에서 공정 이상이 발생한 것으로 판정한다.

3. 결론

본 연구에서는 대량의 공정변수 자료들이 온라인으로 수집되는 경우에 배치 공정을 실시간으로 온라인 모니터링할 수 있는 다변량 관리도를 제안한다. 다변량 관리도의 설정을 위해 관리상태에 있는 공정을 대변하는 통계적 모형을 각 시점별로 수립하는 시점별 모형화 방법을 제안한다. 이러한 시점별 모형화 방법을 사용하면 기존의 연구에서 경험적인 방법에 의존했던 관리한계선을 정확하게 설정할 수 있는 장점이 있다. 또한 본 연구에서 제안한 lagged PCA는 변수들간의 상관관계뿐만 아니라 시간적인 상관관계도 함께 고려할 수 있어 배치 공정의 온라인 모니터링에 적합한 방법이다.

참고문헌

- [1] Geladi, P. and B. R. Kowalski, "Partial Least Squares Regression: A Tutorial," *Analytica Chimica Acta*, Vol. 185, pp. 1-17, 1986.
- [2] Jackson, J. E., *A User's Guide to Principal Components*, John Wiley & Sons, Inc., New York, 1991.
- [3] Jackson, J. E. and G. S. Mudholkar, "Control Procedures for Residuals Associated With Principal Component Analysis," *Technometrics*, Vol. 21, No. 3, pp. 341-349, 1979.
- [4] Kourti, T. and J. F. MacGregor, "Process Analysis, Monitoring and Diagnosis, using Multivariate Projection Methods," *Chemometrics and Intelligent Laboratory Systems*, Vol. 28, pp. 3-21, 1995.
- [5] Kourti, T. and J. F. MacGregor, "Multivariate SPC Methods for Process and Product Monitoring," *Journal of Quality Technology*, Vol. 28, pp. 409-428, 1996.
- [6] Kresta, J., J. F. MacGregor, and T. E. Marlin, "Multivariate Statistical Monitoring of Process Operating Performance," *Canadian Journal of Chemical Engineering*, Vol. 69, pp. 35-47, 1991.
- [7] Ku, W., R. H. Storer, and C. Georgakis, "Disturbance Detection and Isolation by Dynamic Principal Component Analysis," *Chemometrics and Intelligent Laboratory Systems*, Vol. 30, pp. 179-196, 1995.
- [8] MacGregor, J. F. and T. Kourti, "Statistical Process Control of Multivariate Processes," *Control Engineering Practice*, Vol. 3, No. 3, pp. 403-414, 1995.
- [9] Nomikos, P. and J. F. MacGregor, "Monitoring of Batch Process Using Multiway Principal Component Analysis," *Journal of the American Institute of*

- Chemical Engineers*, Vol. 40, pp. 1361-1375, 1994.
- [10] Nomikos, P. and J. F. MacGregor, "Multivariate SPC Charts for Monitoring Batch Processes," *Technometrics*, Vol. 37, No. 1, pp. 41-59, 1995.
- [11] Russell, E. L., L. H. Chiang, and R. D. Braatz, "Fault Detection in Industrial Processes using Canonical Variate Analysis and Dynamic Principal Component Analysis," *Chemometrics and Intelligent Laboratory Systems*, Vol. 51, pp. 81-93, 2000.
- [12] Tracy, N. D., J. C. Young, and R. L. Mason, "Multivariate Control Charts for Individual Observations," *Journal of Quality Technology*, Vol. 24, pp. 88-95, 1992.
- [13] Wold, S., K. Esbensen, and P. Geladi, "Principal Component Analysis," *Chemometrics and Intelligent Laboratory Systems*, Vol. 2, pp. 37-52, 1987.