

# 배치평균을 이용한 빅데이터 시대의 관리도 운용 방법

송유진<sup>1</sup> · 주혜진<sup>1</sup> · 동승훈<sup>2</sup> · 변재현<sup>1\*</sup>

<sup>1</sup>경상국립대학교 산업시스템공학부 / <sup>2</sup>삼성전자공과대학교

## Implementing Batch Means Control Charts for Manufacturing Big Data

Yujin Song<sup>1</sup> · Hyejin Ju<sup>1</sup> · Seung Hoon Tong<sup>2</sup> · Jai-Hyun Byun<sup>1</sup>

<sup>1</sup>Department of Industrial and Systems Engineering, Gyeongsang National University

<sup>2</sup>Samsung Institute of Technology

Control chart using big data collected from sensors can detect small shift very effectively. However, applying the Shewart chart directly to these data leads to many false alarms, since the process big data is auto-correlated. This paper presents a method to construct batch means control charts that can be easily applied to process big data with autocorrelation. Through a simulation study, this paper presents best control chart plans according to the degree of autocorrelation in terms the number of observations spaced between batches and batch size. The applicability of the results of this study was confirmed by a practice case study of acceleration data using a 'physics toolbox' application on a smartphone. Opinions on further big data control chart education are also presented.

**Keywords:** Control Chart, Big Data, Autocorrelation, Batch Means, Average Run Length

### 1. 서론

관리도는 공정이 진행되는 과정에서 수집한 공정변수 데이터를 통계적으로 분석하여 공정이 관리상태에 있는지 판단하고, 공정이 이상상태로 돌입하면 가능한 한 신속하게 이를 탐지하는 기법이다. 슈워트(Shewhart)가 1924년에 관리도를 개발하였을 때는 컴퓨터는 물론 전자계산기도 없는 시대였다. 그래서 한 부분군에서 5개의 샘플을 얻어 공정의 중심위치는 평균을, 산포는 손으로 계산하기 힘든 표준편차 대신 범위를 이용하여 공정의 관리상태를 판정하는 방법을 제시하였다. 슈워트 관리도는 지금도 유용하게 산업현장에서 활용되고 있다. 과거에는 공정 데이터를 측정하는 데 시간과 비용이 많이 들어 공정이 진행되는 과정에서 간헐적으로 5개 이하의 샘플을 취하여 관리도를 운영하였다. 그런데 지금은 센서 등 수집 기술이

발달하여 추가 비용 없이 데이터를 1초 또는 0.1초 또는 그보다 짧은 시간에 한 개씩 자동으로 수집하여 저장할 수 있다. 문제는 대량으로 수집되는 데이터의 개별 관측치가 자기상관 (autocorrelation)을 가지게 되어 슈워트 관리도 적용의 기본가정인 관측치 간 독립성을 위배한다는 것이다. 빅데이터 시대에 효과적인 공정관리를 위해서는 실제 현장에서 나타나는 자기상관성이 높은 대용량 공정변수 데이터에 기반하여 관리도 교육도 변화되어야 한다.

자기상관을 가지는 데이터에 대한 관리도를 작성하기 위해, VanBrackle and Reynolds(1997)는 적합된 모형에 기반하여 관리한계선을 조정하는 방법을, Montgomery and Mastrangelo (1991)는 데이터를 모형에 적합한 다음에 구한 잔차의 지수가중이동평균(Exponentially Weighted Moving Average: EWMA)을 이용하여 관리도를 작성하는 방법을 제시했다. 이후에

본 논문은 정부(산업통상자원부)의 재원으로 한국에너지기술연구원 지원을 받아 수행된 자원순환(제제조) 산업 고도화 인력양성사업(No. 20214000000520)의 연구 결과입니다.

\* 연락처 : 변재현 교수, 52828 경남 진주시 진주대로 501 경상국립대학교, Tel: 055-772-1692, Fax: 055-772-1699, E-mail: jbyun@gnu.ac.kr  
2024년 12월 30일 접수; 2025년 1월 27일 수정본 접수; 2025년 2월 3일 게재 확정.

Montgomery(2019)는 표본추출간격을 늘리면 공정 데이터의 자기상관성을 어느 정도 해결할 수 있다고 밝혔다. Alwan and Radson(1992)은 일정한 간격을 두고 수집한 4~6개의 데이터를 하나의 배치로 묶어 배치의 평균을 이용하여 개별관리도를 작성하는 방법을 제시했다. 예를 들어, 초당 1개씩 데이터를 수집하는 공정에서 배치 간격  $w=10$ , 배치 크기  $b=4$ 이면, 1초에서 4초까지 수집한 4개 데이터의 평균을 첫 번째 타점, 5초부터 14초까지 데이터는 건너뛰고, 이후 15초에서 18초까지의 데이터 평균을 두 번째 타점으로 사용하는 것이다. Runger and Willemain(1996)은 관측값을 건너뛰어 표본을 추출하는 것보다는 배치 크기를 늘려 그 평균값을 이용한 개별관리도 작성 방법이 더 효과적이라고 주장하고, 이 방법을 화학제품을 생산하는 연속공정이나 자동 검사 등 데이터가 빈번하게 수집되는 공정에 적용했다. 최근에 Zwetsloot and Woodall(2021)은 배치평균을 이용한 방법이 프로세스 모니터링 성능에 미치는 영향에 관한 문헌을 검토하고 실제적인 조건과 연구 방향을 제안하였다. 그들은 우선 배치평균 관리도는 시계열 모델 없이 자기상관 데이터를 처리할 수 있는 실용적 기법이라는 것을 강조하고, 가공되지 않은 원시 공정 데이터를 활용할 것을 권장했다. 추후 연구 방향으로는 다변량 관리도로 확장, 시공간(spatiotemporal) 데이터와 이미지 데이터에 적용할 것을 제시했다.

본 논문에서는 자기상관을 가진 빅데이터를 이용한 배치평균 관리도를 운용하는 방법을 제안하고자 한다. 논문의 구성은 다음과 같다. 제2장에서는 부분군 크기에 따른 관리도의 성능과 공정 빅데이터가 자기상관을 가질 때 문제점을 제시하고, 이러한 자기상관 데이터를 이용하여 관리도를 작성하는 방법들을 제3장에 요약하였다. 제4장에서는 배치평균 관리도 활용하는 방법을, 제5장에는 배치평균 관리도 실습사례를 제시한다. 제6장에는 배치평균 관리도를 확장하기 위한 의견을 제안하고, 본 논문의 결론은 제7장에 기술한다.

## 2. 공정 데이터와 관리도

### 2.1 부분군 크기에 따른 관리도 성능

관리도는 데이터로부터 관리상한선(Upper Control Limit: UCL)과 관리하한선(Lower Control Limit: LCL)을 구하여 공정이 안정 상태에 있으면 타점이 UCL과 LCL 사이에 있고, 공정의 중심위치나 산포에 변화가 생기면 가능한 한 이른 시점에 UCL 위나 LCL 아래에 점이 찍히게 하여 관리상태를 벗어났다는 판정을 내리게 하는 기법이다. 중심위치의 관리를 위한  $\bar{X}$  관리도는 데이터가 서로 독립적이고 분산이 같은 정규분포를 따른다는 가정하에 작성하고, UCL과 LCL은 각각  $\bar{X}$  통계량의 평균에서  $3\sigma$ 를 더하고 뺀 값이다. 이때 공정 평균에 변화가 없음에도 불구하고  $\bar{X}$  값이 관리한계선을 벗어날 확률인

제1종 오류(생산자위험)는 0.0027이다.  $\bar{X}$  관리도의 성능을 평가하는 척도는 공정 평균이 어떤 값을 유지할 때 관리한계선을 이탈한 시점까지의 타점수의 평균값인 ‘평균 런 길이(Average Run Length: ARL)’인데, ARL은 검정력의 역수로 계산한다. 공정 평균이 변화 없이  $\mu_0$ 를 유지할 때 ARL인  $ARL_0$ 은  $1/0.0027 = 370$ 이다. 평균이  $\mu_1 = \mu_0 + k\sigma$ 로 이동하면, 검정력(power)은 식 (1)을 이용하여 구한다.

$$1 - P(LCL \leq \bar{X} \leq UCL \mid \mu = \mu_1) \tag{1}$$

$$= 1 - \Pr(-3 - k\sqrt{n} \leq Z \leq 3 - k\sqrt{n})$$

식 (1)에서  $n$ 은 부분군의 크기이고,  $Z$ 는 표준정규분포를 따르는 통계량이다.

평균이  $\mu_0 + k\sigma$ 로 이동하면, 부분군 크기  $n$ 에 따라 검정력이 크게 달라진다.  $k=1$ 이면  $n=5, 100$ 에 따라 (검정력, ARL)은 각각 (0.223, 4.5), (1.000, 1)이고,  $k=0.5$ 이면 (검정력, ARL)은  $n=5, 100$ 일 때 각각 (0.03, 33.3), (0.98, 1)로 차이가 아주 크다. 데이터가 서로 독립적이고 분산이 같은 정규분포를 따른다는 가정하에 평균이  $\mu_0 + 0.5\sigma$ 로 작게 이동하면, 부분군의 크기가 5일 때는 평균적으로 33번째 타점까지 기다려야 하지만, 100이라면 바로 다음 타점에서 이를 감지할 수 있다.

### 2.2 공정 빅데이터 자기상관성의 문제점

센서 등 자동화 수집장치를 이용하여 연속적으로 얻은 대량의 공정변수 측정데이터는 서로 독립적이지 않고 자기상관(autocorrelation)을 갖는다(Hahn, 1989). 자기상관을 가진 데이터에 슈윙트 관리도를 적용하면 잘못된 경보인 오경보(false alarm)를 자주 낼 수 있다(Montgomery, 2019).

우선 자기상관의 정도에 따른 오경보 문제를 살펴보기로 한다. <Figure 1>은  $N(80, 1^2)$ 을 따르고 서로 독립인 데이터 1,000개를 랜덤하게 생성하여 부분군 크기가 5인  $\bar{X}$  관리도를 작성한 것으로, 관리한계선을 벗어난 점이 없이 안정된 상태이다.

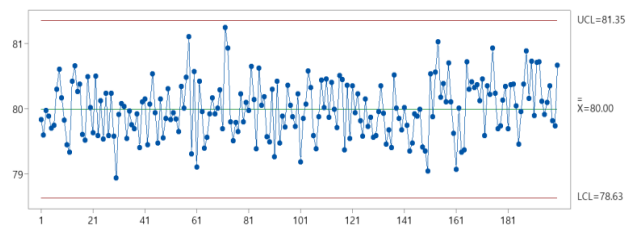


Figure 1.  $\bar{X}$  Control Chart with Independent Observations

<Figure 2>와 <Figure 3>은 각각 자기상관계수가 0.9, 0.25인 데이터를 이용하여 구성한  $\bar{X}$  관리도이다. 자기상관계수가 0.9일 때는 오경보가 아주 심하게 일어나고, 계수가 0.25일 때

도 무시할 수 없을 정도로 오경보가 나타난다.

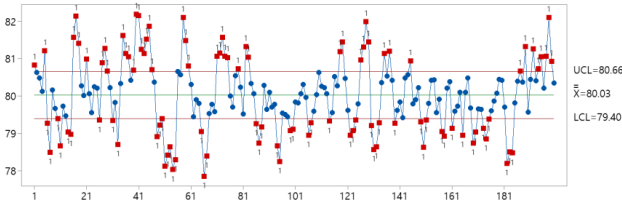


Figure 2.  $\bar{X}$  Control Chart with Observations of Autocorrelation Coefficient 0.9

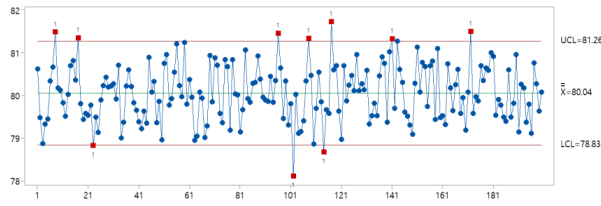


Figure 3.  $\bar{X}$  Control Chart with Observations of Autocorrelation Coefficient 0.25

### 3. 자기상관이 있는 데이터를 이용한 관리도

#### 3.1 자기상관

자동수집장치로 연속 측정되는 공정변수 데이터는 양의 자기상관관계를 갖는다(Hahn, 1989). 즉,  $t$  시점의 데이터는  $(t-1)$  시점의 영향을 받아서  $(t-1)$  시점에서 작은 값이 관측되면  $t$  시점에서도 작은 값이,  $(t-1)$  시점에서 큰 값이 관측되면  $t$  시점에서도 큰 값이 나타난다. 일반적으로 이전 시점 값이 다음 시점 값에 영향을 주는 현상인 자기상관의 정도를 평가하는 자기상관계수(autocorrelation coefficient)  $\phi$ 는 다음 식 (2)로 정의한다.

$$\phi = \frac{Cov(x_t, x_{t-1})}{V(x_t)} \quad (2)$$

식 (2)에서  $Cov(x_t, x_{t-1})$ 은 공분산이고  $V(x_t)$ 는 분산인데, 시점과 관계없이 일정한 값을 가진다고 가정한다.

제2장에서 기술한 바와 같이, 자기상관이 있는 데이터는 전통적인 관리도의 가정을 위배하기 때문에 기존의 관리도에 사용할 수 없다. 자기상관을 가지는 데이터를 다루는 방법에 관한 연구가 다양하게 진행되었는데, 크게 보면 모형 기반 방법과 모형 무관 방법으로 나눌 수 있다.

#### 3.2 모형 기반 방법

첫째, VanBrackle and Reynolds(1997)는 적합한 모형에 기반

하여 관리한계선을 조정하는 방법을 제안하였다. 적절한 시계열 모형, 자기상관계수, 그리고 공정의 표준편차를 알고 있는 경우, 관리한계 계수를 조정하면  $ARL_0$  값이 슈워트 관리도와 비슷하게 나타남을 보였다. 하지만 이 방법은 데이터 간에 자기상관이 크지 않을 때 사용하는 방법이고, 자기상관계수가 높으면 잘 작동하지 않는다(Lu and Reynolds, 1999).

둘째, 자기상관이 있는 데이터를 시계열 모형에 적합하고 예측값과 실제값의 차이인 잔차를 이용하는 것으로서, 평균이 0이고 서로 독립이며 정규분포를 따르는 잔차를 EWMA 관리도에 타점하는 방식으로 관리도를 그린다(Montgomery and Mastrangelo, 1991). 이 방법은 특히 관측치들이 양의 자기상관관계를 가지며 공정 평균의 변화가 클 때 잘 적용되지만, 자기상관이 높지 않고 공정 평균의 변화가 작을 때는 잘 작동하지 않는다(Wardell et al., 1994). 또한 이 방법은 다소 복잡하고, 시간이 오래 걸리며 시계열 모형을 구축하는 방법에 관한 이해가 필요하므로 통계적 방법에 익숙하지 않은 공정담당자가 사용하기에는 어려움이 있다.

#### 3.3 모형 무관 방법

첫 번째는 낮은 빈도로 데이터를 추출하는 가장 단순한 방법으로서, 공정 데이터에서 표본을 추출하는 간격을 비규칙적으로 늘리는 것이다. 간격을 늘리면 자기상관은 크게 줄어들게 되지만, 데이터 일부만 사용하는 것이기 때문에 다소 비효율적이고 모든 데이터를 사용할 때와 비교하면 공정 평균의 변화를 늦게 알아차릴 수 있다는 문제가 있다(Montgomery, 2019).

두 번째 방법은 연속적으로 수집되는 데이터를 배치(batch)로 구분하고, 각 배치의 평균을 이용하여 개별관리도를 작성하는 방식이다. 이는 자기상관이 있는 연속 데이터에 잘 적용할 수 있고 단순하다는 장점이 있다(Runger and Willemain, 1996).  $j$ 번째 배치평균은 식 (3)과 같이 정의한다.  $b$ 는 배치의 크기를 나타내며,  $w$ 는 배치 사이의 간격(또는 건너편 관측 횟수)이다. 자기상관이 있는 데이터의 배치 크기를 키우면 공정 평균의 변화를 신속하게 감지하는 데 유리하지만, 평균의 큰 변화를 감지하는 데에는 시간이 지체되므로, 배치 크기는 배치 평균 간 자기상관계수가 0.1 이하가 되는 최소한의 수로 정하는 것이 바람직하다.

$$\bar{x}_j = \frac{1}{b} \sum_{i=1}^b x_{(j-1)(w+b)+i}, j = 1, 2, 3, \dots \quad (3)$$

3.1절에서 언급했듯이 자기상관이 0.25만 되어도 관리도 상에 문제가 발생하고, 그 기능을 제대로 수행하지 못한다. Runger and Willemain(1996)은 배치평균 간 자기상관계수 값이 0.1 이하로 충분히 작아질 때까지 배치 크기를 두 배로 늘리는 방법을 제안했다. 예를 들어,  $w = 0$ 이고,  $\phi = 0.9$ 일 때, 자기

상관계수가 0.1 이하로 떨어지는 첫 배치( $b$ )의 크기는 60이다. 아래 <Table 1>은  $\phi = 0.9, b = 60, w = 0$ 일 때, 40,000개의 데이터를 생성하는 시뮬레이션을 5번 시행하여 구한  $ARL$  결과와 슈윙트 관리도의 이론적  $ARL$  값을 비교한 표이다. 배치평균 관리도의 값은 5번 시행한 값의 평균이고, 괄호에 있는 것은 표준오차(standard error: s.e.)이다.  $k$ 는 공정 평균의 변화폭을 결정하며, 표준편차( $\sigma$ )의 몇 배만큼 평균이 이동했는지 나타내는 값이다.

**Table 1.** Average Run Length of Batch Means Control Chart Compared to Shewart Chart

$k$	$ARL$	
	batch means chart (s.e.)	Shewart chart
0	374.49 (2.62)	370.44
0.5	152.49 (1.86)	155.22
1	43.11 (0.29)	43.89
2	6.21 (0.13)	6.30
3	1.98 (0.02)	2.00
4	1.18 (0.01)	1.19

위 <Table 1>에서 기존의 슈윙트 관리도와 배치평균 관리도의  $ARL$  값이 거의 같게 나타나서 자기상관이 있는 데이터를 이용한 배치평균 관리도의 탐지 능력이 자기상관이 없는 슈윙트 관리도와 비슷하다는 것을 알 수 있다.

#### 4. 배치평균을 이용한 관리도 제안

##### 4.1 연구수행 방법

Runger and Willemain(1996)은 자기상관계수가 충분히 감소할 때까지 배치 크기( $b$ )를 2배씩 늘리는 방법을 제안했지만, 배치 간격( $w$ )에 관한 기준은 제시하지 않았다. 본 연구는  $b$ 의 값 간격을 좀 더 촘촘히 하고,  $w$ 의 값도 적절한 수준으로 정하는 식으로 시뮬레이션 실험을 시행하여 자기상관계수의 값에 따라 최적의 ( $w, b$ ) 값을 제시하고자 한다. 각각의 ( $w, b$ ) 계획에 따라 시뮬레이션을 500번씩 시행하였다.

시뮬레이션 과정은 다음과 같다. 빅데이터의 자기상관계수( $\phi$ )는 0.5 이상으로 나타날 것으로 보고, 본 논문에서는  $\phi = 0.95, 0.9, 0.8, 0.7, 0.6, 0.5$ 에 대하여, 식 (4)를 사용해  $\phi$  값별로 80,000개의 데이터를 500세트 생성하였다.

$$X_t = \delta(1 - \phi) + \phi X_{t-1} + \epsilon_t \quad (4)$$

식 (4)에서  $\delta$ 는 미지의 상수,  $\epsilon_t$ 는 서로 독립이고  $N(0,1)$ 을 따른다고 가정한다. 분석에 사용할 데이터는 미니탭과 엑셀을 이용하여 생성하고, 주어진  $\phi$  별로 다양한 ( $w, b$ ) 값에 따라 데

이터셋을 만든 다음에 파이썬(Python)을 이용하여  $ARL$ 을 계산하였다.

우선  $w$ 는 0과 1에서 시작하여 20 이하까지 같은 간격으로 5씩 나누어 5, 10, 15, 20으로 정하였다.  $w$ 가 20을 넘어가면 평균 이동을 탐지하는 능력이 좋지 않음을 시뮬레이션을 통해 확인하고 나서 20보다 큰 간격은 고려하지 않았다. 주어진  $w$  별로 배치평균의 자기상관계수가 0.1 이하가 되는  $b$ 를 구하기 위해,  $b$  값을 1에서 시작하여 4를 더해가는 방식으로 간격을 조정하였다.

##### 4.2 수행 결과

공정 평균이 이동하지 않을 때  $ARL$  값인  $ARL_0$ 은 슈윙트 관리도의 370 이상이면서 이에 근접하고, 공정 평균에 변화가 있을 때  $ARL_1$ 은 가장 작게 나오는 ( $w, b$ ) 계획을 탐색했다. <Table 2>와 <Table 3>은 자기상관계수( $\phi$ )가 각각 0.95, 0.90일 때,  $w = 0, 1, 5, 10, 15, 20$ 에 대하여  $ARL_0$ 이 370 이상이면서 370에 가장 가깝고,  $ARL_1$ 은 최소값을 가지는  $b$  값과 배치평균의 자기상관계수( $\phi_1$ )를 정리하여 나타낸 것이다.

**Table 2.** Average Run Length of Batch Means Control Charts with  $\phi = 0.95$  for Different  $w$  and  $b$

$w$	$b$	$\phi_1$	$k$					
			0	0.5	1	1.5	2	3
1	177	0.058	370	154	43.2	14.7	6.19	1.97
5	189	0.088	373	155	43.5	14.8	6.23	1.98
10	145	0.060	375	155	43.5	14.8	6.23	1.98
10	153	0.082	372	154	43.3	14.7	6.20	1.98
10	157	0.060	375	155	43.6	14.8	6.24	1.98
20	125	0.094	371	154	43.4	14.8	6.22	1.98

**Table 3.** Average Run Length of Batch Means Control Charts with  $\phi = 0.9$  for Different  $w$  and  $b$

$w$	$b$	$\phi_1$	$k$					
			0	0.5	1	1.5	2	3
0	81	0.050	375	156	43.9	14.9	6.27	1.99
0	85	0.052	374	156	43.8	14.9	6.27	1.99
1	77	0.035	373	155	43.6	14.8	6.25	1.98
1	81	0.033	379	158	44.3	15.0	6.31	2.00
10	49	0.015	371	155	43.7	14.9	6.27	1.99
20	33	0.016	372	155	43.8	14.9	6.27	1.99

<Table 2>에서 (1, 177), (10, 153), (20, 125)의 탐지 성능이 차이가 거의 없지만,  $ARL_0$  값이 가장 작은 (1, 177) 계획을 최적 계획으로 선정한다. <Table 3>에서는 (1, 77), (10, 49), (20, 33)

계획이  $ARL_1$  측면에서는 거의 차이가 없지만,  $ARL_0$  측면에서는 (10, 49) 계획이 조금 더 낫다.

**Table 4.** Best  $(w, b)$  Plans with Respect to  $\phi$

$\phi$	w	b	$\phi_1$	$ARL_1$					$ARL_0$
				k					
				0.5	1	1.5	2	3	
0.95	1	177	0.058	154	43.2	14.7	6.19	1.97	370
0.90	10	49	0.015	155	43.7	14.9	6.27	1.99	371
0.80	20	9	-0.059	157	44.1	15.0	6.30	2.00	376
0.70	20	5	-0.002	163	45.5	15.4	6.43	2.02	392
0.60	1	9	0.054	157	44.2	15.0	6.32	2.00	375
0.50	0	9	0.085	166	46.3	15.6	6.52	2.04	400

**Table 5.** Recommended  $(w, b)$  Plans with Respect to  $\phi$

$\phi$	$(w, b)$	$\phi$	$(w, b)$	$\phi$	$(w, b)$
0.95	(0, 200)	0.90	(0, 90)	0.80	(0, 40)
	(1, 180)		(1, 90)		(1, 30)
0.70	(0, 20)	0.60	(0, 15)	0.50	(0, 10)
	(1, 20)		(1, 10)		(1, 10)

다양한 자기상관계수( $\phi$ ) 값에 대하여  $ARL_0$ 는 370에 가깝고  $ARL_1$ 은 작게 만드는 조건을 선정할 결과는 <Table 4>와 같다. 공정 데이터 간의 자기상관이 강한 경우, w의 간격은 짧고 b는 큰 것이 좋은 성능을 보였으며, 자기상관이 그리 강하지 않을 때는 w의 간격을 늘리는 것이 좋은 성능을 보였다. 문제는 배치 간격 w가 길면,  $ARL_1$ 이 우수하여도 배치와 배치 사이에서 발생하는 공정 이상을 간과할 수 있다는 것이다.  $\phi$ 가 0.8일 때, 최적의 배치 간격과 크기 조합은  $(w, b) = (20, 9)$ 이다. 20개 데이터는 건너뛰고, 9개를 묶은 배치의 평균을 이용한다. 이런 계획에서 건너뛴 20개 데이터에서 급격한 변화 (abrupt change)가 생기면 이것을 탐지할 수 없다는 결정적 문제가 생긴다. 이러한 문제를 극복하기 위하여 배치평균의 자기상관계수( $\phi_1$ )는 0.1 이하,  $ARL_0$ 는 370 이상이 되면서, 배치 간격 w는 0 또는 1, 배치 크기 b는 10단위이면서, 최적의  $(w, b)$  계획과 가장 가까운 추천 계획을 구하여 <Table 5>에 제시한다.  $\phi = 0.6$ 의 경우에는 (0, 10) 계획을 쓰면 배치평균 자기상관계수  $\phi_1$ 가 0.1보다 크게 나와서 (0, 15) 계획으로 대신한다.

추천된  $(w, b)$  계획의 탐지 능력을 최적 계획과 비교한 것을 <Table 6>에 제시하였다. 주어진  $\phi$ 에 대하여 추천된 계획 중, b가 같을 때는 w가 0인 것이 1인 것에 비해 탐지 성능이 좋으

**Table 6.** Comparison between Best and Recommended  $(w, b)$  Plans with Respect to  $\phi$

$\phi$	Best vs. Recommended	$(w, b)$	$ARL_1$					$ARL_0$	$\phi_1$
			k						
			0.5	1	1.5	2	3		
0.95	Best	(1, 177)	154	43.2	14.7	6.19	1.97	370	0.058
	Recommended	(0, 200)	163	45.6	15.4	6.43	2.02	395	0.036
		(1, 180)	158	44.3	15.1	6.32	2.00	379	0.037
0.90	Best	(10, 49)	155	43.7	14.9	6.27	1.99	371	0.015
	Recommended	(0, 90)	167	46.6	15.7	6.54	2.03	404	0.072
		(1, 90)	175	48.4	16.2	6.70	2.06	426	0.095
0.80	Best	(20, 9)	157	44.1	15.0	6.30	2.00	376	-0.059
	Recommended	(0, 40)	168	46.9	15.8	6.57	2.04	407	0.081
		(1, 30)	165	43.0	15.6	6.51	2.03	398	0.067
0.70	Best	(20, 5)	163	45.5	15.4	6.43	2.02	392	-0.002
	Recommended	(0, 20)	167	46.8	15.8	6.56	2.04	405	0.078
		(1, 20)	186	51.1	16.9	6.94	2.10	456	0.058
0.60	Best	(1, 9)	157	44.2	15.0	6.32	2.00	375	0.054
	Recommended	(0, 15)	184	50.7	16.8	6.91	2.09	452	0.090
		(1, 10)	172	40.6	16.1	6.66	2.05	417	0.082
0.50	Best	(0, 9)	166	46.3	15.6	6.52	2.04	400	0.085
	Recommended	(0, 10)	181	50.0	16.6	6.85	2.09	443	0.035
		(1, 10)	214	57.5	18.7	7.51	2.19	537	0.073

며,  $w$ 가 다를 때는  $b$ 가 작으면서  $w$ 가 1인 것의 성능이 더 좋다. <Table 6>을 보면, 주어진  $\phi$ 별로 두 개의 추천 계획 중 성능이 좋은 것을 택하면 최적 계획과 큰 차이가 없다는 것을 알 수 있다. 가장 크게 성능 차이가 나타나는 것은  $\phi$ 가 0.6이고 평균이  $\mu_0$ 에서  $\mu_0 + 0.5\sigma$ 로 이동했을 때 나타난다. 추천 계획의  $ARL_1$ 이 172로서 최적 계획의 157보다 15만큼 늦게 탐지하게 되어 9.6% 정도 늘어나는 것이다. 하지만  $ARL_0$ 는 375에서 417로 늘어나기 때문에 제1종 오류는 줄어든다.

$\phi$ 별로 <Table 5>에 제시된 2개의 방법 중, 탐지 성능이 조금이라도 좋은 것은 굵은 글씨로 표시하였다. 활용 방법을 제시 하자면 다음과 같다. 1) 예비 데이터를 구하여 공정빅데이터의 자기상관계수  $\phi$ 를 계산한다. 2) 계산된  $\phi$  값과 가장 가까운 것에 맞는  $(w, b)$  계획을 <Table 5>로부터 구하여 적용하는데, 굵은 글씨로 표시된 계획을 우선하여 쓴다.

### 5. 배치평균 관리도 실습사례

본 논문에서 제시한 배치평균 관리도가 실제로 적용되는지 확인하기 위해 스마트폰 앱에 있는 센서로 자동차의 흔들림 데이터를 측정하여 실습한 내용을 소개한다. 주행 시 차량이 좌우로 흔들리는 현상은 차체 하부 부품의 마모나 느슨함, 타이어의 부적절한 공기압, 또는 휠 얼라인먼트의 불량 등 여러 요인에 의해 발생할 수 있다. 이러한 좌우 흔들림은 주행 중 운전자의 불안감을 높이고 차량의 안전성에 직접적인 영향을 미칠 수 있다. 또한, 이러한 문제들은 초기에는 큰 영향을 미치지 않을 수 있으나, 계속해서 발생하면 사고로 이어질 가능성이 커지므로 사전에 점검하고 수리하는 것이 필요하다. 본 실습사례는 차량이 저속으로 주행할 때의 좌우 흔들림 정도를 ‘가속도’ 물리량으로 측정하고 배치평균 관리도에 근거하여 관리상태 여부를 평가하고자 한다.

실험 대상은 8년 전 출고된 스포츠 유틸리티 차량(SUV)이며, 가속도를 측정하기 위해 스마트폰에서 이용할 수 있는 ‘physics toolbox’를 활용했다. 이 애플리케이션은 중력 센서,

가속도 센서 등을 이용하여 중력, 가속도 등 다양한 물리량을 측정하는 데 쓰인다. 측정된 데이터는 시간이 지남에 따라 그래프와 수치로 기록되며, .csv 파일 형식으로 저장되므로 빅데이터를 이용한 배치평균 관리도 작성에 적합하다고 판단했다. 본 사례에서는 진동이나 충격 등에 의해 발생하는 가속도를 측정하기 위해 이 애플리케이션에 있는 선형 가속도계(linear accelerometer)를 사용했다.

대학교 구내에서 시속 25km로 주행하면서 1초마다 측정된 가속도를 60분 동안 수집하여 총 3,600개의 데이터를 얻었다. <Table 7>은 수집한 데이터 3,600개 중 100개만 추출하여 나타낸 표이다. 가속도 단위는  $m/sec^2$ 이다.

수집된 데이터를  $\bar{X} - R$  관리도(부분군 5)로 작성하면 다음의 <Figure 4>와 같다.

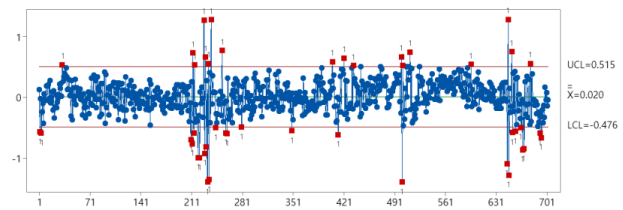


Figure 4.  $\bar{X}$  Control Chart of the Acceleration Data from Practice Case Study

데이터의 자기상관계수( $\phi$ )를 계산하니 0.65이다. 이렇게 자기상관이 높은 데이터를 이용해  $\bar{X}$  관리도를 작성하니 많은 오경보가 발생했다. <Table 6>을 보면,  $\phi$  값이 0.65보다 크면 가장 가까운 0.7일 때 최적의 배치평균 관리도 계획은  $(w, b) = (20, 5)$ 이고, 추천 계획은  $(w, b) = (0, 20), (1, 20)$ 이다. 이 계획을 적용한 배치평균 관리도를 <Figure 5>에 나타내었는데, 차량의 좌우 흔들림을 나타내는 가속도 물리량은 관리상태에 있음을 알 수 있다.

본 실습사례는 가속도 센서를 이용한 것이다. 스마트폰이나 다른 기기에서 센서를 통하여 빅데이터를 얻을 수 있으면, 우선 초기 관측 데이터를 이용하여 자기상관계수를 구하고,

Table 7. One Hundred Acceleration Data for Practice Case Study

-0.381	0.283	0.395	0.398	-0.025	-0.461	-0.655	-0.799	-0.583	-0.318
-0.318	-0.140	0.161	0.403	0.001	-0.518	-0.989	-1.104	-0.515	-0.255
-0.064	0.245	-0.045	-0.290	-0.435	-0.619	-0.520	-0.087	0.284	-0.068
-0.104	-0.175	-0.447	-0.360	-0.579	-0.569	-0.020	0.643	0.212	-0.205
-0.331	-0.536	-0.180	-0.002	0.159	0.388	0.146	-0.155	-0.045	-0.045
-0.124	0.002	0.443	0.139	0.066	0.317	-0.097	-0.203	0.228	-0.188
0.053	0.745	0.426	-0.178	0.093	-0.139	-0.135	0.433	0.313	-0.003
0.471	0.559	0.017	-0.371	-0.337	0.068	0.388	0.419	0.365	0.047
0.200	0.150	-0.217	-0.310	-0.198	0.171	0.314	0.395	-0.020	-0.478
-0.180	-0.177	-0.181	0.111	0.258	0.454	0.271	-0.051	-0.550	-0.548

<Table 5>으로부터 그에 맞는 배치평균 관리도 작성계획( $w, b$ )를 찾아서 적용하는 식으로 다양한 실습을 할 수 있겠다.

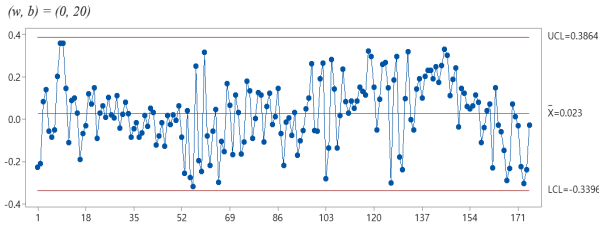


Figure 5. Batch Means Control Chart of the Acceleration Data from Practice Case Study

### 6. 빅데이터를 이용한 관리도 교육 확장을 위한 의견

Byun *et al.* (2024)은 대한산업공학회지(JKIE)와 미국산업및시스템공학회(Institute of Industrial and Systems Engineers: IISE)의 학회지인 IISE Transactions에 지난 20년간 출판된 품질 분야 연구논문을 조사한 결과, 관리도 논문이 JKIE 5편, IISE Transactions 88편임을 알아내었다. 관리도는 양산공정의 품질 관리와 개선을 위한 중요한 기법인데, 빅데이터 시대에 산업현장에 필요한 국내 관리도 연구가 아주 미흡함을 알 수 있다. 본 장에서는 학생들이 산업현장에 적응하는 데 도움이 될 수 있는 확장된 관리도 교육 내용을 제안하고자 한다.

첫째, 본 논문은 공정변수가 일정한 값을 유지해야 하는 데 필요한 배치평균 관리도 작성 방법을 제시하였다. 하지만 공정빅데이터를 수집하는 반도체나 화학 공정에서 공정변수가 구간별로 다른 값을 가져야 하는 경우가 있다. 이때에는  $i$ 번째 구간의  $j$ 번째 관측값( $O_{ij}$ )에서 구간별 목표값( $T_i$ )을 빼 것( $D_{ij}$ )을 이용하여 배치평균 관리도를 이용할 수 있다. 이때 공정변수의 값은 목표값을 유지하는 것이 바람직하므로  $D_{ij}$ 의 중심선(center line)은 0이다.

둘째, 공정변수의 값이 프로파일(profile) 형태로 나타나는 경우이다. Qiu(2017)는 품질특성이 선형 또는 비선형의 프로파일 형태로 나타날 때 관리도 활용 방법을 제시했다. 하지만 본 연구는 품질특성이 아니고 공정변수의 관리도를 작성하는 것을 다루고자 했다. 간단하게 공정변수가 선형 프로파일을 따르는 문제를 보자. 예를 들어, 온도를 처음에는 선형으로 올리고, 그다음 일정한 시간 동안 온도를 유지하며, 나중에는 온도를 서서히 선형으로 내리는 것이다. 온도를 올리거나 내릴 때 공정변수의 이상적인 프로파일이 선형이라고 가정하고, 온도를 올리는 구간을 보기로 한다.  $j$ 번째 시간( $x_j$ )에서 측정된 온도( $y_j$ ) 데이터를 생각하면  $N$ 개의 ( $x_j, y_j$ ) 데이터를 수집하게 된다. 이 데이터를 이용하여 선형 회귀식  $\hat{y}_j = a + bx_j$  을 구하면,  $j$ 번째 시점의 잔차(residual)는  $r_j = \hat{y}_j - y_j$ 이다. 그러면 잔차  $r_j$ 를 대상으로 배치평균 관리도를 작성할 수 있다. 다

만, 산업현장에서는 이러한 온도 상승구간이 종료되는 시점 이후에도 바로 일정한 온도를 유지하지 않고, 비선형적으로 증감 진폭을 가지면서 어느 정도 시간이 흐른 다음에 일정한 온도로 수렴한다는 것을 염두에 둘 필요가 있다.

셋째, 대규모 설비를 활용하고 있는 반도체산업 등에 있어서는 공정이 매일 열화(degradation) 현상 등으로 인하여 조금씩 변화한다. 이러한 현상에 대응하기 위하여 사전에 정한 주기와 조건에 따라 설비의 주요 변수를 조정하게 되는데, 이때 조정에 의한 공정변수 값의 변화는 감내할 수 있는 정상적 이동이라고 볼 수 있다. 이런 정상적 변화에 의해서는 경보가 일어나지 않도록 관리도를 운용하는 현실적 방법이 필요하다.

마지막으로, 반도체 생산공정은 다단계 공정으로서 단계 수는 크게 나누어도 수백 개 이상이고, 공정변수는 수천 또는 수만 개에 달하며, 한 설비에 부착된 센서의 수도 수백 개에 이른다. <Figure 6>은 다단계 공정에서 여러 변수 간 상호상관(cross-correlation)이 있고, 각 변수는 시간의 흐름에 따라 자기상관을 갖는 데이터를 나타낸 그림이다(Ding *et al.*, 2002). 우선 이렇게 많은 공정변수 중 중간 또는 최종 제품의 주요 품질 특성에 영향을 미치는 것들을 파악해야 한다. 이후에는 품질 특성에 문제가 생길 때, 어떤 공정변수의 관리도에 관리상태를 벗어났다는 경보가 발생했는지를 추적해야 한다. 실질적인 관리도 교육을 위하여, 여러 단계에서 몇 개의 변수를 대상으로 실제보다 단순한 상황을 가정하여 시스템을 구성한 다음 시뮬레이션 실습을 할 수 있게 하면, 학생들의 현장 응용력을 높일 수 있다고 생각한다. 우선 간단하게 3개의 단계, 각 단계에서 3개의 변수를 대상으로 각 단계의 변수 간에는 상호상관이 있고, 각 변수는 자기상관을 갖는 공정 빅데이터를 생성하여 학생들이 실습할 수 있게 하는 것을 제안한다.

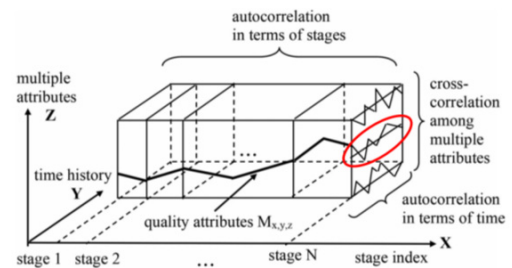


Figure 6. A Diagram of the Multi-stage, Multiple Attributes Process with Complex Data Relationships (Adapted from Ding *et al.*, 2002)

### 7. 결론

본 논문은 자기상관을 가진 공정 빅데이터에 쉽게 적용할 수 있는 배치평균 관리도를 구성하여 운용하는 방법을 제시한다. 관리도는 공정이 진행되는 과정에서 공정변수 데이터를 측정하고 통계적으로 분석하여 공정이 관리상태에 있는지 판단하

고, 공정이 이상상태로 돌입하면 가능한 한 신속하게 이를 탐지하는 기법이다. 최근 센서 등 수집 기술이 발달하여 많은 양의 데이터를 실시간으로 수집하여 저장할 수 있다. 과거처럼 샘플링에 의하지 않고 빅데이터로 관리도를 작성하여 공정 평균의 작은 이동도 신속하게 탐지할 수 있다. 하지만 이러한 빅데이터가 자기상관을 갖기 때문에 전통적인 슈위트 관리도를 적용하면 많은 오경보가 발생한다. 본 논문은 시뮬레이션 연구를 통해, 자기상관의 정도에 따라 활용하기 쉬운 관리도 계획을 수집 간격( $w$ )과 수집 크기( $b$ ) 측면에서 제시하였다. 스마트폰에 있는 센서를 이용하여 ‘가속도’ 물리량 데이터를 수집하여 자기상관계수를 구하고, 추천된 ( $w, b$ ) 계획에 따라 배치 평균 관리도를 작성하여 본 연구 결과의 적용성을 확인하였다. 실제 공정 빅데이터를 수집하는 반도체 공정 등에서 필요한 관리도 교육에 관한 의견도 제시하였다.

관리도는 통계적 공정관리의 핵심 기법이다. 본 논문에서 제시하는 현장에서 활용하기 쉬운 배치평균 관리도 작성과 운용방법이 빅데이터 시대에 제조산업의 통계적 공정관리에 도움이 되기를 바란다. 아울러 빅데이터 시대에 맞는 관리도 교육에 관한 다양한 의견이 공유되기를 기대한다.

## 참고문헌

- Alwan, L. C. and Radson, D. (1992), Time-Series Investigation of Subsample Mean Charts, *IIE Transactions*, **24**(5), 66-80.
- Byun, J. -H., Park, J. W., Kim, Y. S., Kim, D., and Lee, D. -H. (2024), Trends, Industry Application and Future Directions of Quality Control and Reliability Engineering, *Journal of the Korean Institute of Industrial Engineers*, **50**(6), 287-406.
- Ding, Y., Shi, J., and Ceglarek, D. (2002), Diagnosability Analysis of Multi-station Manufacturing Processes, *ASME Transactions, Journal of Dynamic Systems, Measurement, and Control*, **124**(1), 1-13.
- Hahn, G. J. (1989), Statistics-Aided Manufacturing: A Look into the Future, *The American Statistician*, **43**(2), 74-79.
- Lu, C.-W. and Reynolds, Jr., M. R. (1999), Control Charts for Monitoring the Mean and Variance of Autocorrelated Processes, *Journal of Quality Technology*, **31**(3), 259-274.
- Montgomery, D. C. (2019), *Statistical Quality Control*, 8th Ed, John Wiley & Sons, NJ.
- Montgomery, D. C. and Mastrangelo, C. M. (1991), Some Statistical Process Control Methods for Autocorrelated Data, *Journal of Quality Technology*, **23**(3), 179-193.
- Qiu, P. (2017), Statistical Process Control Charts as a Tool for Analyzing Big Data, *Big and Complex Data Analysis* (S. Ejaz Ahmed, ed), 123-138, Springer.
- Runger, G. C. and Willemain, T. R. (1996), Batch Means Control Charts for Autocorrelated Data, *IIE Transactions*, **28**(6), 483-487.
- VanBrackle III, L. N. and Reynolds, Jr., M. R. (1997), EWMA and CUSUM Control Charts in the Presence of Correlation, *Communications in Statistics - Simulation and Computation*, **26**(3), 979-1008.
- Zwetsloot, I. M. and Woodall, W. H. (2021), A Review of Some Sampling and Aggregation Strategies for Basic Statistical Process Monitoring, *Journal of Quality Technology*, **53**(1), 1-16.

## 저자소개

**송유진**: 경상국립대학교에서 산업공학 학사학위를 받았고, 산업시스템공학과 석사과정 학생이다. 관심 분야는 품질공학, 실험계획법, 빅데이터 분석이다.

**주혜진**: 경상국립대학교에서 산업공학 학사학위를 받았고, 산업시스템공학과 석사과정을 수료하였다. 관심 분야는 실험계획법, 품질공학, 품질빅데이터 분석이다.

**동승훈**: 고려대학교에서 산업공학 학사, KAIST에서 산업공학 석사 및 박사 학위를 취득하였고, 삼성전자에서 근무하였으며, 현재 SSIT, HYCU 반도체공학과에서 자문교수역을 수행하고 있다. 관심 분야는 반도체 산업에서의 품질공학, 신뢰성공학, 데이터사이언스 공학통계 기술의 실무적 응용연구이다.

**변재현**: 서울대학교에서 산업공학 학사, KAIST에서 산업공학 석사 및 박사 학위를 취득하였고, 현재 경상국립대학교 산업시스템공학부에서 교수로 근무하고 있다. 관심 분야는 실험계획법, 품질경영, 데이터 분석공학이다.