

K. 실시간 품질관리

포항공과대학교 화학공학과
한 종 훈

실시간 품질 관리

1. 서론

최근에 공정 자동화에 대한 높은 관심과 막대한 투자로 인하여 많은 화학 공장들에 DCS(Distributed Control System)들이 설치됨으로써 공정에 대한 막대한 양의 데이터들을 실시간으로 받아들일 수 있게 되었다. 따라서 많은 센서들로부터 매초 매분 또는 매시간 단위로 수집되고 있는 많은 양의 상관된 데이터들을 어떻게 처리하여 공정의 조업 상황을 모니터하고 이상을 조기에 발견(fault detection), 진단(diagnosis)하여 안정된 조업의 달성과 고품질 제품의 생산성을 높이는가 하는 것이 큰 관심거리로 대두되게 되었다.

이와 같은 상황으로 인하여 현재까지 많은 공정 모니터링 방법들이 발표되었는데 그들은 1) 실제 결과들을 어떤 수식적인 모델로부터의 예측과 비교하는 방법, 2) 지식-기반 모델을 사용하는 방법 그리고 3) Shewart charts, CUSUM charts 또는 EWMA charts 등의 통계 공정 제어 (Statistical Process Control; SPC) 도표들을 사용하여 공정의 현재 상태를 표준 조업 상태들에 대해 비교해 보는 방법 등 크게 세 가지로 나눌 수 있다.

먼저 수식적인 접근 방법은 이론적으로는 가장 직접적인 감시 방법이라 할 수 있으나 실제로 그 모델이 효과적이라면 가능한 모든 오류들과 정상적인 가동으로부터의 탈선 원인들을 내포한 수식 모델을 구성해야 하는 큰 결점을 가진다. 설령 어떤 자세한 수식적인 모델을 얻었을지라도 그 모델의 매개변수들이 불확실하고 때때로 이 매개변수들을 실시간으로 갱신(更新)해야 하는 경우도 많다.

지식-기반 접근 방법으로는 전문가 시스템과 인공 신경 회로망을 사용하는 방법들이 있는데 이들은 둘 다 공정에 대한 자세한 수식적인 모델을 필요로 하지 않는다는 장점이 있다. 그러나 전문가 시스템에서는 공정에 대해 잘 아는 전문가나 조업자의 경험적 지식이 주관적일 가능성이 많고 전문가들 사이의 커뮤니케이션이 잘 이루어지지 않으면 개발 시간 및 비용이 많이 소요되기 쉽다는 단점이 있고 인공 신경 회로망을 사용하는 방법은 모델을 구성하기 위해 오류 및 이상 들을 포함한 대량의 학습 데이터들이 있어야 하나 실제 공정으로부터 이런 대량의 오류, 이상의 데이터들을

얻기란 굉장히 어렵다는 단점이 있다. 또한 이 두 가지 지식-기반 접근 방법들 모두가 데이터를 해석하기 위한 통계적인 기반이 부족하다는 것과 방법들 자체가 많은 변수들을 다룰 때에는 너무 복잡해진다는 결점들을 갖고 있다.

반면에 통계적 방법을 이용한 모니터 방법들은 공정의 데이터만 주어진다면 이 데이터를 직접 그리고 간단한 절차를 거쳐 처리하여 공정을 모니터하는 데 이용할 수 있고 또한 공정의 데이터를 분석할 수 있는 도구를 제공한다는 점에서 매우 유용하다.

이 SPC의 전략 중 가장 기본이 되는 것이 조기에 공정이나 생산 제품의 이상을 감지하여 대응 조치를 취할 수 있도록 하는 모니터링이다. 이러한 개념의 모니터링은 단순히 공정의 온도나 압력 등을 화면에 보여 주기만 했던 기존의 개념으로부터 벗어나 발전된 것으로 이러한 모니터링 도표의 일반적인 구조가 그림1.에 보여 진다. 그러나 기존의 SPC 기법들은 한 두 가지 품질 변수를 모니터링하는 데는 익숙해 있으나 관리해야 할 변수의 수가 많아지면 비정상적인 상태를 잘 감지하지 못하게 된다. 이런 문제 때문에 최근에는 많은 변수들의 상관관계를 고려할 수 있는 chemometrics 방법들(Principal Component Analysis:PCA나 Partial Least Squares:PLS 등)을 사용한 모니터링 방법들이 크게 주목을 받고 있는데 Fevotte와 McKenna[24]은 PCA를 polymerisation reactors의 parameter estimation과 on-line monitoring등에 적용하였고 Nomikos과 Kosanovich등[15,16]은 Multiway PCA를 회분식과 반 회분식 공정의 모니터링에 적용하였다. 또 Piovoso는 PLS를 산업용 고분자 공정의 모니터링에, Miller는 photographic paper sensitization에서의 진단에 적용하였고 Koutoudi등[19]은 Multiblock PLS를 연속 공정인 LDPE반응기에, Kresta 등[20]은 fluidized bed reactor와 extractive distillation column의 모니터링에 적용하였다.

국내에서는 한 등이 chemometric기술과 관련하여 제철 공장의 고로 운전의 모니터링에 PCA[26,30,37]를 적용하였으며, 계층적 분해방법에 PCA를 접목시켜 공장규모의 실시간 모니터링 및 진단기술[27]을 개발하였다. Multiway PCA를 이용해 고로의 열풍로의 효율을 개선하였고, PVC 공정의 모니터링[29,32,33]에 적용하였다. 또한 PLS를 이용해 MMA-VA 공중합 공정에 대한 품질 변수의 모니터링[36]과 증류공정의 제품 농도 추정 소프트웨어 센서[28,31,34], 그리고 spectroscopy 데이터의 조성 예측 기술[38]을 개발하였다.

2. 본론

2.1. Statistical Process Control(SPC)

통계 공정 제어(Statistical Process Control, SPC)란 넓은 의미로 통계적 방법들을 사용하여 공정의 생산성과 제품의 품질을 향상시키는 것을 뜻한다[4,5,6].

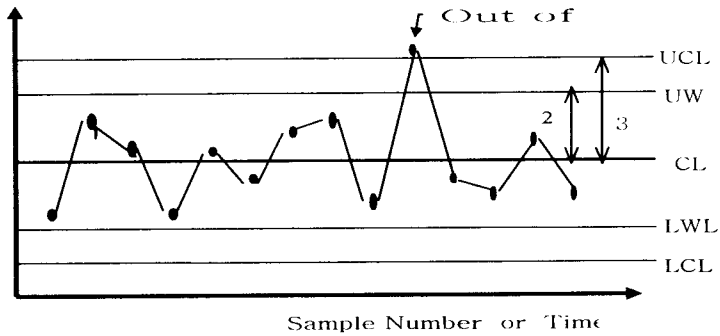
통계적 제어 도표를 사용하여 제조공정들을 모니터링하는 SPC방법들에는 여러가지가 있지만 1930년대 초에 Walter Shewhart[1]가 제시한 Shewhart 도표를 이용한 SPC방법이 가장 널리 사용되고 있으며 이 방법을 이용하여 모니터링하는 방법은 다음과 같다.

우선 모니터링하는 변수는 다음과 같이 일정한 평균과 분산을 가진다고 가정한다.

$$Y_t = \mu + \epsilon_t$$

(Y_t 는 모니터링하는 변수, μ 는 평균 그리고 ϵ_t 는 분산 σ^2 을 가진 독립적인 random error)

그러면 실제 공정으로부터 오는 데이터를 이용하여 통계적 방법으로 그림 1.에서와 같이 99%, 95% 신뢰도를 가지는 제어 선들(control lines; UCL, UWL, CL, LCL, LWL)을 구할 수 있다. 일단 제어 선들이 구해지면 그림 1.에서와 같이 모니터링하는 변수가 아래 위 두 제어 선 사이에 있으면 통계적 제어 상태(in statistical control state)이고 벗어나면 이상 상태(out of control)라는 식으로 모니터링 한다.



<그림 1 shewhart Chart>

일단 공정이 평균을 이동시키거나 분산을 증가시키는 어떤 특별한 사건 때문에 이상 상태가 발생하면 다음과 같은 세단계로 이상을 수정하게 된다.

(i) 그림1.과 같은 Shewhart 도표를 구하여 이상 상태를 발견한다.

(Fault detection)

(ii) 이상의 가능한 원인을 찾는다.

(Diagnosis)

e.g. tank leaking, valve malfunction, controller malfunction....

(iii) 원인을 제거함으로써 공정을 교정하여 목표한 데로 되돌린다.

(Correction)

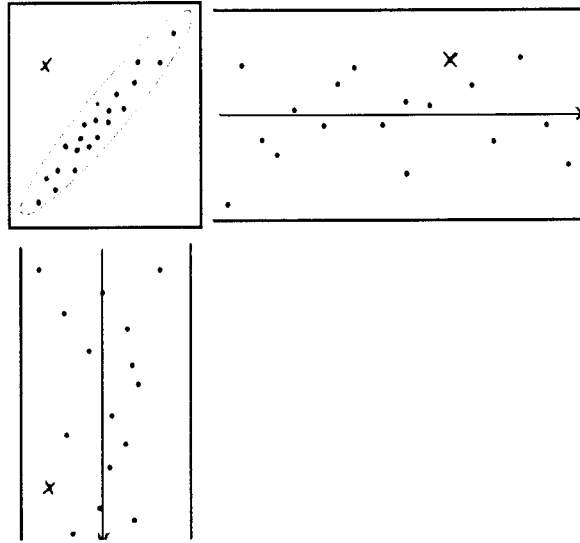
위와 같은 SPC 방법은 공정을 연속적으로 모니터링 하면서 이상이 발생하면 조기에 그 이상 원인들을 제거하여 공정을 지속적으로 안정시킬 수 있으므로 생산성을 증가시킬 수 있다.[4]

2.2. Multivariate Statistical Process Control(MSPC)[6]

기존의 통계 제어 도표 방법들(단 변량 SPC)을 사용하면 수식적인 모델이 필요 없고 오히려 가동 데이터를 직접 사용함으로써 쉽게 적용할 수 있고 해석할 수 있다는 장점이 있다. 그러나 이런 단 변량 SPC 방법들의 주된 결점은 아주 크게 상관된 변수들이 많은 다 변량(multivariate) 문제들

에는 적용시키기 어렵다는 것이다. 실제로 다변량 공정들에서도 변수들이 크게 상관되어 있지 않은 경우에 조업자가 존재하는 작은 상관성을 무시한 채 중요 변수들에 대해서만 Shewart 도표를 사용하여 공정을 모니터링하는 식으로 단 변량 SPC가 적용되고 있지만 변수들이 크게 상관되어 있는 경우에는 이러한 접근 방법이 오히려 더 잘못된 결과를 유발시킬 수 있다. 다음과 같은 예를 들면 쉽게 이해할 수 있다.

기상 반응이 일어나는 연속 반응기를 생각해 보자. 기상 반응이므로 만약 반응기 내의 압력이 올라가면 반응기 내의 온도도 올라가는 것은 당연하며 그 역도 마찬가지이다. 연속 반응기의 반응기 내의 조건은 생성물의 조건과 같으므로 생성물에 있어서도 온도와 압력은 반응기 내에서와 같은 상호관계를 가질 것이다. 그런데 이들에 대해 단 변량 SPC때처럼 상관관계를 무시하고 두 변수 온도(T)와 압력(P) 각각에 대해 독립적으로 Shewart 도표를 구성하여 모니터링 하면 오히려 잘못된 결과를 초래할 수 있다. 그림 2.는 온도(T)와 압력(P) 각각에 대한 Shewart 도표를 보여주고 있다. 이때 타원 안에 있는 모든 정상적인 점들은 그 제어 한계 내에 들어 온다. 그러나 x는 타원 영역 밖에 놓여 있으므로 비 정상적인 점이다. 다시 설명하면 이 생성물 샘플은 온도는 평균보다 더 높으면서도 압력은 평균보다 더 낮은 기본 물리 법칙(상태 방정식)에 어긋나는 비정상적인 사건이다. 그러나 각각의 Shewart 도표를 이용해 모니터링 할 때는 이 x점은 두 도표의 제어 한계 내에 들기 때문에 정상적인 사건으로 판명된다.



<그림 2 상관관계를 가진 데이터들에 대한 Bivariate Plot>

결론적으로 실질적 의미를 가지는 온도와 압력의 물리적 상관관계를 무시하고 각각의 변수에 대해 독립적으로 모니터링 한다면 결과도 물리적 상관관계가 무시된 실질적으로 의미 없는 제품이 이상으로 감지되지 않으므로 그 사건에 대해 아무 조치도 없이 계속 잘못된 제품을 생산하게 되는 엄청난 결과를 초래 할 수도 있을 것이다.

상관되어있는 데이터들에 대해 위와 같은 난점들을 해결하기 위해서는 다변량 통계 방법들에 기초하는 Principal Component Analysis (PCA)와 Partial Least Squares (PLS)라는 더 효과적인 접근 방법을 사용해야만 한다.

$$X = \begin{bmatrix} X'_{11} & X'_{12} & \dots & \dots & \dots & X'_{1K} \\ X'_{21} & X'_{22} & \dots & \dots & \dots & X'_{2K} \\ \vdots & \vdots & \vdots & & & \vdots \\ \vdots & \vdots & & \vdots & & \vdots \\ \vdots & \vdots & & & \vdots & \vdots \\ X'_{N1} & X'_{N2} & \dots & \dots & \dots & X'_{NK} \end{bmatrix}$$

<그림 3. 자료행렬 matrix>

2.3. Principal Component Analysis (PCA)

원래 Pearson[2]에 의해 개발된 PCA는 Hotelling[3]이 다변량 문제들에 있어서 변수들의 상호 관련 구조들을 분석하기 위하여 많이 사용하였다. 이 방법은 다변량 통계 기법에 관한 많은 책들에서 설명되고 있으며 Wold[9,14]의 review에도 자세히 설명되어 있다.

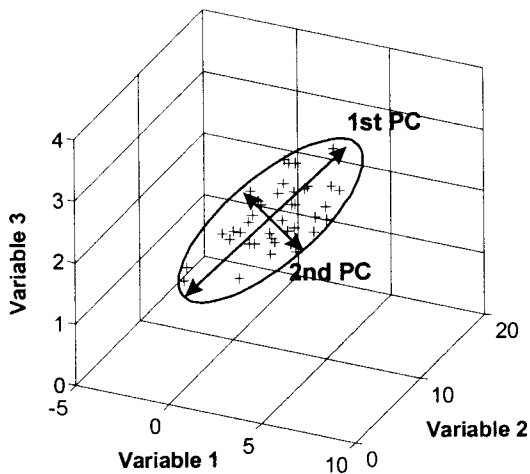
PCA와 같은 모든 다변량 자료 해석에 있어서 출발점은 다음과 같은 X로 나타나지는 자료 행렬(data matrix)에서 시작된다.

이 행렬에서 N개의 행들은 사건(object)들이라 하는데 예를 들면 보통 화학적, 지질학적 샘플이거나 어떤 한 공정에서의 샘플들 등에 해당한다. 그리고 K개의 열들은 변수(variable)들이라 하는데 이는 각 샘플의 측정 항목이며 예를 들면 어떤 공정에서의 온도, 압력, pH등의 공정 변수들 등에 해당한다.

PCA를 이해하는 방법에는 기하학적 방법과 수학적 방법이 있다.

먼저 기하학적으로 PCA를 이해한다는 것은 PC들과 그에 대한 정사영(projection) 그리고 상관 관계를 포함하여 이용되는 수학적 연산까지도 공간에서 일어나는 것으로 이해하는 것이다.

PCA는 그림4.와 같이 데이터를 발생시킨 system을 분석하기 위하여 데이터 공간(데이터가 분포하는 영역)에 대해 분포가 넓은 축부터 차례대로 서로 직교도록 새롭게 축들을 하나씩 정의하고 이들에 대한 정사영 값들을 구하는 것에서 출발한다. 이때 데이터의 상관 관계를 더 쉽게 분석하기 위해 새롭게 정의된 새로운 축들을 PCs라하며 데이터를 이들 축에 투영시켜 얻은 정사영 값들을 그 축에 대한 score vector라고 한다. 그림4.는 데이터 공간에서 PC들이 정의되는 방법을 설명하고 있다. 그런데 PCA가 실질적으로 적용될 때는 정의될 수 있는 모든 축들(PCs)과 그들에 정사영 값들(score vectors)을 이용하여 system을 분석하는 것이 아니라 a개의 축들과 그것들에 투영된 정사영 값들만을 가지고 그들의 linear combination으로 system을 근사한 후 이 근사된 system을 분석하게 된다.



<그림 4. PCA의 이해>

이런 a개의 PC성분으로 근사하는 것을 $N \times K$ 자료 행렬에 대해 수식적으로 표현하면 그림5.와 같다.

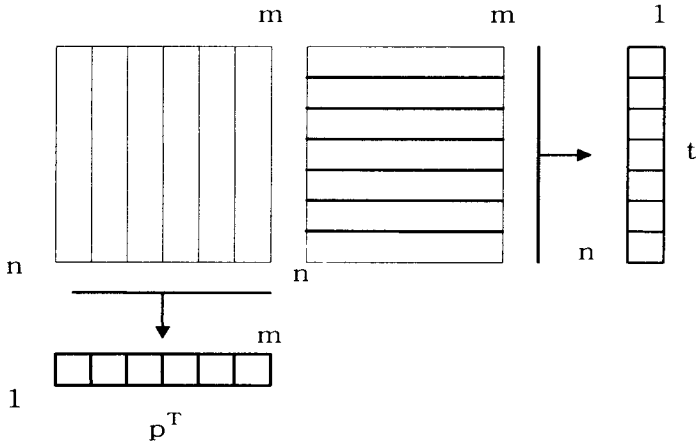
$$X = M_1 + M_2 + M_3 + \dots + M_a$$

$$\begin{array}{c}
 \begin{array}{c} \boxed{X} \\ n \end{array}^m = \begin{array}{c} \boxed{M_1} \\ n \end{array}^m + \begin{array}{c} \boxed{M_2} \\ n \end{array}^m + \begin{array}{c} \boxed{M_3} \\ n \end{array}^m + \dots + \begin{array}{c} \boxed{M_a} \\ n \end{array}^m \\
 \\
 \begin{array}{c} \boxed{X} \\ n \end{array}^m = \begin{array}{c} \boxed{t_1} \\ n \end{array}^1 \begin{array}{c} \boxed{p_1^T} \\ 1 \end{array}^m + \begin{array}{c} \boxed{t_2} \\ n \end{array}^1 \begin{array}{c} \boxed{p_2^T} \\ 1 \end{array}^m + \dots + \begin{array}{c} \boxed{t_a} \\ n \end{array}^1 \begin{array}{c} \boxed{p_a^T} \\ 1 \end{array}^m = \begin{array}{c} \boxed{T_a} \\ n \end{array}^a \begin{array}{c} \boxed{P_a^T} \\ a \end{array}^m
 \end{array}$$

<그림 5. 자료행렬의 근사>

이를 이해하기 위해 PC와 score vector의 기하학적 의미를 알아보자. 그림6.을 보면 X에 존재하는 변수들의 모든 관계를 하나의 대표적인 관계로

근사한 것이 첫번째 PC라는 것을 알 수 있다. 만약 a개의 PC들이 있다면 첫번째 PC는 X에 존재하는 변수들의 모든 관계를 가장 잘 설명할 수 있는 변수들의 관계를 의미하는 벡터이고 두번째 PC는 그 다음으로 잘 설명하는 관계를 의미하며 a번째 PC는 a번째로 변수들의 관계를 잘 설명할 수 있는 벡터를 의미한다. 이때 score vectors(t)도 PC와 마찬가지로 observation들 간의 a개의 대표적인 관계로도 해석될 수 있지만 t는 a개의 PC들에 투영시킨 값들로 해석하는 것이 일반적이다. 따라서 a가 충분히 크다면 이들 변수들 간의 가장 대표적인 a개의 관계들과 score 값들을 가지고 이것들의 linear sum으로서 X를 근사할 수 있다는 것이 그림5의 이론적 바탕이다. 예를 들어 3차원 상에 사람이 방에 누워 있는 상태(system)를 분석하기 위해서 위와 같은 PCA를 이용한다면 우선 첫번째 PC를 데이터의 분포가 가장 큰 머리에서 발로 향하게 잡아 이것에 투영시키면 위에서 빛을 비춰 방바닥에 생긴 그림자와 같을 것이다. 그리고 두번째 PC는 첫번째 PC와 수직이 되도록 옆으로 귀에서 발목쪽으로 잡으면 이번의 정사영은 누워있는 옆 모습이 될 것이다. 물론 세번째 PC로 머리에서 바라본 측면도도 얻어 system을 완벽하게 분석할 수 있지만 세번째 PC는 사람의 누워있는 모습을 분석하는 데 기여하는 부분이 상대적으로 적으므로 무시하고 첫번째, 두번째인 두개의 PC(a개의 PC) 만 가지고 그들의 linear sum으로서 system을 근사할 수 있는 데 이것이 PCA이다.



<그림 6. Score vectors와 loading vectors>

두번째로 PCA를 수학적으로 이해하려면 선형 대수적으로 설명해야 한다. 자료 행렬의 모든 데이터 값에서 일정한 평균 값을 빼고 그 값을 표준편차로 나누면 자료 행렬(X)의 모든 값은 모두 0과 1 사이를 움직이게 되는데 이런 작업들을 각각 centering 과 scaling 이라 한다.

Centering 과 scaling 이 된 자료 행렬(X)을 얻으면 $X^T X$ 와 같은 행렬 곱을 하여 covariance 행렬(S)를 얻을 수 있는데 이 covariance 행렬의 열공간(column space)이 바로 데이터 공간이 된다. 따라서 이 covariance 행렬($S=X^T X$)의 열 공간을 span하는 고유 벡터들(eigenvectors)이 바로 데이터 공간을 span하는 PC들이 되며 각각의 고유 벡터(PC)에 대한 고유 값(eigenvalue)가 그 PC 방향으로 움직임의 대소를 결정한다. 다시 말하면 고유 값이 크면 데이터 공간(covariance 행렬의 열 공간)에서 고유 값(PC) 방향으로 데이터 분포의 분산이 크다는 것을 의미하므로 차례대로 축(PC)들을 정의 할 때 고유 값이 큰 것부터 첫번째 PC, 두번째 PC등과 같이 정하게 된다. 이렇게 해서 고유 값이 큰 것부터 우리가 원하는 a개의 PC를 구해 그림5.같이 근사하면 된다. 이것을 수식적으로 표현하면 다음과 같다.

$$X = T_a P_a^T + E \quad (E : \text{residual})$$

$$E = X - T_a P_a^T$$

$$SSQ = \|E\| = \sqrt{\sum_{n=1}^N \sum_{k=1}^K E(n, k)^2}$$

실제로 이와 같이 PCA를 적용하여 system을 근사할 때 중요하게 대두되는 문제 중에 하나가 몇 개(a개)까지의 PCs를 사용할 것인가 하는 것이다. 이때는 매번 PC가 하나씩 추가될 때마다 통계에서 모델 validation시 많이 사용하는 F-test를 하여 residual(E)이 random error라고 판단될 때까지의 PCs의 수를 구하면 된다.

그러나 이런 수학적 방법에 의해 PCA를 적용하는 것은 차원이 증가하면서 컴퓨터를 이용하는 계산상에 문제가 제기되므로 Nonlinear iterative partial least squares(NIPALS)와 같은 간단하면서도 더 빠르고 정확한 알고리즘이 가장 많이 이용되고 있다.

NIPALS 알고리즘은 중요도에 따라 순서대로 원하는 a개의 PC를 구할 수 있는 알고리즘이지만 a개의 PC를 한번에 구하는 것은 아니고 일단 X로부터 t_1 과 p^T_1 을 구해내면 residual E_1 을 X에서 $t_1 p^T_1$ 을 빼냄으로서 구할 수 있고 이 E_1 으로부터 t_2 과 p^T_2 을 구하는 식으로 계속해서 E가 random error가 될 때까지 a개의 PCs를 구한다.

$E_1 = X - t_1 p_1^T, E_2 = E_1 - t_2 p_2^T, \dots, E_h = E_{h-1} - t_h p_h^T$ 이고 각 단계에서 NIPALS 알고리즘은 다음과 같다.

1. X로부터 임의의 column 벡터 x_j 를 취하여 t_h 라 둔다 : $t_h = x_j$
2. $p_h^T = t_h^T / t_h^T t_h$
3. normalize p_h^T to length 1 : $p_{h\text{new}}^T = p_h^T / \|p_h^T\|$
4. $t_h^T = X p_{h\text{new}} / p_{h\text{new}}^T p_{h\text{new}}$
5. compare the t_h used in step 2 with that obtained in step 4. If they are the same, stop (the iteration has converged). If they still differ, go to step 2.
6. $E = X - t_h p_h^T$; $X = E$
7. go to step 1

또한 PC들이 구해지면 모니터링을 위해 SPC방법을 사용하게 되는데 이때 중요한 것이 우리가 모니터하는 어떤 변수가 어느 정도의 한계를 벗어나야 이를 이상 또는 오류라고 판단 내릴 수 있을까 하는 것이다. 그 한계를 나타내는 제어 선의 값을 정해 주어야 하는데 이때 이용하는 것이 Hotelling's T^2 statistic이다.

정규 분포를 가지는 단 변량 문제에서 어떤 사건이 정상인지 아닌지를 알 수 있는 척도는 평균에서의 거리가 된다. 2개 이상의 변수가 포함된 다변량 문제에서도 각 변수가 서로 독립적인 것이라면 어떤 사건에 대해 각각의 변수 값과 각각의 평균 사이의 거리가 그 척도가 될 수 있다. 하지만 변수들 사이에 큰 상관관계가 존재하는 경우에는 전체 데이터의 상관관계를 고려한 척도가 필요하게 되는데 이 때 사용할 수 있는 척도가 Hotelling's T^2 statistic이다.[3]

먼저 과거 n 개의 다변량 표본으로부터 다음과 같이 표본 covariance 행렬 S 를 구한다. 여기서 n 은 i 번째 사건의 값을 담고 있는 열 벡터이고 \bar{x} 는 평균 벡터(mean vector)이다.

$$S = (n-1)^{-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T$$

다음으로 새로운 다변량 사건 \mathbf{x} 가 얻어지면 Hotelling's T^2 statistic은 다음과 같이 구해지며 통계량에 대한 구조적 해석은 Jackson[2]에 설명되어 있다.

$$T^2 = (x - \tau)^T S^{-1} (x - \tau)$$

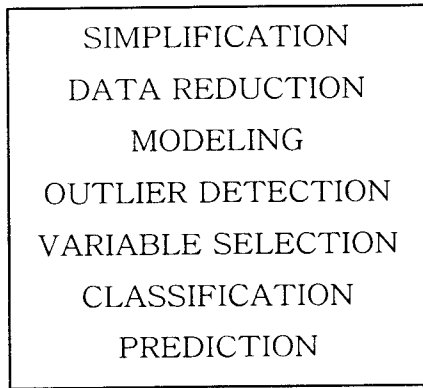
여기서 τ 는 목표값(target value)이다.

이와 같이 구해진 T^2 는 시간에 따른 도표로 그려지게 되는데 이 도표에서 주어지는 상위 제어 선(UCL)은 다음과 같이 주어진다.

$$T^2_{UCL} = \frac{(n-1)(n+1)a}{n(n-a)} F_a(a, n-a)$$

여기서 n 은 a 와 $n-a$ 의 자유도를 가지는 F 분포의 upper 100 % critical point이며 n 은 사건의 수, a 는 PC의 수이며 a 는 신뢰도이다. 따라서 PC의 수 p 가 2개이고 sample 수 n 이 15이고 백분율 신뢰도 $1-a$ 가 95% ($a=0.05$) 일 때 $F_{2,13,0.05}=3.81$ 이므로 T^2 의 UCL 값은 8.21이 된다. 많은 변수들이 존재하는 공정에서도 이 한 값만을 모니터링으로써 공정의 이상 여부를 판단할 수 있게 된다.

그림7.은 PCA로 자료 행렬을 분석하는데 있어 가질 수 있는 여러 목표들을 보이고 있다.



<그림 7 PCA의 여러 가지 목표>

PCA에서 자료 행렬을 a 개만의 축(PC)을 사용하여 system을 근사(또는 모델링)하므로 데이터 전체를 다룰 때에 비해 Simplification과 Data reduction가 일어나는 것은 당연하다. 이렇게 모델이 생기면 새로운 input

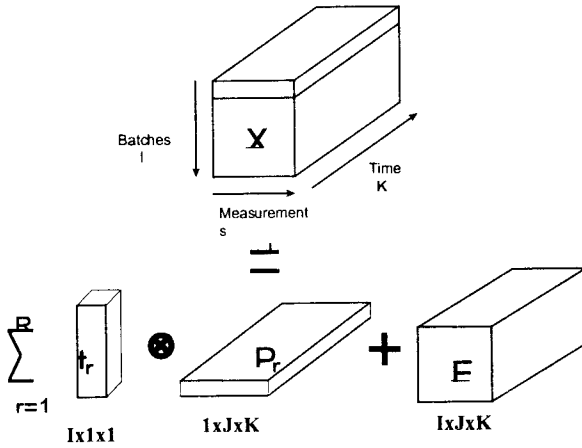
에 대해 output을 예측(prediction)할 수 있다. 또한 score plot(t_1 vs t_2)을 해보면 간단한 classification이 일어나고 그때 어느 group에도 속하지 않는 outlier를 찾아낼 수 있으며 이것을 loading vector plot(pT_1 vs pT_2)과 비교해보면 어느 변수에 의해 이러한 outlier가 일어났는지 알 수 있다. 이것이 바로 outlier detection과 variable selection이다.[8, 9]

PCA의 적용 사례는 Dunia[10]가 센서 고장의 확인에 Tong[11]은 지속되는 총계 오차의 감지에, Heyen[23]은 measurement reconciliation 시 Sensitivity calculations과 variance analysis에, Fevotte와 McKenna[24]은 polymerisation reactors의 parameter estimation과 on-line monitoring등에 적용한 예들이 있다.

2.4. Multiway Principal Component Analysis (MPCA)

위에 설명된 PCA는 순차적으로 수집된 data를 해석기에는 적합하지 않은 면이 있다. 순서를 바꾸어 다시 자료 행렬을 꾸며도 일반 PCA로는 바꾸기 전과 같은 결과가 나오기 때문이다. 이런 경우 MSPC를 적용하기 위해 PCA처럼 도입하는 것이 MPCA이다.[12, 15, 16]

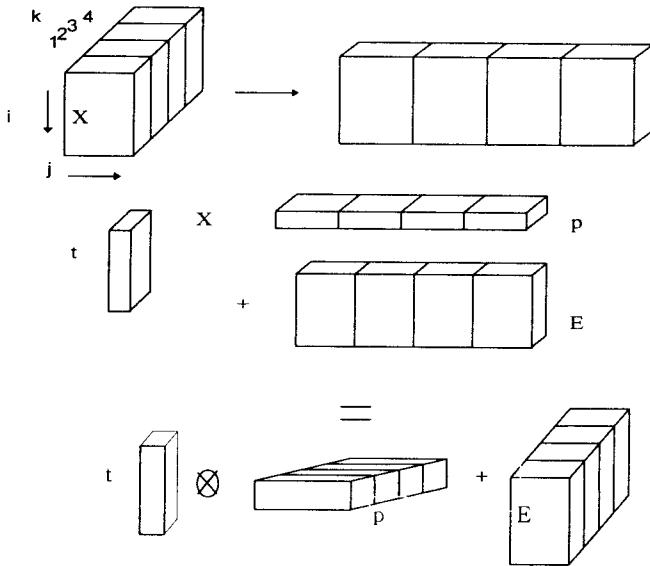
화학, 심리 측정학, image 분석 그리고 회분식 반응과 같은 분야들의 실험으로부터 얻어지는 데이터들이 이런 순차적으로 수집되는 데이터들에 해당한다. 회분식 공정의 예를 들어보자. 매번 $batct(I=1, 2, \dots, I)$ 때마다 시간($j = 1, 2, \dots, J$) 대에 측정되는 k 개의 변수($k=1, 2, \dots, K$) 측정치들이 있다고 하자. 이러한 데이터들은 같은 시간대에서 측정되는 batch들의 k 측정치에 대해 비슷한 값을 가질 것이므로 그림8.같은 삼차원 배열($I \times K \times J$)로 나타내어 질 수 있다.



<그림 8 MPCA에 의한 Three-way array의 arrangement와 decomposition>

그림8.은 다른 batch run들이 수직면을 따라 측정 변수들이 수평면을 따라 그리고 이들의 시간에 따른 변화가 세 번째 차원을 차지하고 있다. 이 배열을 수평으로 자른 각각은 한 batch (i)의 모든 변수(j)들의 시간에 따른 변이나 궤적을 나타내는 $K \times J$ 행렬이 되고 수직으로 자른 각각은 time interval (k)에서의 모든 batch들에서의 모든 변수들의 값들을 나타내는 $N \times K$ 행렬이 된다. 이와 같은 데이터를 다룰 수 있는 방법은 MPCA로써 통계적으로나 알고리즘적으로 PCA와 일치하며 같은 목표와 이점들을 가지고 있다.

MPCA 와 PCA 사이의 관계는 MPCA는 삼차원 배열(three-way array)인 X 를 세 개의 방향 중 하나로 펼쳐 놓음으로써(unfolding) 만들어지는 큰 이차원 행렬에 보통의 PCA를 실행하는 것과 동일하다. 보통은 X 를 펼치는 데 있어 그림9.에서와 같이 X 의 수직으로 자른 면 ($N \times K$) 행렬을 순서대로 오른쪽에 나란히 놓아가는 방식을 사용한다. 결과적으로 이차원 행렬은 ($N \times KJ$)의 차원을 가진다. 이런 펼침은 우리가 변수들과 시간에 따른 변이 모두에 대하여 데이터에서의 정보를 요약함으로써 X 에서의 batch들 사이의 variability를 분석할 수 있게 해 준다.



<그림 9. MPCA 알고리즘>

MPCA는 three-way array인 X 를 그림6.에서처럼 unfolding한 뒤 PCA의 원리에 따라 score vector (t_r)와 loading matrix(p_r)들로 구성된 systematic part ($\sum t_r \otimes p_r$)와 residual part (E)의 더해진 형태로 분해하는 것이다.

이런 MPCA 방법을 이용하여 Nomikos와 MacGregor[15, 16]는 회분식 공정에 대해 보통 SPE를 이용하여 모니터링할 때는 정상으로 나타나는 것이 score 도표를 이용하면 실제 공정은 비정상 상태를 밝혀 내었다. 그 밖에 MPCA를 적용하고 있는 사례들은 Kosanovich(1996)이 회분식 또는 반 회분식 공정의 모니터링에 적용하였고 Gallagher N.B.와 B.M. Wise[12]도 핵폐기물 저장 탱크의 모니터링에 적용하였다.

2.5. Partial Least Squares or Projection to Latent Structures (PLS) [14, 17, 18]

PCA는 하나의 자료 행렬(X)에서 변수들의 관계를 잘 대표하는 PC들을 찾고 제어 선들을 구해내어 자료 행렬의 공정 데이터들을 발생시키는 system을 모니터링할 때 이용한다. 그러나 실질적으로 더 중요하면서도 모

니터하고 싶은 것이 product의 품질일 경우일 때가 있는데 이때 사용되는 것이 PLS이다. PCA가 공정 변수들을 모니터하기 위해 공정 변수들의 측정치로 구성된 자료 행렬(X)을 이용했듯이 PLS도 품질 변수들의 측정치로 구성된 행렬(Y)을 사용하여 품질 변수를 모니터할 수 있다. 그러나 불행하게도 이 변수들은 공정 변수들보다는 훨씬 더 낮은 빈도로 측정되므로 행렬(Y)을 구성하는 것이 어렵다. 따라서 우리는 공정 변수들(X)에 포함되어 있는 정보를 이용하여 품질 변수들(Y)에서의 변화를 예측, 모니터, 감지하는 방법을 사용하는데 이때 PCA에 이론적 바탕을 두고 X와 Y의 관계를 짓는 방법이 바로 PLS이다. X와 Y를 mapping하는 transfer matrix를 구성하는 방법에는 PLS외에도 Multiple Linear Regression(MLR)와 Principal Component Regression(PCR)이 있다. 그러나 가장 널리 응용되고 있는 MLR은 centering되고 scaling된 자료 행렬에 대해 collinearity와 singularity (Wold, S. et al. , 1984)의 문제를 안고 있고 PCR은 X의 score vectors(or PCs)에 대해 y의 변수들 각각을 regression시킴으로 collinearity와 singularity 문제는 발생하지 않지만 X를 가장 잘 설명하는 첫번째 PC에 대한 score vector가 Y를 가장 잘 설명하리라는 보장은 없으므로 PCR은 한계가 있다. 다시 말하면 Y를 구성하는 데이터들이 크게 상관 관계를 가진다면 PCR에서처럼 X에 대한 데이터 공간의 상관 관계를 가장 잘 설명한다고 해서(첫번째 PC) Y에 대한 데이터 공간의 상관 관계까지도 가장 잘 설명한다고 볼 수는 없다. 오히려 X에 대한 첫번째 PC가 가장 못 설명할 수도 있다.

반면 Partial Least Squares(PLS)는 위에서 언급한 문제들을 가장 잘 해결할 수 있는 방법이다. 먼저 X와 Y 공간 상에서 각각에 대해 PCA를 적용해서 그림10과 같이 분해한다. 이런 관계들을 각각에 대한 outer relation이라 한다.

$$\begin{array}{c}
 \begin{array}{ccc}
 \begin{array}{|c|} \hline X \\ \hline \end{array} & = & \begin{array}{|c|} \hline T \\ \hline \end{array} \begin{array}{|c|} \hline P^T \\ \hline \end{array} + \begin{array}{|c|} \hline E \\ \hline \end{array} \\
 \begin{array}{ccc}
 n & & n
 \end{array}
 \end{array} \\
 \\
 \begin{array}{ccc}
 \begin{array}{|c|} \hline Y \\ \hline \end{array} & = & \begin{array}{|c|} \hline U \\ \hline \end{array} \begin{array}{|c|} \hline Q^T \\ \hline \end{array} + \begin{array}{|c|} \hline F^* \\ \hline \end{array} \\
 \begin{array}{ccc}
 n & & n
 \end{array}
 \end{array}
 \end{array}$$

<그림 10. X와 Y에 대한 Outer relations>

PLS가 PCR보다 나은 점은 X의 score vector(t_h)와 Y의 score vector(u_h) 사이에 $u_h = b_{ht}t_h$ 와 같은 inner relation을 만들어 서로에 대한 정보를 공유한다는 점이다. 이 inner relation은 X의 score vector들을 Y의 데이터 공간을 잘 설명할 수 있도록 회전시키는 것을 의미하므로 X공간에서의 평면이 Y를 더 잘 예측할 수 있도록 기울어진다는 것을 의미하기도 한다. 그런 다음 X의 score vector들에 대해 weight(w_h)를 주어 그들 각각이 Y의 데이터 공간을 설명하는 정도를 정해준다. 이런 PLS를 이용하면 score vector들로 regression하므로 collinearity와 singularity 문제도 발생하지 않을 뿐더러 X와 Y에게 서로 정보를 공유하고 contribution의 정도를 weight를 주어 조정하므로 PCR에서 생겼던 문제도 발생하지 않는다. 수학적으로 PCA에서의 loading vector들은 covariance 행렬($S = X^T X$)의 고유 벡터들이듯이 PLS의 loading vector들은 $(X^T Y)(Y^T X)$ 의 고유 벡터들이다.

PLS를 실제 적용할 때는 PCA에서 사용했던 NIPALS 알고리즘과 비슷한 다음과 같은 알고리즘을 이용한다. 먼저 다음과 같은 outer relation을 PCA를 적용해 구한 다음 알고리즘을 적용한다.

$$X = TP^T + E, Y = UC^T + F, U = TB + G \text{ 이고 이때 } B = (T^T T)^{-1} T^T U \text{이다.}$$

1. Start: set u equal to a column of Y
2. $w^T = u^T X / u^T u$ (regress columns of X on u)
3. Normalize w to unit length
4. $t = Xw / w^T w$ (calculate the scores)
5. $q^T = t^T Y / t^T t$ (regress columns of Y on t)
6. Normalize q to unit length

7. $u = Yq / q^T q$ (calculate new u vector)
8. Check convergence: if YES to 9, if NO to 2
9. X loadings: $p = X^T t / t^T t$
10. Regression: $b = u^T t / t^T t$
11. Calculate residual matrices: $E = X - tp^T$ and $F = Y - btq^T$
12. To calculate the next set of latent vectors replace X & Y by E and F and repeat.

이 알고리즘을 잘 보면 t 가 1-8번 과정에도 나타나고 9번 과정에서도 나타나는 것을 알 수 있다. 1-8번 과정에서의 t 는 서로 정보를 교환함으로써 PCA의 NIPALS에서와 달리 직교하지 않는 t 이지만 9번 과정에서의 t 는 이들을 다시 직교하도록 만들어 준 것이다. 직교하는 t 는 꼭 있어야 하는 것은 아니지만 PCR과의 비교를 용이하게 하며 PCA와의 연결이 쉽다는 장점이 있으므로 이 과정을 알고리즘에 포함시키는 것이 일반적이다. 또한 위 알고리즘은 선형 회귀 법(10번 과정)을 사용한 것이므로 원래 변수들을 변형시키거나 10번 과정에 알고 있는 비선형 관계를 집어 넣음으로써 PLS 모델에 비선형성을 줄 수도 있다.

PLS도 PCA에서처럼 a 개의 PLS성분을 결정해야 하는데 이때 많이 사용하는 방법은 cross validation과 PRESS같은 통계적 방법이다.

PLS를 모니터링에 응용 사례[17, 19]는 Wise가 industrial ceramic melter의 분석 및 모니터링에, Piovoso가 산업용 고분자 공정의 모니터링에, Slama가 정제소의 fluid bed catalytic cracking 및 fractionation section의 가동 행태의 분석에, Miller가 photographic paper sensitization에서의 진단에, Dayal가 industrial pulp digester에서의 가동시 문제점의 분석에, Hodouin가 mineral crushing, grinding, and flotation circuit들의 가동 행태의 분석등에 적용한 예가 있으며 이 외에도 철강, 타이어, 고무, 의약품 및 화학 약품 산업 등에 널리 응용되었다.

2.6. Multiblock Partial Least Squares or Projection to Latent Structures (MPLS)[19]

실제 대형 화학 공정에 위에서 설명한 PLS 같은 것을 적용하려면 공정 전체를 하나로 다루는 것과 함께 모니터링과 진단을 모두 용이하게 하기 위해 공정 전체를 몇 개의 하위 블록으로 나누어 각각의 하위 블록에 PLS

를 적용하는 MPLS를 사용하게 된다. 왜냐하면 전체의 모니터링에 문제가 생기면 그 원인을 찾을 때 각각의 하위 블록들로 내려가서 이상 여부를 살펴보면 어느 블록에서 이상이 생겼는 지 알 수 있고 그렇게 되면 그 문제의 블록 안에서 이상 원인을 밝혀내면 되므로 진단에 있어서 더욱 효과적이 때문이다. 이러한 multiblock projection method에는 앞서 설명한 multiway PCA와 multiblock PLS가 있는데 모두 계층적인 구조를 가지고 있어서 전체 블록과 하위 블록들 사이에 서로 보완적인 관계를 갖는다.

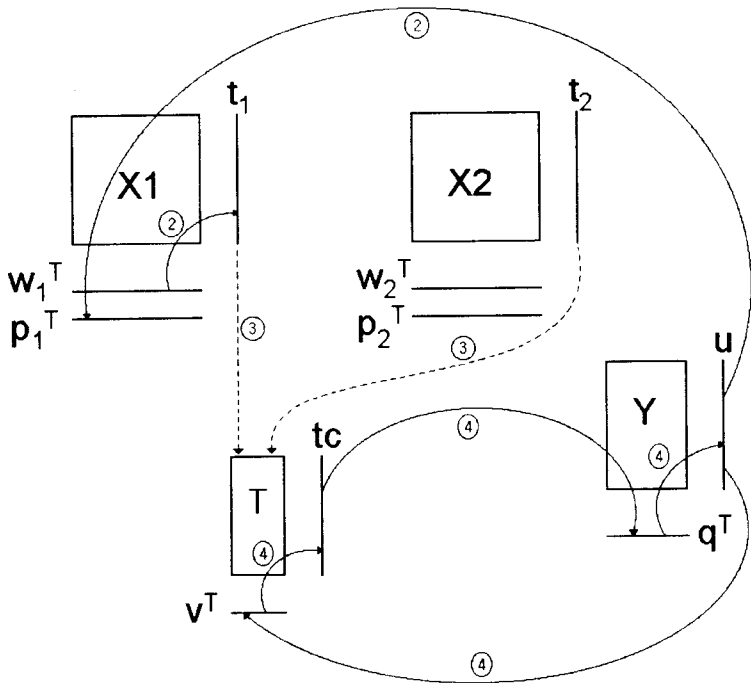
MPCA가 PCA와 multiblock방법을 결합한 것처럼 MPLS도 PLS와 multiblock방법(Herman Wold[4]와 Svante Wold[4])을 결합하고 있다. 다만 MPCA에서는 데이터가 순차적으로 수집되므로 그 순서대로 unfolding 하면 되지만 MPLS는 전체 공정을 어떻게 나누는가에 따라 결과가 많이 틀려지므로 어떻게 블록화 할 것인가가 중요하다. 하지만 이것에 대한 일반적으로 적용되는 규칙은 없는 것으로 알려져 있다. 다만 같은 블록 내에서 각 변수들간의 상호작용이 크고 다른 블록들 사이에서 최대한 작아지도록 블록화 하는 것이 일반적이다. 화학 공정에 대해서는 대부분이 단위 공정들이 모여 전체 공정을 이루므로 단위 공정들 별로 블록화 하는 것이 일반적이다.

MPLS 알고리즘은 기본적으로 NIPALS에 바탕을 두고 있으며 PLS 알고리즘과 많은 면에서 비슷하다. 다만 여러 개의 블록에서 매 iteration시 나오는 score vector들은 모두 모아 하나의 score 행렬로 구성한 후 그것을 다시 projection 시켜 하나의 대표 score vector(tc)로 나타내면 그 다음을 구해낸 tc를 Y의 score vector의 초기값으로 하는 PLS 알고리즘과 같다.

이 MPLS 알고리즘을 반응기와 증류 탑으로 구성된 전체 공정에 대해 실제 off-line으로 측정되는 process yield(품질 변수)를 모니터링하는 문제의 예를 통해 설명해 보자. 하나의 단위 공정이 하나의 블록을 이룬다고 했으므로 2개의 X블록과 1개의 Y블록이 있는 문제이다. 이 문제에 대한 M PLS 알고리즘과 그림은 다음과 같다.

- 1) Y의 임의의 한 컬럼을 u라 한다.
- 2) (w_1, t_1) , (w_2, t_2) 을 구하기 위해서, X_1, X_2 각각의 블록에 PLS를 적용한다.
- 3) 모든 score vectors t_1, t_2 를 모아서 T matrix라 한다.
- 4) T matrix의 loading vector v, score vector tc, Y matrix의 loading vector q, new score vector u를 구하기 위하여 T matrix를 X로 간주하고 PLS를 적용한다.

- 5) u 의 수렴 여부를 결정하여 수렴하지 않으면 (2)로 돌아간다.
- 6) 각각의 블록 X_1, X_2 에 대한 loading vector를 구한다.
 $(p_1 = X_1^T t_1 / t_1^T t_1, p_2 = X_2^T t_2 / t_2^T t_2)$
- 7) Residual matrix를 구한다.
 $(E_1 = X_1 - t_1 p_1^T, E_2 = X_2 - t_2 p_2^T, F = Y - t c q^T)$
- 8) X_1, X_2, Y 를 E_1, E_2, F 로 대체한 후 (1)로 돌아간다.



<그림 11. Multiblock PLS 알고리즘>

이들을 a 개의 PLS 성분을 이용하여 수식적으로 표현하면 다음과 같다.

$$X1 = \sum_{a=1}^A t1_a p1_a^T + E1 \quad X2 = \sum_{a=1}^A t2_a p2_a^T + E2$$

$$Y = \sum_{a=1}^A tc_a q_a^T$$

Macgregor[19]등은 LDPE 반응기를 2개의 블록으로 나누어 불순물의 비정상적인 증가를 모니터링하고 특정 블록 내에서의 contribution plot를 통하여 그 원인을 찾아 내었으며(1) batch polymerization 공정에 대해서도 batch data의 특성을 고려하여 batch set-up data와 batch trajectory data를 2개의 블록으로 나누어 모델에 포함된 정상적인 batch와 비교하여 원하지 않는 batch를 실시간으로 모니터링하고 그 원인을 규명해내었다(2). 또한 Kresta[20]등(1991)은 fluidized bed reactor와 extractive distillation column의 모니터링에 적용하기도 하였다.

3. 결론

지금까지 실시간 품질관리에 대한 PCA, PLS와 같은 chemometrics적 접근 방법의 개념과 알고리즘 그리고 응용 사례들에 대해 설명하였다. 다시 한번 정리하면 chemometrics 방법(PCA, PLS)은 특히 큰 상관 관계를 가지는 변수들을 다룰 때 데이터의 구조 분석을 쉽게 하고 데이터 compression과 complexity reduction 그리고 이상 진단[10, 25]에 대한 정보를 제공한다는 점들에서 다른 방법에 비해 큰 강점을 가지며 주어진 데이터들에 대해 같은 모델을 사용한다는 점에서 모델이 deterministic하고 필요한 계산이 간단하다는 이점이 있다. 이런 이점들 때문에 이번에 소개한 실시간 품질관리 분야뿐만 아니라 data reconciliation[23], gross error detection 그리고 parameter estimation[24]등의 분야에 대한 응용이 활발히 연구되고 있다.

세계화 추세에 따라 기업간 경쟁이 더욱 치열해지고 있는 지금 높은 수준의 품질관리는 경쟁에서 승리하기 위한 필수 조건이다. 기존의 다변량 통계 제어 방법보다는 훨씬 빠르고 정확한 다변량 통계적 품질 관리 기법은 앞으로 더욱 빠른 속도로 화학, 환경, 재료 등 다양한 산업분야로 적용되어 나아갈 것이며, 아울러 더욱 많은 연구 개발이 필요한 것으로 예상된다.

참고 문헌

1. Shewhart, W. A., "Economic Control of Quality of Manufactured Product", Van Nostrand, Princeton, NJ (1931)
2. Pearson, K., On lines and planes of closest fit to systems of points in space, *Phil Mag*, ser 6, 2, pp.559-572, (1901)
3. Hotelling, H., Analysis of a complex of statistical variables into principal components, *J. Educat Psychol.*, 24, pp. 417-441, (1933)
4. MacGregor, J. F., Statistical Process Control for the Process Industries, the 4th International Symp. on PSE, Montebello, Quebec, Canada August 5-9 (1991)
5. Montgomery D. C., Introduction to Statistical Quality Control, 2nd Ed., John Wiley & Sons, INC., New York (1991).
6. Kourti T., J. Lee and J.F. MacGregor, Experiences with industrial applications of projection methods for multivariate statistical process control, *Comp. Chem. Eng.* Vol. 20, suppl., pp. S745-S750 (1996)
7. Neter, J., W. Wasserman and M.H. Kutner, Applied Linear Statistical Models, 3rd ed., Richard D. IRWIN, INC., (1990)
8. Jackson J. E., A User's Guide To Principal Components, John Wiley & Sons, INC., (1991)
9. Wold, S. K. Esbensen and P. Geladi, Principal Component Analysis, *Chemometrics and Intel. Lab. Sys.* 2, 37-52 (1987)
10. Dunia R., S. J. Qin, T. F. Edgar and T. J. McAvoy, Use of principal component analysis for sensor fault identification, *Comp. Chem. Eng.* Vol. 20, suppl., pp. S713-S718,(1996)
11. Tong H. and C.M. Crowe, Detection persistent gross errors by sequential analysis of principal components, *Comp. Chem. Eng.* Vol. 20, suppl., pp. S733-S738,(1996)
12. Gallagher N.B. and B.M. Wise, Application of multi-way principal components analysis to nuclear waste storage tank monitoring, *Comp. Chem. Eng.* Vol. 20, suppl., pp. S739-S744,(1996)
13. Martin E.B., A.J. Morris, M.C. Papazoglou and C. Kiparissides, Batch process monitoring for consistent production, *Comp. Chem. Eng.* Vol. 20, suppl., pp. S599-S604,(1996)
14. Wold, S., P. Geladi, K. Esbensen and J. Ohman, Multi-way Principal Components and PLS-Analysis, *J. Chemometrics* 1,41-56 (1987)

15. Nomikos, P. and J. F. MacGregor, Monitoring Batch Processes Using Multiway Principal Component Analysis , *AIChE J.* 40, 8, 1361-1375(1994)
16. Nomikos, P. and J.F. MacGregor, Multivariate SPC Charts for Monitoring Batch Processes, *Technometrics* 37,1, 41-59(1995)
17. Wold, S., A. Ruhe, H. Wold and W. J. Dunn III, The Collinearity Problem in Linear Regression. The Partial Least Squares(PLS) Approach to Generalized Inverses , *SIAM J. Sci. Stat. Comput.* 5, 3(1984)
- 18 Geladi, P. and B. R. Kowalski, Partial Least-Squares Regression: A Tutorial, *Analytica Chimica Acta* 185,1-17 (1986)
19. MacGregor J. F., C. Jaeckle, C. Kiparissides and M. Koutoudi, Process Monitoring and Diagnosis by Multi-block PLS Methods, *AIChE J.* 40, 5, 826-838 (1994)
20. Kresta J. V., J. F. MacGregor and T. E. Marlin, Multivariate Statistical Monitoring of Process Operating Performance, *The Can. J. Chem. Eng.* 69, 35-47 (1991)
21. Zullo L., Validation and verification of continuous plants operating modes using multivariate statistical methods, *Comp. Chem. Eng.* Vol. 20, suppl., pp. S683-S688,1996
22. Bandoni J.C.A. and J.A. Romagnoli, Robust statistical process monitoring, *Comp. Chem. Eng.* Vol. 20, suppl., pp. S497-S502 (1996)
23. Heyen G., E. Marechal and B. Kalitventzeff, Sensitivity calculations and variance analysis in plant measurement reconciliation, *Comp. Chem. Eng.* Vol. 20, suppl., pp. S539-S544 (1996)
24. Fevotte G., I. Varudio and T.F. McKenna, Computer-aided parameter estimation and on-line monitoring of emulsion and solution polymerisation reactors, *Comp. Chem. Eng.* Vol. 20, suppl., pp. S581-S586 (1996)
25. Raich, A. and A. Cinar, Statistical Process Monitoring and Disturbance Diagnosis in Multivariable Continuous Processes," *AIChE J.*, 42, 4, pp.995-1009 (1996)
26. Hwang, D.H. and C. Han, Non-Experimental Process Improvement based on multivariate statistical analysis, *AIChE Journal*, (1999)

- (submitted) 26.
27. Cho, H.W. and C. Han, Hierarchical plant-wide monitoring and triangular representation based diagnosis, *Control Engineering Practice*, (1999) (under revision).
 28. Hong, S.J., J.H. Jung and C. Han, A design methodology of a soft sensor based on local models, *Computers chem. Engng*, 23, pp.S343-346, (1999).
 29. Lee, Y.H., D.Y. Lee and C. Han, RMBatch: intelligent real-time monitoring and diagnosis system for batch processes, *Computers chem. Engng*, 23, pp.S689-692, (1999).
 30. Hwang, D.H. and C. Han, Real-time monitoring for a process with multiple operation modes, *Control Engineering Practice*, Vol.7, No.7, pp.891-902, (1999).
 31. 홍선주, 허창구, 한종훈, PLS를 이용한 증류 공정의국부적인 조성 추정 소프트센서, *화학공학*, 제37권 제3호, pp.445-452, (1999).
 32. 이영학, 조현우, 한종훈, 시간대에 따른 국부모델을 이용한 회분식 공정의 실시간 감시 및 진단, *화학공학*, 제37권 제3호, pp.402-410, (1999).
 33. 이영학, 한종훈, 다변량 데이터 압축법을 이용한 회분식 PVC 중합 공정의 실시간 감시, *화학공학*, 제37권 제2호, pp.319-329, (1999).
 34. 홍선주, 한종훈, 증류공정의 감시, 진단 및 제어를 위한 데이터 기반 소프트센서 설계, *화학공업과 기술*, 제17권 2호, pp.172-181, (1999).
 35. 윤상엽, 철강공정에서의 전로조업 자동화를 위한 Neuro-Expert System 개발, 포항공대 박사학위 논문, (1996).
 36. 이희두, 다변량 통계 기법을 이용한 연속 교반 MMA-VA 공중합 공정 품질 변수 온라인 모니터링, 포항공대 석사학위 논문, (1997).
 37. 안태진, 고로공정에서의 송풍 조건 변화에 따른 연소대 해석, 포항공대 석사학위 논문, (1998).
 38. 정재현, 변수의 선별적 사용과 Wavelet 변환을 통한 강건한 NIR 검량 모델 구축, 포항공대 석사학위 논문, (1999).