

다변량 통계 분석 방법을 이용한 연속 교반 MMA-VA 공중합 공정 품질 변수 온라인 모니터링

이희두 · 이무호 · 조현우 · 한종훈[†] · 장근수

공정산업의 지능자동화 연구센터,
포항공과대학교 화학공학과
(1996년 11월 21일 접수, 1997년 7월 16일 채택)

Online Quality Monitoring Using Multivariate Statistical Methods in Continuous-stirred MMA-VA Copolymerization Process

Hee Du Lee, Moo Ho Lee, Hyun Woo Cho, Chonghun Han[†] and Kun Soo Chang

Automation Research Center, Department of Chemical Engineering, POSTECH
(Received 21 November 1996; accepted 16 July 1997)

요 약

본 연구에서는 공정의 과거 조업 데이터만 있으면 적용 가능한 다변량 통계 분석 방법인 PLS와 PCA를 사용하여 연속 공정의 품질 변수를 온라인으로 모니터링하는 방법을 개발하였다. 선형 회귀 방법들 중 가장 뛰어난 성능을 나타내는 PLS(partial least squares)를 사용하여 온라인으로 측정되고 있는 공정 변수들로부터 생산될 제품의 품질 변수들을 추정하였고 주성분 분석(principal component analysis, PCA)을 사용하여 추정된 품질 변수들이 비정상 상태(out-of-control state) 인지를 판별하는데 사용할 품질 변수 모니터링 모델을 구성하였다. 이 방법을 강한 비선형성이 포함된 연속 교반 MMA-VA 공중합 공정의 수식 모델로부터 생성된 데이터에 적용하였다. 결과로 개발된 PLS 품질 변수 추정기는 정적 추정기로서 현재 품질 변수 값으로부터 공정 변수들의 변화로 인한 새로운 품질 변수 값으로 수렴하기까지의 동역학은 나타내지 못하고 있지만 정상 조업 조건에 가까운 조건하에서는 PLS 품질 추정기가 $\pm 3\%$ 범위내의 오차로 정상 상태(steady state)로 수렴할 때의 품질 변수들을 추정해 내고 있으며 품질 변수들이 비정상인지를 각 변수의 상관 관계를 고려하여 판별하는 T^2 값의 실제 값과 추정 값도 거의 같게 나타나고 있다. 따라서 이 방법은 장래의 비정상 상태를 감지하여 공정을 개선하는 도구로써 사용될 수 있다.

Abstract—A scheme for online monitoring the quality of continuous-stirred MMA-VA copolymerization process using multivariate statistical methods such as PLS(partial least squares) and PCA(principal component analysis) is developed. PLS is used to estimate future steady-state quality variables and PCA is used to construct quality monitoring model to determine if estimated future steady-state quality variables are out-of-control state. While the dynamics of the process is not well described in the developed PLS quality estimator, future steady-state quality values is well predicted. Therefore, this scheme can be used as a tool to detect future abnormal states and improve the process.

Key words: Online Quality Monitoring, Multivariate Statistical Methods, PLS, PCA

1. 서 론

현대 국제적인 경쟁 사회에 있어서 공정 산업의 경쟁력에는 여러 가지가 있겠지만 그 중 가장 중요한 것 중의 하나는 생산 제품의 고품질화이다. 그리고 최근에는 고품질의 의미가 단순히 정해진 목표치를 평균적으로 달성하는데서부터 얼마나 균일하게 그 목표치를 달성하느냐 하는 것으로 변화하고 있고 이러한 제품 품질의 균일성을 이루기 위한 노력에 대한 관심이 급증하고 있다[1]. 이러한 제품의 고품질화를 이루게 되면 그 제품의 경쟁력 뿐만 아니라 off-spec.의 처리 비용을 절감하게 되어 공정의 생산성을 크게 높일 수 있게 된다.

이러한 제품의 고품질화를 이루기 위해서 각 산업 현장에서 사용되는 것이 통계적 공정 관리 또는 통계적 품질 관리[Statistical Process Control(SPC) or Statistical Quality Control(SQC)]의 기법이다. 이 SPC의 전략 중 가장 기본이 되는 것이 초기에 공정이나 생산 제품의 이상을 감지하여 대응 조치를 취할 수 있도록 하는 모니터링이다. 그러나 기존의 SPC 기법들은 한두 가지 품질 변수를 모니터링하는데는 익숙해 있으나 관리해야 할 변수의 수가 많아지면 비정상적인 상태를 잘 감지하지 못하게 된다. 최근에는 많은 변수들을 한꺼번에 그리고 서로 간의 상호관계를 고려할 수 있는 principal component analysis(PCA)나 partial least squares(PLS) 등과 같은 chemomet-

rics 방법들을 사용한 모니터링 방법들이 도입되어 크게 주목을 받고 있는데 Skagerberg 등(1992)은 연속 공정인 LDPE반응기에, Kresta 등(1991)은 fluidized bed reactor와 extractive distillation column의 모니터링에 적용하고 있으며 Nomikos 등(1994, 1995)과 Kosanovich 등(1996)은 회분식 또는 반회분식 공정의 모니터링에 이 chemometrics 방법들을 적용하고 있다. 또한 96년 5월에 그리스에서 열린 European Symposium on Computer Aided Process Engineering에서 10여 편이나 되는 관련 연구 및 응용 사례들이 발표되었다. 한편, 국내에서도 관심이 고조되면서 최근 연구 결과들이 나오고 있다[33,34].

SPC의 기법에서는 크게 네 가지 단계를 일관되게 반복적으로 수행하게 되는데 첫 번째 단계는 공정이나 품질의 이상을 조기에 감지하는 단계로 이 연구에서 수행한 모니터링 단계에 해당하고 두 번째 단계는 이상의 원인을 진단하는 단계이며 세 번째로는 이상의 원인을 제거하거나 공정을 개선하는 단계이고 마지막으로 고쳐진 공정을 검증하게 된다[1]. 그러나 일반적으로는 제품의 품질을 온라인으로 즉각 측정하지 못하고 상대적으로 드문 샘플링을 통해 실험실에서 각 특성들을 측정하게 되므로 이상이 발생했을 때 즉시 대응 조치를 취하기가 어렵다. 따라서 공정에서 실시간으로 측정되고 있는 공정 변수들로부터 품질 변수들을 추정하는 예측기[혹은 추정기(estimator)]를 구성할 필요성이 생긴다. 직접 측정하지 못하는 상태 변수를 추정하기 위한 추정기 설계에 있어서 기존[25-28]에는 주로 Kalman filter를 사용하였다. 그러나 이 방법은 많은 장점에도 불구하고 공정의 정확한 수식 모델을 필요로 한다는 점에서 한계가 있다. 이에 반해 통계적 방법을 이용한 추정기 설계는 실제 공정의 데이터를 이용하여 그리 어렵지 않은 절차를 통해 만족할 만한 결과를 보이는 추정기를 구성할 수 있으므로 실제 현장의 조업자들에게 매력적인 방법이 될 수 있다.

본 연구는 비선형성이 포함되어 실제 공정이라 가정할 수 있는 연속 교반 MMA-VA(methyl methacrylate and vinyl acetate) 공중합 공정의 수식 모델[22]을 대상으로 하였다. 현재까지는 측정 가능한 공정 변수들만을 사용하여 모니터링 모델을 구성하였으나[2, 7, 8, 10] 본 연구에서는 이 공정에 PLS를 이용하여 실시간으로 측정되고 있는 공정 변수들로부터 품질 변수들을 추정하는 정적 추정기(static estimator)를 구성하였으며 품질 변수들의 미리 설정된 요구 조건에 맞는 데이터를 사용하여 PCA로 모니터링 모델을 구성한 후 PLS 추정기로 추정된 품질 변수를 이 모델에 비교하여 이상을 감지할 수 있게 하였다. 그 결과로 추정기로 추정된 품질 변수들과 수식 모델상의 실제 값이 비교되고 있고 이상 여부를 확인하기 위한 T^2 도표에서 실제 값을 사용했을 때와 추정된 값을 사용했을 때의 T^2 값이 비교되고 있으며 그 결과에 대한 분석이 주어진다.

2. 이 론

본 연구에서 사용한 PCA와 PLS는 상관 관계가 높은(highly correlated) 다변량 데이터를 분석하기 위한 다변량 통계 분석 방법으로서 실제 공정 상에 많은 서로 상관성이 크면서도 redundant한 정보들 가운데서 중요한 정보만을 추출하고 전체 데이터의 차원을 줄일 수 있는 방법이다.

2-1. Principal Component Analysis(PCA)

PCA는 원래 직교 최소 자승법(orthogonal least squares)에 의해 평면들을 fitting하는 방법으로써 Pearson[23]에 의해 개발되었으나 그 후에 Hotelling[24]은 상호 관련 구조들을 분석하기 위한 목적으로 이 PCA를 사용하게 되었다. 이 방법은 많은 다변량 분석에 관한 책들[29-32]에서 설명하고 있는 기본적인 다변량 통계 기법이며 Wold

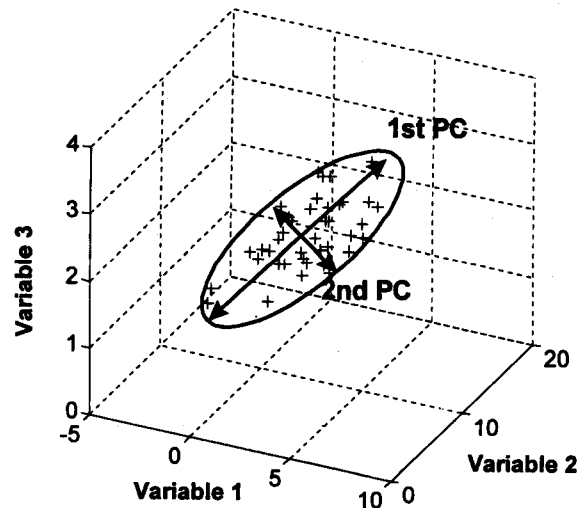


Fig 1. Principal component model of three-dimensional data set lying primarily in a single plane composed of two principal components.

[4]의 review에도 자세히 설명되어 있다. PCA는 변수들끼리 서로 상관성이 많은 자료 행렬(highly correlated data matrix) X 를 분해하여 서로 통계적으로 독립적인 새로운 변수들[주성분 혹은 잠재 변수, Principal Components(PCs) or Latent variables]로 변환시킴으로써 데이터 분포의 정보는 그대로 사용하면서도 처리 및 해석을 더욱 쉽게 해 줄 수 있게 하는 것이다. 그리고 대부분의 correlated된 데이터 공간에서는 그 변수의 차원을 줄일 수 있는 경우가 많은데 Fig. 1에서 보는 바와 같이 3차원으로 분포되어 있는 경우 1st PC(Principal Component)와 2nd PC로 구성된 2차원 평면상에 대부분의 데이터가 존재하고 있으므로 이 두 PC들로 구성된 PC 모델로 전체 구조를 나타낼 수 있게 되며 이 평면에 수직 방향인 세 번째 PC는 공정 잡음으로 간주될 수 있다.

PCA는 아래 (1)식과 같이 자료 행렬 X 로부터 T (score matrix)와 P (loading matrix)를 구하는 것으로 X 의 표본 공분산 행렬(sample covariance matrix)의 Singular Value Decomposition에 의해 구하거나 NIPALS(Nonlinear Iterative Partial Least Squares) 알고리즘에 의해 원하는 수만큼의 PC를 중요한 PC 순으로 구해 줄 수 있다[2-4, 21].

$$X = M_1 + M_2 + \dots + M_k$$

$$M_i = t_i p_i^T \quad (1)$$

where t_i : score vector(변환된 좌표축에서의 좌표값)

p_i : loading vector(변환된 좌표축의 방향 벡터)

$$X = t_1 p_1^T + t_2 p_2^T + t_3 p_3^T + \dots + t_a p_a^T + E = TP^T + E \quad (2)$$

where $a \leq k$

2-1-1. Hotelling's T^2 statistic

정규 분포를 가지는 단변량 문제에서 어떤 observation이 정상인지 아닌지를 알 수 있는 척도는 평균에서의 거리가 된다. 2개 이상의 변수가 포함된 다변량 문제에서도 각 변수가 서로 독립적인 것이라면 어떤 observation에 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 statistic이 된다[2].

먼저 과거 n 개의 다변량 표본으로부터 다음과 같이 표본 공분산 행렬 S 를 구한다. 여기서 x_i i 번째 observation의 값을 담고 있는 column vector이고 \bar{x} 는 평균 벡터(mean vector)이다.

$$S = (n-1)^{-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T \quad (3)$$

다음으로 새로운 다변량 observation x 가 얻어진다면 Hotelling's T^2 statistic은 다음과 같이 구해지며 통계량에 대한 구조적 해석은 Jackson[2]에 설명되어 있다.

$$T^2 = (x - \bar{x})^T S^{-1} (x - \bar{x}) \quad (4)$$

여기서 \bar{x} 는 목표값(target value)이다.

이와 같이 구해진 T^2 는 시간에 따른 도표로 그려지게 되는데 이 도표에서 주어지는 upper control limit(UCL)은 다음과 같이 주어진다.

$$T_{UCL}^2 = \frac{(n-1)(n+1)a}{n(n-a)} F_{\alpha}(a, n-a) \quad (5)$$

여기서 $F_{\alpha}(a, n-a)$ 는 a 와 $n-a$ 의 자유도를 가지는 F 분포의 upper 100 α % critical point이며 n 은 observation 수, a 는 component 수이며 α 는 신뢰도이다.

2-2. Partial Least Squares(PLS)

X와 Y 두 자료 행렬 사이의 선형 관계를 분석하는데 있어서 주로 사용되어 온 방법은 다중 선형 회귀법[Multiple Linear Regression (MLR)]이다. 그러나 각 행렬 내에 서로 상관성이 많은 변수들(highly correlated variables)이 포함된 경우 그 예측력은 상당히 떨어지게 된다. 이 문제를 공선형 문제(collinearity problem)라 부르는데 PLS는 이 문제를 효과적으로 처리하여 예측력이 뛰어나고 공정 잡음과 어느 정도의 센서 고장 시에도 강건한 모델을 제공할 수 있으며 최근 회귀 분석의 주된 방법으로 자리를 굳혀 가고 있다[3, 18]. PLS에서는 앞에서의 PCA를 이용하여 먼저 각 행렬의 PC들을 구한 후에 새로운 이 두 블록 사이를 선형 관계로 분석하여 서로 연관이 많은 변수들로 이루어진 블록 사이에서의 더 정확한 관계를 얻게 된다[3, 5].

PLS 모델의 개발은 다음과 같다. 우선 모니터링되어야 할 응답 변수들(여기서는 품질 변수들)의 데이터들로 $N \times M$ 응답 자료 행렬(response data matrix) Y를 구성한다. 여기서 N은 observation 수이고 M은 응답 변수(response variable)들의 수이다. 그런 후에 같은 observation 들의 집합에 대해서 관계된 예측자 변수(predictor variable)들(여기서는 온라인으로 측정되고 있는 공정 변수들)로 $N \times K$ 예측자 행렬(predictor matrix) X를 구성한다. PLS의 기하학적인 표현은 Fig. 2 와 같은데 observation들이 X 공간에서는 K 차원의 점들로 Y 공간에서는 M 차원의 점들로 보여지게 된다. 개념적으로 PLS는 (6), (7) 식과 같이 X, Y 각각에 PCA를 적용한 후 X와 Y 사이를 (8) 식과 같은 내적 관계(inner relation)로 관계 짓는다.

$$X = TP^T + E \quad (6)$$

$$Y = UC^T + F \quad (7)$$

$$U = TB + G \quad (8)$$

$$\text{where } B = (T^T T)^{-1} T^T U$$

여기서 E, F, G는 잔차 행렬(residual matrix)들이고 T, U는 $N \times A$, P는 $K \times A$, C는 $M \times A$ 그리고 B는 $A \times A$ 이며 여기서 A는 PLS component들의 수이다.

2-2-1. VIP(Variable Influence on Projection or Variable Importance in the Projection)

PLS 모델이 만들어지면 어떤 X 변수가 Y를 모델링하는데 중요한지를 PLS weight들의 정량적인 양에 의해서 알 수 있다. 그러나 고려해야 할 PLS component의 수가 3이상일 때는 PLS weight들 하나

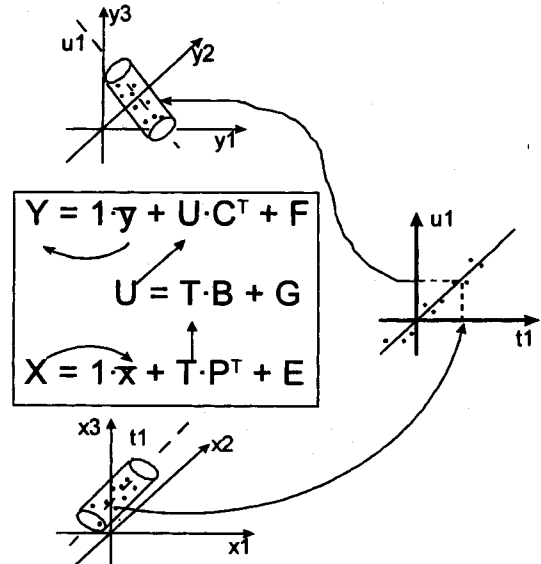


Fig. 2. A geometrical representation of PLS[18].

하나를 살펴 보는 것이 어렵게 된다. 이 때 PLS는 VIP라는 매개 변수를 도입하여 모든 차원에서 그리고 모든 Y 변수들에 대해 각 X 변수들의 중요도에 대한 정보를 제공한다[18, 19].

2-2-2. The Number of PLS Components, A[18, 19]

PLS 모델의 정확한 모델 차원을 결정하는 것은 기본적인 것이다. X 변수들이 많이 포함될수록 주어진 데이터를 더 잘 fitting할 수는 있으나 overfitting이 되어 실제로 예측력이 거의 없게 될 수도 있으므로 연속적으로 구해지는 각각의 PLS component들이 통계적으로 의미 있는 것인지에 대한 엄격한 검증 방법이 필요하다. 이 방법을 통해 몇 번째 component부터 의미 없는 것이 되는지 따라서 언제 component 구하는 절차를 멈추어야 하는지를 결정해야 한다.

Cross-validation(CV)는 PLS 분석에서 기본적인 것이 된 실제적이면서도 신뢰할 수 있는 방법이다[18-20]. 기본적으로 CV는 데이터를 많은 그룹들(예를 들면 5-10개의 그룹들)로 나눈 후에 그룹들 중에서 하나를 제외시킨 데이터를 여러 세트들로 만들게 된다. 이 세트들로 각각 PLS 모델을 만든 후에 각 모델을 만들 때 제외된 그룹이 각 모델의 검증 그룹이 되어 이 검증 그룹에 대한 Y 변수들의 실제 값과 예측 값의 차이를 계산하게 된다. 모든 동등한 모델로부터 이 차이들의 SS(sum of squares)를 계산하고 합산하게 되면 그 모델의 예측력에 대한 척도인 PRESS(predictive residual sum of squares)가 된다. 보통 PRESS는 Q^2 (the "cross-validated R^2 ")로 재표현되는데 이 값은 $(1 - \text{PRESS}/\text{SS})$ 이며 여기서 SS,는 Y의 평균에 대한 Y의 SS이다. 이 값은 $R^2 = (1 - \text{RSS}/\text{SS})$ 와 비교할 수 있는데 이 R^2 는 0에서 1 사이의 값을 가지며 1은 완전한 모델임을, 0은 전혀 관련이 없는 모델임을 나타낸다. 여기서 RSS(residual sum of squares)는 모델이 설명할 수 없는 residual들의 SS이다. 일반적으로 Q^2 도 0에서 1사이의 값을 가지게 되며 경험적으로는 R^2 가 Q^2 보다 보통 5-20% 정도 더 높게 나타나는데 실제로 X 변수 중 Y와 상관 없는 변수가 많이 포함되어 overfitting이 될수록 이 차이가 더 커지게 된다[18].

3. 적 용

3-1. Process Description and Data Generation

3-1-1. 대상 공정

PLS와 PCA를 사용한 온라인 모니터링 기법을 적용한 곳은 연속

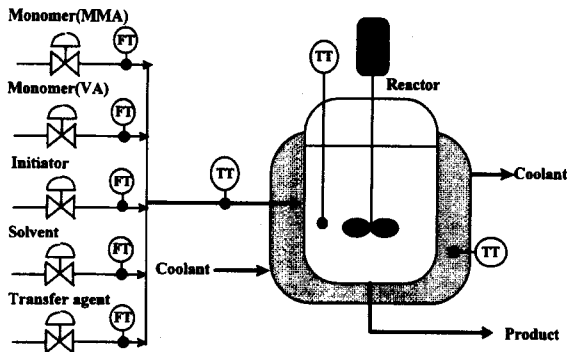


Fig. 3. Schematic diagram and sensor location of continuous-stirred MMAVA copolymerization reactor.

교반 반응기에서의 MMA-VA 용액 공중합 공정의 수식 모델이다. 이 공중합 공정의 개략도는 Fig. 3에 나타나 있는 바와 같고 이 공정의 수식 모델은 H. Ray 등[22]이 제안한 것을 사용하였다. 그리고 이 공정의 정상 상태 정상 조업 조건들(Steady State Operating Conditions)은 Table 1에 나타나 있다. 이 공정에서는 무게 평균 분자량, 수 평균 분자량, 둘 이상의 모노머가 반응하는 공중합 공정에서 중요한 품질 변수인 조성비, 그리고 전환율과 같이 제품 품질을 평가하는데 있어서 중요한 변수들이 많이 존재하고 있지만 이 중 어느 것도 쉽게 그리고 자주 측정하기가 쉽지 않으며 무엇보다도 온라인으로 이 변수들을 측정할 수 있는 센서가 없는 경우이다.

3-1-2. 모델 데이터 생성 방법

Table 2에는 이 공정에서 조절할 수 있는 조절 변수들과 이 조절 변수의 변화에 종속되어 있는 비조절 변수들을 보여 주고 있는데 이 연구에서 추정기 모델과 모니터링 모델 개발에 사용한 데이터는 조절 변수들의 정상 상태 조업 조건들에 각각 일정한 범위의 균일 분포 임의의 변이(uniform distribution random variation)를 더해 준 후 각각의 값에 대해 정상 상태로 수렴되는 비조절 변수들의 값을 수식 모델로 구하였다.

이 연구에서는 두 단계의 과정을 거쳐 데이터를 생성하였는데 우선은 정상 상태 조업 조건을 중심으로 하는 임의의 조업 조건들에서의 품질을 PLS로 추정하기 위하여 Table 3에서 보는 바와 같은 분포를 가지는 데이터를 구하였다. 각 변수의 분포는 각 변수 고유의 변화 정도를 고려하였는데 조절 변수들 중 온도의 변화 정도는 유량의 변화

Table 1. Steady-state operating conditions

On-line measured variables(Predictors), X	
Monomer A(MMA) feed rate	F(MMA) = 18.0 kg/h
Monomer B(VA) feed rate	F(VA) = 72.0 kg/h
Initiator feed rate	F(Init) = 0.18 kg/h
Solvent feed rate	F(Solv) = 36.0 kg/h
Chain transfer agent(CTA) feed rate	F(CTA) = 2.7 kg/h
Reactor feed temperature	T_f = 353.15 K
Reactor jacket temperature	T_j = 330.00 K
Reactor temperature	T_r = 341.35 K
Reactor parameters	
Reactor volume	V_r = 1.0 m ³
Reactor heat transfer area	S_r = 4.6 m ²
Product quality variables monitored(Response variables), Y	
Melt index	MI = 5.321
Weight average molecular weight	M_w = 43,950
Number average molecular weight	M_n = 29,000
Weight fraction of B in polymer	Y_w = 0.254
Conversion	Conv = 0.235

Table 2. Manipulated and unmanipulated(dependent) variables

Manipulated variables	
Monomer A(MMA) feed rate	F(MMA)
Monomer B(VA) feed rate	F(VA)
Initiator feed rate	F(Init)
Solvent feed rate	F(Solv)
Chain transfer agent(CTA) feed rate	F(CTA)
Reactor feed temperature	T_f
Reactor jacket temperature	T_j
Unmanipulated(dependent) variable	
Melt index	MI
Weight average molecular weight	M_w
Number average molecular weight	M_n
Weight fraction of B in polymer	Y_w
Conversion	Conv
Reactor temperature	T_r

Table 3. Statistics for X & Y observations used in PLS modeling

	Variable	Min	Max	Mean	Median	StdDev
X	F1(MMA)	14.4	21.6	18.18	18.3	2.129
	F2(VA)	57.6	86.4	72.38	72.7	8.154
	F3(Init)	0.144	0.216	0.1767	0.174	0.02109
	F4(Solv)	28.8	43.2	36.01	35.8	4.078
	F5(CTA)	2.16	3.24	2.686	2.67	0.3107
	T_f	336	371	352.7	353	10.21
	T_j	322	346	332	332	4.846
	T_r	335	358	343.1	343	3.973
Y	MI	3.14	9.66	5.945	5.9	1.634
	M_w	38,000	50,000	43,190	42,700	3,403
	M_n	25,100	33,000	28,500	28,200	2,223
	Y_w	0.196	0.34	0.2573	0.256	0.02338
	Conv	0.162	0.329	0.2387	0.237	0.03046

Table 4. Statistics for Y observations used in PCA modeling

Variable	Min	Max	Median	Mean	StdDev
MI	4.22	5.71	5.045	5.007	0.424
M_w	4,3100	4,6900	4,4500	44,700	1071
M_n	2,8500	3,0900	2,9400	2,9500	698.6
Y_w	0.241	0.263	0.251	0.2517	0.00438
Conv	0.219	0.243	0.231	0.2311	0.004795

정도보다 더 작도록 하여 정상 상태 조업 조건을 중심으로 임의로 생성되었다. 그 다음으로 이 데이터 중 생산자가 원하는 사양(specification)에 맞는 품질 변수의 범위에 속하는 것만 골라서 Table 4와 같은 통계량을 나타내는 데이터를 구하였는데 이 두 번째 데이터로 PCA를 이용한 모니터링 모델을 구성하게 된다. 여기서 제품의 사양은 경우에 따라 달라질 수 있으며 본 연구에서는 무게 평균 분자량이 42,900에서 46,900 사이이고 조성비가 0.2323에서 0.2723사 이의 값을 나타낼 때의 데이터만 모아서 Table 4와 같은 모니터링 모델 데이터를 얻었다. 실제 공정에서는 Table 3에서와 같이 정상 상태 조업 조건을 중심으로 넓은 영역에 걸쳐 균일한 분포를 가지는 데이터를 얻기란 조금 힘든 일이 될 수도 있으나 이용 가능한 모든 데이터를 사용하면 비슷한 결과를 얻을 수 있다. 단 정상 상태 조업 조건에서 너무 많이 벗어나는 데이터를 포함시키게 되면 PLS 추정기의 예측력이 현저히 떨어지게 될 수도 있는데 이는 PLS가 선형 회귀 분석 방법의 한 종류이므로 비선형 공정의 넓은 영역에 대해서 적용하기는 어렵기 때문이다.

3-2. 품질 변수 추정을 위한 PLS 모델의 구성

이 연구에서는 우선 온라인으로 측정되고 있는 공정 변수들로부터

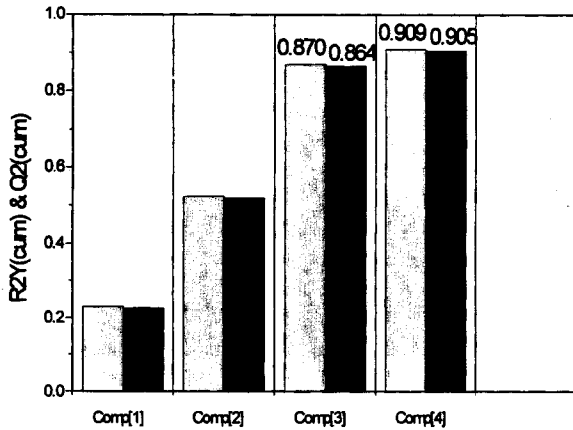


Fig. 4. Cumulative R^2Y & Q^2 explained by PLS model.

[□ : $R^2Y(cum)$, ■ : $Q^2(cum)$].

터 최종 제품의 품질을 추정하기 위하여 PLS를 이용하였다. 이론에서 설명한 바와 같이 PCA나 PLS는 많은 서로 상관 관계를 이루는 정보(여기서는 공정 변수들 또는 품질 변수들)로부터 그 시스템을 잘 설명하면서도 차원(변수 수)을 줄일 수 있는 방법들이다. 본 연구에서 품질 변수들을 온라인으로 측정되는 공정 변수들로 추정하는 추정 모델을 PLS로 모델링한 결과는 Fig. 4에서 보는 바와 같이 나타난다. 이 그림에서는 PLS component의 수를 1에서부터 증가시킬 때마다 축적되는 R^2Y 와 Q^2 값을 보여 주고 있다. PLS component 4개를 사용한 PLS 모델이 모델을 구성하기 위해 주어진 데이터의 90.9%(R^2Y)를 설명하고 있으며 예측력을 나타내는 Q^2 값은 0.905로 이 모델이 품질 변수를 잘 예측하고 있음을 보여 준다. Fig. 5에서는 공정 변수들의 VIP 값을 보여 주고 있는데 이 값이 1이상인 변수들이 PLS 모델링 시 유의한 변수들이다. 이 그림에서 보는 바와 같이 MMA(methyl methacrylate), VA(vinyl acetate), 개시제(initiator)의 유량 그리고 반응기의 온도의 순서로 품질 변수에 영향을 미치는 정도가 큰 것을 알 수 있다. Fig. 6은 각 품질 변수에 대한 정규화된 회귀 계수를 보여 주고 있는데 각 공정 변수가 각 품질 변수에 미치는 영향을 표준화된 수치로 비교해 볼 수 있다. 예를 들어 무게 평균 분자량과 각 공정 변수 사이의 관계를 보면 MMA와 VA의 유량은 양의 상관관계를 가지고 있으며 MMA의 유량이 VA의 유량보다 무게 평균 분자량에 미치는 영향이 크다. 개시제의 유량은 MI, 무게 평균 분자량, 수 평균 분자량에 미치는 영향이 크게 나타나고 있으며 용매의 유량이나 CTA의 유량 변화는 품질 변수에 미치는 영향이 상대적으로 적음을 확인할 수 있다.

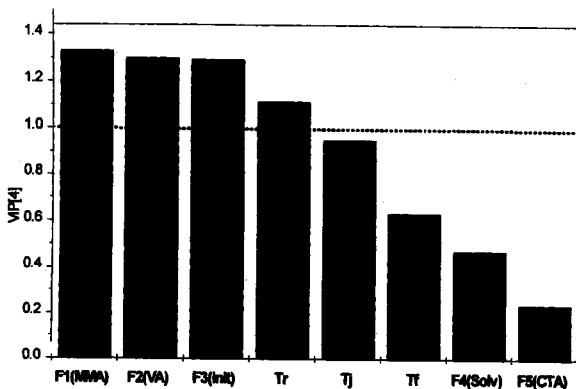


Fig. 5. VIP(Variable Influence on Projection) plot in PLS modeling.

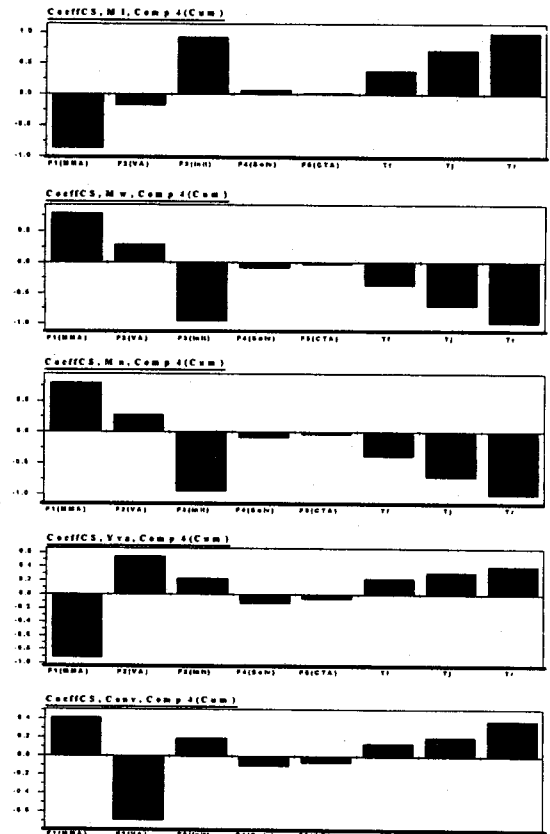


Fig. 6. Centered & scaled regression coefficients of PLS model.

3-3. 품질 변수 모니터링을 위한 PCA 모델의 구성

품질 변수의 off-spec.을 감지하기 위한 PCA 모니터링 모델의 개관이 Fig. 7에 나타나 있다. 그림에서 보는 바와 같이 3개의 component가 전체 품질 변수 데이터 정보의 99.9%를 설명하고 있다. 따라서 이 3개의 component만을 사용하여 품질 변수들을 모니터링하는 도표와 T^2 제어 한계를 구하였다. 품질 변수의 비정상 상태를 판별할 기준이 되는 T^2 제어 한계는 (5) 식에 의해 95% 신뢰도의 T^2 제어 한계는 8.527이고, 99% 신뢰도의 T^2 제어 한계는 12.69이다.

3-4. 적용 결과 및 고찰

본 연구에서는 3-2에서 만든 추정 모델로부터 각 입력 변수들에 다양한 크기의 계단형 외란이 들어왔을 때의 제품의 품질 변수들을 추정하고 이들의 이상 여부를 PCA로 구성한 모니터링 모델로 비교

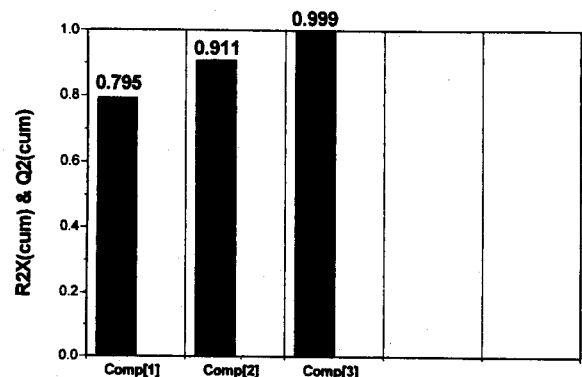


Fig. 7. Cumulative R^2X explained by PCA model[■ : $R^2X(cum)$].

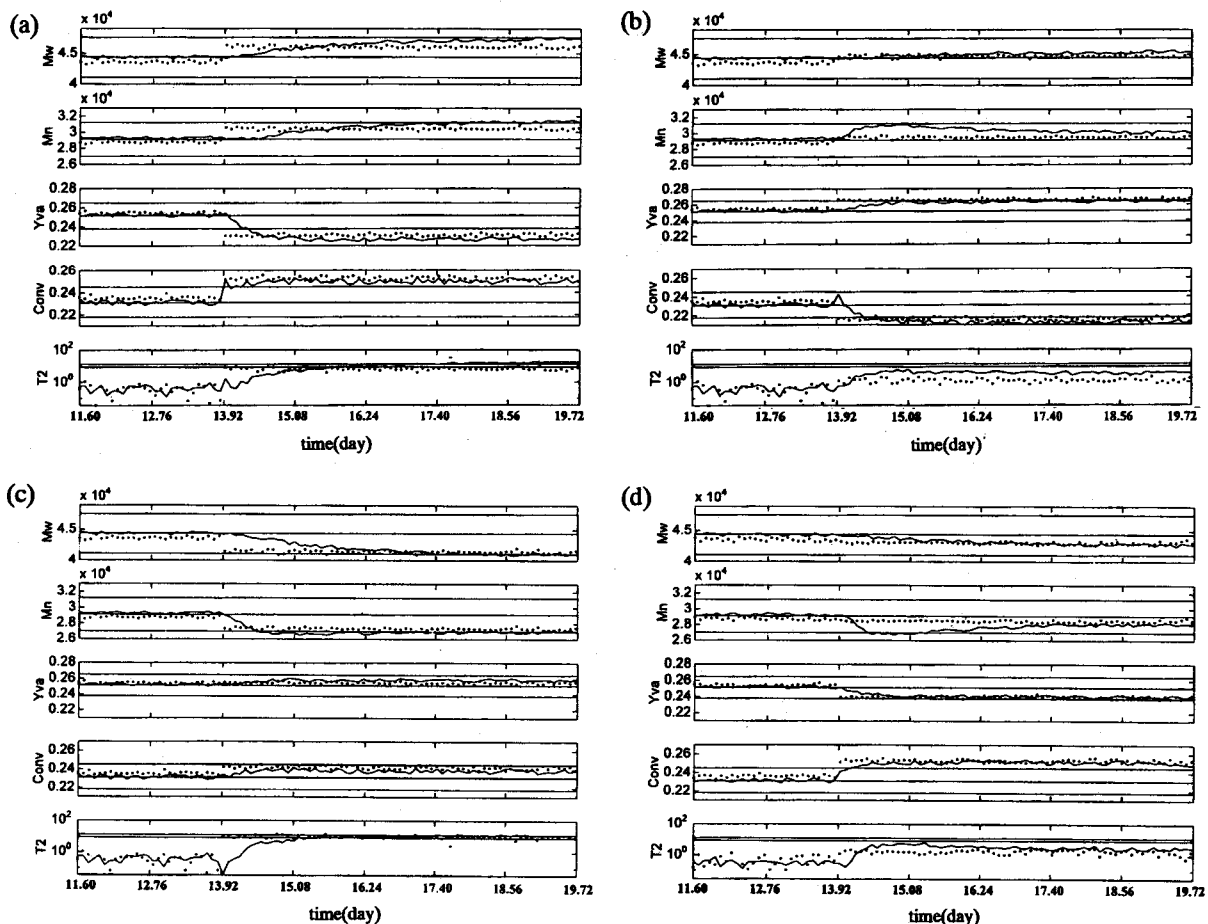


Fig. 8. Quality estimation and T^2 chart monitoring results (a) step disturbance of 2 kg/hr to NOC(normal operation condition) occurred in F(MMA) at $t=1.2 \times 10^6$ s; (b) step disturbance of 4.2 kg/hr in F(MMA); (c) step disturbance of 8 kg/hr in F(VA); (d) step disturbance of 16.3 kg/hr in F(VA); (e) step disturbance of 0.02 kg/hr in F(Init); (f) step disturbance of 0.0042 kg/hr in F(Init); (g) step disturbance of 2 kg/hr in F(MMA) and 0.02 kg/hr in F(Init); (h) step disturbance of 4.2 kg/hr in F(MMA) and 0.04 kg/hr in F(Init). (• : estimated value, — : actual value).

해 보았다. Fig. 8(a)-(d)는 Fig. 5의 VIP plot에서 PLS 모델링 시 가장 영향을 많이 끼치는 것으로 나타난 MMA와 VA, 그리고 initiator의 유량에 각각 그림에 나타난 것과 같은 계단형 외란이 들어왔을 때 생산될 제품의 품질의 실제 공정의 응답과 PLS로 추정된 공정 응답을 나타낸 것이다. 각각의 그림에서 위 4개의 plot은 각각 무게 평균 분자량, 수 평균 분자량, 조성비, 전환율의 실제 값과 추정 값을 비교하여 나타내고 있으며 각 그림의 최하단에 있는 도표는 시간에 따른 T^2 를 로그 스케일로 보여 주고 있는데 위와 동일하게 실제 품질 변수들을 사용하여 구한 값과 추정한 품질 변수 값을 사용하여 구한 T^2 값이 비교되어 보여지고 있다. 각 그림에서 실선은 실제 값을 나타내고 점선은 PLS 추정기에 의해 추정된 품질 변수 값을 나타내는데 실제 값의 $\pm 3\%$ 범위로 추정하고 있으며 각 공정 변수들 즉 MMA, VA 그리고 개시제의 유량의 정상 상태 조업 조건에서 많이 벗어나는 외란이 발생할수록 이 실제 값과 추정 값 사이의 차이는 커진다. 이는 PLS가 목표치를 중심으로 공정을 선형화하여 추정하는 방법이기 때문에 발생하는 현상이다. 그리고 공정 모니터링 시에 이상 여부를 판단하는데 사용되는 T^2 도표가 각 그림에서 맨 하단에 나타나 있는데 실제 값을 사용하여 구한 것과 추정된 값을 사용한 것이 거의 비슷하게 나타남을 볼 수 있다. 앞의 VIP plot에서 살펴 보았듯이 품질 변수에 가장 영향을 미치는 변수들은 MMA, VA, 개시제로서 Fig. 8에서도 그 영향의 크기를 살펴볼 수 있다. 즉, Fig. 8(a)와 (b)를 (c)와 비교해 보면 M_w 와 M_n 변수에서 특

히 오차가 증가하게 되는데 이것은 MMA와 VA가 개시제보다 품질 변수에 미치는 영향이 더 큰 결과로 보여진다. 반면에 MMA와 개시제 모두에 외란이 들어간 Fig. 8(d)의 경우에는 품질 변수에 영향을 미치는 두 개의 변수에 외란이 동시에 들어갔음에도 불구하고 (a), (b), (c)와 비교해서 아주 작은 오차를 보이고 있다. 이것은 Fig. 6에서 알 수 있듯이 M_w 와 M_n 에 미치는 MMA와 개시제의 계수가 서로 다른 부호를 가져 외란이 상쇄되는 효과로 인한 것을 알 수 있다.

본 연구에서 구성한 PLS 추정기는 정상 상태에서의 데이터들로 구성하였으므로 실제 공정의 동역학이 존재하는 부분은 잘 추정하지 못하고 있는데 이는 Fig. 8의 각 그림에서 실제 값과 추정 값 사이를 비교해 보면 알 수 있는 것처럼 결국 PLS로 추정하는 값은 현재의 외란이 미래에도 계속된다는 가정하에서 미래의 새로운 정상 상태에서의 수렴 값을 계산한 것이 된다. 이는 오히려 우리가 PLS로 추정된 값들을 모니터링하고 있을 때 T^2 도표에서 이상이 발생되었다고 나타나더라도 그 이상이 당장에 나타난 것은 아닐 수 있으므로 이상 상태로 가기 전에 공정을 교정하여 미리 이상 상태를 방지해 줄 수 있는 역할을 할 수 있을 것이다.

4. 결 론

본 연구에서는 연속 교반 MMA-VA 공중합 공정에서 대표적인 다변량 통계 분석 방법들 중 하나인 PLS(Partial Least Squares)를 통

하여 온라인으로 측정하고 있는 공정 변수들로서 온라인으로 측정하기 어려운 품질 변수들을 추정하는 PLS 추정기를 구성하였고 이 추정기를 통하여 추정된 품질 변수들을 미리 사양에 맞는 또는 정상 조업 조건(normal operating condition)에 맞는 값들을 사용하여 PCA로 구성된 모니터링 모델과 비교하여 현 공정의 이상 여부를 판단할 수 있도록 하였다.

개발된 추정기는 공정의 동특성은 나타낼 수 없었지만 정상 상태로 수렴할 때의 값을 잘 예측할 수 있었으며 이로써 장차 발생하게 될 이상 상태를 미리 감지하여 조치할 수 있는 도구로서 쓰일 수 있다.

감 사

본 연구는 한국과학재단 지정 우수연구센터인 공정산업의 지능자 동화 연구센터와 교육부 지정 포항공과대학교 환경공학부의 지원으로 이루어진 것으로 이에 감사를 드립니다.

사용기호

A	: number of PLS components retained
a	: PCA model dimension
B	: regression coefficient in PLS inner relation
Conv	: conversion
E	: residual error in modeling data matrix X
F	: F distribution value, residual error in modeling data matrix Y
G	: residual error in modeling data matrix U
k	: number of columns in data matrix X
MMA	: methyl methacrylate
M_n	: number average molecular weight
M_w	: weight average molecular weight
n	: number of observations
P	: PCA or PLS model loadings for change of coordinates
p	: loadings for one dimension of PCA or PLS model
PC	: principal component
Q^2	: cross-validated coefficient of determination
R^2	: coefficient of determination
R^2_X	: coefficient of determination in X
R^2_Y	: coefficient of determination in Y
RSS	: residual sum of squares
SPC	: statistical process control
SQC	: statistical quality control
SS	: sum of squares
StdDev	: standard deviation
T	: PCA or PLS score matrix for process measurements
T^2	: Hotelling's T-squared statistic
t	: PCA score for an observation as a column vector
UCL	: upper control limit
VA	: vinyl acetate
VIP	: variable influence on projection or variable importance in the projection
X	: process data matrix
x	: process data observation as a column vector
\bar{x}	: sample mean as a column vector
Y	: quality data matrix

그리이스 문자

α	: confidence level
τ	: target value

상첨자

T	: transpose
---	-------------

하첨자

f	: feed to the reactor
i	: initiator, instantaneous
j	: cooling jacket, stream counter
k	: component counter
n	: number average polymer property
p	: propagation, dead polymer
q	: number of B units in polymer chain
r	: reactor
s	: solvent
w	: weight average polymer property

참고문헌

1. Montgomery D. C.: "Introduction to Statistical Quality Control", 2nd Ed., John Wiley & Sons, inc., New York(1991).
2. Jackson, J. E.: "A User's Guide to Principal Components", John Wiley & Sons, Inc., New York(1991).
3. Geladi, P. and Kowalski, B. R.: *Analytica Chimica Acta*, **185**, 1 (1986).
4. Wold, S.: *Chemometrics and Intelligent Laboratory Systems*, **2**, 37(1987).
5. Hoskuldsson, A.: *J. Chemometrics*, **2**, 211(1988).
6. Skagerberg, B., MacGregor, J. F. and Kiparissides, C.: *Chemometrics and Intelligent Laboratory Systems*, **14**, 341(1992).
7. Nomikos, P. and MacGregor, J. F.: *AIChE J.*, **40**(8), 1361(1994).
8. Nomikos, P. and MacGregor, J. F.: *Technometrics*, **37**(1), 41(1995).
9. Kusanovich, K. A., Dahl, K. S. and Piovoso, M. J.: *Ind. Eng. Chem. Res.*, **35**, 138(1996).
10. Kresta, J. V., MacGregor, J. F. and Marlin, T. E.: *The Canadian Journal of Chemical Engineering*, **69**, 35(1991).
11. Gallagher, N. B. and Wise, B. M.: *Comp. Chem. Eng.*, **20**, suppl., S739(1996).
12. Martin, E. B., Morris, A. J., Papazoglou, M. C. and Kiparissides, C.: *Comp. Chem. Eng.*, **20**, suppl., S599(1996).
13. Bandoni, J. C. A. and Romagnoli, J. A.: *Comp. Chem. Eng.*, **20**, suppl., S497(1996).
14. Kourti, T., Lee, J. and MacGregor, J. F.: *Comp. Chem. Eng.*, **20**, suppl., S745(1996).
15. Zullo, L.: *Comp. Chem. Eng.*, **20**, suppl., S683(1996).
16. Dunia, R., Qin, S. J., Edgar, T. F. and McAvoy, T. J.: *Comp. Chem. Eng.*, **20**, suppl., S713(1996).
17. Tong, H. and Crowe, C. M.: *Comp. Chem. Eng.*, **20**, suppl., S 733(1996).
18. Eriksson, L., Hermens, J. L. M., Johansson, E., Verhaar, H. J. M. and Wold, S.: *Aquatic Sciences*, **57**(3), 1015(1995).
19. SIMCA P 2.1 manual, Umetri AB, P.O. Box 7960, 90719 Umeå, Sweden, 1994.
20. Wold, S.: *Technometrics*, **20**, 387(1978).
21. Zhang, J., Martin, E. B. and Morris, A. J.: *Trans IChemE*, **74**,

- 89(1996).
22. Congalidis, J. P., Richards, J. R. and Ray, W. H.: *AIChE J.*, **35**(6), 891(1989).
23. Pearson, K.: *Phil. Mag.*, ser 6, 559(1901).
24. Hotelling, H.: *J. Educat. Psychol.*, **24**, 417(1933).
25. McAuley, K. B. and MacGregor, J. F.: *AIChE J.*, **39**(5), 855(1993).
26. Lee, J. H. and Datta, A. K.: *AIChE J.*, **40**(1), 50(1994).
27. McAuley, K. B. and MacGregor, J. F.: *AIChE J.*, **37**(6), 825(1991).
28. Ellis, M. F., Taylor, T. W. and Jensen, K. F.: *AIChE J.*, **40**(3), 445(1994).
29. Johnson, R. A. and Wichern, D. W.: "Applied Multivariate Statistical Analysis", 3rd edition, Prentice Hall(1992).
30. Anderson, T. W.: "Introduction to Multivariate Statistical Analysis", 2nd edition, Wiley, New York(1984).
31. Marida, K. V., Kent, J. T. and Bibby, J. M.: "Multivariate Analysis", Academy Press, London(1982).
32. Morrison, D. F.: "Multivariate Statistical Methods", 2nd edition, McGraw-Hill(1976).
33. 황대회, 조현우, 한종훈, 김재연: *화학공업과 기술*, **15**(3), 247 (1997).
34. 조현우, 장근수, 한종훈: *한국화학공학회 춘계학술발표회초록집* (1997).