

공정변수를 갖는 혼합물 실험 자료의 분석

임 용 빈[†]

이화여자대학교 자연과학대학 통계학과

Analysis of mixture experimental data with process variables

Yong B. Lim[†]

Department of Statistics, Ewha Womans University

Purpose: Given the mixture components - process variables experimental data, we propose the strategy to find the proper combined model.

Methods: Process variables are factors in an experiment that are not mixture components but could affect the blending properties of the mixture ingredients. For example, the effectiveness of an etching solution which is measured as an etch rate is not only a function of the proportions of the three acids that are combined to form the mixture, but also depends on the temperature of the solution and the agitation rate. Efficient designs for the mixture components - process variables experiments depend on the mixture components - process variables model which is called a combined model. We often use the product model between the canonical polynomial model for the mixture and process variables model as a combined model.

Results: First we choose the reasonable starting models among the class of admissible product models and practical combined models suggested by Lim(2011) based on the model selection criteria and then, search for candidate models which are subset models of the starting model by the sequential variables selection method or all possible regressions procedure.

Conclusion: Good candidate models are screened by the evaluation of model selection criteria and checking the residual plots for the validity of the model assumption. The strategy to find the proper combined model is illustrated with examples in this paper.

Key Words : Mixture Components - Process Variables Experiments, Strategy to Find Proper Combined Model, Model Selection Criteria

1. 서 론

반도체와 LCD 분야에서의 식각용액의 혼합비, 식품
 산업에서의 과즙음료의 혼합 비율의 결정, 표백력과 세

탁력을 최적화하는 세탁제의 혼합비, 알약의 코팅에 사
 용되는 유액의 혼합비 등 혼합물을 활용하는 제조업의
 분야는 광범위하다. 혼합물 실험의 특징은 혼합물 성분
 들의 비율의 합은 1이고, 제품의 품질 특성과 연계된
 반응변수는 혼합물의 총량에는 영향을 받지 않고, 성분
 들의 혼합비율만의 함수로 가정된다. 예를 들면, 사과,
 파인애플, 오렌지 주스를 섞어서 과일주스를 만드는 경
 우에, 일반적으로 과일주스의 맛은 3 종류의 주스의 혼
 합 비율에만 종속하고, 만들어진 주스의 총량에는 영향
 을 받지 않는다. 즉, 혼합비율만 동일하면, 과일주스를
 1리터 만들든지 2리터 만들든지 상관없이 맛은 동일하

[†] Corresponding Author

Department of Statistics, Ewha Womans University,
 Daehyundong, Seodaemoongu, Seoul 120-750, Korea
 Tel:+ 82.2.3277.2302 Fax: + 82.2.3277.3607
 E-mail: yblim@ewha.ac.kr

※ 이 논문은 2010년도 정부(교육과학기술부)의 재원으로 한국
 연구재단의 기초연구사업 지원을 받아 수행된 것임
 (2011-0026076)

리라 가정한다. 많은 혼합물 실험에서 제품의 품질 특성은 혼합물 성분들의 비율에 따라서 영향을 받을 뿐만 아니라, 혼합물을 처리하는 공정의 조건에 따라서도 달라진다. 공정변수는 혼합물의 성분과는 관련이 없는 인자이지만, 수준 값이 달라지면 혼합물의 혼합성질에 영향을 줄 수 있다. 즉, 혼합 비율이 동일한 혼합물일 지라도, 공정의 조건을 달리하면, 제조된 혼합물 완제품의 품질특성인 반응변수의 값이 달라질 수 있다. Cornell (1990)에서 논의된 예를 들면, 생선 패티 혼합물이 민어, 송어, 도미를 혼합하여 구성되는 경우에, 생선 까스의 맛은 생선의 혼합 비율에 따라서도 영향을 받지만, 생선 패티 혼합물의 가공 조건을 결정하는 기름에 튀기는 시간, 요리 온도, 요리 시간에 따라서도 영향을 받는다. 이와 같이 혼합물 성분-공정변수 실험은 나노화학, 식품공학, 제약 분야 등 소비자 제품을 생산하는 많은 산업 분야에서 흔히 접하게 된다.

혼합물 성분비와 공정변수들 간의 관계를 설명하는 혼합물 성분-공정변수의 모형을 결합모형(combined model)이라고 하는데, 이 결합모형으로는 혼합물 모형과 공정변수 모형의 교적모형(product model)을 활용할 수 있고, 결합설계(combined design)로는 혼합물 설계와 공정변수 설계의 교적설계를 활용한다. 그런데, 혼합물 성분들의 개수인 q 와 공정변수들의 개수인 p 가 증가함에 따라서 교적모형의 항들의 개수와 혼합물 성분-공정변수 결합설계의 크기는 굉장히 커진다. Kowalski 등 (2000), Myers & Montgomery (2009)에서는 실용적인 모형으로 교적모형의 일부 항들로 구성된 부분모형을 제안하고, 실험설계로는 심플렉스 중심배열법과 2수준 일부실시법의 교적설계에 의한 결합설계의 부분 실시법을 제안한다. 임(2011)에서는 공정변수의 곡선효과와 혼합물 성분과 공정변수간의 3차 교호작용 효과 까지를 반영하는 실용적인 모형을 제시하고, 축점을 포함하는 심플렉스중심설계와 중심합성설계의 교적설계의 실험점들을 실용적인 실험설계의 후보 실험점으로 간주하여 엔지니어들이 허용할 수 있는 전체 실험의 크기가 주어 졌을 때에 D-, G-, I- 최적기준에 강건한 실용적인 실험설계를 후보 실험점들로 부터 선택하는 방법과 실용적인 실험 설계들이 제시된다.

이 논문의 목적은 혼합물 성분비와 공정변수들에 관한 실험 자료가 주어졌을 때에, 주어진 실험 자료를 잘 설명하는 적절한 결합모형을 찾는 것이다. 우선 모형 선택 기준에 부합하는 시작모형의 후보들을 교적모형의 범주와 임(2011)에서 제시된 실용적인 모형들 중에

서 찾는다. 다음으로 선택된 시작모형을 완전모형으로 간주하여, 모형의 간결성의 원칙에 따라서 완전모형의 부분모형으로 구성된 적절한 결합모형에 대한 후보모형들의 묶음을 구성하는 것이다. 완전모형의 부분모형은 축차적인 변수 선택법이나 모든 가능한 회귀 방법에 의한 변수 선택법(All possible regressions procedure)에 의해서 선별되고, 적절한 후보모형들의 묶음에 포함된다. 묶음에 있는 모형들의 모형 성능 통계량의 값의 비교와 표준화 잔차 그림의 검토를 통해서 오차의 정규성과 등분산성을 확인한 후에 최종적으로 적절한 모형을 복수로 추천한다. 2절에서는 혼합물 성분-공정변수의 교적모형과 임(2011)이 제안한 세 가지 형태의 혼합물 성분-공정변수에 관한 시작모형을 소개한다. 3절에서는 주어진 혼합물 성분비와 공정변수들에 관한 실험자료를 잘 설명하는 적절한 결합모형의 선택을 위한 전략이 논의된다. 4절에서는 잘 알려진 혼합물성분-공정변수 실험 자료에 대해서 Design Expert 8.0을 활용한 적절한 결합모형 찾기가 예시된다.

2. 혼합물 성분-공정변수 결합모형

q 개 성분의 혼합물에 있어서 x_i 를 i 번째 성분의 혼합 비율이라고 하면, 혼합비율들은 다음과 같은 관계식을 만족해야 하기에,

$$x_1 + x_2 + \dots + x_q = 1, \quad x_i \geq 0, \quad i = 1, 2, \dots, q \quad (1)$$

혼합물 실험의 실험 공간은 $q-1$ 차원 심플렉스 공간이 된다. 혼합물 실험공간에서의 모형은 혼합비율들의 합이 1이라는 식 (1)의 제약조건 때문에 상수항과 순수 2차항들을 포함하지 않는 Scheffe의 정준 다항식을 가정하고, 곡선효과를 반영하는 가장 간단한 모형인 이차 모형은

$$E(y) = \sum_{i=1}^q \beta_i x_i + \sum_{i < j}^q \beta_{ij} x_i x_j \quad (2)$$

이고 삼차모형은

$$E(y) = \sum_{i=1}^q \beta_i x_i + \sum_{i < j}^q \beta_{ij} x_i x_j + \sum_{i < j}^q \delta_{ij} x_i x_j (x_i - x_j) + \sum \sum_{i < j < k}^q x_i x_j x_k \quad (3)$$

이다. 공정변수들을 갖는 혼합물 실험에서는 q 개의 혼합물 성분을 갖는 혼합물 실험이 p개의 공정변수를 갖는 공정조건에서 수행된다. 코우드화된 공정변수의 실험공간은 $-1 \leq z_j \leq 1, j=1, \dots, p$ 을 포함하는 p차원의 하이퍼 큐브(hyper-cube)이고, 공정변수만의 모형은 가장 간단한 선형효과만을 가정한 경우에는 일차항 z_j 를 포함하고(L로 표시), 교호작용들이 존재하는 경우에는 $z_j z_k$ 항들을 추가로 포함하고(2FI로 표시), 공정변수들의 곡선효과를 모형에 반영하는 경우에는 순수 이차항인 z_j^2 들을 추가로 포함한다(Q로 표시). 이제 q 개의 혼합물 성분과 p개의 공정변수를 갖는 혼합물 성분-공정변수 실험에 대한 결합모형(combined model)을 생각해 보자. 일반적으로 혼합물 성분-공정변수 실험에서 주된 관심사는 혼합물 성분 비율들이 반응변수에 미치는 영향이기에, 결합모형에서 혼합물에 대한 모형으로는 혼합물 성분비율들의 곡선효과를 반영한 2차 모형, 3개의 성분들간의 교호작용효과인 $x_i x_j x_k$ 를 포함하는 특별3차 모형 혹은 3차 모형을 가정한다.

혼합물-공정변수 실험 자료의 분석을 위한 시작모형으로 혼합물 모형과 공정변수 모형의 교적모형(product model)을 고려할 수 있다. 예를 들면, 혼합물 모형으로 2차 모형을 가정하고, 공정변수 모형으로 일차모형을 가정한 경우에 교적모형은 다음과 같이 주어진다.

$$E(y) = \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^{q-1} \sum_{i < j}^q \beta_{ij} x_i x_j + \sum_{k=1}^p \left[\sum_{i=1}^q \alpha_{ik} x_i + \sum_{i=1}^{q-1} \sum_{i < j}^q \alpha_{ijk} x_i x_j \right] z_k \quad (4)$$

다음으로 교적모형의 일부 항들로 구성된 간단한 형태의 결합모형을 생각해 보자. 첫 번째 형태의 결합모형인 모형1은 혼합물 성분-공정변수에 관한 가장 간단한 모형으로 2차의 혼합물모형과 일차의 공정변수 모형을 고려하여, 교적모형의 혼합물 성분비율과 공정변수와의 교호작용효과에서 혼합물 성분비율의 일차항들과 각각의 공정변수와의 교호작용인 이차교호작용효과만 인정하는 것이다. 즉, 결합모형의 간결성과 항들에 관한 간단한 해석을 위하여 혼합물 성분비율과 공정변수들간의 3차 이상의 교호작용효과는 무시한 모형으로 아래와 같이 주어진다.

$$E(y) = \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^{q-1} \sum_{i < j}^q \beta_{ij} x_i x_j + \sum_{k=1}^p \sum_{i=1}^q \alpha_{ik} x_i z_k \quad (5)$$

$$E(y) = \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^{q-1} \sum_{i < j}^q \beta_{ij} x_i x_j + \sum_{k=1}^p \sum_{i=1}^q \alpha_{ik} x_i z_k + \sum_{k=1}^p \sum_{i=1}^{q-1} \sum_{i < j}^q \alpha_{ijk} x_i x_j z_k + \sum_{k=1}^p \sum_{i=1}^q \alpha_{ikk} x_i z_k^2 \quad (6)$$

세 번째 형태의 모형인 모형3은 공정변수들의 이차 교호작용효과와 혼합물 성분의 선형 효과들간의 교호작용인 3차 교호작용 효과의 항들을 두 번째 형태의 모형에 추가하는 것으로 다음과 같이 주어진다.

$$E(y) = \sum_{i=1}^q \beta_i x_i + \sum_{i=1}^{q-1} \sum_{i < j}^q \beta_{ij} x_i x_j + \sum_{k=1}^p \sum_{i=1}^q \alpha_{ik} x_i z_k + \sum_{k=1}^p \sum_{i=1}^{q-1} \sum_{i < j}^q \alpha_{ijk} x_i x_j z_k + \sum_{k=1}^p \sum_{i=1}^q \alpha_{ikk} x_i z_k^2 + \sum_{k=1}^{p-1} \sum_{l=k+1}^p \sum_{i=1}^q \delta_{ikl} x_i z_k z_l \quad (7)$$

적절한 결합모형의 후보는 앞에서 언급된 시작모형의 후보인 교적모형, 모형1, 모형2, 모형3의 부분모형으로, 항들에 대한 위계 질서가 존중된 모형이다. 즉, 삼차항인 $x_1 x_2 z_1$ 이 모형에 선택된 경우에는 관련된 2차항인 $x_1 x_2, x_1 z_1, x_2 z_1$ 는 추가 설명력의 여부에 관계없이 모형에 포함된다. 부분모형에서 항들에 대한 위계질서가 존중된 모형만을 고려하는 이유는 위계질서가 존중된 모형에서는 혼합물 성분들이 제약조건이 있는 경우에 코우드화된 성분인 pseudocomponents와 공정변수들의 코우드화 변수들에 대한 모형과 원래 혼합물 성분과 원래 공정변수에 대한 모형이 서로 동등하기 때문에, pseudocomponents와 공정변수들의 코우드화 변수들에 대한 부분모형의 비교를 통해서 적절한 모형을 추천할 수 있기 때문이다.

3. 적절한 결합 모형 선택

혼합물-공정변수 실험 자료가 주어질 경우에, 혼합물-공정변수 실험 자료의 분석을 위해서 주어진 실험 자료를 잘 설명하는 적절한 모형을 찾는 것이 중요하다. 이를 위한 우리의 전략은 우선 분석을 위한 시작모형인 완전모형을 효율적으로 결정하는 방법을 제시하고, 모형의 간단성의 원칙에 따라서, 추가 설명력이 떨어지는 불필요한 항들을 완전모형에서 제거하여 얻어지는, 항들에 대한 위계 질서가 존중된 바람직한 부분

모형들로 구성된 묶음을 구한다. 묶음에 있는 각각의 후보모형에 대해서 결정계수인 R^2 에 모형에 있는 항들의 개수를 반영한 통계량인 $adj-R^2$, 모형의 예측 능력을 평가하기 위해서 R^2 의 정의에서 잔차제곱합 대신에 press 잔차들의 제곱합으로 대체한 $pred-R^2$ 과 모형에 있는 항들의 개수 기준에서 우열성을 살펴보고, 각각의 모형에 대한 우선순위를 결정한다. 상위에 해당되는 모형들에 대해서 모형 선택 기준 통계량인 C_p 와 press 잔차들의 절대값의 합을 추가적으로 비교하고, 선별된 후보모형의 오차의 정규성과 등분산성을 검토하여 적절한 모형을 복수로 추천한다.

먼저 시작모형을 결정하는 합리적인 전략을 생각해 보자. 우선 혼합물 다항모형의 차수와 공정변수 다항모형의 차수가 증가하는 순서로, 축차적인 모형변동에 의한 제곱합의 분할에 근거하여 계산되는 혼합물 성분과 공정변수 각각의 차수 항에 대한 축차적인 p-값 (sequential p-value), 주어진 차수의 모형에 대한 $adj-R^2$, $pred-R^2$ 값들로 구성된 혼합물 공정변수 결합모형 적합 결과 요약표(combined model mixture process fit summary table)를 만든다. 적합 결과 요약표로부터 잘 알려진 모형 선택 기준 통계량에 근거하여 시작모형을 결정하는 방법을 다음과 같이 제시한다.

- (1) 적합 결과 요약표로부터 혼합물 차수와 공정변수 차수에 대한 축차적인 p-값이 모두 0.05 (혹은 0.1) 이하인 차수들 중에서 차수가 가장 큰 모형의 교적모형을 시작모형으로 결정한다.
- (2) 적합 결과 요약표로부터 혼합물 차수와 공정변수 차수에 대한 $adj-R^2$ 값이 가장 큰 모형의 교적모형을 시작모형으로 결정한다.
- (3) 적합 결과 요약표로부터 혼합물 차수와 공정변수 차수에 대한 $pred-R^2$ 값이 가장 큰 모형의 교적모형을 시작모형으로 결정한다.
- (4) 교적모형에서 혼합물 성분과 공정변수들 간의 교차의 교호작용효과를 무시한 모형인 모형1, 모형2와 모형3의 $adj-R^2$ 과 $pred-R^2$ 값을 비교하여, 가장 바람직한 모형을 시작모형으로 결정한다.

다음으로 주어진 시작모형을 완전모형으로 가정하고, 완전모형보다 우월한 간결한 모형을 후보모형으로 결정하고, 이 모형을 바람직한 후보모형들의 묶음에 포함시키기 위한 방법을 다음과 같이 제시한다.

- ① 모든 가능한 회귀 방법에 의한 변수 선택법에

Mallow의 C_p , 수정된 결정계수인 $adj-R^2$ 와 AIC 기준을 적용하여 각각의 기준에서 최적인 모형을 찾아서 후보 모형에 포함시킨다. 물론 이때 선택된 모형은 항들에 대한 위계 질서가 존중된 모형이다.

- ② 후방소거법과 Stepwise 방법을 적용하여, 최적모형을 구하고, 후보 모형에 포함시킨다. 물론 이때 선택된 모형은 항들에 대한 위계 질서가 존중된 모형이다.

이제 묶음에 있는 후보모형들 중에서 적절한 모형을 선별하기 위해서, 후보모형들에 대한 우선순위를 정해 보자. 묶음에 있는 각각의 후보 모형에 대해서 $adj-R^2$ 과 $pred-R^2$ 값에 대한 상대효율(relative efficiency)인 상대효율1과 상대효율2를 각각 구한다. 후보모형들의 묶음에서 $adj-R^2$ 값이 가장 큰 모형의 상대효율1은 1이 되고, 각각의 후보모형에 대한 상대효율1의 값은 이 모형에 대한 $adj-R^2$ 값과 상대효율1 값이 1인 모형의 $adj-R^2$ 값의 비율에 의해서 정의된다. 유사하게 상대효율2의 값도 $pred-R^2$ 값의 비율에 의해서 정의된다. 후보모형의 우열을 가리는 계량적인 값인 상대효율은 상대효율1과 상대효율2의 최소값으로 정의하거나, 상대효율1과 상대효율2의 평균으로 정의할 수도 있다. 상대효율의 우선순위를 존중하고, C_p , press 잔차들의 절대값의 합, 모형에 있는 항들의 개수 기준에서 우열성을 살펴보고, 적절한 모형에 대한 우선순위를 결정한다.

우선순위의 상위에 있는 2, 3개의 모형에 대해서, 표준화 잔차의 정규 확률그림이 직선모양을 따르는지의 여부와 표준화된 잔차의 산점도가 수평 밴드에서 랜덤하게 분포하고 있는지를 확인하여, 오차의 정규성과 등분산성에 문제가 없는 모형을 적절한 모형으로 추천한다.

4. 사례분석

사례분석 1. Nas 등(1998)에 소개된 3개의 혼합물 성분과 2개의 공정변수를 갖는 혼합물-공정변수 실험 자료를 분석한다. 여기서 실험실 제빵 실험은 세 가지 밀가루 상표인 Tjalve, Folke와 Hard Red Spring를 혼합한 10개의 서로 다른 혼합비를 갖는 밀가루 혼합물을 3수준의 혼합시간(5, 15, 25분)과 교정시간(35, 47.5, 50분)으로 가공하여 만들어진 빵 덩어리의 부피(loaf volume)를 측정하는 것이다.

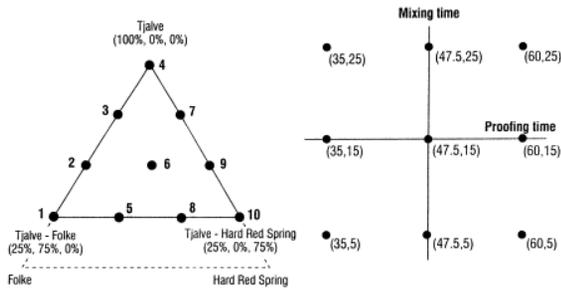


Fig. 1. Illustration of product design for the flour blends and process variables

교적설계를 위한 혼합물 성분과 공정변수들에 대한 실험점들이 <그림 1>에 도시되어 있고, 각각의 실험점에서의 반응변수들의 값은 Nas 등(1998)에 주어진다. 이 혼합물 성분-공정변수 실험 자료에 대한 혼합물 공정변수 결합모형 적합 결과 요약 표가 <표 1>에 주어진다. 여기서 Process Order가 2FI 모형은 일차모형에 이인자교호작용인 z_1z_2 가 추가된 모형이다.

- (1) 적합 결과 요약표로부터 혼합물 차수와 공정변수 차수에 대한 축차적인 p-값이 모두 0.05 (혹은 0.1) 이하인 차수는 혼합물 1차 모형과 공정변수 2차모형의 교적모형인 Linear&Quadratic과 혼합물 3차 모형과 공정변수 2차모형의 교적모형인 Cubic&Quadratic이다. 그 중에서 고차인 모형은 Cubic&Quadratic이다.
- (2) 적합 결과 요약표로부터 혼합물 차수와 공정변수 차수에 대한 $adj-R^2$ 값이 가장 큰 교적모형은

혼합물 3차 & 공정변수 2차인 교적모형으로 축차적인 p-값에 의한 시작모형의 선택결과와 일치한다.

- (3) 적합 결과 요약표로부터 혼합물 차수와 공정변수 차수에 대한 $pred-R^2$ 값이 가장 큰 교적모형은 혼합물 1차 모형과 공정변수 2차모형의 교적모형인 Linear&Quadratic이다.
- (4) 교적 모형의 일부 항들로 구성된 간단한 형태의 모형인 모형1, 모형2, 모형3 중에서 $adj-R^2$ 값이 가장 큰 모형인 모형3과 $pred-R^2$ 값이 가장 큰 모형인 모형2를 시작모형으로 결정한다.

1) 시작모형이 Cubic&Quadratic 경우

시작모형인 Cubic&Quadratic를 완전모형으로 간주하고, 후방 소거 방법에 의해서 적절한 모형을 선택한다. 선택된 모형에 대해서는 모형의 위계질서를 존중하여, 고차의 항이 선별되는 경우에는 관련된 저차의 항들의 p-값이 큰 경우에도 저차의 항들을 모형에 포함시키는데, 이 경우의 우리의 관심사는 저차의 항들을 포함한 경우에도 여전히 고차의 항이 추정 설명력이 있는지를 검토하는 것이다. 항들의 제거를 위한 유의수준 값인 $\alpha-out$ 값을 0.1로 하여 후방 소거 방법을 실행하고, 얻어진 모형에 모형의 위계질서를 존중하여 고차의 교호작용효과와 관련이 있는 저차의 항들을 추가시킨 모형을 cand1이라 하자.

cand1 모형의 각각의 항들에 대한 p-값을 검토할 때에, 고차의 교호작용항인 $x_1x_2z_2^2(x_1-x_2)$ 항의 p-값이

Table 1. Combine model mixture process fit summary table

Mixture Order	Process Order	Sequential P-value		Summary Statistics	
		Mix	Process	$adj-R^2$	$pred-R^2$
L	L	< 0.0001	< 0.0001	0.8250	0.8059
L	2FI	< 0.0001	0.5673	0.8229	0.7919
L	Q	< 0.0001	< 0.0001	0.8922	0.8608
Q	L	0.8755	< 0.0001	0.8145	0.7679
Q	2FI	0.8615	0.6288	0.8102	0.7140
Q	Q	0.3633	< 0.0001	0.8952	0.7892
C	L	0.7551	< 0.0001	0.8023	0.6847
C	2FI	0.8433	0.7653	0.7899	0.4343
C	Q	0.0346	< 0.0001	0.9273	0.7283

0.21로 유의하지 않다. 즉, 모형의 위계질서 존중을 위하여 관련된 저차의 항들을 모형에 추가 시킨 결과, 고차의 교호작용효과인 $x_1x_2z_2^2(x_1-x_2)$ 항의 추가 설명력이 떨어져서, 모형에 더 이상 필요 없음을 의미한다. 이제, 유사하게 cand1 모형에서 추가설명력이 없는 $x_1x_2z_2^2(x_1-x_2)$ 항을 제거한 모형에, 후방 소거 방법을 적용시키고 모형의 위계질서를 존중하여 얻은 모형을 cand2라 하자. cand2 모형의 각각의 항들에 대한 p-값을 검토할 때에, 고차의 교호작용항인 $x_1x_2x_3z_2^2$ 항의 p-값이 커서, 이 항을 cand2 모형에서 제거한 모형에 유사하게 후방 소거 방법을 적용하여 얻은 모형을 cand3 라 한다.

cand3 모형에서 고차의 교호작용효과들 중에서 유의하지 않은 항인 $x_1x_2x_3z_2$ 항을 제거하고, 후방 소거 방법에 의해서 얻어진 모형을 cand4라 하자. cand4 모형에 있는 각각의 항들의 p-값이 0.1보다 작아서 추가 설명력이 있다고 판단된다. 항들의 제거와 삽입을 위한 유의수준 값을 0.1로 하여 Stepwise 방법을 실행하고, 위계질서를 존중하여 고차의 교호작용효과와 관련이 있는 저차의 항들을 추가시키고, 고차의 교호 작용항의 추가설명력을 검토하여 결정된 최종모형은 cand5이다.

2) 시작모형이 Linear&Quadratic인 경우

시작모형인 Linear&Quadratic을 완전모형으로 간주하고, 항들의 제거를 위한 유의수준 값인 $\alpha-out$ 값을 0.1로 하여 후방 소거 방법을 실행하고, 위계질서를 존중하여 고차의 교호작용효과와 관련이 있는 저차의 항들을 추가시킨 모형을 cand6라 하자. cand6 모형에 있는 각각의 항들의 p-값이 모두 작아서 추가 설명력이 있다고 판단된다.

3) 시작모형이 모형3인 경우

시작모형인 모형3을 완전모형으로 간주하고 유의수준 값인 $\alpha-out$ 값을 0.1로 하여 후방 소거 방법을 실행하고, 위계질서를 존중하여 고차 항들의 추가설명력을 축차적으로 검토하여 결정한 최종 후보 모형을 cand7이라 하자. 동일한 유의수준으로, Stepwise 방법을 적용하여 결정한 최종 후보 모형은 cand6와 동일하다. cand7에 있는 각각의 항들의 p-값이 모두 작아서 추가 설명력이 있다고 판단된다.

4) 시작모형이 모형2인 경우

시작모형인 모형2를 완전모형으로 간주하고 유의수준 값인 $\alpha-out$ 값을 0.1로 하여 후방 소거 방법을 실행

하고, 위계질서를 존중하여 고차 항들의 추가설명력을 축차적으로 검토하여 결정한 최종 후보 모형을 cand8이라 하자. Stepwise 방법에 의한 최종모형은 cand6와 동일하다.

이제 앞에서 추천된 4개의 시작모형으로부터 선별된 8개의 후보모형이 후보모형의 묶음이 되고, 묶음에 있는 각각의 모형에 대해서 모형 선택 기준인 $adj-R^2$, $pred-R^2$ 값들과 모형의 간단성을 판단하는 지표가 될 수 있는 각각의 모형에 있는 항들의 개수가 <표 2>에 요약되어 있다. $adj-R^2$ 를 가장 크게 하는 모형은 cand4 모형이고, 예측력을 평가하는 기준인 $pred-R^2$ 를 가장 크게 하는 모형은 cand6 모형이다. 이 값들을 기준으로 $adj-R^2$ 의 비율인 상대효율1과 $pred-R^2$ 의 비율인 상대효율2를 각각 구한다. 상대효율1, 상대효율2, 상대효율1과 상대효율2의 최소값과 평균이 <표 2>에 요약되어 있다. cand1, cand2, cand3 모형과 cand4 모형을 비교하면, cand4 모형이 항들의 개수도 적고, $adj-R^2$, $pred-R^2$ 값들도 크기 때문에, cand4 모형이 이 모형들보다 우월하다. 따라서 후보모형을 선택할 때에는 모형의 위계질서를 존중하여 고차의 항들과 관련된 저차의 항들을 추가한 후에, 고차의 항들의 추가 설명력이 유지되는 것을 검토하여, 축차적으로 고차의 항들이 유의할 때 까지 후방소거법을 적용하여 후보모형을 결정한다. 그런데, cand7 모형은 cand8모형에 3차의 교호작용효과인 $x_1z_1z_2$ 항이 추가된 모형이다. cand7 모형에서 $x_1z_1z_2$ 항의 p-값이 0.0965로 추가 설명력이 있기 때문에 이 항이 cand7 모형에서 필요하고, cand7 모형이 cand8 모형보다 우월하다. 따라서 열등한 모형을 제외한 나머지 모형인 cand4, cand5, cand6, cand7 모형이 우리의 관심 대상이 될 수 있다. 그런데, 상대효율의 최소값으로 우월한 세 개의 모형은 cand4, cand5와 cand7이고, 상대효율의 평균으로 우월한 세 개의 모형은 cand4, cand5와 cand6이기에 우선 4개의 모형인 cand4, cand5, cand6, cand7 모형이 우리의 관심대상이 된다는 사실은 제시된 상대효율의 크기에 의한 모형선별 방법의 타당성을 뒷받침 해 준다. 이제 4개의 후보모형인 cand4, cand5, cand6, cand7의 잔차 그림을 살펴보자. 각각 모형의 표준화 잔차의 정규 확률그림과 표준화 잔차와 예측치의 산점도가 <그림 2>에 주어진다. 표준화 잔차의 정규 확률그림은 모두 직선 형태를 보이고, 표준화 잔차와 예측치의 산점도도 모두 수평 밴드내에서 랜덤한 분포를 보이고 있어서, 모형의 정규성과 등분산성 가정에는 문제가 없어 보인

Table 2. Model selection criteria values and relative efficiency for the candidate models in the pool

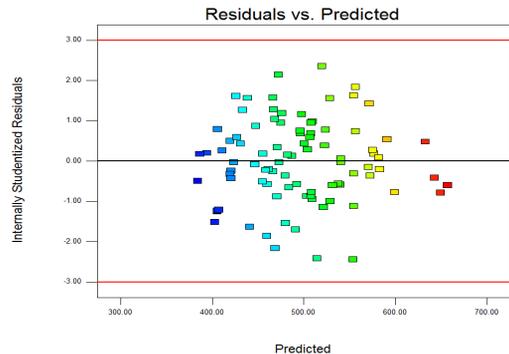
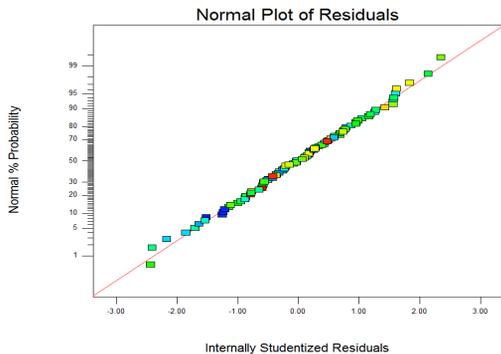
Candidate Model	number of terms	$adj-R^2$	$pred-R^2$	relative eff.1	relative eff.2	min. of relative eff.	ave. of relative eff.
cand1	45	0.9272	0.8301	0.9999	0.9445	0.9445	0.9722
cand2	44	0.9262	0.8322	0.9988	0.9469	0.9469	0.9728
cand3	38	0.9255	0.8503	0.9981	0.9675	0.9675	0.9828
cand4	36	0.9273	0.8638	1.0000	0.9828	0.9828	0.9914
cand5	22	0.9079	0.8682	0.9791	0.9878	0.9791	0.9835
cand6	11	0.8955	0.8789	0.9657	1.0000	0.9657	0.9829
cand7	17	0.9002	0.8737	0.9708	0.9941	0.9708	0.9824
cand8	16	0.8978	0.8764	0.9682	0.9972	0.9682	0.9827

다. 따라서 적절한 모형으로 상대효율의 최소값과 평균이 가장 큰 cand4 모형을 추천할 수 있지만, cand4 모형에는 혼합물의 3차 효과와 공정변수의 2차 효과의 교호작용항인 $x_1x_2(x_1-x_2)z_1^2$, $x_2x_3(x_2-x_3)z_1z_2$ 항 등을 포함하고 있어서, 모형이 복잡하다. cand5 모형도 cand4 모형 보다는 간단하지만 혼합물의 3차 효과와 공정변수의 일차효과와 교호작용항인 $x_1x_2(x_1-x_2)z_1$ 항을 포함하고 있다. 모형의 간단성이 특별히 요구되는 경우에는 혼합물 성분의 선형효과와 공정변수의 선형

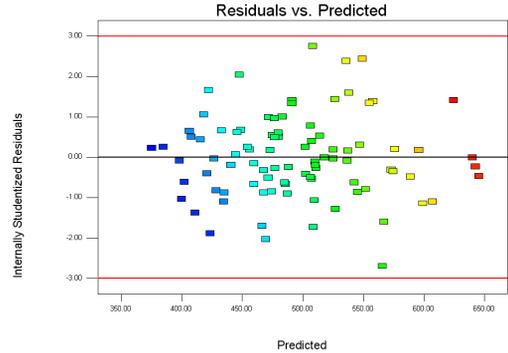
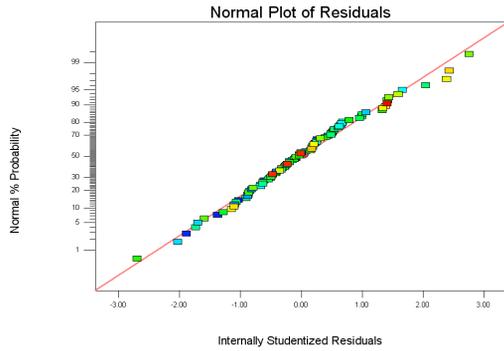
효과 혹은 공정변수의 순수 2차 효과의 교호작용효과만을 포함하는 cand6 모형이나 혼합물 성분의 곡선효과인 2차 효과와 공정변수의 선형효과, 혼합물 성분의 선형효과와 공정변수의 2차 효과의 교호작용 항들을 포함하는 cand7도 추천될 수 있다. cand4, cand5, cand6, cand7 모형에 대한 모형 선택 기준 통계량들이 <표 3>에 요약되어 있다. 참고로 C_p 통계량 관점에서는 cand4 모형이 추천되고, PRESS 잔차들의 절대치의 합을 작게 하는 모형은 cand7 모형이다.

Table 3. Comparison of the performance of recommended four candidate models

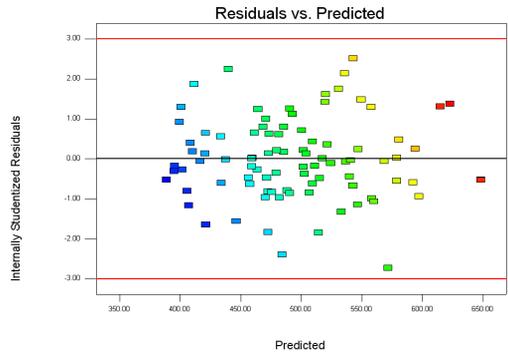
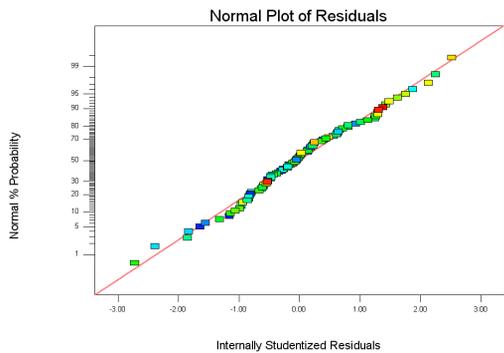
Candidate Model	number of terms	$adj-R^2$	$pred-R^2$	$\sum e_{i,-i} $	C(p)
cand4	36	0.9273	0.8638	1673.542	38.05
cand5	22	0.9079	0.8682	1591.071	42.19
cand6	11	0.8955	0.8789	1544.248	47.55
cand7	17	0.9002	0.8737	1542.276	46.21



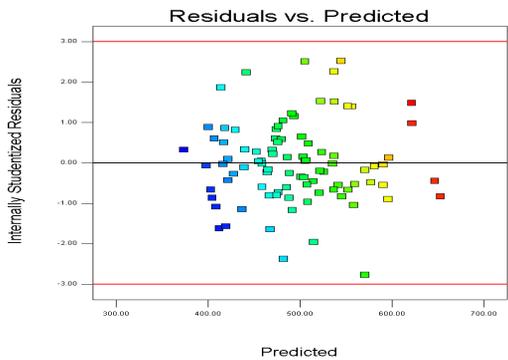
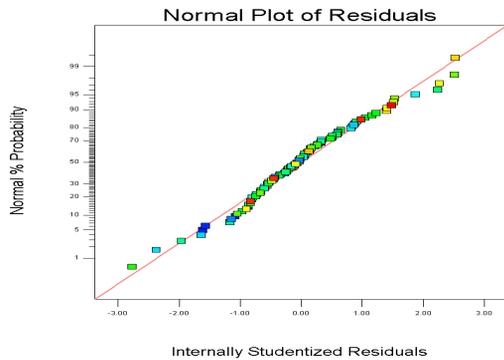
(a) cand4 model



(b)cand5 model



(c) cand6 model



(d) cand7 model

Fig. 2. Residual plots for cand4, cand5, cand6 and cand7 model

사례분석 2. Myers 등(2008)에 소개된 3개의 혼합물 성분과 2개의 공정변수를 갖는 혼합물 성분-공정변수 실험 자료를 분석한다. 항공 우주 산업에 적용되는 접착제는 수지(resin)인 x_1 과 두 종류의 가교제(cross-linkers)인 x_2 와 x_3 를 혼합하여 만들며, 접착제가 부품에 적용되고 전체 조립체는 12시간 동안 제어된 온도와 습도에서 회복된다. 혼합물인 각각의 접착제가 적용된 전체 조립체는 2수준의 온도(40, 100°F)와 습도(15,

85%)에서 12시간 동안 회복된 후에 조립체를 분리시키기 위해서 가해지는 힘인 접착력을 측정한다. 혼합물 2차 모형과 공정변수 일차모형의 교적모형을 가정하여 D-최적기준에 따라서 Design Expert를 활용하여 5개의 반복된 실험점을 포함한 29개의 서로 다른 실험점들이 선택된다.

이 혼합물-공정변수 실험 자료에 대한 혼합물 공정변수 결합모형 적합 결과 요약 표가 <표 4>에 주어진다.

Table 4. combine model mixture process fit summary table

Mixture Order	Process Order	Sequential P-value			Summary Statistics	
		Mix	Process	Lack of fit	$adj-R^2$	$pred-R^2$
L	L	0.0002	< 0.0001	0.6761	0.9362	0.9020
L	2FI	0.0006	0.4021	0.6761	0.9364	0.8814
Q	L	0.2618	< 0.0001	0.7706	0.9445	0.8821
Q	2FI	0.3997	0.5333	0.7529	0.9422	-120.3955

- (1) 적합 결과 요약표로부터 혼합물 차수와 공정변수 차수에 대한 축차적인 p-값이 모두 0.05 (혹은 0.1) 이하인 차수는 혼합물 1차 모형과 공정변수 1차 모형의 교적모형인 Linear&Linear이다.
- (2) 적합 결과 요약표로부터 혼합물 차수와 공정변수 차수에 대한 $adj-R^2$ 값이 가장 큰 교적모형은 혼합물 2차 모형과 공정변수 1차모형의 교적모형인 Quadratic &Linear이다.
- (3) 적합 결과 요약표로부터 혼합물 차수와 공정변수 차수에 대한 $pred-R^2$ 값이 가장 큰 교적모형은 혼합물 1차 모형과 공정변수 1차 모형의 교적모형으로 p-값에 의한 시작모형의 선택결과와 일치한다.
- (4) 교적 모형의 일부 항들로 구성된 간단한 형태의 모형인 모형1, 모형2, 모형3 중에서 $adj-R^2$ 값과 $pred-R^2$ 값이 가장 큰 모형인 모형1을 시작모형으로 결정한다.

1) 시작모형이 Linear&Linear 경우

시작모형인 Linear&Linear 의 각각의 항들의 p-값이 0.05 보다 작아서 유의하다. 따라서 이 모형을 후보 모형으로 추천한다.

2) 시작모형이 Quadratic&Linear 인 경우

시작모형인 Quadratic&Linear를 완전모형으로 간주하고, 모든 가능한 회귀 방법에 Mallow의 C_p , 수정된 결정계수인 $adj-R^2$, AIC 기준을 적용하여 구한 최적 모형은 동일하게 $x_1, x_3, x_1x_2, x_1x_3, x_2x_3, x_1z_1, x_1z_2, x_2z_1, x_2z_2, x_3z_1, x_3z_2, x_1x_3z_1$ 항들을 포함한다. 이 모형에 항들의 위계질서를 존중하기 위해서 x_2 항이 추가된 모형이 cand1이다. cand1 모형에 유의수준 $\alpha-out$ 값을 0.1로 하여 후방 소거 방법을 적용하여 얻은 모형이 cand2-1이고, 유의수준 $\alpha-out$ 값을 0.15로 하여 후

방 소거 방법을 적용하여 얻은 모형이 cand2-2이다. cand2-2 모형은 cand2-1 모형에 $x_1x_3z_1$ 항이 추가된 모형인데, 이 항에 대한 p-값이 0.104이고, 관련된 항인 x_1x_3 의 p-값이 0.06 임을 고려할 때에 적절한 모형으로 선택될 가능성이 높은 후보모형으로 판단된다.

시작모형인 Quadratic&Linear를 완전 모형으로 간주하고 유의수준 $\alpha-out$ 값을 0.1로 하여 후방 소거 방법을 실행하고, 모형의 위계질서를 존중하여 고차의 교호 작용항들과 관련된 저차의 항들이 추가된 모형을 cand3라 하자. cand3 모형이 Myers 등(2008)에서 추천된 모형이다. cand3 모형에서 고차의 교호작용 항에 대한 추가 설명력이 유지되는 지를 검토하여, 축차적으로 고차의 교호작용항들이 유의할 때 까지 후방 소거 방법을 적용하여 얻은 최종 모형이 cand2-1 모형이다. Stepwise 방법을 적용하여 결정한 최종후보모형도 동일하게 cand2-1 모형이다.

3) 시작모형이 모형1 인 경우

시작모형인 모형1을 완전모형으로 간주하고, 모든 가능한 회귀 방법에 Mallow의 C_p , 수정된 결정계수인 $adj-R^2$, AIC 기준을 적용하여 구한 최적모형은 동일하게 모형1에서 x_2 가 제외된 모형이다. 따라서 모형의 위계질서가 존중된 최적모형은 모형1이다. 모형1을 완전모형으로 간주하고 유의수준 값을 0.1로 하여 후방 소거 방법을 실행하고, 위계질서를 존중하여 고차 항들의 추가설명력을 축차적으로 검토하여 결정한 최종 후보 모형은 cand2-1이다. Stepwise 방법에 의한 최종 모형도 cand2-1과 동일하다.

이제 앞에서 추천된 3개의 시작모형으로부터 추천된 6개의 후보모형이 후보모형의 묶음이 되고, 묶음에 있는 각각의 모형에 대해서 모형에 있는 항들의 개수, 모형 선택 기준 통계량 및 상대효율이 <표 5>에 요약되어 있다. cand2-2 모형은 cand3 모형과 모형1과 비교해

Table 5. Model selection criteria values and relative efficiency for the candidate models in the pool

Candidate Model	number of terms	$adj-R^2$	$pred-R^2$	relative eff.1	relative eff.2	min. of relative eff.	ave. of relative eff.
Linear & Linear	9	0.9362	0.9020	0.9786	0.9708	0.9708	0.9747
cand1	13	0.9567	0.9266	1.0000	0.9973	0.9973	0.9987
cand2-1	10	0.9513	0.9259	0.9944	0.9966	0.9944	0.9955
cand2-2	11	0.9548	0.9291	0.9980	1.0000	0.9980	0.9990
cand3	14	0.9489	0.9109	0.9918	0.9804	0.9804	0.9861
모형1	12	0.9522	0.9214	0.9953	0.9917	0.9917	0.9935

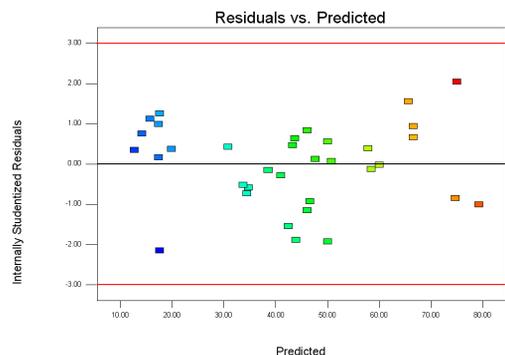
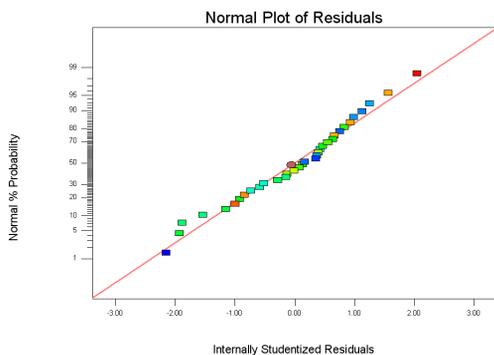
서 항들의 개수가 적어 간단한 모형이고, $adj-R^2$ 과 $pred-R^2$ 값들이 크기 때문에 우월하다. 상대효율의 최소 값과 평균을 동시에 가장 크게 하는 모형은 cand2-2 모형이다. 상대효율의 크기로 선별된 3개의 모형은 cand1, cand2-1과 cand2-2 모형이다.

선별된 모형들에 대한 표준화 잔차의 정규 확률그림과 표준화 잔차와 예측치의 산점도가 <그림 3>에 주어진다. 표준화 잔차의 정규 확률그림은 모두 직선 형태를 보이고, 표준화 잔차와 예측치의 산점도도 모두 수평 밴드 내에서 랜덤한 분포를 보이고 있어서, 선별된 모형의 정규성과 등분산성 가정에는 문제가 없어 보인다.

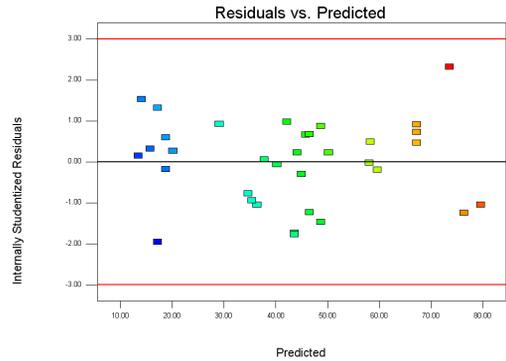
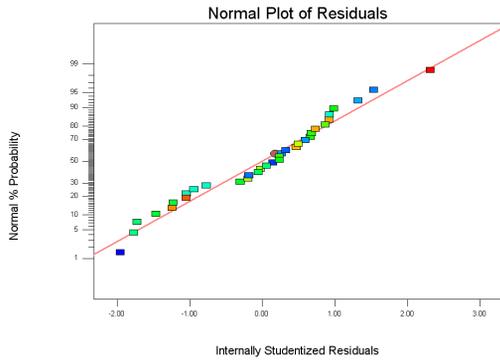
선별된 3개 모형들에 대한 PRESS 잔차들의 절대치의 합과 C_p 를 포함한 모형 선택 기준 통계량의 값들이 <표 6>에 요약되어 있다. C_p 통계량 기준에서도 상대효율의 크기에서 가장 우월한 cand2-2 모형이 추천되고, PRESS 잔차들의 절대치의 합 기준에서는 cand1 모형과 cand2-2 모형이 대등하다. 그런데 cand2-2 모형은 cand1 모형에서 x_1x_2 와 x_2x_3 항이 제거된 모형이고, cand1 모형에서 이 항들에 대한 p-값은 각각 0.22와 0.16으로 추가 설명력이 없다고 판단된다. 따라서 이 사례에 대한 적절한 모형으로 cand2-2 모형을 추천한다.

Table 6. Comparison of the performance of recommended three candidate models

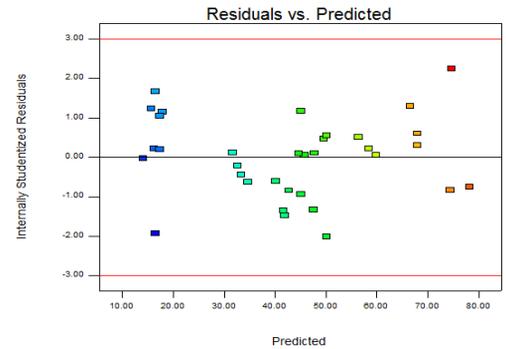
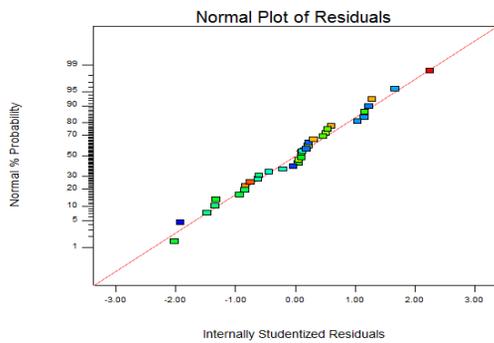
Candidate Model	number of terms	$adj-R^2$	$pred-R^2$	$\sum e_{i,-i} $	C(p)
cand1	13	0.9567	0.9266	139.1437	10.38
cand2-1	10	0.9513	0.9259	142.7619	9.06
cand2-2	11	0.9548	0.9291	139.5284	8.72



(a) cand1 model



(b) cand2-1 model



(c) cand2-2 model

Fig. 3. Residual plots for cand1, cand2-1 and cand2-2 model

5. 요약 및 결론

혼합물 성분비와 공정변수들에 관한 실험 자료가 주어졌을 때에, 주어진 실험 자료를 잘 설명하는 적절한 결합모형을 찾기 위한 체계적인 방법이 제시된다. 우선 모형 선택 기준인 $adj-R^2$ 과 $pred-R^2$ 를 가장 크게 하는 시작모형의 후보들을 교적모형의 범주와 임(2011)에서 제시된 실용적인 모형들 중에서 찾는다. 다음으로 선택된 시작모형을 완전모형으로 간주하여, 모형의 간결성의 원칙에 따라서 완전모형의 부분모형으로 구성된 적절한 결합모형에 대한 후보모형들의 묶음을 구성하는 것이다. 완전모형의 부분모형은 축차적인 변수 선택법이나 모든 가능한 회귀 방법에 의한 변수 선택법에 의해서 선별된다. 묶음에 있는 후보모형들의 $adj-R^2$ 과 $pred-R^2$ 기준에 근거한 각 모형에 대한 유효효율을 정의하고, 유효효율의 크기에 따라서, 후보모형들의 우선순위를 결정한 후에, 후보모형들을 추가적으로 선별한다. 선별된 후보모형들에 대한 모형 선택 기준 통계량의 값인 C_p 와 PRESS 잔차들의 절대치의 합의 추

거적인 비교와 잔차 그림의 검토를 통해서 오차의 정규성과 등분산성을 확인한 후에 최종적으로 적절한 모형을 복수로 추천한다. 잘 알려진 사례분석들을 통해서 적절한 결합모형을 찾기 위한 체계적인 방법이 제시되고, 제시된 사례는 유효효율의 크기에 의한 후보모형의 선별에 대한 타당성을 뒷받침한다.

참고문헌

- [1] Lim, Y.(2010), "Mixture response surface methodology for improving the current operating condition", *Journal of Korean Society for Quality Management*, 38, pp. 413-424.
- [2] Lim, Y.(2011), "Practical designs for mixture component-process experiments", *Journal of Korean Society for Quality Management*, 39, pp. 400-411.
- [3] Cornell, J. A.(2002), *Experiments with Mixtures, 3rd ed.*, New York: Wiley.
- [4] Cornell, J. A. and Gorman, J. W.(1984), "Fractional design plans for process variables in mixture ex-

- periments”, *Journal of Quality Technology*, 16, pp. 20-38.
- [5] Kowalski S., Cornell, J. A. and Vining, G. G.(2000), “A new model and class of designs for mixture experiments with process variables”, *Commun. Statist. -Theory Meth*, 29, pp. 2255-2280.
- [6] Myers, R. H. and Montgomery, D. C.(2009), *Response Surface Methodology, 3rd ed.*, New York: Wiley.
- [7] Nas, T., Fargestad, E. M. and Cornell, J. A.(1998), “A comparison of methods for analyzing data from a three component mixture experiment in the presence of variation created by two process variables”, *Chemometrics and Intelligent Laboratory Systems*, 41, pp. 221-235.

2012년 8월 17일 접수, 2012년 9월 13일 수정, 2012년 9월 14일 채택