Journal of Mathematics and Statistics 3 (3): 93-99, 2007
ISSN 1549-3644

A Note on Model Selection in Mixture Experiments

Kadri Ulaş AKAY
University of Marmara, Departments of Mathematics, Göztepe Kampüsü
Kadıköy, 34722 Istanbul Turkey

Abstract: In mixture experiments, determination of the best model for modeling the mixture system is significant in both understanding and interpreting the system. For obtaining the best model in mixture experiments, different methods have been used. Most commonly used methods are the stepwise type methods. However, the models obtained with these methods are not always the best model depending on the chosen criteria. As the models obtained with these methods can be affected by collinearity, in this paper, an alternative approach is used for the determination of the models taken into account in the modeling of the mixture surface, which is obtained on the experimental region. This approach depends on the examination of all possible subset regression models obtained for the mixture model. To determine the best subset model, the condition numbers of models and the model control graphs are also taken into account. Then, proposed approach has been investigated on flare data set, which is widely known in literature.

Key words: Mixture model, all possible subset selection, variable selection, model reduction, mixture components, collinearity, condition number

INTRODUCTION

In mixture experiments, the measured response is assumed to depend only on the proportions of ingredients present in the mixture and not on the amount of mixture. For example, the response might be the tensile strength of stainless steel which is a mixture of iron, nickel, copper and chromium, or, it might be octane rating of a blend of gasoline. The purpose of mixture experiments is to build an appropriate model relating the response(s) to mixture components.

All of the work on mixture models has been based on response surface concepts. A model is fitted to data by an experimental design. Various mixture models can be used in the analysis of mixture experiments. However, the determination of the best model for modeling the mixture system is important in both understanding and interpreting the mixture system since the fitted models are used to screen the components, predict the response(s), determine the effects of components on the response(s), or optimize the response(s) over the experimental region.

In general, computer-based methods for choosing the best subset regression have been suggested for determining the best model in mixture experiments. Cornell[1], Piepel *et al.*[2-4] and Martin *et al.*[5] used the stepwise regression for choosing the model in mixture experiments. In addition, Draper and St. John[6] used the backward elimination regression procedure. Different from these methods, there are many various methods which examine all possible subset regression models. One of these methods is "RSQUARE procedure" in SAS. This approach was used in mixture experiments, by Khuri[7] and Cornell[1]. Khuri revised the work done by Cornell[1] and he gave some collinearity diagnostic measures for each p-parameter submodel with the highest R^2 value. However, for each submodel with p-parameter, the models including different terms should not be ignored as an alternative to the model with the highest R^2 value. Using the model with the highest R^2 value may not be suitable for the interpretation of the mixture model, as it can be affected by collinearity compared to other models.

The purpose of this study was to get attention on the results obtained by using an alternative approach, in choosing a model in mixture experiments. This approach depends on the examination of all possible subset regression models obtained for the mixture model. By comparing all possible subsets, an investigator can not only determine the best reduced models according to the selected criteria such as R_A^2, but also identify alternatives to the best ones. In addition, extra criteria will also be taken into account

Corresponding Author: Kadri Ulaş Akay, University of Marmara, Departments of Mathematics, Göztepe Kampüsü Kadıköy, 34722 Istanbul Turkey

for the determination of alternative models. In this way, with the help of models including different interaction terms, the mixture system can be interpreted much better and the role of components in the system can be understood much easier.

Mixture experiments: A mixture experiment involves mixing various proportions of two or more components to make different compositions of an end product. In a *q*-components mixture in which x_i represents the proportion of the *i*th components present in mixture,

$$0 \le x_i \le 1 \quad i = 1,2,...,q \quad \sum_{i=1}^{q} x_i = 1 \tag{1}$$

The composition space of the q components takes the form of a regular $(q-1)-$dimensional simplex. Physical, theoretical, or economic considerations often impose additional constraints on individual components

$$0 \le L_i \le x_i \le U_i \le 1 \quad i = 1,2,...,q \tag{2}$$

where L_i and U_i denote lower and upper bounds, respectively. In general, restriction (2) reduce the constraint region given by (1) to an irregular $(q-1)-$dimensional hyperpolyhedron. When the component proportions are restricted by lower and upper bounds, collinearity appears all too frequently[8].

It is assumed that the response or property of interest, denoted by η, is to be expressed in terms of a suitable function *f* of the mixture variables x_i,

$$\eta = f\left(x_1, x_2, ..., x_q\right) \tag{3}$$

A typical model may thus be written

$$y_i = \eta_i + \varepsilon_i \tag{4}$$

where ε_i is assumed that $\varepsilon_i \sim \text{NID}\left(0, \sigma^2\right)$. Mixture model forms most commonly used in fitting data are the canonical polynomials introduced by Scheffé[9] in the form

$$E(Y) = \eta = \sum_{i=1}^{q} \beta_i x_i + \sum_{i<j}^{q} \sum^{q} \beta_{ij} x_i x_j \tag{5}$$

For modeling well-behaved systems, generally the Scheffé polynomials are adequate. For some situations, however, there are better modeling forms than Scheffé polynomials which could be used. For example, as an alternative to Scheffé mixture models, models including inverse term are used in order to model an extreme change in the response behavior of one or more components, which are close to boundary of the simplex region[6]. Following, quadratic model including an inverse term has been proposed by Draper and St. John,

$$E(Y) = \sum_{i=1}^{q} \beta_i x_i + \sum_{i<j}^{q} \sum^{q} \beta_{ij} x_i x_j + \sum_{i=1}^{q} \beta_{-i} x_i^{-1} \tag{6}$$

Scheffé polynomial models fails to satisfy the modeling of additive effect of one component and at the same time accommodate the curvilinear blending effects of the remaining components. To model these effects jointly, Becker has developed a set of mixture models which are homogeneous of degree one[10]. They provide alternatives to the Scheffé polynomials. Becker's three second order models are of the form

$$\begin{aligned}
&\text{H1:}\ \eta = \sum_{i=1}^{q} \beta_i x_i + \sum_{i<j}^{q} \sum^{q} \beta_{ij} \min(x_i, x_j) + ... + \beta_{12...q} \min(x_1, x_2, ..., x_q) \\
&\text{H2:}\ \eta = \sum_{i=1}^{q} \beta_i x_i + \sum_{i<j}^{q} \sum^{q} \beta_{ij} \frac{x_i x_j}{x_i + x_j} + ... + \beta_{12...q} \frac{x_1 x_2 ... x_q}{(x_1 + ... + x_q)^{q-1}} \\
&\text{H3:}\ \eta = \sum_{i=1}^{q} \beta_i x_i + \sum_{i<j}^{q} \sum^{q} \beta_{ij} (x_i x_j)^{1/2} + ... + \beta_{12...q} (x_1 x_2 ... x_q)^{1/q}
\end{aligned} \tag{7}$$

In the H2 model, $x_i x_j \big/ \left(x_i + x_j\right) = 0$ whenever $\left(x_i + x_j\right) = 0$.

As usual, we can represent the Scheffé canonical polynomial models, mixture models with inverse terms and Becker Homogenous models in matrix form by

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \tag{8}$$

where $\mathbf{Y}$ is $n \times 1$ vector of observations on the response variable, $\mathbf{X}$ is $n \times p\left(\ge q\right)$ matrix, where *p* is number of terms in the model, $\boldsymbol{\beta}$ is the $p \times 1$ vector of parameters to be estimated and $\boldsymbol{\varepsilon}$ is $n \times 1$ vector of errors. It was assumed that the errors have the property

$$E(\boldsymbol{\varepsilon}) = 0, \ E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}') = \sigma^2 \mathbf{I}_n \tag{9}$$

where $\mathbf{I}_n$ is identity matrix and σ^2 is the error variance. Hence $E(\mathbf{Y}) = \boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$ where $\boldsymbol{\mu}$ is column vector of all expected responses. The least squares estimator for $\boldsymbol{\beta}$ is $\mathbf{b} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$ and variance-covariance matrix of $\mathbf{b}$ is $\text{var}(\mathbf{b}) = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\sigma^2$. A comprehensive reference on the design and analysis of mixture data is given by Cornell[11,12].

Determination and comparison of mixture models: In mixture experiments, reduction of the model is as important as choosing between different mixture model forms, since it is not a very good approach to add all the terms of the chosen model to itself. In a situation like this, the model may include correlated terms. It may also be hard to make comments on the mixture system as the parameter values may be affected.

In mixture experiments, determination of the best model for modeling the mixture system is significant in

both understanding and interpreting the system. There are various methods for choosing a regression model such as forward selection, backward elimination and stepwise regression when there are many candidate model terms. In addition, Cornell[11] mentioned that the stepwise regression model can be investigated for various models in mixture experiments. The objective is to obtain a model form that not only contains an adequate amount of information about the mixture system under investigation but whose form also makes sense. There are serious problems with stepwise type methods since they do not give the best model (based on the selected criteria, for example R_A^2). This is because they handle variables one at a time. In addition, only one model is obtained with these methods. Therefore, there is a possibility of missing better models.

Mixtures problems are particularly prone to ill-conditioning (or collinearity) because of constraint (1). Collinearity between the terms may lead to inconsistent or confusing conclusions when comparing the result of different stepwise regression procedures. This inconsistency makes variable selection a potentially misleading process when collinearity is present. Ill-conditioning may not be a problem if the goal is prediction within the experimental region. However, ill-conditioning can be a serious problem if interpretation of the coefficients is the objective[8].

Standard variable selection methods do not perform well when the data is highly multi-collinear. For this reason, in order to obtain the best reduced model, all the possible subset regression models should be examined. The sequential model fitting methods proposed by Draper and St. John[6] for mixture experiments can be useful. But, if there are many terms, it can require too much labor. A more preferable method than these methods is to fit all possible regression models, and to evaluate these according to some criterion. In this way a number of best regression models can be selected.

In order to find the best subset regression model "RESEARCH procedure" on GENSTAT was used[13]. While using this procedure, three criteria will be taken into account in determining the best models. First of all, linear mixture terms $\left(x_1, x_2, ..., x_q\right)$ were kept in the model and all possible combinations for the rest of the terms were added to the linear mixture terms. The reason for keeping the linear mixture terms in the model is that the model proves the hierarchy principle. Hierarchy principle is important for the equivalence of the models[7]. As an addition to linear mixture terms, the number of models with the term t $\left(1 \le t \le p-q\right)$ is $\binom{p-q}{t}$. Therefore, for different t values in the mixture system, as a total of $\left(2^{p-q}-1\right)$ subset regression models will be obtained.

For the models obtained, the terms which have $p-value$ smaller than 0.05 according to F- statistics are meaningful. However this situation can affect because of the collinearity and therefore, some important terms for the mixture system can be ignored. In this situation, instead of taking models with meaningful terms into account, the VIF values of the terms should be taken into account. The condition numbers of the models can also be used for comparing the reduced models. A useful measure of collinearity is the condition number, κ, defined by

$$\kappa = \left[\frac{\lambda_{\max}}{\lambda_{\min}}\right]^{\frac{1}{2}}$$

where $\lambda_{\max}$ and $\lambda_{\min}$ denote the largest and the smallest of the eigenvalues of $\mathbf{X}'\mathbf{X}$ (the columns of $\mathbf{X}$ have been scaled to unit length), respectively. Smaller condition numbers indicate more stability (better conditioning) in the least squares estimates than indicated by larger condition numbers. In this study, subset regression models with a condition number less than 40 will be taken into account. In some situations, the condition number of the model with the highest R^2 can be greater than those for other models and this affects the parameter values which may cause misinterpretations about the system. The condition number being less than 40 does not guarantee that VIF values of the terms are less than 100. For this reason, the models with condition numbers less than 40 and the terms with VIF values less than 100 will be taken into account.

Thirdly, in order to examine which of the models are adequate, model control graphs should be obtained. For the models whose model control graphs are adequate, a final decision can be made by looking at R_A^2 and MSE values of the models. The proposed approach will be examined in the following part over the flare data set.

Flare experiment: McLean and Anderson[14] presented an example to illustrate their extreme-vertices design. A flare is manufactured by mixing magnesium $\left(x_1\right)$, sodium nitrate $\left(x_2\right)$, strontium nitrate $\left(x_3\right)$, and

Table 1: Components proportions and illumination response values for flare experiment

Blend No	Component Proportions				Illumination (1000 candles)
	x_1	x_2	x_3	x_4	
1	0.40	0.10	0.47	0.03	75
2	0.40	0.10	0.42	0.08	180
3	0.60	0.10	0.27	0.03	195
4	0.60	0.10	0.22	0.08	300
5	0.40	0.47	0.10	0.03	145
6	0.40	0.42	0.10	0.08	230
7	0.60	0.27	0.10	0.03	220
8	0.60	0.22	0.10	0.08	350
9	0.50	0.1000	0.3450	0.055	220
10	0.50	0.3450	0.1000	0.055	260
11	0.40	0.2725	0.2725	0.055	190
12	0.60	0.1725	0.1725	0.055	310
13	0.50	0.2350	0.2350	0.030	260
14	0.50	0.2100	0.2100	0.080	410
15	0.50	0.2225	0.2225	0.055	425

binder (x_4) under the following constraints,

$0.40 \le x_1 \le 0.60 \qquad 0.10 \le x_3 \le 0.47$

$0.10 \le x_2 \le 0.47 \qquad 0.03 \le x_4 \le 0.08$

The component proportions for design points as well as the measured illumination values are given in Table 1.

Snee[15] used Homogenous mixture models (7) for the modeling of the flare data set. Draper and St. John[6] made a comparison of the mixture models (6) and (7) for the flare data set. On the other hand, Piepel and Cornell[16] gave a summary of the models proposed for the flare data set till now. When these models are examined, it can be seen that they have three terms as an addition to linear mixture terms and they also have the highest R^2 values. However, as the experimental region is restricted, the parameter predictions are affected due to collinearity. For this reason, comparing the models obtained to their condition numbers and using the models with small condition number are more accurate for interpreting the system. On the other hand, Snee[15] and Draper and St. John[6] considered pseudo-components for flare data set. In this paper, subset regression model for actual components will be given by using Scheffé, Homogenous H2 and Models including inverse term.

First of all, let's take the second degree Scheffé models into account. For Scheffé model, as an addition to linear mixture terms with a condition number less than 40, models with one and two terms are obtained. The models with one term have the terms x_2x_3, x_3x_4, and x_2x_4. The condition numbers of these models are 14.6, 18.8 and 18.8, respectively. The models with two terms have the terms x_2x_3, x_3x_4, and x_2x_3, x_2x_4 and their condition numbers being equal to each other, is 21. However, the model recommended for the second degree Scheffé model have the terms x_1x_2, x_1x_3 and x_2x_3 [16]. The condition number of this model is 112.15. In addition, when the stepwise regression and forward selection is used for the Scheffé model taken into account, the model with only x_2x_3 is obtained but with backward elimination method, model with x_1x_2 and x_1x_3 is obtained. The condition numbers for these models obtained are 14.6 and 99.8, respectively. Although the model obtained with backward elimination method has the highest R^2 value $\left(R^2 = 73.59\right)$, this model is the most affected by the collinearity among other models with two terms.

When the homogenous H2 model is examined, five models with one term and condition numbers less than 40 are obtained. These models include the terms except for $x_1x_4/(x_1+x_4)$. Among these models the model with only $x_2x_3/(x_2+x_3)$ term has the terms with VIF values less than 100. As some of the terms in the models with two terms have VIF values greater than 100, they are ignored. On the other hand, for the homogenous H2 model, a model with $x_1x_2/(x_1+x_2)$ and $x_1x_3/(x_1+x_3)$ is obtained by using backward elimination method. In contrast to this, with forward selection and stepwise regression a model with $x_2x_3/(x_2+x_3)$ is obtained. The condition numbers for these models are 45.3 and 12.7, respectively. The model control graphs of the Scheffé and Homogenous H2

subset regression models obtained show that these models are not adequate.

Now let's take mixture models including inverse term into account. If the model including inverse term is from the first degree, then the models with condition number less than 40 have x_2^{-1}, x_3^{-1} and x_4^{-1} with condition numbers 16.8, 16.7 and 39.4, respectively. As the VIF value of x_4^{-1} is 106, this model was ignored. Although the model with x_1^{-1} has the highest R^2 value $\left(R^2 = 78.1\right)$, it is the model the most affected from the collinearity with a condition number of 137. The model control graphs of the models with x_2^{-1} and x_3^{-1} are adequate. The model with two terms, only has x_2^{-1} and x_3^{-1} and the condition number of this model is 22.4. On the other hand, with backward elimination and stepwise regression, the model with x_1^{-1} and x_2^{-1} is obtained. Although this model is the model with the highest R^2, its condition number is 160.77. With forward selection, the model with the terms x_1^{-1}, x_2^{-1} and x_3^{-1} is obtained. This model with highest R^2 value was also used by Piepel and Cornell[16]. The VIF value of x_1^{-1} in the model is 3587.9 and the condition number is 179.2. Therefore, the models that can be recommended for the model including inverse term are only the models with x_2^{-1} and x_3^{-1} from the models with one term. The summary statistics of these models are $R_A^2 = 64.1$, $MSE = 3279$ and $R_A^2 = 60.2$, $MSE = 3638$, respectively.

If the model including inverse term is from the second degree, there is a linear relation between the terms. For this reason, some of the terms in the model are ignored. In this example, the terms x_1^{-1} and x_4^{-1} are kept outside the model as they are the linear composition of other terms. In this situation, six models with two terms and condition numbers less than 40 are obtained. These models include the term couples $\left(x_2x_3, x_2^{-1}\right)$, $\left(x_2x_4, x_2^{-1}\right)$, $\left(x_3x_4, x_2^{-1}\right)$, $\left(x_2x_3, x_3^{-1}\right)$, $\left(x_2x_4, x_3^{-1}\right)$ and $\left(x_3x_4, x_3^{-1}\right)$. The condition numbers of these models are 35.6, 20.9, 20.6, 35.5, 20.6, and 20.9, respectively. The model control graphs of all models except for the last model also show that the models are also adequate. However, as some linear mixture terms have VIF values close to or greater than 100 in models with $\left(x_2x_3, x_2^{-1}\right)$ and $\left(x_2x_3, x_3^{-1}\right)$, they were ignored. VIF values are equal and 99.3 for the x_3 term in the first model, and the x_2 for the second model. Therefore, the models that can be recommended for the second degree models including inverse term include the terms $\left(x_2x_4, x_2^{-1}\right)$, $\left(x_3x_4, x_2^{-1}\right)$ and $\left(x_2x_4, x_3^{-1}\right)$. The summary statistics of these models are $R_A^2 = 60.7$, $MSE = 3587$; $R_A^2 = 60.2$, $MSE = 3634$ and $R_A^2 = 55.8$, $MSE = 4031$, respectively.

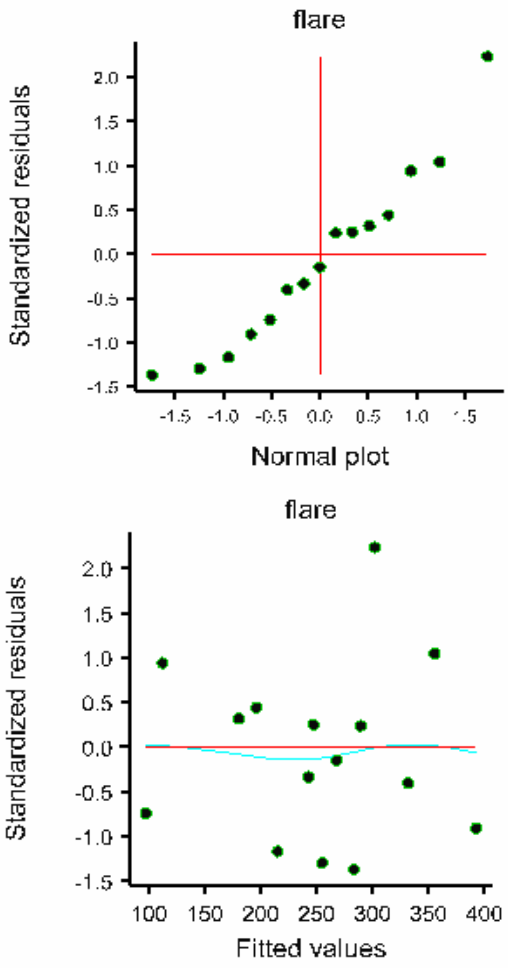

Fig. 1: Model control graphs of model including inverse terms

For the models including inverse term from the second degree, six models can also be obtained with three terms and condition numbers less than 40. The models with only adequate control graph include the terms $\left(x_2x_3, x_2x_4, x_2^{-1}\right)$ and $\left(x_2x_3, x_3x_4, x_2^{-1}\right)$. The condition numbers of these models are 38.5 and 38.3,

respectively. However, as VIF values of the x_3 term in these models are 99.8 and 106.5, respectively, these models were ignored. In addition, although the model with three terms and the highest R^2 value include the terms $\left(x_2x_3, x_2^{-1}, x_3^{-1}\right)$, VIF values of these terms are 223.1, 139.2 and 139.2, respectively. The condition number of this model is also 70.8. In models with three terms, the model with the smallest condition number $\left(\kappa = 24.3\right)$ includes the terms $\left(x_3x_4, x_2^{-1}, x_3^{-1}\right)$. On the other hand, for the model including inverse term from the second degree, only a model with the couple $\left(x_1x_3, x_2^{-1}\right)$ is obtained by using backward elimination, forward selection and stepwise regression. In this model, VIF value for x_1x_3 is 261.6 and condition number is 47.4.

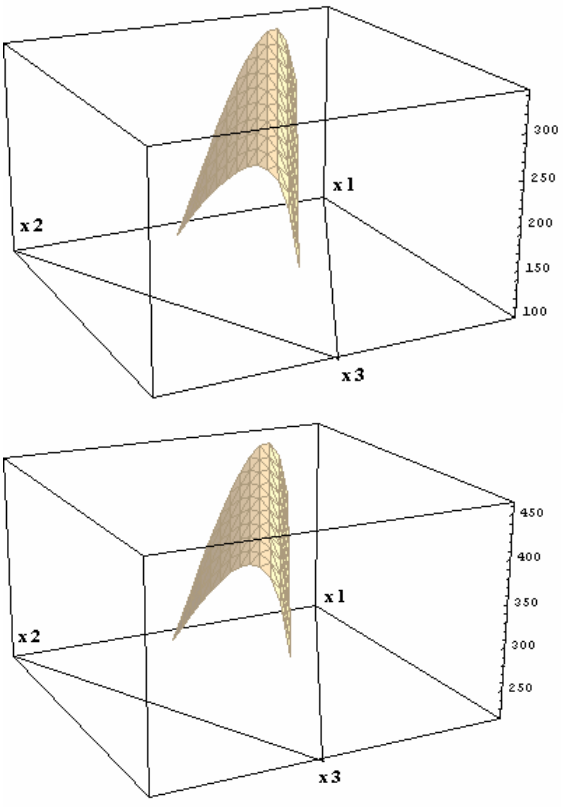

Fig. 2: Mixture surfaces obtained for model including inverse terms

The models including inverse term are better for the interpretation of the mixture system than Scheffé and Homogenous H2 mixture models due to their condition numbers and adequate model control graphs. For this reason, the investigator can choose the best model among the models with one or two terms. For example, the model control graph of the model with terms $\left(x_3x_4, x_2^{-1}\right)$ is given in Fig. 1.

The mixture surface for $x_4 = 0.03$ and $x_4 = 0.08$ on the experimental region for the model is shown respectively in Fig. 2.

CONCLUSION

In this study, comparisons of the results, which were obtained by different subset selection methods for the flare data set, were done. The models obtained by backward elimination, forward selection and stepwise regression methods are one of the models obtained by all possible subset selection. By using all possible subset selection, we can obtain models according to criteria we want. On the other hand, due to effect by collinearity in the models with the highest R^2 value obtained by "RSQUARE procedure", extra criteria were taken into account in choosing the models that can be used in interpreting the mixture system. These criteria are the comparison of the condition numbers and the investigation of the model control graphs of the models with small condition numbers. As a result of comparing the condition numbers of the models, models with more consistent parameter values were obtained. Therefore the condition number of the each model should be taken into account when the all possible subset regression models for the determination of the mixture model are investigated.

As an addition, linear mixture terms of the model were kept in the model to prove the hierarchy principle. Hierarchy principle in mixture experiments is essential for the models, obtained by pseudo-components and actual components, to be equivalent. Expressing the components in terms of pseudo-components will also alleviate the problems due to the correlations among the coefficients. However, this property due to the structure of the models including inverse term and that of homogenous H2 mixture models has to be investigated for both the pseudo-components and the actual components. In addition, in the presence of collinearity, ridge trace can be used as an alternative approach in choosing the model for mixture experiments.

REFERENCES

1. Cornell, J. A., 2000. Fitting a Slack-Variable Model to Mixture Data: Some Questions Raised. J. Quality Tech., 32, 2: 133-147
2. Piepel, G. and T. Redgate, 1997. Mixture experiment techniques for reducing the number of components applied for modeling waste glass sodium release. J. Am. Ceramic Soc., 80: 3038-3044.
3. Piepel, G. and T. Redgate, 1998. A mixture experiment analysis of the hald cement data. The Am. Stat., 52, 1: 23-30.
4. Piepel, G. F., J.M. Szychowski and J.M. Loeppky, 2002. Augmenting Scheffé linear mixture models with squared and/or crossproduct terms. J. Quality Tech., 34, 3: 297-314.
5. Martin, R. J., L.M. Platts, A.B. Seddon and E.C. Stillman, 2003. The design and analysis of a mixture experiment on glass durability. Australian & New Zealand J. Stat., 45: 19-27.
6. Draper, N. R. and R. C. St. John, 1977. A mixtures model with inverse terms. Technometrics, 19: 37-46.
7. Khuri, A.I. 2005. Slack-variable models versus Scheffé's mixture models. J. Applied Stat. 32, 9: 887-908.
8. St. John, R. C., 1984. Experiments with Mixtures, Ill-Conditioning, and ridge regression. J. Quality Tech., 16, 2: 81-96.
9. Scheffé, H., 1958. Experiments with mixtures. J. Royal Stat. Soc., B, 20: 344-360.
10. Becker, N. G., 1968. Models for the response of a mixture. J. Royal Statistical Soc., B, 30: 349-358.
11. Cornell, J.A., 2000. Developing Mixture Models, Are we done?. Journal of Statistical Computation and Simulation, 66: 127-144.
12. Cornell, J.A., 2002. Experiments with mixtures, 3 rd. Ed. Wiley-Interscience.
13. GENSTAT, Release 7.1, 2003. The Guide to Genstat Release 7.1: Part 2 Statistics.
14. McLean, R. A. and V. L. Anderson, 1966. Extreme vertices design of mixture experiments. Technometrics, 8: 447-454.
15. Snee, R. D., 1973. Techniques for the analysis of mixture data. Technometrics, 15: 517-528.
16. Piepel, G.F. and J.A. Cornell, 1994. Mixture experiment approaches: examples, discussion, and recommendations. J. Quality Tech., 26: 177-196.