Technometrics

ISSN: 0040-1706 (Print) 1537-2723 (Online) Journal homepage: https://www.tandfonline.com/loi/utch20

Additive Heredity Model for the Analysis of Mixture-of-Mixtures Experiments

Sumin Shen, Lulu Kang & Xinwei Deng

To cite this article: Sumin Shen, Lulu Kang & Xinwei Deng (2019): Additive Heredity Model for the Analysis of Mixture-of-Mixtures Experiments, Technometrics, DOI: 10.1080/00401706.2019.1630010

To link to this article: https://doi.org/10.1080/00401706.2019.1630010

TECHNOMETRICS
2019, VOL. 00, NO. 0, 1–12
https://doi.org/10.1080/00401706.2019.1630010

Taylor & Francis
Taylor & Francis Group

Additive Heredity Model for the Analysis of Mixture-of-Mixtures Experiments

Sumin Shen[a], Lulu Kang[b], and Xinwei Deng[a]

[a]Department of Statistics, Virginia Tech, Blacksburg, VA; [b]Department of Applied Mathematics, Illinois Institute of Technology, Chicago, IL

ABSTRACT

The mixture-of-mixtures (MoM) experiment is different from the classical mixture experiment in that the mixture component in MoM experiments, known as the major component, is made up of subcomponents, known as the minor components. In this article, we propose an additive heredity model (AHM) for analyzing MoM experiments. The proposed model considers an additive structure to inherently connect the major components with the minor components. To enable a meaningful interpretation for the estimated model, the hierarchical and heredity principles are applied by using the nonnegative garrote technique for model selection. The performance of the AHM was compared to several conventional methods in both unconstrained and constrained MoM experiments. The AHM was then successfully applied in two real-world problems studied previously in the literature. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received August 2018
Accepted May 2019

KEYWORDS
Mixture experiments; Model selection; Nonnegative garrote method

1. Introduction

In the mixture experiment (Cornell 2002), the typical design variables are the proportions of mixture components in a blend, with their summation equal to unity. As a special case of mixture experiments, the mixture-of-mixtures (MoM) experiment considers the case that each mixture component (so-called "major component") is also made up by a mixture of subcomponents (so-called "minor components"). Clearly, the proportions of both major and minor components are varied (Cornell and Ramsey 1998; Piepel 1999) in a constraint experimental region. Such an inherent structure of experiments poses an intriguing challenge to appropriate modeling and analysis of MoM experiments. There can be other variables, such as process variables and the total amount of a blend, to be considered in MoM experiments. In this work, we mainly focus on the analysis of MoM experiments with design variables only including the proportions of major components and their corresponding portions of minor components.

Mixture experiment is one of the classic topics in the design and analysis of experiments area. It has been applied widely in food, medicine, and chemistry industries. The MoM experiments, as a special case of mixture experiments, also frequently appears in applications. See Piepel (1999), Dingstad, Egelandsdal, and Næs (2003), Borges et al. (2007), Didier et al. (2007), and Coetzer and Haines (2013) for interesting MoM case studies. One of the recent MoM examples was studied by Dingstad, Egelandsdal, and Næs (2003). The MoM experiment was conducted to understand the effects of the protein source on the firmness of sausages. There are three major components, dark beef muscle, bright beef muscle, and bright pork muscle, whose proportion are denoted by c_1, c_2, and c_3. Each major component k is composed of three minor components protein, connective tissue, and fat with proportion x_{k1}, x_{k2}, and x_{k3} with respect to c_k, $k = 1, 2, 3$. The firmness of sausages is the response variable of interest. More details can be found in Dingstad, Egelandsdal, and Næs (2003).

Various regression models have been proposed in the literature to analyze mixture experiments, such as the Scheffé model (Scheffé 1958), the Cox model (Cox 1971), the slack-variable model (Snee and Rayner 1982; Khuri 2005; Kang, Salgado, and Brenneman 2016), the Kronecker model (Draper and Pukelsheim 1998; Prescott et al. 2002), the component-slope-linear model (Piepel 2007), and a general blending model by Brown, Donev, and Bissett (2015). The idea of the Scheffé model has been extended to the multiple-Scheffé model (Lambrakis 1968, 1969; Cornell and Ramsey 1998) for analyzing MoM experiments. The multiple-Scheffé model considers a product of the Scheffé models for both major and minor components to model the major–minor interactions. However, such a modeling strategy involves a large number of interaction terms and needs a large number of experimental observations. Another limitation is that the minor components can still be in the model even when its major component is absent in the model. To address limitations of the multiple-Scheffé model, Kang, Joseph, and Brenneman (2011) developed a so-called "major–minor model" to analyze MoM experiments. Their key idea is to use the Scheffé model to capture the relationship between the mean response and the major components, while the coefficients of major components and their interactions are modeled as a function of their respective minor components. The major–minor model can have a smaller model size than the multiple-Scheffé model. Furthermore, when the major component is absent, all of its corresponding minor components are absent from the major–minor model. On the other hand, the complexity of the major–minor model depends on the degrees of the Scheffé models in the levels of major and minor components. When a linear

CONTACT Lulu Kang lkang2@iit.edu Department of Applied Mathematics, Illinois Institute of Technology, Chicago, IL 60616.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/TECH.

Scheffé model is used for both major and minor components, the major–minor model does not have interactions between the major components and between the minor components. When a quadratic Scheffé model is assumed for both major and minor components, the resultant terms can be too complex to interpret because of some high-order interaction terms.

To address these limitations, we propose an additive heredity model (AHM) to better capture the major–minor structure of MoM experiments with meaningful interpretation. Similar to the major–minor model, the AHM also assumes two Scheffé models for the major and minor components, respectively. But different from the major–minor model, the two models are added together to form the final heredity model. The detailed development is shown in Section 2. The additive model is simple and paves an easy way to study the contribution of each design variable in the model. Moreover, by imposing the coefficients of the minor components to be functions of their respective major components, the minor components exist only when the corresponding major component is presented in the model. The nonnegative garrote technique (Breiman 1995; Yuan and Lin 2007; Xiong 2010) is used in the model estimation to enable the hierarchical and heredity principles for major and minor components. The proposed AHM has several advantages. First, it provides a meaningful model interpretation of the major–minor structure such that the coefficients of minor components are dependent upon the corresponding major components. In MoM experiments, the minor components will be presented in the model only if their corresponding major component is included in the model (Yuan, Joseph, and Zou 2009; Kang, Joseph, and Brenneman 2011). Second, it can explicitly quantify the contributions of the individual major and minor components through the additive form. Third, the AHM can control the model complexity via variable selection, which is achieved by the nonnegative garrote method. Furthermore, the additive structure is flexible to include terms of interest based on practitioners' objectives.

The rest of article is organized as follows. In Section 2, we detail the proposed AHM. In Section 3, we present the estimation procedures by the heredity constrained nonnegative garrote method. The simulation and real case studies are conducted in Sections 4 and 5. We conclude this work with some discussion in Section 6.

2. Additive Heredity Model

In an MoM experiment, assume that there are q major components, and let c_k be the proportion of the kth major component such that

$$\sum_{k=1}^{q} c_k = 1, \quad 0 \le c_k \le 1, \quad k = 1, \ldots, q. \tag{1}$$

Moreover, each major component is composed of m_k minor components, whose proportions with respect to c_k are x_{kl},

$$\sum_{l=1}^{m_k} x_{kl} = 1, \quad 0 \le x_{kl} \le 1, \quad l = 1, \ldots, m_k. \tag{2}$$

To flexibly quantify the effects of major and minor components on the response y, we consider an additive modeling strategy to incorporate the major and minor structure relations in the model. We propose an AHM as

$$y = \sum_{k=1}^{q} \gamma_k c_k + \sum_{k<j} \gamma_{kj} c_k c_j + \sum_{k=1}^{q} \sum_{l=1}^{m_k} \delta_l^{(k)} x_{kl} + \sum_{k=1}^{q} \sum_{l<l'} \delta_{l,l'}^{(k)} x_{kl} x_{kl'} + \epsilon, \text{ with } \epsilon \sim N(0, \sigma^2), \tag{3}$$

where γ_k is the coefficient for the major component proportion c_k, γ_{kj} is the coefficient for the interactions between c_k and c_j. The $\delta_l^{(k)}$ is denoted as the coefficient for the minor component proportion x_{kl}, and $\delta_{l,l'}^{(k)}$ is denoted as the coefficient for the interaction between x_{kl} and $x_{kl'}$. To ensure the contribution of the minor component to the response depends upon the corresponding major component, we consider $\delta_l^{(k)}$ to be a function of the major component c_k. A monotonic and bounded mapping from the major component c_k to $\mathbb{R}$ is a proper choice for $\delta_l^{(k)}$ because the larger the c_k is, intuitively, the more influential the minor components of this major component should be to the response of the whole mixture. Under this consideration, we consider to use a power function as

$$\delta_l^{(k)} = \zeta_l^{(k)} c_k^h,$$

where $\zeta_l^{(k)}$ is the coefficient and h is the power parameter. Similarly we consider $\delta_{l,l'}^{(k)} = \zeta_{l,l'}^{(k)} c_k^{2h}$ with $\zeta_{l,l'}^{(k)}$ being the coefficient. Clearly, the power function c_k^h is bounded on the domain of c_k. The hyperparameter h is the power index of c_k and set in the range of $(0, 2)$. For the c_k in $(0, 1)$, c_k^h is decreasing with respective to h. Thus, the larger the h is, the less role the major component would play in the minor components' effects, including $\delta_l^{(k)} x_{kl}$ and $\delta_{l,l'}^{(k)} x_{kl} x_{kl'}$. The hyperparameter h is estimated from the data via cross-validation as described in Algorithm 1. We set the upper bound of h to be 2, which is shown to be sufficient in our study. Readers can choose any values that are larger than zero based on their understanding of the major components' influence.

Algorithm 1

1: Input: data
2: **for** a sequence of h **do**
3: Obtain initial estimators $\gamma_k^{\text{init}}, \gamma_{kj}^{\text{init}}, (\delta_l^{(k)})^{\text{init}}$, and $(\delta_{ll'}^{(k)})^{\text{init}}$ from ridge estimators of (3).
4: **for** a sequence of M **do**
5: Solve the constrained optimization problem (6) and save the scaling factors $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$.
6: Compute the GCV value at each M.
7: **end for**
8: $M_{\text{sel}} = \arg\min_M \text{GCV}(M)$
9: Compute the mean squared cross-validation (MSCV), defined later, at M_{sel}.
10: **end for**
11: $h_{\text{sel}} = \arg\min_h \text{MSCV}(h)$
12: Output: the estimated model and MSCV at h_{sel}

The AHM specifies the effects of major and minor components in an additive form. Moreover, the AHM has a flexible

major–minor structure relationship by assuming the coefficient of minor components varies as a function of a major component. This varying-coefficient property is useful to flexibly accommodate certain dependence relationships between the major and minor components. The AHM thus can provide a meaningful model interpretation of the major–minor structure. Lastly, various Scheffé models of appropriate order are applicable in the AHM framework. For example, a quadratic Scheffé model for the major component and a linear Scheffé model for the minor component. We choose the quadratic Scheffé model for both major and minor components in this study because we are interested in both the main effects and the two-factor interactions of the major components, the main effects of the minor components, and the two factor interactions of the minor components from the same major component.

2.1. Connection to the Major–Minor Model

In this section, we make a connection between the proposed AHM and the major–minor model in Kang, Joseph, and Brenneman (2011). The major–minor model considers the coefficients of major components as the Scheffé model on the corresponding minor components to incorporate the major–minor structure relations. Take the example of having two major components c_1 and c_2, each with two minor components x_{k1}, x_{k2}, $k = 1, 2$. Assuming the quadratic Scheffé model for both the major and minor components, the major–minor model is expressed as

$$\begin{aligned} y = {} & (\gamma_1 x_{11} + \gamma_2 x_{12} + \gamma_3 x_{11} x_{12}) c_1 + (\gamma_4 x_{21} + \gamma_5 x_{22} + \gamma_6 x_{21} x_{22}) c_2 \\ & + (\gamma_7 x_{11} x_{21} + \gamma_8 x_{11} x_{22} + \gamma_9 x_{12} x_{21} + \gamma_{10} x_{12} x_{22} \\ & + \gamma_{11} x_{11} x_{12} x_{21} + \gamma_{12} x_{11} x_{12} x_{22} + \gamma_{13} x_{11} x_{21} x_{22} \\ & + \gamma_{14} x_{12} x_{21} x_{22} + \gamma_{15} x_{11} x_{12} x_{21} x_{22}) c_1 c_2 + \epsilon. \end{aligned} \tag{4}$$

Therefore, when a major component proportion $c_k = 0$, the corresponding minor components no longer exist in the model. As a major component's proportions increase, the contribution from the corresponding minor components also increases.

Nevertheless, the use of the Scheffé model on minor components as the coefficients for the corresponding major component is not the only way to represent the major–minor structure relations. For instance, we can use a so-called "minor-major" model at the minor level to describe the relationship between the response, the major, and the minor components. That is, the coefficients of minor components are a function of the respective major components, indicating that the contribution from minor components is dependent upon their major components. Assuming the quadratic Scheffé model for minor components, the minor-major model can be expressed as

$$\begin{aligned} y = {} & (\phi_1 + \phi_2 c_1) x_{11} + (\phi_3 + \phi_4 c_1) x_{12} + (\phi_5 + \phi_6 c_1 + \phi_7 c_1^2) x_{11} x_{12} \\ & + (\phi_8 + \phi_9 c_2) x_{21} + (\phi_{10} + \phi_{11} c_2) x_{22} + (\phi_{12} + \phi_{13} c_2 + \phi_{14} c_2^2) x_{21} x_{22} \\ & + (\phi_{15} c_1 + \phi_{16} c_2 + \phi_{17} c_1 c_2) x_{11} x_{21} + (\phi_{18} c_1 + \phi_{19} c_2 + \phi_{20} c_1 c_2) x_{11} x_{22} \\ & + (\phi_{21} c_1 + \phi_{22} c_2 + \phi_{23} c_1 c_2) x_{12} x_{21} + (\phi_{24} c_1 + \phi_{25} c_2 + \phi_{26} c_1 c_2) x_{12} x_{22} + \epsilon, \end{aligned} \tag{5}$$

where the coefficients of minor components are assumed to be dependent only upon their corresponding major component.

Note that the major–minor model in (4) contains terms that are complicated and difficult to interpret, such as $c_1 c_2 x_{11} x_{21} x_{22}$. In contrast, such terms do not appear in the AHM (3) when a quadratic model is used for both minor and major components. Through heredity principle incorporated in the nonnegative garrote method, the AHM can control the model complexity. We also like to remark that the major–minor model can be rewritten in the form of additive models. For example, if we assume the linear Scheffé model for both major and minor components, the major–minor model is equivalent to the AHM assuming $\delta_l^{(k)} = \zeta_l^{(k)} c_k$, $\delta_{l,l'}^{(k)} = 0$, and $\gamma_{kj} = 0$ in model (3). In the supplemental materials, we show that any major–minor model can be expressed in some forms of additive models.

3. Model Estimation

To estimate the parameters in the proposed AHM, we employ the nonnegative garrote method (Breiman 1995; Yuan and Lin 2007; Xiong 2010) to pursue a parsimonious and structured model. The nonnegative garrote estimate of a parameter is expressed as $\theta^{nng} = \theta^{(0)} \alpha_s$, where $\theta^{(0)}$ is the initial estimate and $\alpha_s \geq 0$ is a nonnegative scaling factor. The key idea of the nonnegative garrote method is to scale the initial estimates via scaling factors. One feature of the nonnegative garrote method is the flexibility to adapt the hierarchical and heredity principles in the form of linear constraints (Yuan and Lin 2007). The hierarchical principle between the major and corresponding minor components is to require the minor components being present only if the corresponding major component is present in the model. The heredity principle requires that the interaction terms can appear in the model only if one of its main effects appears in the model. By imposing such principles, it can make the proposed model more meaningful and interpretable. Besides the nonnegative garrote method, the least angle regression selection (LARS) method can also be modified to incorporate the heredity principle, as shown in Yuan and Lin (2007). But Yuan, Joseph, and Zou (2009) pointed out that the nonnegative garrote method is more efficient in computation and much more flexible and easier to adopt any kind of constraints between the effects. For the proposed additive model, the hierarchical and heredity principle is more complex than the regular regression model and thus we choose the nonnegative garrote method.

The nonnegative garrote method with the weak heredity principle can be expressed in (6). The response y is a simplified notation y_i without the observation index i. Let n be the total number of observations in the MoM experiment. The γ_k^{init}, $\gamma_{kj}^{\text{init}}$, $(\zeta_l^{(k)})^{\text{init}}$, $(\zeta_{l,l'}^{(k)})^{\text{init}}$ are the initial estimates of the parameters γ_k, γ_{kj}, $\zeta_l^{(k)}$, and $\zeta_{l,l'}^{(k)}$, respectively. We denote α_k to be the scaling factor for the major component c_k, α_{kj} to be the scaling factor for the interaction between major components c_k and c_j, $\beta_l^{(k)}$ to be the scaling factor for the minor component l within the major component k, and $\beta_{l,l'}^{(k)}$ the scaling factor for the interaction between minor components l and l'. The constrained optimization of estimating parameters is given as follows

$$\begin{aligned} \min_{\boldsymbol{\alpha},\boldsymbol{\beta}} \sum_{i=1}^{n} \Bigg\{ y - \Bigg\{ & \sum_{k=1}^{q} \gamma_k c_k + \sum_{k<j} \gamma_{kj} c_k c_j + \sum_{k=1}^{q} \sum_{l=1}^{m_k} \delta_l^{(k)} x_{kl} \\ & + \sum_{k=1}^{q} \sum_{l<l'} \delta_{ll'}^{(k)} x_{kl} x_{kl'} \Bigg\} \Bigg\}^2, \end{aligned} \tag{6}$$

$$
\begin{aligned}
&\text{s.t.}\gamma_k = \gamma_k^{\text{init}}\alpha_k, \gamma_{kj} = \gamma_{kj}^{\text{init}}\alpha_{kj}; \delta_l^{(k)} = (\zeta_l^{(k)})^{\text{init}} c_k^h \beta_l^{(k)},\\
&\delta_{ll'}^{(k)} = (\zeta_{ll'}^{(k)})^{\text{init}} c_k^{2h} \beta_{l,l'}^{(k)};\\
&\alpha_k \geq 0, \alpha_{kj} \geq 0, \beta_l^{(k)} \geq 0, \beta_{l,l'}^{(k)} \geq 0;\\
&\sum_{k=1}^{q}\left\{\sum_{j=1,k<j}^{q}(\alpha_k+\alpha_{kj}) + \sum_{l<l'}^{m_k}(\beta_l^{(k)}+\beta_{l,l'}^{(k)})\right\} \leq M;\\
&\beta_l^{(k)} \leq \alpha_k; \beta_{l'}^{(k)} \leq \alpha_k;\\
&\alpha_{kj} \leq \alpha_k + \alpha_j, \text{ for } k \neq j; \qquad (6a)\\
&\beta_{l,l'}^{(k)} \leq \beta_l^{(k)} + \beta_{l'}^{(k)}, \text{ for } l \neq l', \qquad (6b)
\end{aligned}
$$

where the constraints (6a) and (6b) are specially applied when the weak heredity principle is assumed. The weak and strong heredity principles select the interaction terms based on their parent terms. The weak heredity only requires one of its parent terms to be significant. For example, in (6a), α_{kj} will be forced to be zero unless at least one of α_k and α_j is strictly positive. This reflects the weak heredity principle, as it indicates $c_k c_j$ would be significant if one of its parent terms c_k and c_j is significant. The strong heredity selects a two-factor interaction only if both its parent terms are significant. When the strong heredity principle is assumed, these two constraints (6a) and (6b) can be replaced by

$$
\begin{aligned}
&\alpha_{kj} \leq \alpha_k;\ \alpha_{kj} \leq \alpha_j, \text{ for } k \neq j;\\
&\beta_{l,l'}^{(k)} \leq \beta_l^{(k)};\ \beta_{l,l'}^{(k)} \leq \beta_{l'}^{(k)}, \text{ for } l \neq l'.
\end{aligned}
$$

For example, α_{kj} would be strictly positive if both α_k and α_j are larger than 0, and accordingly, $c_k c_j$ is significant if both c_k and c_j are significant. If no heredity principle is assumed, the two constraints (6a) and (6b) can be removed. Here M is a tuning parameter to control the general sparsity in the model. Note that the objective in the above optimization can be expressed as $(\boldsymbol{y} - \boldsymbol{X\eta})^T(\boldsymbol{y} - \boldsymbol{X\eta})$, where $\boldsymbol{y} = (y_1, \ldots, y_n)$ is the response vector, $\boldsymbol{\eta} = (\alpha_1, \ldots, \beta_{q-1,q})'$ is the parameter vector containing all scaling factors α and β, and $\boldsymbol{X}$ is the corresponding regression matrix. We adopt the generalized cross-validation (GCV) for finding an optimal value of M, which is given by

$$
\text{GCV} = \frac{(\hat{\boldsymbol{y}} - \boldsymbol{y})^T(\hat{\boldsymbol{y}} - \boldsymbol{y})}{n(1 - tr(\boldsymbol{H})/n)^2},
$$

where $\boldsymbol{H}$ is the hat matrix, and $\text{tr}(\boldsymbol{H}) = \text{tr}(\boldsymbol{X}\text{diag}(\hat{\boldsymbol{\eta}})(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T) = \text{tr}(\text{diag}(\hat{\boldsymbol{\eta}}))$. Good justification of using GCV for tuning parameter selection in the nonnegative garrote problem can be found in Xiong (2010).

The computational algorithm for model estimation (6) is summarized in Algorithm 1. Because of the nonnegative garrote method with proper constraints, the model size of the AHM can be much smaller compared to the multiple-Scheffé model. It is worth pointing out that the performance of the nonnegative garrote estimate relies on the choice of the initial estimate. Yuan and Lin (2007) argued that the nonnegative garrote method can be used with initial estimators from the ridge regression, the LASSO, and the elastic net. Here we use the ridge regression estimate with the regularization parameter λ determined by the leave-one-out cross-validation for the initial estimators of the nonnegative garrote method. Other than leave-one-out cross-validation, the user can also consider using other criteria such as AICc (Draguljić et al. 2014), especially when the data are collected from designed experiments with very limited runs.

4. Simulation

In this section, we evaluate the performance of the proposed AHM in both unconstrained and constrained MoM experiments. We consider two different types of MoM experiments: (a) each major component contains the same number of minor components and (b) each major component contains a different number of minor components, and one major component has a single minor component. Without loss of generality, for both (a) and (b) we assume there are only three major components, that is, c_1, c_2, and c_3. In type (a), each major component c_k has two minor components x_{k1}, x_{k2}. In type (b), the numbers of minor components corresponding to c_1, c_2, and c_3 are three, two, and one, respectively. The simulation results for (a) are shown in this section. The results for (b) are in supplement due to space limitations.

In case (a), there are five underlying models to be considered for generating the data

$$
\begin{aligned}
\text{I}:\ & y = 10c_1 + 30c_2 + 20c_3 + 18c_1c_2 + \epsilon,\\
\text{II}:\ & y = 15c_1x_{11} + 12.5c_1x_{12} + 22.5c_2x_{21} + 20c_2x_{22} + 15c_3x_{31}\\
& \quad + 17.5c_3x_{32} + \epsilon,\\
\text{III}:\ & y = 10c_1 + 30c_2 + 20c_3 + 15c_1^h x_{11} + 27.5c_2^h x_{21} + \epsilon,\\
& \quad \text{where } h = 0.5,\\
\text{IV}:\ & y = 10c_1 + 30c_2 + 20c_3 + 25c_2x_{11} + 22.5c_3x_{21} + \epsilon,\\
\text{V}:\ & y = 10c_1 + 30c_2 + 20c_3 + 7c_2c_3 + 13.75c_1^2x_{11}x_{12} + \epsilon,
\end{aligned}
$$

where the noise $\epsilon \sim N(0, \sigma^2)$ is independent of the component proportions. The noise variance is chosen such that the signal-to-noise (SN) ratio (Wu and Hamada 2009) is three.

One benchmark method for comparison is the multiple-Scheffé model. Assuming the linear Scheffé model for both major and minor components, the corresponding multiple-Scheffé model is

$$
y = (\alpha_1c_1 + \alpha_2c_2 + \alpha_3c_3) \times \prod_{k=1}^{3}(\beta_{k1}x_{k1} + \beta_{k2}x_{k2}) + \epsilon,
$$

which contains 24 regression coefficients. Other methods used in comparison include the major-only linear Scheffé model (7) and the major-only quadratic Scheffé model (8),

$$
\begin{aligned}
y &= \gamma_1c_1 + \gamma_2c_2 + \gamma_3c_3 + \epsilon, \qquad (7)\\
y &= \gamma_1c_1 + \gamma_2c_2 + \gamma_3c_3 + \gamma_4c_1c_2 + \gamma_5c_1c_3 + \gamma_6c_2c_3 + \epsilon. \qquad (8)
\end{aligned}
$$

Both (7) and (8) naively ignores the information on the minor components. We also make comparison with the major–minor models (Kang, Joseph, and Brenneman 2011) assuming the linear Scheffé model for minor components, linear or quadratic Scheffé model for major components, defined in (9) and (10),

respectively.

$$y = (\gamma_1 + \gamma_2 x_{11})c_1 + (\gamma_3 + \gamma_4 x_{21})c_2 + (\gamma_5 + \gamma_6 x_{31})c_3 + \epsilon, \tag{9}$$

$$\begin{aligned} y = {} & \gamma_1 c_1 + \gamma_2 c_2 + \gamma_3 c_3 + \gamma_4 x_{11} c_1 + \gamma_5 x_{21} c_2 + \gamma_6 x_{31} c_3 \\ & + \gamma_7 c_1 c_2 + \gamma_8 c_1 c_3 + \gamma_9 c_2 c_3 + \gamma_{10} x_{11} c_1 c_2 + \gamma_{11} x_{11} c_1 c_3 \\ & + \gamma_{12} x_{21} c_1 c_2 + \gamma_{13} x_{21} c_2 c_3 + \gamma_{14} x_{31} c_1 c_3 + \gamma_{15} x_{31} c_2 c_3 \\ & + \gamma_{16} x_{11} x_{21} c_1 c_2 + \gamma_{17} x_{11} x_{31} c_1 c_3 + \gamma_{18} x_{21} x_{31} c_2 c_3 + \epsilon. \end{aligned} \tag{10}$$

A summary list of compared models is here:

a. the multiple-Scheffé model (MultipleScheffe),
b. the major-only linear Scheffé model (MajorLinear),
c. the major-only quadratic Scheffé model (MajorQuad),
d. the major–minor model assuming the linear Scheffé model for the major components (1st_MM),
e. the major–minor model assuming the quadratic Scheffé model for the major components (2nd_MM),
f. the additive heredity model with weak heredity constraints (AHM).

To evaluate the model performances, we use the metrics including the R^2, the small-sample-size corrected version of Akaike information criterion (AICc), mean squared error (MSE), mean squared cross-validation (MSCV), normalized MSCV (MSCVnorm), and model size. The R^2, AICc, and MSE measure the fitting performance of models. Note that $R^2 = 1 - \sum_{i=1}^{n}(y_i - \hat{y}_i)^2 / \sum_{i=1}^{n} y_i^2$ is adapted for the Scheffé models that do not contain intercept. The AICc (Hurvich and Tsai 1989; Burnham and Anderson 2002; Draguljić et al. 2014) is calculated via AICc = $n\log(\mathrm{RSS}/n) + 2\tilde{p}n/(n - \tilde{p} - 1)$, where RSS is the residual sum of squares, $\tilde{p}$ is the number of nonzero parameters. The MSCV is calculated via MSCV $= \sum_{i=1}^{n}(y_i - \hat{y}_{-i})^2/n$, where $\hat{y}_{-i}$ is the prediction at the ith input by the model fitted without the ith data point. The smaller MSCV value indicates better prediction performance. The MSCVnorm is calculated by MSCVnorm = $\mathrm{MSCV}(n-1)/(\sum_{i=1}^{n}(y_i - \bar{y})^2)$. The AHM has a varied model size because of the nonnegative garrote technique employed.

4.1. Unconstrained MoM Experiments

The unconstrained mixture experiment is a typical situation where each component proportion can take any value in [0, 1]. For the major components, we consider two different designs: the I-optimal design (Laake 1975; Goos, Jones, and Syafitri 2016) and the maximin distance design (Johnson, Moore, and Ylvisaker 1990). For the minor components, because each major component has two minor components, we choose three design point levels: the two endpoints and the middle point in the domain. For example, the design points for the minor components, (x_{11}, x_{12}), are $(1, 0)$, $(\frac{1}{2}, \frac{1}{2})$, and $(0, 1)$. We apply the idea of crossed design to combine the designs for the major and minor components (Cornell and Ramsey 1998; Dingstad, Egelandsdal, and Næs 2003; Kang, Joseph, and Brenneman 2011). That is, corresponding to every treatment combination of the major components, all possible settings of the minor components are included in the design.

We first use the I-optimal design for the major components. The I-optimal criterion is to minimize the average prediction variance over the experimental region. For the simulation setting described above, an I-optimal design for a quadratic Sheffé model of three major components is the simplex-centroid design (Scheffé 1963; Lambrakis 1968, 1969; Cornell 2002) containing three vertices, three middle points, and the overall centroid of the triangular constrained region by $c_1 + c_2 + c_3 = 1$, as shown in Figure 1(A). The plots are generated by the mixexp package (Lawson and Willden 2016) in R software. Apparently, the design for two minor components, containing $(1, 0)$, $(0, 1)$, and $(1/2, 1/2)$, is also a simplex-centroid for any mixture experiment with two components. Thus, the overall design is a crossed design of simplex-centroid for major components and simplex-centroid design for the minor components.

Table 1 shows the simulation results in terms of R^2, AICc, MSE, MSCV, MSCVnorm, and model size of the different models under comparison. The proposed AHM generally outperforms the other models in prediction accuracy measured by MSCV and MSCVnorm for all simulation models but IV. For the simulation model I, which only contains the major components, the AHM has comparable prediction performance with the MajorQuad model. The simulation model II is essentially a linear Scheffé model disregarding the MoM structure because it is an additive model of all the terms $c_k x_{k1}$ and $c_k x_{k2}$, which are all the minor components proportions with respect to all the entire mixture. For this model, the AHM has competitive prediction performance comparable with 1st_MM, but better prediction performance than the MajorLinear and the MajorQuad. We also notice that, in the simulation model III containing the interactions between the major and corresponding minor components, both the AHM and the 2nd_MM have good prediction performance. For the simulation model IV, which contains the interaction terms between the major and non-corresponding minor components, the AHM and the 2nd_MM do not have as competitive prediction performance as the MultipleScheffe model. One possible explanation is that the crossed-component interaction terms are not included in the proposed AHM in this study. The simulation model V contains one interaction term between c_2 and c_3, and one between minor components $c_1 x_{11}$ and $c_1 x_{12}$, and the prediction performance of the AHM is best and close to that of the true model.

In terms of model fitting, the measures R^2, AICc, and MSE values in Table 1 indicate that the AHM performs satisfactorily. The model size of the AHM varies across different settings because of the variable selection performed via the nonnegative garrote method but is often larger than that of 1st_MM but smaller than that of 2nd_MM. Note that throughout this work, the best (or close to best) values are in bold font in all the tables.

The idea of the maximin design is to spread design points in the constrained space by maximizing the minimum distance between all pairs of points. Although originally proposed for computer experiments (Johnson, Moore, and Ylvisaker 1990), the maximin design can be applied to mixture experiments too, except that the experimental space is confined to a polyhedron, more complicated than the typical cubic for the computer experiments. We used a stochastic search strategy

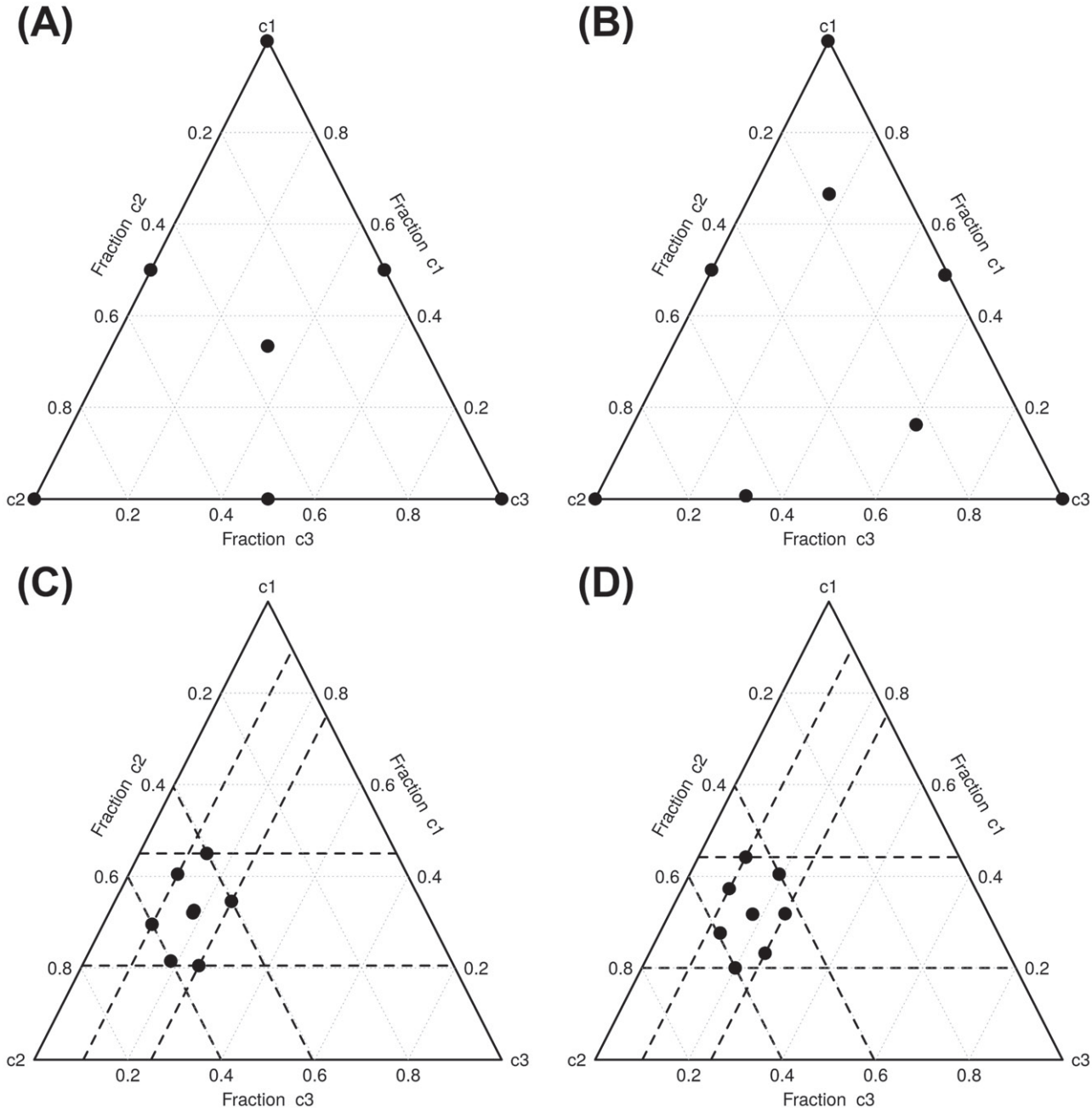

Figure 1. Designs for the major components. In the unconstrained mixture experiment are (A) the I-optimal design with 7 design points and (B) the maximin distance design with 8 design points. In the constrained mixture experiment are (C) the I-optimal design with 8 design points and (D) the maximin distance design with 8 design points. In (C) and (D) the dashed lines represent the upper and lower constraints for each mixture component.

to find the optimal design under the maximin distance criterion. Due to the nature of the stochastic search, the optimal design varies somewhat in each search, which is different from the simplex designs. Thus in each replicate of the simulation, the maximin design might be slightly different. The algorithm to generate the maximin distance design in the mixture experiment is available in the supplemental materials. Figure 1(B) shows the maximin distance design for the three major components.

Table 2 compares the performances of different models in terms of the same measurements as above. The results are similar to the ones in Table 1. The AHM has a better prediction performance than other methods in terms of MSCV and MSCVnorm under different true simulation models but IV. The R^2, AICc, and MSE values show that the AHM ranks with the best fitting models in all scenarios. For the model generating models I, IV, and V, the 2nd_MM have comparable fitting performance as the AHM. The model size of the AHM is larger than that of the 1st_MM but smaller than that of the 2nd_MM.

4.2. Constrained MoM Experiments

There have been many mixture experiments with additional constraints imposed on the components. Here we also consider simulations where certain lower and upper bounds are placed on

Table 1. Performance comparisons of models under the unconstrained MoM experiment using I-optimal design for major components from 100 simulation replications (means and standard errors (in parenthesis)).

	Model	R^2	MSE	MSCV	MSCVnorm	AICc	Size
I	1st_MM	0.970 (0.002)	14.98 (0.86)	15.47 (0.90)	0.30 (0.02)	519.8 (11.0)	6.0 (0.0)
	2nd_MM	0.976 (0.001)	12.86 (0.42)	14.28 (0.52)	0.28 (0.02)	506.1 (6.3)	18.0 (0.0)
	AHM	0.975 (0.001)	12.65 (0.30)	13.20 (0.28)	**0.26** (0.02)	**491.1** (4.4)	8.5 (1.5)
	MajorLinear	0.970 (0.002)	14.98 (0.82)	15.21 (0.83)	0.30 (0.02)	516.5 (10.3)	3.0 (0.0)
	MajorQuad	0.975 (0.001)	12.84 (0.22)	13.26 (0.23)	**0.26** (0.02)	**490.9** (3.3)	6.0 (0.0)
	MultipleScheffe	0.973 (0.002)	15.23 (0.98)	17.85 (1.37)	0.35 (0.03)	546.7 (12.1)	24.0 (0.0)
	trueModel	0.974 (0.001)	12.85 (0.15)	13.13 (0.15)	0.26 (0.02)	488.9 (2.2)	4.0 (0.0)
II	1st_MM	0.988 (0.000)	16.56 (0.24)	17.10 (0.25)	**0.26** (0.02)	**539.0** (2.8)	6.0 (0.0)
	2nd_MM	0.989 (0.000)	16.54 (0.55)	18.34 (0.64)	0.28 (0.02)	553.7 (6.4)	18.0 (0.0)
	AHM	0.988 (0.000)	16.37 (0.41)	17.09 (0.41)	**0.26** (0.02)	**539.8** (4.5)	8.5 (1.3)
	MajorLinear	0.980 (0.001)	27.30 (1.62)	27.81 (1.66)	0.42 (0.03)	629.9 (11.3)	3.0 (0.0)
	MajorQuad	0.980 (0.001)	27.46 (1.67)	28.41 (1.74)	0.43 (0.03)	634.2 (11.6)	6.0 (0.0)
	MultipleScheffe	0.989 (0.000)	16.62 (0.57)	19.34 (0.80)	0.29 (0.02)	563.4 (6.6)	24.0 (0.0)
	trueModel	0.988 (0.000)	16.56 (0.24)	17.10 (0.25)	0.26 (0.02)	539.0 (2.8)	6.0 (0.0)
III	1st_MM	0.922 (0.005)	159.18 (8.99)	164.45 (9.31)	0.31 (0.02)	966.5 (10.7)	6.0 (0.0)
	2nd_MM	0.939 (0.003)	133.82 (4.59)	148.35 (5.47)	0.28 (0.02)	948.9 (6.5)	18.0 (0.0)
	AHM	0.937 (0.003)	132.72 (3.37)	138.92 (3.34)	**0.26** (0.02)	**936.5** (4.7)	9.5 (1.4)
	MajorLinear	0.816 (0.011)	371.83 (23.25)	377.82 (23.64)	0.71 (0.03)	1123.4 (11.9)	3.0 (0.0)
	MajorQuad	0.825 (0.011)	359.80 (22.41)	371.58 (23.16)	0.70 (0.04)	1120.5 (11.9)	6.0 (0.0)
	MultipleScheffe	0.930 (0.005)	160.31 (9.95)	189.79 (13.42)	0.36 (0.03)	991.5 (11.8)	24.0 (0.0)
	trueModel	0.935 (0.003)	133.50 (1.74)	137.12 (1.83)	0.26 (0.02)	932.4 (2.5)	5.0 (0.0)
IV	1st_MM	0.840 (0.011)	289.58 (21.96)	302.55 (22.87)	0.63 (0.03)	1079.3 (14.4)	6.0 (0.0)
	2nd_MM	0.861 (0.011)	270.02 (21.18)	296.74 (22.96)	0.62 (0.04)	1081.1 (14.9)	18.0 (0.0)
	AHM	0.845 (0.012)	281.78 (21.85)	293.94 (22.50)	0.61 (0.03)	1075.1 (14.5)	6.9 (1.2)
	MajorLinear	0.834 (0.012)	296.36 (21.88)	302.38 (22.32)	0.63 (0.03)	1080.4 (14.0)	3.0 (0.0)
	MajorQuad	0.835 (0.012)	299.17 (22.29)	309.86 (23.07)	0.64 (0.03)	1085.5 (14.1)	6.0 (0.0)
	MultipleScheffe	0.939 (0.003)	123.07 (4.39)	144.79 (5.62)	**0.30** (0.02)	**941.8** (6.8)	24.0 (0.0)
	trueModel	0.932 (0.003)	122.95 (1.91)	126.35 (2.03)	0.26 (0.02)	916.8 (3.0)	5.0 (0.0)
V	1st_MM	0.964 (0.003)	18.66 (1.51)	19.27 (1.56)	0.44 (0.04)	561.0 (15.5)	6.0 (0.0)
	2nd_MM	0.967 (0.003)	18.52 (1.57)	19.91 (1.63)	0.46 (0.04)	574.6 (16.2)	18.0 (0.0)
	AHM	0.980 (0.001)	10.78 (0.29)	11.28 (0.30)	**0.26** (0.02)	**463.1** (5.0)	10.5 (1.2)
	MajorLinear	0.964 (0.003)	18.55 (1.49)	18.96 (1.54)	0.43 (0.04)	556.5 (15.4)	3.0 (0.0)
	MajorQuad	0.965 (0.003)	18.02 (1.44)	18.68 (1.50)	0.43 (0.03)	554.4 (15.2)	6.0 (0.0)
	MultipleScheffe	0.966 (0.003)	19.52 (1.66)	22.99 (2.09)	0.53 (0.05)	593.2 (16.3)	24.0 (0.0)
	trueModel	0.979 (0.001)	10.82 (0.17)	11.12 (0.19)	0.26 (0.02)	457.4 (3.0)	5.0 (0.0)

both major and minor components. Specifically, we assume that the major and minor components have to satisfy the following constraints:

$$\begin{aligned} c_1 + c_2 + c_3 = 1,\ \ 0.2 \le c_1 \le 0.45,\\ 0.4 \le c_2 \le 0.6,\ \ 0.1 \le c_3 \le 0.25,\\ x_{11} + x_{12} = 1,\ \ 0.5 \le x_{11} \le 0.85,\\ x_{21} + x_{22} = 1,\ \ 0.73 \le x_{21} \le 0.95,\\ x_{31} + x_{32} = 1,\ \ 0.68 \le x_{31} \le 0.92. \end{aligned} \tag{11}$$

Figure 1(C) shows the I-optimal design for the second-order Scheffé model for the major components in the constrained mixture experiment. The design is generated by the AlgDesign package in R software. Figure 1(D) shows the maximin distance design for the major components in the constrained mixture experiment. The comparison results for using the I-optimal design and the maximin distance design for the major components are reported in Tables 3 and 4, respectively. From both tables, we can conclude that the AHM, the 1st_MM, and the 2nd_MM all have competitive prediction performance.

It is worth noting that in the simulation with data generating model IV, the AHM has a comparable prediction performance as the true model, which is an improvement from the unconstrained MoM case. This phenomenon is likely due to the additional constraints which make the design space more complicated, and thus the flexibility of the AHM is more advantageous than its counterparts for the true underlying model IV. The AHM has similar R^2, AICc, and MSE values as the 1st_MM for simulation models I–V. The model size of the AHM is larger than that of the 1st_MM but smaller than that of the 2nd_MM. Note that the simulation results and conclusions for the type (b) MoM experiments in the supplemental materials are quite similar to those of the type (a) MoM experiments.

5. Real-Data Analysis

In this section, we analyze two real-data problems studied previously in the literature, the photoresist-coating experiment (Cornell and Ramsey 1998) and the Pringles experiment (Kang, Joseph, and Brenneman 2011), to evaluate the model performance of the proposed AHM.

5.1. Photoresist-Coating Experiment

The objective of the photoresist-coating experiment is to determine the effect of proportions of resins in the formulation on the photoresist material's characteristic of interest (Cornell and Ramsey 1998). The two major components (c_1 and c_2) are the base resin types, and the minor components are the minor resins

Table 2. Performance comparisons of models under the unconstrained MoM experiment using the maximin distance design for major components from 100 simulation replications (means and standard errors (in parenthesis)).

	Model	R^2	MSE	MSCV	MSCVnorm	AICc	Size
I	1st_MM	0.973 (0.004)	13.54 (2.19)	13.93 (2.25)	0.30 (0.02)	567.9 (40.2)	6.0 (0.0)
	2nd_MM	0.978 (0.003)	11.80 (1.73)	12.90 (1.89)	0.28 (0.02)	553.4 (36.8)	18.0 (0.0)
	AHM	0.977 (0.003)	11.66 (1.69)	12.13 (1.75)	**0.26** (0.01)	**539.5** (36.7)	8.8 (1.5)
	MajorLinear	0.973 (0.004)	13.51 (2.17)	13.70 (2.20)	0.29 (0.02)	564.2 (39.9)	3.0 (0.0)
	MajorQuad	0.976 (0.003)	11.80 (1.69)	12.13 (1.74)	**0.26** (0.01)	**538.8** (36.1)	6.0 (0.0)
	MultipleScheffe	0.975 (0.004)	13.67 (2.27)	15.69 (2.67)	0.33 (0.03)	592.8 (41.2)	24.0 (0.0)
	trueModel	0.976 (0.003)	11.80 (1.69)	12.02 (1.72)	0.26 (0.01)	536.6 (36.1)	4.0 (0.0)
II	1st_MM	0.989 (0.001)	15.24 (2.23)	15.68 (2.30)	**0.26** (0.02)	**593.9** (37.9)	6.0 (0.0)
	2nd_MM	0.990 (0.001)	15.25 (2.27)	16.66 (2.48)	0.27 (0.02)	608.6 (38.6)	18.0 (0.0)
	AHM	0.989 (0.001)	15.10 (2.22)	15.69 (2.31)	**0.26** (0.02)	**594.7** (38.3)	8.4 (1.4)
	MajorLinear	0.981 (0.002)	25.56 (3.62)	25.96 (3.67)	0.43 (0.03)	702.8 (34.2)	3.0 (0.0)
	MajorQuad	0.981 (0.002)	25.71 (3.64)	26.48 (3.74)	0.44 (0.03)	707.3 (34.2)	6.0 (0.0)
	MultipleScheffe	0.990 (0.001)	15.26 (2.29)	17.38 (2.60)	0.29 (0.02)	617.0 (38.6)	24.0 (0.0)
	trueModel	0.989 (0.001)	15.24 (2.23)	15.68 (2.30)	0.26 (0.02)	593.9 (37.9)	6.0 (0.0)
III	1st_MM	0.929 (0.007)	149.03 (15.58)	153.36 (16.05)	0.31 (0.02)	1088.1 (23.6)	6.0 (0.0)
	2nd_MM	0.943 (0.004)	127.12 (9.87)	139.07 (10.99)	0.28 (0.02)	1069.0 (17.4)	18.0 (0.0)
	AHM	0.942 (0.004)	123.48 (8.62)	128.68 (8.96)	**0.26** (0.02)	**1052.2** (15.4)	9.5 (1.5)
	MajorLinear	0.827 (0.012)	359.67 (34.80)	364.74 (35.17)	0.73 (0.04)	1275.4 (21.0)	3.0 (0.0)
	MajorQuad	0.834 (0.011)	350.05 (33.99)	360.07 (34.70)	0.72 (0.05)	1272.8 (20.9)	6.0 (0.0)
	MultipleScheffe	0.937 (0.007)	146.10 (16.10)	168.50 (19.52)	0.34 (0.03)	1106.6 (24.8)	24.0 (0.0)
	trueModel	0.941 (0.004)	124.18 (8.67)	127.12 (8.88)	0.25 (0.01)	1048.3 (15.8)	5.0 (0.0)
IV	1st_MM	0.846 (0.013)	275.88 (30.50)	286.01 (31.66)	0.62 (0.05)	1221.0 (25.2)	6.0 (0.0)
	2nd_MM	0.872 (0.015)	243.74 (34.31)	265.99 (36.78)	0.58 (0.06)	1207.9 (33.3)	18.0 (0.0)
	AHM	0.867 (0.016)	241.91 (34.45)	252.98 (35.77)	0.55 (0.06)	1194.4 (32.3)	8.4 (1.4)
	MajorLinear	0.839 (0.011)	285.12 (27.88)	289.93 (28.40)	0.63 (0.05)	1225.1 (21.8)	3.0 (0.0)
	MajorQuad	0.840 (0.011)	287.56 (28.17)	296.60 (29.14)	0.65 (0.05)	1230.3 (21.8)	6.0 (0.0)
	MultipleScheffe	0.941 (0.006)	115.03 (11.83)	131.51 (13.60)	**0.29** (0.02)	**1055.1** (23.9)	24.0 (0.0)
	trueModel	0.936 (0.006)	114.74 (10.66)	117.50 (10.90)	0.26 (0.02)	1030.7 (21.8)	5.0 (0.0)
V	1st_MM	0.967 (0.006)	17.03 (3.03)	17.49 (3.11)	0.43 (0.05)	616.5 (45.6)	6.0 (0.0)
	2nd_MM	0.969 (0.006)	16.90 (3.09)	18.08 (3.25)	0.45 (0.05)	629.4 (46.6)	18.0 (0.0)
	AHM	0.981 (0.002)	9.96 (1.36)	10.37 (1.42)	**0.26** (0.02)	**508.0** (33.7)	10.7 (1.4)
	MajorLinear	0.967 (0.006)	16.92 (3.00)	17.23 (3.05)	0.43 (0.05)	612.0 (45.3)	3.0 (0.0)
	MajorQuad	0.968 (0.006)	16.50 (2.98)	17.04 (3.08)	0.42 (0.05)	609.6 (46.3)	6.0 (0.0)
	MultipleScheffe	0.969 (0.006)	17.67 (3.19)	20.19 (3.63)	0.50 (0.06)	647.4 (45.9)	24.0 (0.0)
	trueModel	0.980 (0.002)	10.00 (1.34)	10.24 (1.37)	0.25 (0.02)	502.4 (33.1)	5.0 (0.0)

possessing different dissolution rates (slow and fast) denoted as x_{11}, x_{12}, and x_{21}, x_{22}, respectively. The range of values for all components is $[0, 1]$. All possible settings of (c_1, c_2) are $(0.75, 0.25)$, $(0.5, 0.5)$, and $(0.25, 0.75)$. For (x_{k1}, x_{k2}) with $k = 1, 2$, three settings $(1, 0)$, $(0.5, 0.5)$, and $(0, 1)$ are chosen. The overall design is a crossed design of the three designs containing $3 \times 3 \times 3 = 27$ design points. In total, 42 measurements are observed with replications at certain design points.

Table 5 compares the performances of different models. The AHM has the smallest MSCV and AICc values among all. The MajorLinear and MajorQuad model have the worst prediction performance, indicating that the minor components play an important role. The 2nd_MM has better prediction and model fitting than the 1st_MM and the MultipleScheffe. The model size of the AHM is larger than that of the 1st_MM and 2nd_MM, but smaller than that of the MultipleScheffe.

The fitted AHM is

$$\begin{aligned} y = {} & 25.919c_1 + 29.21c_2 - 6.536c_1^{1.1}x_{11} + 23.616c_1^{1.1}x_{12} \\ & - 5.58c_2^{1.1}x_{21} + 30.706c_2^{1.1}x_{22} - 38.974c_1c_2 \\ & - 18.818(c_1^{1.1})^2x_{11}x_{12} - 19.363(c_2^{1.1})^2x_{21}x_{22}. \end{aligned}$$

Based on the estimated parameters in the fitted AHM, the major components have significant effects on the response, and both main and interaction effects of the minor components depend on their respective major components. These results are consistent with the findings in the article (Cornell and Ramsey 1998). However, the AHM reveals that the inter-major-component blending property exists via the interaction term between the major components, c_1c_2, which is different from the multiple Scheffé model results (Cornell and Ramsey 1998). The multiple Scheffé model assumes that the blending properties of the minor components of one major component also depend on the presence of minor components of other major components. The reason for this different interpretation is that AHM considers all the inter-major-component interactions only at the major-component level, and all the inter-minor-component interactions are restricted within the minor components nested under the same major component. As a result, no inter-minor-component-interactions are considered for the minor components nested under different major components.

5.2. Pringles Experiment

The goal of Pringles® experiment is to develop a new kind of Pringles® potato crisp such that the percentage of fat and the hardness in the potato crisps are optimized. There are three major components A, B, and C, whose proportion are denoted by c_1, c_2, and c_3. The major component A is composed of two minor components A_1 and A_2 with proportion x_{11} and x_{12} with

Table 3. Performance comparisons of models under the constrained MoM experiment using I-optimal design for major components from 100 simulation replications (means and standard errors (in parenthesis)).

	Model	R^2	MSE	MSCV	MSCVnorm	AICc	Size
I	1st_MM	0.999 (0.000)	0.47 (0.01)	0.49 (0.01)	**0.26** (0.02)	−152.8 (4.9)	6.0 (0.0)
	2nd_MM	0.999 (0.000)	0.47 (0.01)	0.51 (0.02)	0.27 (0.02)	−142.1 (6.6)	18.0 (0.0)
	AHM	0.999 (0.000)	0.47 (0.01)	0.48 (0.01)	**0.26** (0.02)	−153.5 (6.1)	8.9 (2.0)
	MajorLinear	0.999 (0.000)	0.47 (0.01)	0.48 (0.01)	**0.26** (0.02)	**−156.3** (3.8)	3.0 (0.0)
	MajorQuad	0.999 (0.000)	0.47 (0.01)	0.48 (0.01)	**0.26** (0.02)	**−156.5** (3.9)	6.0 (0.0)
	MultipleScheffe	0.999 (0.000)	0.47 (0.02)	0.54 (0.02)	0.29 (0.02)	−130.1 (7.8)	24.0 (0.0)
	trueModel	0.999 (0.000)	0.47 (0.01)	0.48 (0.01)	0.26 (0.02)	−158.5 (3.0)	4.0 (0.0)
II	1st_MM	0.999 (0.000)	1.14 (0.02)	1.17 (0.02)	**0.26** (0.02)	**36.6** (3.3)	6.0 (0.0)
	2nd_MM	0.999 (0.000)	1.14 (0.03)	1.24 (0.03)	0.27 (0.02)	50.8 (5.4)	18.0 (0.0)
	AHM	0.999 (0.000)	1.13 (0.02)	1.17 (0.02)	**0.26** (0.02)	38.6 (4.9)	8.8 (1.4)
	MajorLinear	0.999 (0.000)	1.68 (0.11)	1.70 (0.12)	0.37 (0.02)	116.3 (14.7)	3.0 (0.0)
	MajorQuad	0.999 (0.000)	1.68 (0.12)	1.73 (0.12)	0.38 (0.02)	120.4 (15.0)	6.0 (0.0)
	MultipleScheffe	0.999 (0.000)	1.14 (0.04)	1.30 (0.05)	0.28 (0.02)	58.8 (7.4)	24.0 (0.0)
	trueModel	0.999 (0.000)	1.14 (0.02)	1.17 (0.02)	0.26 (0.02)	36.6 (3.3)	6.0 (0.0)
III	1st_MM	0.998 (0.000)	8.53 (0.20)	8.78 (0.21)	**0.26** (0.01)	**471.5** (5.2)	6.0 (0.0)
	2nd_MM	0.998 (0.000)	8.33 (0.25)	9.10 (0.28)	0.27 (0.02)	480.9 (6.7)	18.0 (0.0)
	AHM	0.998 (0.000)	8.37 (0.18)	8.53 (0.17)	**0.26** (0.01)	**470.9** (5.1)	9.2 (1.3)
	MajorLinear	0.994 (0.000)	26.73 (1.49)	27.11 (1.51)	0.81 (0.03)	714.6 (12.2)	3.0 (0.0)
	MajorQuad	0.994 (0.000)	26.96 (1.49)	27.73 (1.54)	0.83 (0.03)	719.7 (12.1)	6.0 (0.0)
	MultipleScheffe	0.998 (0.000)	8.35 (0.31)	9.50 (0.39)	0.29 (0.02)	489.6 (8.1)	24.0 (0.0)
	trueModel	0.998 (0.000)	8.35 (0.10)	8.55 (0.10)	0.26 (0.01)	465.7 (2.6)	5.0 (0.0)
IV	1st_MM	0.994 (0.000)	12.11 (0.56)	12.48 (0.58)	0.30 (0.02)	547.0 (10.0)	6.0 (0.0)
	2nd_MM	0.995 (0.000)	10.42 (0.30)	11.39 (0.35)	0.28 (0.02)	529.2 (6.3)	18.0 (0.0)
	AHM	0.995 (0.000)	10.71 (0.30)	11.01 (0.31)	**0.27** (0.02)	**523.6** (6.4)	8.7 (1.1)
	MajorLinear	0.989 (0.001)	24.46 (1.68)	24.81 (1.70)	0.60 (0.03)	695.2 (14.8)	3.0 (0.0)
	MajorQuad	0.989 (0.001)	24.65 (1.70)	25.36 (1.74)	0.61 (0.03)	700.2 (14.8)	6.0 (0.0)
	MultipleScheffe	0.996 (0.000)	10.33 (0.30)	11.76 (0.41)	0.29 (0.02)	535.7 (6.3)	24.0 (0.0)
	trueModel	0.995 (0.000)	10.31 (0.14)	10.55 (0.15)	0.26 (0.02)	511.3 (3.0)	5.0 (0.0)
V	1st_MM	0.999 (0.000)	0.39 (0.01)	0.41 (0.01)	0.27 (0.02)	−192.4 (7.2)	6.0 (0.0)
	2nd_MM	0.999 (0.000)	0.38 (0.02)	0.42 (0.02)	0.28 (0.02)	−185.9 (8.6)	18.0 (0.0)
	AHM	0.999 (0.000)	0.37 (0.01)	0.38 (0.01)	**0.26** (0.02)	**−200.7** (4.6)	9.8 (1.4)
	MajorLinear	0.999 (0.000)	0.50 (0.03)	0.51 (0.03)	0.34 (0.02)	−145.6 (14.9)	3.0 (0.0)
	MajorQuad	0.999 (0.000)	0.49 (0.03)	0.51 (0.04)	0.34 (0.02)	−145.0 (15.3)	6.0 (0.0)
	MultipleScheffe	0.999 (0.000)	0.39 (0.02)	0.45 (0.02)	0.30 (0.02)	−171.8 (10.1)	24.0 (0.0)
	trueModel	0.999 (0.000)	0.37 (0.01)	0.38 (0.01)	0.26 (0.02)	−206.3 (3.0)	5.0 (0.0)

respect to c_1, and B is composed of two minor components B_1 and B_2 with proportion x_{21} and x_{22} with respect to c_2. Component C is pure material which can be considered to have only a single minor component. The constraints on the components are given by

$$\begin{aligned} c_1 + c_2 + c_3 = 1,\ \ & 0.601 \le c_1 \le 0.643, \\ 0.34 \le c_2 \le 0.38,\ \ & 0.017 \le c_3 \le 0.019, \\ x_{11} + x_{12} = 1,\ \ & x_{21} + x_{22} = 1, \\ 0.835 \le x_{11} \le 0.905,\ \ & 0.095 \le x_{12} \le 0.165, \\ 0.9 \le x_{21} \le 0.98,\ \ & 0.02 \le x_{22} \le 0.1. \end{aligned}$$

The hardness of the potato chip (Hardness) and the percentage of fat (%Fat) are the two response variables. The experimental design is illustrated in Kang, Joseph, and Brenneman (2011) in details.

Table 6 compares the model performances for the Pringles experiment. For both responses "Hardness" and "%Fat," compared to the 1st_MM, the AHM has a smaller MSCV value, but a larger AICc value, suggesting better prediction performance but worse fitting performance. The MajorLinear and MajorQuad model have the largest MSCV values, indicating that the minor components play an important role in this study. The MultipleScheffe model and the 2nd_MM has largest AICc values. The model size of the AHM is larger than that of the 1st_MM, but smaller than that of the 2nd_MM.

We use the fitted AHM to find the optimal settings to maximize the response "Hardness." The fitted AHM is

$$\begin{aligned} \hat{y}_{\text{hardness}} = {} & 9.745c_1 - 5.115c_2 + 6.916c_1^{1.3}x_{11} - 11.184c_2^{1.3}x_{21} \\ & + 27.203c_2^{1.3}x_{22} + 21.176(c_2^{1.3})^2 x_{21}x_{22}. \end{aligned}$$

The 1st_MM proposed by Kang, Joseph, and Brenneman (2011) is

$$\begin{aligned} \hat{y}_{\text{hardness}} = {} & 8.786c_1 + 20.966c_2 + 13.506c_3 \\ & + 8.658c_1x_{11} - 37.641c_2x_{21}. \end{aligned}$$

Compared to the 1st_MM, the fitted AHM does not contain the third major component c_3. Similarly, we can also use the fitted AHM to find the minimizer of "%Fat." Table 7 shows the optimal settings to minimize the response "%Fat" and to maximize the response "Hardness," respectively. The optimization can be performed using the constrained nonlinear optimization in R software. These optimal settings agree well with Kang, Joseph, and Brenneman (2011). This experiment is a preliminary study, and in the follow-up experiments, larger experiments should be conducted around the optimal settings to find better formulations.

Table 4. Performance comparisons of models under the constrained MoM experiment using the maximin distance design for major components from 100 simulation replications (means and standard errors (in parenthesis)).

	Model	R^2	MSE	MSCV	MSCVnorm	AICc	Size
I	1st_MM	0.999 (0.000)	0.45 (0.05)	0.46 (0.05)	**0.26** (0.02)	−167.4 (24.0)	6.0 (0.0)
	2nd_MM	0.999 (0.000)	0.44 (0.05)	0.48 (0.05)	0.27 (0.02)	−155.3 (24.3)	18.0 (0.0)
	AHM	0.999 (0.000)	0.44 (0.05)	0.45 (0.05)	**0.26** (0.02)	−167.4 (24.6)	9.0 (1.8)
	MajorLinear	0.999 (0.000)	0.45 (0.05)	0.45 (0.05)	**0.26** (0.01)	**−170.9** (24.0)	3.0 (0.0)
	MajorQuad	0.999 (0.000)	0.44 (0.05)	0.45 (0.05)	**0.26** (0.01)	**−170.2** (24.2)	6.0 (0.0)
	MultipleScheffe	0.999 (0.000)	0.45 (0.05)	0.51 (0.06)	0.29 (0.02)	−141.0 (25.5)	26.5 (1.1)
	trueModel	0.999 (0.000)	0.44 (0.05)	0.45 (0.05)	0.26 (0.01)	−172.4 (24.2)	4.0 (0.0)
II	1st_MM	0.999 (0.000)	1.07 (0.09)	1.10 (0.10)	**0.26** (0.02)	**21.6** (19.6)	6.0 (0.0)
	2nd_MM	0.999 (0.000)	1.06 (0.09)	1.16 (0.10)	0.27 (0.02)	35.3 (19.9)	18.0 (0.0)
	AHM	0.999 (0.000)	1.06 (0.09)	1.09 (0.10)	**0.26** (0.02)	23.9 (20.0)	8.9 (1.4)
	MajorLinear	0.999 (0.000)	1.59 (0.16)	1.62 (0.16)	0.38 (0.03)	104.9 (21.4)	3.0 (0.0)
	MajorQuad	0.999 (0.000)	1.60 (0.16)	1.65 (0.16)	0.39 (0.03)	109.2 (21.7)	6.0 (0.0)
	MultipleScheffe	0.999 (0.000)	1.07 (0.10)	1.22 (0.12)	0.29 (0.02)	47.8 (21.4)	26.5 (1.1)
	trueModel	0.999 (0.000)	1.07 (0.09)	1.10 (0.10)	0.26 (0.02)	21.6 (19.6)	6.0 (0.0)
III	1st_MM	0.998 (0.000)	8.44 (0.31)	8.68 (0.32)	**0.26** (0.02)	**469.0** (7.9)	6.0 (0.0)
	2nd_MM	0.998 (0.000)	8.32 (0.31)	9.10 (0.35)	0.28 (0.02)	480.6 (8.2)	18.0 (0.0)
	AHM	0.998 (0.000)	8.35 (0.28)	8.50 (0.28)	**0.26** (0.02)	**470.4** (7.3)	9.3 (1.0)
	MajorLinear	0.994 (0.000)	26.70 (1.80)	27.08 (1.82)	0.82 (0.03)	714.2 (14.7)	3.0 (0.0)
	MajorQuad	0.994 (0.000)	26.96 (1.82)	27.73 (1.87)	0.84 (0.03)	719.5 (14.7)	6.0 (0.0)
	MultipleScheffe	0.998 (0.000)	8.34 (0.36)	9.65 (0.49)	0.29 (0.02)	493.0 (9.8)	26.5 (1.1)
	trueModel	0.998 (0.000)	8.31 (0.24)	8.51 (0.25)	0.26 (0.02)	464.6 (6.4)	5.0 (0.0)
IV	1st_MM	0.995 (0.000)	11.11 (0.81)	11.44 (0.83)	0.30 (0.02)	528.0 (16.1)	6.0 (0.0)
	2nd_MM	0.996 (0.000)	9.63 (0.54)	10.53 (0.59)	0.28 (0.02)	512.1 (12.2)	18.0 (0.0)
	AHM	0.995 (0.000)	9.92 (0.58)	10.21 (0.60)	**0.27** (0.02)	**506.8** (12.9)	8.7 (1.2)
	MajorLinear	0.989 (0.001)	23.49 (1.62)	23.82 (1.64)	0.62 (0.04)	686.4 (15.0)	3.0 (0.0)
	MajorQuad	0.989 (0.001)	23.68 (1.65)	24.36 (1.70)	0.64 (0.04)	691.5 (15.1)	6.0 (0.0)
	MultipleScheffe	0.996 (0.000)	9.57 (0.56)	11.01 (0.69)	0.29 (0.02)	522.7 (12.8)	26.5 (1.1)
	trueModel	0.996 (0.000)	9.58 (0.46)	9.80 (0.47)	0.26 (0.02)	495.1 (10.6)	5.0 (0.0)
V	1st_MM	0.999 (0.000)	0.35 (0.03)	0.36 (0.03)	0.27 (0.02)	−218.3 (20.4)	6.0 (0.0)
	2nd_MM	0.999 (0.000)	0.34 (0.03)	0.37 (0.03)	0.28 (0.02)	−209.1 (20.5)	18.0 (0.0)
	AHM	0.999 (0.000)	0.33 (0.03)	0.34 (0.03)	**0.26** (0.02)	**−225.7** (20.4)	9.8 (1.5)
	MajorLinear	0.999 (0.000)	0.46 (0.04)	0.47 (0.05)	0.35 (0.02)	−162.7 (21.2)	3.0 (0.0)
	MajorQuad	0.999 (0.000)	0.46 (0.04)	0.47 (0.05)	0.36 (0.03)	−161.3 (21.3)	6.0 (0.0)
	MultipleScheffe	0.999 (0.000)	0.35 (0.03)	0.40 (0.04)	0.30 (0.02)	−193.5 (20.9)	26.5 (1.1)
	trueModel	0.999 (0.000)	0.33 (0.03)	0.34 (0.03)	0.26 (0.01)	−231.5 (19.9)	5.0 (0.0)

Table 5. Performance comparisons of models in the photoresist-coating experiment.

Model	R^2	MSE	MSCV	AICc	Size
AHM	0.998	1.929	**2.324**	**44.552**	9
MultipleScheffe	1.000	0.159	35.193	60.378	27
MajorLinear	0.902	90.321	95.529	193.724	2
MajorQuad	0.903	91.312	98.336	195.569	3
1st_MM	0.995	4.591	5.294	71.477	4
2nd_MM	0.998	2.425	2.860	51.952	8

6. Discussion

The intrinsic relationship between the major and minor components is a key feature in the MoM experiment. This work proposes an AHM with a meaningful interpretation of the model structure for MoM experiments. The AHM considers the effects of major and minor components in an additive fashion and employs the hierarchical and heredity principles by the nonnegative garrote technique for model selection. The AHM incorporates the dependence between the major and minor components via the coefficients of minor components. The coefficient functions represent various types of knowledge. For example, when one major component is not included, all of its corresponding minor components are excluded from the model. According to the numerical studies, the AHM provides superior prediction performances compared to the benchmark models.

It is worth remarking that the MoM experiment is closely related to the multilevel model. The multilevel model usually has two types of variables, the group-level variables and the individual-level variables (Dedrick et al. 2009). The individual-level variable has a direct effect on the response, while the group-level variable contributes to the response in both direct and indirect ways. For the example of using the random intercept in the multilevel model, the intercept term will be dependent upon the group-level variable, representing the direct contribution of the group-level variable. The interaction between the individual-level variable and the group-level variable represents the indirect contribution of the group-level variable. However, in the proposed AHM, the contribution of the minor components (the individual-level variable) is made through a function depending on the major components (the group-level variable).

The proposed AHM can be extended to the general varying-coefficient models (Hastie and Tibshirani 1993; Fan and Zhang 1999). For example, one can consider the coefficients for the minor components to be nonparametric functions of the corresponding major components. In this work, we adopted a parametric power function of order h to express the structural dependence between the major and minor components. This power function is monotonic and bounded on the domain of

Table 6. Comparison between proposed models.

Response	Model	R^2	MSE	MSCV	AICc	Size
%Fat	AHM	1.000	0.261	**0.362**	10.048	8
	MultipleScheffe	1.000	0.230	0.794	101.560	12
	MajorLinear	0.999	1.426	1.757	14.068	3
	MajorQuad	0.999	1.569	3.508	22.131	5
	1st_MM	1.000	0.296	0.421	**−6.202**	5
	2nd_MM	1.000	0.102	0.590	87.772	12
Hardness	AHM	0.996	0.157	**0.174**	−12.439	6
	MultipleScheffe	0.999	0.069	0.235	81.138	12
	MajorLinear	0.985	0.487	0.600	−4.194	3
	MajorQuad	0.986	0.546	0.856	4.195	5
	1st_MM	0.997	0.128	0.183	**−20.474**	5
	2nd_MM	0.999	0.130	0.428	91.823	12

Table 7. Optimal settings from the AHMs.

Response	c_1	c_2	c_3	x_{11}	x_{12}	x_{21}	x_{22}
%Fat	0.641	0.34	0.019	0.892	0.108	0.9	0.1
Hardness	0.643	0.34	0.017	0.905	0.095	0.9	0.1

c_k. It will be interesting to investigate how to incorporate an appropriate nonparametric form, allowing flexible structures to describe the structural dependence of minor components on their corresponding major components.

Supplementary Materials

The online supplementary materials include the lemma for the relationship between the major–minor model and the additive model, the algorithm of generating maximin distance designs for major components, and simulation results for the type (b) MoM experiments in Section 4 of the main article.

Acknowledgments

The authors acknowledge the editor, associate editor, and reviewers for their insightful comments to revise this article.

Funding

This work was supported by the National Science Foundation CMMI-1435996.

References

Borges, C., Bruns, E. R., Almeida, A. A., and Scarminio, I. S. (2007), "Mixture-Mixture Design for Fingerprint Optimization of Chromatographic Mobile Phases and Extraction Solutions for Camellia Sinensis," *Analytica Chimica Acta*, 595, 28–37. [1]

Breiman, L. (1995), "Better Subset Regression Using the Nonnegative Garrote," *Technometrics*, 37, 373–384. [2,3]

Brown, L., Donev, A. N., and Bissett, A. C. (2015), "General Blending Models for Data From Mixture Experiments," *Technometrics*, 57, 449–456. [1]

Burnham, K. P., and Anderson, D. R. (2002), *Model Selection and Multimodel Inference* (2nd ed.), New York: Springer. [5]

Coetzer, R. L., and Haines, L. M. (2013), "Optimal Designs for Multiple-Mixture by Process Variable Experiments," in *mODa 10—Advances in Model-Oriented Design and Analysis*, Contributions to Statistics, eds. D. Ucinski, A. Atkinson, and M. Patan, Heidelberg: Springer. [1]

Cornell, J. A. (2002), *Experiments With Mixtures: Designs, Models, and the Analysis of Mixture Data* (3rd ed.), New York: Wiley. [1,5]

Cornell, J. A., and Ramsey, P. J. (1998), "A Generalized Mixture Model for Categorized-Components Problems With an Application to a Photoresist-Coating Experiment," *Technometrics*, 40, 48–61. [1,5,7,8]

Cox, D. R. (1971), "A Note on Polynomial Response Functions for Mixtures," *Biometrika*, 58, 155–159. [1]

Dedrick, R. F., Ferron, J. M., Hess, M. R., Hogarty, K. Y., Kromrey, J. D., Lang, T. R., Niles, J. D., and Lee, R. S. (2009), "Multilevel Modeling: A Review of Methodological Issues and Applications," *Review of Educational Research*, 79, 69–102. [10]

Didier, C., Etcheverrigaray, M., Kratje, R., and Goicoechea, H. C. (2007), "Crossed Mixture Design and Multiple Response Analysis for Developing Complex Culture Media Used in Recombinant Protein Production," *Chemometrics and Intelligent Laboratory Systems*, 86, 1–9. [1]

Dingstad, G., Egelandsdal, B., and Næs, T. (2003), "Modeling Methods for Crossed Mixture Experiments—A Case Study From Sausage Production," *Chemometrics and Intelligent Laboratory Systems*, 66, 175–190. [1,5]

Draguljić, D., Woods, D. C., Dean, A. M., Lewis, S. M., and Vine, A. E. (2014), "Screening Strategies in the Presence of Interactions," *Technometrics*, 56, 1–16. [4,5]

Draper, N. R., and Pukelsheim, F. (1998), "Mixture Models Based on Homogeneous Polynomials," *Journal of Statistical Planning and Inference*, 71, 303–311. [1]

Fan, J., and Zhang, W. (1999), "Statistical Estimation in Varying Coefficient Models," *Annals of Statistics*, 27, 1491–1518. [10]

Goos, P., Jones, B., and Syafitri, U. (2016), "I-Optimal Design of Mixture Experiments," *Journal of the American Statistical Association*, 111, 899–911. [5]

Hastie, T., and Tibshirani, R. (1993), "Varying-Coefficient Models," *Journal of the Royal Statistical Society*, Series B, 55, 757–796. [10]

Hurvich, C. M., and Tsai, C. L. (1989), "Regression and Time Series Model Selection in Small Samples," *Biometrika*, 76, 297–307. [5]

Johnson, M. E., Moore, L. M., and Ylvisaker, D. (1990), "Minimax and Maximin Distance Designs," *Journal of Statistical Planning and Inference*, 26, 131–148. [5]

Kang, L., Joseph, V. R., and Brenneman, W. A. (2011), "Design and Modeling Strategies for Mixture-of-Mixtures Experiments," *Technometrics*, 53, 125–136. [1,2,3,4,5,7,9]

Kang, L., Salgado, J. C., and Brenneman, W. A. (2016), "Comparing the Slack-Variable Mixture Model With Other Alternatives," *Technometrics*, 58, 255–268. [1]

Khuri, A. I. (2005), "Slack-Variable Models Versus Scheffé's Mixture Models," *Journal of Applied Statistics*, 32, 887–908. [1]

Laake, P. (1975), "On the Optimal Allocation of Observations in Experiments With Mixtures," *Scandinavian Journal of Statistics*, 2, 153–157. [5]

Lambrakis, D. P. (1968), "Experiments With Mixtures: A Generalization of the Simplex-Lattice Design," *Journal of the Royal Statistical Society*, Series B, 30, 123–136. [1,5]

——— (1969), "Experiments With Mixtures: Estimated Regression Function of the Multiple-Lattice Design," *Journal of the Royal Statistical Society*, Series B, 31, 276–284. [1,5]

Lawson, J., and Willden, C. (2016), "Mixture Experiments in R Using mixexp," *Journal of Statistical Software*, 72, 1–20. [5]

Piepel, G. F. (1999), "Modeling Methods for Mixture-of-Mixtures Experiments Applied to a Tablet Formulation Problem," *Pharmaceutical Development and Technology*, 4, 593–606. [1]

——— (2007), "A Component Slope Linear Model for Mixture Experiments," *Quality Technology & Quantitative Management*, 4, 331–343. [1]

Prescott, P., Dean A. M., Draper, N. R., and Lewis, S. M. (2002), "Mixture Experiments: Ill-Conditioning and Quadratic Model Specification," *Technometrics*, 44, 260–268. [1]

Scheffé, H. (1958), "Experiments With Mixtures," *Journal of the Royal Statistical Society*, Series B, 20, 344–360. [1]

——— (1963), "The Simplex-Centroid Design for Experiments With Mixtures," *Journal of the Royal Statistical Society*, Series B, 25, 235–263. [5]

Snee, R. D., and Rayner, A. A. (1982), "Assessing the Accuracy of Mixture Model Regression Calculations," *Journal of Quality Technology*, 14, 67–79. [1]

Wu, C. F. J., and Hamada, M. S. (2009), *Experiments: Planning, Analysis, and Optimization* (2nd ed.), Hoboken, NJ: Wiley. [4]

Xiong, S. (2010), "Some Notes on the Nonnegative Garrote," *Technometrics*, 52, 349–361. [2,3,4]

Yuan, M., Joseph, V. R., and Lin, Y. (2007), "An Efficient Variable Selection Approach for Analyzing Designed Experiments," *Technometrics*, 49, 430–439. [2,3,4]

Yuan, M., Joseph, V. R., and Zou, H. (2009), "Structured Variable Selection and Estimation," *The Annals of Applied Statistics*, 3, 1738–1757. [2,3]

Yuan, M., and Lin, Y. (2007), "On the Non-Negative Garrotte Estimator," *Journal of the Royal Statistical Society*, Series B, 69, 143–161. [2,3,4]