

Efficient Mixture/Process Experiments Using Nonlinear Models
9/2015

Ronald D. Snee, Snee Associates, Newark, DE
Roger W. Hoerl, Union College, Schenectady, NY
Gabriella Bucci, Out of Door Academy, Sarasota, FL

Copyright: Ronald D. Snee, Roger W. Hoerl, and Gabriella Bucci 2015

Note: This paper is currently under editorial review under the title: “A Statistical Engineering Approach to Mixture Experiments With Process Variables”

Abstract: Experimenting with both mixture components and process variables, especially when there is likely to be interaction between these two sets of variables is discussed. We consider both design and analysis questions within the context of addressing an actual mixture/process problem. We focus on a strategy for attacking such problems, as opposed to finding the best possible design or best possible model for a given set of data. In this sense, a statistical engineering framework is used. In particular, when we consider the potential of fitting parsimonious non-linear models as opposed to larger linearized models, we find significant opportunity to reduce the size of experimental designs. It is difficult in practice to know what type of model will best fit the resulting data. Therefore, an integrated, sequential design and analysis strategy is recommended. Using two published data sets and one new data set, we find that in some cases non-linear models, or linear models ignoring interaction, enable reduction of experimentation on the order of 50%. In other cases, non-linear models may not suffice. We therefore provide guidelines as to when such an approach is likely to succeed, and propose an overall strategy for such problems.

Keywords: mixtures, mixture models, mixture experimentation, non-linear models

Introduction

Mixture problems – where the ingredient proportions must sum to 1.0 - occur in many industries and application areas, such as the development of foods, pharmaceuticals, and chemicals, to name just a few. In many mixture applications, it is not just the formulation that impacts the final

response, but also process variables. For example, in producing a blended wine the final taste and other important characteristics depend not only the proportions of the grape varieties used, but also on process variables, such as climatic conditions of the growing season, when the grapes are harvested, temperatures at various stages of the fermentation process, type of aging barrels used, and so on.

In some cases, the mixture components and process variables do not interact. In the wine example, this would mean that the fermentation and other processing could be optimized without consideration of what grape varieties were used. This scenario would lead one to utilize additive models combining a mixture model plus a process variable model. In other cases, this additive scenario will not apply, because the best processing conditions might depend on the varieties used. In that case, there would be interaction between the mixture and process variables. The interaction scenario would lead to models that incorporate terms to account for interaction between mixture and process variables.

The experimental designs typically used when such interaction is anticipated tend to be large, in order to accommodate the numbers of terms in linear models incorporating interaction (Cornell 2002, 2011). Of course, if one knows from subject matter knowledge or previous experimentation what the final model form will be, one may design an optimal or near optimal experiment to estimate that specific model. We consider the more common situation where the experimenters suspect interaction, but are not confident in specifying the final model form prior to running the design and obtaining the subsequent data. Our focus is therefore neither on the optimal design for a given model, nor on the optimal model for a given set of data. Rather, we focus on a strategy for attacking such problems, considering alternative designs, alternative models, and factoring in the need for efficiency in addressing problems in practice.

In this sense, we take a statistical engineering approach (Hoerl and Snee 2010, Anderson-Cook and Lu 2012) to identify the most effective and efficient approaches to address mixture/process applications occurring in business, industry, research or government. In particular, we find that serious consideration of models non-linear in the parameters may provide an efficient alternative to traditional models linear in the parameters. While fitting non-linear models is more

challenging than fitting linear models, modern software, such as, R, JMP, Minitab, or SAS, allows users to estimate and interpret non-linear models fairly easily, at least compared with commercial software of 10 to 20 years ago.

Non-linear models have fewer parameters to estimate for the same mixture/process problem. For example, in Cornell's fish patty application (Cornell 2002) there were 56 terms to estimate in the linear model versus 14 terms in the non-linear model. Therefore, should non-linear models provide a viable alternative, much more aggressive fractional design strategies could be taken to significantly reduce the initial experimentation. We acknowledge that the theory underlying non-linear models, for such things as standard errors of coefficients, is less developed than for linear models, and often relies on asymptotic or approximate results (Montgomery 2012).

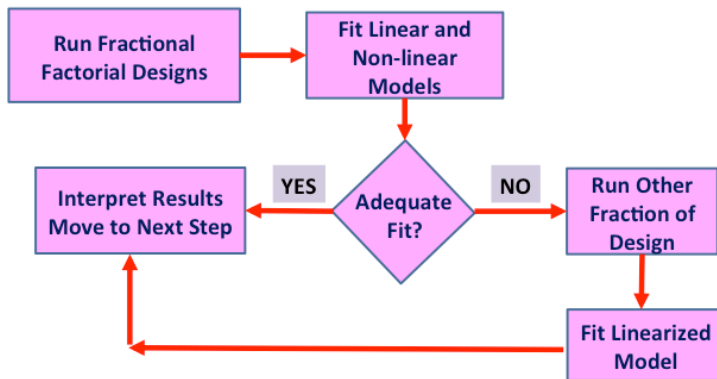
In other words, since non-linear models require fewer parameters to accommodate interaction between process and mixture variables, much smaller fractional designs can be run initially. Obviously, one does not need 56 runs to estimate 14 terms, assuming the cost of experimentation is non-negligible. If the non-linear model, or even an additive linear model without interaction, proves sufficient, then the problem may be solved with the initial design.

We have much to gain and little, if anything, to lose by considering non-linear models. If these smaller models do not provide sufficient accuracy, then additional experimentation can be performed in a sequential manner. In this way, the best-case scenario is that we reduce the experimental effort required to solve the problem by up to 50%, while the worst-case scenario is that we eventually have to run all the experimentation that a typical linear model with interaction would require. This overall strategy is illustrated in Figure 1. Emphasizing the best overall approach to attack the problem, rather than focusing on the best individual design for a given model, or the best individual model for a given data set, is at the heart of statistical engineering, in our view.

The remainder of the article is organized as follows; we first review the typical models and designs used in mixture experiments with process variables, including those with interaction. Next, we review the fundamentals of statistical engineering, and discuss how they apply to this

specific problem. We then introduce three data sets, two previously published, and one new data set. Using these data sets, we compare linear additive, linear interactive, and non-linear models, in terms of fitting the data. Further, we note potential interpretations of the coefficients of the non-linear model. Next, we consider the hypothetical results we might have found with fractional designs for each of these problems. That is, we analyze a subset of the actual data according to a fractional design, fit the different models, and use the models to predict the data not used by the model. Finally, we consider the implications of this analysis, and suggest a strategy for addressing mixture/process problems where interaction is suspected.

FIGURE 1
Potential Statistical Engineering Strategy for Mixture-
Process Problems



Typical Models Used in Mixture/Process Problems

There are many potential designs and models that are applied to mixture problems with process variables. Here we highlight the most common models and designs used in applications, particularly Scheffé models (Scheffé 1963, Cornell 2002), without attempting to provide a comprehensive review. We first consider mixture models themselves, then mixture models with process variables without interaction, then with interaction. Once we finish with models, we discuss the designs typically run to enable fitting of these models.

Because of the constraint that the formulation components sum to 1.0, resulting in the loss of one degree of freedom, it is common to simply drop the constant term from a standard linear

regression model. For example, with q mixture components the linear Scheffé model would be written:

$$E(y) = \sum_{i=1}^q b_i x_i$$

Because the components sum to 1.0, it is not the absolute value of the coefficients that is most meaningful, but rather the relationships between the coefficients. For example, if $b_1=10$, $b_2=20$, and $b_3=30$, with three components, then we would say that x_1 has a negative impact on the response, because to increase x_1 we would have to decrease either x_2 or x_3 by an equal amount, hence the net impact of increasing x_1 would be to decrease $E(y)$. Further information on interpretation of mixture coefficients can be found in Cornell (2002).

Note that for all models presented in this paper we assume an additive random error term, independent and identically distributed normal with mean 0 and standard deviation σ . While error terms are obviously important in any statistical model, this is not a key point of the current article, hence we will say little about this error term moving forward.

There are other mixture models that could be considered, such as “slack variable” models, in which we drop one of the mixture ingredients to avoid the mixture constraint, and add the constant term back into the model. When curvature is present, it is common to fit a quadratic Scheffé model, which incorporates linear terms plus all possible cross product terms between components:

$$E(y) = \sum_{i=1}^q b_i x_i + \sum \sum_{1 \leq i < j}^q b_{ij} x_i x_j$$

Because each component can be viewed as one minus the sum of all other components, we would not typically consider the cross product terms as modeling interaction in the traditional sense of the word, but rather modeling non-linear blending in general. When additional non-linear blending is present, and sufficient degrees of freedom are available, we may fit the special cubic Scheffé model:

$$E(y) = \sum_{i=1}^q b_i x_i + \sum \sum_{1 \leq i < j}^q b_{ij} x_i x_j + \sum \sum \sum_{1 \leq i < j < k}^q b_{ijk} x_i x_j x_k$$

This is considered a “special” cubic in the sense that it does not include all third-order terms, but is rather the quadratic model plus all possible cross products between three terms. For example, the special cubic does not include the term $x_1^2 x_2$. A full cubic Scheffé model would be:

$$E(y) = \sum_{i=1}^q b_i x_i + \sum \sum_{1 \leq i < j}^q b_{ij} x_i x_j + \sum \sum_{1 \leq i < j}^q d_{ij} x_i x_j (x_i - x_j) + \sum \sum \sum_{1 \leq i < j < k}^q b_{ijk} x_i x_j x_k,$$

where the d_{ij} are cubic coefficients of the binary excess or “synergism” (Cornell 2011 p. 31).

When there are process variables in our model, we will use x_i for a mixture variable, z_i for a process variable, b_i for a mixture coefficient, and a_i for a process variable coefficient. We do not consider d_{ij} terms. Further, the generic term $f(x)$ will be used for a model with only mixture ingredients, while $g(z)$ will be used for models with only process variables. Combined models that incorporate both will be written $c(x,z)$.

Again, there are many potential models that can be applied to process variables, and to combinations of mixture ingredients and process variables. Here we focus on standard designs most commonly applied in applications. A linear process variable model would be:

$$E(y) = a_0 + \sum_{i=1}^r a_i z_i,$$

where we have r independent variables. A second order (full quadratic) process variable model would be:

$$E(y) = a_0 + \sum_{i=1}^r a_i z_i + \sum \sum_{1 \leq i \leq j}^r a_{ij} z_i z_j$$

Note that the double summation includes i=j, hence the full quadratic model includes squared (quadratic) terms in the model. Obviously, fitting such a model requires more than two levels in the design. With two-level designs, a factorial model - linear terms and second order interactions, but no quadratic (squared) terms - may be used:

$$E(y) = a_0 + \sum_{i=1}^r a_i z_i + \sum \sum_{1 \leq i < j}^r a_{ij} z_i z_j$$

With full factorial designs, of course, we could estimate higher order interactions as well.

In situations with both mixture and process variables, a key consideration is the degree to which interaction between the mixture and process variables is anticipated. If we have theory or experience that suggests that the mixture and process variables do not interact with one another, then we can fit the mixture and process models separately. In other words, the overall model would become:

$$c(x,z) = f(x) + g(z) \tag{1}$$

We refer to Equation 1 as the linear additive model. In this case, $f(x)$ would be whatever model was found appropriate for the mixture components, and $g(z)$ would be whatever model was found appropriate for the process variables. In many situations, however, the mixture components and process variables interact. In other words, the best time and temperature to bake a cake may depend on the ingredients in the cake. A red wine is not typically served as chilled as a white wine, and so on. In these cases, many authors, such as Cornell (2002, 2011) and Prescott (2004), suggest multiplying the mixture and process variable models, as opposed to adding them. In other words, to accommodate mixture-process interaction we may use the model:

$$c(x,z) = f(x)*g(z) \tag{2}$$

This multiplicative model is not linear in the parameters (b_i and a_i), hence the least squares solution cannot be obtained in closed form, or by standard linear regression software. One approach to address this problem is to multiply out the individual terms in $f(x)*g(z)$. Consider the case where there are three mixture variables and two process variables. For simplicity, assume that we know these models will be of the form:

$$f(x) = b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3, \text{ and}$$

$$g(z) = a_0 + a_1z_1 + a_2z_2 + a_{12}z_1z_2,$$

which would be a special cubic mixture model and a factorial process variable model. Therefore, $f(x)$ has seven terms to estimate, and $g(z)$ has four. If we multiply each term in $f(x)$ by each term in $g(z)$, this would produce 28 terms in a linearized regression model. We refer to this model as linearized because Equation 2 is, in fact, a non-linear model. By multiplying out all the terms, we create a linear model in 28 terms. That is, we have:

$$\begin{aligned} c(x,z) = f(x)*g(z) = & b_1x_1*a_0 + b_1x_1*a_1z_1 + b_1x_1*a_2z_2 + b_1x_1*a_{12}z_1z_2 \\ & + b_2x_2*a_0 + b_2x_2*a_1z_1 + b_2x_2*a_2z_2 + b_2x_2*a_{12}z_1z_2 \\ & ++ b_{123}x_1x_2x_3*a_{12}z_1z_2 \end{aligned} \tag{3}$$

Note that Equation 3 is linear in the parameters if we consider b_1a_0 to be one parameter, b_1a_1 to be another, and so on. Equation 3 can therefore be estimated with standard linear regression software, as can Equation 1. Of course, when the least squares estimates are obtained, the estimated parameter for x_1 , listed as “ b_1a_0 ” in Equation 3, will with probability zero equal the estimated coefficient for x_1 in $f(x)$ from Equation 2 (\hat{b}_1) times the estimated intercept in $g(z)$ from Equation 2 (\hat{a}_0), if these individual terms were estimated directly from non-linear least squares. That is, the least squares estimates from Equation 3 cannot be used to determine the non-linear least squares estimates from Equation 2.

An important point to which we return later is the comparison of Equation 2 with Equation 3. Equation 2 is a model non-linear in the parameters. If it is estimated as such, only 11 parameters need to be estimated, seven for $f(x)$ and four for $g(z)$. In practice, only 10 parameters will be required, since Equation 2 is actually over-specified; there can be no unique least squares solution to this equation. This can be immediately seen if one considers that by multiplying $f(z)$ by an arbitrary non-zero constant, say m , and dividing $g(z)$ by m , the value of $c(x,z)$ would remain unchanged, because the m and $1/m$ would cancel. Since there are an infinite number of such constants m that could be multiplied by each coefficient in $f(x)$, each producing the same residual sum of squares, there can be no unique least squares solution to Equation 2. We address this over-specification by setting $a_0 = 1.0$. We make no claims that this is the best or only way to address the over-specification, although it does aid interpretation of the other coefficients, as we shall see shortly.

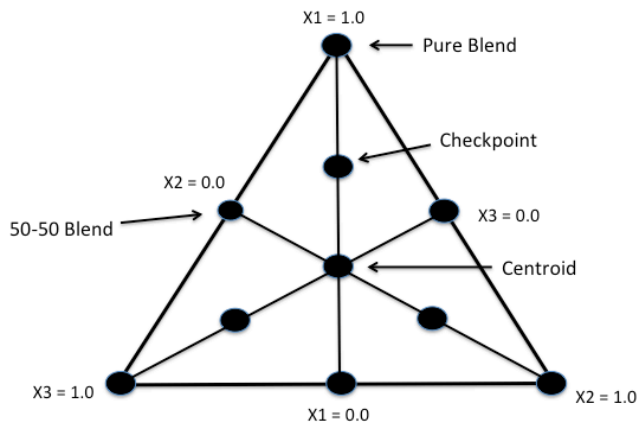
The key point, of course, is that if one can estimate Equation 2 directly as a non-linear least squares problem, then only 10 degrees of freedom for estimated parameters would be required, not the 28 that are required for the linearized version. This has obvious implications for experimental design when interaction is anticipated. Of course, if no interaction is anticipated, then Equation 1, the linear additive model, can be estimated. This equation is also over-specified, in that the a_0 term is not needed, and in fact cannot be estimated, since the intercept is already incorporated into the mixture model $f(x)$. By dropping the intercept, we again have 10 parameters to estimate, as we do in the non-linear model.

Typical Experimental Designs Used in Mixture/Process Problems

Designs and models obviously go hand in hand, in that we typically have specific models in mind when we design an experiment, even if we are not confident in the exact form. In mixture experimentation, there are models suitable for estimating linear Scheffé models, and of course, larger designs suitable for estimating quadratic or special cubic models.

Simplex designs are commonly utilized in mixture experimentation. These are typically represented in $q-1$ dimensions when we have q components. For example, if we have three components, then because of the mixture constraint, there are only two degrees of freedom, hence the design can be displayed in two dimensions, as in Figure 2. Note that all three components are represented, in that each has an axis. In an unconstrained mixture problem, each component can be varied from a 0 to 1.0, hence the full simplex goes from 0 to 1 on each axis. The 0 point is represented by the base of the simplex, and then 1.0 would be at the simplex point opposite the base. Note that if $x_1 = 1.0$, then $x_2 = x_3 = 0$, hence this point corresponds to the base of the x_2 and x_3 axes. These points are referred to as “pure blends” for obvious reasons. The pure blends, and other common points run with simplex designs, are identified in Figure 2.

FIGURE 2
 Simplex Design With Pure Blends, Centroid, 50-50 Blends and Checkpoints



We point out that the centroid, corresponding to the center point in factorial designs, corresponds to the point $1/3, 1/3, 1/3$, which is in the middle of the simplex. Other common points run with simplex designs are so-called 50-50 blends, which have proportions of .5 for two components and 0 for the third. These 50-50 blends correspond to midpoints along each base of the simplex. Another common point often used in simplex designs is the “checkpoint”, typically placed halfway between the centroid and the pure blend. For a three component simplex, these would be at $2/3, 1/6, 1/6$, for x_1 , $1/6, 2/3, 1/6$, for x_2 , and $1/6, 1/6, 2/3$ for x_3 . These checkpoints allow verification that the model fits the data at more than just the centroid and boundaries of the simplex. In other words, fitting the checkpoints well would suggest that it is relatively safe to interpolate within the simplex.

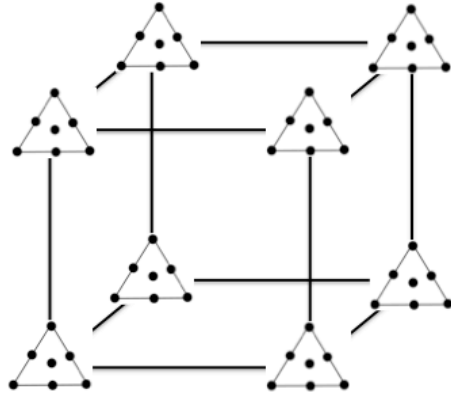
As explained in Cornell (2002), more elaborate simplex-lattice designs that position points uniformly over the entire simplex are also used in practice. In experiments beyond four components, these designs are difficult to illustrate, but the same types of designs can be formed, as is the case with factorial designs. Often, of course, there will be constraints on the levels of the

formulation ingredients. For example, few consumers would like a salad dressing composed of 100% vinegar. In such cases, one may create pseudocomponents (Cornell 2002), which are transformations of the original components, or utilize computer-aided designs in the abnormal design spaces created by the constraints. Such designs are not the focus of the current article. We also assume that the reader is familiar with standard factorial designs used for the process variables.

In mixture designs with process variables, some hybrid design is typically used, so that mixture effects and models can be estimated, as well as process effects and models. If the experimenters suspect interaction, then the ability to estimate all terms in the linearized version of Equation 2 is often desired. That is, the experimental design chosen should allow estimation of all effects that are obtained by multiplying each term in the mixture model with each term in the process model. In the example noted above, one would want a design that allows estimation of all 28 terms in Equation 3.

Such designs typically cross a factorial, or fractional factorial, with a mixture design. For example, Cornell's fish patty experiment (Cornell 2002) crossed a three component simplex design involving pure blends, a centroid, and 50-50 blends, with a 2^3 factorial design in three process variables. This design is shown in Figure 3. Note that at every point of the 2^3 factorial design, the full simplex design was run. Alternatively, of course, one could portray this as a full factorial design at each point of the simplex. Because the simplex design had 7 points and the factorial design 8 points, the full design was 56 points, run in random order to avoid split plotting (Cornell 2002, 2011).

FIGURE 3
Cornell's Fish Patty Design



If interaction is not anticipated, then of course the designs could be much smaller. For example, in the Cornell fish patty example, a special cubic mixture model would have 7 terms (3 linear blending terms, three 2-factor cross products, and the special cubic 3-factor cross product), and a full factorial model would have 8 terms (1 intercept, 3 main effects, 3 2-factor interactions, and 1 3-factor interaction). Because the mixture model $f(x)$ incorporates an intercept, we could drop the intercept from $g(z)$, leaving $7 + 8 - 1 = 14$ terms to estimate in a linear additive model with no interaction, i.e., $c(x,z) = f(x) + g(z)$. Clearly a much smaller fractional design could be run, or since there is no interaction in the model, the mixture and process variables could be investigated independently.

The Role of Statistical Engineering in Mixture Models With Process Variables

Hoerl and Snee (2010) defined statistical engineering as: “The study of how to best utilize statistical concepts, methods, and tools, and integrate them with information technology and other sciences, to generate improved results.” Conversely, statistical science can be viewed as the

study and advancement of the underlying theory and application of individual statistical methods. The key distinction is that statistical science focuses on advancing the fundamental knowledge of tools themselves, as well as their application. Statistical engineering, on the other hand, focuses on how to take existing methods and tools, that is, existing statistical science, and creatively link and integrate these tools, often with non-statistical tools, in order to produce greater impact. Creative utilization of existing science to achieve greater impact is the essence of any field of engineering.

Obviously, both statistical science and statistical engineering are needed for the field of statistics to grow, and enhance its impact on society. We argue that while a balanced approach is needed, the current focus of the profession tends to be biased towards statistical science (Snee and Hoerl 2011). In particular, we feel that greater emphasis is needed on developing an underlying theory of how to attack large, complex, unstructured problems – problems too large for any one tool to solve. Such problems obviously require the creative integration of multiple statistical methods, as well as creative use of information technology or other sciences. Clearly, an engineering approach is called for. See the Special Edition of Quality Engineering on Statistical Engineering (Anderson-Cook and Lu 2012) for further background on the discipline of statistical engineering.

We have noted that when attacking mixture problems with process variables, the choice of design and choice of model go hand in hand. That is, the models we can estimate are determined by the design, and the appropriate design depends on the models we wish to estimate. Further, if we knew the specific model in advance, we could construct an optimal design to estimate that model. However, in many practical situations our subject matter knowledge is limited, and we

are not certain of the final model. Typically, this results in an iterative approach to design and analysis, as emphasized by Box et al. (2005), Montgomery (2012), and many others.

The key question in such situations is how to best approach the problem when we anticipate interaction between the mixture and process variables, but are not sure. We note that this is a very different question than asking what the optimal design is for a given model, and also different from asking what the optimal model is to fit a given set of data. The latter are more statistical science questions, because they deal with the theory and practice of individual tools. In addition, one can determine the “correct” answer analytically, based on objective criteria. The question of how to attack a complex problem is not a statistical science question, however, because we have few givens, and no agreed-upon objective metric for success. In other words, the problem is unstructured.

We argue that this strategy question is better suited for statistical engineering because we require an overall approach incorporating both design and analysis, perhaps in a sequential manner, in order to obtain better results. Integration of multiple tools is key. In this case, “better results” may be based on statistical criteria, or based on saving experimental effort, including time and money, or both. In other words, it may be based on finding the best business solution, as opposed to finding the best statistical solution. This is a more complex and less structured problem than obtaining the best design or best model under specified conditions. The statistical engineering viewpoint therefore guides our study and recommended strategy for mixture problems with process variables.

Three Case Studies in Design and Analysis

Cornell (2002) presented data from an experiment on three species of fish patties: mullet, sheepshead, and croaker. In addition, each patty was cooked at different baking temperatures, baking times, and frying times (the patties were both baked and fried). The response of interest was the texture of the resulting patties. The design used was a seven-run simplex in the mixture variables, including pure blends, 50-50 blends, and the centroid. For the process variables, an eight-run 2^3 was run. Cornell crossed the mixture design with the process design, as listed in Table 1 and illustrated in Figure 3. This design has $7*8 = 56$ runs total. Cornell fit a linearized model, where:

$$f(x) = b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3,$$

$$g(z) = a_0 + a_1z_1 + a_2z_2 + a_3z_3 + a_{12}z_1z_2 + a_{13}z_1z_3 + a_{23}z_2z_3 + a_{123}z_1z_2z_3, \text{ and}$$

$$c(x,z) = f(x)*g(z)$$

TABLE 1 Cornell's Fish Patty Texture Data

Blend			Fry Time							
			25				40			
			Oven Time							
			25		40		25		40	
			Oven Temp							
Mullet	Sheepshead	Croaker	325	425	325	425	325	425	325	425
			Texture							
1	0	0	1.84	2.86	3.01	4.13	1.65	2.32	3.04	4.13
0	1	0	0.67	1.10	1.21	1.67	0.58	0.97	1.16	1.30
0	0	1	1.51	1.60	2.32	2.57	1.21	2.12	2.00	2.75
0.5	0.5	0	1.29	1.53	1.93	2.26	1.18	1.45	1.85	2.06
0.5	0	0.5	1.42	1.81	2.57	3.15	1.45	1.93	2.39	2.82
0	0.5	0.5	1.16	1.50	1.83	2.22	1.07	1.28	1.60	2.10

1/3	1/3	1/3	1.59	1.68	1.94	2.60	1.41	1.54	2.05	2.32
-----	-----	-----	------	------	------	------	------	------	------	------

Since $f(x)$ has 7 terms and $g(z)$ has 8 terms, when the individual terms are multiplied out to create a linearized model, $c(x,z)$ has $7*8 = 56$ terms. There were obviously no degrees of freedom left to estimate experimental error with 56 runs in the design, producing a root mean squared error (RMSE) of 0, and no ability to estimate of the standard deviation of the error term (ϵ), i.e., the standard error. Further, there was no opportunity to evaluate residuals for evidence of model inadequacy, or to perform lack of fit tests. Cornell therefore used variation between replicate measurements (not true replicates) as an estimate of the standard error. Obviously, this within-run variation does not allow one to estimate between-run sources of variation, so it cannot be considered an estimate of the standard error equivalent to that estimated from true replicates, or even from residual degrees of freedom.

For our purposes of comparing models, we applied Lenth's PSE (pseudo standard error) method (Lenth 1989) to obtain an estimate of the standard error for Cornell's model. Lenth's method first determines the effects that appear to be real, and then drops the remaining variables, using the degrees of freedom gained to estimate the standard error. It is a pseudo standard error in that it is dependent on dropping the appropriate terms from the model.

Prescott (2004) tested bread loaf volumes using three flours, Tjalve, Folke, which are both Norwegian, and Hard Red Spring, which is American. There were also two process variables, the proofing time and the mixing time. There were constraints on the levels of three flours: $.25 \leq x_1 \leq 1$, $0 \leq x_2 \leq .75$, and $0 \leq x_3 \leq .75$, where Tjalve is x_1 , Folke is x_2 , and Hard Red Spring is x_3 . When applied to the simplex mixture space, this results in a smaller full simplex in a reduced

space. By converting to pseudocomponents (Cornell 2002), a standard simplex design can still be utilized. For the purposes of model comparisons, and without loss of generality, we use the pseudocomponent version of Prescott's mixture design. Prescott used a 10-run simplex-lattice design for the mixture variables, and a 3^2 , i.e., a three level factorial in two variables, for the process design. This design has 9 runs, of course.

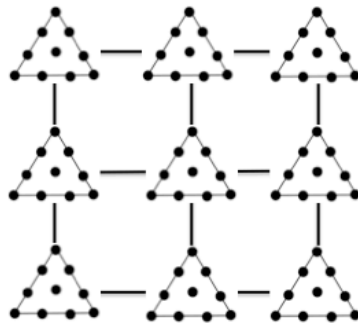
Again crossing the mixture design with the process design, Prescott created a $10 \times 9 = 90$ run design, which is listed in Table 2, and illustrated in Figure 4. For the mixture model $f(x)$, Prescott considered the same special cubic model with 7 terms that Cornell used, and for the process model, $g(z)$, he considered a full quadratic model with 6 terms:

$$g(z) = a_0 + a_1z_1 + a_2z_2 + a_{12}z_1z_2 + a_{11}z_1^2 + a_{22}z_2^2$$

TABLE 2 Prescott's Loaf Volume Data

Pseudocomponent Mixture			Proofing Time (Minutes)								
			35			47.5			60		
			Mixing Time (Minutes)								
Tjalve	Folke	Hard Red Spring	5	15	25	5	15	25	5	15	25
			Loaf Volume								
.25	.75	0	378.89	396.67	392.22	445.56	452.22	487.78	457.22	500.56	472.78
.50	.50	0	388.89	423.33	416.11	460.00	488.89	475.78	472.78	478.00	506.11
.75	.25	0	426.11	483.33	389.44	474.44	514.44	462.78	506.67	591.67	522.22
1.0	0	0	386.11	459.11	423.33	458.33	506.11	514.44	545.56	522.22	551.11
.25	.50	.25	417.78	437.22	444.56	484.44	490.00	495.00	497.78	531.11	577.78
.50	.25	.25	389.44	447.22	415.00	490.89	528.89	507.78	517.78	567.22	538.33
.75	0	.25	448.33	459.44	455.56	436.00	535.00	552.22	507.44	578.89	590.00
.25	.25	.50	413.89	485.56	462.22	483.89	529.44	540.00	565.00	598.89	580.56
.50	0	.50	415.56	514.44	437.78	493.89	583.33	578.89	524.44	649.44	640.00
.25	0	.75	432.78	498.33	517.22	474.44	568.33	579.44	541.11	638.89	638.89

FIGURE 4 Prescott's Bread Design



Note that the quadratic terms are estimable because the process design was a three-level design. Prescott ran several linearized interactive models, i.e., of the form $c(x,z) = f(x)*g(z)$, which had $7*6 = 42$ terms total. With 90 runs there were sufficient degrees of freedom to estimate the full 42-term model, and still have degrees of freedom left to estimate the standard error. After considering several models, the final model suggested by Prescott for this data was a reduced model of the form:

$$y = 522.8x_1 + 448.1x_2 + 599.3x_3 + 13.0x_1z_1 + 1.7x_2z_1 + 54.3x_3z_1 + 56.3x_1z_2 + 37.2x_2z_2 + 73.8x_3z_2 - 39.4x_1z_1^2 + 3.7x_2z_1^2 - 46.0x_3z_1^2 - 10.2x_1z_2^2 + 28.4x_2z_2^2 + 1.0x_3z_2^2$$

Note that some terms were dropped from the model, such as terms involving any cross products of the x_i , or the interaction of z_1 and z_2 . In fact, only 15 of the possible 42 terms were included in this final model. We return to this point when comparing the linearized with non-linear models.

Wang, et al. (2013) ran a mixture/process experimental design as part of a class project for a statistics course taught by the first author at Rutgers University in 2013. This experiment includes three mixture variables, water, milk, and orange juice, and two process variables, the addition of sugar (0 or 2 grams) and the type of milk used (2% or whole). The response was the melting time in minutes of ice produced from the mixture. The mixture design was a 10-run simplex design with pure components, 50-50 blends, checkpoints, and a centroid, in pseudocomponent form. This is the design illustrated in Figure 2. For the process design, Wang ran a 2^2 factorial in four runs. The crossed design with $10 \times 4 = 40$ runs and resulting responses are given in Table 3, and illustrated in Figure 5.

FIGURE 5
Wang's Ice Melting Design

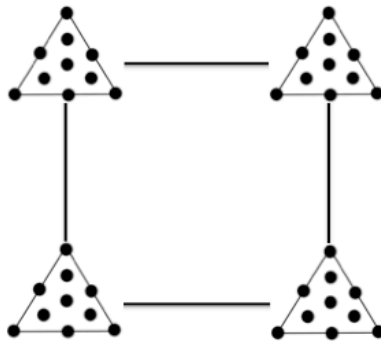


TABLE 3 Wang et al. Melting Time Data

Mixture			Milk Type			
			2%		Whole	
			Sugar			
			0g	2g	0g	2g
Water	Milk	Juice	Melting Time			
1/6	1/6	2/3	29.43	24.31	34.94	32.97
0	0.5	0.5	37.36	32.89	45.69	39.06
0.5	0.5	0	33.32	28.34	41.78	35.84
1	0	0	19.13	17.73	23.7	22.16
2/3	1/6	1/6	24.11	22.49	29.76	25.43
1/3	1/3	1/3	32.08	26.29	36.97	30.13

0	0	1	30.41	23.25	31.92	29.45
0.5	0	0.5	21.86	22.88	27.09	26.21
1/6	2/3	1/6	37.35	28.43	42.92	39.95
0	1	0	48.64	33.00	45.08	47.35

Wang, et al. (2013) fitted a linearized model using a 7-term special cubic for $f(x)$, and a 4-term factorial process model for $g(z)$, which had an intercept, two main effects, and the 2-factor interaction. Thus, when multiplying out all terms, $c(x,z)$ is the 28-term linearized model given in Equation 3. With 40 runs and 28 terms to estimate, there were 12 degrees of freedom left to estimate the standard error.

Linear and Non-Linear Models and Results

For each of these designs and data sets, one may also consider much smaller models, such as the linear additive model, $c(x,z) = f(x) + g(z)$, which of course has no terms to estimate interaction between mixture and process variables, and also the non-linear model. To estimate the non-linear model one finds the least squares solution to $c(x,z) = f(x)*g(z)$ directly, without multiplying out the terms. As noted previously, this model is non-linear in the parameters. For example, in the simplistic case with two mixture variables and one process variable:

$$c(x,z) = (b_1x_1 + b_2x_2 + b_{12}x_1x_2)*(a_0 + a_1z) \quad (4)$$

Assuming we wish to estimate each of these 5 coefficients, then we can consider Equation 4 as a non-linear equation with 5 coefficients, and use non-linear least squares to find a solution. As noted previously, in practice we only have 4 coefficients to estimate because Equation 4 is over determined, and will have no unique least squares solution. By

arbitrarily setting $a_0 = 1.0$, we can generally obtain a unique least squares solution in the other parameters. Using some other non-zero value for a_0 would not change model metrics, such as R^2 , standard error, and so on. Of course, with non-linear least squares, no unique solution is guaranteed.

To fit the non-linear model we use the non-linear regression algorithm in JMP Pro 10, run on a standard MacBook Pro laptop. As noted previously, running non-linear models is more challenging than linear models, but with modern software it is certainly feasible, and we suggest that it is a viable option for these types of problems. Before discussing the results obtained fitting linear additive and non-linear models to these three data sets, we briefly discuss interpretation of the coefficients in the non-linear model. We do not discuss the linear additive model, as interpretation of mixture terms has been discussed extensively in the literature, such as Scheffé (1963), Cornell (2002), and many others.

After estimating the terms in the non-linear model in Equation 4, we are free to multiply out the estimated terms to aid interpretation. Note that this produces a very different model from that obtained by multiplying out the terms prior to estimating the coefficients, which would produce the linearized model. Also, the decision to set $a_0 = 1$ rather than some other value makes the interpretation of the model more straightforward. Once we multiply out the estimated terms in the non-linear approach to Equation 4 we obtain:

$$\begin{aligned}\hat{y} &= (\hat{b}_{1X_1} + \hat{b}_{2X_2} + \hat{b}_{12X_1X_2}) * (1 + \hat{a}_{1Z}) \\ &= (\hat{b}_{X_1} + \hat{b}_{X_2} + \hat{b}_{12X_1X_2}) + (\hat{b}_{X_1} + \hat{b}_{X_2} + \hat{b}_{12X_1X_2}) * \hat{a}_{1Z}\end{aligned}\tag{5}$$

Without loss of generality we may assume that z is scaled to have a mean of 0. Therefore, when z is at its mean value, Equation 5 simplifies to just the left part, which is the pure mixture model, $f(x)$. Therefore, we can interpret the mixture coefficients as the blending we estimate at the midpoint of the process variable, or alternatively, as the average blending across the entire design. This interpretation holds whether or not there is interaction with z . The interpretation of \hat{a}_1 is more complicated. Note that the only place z appears in Equation 5 is through a set of interactions.

This implies inability of the nonlinear model to estimate the main effect of z uniquely from the interactions of z with the mixture variables. This is because the only coefficient available to incorporate the main effect of z and also its interactions with the x_i is \hat{a}_1 . While $f(x)$ could be rewritten as a slack variable model with an intercept, we would still only have one degree of freedom to incorporate both the main effect of z and its interactions with the mixture variables. We view this as a drawback of the non-linear approach, and speculate that the practical result will be to have some difficulty fitting surfaces with both large process variable main effects and severe interaction between mixture and process variables.

We can extend this interpretation to larger non-linear models, such as the 7-term special cubic mixture model and 8-term process variable model from his fish patty experiment. Should we estimate $c(x,z) = f(x)*g(z)$ directly as a non-linear model, we would estimate $7 + 8 - 1 = 14$ coefficients. Recall that we set $a_0 = 1$. After solving for these 14 coefficients using

non-linear least squares, we can then multiply out terms to aid interpretation. Doing so produces:

$$\hat{f}(x) * (1 + \hat{a}_1 z_1 + \hat{a}_2 z_2 + \hat{a}_3 z_3 + \hat{a}_{12} z_1 z_2 + \hat{a}_{13} z_1 z_3 + \hat{a}_{23} z_2 z_3 + \hat{a}_{123} z_1 z_2 z_3),$$

which is 1 times $\hat{f}(x)$, plus $\hat{a}_1 z_1$ times $\hat{f}(x)$, plus $\hat{a}_2 z_2$ times $\hat{f}(x)$, and so on. In other words, all other terms in the model besides $\hat{f}(x) * 1$ involve \hat{a}_i times one or more z_i , times $\hat{f}(x)$. Therefore, none of the main effects of the z_i can be estimated uniquely from the interaction of that z_i with the mixture variables (and their interactions), analogous to the situation in Equation 5. Similarly, interactions between the z_i , such as $z_1 z_2$, for example, cannot be estimated uniquely from the interactions between this term and the mixture terms.

From a practical point of view, inability to estimate individual coefficients, for z_1 and $x_1 z_1$, for example, is of greater concern than inability to estimate individual coefficients for the interactions such as $z_1 z_2 z_3$ and $x_1 x_2 x_3 z_1 z_2 z_3$, in the sense of resolution of the design (Box et al. 2005). That is, practitioners would rarely be concerned about the practical importance of a six-factor interaction.

Fitting both the linear additive model, $c(x,z) = f(x) + g(z)$, and the nonlinear model, $c(x,z) = f(x) * g(z)$, to the three data sets discussed previously, we obtain the results shown in Table 4. For the Cornell and Wang data sets, we fit complete models with no attempt at subset selection. With the Prescott data, for both the linear additive and non-linear models, we incorporate only the terms in $f(x)$ and $g(z)$ that were selected by Prescott to be in his final

linearized model. Specifically, for $f(x)$ we use only x_1 , x_2 , and x_3 . For $g(z)$ we include z_1 , z_2 , z_1^2 , and z_2^2 . In this sense, the Cornell and Wang comparisons are “apples to apples” with no subset selection, and the Prescott comparison is “apples to apples” with the subset selection applied by Prescott.

The numerical criterion we are using is estimated standard error, or root mean squared error (RMSE). Obviously, there is no single metric that can determine the “goodness” of a model. To develop appropriate models in practice one needs to look at several numerical criteria, evaluate residual plots, compare the results to solid subject matter theory, and so on. The reader should not infer that we recommend focusing solely on RMSE in model building. For the purposes of model comparison, however, we use estimated standard error simply as a convenient objective criterion to measure the degree to which the model is able to fit the data.

From Table 4 we see that in two cases the linearized model provides the smallest estimated standard error, and is nearly smallest for the Cornell data. This is not surprising, in that the linear additive model does not accommodate any interaction, and the non-linear model has the limitations noted previously. We conclude from these three analyses that if sufficient degrees of freedom are available, the linearized model is the preferred approach, based on these three models applying this criterion.

TABLE 4 RMSE of Linearized, Linear Additive, and Non-Linear Models

	<u>Cornell Data</u>	<u>Prescott Data</u>	<u>Wang Data</u>
Linearized	0.17*	21.0	1.36

Linear Additive	0.24	24.3	2.38
Non-Linear	0.16	23.1	2.05

*Calculated using Lenth's method because there are no df for error.

Upon closer examination, however, we notice that for two of the data sets, the Cornell data and the Prescott data, the differences between the models are not as large as one might expect, and in fact, the non-linear model RMSE is slightly smaller than the linearized model RMSE for Cornell's data. Recall that because of no degrees of freedom left to estimate error, the linearized model RMSE is calculated using Lenth's method. The performance of the non-linear approach is impressive given the magnitude of the differences in number of estimated coefficients in these models: 14 (linear additive or non-linear) versus 56 (linearized) for Cornell's model, 12 versus 42 for Prescott's full model (7 versus 15 if we use only Prescott's reduced model terms), and 10 versus 28 for Wang's model. The non-linear and linear additive models are much more parsimonious.

One possible explanation for why the linear additive and non-linear approaches do not work well for the Wang data is that there is highly significant interaction between the mixture and process variables with this data. For example, the terms in the linearized model for milk*sugar*milk type ($x_2z_1z_2$), and milk*juice*sugar*milk type ($x_2x_3z_1z_2$), are both statistically significant at the .01 level. The term for water*milk*sugar*milk type ($x_1x_2z_1z_2$), is nearly significant at the .01 level (p value = .0101). In our experience, it is very unusual to find four factor interactions so highly significant. The linear model, not including any terms for interaction between process and mixture variables, will obviously be unable to effectively model such interaction. The non-linear model simply does not have enough

terms to provide a good approximation. With such significant higher order interaction, it is likely that the linearized approach will be required.

Given the relatively low differences seen in Table 4, it seems plausible that we may be able to save significant time and money by reducing the amount of initial experimentation. Certainly, 90 runs are not needed to estimate 12 terms in the Prescott non-linear model using all terms, nor 56 runs to estimate the 14 terms in the Cornell non-linear model. In order to answer this question, we evaluate how each of these models would fit a fraction of the designs actually run.

Results With Fractional Designs

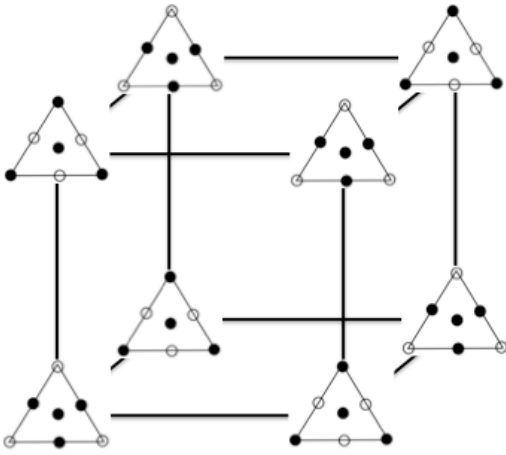
Each of the designs we have discussed was produced by crossing the mixture design with the process design. Therefore, the total number of runs was $n_1 * n_2$, where n_1 was the number in the mixture design, and n_2 the number in the process design. For Cornell's fish patty design, this was $7 * 8 = 56$. Given the results shown above for non-linear models, a logical question to ask is whether all 56 runs were actually needed, or if Cornell might have been able to run a much smaller design and obtain nearly the same results? One of the justifications given by Cornell (2002) for running the 56-run design was the need to estimate all the coefficients in the linearized model.

Cornell (2002) also noted that if experimenters were unable to run the full design, they could create a fractional design by running a fraction of the process design at each point of the mixture space. In order to ensure that as many terms as possible would be estimable,

Cornell suggested switching which fraction of the process design was run at different points of the mixture design. For example, in the case of three process variables, we might run the $z_1z_2z_3 = 1$ half fraction at roughly half of the mixture design points, and $z_1z_2z_3 = -1$ at the other half. Cornell further recommended running the full process design at the centroid of the mixture space, if possible. He noted that if this fractional design were run, the experimenter would not be able to estimate the full linearized model, but would be restricted to estimating a subset of these terms.

In the fish patty case, the mixture design had 7 points, and the process design 8. Running all 8 points at the centroid, and then 4 (a half fraction) at all other points would result in a fractional design with 32 points. This is not enough points to estimate all terms in the linearized fish patty model, but is more than enough to estimate the 14 terms in the non-linear or linear additive model. In Figure 6 we show the fractional design with $z_1z_2z_3 = 1$ run at the pure blends, and $z_1z_2z_3 = -1$ run at the 50-50 blends. To ease graphical interpretation, and to enable comparisons with Figure 3, we again show the mixture points run at each point of the process design, although the method of construction was as noted above.

FIGURE 6
Fractional Cornell Design 1



In addition to seeing how well the non-linear model fits the reduced data set, we also have a natural hold-out sample to better evaluate the model. That is, only 32 runs are used to estimate the non-linear model. Therefore, once we have estimated the model terms, we can use this model to predict the 24 runs that were not used to estimate this model. Such a hold-out sample is generally considered a more valid criterion to evaluate an estimated model, since the model has not seen this data, and cannot be influenced by it during the fitting process. Table 5 shows the results of first fitting both the linear additive and non-linear models to the fractional design shown in Figure 6, and also the prediction standard error (RMSE) when predicting the hold-out data. Note that Cornell's model had too many terms to estimate with a fractional design, so the linearized model is not an option.

Obviously, one could arbitrarily reverse the fractional design run at the pure blends versus the 50-50 blends. Therefore, for completeness we consider the design shown in Figure 6 as fractional design 1, and define fractional design 2 to be that in which $z_1z_2z_3 = -1$ is run at

pure blends, and $z_1z_2z_3 = 1$ is run at 50-50 blends. Again, the full 8 process design points are run at the centroid.

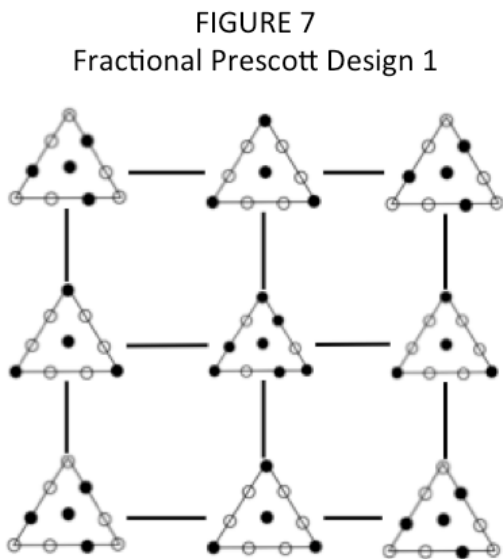
TABLE 5 Prediction Results With Cornell Fractional Designs

	Full Data: Linearized	Full Data: Linear	Full Data: Non-Linear	Fractional Design 1: Linear	Fractional Design 2: Linear	Fractional Design 1: Non- Linear	Fractional Design 2: Non-Linear
RMSE	0.17*	0.24	0.16	0.27	0.24	0.16	0.16
RMSE of Hold- Out Data	NA	NA	NA	0.25	0.28	0.22	0.22
Design Size	56	56	56	32	32	32	32
Excluded Points	0	0	0	24	24	24	24

*Calculated using Lenth’s method because there are no df for error.

Note in Table 5 that for both fractional designs the non-linear models do a good job of fitting the “training” data, and fit the hold-out data reasonably well. The prediction RMSE is 0.22 for both fractional designs, which is higher than the original 0.16 RMSE, but within reason, considering we only used 32 of the original 56 runs, which is a 43% reduction. The linear additive model also performs consistently with its fit of the entire data set. As noted previously, if our sole objective is to maximize model fit, then a full crossed design with the linearized model is preferred. However, if time and cost are important considerations, then a non-linear model with lower prediction accuracy, but obtained with much less experimental effort, should be considered.

Prescott's mixture design was a simplex-lattice with two edge points between each pure blend, and a centroid. This was crossed with a 3^2 process design. Such a combination presents more challenge in determining a fractional design that would be balanced in a heuristic sense. After considering various options, we define fractional design 1 to be the full 3^2 at the centroid, the points where $z_1z_2 = 0$ at the pure blends, and the points where $z_1 = z_2 = 0$ and also where $z_1z_2 = +/-1$ at the following three edge points: $(2/3, 1/3, 0)$, $(0, 2/3, 1/3)$, and $(1/3, 0, 2/3)$. This results in a 39 run fractional design $(9 + 5*3 + 5*3)$, illustrated in Figure 7. Note that the fractional design has less than half as many points as the original 90 run design. Fractional design 2 was obtained by again running the full 3^2 at the centroid, but reversing the process points run at the pure blends versus the three edge points.



The prediction results for the Prescott fractional designs are shown in Table 6. Note that for the Prescott linearized model, only the 15 terms kept by Prescott were used in the final model, hence we can estimate this model with the fractional design. This allows us to

compare the linearized model with the linear additive and non-linear models with the fractional data. Of course, if Prescott had run a full special cubic mixture model in 7 terms with a factorial process design in 6 terms, we would not have the 42 degrees of freedom necessary to estimate all terms.

We can see from Table 6 that the linearized model does extremely well with the fractional designs using the training data. In fact, the RMSE for both designs are lower than the RMSE obtained with the full data set. This can be partially explained by the fact that Prescott determined the parameters needed in the linearized model from the full data set. In this sense, we are using information from the full data set when we fit a linearized model to the fractional data. However, we cannot estimate all terms in the full linearized model with these 39 data points. Note, however, that the linearized model does much worse in predicting the hold-out data. In fact, for both fractional designs the linearized prediction RMSE is larger than those of both the linear and non-linear models.

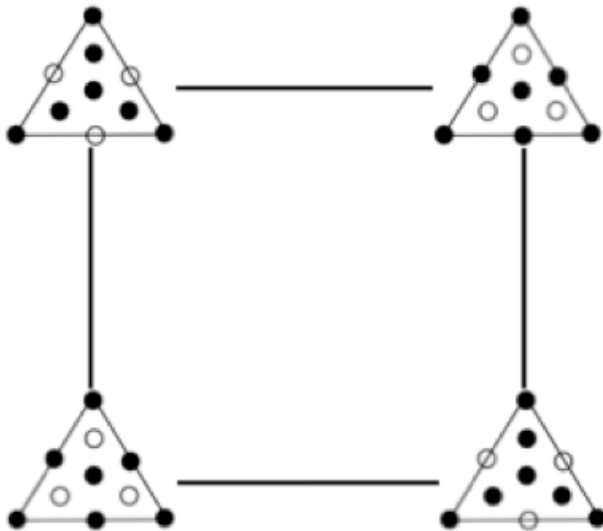
Again, the linear additive and non-linear models provide prediction RMSE that is reasonably close to the RMSE from the training set, for both designs. As was the case with Cornell's data, we see that fitting a linear additive or non-linear model with a fractional design provides a viable alternative to running the full 90-run design.

TABLE 6 Prediction Results With Prescott Fractional Designs

	Full Data: Linearized	Full Data: Linear	Full Data: Non-Linear	Fractional Design 1: Linearized	Fractional Design 2: Linearized	Fractional Design 1: Linear	Fractional Design 2: Linear	Fractional Design 1: Non-Linear	Fractional Design 2: Non-Linear
RMSE	21.0	24.3	23.1	20.3	17.2	24.9	23.9	22.6	23.6
RMSE of Hold- Out Data	NA	NA	NA	27.2	30.3	25.3	26.8	25.9	24.3
Design Size	90	90	90	39	39	39	39	39	39
Excluded Points	0	0	0	51	51	51	51	51	51

Wang's original design was a 10-run simplex with 50-50 edge points and checkpoints, crossed with a 2^2 process design, resulting in $10 \times 4 = 40$ runs. We obtain fractional design 1 by running a full 2^2 at the centroid and pure blends, the half fraction $z_1 z_2 = 1$ at the 50-50 edge points, and the half fraction $z_1 z_2 = -1$ at the checkpoints. This results in $4 \times 4 + 2 \times 3 + 2 \times 3 = 28$ runs, as shown in Figure 8. Fractional design 2 was obtained by reversing the fractions run at the 50-50 edge points and checkpoints.

FIGURE 8
Fractional Wang Design 1



Using these fractional designs, we again compare the linearized, linear additive, and non-linear models in terms of prediction RMSE. Since the linearized model run by Wang has 28

parameters, we are able to estimate this full model, although with no remaining degrees of freedom to estimate error. Therefore, for comparison purposes we again use Lenth’s pseudo standard error in place of RMSE for the linearized model. The results of these comparisons are shown in Table 7.

TABLE 7 Prediction Results With Wang Fractional Designs

	Full Data: Linearized	Full Data: Linear	Full Data: Non-Linear	Fractional Design 1: Linearized	Fractional Design 2: Linearized	Fractional Design 1: Linear	Fractional Design 2: Linear	Fractional Design 1: Non-Linear	Fractional Design 2: Non-Linear
RMSE	1.36	2.38	2.05	0.90*	3.52*	2.47	2.67	2.02	2.23
RMSE of Hold- Out Data	NA	NA	NA	3.46	4.02	2.93	2.30	3.01	2.44
Design Size	40	40	40	28	28	28	28	28	28
Excluded Points	0	0	0	12	12	12	12	12	12

*Calculated using Lenth’s method because there are no df for error.

Perhaps most noteworthy from Table 7 are the poor predictions from the linearized model. In large part, this is likely due to the fact that there are zero degrees of freedom to estimate error with the fractional designs. In our view, a key lesson from this is the danger in using regression models with no or few degrees of freedom for error to predict new data. Model evaluation and validate are difficult tasks; doing so with few or no degrees of freedom for error are likely to result in poor predictions. The linear additive and non-linear models perform better, roughly equivalent to one another in terms of out of sample prediction. Note that neither the linear additive nor non-linear models produced a RMSE close to the linearized model with the full data set. However, when predicting out of sample they both perform better than the linearized model. Again, we argue that this provides some evidence that linear additive and non-linear models, run on fractional designs, are viable alternatives to consider in practice.

What Have We Learned?

We have compared linearized, linear additive, and non-linear models using two published mixture-process data sets and one new data set. Further, we have compared these approaches not only in terms of fitting the original data sets, but also in fitting potential fractional designs that could have been run instead of the full crossed (mixture times process) designs. We repeat that our purpose is not to find the optimal design or best model in an absolute sense. Rather, we propose a statistical engineering approach to determine the best strategy for attacking mixture-process problems in the face of potential interaction. Such a strategy would consider the practical benefits of fractional designs and sequential approaches.

In short, we argue that we have learned the following from these comparisons:

1. If sufficient degrees of freedom are available, linearized models, obtained by multiplying the terms of the mixture model with those of the process model, are likely to fit the data best, and provide the best predictions of out of sample data.
2. Linear additive models, ignoring mixture-process interaction, and non-linear models, offer viable alternatives to the linearized model, and offer significant advantages in number of terms to estimate.
3. When experimenters consider the option of linear additive or non-linear models, a sequential strategy beginning with fractional designs offers the potential to significantly reduce the total amount of experimentation required to develop a useful model.

4. If the linear additive and non-linear models do not fit the original fractional designs well enough, experimenters can then run the remaining full crossed design, and fit the linearized model. In this sense, experimenters have much to gain and very little to lose by considering a sequential strategy.
5. An additional point not to be overlooked is that models with few or no degrees of freedom for residual error may fit sample data well, but are risky to use predicting out of sample data.

The strategy we suggest was illustrated in Figure 1. By running the fractional design initially, we have the potential to significantly reduce the amount of experimentation required to fit an adequate model. The different fractional designs and the number of coefficients in the various models are summarized in Table 8 for comparison.

TABLE 8 Size of Full and Fractional Designs and Number of Coefficients in the Models

Data Set	No. Runs Full Design	No. Runs Fractional Design	Reduction	No. Linearized Model Coefficients	No. Linear Additive Model Coefficients	No. Nonlinear Model Coefficients
Cornell	56	32	43%	56	14	14
Prescott	90	39	57%	42	12	12
Wang et al.	40	28	30%	28	11	11

We suggest running both the linear additive and non-linear models on the fractional data. If the non-linear model fits appreciably better, this would indicate noteworthy interaction between the mixture and process variables. Of course, there is no guarantee that either the

linear additive or non-linear model will fit well. In particular, with high degrees of interaction between mixture and process variables, linearized models will likely be needed.

If experimenters are not satisfied with the models generated from the fractional design, they have the option of running the other fraction – the fraction not run initially, to complete the full design. This option should be part of the strategy adopted at the beginning of the study. This approach would enable estimation of the full linearized model. One degree of freedom would be required to effectively block the design over time, between the first fraction and second fraction. Otherwise, there is no real downside to this approach, assuming it is possible to experiment sequentially. We again acknowledge that the theory underlying non-linear models is not as developed as that for linear models.

The model adequacy step in Figure 1 should not be limited to evaluating RMSE, of course. This decision step is meant to include a complete evaluation of model adequacy, including careful scrutiny of residuals, evaluation of potential outliers, comparison with current domain knowledge about this problem, and so on. Such evaluation of the linearized model is assumed, hence we have not included this as another step in Figure 1. Our goal in this figure is to show how to evaluate the different models in a sequential fashion, as opposed to describing the details of sound statistical model building, as discussed in Box et al. (2005), for example.

The reader may note that in this article we do not investigate the potential of computer-aided designs to select an optimal fraction to fit the non-linear design. We believe that

research into application of computer designs for non-linear models in mixture-process problems could provide even better results than those we present. We leave this to future research.

A Workable Strategy Has Been Developed, Tested and Illustrated

We consider the problem of designing experiments and fitting models with both mixture and process variables, which are considered likely to interact with one another. A common approach to such problems is to run a full mixture design crossed with a full factorial design in the process variables. Such a design allows estimation of all coefficients in a linearized model – a model created by crossing all the terms in the mixture model with all those in the process model. However, both the models and the designs required to estimate them tend to be quite large. For example, in Prescott's bread making example, a 90-run design was used. Cornell's fish patty design required 56 runs.

We consider the case where experimenters would like a good model, perhaps not the optimal one, and would like to obtain such a model with much less experimentation than 90 or 56 runs. In these situations, we argue that a statistical engineering strategy using sequential experimentation, much smaller fractional designs initially, and consideration of linear additive or non-linear models, provides a better overall approach. To repeat, we do not claim that the non-linear model is likely to be superior to the linearized model in terms of RMSE or prediction RMSE. Neither do we claim that fractional designs contain as much information as full crossed designs. In this sense, we are not proposing a specific design per se, nor are we proposing a specific model per se.

Rather, we propose a strategy for attacking such problems that we believe is superior overall to the approaches suggested in current literature. In this sense, our proposed strategy is a statistical engineering approach. Based on the results shown here, we conclude that this approach is superior because it offers the potential to save significant experimental time and effort, while having virtually no down side, assuming a sequential approach is feasible. In particular, non-linear models are fairly straightforward to estimate with modern statistical software, such as R, JMP, Minitab, or SAS, hence software need not be a hindrance to consideration of non-linear models. We further believe that similar approaches may be of more general use in other response surface applications – beyond mixture/process variable models, but this is a topic for future research.

References

Anderson-Cook, C.M., and Lu, L. (Guest Editors) (2012). “Special issue on statistical engineering”, Quality Engineering, 24(2).

Box, G.E.P., Hunter, J.S., and Hunter, W.G. (2005). Statistics for experimenters, 2nd edition, John Wiley and Sons, Hoboken, NJ.

Cornell, J.A. (2002). Experiments with mixtures: Designs, models, and analysis of mixture data, 3rd Edition, John Wiley and Sons, Hoboken, NY.

Cornell, J.A. (2011). A primer on experiments with mixtures, John Wiley and Sons, Hoboken, NJ.

Hoerl, R.W., and Snee, R.D. (2010). “Statistical thinking and methods in quality improvement: A look to the future,” Quality Engineering, 22(3): 119-139.

Lenth, R.V. (1989). “Quick and easy analysis of unreplicated factorials”, Technometrics, 31(4): 469-473.

Montgomery, D.C. (2012). Design and analysis of experiments, 8th edition, John Wiley and Sons, Hoboken, NJ.

Prescott, P. (2004). "Modelling in mixture experiments including interactions with process variable." Quality Technology & Quantitative Management,1(1): 87-103.

Scheffé, H. (1963). "The simplex-centroid design for experiments with mixtures", Journal of the Royal Statistical Society B, 25(2): 235-263.

Snee, R.D., and Hoerl, R.W. (2011). "Proper blending; The right mix between statistical engineering and applied statistics," Quality Progress, June: 46-49.

Wang, L., Xu, J., Wu, S., and Zhang, K. (2013). Study of melting rate of self-made ice cubes, Unpublished project report.