Technical

Mixture DOE uncovers formulations quicker

By Mark J. Anderson and
Patrick J. Whitcomb
Stat-Ease Inc.

Executive summary

This is the second article of a series on design of experiments (DOE). It shows how to uncover "sweet spots" where multiple product specifications can be met in a most desirable way. The first publication provided tools for process breakthroughs via two-level factorial designs.[1] This follow-up article offers a simple case study that illustrates how to put rubber or plastics formulations to the test by using powerful statistical methods for mixture design and analysis.

Design of experiments (DOE) techniques provide an efficient means for you to optimize your process. But you shouldn't restrict your studies only to process factors. Adjustments in the formulation may prove to be beneficial as well. A simple, but effective, strategy of experimentation involves:

1. finding an ideal formulation via mixture design; and
2. optimizing the process with factorial design and response surface methods.

This article shows you how to apply design of experiment methods to your formulation. Two case studies give you templates for action.

TECHNICAL NOTEBOOK
Edited by Harold Herzlich

Why factorial methods don't work well for mixtures

Industrial experimenters typically turn to two-level factorials as their first attempt at DOE. These designs consist of all combinations of each factor at its high and low levels. With large numbers of factors, only a fraction of the runs needs to be completed to produce estimates of main effects and simple interactions. However, when the response depends on proportions of ingredients, such as in chemical or food formulations, factorial designs may not make sense. For example, look at what happens with experiments on lemonade (**Table I**) that vary the number of lemons vs. cups of sugar water. Standard orders 1 (both factors low) and 4 (both factors high) taste the same. It makes more sense to look at taste as a function of the proportion of lemons to water, not the amount. Mixture design accounts for the dependence of response on proportionality of ingredients. If you experiment on formulations where proportions matter, not the amount, factorials won't work. Use a mixture design.

Case study illustrates a simple mixture design

To illustrate how to apply mixture design, we present a relatively simple study that involves three solvents.[2] The experimenters measured the effects of three solvents known to dissolve a particular family of complex organic chemicals. They previously had discovered a new compound in this family. It needed to be dissolved for purification purposes, so they needed to find the optimal blend of solvents.

Table II shows the experimental design in a convenient layout that identifies the blends by type. The actual run order for experiments like this always should be randomized to counteract any time-related effects due to aging of material, etc. Also, we recommend that you always replicate at least four blends to get a measure of error. In this case, it would have been helpful to do two each of the pure materials and also replicate the three-part blend, called the "centroid."

The geometry of the experimental region can be seen in **Fig. 1**. In this triangular layout, the apexes (point 1, 2 and 3) represent pure component blends. Binary blends, which provide estimates of second order effects, occur at the midpoints of the sides on the triangle (points 4, 5 and 6). The centroid (point 7) contains equal amounts of all three ingredients. The individual proportions go from zero to one from base to apex in each of the three axes. The pattern of the points 1 through 7 in the mixture "space" (shown by **Fig. 1**) forms a textbook design called an augmented simplex-centroid.[3] The term "simplex" relates to the geometry—the simplest figure with one more vertex than the number of dimensions. In this case only two dimensions are needed to graph the three components on to an equilateral triangle. A four-component mixture experiment requires another dimension in simplex geometry—a tetrahedron (like a pyramid, but with three sides, not four). Let's keep things really simple by only

See **Mixture,** *page 18*

Table I. Misleading factorial design for lemonade.

Std Order	Lemons	Sugar-Water (cups)	Ratio Lemons/Water	Taste
1	1	1	1.0	Good
2	2	1	2.0	Sour
3	1	2	0.5	Weak
4	2	2	1.0	Good

Table II. Design matrix and data for solvent study.

Blend #	A MEK	B Toluene	C Hexane	Blend Type	Solubility (g/l)
1	100	0	0	Pure A	121
2	0	100	0	Pure B	164
3	0	0	100	Pure C	179
4	50	50	0	Binary AB	140
5	0	50	50	Binary BC	180
6	50	0	50	Binary AC	185
7	33.3	33.3	33.3	Centroid	199
8	66.6	16.7	16.7	Check	175
9	16.7	66.6	16.7	Check	186
10	16.7	16.7	66.6	Check	201

Fig. 1. Simplex-centroid mixture design augmented with check blends.

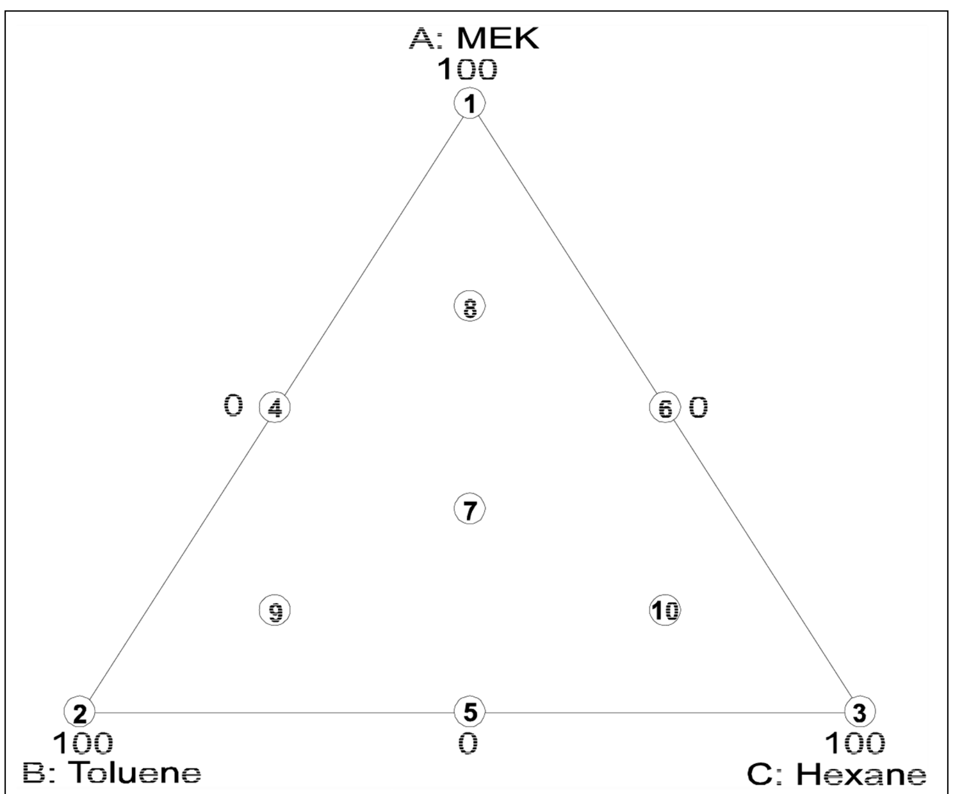
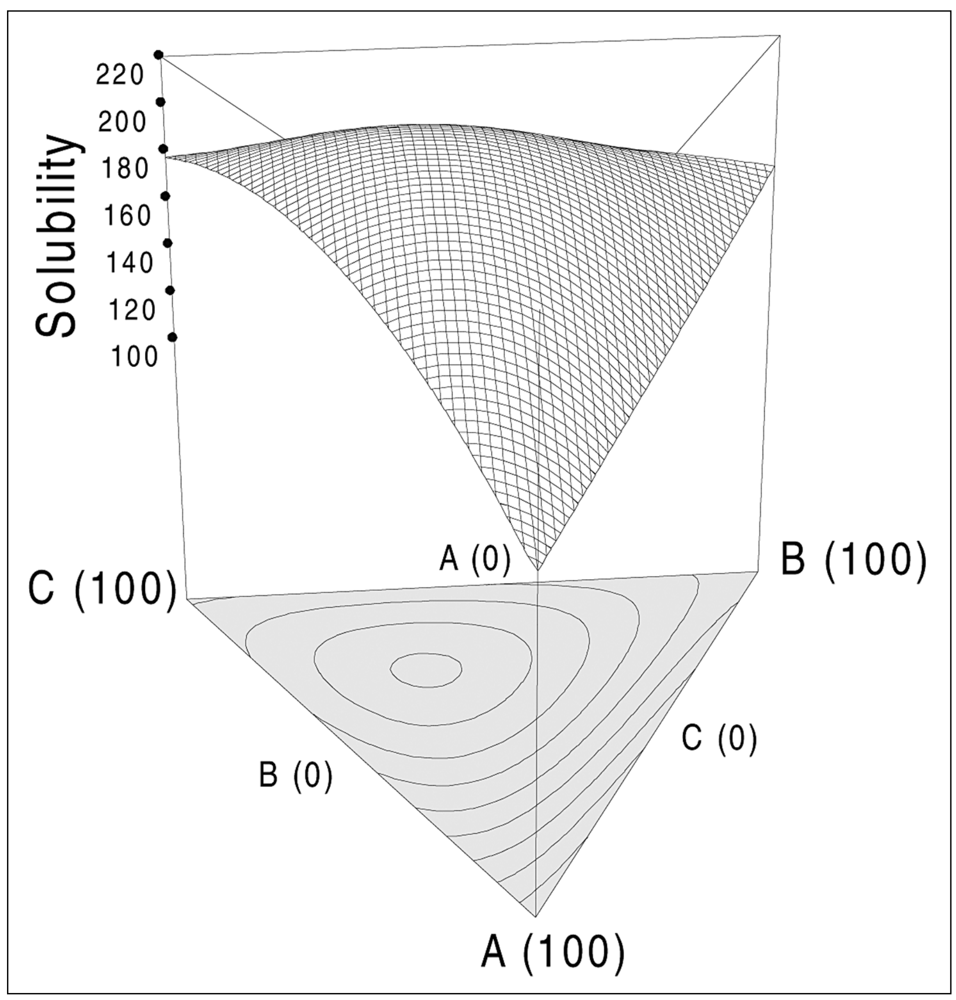


Fig. 2. Response surface graph of solubility.



Mixture

Continued from page 16

discussing three-component problems.

The points in the interior of **Fig. 1** between the centroid and each apex (8, 9 and 10) do not come from the simplex-centroid design. The formulators added these three-part blends, made up of two-thirds of each respective component and one-sixth each of the other two components, to provide better coverage of the experimental region. The interior points, which augment the textbook design to make it more practical, are called "axial check blends."

Creating a mathematical model

As shown, the solubility response data were fitted via least squares regression to a special form of polynomial equation developed by Scheffe.[4]

Technical

$$\hat{Y} = 122A + 165B + 178C - 6AB + 141AC + 35BC + 799ABC$$

We call this simply a "mixture model." Notice that this equation, unlike ones used to graph responses from a process, contains no intercept term, thus accounting for the overall constraint that all mixture components must sum to one. The "$\hat{Y}$" (referred to by statisticians as "Y-hat") represents the predicted response. It's the dependent variable. The independent variables (A, B, C), sometimes represented mathematically by Xs, have been transformed from their original metric of 0 to 100 percent to a coded format from 0 to 1, thus facilitating interpretation of the resulting coefficients (rounded).

For experiments like this one, where each ingredient can be put in at 100 percent, the first-order mixture-model coefficients predict the response from the pure components. For example, component A (methyl-ethyl-ketone or MEK) is the poorest solvent of the three tested—only 122 grams per liter (g/l) of the new substance went into solution vs. 165 and 178 for components B (toluene) and C (hexane), respectively.

The second order terms, such as AB, reveal interactions. For responses such as solubility, where higher is better, positive interaction coefficients indicate synergism. In this case the combination of A (MEK) and C (hexane) proved to be most synergistic according to the big positive coefficient (141) for AC. Together these two solvents work better than either one alone. On the other hand, negative interaction coefficients show antagonism between ingredients. For example, if the coefficient of -6 for the AB interaction was statistically significant (it's not), one could conclude that ingredients A and B work against each other to make the substance less soluble.

In this case, by augmenting their design with check blends, the experimenters made a sufficient number of unique formulations to allow estimation of a third-order term: ABC. This term, called a "special cubic," reveals the three-component interaction, if any. The coefficient of 799 does achieve statistical significance, thus providing solid evidence that all three solvents work together to be most efficacious. When you work with chemical formulations, be prepared for complex interactions of this degree.

Response surface graphs tell the story

The mixture models become the basis for response surface graphs, which can be generated from specialized software for mixture DOE.[5] Don't get bogged down in the mathematics—let the pictures tell the story. The graphs provide valuable insights about your formulation. **Fig. 2** shows a 3-D representation of the solubility response, with 2-D contours projected below it, as function of the three solvents. As you might expect from the data and discussion so far, the peak solubility is predicted when all three solvents are mixed together. A computer-aided search reveals an optimum composition of 27.58 percent MEK, 25.56 percent toluene and 46.85 percent hexane producing a predicted solubility of nearly 208 g/l. Subsequent confirmation blends performed within the normal range of this predicted response, so this mixture experiment proved to be successful.

What if you can't allow each ingredient to go in at 100 percent?

In many cases it will be unreasonable to vary each ingredient over a range of 0 to 100 percent. You must impose constraints on one or more of the ingredients, or on some combination of ingredients. Your constraints may form complex regions that cannot be covered by the standard mixture designs such as the simplex centroid. However, a number of statistical software packages can generate optimal designs that fit whatever degree of polynomial, such as the one used for the solubility case, that you think you need to adequately model your response. For example, consider making play putty as a kitchen chemistry experiment.[6] In this formulation, a chemical reaction occurs between a polymer (polyvinyl acetate in white glue) and a crosslinker (borax). Water participates as a solvent and modifies the physical properties (rebound and deformability) of the resulting dilatant material. Obviously the borax should be constrained to a narrow range of composition (1-3 percent) relative to the glue (40-59 percent) and water (40-59 percent). **Table III** shows a design of experiments to study the behavior of play putty as these ingredients vary. It was constructed by filling the mixture space with blends spaced at relatively even intervals (see **Fig. 3**), with enough of them to fit a special cubic model such as the one used in the previous example. Notice that some points are labeled with the number "2." These represent blends to be replicated (in random run order) for estimation of pure error.

Conclusion

Design of experiment methods can be applied to formulations if you account for the unique aspects of mixtures. By using appropriate designs, you greatly accelerate your exploration of alternative blends. Then with the aid of response surface graphics based on mixture models, you will discover the winning component combination.

References

1. Anderson, M.J., Whitcomb, P.J., Breakthrough Improvements with Experiment Design, *Rubber and Plastics News*, June 16, 1997, pp. 14-15.
2. Del Vecchio, R.J., Design of Experiments, Hanser/Gardner Pub., Cincinnati, 1997, pp. 100-101.
3. Cornell, J.A., Experiments with Mixtures, 3rd ed., John Wiley & Sons Inc., New York, 2002, pp. 60-66.
4. Scheffe, H., Experiments with Mixtures, Journal of the Royal Statistical Society, B, Vol. 20, 1958, pp. 344-360.
5. Helseth, T.J., et al, Design-Expert, Version 6 Software for Windows, Stat-Ease Inc., Minneapolis.
6. Anderson, M.J., Play Putty Problem, 3 pages of instruction and background, Stat-Ease Inc., 2002.
7. Anderson, M.J., Whitcomb, P.J., DOE Simplified: Practical Tools for Effective Experimentation, Productivity Inc., Portland, Ore., (2000).

The authors

Mark J. Anderson and Patrick J. Whitcomb are principals of Stat-Ease Inc. The chemical engineers co-authored "DOE Simplified: Practical Tools for Effective Experimentation."[7] Anderson and Whitcomb also have collaborated on numerous articles on design of experiments, many of which can be seen or ordered as reprints from www.StatEase.com.

Table III. Design of experiments for play putty.

Blend #	A Glue	B Water	C Borax	Blend Type
1	59	40	1	Vertex
1	59	40	1	Vertex
2	40	59	1	Vertex
2	40	59	1	Vertex
3	57	40	3	Vertex
3	57	40	3	Vertex
4	40	57	3	Vertex
4	40	57	3	Vertex
5	53	46	1	Edge
6	46	53	1	Edge
7	51	46	3	Edge
8	46	51	3	Edge
9	49	49	2	Centroid
9	49	49	2	Centroid
10	43	55	2	Check
11	55	43	2	Check

Fig. 3. Optimal design for a complex mixture (play putty).