Use of Simplex Experimental Designs in Detergent Formulation

J.P. NARCY and J. RENAUD, Colgate-Palmolive, Research and Development Paris Laboratory, 5 boulevard de la Mission-Marchand, 92 Courbevoie, France

ABSTRACT

Research in the field of detergent formulation is deeply involved in the optimization of numerous detergent systems. The use of experimental designs in looking for optimal formulations can save a large amount of laboratory time. This paper describes a coherent strategy of use for a special class of simplex designs adapted from Scheffé's work. The technique applies to multicomponent systems, as they are generally met in detergent formulation. Proper designs provide polynomial equations and, in addition, convenient graphical representations that enable the chemist to predict responses for a wide range of mixtures. Calculations are easy. A computer may help for routine application of these designs, but is not strictly required. The method is illustrated by a problem in the solubilization of light duty liquid detergents and at the same time provides arguments for better use of statistics in formulation. Finally, this simplex method is positioned among some other statistical techniques dealing with mixture designs.

INTRODUCTION

It is well established that statistics, and experimental design in particular, may provide answers of demonstrated worth for many formulation problems. However several pitfalls may discourage potential users, thus depriving them of a powerful tool. On the one hand, most of the existing statistical designs are general purpose techniques and cannot be applied directly to specific problems faced by the chemist. Other techniques require the assistance of a nonchemist—a computer programer or a statistician—for choosing the proper design and conducting subsequent data analysis.

In this paper we describe a method that overcomes these difficulties. The technique is devised especially for a specific problem: the study of multicomponent systems. Except for routine application of the method, a computer is not required; a chemist with a good knowledge in statistics can do without a professional statistician, for the paper offers a complete and flexible strategy.

The method applies to multicomponent systems; that is, mixtures of several components, the properties of which depend only on the relative percentages of the components and possibly on their total amount. This is typically the case with detergents for definite usage conditions wherein one is interested in performance or physical characteristics. Proper designs lead to polynomial equations and graphical representations which enable the chemist to easily predict responses for a wide range of different mixtures. Moreover requirements for several different responses may be combined. This is a useful application of the method in reaching pertinent marketing targets.

The method consists of a coherent strategy of use that we have built for a special class of "simplex" designs originally by Scheffé (1,2). After a presentation of simplex designs, we illustrate the method on a solubilizing problem in the field of the light duty liquid detergents (LDLD). For a fixed active ingredients system, we search for the solubilizers mixtures that lead to preassigned values of viscosity and cold weather clarity (clear point). It is unnecessary to insist on the interest of these two properties in the case of clear liquids that are sold in transparent packs.

SIMPLEX DESIGNS

Response Surface Designs

The use of experimental design in formulation work may arise for two different types of experimental objectives: to find combinations of components that optimize the property under study, or to obtain an empirical equation for that property in a domain of interest. The second objective, which includes the first, needs more elaborate designs, e.g., the response surface designs. The equation is generally a low order polynomial and consists of only a few terms; very often adding more terms improves the fit of the model.

Example: $Y = a_1 x_1 + a_2 x_2 + a_3 x_3 + a_4 x_1 x_2 + a_5 x_1 x_3 + a_6 x_2 x_3 + \epsilon$, where the a_i's are the coefficients, the x_i's the components proportions and ϵ an infinitesimal Gaussian variate. Taking out ϵ, one gets $\hat{Y}$ which is an estimator of the response Y, and the corresponding equation represents the response surface of the phenomenon. With three variables, response surfaces have no graphical representation in general, but for a particular case, the simplex design, visualization is possible. It concerns the domain of (x_1, x_2, x_3) mixtures such that the total content $x_1 + x_2 + x_3$ is fixed. In fact there exists a way of overriding this constraint, which will be used in this paper.

The usual mathematical definition of the simplex is:

$$\begin{cases} x_1 + x_2 + x_3 = 1 \\ x_1 \geqslant 0, x_2 \geqslant 0, x_3 \geqslant 0 \end{cases}$$

The representation is shown in Figure 1.

The effect of the constraint is to reduce the parameter space dimensionality from three to two; then all (x_1, x_2, x_3) mixtures lie inside a triangle.

With four components we get a tetrahedron, and graphical representation remains possible for faces and cuts of the tetrahedron.

Scheffé Designs

Several statisticians have provided simplex designs, but those by Scheffé (1,2) and Gorman and Hinman (3) are particularly relevant here. The Scheffé method consists in exploring the components space at points corresponding to an ordered arrangement known as a lattice. Two families of designs are available. First, Scheffé developed the simplex lattice design (1), which is determined by two figures: q for the number of components, and m for the degree of the polynomial response. Design points consist of pure components and mixtures of two or more components (up to m). For each mixture the proportions of the components are simple fractional numbers, design points symmetrically covering the simplex. When there is only one measurement per point (one replication), the sample size is given by the formula: $n = (m + q - 1)\,!\ /\ m!\ (q - 1)\,!$ Examples of representations of several (q, m) simplex-lattice designs are given in Figure 2.

Five years later, Scheffé published a new paper introducing the simplex-centroid. Contrary to the lattice design, in which there is no mixture with more than m components, centroid designs are such that for any number of components between 1 and q there exist one or several design points corresponding to these mixtures. In particular there is one point with all components present in equal

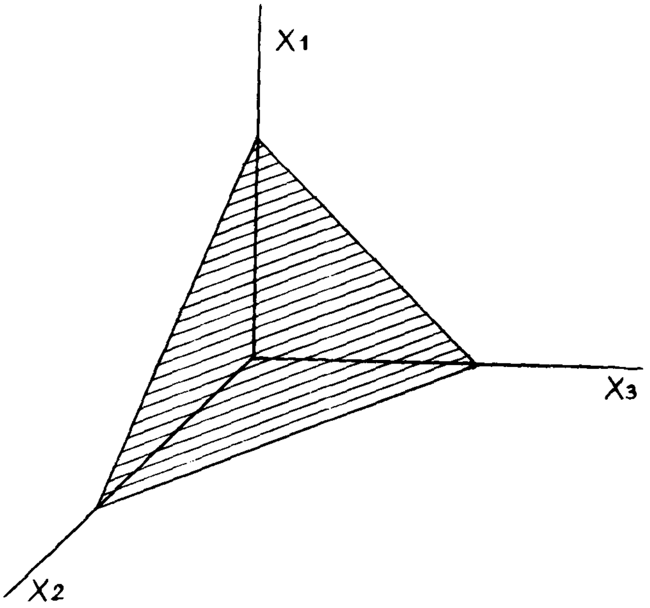

FIG. 1. Simplex representation.

proportions. The sample size for one replication is $(2^q - 1)$. Examples of representation can be seen in Figure 3.

We would say that the simplex centroid is relevant for mixtures of many components when response is simple (low order). The simplex lattice would apply to higher order variations but with not too many components, because estimation of regression coefficients becomes tedious when q is large.

Both families provide polynomial equations, the coefficients of which are simple functions of the measured responses at the experimental points.

EXPERIMENTAL AND STATISTICAL PROCEDURES

Starting Materials

Experiments were run on two different active ingredients systems denoted by A and B, respectively. The A system has been used to test the strategy of handling the designs. The B system was devoted to the illustration of a restricted design.

In both cases the solubilizing system consisted of three classical solubilizers: ethyl alcohol, urea and sodium xylene sulfonate (SXS). Obviously viscosity and clear point depend not only on the relative proportions but also on the total amount of these solubilizers. Therefore we introduced part of the water content as a pseudocomponent, the only purpose of which is to allow the total amount of the three solubilizers to vary. Let us denote x_1, x_2, x_3 and x_4 as the respective proportions of: the pseudocomponent part of water, alcohol, urea and SXS. Only $x_1 + x_2 + x_3 + x_4$ was fixed, say at p per cent, p being determined after some initial trials. Thus the pure water point corresponds to $x_2 + x_3 + x_4 = 0$, $x_1 = p$, but in the total formula water content is of course greater. The no-water face corresponds to $x_1 = 0$ and again, of course, there remains water in the formula; on this face $x_2 + x_3 + x_4 = p$ reaches its maximum. Between these two limiting cases $x_2 + x_3 + x_4$ takes intermediate values.

Besides water, alcohol content is split into two parts with one part fixed, while the other part (x_2) varies between 0 and p. The fixed part equals the quantity of alcohol necessary to obtain a "clear product" at the pure water point. It is therefore the minimum alcohol content over all the design points. There is no objection to doing so, since for usual formulas the alcohol content is generally higher than the content of the other solubilizers.

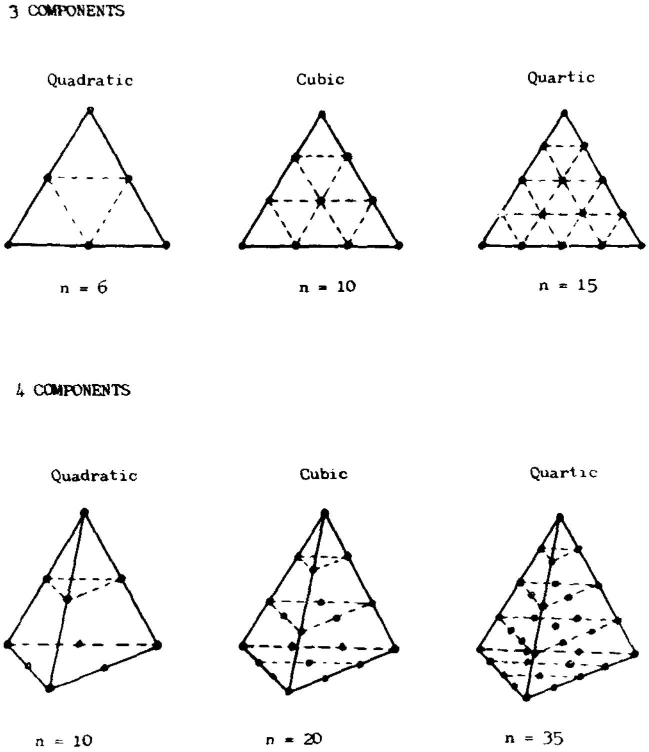

FIG. 2. Simplex lattice designs.

Methods of Measurement

All formulas were tested 24 hr after their formulation.

Viscosity: All measurements were made with a Brookfield viscometer model LVF, at 30 rpm, spindle $n°$ 1 or 2. Each sample was kept at 22 C for ca. 6 hr and then warmed to 25.5 C. At that moment the viscometer-spindle was put in the solution, the measurement being done when the sample is again at 25 C.

Clear point: The method was aimed at determining the temperature such that the solution became clear again after a previous freezing. About 15 ml of solution were poured into a glass test tube and left at -18 C for 12 hr. Then the test tube was left at room temperature and a thermometer was introduced in the frozen solution. While the temperature was going up, the thermometer was turned slowly inside the test tube for homogenization purposes. The temperature at which the solution became perfectly clear was taken as the clear point. For samples with a clear point under 7 C, a slight alteration was adopted to keep the experimental error level. In the latter cases clear point would be reached too fast with the above procedure. Therefore, after the frozen stage, the test tube was placed in an ice water bath instead of being left at room temperature.

Experimental error: The estimated error is the sum of two components: the bias error and the experimental error. The first one is bound to the lack of fit of the polynomial model and is studied further. The second one is familiar to all experimenters: it depends on the kind of variable measured, on the operator and on the chosen procedure. For both viscosity and clear point the assumption of a constant experimental error was valid here: $\eta \pm 20$ cps for viscosity and $\theta \pm 1.5$ C for clear point.

Response Surface Coefficients

References 1-3 provide the exact forms of the response surfaces together with the linear equations that give polynomial coefficients as functions of the responses at the design points. In these papers this information is supplied for simplex centroid and simplex lattice designs for systems

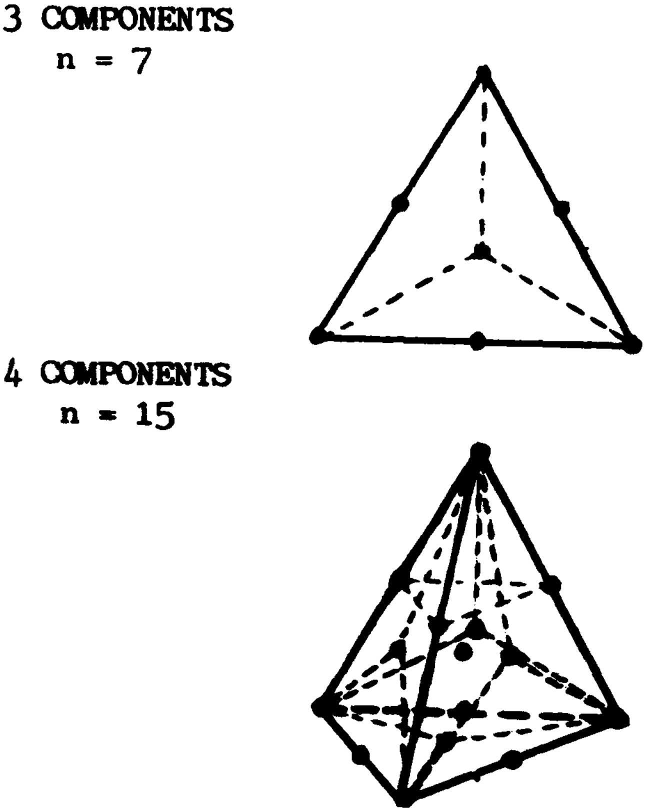

FIG. 3. Simplex centroid designs.

of three and four components. For details we therefore refer readers to References 1-3. The following example shows the simplicity of the equations. Let us contemplate the cubic simplex lattice design with three components. There are 10 design points, and 10 coefficients are to be estimated. Let Y_i denote the responses, with i as indicated on Figure 4.

The response surface is:

$$Y = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_{12} x_1 x_2 + \beta_{13} x_1 x_3 + \beta_{23} x_2 x_3 + \gamma_{12} x_1 x_2 (x_1 - x_2) + \gamma_{13} x_1 x_3 (x_1 - x_3) + \gamma_{23} x_2 x_3 (x_2 - x_3) + \beta_{123} x_1 x_2 x_3.$$

β_i's and γ_i's respectively are determined as follows:

$$\begin{gathered}
\beta_1 = Y_1 \\
\beta_2 = Y_2 \\
\beta_3 = Y_3 \\
\beta_{12} = (9/4)(Y_4 + Y_5 - Y_1 - Y_2) \\
\beta_{13} = (9/4)(Y_8 + Y_9 - Y_2 - Y_3) \\
\beta_{23} = (9/4)(Y_6 + Y_7 - Y_2 - Y_3) \\
\gamma_{12} = (9/4)(3Y_4 - 3Y_5 - Y_1 + Y_2) \\
\gamma_{13} = (9/4)(3Y_9 - 3Y_8 - Y_1 + Y_3) \\
\gamma_{23} = (9/4)(3Y_6 - 3Y_7 - Y_2 + Y_3) \\
\beta_{123} = 27Y_{10} - (27/4)(Y_4 + Y_5 + Y_6 + Y_7 + Y_8 + Y_9) \\
+ 9/2\ (Y_1 + Y_2 + Y_3).
\end{gathered}$$

Test of Fit

The response surface being obtained, the next step is to check its adequacy to the true responses. This is done with tests of fit, which deliver proper measures comparing the bias error to the experimental error. Throughout the paper we rely on a test, originally by Behrens (4).

By construction of this type of design, there is a perfect fit of the model at all design points. Consequently check points are to be different from these points. Some of them will correspond to mixtures of a particular interest, whereas others will consist of design points of different simplex models, thus permitting economy in experimentation.

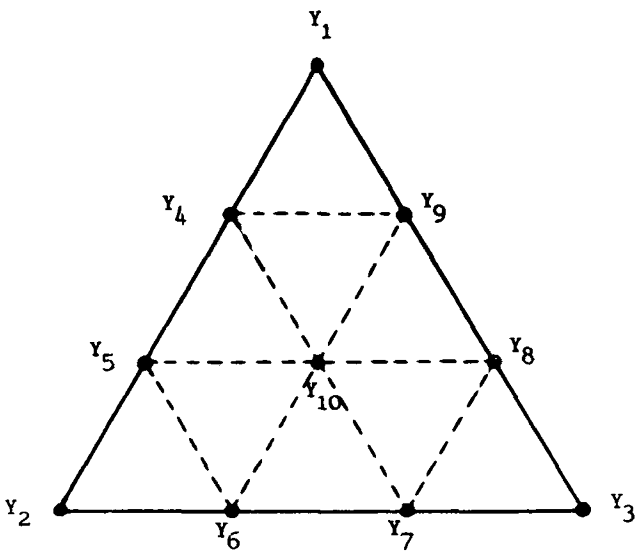

FIG. 4. Cubic simplex lattice design with three components.

The number of check points necessary depends on the degree of the response polynomial and will be accounted for later. One must perform the Behrens test for each check point and report the individual results, fit or no fit, on the graphical representation of the design.

Then by looking at the graph one gets an overall mapping of the adequacy of the tested model. We strongly recommend this way of proceeding, as opposed to using global tests such as the Fisher one, for the sake of visualization together with a more detailed information on the lack of fit.

We now describe the test procedure to use for each check point. Suppose we run r different measurements at the check mixture, $\bar{Y}$ denoting the average measured response. Let $\hat{Y}$ denote the predicted response at that point, as given by the model under test. What follows relies on two basic assumptions: first, at each of the n design points, all observations have been equally replicated, say k times. Scheffé's theory may apply to a more general case where observations can be replicated differently, according to the position of the mixture, vertex, mid point, centroid etc. However, for the operator's convenience and simplicity in calculation, we recommend equal replication. The second basic assumption concerns the experimental arror, denoted by σ^2. We assume that σ is homogenous enough to be considered constant over the domain of the study. For viscosity and clear point, as said before, this assumption is realistic; when it is not, refer to References 1-3. The test procedure consists in comparing the characteristic $t_{f,g,\theta}$ with the critical value given in Behrens tables (4) which are read for suitable entries f, g and θ. t is expressed by:

$$t_{f,g,\theta} = |\bar{Y} - \hat{Y}| / \sigma\sqrt{1/r + \xi/k}$$

where f and g are the numbers of degrees of freedom, respectively, associated with $\bar{Y}$ and $\hat{Y}$: $f = r - 1$; $g = nk - 1$; and θ: $\tan\theta = \sqrt{\xi (r/k)}$.

ξ is an important parameter which is a symmetric function in the x_i's and only depends on the position of the check point. Contour plots for ξ are given in Reference 3. We use the Behrens test rather than the usual student test for paired comparisons, because the errors on $\bar{Y}$ and $\hat{Y}$ arise from different sources. Only Behrens test properly accounts for this situation, by introducing an additional parameter, $\tan\theta$, which is the ratio of two variances:

TABLE I

Number of Design Points

q	Quadratic	Centroid	Cubic	Quartic
3	6	7	10	15
4	10	15	20	35
5	15	31	35	70

$$\tan\theta = \sqrt{\operatorname{var}\hat{Y}/\operatorname{var}\overline{Y}} \text{ with: } \operatorname{var}\hat{Y} = \xi\,(\sigma^2/k)$$
$$\operatorname{var}\overline{Y} = \sigma^2/r$$

When $f = g$ and $\theta = 45°$, Behrens and Student procedures are identical. With $f = g$ but $\theta \neq 45°$, the Student test would be too severe. With $f<g$, as is the case in our illustration about LDLD solubilization, the Student test would be too severe for $\theta<45°$ and not severe enough for $\theta>45°$.

Possibility of Handling Restricted Designs

After Behrens test has been performed on all check points we have a good idea of the fit of the model. A frequent cause of misfit is the presence of one or several extreme values, very different from the response at neighboring points, on the boundaries of the domain. In this case handling the design remains possible, with appropriate changes in coordinates. These changes act as if the domain were restricted, as drawn in Figure 5. The shaded areas represent the reduced domains, which are simplexes in x'_i.

Numerical support will be given only for case A from Figure 5. when dealing with experimental results.

We generalized a technique, first recommended by Scheffé but for different purposes, which we called the pseudovertex method. We present with some details case A, the simplest one, and more briefly cases B and C.

If the extreme value corresponds to the third component, $x_3 = 1$, we search for a point, with proportion h of this component, such that the response becomes reasonable, by gradually reducing x_3. Several trials may be necessary to find the most appropriate proportion h.

Coordinates of the pseudo-vertex are:

$$\begin{cases} x_1 = x_2 = (1-h)/2,\ x_3 = h \\ 0 \leqslant x_3 \leqslant h < 1 \end{cases}$$

This leads to the following change in coordinates:

$$\text{(I)} \begin{cases} x_1 = x'_1 + \left[(1-h)\,x'_3/2\right] \\ x_2 = x'_2 + \left[(1-h)\,x'_3/2\right] \\ x_3 = hx'_3 \end{cases}$$

$$\text{(II)} \begin{cases} x'_1 = x_1 - \left[(1-h)\,x_3/2h\right] \\ x'_2 = x_2 - \left[(1-h)\,x_3/2h\right] \\ x'_3 = x_3/h \end{cases}$$

One can check that $x'_1 + x'_2 + x'_3 = x_1 + x_2 + x_3$. Consequently the reduced design is still a simplex, but in (x'_1, x'_2, x'_3).

System I is first used, in order to determine the true mixtures to be run for the new design.

System II is used for testing the lack of fit and for prediction of responses at mixtures of interest.

Case B, in Figure 5, can be solved with two successive changes in coordinates the resultant of which is:

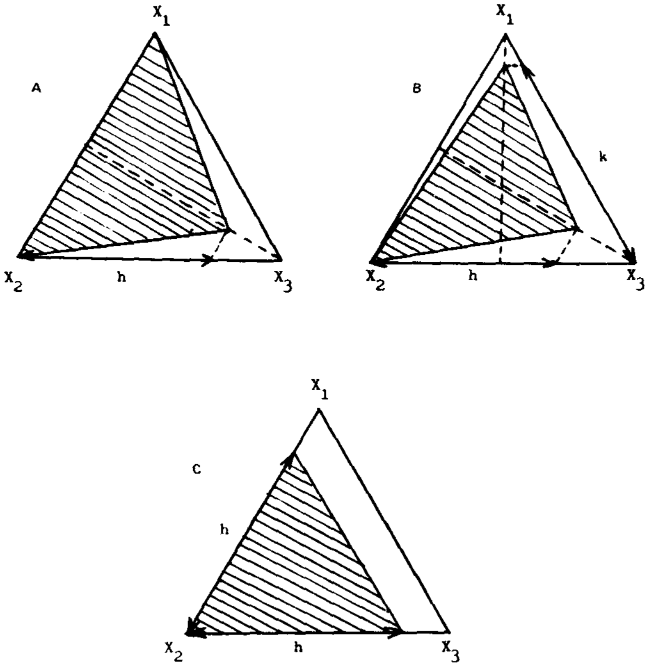

FIG. 5. Restricted designs. A: one extreme vertex; B: two extreme vertices; C: one extreme edge.

$$\begin{cases} x'_1 = (x_1/k) - \left[(1-h)\,x_3/2hk\right] \\ x'_2 = x_2 - \left[(1-k)\,x_1/2k\right] - \left[(1-h)\,x_3/2h\right] + \left[(1-k)\,(1-h)x_3/4hk\right] \\ x'_3 = (x_3/h) - \left[(1-k)x_1/2k\right] + \left[(1-k)\,(1-h)x_3/4hk\right] \end{cases}$$

The above system still can easily be worked out through hand calculations.

Finally, for case C, we have:

$$\begin{cases} x'_1 = x_1/h \\ x'_2 = x_2 - \left[(1-h)\,(x_1 + x_3)\,/h\right] \\ x'_3 = x_3/h \end{cases}$$

The latter system is about as simple as for case A, but it may be handled in a different way. As it appears on Figure 5C the restricted simplex is an equilateral triangle. With a suitable h we may avoid the change in coordinates. For example let us consider a quartic design; if h = 0.75, the

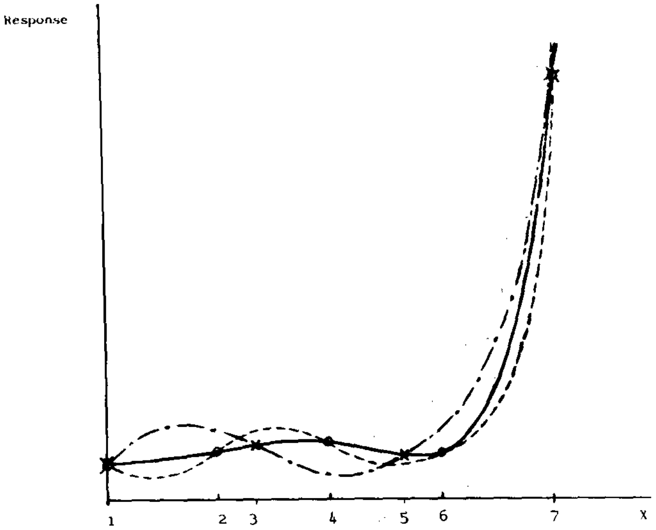

FIG. 6. Effect of an extreme point on the polynomial degree. o—x = Experimental; x—·—x = cubic; o– – –o = quartic.

TABLE II

Cubic Simplex Lattice, Four Components

Design points	Viscosity	Clear point		
		First measurement	Second measurement	Average
6% Water	362.5	35	38	36.5
6% Alcohol	78	11.3	11.6	11.5
6% SXS	365	12	12.6	12.3
6% Urea	1630	23.5	23.8	23.7
2% Water + 4% alcohol	129	7.2	7.4	7.3
2% Water + 4% SXS	332	11.8	12	11.9
2% Water + 4% urea	497	8.2	9.6	8.9
2% Alcohol + 4% SXS	166	14.3	14	14.2
2% Alcohol + 4% urea	232	8	9.2	8.6
2% SXS + 4% urea	427	11.5	10	10.8
4% Water + 2% alcohol	199	7.8	8.2	8
4% Water + 2% SXS	341	7.2	6.8	7
4% Water + 2% urea	431	4	6.5	5.3
4% Alcohol + 2% SXS	98	12.8	12.6	12.7
4% Alcohol + 2% urea	119	8	8.5	8.3
4% SXS + 2% urea	358	14	13.2	13.6
2% Water + 2% alcohol + 2% SXS	200	10.8	11.8	11.3
2% Water + 2% alcohol + 2% urea	265	5	5	5
2% Water + 2% SXS + 2% urea	492.5	8.7	8.4	8.6
2% Alcohol + 2% SXS + 2% urea	235	10	9.9	10

remaining design points correspond to the cubic model for the restricted design.

Besides, it may happen that the domain of interest is very different from an equilateral triangle. But very often this domain can be split into several adjacent equilateral triangles. In particular, an isosceles trapezium such that one base line is twice the other can be split into three equilateral triangles. Actually this case arises when we cannot formulate the upper half of a Scheffé's simplex, since the other half is an isosceles trapezium. This is a generalization of some ideas contained in Reference 17.

Sequential Search for Proper Simplex Designs

Depending on the types of applications, the number of components can vary from three to about ten. When this number becomes large, above five for instance, only quadratic fit is reasonable. This is the case in the petroleum industry, where problems such as octane-blending vs. performance characteristics are encountered. In detergent formulation chemists rarely work on more than five components. Accordingly we deal only with systems composed of three to five materials, and we suggest a maximum of the fourth degree for the approximating polynomials. Beyond this we believe reducing the domain is better than attempting quintic equations. Table I gives *n*, the number of design points, for designs relevant to these conditions.

In Table I two cells are crossed, because we do not recommend using the corresponding designs: in one case the number of experiments is probably prohibitive; in the other, the crossed design is identical to the next one, with the exception of one point. It should not be thought that increasing the degree is necessarily an improvement. In particular, there is a lot of hesitation between cubic and quartic models, and we wish to illustrate this difficulty with the example in Figure 6.

TABLE III

Quartic Simplex Lattice, Three Components[a]

Design points	Viscosity	Clear point		
		First measurement	Second measurement	Average
6% Water	362.5	35	38	36.5
6% Alcohol	78	11.3	11.6	11.5
6% Urea	1630	23.5	23.8	23.7
3% Water + 3% alcohol	165	12.8	13	12.9
3% Water + 3% urea	537.5	6	5.8	5.9
3% Alcohol + 3% urea	202.5	5	5.4	5.2
1.5% Water + 4.5% alcohol	96.5	11	11.8	11.4
1.5% Water + 4.5% urea	560	16.4	15.2	15.8
1.5% Alcohol + 4.5% urea	310	7.6	8	7.8
4.5% Water + 1.5% alcohol	230	14.8	13.4	14.1
4.5% Water + 1.5% urea	455	3.6	3.3	3.5
4.5% Alcohol + 1.5% urea	108.5	7.6	9.4	8.5
3% Water + 1.5% alcohol + 1.5% urea	257.5	7.6	7	7.3
3% Alcohol + 1.5% water + 1.5% urea	167	8.4	9.4	8.9
3% Urea + 1.5% water + 1.5% alcohol	275	3.3	3	3.2

[a] Polynomial responses: Where x_1, x_2, x_3, respectively, represent water, alcohol and urea. Viscosity: $\eta = 362.5x_1 + 78x_2 + 1630x_3 - 221x_1x_2 - 1835x_1x_3 - 2606x_2x_3 - 46.666\,x_1x_2\,(x_1 - x_2) + 2820x_1x_3(x_1 - x_3) + 3064x_2x_3(x_2 - x_3) - 332x_1x_2(x_1 - x_2)^2 - 3086.66x_1x_3(x_1 - x_3)^2 - 3330.666x_2x_3(x_2 - x_3)^2 - 4949.333x_1^2x_2x_3 - 2366.667x_1x_2^2x_3 + 13438.667x_1x_2x_3^2$. Clear point: $\theta = 36.5x_1 + 11.5x_2 + 23.7x_3 - 44.4x_1x_2 - 96.8x_1x_3 - 49.6x_2x_3 - 52.3x_1x_2(x_1 - x_2) - 99.733x_1x_3(x_1 - x_3) + 36.266x_2x_3(x_2 - x_3) - 62.4x_1x_2(x_1 - x_2)^2 - 49.066x_1x_3(x_1 - x_3)^2 - 3.2x_2x_3(x_2 - x_3)^2 + 555.733x_1^2x_2x_3 + 174.933x_1x_2^2x_3 - 438.933x_1x_2x_3^2$.

TABLE IV

Tests of Fit for Quartic Model

A. Behrens Test for Viscosity

Check point	Computed values	Observed values	$\lvert d \rvert$	ξ	$t_{f, g, \theta}$
a	207	199	8	0.67	1.1
b	119	129	10	0.67	1.3
c	161	119	42	0.67	5.7 **
d	225	232	7	0.67	0.9
e	515	497	18	0.67	2.4
f	510	431	79	0.67	10.6 **
g	248	245	3	0.50	0.45
h	133.6	130	3.6	1.3	0.37
i	223	219	4	1	0.46
j	339.4	327	12.4	1.05	1.38
k	236.2	235	1.2	1.05	0.13

B. Student Test for Clear Point

Check point	Computed values	Average of the two observed values	$\lvert d \rvert$	ξ	t
a	12.9	8	4.9	0.67	7.1 **
b	12.3	7.3	5	0.67	7.3 **
c	7.2	8.3	0.9	0.67	1.3
d	5	8.6	3.6	0.67	5.2 **
e	12.7	8.9	4.8	0.67	7.0 **
f	2.1	5.3	3.2	0.67	4.6 **
g	6.3	5	1.3	0.50	2.0
h	8	7.5	0.5	1.3	0.6
i	2.8	3.4	0.6	1	0.8
j	5.2	5.2	0	1.05	0
k	2.8	3.4	0.6	1.05	0.8

TABLE V

Quartic Model on Restricted Design, Three Components

		Clear point		
Simplex points	Viscosity, cps	First measurement	Second measurement	Average
6% Water	410	8.1	7	7.6
6% Alcohol	82	10.8	11.6	11.2
6% Urea	7600	39	38	38.5
3% Water + 3% alcohol	176.5	6.7	7	6.9
3% Water + 3% urea	480	2	2.8	2.4
3% Alcohol + 3% urea	197.5	6.8	6.8	6.8
1.5% Water + 4.5% alcohol	116	8.8	8.8	8.8
1.5% Water + 4.5% urea	560	3	3	3
1.5% Alcohol + 4.5% urea	332.5	4.8	5	4.9
4.5% Water + 1.5% alcohol	265	3.6	4.2	3.9
4.5% Water + 1.5% urea	445	2.4	2	2.2
4.5% Alcohol + 1.5% urea	127	8.2	9	8.6
3% Water + 1.5% alcohol + 1.5% urea	285	4.6	4.8	4.7
3% Alcohol + 1.5% water + 1.5% urea	186	6.9	6.6	6.8
3% Urea + 1.5% water + 1.5% alcohol	325	4.8	4.9	4.9

TABLE VI

Transformed Design Points for the Restricted Simplex

Design points	*a*	*b*	*c*	*d*	*e*	*f*	*g*	*h*	*i*	*j*	*k*	*l*	*m*	*n*	*o*
Viscosity values, cps	410	265	176.5	116	82	115	183	295	1440	465	440	430	273	176	287

TABLE VII

Check Points for the Restricted Simplex

	True proportions			Pseudoproportions			Viscosity values		
Design points	x_1 Water	x_2 Alcohol	x_3 Urea	x'_1 Water	x'_2 Alcohol	x'_3 Urea	Experimental	Usual quartic	Transformed quartic
p	1:10	13:20	1:4	.072	.621	.305	154	146	158.9
q	3:20	1:4	3:5	.084	.184	.732	297.5	104	320
r	3:4	3:20	1:10	.739	.139	.122	310	35.4	290.4
s	1:2	7:20	3:20	.484	.334	.183	230	-4.1	210.3

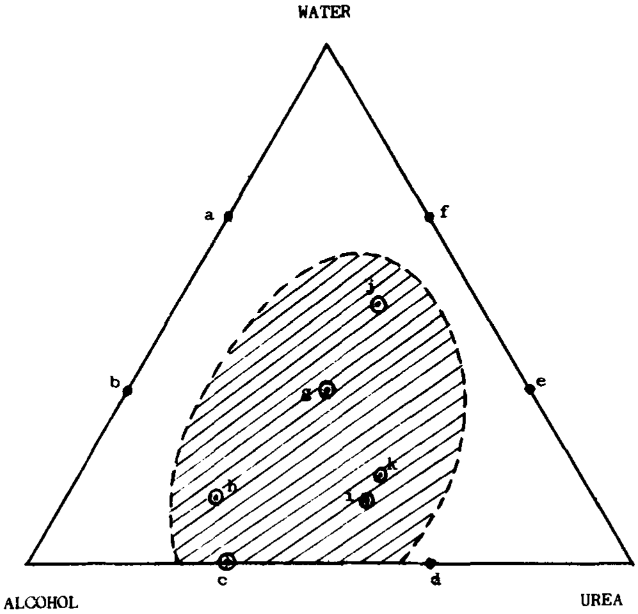

FIG. 7. Fit of the quartic model for clear point.

The X-axis represents design points on the edge of a ternary diagram. Points 1 and 7 corresponding to vertices are common to cubic and quartic models, points 3 and 5 belong to the cubic and 2, 4 and 6 to the quartic. An experimental curve was drawn by assuming the simplest variation between two consecutive points; cubic and quartic responses were also drawn.

This example leads to several remarks: cubic and quartic deviations from the experimental curve are of opposite sign; the quartic model is closer to the experimental situation than the cubic one; the distance between cubic and quartic curves is enlarged by the presence of point 7, which has an extreme value; and by taking off point 7 (restricted simplex) a quadratic model would fit best.

In practice the experimenter, facing a lack of fit situation with the cubic model, would have tried the quartic one with only a very small benefit. As a last remark about these two models, a cubic design should be tried when a minimum and a maximum are expected, whereas a quartic design should be run for three extrema. We are now in a position to list the steps required for the sequential search of proper simplex designs: (a) The first step consists in defining the domain to be considered. The limits of the domain are subject to three constraints: formulation

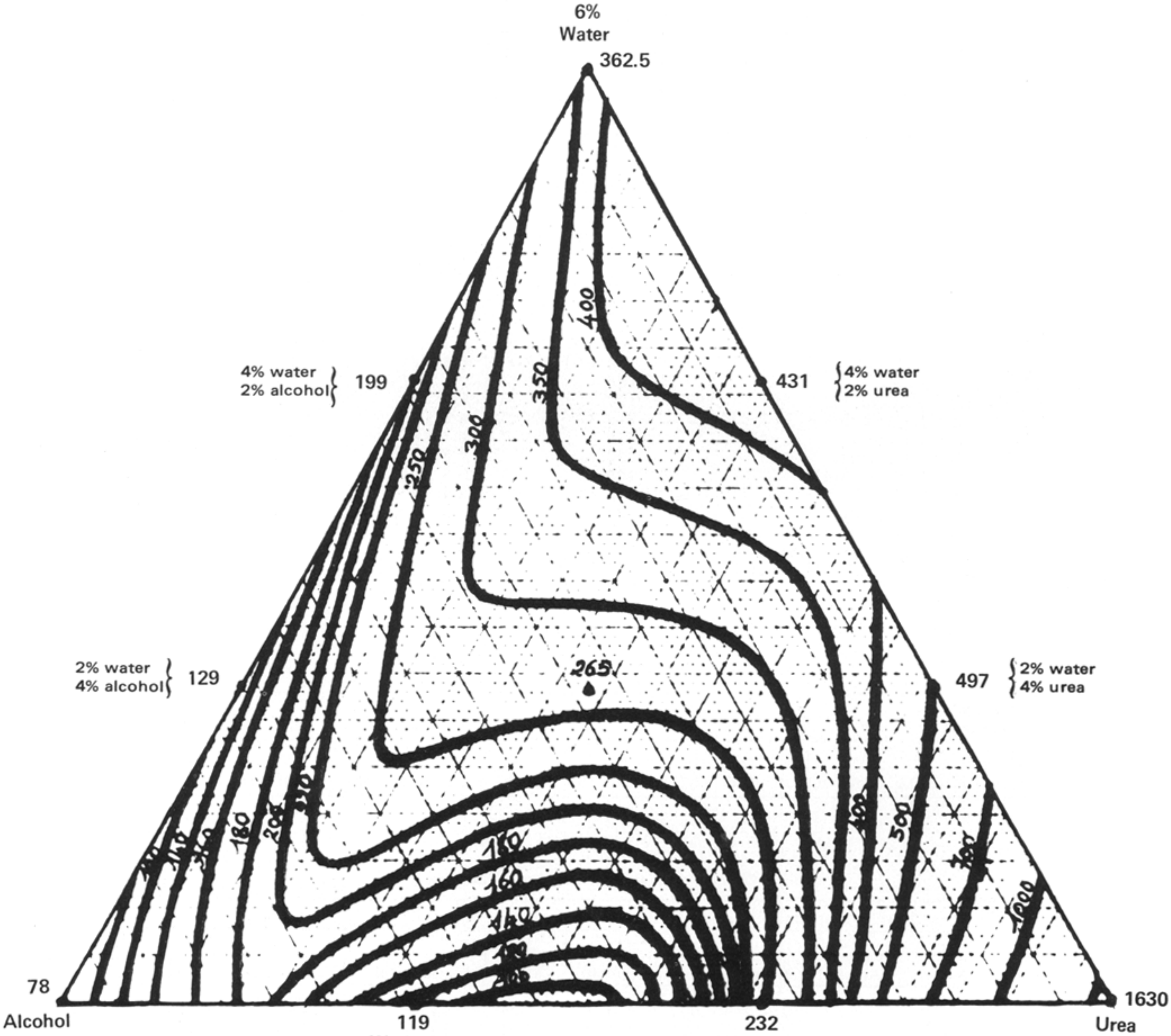

FIG. 8. Viscosity, cubic model.

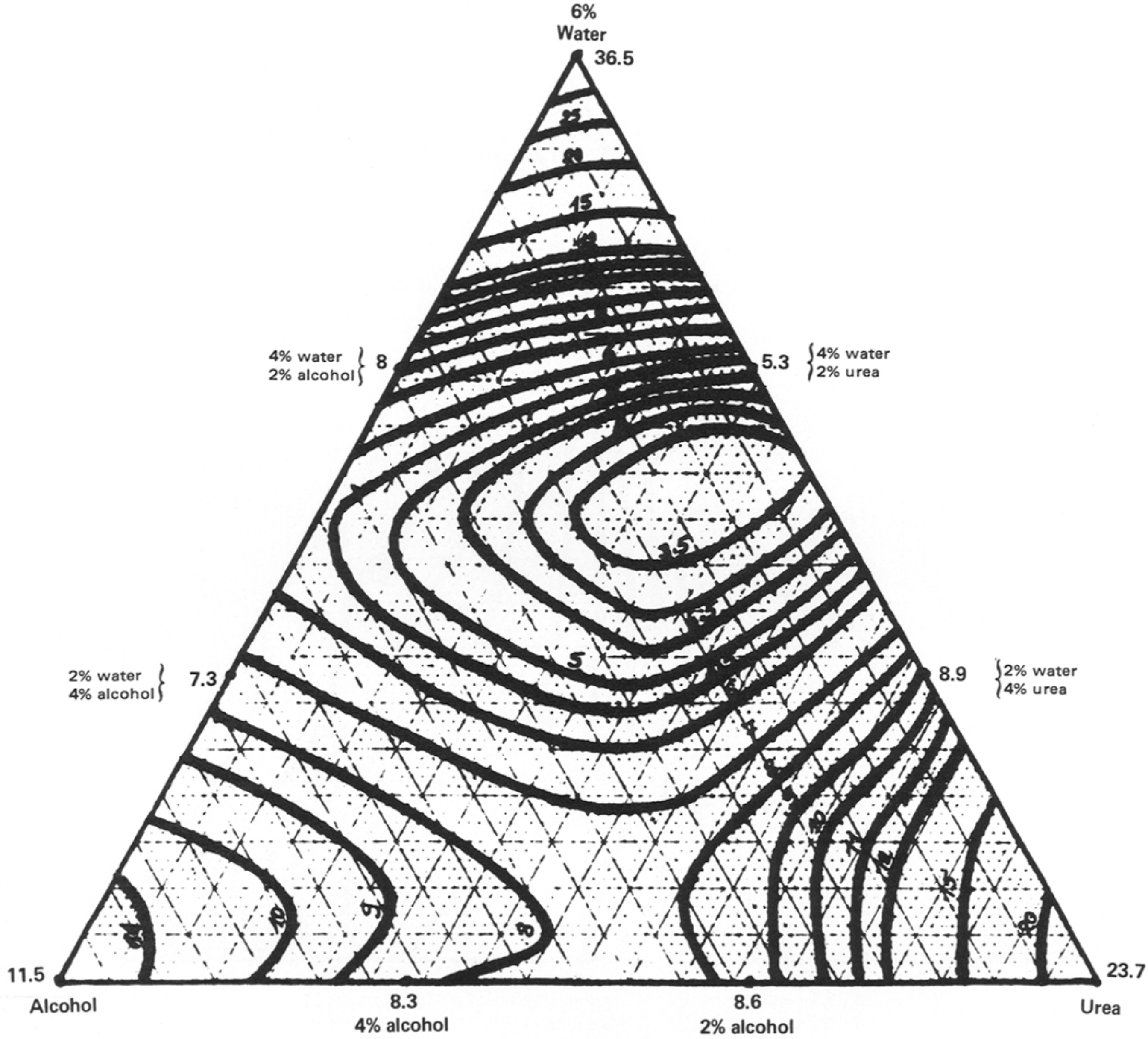

FIG. 9. Clear point, cubic model.

feasibility, mixtures of interest from an economic point of view, and responses in the ranges of interest. A few pilot experiments are necessary. (b) A second step is to measure the responses at the vertices: the pseudovertex feature is used in case of extreme values. (c) A third step is to run a simple model, that is the quadratic lattice or the centroid. Measurements at the design points are replicated to reach a low variance for polynomial coefficients and we establish the polynomial equation of the model. Check points are chosen, lying inside the domain, such as the inner points of the quartic model. If fit is correct, graphical representations are drawn for systems of three components. Optimal areas are isolated. If we have to consider several responses, an "optimum optimorum" is obtained by superposition of the respective graphs.

If the model is inadequate, several decisions can be made: if many discontinuities are encountered, either reduce the domain or split it in several parts; if values appear to be very abnormal, reconsider the problem, gain knowledge on the phenomenon and possibly try an "explicative" model, perhaps nonlinear. Otherwise, move to the cubic lattice design: Proceed similarly for cubic and eventually for quartic. Check points for the cubic model will be chosen roughly half among design points of the previous model, half among these of the quartic model. This recommendation has an obvious interpretation in terms of economy in experimentation. As for the check points of the quartic model, try all cubic points that do not belong to the quartic design, together with some inner points of interest. The examples presented in the figures very closely illustrate this procedure.

Several statisticians have also provided designs for multicomponent systems, and we have included a fairly complete bibliography on this subject (5-16). Roughly all designs can be split into two families: one relies on Box and Draper principles (6), whereas the other one follows Scheffé's ideas. When one is worried about having too many inner points in the design we suggest that Box and Draper be used. However we believe that methods similar to those of Scheffé's are preferable in most cases, above all because of convenience, ease of computation, and graphical representation.

RESULTS

As mentioned before, we first considered the LDLD based on the A system of active ingredients. This allowed us to display the step by step progression for reaching the

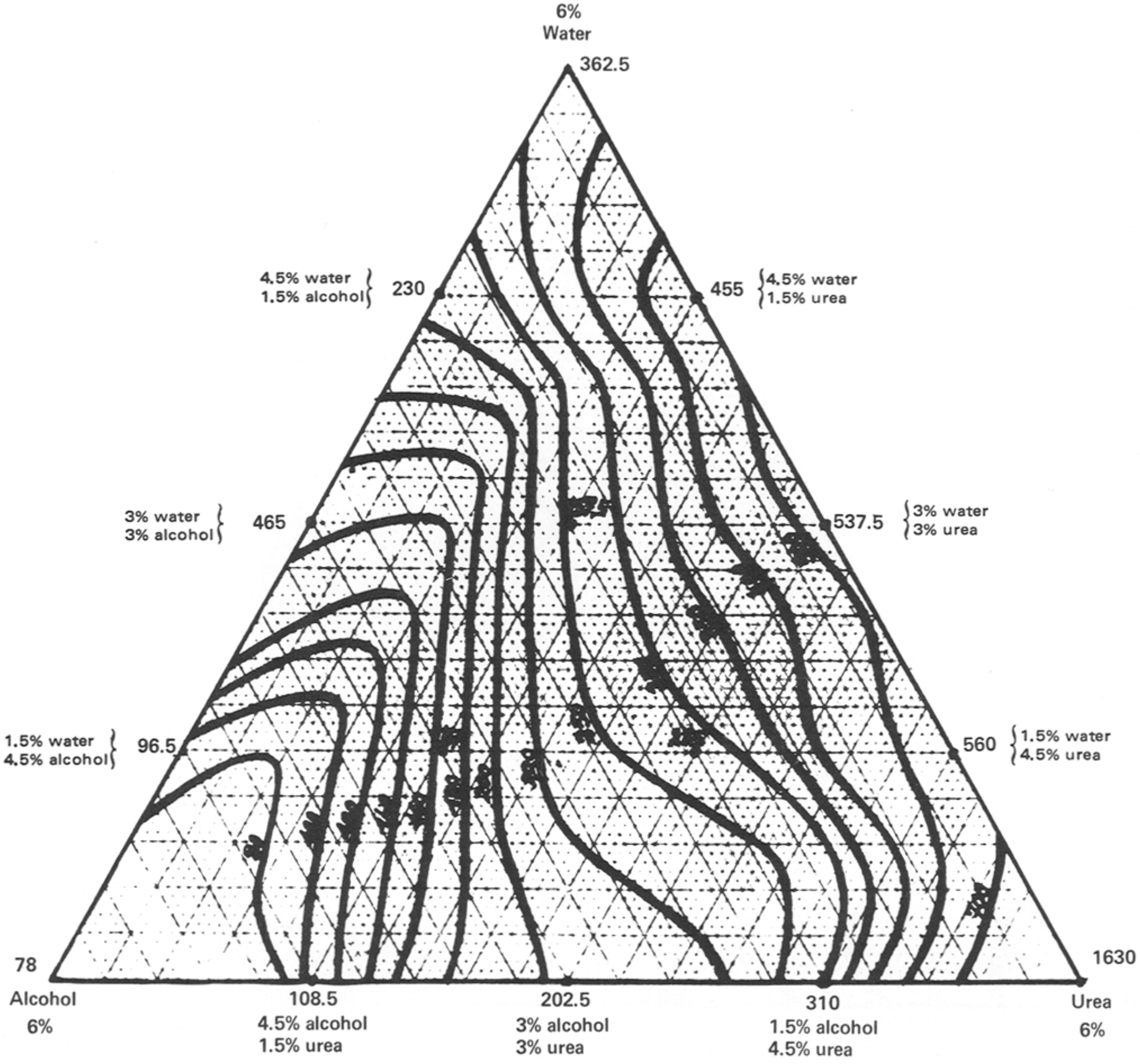

FIG. 10. Viscosity, quartic model.

simplex model that fits the solubilization system, with respect to both viscosity and clear point. Then we used the formula based on the B system of active ingredients for numerical support of the pseudovertex technique. For both systems of active ingredients, designs were aimed at better control of the solubilization system compositions in terms of viscosity and clear point values. There were several marketing requests asking for a clear point either under 5 C or around 10 C, with a medium viscosity of 180-200 cps in all cases.

Definition of the Domain

Nine pilot experiments were run with points belonging to a simplex centroid design. There were four components: water, alcohol, urea and SXS, the sum of which corresponded to 12% of the whole formula. The fixed part of the alcohol content was set at 4% of the whole formula. Out of the nine experiments, four consisted of pure components, four of mixtures of three components in equal proportions, the last one being the grand centroid, i.e., the point including all four components in equal proportions. These points were spread well from a geometrical point of view. The fact that we chose pilot experiments within a simplex design suggests that we were confident about the target domain. In fact results proved that reality was different; more precisely, all nine clear point values were above 10 C, which left little hope for finding mixtures at 5 C.

Our feeling was that lowering the total amount of solubilizers was a possible improvement. Some trials, this time out of any design, indicated that this was correct, so we repeated the nine pilot experiments but with the sum of the solubilizer proportions kept at 6% rather than 12% and with only 2% of fixed alcohol. Out of the nine points, seven had a clear point between 5 C and 12 C and a viscosity between 80 cps and 500 cps: the domain appeared to be well centered, with no extreme values at the vertices.

Simple Model

For this purpose we ran the six remaining points of the centroid design. Polynomial equations for response were obtained, and three check points showed the poorness of the fit.

Cubic Model

The sum of the proportions of the solubilizers was kept at 6% with, again, 2% of fixed alcohol. Since vertices and centers of faces are common to centroid and cubic designs,

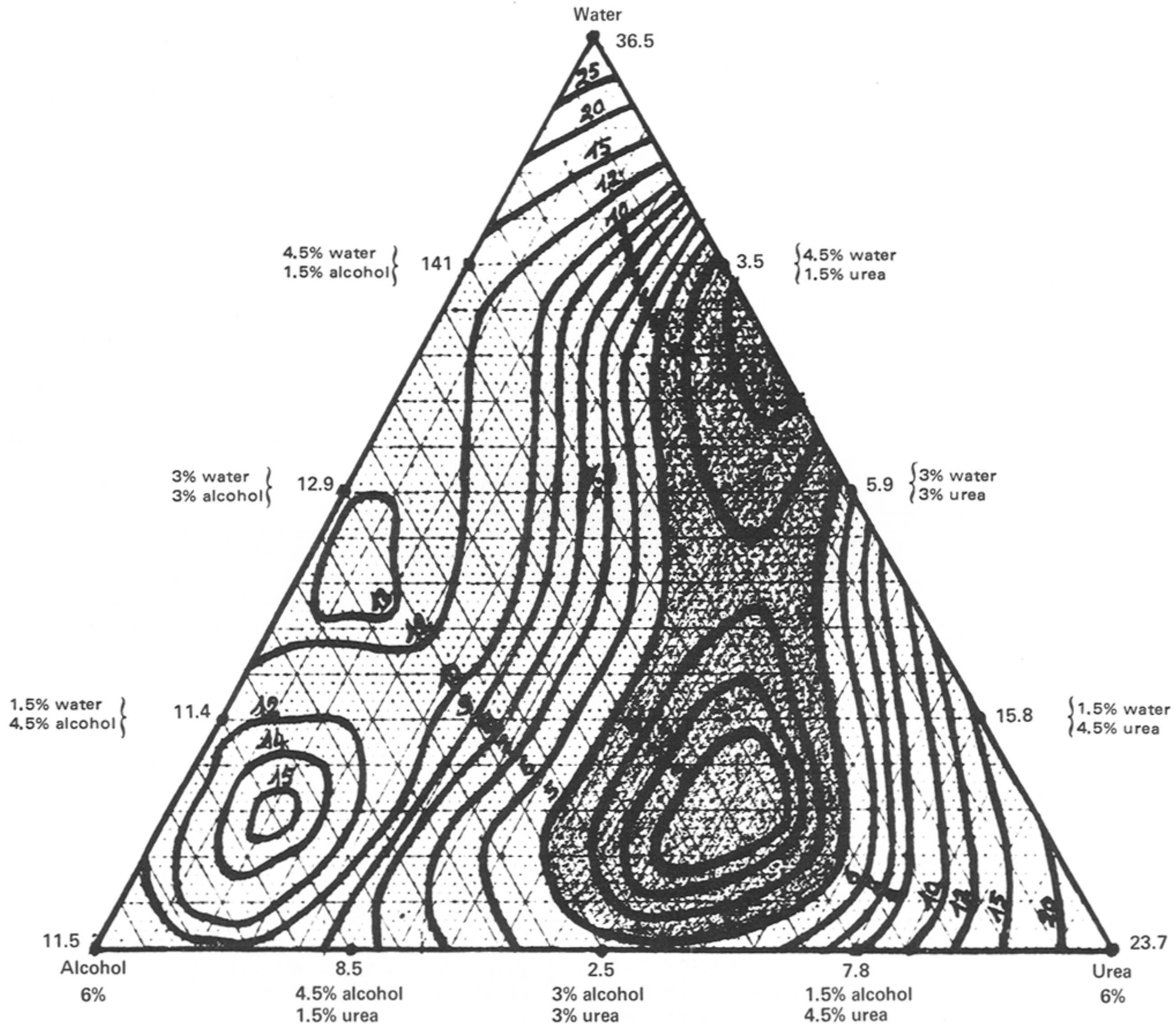

FIG. 11. Clear point, quartic model.

out of the 20 points of the cubic simplex, 8 were known, and only 12 had to be run. Results are shown in Table II.

A rapid examination of data showed that mixtures with SXS always give higher clear points than corresponding mixtures with urea. Therefore check points were preferably chosen among ternary mixtures with water, alcohol and urea. We used 16 check points, seven belonging to the previous design—grand centroid and midpoints of the edges, and nine belonging to the quartic model, all taken on the water-alcohol-urea face.

For viscosity only 5 points out of 16 fit reality at the 95% level of confidence. For clear point only two points fit, which reflects the expected greater complexity of the clear point response. Obviously we must turn to a more complex model.

Quartic Model

We had now accumulated enough data to be able to think that we could match all marketing requests by only considering water, alcohol and urea. We immediately remarked that all of the 15 quartic points had already been run, whether as points of previous designs or check points.

Grouped results, together with polynomial responses, are presented in Table III. Eleven check points were chosen: seven, labeled from *a* to *g*, belonged to the cubic design, including the grand centroid; and four, labeled from *h* to *k*, represented inner points of interest.

Viscosity

Throughout this study viscosity measurements at the design points have not been replicated, because previous experimental work has displayed a fairly stable error. Numerical results for check points are given in Table IV in which $|d|$ represents $|\bar{Y} - \hat{Y}|$.

In order to apply Behrens formula, as described above, five replications were done at the grand centroid and $\sigma^2 = 64.38$, with four degrees of freedom. The remaining parameters are: $r = 4$, $k = 1$, $f = 4$, $g = 14$ and $\theta \simeq 75°$.

Only for points *c* and *f*, observed and computed values contradict at the 95% level of confidence. For all other points fit is very good; consequently the quartic lattice model can be used as the surface response for viscosity.

Clear Point

For all clear point measurements in this paper, two replications were done, because they are more subject to interpretation by experimenters. Again Behrens test was used, but here it reduced to the usual Student's *t* test.

First, causes for error for check points and design points were similar, since the former were well spread belonging to other designs. Second, and as a consequence of the first reason, the numbers of degrees of freedom are high. Results

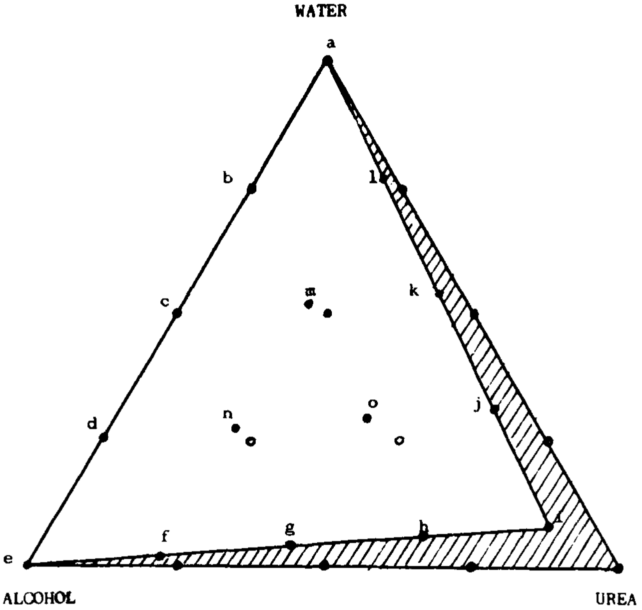

FIG. 12. Restricted simplex.

are shown in Table IV. Five out of eleven points present significant differences, but, as seen on Figure 7, the most interesting part of the domain is well fitted. This area, shaded on the figure, corresponds to an intermediate level of urea, and clear point variation is continuous inside it; in this region the quartic lattice model can be used as the surface response for clear point. It is interesting to note that, for extreme levels of urea, the profile of variations explains the poorer fit by a quartic equation. This is understandable technically: when we increase urea concentration from zero, the influence on clear point is first fairly stable; then clear point is lowered, and, finally for high urea levels, clear point goes up very quickly.

Graphic Representation

As stated above, Scheffé designs allow graphic representation for ternary mixtures. As examples we give four graphs: in Figures 8 and 9 viscosity and clear point are represented for the cubic model, and in Figures 10 and 11 for the quartic one.

Using the two graphs corresponding to the best design – quartic, we could define the respective domains that met the required objectives. For instance, in Figure 11, the shaded area consists of mixtures with clear point below 5 C. To see if viscosity and clear point objectives are compatible, we only have to superpose the two corresponding diagrams.

Illustration of a Restricted Design

For this purpose we worked on the formula based on the B system, using the same solubilizers as before: alcohol and urea. By changing only the active ingredients system, we do not modify the degree of complexity of the solubilization properties. Therefore, taking benefit of the previous series of experiments, we directly adjusted a quartic model. Results are presented in Table V.

Fit was very good for clear point. As for viscosity, a first trial of adjusting the corresponding quartic equation proved poor: we even obtained some negative values for predicted viscosities! This was due to the pure urea point, the viscosity of which is 7600 cps; that is, we had encountered an extreme value. The period of that experimentation was prior to our implementation of the pseudovertex technique: indeed we should have checked for extreme values before determination of the equations.

Two attempts were necessary to get the proper pseudovertex. With $h = 0.88$ viscosity was still too high, 4600 cps, but with $h = 0.82$, it goes down to 1440 cps, and this point was retained as the pseudovertex point.

In Figure 12 we can see that among the 15 points of the restricted design, the five points of the water-alcohol edge did not have to be rerun. For the nine new points, graphical interpolation was performed in the ternary diagram of the complete design. Four check points, labeled from *p* to *s*, assessed the validity of the transformed quartic. Results are presented in Tables VI and VII.

It is this series of experiments that first permitted us to develop the strategy of searching the best simplex design.

REFERENCES

1. Scheffé, H., J. R. Statist. Soc. B, 20:344 (1958).
2. Scheffé, H., Ibid. 25:235 (1963).
3. Gorman, J.W., and J.E. Hinman, Technometrics 4:463 (1962).
4. Fisher, R.A. and F. Yates, Statistical Tables, Sixth edition, Oliver and Boyd, 1967.
5. Cornell, J.A., Technical Report No. 5, University of Florida, 1970.
6. Box, G.E.P., and N.R. Draper, J. Amer. Statist. Ass., 54:622 (1959).
7. Lambrakis, D.P., J.R. Statist. Soc. B, 30:123 (1968).
8. Lambrakis, D.P., Ibid. 30:137 (1968).
9. Lambrakis, D.P., Ibid. 31:234 (1969).
10. Lambrakis, D.P., Ibid. 31:276 (1969).
11. Draper, N.R., and W. Lawrence, Ibid. 27:450 (1965).
12. Draper, N.R., and W. Lawrence, Ibid. 27:473 (1965).
13. Thomson, W.O., and R.H. Myers, Technometrics 10:739 (1968).
14. Kiefer, J., Ann. Math. Statist. 32:298 (1961).
15. Murty, J.S., and M.N. Das, Ann. Math. Statist 39:1517 (1968).
16. McLean, R.A., and V.L. Anderson, Technometrics 8:447 (1966).
17. Spendley, W., G.R. Hext and F.R. Himworth, Ibid. 4:441 (1962).

[Received October 28, 1971]