Applications of NonLinear Programming to Detergent Formulations[1]

EDMUND C. STEINLE, CHARLES D. HENDRIX and ROBERT R. FIELDS,
Union Carbide Corporation, Chemicals and Plastics, Research and Development Department, P.O. Box 8361, South Charleston, West Virginia 25303

ABSTRACT

Designed experiments have been used to secure orthogonal data in a typical light duty liquid detergent system. Regression analyses of the data have provided prediction equations relating composition to performance and physical properties for formulations in this system. The use of these equations in a nonlinear programming computer optimization program is described, illustrated by examples taken from a limited subsystem wherein two of the six components are held constant. Costs, types and uses of such optimization procedures are considered.

INTRODUCTION

This paper discusses one strategy for the optimization of detergent systems. This strategy involves a three-step sequence: (a) statistically designed experiments capable of estimating linear, quadratic and interactive effects; (b) stepwise multiple regression analyses; and (c) function optimization. Recognizing that one or more of these expressions may be unfamiliar to some, it is appropriate to begin with a brief description of each.

Statistically Designed Experiments

Multifactor experiments which have been arranged in certain optimum ways are called "statistically designed experiments". The objective here is to obtain the most information from a rather small number of experiments. Typically, one begins with a factorial design which has been augmented with certain points to permit the estimation of curvature. As a rule, about six experiments will be required for every variable in the design. Numerous papers have been published on the merits of such designed experiments. A good review is provided by Davies (1).

Stepwise Multiple Regression Analysis

Once the data have been obtained from a designed experiment, it is convenient to express the results as prediction equations. One way to obtain such equations is to submit the data to multiple regression analysis. Of the available methods for multiple regression analysis, one of the more popular is "stepwise analysis," which introduces terms into linear equations in a sequential fashion. A good "stepwise program" not only estimates regression coefficients, but also tests these coefficients for statistical significance and suggests which if any of the observations might be inconsistent with the remaining data. Substantially all stepwise regression programs are based on a scheme first proposed by Efroymson (2).

Function Optimization

Once prediction equations have been obtained, they serve to permit prediction of responses at combinations of conditions which have not yet been studied in the laboratory. This is accomplished by merely substituting into the equations the proposed conditions and computing the predicted result. One may have quite a few equations which represent a detergent system.

The wealth of information revealed by these equations is a source of some frustration, for one can spend many hours evaluating proposed detergent systems by laboring over a desk calculator. One possible solution to this problem is to convert the equations to contour plots with the aid of a digital computer. Contour plots for a light duty liquid system have been published by Huggins (3). Limitations of this approach stem from the sheer complexity of depicting, comprehending and searching the multi-dimensional space for each response to reach the best compromise in a many faceted problem. For instance, if a system composed of five components also exhibits five responses (including cost), one is forced to generate and manipulate perhaps 20 or 30 contour plots in order to find a good compromise. Even so, the derived compromise is somewhat arbitrary.

An alternative to contour plots has become available with the advent of computer search routines which can locate an optimum combination of independent variables. Such programs have been discussed in the operations research literature for a number of years. But only recently have they been structured for routine application by casual users of computers. A nontechnical review of some of these programs is provided by Barneson (4). Reports of application to specific fields of interest are beginning to appear, rubber formulation (5,6), for example.

The Strategy

The proposed strategy, then, is to obtain an adequate collection of multifactor data which covers essentially the full range of the independent variables, to fit the resulting data to multivariable equations which are linear in the coefficients, and to wed these prediction equations to a viable search program capable of optimizing a system subject to constraints.

Some homework must be done before undertaking such a venture. One begins by selecting a system of independent variables (in this instance, one must decide which of many possible components will be included in the proposed detergent system). These variables must be selected from prior knowledge of markets and applications. Further, the ranges over which these variables are to be varied must also be anticipated. These ranges are chosen from prior experimentation and screening experiments. Although not discussed here, we have found self-directing optimization to be very helpful in determining where to position an experimental design. A recent discussion of self-directing optimization is given by Himmelblau (7).

Applications

Any efficient system of investigation will be most useful in application to a new system. However, perturbations to a system such as price changes or modification of ingredients can provide new ground to cover. Such an application is illustrated here—the search to understand and utilize the capabilities of new light duty liquid systems based on completely biodegradable surfactants.

---

[1]One of seven papers being published from the Symposium, "Computer Systems and Applications in the Oil and Fat Industry" presented at the ISF-AOCS World Congress, Chicago, September 1970

TABLE I

Central Composite Experimental Design

No.	LAS[a]	AES	NON	CDA	SXS
1	15	2	8	2	2
2	30	2	8	2	5
3	15	8	8	2	5
4	30	8	8	2	2
5	15	2	2	2	5
6	30	2	2	2	2
7	15	8	2	2	2
8	30	8	2	2	5
9	15	2	8	5	5
10	30	2	8	5	2
11	15	8	8	5	2
12	30	8	8	5	5
13	15	2	2	5	2
14	30	2	2	5	5
15	15	8	2	5	5
16	30	8	2	5	2
17-20	22.5	5	5	3.5	3.5
21	35	5	5	3.5	3.5
22	10	5	5	3.5	3.5
23	22.5	0	5	3.5	3.5
24	22.5	10	5	3.5	3.5
25	22.5	5	0	3.5	3.5
26	22.5	5	10	3.5	3.5
27	22.5	5	5	0	3.5
28	22.5	5	5	7	3.5
29	22.5	5	5	3.5	0
30	22.5	5	5	3.5	7

[a]Abbreviations: LAS, linear alkylate sulfonate, type 11; AES, alcohol ethoxylate sulfate 25-L-3A; NON, Tergitol nonionic 15-S-9; CDA, coconut diethanolamide; and SXS, sodium xylene sulfonate.

PROCEDURES

We selected for our initial application of computer optimization the light duty liquid system using the following ingredients over the range (per cent actives) shown: linear alkylate sulfonate, type 11 (LAS), 10-35%; alcohol ethoxylate sulfate 25-L-3A (AES), 0-10%;Tergitol nonionic 15-S-9 (NON), 0-10%; coconut diethanolamide (CDA), 0-7%; sodium xylene sulfonate (SXS), 0-7%; and SD3A ethanol (EtOH), 0-7%. The first four surfactants were selected because of their known utility in light duty liquid performance; the last two ingredients are known hydrotropes and viscosity controllers.

Experimental Design

Initial experiments were done on the simpler five-component system not including ethanol. For a system of this size, satisfactory coverage of the variables and orthogonal data for analysis were obtained from the formulations shown in Table I. The first 16 experiments provide a two-level fractional factorial design in five variables, sufficient to show main effects and interactions. To this were added four center point replicates (numbers 17-20) and formulas testing outlying values of each variable (numbers 21-30). These 30 formulations provided a data base sufficient to show main effects, two-factor interactions, and curvature for each response tested.

Dependent Variables

Each of the formulations above was tested for those physical properties and performance responses considered important.

Spangler Dynamic Foam Test. Foam stability in the presence of soil ("Crisco") was determined using the Terg-O-Tometer test procedure published by W.G. Spangler (8). Test conditions were 0.0125% by weight detergent on an "as is" basis, 50 C initial bath temperature, varying water hardness and 75 rpm agitator speed.

Modified CSMA Dishwashing Test. Foam stability in the presence of soil (solid vegetable shortening–"Crisco", "Spry", etc.–flour, oleic acid mixture) was estimated by the CSMA Hand Dishwashing Procedure (9). Test conditions were 0.1% by weight detergent on an "as is" basis, 46 C initial bath temperature, and water of varying hardness.

Ross-Miles Flash Foam Test. The Ross-Miles foam test was conducted according to the ASTM D1173-53 procedure at 0.1% detergent concentration, 150 ppm water hardness and 50 C.

Viscosity. Viscosity was measured with a Cannon-Fenske viscometer tube in a constant temperature bath set at 25 C (ASTM D445).

Haze-Clear Points. Haze-clear points were determined as a measure of the low temperature solubility limits of the formulated detergent. A 20 ml sample of detergent was placed in a test tube, stirred gently with a thermometer to avoid air bubbles, and cooled slowly in a dry ice-acetone bath until clouding or haziness persisted. This temperature was recorded as the haze point. The sample was cooled further until completely cloudy and then allowed to warm, with stirring, until the detergent was clear. This temperature was recorded as the clear point. The average of these is recorded.

Regression Analysis

The data for each response were analyzed by stepwise multiple regression analysis to yield significant linear, curvilinear and interaction terms for the independent variables. The procedure for the regression analysis is similar to that given by Efroymson (2).

Transformation of the dependent variables was found

TABLE II

Analysis of Data From Designed Experiment

	Data		Fit of regression equations[a]		
Response	Mean	Standard deviation	Standard error	Multiple R^2	Overall F
Ross-Miles foam, mm, 150 ppm	145	29	9.3	.92	41
CSMA plates, 50 ppm	16.8	2.4	1.1	.82	50
Spangler, swatches					
50 ppm	11.6	2.8	0.9	.92	69
150 ppm	13.2	2.4	0.6	.94	65
Viscosity, cs	275	151	47	.93	39
Haze-clear point, C	3.1	12.5	4.7	.88	32

[a]Equations shown in Table III.

TABLE III

Prediction Equations When NON = 3.0

$CSMA_{50}$, plates	=	17.23 + .313 (LAS[a]-22.5) + .213 (AES-5) + .356 (CDA-3.5)
SP_{50}, swatches	=	10.31 + .268 (LAS-22.5) + .32 (AES-5) + .47 (CDA-3.5)
SP_{150}, swatches	=	12.6 + .235 (LAS-22.5) + .32 (AES-5) + .32 (CDA-3.5) -.067 (LAS-22.5)(CDA-3.5) - .0096 $(LAS-22.5)^2$
RM_{150}	=	163.3 + 1.79 (LAS-22.5) + 3.69 (AES-5) + 2.06 (CDA-3.5) -.246 (LAS-22.5)(AES-5) -1.25 (AES-5)(CDA-3.5) -1.15 $(AES-5)^2$
log VISC	=	2.48 + .0415 (LAS-22.5) + .0156 (AES-5) + .0566 (CDA-3.5) -.1047 (SXS-3.5) - .00174 (LAS-22.5)(AES-5) -.00591 (LAS-22.5)(CDA-3.5) + .0125 (LAS-22.5)(SXS-3.5) +.01 (AES-5)(SXS-3.5) + .0152 (CDA-3.5)(SXS-3.5) -.00122 $(LAS-22.5)^2$
log (HC + 50)	=	1.67 + .0057 (LAS-22.5) - .00394 (AES-5) - .0199 (SXS-3.5) -.00262 (LAS-22.5)(SXS-3.5) + .00105 $(LAS-22.5)^2$

[a]Abbreviations: see Table I and Table IV.

desirable for the viscosity and haze-clear point responses. For viscosity, the logarithm was used. For haze-clear point, best results were found when the logarithm of the value plus the constant 50 was used.

Optimization Program

The optimization program we use is called "Sidewinder" and was developed at Union Carbide by C.D. Hendrix. Although it is proprietary, we offer here a short description and classification so that the reader may estimate what can be accomplished with programs of its type.

The main program provides for the setting up and optimization of an experimental design in any number of independent variables. It is generally applicable to any optimization problem for which the necessary descriptive equations are available. Because these equations often contain nonlinear terms, the general program fits into the area of nonlinear programming. The program begins exploration of the territory at any point designated by the experimenter. Using the "Simplex" method of steepest ascent optimization, it does experiments, refers each of them to the objective subprogram for evaluation, and then ranks the experiments. Based on the ranking, new runs are calculated in the direction of better scores until no further progress is made. The program automatically expands and contracts the step size as it progresses to the optimum. At each point in the progress, the objective program is consulted for an evaluation of the formulations, scored according to overall worth or desirability. A description of the method and comparison with several other techniques is given by Barneson (4). The original scheme was proposed by Spendley, Hext and Himsworth (10) and modified by Nelder and Mead (11).

The Objective Subprogram

The objective subprogram is the responsibility and property of the user of "Sidewinder." It contains the predictive equations and other mathematical or logical statements which allow the overall worth of the individual trials made by "Sidewinder" to be calculated. As used in this light duty liquid investigation, it contains the following: (a) Means to encode the independent variable value into deviation from mean to fit the equations; (b) prediction equations for the performance tests and for physical properties; (c) straightforward objective elements such as cost equations; and (d) a discretely specified routine for calculating the score or worth of each trial.

Scoring

The problem of scoring is a general and important one. We must decide in advance how to rate any possible combination of properties, performance, cost, etc. This presupposes that we know the true worth of each increment of each response. The assignment of linear or nonlinear rating scales for all possible values of each response may be done so that the best response equals 1 and the worst equals 0. When these ratings are cross multiplied, the scores can range from 0 up to a perfect rating of 1. A discussion of this system is contained in the work by Derringer (5).

But we usually don't know the relative worth of the responses. One simple way to start is to cross-multiply all responses which we wish to be large by the reciprocal of responses we wish to minimize. This implies that a 10% gain in any response equals a 10% gain in any other response. Where the responses are not linear, this may not give the desired results. For example, CSMA plates/penny is not a

TABLE IV

Sensitivity Analysis, Viscosity = 200-250

Limits				Predicted performance					Formulation					
RM[a]	CSMA	SP	H/C	RM	CSMA	SP	H/C	VISC	LAS[a]	AES	NON	CDA	SXS	Cost
0	16	0	0	143	16	8.7	-1.3	250	22.2	3.2	3	1.3	2.9	4.70
160	16	0	---	160	16	8.8	-1.9	249	22.0	5.6	3	0.1	2.6	4.83
0	17	0	---	157	17	9.6	-3.1	250	25.4	4.3	3	0.7	5.1	5.39
160	17	0	0	160	17	10.3	-3.5	250	20.4	3.9	3	5.3	3.9	5.68
0	16	9	---	142	16	9	-1.4	250	21.7	3.0	3	2.0	2.9	4.80
160	16	9	0	160	16	9.1	-3.2	249	20.1	5.9	3	1.6	2.7	5.01
0	17	9	0	157	17	9.6	-3.1	250	25.4	4.3	3	0.7	5.1	5.39
160	17	9	---	160	17	10.3	-3.5	250	20.4	3.9	3	5.3	3.9	5.68

[a]Abbreviations: RM, Ross-Miles Flash Foam Test, ml; CSMA, dishwashing test, plates; SP, Spangler dynamic foam test, swatches; H/C, haze-clear point, C; VISC, viscosity, cs. See also Table I.

linear response; more than twice the cost is needed to double the plates. Such a score would lead only to lowest cost formulations. The price may be fixed at the desired level, but this may still lead to products in which some property is unacceptably low.

An easier problem is the duplication of a successful product performance spectrum at lowest possible cost. Many of our studies are guided in this fashion. The score is set equal to negative cost minus penalties assigned for violation of minimum performance thresholds or other constraints. No credit is given for exceeding the minimum performance. By assigning steep penalties to guide out of contrained areas, we channel the optimization into acceptable territory where lowest cost may be sought. This same method of penalty scoring is used to keep the independent variables out of forbidden territory, e.g., away from negative amounts of ingredients.

RESULTS

We have two areas of results to consider, first, results of the designed experiment and regression analyses, and second, the operation of our optimization program. Both of these tend to reveal proprietary information because they deal with the hidden strengths and weaknesses of a system which is competitive in the marketplace. For this reason we have simplified the system. Results are presented for the artificially restricted slice through six-dimensional space which involves only those formulations which contain 3% nonionic and no alcohol. This encompasses a useful and competitive area which illustrates the operation of the larger system.

Analysis of Data

The experimental design gave data whose means and standard deviations are shown in Table II. Also listed are the results of the multiple regression analyses showing the multiple R^2 (fraction of variance accounted for), residual standard error of prediction, and overall F levels achieved by the regression equations. You will note that the viscosity is the most difficult response to predict with accuracy. This is most likely because of higher order interactions. The equations, corrected to correspond to a constant nonionic level of 3%, are listed in Table III. Some equations are rather straightforward lists of main effects while others deal in curvilinear terms and interactions. The main effects of the nonionic have been absorbed into the main constant; nonionic interactions are absorbed in the constants of the other variables.

Illustrated Use of Optimization Program

We use "Sidewinder" on a time-shared computer system with remote terminal or on the larger batch-processing computer such as the IBM 360/65. A typical optimization on the time-shared system might take 10 min operator time and 1 min computer time. Scheme 1 illustrates the progress of such a problem. The underlined entries are supplied by the operator, i.e., low thresholds or limits, high limits, ingredient prices and starting position (high and low levels of each variable). The computer then prints its progress every 100 steps. The first line is the formula, LAS, AES, NON, CDA, SXS, EtOH. The second line lists the results predicted, Viscosity, Ross-Miles (150 ppm), CSMA plates (50 ppm), haze clear point, cost, Spangler swatches (150 ppm), and score.

```
6 VARIABLES, RESULTS LISTED EACH 100 CYCLES
  LOW LIMITS: VISC, RM150, CSMA50, SP50, LAS , AES, NON, CDA, SXS, ALC
              200    0      17      0    0    0    3    0    0    0

  HIGH LIMITS: VISC, H/C, COST
               250   50   10

  INGRED PRICES: LAS    AES    NON    CDA   SXS   ALC
                 .1365  .2041  .1425  .25   .085  .062

  STARTING LOWS AND HIGHS
             19 20   5. 5.5   3 3   2 2.1   2 2.1   0 0

EQUATIONS 9991 EXC SPANGLER 999B AND AHC, VISC 20ECS21B-CORR PROGRAM
20ECS33
VALUE STUDY 20ECS34

INCOMPLETE

 21.098     4.883     3.000    0.936     2.473   0.0
248.636   154.655    15.853   -2.002     4.748   8.691     -678.831

INCOMPLETE

 21.183     7.204     3.000    0.000     2.681   0.0
249.772   166.619    16.041   -3.257     5.017   9.016     -568.534

INCOMPLETE

 24.197     6.891     3.000    0.351     3.770   0.0
248.325   169.358    17.043   -2.334     5.545   9.889       -5.545

NORMAL STOP

 24.185     6.778     3.000    0.314     3.712   0.0
249.989   169.012    17.002   -2.206     5.506   9.832       -5.506
```

SCHEME 1

We have illustrated an example of slow progress. There was initial difficulty in finding a combination of actives yielding 17 plates without violating the upper viscosity limit of 250 cs. Hence, the very large negative scores in the first two entries. Once acceptable formulations were found yielding no penalties, the score became equal to negative cost and the cost was reduced. Further searching gave the lower cost formulation of Table IV, example 3, with reduced Ross-Miles performance.

The first successful optimization is exhilarating—what could be better than an optimum? But horizons are quickly broadened. The question becomes that of seeking the most appropriate answer out of hundreds of optimums which can be sought as the elements of the question are changed. Let us list some of the obvious variables: the prices charged for raw materials, the floors under acceptable performance, and the constraints placed on physical properties.

It becomes apparent that each optimization is an experiment in territory where the answers are unknown. There are many variables subject to manipulation. It is desirable to methodically explore the sensitivity of the system cost to changes in these variables.

A limited illustration of this approach is shown in Table IV. The area of interest is the cheapest formulation to yield at least 16 CSMA plates, 160 mm Ross-Miles foam, 9 Spangler swatches, and haze-clear point of 0 C or lower, with viscosity in the range of 200-250 cs. The group of eight experiments shown in Table IV is an experiment in four variables to show how changes in the Ross-Miles, CSMA, Spangler and haze-clear point demands affect the price.

The following information is apparent by simple inspection of the results: (a) In no case was the haze-clear point limiting. The upper limit on viscosity was constraining so that sufficient hydrotrope was necessary in each case to meet the haze-clear point requirement. (b) The 17th plate cost on the average about 0.7 cents. (c) The nine-swatch Spangler requirement costs nothing extra on the 17-plate level. At the 16-plate level it costs about 0.15 cents. (d) The average incremental cost of meeting the 160 Ross-Miles demand was 0.23, less at the 17-CSMA plate level and more at the 16-CSMA plate level. Such information shows some of the effects and interactions available within this system at the point of the investigation.

DISCUSSION

Rapid progress is being made in the ability to combine the strengths of experimental design, regression analysis and computer optimization programs in an overall strategy to meet the needs of product managers. The apparent strengths and weaknesses of light duty liquid systems in the marketplace will depend on the relative values placed on the many tests of performance and aesthetic qualities. By the study of optimized systems, managers of markets and products may explore the relative costs of each aspect of product acceptability. The understanding gained should aid in the design of products which combine the most useful and appealling features at lowest cost commensurate with internal profitability.

REFERENCES

1. Davies, O.L., editor, "The Design and Analysis of Industrial Experiments," Hafner Publishing Company, New York, 1960.
2. Efroymson, M.A., in "Mathematical Methods for Digital Computers," Edited by A. Ralston and H. Wilf, John Wiley & Sons, Inc., New York, 1960.
3. Huggins, J.M., Chem. Spec. Mfr. Assoc., Proc. Mid-Year Meet. 53:113-118 (1967).
4. Barneson, R.A., N.F. Brannock, J.G. Moore and C. Morris, Chem. Eng. 77(16):132 (1970).
5. Derringer, G.C., "A Powerful Digital Method for Rubber Compound Optimization," presented at the Meeting of the Division of Rubber Chemistry, A.C.S., Washington, D.C., May 1970.
6. Davidson, A.R., Rubber Age 102(1):83 (1970).
7. Himmelblau, D.M., "Process Analysis by Statistical Methods," John Wiley & Sons, Inc., New York, 1970.
8. Spangler, W.G., JAOCS 41:300 (1964).
9. Harris, J.C., Chem. Spec. Mfr. Ass. Proc. Annu. Meet. 45:190-192 (1958).
10. Spendley, W., G.R. Hext and F.R. Himsworth, Technometrics 4(4):441 (1962).
11. Nelder, A.J., and R. Mead, The Comput. J. 7:308 (1965).

[Received October 1, 1970]