1986

The Use of Scheffe-Equivalent Equations to Predict Physical Properties of Neoprene

Cheng Y. Loh
University of Central Florida

THE USE OF SCHEFFE-EQUIVALENT EQUATIONS TO PREDICT PHYSICAL PROPERTIES OF NEOPRENE

BY

CHENG YEE LOH

B.A., Randolph-Macon Woman's College, 1982

RESEARCH REPORT

Submitted in partial fulfillment of the requirements
for the Master of Science degree in Industrial Chemistry in the
Graduate Studies Program of the College of Arts and Sciences
University of Central Florida
Orlando, Florida

Spring Term
1986

ABSTRACT

The goal of this study is to find a more organized and directed approach to build models for mixture systems. An attempt is made to generate and then compare Scheffe (mixture) models with those generated by McGee using the 'conventional' method for neoprene data. The models are judged on their ability to predict physical properties of neoprene by comparing the following: predicted and actual values by inspection; the calculated % error of prediction; the squared multiple correlation coefficients; adjusted squared multiple correlation coefficients; the Fisher statistic and significance probability. Scheffe models do not have an intercept term and test statistics which appear on the computer printout are inflated. Pseudocomponents and Scheffe-equivalent models are procedures used to obtain accurate test statistics to describe the selected Scheffe models. The effectiveness of these two procedures is evaluated. Results indicate that Scheffe models are better predictors for the physical properties of neoprene than those generated by McGee using the 'conventional' method in 1980. Scheffe-equivalent equations are found to be more reliable than pseudocomponents for generating accurate test statistics to describe the selected Scheffe models.

ACKNOWLEDGEMENTS

I would like to express my sincere appreciation to Dr. William McGee for his unlimited patience, guidance and help throughout my graduate research. I also wish to extend my thanks to Dr. Chris Clausen and Dr. Guy Mattson for their time and effort as members of my graduate committee, and to Mrs. Claire McClure for her help in preparing this report.

Finally I would like to express my gratitude to my husband Richard, and my father, Mr. Loh Hin Foon for their unwavering support and encouragement throughout my academic years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

Regression analysis has been used to generate model equations which are used to describe a wide variety of chemical and physical data of synthetic rubbers since 1980. At that time, McGee[1] found strong correlations for physical properties ands weak correlations for acoustic properties. The ability to construct model equations was limited by the knowledge of statistical methods available and variables in the model equations were selected empirically. The goal of this study is to develop a more organized and directed approach toward model equation building and testing. In this study an attempt will be made to generate and then compare Scheffe (mixture) models with those generated by the 'conventional' way using neoprene data.

Regression analysis is the fitting of an equation to data. The equation is usually obtained using a least squares fit which fits the 'best' straight line through the data points by minimizing the difference between the actual (the data point) and the predicted value (the line). In order to begin the least squares fit, variables to be included in the equation must first be selected/specified. The two general ways to do this are the conventional method and the mixture (Scheffe)[2].

In the conventional method, new variables are obtained by generating numerous functions from the original variables. The functions together with the original variables then undergo Forward, Backward or Stepwise selection (pg. 11) to filter out the unimportant

variables and arrive at the final equation. In the Scheffe method, there is an established scheme to decide which variables and functions to include in the equation and no filtering is necessary. The question here is not the choice of variables but in fact the order of the desired equation e.g., linear, quadratic, special or full cubic Scheffe equations. Although Scheffe equations are quite popularly used to describe mixture systems, test statistics (R^2, R^2_{adj}, F, P>F) generated in the computer output are inaccurate.[3-5] Pseudocomponents (pg. 21) and Scheffe-equivalent (pg. 24) procedures are two procedures to generate accurate test statistics to describe the Scheffe equation of interest.

Model equations generated by McGee will be used to represent models generated the 'conventional' way. Scheffe models represent mixture models. All regression will be performed using SAS (Statistical Analysis System). The models generated will be judged on their ability to predict nine physical properties of cured neoprene rubber. This is done by comparing predicted values with the actual by inspection and the calculated % error of prediction. Other criteria for judging models are the squared multiple correlation coefficient (R^2) and the adjusted squared multiple coefficient (R^2_{adj}), the Fisher statistic (F value), significance probability (P>F) for the model.

Figure 1 summarizes the relationship between regression analysis and variable selection.

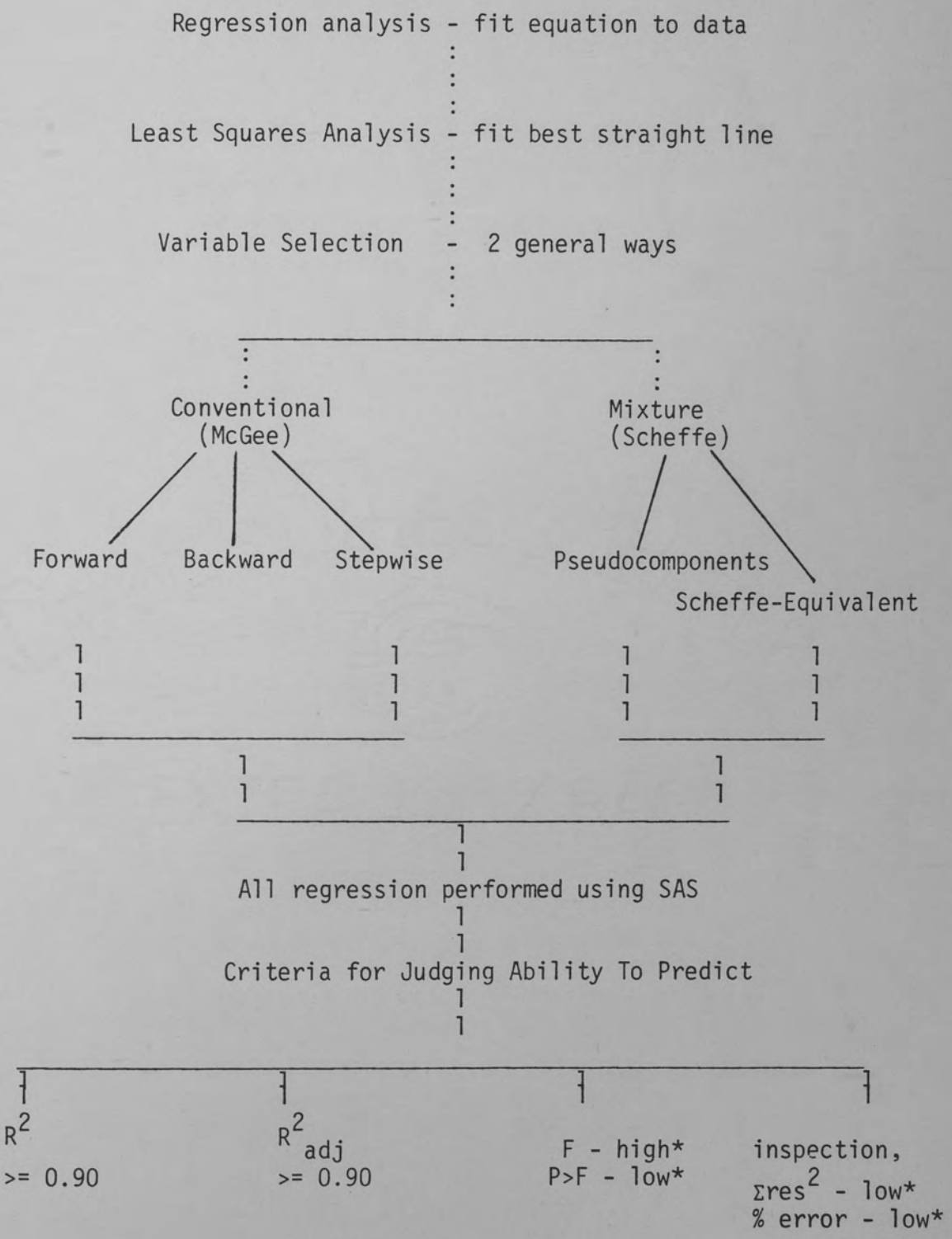

Figure 1. Relationship between regression analysis and variable selection.

* high and low are relative to the model of comparison.

MODEL BUILDING

Regression analysis is the fitting of an equation to a set of values. The equation predicts a response variable Y from a function of regressor variables X and parameters. One method used to estimate the parameters is called the least squares fit.

Least Squares Analysis

Credit is usually given to C.F. Gauss (about 1809) for the method of least squares. This method fits the 'best' straight line to the data set by minimizing the sum of squared residuals (the difference between the actual and predicted response). The criterion is:

$$(Y_i - \hat{Y})^2 = \text{minimum}$$

where Y_i is the actual response and $\hat{Y}$ is the predicted response. The smaller the deviations of the observed values from the fitted line, the better the fit.

Assumptions for Least Squares Analysis

Some assumptions are made about the data when this method is used:[6]

1. The probability distributions of Y for a given X have the same variance σ^2 for all X_i (e.g., if it takes person A 5 ± 1 min to run a mile and it takes person B 7 min, B's running time must then be 7 ± 1 min).
2. The means of Y_i, μ_i lie on a straight line known as the true regression line.

$$\mu_i = b_0 + b_i x_i$$

The population parameters b_o and b_i specify the line and are to be estimated from sample information (e.g., each response has a distribution and the means of the responses will lie on the best straight line). The intercept and coefficients are estimated from the data.

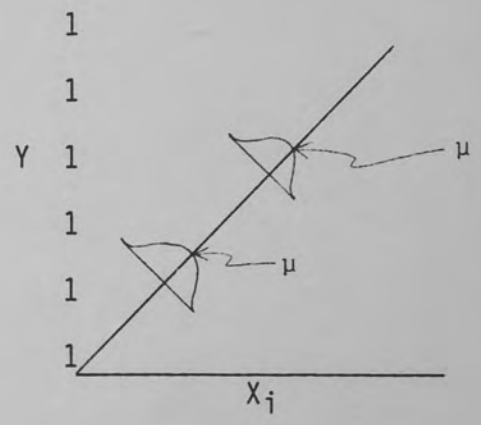

Figure 2. Illustration for assumption two.

3. The random variables Y_i are statistically independent with variance equal to σ^2 (e g., the time it takes A to run 1 mile will not affect the time it takes B to do the same).

4. The deviation of Y_i from its expected value is the error term E_i so

$$Y_i = b_o + b_i X_i + E_i$$

where E_i are independent random variables with mean equal to 0 and variance equal to σ^2. The distributions of Y and E are identical except their means differ. The distribution of E is just the distribution of Y translated onto a zero mean (e.g., the difference between the actual and the predicted value is the error term E; the errors are independent of each other). Some of the errors are positive while some are negative; and the average error is zero.

In the 150 years that followed Gauss' invention of least squares analysis, many developments occurred in the theory of linear statistical inference. However, regression methodology did not change much due to the lack of high speed computing equipment.[7]

By the early 1960s the examination of residuals became part of regression analysis when some of the early computer programs had the option of computing residuals. The residual is the difference between the observed and predicted values. The purpose of residual analysis is to assess the appropriateness of a model in terms of the behavior of the set of residuals. The most direct and revealing way to examine residuals is to use a scatter diagram. Residuals are plotted against predicted values with the mean ordinate being zero. Each observation is represented by a point. If the model is good, the points should be a random scatter with approximately half of them being evenly distributed above and below the zero reference line.

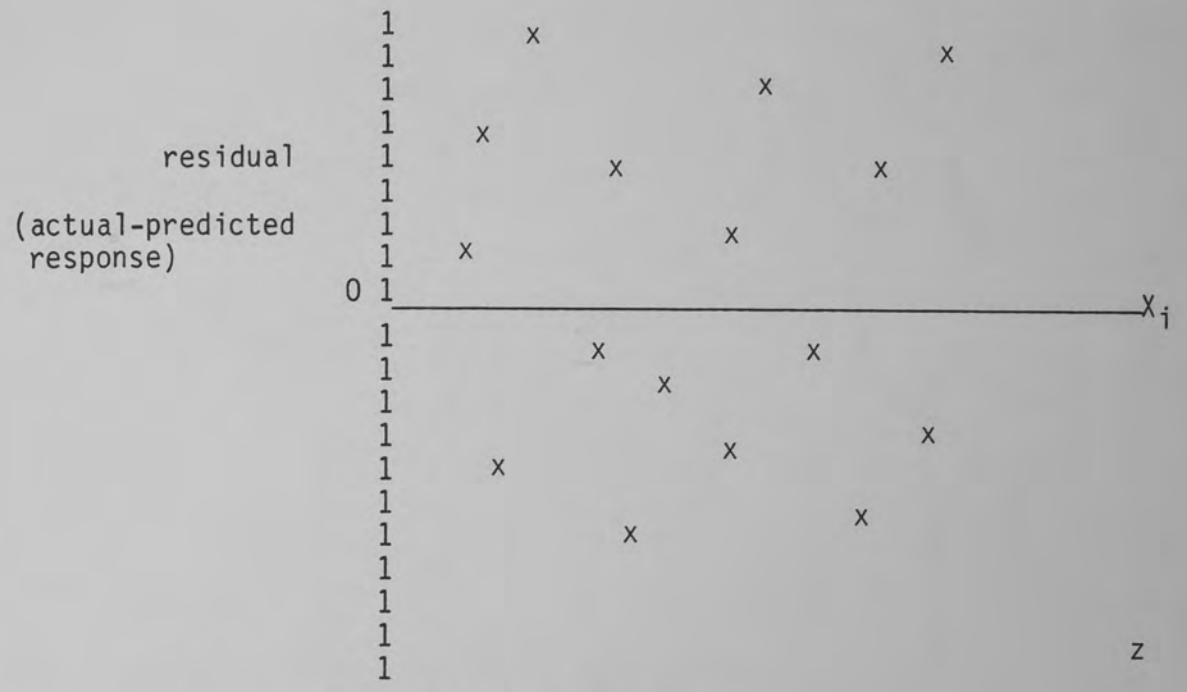

Figure 3. Scatter diagram (Z is an outlier).

Outliers

An outlier can be seen as an isolated point (z) on the scatter diagram in Fig. 3. Outliers can lie as far as 3 or more standard deviations from the mean of the residuals. The least squares estimates can be affected by such extreme values so it is important to evaluate whether the outlier should be discarded. The decision can only be made after a thorough evaluation of the experimental conditions, data collection process and data. The data point is discarded if it is due to mistakes in recording or instrument malfunction (determinate error). Sources of outliers could be misreading the scale; errors in conducting the experiment; or the correct observation of unusual phenomena or samples with longer-tailed distributions of errors (Fig. 4). If error in conducting the experiment is suspected, the observation is checked before discarding. If the outliers are believed to be an observation of 'irregular' or 'regular but long-tailed' phenomena

(Fig. 4) it may be modified by Winsorization to be given reduced weight. Winsorization involves replacing the suspect value with the nearest non-suspect value.

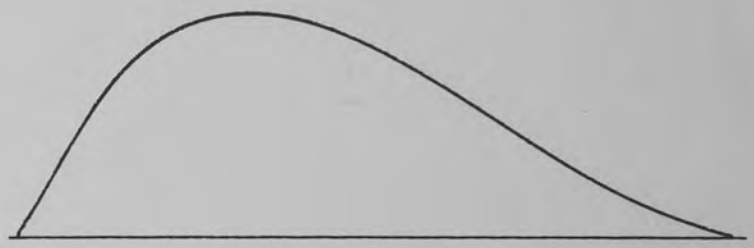

Figure 4. Long-tailed or skewed distribution of data.

The Price of Rejecting Data

Sometimes definite rejection rules are used for outliers. Treating any rejected observation as missing (eliminating the data point) and applying the conventional analysis to the remaining observations modifies simple least squares analysis by giving rejected observations zero weight. Discarding the data point protects against gross errors but the rule can occasionally lead to rejection of good observations which satisfy the ideal conditions. Fewer data points result in an increase of the average error variance of the parameter estimates which might be regarded as insurance against bad observations.

Variable Selection

The normal starting point when building models is variable selection. There are two general ways of variable selection: (1) Conventional method, and (2) Mixture (Scheffe) method. In conventional model building it is a common practice to consider a large number of variables or functions of variables to be included in the data set for generating a model. A subset of variables which appear most

relevant by intuition is then selected and a model specified based on the selected terms. The initial inclusion of a large number of variables is justified since omission of essential variables may produce biased estimates while the inclusion of extraneous variables does not. Three procedures which have been widely used for testing for the inclusion of a variable in the model equation are: 1. Backward Elimination (step down); 2. Forward Selection (step up); and 3. Stepwise Selection.

Backward Elimination

This is a one-at-a-time elimination or step down procedure. An initial analysis is peformed using all variables from the initial set for examination. The t statistic for significance of each coefficient or the partial F statistic for each variable is calculated and the variable associated with the least significant value is deleted. Significance statistics are again computed for the new regression until all remaining coefficients are statistically significant at some predetermined probability level. Once a variable has been deleted it may not re-enter.

Forward Selection

This is sometimes called a step up procedure. In contrast to backward elimination, this procedure begins with the 'best' one-variable equation, adding one variable at a time. The independent variable for the best one-variable equation is obtained by choosing the one most highly correlated with the dependent variable. If the F or t statistic is not significant based on a predetermined probability level, the

procedure is stopped and it is concluded that no independent variables are important predictors. At each step, F or t values and significance probabilities (to be discussed later) are determined for those variables not already in the model. The variable with the largest F or t value is added to the model if it is statistically significant. The process is terminated if the largest F or t is not significant at some predetermined test level. Once added selected variables may not be dropped and can therefore lead to nonoptimal variable selection.

Stepwise Selection

This procedure is a refinement of the forward selection procedure and overcomes some of the major deficiencies of the forward and backward methods. Many consider it the 'best' of the three procedures but some still prefer backward elimination which allows the user to see the order of elimination in getting to the final equation. The final results of stepwise selection are dependent on choices of significance levels for the forward selection and backward elimination.

In stepwise, before the determination of the next variable to be added to the equation, significance statistics of the already chosen variables are examined to see if any can be eliminated. This procedure may be described as a step up procedure with a step down adjustment, and is terminated when neither forward selection or backward elimination is allowed according to chosen significance levels.

Advantages and Disadvantages of Conventional Model Building.

Backward Elimination

Advantage:

This method allows the user to see the equation starting with all the variables progress through the various steps of elimination until the final equation is obtained. This is particularly useful if the user decides to choose an intermediate rather than the final equation due to special insight/knowledge relating to the problem.

Disadvantage:

Once a variable is deleted it cannot reenter. This might prevent the finding of an optimum combination of variables when there is a high degree of collinearity between the variables. When there is a high degree of collinearity, a deleted variable may become important once another variable is deleted.

Forward Selection

Advantage:

Collinearity does not affect forward selection. Although the XX' matrix (see Appendix C for matrices) gets larger as more variables are added to the model, the XX' matrix containing all the variables will never be inverted to obtain the coefficient estimates since the procedure would have terminated before all the variables are included in the model.

Disadvantages:

Once a variable is selected it cannot be dropped. It is not possible to investigate the effectiveness of an equation containing all

the variables because the procedure would usually have terminated before the inclusion of all the variables.

Stepwise Selection

Advantages:

It combines the forward and backward procedures. Each time a variable is included in the equation, new significance statistics are computed for each variable already in the equation to see if any can be dropped before adding another. It is not affected by collinearity for the same reason as in forward selection.

Disadvantage:

If only a few variables are acceptable in the model according to the chosen significance, the user will not have a chance to see other equations that might be more appropriate in describing his problem based on his own knowledge of the subject.

Test Statistics Used to Judge Models

The following are test statistics used to judge predictor models.

1. R^2, squared multiple correlation coefficient. This statistic measures how efficiently the variation in the dependent variable is accounted for by the model. Values are from 0 to 1.0. There is no actual cutoff point for the acceptability of a model but in this study the preferred value should be greater than or equal to 0.90.
2. R^2_{adj}, adjusted squared multiple correlation coefficient. This statistic is similar to R^2. It adjusts R^2 downward to take into account the large number of variables (k) in the equation when k is approximately equal to the number of observations, n. If n <= (k+1), a perfect fit will be obtained no matter how ludicrous the

hypothesized relationship. When n is much larger than k, R^2 and R^2_{adj} will be almost equal.

3. F, Fisher statistic. F measures how well the model as a whole accounts for the behavior of the independent variable. Model is significant if P>F, the significance probability associated with the F value, is small.

4. P>F, significance probability associated with F value. (1 - P>F) describes the chances that user is right in assuming the model is significant.

5. Σres^2, sum of squared residuals. This is the sum of the squares of error for each prediction. This statistic should be used to calculate the % error of prediction which gives a clearer picture of how much error there is in the predicted value relative to the mean of the actual values.

6. % error of prediction. Computed by taking the square root of the average squared residual, dividing it by the mean of the actual values, $\bar{Y}$, and multiplying the results by 100.

$$(\Sigma res^2/n)^{\frac{1}{2}}/\bar{Y} \times 100$$

7. VIF, Variance Inflation Factor. The VIF is used to detect collinearity. It tells the factor by which the variance of the coefficient estimates are inflated. Some[8] prefer the VIF to be less than 10 while others limit it to 100. High VIF values might lead to the coefficients having incorrect signs and thus may be misleading in interpreting the coefficient estimates. Collinearity does not affect prediction. It is not always possible to obtain low VIF values. No limit will be used in this study but instead the values will be evaluated relative to the model under study.

More About Collinearity

Collinearity occurs when two variables measure nearly the same thing. There exists a nearly linear relation among the predictor variables. This leads to large variance and hence, broad confidence levels making it difficult to establish that an individual regressor influences the response, Y. When two regressors are nearly the same, the influence of one of them on Y might erroneously be attributed to the other.

Correlation analysis (inspection of the correlation table) can be used to confirm collinearity. Two predictor (X) variables that are highly correlated to each other suggests collinearity. However the variance inflation factor (VIF) is the statistic most often used to detect collinearity because it gives information on the extent of the problem.[9]

$$VIF = (1-R_i^2)^{-1}$$

where R_i is the multiple correlation of the regression of X_i on all other X_j ($i \neq j$). The variance of the i^{th} regression coefficient b_i in a model with a constant term b_o is:

$$V(b_i) = S^2 VIF_i/(n-1)S^2_i$$

where S^2 is the observation error variance and

n is the number of observations and S_i is the variance of X_i.

Collinearity may be reduced by centering (the mean is subtracted from each data point) or standardizing (centered value/standard deviation) the data. Centering and standardizing merely disguise the problem because a new set of numbers is obtained by manipulating the data. Relationships obtained will hold only with these numbers and are not in terms of the original data.

Recent Developments in Model Building

In the past 15 years an interest in mixture* models has grown within the statistical community. Scheffe[2] first published his paper on mixture theory which describes the simplex experimental design and the derivation of mixture models in 1958. The paper did not receive much attention until the mid-sixties when scientists and statisticians began to apply his theory to the data of mixtures.

Scheffe's equations are widely accepted as a means to describe mixture systems. The Scheffe or mixture model equation does not have an intercept term. Even though most statistical packages have the capability to fit least squares to an equation by suppressing the intercept, test statistics printed in the output are inaccurate. Two procedures have been developed in the attempt to obtain more accurate test statistics for the Scheffe model. The first procedure was developed by Kurotori[10] in 1966 and involves the use of pseudocomponents.[11] The second procedure is relatively new and was developed by Snee[8] in 1982. It involves the use of Scheffe-equivalent equations. Both procedures will be described in detail in this report. Scheffe, pseudocomponents and Scheffe-equivalent equations all give the same predictions because the latter two are derived from the Scheffe form. The development of Scheffe models for mixture systems is discussed in the following section.

*Mixtures are systems in which the response to a blend depends only on the relative proportions of the ingredients present in the blend, e.g., cakes, rubber, soap formulation. The ingredients when expressed as proportions based on weight or volume will always sum up to one.

Development of Scheffe Forms

In mixture problems such as a compounded elastomer, the response depends only on the relative proportions of the ingredients present in the blend and not on the total weight or volume of the blend. The controllable variables are the proportions or fractions of the mixture. These proportions are non-negative and when expressed as fractions of the components in the mixture, they sum up to one.

$$0 <= x_i <= 1$$

$$\sum_{i=1}^{q} x_i = 1$$

where x_i is the i^{th} component

and q is the total number of components.

Equations obtained using least squares analysis contain an intercept term. In theory, the intercept is the expected response if all components were set to zero. In reality, this is not possible for if all components were zero, there would be no response. It is desired that the response be expressed as a function of the components involved using a model with no intercept term.

The response, Y is often approximated over the experimental range by a polynomial derived from a second-order Taylor series. For two variables, X_1 and X_2 the model will be:

$$Y = b_o + b_1X_1 + b_2X_2 + b_{12}X_1X_2 + b_{11}X_1^2 + b_{22}X_2^2$$

The sum of the proportions x_i of the components must be 1 in a mixture. Consequently, for the two component case:

$$X_1^2 = X_1X_1 = X_1(1-X_2) = X_1 - X_1X_2$$

$$X_2^2 = X_2X_2 = X_2(1-X_1) = X_2 - X_1X_2$$

so

$$Y = b_o(X_1 + X_2) + b_1X_1 + b_2X_2 + b_{12}X_1X_2$$

$$+ b_{11}(X_1 - X_1X_2) + b_{22}(X_2 - X_1X_2)$$

$$= b_1^*X_1 + b_2^*X_2 + b_{12}^*X_1X_2$$

where $b_1^* = b_o + b_1 + b_{11}$

$b_2^* = b_o + b_2 + b_{22}$

$b_{12}^* = b_{12} - b_{11} - b_{22}$

Similarly the response for a three variable quadratic is:

$$Y = b_1X_1 + b_2X_2 + b_3X_3 + b_{12}X_1X_2 + b_{13}X_1X_3 + b_{23}X_2X_3$$

The cubic forms describe a third order polynomial derived from the Taylor series. The special cubic form takes into account the inclusion of each variable raised to the third power and also the crossproducts obtained by multiplying X_1, X_2 and X_3 together ($X_1X_2X_3$). The full cubic form includes the γ terms. These are coefficients that correspond to the crossproducts of different combinations of two variables which are raised to the total power of three, taking into account terms such as $X_1^2X_2$, $X_1^2X_3$, $X_2^2X_1$, $X_2^2X_3$, $X_3^2X_1$ and $X_3^2X_2$ of the third order Taylor series polynomial. The response for a three variable special cubic and full cubic form, respectively, are:

$$Y = b_1X_1 + b_2X_2 + b_3X_3 + b_{12}X_1X_2 + b_{13}X_1X_3 + b_{23}X_2X_3 + b_{123}X_1X_2X_3$$

$$Y = b_1X_1 + b_2X_2 + b_3X_3 + b_{12}X_1X_2 + b_{13}X_1X_3 + b_{23}X_2X_3 + b_{123}X_1X_2X_3$$
$$+ \gamma_{12}X_1X_2(X_1-X_2) + \gamma_{13}X_1X_3(X_1-X_3) + \gamma_{23}X_2X_3(X_2-X_3).$$

The mixture model thus has fewer coefficients than the usual second-order polynomial and has no constant term. In general, Scheffe suggested the following canonical forms for mixture models:

Linear: $$Y = \sum_{i=1}^{q} b_iX_i$$

Quadratic: $$Y = \sum_{i=1}^{q} b_iX_i + \sum_{1<=i<j}^{q} b_{ij}X_iX_j$$

Special Cubic:

$$Y = \sum_{i=1}^{q} b_i X_i + \sum_{1<=i<j}^{q} b_{ij} X_i X_j + \sum_{1<=i<j<k}^{q} b_{ijk} X_i X_j X_k$$

Full Cubic:

$$Y = \sum_{i=1}^{q} b_i X_i + \sum_{1<=i<j}^{q} b_{ij} X_i X_j + \sum_{1<=i<j<k}^{q} b_{ijk} X_i X_j X_k + \sum_{1<=i<j}^{q} \gamma_{ij} X_i X_j (X_i - X_j)$$

The linear coefficients b_i are the average responses associated with each of the components in the mixture if the independent variables are centered ($x_i = 1$, $x_j = 0$, $i \neq j$). If linear coefficients are used (e.g., $X_1, \ldots X_n$), the components act additively so that the response surface is a plane. Cubic and higher-order terms describe the deviations of the response surface from a plane. The number of coefficients in the various models as a function of components in the mixture, is shown in Table I. Application of the canonical forms to a three component example is shown below the table. The number of coefficients for the Scheffe models is less than that for a 'conventional' polynomial with the same number of components and of the same order.

Table I

Number of coefficients in Scheffe models as a function of number of components.[12]

Number of Components	Model			
	Linear	Quadratic	Special Cubic	Full Cubic
2	2	3	-	-
3	3	6	7	10
4	4	10	14	20
5	5	15	25	35
6	6	21	41	56
7	7	28	63	84
8	8	36	92	120
9	9	45	129	165
10	10	55	175	220

Three component example:

Let the mixture be a sponge cake made only of eggs (E), flour (F) and sugar (S). The texture (T) of the cake may be expressed as follows:

Linear: $T = b_1E + b_2F + b_3S$

Quadratic: $T = b_1E + b_2F + b_3S + b_{12}EF + b_{13}ES + b_{23}FS$

Special cubic: $T = b_1E + b_2F + b_3S + b_{12}EF + b_{13}ES + b_{23}FS + b_{123}EFS$

Full cubic: $T = b_1E + b_2F + b_3S + b_{12}EF + b_{13}ES + b_{23}FS + b_{123}EFS + \gamma_{12}EF(E-F) + \gamma_{13}ES(E-S) + \gamma_{23}FS(F-S)$

Advantages and Disadvantages of Scheffe Models

Advantages:

Scheffe models are more appropriate for mixture systems since the sum of the proportions of components adds up to a constant of one. Scheffe has established a scheme for the user to follow in building models, i.e., linear, quadratic, special, and full cubic forms. These

models contain fewer terms than a polynomial of the same order. Most statistical packages now have the capability to fit least squares while suppressing the intercept.

Disadvantages:

Although statistical packages can fit least squares without an intercept, test statistics printed in the output are inaccurate[3-5] due to computer roundoff errors (ill-conditioning) although the predictor equations are good. By forcing all the proportions to sum to one, additional collinearity may be induced. Collinearity may result in models that are not full-rank, i.e., there is no unique solution but rather many equally possible equations.

From Scheffe to Scheffe-Equivalent Models

Although statistical packages can fit least squares without the intercept for the Scheffe equations, test statistics printed in the output are inaccurate (see Table VII). Attempts have been made to overcome this problem. Kurotori[10] proposed the use of pseudocomponents in the late sixties and Snee[8] later proposed the Scheffe equivalent model. Figure 5 shows the relationship between these two models and the Scheffe model. Both pseudocomponents and Scheffe-equivalent models will be examined in detail in the following pages.

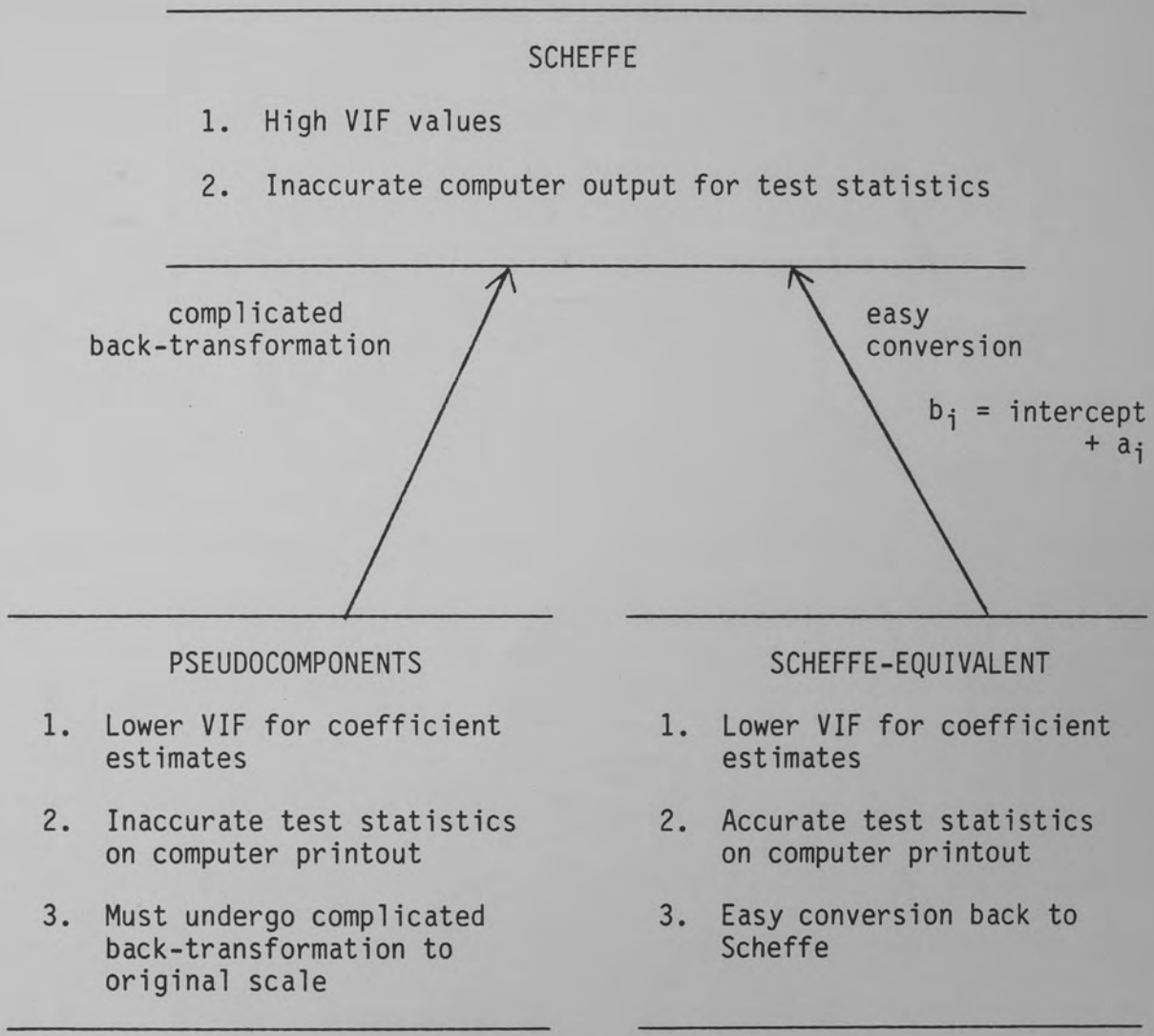

Figure 5. The search for true test statistics for the Scheffe Model

a_i - coefficient of linear term X_i in Scheffe-equivalent equation.

b_i - coefficient of linear term X_i in Scheffe equation.

Predicted values obtained by all three methods are the same.

Pseudocomponents and the Simplex

In Scheffe's paper[2] on mixture theory he proposed the simplex design. This is an experimental design that selects a set of points in the mixture space at which data are gathered to fit the Scheffe/mixture equations. The simplex is a figure to describe the experimental space

of a mixture system. It has (q-1) dimensions where q is the number of components in the mixture. For example, a three-component mixture will be described by a triangle (a two-dimensional figure) and a four-component mixture will be described as a tetrahedron (a three-dimensional figure).

Illustration of the Relationship Between Pseudocomponents and the Simplex

If a sponge cake is made from only eggs (E), flour (F) and sugar (S), its mixture space can be described by the equilateral triangle in Fig. 6. It is impossible to make a cake that consists of 0% or 100% of any one component. Restraints exist on the composition of the cake. Assuming that the restraints on the proportions (by weight) are:

$0.33 <= E <= 0.50 \qquad 0.33 <= F <= 0.75 \qquad 0.20 <= S <= 0.40$

The actual experimental area is only the small shaded area within the simplex in Fig. 6. This region is redefined by a mathematical transformation and the new numbers obtained are called pseudo-componments. Pseudocomponents 'magnify' the actual experimental region by changing the coordinates of the points to bring the origin near the observations in the mixture space and is done as follows.

$$E_p = (E - \text{lower bound of } E)/L$$

$$F_p = (F - \text{lower bound of } F)/L$$

$$S_p = (S - \text{lower bound of } S)/L$$

where L = 1 - sum of lower bounds of E, F and S

and the subscript p stands for pseudocomponent.

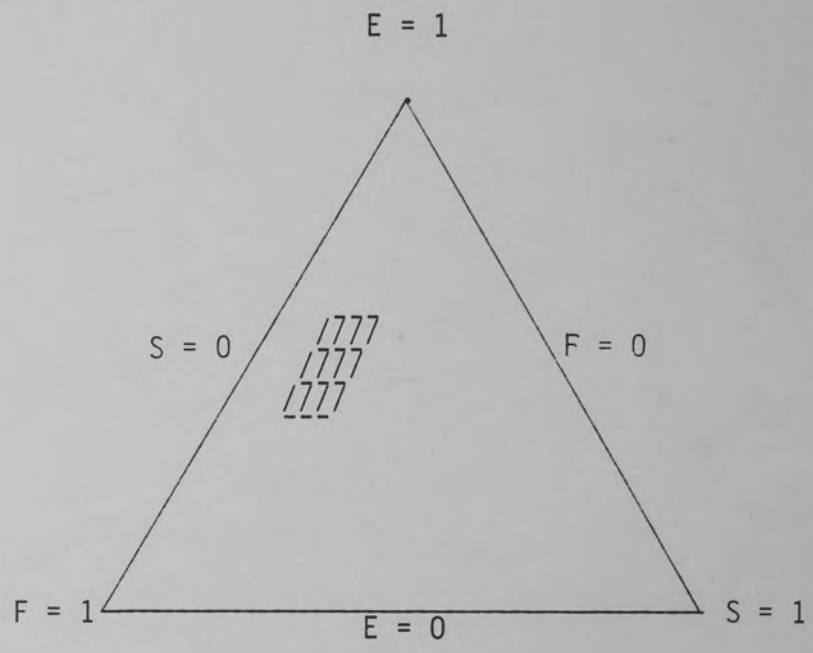

Figure 6. Simplex for the cake model

Restraints: $0.33 <= E <= 0.50$

$0.33 <= F <= 0.75$

$0.20 <= S <= 0.40$

E_p = (E - lower bound of E)/L

F_p = (F - lower bound of F)/L

S_p = (S - lower bound of S)/L

where L = 1 - sum of lower bounds of E, F and S and the subscript p stands for pseudocomponent.

Advantages and Disadvantages of Pseudocomponents

Advantage:

Lower VIF[13] values are obtained for the coefficients of the equation using transformed data.

Disadvantages:

The equation obtained is not expressed in terms of the original data. To get it in terms of the original values the user must perform a

complicated back-transformation. Pseudocomponents do not give accurate values for the test statistics of the Scheffe model.

Development of the Scheffe-Equivalent Model

Snee introduced the Scheffe-equivalent model. This model does not require additional data manipulation beyond transforming the data. The proportions of the components sum to one, i.e., to mixture form. It gives the same prediction as the Scheffe model. Using a three variable linear model as an example, the Scheffe form is:

$$Y = b_1X_1 + b_2X_2 + b_3X_3$$

The Scheffe-equivalent form is obtained by omitting one of the linear terms e.g., X_3 so:

$$Y = a_o + a_1X_1 + a_2X_2$$

The equivalence between the two equations can easily be verified by multiplying the intercept of the latter a_o by $(X_1 + X_2 + X_3)$ i.e., by one.

$$\begin{aligned} Y &= a_o(X_1 + X_2 + X_3) + a_1X_1 + a_2X_1 + a_2X_2 \\ &= a_oX_1 + a_oX_2 + a_oX_3 + a_1X_1 + a_2X_2 \\ &= (a_o + a_1)X_1 + (a_o + a_2)X_2 + a_oX_3 \end{aligned}$$

so

$$\begin{aligned} b_1 &= a_o + a_1 \\ b_2 &= a_o + a_2 \\ b_3 &= a_o \end{aligned}$$

To illustrate further using a three variable quadratic model, the Scheffe form is:

$$Y = b_1X_1 + b_2X_2 + b_3X_3 + b_{12}X_1X_2 + b_{13}X_1X_3 + b_{23}X_2X_3$$

The Scheffe-equivalent form with X_3 omitted from the linear portion of

the equation is therefore:

$$Y = a_o + a_1X_1 + a_2X_2 + a_{12}X_1X_2 + a_{13}X_1X_3 + a_{23}X_2X_3$$

To confirm equivalence, a_o is again multiplied by $(X_1 + X_2 + X_3)$ i.e., by 1.0.

$$Y = a_o(X_1 + X_2 + X_3) + a_1X_1 + a_2X_2 + a_{12}X_1X_2 + a_{13}X_1X_3 + a_{23}X_2X_3$$

$$= (a_o + a_1)X_1 + (a_o + a_2)X_2 + a_oX_3 + a_{12}X_1X_2 + a_{13}X_1X_3 + a_{23}X_2X_3$$

where $b_1 = a_o + a_1$ $\qquad$ $b_{12} = a_{12}$

$b_2 = a_o + a_2$ $\qquad$ $b_{13} = a_{13}$

$b_3 = a_o$ $\qquad$ $b_{23} = a_{23}$

Advantages and Disadvantages of Scheffe-Equivalent Models

Advantages:

This equivalent form is easily obtained by omitting one component from the linear portion of the corresponding Scheffe equation. Predictions are the same as when using the Scheffe equation since the two equations are equivalent. Accurate test statistics are obtained for describing the corresponding Scheffe equation in one step. The intercept is justified since this form was derived from the Scheffe equations. However, if the user prefers to have a final equation without the intercept, the intercept can easily be eliminated by multiplying by one and regrouping the coefficients to form the corresponding Scheffe equation.

Disadvantage:

This equivalent form has an intercept which might displease the user who is accustomed to seeing an equation without an intercept. However, this is easily remedied by multiplying the intercept by one and converting back the the corresponding Scheffe form.

The Neoprene Rubber Problem

History of Neoprene Rubber Data

In 1977, the U.S. Navy began a research program designed to improve the quality of the elastomers purchased for special defense applications. This program was prompted by an alarmingly high rate of failure of components made from the elastomers. From 1977 to 1980, the University of Central Florida (UCF) was involved in the development of analytical methods to monitor the components in a number of compounded elastomer systems. In 1980 and 1981, the U.S. Navy conducted an in-house single blind test of the procedures. The goal of this study was two-fold: one, to evaluate the compositional analysis procedures and two, to determine the effects of variations in the formulation on specific physical test properties. In this study, specification and off-specification samples of uncured, compounded neoprene rubber were prepared and rolled into sheets. Sixteen samples were prepared. A small portion of each sample was delivered to UCF where a compositional analysis was performed. A portion of each remaining sample was cured under rigidly controlled conditions and subjected to a number of physical tests. At the conclusion of this study when all data from the 16 samples was available (compositional analysis and physical test data), an attempt was made to develop regression equations which described the physical test data in terms of compositional parameters. This work was moderately successful. Equations were developed which have been used to predict the performance of the cured neoprene components since 1982.

Neoprene is a Mixture

Neoprene is a mixture since the response/properties of the blend depends only on the relative proportions of the ingredients present and not on the total weight or volume of the blend. The sum of the proportions of the ingredients is one. Neoprene is made by blending the elastomer; Altax; Octamine; carbon black; and lead oxide, Pb_3O_4. Altax (alt) contains sulfur and is used to initiate crosslinking. Octamine (oct) is an antioxidant; lead oxide (pb) provides water resistance; and carbon black (c) adds durability to the rubber.

The raw data given are shown in Table II. Ingredient concentrations (weight %) are shown as parts per hundred (pph) in Table IIa. The nine physical properties are shown in Table IIb. Explanations for each abbreviated name for the physical properties are shown below the table. Abbreviated names which end with 'F' represent the change in the property upon aging a fortnight (2 weeks).

Table IIa

Table of Raw Data. Molecular Weights and Components (pph).

OBS	MW*	MN*	ALT	OCT	C	PB
1	356000	97500	0.228	1.26	30.2	8.92
2	367000	94400	0.237	1.22	30.2	4.26
3	387000	88000	0.190	1.19	30.2	6.58
4	335000	95400	0.213	0.948	28.7	11.0
5	517000	89400	0.243	1.35	19.1	9.19
6	367000	77200	0.172	0.948	29.5	8.18
7	385000	87400	0.226	1.10	30.3	7.95
8	371000	80500	0.169	0.00	26.9	7.23
9	309000	88800	0.179	0.236	29.2	8.52
10	364000	96800	0.185	0.511	29.1	7.76
11	286000	96300	0.190	1.78	28.5	8.26
12	303000	82700	0.0988	1.05	26.7	7.23
13	435000	92500	0.542	1.04	27.9	8.34
13	404000	84800	0.811	1.10	27.0	7.43
15	539000	90700	0.317	1.21	26.9	8.24
16	296000	66300	0.734	1.22	28.6	7.53

* - as measured and reported by McGee[1]

MW - weight average molecular weight

MN - number average molecular weight

Table IIb

Physical Properties

OBS	HARD	HARDF	SPD	TEN	TENF	BRK	BRKF	ULT	ULTF
1	65.0	4.5	1580	1549	1886	2548	2413	714	541
2	64.9	0.8	1595	2859	1919	5018	2524	790	572
3	69.8	-3.1	1593	3262	1941	5144	2532	735	548
4	66.5	0.3	1539	1543	1257	2585	2434	763	569
5	50.2	5.3	1525	1299	1777	*	*	*	*
6	66.0	0.3	1580	2980	1747	4987	2427	698	561
7	74.0	2.2	1602	4003	2314	5329	2549	584	430
8	69.2	2.3	1588	3515	2036	5347	2626	697	491
9	68.4	0.8	1571	1758	2120	2729	2586	694	517
10	66.4	2.9	1577	1710	2036	2684	2569	687	520
11	67.7	-3.2	1586	1593	1923	2722	2949	712	547
12	67.3	1.7	1583	1689	2077	2671	2683	691	550
13	66.2	1.1	1577	1618	1800	2579	2403	817	577
14	67.3	3.4	1582	1265	1620	2406	2385	825	686
15	63.9	0.6	1574	1433	1891	2632	2542	827	569
16	65.4	0.8	1568	1709	2190	2224	2254	556	763

HARD Shore Hardness
HARDF Change in Shore Hardness Upon Aging
SPD Soundspeed
TEN Tensile Modulus @ 300% Elongation
TENF Change in Tensile Modulus @ 300% Elongation Upon Aging
BRK Tensile Strength at Break
BRKF Change in Tensile Strength at Break Upon Aging
ULT Ultimate Elongation
ULTF Change in Ultimate Elongation Upon Aging
"F" aging period - two weeks
* missing value

Conversion of Molecular Weight Parameters to Elastomer Concentration

The weight percent of elastomer (elas) was not given by McGee (Table II). This weight percent cannot be estimated by difference or errors in compositional analysis will be compounded. Elas will have to

be calculated from experimental data. The number average molecular weight (MN) and weight average molecular weight (MW) were the only parameters available to use in finding the elastomer concentration. No attempt was made in the previous study to measure the elastomer concentration. Weight percent for the elastomer must be calculated from the chromatograms used to determine MN and MW.[1] A calibration curve must first be prepared to find response factors as a function of the weight average molecular weights (MW). When an equation has been established relating the response factor to the weight average molecular weight, response factors can be calculated for each sample given in Table II since MW for each is given. Once this is done, the weight % elastomer can be calculated from the area of each sample peak, injection size, actual sample size and sample weight. The procedure is described in detail below.

Calibration

A calibration curve is first prepared using polystyrene standards with weight average molecular weights in the 10^5 range since those of the neoprene data fall between 20,000 and 50,000 (Table II). Approximately 10 mg of each standard is measured accurately (to 0.1 mg) and dissolved in 10.0 ml of tetrahydrofuran (THF). Accurate injections of 100, 120, or 150 µl of MW standards are used to obtain response factors as a function of MW for each standard. Individual peak heights of the chromatogram are summed at two mm increments along the horizontal axis (Fig. 7) to obtain Σh. Peak heights are measured by superimposing a grid of two mm squares on the chromatogram. Response factors are expressed as mg/Σh.

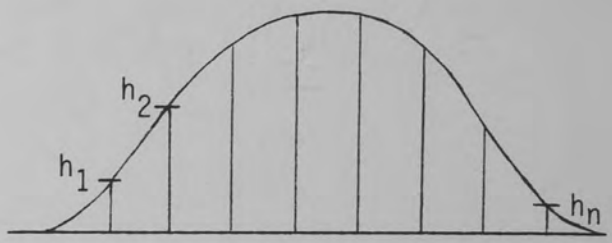

Figure 7. Summing peak heights of the chromatogram at two mm increments.

$$\Sigma h = h_1 + h_2 + \ldots + h_n.$$

The calibration equation describing the response factor was obtained using a least squares fit.

Response factor, $mg/\Sigma h = 1.771125\ E\text{-}04 + (2.4678\ E\text{-}11)MW$

Using the calibration curve, response factors can be predicted for each MW given assuming that the whole sample has that weight average molecular weight rather than a distribution. This act is consistent with current practice of describing a polymer by a weight or number average molecular weight.

Calculation of % Elastomer (pph)

When McGee[1] analyzed the samples for MW, he made 10.00 ml of each sample. The injection size used was 500 µl* which is equivalent to 1/20 of the total sample volume. Having obtained the response factor from

*Although McGee used 500 µl sample injections, it is felt that injections for the standards need not be so large since the response factor is expressed as $mg/\Sigma h$. The larger the injection size for the standard the greater the Σh and the more tedious the summation of the peak heights for calibration.

the previous equation, the weight percent (pph) of elastomer can be calculated as follows:

$$\text{pph elas in 10 ml sample} = \frac{(\text{response factor} \times \Sigma h_{\text{sample}} \times 20)}{\text{weight sample in mg}} \times 100$$

Table III shows data for the concentration, injection volume and weight, the sum of the peak heights and the response factor calculated for each of the four MW standards used for calibration. The conditions for calibration are shown below the table. Table IV shows the weight % elastomer in pph calculated from each corresponding sample MW, response factor, sample weights and the sum of the peak heights.

Table III

Calibration Data

MW	RESPONSE FACTOR mg/Σh	CONC,C mg/μl	INJ VOL, V μl	MG INJ C X V	ΣH
110000	1.786 E-04	1.00 E-03	100	0.1000	561
223000	1.849 E-04	1.02 E-03	120	0.1224	662
390000	1.855 E-04	1.03 E-03	120	0.1236	667
600000	1.921 E-04	1.00 E.03	150	0.1500	782

The conditions for the calibration are:

Liquid chromatograph	Waters Associates (Milford, MA) M-6000A pump Waters U6K injector - 500 μl loop
Column banks	10^3A (ASI, Santa Clara, CA) 10^4, 10^5A (Waters Associates)
Mobile phase	U.V. grade THF (Burdick & Jackson, American Scientific Products Co., Miami, FL)
Detector	Waters Model ALC/GPC 201 Refractive index detector
Attenuation	2 X
Flow rate	2.0 ml/min THF

Table IV

Data Involved in the Calculation of Weight Percent (pph) Elastomer

MW*	RESPONSE FACTOR, mg/ΣH	ΣH	SAMPLE WEIGHT, mg	ELAS pph
356000	1.85898 E-04	736	10.015	27.3232
367000	1.86170 E-04	846	10.000	31.4999
387000	1.86663 E-04	716	10.018	26.4446
335000	1.85380 E-04	804	10.025	29.7348
517000	1.89871 E-04	954	10.051	36.0436
367000	1.86170 E-04	832	10.016	30.9291
385000	1.86614 E-04	705	10.007	26.2941
371000	1.86268 E-04	871	10.044	32.3058
309000	1.84738 E-04	807	10.029	29.7305
364000	1.86096 E-04	783	10.051	28.9947
286000	1.84171 E-04	783	10.006	28.8283
303000	1.84590 E.04	810	10.009	29.8767
435000	1.87848 E-04	704	9.989	28.1706
404000	1.87083 E-04	785	10.076	29.1504
539000	1.90414 E-04	980	10.099	36.9554
296000	1.84417 E-04	605	10.010	22.2922

* - as measured and reported by McGee.[1]

Conversion of Data to Mixture Form

The sum of the proportions of the ingredients in a mixture must equal one. The raw data from Table IIa must be transformed to the mixture form. This is done by adding the weights of the individual components and dividing each by the sum (tot) since they do not add up

to 100% in the raw data. The reason for this discrepancy is probably due to the technical grade of the ingredients used, the omission of stearic acid (an inactive ingredient which helps assure good blending) from the recipe, and error in the compositional analysis. Table V shows the transformed data for the ingredients. Equations for transformation are shown below the table. All model building will be done using the transformed data and the subscript t will be omitted in later equations for convenience.

Table V

Transformed Component Data

ALT	OCT	C	PB	ELAS
0.0033563	0.0185482	0.444567	0.131309	0.402219
0.0035154	0.0180964	0.447959	0.063189	0.467240
0.0029410	0.0184198	0.467459	0.101850	0.409330
0.0030172	0.0134286	0.406540	0.155817	0.421197
0.0036859	0.0204773	0.289716	0.139397	0.546724
0.0024667	0.0135955	0.423066	0.117311	0.433561
0.0034310	0.0166995	0.459996	0.120692	0.399182
0.0025374	0.0000000	0.403875	0.108551	0.485037
0.0026376	0.0034775	0.430263	0.125542	0.438080
0.0027798	0.0076784	0.437261	0.116603	0.435678
0.0028126	0.0263494	0.421886	0.122273	0.426679
0.0015187	0.0161401	0.410419	0.112673	0.459250
0.0082130	0.0157593	0.422775	0.126378	0.426875
0.0123833	0.0167961	0.412268	0.113450	0.445103
0.0043058	0.0164352	0.365378	0.111923	0.501958
0.0012292	0.0204302	0.478937	0.126098	0.373306

Equations for transforming ingredient data from Table IIa:

$$tot = elas + alt + oct + c + pb$$

$$elas_t = elas/tot \qquad c_t = c/tot$$

$$alt_t = alt/tot \qquad pb_t = pb/tot$$

$$oct_t = oct/tot$$

where t stands for transformed (not to be confused with centered which expresses data as a deviation from the mean).

Generation of Scheffe Models for Neoprene Using Scheffe-Equivalent Equations

The simplest model is the linear model:

$$Y = b_1 elas + b_2 alt + b_3 oct + b_4 pb + b_5 c$$

Since all five components now sum to one, collinearity[3-5] is induced and will result in inaccurate test statistics. Omission of one component breaks the collinearity and gives an alternative intercept model. The linear term is replaced by a constant term. C is deleted because it has the largest range of all the components.

$$Y = a_0 + a_1 elas + a_2 alt + a_3 oct + a_4 pb$$

These two equations are equivalent because the former can be obtained by multiplying a_0 by elas + alt + oct + pb + c (i.e., by one).

$$Y = a_0(elas+alt+oct+pb+c) + a_1 elas + a_2 alt + a_3 oct + a_4 pb$$

$$= (a_0 + a_1)elas + (a_0 + a_2)alt + (a_0 + a_3)oct + (a_0 + a_4)pb + a_0 c$$

The relationship between the parameters of the two models are as follows:

$$b_1 = a_0 + a_1 \qquad b_4 = a_0 + a_4$$

$$b_2 = a_0 + a_2 \qquad b_5 = a_0$$

$$b_3 = a_0 + a_3$$

Although the linear model is simple, it is insufficient for prediction. Test parameters predicted by the model equations differ from the observed too much to be of use. This is shown by the large % error of prediction in Table X. The next choice is the quadratic

model. Scheffe-equivalent five variable quadratic:

$$Y = a_0 + a_1 elas + a_2 alt + a_3 oct + a_4 pb$$
$$+ a_{12} elas*alt + a_{13} elas*oct + a_{14} elas*pb + a_{15} elas*c$$
$$+ a_{23} alt*oct + a_{24} alt*pb + a_{25} alt*c$$
$$+ a_{34} oct*pb + a_{35} oct*c + a_{45} pb*c$$

This Scheffe-equivalent quadratic model has 14 terms and an intercept. Equivalence and conversion back to the Scheffe form is easily achieved by multiplying a_0 by (elas + alt + oct + pb + c), i.e., one.

$$Y = a_0(elas+alt+oct+pb+c) + a_1 elas + a_2 alt + a_3 oct + a_4 pb$$
$$+ a_{12} elas*alt + a_{13} elas*oct + a_{13} elas*pb + a_{15} elas*c$$
$$+ a_{23} alt*oct + a_{24} alt*pb + a_{25} alt*c$$
$$+ a_{34} oct*pb + a_{35} oct*c + a_{45} pb*c$$
$$= (a_0 + a_1) elas + (a_0 + a_2) alt + (a_0 + a_3) oct + (a_0 + a_4) pb + a_0 c$$
$$+ a_{12} elas*alt + a_{13} elas*oct + a_{14} elas*pb + a_{15} elas*c$$
$$+ a_{23} alt*oct + a_{24} alt*pb + a_{24} alt*c$$
$$+ a_{34} oct*pb + a_{35} oct*c + a_{45} pb*c$$

where $b_1 = a_0 + a_1$ $\quad b_{12} = a_{12}$ $\quad b_{24} = a_{24}$

$b_2 = a_0 + a_2$ $\quad b_{13} = a_{13}$ $\quad b_{25} = a_{25}$

$b_3 = a_0 + a_3$ $\quad b_{14} = a_{14}$ $\quad b_{34} = a_{34}$

$b_4 = a_0 + a_4$ $\quad b_{15} = a_{15}$ $\quad b_{35} = a_{35}$

$b_5 = a_0$ $\quad b_{23} = a_{23}$ $\quad b_{45} = a_{45}$

Some components may be combined to form reduced quadratic models with fewer terms. Two components are combined based on similarity in coefficient signs using a linear model or based on knowledge of similar functions. If two components are combined (e.g., elas, c+pb, oct, alt and crossproducts*) the result is a reduced four variable quadratic

which will have five terms less than the previous five variable quadratic. A reduced three variable quadratic model can be obtained by combining three original variables or another two components (e.g., elas, c+pb, oct, alt and crossproducts*; or elas+c+pb, alt, oct and crossproducts*).

If the quadratic model is not satisfactory, e.g., based on inspection of predicted and actual values (Table IX); or if based on a preselected R^2 of not less than 0.90, P>F value of 0.15 or less and the % error in prediction (Table X), the next attempt should be a cubic model.

There are two types of cubic models: the special and the full cubic. Special and full cubic equations based on five variables will contain 24 and 34 terms, respectively. Regression for five variable special and full cubic models is not possible since the number of terms in the equation exceeds the number of data points (25, 35 vs 16, respectively). Analysis of variance (ANOVA) cannot be performed when the number of degrees of error is zero. Reduced cubic models based on four and three variables (e.g., combinations used above) are easier to manage.

*crossproducts are obtained by multiplying different combinations of two terms for the quadratic models and combinations of two or three terms for cubic. Actual examples are shown on the following pages.

The special cubic forms for the four variable (two original components combined, e.g., based on elas, c+pb, oct, alt) and three variable (three original components, e.g., elas, c+pb, oct+alt or two more original combined, e.g., elas+c+pb, oct, alt) are shown below. Note that one variable is omitted from the linear portion to reduce collinearity.

Note: X_1, X_2, X_3 and X_4 are the variables remaining after combining two components. The linear term with the largest range, assigned the symbol X_4 is omitted from the linear portion of the equation to reduce collinearity. Special cubic four variable:

$$\begin{aligned} Y = {} & a_o + a_1X_1 + a_2X_2 + a_3X_3 \\ & + a_{12}X_1X_2 + a_{13}X_1X_3 + a_{14}X_1X_4 \\ & + a_{23}X_2X_3 + a_{24}X_2X_4 + a_{34}X_3X_4 \\ & + a_{123}X_1X_2x_3 + a_{124}X_1X_2X_4 + a_{134}X_1X_3X_4 \\ & + a_{234}X_2X_3X_4 \end{aligned}$$

The relationships between the coefficients a and b are as before and

$b_{123} = a_{123}$ $b_{124} = a_{124}$ $b_{134} = a_{134}$ $b_{234} = a_{234}$

Special cubic four variable:

Example: Shore Hardness

X_1 = alt X_2 = oct X_3 = elas+pb X_4 = c

$$\begin{aligned} \text{Shore Hardness} = {} & a_o + a_1\text{alt} + a_2\text{oct} + a_3\{\text{elas+pb}\} \\ & + a_{12}(\text{alt*oct}) + a_{13}(\text{alt} * \{\text{elas+pb}\}) + a_{14}(\text{alt*c}) \\ & + a_{23}(\text{oct} * \{\text{elas+pb}\}) + a_{24}(\text{oct*c}) + a_{34}(\{\text{elas+pb}\} * \text{c}) \\ & + a_{123}(\text{alt*oct*} \{\text{elas+pb}\}) + a_{124}(\text{alt*oct*c}) \\ & + a_{134}(\text{alt*} \{\text{elas+pb}\} * \text{c}) + a_{234}(\text{oct*} \{\text{elas+pb}\} *\text{c}) \end{aligned}$$

$b_{\text{alt*oct*}\{\text{elas+pb}\}} = a_{123}$ $b_{\text{alt*}\{\text{elas+pb}\}*\text{c}} = a_{134}$

$b_{\text{alt*oct*c}} = a_{124}$ $b_{\text{oct*}\{\text{elas+pb}\}*\text{c}} = a_{234}$

Note: X_1, X_2 and X_3 are the variables remaining after combining another two components in addition to the previous combination (for the four variable equation) or after combining three original variables. Examples of such combinations are shown on the previous page. The linear term with the largest range, assigned the symbol X_3 is omitted from the linear portion of the equation to reduce collinearity.

Special cubic three variable:

$$Y = a_o + a_1X_1 + a_2X_2 + a_{12}X_1X_2 + a_{13}X_1X_3 + a_{23}X_2X_3 + a_{123}X_1X_2X_3$$

The relationship between the coefficients a and b are as before and $b_{123} = a_{123}$.

Example: Shore Hardness

X_1 = oct X_2 = elas+pb X_3 = c+alt

$$\text{Shore Hardness} = a_o + a_1\text{oct} + a_2\{\text{elas+pb}\} + a_{12}(\text{oct}^*\{\text{elas+pb}\}) + a_{13}(\text{oct}^*\{\text{c+alt}\}) + a_{23}(\{\text{elas+pb}\} * \{\text{c+alt}\}) + a_{123}(\text{oct}^*\{\text{elas+pb}\}^*\{\text{c+alt}\})$$

$b_{\text{oct}^*\{\text{elas+pb}\}^*\{\text{c+alt}\}} = a_{123}$

Finally, the Scheffe equivalent form for the three variable full cubic model is:

$$Y = a_o + a_1X_1 + a_2X_2 + a_{12}X_1X_2 + a_{13}X_1X_3 + a_{23}X_2X_3 + a_{123}X_1X_2X_3 + \gamma_{12}X_1X_2(X_1-X_2) + \gamma_{13}X_1X_3(X_1-X_3) + \gamma_{23}X_2X_3(X_2-X_3)$$

where $b_{123} = a_{123}$

and $\gamma_{12,13,23}$ are the same for both the Scheffe and its equivalent form.

Example: Shore Hardness

X_1 = oct X_2 = elas+pb X_3 = c+alt

Shore hardness = $a_0 + a_1 oct + a_2\{elas+pb\}$

$+ a_{12}$ (oct* {elas+pb}) + a_{13}(oct* {c+alt})

$+ a_{23}$({elas+pb} * {c+alt}) + a_{123}(oct* {elas+pb}*{c+alt})

$+ \gamma_{12}$(oct* {elas+pb})(oct-{elas+pb})

$+ \gamma_{13}$(oct* {c+alt})(oct-{c+alt})

$+ \gamma_{23}$({elas+pb} * {c+alt})({elas+pb}-{c+alt})

$b_{oct*\{elas+pb\}*\{c+alt\}} = a_{123}$

$\gamma_{12,13,23}$ are the same for both Scheffe and Scheffe-equivalent forms.

RESULTS AND DISCUSSION

Selection and Use of SAS Software

The letters SAS are an acronym for Statistical Analysis System. This computer software system for data analysis began in 1966 in Cary, North Carolina and has grown into an all-purpose data analysis system. The decision to use SAS was based on the fact that it has more options and can handle larger data sets than other statistical packages available at the University of Central Florida such as Minitab and SPSS (Statistical Package for the Social Sciences). Another deciding factor was that McGee had used SAS in his earlier study.

SPSS was not chosen because it is tailored for use in the social sciences. Minitab, although a general purpose statistical computing system does not have all the options that SAS has and can only take a limited number of data points.

SAS Procedures

SAS procedures are computer programs that read the SAS data set, perform various computations and print the results. The statements that ask SAS to run a procedure always begin with PROC. Examples of some of the commonly used procedures are given below.

Example 1:

```
PROC PRINT DATA = XYZ.DATA; VAR A B;
```

tells SAS to print all the values of only A and B from the data set XYZ.DATA. If the entire data set is desired, the VAR statement is omitted.

Example 2:

PROC FSEDIT DATA = XYZ.DATA;

in interactive SAS tells SAS that one desires to edit the specified data set. See Appendix A for more details.

Example 3:

PROC PLOT DATA = XYZ.DATA; PLOT A*B;

will plot A against B with A on the vertical axis and B on the horizontal using data from the specified data set.

Example 4:

PROC MEANS DATA = XYZ.DATA

finds summary statistics such as the number of non-missing observations, mean, standard deviation, minimum and mazimum values and the standard error of the mean for all variables in the data set.

Example 5:

PROC REG DATA = XYZ.DATA; MODEL Y = A B C^2/P;

fits least squares estimates to the linear regression model of Y as a function of A, B and C^2. P tells SAS to include the predicted value for each point in the output.

SAS programs used to input data, define crossproduct variables and perform regression for this study are shown in Appendix B.

Using data from Table IIb (physical property data) and the transformed component data from Table V, pseudocomponent equations (according to the illustration on pp 22-23) and Scheffe-equivalent equations (developed on pp 24-25) are generated for neoprene using the PROC REG procedure of SAS. Exact SAS programs used are shown in Appendix B. Coefficients for each variable are shown in Table XI.

Comparison of Pseudocomponents and Scheffe-Equivalent Equations

The goal of this study is to find a more organized way of building models for mixture systems. Scheffe models provide us with a different way of looking at such systems. Scheffe provides an established scheme of building models of increasing complexity. Using this scheme, linear, quadratic, special and full cubic models were generated from the neoprene data for the nine physical test properties. Of the four model types, the quadratic Scheffe equations were selected because predicted values were very close, and in most cases, identical to the actual values (by inspection, Table IX). Since the criteria for model judgement involves the use of test statistics such as R^2, adjusted R^2, F and P>F, the chosen quadratic Scheffe models should be compared to McGee's on the same basis. Test statistics in the computer output for the Scheffe equations are inaccurate. Scheffe equations do not have an intercept. Even though most statistical packages are capable of fitting least squares equations without an intercept, test statistics that appear in the computer printout are inflated. It has long been believed[10] that the inflation of test statistics is due to collinearity in the data which causes extreme roundoff errors in the computer. Two procedures which have been used to obtain accurate test statistics to describe the Scheffe equation will be evaluated.

The first attempt is the use of pseudocomponents which has been described by Kurotori.[10] To generate pseudocomponents, mixture data must undergo further mathematical manipulation. Pseudocomponent equations are expressed in terms of the new set of data generated (see cake model on pg. 22). In order to obtain an equation in terms of the

original mixture data, the pseudocomponent equation must be back-transformed.

It is believed that VIF values in equations must be low in order to produce more accurate test statistics. Table VI shows VIF values for coefficients of the Scheffe, pseudocomponents and Scheffe-equivalent linear and quadratic model equations for the neoprene data. VIFs for coefficients of the linear model are examined first. As seen in Table VI, VIFs obtained for coefficients of pseudocomponents (column 2) and Scheffe-equivalent (column 3) equations are lower than those for the corresponding Scheffe (column 1) form. For the linear model, the lowest coefficient VIFs are obtained using the Scheffe-equivalent equation. The situation is slightly different for the quadratic shown in Table VI. Pseudocomponents (column 2) give the lowest coefficient VIFs when compared to the Scheffe (column 1) and Scheffe-equivalent (column 3) quadratic equations. In both the linear and quadratic cases, VIFs are reduced by the use of pseudocomponents and Scheffe-equivalent equations. This suggests that collinearity between data can be reduced by pseudocomponents and Scheffe-equivalent equations. Table VII compares inaccurate test statistics for the quadratic Scheffe and pseudocomponent model equations to those obtained for the quadratic Scheffe-equivalent model equations. Test statistics obtained using quadratic pseudocomponent equations are identical to those obtained with the quadratic Scheffe model and are combined and shown together with the Scheffe statistics. Since those from the Scheffe printout are inaccurate, it can be concluded from the data in Table VII that pseudocomponents do not always give accurate test statistics even with low VIFs as previously believed.[10]

Test statistics (Table VII) obtained with the quadratic Scheffe-equivalent equation to describe the corresponding Scheffe equation are different. The R^2, adjusted R^2 and F are lower in magnitude and P>F is higher than those on the Scheffe printout (Table VII). The values are less exaggerated and therefore more believable. The issue is not whether pseudocomponents or Scheffe-equivalent equations give better predicted values but which one gives accurate R^2, adjusted R^2, F and P>F values for the chosen Scheffe models. The Scheffe-equivalent model has an intercept which when multiplied by (alt+oct+c+pb+elas) converts to the Scheffe model. Since this model has an intercept, the computer handles the computation of test statistics for the Scheffe-equivalent form better than for pseudocomponents. With accurate test statistics, the selected quadratic equations can now be compared with McGee's equations more effectively. The Scheffe-equivalent model is the choice of this study since it generates accurate test statistics for the corresponding Scheffe equation without requiring additional transformation of the data. Pseudocomponents not only require further data manipulation but test statistics generated in the computer printout are still inaccurate.

Comparison of Conventional (McGee) and Mixture (Scheffe) Models

Table VIII summarizes the test statistics for all the Scheffe models generated for the nine physical properties of neoprene rubber. Test statistics for the five variable quadratic models with five, four and three components for BRK, BRKF, ULT and ULTF are obtained by repeating data point #8 in the transformed data set (Table V) so the degrees of freedom of error will equal one. Data point #8 was chosen because the predicted values were excellent for all four properties in

an earlier run. Statistics for the equations developed previously by McGee are also shown for comparison. McGee's original equations were generated using the combined 1979 and 1980 data. Since the author worked exclusively with the 1980 data, McGee's original equations were modified for comparison. The variables selected for constructing his equations were retained but the original coefficients were substituted with a new set of coefficients obtained by a least squares fit using only the 1980 data. The modified equations will thus give the best fit for the 1980 data using the variables McGee had chosen. McGee's original models were built using standardized data. In order to obtain meaningful numbers for the comparison of predicted values and the sum of squared residuals, the predicted values from his modified equations had to be 'unstandardized'. McGee's models for HARD, TEN, BRK and ULT have high correlation coefficients (>0.90). This is misleading since the sum of the squared residuals (Table VIII) and thus the % error of prediction (Table X) for McGee's models are much higher than those for the selected Scheffe models for these physical properties. The author's models are almost perfect predictors for seven of the nine physical properties of interest, except for TEN and TENF as shown in Table IX.

Models are chosen based on low sum of squared residuals. The selected Scheffe model for all nine physical properties is the five variable quadratic:

physical property
(e.g., Shore Hardness) = f(alt, oct, pb, elas, c
elas*alt, elas*oct, elas*pb, elas*c
alt*oct, alt*c, alt*pb
oct*c, oct*pb, pb*c)

Although the five variable Scheffe quadratic gives the lowest sum of squared residuals of all the models tried for TEN (Table VIII), it is far from satisfactory. The adjusted R^2 is too low (-0.38) and the P>F is too high (0.75). The P>F value of 0.75 corresponds to a significance probability of only 25%. This means that there is only a 25% chance of the user being right in assuming that the combination of variables describes the physical property significantly.

Table XI shows the coefficients for the 15 variables in the quadratic Scheffe equations for each of the nine physical properties. Equation coefficients for BRK, BRKF, ULT and ULTF should be used with caution. Repetition of data point #8 alters the distribution of the data set. Residuals for the two duplicate points (by inspection of Table IX) are larger than that of the other points. One of the residuals is positive while the other is negative in order to compensate for each other (Table IX).

Table VI

Comparison of Variance Inflation Factors

Variable	LINEAR MODEL EQUATION MODEL TYPE: Scheffe	Pseudocomponents	Scheffe-Equivalent
ALT	3.10	1.96	1.02
OCT	6.62	6.18	1.03
C	34.30	5.43	OMITTED
PB	31.69	6.40	1.02
ELAS	41.30	3.03	1.05

Variable	QUADRATIC MODEL EQUATION Scheffe	Pseudocomponents	Scheffe-Equivalent
ALT	998,620	43,829	329,025
OCT	241,113	19,763	36,334
C	14,983	5,103	OMITTED
PB	53,371	570	1,099
ELAS	82,643	668	1,044
ALTOCT	23,581	15,086	9,584
ALTC	319,588	18,777	104,554
ALTPB	30,900	4,277	10,267
ELALT	244,645	4,421	81,006
OCTC	45,069	4,891	7,191
OCTPB	11,265	2,685	1,941
ELOCT	38,451	1,301	6,104
PBC	29,716	701	832
ELC	99,051	205	332
ELPB	80,105	642	2,584

OMITTED - This term has the largest range of all the components and is omitted to reduce collinearity.
VIF values for both pseudocomponents and Scheffe-equivalent models are relatively low compared to Scheffe's. Scheffe-equivalent equations give lowest VIFs for the linear model while pseudocomponents give lowest VIFs for the quadratic model. Collinearity can be reduced with pseudocomponents and Scheffe-equivalent equations.

Table VII

Comparison of Inaccurate and Accurate Test Statistics for the Scheffe Quadratic Models

Independent Variable	Inaccurate Using Quadratic Scheffe or Pseudocomponents				Accurate Using Quadratic Scheffe-Equivalent			
	R^2	R^2_{adj}	F	P>F	R^2	R^2_{adj}	F	P>F
HARD	1.0	1.0	29134.48	0.0046	0.9995	0.9932	158.561	0.0622
ΔHARDF	0.9964	0.9464	18.579	0.1803	0.9952	0.9280	14.803	0.2013
SPD	1.0	1.0	575101.310	0.0010	0.9992	0.9880	89.057	0.0829
TEN	0.9869	0.8038	5.029	0.3380	0.9078	-0.3826	0.703	0.7470
ΔTENF	1.0	0.9995	2215.905	0.0167	0.9981	0.9708	36.608	0.1289
BRK	1.0	1.0	167481.755	0.0019	1.0	1.0	21421	0.0054
ΔBRKF	1.0	1.0	999999.99	0.0008	1.0	1.0	218802	0.0017
ULT	1.0	1.0	999999.99	0.0008	1.0	1.0	972187	0.0008
ΔULTF	1.0	1.0	999999.99	0.0008	1.0	1.0	29827	0.0014

HARD Shore Hardness
HARDF Shore Hardness Upon Aging*
SPD Soundspeed
TEN Tensile Modulus @ 300% Elongation
TENF Tensile Modulus @ 300% Elongation Upon Aging*
BRK Tensile Strength at Break
BRKF Tensile Strength at Break Upon Aging*
ULT Ultimate Elongation
ULTF Ultimate Elongation Upon Aging*

Δ - change in ···

* - aging period is two weeks

Table VIII

Summary of Test Statistics

DEPENDENT VARIABLE: HARD - SHORE HARDNESS							
MODEL	NUMBER OF COMPONENTS	TERMS	R^2	R^2_{adj}	F	P>F	ΣRes^2
Linear	5	5	0.6912	0.5789	6.155	0.0075	110.39
Quadratic	5	15	0.9995	0.9932	158.561	0.0622	0.16
Reduced Quadratic							
a	4	10	0.8181	0.5452	2.988	0.0973	65.03
b	3	6	0.7989	0.6984	7.947	0.0029	71.88
Special Cubic							
a ***	4	14	0.9343	0.6717	3.558	0.1619	23.47
b	3	7	0.8660	0.7767	9.695	0.0017	47.90
Full Cubic	3	10	0.9330	0.8326	9.288	0.0067	23.94
McGee		11	0.9622	0.8584	9.267	0.0227	13.50

* - in Scheffe-equivalent intercept, excluding intercept

** - test statistics estimated by repeating 1 data point (#8)

*** - model not full rank

R^2_{adj} - R-square adjusted for degrees of freedom

Table VIII (continued)

DEPENDENT VARIABLE: HARDF - CHANGE IN SHORE HARDNESS UPON AGING TWO WEEKS							
MODEL	NUMBER OF COMPONENTS	NUMBER OF TERMS	R^2	R^2_{adj}	F	P>F	ΣRes^2
Linear	5	5	0.2819	0.0208	1.080	0.4126	56.06
Quadratic	5	15	0.9952	0.9280	14.803	0.2013	0.38
Reduced Quadratic							
a	4	10	0.5886	-0.0284	0.954	0.5445	32.12
b	3	6	0.3174	-0.0238	0.930	0.5009	53.29
Special Cubic							
a ***	4	14	0.6773	-0.6135	0.525	0.8175	25.19
b	3	7	0.3178	-0.1369	0.699	0.6582	53.26
Full Cubic	3	10	0.4675	-0.3312	0.585	0.7745	41.57
McGee		4	0.3150	0.0659	1.264	0.3411	53.48

DEPENDENT VARIABLE: SPD - SOUNDSPEED							
MODEL	NUMBER OF COMPONENTS	NUMBER OF TERMS	R^2	R^2_{adj}	F	P>F	ΣRes^2
Linear	5	5	0.7271	0.6278	7.326	0.0040	1569.53
Quadratic	5	15	0.9992	0.9880	89.057	0.0829	4.61
Reduced Quadratic							
a	4	10	0.9692	0.9230	0.985	0.0007	177.08
b	3	6	0.6353	0.4530	3.484	0.0440	2097.37
Special Cubic							
a	4	14	0.9942	0.9568	26.546	0.0369	33.14
b	3	7	0.6663	0.4438	2.994	0.0680	1919.39
Full Cubic ***	3	10	0.9150	0.8178	9.416	0.0039	488.97
McGee		6	0.1697	-0.3839	0.307	0.9182	4791.24

Table VIII (continued)

DEPENDENT VARIABLE: TEN - TENSILE MODULUS @ 300% ELONGATION

MODEL		NUMBER OF COMPONENTS	NUMBER OF TERMS	R^2	R^2_{adj}	F	P>F	ΣRes^2
Linear		5	5	0.3972	0.1780	1.812	0.1968	7113186
Quadratic		5	15	0.9078	-0.3826	0.703	0.7470	1087685
Reduced Quadratic								
a		4	10	0.6337	0.0844	1.154	0.4468	4321823
b		3	6	0.3659	0.0488	1.154	0.3946	7482663
Special Cubic								
a		4	14	0.8604	-0.0473	0.948	0.6238	1647725
b		3	7	0.4383	0.0638	1.170	0.3992	6628648
Full Cubic	***	3	10	0.4567	-0.1642	0.736	0.6638	6410784
McGee			9	0.7175	0.2938	1.693	0.2684	3041735

DEPENDENT VARIABLE: TENF - CHANGE IN TENSILE MODULUS @ 300% ELONGATION UPON AGING TWO WEEKS

MODEL		NUMBER OF COMPONENTS	NUMBER OF TERMS	R^2	R^2_{adj}	F	P>F	ΣRes^2
Linear		5	5	0.3445	0.1062	1.445	0.2836	599330
Quadratic		5	15	0.9981	0.9708	36.608	0.1289	1780
Reduced Quadratic								
a		4	10	0.3798	-0.0336	0.919	0.5236	567036
b		3	6	0.5656	0.3484	2.604	0.0928	397164
Special Cubic								
a		4	14	0.7149	0.1446	1.254	0.4244	260698
b		3	7	0.6258	0.3764	2.509	0.1039	342128
Full Cubic	***	3	10	0.7577	0.4807	2.736	0.1012	221573
McGee			3	0.0638	-0.1702	0.273	0.8440	855987

Table VIII (continued)

DEPENDENT VARIABLE: BRK - TENSILE STRENGTH AT BREAK							
MODEL	NUMBER OF COMPONENTS	TERMS	R^2	R^2_{adj}	F	P>F	ΣRes^2
Linear	5	5	0.4248	0.1947	1.846	0.1967	12869670
Quadratic**	5	15	1.0000	0.9999	21421	0.0054	86
Reduced Quadratic							
a	4	10	0.6721	0.0820	1.139	0.4675	7335330
b	3	6	0.4118	0.0851	1.260	0.3587	13159465
Special Cubic							
a	4	14	0.9046	-0.3353	0.730	0.7374	2134030
b	3	7	0.6538	0.3942	2.518	0.1135	7745279
Full Cubic ***	3	10	0.7330	0.2524	1.525	0.3344	5973829
McGee		11	0.9653	0.8380	7.584	0.0609	776625

DEPENDENT VARIABLE: BRKF - CHANGE IN TENSILE STRENGTH AT BREAK UPON AGING TWO WEEKS							
MODEL	NUMBER OF COMPONENTS	TERMS	R^2	R^2_{adj}	F	P>F	ΣRes^2
Linear	5	5	0.2962	0.0147	1.052	0.4284	243269.3
Quadratic**	5	15	1.0000	1.0000	218802	0.0017	0.1
Reduced Quadratic							
a	4	10	0.8986	0.7160	4.992	0.0470	36500.1
b	3	6	0.4479	0.1411	1.460	0.2926	198698.3
Special Cubic							
a	4	14	0.9680	0.5522	2.328	0.4764	11511.5
b	3	7	0.4765	0.0839	1.214	0.3888	188389.7
Full Cubic	3	10	0.4773	-0.2196	0.685	0.6798	188102.6
McGee		6	0.1821	-0.4313	0.297	0.9219	294335.7

Table VIII (continued)

DEPENDENT VARIABLE: ULT - ULTIMATE ELONGATION							
MODEL	NUMBER OF COMPONENTS	TERMS	R^2	R^2_{adj}	F	P>F	ΣRes^2
Linear	5	5	0.7020	0.5828	5.890	0.0106	26211.59
Quadratic**	5	15	1.0000	1.0000	972187	0.0008	0.007
Reduced Quadratic							
a	4	10	0.9085	0.7438	5.515	0.0373	8049.70
b	3	6	-	-	-	-	-
Special Cubic							
a ***	4	14	0.9783	0.8478	7.498	0.1236	1912.59
b	3	7	-	-	-	-	-
Full Cubic	3	10	-	-	-	-	-
McGee		8	0.9512	0.8022	8.097	0.0100	7457.46

DEPENDENT VARIABLE: ULTF - CHANGE IN ULTIMATE ELONGATION UPON AGING TWO WEEKS							
MODEL	NUMBER OF COMPONENTS	TERMS	R^2	R^2_{adj}	F	P>F	ΣRes^2
Linear	5	5	0.2301	-0.0799	0.747	0.5817	64250.31
Quadratic**	5	15	1.0000	1.0000	298727	0.0014	0.02
Reduced Quadratic							
a	4	10	0.9021	0.7258	5.117	0.0435	8173.40
b	3	6	0.7336	0.5856	4.956	0.0186	22232.95
Special Cubic							
a	4	14	0.9812	0.7367	4.013	0.3740	1569.48
b	3	7	0.7354	0.5369	3.705	0.0459	22085.06
Full Cubic	3	10	0.7911	0.5821	3.786	0.0500	17435.86
McGee		9	0.8044	0.4524	2.285	0.1880	16321.57

Table IX

Comparison of Predicted and Observed Values

HARD OBS	SCHEFFE*	MCGEE	HARDF OBS	SCHEFFE*	MCGEE	SPD OBS	SCHEFFE*	MCGEE
65.0	65.0	66.5	4.5	4.4	1.9	1580	1580	1573
64.9	64.9	64.4	0.8	0.9	1.8	1595	1595	1571
69.8	69.7	69.8	-3.1	-3.3	-3.1	1593	1594	1571
66.5	66.5	66.3	0.3	0.2	1.5	1539	1539	1573
50.2	50.2	50.7	5.3	5.3	1.2	1525	1525	1570
66.0	66.3	68.0	0.3	0.8	1.0	1580	1578	1572
74.0	74.1	74.0	2.2	2.3	1.1	1602	1602	1601
69.2	69.2	68.7	2.3	2.3	1.6	1588	1588	1572
68.4	68.5	68.3	0.8	0.9	1.7	1571	1571	1574
66.4	66.2	66.0	2.9	2.7	1.6	1577	1578	1568
67.7	67.7	66.4	-3.2	-3.2	1.5	1586	1586	1581
67.3	67.2	65.7	1.7	1.6	1.5	1583	1584	1581
66.2	66.2	66.4	1.1	1.2	1.8	1577	1577	1575
67.3	67.3	68.7	3.4	3.4	3.3	1582	1582	1581
63.9	63.8	63.8	0.6	0.5	1.4	1574	1574	1569
65.4	65.4	64.6	0.8	0.9	1.0	1568	1568	1575

TEN OBS	SCHEFFE*	MCGEE	TENF OBS	SCHEFFE*	MCGEE
1549	1706	1570	1886	1880	1966
2859	2776	2856	1919	1922	1966
3262	3547	3030	1941	1930	1966
1543	1662	1608	1257	1252	1897
1299	1257	1295	1777	1779	1747
2980	2184	3233	1747	1778	1945
4003	3851	3444	2314	2320	1969
3515	3517	2503	2036	2036	1920
1758	1617	2526	2120	2126	1933
1710	2126	1401	2036	2019	1928
1593	1566	1296	1923	1924	1859
1689	1956	2409	2077	2066	1915
1618	1527	1203	1800	1804	1797
1265	1290	1550	1620	1619	1923
1433	1580	1627	1891	1885	1920
1709	1623	1716	2190	2194	1882

HARD - Shore Hardness
HARDF - Change in Shore Hardness Upon Aging for Two Weeks
SPD - Soundspeed
TEN - Tensile Modulus @ 300% Elongation
TENF - Change in Tensile Modulus @ 300% Elongation Upon Aging for Two Weeks

*Scheffe, pseudocomponents and Scheffe-equivalent models will give the same prediction. Latter two are derived from Scheffe models.

Table IX (continued)

BRK OBS	SCHEFFE*	MCGEE	BRKF OBS	SCHEFFE*	MCGEE
2548	2548	2924	2413	2413	2430
5018	5018	4348	2524	2524	2473
5144	5144	5194	2532	2532	2453
2585	2585	2615	2434	2434	2523
.	-36488	90477	.	5413	3909
4897	4897	4781	2427	2427	2520
5329	5329	5585	2549	2549	2582
5347	5354	5443	2626	2625	2620
2729	2729	2922	2586	2586	2552
2684	2684	2758	2569	2569	2523
2722	2722	2639	2949	2949	2604
2671	2671	2573	2683	2683	2606
2579	2579	2641	2403	2403	2401
2406	2406	2230	2385	2385	2538
2632	2632	2633	2542	2542	2467
2224	2224	2231	2254	2254	2586

ULT OBS	SCHEFFE*	MCGEE	ULTF OBS	SCHEFFE*	MCGEE
714	714	728	541	541	532
790	790	760	572	572	553
735	735	727	548	548	583
763	763	721	569	569	551
.	1830	3630	.	206	800
698	698	705	561	561	547
584	584	595	430	430	422
697	697	690	491	491	498
694	694	682	517	517	548
687	687	747	520	520	531
712	712	718	547	547	568
691	691	688	550	550	597
817	817	799	577	577	562
825	825	836	686	686	593
827	827	835	569	569	605
556	556	559	763	763	753

BRK - Tensile Strength at Break
BRKF - Change in Tensile Strength at Break Upon Aging for Two Weeks
ULT - Ultimate Elongation
ULTF - Change in Ultimate Elongation Upon Aging for Two Weeks

. - missing value
*Scheffe, pseudocomponents or Scheffe-equivalent model will give the same prediction. Latter two are derived from Scheffe models.

Table X

Comparison of % Error in Prediction

VARIABLE	SCHEFFE LINEAR	SCHEFFE QUADRATIC	MCGEE
HARD	3.43	0.151	1.40
HARDF	89.1	7.33	8.71
SPD	0.629	0.0341	1.098
TEN	31.6	12.4	20.64
TENF	10.1	0.556	12.11
BRK	25.2	0.0653	6.19
BRKF	4.9	0.00313	5.37
ULT	5.6	0.00291	3.01
ULTF	11.4	0.00634	5.72

HARD - Shore Hardness
HARDF - Change in Shore Hardness Upon Aging for Two Weeks
SPD - Soundspeed
TEN - Tensile Modulus @ 300% Elongation
TENF - Change in Tensile Modulus @ 300% Elongation Upon Aging for Two Weeks
BRK - Tensile Strength at Break
BRKF - Change in Tensile Strength at Break Upon Aging for Two Weeks
ULT - Ultimate Elongation
ULTF - Change in Ultimate Elongation Upon Aging for Two Weeks

% error - (mean residual/mean response) x 100 = $(\Sigma res^2/n)^{\frac{1}{2}}/\bar{Y}$ x 100
means are calculated using absolute values

Table XI

Coefficients for Each Selected Scheffe (Quadratic) Equation

MODEL EQUATION VARIABLE	SHORE HARDNESS PARAMETER ESTIMATE	Δ SHORE HARDNESS UPON AGING TWO WEEKS PARAMETER ESTIMATE	SOUNDSPEED PARAMETER ESTIMATE
ALT	-115131.39	126809.46	-363607.10
OCT	15729.852	-15033.61127	23556.20889
C	-522.12962	198.39312	467.48827
PB	-3506.13995	1351.70168	-10483.13494
ELAS	403.36627	555.01999	3230.19268
ALTOCT	-2458074.65	659300.37	-2921781.83
ALTC	378206.53	-159670.03	862905.71
ALTPB	-384370.71	-49340.65764	-528487.89
ELALT	99979.27464	-149586.11	269056.39
OCTC	7527.80929	1876.35713	3706.18944
OCTPB	-37301.71026	35548.03619	-20889.11799
ELOCT	-16141.38342	16786.87908	-24299.26142
PBC	5498.23139	212.26877	20424.16533
ELC	-1669.34477	-845.95189	-4477.54026
ELPB	8325.82011	-5656.04060	15296.67192

MODEL EQUATION VARIABLE	TENSILE MODULUS @ 300% PARAMETER ESTIMATE	ΔTENSILE MODULUS @ 300% UPON AGING TWO WEEKS PARAMETER ESTIMATE	TENSILE STRENGTH AT BREAK PARAMETER ESTIMATE
ALT	-53804867.15	6864881.31	-2348744686
OCT	-3143458.49	2465622.18	-15673217.64
C	-44804.98566	672.55883	-878590.04
PB	-580277.20	-186415.88	-2450328.82
ELAS	57023.54299	68367.88476	-259493.28
ALTOCT	-442291651	-94476632.63	-789806221
ALTC	108675987	2078835.14	328188029
ALTPB	-55551867.19	-19160589.41	12968696.46
ELALT	51346066.30	9112076.32	253057507
OCTC	8008707.68	-2448851.45	34957105.04
OCTPB	3570842.32	-3106381.34	308343.81
ELOCT	1734246.17	-1961976.30	10522110.60
PBC	359029.48	456844.49	4527564.74
ELC	-405043.20	-191747.34	1193060.56
ELPB	1857670.23	62970.16846	5023024.32

Δ - change in ...

Table XI (continued)

MODEL EQUATION VARIABLE	ΔTENSILE STRENGTH @ BREAK UPON AGING TWO WEEKS PARAMETER ESTIMATE	ULTIMATE ELONGATION PARAMETER ESTIMATE	ΔULTIMATE ELONGATION UPON AGING TWO WEEKS PARAMETER ESTIMATE
ALT	8940640.24	1756852.44	3933928.89
OCT	2860411.73	-52099.48603	-216718.83
C	47246.41829	24425.16308	-1479.03358
PB	39477.37848	82765.41125	29814.02766
ELAS	20193.62830	-3978.24275	-9292.88768
ALTOCT	-38632557.41	50809472.70	16204956.37
ALTC	-4923080.09	-6216748.03	-7383742.52
ALTPB	-13802233.13	7348086.85	2217936.76
ELALT	-10806879.48	-1984849.82	-3150630.63
OCTC	-3596417.06	-358578.18	208840.35
OCTB	-1846360.17	200530.79	69003.25418
ELOCT	-2483596.62	48954.46506	164777.80
PBC	-199563.46	-188814.02	-32484.79939
ELC	-123706.21	1862.31952	42425.85362
ELPB	36230.88020	-113532.79	-69818.24419

Δ - change in ...

Example:

Δ Tensile Strength @ Break Upon Aging Two Weeks

= 8940640.24 alt + 2860411.73 oct + 47246.41829 c

+ 39477.37848 pb + 20193.62830 elas - 38632557.41 altoct

- 4923080.09 altc - 13802233.13 altpb - 10806879 elalt

- 3596417.06 octc - 1846360.17 octpb - 2483596.62 eloct

- 199563.46 pbc - 123706.21 elc + 36230.88020 elpb

CONCLUSIONS AND RECOMMENDATIONS

It can be seen from Table X that the Scheffe quadratic models predict better (as judged by the % error in prediction) than the McGee/ conventional models. For seven of the nine physical properties investigated, the error in prediction for the Scheffe models is less than 1%. The error range for the Scheffe models is 0.002 to 12.4% while that for the McGee models is 1.40 to 87.1%. Predicted values using the McGee and Scheffe equations and actual values can also be compared by inspection in Table X. It can be seen that the Scheffe quadratic models are almost perfect predictors. When comparing the R^2 for the models in Table VIII, all nine Scheffe quadratic models have R^2 greater than 0.90. The R^2 range for the McGee models is from 0.06 to 0.96 and in only three of the nine models are the R^2 actually greater than 0.90. Since the quadratic equation contains almost as many variables as the number of observations (15 vs. 16), the adjusted R^2 should also be compared. Eight of the nine quadratic equations have adjusted R^2 values of 0.90 or greater while none of McGee's equations satisfy that level.

The smaller the values are for the significance probability P>F, the better the model is. Assuming at least 90% significance is desired (P>F must be <= 0.1), six of the nine Scheffe quadratic models are significant while only three of the nine McGee models meet that criterion. The P>F range for the Scheffe quadratic models is 0.7470 to 0.0008 corresponding to 25.3 to 99.25 significance and that for the McGee models is 0.9219 to 0.0100 corresponding to 7.8 to 90.0% significance.

Summary

Mixture models provide a new way of looking at data of systems in which the response to the system depends only on the proportions of the ingredients in the system. The mixture model would appear to be more appropriate for such systems because the sum of the independent variables is one. The mixture model is easier to build than the conventional models. The conventional method involves generating different new variables using functions of the independent variables and then using stepwise regression to filter out the insignificant variables. Scheffe (mixture model), on the other hand gives the user an established scheme to follow in model building. One begins with the simple linear model, progressing through the quadratic to the special and full cubic models. In most cases the quadratic model is found to be sufficient for prediction as shown by the almost perfect predictions in Table X. Scheffe's models also contain fewer terms than the conventional models of the same order. The proposed Scheffe and Scheffe-equivalent models do well with data not designed using the simplex method. Good predictions are not due to the large number of variables in the model because the adjusted R^2 for eight of the nine Scheffe quadratic models are greater than 0.90.

Recommendations

This study has demonstrated the advantages of using Scheffe rather than the conventional (McGee) models for mixtures with five components or less. When the number of components is greater than five, component reduction[15-17] should be undertaken for easier handling of the data. It is recommended that the components be combined based on similarity in coefficient signs using a preliminary linear model or based on the

coefficient signs using a preliminary linear model or based on the knowledge that certain components have similar functions.

Although the proposed Scheffe and Scheffe-equivalent models do well with data not designed using the simplex method, the simplex design should be incorporated into the data collection process whenever possible. If the data set is big enough (e.g., more than 40 data points) it should be divided into two groups. One group should be used for model building while the other is used for equation testing. The equations generated should not be used to predict for component proportions outside the range of the data set used to generate them.

A list of suggested reading has been assembled for future readers. These articles will give the reader basic information on model building. The list is arranged in alphabetical order according to topic.

APPENDIX A

INSTRUCTIONS TO INPUT AND EDIT DATA USING INTERACTIVE SAS

Statements that the user puts in will be shown in capital letters.

Keys are shown by underlining.

To input raw neoprene rubber data into a SAS data set named raw.data:

SAS enter

GO enter

When screen asks for file definitions, press enter

An empty screen with line numbers will appear.

DATA RAW.DATA; INPUT MW MN ALT OCT C PB HARD HARDF SPD TEN TENF BRK BRKF ULT ULTF; RUN; ALT 3

Ignore error message which appears.

PROC FSEDIT DATA = RAW.DATA; RUN; ALT 3

Screen with blank spaces next to variable names will appear. * If cursor is not on the first blank, bring it down using the downward arrow key. Type in each value and depress the enter key each time when finished. When the value for ULTF has been entered, depress PF 21. Repeat from * for each set of observations.

When the entire data set has been entered, depress PF 15 to save.

The blank screen with line numbers will return.

To exit interactive SAS, type:

ENDSAS; ALT 3

Helpful hints:

To edit only part of a large data set, use the VAR statement.

example:

PROC FSEDIT DATA = RAW.DATA; VAR MW BRK; RUN; ALT 3

To move screen backward use PF 19.

To move screen forward use PF 20.

To get the edit screen to display the observation desired, move the cursor to the COMMAND line at the top and type in the observation number.

APPENDIX B

SAS PROGRAMS

SAS PROGRAMS TO GENERATE DATA FOR
SCHEFFE & SCHEFFE-EQUIVALENT MODELS

```
data neo.data1; set raw.data;
                ** compute response factor **
response = (2.46788 E-11 * mw) + 1.771125 E-04;
              ** compute pph elastomer/transform **
pphelas = (response * h * 2000)/wt;
tot = pph + alt + oct + c + pb;
elas = pph/tot; alt = alt/tot; oct = oct/tot; c = c/tot;
pb = pb/tot;
                   ** define crossproducts **
elalt = elas * alt; eloct = elas * oct; elc = elas * c;
elpb = elas * pb;
altoct = alt * oct; altc = alt * c; altpb = alt * pb;
octc = oct * c; octpb = oct * pb; pbc = pb * c; run;

data neo.data2; set neo.data1;
               ** define combined variables **
              ** for reduced quad 4,3 var model **
rcel = c + elas; rocpb = oct + pb; rcpb = c + pb;
eloc = oct + elas; raloc = alt + oct; rceo = rcel * oct;
ropa = rocpb * alt; rcea = rcel * alt; reoa = reloc * alt;
rcpa = rcpb * alt; reoc = reloc * c; raoe = raloc * elas;
rcpe = rcpb * elas; rcpo = rcpb * oct; rcep = rcel * pb;
raoc = raloc * c;raop = raloc * pb; ropc = rocpb * c;
rope = rocpb * elas;reop = reloc * pb;

rpel = pb + elas; rcoct = oct + c;
rpea = alt * rpel; rpec = c * rpel; rpeo = oct * rpel;
rcoa = alt * rcoct; rcob = rcoct * pb; rcoe = rcoct * elas;

rocelc = oct + c + elas; raocpb = alt + oct + pb;
roeca = rocelc * alt; roecp = rocelc * pb; raope = raocpb *
elas;
roprce = rocpb * rcel; roercp = reloc * rcpb; ralo = ralc *
oct;
ralp = ralc * rpel; rcopel = rcoct * rpel;
rcpeal = alt * rcpbel; rcpoel = oct * rcpbel;
              **  for special cubic 3 var model **

xshard = ralc * oct * rpel; xshardf = alt * rcob * rpel;
xsspd = alt * rocelc * pb; xsten = alt * rocpb * rcel;
xstenf = xsten; xsbrk = raocpb * c * elas;
xsbrkf = alt * reloc * rcpb; xsultf = alt * oct * rcpbel;
```

```
** for special cubic 4 var model **

xsh1 = altoct * rpel; xsh2 = altc * rpel;
xsh3 = octc * rpel;
xsh4 = altoct * c;
xshf1 = altpb * rcoct; xshf2 = rcoct * elalt;
xshf3 = rcoct * elas;
xshf4 = altpb * elas;
xss1 = altc * reloc; xss2 = altc * pb; xss3 = pbc * reloc;
xss4 = altpb * reloc; xst1 = altc * rocpb;
xst2 = altc * elas; xst3 = elalt * rocpb;
xst4 = elc * rocpb;
xstf1 = altoct * pb; xstf2 = altoct * rcel;
xstf3 = altpb * rcel; xstf4 = octpb * rcel;
xsb1 = raloc * pbc; xsb2 = raloc * elpb;
xsb3 = raloc * elc; xsb4 = pbc * elas;
xsbf1 = altoct * rcpb; xsbf2 = altoct * elas;
xsbf3 = elalt * rcpb; xsbf4 = eloct * rcpb;
xsu1 = xss1; xsu2 = xss2; xsu3 = xss3; xsu4 = xss4;
xsuf1 = xsbf1; xsuf2 = xsbf2; xsuf3 = xsbf3; xsuf4 = xsbf4;

** for full cubic 3 var model **

xfh1 = ralo * (ralc - oct); xfh2 = ralp * (ralc - rpel);
xfh3 = rpeo * (oct - rpel);
xfhf1 = rcoa * (alt - rcob); xfhf2 = rpea * (alt -rpel);
xfhf3 = rcopel * (rcob - rpel);
xfs1 = roeca * (alt - rocelc); xfs2 = altpb * (alt - pb);
xfs3 = roecp * (pb - rocelc);
xft1 = ropa * (alt - rocpb); xft2 = rcea * (alt - rcel);
xft3 = roprce * (rocpb - rcel);

xftf1 = xft1; xftf2 = xft2; xftf3 = xft3;
xfb1 = raopc * (c - raocpb);
xfb2 = raope * (elas - raocpb);
xfb3 = elc * (c - elas);
xfbf1 = reoa * (alt - reloc); xfbf2 = rcpa * (alt - rcpb);
xfbf3 = reorcp * (reloc - rcpb);
xfuf1 = altoct  * (alt - oct );
xfuf2 = rcpeal * (alt - rcpbel);
xfuf3 = rcpoel * (oct - rcpbel); run;
```

SAS PROGRAMS FOR REGRESSION / SCHEFFE MODELS

```
                    ** linear model **
proc reg data = neo.data2;
    model y* = alt oct c pb elas/p vif noint;
                ** quadratic 5 var model **
 proc reg data = neo.data2;
    model y* = alt oct c pb elas
               elalt eloct elc elpb
               altoct altc altpb
               octc octpb pbc/p vif noint;
```

SAS PROGRAMS FOR REGRESSION / SCHEFFE-EQUIVALENT MODELS

```
                    ** linear model **
proc reg data = neo.data2;
    model y* = alt oct pb elas/p vif;
                ** quadratic 5 var model **
 proc reg data = neo.data2;
    model y* = alt oct pb elas
               elalt eloct elc elpb
               altoct altc altpb
               octc octpb pbc/p vif;

* - substitute desired physical property symbol for y**

            ** reduced quadratic 4 var model **
proc reg data = neo.data2;
    model hard = alt oct rpel altoct altc rpea rpec octc
    rpeo/p;
proc reg data = neo.data2;
    model hardf = alt pb elas rcoa altpb elalt rcob rcoe
    elpb/p;
proc reg data = neo.data2;
    model spd = alt reloc pb altc reoa altpb reoc pbc
    reop/p;
proc reg data = neo.data2;
    model ten = alt rocpb elas ropa altc elalt rope ropc
    elc/p;
proc reg data = neo.data2;
    model tenf = alt oct rcel altoct altpb rcea octpb rceo
    rcep/p;
proc reg data = neo.data2;
    model brk = raloc pb elas raoc raop raoe pbc elc
    elpb/p;
proc reg data = neo.data2;
    model brkf = alt oct elas altoct rcpa elalt eloct rcpe
    rcpo/p;
```

```
proc reg data = neo.data2;
    model ult = alt reloc pb reoa altc altpb reoc reop
    pbc/p;
proc reg data = neo.data2;
    model ultf = alt oct elas altoct rcpa elalt eloct rcpe
    rcpo/p;

quadratic 3 var model **
proc reg data = neo.data2;
    model hard = oct rpel ralo ralp rpeo/p;
proc reg data = neo.data2;
    model hardf = alt rcob rcoa rpea rcopel/p;
proc reg data = neo.data2;
    model spd = alt rocelc roeca altpb roecp/p;
proc reg data = neo.data2;
    model ten = alt rcel ropa rcea roprce/p;
proc reg data = neo.data2;
    model tenf = alt rcel ropa rcea roprce/p;
proc reg data = neo.data2;
    model brk = raocpb elas raopc raope elc/p;
proc reg data = neo.data2;
    model brkf = alt reloc reoa rcpa reorcp/p;
proc reg data = neo.data2;
    model ultf = alt oct altoct rcpeal rcpoel/p;
              ** special cubic 4 var model **
proc reg data neo.data2;
    model hard = alt oct rpel altoct altc rpea rpec octc
    rpeo xsh1 xsh2 xsh3 xsh4/p;
proc reg data neo.data2;
    model hardf = alt pb elas rcoa altpb elalt rcob rcoe
    elpb xshf1 xshf2 xshf3 xshf4/p;
proc reg data neo.data2;
    model spd = alt reloc pb altc reoa altpb reoc pbc reop
   xss1 xss2 xss3 xss4/p;
proc reg data neo.data2;
    model ten = alt rocpb elas ropa altc elalt rope ropc
    elc xst1 xst2 xst3 xst4/p;
proc reg data neo.data2;
    model tenf = alt oct rcel altoct altpb rcea
   xstf1 xstf2 xstf3 xstf4/p;
proc reg data neo.data2;
    model brk = raloc pb elas raoc raop raoe pbc elc elpb
  xsb1 xsb2 xsb3 xsb4/p
proc reg data neo.data2;
    model brkf = alt oct  elas altoct rcpa elalt eloct rcpe
    rcpo xsbf2 xsbf2 xsbf3 xsbf4/p;
proc reg data neo.data2;
    model ult = alt reloc pb reoa altc altpb reoc reop pbc
   xsu1 xsu2 xsu3 xsu4/p;
proc reg data neo.data2;
    model ultf = alt oct elas altoct rcpa elalt eloct rcpe
    rcpo xsuf1 xsuf2 xsuf3 xsuf4/p;
```

```
                ** special cubic 3 var model **
proc reg data neo.data2;
    model hard = oct rpel ralo ralp rpeo xshard/p;
proc reg data neo.data2;
    model hardf = alt rcob rcoa rpea rcopel xshardf/p;
proc reg data neo.data2;
    model spd = alt rocelc roeca altpb roecp xsspd/p;
proc reg data neo.data2;
    model ten = alt rcel ropa rcea roprce xsten/p;
proc reg data neo.data2;
    model tenf = alt rcel ropa rcea roprce xstenf/p;
proc reg data neo.data2;
    model brk = raocpb elas raopc raope elc xsbrk/p;
proc reg data neo.data2;
    model brkf = alt reloc reoa rcpa reorcp xsbrkf/p;
proc reg data neo.data2;
    model ultf = alt oct altoct rcpeal rcpoel xsultf/p;
                ** full cubic 3 var model **
proc reg data=neo.data2;
    model hard = oct rpel ralo ralp rpeo xshard xfh1 xfh2
    xfh3/p;
proc reg data=neo.data2;
    model hardf = alt rcob rcoa rpea rcopel xshardf xfhf1
    xfhf2 xfhf3/p;
proc reg data=neo.data2;
    model spd = alt rocelc roeca altpb roecp xsspd xfs1
    xfs2 xfs3/p;
proc reg data=neo.data2;
    model ten = alt rcel ropa rcea roprce xsten xft1 xft2
    xft3/p;
proc reg data=neo.data2;
    model tenf = alt rcel ropa rcea roprce xstenf xftf1
    xftf2 xftf3/p;
proc reg data=neo.data2;
    model brk = raocpb elas raopc raope elc xsbrk xfb1 xfb2
    xfb3/p;
proc reg data=neo.data2;
    model brkf = alt reloc reoa rcpa reorcp xsbrkf xfbf1
    xfbf2 xfbf3/p;
proc reg data=neo.data2;
    model ultf = alt oct altoct rcpeal rcpoel xsultf xfuf1
    xfuf2 xfuf3/p;
```

SAS PROGRAM TO GENERATE PSEUDOCOMPONENTS

```
            ** transform to pseudocomponents **
data pseu.data; set neo.data1;
palt = (alt - 0.00122916) / 0.27256;
poct = oct / 0.27256;    pc = (c - 0.28971593) /
0.27256;          ppb = (pb - 0.06318890) / 0.27256;
pelas = (elas - 0.37330631) / 0.27256;

                  ** define crossproducts **
pelalt = pelas * palt; peloct = pelas * poct;
pelc = pelas * pc; pelpb = pelas * ppb;
paltoct = palt * poct; paltc = palt * pc;
paltpb = palt * ppb; poctc = poct * pc;
poctpb = poct * ppb; ppbc = ppb * pc; run;
```

SAS PROGRAM FOR REGRESSION / PSEUDOCOMPONENT MODELS

```
                     ** linear model **
proc reg data = pseu.data;
    model y* = palt poct pc ppb pelas/p vif noint;

                  ** quadratic 5 var model **
 proc reg data = neo.data2;
    model y* = palt poct pc ppb pelas
               pelalt peloct pelc pelpb
               paltoct paltc paltpb
               poctc poctpb ppbc/p vif noint;
```

* - substitute desired physical property symbol for y

```
    PROGRAMS TO STANDARDIZE DATA / DEFINE NEW VARIABLES
     UNSTANDARDIZE PREDICTED VALUES FOR MCGEE MODELS **

          ** obtain mean and standard deviation **
proc means data = raw.data;

                  ** standardize raw data **
data std.data1; set raw.data;
mw = (mw - 376312.500000)/71946.9422561;
mn = (mn - 88043.750000)/8346.1742733;
alt = (alt - 0.254637)/0.1807276;
oct = (oct - 1.010187)/0.4340633;
c = (c - 28.062500)/2.7089666;
pb = (pb - 7.920000)/1.3989377;
hard = (hard - 66.137500)/4.8817859;
hardf = (hardf - 1.293750)/2.2813647;
spd = (spd - 1576.250000)/19.5806026;
ten =(ten - 2111.562500)/886.9432127;
tenf = (tenf - 1908.375000)/246.8913054;
brk = (brk - 3434.333333)/1264.1672960;
brkf = (brkf - 2525.066667)/160.3271061;
ult = (ult - 719.333333)/79.2650767;
ultf = (ultf - 562.733333)/77.2070192; run;

                  ** define new variables **
data std.data2; set std.data1;
alt2 = alt * alt; alt3 = alt2 * alt;
c2 = c * c; c3 = c2 * c;
mn2 = mn * mn; mn3 = mn2 * mn; mw2 = mw * mw; mw3 = mw2 *
mw;
mwn = mw * mn; mw2n = mw2 * mn; mwn2 = mw * mn2;
mnpb = mn * pb; pbmw = pb * mw;
cmw =c * mw; cmn = c * mn; c2mn = c2 * mn; c2mw = c2 * mw;
cmn2 = c * mn2;
vmn = 1/mn; vmw = 1/mw; vpb = 1/pb;
vmwbld = vmw; vpbbld = vpb; vmnbld = vmn;
vc = 1/c; vc2 = 1/c2; vpb3 = 1/(pb * pb * pb);
vmn3 = 1/mn3; vmw2 = 1/mw2; vmw3 = 1/mw3;
vpbc = 1/(pb * c);
vmw2n = 1/mw2n; vmwn2 = 1/mwn2; valt2 = 1/alt2;
valt3 = 1/alt3; run;

data un.data; set pred.data1;
            ** unstandardize predicted values **
uhard = (phard * 4.8817859) + 66.137500;
uhardf = (phardf * 2.2813647) + 1.293750;
uspd = (pspd * 19.5806026) + 1575.250000;
uten = (pten * 886.9432127) + 2111.562500;
utenf = (ptenf * 246.8913054) + 1908.375000;
ubrk = (pbrk * 1264.1672960) + 3434.333333;
ubrkf = (pbrkf * 160.3271061) + 2525.066667;
uult = (pult * 79.2650767) + 719.333333;
uultf = (pultf * 77.2070192) + 562.733333
```

```
                    ** compute residuals **
rhard = hard - uhard; rhardf = hardf - uhardf;
rspd = spd - uspd;
rten = ten - uten; rtenf = tenf - utenf; rbrk = brk - ubrk;
rbrkf = brkf - ubrkf; rult = ult - uult;
rultf = ultf - uultf;
                  ** compute squared residuals **
rhard2 = rhard * rhard; rhardf2 = rhardf * rhardf;
rspd2 = rspd * rspd;
rten2 = rten * rten; rtenf = rtenf * rtenf;
rbrk2 = rbrk * rbrk;
rbrkf2 = rbrkf * rbrkf; rult2 = rult * rult;
rultf2 = rultf * rultf;
             ** obtain sum of squared residuals **
proc means sum data = un.data;
var rhard2 rhardf2 rspd2 rten2 rtenf2 rbrk2 rbrkf2 rult2
rultf2;
```

SAS PROGRAM FOR REGRESSION / MCGEE MODELS

```
proc reg data = std.data1;
    model hard = c oct mn vmn vmn3 vmw vmw2 vmw2n vmwn2
    pbmw mnpb/p;
proc reg data = std.data1;
    model hardf = alt alt3 cmn vmnbld/p;
proc reg data = std.data1;
    model spd = alt2 alt3 mn vpbc vpb3 vc2/p;
proc reg data = std.data1;
    model ten = c c2 c3 mn mn2 mn3 alt valt2 valt3/p;
proc reg data = std.data1;
    model tenf = c vc vc2/p;
proc reg data = std.data1;
    model brk = mw mw2 mn2 mwn mw2n mwn2 c2 c2mw cmn c2mn
    cmn2/p;
proc reg data = std.data1;
    model brkf = mw c cmw vmw vmw2 vmw3/p;
proc reg data = std.data1;
    model ult = alt mn c c2 cmn c2mn vmwbld vpbbld/p;
proc reg data = std.data1;
    model ultf = vmw vmw2 vmw3 vpbc cmw cmn mnpb pb c/p;
```

APPENDIX C

MATRICES INVOLVED IN OBTAINING COEFFICIENT ESTIMATES IN LEAST SQUARES ANALYSIS

In least squares analysis, the coefficient estimate b is:

$$b = (X'X)^{-1} X'Y$$

where $X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$

and $X' = \begin{bmatrix} x_{11} & \cdot & \cdot & x_{m1} \\ x_{12} & \cdot & \cdot & x_{m2} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ x_{1n} & \cdot & \cdot & x_{mn} \end{bmatrix}$

and $Y = \begin{bmatrix} y_1 \\ \cdot \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix}$

The X matrix consists of n observations of m independent variables. The Y matrix consists of n responses. The X' matrix is the transpose of X where the rows and columns of X becomes interchanged in X'.

APPENDIX D

LIST OF SUGGESTED READING

General Experimental Design

1. Box, G.; Hunter, W. A useful method for model-building. Technometrics, 1962, 4, (3), 301-318.
2. Box, G.; Hunter, W. The $2^{(k-p)}$ fractional designs part II. Technometrics, 1961, 3, (4), 449-458.
3. Cornell, J. Discussion. Technometrics, 1984, 26, (2), 19-122.
4. Cornell, J.; Gorman, J. Fractional design plans for process variables in mixture experiments. Journal of Quality Technology, 1984, 16, (1), 20-38.
5. Deming, S.; Morgan, S. Teaching the fundamentals of experimental design. Anal. Chim. Acta, 1983, 150, 183-198.
6. Draper, N.; Stoneman, D. Factor changes and linear trends in eight-run two-level factorial designs. Technometrics, 1968, 10, (2), 301-311.
7. Kennard, R.; Stone, L. Computer aided design of experiments. Technometrics, 1969, 11, (1), 137-149.
8. Mitchell, T. An algorithm for the construction of 'D-optimal' experimental designs. Technometrics, 1974, 16, (2), 203-210.
9. Steinberg, D.; Hunter, W. Experimental design: review and comment. Technometrics, 1984, 26, (2), 71-97.

Linear Regression

1. Andrews, D. A robust method for multiple linear regression. Technometrics, 1974, 16, (4), 523-531.
2. Baskerville, J.; Toogood, J. Guided regression modeling for prediction and exploration of structure with many exploratory variables. Technometrics, 1982, 24, (1), 9-17.
3. Bradley, R.; Srivastava, S. Correlation in polynomial regression. The American Statistician, 1979, 33, (1), 11-14.

4. Hinchen, J. Multiple regression in process development. Technometrics, 1968, 10, (2), 257-269.

5. Hocking, R. Developments in linear regression methodology: 1959-1982. Technometrics, 1983, 25, (3), 219-230.

6. Schatzoff, M.; Tsao, R.; Fienberg, S. Efficient calculation of all possible regressions. Technometrics, 1968, 10, (4), 769-779.

7. Webster, J.; Gunst, R. Latent root regression analysis. Technometrics, 1974, 16, (4), 513-522.

8. Furnival, G. All possible regressions with less computation. Technometrics, 1971, 13, (2), 403-408.

Model Reduction

1. Gorman, J.; Cornell, J. A note on model reduction for experiments with both mixture components and process variables. Technometrics, 1982, 24, (3), 243-247.

2. Snee, R. Developing blending models for gasoline and other mixtures. Technometrics, 1981, 23, (2), 119-130.

Mixture Models

1. Becker, N. Models for the response of a mixture. Journal of the Royal Statistical Society, Series B, 1968, 30, 349-358.

2. Cornell, J. Experiments with mixtures: an update and bibliography. Technometrics, 1979, 21, (1), 95-106.

3. Cornell, J. Experiments with mixtures: a review. Technometrics, 1973, 15, (3), 437-455.

4. Cornell, J.; Good, I. The mixture problem for categorized components. Technometrics, 1970, 65, 339-355.

5. Draper, N. A mixtures model with inverse terms. Technometrics, 1977, 19, (1), 37-46.

6. Quenouille, M. Experiments with mixtures. Journal of The Royal Statistical Society, Series B, 1959, 21, 201-202.

7. Scheffe, H. Experiments with mixtures. Journal of the Royal Statistical Society, Series B, 1958, 20, 344-360.

8. Snee, R. Experimenting with mixtures. Chemtech, 1979, 702-710.

9. Snee, R. Techniques for the analysis of mixture data. Technometrics, 1973, 15, (3), 517-528.

Response Surfaces

1. Deming, S.; Morgan, S. Simplex optimization of variables in analytical chemistry. Anal. Chem., 1973, 45, (30), 278-283A.

2. Draper, N.; Stoneman, D. Response surface designs for factors at 2 or 3 levels and at 2 and 4 levels. Technometrics, 1968, 10, (1), 177-192.

3. Hare, L.; Brown, P. Plotting response surface contours for three-component mixtures. Journal of Quality Technology, 1977, 9, (4), 193-197.

4. Hendrix, C. Through the response surface with test tube and pipe wrench. Chemtech, 1980, 488-497.

5. Thompson, W.; Myers, R. Response surface designs for experiments with mixtures. Technometrics, 1968, 10, (4), 739-756.

6. Snee, R. Discussion. Technometrics, 1975, 17, (4), 425-430.

Ridge Regression

1. Marquardt, D. Comment. Journal of the American Statistical Association, 1980, 75, (369), 87-91.

2. St. John, R. Experiments with mixtures, ill-conditioning and ridge regression. Journal of Quality Technology, 1984, 16, (2), 81-96.

Simplex Designs

1. Crosier, R. Mixture experiments: geometry and pseudocomponents. Technometrics, 1984, 26, (3), 209-216.

2. Gorman, J. Fitting equations to mixture data with restraints on compositions. Journal of Quality Technology, 1970, 2, (4), 186-194.

3. Gorman, J.; Hinman, J. Simplex lattice designs for multicomponent systems. Technometrics, 1962, 4, (4), 463-487.

4. Hare, L. Mixture designs applied to food formulation. Food Technology, 1974, 50-57.

5. Kurotori, I. Experiments with mixtures of components having lower bounds. Industrial Quality Control, 1966, 592-596.

6. McLean, R.; Anderson, V. Extreme vertices design of mixture experiments. Technometrics, 1966, 8, (3), 447-454.

7. Morgan, S.; Deming, S. Simplex optimization of analytical chemical methods. Anal. Chem., 1974, 46, (9), 1170-1181.

8. Narcy, J.; Renaud, J. Use of simplex experimental designs in detergent formulation. Journal of the American Oil Chemists' Society, 1972, 49, 598-608.

9. Nigam, A.; Gupta, S.; Gupta, S. A new algorithm for the extreme vertices design for linear mixture models. Technometrics, 1983, 25, (4), 367-371.

10. Piepel, G. Calculating centroids in constrained mixture experiments. Technometrics, 1983, 25, (3), 279-283.

11. Snee, R. Experimental designs for mixture systems with multicomponents constraints. Communications In Statistics, Part A, 1979, A8, (4), 303-326.

12. Snee, R. Experimental designs for quadratic models in constrained mixture spaces. Technometrics, 1975, 17, (2), 149-159.

13. Snee, R. Marquardt, D. Extreme vertices designs for linear models. Technometrics, 1974, 16, (3), 399-408.

14. Snee, R.; Marquardt, D. Screening concepts and designs for experiments with mixtures. Technometrics, 1976, 18, (1), 19-29.

15. Spendley, W.; Hext, G.; Himsworth, F. Sequential application of simplex designs in optimisation and evolutionary operation. Technometrics, 1962, 4, (4), 441-461.

16. Long, D. Simplex optimization of the response from chemical systems. Anal. Chim. Acta, 1969, 46, 193-206.

Test Statistics and Residual Analysis

1. Allen, D. Mean square error of prediction as a criterion for selection variables. Technometrics, 1971, 13, (3), 469-475.

2. Anscombe, F.; Tukey, J. The examination and analysis of residuals. Technometrics, 1963, 5, (2), 141-160.

3. Bross, I. Outliers in patterned experiments: a strategic appraisal. Technometrics, 1961, 3, (1), 91-102.

4. Kennard, R. A note on the Cp statistic. Technometrics, 1971, 13, (4), 899-900.

5. Marquardt, D.; Snee, R. Test statistics for mixture models. Technometrics, 1974, 16, (4), 533-537.

6. Morgan, J.; Tatar, J. Calculation of the residual sum of squares for all possible regressions. Technometrics, 1972, 14, (2), 317-325.

7. Pope, P.; Webster, J. The use of an F-statistic in stepwise regression procedures. Technometrics, 1972, 14, (2), 327-340.

8. Snee, R.; Rayner, A. Assessing the accuracy of mixture model regression calculations. Technometrics, 1982, 14, (2), 67-79.

9. Weisberg, S. Discussion. Technometrics, 1983, 25, (3), 240-244.

10. Willan, A.; Watts, D. Meaningful multicollinearity measures. Technometrics, 1978, 20, (4), 407-412.

Variable Selection and Transformation

1. Berk, K. Comparing subset regression procedures. Technometrics, 1978, 20, (1), 1-6.

2. Box, G.; Tidwell, P. Transformation of the independent variables. Technometrics, 1962, 4, (4), 531-550.

3. Gorman, J.; Toman, R. Selection of variables for fitting equations to data. Technometrics, 1966, 8, (1), 27-51.

4. Mantel, N. Why stepdown procedures in variable selection. Technometrics, 1970, 12, (3), 621-625.

Miscellaneous

1. Bechhofer, R. Discussion. Technometrics, 1984, 26, (2), 125-127.

2. Cox, D. Discussion. Technometrics, 1984, 26, (2), 116.

3. Hendrix, C. What's the number mean. Chemtech, 1983, 598-605.

4. Hocking, R. Response. Technometrics, 1983, 25, (3), 248-249.

5. Helms, R. Discussion. Technometrics, 1971, 13, (3), 477-481.

6. Joiner, B. Discussion. Technometrics, 1984, 26, (2), 123-124.

7. Meites, L. CRC Critical Reviews In Analytical Chemistry, 1979, 1-53.

8. Snee, R. Discussion. Technometrics, 1983, 25, (3), 230-237.

9. Steinberg, D.; Hunter, W. Response. Technometrics, 1984, 26, (2), 128-130.

10. Vandeginste, B. Teaching Chemometrics. Anal. Chim. Acta, 1983, 150, 199-206.

11. Ziegel, E. Discussion. Technometrics, 1984, 26, (2), 98-104.

SAS

1. SAS Institute Inc. SAS Introductory Guide. Revised edition. Cary, NC: SAS Institute Inc., 1983.

2. SAS Institute Inc. SAS For Linear Models. Cary, NC: SAS Institute Inc., 1981.

3. SAS Institute Inc. SAS User's Guide: Statistics, 1982 Edition. Cary, NC: SAS Institute Inc., 1982.

APPENDIX E

GLOSSARY

ANOVA - analysis of variance, partitioning of total corrected sum of squares into its component parts - sum of squares of model and sum of squares of residuals (also known as sum of squares about regression).

canonical - conforming to a general rule.

centering - subtracting the mean from the independent (predictor) variables. $(x_j - \bar{x})$.

collinearity/multicollinearity - high degree of correlation among independent variables. Might lead to unstable estimates of regression coefficients.

correlation coefficient - measure of strength of association between 2 variables on a scale of -1 to +1. -1 implies perfect negative linear association, +1 implies perfect positive linear association. Zero or near zero implies randomness or lack of association.

C.V. - coefficent of variation. Expresses standard deviation as a percentage of the mean.
Std.dev(Y) x 100 / mean Y or root MSE x 100 / mean Y.

degrees of freedom - df, units of information contained in mean square.

error mean square - MSE, variance of residuals or errors. sum of squares of residuals/df $_{error}$.

full-rank model - model that has a unique solution.

intercept - predicted value if all model variables set to 0.

lack of fit -present when standard deviation of residuals is much larger than that of replicated experiment. Implies important feature of the data or model overlooked.

mean square - variance. sum of squares/df.

model F - Fisher statistic. Measures how well model as a whole (after adjusting for the mean) accounts for the behavior of the dependent variable; significant if P>F is small; equal to (mean square of model)/(mean square of error). For testing individual variable, see t-value.

multicollinearity - see collinearity.

multiple correlation coefficient - R measures efficiency of model, measures variation in dependent variable accounted for by model. R-square is from 0 to 1. Equals (sum of squares of model/sum of squares of corrected total). $R^2 = (SST - SSE)/SST = SSR/SST$
$= [\Sigma(Y_i - \bar{Y})^2 - \Sigma(Y_i - \hat{Y})^2]/\Sigma(Y_i - \bar{Y})^2$.

multiple regression - regression against more than one basic variable [e.g., $Y = f(X_1 \ X_2 \ . \ . \ . \ X_n)$].

not significant - effect is so small that it cannot be detected in the data. Does not mean variable has no effect, only that effect has not been demonstrated.

null hypothesis - a preconceived idea about the value of a parameter.

polynomial regression - regression against 1 basic variable X and other "independent variables" which are functions of X [e.g., $Y = f(X \ X^2 \ X^3 \ \sqrt{X} \ 1/X)$].

residual - error. Difference between observed and predicted value; difference between what you see and what you get. $(Y_i - \hat{Y})$.

root mean square - standard deviation.

significance - if variable is significant, its effect is large enough to be detected in the data. Observed variation is not just due to random scatter; thus data contains real information.

standardize - data values adjusted to similar magnitude for easy comparison. standardized value = (original-mean)/standard deviation or centered value/standard deviation.

Student's t (t value) - equals square root of F. A form of signal-to-noise ratio. Use is similar to F except that t is used to evaluate individual variables rather than the model.
mean x/standard deviation of x.

simplex design - set of selected data points over the mixture space at which data is gathered to fit an assumed response equation.

SS - sum of squares.

SSE - sum of squares of error, sum of squares of residuals, sum of squared residuals, sum of squares about regression. SST-SSR = $\Sigma (Y_i - \hat{Y})^2$.

SSR - sum of squares of model, sum of squares of regression. SST-SSE.

SST - total sum of squares. SSE + SSR = $\Sigma(Y_i - \bar{Y})^2$.

Type I SS - sequential SS - SS attributed to the independent variable if it's the first and only independent variable entered into the model excluding the intercept.

Type II SS - Contribution of the coefficient over and above that provided by all other coefficients in the model. Appropriate in situations where no interaction present between factors. Adjusted for crossproduct terms in full-rank models.

Type III SS - also called complete least square analysis. Corresponds to Yates' weighted square of means analysis. Principal use in situations which require comparison of main effects in the presence of interaction. Same as Type II SS if only main effects contained in model.

Type IV SS - called partial or adjusted SS - same as Type I for each coefficient if it is the last coefficient specified in model. Independent of order variables are presented in model.

undesigned data - in this study it specifically refers to data not gathered following the simplex design.

VIF - variance of inflation; factor by which variance of the estimated coefficient is inflated. $1/(1 - R^2)$.

X' matrix - tranpose of X. Rows and column of X are interchanged in X'.

XX' matrix - matrix that must be inverted in order to estimate coefficient. If collinearity is high, there is no unique solution for the inverse.

REFERENCES

1. Final Project Report N00014-81-L-2203: Analytical Procedures for Monitoring Sulfur and Metal Salts, and Final Project Report N00014-81-L-2243: Development of Automated Integrated Analysis Equipment. W.W. McGee, May 1982.

2. Scheffe, H. Journal of the Royal Statistical Society, Series B, 1958, 20, 344-460.

3. St. John, R. Journal of Quality Technology, 1984, 16, (2), 81-96.

4. Marquardt, D.; Snee, R. Technometrics, 1974, 16, (4), 533-537.

5. Willan, A; Watts, D. Technometrics, 1978, 20, (4), 407-412.

6. Kleinbaum, D.; Kupper, L. "Applied Regression Analysis and Other Multivariate Methods"; Duxbury Press: North Scituate, MA, 1978; p 44.

7. Hocking, R. Technometrics, 1983, 25, (3), 219-230.

8. Snee, R.; Rayner, A. Technometrics, 1982, 14, (2), 67-79.

9. Snee, R. Technometrics, 1983, 25, (3), 230-237.

10. Kurotori, I. Industrial Quality Control, 1966, 592-596.

11. Crosier, R. Technometrics, 1984, 26, (3), 209-216.

12. Snee, R. Chemtech, 1979, 702-710.

13. Gorman, J. Journal of Quality Technology, 1970, 2, (4), 186-194.

14. SAS Institute Inc. "SAS Introductory Guide". Revised edition. Cary, NC: SAS Institute Inc., 1983. 95 pp.

15. Gorman, J.; Cornell, J. Technometrics, 1982, 24, (3), 243-247.

16. Snee, R. Technometrics, 1981, 23, (2), 119-130.

17. Snee, R. Technometrics, 1973, 15, (3), 517-528.