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

1986

Cheng Y. Loh

University of Central Florida

Find similar works at: http://stars.library.ucf.edu/rtd

University of Central Florida Libraries http://library.ucf.edu

Part of the Other Chemistry Commons

STARS Citation


This Masters Thesis (Open Access) is brought to you for free and open access by STARS. It has been accepted for inclusion in Retrospective Theses and Dissertations by an authorized administrator of STARS. For more information, please contact lee.dotson@ucf.edu.
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
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

ACKNOWLEDGEMENTS ........................................................... iii
LIST OF TABLES ................................................................. vii
LIST OF FIGURES ............................................................... viii
INTRODUCTION ........................................................................ 1
MODEL BUILDING ....................................................................... 4
Least Squares Analysis .......................................................... 4
Assumptions for Least Squares Analysis ............................... 4
Outliers ................................................................................. 7
The Price of Rejecting Data ................................................... 8
Variable Selection .................................................................... 8
Backward Elimination ......................................................... 9
Forward Selection .................................................................. 9
Stepwise Selection ................................................................ 10
Advantages and Disadvantages of Conventional Model Building
Backward Elimination ......................................................... 11
Forward Selection ............................................................... 11
Stepwise Selection .............................................................. 12
Test Statistics Used to Judge Models ................................. 12
More About Collinearity ..................................................... 14
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recent Developments in Model Building</td>
<td>15</td>
</tr>
<tr>
<td>Development of Scheffe Forms</td>
<td>16</td>
</tr>
<tr>
<td>Advantages and Disadvantages of Scheffe Models</td>
<td>19</td>
</tr>
<tr>
<td>From Scheffe to Scheffe-Equivalent Models</td>
<td>20</td>
</tr>
<tr>
<td>Pseudocomponents and the Simplex</td>
<td>21</td>
</tr>
<tr>
<td>Illustration of the Relationship Between Pseudocomponents and the Simplex</td>
<td>22</td>
</tr>
<tr>
<td>Advantages and Disadvantages of Pseudocomponents</td>
<td>23</td>
</tr>
<tr>
<td>Development of the Scheffe-Equivalent Model</td>
<td>24</td>
</tr>
<tr>
<td>Advantages and Disadvantages of Scheffe-Equivalent Models</td>
<td>25</td>
</tr>
<tr>
<td>The Neoprene Rubber Problem</td>
<td>26</td>
</tr>
<tr>
<td>History of Neoprene Rubber Data</td>
<td>26</td>
</tr>
<tr>
<td>Neoprene is a Mixture</td>
<td>27</td>
</tr>
<tr>
<td>Conversion of MW Parameters to Elastomer Concentration</td>
<td>30</td>
</tr>
<tr>
<td>Calibration</td>
<td>31</td>
</tr>
<tr>
<td>Calculation of % Elastomer</td>
<td>33</td>
</tr>
<tr>
<td>Conversion of Data to Mixture Form</td>
<td>33</td>
</tr>
<tr>
<td>Generation of Scheffe Models for Neoprene Using Scheffe-Equivalent Equations</td>
<td>35</td>
</tr>
<tr>
<td>RESULTS AND DISCUSSION</td>
<td>41</td>
</tr>
<tr>
<td>Selection and Use of SAS Software</td>
<td>41</td>
</tr>
<tr>
<td>SAS Procedures</td>
<td>41</td>
</tr>
<tr>
<td>Comparison of Pseudocomponents and Scheffe-Equivalent Equations</td>
<td>43</td>
</tr>
<tr>
<td>Comparison of Conventional (McGee) and Mixture (Scheffe) Models</td>
<td>45</td>
</tr>
</tbody>
</table>
LIST OF TABLES

I. Number of coefficients in Scheffe models as a function of number of components ................. 19

II. Table of raw data (1980)
   a) Molecular weights and components (pph) .............. 28
   b) Physical properties .................................. 29

III. Calibration data ........................................... 32

IV. Data involved in the calculation of weight percent (pph) elastomer ............................. 33

V. Transformed component data ................................. 34

VI. Comparison of variance inflation factors ....................... 48

VII. Comparison of inaccurate and accurate test statistics for the Scheffe quadratic models ............... 49

VIII. Summary of test statistics ................................. 50

IX. Comparison of predicted and observed values ................. 55

X. Comparison of % error in prediction ............................ 57

XI. Parameters for each selected Scheffe (quadratic) equation ................................. 58
LIST OF FIGURES

1. Relationship between regression analysis and variable selection ........................................ 3
2. Illustration for assumption two .................................................. 5
3. Scatter diagram ................................................................. 7
4. Long-tailed or skewed distribution of data ........................................ 8
5. The search for true test statistics for the Scheffe model ................................. 21
6. Simplex for the cake model ..................................................... 23
7. Summing of peak height of the chromatogram at two mm increments ......................... 31
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, McGeel found strong correlations for physical properties and 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.
Regression analysis - fit equation to data

Least Squares Analysis - fit best straight line

Variable Selection - 2 general ways

Conventional (McGee)
- Forward
- Backward
- Stepwise

Mixture (Scheffe)
- Pseudocomponents
- Scheffe-Equivalent

All regression performed using SAS

Criteria for Judging Ability To Predict

- $R^2 \geq 0.90$
- $R_{\text{adj}}^2 \geq 0.90$
- $F$ - high*
- $P>F$ - low*
- $\sum \text{res}^2$ - low*
- % error - low*

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:

1. The probability distributions of $Y$ for a given $X$ have the same variance $\sigma^2$ 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 $Y_i$, $\mu_i$ lie on a straight line known as the true regression line.

$$\mu_i = b_0 + b_1x_i$$
The population parameters $b_0$ 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 $\sigma^2$ (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

$$Y_i = b_0 + b_i X_i + E_i$$

where $E_i$ are independent random variables with mean equal to 0 and variance equal to $\sigma^2$. 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
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 performed 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 $n \leq (k+1)$, 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. \( \Sigma res^2 \), 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.

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

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 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 \(b_0\) is:

\[
V(b_i) = \frac{S^2 VIF_i}{(n-1)S_i^2}
\]

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.

\[ 0 \leq x_i \leq 1 \]
\[ q \]
\[ \sum_{i=1}^{q} x_i = 1 \]

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:

\[ Y = b_0 + b_1 X_1 + b_2 X_2 + b_{12} X_1 X_2 + b_{11} X_1^2 + b_{22} X_2^2 \]

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

\[ X_1^2 = X_1 X_1 = X_1(1-X_2) = X_1 - X_1 X_2 \]
\[ X_2^2 = X_2 X_2 = X_2(1-X_1) = X_2 - X_1 X_2 \]

so

\[ Y = b_0(X_1 + X_2) + b_1 X_1 + b_2 X_2 + b_{12} X_1 X_2 \]
\[ + b_{11}(X_1 - X_1X_2) + b_{22}(X_2 - X_1X_2) \]
\[ = b_1X_1 + b_2X_2 + b_{12}X_1X_2 \]

where \( b_1^* = b_0 + b_1 + b_{11} \)
\( b_2^* = b_0 + b_2 + b_{22} \)
\( b_{12}^* = b_{12} - b_{11} - b_{22} \)

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 \( \gamma \) 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:
\[ Y = \sum_{i=1}^{q} b_i X_i \]

Quadratic:
\[ Y = \sum_{i=1}^{q} b_i X_i + \sum_{1 \leq i < j}^{q} b_{ij} X_i X_j \]
Special Cubic: \[ Y = \sum_{i=1}^{q} b_i x_i + \sum_{1\leq i<j}^{q} b_{ij} x_i x_j + \sum_{1\leq i<j<k}^{q} b_{ijk} x_i x_j x_k \]

Full Cubic: \[ Y = \sum_{i=1}^{q} b_i x_i + \sum_{1\leq i<j}^{q} b_{ij} x_i x_j + \sum_{1\leq i<j<k}^{q} b_{ijk} x_i x_j x_k + \sum_{1\leq i<j}^{q} g_{ij} x_i x_j (x_i - x_j) \]

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

<table>
<thead>
<tr>
<th>Number of Components</th>
<th>Linear</th>
<th>Quadratic</th>
<th>Special Cubic</th>
<th>Full Cubic</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>2</td>
<td>3</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>6</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>10</td>
<td>14</td>
<td>20</td>
</tr>
<tr>
<td>5</td>
<td>5</td>
<td>15</td>
<td>25</td>
<td>35</td>
</tr>
<tr>
<td>6</td>
<td>6</td>
<td>21</td>
<td>41</td>
<td>56</td>
</tr>
<tr>
<td>7</td>
<td>7</td>
<td>28</td>
<td>63</td>
<td>84</td>
</tr>
<tr>
<td>8</td>
<td>8</td>
<td>36</td>
<td>92</td>
<td>120</td>
</tr>
<tr>
<td>9</td>
<td>9</td>
<td>45</td>
<td>129</td>
<td>165</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>55</td>
<td>175</td>
<td>220</td>
</tr>
</tbody>
</table>

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\textsuperscript{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.

\textbf{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\textsuperscript{10} proposed the use of pseudocomponents in the late sixties and Snee\textsuperscript{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.
SCHEFFE

1. High VIF values
2. Inaccurate computer output for test statistics

complicated back-transformation

SCHERFFE-EQUIVALENT

1. Lower VIF for coefficient estimates
2. Accurate test statistics on computer printout
3. Easy conversion back to Scheffe

PSEUDOCOMPONENTS

1. Lower VIF for coefficient estimates
2. Inaccurate test statistics on computer printout
3. Must undergo complicated back-transformation to original scale

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

\[ a_i \] - coefficient of linear term \( X_i \) in Scheffe-equivalent equation.

\[ b_i = \text{intercept} + a_i \]

\[ b_i \] - 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:

\[
0.33 < E < 0.50 \\
0.33 < F < 0.75 \\
0.20 < S < 0.40
\]

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 pseudocomponents. 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 \(L = 1 - \text{sum of lower bounds of } E, F \text{ and } S\) and the subscript \(p\) stands for pseudocomponent.
Figure 6. Simplex for the cake model

Restraints: \( 0.33 \leq E \leq 0.50 \)
\( 0.33 \leq F \leq 0.75 \)
\( 0.20 \leq S \leq 0.40 \)

\( E_p = (E - \text{lower bound of } E)/L \)
\( F_p = (F - \text{lower bound of } F)/L \)
\( S_p = (S - \text{lower bound of } S)/L \)

where \( L = 1 - \text{sum of lower bounds of } E, F \text{ and } S \) and the subscript \( p \) stands for pseudocomponent.

Advantages and Disadvantages of Pseudocomponents

Advantage:

Lower VIF 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:

\[ Y = a_0 + a_1X_1 + a_2X_2 \]

The equivalence between the two equations can easily be verified by multiplying the intercept of the latter \( a_0 \) by \((X_1 + X_2 + X_3)\) i.e., by one.

\[ Y = a_0(X_1 + X_2 + X_3) + a_1X_1 + a_2X_1 + a_2X_2 = a_0X_1 + a_0X_2 + a_0X_3 + a_1X_1 + a_2X_2 = (a_0 + a_1)X_1 + (a_0 + a_2)X_2 + a_0X_3 \]

so

\[ b_1 = a_0 + a_1 \]
\[ b_2 = a_0 + a_2 \]
\[ b_3 = a_0 \]

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:
\[ Y = a_0 + a_1X_1 + a_2X_2 + a_{12}X_1X_2 + a_{13}X_1X_3 + a_{23}X_2X_3 \]
To confirm equivalence, \( a_0 \) is again multiplied by \( (X_1 + X_2 + X_3) \), i.e., by 1.0.
\[ Y = a_0(X_1 + X_2 + X_3) + a_1X_1 + a_2X_2 + a_{12}X_1X_2 + a_{13}X_1X_3 + a_{23}X_2X_3 \]
\[ = (a_0 + a_1)X_1 + (a_0 + a_2)X_2 + a_0X_3 + a_{12}X_1X_2 + a_{13}X_1X_3 + a_{23}X_2X_3 \]
where \( b_1 = a_0 + a_1 \) \( b_{12} = a_{12} \)
\( b_2 = a_0 + a_2 \) \( b_{13} = a_{13} \)
\( b_3 = a_0 \) \( b_{23} = a_{23} \)

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$_3$O$_4$. 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>
<thead>
<tr>
<th>OBS</th>
<th>MW*</th>
<th>MN*</th>
<th>ALT</th>
<th>OCT</th>
<th>C</th>
<th>PB</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>356000</td>
<td>97500</td>
<td>0.228</td>
<td>1.26</td>
<td>30.2</td>
<td>8.92</td>
</tr>
<tr>
<td>2</td>
<td>367000</td>
<td>94400</td>
<td>0.237</td>
<td>1.22</td>
<td>30.2</td>
<td>4.26</td>
</tr>
<tr>
<td>3</td>
<td>387000</td>
<td>88000</td>
<td>0.190</td>
<td>1.19</td>
<td>30.2</td>
<td>6.58</td>
</tr>
<tr>
<td>4</td>
<td>335000</td>
<td>95400</td>
<td>0.213</td>
<td>0.948</td>
<td>28.7</td>
<td>11.0</td>
</tr>
<tr>
<td>5</td>
<td>517000</td>
<td>89400</td>
<td>0.243</td>
<td>1.35</td>
<td>19.1</td>
<td>9.19</td>
</tr>
<tr>
<td>6</td>
<td>367000</td>
<td>77200</td>
<td>0.172</td>
<td>0.948</td>
<td>29.5</td>
<td>8.18</td>
</tr>
<tr>
<td>7</td>
<td>385000</td>
<td>87400</td>
<td>0.226</td>
<td>1.10</td>
<td>30.3</td>
<td>7.95</td>
</tr>
<tr>
<td>8</td>
<td>371000</td>
<td>80500</td>
<td>0.169</td>
<td>0.00</td>
<td>26.9</td>
<td>7.23</td>
</tr>
<tr>
<td>9</td>
<td>309000</td>
<td>88800</td>
<td>0.179</td>
<td>0.236</td>
<td>29.2</td>
<td>8.52</td>
</tr>
<tr>
<td>10</td>
<td>364000</td>
<td>96800</td>
<td>0.185</td>
<td>0.511</td>
<td>29.1</td>
<td>7.76</td>
</tr>
<tr>
<td>11</td>
<td>286000</td>
<td>96300</td>
<td>0.190</td>
<td>1.78</td>
<td>28.5</td>
<td>8.26</td>
</tr>
<tr>
<td>12</td>
<td>303000</td>
<td>82700</td>
<td>0.0988</td>
<td>1.05</td>
<td>26.7</td>
<td>7.23</td>
</tr>
<tr>
<td>13</td>
<td>435000</td>
<td>92500</td>
<td>0.542</td>
<td>1.04</td>
<td>27.9</td>
<td>8.34</td>
</tr>
<tr>
<td>13</td>
<td>404000</td>
<td>84800</td>
<td>0.811</td>
<td>1.10</td>
<td>27.0</td>
<td>7.43</td>
</tr>
<tr>
<td>15</td>
<td>539000</td>
<td>90700</td>
<td>0.317</td>
<td>1.21</td>
<td>26.9</td>
<td>8.24</td>
</tr>
<tr>
<td>16</td>
<td>296000</td>
<td>66300</td>
<td>0.734</td>
<td>1.22</td>
<td>28.6</td>
<td>7.53</td>
</tr>
</tbody>
</table>

* - as measured and reported by McGee$^1$

MW - weight average molecular weight

MN - number average molecular weight
## Table IIb
### Physical Properties

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

**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. 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 $\Delta h$. Peak heights are measured by superimposing a grid of two mm squares on the chromatogram. Response factors are expressed as mg/$\Delta h$. 
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, \( \text{mg/} \Sigma h = 1.771125 \times 10^{-4} + (2.4678 \times 10^{-11})\text{MW} \)

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 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 \( \Sigma h \) 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:

\[
\text{pph elas} = \frac{(\text{response factor} \times \text{Eh}_{\text{sample}} \times 20)}{\text{weight sample in mg}} \times 100
\]

in 10 ml sample

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

<table>
<thead>
<tr>
<th>MW</th>
<th>RESPONSE FACTOR mg/zh</th>
<th>CONC C mg/µl</th>
<th>INJ VOL, V µl</th>
<th>MG INJ C X V</th>
<th>ΣH</th>
</tr>
</thead>
<tbody>
<tr>
<td>110000</td>
<td>1.786 E-04</td>
<td>1.00 E-03</td>
<td>100</td>
<td>0.1000</td>
<td>561</td>
</tr>
<tr>
<td>223000</td>
<td>1.849 E-04</td>
<td>1.02 E-03</td>
<td>120</td>
<td>0.1224</td>
<td>662</td>
</tr>
<tr>
<td>390000</td>
<td>1.855 E-04</td>
<td>1.03 E-03</td>
<td>120</td>
<td>0.1236</td>
<td>667</td>
</tr>
<tr>
<td>600000</td>
<td>1.921 E-04</td>
<td>1.00 E-03</td>
<td>150</td>
<td>0.1500</td>
<td>782</td>
</tr>
</tbody>
</table>

The conditions for the calibration are:

- Liquid chromatograph: Waters Associates (Milford, MA) M-6000A pump, Waters U6K injector - 500 µl loop
- Column banks: 10³A (AŠI, Santa Clara, CA), 10⁴, 10⁵A (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

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

* - 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

<table>
<thead>
<tr>
<th>ALT</th>
<th>OCT</th>
<th>C</th>
<th>PB</th>
<th>ELAS</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0033563</td>
<td>0.0185482</td>
<td>0.444567</td>
<td>0.131309</td>
<td>0.402219</td>
</tr>
<tr>
<td>0.0035154</td>
<td>0.0180964</td>
<td>0.447959</td>
<td>0.063189</td>
<td>0.467240</td>
</tr>
<tr>
<td>0.0029410</td>
<td>0.0184198</td>
<td>0.467459</td>
<td>0.101850</td>
<td>0.409330</td>
</tr>
<tr>
<td>0.0030172</td>
<td>0.0134286</td>
<td>0.406540</td>
<td>0.155817</td>
<td>0.421197</td>
</tr>
<tr>
<td>0.0036859</td>
<td>0.0204773</td>
<td>0.289716</td>
<td>0.139397</td>
<td>0.546724</td>
</tr>
<tr>
<td>0.0024667</td>
<td>0.0135955</td>
<td>0.423066</td>
<td>0.117311</td>
<td>0.433561</td>
</tr>
<tr>
<td>0.0034310</td>
<td>0.0166995</td>
<td>0.459996</td>
<td>0.120692</td>
<td>0.399182</td>
</tr>
<tr>
<td>0.0025374</td>
<td>0.0000000</td>
<td>0.403875</td>
<td>0.108551</td>
<td>0.485037</td>
</tr>
<tr>
<td>0.0026376</td>
<td>0.0034775</td>
<td>0.430263</td>
<td>0.123542</td>
<td>0.438080</td>
</tr>
<tr>
<td>0.0028126</td>
<td>0.0076784</td>
<td>0.437261</td>
<td>0.116603</td>
<td>0.435678</td>
</tr>
<tr>
<td>0.0028126</td>
<td>0.0063494</td>
<td>0.421886</td>
<td>0.122273</td>
<td>0.426679</td>
</tr>
<tr>
<td>0.0015187</td>
<td>0.0161401</td>
<td>0.410419</td>
<td>0.112673</td>
<td>0.459250</td>
</tr>
<tr>
<td>0.0082130</td>
<td>0.0157593</td>
<td>0.422775</td>
<td>0.126378</td>
<td>0.426875</td>
</tr>
<tr>
<td>0.0123833</td>
<td>0.0167961</td>
<td>0.412268</td>
<td>0.113450</td>
<td>0.445103</td>
</tr>
<tr>
<td>0.0043058</td>
<td>0.0164352</td>
<td>0.365378</td>
<td>0.111923</td>
<td>0.501958</td>
</tr>
<tr>
<td>0.0012292</td>
<td>0.0204302</td>
<td>0.478937</td>
<td>0.126098</td>
<td>0.373306</td>
</tr>
</tbody>
</table>
Equations for transforming ingredient data from Table Ila:

\[ \text{tot} = \text{elas} + \text{alt} + \text{oct} + c + \text{pb} \]

\[ \text{elas}_t = \text{elas}/\text{tot} \quad \text{ct} = c/\text{tot} \]

\[ \text{alt}_t = \text{alt}/\text{tot} \quad \text{pb}_t = \text{pb}/\text{tot} \]

\[ \text{oct}_t = \text{oct}/\text{tot} \]

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:

\[ Y = b_1\text{elas} + b_2\text{alt} + b_3\text{oct} + b_4\text{pb} + b_5c \]

Since all five components now sum to one, collinearity 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.

\[ Y = a_0 + a_1\text{elas} + a_2\text{alt} + a_3\text{oct} + a_4\text{pb} \]

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

\[ Y = a_0(\text{elas}+\text{alt}+\text{oct}+\text{pb}+c) + a_1\text{elas} + a_2\text{alt} + a_3\text{oct} + a_4\text{pb} \]

\[ = (a_0 + a_1)\text{elas} + (a_0 + a_2)\text{alt} + (a_0 + a_3)\text{oct} + (a_0 + a_4)\text{pb} + a_0c \]

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

\[ b_1 = a_0 + a_1 \quad b_4 = a_0 + a_4 \]

\[ b_2 = a_0 + a_2 \quad b_5 = a_0 \]

\[ b_3 = a_0 + a_3 \]

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:

\[ Y = a_0 + a_1 \text{elas} + a_2 \text{alt} + a_3 \text{oct} + a_4 \text{pb} \\
+ a_{12} \text{elas} \times \text{alt} + a_{13} \text{elas} \times \text{oct} + a_{14} \text{elas} \times \text{pb} + a_{15} \text{elas} \times c \\
+ a_{23} \text{alt} \times \text{oct} + a_{24} \text{alt} \times \text{pb} + a_{25} \text{alt} \times c \\
+ a_{34} \text{oct} \times \text{pb} + a_{35} \text{oct} \times c + a_{45} \text{pb} \times c \]

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 \( (\text{elas} + \text{alt} + \text{oct} + \text{pb} + c) \), i.e., one.

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

= \( (a_0 + a_1) \text{elas} + (a_0 + a_2) \text{alt} + (a_0 + a_3) \text{oct} + (a_0 + a_4) \text{pb} + a_0 \text{c} \\
+ a_{12} \text{elas} \times \text{alt} + a_{13} \text{elas} \times \text{oct} + a_{14} \text{elas} \times \text{pb} + a_{15} \text{elas} \times c \\
+ a_{23} \text{alt} \times \text{oct} + a_{24} \text{alt} \times \text{pb} + a_{24} \text{alt} \times c \\
+ a_{34} \text{oct} \times \text{pb} + a_{35} \text{oct} \times c + a_{45} \text{pb} \times c \]

where \( b_1 = a_0 + a_1 \)  \\
\( b_2 = a_0 + a_2 \)  \\
\( b_3 = a_0 + a_3 \)  \\
\( b_4 = a_0 + a_4 \)  \\
\( b_5 = a_0 \)

\( b_{12} = a_{12} \)  \\
\( b_{13} = a_{13} \)  \\
\( b_{14} = a_{14} \)  \\
\( b_{15} = a_{15} \)  \\
\( b_{23} = a_{23} \)  \\
\( b_{24} = a_{24} \)  \\
\( b_{25} = a_{25} \)  \\
\( b_{34} = a_{34} \)  \\
\( b_{35} = a_{35} \)  \\
\( b_{45} = a_{45} \)

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:

$$Y = a_0 + a_1X_1 + a_2X_2 + a_3X_3$$
$$+ a_{12}X_1X_2 + a_{13}X_1X_3 + a_{14}X_1X_4$$
$$+ a_{23}X_2X_3 + a_{24}X_2X_4 + a_{34}X_3X_4$$
$$+ a_{123}X_1X_2X_3 + a_{124}X_1X_2X_4 + a_{134}X_1X_3X_4$$
$$+ a_{234}X_2X_3X_4$$

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

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

Special cubic four variable:

Example: Shore Hardness

$$X_1 = \text{alt} \quad X_2 = \text{oct} \quad X_3 = \text{elas+pb} \quad X_4 = c$$

Shore Hardness = $a_0 + a_1\text{alt} + a_2\text{oct} + a_3\text{elas+pb}$
$$+ a_{12}(\text{alt} \cdot \text{oct}) + a_{13}(\text{alt} \cdot \text{elas+pb}) + a_{14}(\text{alt} \cdot c)$$
$$+ a_{23}(\text{oct} \cdot \text{elas+pb}) + a_{24}(\text{oct} \cdot c) + a_{34}(\text{elas+pb} \cdot c)$$
$$+ a_{123}(\text{alt} \cdot \text{oct} \cdot \text{elas+pb}) + a_{124}(\text{alt} \cdot \text{oct} \cdot c)$$
$$+ a_{134}(\text{alt} \cdot \text{elas+pb} \cdot c) + a_{234}(\text{oct} \cdot \text{elas+pb} \cdot c)$$

$$b_{\text{alt} \cdot \text{oct} \cdot \text{elas+pb}} = a_{123} \quad b_{\text{alt} \cdot \text{elas+pb} \cdot c} = a_{134}$$
$$b_{\text{alt} \cdot \text{oct} \cdot c} = a_{124} \quad b_{\text{oct} \cdot \text{elas+pb} \cdot c} = a_{234}$$
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:

$$Y = a_0 + a_1X_1 + a_2X_2 + a_{12}X_1X_2 + a_{13}X_1X_3$$
$$+ a_{23}X_2X_3 + a_{123}X_1X_2X_3$$

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

Example: Shore Hardness

$$X_1 = \text{oct} \quad X_2 = \text{elas} + \text{pb} \quad X_3 = \text{c} + \text{alt}$$

Shore Hardness = $a_0 + a_1\text{oct} + a_2(\text{elas} + \text{pb})$
$$+ a_{12}(\text{oct} \ast \{\text{elas} + \text{pb}\}) + a_{13}(\text{oct} \ast \{\text{c} + \text{alt}\})$$
$$+ a_{23}(\{\text{elas} + \text{pb}\} \ast \{\text{c} + \text{alt}\})$$
$$+ a_{123}(\text{oct} \ast \{\text{elas} + \text{pb}\} \ast \{\text{c} + \text{alt}\})$$

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

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

$$Y = a_0 + a_1X_1 + a_2X_2 + a_{12}X_1X_2$$
$$+ a_{13}X_1X_3 + a_{23}X_2X_3 + a_{123}X_1X_2X_3$$
$$+ \gamma_{12}X_1X_2(X_1 - X_2) + \gamma_{13}X_1X_3(X_1 - X_3)$$
$$+ \gamma_{23}X_2X_3(X_2 - X_3)$$

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

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

\[ X_1 = \text{oct} \quad X_2 = \text{elas+pb} \quad X_3 = \text{c+alt} \]

Shore hardness = \( a_0 + a_1 \text{oct} + a_2 (\text{elas+pb}) \)

\[ + a_{12} (\text{oct} \ast (\text{elas+pb})) + a_{13} (\text{oct} \ast (\text{c+alt})) \]

\[ + a_{23} ((\text{elas+pb}) \ast (\text{c+alt})) + a_{123} (\text{oct} \ast (\text{elas+pb}) \ast (\text{c+alt})) \]

\[ + \gamma_{12} (\text{oct} \ast (\text{elas+pb}) \ast (\text{oct} - \text{elas+pb})) \]

\[ + \gamma_{13} (\text{oct} \ast (\text{c+alt}) \ast (\text{oct} - \text{c+alt})) \]

\[ + \gamma_{23} ((\text{elas+pb}) \ast (\text{c+alt}) \ast ((\text{elas+pb}) - (\text{c+alt}))) \]

\[ b_{\text{oct} \ast (\text{elas+pb}) \ast (\text{c+alt})} = a_{123} \]

\( \gamma_{12,13,23} \) 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 maximum 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:

\[ \text{physical property} = f(\text{alt}, \text{oct}, \text{pb}, \text{elas}, c) \]

\[ \text{elas*alt}, \text{elas*oct}, \text{elas*pb}, \text{elas*c} \]

\[ \text{alt*oct}, \text{alt*c}, \text{alt*pb} \]

\[ \text{oct*c}, \text{oct*pb}, \text{pb*c} \]
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

<table>
<thead>
<tr>
<th>Variable</th>
<th>Linear Model Equation</th>
<th>Quadratic Model Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MODEL TYPE:</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Scheffe</td>
<td>Pseudocomponents</td>
</tr>
<tr>
<td>ALT</td>
<td>3.10</td>
<td>1.96</td>
</tr>
<tr>
<td>OCT</td>
<td>6.62</td>
<td>6.18</td>
</tr>
<tr>
<td>C</td>
<td>34.30</td>
<td>5.43</td>
</tr>
<tr>
<td>PB</td>
<td>31.69</td>
<td>6.40</td>
</tr>
<tr>
<td>ELAS</td>
<td>41.30</td>
<td>3.03</td>
</tr>
<tr>
<td>ALT</td>
<td>998,620</td>
<td>43,829</td>
</tr>
<tr>
<td>OCT</td>
<td>241,113</td>
<td>19,763</td>
</tr>
<tr>
<td>C</td>
<td>14,983</td>
<td>5,103</td>
</tr>
<tr>
<td>PB</td>
<td>53,371</td>
<td>570</td>
</tr>
<tr>
<td>ELAS</td>
<td>82,643</td>
<td>668</td>
</tr>
<tr>
<td>ALTOCT</td>
<td>23,581</td>
<td>15,086</td>
</tr>
<tr>
<td>ALTC</td>
<td>319,588</td>
<td>18,777</td>
</tr>
<tr>
<td>ALTPB</td>
<td>30,900</td>
<td>4,277</td>
</tr>
<tr>
<td>ELALT</td>
<td>244,645</td>
<td>4,421</td>
</tr>
<tr>
<td>OCTC</td>
<td>45,069</td>
<td>4,891</td>
</tr>
<tr>
<td>OCTPB</td>
<td>11,265</td>
<td>2,685</td>
</tr>
<tr>
<td>ELOCT</td>
<td>38,451</td>
<td>1,301</td>
</tr>
<tr>
<td>PBC</td>
<td>29,716</td>
<td>701</td>
</tr>
<tr>
<td>ELC</td>
<td>99,051</td>
<td>205</td>
</tr>
<tr>
<td>ELPB</td>
<td>80,105</td>
<td>642</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>R²</th>
<th>R² adj</th>
<th>F</th>
<th>P&gt;F</th>
<th>R²</th>
<th>R² adj</th>
<th>F</th>
<th>P&gt;F</th>
</tr>
</thead>
<tbody>
<tr>
<td>HARD</td>
<td>1.0</td>
<td>1.0</td>
<td>29134.48</td>
<td>0.0046</td>
<td>0.9995</td>
<td>0.9932</td>
<td>158.561</td>
<td>0.0622</td>
</tr>
<tr>
<td>ΔHARDF</td>
<td>0.9964</td>
<td>0.9464</td>
<td>18.579</td>
<td>0.1803</td>
<td>0.9952</td>
<td>0.9280</td>
<td>14.803</td>
<td>0.2013</td>
</tr>
<tr>
<td>SPD</td>
<td>1.0</td>
<td>1.0</td>
<td>575101.310</td>
<td>0.0010</td>
<td>0.9992</td>
<td>0.9880</td>
<td>89.057</td>
<td>0.0829</td>
</tr>
<tr>
<td>TEN</td>
<td>0.9869</td>
<td>0.8038</td>
<td>5.029</td>
<td>0.3380</td>
<td>0.9078</td>
<td>-0.3826</td>
<td>0.703</td>
<td>0.7470</td>
</tr>
<tr>
<td>ΔTENF</td>
<td>1.0</td>
<td>0.9995</td>
<td>2215.905</td>
<td>0.0167</td>
<td>0.9981</td>
<td>0.9708</td>
<td>36.608</td>
<td>0.1289</td>
</tr>
<tr>
<td>BRK</td>
<td>1.0</td>
<td>1.0</td>
<td>167481.755</td>
<td>0.0019</td>
<td>1.0</td>
<td>1.0</td>
<td>21421</td>
<td>0.0054</td>
</tr>
<tr>
<td>ΔBRKF</td>
<td>1.0</td>
<td>1.0</td>
<td>999999.99</td>
<td>0.0008</td>
<td>1.0</td>
<td>1.0</td>
<td>218802</td>
<td>0.0017</td>
</tr>
<tr>
<td>ULT</td>
<td>1.0</td>
<td>1.0</td>
<td>999999.99</td>
<td>0.0008</td>
<td>1.0</td>
<td>1.0</td>
<td>972187</td>
<td>0.0008</td>
</tr>
<tr>
<td>ΔULTF</td>
<td>1.0</td>
<td>1.0</td>
<td>999999.99</td>
<td>0.0008</td>
<td>1.0</td>
<td>1.0</td>
<td>29827</td>
<td>0.0014</td>
</tr>
</tbody>
</table>

HARD  Shore Hardness
HARDF  Shore Hardness Upon Aging*
SPD  Soundspeed
TEN  Tensile Modulus @ 300% Elongation
TENF  Tensile Modulus @ 300% Elongation Upon Aging*

Δ - change in ...

* - aging period is two weeks
Table VIII
Summary of Test Statistics

<table>
<thead>
<tr>
<th>MODEL</th>
<th>NUMBER OF COMPONENTS</th>
<th>TERMS</th>
<th>R²</th>
<th>R²adj</th>
<th>F</th>
<th>P&gt;F</th>
<th>ΣRes²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>5</td>
<td>5</td>
<td>0.6912</td>
<td>0.5789</td>
<td>6.155</td>
<td>0.0075</td>
<td>110.39</td>
</tr>
<tr>
<td>Quadratic</td>
<td>5</td>
<td>15</td>
<td>0.9995</td>
<td>0.9932</td>
<td>158.561</td>
<td>0.0622</td>
<td>0.16</td>
</tr>
<tr>
<td>Reduced Quadratic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>4</td>
<td>10</td>
<td>0.8181</td>
<td>0.5452</td>
<td>2.988</td>
<td>0.0973</td>
<td>65.03</td>
</tr>
<tr>
<td>b</td>
<td>3</td>
<td>6</td>
<td>0.7989</td>
<td>0.6984</td>
<td>7.947</td>
<td>0.0029</td>
<td>71.88</td>
</tr>
<tr>
<td>Special Cubic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a ***</td>
<td>4</td>
<td>14</td>
<td>0.9343</td>
<td>0.6717</td>
<td>3.558</td>
<td>0.1619</td>
<td>23.47</td>
</tr>
<tr>
<td>b</td>
<td>3</td>
<td>7</td>
<td>0.8660</td>
<td>0.7767</td>
<td>9.695</td>
<td>0.0017</td>
<td>47.90</td>
</tr>
<tr>
<td>Full Cubic</td>
<td>3</td>
<td>10</td>
<td>0.9330</td>
<td>0.8326</td>
<td>9.288</td>
<td>0.0067</td>
<td>23.94</td>
</tr>
<tr>
<td>McGee</td>
<td>11</td>
<td></td>
<td>0.9622</td>
<td>0.8584</td>
<td>9.267</td>
<td>0.0227</td>
<td>13.50</td>
</tr>
</tbody>
</table>

* - in Scheffe-equivalent intercept, excluding intercept
** - test statistics estimated by repeating 1 data point (#8)
*** - model not full rank
R²adj - R-square adjusted for degrees of freedom
**Table VIII (continued)**

| DEPENDENT VARIABLE: HARDF - CHANGE IN SHORE HARDNESS UPON AGING TWO WEEKS |
|----------------------------------|------------------|-------------|--------|---------|-------|
| MODEL                            | COMPONENTS       | TERMS       | R²     | R²_adj  | F     | P>F   | Res² |
| 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                | 4                | 10          | 0.5886 | -0.0284 | 0.954 | 0.5445| 32.12|
| Special Cubic                    | 3                | 6           | 0.3174 | -0.0238 | 0.930 | 0.5009| 53.29|
| McGee                            | 3                | 10          | 0.4675 | -0.3312 | 0.585 | 0.7745| 41.57|
|                                   |                 |             | 0.3150 | 0.0659  | 1.264 | 0.3411| 53.48|

| DEPENDENT VARIABLE: SPD - SOUNDSPEED |
|--------------------------------------|------------------|-------------|--------|---------|-------|
| MODEL                                | COMPONENTS       | TERMS       | R²     | R²_adj  | F     | P>F   | Res² |
| 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                   | 4                | 10          | 0.9692 | 0.9230  | 0.985 | 0.0007| 177.08|
| Special Cubic                       | 3                | 6           | 0.6353 | 0.4530  | 3.484 | 0.0440| 2097.37|
| McGee                               | 3                | 10          | 0.9150 | 0.8178  | 9.416 | 0.0039| 488.97|
|                                   |                 |             | 0.1697 | -0.3839 | 0.307 | 0.9182| 4791.24|
Table VIII (continued)

**DEPENDENT VARIABLE: **TEN - TENSILE MODULUS @ 300% ELONGATION

<table>
<thead>
<tr>
<th>MODEL</th>
<th>COMPONENTS</th>
<th>TERMS</th>
<th>$R^2$</th>
<th>$R^2_{adj}$</th>
<th>F</th>
<th>P&gt;F</th>
<th>$\Sigma$Res$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>5</td>
<td>5</td>
<td>0.3972</td>
<td>0.1780</td>
<td>1.812</td>
<td>0.1968</td>
<td>7113186</td>
</tr>
<tr>
<td>Quadratic</td>
<td>5</td>
<td>15</td>
<td>0.9078</td>
<td>-0.3826</td>
<td>0.703</td>
<td>0.7470</td>
<td>1087685</td>
</tr>
<tr>
<td>Reduced Quadratic</td>
<td>10</td>
<td></td>
<td>0.6337</td>
<td>0.0844</td>
<td>1.154</td>
<td>0.4468</td>
<td>4321823</td>
</tr>
<tr>
<td>Special Cubic</td>
<td></td>
<td>6</td>
<td>0.3659</td>
<td>0.0488</td>
<td>1.154</td>
<td>0.3946</td>
<td>7482663</td>
</tr>
<tr>
<td>Full Cubic</td>
<td>10</td>
<td>0.4567</td>
<td>-0.1642</td>
<td>0.736</td>
<td>0.6638</td>
<td>6410784</td>
<td></td>
</tr>
<tr>
<td>McGee</td>
<td>9</td>
<td>0.7175</td>
<td>0.2938</td>
<td>1.693</td>
<td>0.2684</td>
<td>3041735</td>
<td></td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>MODEL</th>
<th>COMPONENTS</th>
<th>TERMS</th>
<th>$R^2$</th>
<th>$R^2_{adj}$</th>
<th>F</th>
<th>P&gt;F</th>
<th>$\Sigma$Res$^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>5</td>
<td>5</td>
<td>0.3445</td>
<td>0.1062</td>
<td>1.445</td>
<td>0.2836</td>
<td>599330</td>
</tr>
<tr>
<td>Quadratic</td>
<td>5</td>
<td>15</td>
<td>0.9981</td>
<td>0.9708</td>
<td>36.608</td>
<td>0.1289</td>
<td>1780</td>
</tr>
<tr>
<td>Reduced Quadratic</td>
<td>10</td>
<td></td>
<td>0.3798</td>
<td>-0.0336</td>
<td>0.919</td>
<td>0.5236</td>
<td>567036</td>
</tr>
<tr>
<td>Special Cubic</td>
<td></td>
<td>6</td>
<td>0.5656</td>
<td>0.3484</td>
<td>2.604</td>
<td>0.0928</td>
<td>397164</td>
</tr>
<tr>
<td>Full Cubic</td>
<td>10</td>
<td>0.7149</td>
<td>0.1446</td>
<td>1.254</td>
<td>0.4244</td>
<td>260698</td>
<td></td>
</tr>
<tr>
<td>McGee</td>
<td>3</td>
<td>0.6258</td>
<td>0.3764</td>
<td>2.509</td>
<td>0.1039</td>
<td>342128</td>
<td></td>
</tr>
</tbody>
</table>

**DEPENDENT VARIABLE: **TEN - TENSILE MODULUS @ 300% ELONGATION

**DEPENDENT VARIABLE: **TENF - CHANGE IN TENSILE MODULUS @ 300% ELONGATION UPON AGING TWO WEEKS
### Table VIII (continued)

<table>
<thead>
<tr>
<th>DEPENDENT VARIABLE:</th>
<th>BRK - TENSILE STRENGTH AT BREAK</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MODEL</strong></td>
<td><strong>NUMBER OF COMPONENTS</strong></td>
</tr>
<tr>
<td>Linear</td>
<td>5</td>
</tr>
<tr>
<td>Quadratic**</td>
<td>5</td>
</tr>
<tr>
<td>Reduced Quadratic</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>4</td>
</tr>
<tr>
<td>b</td>
<td>3</td>
</tr>
<tr>
<td>Special Cubic</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>4</td>
</tr>
<tr>
<td>b</td>
<td>3</td>
</tr>
<tr>
<td>Full Cubic***</td>
<td>3</td>
</tr>
<tr>
<td>McGee</td>
<td>11</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>DEPENDENT VARIABLE:</th>
<th>BRKF - CHANGE IN TENSILE STRENGTH AT BREAK UPON AGING TWO WEEKS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>MODEL</strong></td>
<td><strong>NUMBER OF COMPONENTS</strong></td>
</tr>
<tr>
<td>Linear</td>
<td>5</td>
</tr>
<tr>
<td>Quadratic**</td>
<td>5</td>
</tr>
<tr>
<td>Reduced Quadratic</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>4</td>
</tr>
<tr>
<td>b</td>
<td>3</td>
</tr>
<tr>
<td>Special Cubic</td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>4</td>
</tr>
<tr>
<td>b</td>
<td>3</td>
</tr>
<tr>
<td>Full Cubic</td>
<td>3</td>
</tr>
<tr>
<td>McGee</td>
<td>6</td>
</tr>
</tbody>
</table>
Table VIII (continued)

<table>
<thead>
<tr>
<th>DEPENDENT VARIABLE: ULT - ULTIMATE ELONGATION</th>
<th>NUMBER OF MODEL COMPONENTS TERMS</th>
<th>$R^2$</th>
<th>$R^2_{adj}$</th>
<th>$F$</th>
<th>$P&gt;F$</th>
<th>$\Sigma Res^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear</td>
<td>5</td>
<td>5</td>
<td>0.7020</td>
<td>0.5828</td>
<td>5.890</td>
<td>0.0106</td>
</tr>
<tr>
<td>Quadratic**</td>
<td>5</td>
<td>15</td>
<td>1.0000</td>
<td>1.0000</td>
<td>972187</td>
<td>0.0008</td>
</tr>
<tr>
<td>Reduced Quadratic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>4</td>
<td>10</td>
<td>0.9085</td>
<td>0.7438</td>
<td>5.515</td>
<td>0.0373</td>
</tr>
<tr>
<td>b</td>
<td>3</td>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Special Cubic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a ***</td>
<td>4</td>
<td>14</td>
<td>0.9783</td>
<td>0.8478</td>
<td>7.498</td>
<td>0.1236</td>
</tr>
<tr>
<td>b</td>
<td>3</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Full Cubic</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>a</td>
<td>4</td>
<td>14</td>
<td>0.9812</td>
<td>0.7367</td>
<td>4.013</td>
<td>0.3740</td>
</tr>
<tr>
<td>b</td>
<td>3</td>
<td>7</td>
<td>0.7354</td>
<td>0.5369</td>
<td>3.705</td>
<td>0.0459</td>
</tr>
<tr>
<td>McGee</td>
<td>8</td>
<td>10</td>
<td>0.9044</td>
<td>0.4524</td>
<td>2.285</td>
<td>0.1880</td>
</tr>
</tbody>
</table>

| DEPENDENT VARIABLE: ULTF - CHANGE IN ULTIMATE ELONGATION UPON AGING TWO WEEKS |
|-----------------------------------------------|---------------------------------|-------|-------------|-----|-------|----------------|
| MODEL COMPONENTS TERMS                       | R^2 | R^2_{adj} | F   | P>F | $\Sigma Res^2$ |
| 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

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

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

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)

<table>
<thead>
<tr>
<th>BRK OBS</th>
<th>SCHEFFE*</th>
<th>MCGEE</th>
<th>BRKF OBS</th>
<th>SCHEFFE*</th>
<th>MCGEE</th>
</tr>
</thead>
<tbody>
<tr>
<td>2548</td>
<td>2548</td>
<td>2924</td>
<td>2413</td>
<td>2413</td>
<td>2430</td>
</tr>
<tr>
<td>5018</td>
<td>5018</td>
<td>4348</td>
<td>2524</td>
<td>2524</td>
<td>2473</td>
</tr>
<tr>
<td>5144</td>
<td>5144</td>
<td>5194</td>
<td>2532</td>
<td>2532</td>
<td>2453</td>
</tr>
<tr>
<td>2585</td>
<td>2585</td>
<td>2615</td>
<td>2434</td>
<td>2434</td>
<td>2523</td>
</tr>
<tr>
<td>.</td>
<td>-36488</td>
<td>90477</td>
<td>.</td>
<td>5413</td>
<td>3909</td>
</tr>
<tr>
<td>4897</td>
<td>4897</td>
<td>4781</td>
<td>2427</td>
<td>2427</td>
<td>2520</td>
</tr>
<tr>
<td>5329</td>
<td>5329</td>
<td>5585</td>
<td>2549</td>
<td>2549</td>
<td>2582</td>
</tr>
<tr>
<td>5347</td>
<td>5354</td>
<td>5443</td>
<td>2626</td>
<td>2625</td>
<td>2620</td>
</tr>
<tr>
<td>2729</td>
<td>2729</td>
<td>2922</td>
<td>2586</td>
<td>2586</td>
<td>2552</td>
</tr>
<tr>
<td>2684</td>
<td>2684</td>
<td>2758</td>
<td>2569</td>
<td>2569</td>
<td>2523</td>
</tr>
<tr>
<td>2722</td>
<td>2722</td>
<td>2639</td>
<td>2586</td>
<td>2586</td>
<td>2604</td>
</tr>
<tr>
<td>2671</td>
<td>2671</td>
<td>2573</td>
<td>2663</td>
<td>2663</td>
<td>2606</td>
</tr>
<tr>
<td>2579</td>
<td>2579</td>
<td>2641</td>
<td>2403</td>
<td>2403</td>
<td>2401</td>
</tr>
<tr>
<td>2406</td>
<td>2406</td>
<td>2230</td>
<td>2385</td>
<td>2385</td>
<td>2538</td>
</tr>
<tr>
<td>2632</td>
<td>2632</td>
<td>2633</td>
<td>2542</td>
<td>2542</td>
<td>2467</td>
</tr>
<tr>
<td>2224</td>
<td>2224</td>
<td>2231</td>
<td>2254</td>
<td>2254</td>
<td>2586</td>
</tr>
</tbody>
</table>

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

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

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

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 - \((\text{mean residual}/\text{mean response}) \times 100 = \left(\frac{\sum \text{res}^2}{n}\right)^{1/2}/Y \times 100\)
Table XI

coefficients for Each Selected Scheffe (Quadratic) Equation

<table>
<thead>
<tr>
<th>MODEL EQUATION</th>
<th>SHORE HARDNESS</th>
<th>( \Delta ) SHORE HARDNESS</th>
<th>SOUNDSPEED</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>VARIABLE</td>
<td>UPON AGING TWO WEEKS</td>
<td>PARAMETER</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ESTIMATE</td>
<td>ESTIMATE</td>
</tr>
<tr>
<td>ALT</td>
<td>-115131.39</td>
<td>126809.46</td>
<td>-363607.10</td>
</tr>
<tr>
<td>OCT</td>
<td>15729.852</td>
<td>-15033.6127</td>
<td>23567.2089</td>
</tr>
<tr>
<td>C</td>
<td>-522.12962</td>
<td>198.203132</td>
<td>467.14827</td>
</tr>
<tr>
<td>PB</td>
<td>-3506.13995</td>
<td>1351.70168</td>
<td>-10483.13494</td>
</tr>
<tr>
<td>ELAS</td>
<td>403.36627</td>
<td>555.01999</td>
<td>3230.19268</td>
</tr>
<tr>
<td>ALTOCT</td>
<td>-2458074.65</td>
<td>659300.37</td>
<td>-2921781.83</td>
</tr>
<tr>
<td>ALTC</td>
<td>378206.53</td>
<td>-159670.03</td>
<td>862905.71</td>
</tr>
<tr>
<td>ALTPB</td>
<td>-384370.71</td>
<td>-49340.65764</td>
<td>-528487.89</td>
</tr>
<tr>
<td>ELALT</td>
<td>99979.27464</td>
<td>-149586.11</td>
<td>269056.39</td>
</tr>
<tr>
<td>OCTC</td>
<td>7527.8029</td>
<td>1876.35713</td>
<td>3706.1894</td>
</tr>
<tr>
<td>OCTPB</td>
<td>-37301.71026</td>
<td>35548.03619</td>
<td>-20889.11799</td>
</tr>
<tr>
<td>ELOCT</td>
<td>-16141.38342</td>
<td>16786.87908</td>
<td>-24299.26142</td>
</tr>
<tr>
<td>PBC</td>
<td>5498.23139</td>
<td>212.26877</td>
<td>20424.15633</td>
</tr>
<tr>
<td>ELC</td>
<td>-1669.34477</td>
<td>-845.95189</td>
<td>-4477.54026</td>
</tr>
<tr>
<td>ELPB</td>
<td>8325.82011</td>
<td>-5656.04060</td>
<td>15296.67192</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>MODEL EQUATION</th>
<th>TENSILE MODULUS</th>
<th>( \Delta ) TENSILE MODULUS</th>
<th>TENSILE STRENGTH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>@ 300%</td>
<td>UPON AGING TWO WEEKS</td>
<td>AT BREAK</td>
</tr>
<tr>
<td></td>
<td>VARIABLE</td>
<td>PARAMETER ESTIMATE</td>
<td>PARAMETER ESTIMATE</td>
</tr>
<tr>
<td>ALT</td>
<td>-53804867.15</td>
<td>6864881.31</td>
<td>-2348744686</td>
</tr>
<tr>
<td>OCT</td>
<td>-3143458.49</td>
<td>2465622.18</td>
<td>-15673217.64</td>
</tr>
<tr>
<td>C</td>
<td>-44804.98566</td>
<td>672.55883</td>
<td>-878590.04</td>
</tr>
<tr>
<td>PB</td>
<td>-580277.20</td>
<td>-186415.88</td>
<td>-2450328.82</td>
</tr>
<tr>
<td>ELAS</td>
<td>57023.54299</td>
<td>68367.88476</td>
<td>-259493.28</td>
</tr>
<tr>
<td>ALTOCT</td>
<td>-442291651</td>
<td>-94476632.63</td>
<td>-789806221</td>
</tr>
<tr>
<td>ALTC</td>
<td>108675987</td>
<td>2078835.14</td>
<td>328188029</td>
</tr>
<tr>
<td>ALTPB</td>
<td>-55551867.19</td>
<td>-19160589.41</td>
<td>1296896.46</td>
</tr>
<tr>
<td>ELALT</td>
<td>51346066.30</td>
<td>9112076.32</td>
<td>253057507</td>
</tr>
<tr>
<td>OCTC</td>
<td>8008707.68</td>
<td>-2448851.45</td>
<td>34957105.04</td>
</tr>
<tr>
<td>OCTPB</td>
<td>3570842.32</td>
<td>-3106381.34</td>
<td>308343.81</td>
</tr>
<tr>
<td>ELOCT</td>
<td>1734246.17</td>
<td>-1961976.30</td>
<td>10522110.60</td>
</tr>
<tr>
<td>PBC</td>
<td>359029.48</td>
<td>456844.49</td>
<td>4527564.74</td>
</tr>
<tr>
<td>ELC</td>
<td>-405043.20</td>
<td>-191747.34</td>
<td>1193060.56</td>
</tr>
<tr>
<td>ELPB</td>
<td>1857670.23</td>
<td>62970.16846</td>
<td>5023024.32</td>
</tr>
</tbody>
</table>

\( \Delta \) - change in ...
Table XI (continued)

<table>
<thead>
<tr>
<th>MODEL EQUATION</th>
<th>ΔTENSILE STRENGTH @ BREAK UPON AGING TWO WEEKS PARAMETER</th>
</tr>
</thead>
<tbody>
<tr>
<td>VARIABLE</td>
<td>ESTIMATE</td>
</tr>
<tr>
<td>ALT</td>
<td>8940640.24</td>
</tr>
<tr>
<td>OCT</td>
<td>2860411.73</td>
</tr>
<tr>
<td>C</td>
<td>47246.41829</td>
</tr>
<tr>
<td>PB</td>
<td>39477.37848</td>
</tr>
<tr>
<td>ELAS</td>
<td>20193.62830</td>
</tr>
<tr>
<td>ALTOCT</td>
<td>-38632557.41</td>
</tr>
<tr>
<td>ALTC</td>
<td>-4923080.09</td>
</tr>
<tr>
<td>ALTPB</td>
<td>-13802233.13</td>
</tr>
<tr>
<td>ELALT</td>
<td>-10806879.48</td>
</tr>
<tr>
<td>OCTC</td>
<td>-3596417.06</td>
</tr>
<tr>
<td>OCTB</td>
<td>-1846360.17</td>
</tr>
<tr>
<td>ELOCT</td>
<td>-2483596.62</td>
</tr>
<tr>
<td>PBC</td>
<td>-199563.46</td>
</tr>
<tr>
<td>ELC</td>
<td>-123706.21</td>
</tr>
<tr>
<td>ELPB</td>
<td>36230.88020</td>
</tr>
<tr>
<td>ΔULTIMATE ELONGATION UPON AGING TWO WEEKS PARAMETER</td>
<td></td>
</tr>
<tr>
<td>ESTIMATE</td>
<td></td>
</tr>
<tr>
<td>ALT</td>
<td>1756852.44</td>
</tr>
<tr>
<td>OCT</td>
<td>-52099.48603</td>
</tr>
<tr>
<td>C</td>
<td>24425.16308</td>
</tr>
<tr>
<td>PB</td>
<td>82765.41125</td>
</tr>
<tr>
<td>ELAS</td>
<td>-3978.24275</td>
</tr>
<tr>
<td>ALTOCT</td>
<td>50809472.70</td>
</tr>
<tr>
<td>ALTC</td>
<td>-6216748.03</td>
</tr>
<tr>
<td>ALTPB</td>
<td>7348086.85</td>
</tr>
<tr>
<td>ELALT</td>
<td>-1984849.82</td>
</tr>
<tr>
<td>OCTC</td>
<td>-358578.18</td>
</tr>
<tr>
<td>OCTB</td>
<td>200530.79</td>
</tr>
<tr>
<td>ELOCT</td>
<td>48954.46506</td>
</tr>
<tr>
<td>PBC</td>
<td>-188814.02</td>
</tr>
<tr>
<td>ELC</td>
<td>1862.31952</td>
</tr>
<tr>
<td>ELPB</td>
<td>-113532.79</td>
</tr>
</tbody>
</table>

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 $\leq 0.1$), 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 PE_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 PE 12.
To move screen forward use PE 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;
altoc = alt * oct; altc = alt * c; altpb = alt * pb;
ocrt = 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; rcpb = oct + pb; rcoc = c + pb;
eloc = oct + elas; raloc = alt + oct; rceo = rcel + oct;
ropa = rcpb + alt; roea = rcel + alt; roea = reoc + alt;
rcpa = rcpb + alt; rceo = reoc + c; rceo = raloc + elas;
rceo = rcpb + raloc + elas;
raoc = raloc + c: raoc = raloc + pb; rocp = raloc + c;
rope = raloc + pb; rope = reoc + pb;
** for special cubic 3 var model **
xshard = ralc * oct * rpel; xshardf = alt * rcoct * rpel;
xsspdp = alt * ralc * pb; xsten = alt + rcoct * rcel;
xstenf = xsten; xsbrk = raloc + c + elas;
xsbkf = alt + reloc + rcoc; xsultf = alt + oct * rcpbel;
run;
** for special cubic 4 var model **

\[ \begin{align*}
\text{xsh1} &= \text{altoct} \times \text{rpel}; \\
\text{xsh2} &= \text{altc} \times \text{rpel}; \\
\text{xsh3} &= \text{octc} \times \text{rpel}; \\
\text{xsh4} &= \text{altoct} \times \text{c}; \\
\text{xshf1} &= \text{altpb} \times \text{rcoct}; \\
\text{xshf2} &= \text{rcoct} \times \text{elalt}; \\
\text{xshf3} &= \text{rcoct} \times \text{elas}; \\
\text{xshf4} &= \text{altpb} \times \text{elas}; \\
\text{xss1} &= \text{altc} \times \text{reloc}; \\
\text{xss2} &= \text{altc} \times \text{pb}; \\
\text{xss3} &= \text{pbc} \times \text{reloc}; \\
\text{xss4} &= \text{altpb} \times \text{reloc}; \\
\text{xst1} &= \text{altc} \times \text{rcpb}; \\
\text{xst2} &= \text{altc} \times \text{elas}; \\
\text{xst3} &= \text{elalt} \times \text{rcpb}; \\
\text{xst4} &= \text{elc} \times \text{rcpb}; \\
\text{xstf1} &= \text{altoct} \times \text{pb}; \\
\text{xstf2} &= \text{altoct} \times \text{rcel}; \\
\text{xstf3} &= \text{altpb} \times \text{rcel}; \\
\text{xstf4} &= \text{octpb} \times \text{rcel}; \\
\text{xsbf1} &= \text{altoct} \times \text{rcpb}; \\
\text{xsbf2} &= \text{altoct} \times \text{elas}; \\
\text{xsbf3} &= \text{altpb} \times \text{elas}; \\
\text{xsbf4} &= \text{elc} \times \text{rcpb}; \\
\text{xsu1} &= \text{xss1}; \\
\text{xsu2} &= \text{xss2}; \\
\text{xsu3} &= \text{xss3}; \\
\text{xsu4} &= \text{xss4}; \\
\text{xsuf1} &= \text{xsb1}; \\
\text{xsuf2} &= \text{xsb2}; \\
\text{xsuf3} &= \text{xsb3}; \\
\text{xsuf4} &= \text{xsb4}; \\
\text{xsl} &= \text{xss1}; \\
\text{xsl2} &= \text{xss2}; \\
\text{xsl3} &= \text{xss3}; \\
\text{xsl4} &= \text{xss4}; \\
\text{xuf1} &= \text{xsb1}; \\
\text{xuf2} &= \text{xsb2}; \\
\text{xuf3} &= \text{xsb3}; \\
\text{xuf4} &= \text{xsb4}; \\
\text{xsh1} &= \text{ralo} \times (\text{ralc} - \text{oct}); \\
\text{xsh2} &= \text{ralp} \times (\text{ralc} - \text{rpel}); \\
\text{xsh3} &= \text{rpec} \times (\text{oct} - \text{rpel}); \\
\text{xshf1} &= \text{roea} \times (\text{alt} - \text{rcob}); \\
\text{xshf2} &= \text{rpea} \times (\text{alt} - \text{rpel}); \\
\text{xshf3} &= \text{roepel} \times (\text{alt} - \text{rcob}); \\
\text{xshf4} &= \text{roepel} \times (\text{alt} - \text{rpel}); \\
\text{xfs1} &= \text{roeca} \times (\text{alt} - \text{reloc}); \\
\text{xfs2} &= \text{altpb} \times (\text{alt} - \text{pb}); \\
\text{xfs3} &= \text{roecp} \times (\text{alt} - \text{reloc}); \\
\text{xft1} &= \text{roea} \times (\text{alt} - \text{rcel}); \\
\text{xft2} &= \text{roeca} \times (\text{alt} - \text{rcel}); \\
\text{xft3} &= \text{roprce} \times (\text{rcel} - \text{rcel}); \\
\text{xftf1} &= \text{xft1}; \\
\text{xftf2} &= \text{xft2}; \\
\text{xftf3} &= \text{xft3}; \\
\text{xfb1} &= \text{raopc} \times (\text{c} - \text{raocpb}); \\
\text{xfb2} &= \text{raope} \times (\text{elas} - \text{raocpb}); \\
\text{xfb3} &= \text{elc} \times (\text{c} - \text{elas}); \\
\text{xfbf1} &= \text{reoa} \times (\text{alt} - \text{reloc}); \\
\text{xfbf2} &= \text{rcpa} \times (\text{alt} - \text{rcel}); \\
\text{xfbf3} &= \text{reorcp} \times (\text{reloc} - \text{rcel}); \\
\text{xuf1} &= \text{altoc} \times (\text{alt} - \text{oct}); \\
\text{xuf2} &= \text{rcpeal} \times (\text{alt} - \text{rcpeal}); \\
\text{xuf3} &= \text{rcpeal} \times (\text{oct} - \text{rcpeal}); \\
\text{run};
\end{align*} \]
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
           octxct octxpb 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
           octxct octxpb 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 altpc rpec octx
             rpeo/p;
proc reg data = neo.data2;
  model hardf = alt pb elas rcoa altpb elalt 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 rocspb elas rpia altc elalt rope ropc
            elc/p;
proc reg data = neo.data2;
  model tenf = alt oct rce1 altoct altpb rceo octxp rcep/p;
proc reg data = neo.data2;
  model brk = raloc pb elas raoc racp raop roc pbc elc
            elpb/p;
proc reg data = neo.data2;
  model brkf = alt oct elas altoct rcpa elalt eloct rcep
              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 rcoob 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 e1c/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 = 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 rcoeb 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 rcoob elas rcoa altc elalt rope ropc e1c 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 = raocpb elas raopc raope pbc e1c 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 xsspnd /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 reorcpen xsbrkf /p;
proc reg data neo.data2;
  model ultf = alt oct altoct rcpeal rcpoel xsltf /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 xsspnd 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 reorcpen xsbrkf xfbf1 xfbf2 xfbf3 /p;
proc reg data=neo.data2;
  model ultf = alt oct altoct rcpeal rcpoel xsltf xuf1 xuf2 xuf3 /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;
pocctp = 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;
    mn2 = mn * mn;
    c2 = c * c;
    cmn = c * mn;
    cmn2 = c * mn2;
    vmn = 1/mn;
    vmw = 1/mw;
    vpbc = 1/(pb * c);
    vmn3 = 1/mn3;
    vmw2 = 1/mw2;
    valt = 1/alt;
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 * 2525.066667) + 160.3271061;
    uult = (pult * 79.2650767) + 719.333333;
    uultf = (ultf * 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 vpbv 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 vmw1d vpb1d /p;
proc reg data = std.data1;
  model ultf = vmw vmw2 vmw3 vpbv 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 $X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} \end{bmatrix}$

and $X' = \begin{bmatrix} x_{11} & \cdots & x_{m1} \\ \vdots & \ddots & \vdots \\ x_{1n} & \cdots & x_{mn} \end{bmatrix}$

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

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 $X$ become interchanged in $X'$. 
APPENDIX D

LIST OF SUGGESTED READING
General Experimental Design


Linear Regression


Model Reduction


Mixture Models


**Response Surfaces**


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.


**Ridge Regression**


Simplex Designs


**Test Statistics and Residual Analysis**


Variable Selection and Transformation


Miscellaneous


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. \( (X_i - \bar{X}) \).

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. - coefficient of variation. Expresses standard deviation as a percentage of the mean. \( \frac{\text{Std.dev}(Y) \times 100}{\text{mean } Y} \) or root MSE \( \times \frac{100}{\text{mean } Y} \).

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 \( (\text{mean square of model})/(\text{mean square of error}) \). 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). \[ R^2 = \frac{(\text{SST} - \text{SSE})/\text{SST}}{\text{SSR}/\text{SST}} = \frac{\left[ \Sigma(Y_i - \bar{Y})^2 - \Sigma(Y_i - \bar{Y})^2 \right]}{\Sigma(Y_i - \bar{Y})^2}. \]

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

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. \((Y_i - \bar{Y})\).

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 = \( \sum (Y_i - \bar{Y})^2 \).

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

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

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 - transpose 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


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


