A STUDY OF STATISTICAL AND DETERMINISTIC MODELS

FOR THE PREDICTION OF THE COMPOSITION

OF A MIXTURE

by

Raymond H. Myers, B.Sc.

Thesis submitted to the Graduate Faculty of the

Virginia Polytechnic Institute

in candidacy for the degree of

MASTER OF SCIENCE

in

Statistics

May, 1961

Blacksburg, Virginia

TABLE OF CONTENTS

CHAPTER I

INTRODUCTION

1.1 Purpose of Investigation
General Background

The determination of the concentration of known components is one of the most critical problems in large scale production of a chemical product. If a method for producing a material is found to be feasible on a research or development scale, there still lies ahead the difficult problem of controlling to a desired accuracy the output of this material on a production scale. Often there is a comparatively large difference in the results between a product synthesized in a laboratory or pilot plant and material produced commercially, even when it seems as if there is no difference in the physical conditions involved. For this reason, the problem still exists when the material is transferred to a production scale. In some instances, measurements of concentration are quite easy and require no statistical treatment except, perhaps, the preparation of routine calibration charts. However, if direct analyses for individual components are not available, or are very expensive to perform, the problem of analysis and control must be approached by a control study of models which assume relationships among easily measurable physical and chemical properties of the entire system, and the concentration of each component.

The Specific Problem

The purpose of this study is to investigate the possibility of various deterministic and statistical models in the prediction of the composition of the product in the manufacture of a casting solvent system involving: nitroglycerine (NG), 2-nitrodiphenylamine (2NDPA), and triacetin (TA). Nitroglycerine is, of course, the base substance while the latter two are stabilizers. The ultimate goal is to devise, through the study of these models, control charts or graphs which enable the quality control technician to make a quick and accurate determination of the weight per cent of each of the three components in the system. It is assumed that the technician would have data taken on samples from running production. These data will include determinations of physical properties of the entire mixture and, if necessary, analyses of one or more of the components separately.

1.2 Method of Investigation

Data

Data on specific gravity and refractive index for the entire system and spectrophotometer analysis of 2NDPA were supplied by Hercules Powder Company at Radford Arsenal in Radford, Virginia.

Twenty-eight different mixtures were pre-determined in such a way that they would supply adequate coverage of the

range of composition encountered in the process. The components were taken from production grade material. Actual concentrations of the mixtures and the results of physical and chemical analyses have been recorded in Table I.

Approach

One of the stabilizers (2NDPA) can be analyzed most easily and very accurately in a spectrophotometer [1] (See Chapter II). For the determination of the weight per cent of the other two constituents, such simple measures as density and refractive index were considered.

The first endeavor was a study of deterministic models for the prediction of concentration from these simple quantities. The laboratory technicians suggested that if one of the two physical properties were sufficient for prediction purposes, the one should be specific gravity, inasmuch as the equipment available for the determination of refractive index was considered less satisfactory than that for the former. Thus, studies were made to determine if the inclusion of refractive index, in terms of a physically meaningful function, would significantly reduce the error in prediction.

The initial model study included various investigations of deterministic models based on accepted chemical theory. From these models, equations were developed from which one could predict the weight per cent NG from known values of:

TABLE I.   Experimental Data on Specific Gravity, Refractive
           Index, and Spectrophotometer Analysis for Mixtures
           with Known Composition.

Mix-ture	PerCent NG Weighed	PerCent TA Weighed	PerCent 2NDPA Weighed	Specific Gravity 21.1°C	Refrac-tive Index	Spectro-photometer Analysis of 2NDPA
1	79.98	19.85	0	1.4774	1.460	0
2	80.06	18.91	1.00	1.4807	1.465	1.01
3	79.97	17.86	2.00	1.4793	1.470	2.01
4	80.10	16.87	3.00	1.4829	1.473	3.02
5	77.61	22.36	0	1.4664	1.462	0
6	77.60	21.38	1.00	1.4677	1.465	1.00
7	77.63	20.35	2.00	1.4686	1.468	2.00
8	77.34	19.65	2.99	1.4684	1.473	3.02
9	75.02	24.96	0	1.4524	1.457	0
10	75.03	23.95	1.00	1.4537	1.463	1.00
11	74.99	22.99	2.00	1.4549	1.465	2.00
12	74.98	22.00	3.00	1.4565	1.468	3.02
13	72.50	27.47	0	1.4410	1.458	0
14	72.50	26.48	1.00	1.4414	1.460	1.01
15	72.50	25.48	2.00	1.4426	1.463	2.00
16	72.49	24.49	3.00	1.4438	1.468	2.99
17	69.98	29.99	0	1.4279	1.457	0
18	69.98	29.00	1.00	1.4287	1.459	1.00
19	69.99	27.99	2.00	1.4291	1.462	2.03
20	69.99	26.99	3.00	1.4301	1.467	2.98
21	67.51	32.47	0	1.4157	1.455	0
22	67.50	31.47	1.00	1.4172	1.458	1.01
23	67.48	30.50	2.00	1.4183	1.460	2.00
24	67.49	29.49	3.00	1.4188	1.465	3.01
25	65.01	34.97	0	1.4013	1.453	0
26	64.98	34.00	1.00	1.4042	1.457	1.00
27	64.98	33.00	2.00	1.4060	1.460	1.98
28	64.99	31.99	3.00	1.4068	1.463	2.97

specific gravity, spectrophotometer analysis of 2NDPA, and refractive index. Each of these equations were used in "back predicting", i.e. referring them back to the original data and obtaining the error in the prediction of NG content.

The next phase of the study was to investigate various statistical models. Of primary concern were equations identical to the deterministic models except that coefficients, to be estimated by least squares methods, were placed in front of concentration (in weight per cent) in the equations. This seemed appropriate since it remains in keeping with the chemical theory involved. Many deviations from the ideal in physical chemistry are explained, rather empirically, by "activities", [2] which replace concentration, and are usually some constant multiple of the latter. These new models were translated algebraically so that the coefficients could be estimated. The error in estimating the NG content was determined for all of the models in question. Chapter II contains outlines of these models and displays numerical values of these errors.

In further searching for the best statistical model, regression equations were set up by taking both NG and actual 2NDPA concentrations as dependent variables. The latter further established the fact that spectrophotometer analysis alone is by far the most accurate predictor of 2NDPA content. The concomitant variables considered were

various functions of specific gravity, refractive index, and spectrophotometer analysis of 2NDPA. Most of the numerical work was carried out on the IBM 650. The error in estimating NG content was determined for each model (See Chapter III).

The result of the model study indicated that in terms of the standard error in estimation of NG content, the following model was found to be superior to all others:

$$W_{1i} = \alpha + \beta_1\left(\frac{1}{d_i}\right) + \beta_2 W_{3i} + \varepsilon_i , \qquad (1-1)$$

where:

W_{1i} = Weight proportion NG in mixture.

d_i = Specific gravity (or density) of mixture, measured at 21.1°C.

W_{3i} = Spectrophotometer analysis of 2NDPA content.

ε_i = Error.

This model is actually a combination of a statistical and a chemical model, for while it is in the form of a linear regression equation, it was derived from an activities model.

Charts were prepared (See Chapter IV) based on determination of 2NDPA content by spectrophotometer analysis, estimation of NG content from the model in equation (1-1), and TA determination by complementing to 100 per cent. Charts of 99 per cent confidence bounds on NG and 2NDPA determinations are also included.

CHAPTER II

DETERMINISTIC MODELS

2.1 Concentration Models

Density Alone

Functions of density alone were considered first as independent variables in deterministic models to be investigated. The following general equation may be used in ultimately arriving at a model to be investigated:

$$\frac{W_{\alpha i} + W_{\beta i} + W_{\gamma i}}{d_i} = \frac{W_{\alpha i}}{d_\alpha} + \frac{W_{\beta i}}{d_\beta} + \frac{W_{\gamma i}}{d_\gamma} \qquad (2\text{-}1)$$

where:

$W_{\alpha i}$ = weight per cent of component α in mixture i .

$W_{\beta i}$ =   "   "   "   "   "   " β "   "   " .

$W_{\gamma i}$ =   "   "   "   "   "   " γ "   "   " .

d_α = specific gravity of pure component α .

d_β =   "   "   "   "   "   " β .

d_γ =   "   "   "   "   "   " γ .

This equation holds for a completely ideal, non contracting ternary mixture of liquids. The mixture in question, as used in the explosives industry, is a liquid. However, 2NDPA and TA are both powders in the pure state. Therefore equation (2-1) cannot be applied directly for this system.

Since a similar situation was to occur not only in all of the deterministic models under study but also in the statistical models in which activity coefficients are applied, it was necessary to revise the model in terms of certain reference substances which are liquids. The following paragraph outlines the procedure.

From the 28 sample data, three were chosen as "reference" samples. The mixtures corresponding to these samples were assumed to be theoretically "pure" substances. Let us denote these reference components A, B, C. Then, for any one of the 28 observations, one can calculate W_{Ai} (weight per cent A in the i'th mixture), W_{Bi} (weight per cent B in the i'th mixture), and W_{Ci} (weight per cent C in the i'th mixture). It was necessary to find these as functions of W_{1i} for each of the 28 observations in order that equation (2-1) could be tested as a means of predicting NG content. The method used for this purpose is as follows.

Let A, B, C denote three particular mixtures (e.g., three of the substances no. 1 to 28) that were assigned to be reference components.

Let

a_{11} = weight proportion NG in component A.

a_{12} = " " TA " " ".

a_{13} = " " 2NDPA " " ".

a_{21} = weight proportion NG in component B.

a_{22} = "   "   TA "   "   ".

etc.

a_{33} = "   "   2NDPA "   "   C.

(i)   Form the matrix A where

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

(ii)   Find A^{-1}.

(iii) Considering now, a given mixture i, let W_{1i} , W_{2i} , W_{3i} denote weight proportion NG, TA, and 2NDPA respectively in the mixture i.

(iv)   Find W_{Ai} , W_{Bi}, W_{Ci} for observation i in the following manner:

$$[W_{Ai} , W_{Bi} , W_{Ci}] = [W_{1i} , W_{2i} , W_{3i}]A^{-1} \qquad (2-2)$$

These quantities W_{Ai} , W_{Bi} , W_{Ci} , denote the theoretical proportions of components A, B, and C contained in mixture i. Thus with these quantities in mind, and since "components" A, B, C (as combinations of the three basic compounds) are liquids, one can translate back to equation (2-1). Assuming W_{3i} is analyzed directly by the spectrophotometer method, and since $W_{2i} = 1 - W_{1i} - W_{3i}$, then W_{Ai} , W_{Bi} , W_{Ci} can be found from equation (2-2) as functions of W_{1i} . Thus for a

given observation of specific gravity, say d_i, since $W_{Ai} + W_{Bi} + W_{Ci} = 1$, the left hand side of (2-1) is a function of the observation and can predict the right hand side, itself a function of W_{1i}. This can result in a systematic equation for predicting weight proportion NG from the model of equation (2-1). The following is a derivation of such an equation:

Beginning with the matrix equation (2-2) above,

$$[W_{Ai}, W_{Bi}, W_{Ci}] = [W_{1i}, W_{2i}, W_{3i}] \begin{bmatrix} a^{11} & a^{12} & a^{13} \\ a^{21} & a^{22} & a^{23} \\ a^{31} & a^{32} & a^{33} \end{bmatrix}$$

where: $a^{ij} = (ij)$ element of A^{-1}

and replacing W_{2i} with $1 - W_{1i} - W_{3i}$, this equation reduces to:

$$[W_{Ai}, W_{Bi}, W_{Ci}] = [W_{1i}, 1-W_{1i}-W_{3i}, W_{3i}] \begin{bmatrix} a^{11} & a^{12} & a^{13} \\ a^{21} & a^{22} & a^{23} \\ a^{31} & a^{32} & a^{33} \end{bmatrix} \quad (2-3)$$

Thus one can obtain W_{Ai}, W_{Bi}, W_{Ci} in terms of W_{1i}, yielding the following:

$$W_{Ai} = W_{1i} a^{11} + (1 - W_{1i} - W_{3i}) a^{21} + W_{3i} a^{31} .$$

$$W_{Bi} = W_{1i} a^{12} + (1 - W_{1i} - W_{3i}) a^{22} + W_{3i} a^{32} .$$

$$W_{Ci} = W_{1i} a^{13} + (1 - W_{1i} - W_{3i}) a^{23} + W_{3i} a^{33} .$$

Substituting these expressions for W_{Ai}, W_{Bi}, W_{Ci} in the

prediction model (2-1) yields:

$$\frac{1}{d_i} = \frac{1}{d_A}[\hat{W}_{1i}(a^{11}-a^{21})+a^{21}+W_{3i}(a^{31}-a^{21})]+\frac{1}{d_B}[\hat{W}_{1i}(a^{12}-a^{22})$$

$$+a^{22}+W_{3i}(a^{32}-a^{22})]+\frac{1}{d_C}[\hat{W}_{1i}(a^{13}-a^{23})$$

$$+a^{23}+W_{3i}(a^{33}-a^{23})] \ . \tag{2-4}$$

Here W_{1i} has been replaced by $\hat{W}_{1i}$ to indicate that W_{1i} is expressed in terms of other observed quantities. After simplification;

$$\hat{W}_{1i}[(a^{11}-a^{21})\frac{1}{d_A}+(a^{12}-a^{22})\frac{1}{d_B}+(a^{13}-a^{23})\frac{1}{d_C}] = \frac{1}{d_i}$$

$$-[a^{21}+W_{3i}(a^{31}-a^{21})]\frac{1}{d_A}-[a^{22}+W_{3i}(a^{32}-a^{22})]\frac{1}{d_B}$$

$$-[a^{23}+W_{3i}(a^{33}-a^{23})]\frac{1}{d_C} \ , \tag{2-5}$$

and, finally

$$\hat{W}_{1i} = \frac{\frac{1}{d_i} - k_2 - k_3 \ W_{3i}}{k_1} \tag{2-6}$$

where:

$\quad \hat{W}_{1i}$ = prediction of weight proportion NG in a

$\qquad\qquad$ mixture i of unknown composition.

$\quad d_i$ = observed specific gravity of mixture i.

$\quad W_{3i}$ = 2NDPA content.

$\quad k_1 = [(a^{11}-a^{21})\frac{1}{d_A}+(a^{12}-a^{22})\frac{1}{d_B}+(a^{13}-a^{23})\frac{1}{d_C}] \ .$

$$k_2 = a^{21}(\frac{1}{d_A}) + a^{22}(\frac{1}{d_B}) + a^{23}(\frac{1}{d_C}) \ .$$

$$k_3 = (a^{31} - a^{21})\frac{1}{d_A} + (a^{32} - a^{22})\frac{1}{d_B} + (a^{33} - a^{23})\frac{1}{d_C} \ .$$

d_A, d_B, d_C = known specific gravity of reference components
A, B, C.

a^{ij} = (ij) element of A^{-1} .

Then, as stated before, $W_{2i} = 1 - W_{1i} - W_{3i}$. Thus from the
deterministic model of (2-1), the composition of a mixture i
can be predicted from an experimental determination of speci-
fic gravity and the spectrophotometer analysis of 2NDPA in
the mix.

The above model was investigated from three different
sets of reference components. These sets were samples of
those which fairly well covered the entire range of composi-
tion and of density and refractive index. The sets which
were chosen were: 2, 16, 25; 4, 15, 25; 2, 16, 28. The
latter set was found to be superior to the first two. Back
predictions were made in order that the magnitude of the
error could be assessed. Table II gives the result of these
calculations, including the error for each sample mixture
and the estimated error variance for reference components
2, 16, 28.

Density and Refractive Index

Lorenz-Lorentz Law

Density and refractive index in non-polar,non-contractible,

TABLE II. Errors of Prediction of NG content for Deterministic Model (2-1); Results for Reference Components 2, 16, 28.

Mix-ture	W_{1i} (Weight Per Cent)	$\hat{W}_{1i}$ (Weight Per Cent)	$W_{1i}-\hat{W}_{1i}$
1	79.98	79.80	0.18
2	80.06	80.17	-0.11
3	79.97	79.65	0.32
4	80.10	80.08	0.02
5	77.61	77.71	-0.10
6	77.60	77.70	-0.10
7	77.63	77.63	0
8	77.34	77.34	0
9	75.02	75.00	0.02
10	75.03	75.00	0.03
11	74.99	74.99	0
12	74.98	75.05	-0.07
13	72.50	72.76	-0.26
14	72.50	72.59	-0.09
15	72.50	72.57	-0.07
16	72.49	72.56	-0.07
17	69.98	70.13	-0.15
18	69.98	70.05	-0.07
19	69.99	69.88	0.11
20	69.99	69.83	0.16
21	67.51	67.65	-0.15
22	67.50	67.71	-0.21
23	67.48	67.68	-0.20
24	67.49	67.54	-0.05
25	65.01	64.66	0.35
26	64.98	65.02	-0.04
27	64.98	65.14	-0.16
28	64.99	65.06	-0.07

These are errors before mixtures 3 and 25 were eliminated (See Chapter III). However, the following results are recorded for the case in which 3 and 25 are excluded:

$$SSE = 0.3623 \qquad MSE = 0.0129 .$$

transparent mixtures combine in the Lorenz-Lorentz formula
[3]:

$$\frac{n_i^2 - 1}{n_i^2 + 2} = C_i \, d_i \, ,$$

where C_i is a constant depending on the resonance frequencies
of a liquid or solution and n_i is the refractive index of
the substance in question. Then according to this law,
$\frac{\nu_i}{d_i} = C_i$, where $\nu_i = \frac{n_i^2 - 1}{n_i^2 + 2}$. This relationship led to the
investigation of the following model;

$$\frac{\nu_i}{d_i} = W_{Ai} \, \frac{\nu_A}{d_A} + W_{Bi} \, \frac{\nu_B}{d_B} + W_{Ci} \, \frac{\nu_C}{d_C} \qquad (2\text{-}7)$$

where W_{Ai} , W_{Bi} , W_{Ci} are as before and ν_A , ν_B , ν_C are
defined accordingly for the reference components. Recall
that this is similar to the model of equation (2-1) with $\frac{1}{d}$
replaced here by $\frac{\nu}{d}$. Thus (2-7) was investigated as a
possible model for prediction.

Due to its similarity to equation (2-1) for density
alone, (2-7) was treated in the same manner, i.e. reference
components were chosen; W_{Ai} , W_{Bi} , W_{Ci} were found in terms
of W_{1i} ; and the left hand side was used to predict W_{1i} as
before. A prediction equation similar to (2-6) was derived.
The equation is as follows:

$$\hat{W}_{1i} = \frac{\dfrac{\nu_i}{d_i} - k_2' - k_3' \, W_{3i}}{k_1'} \qquad (2\text{-}8)$$

where

$$\nu_i = \frac{n_i^2 - 1}{n_i^2 + 2} \quad . \quad (n_i, \text{ the refractive index of a mixture i)}$$

$$k_1' = [(a^{11}-a^{21})\frac{\nu_A}{d_A} + (a^{12}-a^{22})\frac{\nu_B}{d_B} + (a^{13}-a^{23})\frac{\nu_C}{d_C}] \quad .$$

$$k_2' = a^{21}(\frac{\nu_A}{d_A}) + a^{22}(\frac{\nu_B}{d_B}) + a^{23}(\frac{\nu_C}{d_C}) \quad .$$

$$k_3' = (a^{31}-a^{21})\frac{\nu_A}{d_A} + (a^{32}-a^{22})\frac{\nu_B}{d_B} + (a^{33}-a^{23})\frac{\nu_C}{d_C} \quad .$$

$\hat{W}_{2i}$ can be found as before, i.e. $\hat{W}_{2i} = 1 - \hat{W}_{1i} - W_{3i}$.

Back prediction was carried out for reference components 4, 15, 25 and 2, 16, 28. Once again the results showed that the latter was superior. However it was also revealed that the model (2-1) which included densities only was far more accurate than the one described by equation (2-7). Table III shows the results of testing this model for reference components 2, 16, 28. Note that for some sample mixtures, the error in prediction of NG content is in excess of 2.0 per cent, even for the superior set of reference components.

Separation of Densities and Refractive Indices

Recall that, $\frac{n_A^2 - 1}{n_A^2 + 2} = C_A d_A$ where C_A is a constant

depending on substance A. Thus denoting $\frac{n_A^2 - 1}{n_A^2 + 2}$ as ν_A ,

one can say that, for reference substance A, $\nu_A = C_A d_A$ and

TABLE III. Errors of Prediction for Model (2-7), Using Reference Components 2, 16, 28.

Mixture	W_{1i} (Weight Per Cent)	$\hat{W}_{1i}$ (Weight Per Cent)	$W_{1i}-\hat{W}_{1i}$
1	79.98	80.32	-0.34
2	80.06	80.21	-0.15
3	79.97	78.75	1.22
4	80.10	80.35	-0.25
5	77.61	75.51	2.10
6	77.60	76.39	1.21
7	77.63	77.18	0.45
8	77.34	76.01	1.33
9	75.02	75.40	-0.38
10	75.03	73.83	1.20
11	74.99	75.49	-0.50
12	74.98	76.47	-1.49
13	72.50	71.15	1.35
14	72.50	72.56	-0.06
15	72.50	73.39	-0.89
16	72.49	72.59	-0.10
17	69.98	67.97	2.01
18	69.98	69.49	0.49
19	69.99	70.05	-0.06
20	69.99	69.16	0.83
21	67.51	65.86	1.65
22	67.50	66.74	0.76
23	67.48	68.34	-0.86
24	67.49	67.25	0.24
25	65.01	63.01	2.00
26	64.98	63.47	1.52
27	64.98	64.43	0.55
28	64.99	65.09	-0.10

$$SSE = 26.412 \qquad MSE = 1.019$$

for a given arbitrary mixture i, $v_i = C_i d_i$. Now, from equation (2-1), $\frac{1}{d_i} = W_{Ai} \frac{1}{d_A} + W_{Bi} \frac{1}{d_B} + W_{Ci} \frac{1}{d_C}$ and since $d_i = v_i/C_i$, this equation can be written as;

$$\frac{C_i}{v_i} = W_{Ai} \frac{C_A}{v_A} + W_{Bi} \frac{C_B}{v_B} + W_{Ci} \frac{C_C}{v_C} \; .$$

Thus, if in a limited range of mixtures we assume $C_A = C_B = C_C = C_i$, then the following equation holds:

$$\frac{1}{v_i} = W_{Ai} \frac{1}{v_A} + W_{Bi} \frac{1}{v_B} + W_{Ci} \frac{1}{v_C} \; . \tag{2-9}$$

This is an equation for refractive index only. We can now combine (2-9) and (2-1) in the following manner:

$$\frac{\alpha}{d_i} + \frac{\beta}{v_i} = W_{Ai}(\frac{\beta}{v_A} + \frac{\alpha}{d_A}) + W_{Bi}(\frac{\beta}{v_B} + \frac{\alpha}{d_B}) + W_{Ci}(\frac{\beta}{v_C} + \frac{\alpha}{d_C}) \; .$$

Since α and β are proportionality factors, we can say, without loss of generality that,

$$(\frac{\alpha}{d_i} + \frac{1-\alpha}{v_i}) = W_{Ai}(\frac{\alpha}{d_A} + \frac{1-\alpha}{v_A}) + W_{Bi}(\frac{\alpha}{d_B} + \frac{1-\alpha}{v_B})$$

$$+ W_{Ci}(\frac{\alpha}{d_C} + \frac{1-\alpha}{v_C}) \; . \tag{2-10}$$

which further reduces to the following:

$$[\frac{1}{v_i} + \alpha(\frac{1}{d_i} - \frac{1}{v_i})] = [\frac{1}{v_A} + \alpha(\frac{1}{d_A} - \frac{1}{v_A})]W_{Ai} +$$

$$[\frac{1}{v_B} + \alpha(\frac{1}{d_B} - \frac{1}{v_B})]W_{Bi} + [\frac{1}{v_C} + \alpha(\frac{1}{d_C} - \frac{1}{v_C})]W_{Ci} \; . \tag{2-11}$$

Now since $\frac{1}{v_i} = \frac{W_{Ai}}{v_A} + \frac{W_{Bi}}{v_B} + \frac{W_{Ci}}{v_C}$

from equation (2-9), we have regardless of α, the following deterministic equation:

$$\left(\frac{1}{d_i} - \frac{1}{v_i}\right) = \left(\frac{1}{d_A} - \frac{1}{v_A}\right)W_{Ai} + \left(\frac{1}{d_B} - \frac{1}{v_B}\right)W_{Bi} +$$

$$\left(\frac{1}{d_C} - \frac{1}{v_C}\right)W_{Ci} \quad . \tag{2-12}$$

Equation (2-12) represents a deterministic model derived by separating two equations using density and refractive index, and then combining them to form a model involving both. Notice also that this model is of the same general form as the first two models discussed in this chapter. Therefore back prediction was also carried out for this model. Only reference components 2, 16, 28 were used. Results indicated this model to be inferior to either of the first two. The maximum error observed was 9.24 per cent, and for nine other mixtures, it was in excess of 5 per cent. Hence this model was discarded as a possibility for prediction of the content of the mixture.

2.2 Spectrophotometer Method for Estimation of 2NDPA Content

The descriptions of possible methods of prediction which have been discussed thus far in this chapter have all been made under the assumption that the weight per cent 2NDPA was known. This was based on our knowledge that a very simple and highly accurate analysis is available for this substance by means of a spectrophotometer.

A spectrophotometer measures the relative amount of light transmitted by an absorbing substance at any desired controlled wave length. Readings of either the amount of light transmitted or amount of light absorbed are taken so that the instrument can be calibrated for a particular substance. The existing chart which calibrates the spectrophotometer used in this study for 2NDPA was evaluated by means of a regression line for actual 2NDPA content in terms of weight per cent 2NDPA from the spectrophotometer analysis. The results showed that the error was extremely small.

A chart of 99 per cent confidence bounds on weight per cent 2NDPA was prepared for use as a supplement to the control charts (See Chapter IV, Figure 1). In essence, these results show that all of these confidence bounds are within 0.04 per cent of the predicted value from the calibration chart.

CHAPTER III

STATISTICAL MODELS

3.1 Activity Models

The activity models investigated for this study were based on the same general chemical theory that was discussed in Chapter II. In this case, however, concentrations are replaced by activities which are expressed in the following form:

$$f_{Ai} = C_A W_{Ai}$$

where:

f_{Ai} = activity of component A in mixture i,

C_A = activity coefficient of component A, and

W_{Ai} = weight per cent of component A in mixture i.

Thus, certain of the models investigated in Chapter II were reconsidered, this time as activity models. Consider the following general equation which represents the form in which all of concentration models were expressed:

$$r_i = W_{Ai} r_A + W_{Bi} r_B + W_{Ci} r_C \qquad (3-1)$$

where:

r_i = function of the experimental determination;
$1/d_i$, v/d_i , $(1/v_i - 1/d_i)$ and

r_A, r_B, r_C = function of the same known physical property

for the reference components, i.e. $1/d_A$, $1/d_B$,

$1/d_C$ etc.

Activities were introduced into this general model under the following two categories; 1.) Assuming the activity of the mixture as constant 2.) Assuming the activity of the mixture as not constant, but as the sum of the activities of the three components.

Models Assuming Activity of the Mixture Constant

Equation (3-1) is changed to the following by applying activity coefficients under the assumption that the activity of the mixture is constant:

$$\delta \, r_i = \alpha \, W_{Ai} \, r_A + \beta \, W_{Bi} \, r_B + \gamma \, W_{Ci} \, r_C + \varepsilon_i \qquad (3-2)$$

where α, β, γ, and δ are activity coefficients. Rearranging this equation in order to free the left hand side of unknown constants,

$$r_i = \alpha^* \, W_{Ai} + \beta^* \, W_{Bi} + \gamma^* \, W_{Ci} \quad . \qquad (3-3)$$

where:

$$\alpha^* = \frac{\alpha}{\delta} \, r_A \quad .$$

$$\beta^* = \frac{\beta}{\delta} \, r_B \quad .$$

$$\gamma^* = \frac{\gamma}{\delta} \, r_C \quad .$$

In order that α^*, β^*, γ^* could be estimated by least squares

methods, W_{Ai}, W_{Bi}, W_{Ci} were expressed in terms of W_{1i}, and the equation simplified algebraically. This was done as follows:

From equation (2-3) we know that;

$$W_{Ai} = W_{1i}(a^{11} - a^{21}) + a^{21} + W_{3i}(a^{31} - a^{21}),$$

$$W_{Bi} = W_{1i}(a^{12} - a^{22}) + a^{22} + W_{3i}(a^{32} - a^{22}), \text{ and}$$

$$W_{Ci} = W_{1i}(a^{13} - a^{23}) + a^{23} + W_{3i}(a^{33} - a^{23}),$$

where $a^{ij} = (ij)$ element of A^{-1} as defined in equation (2-3). Substituting these expressions into (3-3) we have;

$$r_i = \alpha*[W_{1i}(a^{11}-a^{21})+a^{21}+W_{3i}(a^{31}-a^{21})]+$$

$$\beta*[W_{1i}(a^{12}-a^{22})+a^{22}+W_{3i}(a^{32}-a^{22})+\gamma*[W_{1i}(a^{13}-a^{23})+$$

$$a^{23}+W_{3i}(a^{33}-a^{23})] . \tag{3-4}$$

Transferring W_{1i} to the left hand side yields;

$$W_{1i}[\alpha*(a^{11}-a^{21})+\beta*(a^{12}-a^{22})+\gamma*(a^{13}-a^{23})] =$$

$$r_i-(a^{21}\alpha*+a^{22}\beta*+a^{23}\gamma*)-W_{3i}[(a^{31}-a^{21})\alpha*+(a^{32}-a^{22})\beta*+$$

$$(a^{33}-a^{23})\gamma*] \tag{3-5}$$

and finally this reduces to;

$$\hat{W}_{1i} = \frac{r_i}{k_1} - \frac{k_2}{k_1} - \frac{k_3}{k_1} W_{3i} \tag{3-6}$$

where:

$$k_1 = (a^{11}-a^{21})\alpha* + (a^{12}-a^{22})\beta* + (a^{13}-a^{23})\gamma* .$$

$$k_2 = a^{21}\alpha* + a^{22}\beta* + a^{23}\gamma* .$$

$$k_3 = (a^{31}-a^{21})\alpha* + (a^{32}-a^{22})\beta* + (a^{33}-a^{23})\gamma* .$$

Equation (3-6) is of the form:

$$W_{1i} = \mu + \beta** X_{1i} + \gamma** X_{2i} + \epsilon_i \qquad (3-7)$$

where:

$$X_{1i} = r_i .$$

$$X_{2i} = W_{3i} .$$

$$\mu = - \frac{a^{21}\alpha* + a^{22}\beta* + a^{23}\gamma*}{(a^{11}-a^{21})\alpha* + (a^{12}-a^{22})\beta* + (a^{13}-a^{23})\gamma*} . \qquad (3-8)$$

$$\beta** = \frac{1}{(a^{11}-a^{21})\alpha* + (a^{12}-a^{22})\beta* + (a^{13}-a^{23})\gamma*} . \qquad (3-9)$$

$$\gamma** = - \frac{(a^{31}-a^{21})\alpha* + (a^{32}-a^{22})\beta* + (a^{33}-a^{23})\gamma*}{(a^{11}-a^{21})\alpha* + (a^{12}-a^{22})\beta* + (a^{13}-a^{23})\gamma*} . \qquad (3-10)$$

Thus μ, $\beta**$, and $\gamma**$ can now be easily estimated by least squares. After this is done, one can replace μ, $\beta**$, and $\gamma**$ of equations (3-8), (3-9), and (3-10) with $\hat{\mu}$, $\hat{\beta}**$, and $\hat{\gamma}**$, their estimates, in order to obtain estimates of $\alpha*$, $\beta*$, and $\gamma*$. Note that this can be done systematically in the following manner: From equations (3-8), (3-9), and (3-10) the following three equations are derived:

$$a^{11} \hat{\alpha}* + a^{12} \hat{\beta}* + a^{13} \hat{\gamma}* = 1/\hat{\beta}** - \hat{\mu}/\hat{\beta}** . \qquad (3-11)$$

$$a^{21} \hat{\alpha}* + a^{22} \hat{\beta}* + a^{23} \hat{\gamma}* = -\hat{\mu}/\hat{\beta}** \ . \tag{3-12}$$

$$a^{31} \hat{\alpha}* + a^{32} \hat{\beta}* + a^{33} \hat{\gamma}* = -\hat{\mu}/\hat{\beta}** - \hat{\gamma}**/\hat{\beta}** \ . \tag{3-13}$$

Thus $\hat{\alpha}*$, $\hat{\beta}*$, and $\hat{\gamma}*$ can be obtained from the following matrix equation:

$$A^{-1} \begin{bmatrix} \hat{\alpha}* \\ \hat{\beta}* \\ \hat{\gamma}* \end{bmatrix} = \begin{bmatrix} \dfrac{1 - \hat{\mu}}{\hat{\beta}**} \\ -\hat{\mu}/\hat{\beta}** \\ \dfrac{-\hat{\gamma}**-\hat{\mu}}{\hat{\beta}**} \end{bmatrix}$$

and thus

$$\begin{bmatrix} \hat{\alpha}* \\ \hat{\beta}* \\ \hat{\gamma}* \end{bmatrix} = A \begin{bmatrix} \dfrac{1 - \hat{\mu}}{\hat{\beta}**} \\ -\hat{\mu}/\hat{\beta}** \\ \dfrac{-\hat{\gamma}**-\hat{\mu}}{\hat{\beta}**} \end{bmatrix} \tag{3-14}$$

After finding these estimates one needs only to divide by r_A, r_B, r_C respectively to obtain estimates of the ratio of activity coefficients, i.e.

$$\hat{\alpha}/\hat{\delta} = \hat{\alpha}*/r_A \tag{3-15}$$

$$\hat{\beta}/\hat{\delta} = \hat{\beta}*/r_B \tag{3-16}$$

$$\hat{\gamma}/\hat{\delta} = \hat{\gamma}*/r_C \tag{3-17}$$

The results for the model in which $r_i = 1/d_i$ are as follows:

$$\hat{\mu} = 3.60788; \quad \hat{\beta}** = -4.15189; \quad \hat{\gamma}** = -0.225008.$$

Applying (3-14), (3-15), (3-16), and (3-17), the estimates

of the ratios of activity coefficients were found to be:

$$\hat{\alpha}/\hat{\delta} = 1.0215 \ .$$
$$\hat{\beta}/\hat{\delta} = 0.9927 \ .$$
$$\hat{\gamma}/\hat{\delta} = 0.9906 \ .$$

These values are close to unity. In the event that the true values of these ratios did not differ significantly from unity, it would be useless to continue since this model would then show no significant improvement over the deterministic model. However, if it is found that they do differ significantly from 1, the activity model would provide a significant reduction in the error above that of the deterministic model, and hence would be an improvement over the latter.

The determination of whether or not the activity model results in a significant reduction in the error of the deterministic model, i.e., to test the hypothesis that the ratio of the activity coefficients are significantly different from unity was carried out in the following manner:

(a) A hypothesis was formulated in terms of μ, β^{**}, and γ^{**}, assuming $\alpha/\delta = \beta/\delta = \gamma/\delta = 1$. Therefore $\alpha^* = r_A$, $\beta^* = r_B$, and $\gamma^* = r_C$.

(b) The hypothesized values of μ, β^{**}, and γ^{**} were found by replacing α^*, β^*, and γ^* with r_A, r_B, r_C in equations (3-8), (3-9), and (3-10), i.e.

$$\beta_o = \frac{1}{(a^{11}-a^{21})r_A+(a^{12}-a^{22})r_B+(a^{13}-a^{23})r_C} \quad ,$$

$$\gamma_o = - \frac{(a^{31}-a^{21})r_A+(a^{32}-a^{22})r_B+(a^{33}-a^{23})r_C}{(a^{11}-a^{21})r_A+(a^{12}-a^{22})r_B+(a^{13}-a^{23})r_C} \quad , \text{ and}$$

$$\mu_o = - \frac{a^{21}r_A + a^{22}r_B + a^{23}r_C}{(a^{11}-a^{21})r_A+(a^{12}-a^{22})r_B+(a^{13}-a^{23})r_C} \quad .$$

(c) The hypothesis H_o: $\mu = \mu_o$

$$\beta^{**} = \beta_o$$

$$\gamma^{**} = \gamma_o$$

was tested against the alternative that at least one of these is not satisfied. The hypothesis in (c) can readily be tested after reformulating the regression equation (3-7) to the following:

$$W_{1i} = \mu^* + \beta^{**}(x_{1i}-\bar{x}_1) + \gamma^{**}(x_{2i}-\bar{x}_2) + \epsilon \qquad (3-18)$$

where:

$$\mu^* = \mu + \beta^{**} \bar{x}_1 + \gamma^* \bar{x}_2 \ .$$

Hence our hypothesis reduces to the following:

H_o: $\mu^* = \mu_o + \beta_o \bar{x}_1 + \gamma_o \bar{x}_2 = \mu_o^* \ .$

$$\beta^{**} = \beta_o \ .$$

$$\gamma^{**} = \gamma_o \ .$$

Now from the following general linear hypothesis

H_o: $C\underline{\xi} = 0 \ ,$

where in our case:

$$C = \text{identity matrix } I_3$$

$$\underline{\xi} = \begin{bmatrix} \mu^* - \mu_o^* \\ \beta^{**} - \beta_o \\ \gamma^{**} - \gamma_o \end{bmatrix} ,$$

SSH reduces to the following computational form:

$$\text{SSH} = N(\hat{\mu}^*-\mu_o^*)^2 + (\hat{\beta}^{**}-\beta_o)^2 \, g_{11} + 2(\hat{\beta}^{**}-\beta_o)$$
$$\cdot (\hat{\gamma}^{**}-\gamma_o) \, g_{12} + (\hat{\gamma}^{**}-\gamma_o)^2 \, g_{22} \qquad (3\text{-}19)$$

with three degrees of freedom; where $g_{rs} = \sum\limits_{i}(x_{ri}-\bar{x}_r)(x_{si}-\bar{x}_s)$.

This work was done after it was decided that mixtures 3 and 25 should be eliminated. The justification for their elimination is given in section (3.2) of this chapter. The numerical results of the test of the hypothesis is as follows:

$$\hat{\mu}^* = 0.72076. \qquad\qquad \mu_o^* = 0.72566880.$$

$$\hat{\beta}^{**} = -4.151894. \qquad\quad \beta_o = -4.117260.$$

$$\hat{\gamma}^{**} = -0.2250077. \qquad \gamma_o = -0.2498421.$$

$$N = 26. \qquad\qquad\qquad g_{11} = 0.003408747.$$

$$g_{12} = 0.00017936512. \qquad g_{22} = 0.00324718962.$$

Hence SSH $= 1.489143 \times 10^{-5}$ with 3 d.f.,

MSH $= 1.489143 \times 10^{-5}/3 = 4.9638 \times 10^{-6}$,

SSE $= 2.2879 \times 10^{-5}$ with 23 d.f.,

MSE $= 2.2879 \times 10^{-5}/23 = 0.99474 \times 10^{-6}$,

and thus F $= 4.9638 \times 10^{-6}/9.99474 \times 10^{-6} = \underline{4.990}$ with

(3, 23) d.f. After comparing this with $F_{0.99}(3, 23) = 4.76$,

it is concluded that the ratios of the activity coefficients
of model (3-2) do differ significantly from unity, that the
reduction in the error of the deterministic model of equation
(3-1) with $d_i = 1/r_i$ is significant, and hence that the
model is significantly improved by applying activity coeffi-
cients. It may be noted that the numerical value of SSE for
the deterministic model is 3.6227×10^{-5} (Table II in
Chapter II) and does not quite agree with
SSH + SSE $= 3.7770 \times 10^{-5}$ for the activities model. The
reason for this slight discrepancy is that the W_{3i} value
used in the activity model was the spectrophotometer analysis
value while the corresponding value in the deterministic
model is the actual 2NDPA content. Table IV gives the
results of the regression analysis study of equation (3-7)
with $r_i = 1/d_i$. This was done after mixture 3 and 25 were
eliminated (See Section 3.2). Note that the results in
Table IV are based on weight per cent while the numerical
values above are based on weight proportions. This change
of scale in the final results was made for ease of notation.

Model Assuming Activity of Mixture Not Constant

If we eliminate the assumption of constant activity for
the mixture, we may reduce the model to one involving two
parameters only, β and γ, which are the ratios of activity
coefficients of reference components B and C to that of A.

TABLE IV. Results of Error Study of Model of Equation
           (3-7) with $d_i = 1/r_i$.

Mix-ture	W_{1i}(Weight Per Cent)	$\hat{W}_{1i}$(Weight Per Cent)	$W_{1i}-\hat{W}_{1i}$
1	79.98	79.76	0.22
2	80.06	80.16	-0.10
4	80.10	80.12	-0.02
5	77.61	77.65	-0.04
6	77.60	77.67	-0.08
7	77.63	77.63	0
8	77.34	77.36	-0.02
9	75.02	74.92	0.10
10	75.03	74.95	0.08
11	74.99	74.97	0.02
12	74.98	75.05	-0.07
13	72.50	72.66	-0.16
14	72.50	72.52	-0.02
15	72.50	72.53	-0.03
16	72.49	72.55	-0.06
17	69.98	70.02	-0.04
18	69.98	69.96	0.02
19	69.99	69.81	0.18
20	69.99	69.80	0.19
21	67.51	67.51	0
22	67.50	67.60	-0.10
23	67.48	67.60	-0.12
24	67.49	67.48	0.01
26	64.98	64.89	0.09
27	64.98	65.04	-0.06
28	64.99	64.99	0

SSE = 2.288×10^{-1}

MSE = 0.99474×10^{-2} (23 d.f.)

In terms of per cents, $\hat{\mu} = 360.788$, $\hat{\beta}** = -415.189$,

$\hat{\gamma}** = -0.225008$, Var $\hat{\beta}** = 2.9267$, $t_{\beta**} = 241.39$,

Var $\hat{\gamma}** = 3.0723 \times 10^{-4}$, $t_{\gamma**} = -12.8617$.

This model is as follows:

$$(W_{Ai} + \beta\, W_{Bi} + \gamma\, W_{Ci}) r_i = W_{Ai}\, r_A + \beta\, W_{Bi}\, r_B +$$

$$\gamma\, W_{Ci}\, r_C + \epsilon_i \tag{3-20}$$

where r_i is taken here as $1/d_i$. One can solve for ϵ_i as follows:

$$\epsilon_i = r_i(W_{Ai} + \beta\, W_{Bi} + \gamma\, W_{Ci}) - W_{Ai}\, r_A - \beta\, W_{Bi}\, r_B - \gamma\, W_{Ci}\, r_C$$

$$= (r_i - r_A) W_{Ai} + \beta(r_i - r_B) W_{Bi} + \gamma(r_i - r_C) W_{Ci} \ . \tag{3-21}$$

$$\epsilon_i^{\,2} = [(r_i - r_A) W_{Ai}]^2 + [\beta(r_i - r_B) W_{Bi}]^2 + [\gamma(r_i - r_C) W_{Ci}]^2$$

$$+ 2\,\beta\, W_{Ai}\, W_{Bi}(r_i - r_A)(r_i - r_B) + 2\,\gamma\, W_{Ai}\, W_{Ci}(r_i - r_A)$$

$$\cdot (r_i - r_C) + 2\,\beta\,\gamma(r_i - r_B)(r_i - r_C) W_{Bi}\, W_{Ci} \ .$$

Differentiating $\Sigma\,\epsilon_i^{\,2}$ with respect to β and γ to obtain least squares estimators yields:

$$\frac{\partial\, \Sigma\, \epsilon_i^{\,2}}{\partial\, \beta} = 2\{\beta\, \Sigma[(r_i - r_B) W_{Bi}]^2 + \Sigma[(r_i - r_B)(r_i - r_A) W_{Ai}\, W_{Bi}]$$

$$+ \gamma\, \Sigma[(r_i - r_B)(r_i - r_C) W_{Bi}\, W_{Ci}]\} \ . \tag{3-22}$$

$$\frac{\partial\, \Sigma\, \epsilon_i^{\,2}}{\partial\, \gamma} = 2\{\gamma\, \Sigma[(r_i - r_C) W_{Ci}]^2 + \Sigma[(r_i - r_C)(r_i - r_A) W_{Ai}\, W_{Ci}]$$

$$+ \beta\, \Sigma[(r_i - r_B)(r_i - r_C) W_{Bi}\, W_{Ci}]\} \ . \tag{3-23}$$

Minimizing $\Sigma\,\epsilon_i^{\,2}$ yields the following normal equations:

$$\hat{\beta}\, \Sigma[(r_i - r_B) W_{Bi}]^2 + \hat{\gamma}\, \Sigma[(r_i - r_B)(r_i - r_C) W_{Bi}\, W_{Ci}]$$

$$= -\Sigma[(r_i - r_B)(r_i - r_A) W_{Ai}\, W_{Bi}] \ . \tag{3-24}$$

$$\hat{\beta} \ \Sigma[(r_i - r_B)(r_i - r_C)W_{Bi} \ W_{Ci}] + \hat{\gamma} \ \Sigma[(r_i - r_C)W_{Ci}]^2$$

$$= -\Sigma[(r_i - r_C)(r_i - r_A)W_{Ai} \ W_{Ci}] \ . \qquad (3\text{-}25)$$

Thus, $\hat{\beta}$ and $\hat{\gamma}$ are found from the following matrix equation:

$$\begin{bmatrix} \Sigma \ x_{1i}{}^2 & \Sigma \ x_{1i} \ x_{2i} \\ \Sigma \ x_{1i} \ x_{2i} & \Sigma \ x_{2i}{}^2 \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} -\Sigma \ x_{1i} \ y_i \\ -\Sigma \ x_{2i} \ y_i \end{bmatrix} \qquad (3\text{-}26)$$

where:

$$y_i = (r_i - r_A)W_{Ai} \ .$$
$$x_{1i} = (r_i - r_B)W_{Bi} \ .$$
$$x_{2i} = (r_i - r_C)W_{Ci} \ .$$

In this model, the observations involving reference component A have been selected, rather arbitrarily, as the dependent variable, while reference components B and C were treated as independent or concomitant variables. As a final statistical estimation procedure, this would certainly not be an acceptable approach. It should be noted, however, that at this stage of the investigation we are merely trying to identify the closest fitting model. More specifically, we are attempting to determine if the application of the model of equation (3-20) results in a significant reduction in the error of the deterministic model (2-1). Furthermore the errors involved with this type of chemical experiment are quite small so that the difference between a regression solution assuming only A as random, and the appropriate principal

axes solution assuming all three as random, would not be expected to be so large as to change our conclusion about the model of closest fit.

The results of the analysis for estimation of β and γ are as follows:

$$\hat{\beta} = 0.997902 \ .$$

$$\hat{\gamma} = 1.001713 \ .$$

$$\text{SSTotal} = 9.058421 \times 10^{-3} \ .$$

$$\text{SSReg} = 9.056187 \times 10^{-3} \ .$$

$$\text{SSE} = 2.235 \times 10^{-6} \ .$$

$$\text{MSE} = 9.069 \times 10^{-8} \ .$$

It can be noted, as in the case of the activities model of the previous section, that these estimates of β and γ are very close to unity. If the actual values of these parameters were not significantly different from 1, the model of equation (3-20) would not result in a significant reduction of the error of prediction below that of the deterministic model. A hypothesis was tested to determine whether or not β and γ differ significantly from unity. This hypothesis was formulated in the following manner:

$$H_o: \quad \beta - 1 = 0$$
$$\gamma - 1 = 0$$

against the alternative that one of the statements in the above hypothesis is not true. On this basis, the sum of

squares due to hypothesis takes on the following form:

$$SSH = [(\hat{\beta}-1),(\hat{\gamma}-1)] \begin{bmatrix} \Sigma\ x_{1i}^2 & \Sigma\ x_{1i}\ x_{2i} \\ \Sigma\ x_{1i}\ x_{2i} & \Sigma\ x_{2i}^2 \end{bmatrix} \begin{bmatrix} (\hat{\beta}-1) \\ (\hat{\gamma}-1) \end{bmatrix} \quad .(3-27)$$

SSH was found to be 5.771211×10^{-8} . The calculation of F was performed in the following manner:

$$F = \frac{SSH/2}{SSE/24} = \frac{2.885605 \times 10^{-8}}{9.069 \times 10^{-8}} = 0.31818$$

with (2, 24) degrees of freedom. This value of F is not significant and hence it is concluded that β and γ do not differ significantly from unity.

On the basis of the preceeding argument, it is concluded that the activities model under the assumption that the activity of the mixture is not constant does not result in a significant improvement over the deterministic model of equation (2-1). Therefore it is further concluded that the activities model discussed in the previous section, i.e. assuming that the activity of the mixture is constant, is superior to the model of equation (3-20).

3.2 Regression Models

On the basis of the foregoing discussion, the statistical model assuming constant activities of the mixture was adopted as the one on which to base the final quality control program. The present section is included in order to show a comparison between a simple linear statistical approach

(based on no physical theory) and the combination of a statistical and chemical approach.

Various multiple linear regression models were studied in which NG and 2NDPA content were dependent variables while functions of density and refractive index were assigned as concomitant variables. The first regression analysis that was investigated is described by the following equation:

$$W_{1i} = \alpha + \beta_1 (1/d_i) + \beta_2 W_{3i} + \epsilon_i . \qquad (3\text{-}28)$$

This model has actually been discussed in section 3.1 for it was this equation which was derived as a method of estimating ratios of activity coefficients (See equation 3-7) assuming constancy in the mixture. Therefore the model of equation (3-28) is a representation of the result of both chemical and statistical theory, and was ultimately chosen as the "best" model, i.e. the model on which to base the control charts.

After the regression analysis of equation (3-28) was conducted and the errors determined, there was suspicion that the errors involved with mixtures 3 and 25 were spurious. Further evidence led to this fact when it was observed that these two errors were consistently large in all deterministic and statistical models under study, and the estimate was always below the actual value. Therefore an attempt was made to show that these two values are outliers.

Thompson's [4] method for isolating spurious results in a population of unknown variance makes use of the fact that a function of

$$u_i = \frac{x_i - \bar{x}}{s}$$

has Student's t-distribution, where s is an estimate of the variance as determined from the sample. The relation is:

$$t = \frac{u_i (n - 2)^{\frac{1}{2}}}{(n - 1 - u_i{}^2)^{\frac{1}{2}}} \quad \text{with n-2 degrees of freedom.}$$

From this method, the conclusion was made that mixtures 3 and 25 are definitely outliers. The numerical results, calculated from the errors associated with the regression analysis of equation (3-28) are as follows:

For mixture 25,

$$u_i = \frac{0.4340 - (-0.029)}{0.147} = 3.16 \quad .$$

$$t = \frac{3.16(26)^{\frac{1}{2}}}{[27-(3.16)^2]^{\frac{1}{2}}} = 3.84 \quad .$$

$$3.84 > t_{0.975} \quad \text{with 26 d.f.}$$

$$> 2.055 \quad .$$

For mixture 3,

$$u_i = 2.218 \quad .$$

$$t = 2.395 \quad .$$

Table V presents the results of the regression and error analysis of equation (3-28) before the exclusion of the

TABLE V.   Results of Error Study of Model of Equation
(3-28); Before Removal of Mixtures 3 and 25.

Mixture	W_{1i}(Weight Per Cent)	$\hat{W}_{1i}$(Weight Per Cent)	$W_{1i}-\hat{W}_{1i}$
1	79.98	79.80	0.18
2	80.06	80.18	-0.12
3	79.97	79.67	0.30
4	80.10	80.11	-0.01
5	77.61	77.69	-0.08
6	77.60	77.70	-0.10
7	77.63	77.63	0
8	77.34	77.36	-0.02
9	75.02	74.97	0.05
10	75.03	74.99	0.04
11	74.99	74.98	0.01
12	74.98	75.05	-0.07
13	72.50	72.72	-0.22
14	72.50	72.56	-0.06
15	72.50	72.55	-0.05
16	72.49	72.55	-0.06
17	69.98	70.08	-0.10
18	69.98	70.00	-0.02
19	69.99	69.84	0.15
20	69.99	69.80	0.19
21	67.51	67.58	-0.07
22	67.50	67.65	-0.15
23	67.48	67.63	-0.15
24	67.49	67.50	-0.01
25	65.01	64.58	0.43
26	64.98	64.94	0.04
27	64.98	65.08	-0.10
28	64.99	65.01	-0.02

SSE = 0.54022   MSE = 0.021609

spurious mixtures. This can be compared to Table IV, the same results after they were removed. It should be noted that the outlier test above applies to only one set of data. The observations for mixtures 3 and 25 were conspicuously in error in all sets of data.

The second regression model considered was the following joint set of equations:

$$W_{1i} = \alpha + \beta_1 d_i + \beta_2 n_i + \beta_3 W_{3i} + \varepsilon_i \qquad (3\text{-}29)$$

$$W_{3ia} = \alpha + \beta_1 d_i + \beta_2 n_i + \beta_3 W_{3i} + \varepsilon_i \qquad (3\text{-}30)$$

where:

W_{3ia} = actual weight proportion 2NDPA in mixture i .

W_{3i} = spectrophotometer analysis of 2NDPA content. This model was chosen because it is the one a statistician would select if he disregarded the theory involved. Each involves the estimation of four parameters. The following results were gathered from this study:

(1) In the prediction of 2NDPA, only the regression coefficient associated with the spectrophotometer analysis is statistically significant. However a study of the data will show that although refractive index was not found to be statistically significant, any systematic change in refractive index is brought about by a corresponding change in 2NDPA content. Refractive index is affected very little by changes in either TA or NG alone or together. Therefore it

is concluded that the lack of significance of refractive index in model (3-30) is a result of the extreme sensitivity of the spectrophotometer analysis to changes in 2NDPA, and further attests to the accuracy of the analysis as a predictor of 2NDPA content.

(2) The prediction of NG content showed significant regression coefficients for density and 2NDPA analysis.

These results verify the conclusion that refractive index provides no significant reduction of error in the prediction of the content of the mixture in question. The results of the study of this model, which are shown in Tables VI and VII, indicate that the error involved with the linear regression of equation (3-29), in which four parameters were estimated, is larger than that of the deterministic model, in which no estimation was made. This leads to the conclusion that one should not allow simple linear regression to overshadow the consideration of the theory involved.

To further establish that refractive index in no way aids in the prediction of NG, the following model was investigated:

$$W_{1i} = \alpha + \beta_1 (\frac{1}{d_i}) + \beta_2 (\frac{1}{v_i}) + \beta_3 W_{3i} + \varepsilon_i \qquad (3-31)$$

where:

$$v_i = \frac{n_i^2 - 1}{n_i^2 + 2} \quad .$$

TABLE VI. Results of Regression Analysis of Model (3-29),

$$W_{1i} = \alpha + \beta_1 d_i + \beta_2 n_i + \beta_3 W_{3i} + \varepsilon_i \,.$$

Mixture	W_{1i}(Weight Per Cent)	$\hat{W}_{1i}$(Weight Per Cent)	$W_{1i}-\hat{W}_{1i}$
1	79.98	79.89	0.09
2	80.06	80.29	-0.23
4	80.10	80.25	-0.15
5	77.61	77.62	-0.01
6	77.60	77.66	-0.06
7	77.63	77.62	0.01
8	77.34	77.31	0.03
9	75.02	74.91	0.11
10	75.03	74.88	0.15
11	74.99	74.92	0.07
12	74.98	75.02	-0.04
13	72.50	72.58	-0.08
14	72.50	72.46	0.04
15	72.50	72.48	0.02
16	72.49	72.46	0.03
17	69.98	69.95	0.03
18	69.98	69.91	0.07
19	69.99	68.77	0.22
20	69.99	69.71	0.28
21	67.51	67.53	-0.02
22	67.50	67.61	-0.11
23	67.48	67.64	-0.16
24	67.49	67.47	0.02
26	64.98	65.00	-0.02
27	64.98	65.15	-0.17
28	64.99	65.09	-0.10

SSE = 0.3479

with 22 d.f.

MSE = 0.0158

$\hat{\beta}_1 = 202.34$	Var $\hat{\beta}_1 = 9.549$	$t_{\beta 1} = 65.696$
$\hat{\beta}_2 = -23.779$	Var $\hat{\beta}_2 = 584.1$	$t_{\beta 2} = -0.983$
$\hat{\beta}_3 = -0.14974$	Var $\hat{\beta}_3 = 4.732 \times 10^{-3}$	$t_{\beta 3} = -2.176$
$\hat{\alpha} = -184.333$		

Note that β_1 and β_3 are significant while β_2 is not.

TABLE VII. Results of Regression Analysis of Model (3-30),

$$W_{3ia} = \alpha + \beta_1 d_i + \beta_2 n_i + \beta_3 W_{3i} + \varepsilon_i .$$

Mixture	W_{3ia} (Weight Per Cent)	$\hat{W}_{3ia}$ (Weight Per Cent)	$W_{3ia} - \hat{W}_{3ia}$
1	0	0.010	-0.010
2	1.00	0.999	0.001
4	3.00	3.006	-0.006
5	0	0.006	-0.006
6	1.00	0.993	0.007
7	2.00	1.991	0.009
8	2.99	3.010	-0.020
9	0	0.004	-0.004
10	1.00	0.996	0.004
11	2.00	1.994	0.006
12	3.00	3.012	-0.012
13	0	0	0
14	1.00	1.008	-0.008
15	2.00	1.996	0.004
16	3.00	2.986	0.014
17	0	0.003	-0.003
18	1.00	1.001	-0.001
19	2.00	2.030	-0.030
20	3.00	2.979	0.021
21	0	0.005	-0.005
22	1.00	1.014	-0.014
23	2.00	2.002	-0.002
24	3.00	3.011	-0.011
26	1.00	1.007	-0.007
27	2.00	1.985	0.015
28	3.00	2.974	0.026

$SSE = 3.86 \times 10^{-3}$ $MSE = 1.75 \times 10^{-4}$

$\hat{\beta}_1 = -0.2615$ $Var \, \hat{\beta}_1 = 0.10624$ $t_{\beta 1} = -0.80215$

$\hat{\beta}_2 = 0.2461$ $Var \, \hat{\beta}_2 = 6.49$ $t_{\beta 2} = 0.09535$

$\hat{\beta}_3 = 0.9982$ $Var \, \hat{\beta}_3 = 7.94 \times 10^{-5}$ $t_{\beta 3} = 110.11$

$\hat{\alpha} = 0.017725$

Note that β_1 and β_2 are not significant.

The results were as expected, i.e., β_2 was non-significant, indicating that the error of prediction is not significantly reduced as compared with the model which does not involve refractive index. The values for β_1 and β_3 were found to differ significantly from zero, as before. The results of this analysis are shown on Table VIII. Note the closeness in results of this model and the one in which $1/v_i$ was removed (See Table V).

From the discussions in Chapter II and III, it was concluded that the activities model of equation (3-2), which ultimately led to the regression equation (3-7), represents the most desirable model of those which were investigated. Therefore the control charts, and confidence bounds to be described in the next chapter, are based on this combination of a statistical and chemical model.

TABLE VIII. Results of Regression Analysis of Model (3-31),

$$W_{1i} = \alpha + \beta_1(\frac{1}{d_i}) + \beta_2(\frac{1}{v_i}) + \beta_3 W_{3i} + \epsilon_i \; .$$

Mixture	W_{1i}(Weight Per Cent)	$\hat{W}_{1i}$(Weight Per Cent)	$W_{1i}-\hat{W}_{1i}$
1	79.98	79.79	0.19
2	80.06	80.17	-0.11
4	80.10	80.12	-0.02
5	77.61	77.63	-0.02
6	77.60	77.66	-0.06
7	77.63	77.62	0.01
8	77.34	77.33	0.01
9	75.02	74.95	0.07
10	75.03	74.94	0.09
11	74.99	74.98	0.01
12	74.98	75.07	-0.09
13	72.50	72.65	-0.15
14	72.50	72.53	-0.03
15	72.50	72.55	-0.05
16	72.49	72.55	-0.06
17	69.98	70.00	-0.02
18	69.98	69.96	0.02
19	69.99	69.82	0.17
20	69.99	69.78	0.21
21	67.51	67.50	0.01
22	67.50	67.60	-0.10
23	67.48	67.62	-0.14
24	67.49	67.47	0.02
26	64.98	64.87	0.11
27	64.98	65.04	-0.06
28	64.99	64.99	0

SSE = 0.22197   MSE = 1.009×10^{-2}

$\hat{\alpha}$ = 354.8048

$\hat{\beta}_1$ = -419.2216   Var $\hat{\beta}_1$ = 27.0412 $t_{\beta 1}$ = -80.619

$\hat{\beta}_2$ = 2.3936   Var $\hat{\beta}_2$ = 8.476 $t_{\beta 2}$ = 0.8223

$\hat{\beta}_3$ = -0.170569   Var $\hat{\beta}_3$ = 4.699×10^{-3} $t_{\beta 3}$ = -2.4905

CHAPTER IV

CONTROL CHARTS*

4.1 Explanation of Charts

After investigating various possible models for use in estimating the content of the casting solvent system, the following was decided on as best in terms of the error in the estimation of weight per cent nitroglycerine:

$$W_{1i} = \alpha + \beta_1 \left(\frac{1}{d_i}\right) + \beta_2 W_{3i} + \varepsilon_i$$

where:

d_i = specific gravity of the mixture, (measured at 21.1°C).

W_{1i} = weight per cent nitroglycerine.

W_{3i} = result of 2NDPA analysis by spectrophotometer.

This model was derived from an equation involving activity coefficients. Thus it is the result of the application of both chemical and statistical theory.

The study was conducted assuming a constant moisture content. The following three steps outline a general procedure for the estimation of the content of the entire system:

---

*Frequent repetition of statements in this chapter with those of previous chapters is due to the fact that this part only was given to the Technical Staff at Hercules Powder Company as an operational manual.

(1) 2NDPA content determination by spectrophotometer method.

(2) Determination of NG concentration from the model above.

(3) Triacetin content as the complement to 100 per cent of the sum of (1) and (2).

It should be noted that if there is reason to believe that the moisture in the system is not negligible (through some type of moisture analysis), then the triacetin content would be obtained by the complement to 100 per cent of the sum of (1), (2) and weight per cent moisture in the sample. More specific instructions will follow.

Figure 1 contains a chart with 99 per cent confidence limits on weight per cent 2NDPA as a function of W_3. A reading on the upper line determines the upper limit and the lower line contains the lower limit. The upper and lower limits define an interval within which it can be said with 99 per cent confidence that the true 2NDPA content lies. The data indicated that the spectrophotometer method in obtaining W_3 is very accurate. It can be noted that the confidence interval is longer than the interval defining the probable error, which gives only 50 per cent confidences, and is frequently found in scientific literature.

Figure 2 contains a chart relating a variable Y with specific gravity. The combination of this chart and figure 3

is a systematic representation of the regression model men-
tioned previously. The equation is as follows:

$$Y = 360.7881 - 415.1894 \left(\frac{1}{d_i}\right)$$

where:

$$Y = W_1 + 0.225 \; W_3 \; .$$

Figure 3 contains the correction factor $0.225 \; W_3$ as a func-
tion of W_3. Subtracting this from Y results in the estimate
of weight per cent NG.

Figure 4 displays 99 per cent confidence intervals for
weight per cent NG as a function of spectrophotometer analysis
and specific gravity. Linear interpolation between curves of
constant per cent 2NDPA is sufficiently accurate. The result
of this chart indicates the interval within which it can be
stated with 99 per cent confidence that the true weight per
cent NG lies.

4.2   Instructions in Use of Charts

The chart on figures 1, 2, 3, and 4 are helpful in the
estimation of the true content of a sample of casting sol-
vent. The following are specific step-by-step instructions
and examples for their use:

(a) Corresponding to the point W_3, on the X-axis of
figure 1, there is a value on the upper and a value on the
lower curve. This interval represents the 99 per cent con-
fidence bounds for the true 2NDPA content.

(b) In figure 2 select, on the basis of the observed specific gravity, that scale along the X-axis which includes the value. Read up to the line having the same Roman number as the scale. Read the corresponding Y value at the left.

(c) Read 2NDPA, as analyzed by the spectrophotometer, along the X-axis of figure 3. The corresponding value on the Y-axis is the correction to be applied to Y from step b.

(d) The difference, Y (from step b) - correction (from step c) results in the estimated content of nitroglycerine in weight per cent.

(e) Obtain the estimate of TA content from the difference,

$$100 - NG - 2NDPA = TA \text{ (in weight per cent)},$$

where

NG = weight per cent NG from step d.

2NDPA = weight per cent 2NDPA as analyzed by spectrophotometer.

TA = weight per cent TA.

The latter value may need reduction by the amount of moisture, depending on whether or not the moisture analysis indicated it to be negligible.

(f) To obtain a confidence interval for the true value of nitroglycerine content, enter the observed specific gravity at the bottom of figure 4. Find the appropriate curve of constant 2NDPA analysis. At the left, a ± value can be

observed which, if added to the estimated NG content, describes the 99 per cent confidence interval.

4.3 Numerical Examples

Example I

Suppose that the spectrophotometer analysis indicates the estimate of 2NDPA content (W_3) to be 2.00 per cent. The specific gravity and moisture content at 21.1°C were found to be 1.4225 and 0.03 per cent respectively. The results are as follows:

(a) From figure 1, the 99 per cent confidence interval on the 2NDPA analysis is found to be 1.96 to 2.04.

(b) From figure 2 with d = 1.4225, the Y value, as read from line I is found to be 68.91 per cent.

(c) From figure 3 with W_3 = 2.00, the correction for Y is 0.45.

(d) Subtracting 0.45 from 68.91, NG is found to be 68.46 per cent.

(e) The estimate of the TA concentration is 100 - (2.00 + 68.46 + 0.03) = 29.51 per cent.

(f) The 99 per cent confidence interval on NG from figure 4 is found to be [68.46 $\pm$ 0.29]. Thus the lower and upper 99 per cent limits respectively are 68.17 and 68.75.

Example II

Suppose W_3 = 1.50, d = 1.4650 and moisture = 0.025.

The results are as follows:

(a) The 99 per cent confidence interval on weight per cent 2NDPA is 1.50 ± 0.039.

(b) Y = 77.34 (from line III of figure 2).

(c) Correction = 0.339.

(d) NG = 77.34 - 0.339 = 77.00 per cent.

(e) The estimate of TA = 100-1.50-77.00-0.025 = 21.475 per cent.

In both of these examples, moisture was determined and taken into consideration in the calculations. However it must be noted that the percent moisture here is entirely within the error of estimation.

Figure 1 - Upper and Lower 99 Per Cent Confidence Limits
on Weight Per Cent 2NDPA
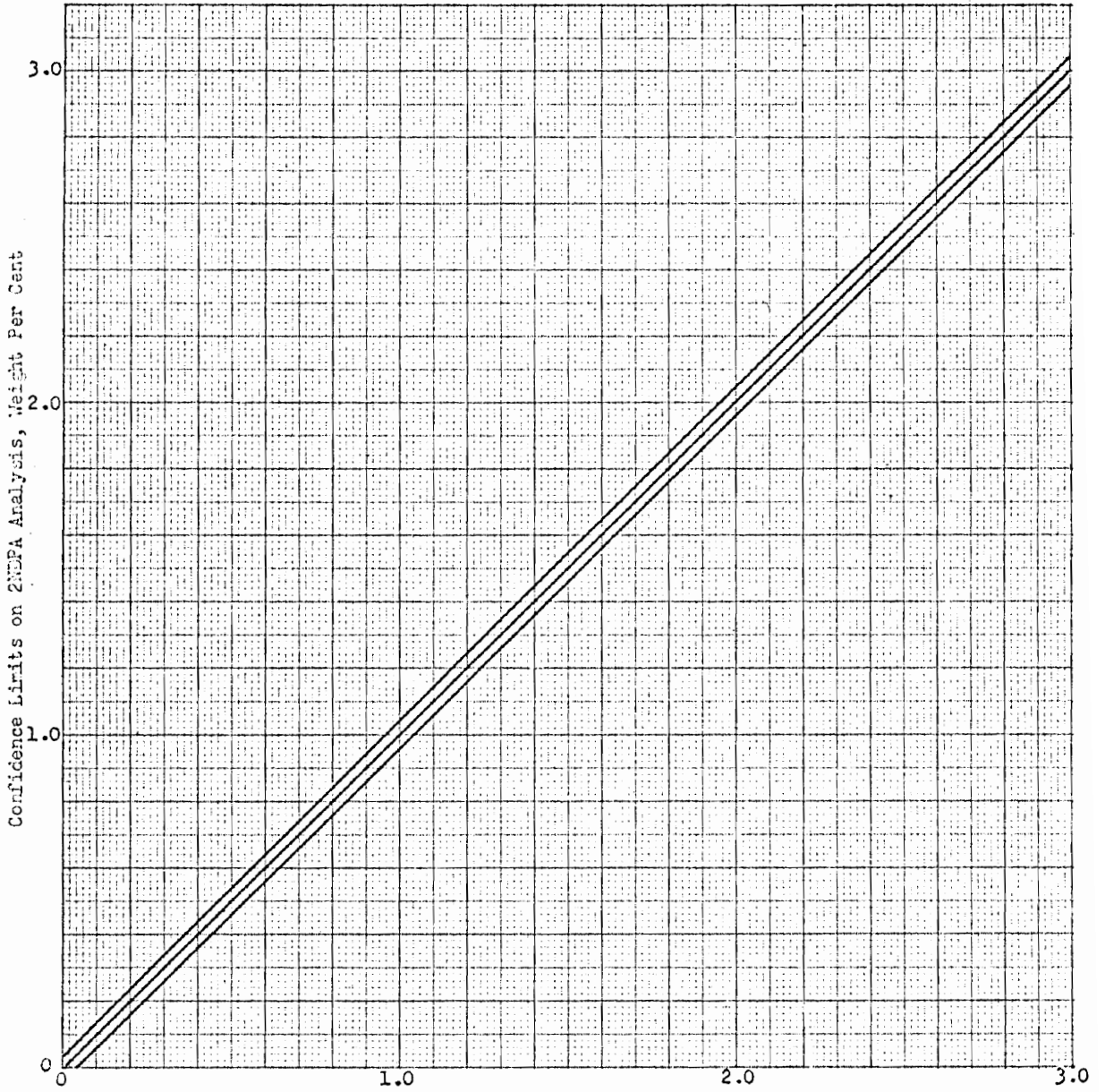
2NDPA Analysis by Spectrophotometer, Weight Per Cent

Figure 2 - Relationship Between Y and Specific Gravity
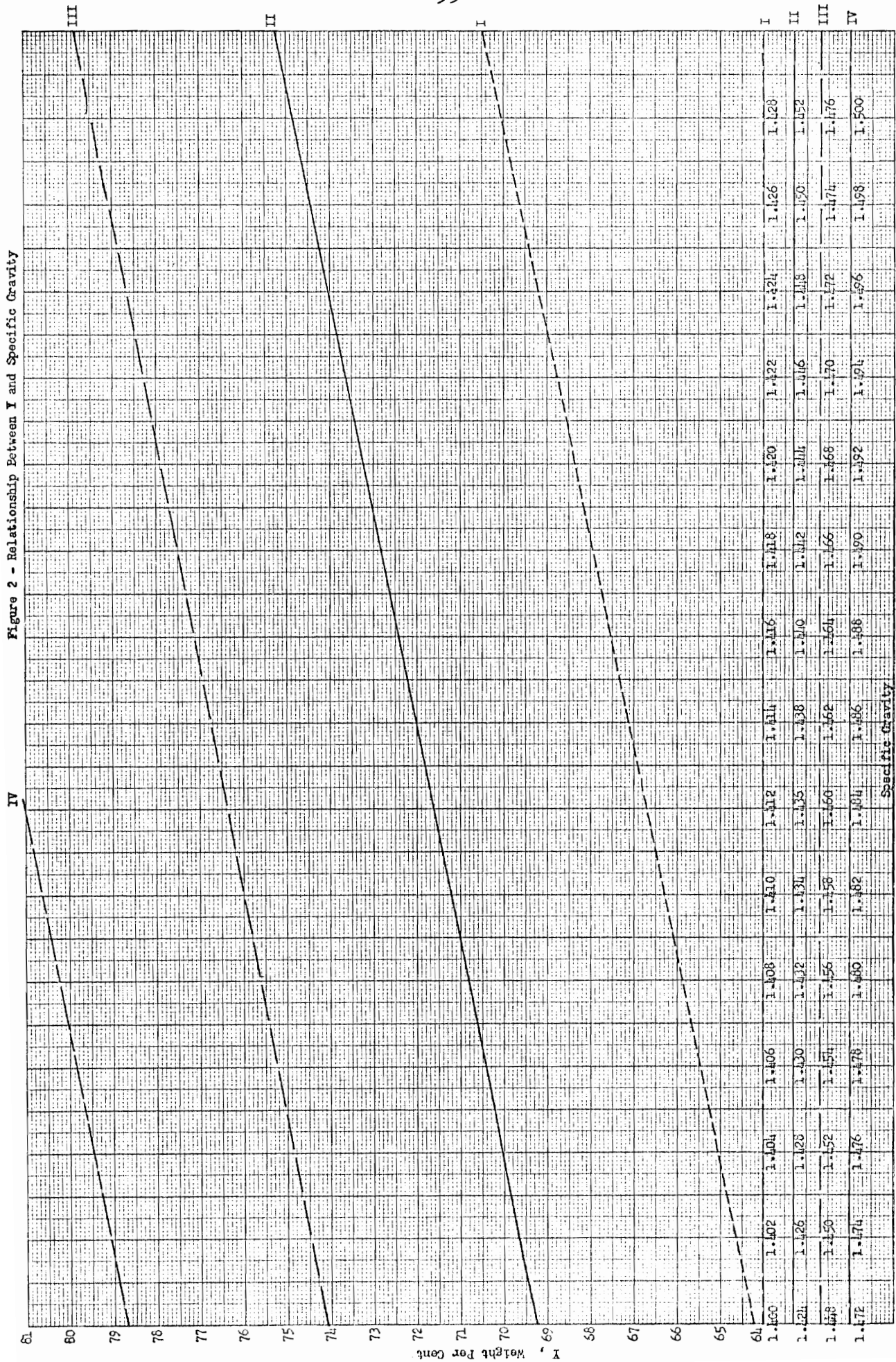
Figure 3 - Relationship Between Correction Factor on Y and Spectrophotometer Analysis
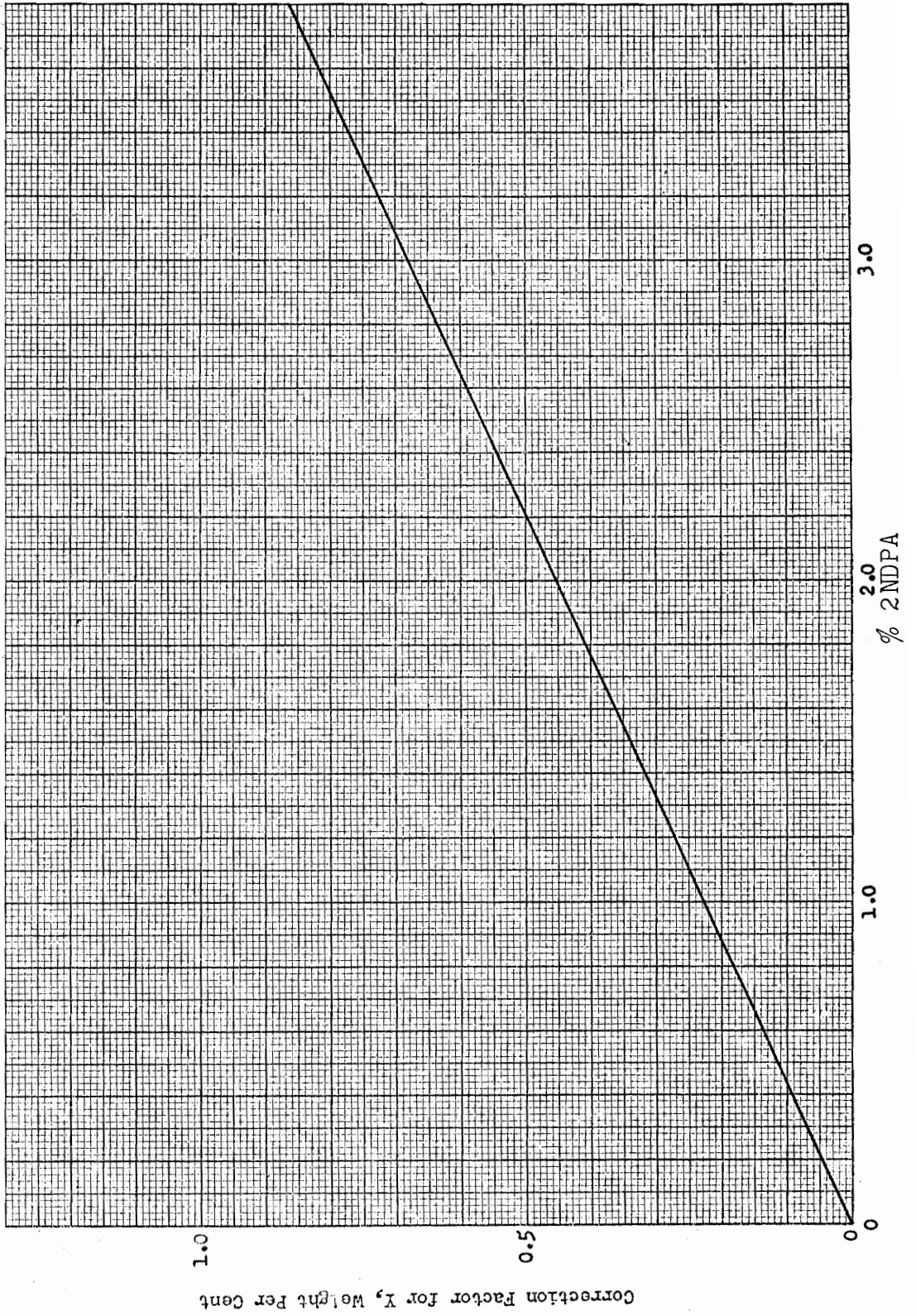
Figure 4 - 99% Confidence Intervals on Weight Per Cent NG as a Function of Specific Gravity and Spectrophotometer Analysis.
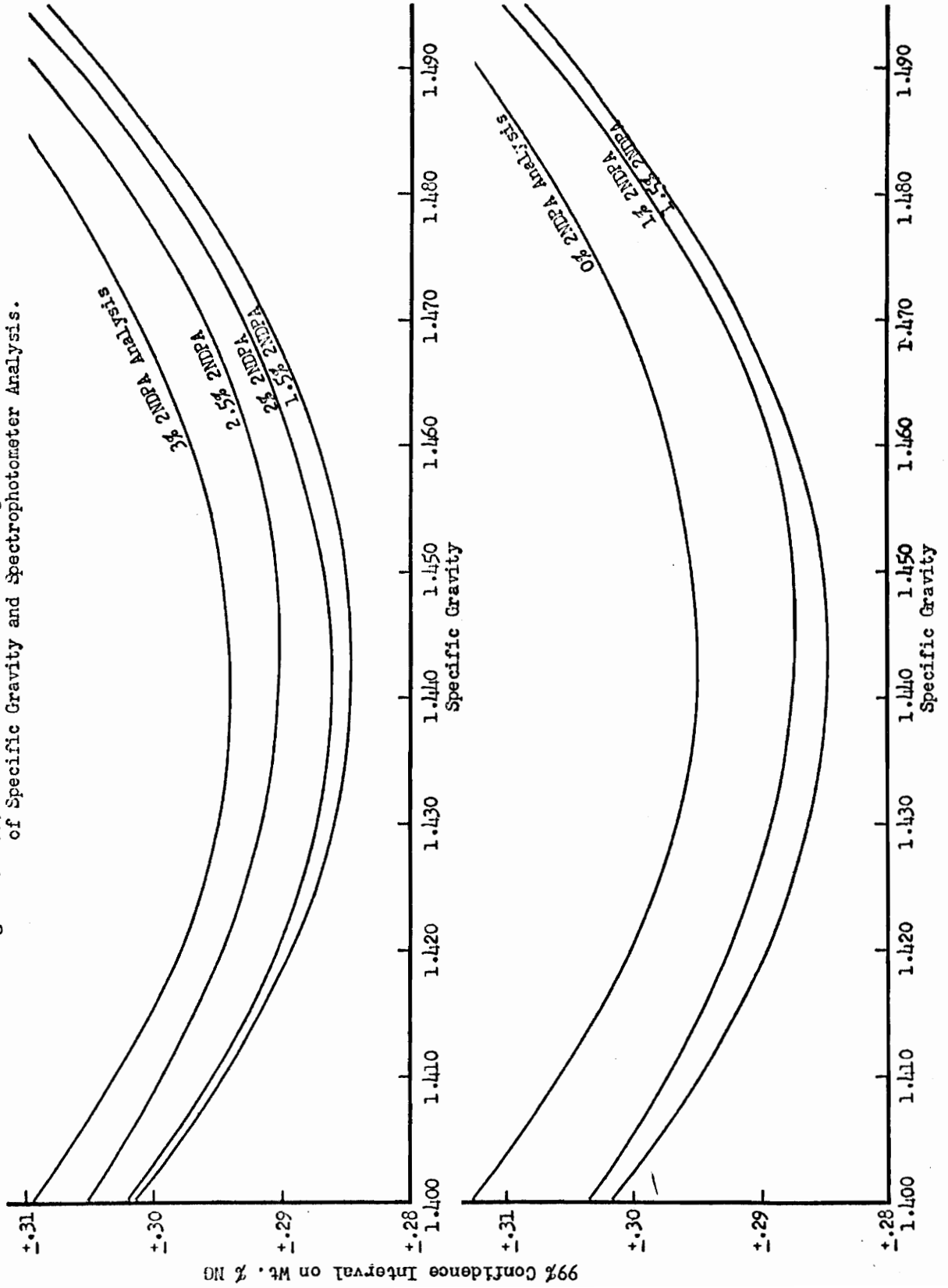
REFERENCES

[1] Hamilton, L. F. and Simpson, S.G (1952). Quantitative Chemical Analysis, page 434, Macmillan Company.

[2] Moore, W. J. (1955). Physical Chemistry, pages 94, 141, Prentice-Hall, Inc.

[3] Slater, J. C. and Frank, N. H. (1947). Electromagnetism, page 110, McGraw-Hill, Inc.

[4] Thompson, W. R. (1935). On a criterion for the rejection of observations and the distribution of the ratio of deviation to sample standard deviation. Ann. Math. Stat., 6, 214-219.

ACKNOWLEDGMENTS

VITA

The author was born on October 13, 1937, in Charleston, West Virginia. He obtained his grammar and high school education in Charleston.

Upon graduation from high school in June, 1955, the author enrolled in Virginia Polytechnic Institute. He received his Bachelor of Science Degree in Chemical Engineering in June, 1959. His undergraduate honors included membership into Phi Kappa Phi Scholastic Honorary and Phi Lambda Upsilon Chemical Honorary Society.

The author enrolled into the Department of Statistics at V.P.I. in September, 1959, to begin work toward a Master of Science Degree. While in graduate school, he was elected into Pi Mu Sigma Mathematics Honorary and Sigma Xi Research Society.

Raymond H. Myers

ABSTRACT

This thesis is a study of various physical and statistical models which may be useful for the prediction of the composition of a ternary liquid mixture. The particular mixture considered in this study was the solvent system consisting of nitroglycerine (NG), 2-nitrodiphenylamine (2NDPA), and triacetin (TA). Several models were investigated for their adequacy and closeness of fit. An attempt has been made to relate the actual composition to a few easily measurable quantities, namely, refractive index, density, and the separate analysis of 2NDPA.

Deterministic models relating the concentration of each component in the mixture with the physical determinations mentioned above have been considered first. These models are based on known theory of physical chemistry. The deterministic model which was chosen as "best" in terms of smallness of error of prediction, estimates the composition from the determination of density and the spectrophotometer analysis of 2NDPA. Since the latter analysis is a quick and accurate determination of the 2NDPA content and since the content of the third component could be determined by complementing to 100 per cent, the models have been formulated in terms of the concentration of only one component, namely NG.

The statistical models under investigation are divided into activity models and regression models. The activity

models is a combination of chemical and statistical theory while the simple regression model represents an approach that a statistician might take if he disregarded the physical or chemical theory involved. Two activity models have been discussed, the first assuming the activity of the mixture constant and the second assuming the activity of the mixture to be a weighted sum of the activities of the three components.

Tests of hypotheses are made to determine whether the activity models result in a significant reduction in error over that of the "best" deterministic model. The investigation that the model formulated assuming constant activity of the mixture results in the smallest error of estimation among all models under study. Thus it has been used as a basis for the preparation of control charts.

The linear regression models, constructed with various functions of density, refractive index, and spectrophotometer analysis as independent variables, produced errors of estimation above those for the deterministic model.

Chapter IV represents the summary of the thesis, and the translation of findings into actual control charts. On the basis of this chapter a technician can easily determine the estimate of the composition of the mixture and the attached 99 per cent confidence bounds. Thus, this Chapter contains these charts combined with instructions in their use and numerical examples.