MATHEMATICAL MODELING OF CHEMICAL REACTIONS

J. R. Kittrell

Chevron Research Company
Richmond, California

I. Introduction 98
A. Mechanisms and Models 98
B. Reaction-Rate Models 99
C. Integral and Differential Models 101
II. Linearly Reducible Models 102
A. Power-Function Models 102
B. Hyperbolic Models 105
III. Parameter Estimation 110
A. Least Squares 111
B. Reparameterization 121
C. Confidence Intervals and Regions 124
D. Multiple-Response Parameter Estimation 129
IV. Tests of Model Adequacy 131
A. Analysis of Variance 131
B. Residual Analysis 137
V. Use of Diagnostic Parameters 142
A. Model Discrimination with Diagnostic Parameters 142
B. Adaptive Model-Building with Diagnostic Parameters 147
VI. Empirical Modeling Techniques 154
A. Response-Surface Methodology 155
B. Transformations of Variables 159
C. Empirical Model Tuning 164
VII. Experimental Designs for Modeling 168
A. Model-Discrimination Designs 171
B. Parameter-Estimation Designs 173
Nomenclature 178
References 181

I. Introduction

A. Mechanisms and Models

For any physical system there exists a precise mathematical and physical representation of all of the phenomena that make up the system. A chemical reaction is no exception, since it does possess some *true* mechanism that is descriptive of every microscopic detail of the reaction. In practice, of course, a complete description of this mechanism cannot be obtained, and approximations must be made. For example, for studies of a fundamental nature on heterogeneously catalyzed reaction systems, this may entail reference to some average property, such as an average pore size, an average size of a particle of metal dispersed on a support, an average adsorbed state, or an average mode of reaction. For a broader problem such as the modeling of the operation of an entire plant, we must make more extensive approximations. For industrial reactions generally we are fortunate if we can characterize the reactor feed in terms of a limited number of major components, if we can precisely define the reaction kinetics of these major components in the industrial scale, and if we can evaluate the residence time distribution of the reactor. Consequently, we will often be concerned with models of the system that are at best quasimechanistic, simply because our reacting system is chosen on the basis of economics rather than simplicity or convenience. This review presents statistical methods for developing mathematical descriptions (either theoretical or empirical) of reacting systems.

In contrast to the above allusion to a mechanism, the term model has been used to describe a wide range of chemical engineering subjects. Chemical engineering modeling activities in general can be thought of as belonging to one of three members of a nested set. The outermost member is plant modeling. This entails the description of, for example, an entire refinery complex, taking into full account topics ranging from crude resources to geographical product distributions. Economic models of such a complex obviously require a knowledge of the behavior of the several processes making up the plant. Hence, modeling activity within the plant-wide modeling is required, that is, process modeling. Process models are capable of predicting the steady-state level of operation at the settings of the process variables, and perhaps the dynamic response of the process to any disturbance. Several process models must be included in the plant model. The development of the process model, in turn, cannot be carried out without recourse to phenomenological or mechanistic modeling. The mechanistic model must be capable of describing adequately the basic physical and chemical steps that take place within the

processes, under the conditions imposed by the settings of the process variables.

Considerable unification of the concepts of plant modeling and process modeling has taken place in recent years, for example, as described in References (F4) and (R3) for process and plant concepts, respectively. Attention has been devoted only intermittently, however, to the development of methodology for mechanistic modeling. For reaction-rate modeling, in particular, there was considerable activity in the 1930's and 1940's and again in the late 1950's and 1960's. The latter activity apparently resulted in large part from the introduction of computer techniques into chemical engineering technology. This review will be concerned specifically with methodology useful in the elucidation of such reaction rate models, treated artificially as two distinct stages. In the specification stage, the problem is to determine the appropriate functional form of a rate model that adequately describes the reaction. Although parameter estimates are often necessary to test the adequacy of any particular functional form, these estimates often need not be very precise. After a model has been specified, it is often advantageous to give particular attention to the estimation of the parameters within the adequate functional form. Once the functionality of the rate equation is known, the estimation of the parameter values is relatively straightforward.

Although examples of the methodology will utilize entirely reaction rates or reactant concentrations, the procedures are equally valid for other model responses. They have been used, for example, with responses associated with catalyst deactivation and diffusional limitations as well as with copolymer reactivity ratios and average polymer molecular weights.

B. Reaction-Rate Models

In the model-specification stage, we concentrate upon two types of reaction-rate models, the power-function model

$$r = kC_A{}^a C_B{}^b \tag{1}$$

and the hyperbolic model exemplified by

$$r = \frac{kC_A C_B}{1 + K_A C_A + K_B C_B} \tag{2}$$

Here, r is a reaction rate, C_A and C_B are measures of concentration, and the remaining terms represent parameters to be estimated.

Equation (1), of course, is simply a representation of the law of mass action, stating that the reaction rate is proportional to a concentration driving force. Such an equation has been applied to many reacting systems, particularly those of considerable complexity.

Equation (2) is generally derived based upon some form of steady state hypothesis. Such models have been applied to gaseous reactions (Lindemann theory), to enzymatic reactions (Michaelis–Menten mechanism), and to gaseous reactions on solid surfaces (Langmuir–Hinshelwood mechanism). Considerable discussion of the derivation of these equations has been presented elsewhere (L2). In general, we refer to this model type as a hyperbolic model or, in the case of solid catalyzed gaseous reactions, as a Hougen–Watson model (to remove restrictions implicit in a Langmuir–Hinshelwood mechanism).

The use of any model more complex than Eq. (1) has frequently been criticized as an attempt to read too much into a set of kinetic data. The implicit results of such a misapplication would be a proliferation of meaningless constants; this would thus lead to unnecessary effort in data fitting and interpretation, or to misleading interpolation or extrapolation.

The latter danger is, of course, potentially present any time any data interpretation is attempted, particularly if nature is assumed always to follow Eq. (1). The only course of action is to attempt to include as much theory in the model as possible, and to confirm any substantial extrapolation by experiment. It is erroneous, however, to presume that kinetic data will always be so imprecise as to be misleading. The use of computers and statistical analyses for any linear or nonlinear reaction rate model allows rather definite statements about the amount of information obtained from a set of data. Hence, although imprecision in analyses may exist, it need not go unrecognized and perhaps become misleading.

However, the amount of error in the data is not generally the limiting factor in data interpretation. Rather, the locations at which the data are taken most severely hinder progress toward a mechanistic model. Reference to Fig. 1 indicates that the decision between the dual- and single-site models would be quite difficult, even with very little error of measurement, if data are taken only in the 2- to 10-atm range. However, quite substantial error can be tolerated if the data lie above 15 atm total pressure (assuming data can be taken here). Techniques are presented that will seek out such critical experiments to be run (Section VII).

The concept of the reaction-rate model should be considered to be more flexible than any mechanistically oriented view will allow. In particular, for any reacting system an entire spectrum of models is possible, each of which fits certain overlapping ranges of the experimental variables. This spectrum includes the purely empirical models, models accurately describing every detail of the reaction mechanism, and many models between these extremes. In most applications, we should proceed as far toward the theoretical extreme as is permitted by optimum use of our resources of time and money. For certain industrial applications, for example, the closer the model approaches

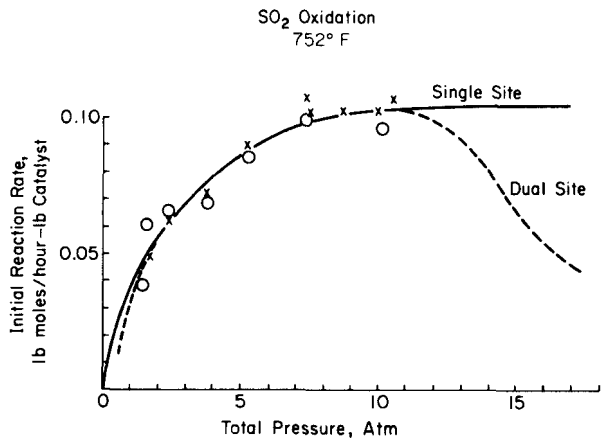

FIG. 1. Initial rate dependence on total pressure for SO_2 oxidation (M3).

the true mechanism, the better the final plant design and operation may become. For example, for a high-throughput process of which many exist and many more are expected to be built, a precise mechanistic model can often be justified. In this area, in fact, recent developments in computing, process dynamics, and large-system design are increasingly demanding more sophisticated models of the mechanistic and transport phenomena of the plant in general and reaction kinetics in particular. On the other hand, the construction of a precise mathematical model can require considerable experimentation and analysis. It would often be more profitable to get a new process on stream with a grossly nonoptimum design than to delay construction for a precise definition of a kinetic model. For most applications, an economic balance must be achieved to allow flexibility for modeling as well as rapid progress toward the process design.

C. Integral and Differential Models

Equations (1) and (2) represent reaction rates and, as such, can represent directly only data from a differential reactor. In many cases, however, data are obtained from an integral reactor. Are the data to be differentiated and compared directly to Eqs. (1) or (2), or are the equations to be integrated with the conservation equations and compared to the integral data?

If the modeling process has entered the estimation stage, then Eqs. (1) or (2) should clearly be integrated and fitted to the integral data. This may be accomplished quite easily, through analytical or numerical integration of the conservation equations combined with one of the estimation procedures of Section III. General programs providing such combinations have been described (H6).

In the specification stage, however, gross elimination of rival models is difficult to accomplish using such generalized techniques with a set of existing data (cf. the example of Section III,A,3). Instead, careful comparisons of the features of the data and the theoretical model surface are required; these are most easily compared in the simplest possible model form, i.e., the differential form (see, for example, Section V). Furthermore, the experimental design procedures of Section VII can consume substantial computer time, a problem made more severe by iterative numerical integration. With the exception of purely empirical modeling (as in Section VI, which is always in the estimation stage), then, it is preferred to carry out the initial model specification with the simplest form of the model with which the data can be made consistent. For example, factors such as severe variation in catalyst-bed temperature can require the integrated equations even in the specification stage.

II. Linearly Reducible Models

Most of the modeling procedures commonly used require that the model first be reduced to a form which is linear in the unknown parameters. This procedure represents very good tactics; the technique will be exploited frequently in this review, particularly in Section V. If the scope of the models to be used or the range of experimental variables to be explored is not limited when applying this philosophy, the procedure also represents good modeling strategy.

For these reasons, a proper balance must be achieved between the linearization tactics and the overall modeling strategy. Both linear and nonlinear methods will be illustrated in the review, along with the problems encountered when relying too heavily on either single approach. First, however, some of the more common linear procedures will be discussed.

A. Power-Function Models

For a single reacting component, Eq. (1) reduces to a form that can easily be compared to rate data to determine (a) if the model is adequate and, if so, (b) the best estimates of the rate constant and the reaction order. For this case, Eq. (1) becomes

$$r = -dC_A/dt = kC_A^a \tag{3}$$

Here, t is a measure of reaction time. To analyze data from certain reactor types using this model, one can use one of the following methods.

1. *The Method of Integration*

Here, the rate equation, such as Eq. (3), is integrated to yield

$$\begin{aligned} C_A^{1-a} - C_{A0}^{1-a} &= (a-1)kt \qquad a \neq 1 \\ \ln (C_A/C_{A0}) &= -kt \qquad a = 1 \end{aligned} \tag{4}$$

Now, appropriate plots of the data are made, which, if linear, would indicate that the assumed model of Eq. (3) is adequate. For example, if $\ln(C_A/C_{A0})$ were linear with t, a first-order model would be adequate. Alternatively, one could assume a model (including the value of the parameter a), calculate the rate constant k at each data point, and tabulate the constants. If these "constants" remain constant, or if there is a reasonable trend of the constants with any independent variable, then the data do not reject the assumed model. For example, the value of $\ln k$ would be expected to be independent of the value of the reaction time and to change linearly with the reciprocal of the absolute temperature.

2. *The Method of Differentiation*

In this method, the reaction rate is read directly, say, as the slope of the concentration–time data. Then, the logarithm of the rate is plotted versus the logarithm of the concentration; if the data lie along a straight line, the slope is equal to the reaction order.

3. *Fractional-Life Methods*

Models may also be tested by utilizing the time required for a given fraction of a reactant to disappear, since this varies with the initial concentration in a fashion characteristic of the reaction order. For example, if the half-life of a reaction is defined as the time required for one-half of the initial amount of reactant to be consumed, then Eq. (4) may be written

$$\begin{aligned} \ln t_{1/2} &= \ln \left(\frac{2^{a-1} - 1}{k(a-1)} \right) - (a-1) \ln C_{A0} \qquad a \neq 1 \\ t_{1/2} &= 0.693/k \qquad a = 1 \end{aligned} \tag{5}$$

Here $t_{1/2}$ is the half-life and C_{A0} the initial concentration of reactant. Now, half-life data (versus C_{A0}) may be analyzed by the same techniques used in the method of integration.

Several other methods are available for analyzing reaction-rate data, such as utilizing linear least squares in the above methods, or such as the method of dimensionless curves. The procedures and their advantages and disadvantages

have been adequately discussed elsewhere (L2); numerous examples of the techniques have also been set forth (B3, F5, H7, L2, W2) and are not discussed further here.

4. *Multicomponent Cases*

For the more usual multicomponent case, as described by Eq. (1), variations of the above methods are used to allow a data analysis. One method is to take raction-rate data at such large concentrations of one reactant that this concentration is effectively a constant during the entire course of the reaction:

$$r = kC_A{}^a C_{B0}^b = k' C_A{}^a \tag{6}$$

The order of the other reactant can be determined by any of the previously discussed methods. This technique, called the isolation method, will allow a determination of the component reaction orders, but it should be kept in mind that a very limited region of the experimental space has been covered in determining these orders. Thus, because the model has not been tested for conditions in which both concentrations are varying, the model should be used with caution here.

A second method of analyzing data to be described by Eq. (1) is to use initial reaction rates. Here, one can vary the initial concentrations of the individual reactants (holding all other reactant concentrations constant), measure the rate at zero time for each case, and analyze the data by the previous methods. An advantage of this technique over the isolation method is that the concentrations of the reactants can be nearly equal instead of some concentrations being in large excess. The method allows reaction rates to be obtained over the entire range of composition with respect to known major reaction participants. However, these rates may not be equal to those obtained from experiments in which extensive conversions are allowed, due to the presence of trace byproducts generated during the reaction that affect reaction rate.

For more complex reacting systems, the simple modeling procedures described above are not sufficient. In these more complex systems, the kinetic model describing the system will more generally consist of a group of coupled, nonlinear differential equations. For specific systems, standard analysis techniques are available (W1, B5); more general procedures have also been made available through recent developments of computing hardware and software, and are discussed in the Sections III–V. One very real danger in the treatment of any such system is the inclusion of extraneous and unnecessary parameters, simply because the descriptive equations become so complex that the contributions of the parameters are hidden. Also, with such multi-

parameter systems, convergence of nonlinear estimation programs can become difficult. Techniques reducing such problems are described.

B. Hyperbolic Models

The hyperbolic model types have very commonly been used in the analysis of kinetic data, as discussed in Section I. Such applications are sometimes justified on the theoretical bases already alluded to, or simply because models of the form of Eq. (2) empirically describe the existing reaction-rate data. Considerably more complex models are quite possible under the Hougen–Watson formalism, however. For example, Rogers, Lih, and Hougen (R1) have proposed the competitive–noncompetitive model

$$r = \frac{\gamma_1 K_1 K_2 p_1 p_2}{(1 + K_1 p_1 + K_2 p_2)^2} + \frac{\gamma_2 K_1 K_2 p_1 p_2}{(1 + K_1 p_1)(1 + K_1 p_1 + K_2 p_2)} \tag{7}$$

for cases in which large and small molecules react over a solid catalyst.

Shabaker (S1), for the hydrogenation of propylene over a platinum alumina catalyst, selected a Hougen–Watson model of the form

$$r = k_1 K_1 K_2 L^2 (C_L/L)^3 p_1 p_2 + k_2 K_1 L (C_L/L)^2 p_1 p_2 \tag{8}$$

where

$$\frac{C_L}{L} = -\frac{(1 + K_2 p_2)}{4K_1 p_1} + \left[+ \frac{(1 + K_2 p_2)^2}{16 K_1^2 p_1^2} + \frac{1}{2K_1 p_1} \right]^{1/2}$$

The adsorption and rate constants in each case exhibit an exponential temperature dependence. The study of these models, in addition to the linearizable models of the form of Eq. (2), have become possible through the use of nonlinear least squares (see Section III).

The usual data analysis procedures for the linearizable models typified by Eq. (2) consist of (1) isolating a class of plausible rival models by means of plots of initial reaction-rate data as a function of total pressure, feed composition, conversion, or temperature; (2) fitting the models passing the screening requirements of the initial rates by linear least squares, and further rejecting models based upon physical grounds.

1. *Dependence of Reaction Rate on Experimental Variables*

The examination of the initial rate dependence upon total pressure is by far the most common means of examining the various classes of rival models in terms of their ability to fit the observed data. Yang and Hougen (Y1) have presented a classification of a multitude of possible models in terms of the dependence of the predicted rate upon total pressure. Generally speaking, the

analysis may be performed by writing each of the candidate models in terms of its dependence upon total pressure. For an alcohol dehydrogenation, for example, two models that might be considered are the dual site

$$r = \frac{\bar{k}\bar{K}_E(p_E - p_A p_H/K)}{(1 + \bar{K}_E p_E + \bar{K}_A p_A + \bar{K}_H p_H)^2} \tag{9}$$

and the single site

$$r = \frac{\bar{k}\bar{K}_E(p_E - p_A p_H/K)}{(1 + \bar{K}_E p_E + \bar{K}_A p_A)} \tag{10}$$

For initial reaction rates r_0 with a pure alcohol feed, all product partial pressures are zero, and $p_{AO} = y_{AO}\pi$, so that for the dual site

$$r_0 = \frac{kK_E\pi}{(1 + K_E\pi)^2} \tag{11}$$

and for the single site

$$r_0 = \frac{kK_E\pi}{(1 + K_E\pi)} \tag{12}$$

At low π, the denominator simplifies to unity in each case and both models are linear in π. For sufficiently high π, the parenthesis in the denominator approaches $K_E\pi$; the initial rate for the dual-site model then approaches zero, and that of the single-site model approaches a constant value. Thus the plot of the experimental data will indicate that the dual-site model is preferable if a maximum exists in the data, or that the single-site model is preferable if a horizontal high-pressure asymptote exists. Hence, for the data of Franckaerts and Froment (F1) shown in Fig. 2, the dual-site model is preferred over the single-site model.

In addition to the maximum point, inflection points of the rate data can be used for testing model adequacy (M6).

The more commonly applied procedure, however, is to linearize the models of Eqs. (11) and (12) to yield

$$\left(\frac{\pi}{r_0}\right)^{1/2} = \frac{1}{(kK_E)^{1/2}} + \frac{K_E}{(kK_E)_{1/2}}\pi \tag{13}$$

$$\left(\frac{\pi}{r_0}\right) = \frac{1}{kK_E} + \frac{1}{k}\pi \tag{14}$$

The data should lie along a straight line when plotted as $(\pi/r_0)^{1/2}$ vs π, if the dual-site model is adequate. If, in addition, the dual-site model is to be preferred over the single-site model, the plot of π/r_0 vs π should be curved

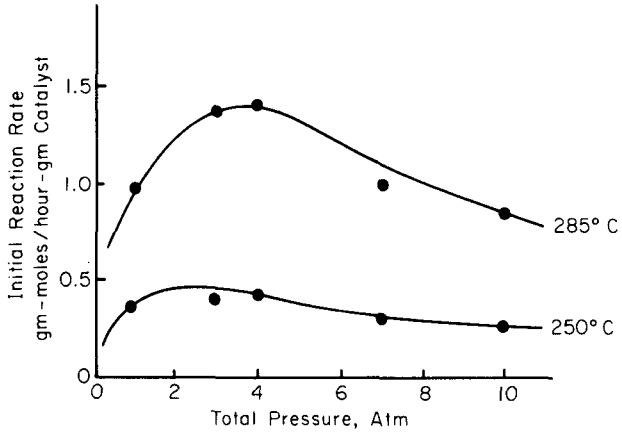

FIG. 2. Initial reaction rate versus total pressure for alcohol dehydrogenation, 250 and 285°C.

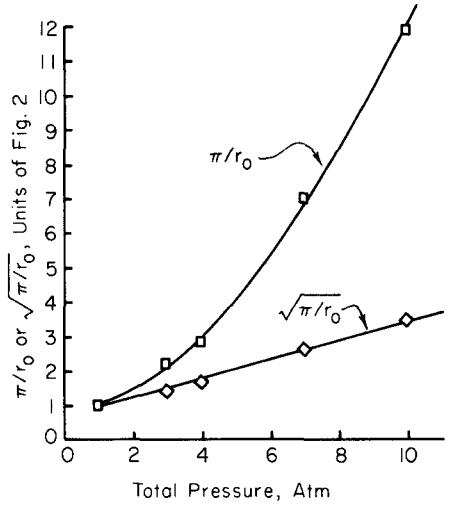

FIG. 3. Linearized initial rate plots for alcohol dehydrogenation, 285°C.

(cf., for example, Section V). The data of Fig. 3, replotted from Fig. 2 according to Eqs. (13) and (14), suggest that the dual-site model is adequate and is preferred over the single-site model.

Often it is advisable to utilize the slopes and intercepts of such plots as well as the above described curvature of the data. Kabel and Johanson (K1), for example, could eliminate an adsorption-controlling model with only 70%

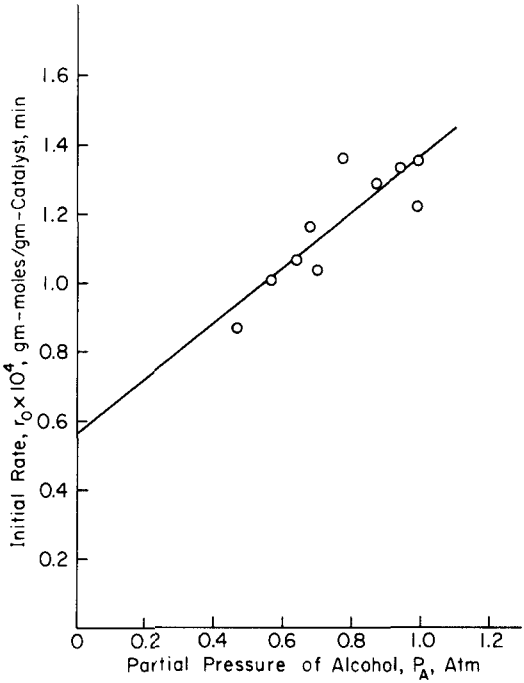

FIG. 4. Linearized initial rate plot for alcohol dehydration, Eq. (16).

certainty using the line curvature for the model representing dehydration of ethanol to ether over an ion-exchange-resin catalyst:

$$r = \frac{kL(p_A - p_E p_W/K)}{1 + [(K_A/K)^2 p_E p_W]^{1/2} + K_W p_W} \tag{15}$$

For ten data points reported at varying alcohol partial pressures (p_A) and zero water partial pressure (p_W), this model becomes

$$r_0 = kLp_A \tag{16}$$

This suggests that a plot of r_0 vs p_A (Fig. 4) or r_0/p_A vs p_A should be linear. It is very difficult to reject this model on the basis of data curvature, even though it is evident that some curvature could exist in Fig. 4. However, Eq. (16) demands that Fig. 4 also exhibit a zero intercept. In fact, the 99.99% confidence interval on the intercept of a least-square line through the data does not contain zero. Hence the model could be rejected with 99.99% certainty.

The dependence of the reaction rate upon conversion, temperature, and composition has not been used as widely as the pressure dependence in model screening. The conversion behavior is discussed in detail in Section V.

The utility of the overall dependence of the reaction rate upon temperature appears to be slight, perhaps in some cases allowing a discrimination between adsorption and surface reaction classes of models. The study of the residuals of various estimated parameters as a function of temperature can clearly indicate model inadequacies (Section IV) and, in some cases, can lead to model modifications correcting these model inadequacies (Section V).

2. *Magnitude of Estimated Parameters*

The second stage of this analysis involves the fitting of the model to the data and examination of the parameter values thus obtained to ensure their physical reasonableness. In particular, (a) the estimated adsorption and rate constants should be positive; (b) a plot of the logarithm of the rate constant with reciprocal absolute temperature should be linear with a negative slope; (c) a plot of the logarithm of the adsorption constant with reciprocal absolute temperature should be linear with a positive slope (exothermic adsorption) and generally a negative intercept (decrease in entropy with adsorption); (d) the model should adequately fit the data (Section IV) with the best estimates of the parameter values. These criteria should, of course, be used with caution. For example, it is possible that adsorption could be endothermic if some of the adsorbed molecules dissociate. The use of these criteria has been discussed at length elsewhere (K7). Much of the present discussion is necessarily deferred until Section III. However, general cautionary measures must include an examination of the confidence interval for any parameter estimate before one rejects a given model. For example, it is illustrated in Section III that although an estimated adsorption constant is negative, its confidence region can easily contain positive values. Also, constants estimated to be negative by unweighted linear least squares can be estimated to be positive by nonlinear least squares or weighted linear least squares (Section III). The general utility of examining the magnitude of the estimated parameters can be illustrated by the data of Kabel and Johanson (K1), which, at zero water partial pressure, can be shown to reject the single-site model

$$r = \frac{kK_A L(p_A^2 - p_E p_W/K)}{(1 + K_A p_A + K_W p_W)} \tag{17}$$

The appropriate linear plot of p_A^2/r_0 vs p_A is shown in Fig. 5. With the amount of scatter in the data for this plot, there is no justification for rejecting the model because of curvature of correlation in the locus of data points. However, the intercept of this graph suggests that the adsorption constant is significantly negative. If a model that possesses theoretical insight into the reacting system is desired, then this model must be rejected.

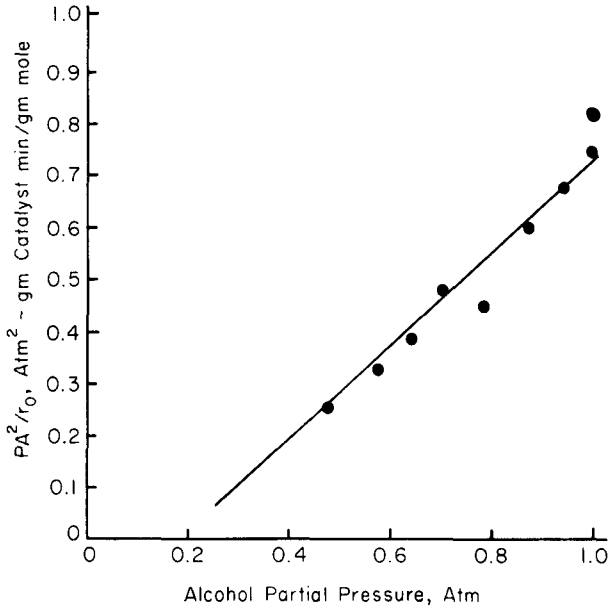

FIG. 5. Linearized initial rate plot for alcohol dehydration, Eq. (17).

An example of a model rejection based upon temperature coefficients can be obtained from the single-site model considered by Ayen and Peters (A3) for the reduction of nitric oxide over a commercial catalyst:

$$r = \frac{kK_{NO}\,p_{NO}\,p_{H_2}}{1 + K_{NO}\,p_{NO} + K_{H_2}\,p_{H_2}} \tag{18}$$

This model was fitted to the data of all three temperature levels, 375, 400, and 425°C, simultaneously using nonlinear least squares. The parameters were required to be exponentially dependent upon temperature. Part of the results of this analysis (K6) are reported in Fig. 6. Note the positive temperature coefficient of this nitric oxide adsorption constant, indicating an endothermic adsorption. Such behavior appears physically unrealistic if NO is not dissociated and if the confidence interval on this slope is relatively small. Ayen and Peters rejected this model also.

III. Parameter Estimation

In any modeling procedure, values of the predicted rate must approximate the values of the observed rate before an adequate model has been established. A simple indication of the comparative shapes of the predicted and

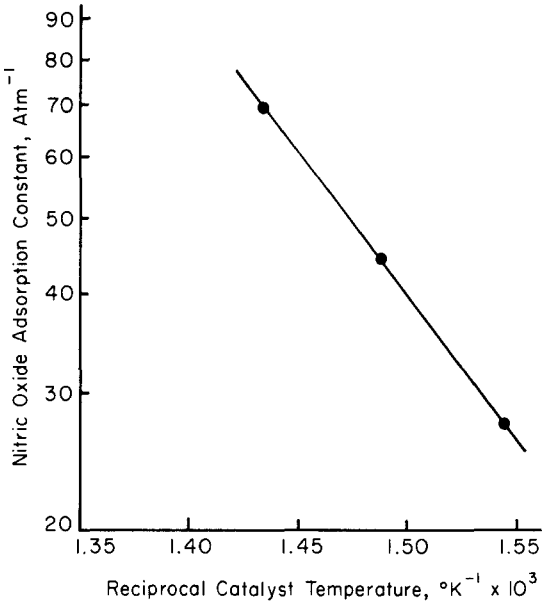

FIG. 6. Temperature dependence of adsorption constant for nitric oxide reduction, Eq. (18).

observed rate surfaces can be obtained by procedures described in Section II. For the final check of the agreement of the model and the data as well as for further use of the model, estimates of the parameters within the model must be obtained. Linear and nonlinear least squares have proven to be of significant value for this purpose. For effective utilization of the parameter estimates thus obtained, the precision of the estimates should also be measured.

A. LEAST SQUARES

1. *Linear Least Squares*

Parameter values producing predicted-model results near the observed results have frequently been selected by minimizing S:

$$S = \sum_{u=1}^{N} (y_u - \hat{y}_u)^2 \tag{19}$$

This method of *least squares* is not only intuitively desirable, but also provides estimates having desirable properties, if certain assumptions are met (D4).

For models that are linear in the parameters, that is, models of the form

$$\hat{y} = \sum_{i=1}^{p} b_i x_i \tag{20}$$

the parameters may be estimated analytically by linear least squares. In using unweighted linear least squares, however, it is desirable that the x's be free of error, that the response y have constant error variance, and that this error be independent and of mean zero.

For a simple linear model

$$\hat{y}_u = b_0 + b_1 x_u \tag{21}$$

the least-squares estimates b_0 and b_1 may be obtained by substituting Eq. (21) into Eq. (19), differentiating, and setting the first derivatives to zero:

$$\frac{\partial S}{\partial b_0} = -2\sum_{u=1}^{N}(y_u - b_0 - b_1 x_u) = 0 \tag{22}$$

$$\frac{\partial S}{\partial b_1} = -2\sum_{u=1}^{N} x_u(y_u - b_0 - b_1 x_u) = 0 \tag{23}$$

Solving these two equations in two unknowns,

$$b_1 = \frac{N\sum x_u y_u - \sum x_u \sum y_u}{N\sum x_u^2 - (\sum x_u)^2} \tag{24}$$

$$b_0 = \frac{\sum x_u^2 \sum y_u - \sum y_u \sum x_u y_u}{N\sum x_u^2 - (\sum x_u)^2} \tag{25}$$

where the summations range from $u = 1$ to $u = N$. In the more general matrix notation, the least-squares estimates for Eq. (20) become

$$\mathbf{b} = (\mathbf{X}^1\mathbf{X})^{-1}\mathbf{X}^1\mathbf{Y} \tag{26}$$

where $\mathbf{b}$ is a $p \times 1$ column vector, $\mathbf{Y}$ is an $N \times 1$ column vector, and $\mathbf{X}$ is an $N \times p$ matrix of the levels of x. For Eq. (21),

$$\mathbf{b} = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} \tag{27}$$

$$\mathbf{Y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix} \tag{28}$$

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ \vdots & \vdots \\ 1 & x_N \end{bmatrix} \tag{29}$$

The equations represented by Eq. (26) are termed the normal equations; they provide parameter estimates possessing the minimum distance from the

data **Y** to the model surface, thus representing a line that is perpendicular (or normal) from the point **Y** to the surface.

Many kinetic equations can be suitably linearized to the form of Eq. (20). For example, Eq. (1) can be transformed logarithmically, or Eq. (2) can be transformed reciprocally. Two equations proposed for describing pentane-isomerization data (C1, J1) are the single site

$$r_1 = \frac{kK_2(p_2 - p_3/K)}{(1 + K_1 p_1 + K_2 p_2 + K_3 p_3)} \tag{30}$$

and the dual site

$$r_2 = \frac{kK_2(p_2 - p_3/K)}{(1 + K_1 p_1 + K_2 p_2 + K_3 p_3)^2} \tag{31}$$

where p_1, p_2, and p_3 are the partial pressures of hydrogen, normal pentane, and isopentane. Rearranging and redefining parameters,

$$\left(\frac{p_2 - p_3/K}{r_1}\right) = b_0 + b_1 p_1 + b_2 p_2 + b_3 p_3 \tag{32}$$

$$\left(\frac{p_2 - p_3/K}{r_2}\right)^{1/2} = b_0 + b_1 p_1 + b_2 p_2 + b_3 p_3 \tag{33}$$

Unweighted linear least-squares estimates of these parameters are shown in Table I. The negative parameter estimates of the single-site model are of no

TABLE I

PARAMETER ESTIMATES FOR PENTANE ISOMERIZATION

	Least-square technique		
Parameters	Unweighted linear	Weighted linear	Unweighted nonlinear
Single Site			
k	189	30.5	36.2
K_1	−0.4	2.0	0.9
K_2	−0.03	1.3	0.5
K_3	−1.08	4.3	2.1
Dual Site			
k	289	73.9	133
K_1	0.03	0.03	0.03
K_2	0.005	0.02	0.02
K_3	0.09	0.07	0.07

great concern, as they arise from a negative estimate of b_0 in Eq. (32) having a confidence interval (Section III,C) so large that positive values of b_0 may be chosen without significantly harming the fit of the data.

Several variations of these concepts (e.g., stepwise regression) have also been proposed (D4).

2. *Weighted Linear Squares*

Thus far, we have assumed that all of the observations are independent and have the same inherent uncertainty. Deviations from such assumptions have been discussed in a generalized fashion elsewhere (H8).

Taking into account deviations from constant variance is a particularly straightforward matter. If, for example, the variable y of Eq. (19) possesses a variance σ_i^2 which varies from data point to data point, then Eq. (19) divided by σ_i^2 is the appropriate sum of squares to minimize. If, in general, we define

$$\mathbf{C} = \begin{bmatrix} \sigma_i^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_N^2 \end{bmatrix} \tag{34}$$

then Eq. (26) becomes

$$\mathbf{b} = (\mathbf{X}^1\mathbf{C}^{-1}\mathbf{X})^{-1}\mathbf{X}^1\mathbf{C}^{-1}\mathbf{Y} \tag{35}$$

or if $\hat{y} = \hat{b}x$, this is simply

$$\hat{b} = \frac{\sum (x_i y_i/\sigma_i^2)}{\sum (x_i^2/\sigma_i^2)} \tag{36}$$

where the sum is taken over all observations. Now, if replication (Section IV) has been carried out at each data point, then the estimate of error variance at each data point, σ_i^2, can be inserted into Eq. (34) and the least-squares estimate calculated from Eq. (35).

More typically, we have an indication that a transformed variable $f(y)$ has constant error variance and will wish to use this information to weight y appropriately. For example, we may suspect $\log y$ has constant error variance and wish to fit y. More typically, we might feel that y has constant error-variance and wish to fit $1/y$.

This problem was solved approximately in 1947 (B2, K9, J1), wherein it was suggested that the transformed function of y be expanded in a linear Taylor series to provide

$$\text{variance}\,[f(y)] = (\partial f(y)/\partial y)^2 \cdot \text{variance}\,(y) \tag{37}$$

For example, if the rate r is thought to have constant error variance and we wished to fit $f(y) = (a/r)^n$ as in Eq. (33), then the quantity

$$\sigma_i^2 = [-nf(y_i)/r_i]^2\sigma_r^2 \tag{38}$$

should be used in the weighting matrix $\mathbf{C}$ of Eq. (34) (the constants n and σ_r^2 are unnecessary). For comparison to the results of the unweighted least-squares analysis of the pentane isomerizations data just discussed, the second column of Table I should be examined. Here, the weighting function of Eq. (38) was used, assuming the rate r had constant error variance [in contrast to the function of r defined in Eqs. (32) and (33) in the unweighted case]. Note the significant shift in the weighted estimates from the unweighted estimates, including the positive single-site parameter estimates. Since the nonlinear least-squares results also assume r has constant error variance, there is a close correspondence in the weighted linear and the nonlinear results.

Equations (34) and (35) can also be generalized to include correlation between observations. In such a case

$$\mathbf{C} = \begin{bmatrix} \sigma_{11}^2 & \sigma_{12}^2 & \cdots & \sigma_{1N}^2 \\ \sigma_{21}^2 & \sigma_{22}^2 & \cdots & \sigma_{2N}^2 \\ \vdots & \vdots & & \vdots \\ \sigma_{N1}^2 & \sigma_{N2}^2 & \cdots & \sigma_{NN}^2 \end{bmatrix} \tag{39}$$

where the $\sigma_{ij}^2 (i \neq j)$ are the covariances between the observation errors. This variance–covariance matrix is again applied through Eq. (35).

3. *Nonlinear Least Squares*

Many of the models encountered in reaction modeling are not linear in the parameters, as was assumed previously through Eq. (20). Although the principles involved are very similar to those of the previous subsections, the parameter-estimation procedure must now be iteratively applied to a nonlinear surface. This brings up numerous complications, such as initial estimates of parameters, efficiency and effectiveness of convergence algorithms, multiple minima in the least-squares surface, and poor surface conditioning.

Iterative Techniques. In estimating parameters in a model that is nonlinear in the parameters

$$r = f(\mathbf{x}; \mathbf{K}) \tag{40}$$

we still minimize S of Eq. (19). Since the partial derivatives similar to Eqs. (22) and (23) do not generally lead to equations that are easily solved, we turn to one of three basic types of methods (K8), even though a wide variety of specific minimization techniques are available (S2).

In the *linearization method*, the nonlinear model of Eq. (40) is linearized by a truncated Taylor expansion:

$$y_u = r_u - f(\mathbf{x}_u ; \mathbf{K}) = \sum_{i=1}^{p} f'_{iu} b_i \tag{41}$$

where

$$f'_{iu} = \frac{\partial f(\mathbf{x}_u ; \mathbf{K})}{\partial K_i} \bigg| \mathbf{K} = \mathbf{K}^\circ \tag{42}$$

Here $\mathbf{K}^\circ$ is the set of initial parameter estimates. Since Eq. (41) is linear, estimates of the correction vector $\mathbf{b}$ may be obtained through Eq. (26), where y_u is now given by Eq. (41) and $\mathbf{X}$ is given by

$$\mathbf{X} = \{f'_{iu}\} \tag{43}$$

Using estimates thus obtained, improved estimates of K_i are given by

$$K_i^{(1)} = b_i + K_i^\circ \tag{44}$$

The procedure is repeated until the procedure converges, that is until the correction vector $\mathbf{b}$ approaches zero. Methods of modifying the size of the correction vector have been developed to improve convergence (B5) and this method, in theory, should always converge (H1). In practice, nonlinearities in the model and poor parameter estimates can prevent convergence. A modification of this method linearizes the normal equations; higher derivatives have also been used.

In the *steepest-descent method* (B10, B18, K8, S2, R2), the sums-of-squares surface is visualized to be a response surface in which the parameters are variables. Steepest-descent procedures are then applied to determine the minimum of the surface. The procedure is, then, to (a) select initial parameter estimates; (b) set up a first-order design (B18) in the parameter space about these latter estimates, calculating the sum of squares at each point; (c) determine the direction of steepest descent; (d) proceed in this direction in the parameter space until the sums of squares begin increasing; (e) if the minimum is not near, go back to step (b); otherwise (f) set up a second-order design for precise location of the minimum, and continue toward the minimum, reparameterizing (Section III,B) as necessary to improve surface conditioning. Although this method converges for nearly any set of initial estimates, its convergence can be agonizingly slow for multiparameter problems (K8).

Experience with fitting many models indicates that the steepest descent is very stable for the initial iterations, while the linearization method is more efficient for the final iterations. A *compromise method* has been suggested (L5, M1, M2) which tends to emphasize steepest descent at the outset and

linearization in the final stages. In this method, the correction vector used in Eq. (44) is calculated from

$$\mathbf{b} = (\mathbf{X}^1\mathbf{X} + \psi\mathbf{I})^{-1}\mathbf{X}^1\mathbf{Y} \tag{45}$$

where y_u is given by Eq. (41) and $\mathbf{X}^1\mathbf{Y}$ is the vector of steepest descent. Thus, a large value of ψ makes the correction vector near the vector of steepest descent while a small ψ makes the correction vector near that obtained by the linearization method. The compromise method, then, incorporates a continually decreasing value of ψ as the iterations proceed.

Numerous examples of applications of nonlinear least squares to kinetic-data analysis have been presented (K7, K8, L3, L4, M7, P2); an exhaustive tabulation of references would, at this point, approach 100 entries. Typical results of a nonlinear estimation and comparison to linear estimates are shown in Table I and discussed in Section III,A,2. Many estimation problems exist, however, as typified in part by Fig. 7. This is the sum-of-squares surface obtained at fixed values of K_s and K_u in the rate equation used for the catalytic hydrogenation of mixed isooctenes (M7)

$$r = \frac{kK_H K_U p_H p_U}{(1 + K_H p_H + K_U p_U + K_S p_S)^2} \tag{46}$$

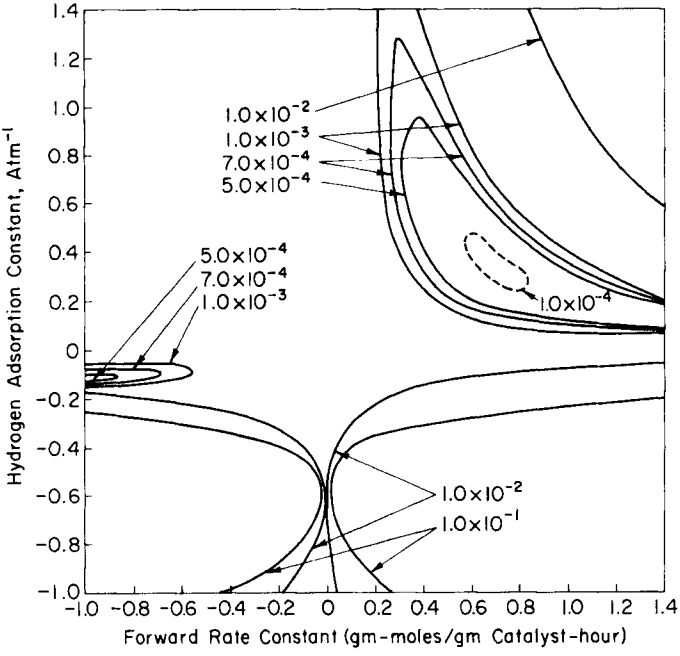

FIG. 7. Contours of sums of squares of residual rates for isooctene hydrogenation, Eq. (46).

First, this figure shows long skewed contours that frequently occur with such models. Also, it should be noted that two minima exist in this plot, in the first and third quadrants. Accordingly, in the estimation of the rate parameters for this model with the standard nonlinear estimation techniques, it would be possible to obtain certain negative parameter estimates. In fact, it can be seen from Eq. (46) that any two of the parameters k, K_H, or K_U can simultaneously be negative without greatly hindering the fit of the equation to the data owing to the fact that they are multiplied together in the numerator. These considerations are of importance in the estimation of the parameter values, since negative estimates of any adsorption constants are generally used as a justification for rejecting the model under consideration. Therefore, one should attempt to examine the sums-of-squares surface in enough detail to ensure that the model should truly be rejected and not just refitted. Measures of nonlinearity of the sum-of-squares surface are available (B3, B8, M2, M7).

Part of the problems encountered with sums-of-squares surfaces such as shown in Fig. 7 can be relieved by obtaining precise initial-parameter estimates for the nonlinear estimation. Employing a pilot or first-stage estimation procedure to obtain values $\mathbf{K}^\circ$ to use in an iterative nonlinear estimation program is often more efficient than simply making rough guesses, even though at first sight it may appear to be more time-consuming. Numerous examples of this type of pilot estimation have been described (K8). Figure 8, for example, presents the sums-of-squares surface for alcohol dehydration data and the model of Eq. (47). As indicated in the figure, any trial value of the two parameters in the first quadrant will, if restricted to the first quadrant, cause the nonlinear estimation procedure to diverge (M7):

$$r = \frac{kK_A(p_A{}^2 - p_E p_W/K)}{(1 + K_A p_A + K_E p_E + K_W p_W)} \tag{47}$$

A brief survey of the sums-of-squares surface, such as shown in this figure, can eliminate many computer hours of fruitless iteration with repeated initial parameter estimates. Other problems related to estimation from Figs. 7 and 8 can be reduced by reparameterization, as discussed in Section III,B.

If the nonlinear estimation procedure is carefully applied, a minimum in the sums-of-squares surface can usually be achieved. However, because of the fitting flexibility generally obtainable with these nonlinear models, it is seldom advantageous to fit a large number of models to a set of data and to try to eliminate inadequate models on the basis of lack of fit (see Section IV). For example, thirty models were fitted to the alcohol dehydration data just discussed (K2). As is evident from the residual mean squares of Table II, approximately two-thirds of the models exhibit an acceptable fit of the data

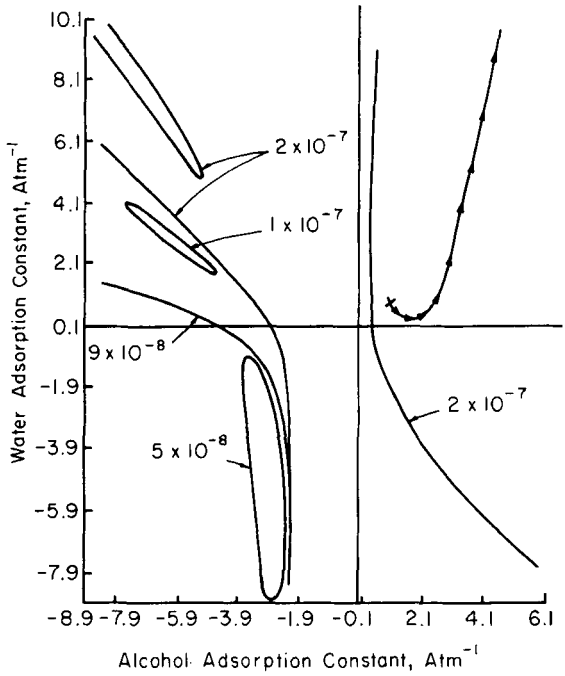

FIG. 8. Contours of sums of squares of residual rates for alcohol dehydration, Eq. (47).

points used (rms $\leqq 1.2 \times 10^{-6}$). The experimental design, as is shown later, plays a major role in determining the success of such a project (Section VII).

For fitting such a set of *existing* data, a much more reasonable approach has been used (P2). For the naphthalene oxidation system, major reactants and products are symbolized in Table III. In this table, letters in bold type represent species for which data were used in estimating the frequency factors and activation energies contained in the body of the table. Note that the rate equations have been reparameterized (Section III,B) to allow a better estimation of the two parameters. For the first entry of the table, then, a model involving only the first-order decomposition of naphthalene to phthalic anhydride and naphthoquinone was assumed. The parameter estimates obtained by a nonlinear-least-squares fit of these data, $\hat{\theta}$, are seen to be relatively precise when compared to the standard errors of these estimates, $s_{\hat{\theta}}$. The residual mean square, using these best parameter estimates, is contained in the last column of the table. This quantity should estimate the variance of the experimental error if the model adequately fits the data (Section IV). The remainder of Table III, then, presents similar results for increasingly complex models, each of which entails several first-order decompositions.

TABLE II: SUMMARY OF PARAMETER ESTIMATES FOR ALCOHOL DEHYDRATION

Model No.	$k \times 10^4$	K_A	K_E	K_W	K_{E1}-	K_{W1}-	K_{A1}-	L	Residual mean square $\times 10^{10}$
1	1.49	0	0	4.09	—	—	—	—	1.09
2	2.26	3.39	0.13	7.31	—	—	—	—	0.70
3	8.63	10^{-8}	—	3.68	10^{-8}	—	—	0.17	1.13
4	5.28	46.5	—	91.4	666	—	—	0.51	0.75
5	8.98	0	0	—	—	40.3	—	0.17	1.36
6	9.72	0	0.19	2.71	10^{-7}	—	—	0.15	1.67
7	16.2	4.26	0.04	1.34	10^{-6}	—	—	0.08	6.71
8	15.8	0	0	8.40	—	8.40	—	0.10	1.45
9	5.0	54.3	0	4.0	—	10^3	—	0.50	1.45
10	1.44	0	0	4.51	—	—	—	—	1.05
11	7.04	2.89	0	6.30	—	—	—	—	0.70
12	1.48	10^4	0	10^4	—	—	—	—	1.12
13	7.6	0	—	4.7	10^{-7}	—	—	0.19	1.08
14	56.0	15.4	—	30.6	52.9	—	—	0.64	0.77
15	8.9	0	0	—	—	45.7	—	0.17	1.29
16	197	4.4	0	—	—	253	—	0.21	0.61
17	5.8	0	0.01	4.03	10^{-5}	—	—	0.25	1.16
18	2.77	2.74	2.98	4.71	10^{-5}	—	—	0.06	78.0
19	8.16	0	0	0.44	—	10^{-8}	—	0.14	5.75
20	9.95	3.69	0	2.01	—	16.6	—	0.90	0.63
21	16.6	—	0	1.86	—	—	10^{-3}	0.30	1.06
22	9.94	—	0	10^4	—	—	10^9	0.15	1.15
23	34.4	—	—	1.74	0	—	0	0.21	35.3
24	9.94	—	—	10^4	0	—	109	0.15	1.15
25	6.5	—	0	—	—	5.4	0	0.48	1.12
26	17.4	—	0	—	—	0	8.51	0.17	5.7
27	38.1	—	0.05	1.03	10^{-6}	—	0	0.19	1.50
28	17.4	—	0	0	10^{-9}	—	8.49	0.17	5.9
29	9.4	—	0	1.15	—	0.90	0	0.40	1.1
30	11.0	—	0	10^5	—	0	10^{10}	0.13	1.2

By slowly increasing the complexity of the models in this fashion, it was hoped that a model could be obtained that was just sufficiently complex to allow an adequate fit of the data. This conscious attempt to select a model that satisfies the criteria of adequate data representation and of minimum number of parameters has been called the principle of parsimonious parameterization. It can be seen from the table that the residual mean squares progressively decrease until entry 4. Then, in spite of the increased model complexity and increased number of parameters, a better fit of the data is not obtained. If the reaction order for the naphthalene decomposition is estimated, as in entry 5, the estimate is not incompatible with the unity order of entry 4. If an additional step is added as in entry 6, no improvement of fit is obtained. Furthermore, the estimated parameter for that step is negative and poorly defined. Entry 7 shows yet another model that is compatible with the data. If further discrimination between these two remaining rival models is desired, additional experiments must be conducted, for example, by using the model discrimination designs discussed later. The critical experiments necessary for this discrimination are by no means obvious (see Section VII).

B. Reparameterization

The problems described in the last subsection can frequently be reduced by improving the surface conditioning through reparameterization. For example, if the two parameters in the first-order model

$$C_B = \exp\{-k_0 \exp(-E/RT)\} \tag{48}$$

are estimated from a set of data (H9), then one contour of the sums-of-squares surface can be represented as shown in Fig. 9. With such "natural" parameterization, an iterative nonlinear least-squares routine will generally converge to the minimum sum of squares quite slowly or not at all. However, writing the model in an equivalent mathematical form, but with different parameters, can lead to better sums-of-squares surface conditioning.

Considerable discussion of reparameterization and examples of its usefulness have been published (B3, B8, B12, G1, G2, M7). Although several specific techniques are useful, one reparametrization of kinetic models often necessary is a redefinition of the independent variables so that the center of the new coordinate system is near that of the experimental design. In particular, the exponential parameter

$$k = k_0 \exp(-E/RT) \tag{49}$$

TABLE III

SUMMARY OF NONLINEAR FITTING OF NAPHTHALENE OXIDATION DATA

Entry	Mechanism[a]		$k_0' = k_0 \exp(-E/RT)$ 1	2	3	4	5	6	$E' = E/R$ 1	2	3	4	5	6	α	$s^2 \times 10^3$
1	*N*—1→**P**	$\hat{\theta}$	8.79	0.842	—	—	—	—	14.5	12.4	—	—	—	—	—	1.06
	N—2→**Q**	$s_{\hat{\theta}}$	0.26	0.162	—	—	—	—	0.4	2.8	—	—	—	—	—	
2	*N*—1→**P**	$\hat{\theta}$	9.23	0.873	0.833	—	—	—	14.6	12.4	9.68	—	—	—	—	0.772
	N—2→**Q**	$s_{\hat{\theta}}$	0.26	0.146	0.151	—	—	—	0.4	2.8	2.34	—	—	—	—	
	N—3→**G**															
3	*N*—1→**P** ↑4	$\hat{\theta}$	8.10	2.16	0.882	18.8	—	—	15.0	11.9	9.80	7.41	—	—	—	0.671
	N—2→**Q**	$s_{\hat{\theta}}$	0.61	0.59	0.142	7.8	—	—	1.0	3.2	2.09	5.60	—	—	—	
	N—3→**G**															
4	*N*—1→**P**—5→**M** ↑4	$\hat{\theta}$	8.10	3.00	0.881	28.0	0.831	—	15.6	13.0	8.86	9.69	16.8	—	—	0.471
	N—2→**Q**	$s_{\hat{\theta}}$	0.73	0.74	0.124	8.9	0.221	—	1.1	2.6	3.96	.80	3.5	—	—	
	N—3→**G**															

5	N—1→P—5→M ↑4 N—2→Q N—3→G	$\hat{\theta}$ $s_{\hat{\theta}}$	8.26 0.88	3.55 1.11	0.926 0.142	31.3 10.4	0.750 0.236	— —	15.9 1.3	13.5 2.8	9.93 1.85	17.6 3.9	9.05 3.8	— —	1.13 0.15	0.475
6	N—1→P—5→M ↑4 N—2→Q ↓6 N—3→G	$\hat{\theta}$ $s_{\hat{\theta}}$	8.09 0.72	2.76 0.74	1.20 0.48	29.2 9.0	0.861 0.227	−4.15 6.10	15.6 1.1	12.8 2.7	11.1 3.2	16.6 3.6	8.84 3.90	10.9 12.1	— —	0.480
7	N—1→P—5→M ↑4 N—2—Q—6→G	$\hat{\theta}$ $s_{\hat{\theta}}$	7.62 1.05	4.16 1.10	— —	32.0 13.1	0.787 0.234	10.7 1.9	15.4 1.5	12.6 2.5	— —	9.60 4.64	17.8 3.9	3.58 2.33	— —	0.497

[a] N, naphthalene; P, phthalic anhydride; M, maleic anhydride; Q, naphthoquinone; G, off gases.
[b] $\propto$ denotes reaction order for naphthalene decomposition steps.

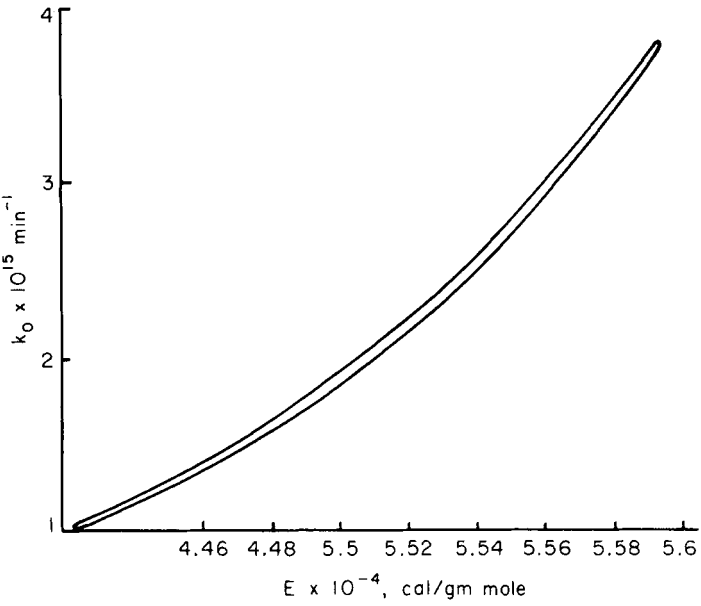

FIG. 9. Approximate 95% confidence region for first-order decomposition model.

should almost invariably be redefined as

$$k = k_0' \exp\left\{-\frac{E}{R}\left(\frac{1}{T} - \frac{1}{\bar{T}}\right)\right\} \tag{50}$$

where

$$k_0' = k_0 \exp(-E/R\bar{T}) \tag{51}$$

In the example above, the sums-of-squares surface is transformed to that shown by Fig. 10. The best point estimates of the parameters of k_0' and E are thus more readily obtained than k_0 and E; the initial-parameter estimates are less critical, and the estimation routine converges more rapidly to the minimum. Although the correlation between the parameter estimates has been reduced by this reparameterization, the size of the confidence region of the original parameters k_0 and E will not change.

C. Confidence Intervals and Regions

In reaction-rate modeling, precise parameter estimates are nearly as essential as the determination of the adequate functional form of the model. For example, in spite of imprecisely determined parameters, an adequate model will still predict the data well over the range that the data are taken,

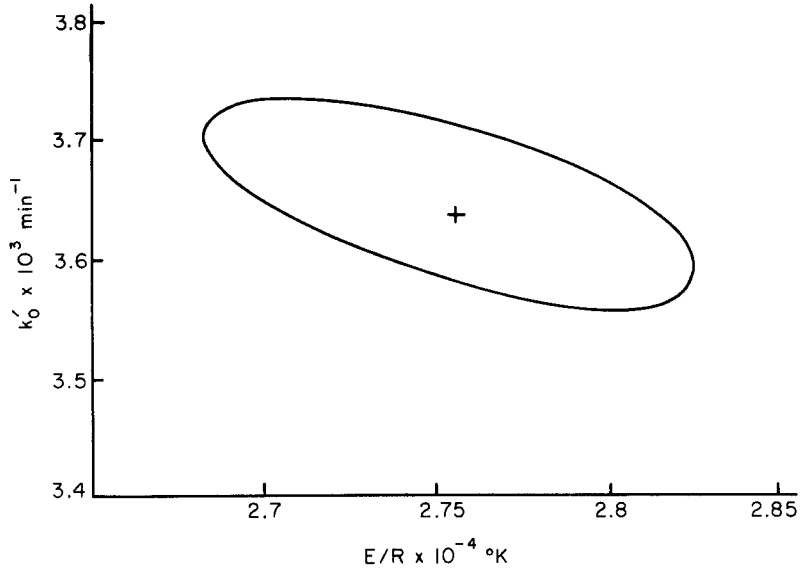

FIG. 10. Approximate 95% confidence region for reparameterized first-order decomposition model.

as shown in Fig. 11 (K11). However, if these parameter estimates are to be subjected to mechanistic interpretation or are to be used in reactor designs (with slight extrapolation), then clearly additional information about the precision of the estimates is required. An example of the first situation is provided by the negative parameter estimates of Table I. As is shown clearly by the confidence region for this entry, positive parameter estimates are quite compatible with the data (K3).

The amount of uncertainty in parameter estimates obtained for the hyperbolic models is particularly large. It has been pointed out, for example, that parameter estimates obtained for hyperbolic models are usually highly correlated and of low precision (B16). Also, the number of parameters contained in such models can be too great for the range of the experimental data (W3). Quantitative measures of the precision of parameter estimates are thus particularly important for the hyperbolic models. (C2).

1. *Linear Models*

For models of the form of Eqs. (32) and (33) [or, more generally, Eq. (20)], the $(1 - \alpha)$ 100% confidence interval for the parameters b_i (not K_i) is given by

$$b_i \pm t_{\nu,\,\alpha/2}[v(b_i)]^{1/2} \tag{52}$$

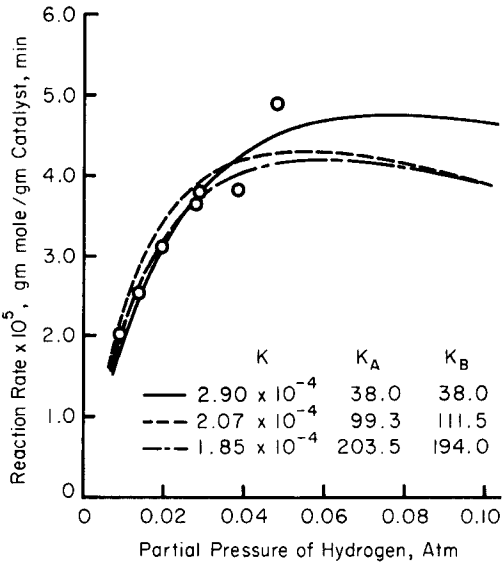

FIG. 11. Dependence of predicted reaction rate upon assumed parameter estimates, Eq. (149).

where $t_\nu, \alpha/2$ is the $(1 - \alpha/2)$ 100% point of the t distribution (D1) with ν degrees of freedom. If the errors in the dependent variable [e.g., left-hand side of Eq. (32)] are normally and independently distributed with constant variance σ^2, then $v(b_i)$ is the ith diagonal element of

$$v(\mathbf{b}) = (\mathbf{X}^1\mathbf{X})^{-1}\sigma^2 \tag{53}$$

If, on the other hand, weighted linear least squares is required, Eq. (53) becomes

$$v(\mathbf{b}) = (\mathbf{X}^1\mathbf{C}^{-1}\mathbf{X})^{-1}\mathbf{X}^1\mathbf{C}^{-1}\sigma^2 \tag{54}$$

The use of these equations allow the calculation of confidence intervals to assess, for example, the significance of negative parameter estimates or the additional experimentation required to estimate the parameters with a particular precision.

As has been described, the correlation of the parameter estimates is also of great interest. This correlation is measured by the off-diagonal terms of $v(\mathbf{b})$, completely ignored by the above equations. If an analysis of variance is conducted, one finds that a joint $(1 - \alpha)$ 100% confidence region is defined by

$$(\boldsymbol{\beta} - \mathbf{b})^1\mathbf{X}^1\mathbf{X}(\boldsymbol{\beta} - \mathbf{b}) = s^2 pF_\alpha(p, \nu) \tag{55}$$

Where $\boldsymbol{\beta}$ is a vector of parameter values being estimated by $\mathbf{b}$, s^2 is an independent estimate of the experimental error variance σ^2, p is the number of

parameters, and $F_\alpha(p, \nu)$ is the $(1 - \alpha)$ 100% point of the F distribution. Alternatively,

$$S_c = S_{\min} + s^2 p F_\alpha(p, \nu) \tag{56}$$

where S_c is the value of the sum of squares on the 100 $(1 - \alpha)$% confidence contour, and $S_{\min}$ is the minimum value of Eq. (19). As is described in Section IV, s^2 can be estimated from replicated data or, if the fitted model is adequate, from

$$s^2 = S_{\min}/(N - p) \tag{57}$$

For two parameters in the model, Eq. (55) simplifies to

$$(\beta_1 - b_1)^2 \sum x_1^2 + 2(\beta_1 - b_1)(\beta_2 - b_2) \sum x_1 x_2 + (\beta_2 - b_2)^2 \sum x_2^2 = s^2 p F_\alpha(p, \nu) \tag{58}$$

where the summations range over N observations. For fixed-data x_i, this equation describes an ellipse in the parameter space. Thus, confidence regions are elliptical (or more generally, ellipsoidal) for linear models.

The covariance (off-diagonal) terms of Eq. (53) are also needed to calculate approximate intervals for parameters in the hyperbolic models. For a model such as

$$\frac{1}{r} = \frac{1}{k} + \frac{K_c}{k} p_c + \frac{K_D}{k} p_D \tag{59}$$

Eq. (53) provides the variance–covariance matrix for the parameter ratios. Letting b_1, b_2, and b_3 denote the respective ratios of Eq. (59), then Eq. (37) may be applied to find

$$v(k) = k^4 v(b_1)$$

where $v(b_1)$ is available from the usual linear least squares Eq. (53). Using a generalized form of Eq. (37), we further obtain

$$v(K_c) = \left(\frac{K_c}{b_2}\right)^2 v(b_2) + \left(\frac{K_c}{b_1}\right)^2 v(b_1) + \left(\frac{K_c}{b_1}\right)\left(\frac{K_c}{b_2}\right) \operatorname{cov}(b_1, b_2) \tag{60}$$

$$v(K_D) = \left(\frac{K_D}{b_3}\right)^2 v(b_3) + \left(\frac{K_D}{b_1}\right)^2 v(b_1) + \left(\frac{K_D}{b_3}\right)\left(\frac{K_D}{b_1}\right) \operatorname{cov}(b_1, b_3) \tag{61}$$

Confidence intervals thus calculated for the rate and adsorption constants have been reported (K7).

2. *Nonlinear Models*

The general approach used with nonlinear models, such as Eq. (40) is to linearize by a Taylor expansion [Eq. (41)] and apply the linear theory of Section III,C,1.

The more approximate method of measuring the precision of parameter

estimates is to use confidence intervals, as defined by Eqs. (52) and (53). Here $\mathbf{X}$ is again given by Eq. (43). This estimate will be erroneous because the Taylor expansion does not perfectly describe the nonlinear model throughout the confidence intervals. Hence, although the intervals are calculated assuming the contours of the sums-of-squares surface are ellipsoidal, they deviate from this shape. An example comparing such contours is shown in Fig. 12, for estimates of the two rate constants in the sequence $A \rightarrow B \rightarrow C$ (B14). In fact, the deviation of the actual sums of squares from an ellipsoidal shape can be used as a measure of the nonlinearity of the model (B3, B8), using latent roots of the $(\mathbf{X}^1\mathbf{X})$ matrix or other indicators (G1, G2, M7).

A better measure of the precision of the parameter estimates for nonlinear models is provided by Eq. (56), which can take into account deviations of the sums-of-squares contours from the ellipsoidal shape. In fact, this equation does provide a rigorous confidence contour for nonlinear models; it is only the exact confidence level, α, associated with this contour that is approximated.

Numerous applications of this theory have been made in calculating confidence intervals for parameter estimates in nonlinear kinetic models, such as typified in Table III (P2). The use of confidence regions is typified in Fig. 13 (M7) for the alcohol dehydration model

$$r = \frac{kK_A(p_A^2 - p_E p_W/K)}{(1 + K_A p_A + K_W p_W)^2} \tag{62}$$

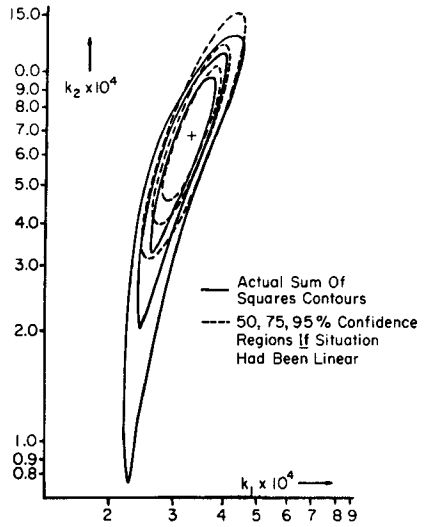

FIG. 12. Contours of actual sum of squares surface and of corresponding linear theory surface, $A \xrightarrow{k_1} B \xrightarrow{k_2} C$.

and Fig. 32 for the nitric oxide reduction model of Section VII. Note the large volume of these regions and the long skewed relationship to the axes (high correlation of estimates).

D. Multiple-Response Parameter Estimation

Parameter estimation through Eq. (19) has been widely used for experimental situations in which only a single response is being measured. Frequently, however, measurements can be made on two or more responses

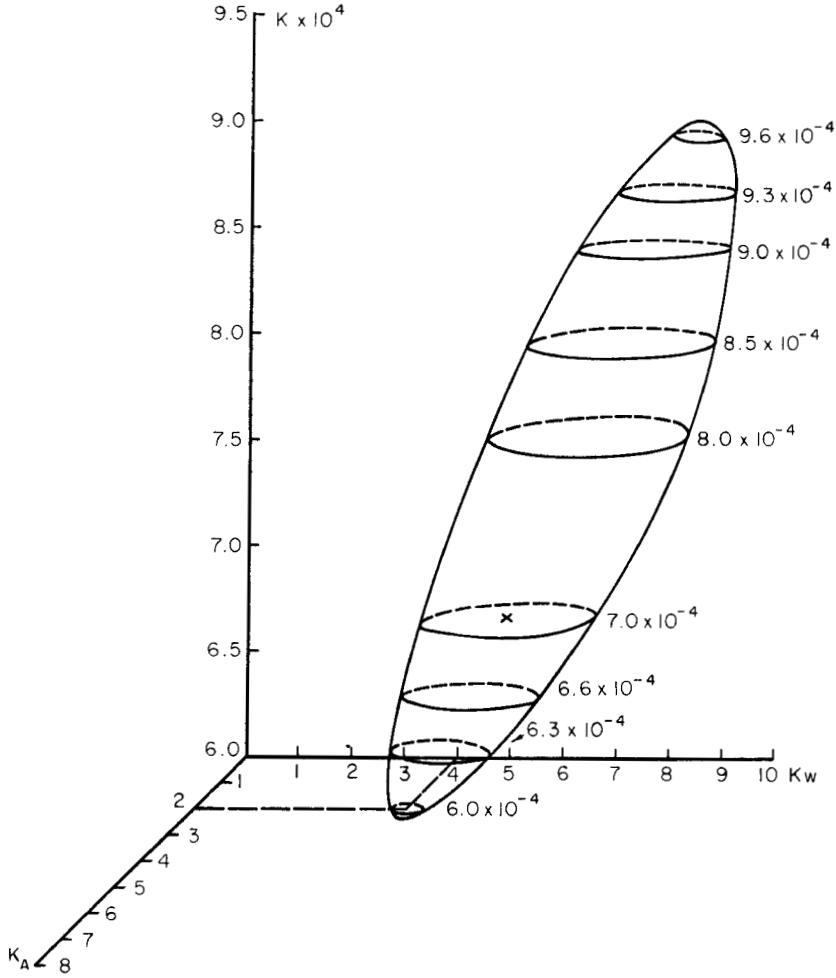

FIG. 13. Approximate 95% confidence region for alcohol dehydration, Eq. (62).

simultaneously. For example, Eq. (19) could be used if only the concentration of B could be measured in the reaction sequence

$$A \rightarrow B \rightarrow C \tag{63}$$

If independent data were available for all three responses, however, all of these data should be used in estimating the two rate constants. One obvious extension of Eq. (19) to this case is to choose parameter estimates minimizing

$$S_r = \sum_{i=1}^{N} (C_{Ai} - \hat{C}_{Ai})^2 + \sum_{i=1}^{N} (C_{Bi} - \hat{C}_{Bi})^2 + \sum_{i=1}^{N} (C_{Ci} - \hat{C}_{Ci})^2 \tag{64}$$

This has been applied previously (B1).

This criterion, and others, can be derived using maximum likelihood arguments (H8). It has been shown that Eq. (64) is applicable (a) when each of the responses has normally distributed error; (b) when the data on each response are equally precise, and (c) when there is no correlation between the measurements of the three responses. These assumptions are rather restrictive.

A much more general criterion has been developed (B12, H8) for which these assumptions have been relaxed. The criterion developed when fitting a model, r, to observations, y, and for v responses is to select parameter values minimizing the determinant

$$S_\Delta = \begin{vmatrix} \sum (y_u^1 - r_u^1)^2 & \sum (y_u^2 - r_u^2)(y_u^1 - r_u^1) & \cdots & \sum (y_u^v - r_v^v)(y_u^1 - r_u^1) \\ \sum (y_u^1 - r_u^1)(y_u^2 - r_u^2) & \sum (y_u^2 - r_u^2)^2 & \cdots & \sum (y_u^v - r_u^v)(y_u^2 - r_u^2) \\ \vdots & \vdots & & \vdots \\ \sum (y_u^1 - r_u^1)(y_u^v - r_u^v) & \sum (y_u^2 - r_u^2)(y_u^v - r_u^v) & \cdots & \sum (y_u^v - r_u^v)^2 \end{vmatrix} \tag{65}$$

where all sums range from $u = 1$ to $u = N$ observations. For the three-response example above, $v = 3$, and the parameters would best be estimated by minimizing the 3×3 determinant of Eq. (65).

The advantages to be gained by following such a procedure are quite striking. The great improvement in the size of the confidence region for the two rate constants (reparameterized logarithmically) of Eq. (63) is shown in Fig. 14 (B12), with the measured responses reproduced thereon. Such precision of parameter estimates not only greatly improves the efficiency of the estimation procedure, but it can also assist in providing hard-to-obtain agreement between thermodynamic and kinetic estimations of kinetic parameters (M4). An application of the procedure is discussed in detail elsewhere (M4).

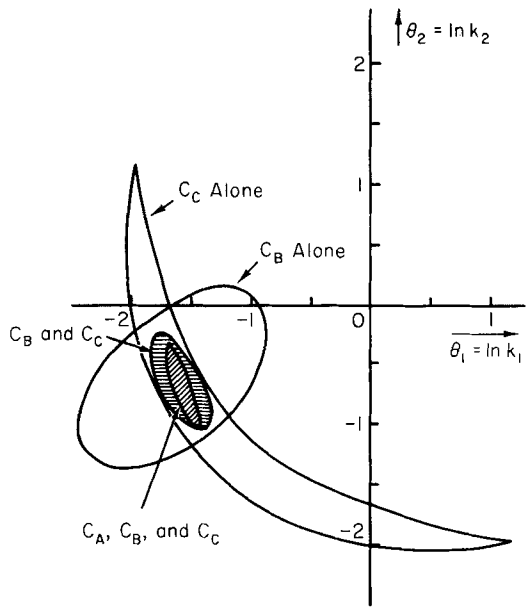

FIG. 14. Approximate 99.75% confidence regions for $A \xrightarrow{k_1} B \xrightarrow{k_2} C$.

IV. Tests of Model Adequacy

The several modeling methods discussed in the accompanying sections are quite useful in testing the ability of a model to fit a particular set of data. These methods do not, however, supplant the more conventional tests of model adequacy of classical statistical theory, i.e., the analysis of variance and tests of residuals.

A. Analysis of Variance

The analysis of variance is used to compare the amount of variability of the differences of predicted and experimental rates with the amount of variability in the data itself. By such comparisons, the experimenter is able to determine (a) whether the overall model is adequate and (b) whether every portion of the model under consideration is necessary.

1. *Significance of the Overall Regression*

For every set of reaction-rate data, a total amount of variability in the data may be expressed as

$$\sum_{i=1}^{N} y_i^2$$

Also, by a least-squares analysis or some other suitable means, estimates of the parameters within a proposed model may be obtained. This allows the calculation of predicted reaction rates at each experimental point and thus an assessment of the total amount of variability which can be accounted for by the proposed model

$$\sum_{i=1}^{N} r_i^2$$

The difference in the predicted and observed rates $(y_i - r_i)$, is termed a residual and is a measure of the inability of the model to describe exactly the experimental data. If the model is entirely correct, in fact, the residual will be a measure of experimental error. A measure of the total amount of variation unaccounted for by the proposed model, then, is

$$\sum_{i=1}^{N} (y_i - r_i)^2$$

In statistical terminology, these quantities are called the crude sum of squares,

$$\sum_{i=1}^{N} y_i^2$$

the sum of squares due to regression,

$$\sum_{i=1}^{N} r_i^2$$

and the residual sum of squares,

$$\sum_{i=1}^{N} (y_i - r_i)^2$$

It is a direct result of the orthogonality property of linear least squares (Section III,A,1) that the crude sum of squares may be broken into two parts as follows:

$$\sum_{i=1}^{N} y_i^2 = \sum_{i=1}^{N} r_i^2 + \sum_{i=1}^{N} (y_i - r_i)^2 \tag{66}$$

Associated with each data point is a certain degree of freedom, which will be used to attribute more information to, say, 500 data points than to 5 data points. In particular, if N data points are used, the total sum of squares is said to possess N degrees of freedom. The predicted rates estimated from a model containing p parameters have p degrees of freedom, and the remaining $N - p$ degrees of freedom are possessed by the residual sum of squares.

If several data points have been taken at the same settings of the independent variables, i.e., replicated data are available, then a knowledge of the

inherent amount of error in the data is measured by the pure-error sum of squares,

$$\sum_{i=1}^{N} (y_i - \bar{y})^2.$$

Here, $\bar{y}$ is the average of all of the replicated data points. If the residual sum of squares is the amount of variation in the data as seen by the model, and the pure-error of squares is the true measure of error in the data, then the inability of the model to fit the data is given by the difference between these two quantities. That is, the lack-of-fit sum of squares is given by

$$\sum_{i=1}^{N} (\bar{y} - r_i)^2 = \sum_{i=1}^{N} (y_i - r_i)^2 - \sum_{i=1}^{N} (y_i - \bar{y})^2 \tag{67}$$

If there are $\tilde{n}$ replications at q different settings of the independent variables, then the pure-error sum of squares is said to possess $(\tilde{n} - 1)$ degrees of freedom (1 degree of freedom being used to estimate $\bar{y}$); while the lack-of-fit sum of squares is said to possess $N - p - q(\tilde{n} - 1)$ degrees of freedom, i.e., the difference between the degrees of freedom of the residual sum of squares and the pure-error sum of squares.

The sums of squares of the individual items discussed above divided by its degrees of freedom are termed mean squares. Regardless of the validity of the model, a pure-error mean square is a measure of the experimental error variance. A test of whether a model is grossly adequate, then, can be made by acertaining the ratio of the lack-of-fit mean square to the pure-error mean square; if this ratio is very large, it suggests that the model inadequately fits the data. Since an F statistic is defined as the ratio of sum of squares of *independent normal* deviates, the test of inadequacy can frequently be stated

$$\frac{\text{lack-of-fit mean square}}{\text{pure-error mean square}} > F_\alpha[N - p - q(\tilde{n} - 1), q(\tilde{n} - 1)] \tag{68}$$

The F statistic is tabulated in many reference texts (D1). More rigorous discussions of the analysis of variance are, of course, available (D4).

One important application of analysis of variance is in the fitting of empirical models to reaction-rate data (cf. Section VI). For the model below, the analysis of variance for data on the vapor-phase isomerization of normal to isopentane over a supported metal catalyst (C1)

$$r = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{11} x_1^2 + b_{22} x_2^2 + b_{33} x_3^2 + b_{12} x_1 x_2 + b_{13} x_1 x_3 + b_{23} x_2 x_3 \tag{69}$$

is reported in Table IV. Here, x_1 is the hydrogen partial pressure, x_2 the normal pentane partial pressure, and x_3 the isopentane partial pressure. It is evident from the F ratios of Table IV that the model adequately fits these data.

TABLE IV

ANALYSIS OF VARIANCE FOR EQ. (69)

Source	Sum of squares	Degrees of freedom	Mean squares	Ratios	F statistics 95%	97.5%
Crude	640.49	24	—	—	—	—
Due to regression	637.35	10	63.74	—	—	—
Residual	3.14	14	0.225	—	—	—
Lack of fit	2.85	11	0.260	2.73	8.8	14.3
Pure error	0.29	3	0.0954			

An analysis of variance can also be used to test the adequacy of more theoretical models. For example, two models considered in Section III for pentane isomerization are the single-site and dual-site models of Eqs. (30) and (31). These were linearized to provide Eqs. (32) and (33). The overall fit of these equations to the data may now be judged by an analysis of variance, reported in Tables V and VI (K3). It is seen that Eq. (33) fits the data quite

TABLE V

ANALYSIS OF VARIANCE FOR EQ. (32)

Source	Sum of squares	Degrees of freedom	Mean squares	Ratios	F statistics 95%	97.5%
Crude	28894.02	24	—	—	—	—
Due to regression	26896.82	4	6,724.2	—	—	—
Residual	1997.20	20	99.86	—	—	—
Lack of fit	1964.56	17	115.56	10.62	8.7	14.3
Pure error	32.65	3	10.88			

TABLE VI

ANALYSIS OF VARIANCE FOR EQ. (33)

Source	Sum of squares	Degrees of freedom	Mean squares	Ratios	F statistics 95%	97.5%
Crude	763.54	24	—	—	—	—
Due to regression	754.29	4	188.6	—	—	—
Residual	9.25	20	0.462	—	—	—
Lack of fit	8.98	17	0.528	5.9	8.7	14.3
Pure error	0.27	3	0.090			

well. However, there is some question as to whether Eq. (32) adequately represents the data, since the ratio of lack-of-fit to pure-error mean square is relatively large (although not so large as to allow a clear rejection of the model, particularly since the y may not be independent normal deviates).

In some cases when estimates of the pure-error mean square are unavailable owing to lack of replicated data, more approximate methods of testing lack of fit may be used. Here, quadratic terms would be added to the models of Eqs. (32) and (33), the complete model would be fitted to the data, and a residual mean square calculated. Assuming this quadratic model will adequately fit the data (lack of fit unimportant), this quadratic residual mean square may be used in Eq. (68) in place of the pure-error mean square. The lack-of-fit mean square in this equation would be the difference between the linear residual mean square [i.e., using Eqs. (32) and (33)] and the quadratic residual mean square. A model should be rejected only if the ratio is very much greater than the F statistic, however, since these two mean squares are no longer independent.

2. *Significance of Terms in the Regression*

Although the overall adequacy of a model to fit the data may be tested as described above, it is frequently desirable to test whether all of the terms in an adequate model need be included. This may be accomplished by a straightforward extension of Section IV,A,1, involving a decomposition of the regression sum of squares into the desired components. Suppose, for example, that we wished to test the necessity of including the quadratic and interaction terms ($b_{ij}x_i x_j$) of Eq. (69). We do this by calculating the crude sum of squares and the regression sum of squares using the fitted linear model

$$r = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 \tag{70}$$

This is, then, the regression sum of squares due to the first-order terms of Eq. (69). Then, we calculate the regression sum of squares using the complete second-order model of Eq. (69). The difference between these two sums of squares is the extra regression sum of squares due to the second-order terms. The residual sum of squares is calculated as before using the second-order model of Eq. (69); the lack-of-fit and pure-error sums of squares are thus the same as in Table IV. The ratio contained in Eq. (68) still tests the adequacy of Eq. (69). Since the ratio of lack-of-fit to pure-error mean squares in Table VII is smaller than the F statistic, there is no evidence of lack of fit; hence, the residual mean square can be considered to be an estimate of the experimental error variance. The ratio

$$\frac{\text{extra mean square due to quadratic terms}}{\text{residual mean square}} > F_\alpha(p, N-p) \tag{71}$$

TABLE VII

COMPLETE ANALYSIS OF VARIANCE FOR EQ. (69)

Source	Sum of squares	Degrees of freedom	Mean squares	Ratios	F statistics 95%	97.5%
Crude	640.49	24	—	—	—	—
Due to first order regression	627.0	4	156.8	—	—	—
Extra due to second order regression	10.34	6	1.72	—	—	—
Residual	3.14	14	0.225	7.66	2.9	3.5
Lack of fit	2.85	11	0.260	—	—	—
				2.73	8.8	14.3
Pure error	0.29	3	0.0954	—	—	—

then tests the necessity of the retaining the six quadratic terms in Eq. (69). Here, the ratio of quadratic to residual mean squares is so large that the quadratic terms must be used for an adequate fit of the rate data. This procedure may be carried out for any number of terms as long as we begin from the bottom of the table as was done here.

There are many cases of interest to kineticists in which the necessity for the inclusion of terms in nonlinear rate models must be tested. This may be done approximately by again calculating the regression sum of squares with and without the added term. Kabel and Johanson (K1), for example, considered the model

$$r = \frac{kK_A L(P_A^2 - P_E P_W/K)}{(1 + K_A P_A + K_E P_E + K_W P_W)^2} \tag{72}$$

in which they suspected K_E was zero. This could be tested by comparing the residual mean squares to the pure-error mean square, with and without the inclusion of the K_E term. Table VIII summarizes the information thus

TABLE VIII

APPROXIMATE ANALYSIS OF VARIANCE FOR EQ. (72)

Situation	kL	K_A, atm^{-1}	K_W, atm^{-1}	K_E, atm^{-1}	Residual mean square
With K_E	7.02×10^{-4}	2.91	6.34	8.7×10^{-3}	7.0×10^{-11}
Without K_E	7.04×10^{-4}	2.89	6.30	—	6.8×10^{-11}

obtained. From the closeness of the residual mean square, there is no reason to suspect that the K_E term is needed. In this case, the parameter estimates

foresee this situation, since the estimated value of K_E is near zero *and* the parameter estimates obtained with the K_E term excluded are near those with the K_E term included. If similar results were reported with the K_A term omitted, the residual mean square without K_A would be greatly inflated over that including the K_A term.

B. Residual Analysis

The analysis of variance techniques of Section IV,A have been seen to provide information about the overall goodness of fit or about testing the importance of the contribution of certain terms in the model toward providing this overall fit of the data. Although these procedures are quite useful, more subtle model inadequacies can exist, even though the overall goodness of fit is quite acceptable. These inadequacies can often be detected through an analysis of the residuals of the model.

A residual is defined as the difference between the observed and predicted values of some response of interest, such as the reaction rate. For example, suppose that an experimentally observed reaction rate y is linearly dependent on two partial pressures x_1 and x_2

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon \tag{73}$$

and that the predicted rate r is obtained using the *correct* model,

$$r = b_0 + b_1 x_1 + b_2 x_2 \tag{74}$$

If the residuals are estimated using some unbiased method such as linear least squares, the b's would be expected to equal the β's, so that the residual becomes

$$y - r = \varepsilon \tag{75}$$

Since the difference between the predicted rate and the observed rate is attributable solely to experimental error, plots of this residual versus any independent variable should exhibit all of the characteristics of this error, such as being random with zero mean. If, on the other hand, the model of Eq. (74) is wrong in that the $b_2 x_2$ term is omitted, the residual should not be random when plotted versus either r or x_2. Numerous suggestions have been made concerning the types of residual plots that are most revealing (A1, A2, B8, D4). These include plotting the residuals in overall plots, against predicted values r, against independent variables x_i, and against time.

1. *Overall Plots*

The overall plots consist of plotting the frequency of occurrence of the rounded values of the residual against the magnitude of the residual. These

plots allow an approximate assessment of the normality of the error disbribution provided by $y - r$ if the model is correct. The plots are also a check of whether the mean of this error distribution is zero, as should be fufilled for linear and nonlinear least-squares applications. Hence, these plots allow an approximate check on some of the least-squares assumptions after the least-squares analysis has been completed. It should be noted that the mean of the residuals *must* be zero if the linear-least-squares calculations have been correctly carried out for linear models containing a constant [e.g., b_0 in Eq. (69)], so only a check on the algebra of linear least squares is obtained here.

2. *Predicted Value Plots*

A plot of the residual versus the predicted value, r, of a model can indicate whether the model truly represents the rate data. For example, residuals that are generally positive at low predicted rates and negative at high predicted rates can indicate a model inadequacy, even though the overall test of an analysis of variance indicates that the model is acceptable.

The analysis of variance for the model of Eq. (32), for example, for the data on the isomerization of normal pentane was shown in Table V; we concluded that the model was marginally acceptable. However, the plot of the residuals of Fig. 15 indicates that this overall fit is achieved by balancing predictions that are too low against predictions that are too high. Hence the

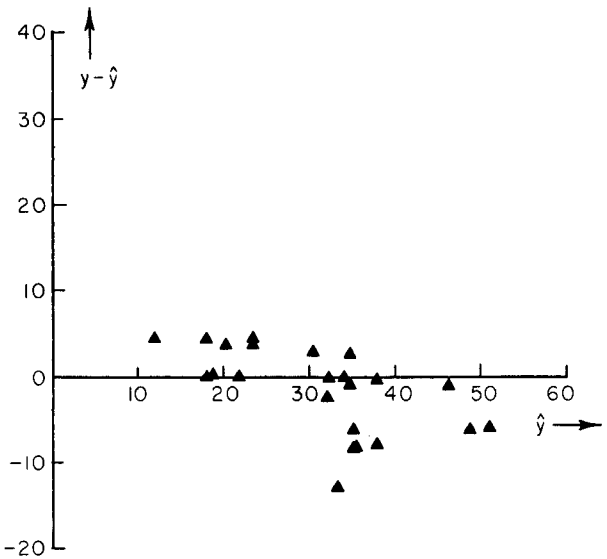

FIG. 15. Residual plot versus predicted response for Eq. (32).

model should be rejected.

These plots can also provide information about the assumption of constant error variance (Section III) made in the unweighted linear or nonlinear least-squares analyses. If the residuals continually increase or continually decrease in such plots, a nonconstant error variance would be evident. Here, either a weighted least-squares analysis should be conducted (Section III,A,2) or a transformation should be found to stabilize the error variance (Section VI).

Figure 16 presents residuals of 72 data points at three temperatures for the hydrogenation of propylene, reported by Shabaker (S1); the model was fitted by us to these data using unweighted nonlinear least squares with Eq. (8). The nonconstant error variance shown here indicates that a weighted nonlinear least-squares analysis would have been more appropriate. The residuals trends were largely removed by Shabaker by fitting the logarithm of the reaction rate rather than the rate itself (a transformation of coordinates such as discussed in Section VI).

3. *Independent Variable Plots*

The plot of the residuals versus the independent variables can yield information as to which of the variables in the equation is causing the residual

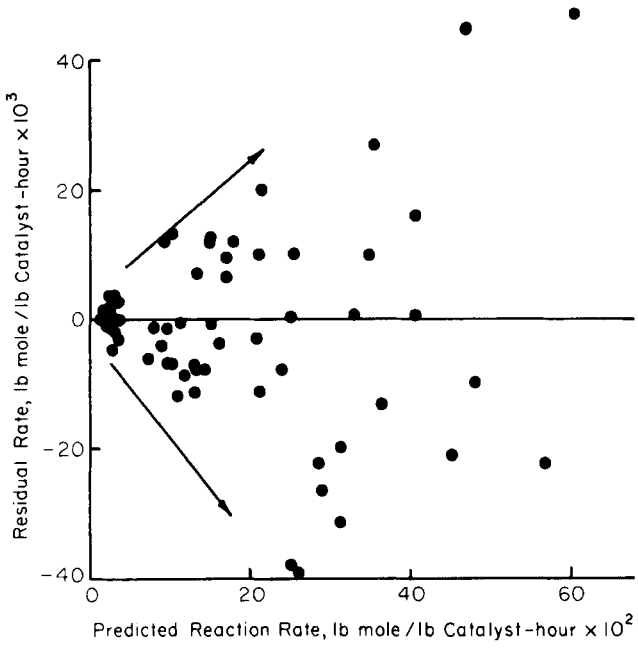

FIG. 16. Plot of residual rates versus predicted rate for Eq. (8).

trends that occur in the predicted value plots. The nonconstant error variance described in the above section also is exhibited in these plots, and can provide information useful in developing a weighting function.

Of particular value in kinetic studies are residual plots using the linearized form of the Hougen–Watson equation. For the model of Eq. (18), for example, we obtain

$$\frac{p_{NO}\, p_{H_2}}{r} = \frac{1}{kK_{NO}} + \frac{1}{k}\, p_{NO} + \frac{K_{H_2}}{kK_{NO}}\, p_{H_2} \tag{76}$$

Reasoning analogous to that of Eqs. (73)–(75) indicates that if this model is wrong only in that adsorbed water should have been included, a plot of residuals versus water partial pressure should be linear. Or, if this model is erroneous only in that the hydrogen should have been considered to be dissociated, the residuals should be independent of the nitric oxide partial pressure *and* nonlinearly dependent upon the hydrogen partial pressure. It should be noted that these observations are strictly valid only if the data have been taken according to an orthogonal (e.g., factorial) design. This subject is treated more comprehensively in Section V; examples of these effects are included there.

4. *Time-Residual Plot*

The plot of residuals versus some measure of the time at which experiments were run can also be informative. If the number of hours on stream or the cumulative volume of feed passed through the reactor is used, nonrandom residuals could indicate improper treatment of catalyst-activity decay. In the same fashion that residuals can indicate variables not taken into account in predicting reaction rates, variables not taken into account as affecting activity decay can thus be ascertained.

5. *Parametric Residuals*

The residuals discussed thus far have been associated with some dependent variable, such as the reaction rate r. It is particularly advantageous in pinpointing the type of defect present in an inadequate model to expand this definition to include parametric residuals. The parametric residual, then, is simply the difference between a value of a given parameter estimated from the data and that predicted from a model. For example, the dots in Fig. 17 represent the logarithm of the alcohol adsorption constants measured in alcohol dehydrogenation experiments from isothermal data at each of several temperature levels (F1). The solid line represents the expectation that these

points can be linearly dependent upon reciprocal temperature. If they do not, a defect exists in the model.

The plot of the residuals of this adsorption constant, using the data of Fig. 17, is shown in Fig. 18. A substantial trending effect is evident (it is also evident from Fig. 17). This is of interest in itself, but it is especially desirable to utilize the information of Fig. 18 to learn how the model must be modified to remove the observed defect. The other residual plots of the section can also assist this objective. This topic, called model-building, is treated in Section V.

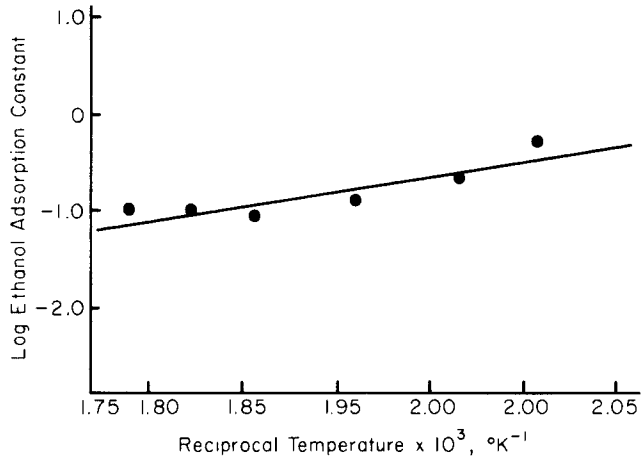

FIG. 17. Dependence of ethanol adsorption constant on temperature—alcohol dehydrogenation.

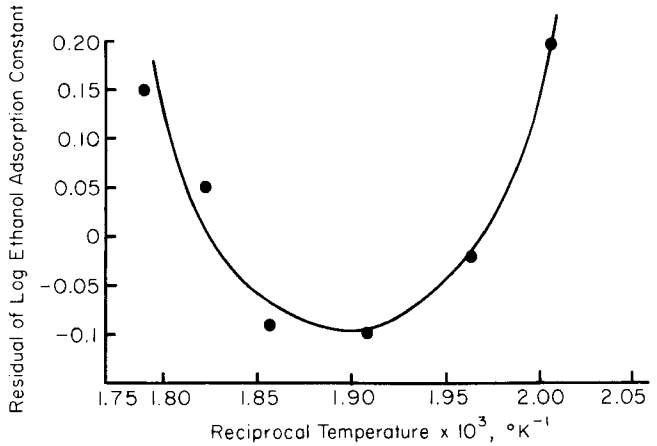

FIG. 18. Dependence of residuals of ethanol adsorption constant on temperature—alcohol dehydrogenation.

V. Use of Diagnostic Parameters

Recently certain diagnostic parameters have been exploited to allow a discrimination among several rival models. These diagnostic parameters can be grouped into two broad classes—those that are inherently present in the model, and those that are introduced solely for the purpose of model discrimination.

There are two primary advantages to the use of diagnostic parameters in reaction-rate modeling. First, the use of these parameters allows an easy analysis of the adequacy of the model. This is accomplished by choosing the diagnostic parameters such that a linear test of a nonlinear model may be obtained. Second, in some cases, the diagnostic analysis will not only indicate a model inadequacy but also can suggest the precise nature of the inadequacy. Then, the model can be appropriately changed to remove this inadequacy.

A. Model Discrimination with Diagnostic Parameters

1. *Nonintrinsic Parameters*

A nonintrinsic parameter is one that is introduced into a model for the purpose of allowing the specification of a preferred model from a larger group of rival models. To be useful, such a parameter must simplify the analysis procedures, allow a broadening of the data base of the analysis, or both. One type of nonintrinsic parameter useful in this regard enters as a multiplier of each of a series of predicted responses (C3, C4). For two models, such a method reduces (W4, W5) to defining a dependent variable

$$z = y - 1/2(r_1 + r_2) \tag{77}$$

This variable is then to be fitted using the equation

$$\hat{z} = \Lambda(r_2 - r_1) \tag{78}$$

where Λ is the nonintrinsic parameter used in specifying which of models r_1 and r_2 is to be preferred. It is to be noted that if

$$y = r_1 + \varepsilon \tag{79}$$

then Λ should be $-1/2$, while if r_2 were adequate Λ should be $+1/2$.

The procedure to be followed, then, is to estimate the parameters $\mathbf{K}$ within each reaction-rate model by some appropriate technique (K8). The intrinsic parameter Λ can then be estimated by linear least squares. Owing to experimental error in the data, this estimate of Λ will typically be neither plus nor minus one-half. Hence the remaining portion of the analysis is to estimate the

confidence interval for this discriminatory parameter estimate. If this confidence interval contains both possible values, this would suggest that no decision could be made about the most adequate model; if neither value is contained, both models should be viewed with suspicion.

In the above analysis, y was considered to be a reaction rate. Clearly, any dependent variable can be used. Note, however, that if the dependent variable, y, is distributed with constant error variance, then the function z will also have constant error variance and the unweighted linear least-squares analysis is rigorous. If, in addition, y has error that is normal and independent, the least-squares analysis would provide a maximum likelihood estimate of Λ. On the other hand, if any transformation of the reaction rate is felt to fulfill more nearly these characteristics, the transformation may be made on y, r_1, r_2 and the same analysis may be applied. One common transformation will be logarithmic.

This method has been applied (M5) for modeling the vapor-phase rate of dehydration of secondary butyl alcohol to the olefin over a commercial silica-alumina cracking catalyst. Integral reactor data are available at 400, 450, and 500°F. Two models considered for describing this reaction are the single site

$$r_1 = \frac{kK_A p_A}{1 + K_A p_A + K_W p_W} \tag{80}$$

and the dual site

$$r_2 = \frac{kK_A p_A}{(1 + K_A p_A + K_{W0} p_W)^2} \tag{81}$$

The usual plots of the initial reaction rate for 500°F are shown in Fig. 19. In

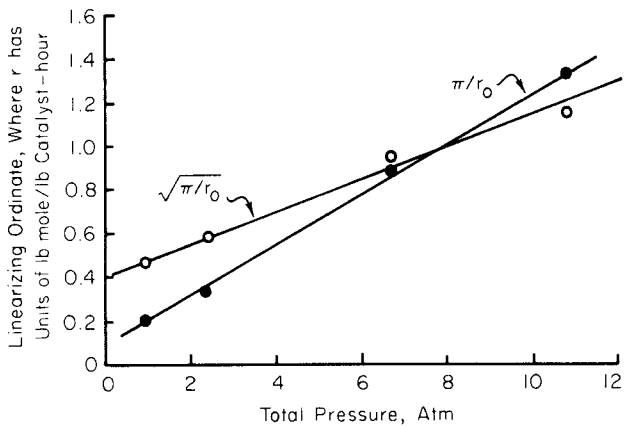

FIG. 19. Linearized initial rate plots for sec-butanol dehydration, Eqs. (80) and (81).

spite of this apparently good fit of the initial rate data by both models, the analysis of the intrinsic parameter Λ will indicate that the single-site model cannot accurately represent the higher conversion data. In this case, the parameter Λ is defined.

$$x - 1/2(\hat{x}_1 + \hat{x}_2) = \Lambda(\hat{x}_2 - \hat{x}_1) \tag{82}$$

where x is the experimentally measured conversion, and $\hat{x}_1$ and $\hat{x}_2$ are the conversions predicted by the integrated forms of Eqs. (80) and (81). The parameters within these latter equations were estimated by nonlinear least squares (M5).

The values of Λ thus obtained are reported in Table IX. It is clear that the

TABLE IX

ESTIMATES OF DISCRIMINATORY PARAMETER, Λ

Temperature °F	Parameter estimate, Λ[a]
400	0.36 ± 0.28
450	0.47 ± 0.23
500	0.46 ± 0.14

[a] The numbers following the $\pm$ signs define the 95% confidence interval on Λ, based upon residual mean squares.

preferred value of Λ is plus one-half, and hence that the data do not reject the dual model r_2. However the probability that a Λ of minus one-half could represent these data is less than 0.0001%. Thus, the single-site model should be rejected. The parameter estimates for the dual-site model are reported elsewhere (M5); they approximately exhibit an exponential temperature dependence, with slopes compatible with the usual predictions of kinetic theory.

2. *Intrinsic Parameters*

An intrinsic parameter is one that is inherently present in or arises naturally from a reaction-rate model. These parameters, which are of a simpler functional form than the entire rate model, facilitate the experimenter's ability to test the adequacy of a proposed model. Using these intrinsic parameters, this section presents a method of preparing linear plots for high conversion data, which is entirely analogous to the method of the initial-rate plots discussed in Section II. Hence, these plots provide a visual indication of the ability of a model to fit the high conversion data and thus allow a more

complete test of a model than does the initial-rate analysis alone. When used in conjunction with the linear initial-rate correlations, the high conversion plots provide estimates of all the parameters in models that exhibit only one adsorbed reactant and one adsorbed product.

When any hyperbolic model is written in terms of fractional conversions instead of partial pressures, two groupings of terms inherently arise within the denominator These two groupings will be called the intrinsic parameters C_1 and C_2. For example, when data are taken for the olefinic dehydration of a pure alcohol feed to a reactor, Eq. (80) becomes

$$r = \frac{kK_A[(1-x)/(1+x)]\pi}{1 + K_A[(1-x)/(1+x)]\pi + K_W[x/(1+x)]\pi} \tag{83}$$

In this formulation, the reaction is assumed to be essentially irreversible; the existence of product partial pressures in the numerator does not alter the method of analysis to be discussed here, however. Equation (83) may be written

$$r = \frac{(1-x)\pi}{(\hat{C}_1 + \hat{C}_2 x)} \tag{84}$$

$$\hat{C}_1 = \frac{1}{kK_A} + \frac{\pi}{k} \tag{85}$$

$$\hat{C}_2 = \frac{1}{kK_A} + \frac{(K_W - K_A)}{kK_A}\pi \tag{86}$$

The $\hat{C}_1$ term, then, is the collection of terms not multiplied by conversion while $\hat{C}_2$ is the collection multiplied by the conversion, x.

The use of the parameter C_1 in Eq. (84) was the subject of a portion of Section II, for Eq. (84) can be written at zero conversion in terms of the initial reaction rate:

$$C_1 = \frac{\pi}{r_0} \tag{87}$$

and Eq. (85) suggests that this variable should be linear with π. From Eq. (84), values of C_2 can be estimated by

$$C_2 = \frac{(1-x)\pi}{xr} - \frac{\pi}{xr_0} \tag{88}$$

where C_1 has been written as shown in Eq. (87). Here, the several values of C_2 (corresponding to several conversion points) at each pressure level can be averaged to give a single value of C_2 at each pressure. Equation (86) requires that these estimates be linear with pressure.

If this linear analysis is to be used, the experimental conversion–space-time data should first be taken at several pressure levels. Using the C_1 analysis alone, then, the plots of C_1 versus total pressure should be made for a preliminary indication of model adequacy. If several models are found to provide near-linear C_1 plots, the complete linear analysis using the C_2 plots should assist in the discrimination among the remaining rival models. If a model is adequate, both the C_1 and C_2 points should be correlatable by a straight line with a *common* intercept, as demanded by Eqs. (85) and (86). If only one model is found to be adequate following the initial C_1 analysis, the complete C_1 and C_2 analysis should still be carried out on this model to verify its ability to fit the high conversion data.

Consider again the 500°F data for the alcohol dehydration just discussed. Equations (84)–(88) arose from Eq. (80). A similar set of equations can be generated from Eq. (81) (K5). Let us examine the two models through an analysis of their intrinsic parameters.

The C_1 (initial reaction rate) plots were presented in Fig. 19. Note again that the data may be correlated well by straight lines for both models. The C_2 values are correlated by the solid lines of Fig 20. Note that the dual-site values can again be correlated by a straight line, but that the single-site values of C_2 show a definite curvature. Alternatively, the 0.975 atm value of the single-site C_2 could be rejected, and the three high-pressure points

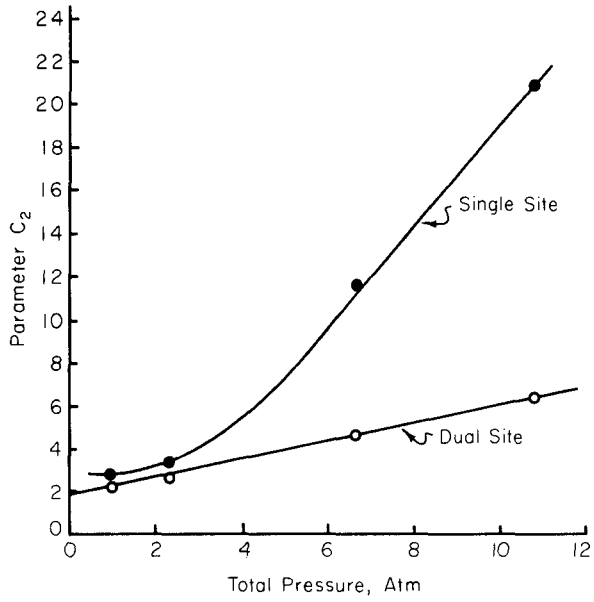

FIG. 20. Plot of parameter C_2 versus pressure for sec-butanol dehydration, Eqs. (80) and (81).

correlated by a straight line. Since the latter procedure results in a negative intercept for C_2 [cf. Eq. (85)], either approach eliminates the single-site model from further consideration. Although the intercept from the dual-site C_2 plot is somewhat higher than that of the dual-site C_1 plot, the 97.5% confidence intervals for these two intercepts overlap.

By using only simple hand calculations, the single-site model has been rejected *and* the dual-site model has been shown to represent adequately both the initial-rate and the high-conversion data. No replicate runs were available to allow a lack-of-fit test. In fact this entire analysis has been conducted using only 18 conversion–space-time points. Additional discussion of the method and parameter estimates for the proposed dual-site model are presented elsewhere (K5). Note that we have obtained the same result as available through the use of nonintrinsic parameters.

B. Adaptive Model-Building with Diagnostic Parameters

The development of an adequate mathematical model representing a physical or chemical system is the object of a considerable effort in research and development activities. A technique has been formalized by Box and Hunter (B14) whereby the functional form of reaction-rate models may be exploited to lead the experimenter to an adequate representation of a given set of kinetic data. The procedure utilizes an analysis of the residuals of a diagnostic parameter to lead to an adequate model with a minimum number of parameters. The procedure is used in the *building* of a model representing the data rather than the postulation of a large number of possible models and the subsequent selection of one of these, as has been considered earlier. That is, the residual analysis of intrinsic parameters, such as C_1 and C_2, will not only indicate the inadequacy of a proposed model (if it exists) but also will indicate *how* the model might be modified to yield a more satisfactory theoretical model.

1. *Hyperbolic Models*

In the following discussion, we shall again separate the terms of a hyperbolic model and identify two parameters C_1 and C_2. As before, each of these two parameters will be a collection of terms, one of which is multiplied by conversion and one not multiplied by conversion. In previous formulations, however, we have oriented the discussion toward a familiar type of experimental design in kinetics: conversion versus space-time data at several pressure levels. Consequently, the parameters C_1 and C_2 were defined to exploit this data feature. Another type of design that is becoming more common is a factorial design in the feed-component partial pressures.

Consequently the definitions of the parameters C_1 and C_2 to be used here will reflect this type of design.

Let us first consider the type of analysis to be used. Suppose that the model, such as Eq. (83), were to predict that $\hat{C}_1$ should be related to the feed-component partial pressures as

$$\hat{C}_1 = b_0 + b_2 p_2 + b_4 p_4 \tag{89}$$

when in fact the true values of C_1 from the data were related as

$$C_1 = \beta_0 + \beta_2 p_2 + \beta_3 p_3 + \beta_4 p_4 + \varepsilon \tag{90}$$

Then, for an orthogonal design and zero expected value of error, the expected value of the residual of C_1 is (H5)

$$E(C_1 - \hat{C}_1) = \beta_3 p_3 \tag{91}$$

Hence, if a correlation of the residual with any independent variable can be detected (e.g., with p_3), then the model generating Eq. (89) is inadequate. Furthermore a modification of the manner in which p_3 enters the model is necessary.

The analysis could easily become more complicated. If, for example,

$$\hat{C}_1 = b_0 + b_1 p_1 + b_3 p_3 \tag{92}$$

then

$$E(C_1 - \hat{C}_1) = -\beta_1 p_1 + \beta_2 p_2 + \beta_4 p_4 \tag{93}$$

In this case, the residual will be correlated with the weighted sum of the three partial pressures. Although a trend may still be detectable, we will not generally be able to determine with which of the variables the residual is correlated and thus will not know how to correct the model; a careful initial model choice circumvents this problem.

Application to Methane Oxidation. This selection of an appropriate initial model can be accomplished as shown here for the complete oxidation of methane. A general representation of the surface reaction model is (K12)

$$r = \frac{k p_1 (1 - x)(p_2 - 2p_1 x)^2}{[1 + K_1 p_1 (1 - x) + K_2 (p_2 - 2p_1 x) + K_3 (p_3 + p_1 x) + K_4 (p_4 + 2p_1 x)]^n} \tag{94}$$

where p_1, p_2, p_3, and p_4 are the feed partial pressures of methane, oxygen, carbon dioxide, and water, respectively, and x is conversion. Collecting terms and using the continuity equation for an integral reactor,

$$\frac{\partial x}{\partial W/F} = \frac{p_1 (1 - x)(p_2 - 2p_1 x)^2}{(C_1 + C_2 x)^n} \tag{95}$$

For this generalized model, then,

$$C_1 = \left[p_1 p_2^2 \left(\frac{\partial W/F}{\partial x} \right)_{x=0} \right]^{1/n} \tag{96}$$

$$C_2 = \frac{p_1 p_2}{n C_1^{n-1}} \left[p_2 \left(\frac{\partial^2 W/F}{\partial x^2} \right)_{x=0} - \frac{C_1^n}{p_1 p_2^2} (p_2 + 4p_1) \right] \tag{97}$$

$$\hat{C}_1 = b_0 + b_1 p_1 + b_2 p_2 + b_3 p_3 + b_4 p_4 \tag{98}$$

$$\hat{C}_2 = \bar{b} p_1 \tag{99}$$

The problem of specifying an adequate model will now be to determine (1) the exponent n by a C_2 analysis and (2) the denominator terms required by a C_1 analysis. Depending upon the particular surface-reaction model considered, the terms within Eq. (98) can change greatly. For any surface-reaction model, however, $\hat{C}_2$ remains the same. Equating Eqs. (97) and (99) provides an equation with two unknown parameters, $\bar{b}$ and n. Estimating n in this way from the reaction data will specify the power of the denominator of Eq. (94). Generally, the selection of any of the models with the appropriate n will eliminate the difficulties represented by Eq. (93) and hence allow an effective C_1 analysis.

For the complete vapor-phase oxidation of methane over a palladium alumina catalyst, conversion–space-time data were taken at 350°C and 1 atm total pressure; the fractional factorial design of Table X (H11) specified the settings of the feed partial pressures of the reacting species.

From these data, n has previously been estimated to be 3 (K12). By starting with models of $n = 3$, the possibility of an ineffective C_1 analysis is diminished.

With $n = 3$, one model with a minimum number of parameters is model 1 of Table XI. By choosing a model with the smaller number of parameters, of course, we tend to prevent including parameters in the models unless their presence is absolutely necessary (principle of parsimonious parameterization). The residuals of the diagnostic parameter, C_1, for this model are shown in Table XII, allowing an analysis such as that suggested earlier. Note that the perfect correlation of the signs of residuals with the water partial pressure of Table X suggests that model 1 should be modified by changing the manner in which the effect of water is taken into account. To maintain $n = 3$, this requires the inclusion of an additional parameter, resulting in model 2 of Table XI. The residuals of model 2 of Table XII, being perfectly correlated with the carbon dioxide level of Table X, suggest that the effect of carbon dioxide is improperly described by model 2.

Keeping the number of parameters to a minimum, we are thus led to model 3 of Table XI. However, the residual trends of model 3 as shown in

TABLE X

EXPERIMENTAL DESIGN

Run No.	$\bar{p}_1$	$\bar{p}_2$	$\bar{p}_3$	$\bar{p}_4$
1	−1	−1	−1	+1
2	+1	−1	−1	−1
3	−1	+1	−1	−1
4	+1	+1	−1	+1
5	−1	−1	+1	−1
6	+1	−1	+1	+1
7	−1	+1	+1	+1
8	+1	+1	+1	−1

$$\bar{p}_1 = \frac{p_1 - 0.015}{0.005} \qquad \bar{p}_2 = \frac{p_2 - 0.120}{0.060}$$

$$\bar{p}_3 = \frac{p_3 - 0.065}{0.035} \qquad \bar{p}_4 = \frac{p_4 - 0.095}{0.055}$$

TABLE XI

MATRIX OF TERMS COMPRISING POSTULATED MODELS[a]

Model No.	Methane p_1	Oxygen p_2	Carbon dioxide p_3	Water, p_4
1	1	1	0	0
2	1	1	0	1
3	1	1	1	0
4	1	0	1	1
5	0	1	1	1

[a] 0 and 1. Gaseous and adsorbed state, respectively.

Table XII would indicate that the effect of water is not described adequately by the model. Utilizing the principle of parsimonious parameterization, one can consider both water and carbon dioxide to be adsorbed and oxygen to be nonadsorbed, resulting in the three-parameter model 4. The residuals in Table XII for model 4, however, are correlated with the oxygen level. Hence model 5 would perhaps be preferable, for it likewise contains only three parameters while allowing adsorbed oxygen. The random residuals of Table XII for model 5 indicate that this model cannot be rejected using the

TABLE XII

RESIDUALS OF PARAMETER C_1 FOR THE POSTULATED MODELS $\times 10^3$

Run No.	Residuals for model 1	2	3	4	5
1	+1.17	−1.38	+2.49	−7.18	−0.29
2	−3.61	−1.07	−2.51	−7.70	+0.50
3	−4.99	−2.21	−3.75	+7.54	−1.08
4	+2.33	−0.54	+3.43	+7.88	+0.93
5	−1.18	+1.81	−2.37	−7.00	+0.24
6	+3.39	+0.55	+2.21	−8.75	−0.51
7	+4.01	+1.42	+2.97	+8.04	−0.08
8	−1.13	+1.42	−2.46	+7.16	+0.29

data presented. Thus, a reaction-rate model adequately describing these experimental data is model 5:

$$r = \frac{k_1 k_2 p_1(1-x)(p_2 - 2p_1 x)^2}{\{1 + K_2(p_2 - 2p_1 x) + K_3(p_3 + p_1 x) + K_4(p_4 + 2p_1 x)\}^3} \tag{100}$$

This model has previously been shown (H11, K12) to have a residual mean square comparing favorably with that expected from pure error, as discussed in Section IV. It is to be noted that we have been led logically from one model to another within the small class of models for which $n = 3$ by the above analysis. For these data, adsorbed methane is not required; however, for data with higher methane concentrations, the adsorbed-methane term may be needed.

The foregoing development represents a substantial deviation from, and in some cases a distinct improvement on, the more usual methods of obtaining a reaction model adequately describing a set of data. In particular, for the hyperbolic models, we have seen that it may be necessary to write a large number of possible models and to search extensively through these to find an adequate model. Common criteria for adequacy include the requirements that the rate data plot linearly with the proper choice of the ordinate, and that the parameter estimates be acceptable. For methane oxidation, more than eighty such models could be thus considered! By contrast, we started with the class of surface reaction controlling models, since experience has indicated that these models are generally found to represent reaction data. The analysis, then, provides first for an estimation of the number of sites participating in the reaction, n, *without* the necessity of the simultaneous estimation of the magnitudes of all of the individual adsorption constants in the model. Then, a model with the appropriate value of n and with the minimum possible number of adsorption constants is considered, and the experimenter is led by

the experimental data and the analysis to a model that, with a minimum number of parameter, adequately represents the data.

2. *Power-Function Models*

A similar procedure may also be used with the power-function models. For example, if we are considering the reaction (B14)

$$A + B \rightarrow F$$
$$A + F \rightarrow G$$

then an initial postulation of the form of the reaction rate model might be

$$-dC_A/dt = k_1 C_B + k_2 C_F \tag{101}$$

$$-dC_B/dt = k_1 C_B \tag{102}$$

$$dC_F/dt = k_1 C_B - k_2 C_F \tag{103}$$

$$dC_G/dt = k_2 C_F \tag{104}$$

These equations might be thought to apply for a high concentration of reactant A, at a given concentration of catalyst C, and at isothermal conditions. To describe measurements on the concentration of F as a function of time at each of the initial conditions of the 2^4 factorial design of Table XIII, we might tentatively entertain the integrated model:

$$C_F = \frac{C_{BO} k_1}{k_1 - k_2} (e^{-k_2 t} - e^{-k_1 t}) \tag{105}$$

This is simply the integrated form of the above equations.

Since several isothermal data points exist as a function of time for each of the 16 run conditions of Table XIII, we can estimate the values of the parameters k_1 and k_2 for each of these 16 conditions. This may be done using Eq. (105) and nonlinear least squares; the parameter estimates are shown in Table XIII (as taken from Reference B14).

Now, if the model of Eqs. (101)–(104) is oversimplified, the estimated values of the parameters k_1 and k_2 will likely be representable as

$$k = C_{AO}^m C_{CO}^n k_0 e^{-E/RT} \tag{106}$$

or

$$\ln k_1 = m_1 \ln C_{AO} + n_1 \ln C_{CO} + \ln k_{01} - E_1/RT \tag{107}$$

$$\ln k_2 = m_2 \ln C_{AO} + n_2 \ln C_{CO} + \ln k_{02} - E_2/RT \tag{108}$$

It should be noted at this point that we are confronted with a highly nonlinear problem involving extremely complex equations [if Eqs. (107) and (108) were

TABLE XIII

EXPERIMENTAL DESIGN AND ESTIMATED PARAMETER VALUES

Run No.	C_{AO}[a]	C_{BO}[a]	C_{CO}[a]	D[a]	$-10 \ln k_1$	$-10 \ln k_2$
1	−	−	−	−	79.74	72.68
2	+	−	−	−	61.99	62.09
3	−	+	−	−	77.49	68.03
4	+	+	−	−	64.56	62.04
5	−	−	+	−	74.80	67.25
6	+	−	+	−	59.91	62.19
7	−	+	+	−	75.99	69.35
8	+	+	+	−	61.01	62.40
9	−	−	−	+	69.48	65.09
10	+	−	−	+	55.70	58.05
11	−	+	−	+	68.78	64.31
12	+	+	−	+	54.82	58.19
13	−	−	+	+	65.71	63.42
14	+	−	+	+	52.04	57.86
15	−	+	+	+	65.22	64.44
16	+	+	+	+	51.76	57.01

[a] Values for pluses and minuses for C_{AO}, C_{BO}, C_{CO}, and D, respectively, are given below:

C_{AO}	C_{BO}	C_{CO}	D, temperature
+ 40 moles/liter	2 moles/liter	1 mM liter	175°C
− 20 moles/liter	1 mole/liter	0.5 mM liter	165°C

inserted into Eqs. (101)–(104)]. Rather, we have broken the problem into one easily solved equation [Eq. (105)] and two linear equations [Eqs. (107) and (108)]. In fact, the problem is entirely analogous to that of the previous subsection, except that the observed values of the parameters are contained in Table XIII and the predicted values are written as Eqs. (107) and (108). If we wished to adopt a solution similar to that used in model-building with hyperbolic models, we would assume the simplest form of Eqs. (107) and (108). Then we would predict parameter values k_1 and k_2 at each run, and compare the parameter residuals to the design of Table XIII. If the residuals were correlated with certain of the variables, we could introduce these variables, and proceed.

To illustrate a slightly different approach, however, let us simply fit Eqs. (107) and (108) to the observed values of the parameters k_1 and k_2 of Table XIII. Thus, we are able to estimate the magnitudes of the desired

parameters of Eqs. (107) and (108). The results of such a linear least-squares fit are contained in Table XIV. This table suggests that Eq. (105) should be

TABLE XIV

PARAMETER ESTIMATE SUMMARY

Parameter	m^a	n^a	$\ln k_0{}^a$	E, kcal/gm-mole^a
k_1	2.08 ± 0.16	0.47 ± 0.16	26.59 ± 4.70	35.2 ± 4.2
k_2	0.99 ± 0.20	0.12 ± 0.20	11.26 ± 0.65	18.5 ± 5.4

[a] The numbers following $\pm$ signs represent 95% confidence intervals.

modified to read

$$C_F = \frac{C_{BO} C_{AO}^2 C_{CO}^{0.5} k_1'}{C_{AO}^2 C_{CO}^{0.5} k_1' - C_{AO} k_2'} (e^{-k_2' C_{AO} t} - e^{-k_1' C_{AO}^2 C_{CO}^{0.5}}) \tag{109}$$

where

$$k_1' = 3.518 \times 10^{11} e^{-35200/RT}$$

$$k_2' = 7.795 \times 10^{4} e^{-18500/RT}$$

That is, the data suggest that Eqs. (101)–(104) should be modified to allow formation of F by a reaction second order in A, first order in B, and half-order in the catalyst concentration C. The disappearance of F appears to be first order in A and in B.

Before the analysis is complete, we must examine the adequacy of the model to fit the data. Since this procedure is described in Section IV as well as in Reference (B14), this aspect of the fitting will not be pursued here. We have simply demonstrated here how the experimenter can be *led* from an inadequate model to an adequate model by the data analysis.

VI. Empirical Modeling Techniques

In every case in which a kinetic model is selected to represent adequately a reaction, the rate surface *predicted* by the model must be compared to the surface *observed* in the data. In the methods discussed in Section II, only one section through the entire rate surface was examined; for example, the dependence of initial rate on total pressure could be investigated when in fact the total rate surface constituted the dependence of rate on several component partial pressures and temperature. The misleading results obtain-

able by using sections of surfaces in this manner have been discussed (K3, K7).

In this section, methods are described for obtaining a quantitative mathematical representation of the entire reaction-rate surface. In many cases these models will be entirely empirical, bearing no direct relationship to the underlying physical phenomena generating the data. An excellent empirical representation of the data will be obtained, however, since the data are statistically sound. In other cases, these empirical models will describe the characteristic shape of the kinetic surface and thus will provide suggestions about the nature of the reaction mechanism. For example, the empirical model may require a given reaction order or a maximum in the rate surface, each of which can eliminate broad classes of reaction mechanisms.

A. Response-Surface Methodology

The two primary features of response surface methodology are the experimental design and the method of data analysis (B7, D2, H3). In the design stages of the study, it is advantageous to use a minimum-variance design, such as the factorial or certain composite designs (H3). These designs not only cover a wide area of the kinetic surface, but they also can provide minimum-variance estimates of the parameters.

The importance of the parameter estimates becomes apparent from the data analysis. Suppose a nonlinear reaction-rate equation contains two independent variables and a set of unknown parameters:

$$r = f(x_1, x_2; \mathbf{K}) \tag{110}$$

The rate surface can be approximately represented, in the region where the data are taken, by a Taylor expansion

$$r = b_0 + b_1 x_1 + b_2 x_2 + b_{12} x_1 x_2 + b_{11} x_1^2 + b_{22} x_2^2 \tag{111}$$

Here, the coefficients **b** represent first and second partial derivatives of the rate expression $f(x_1, x_2; \mathbf{K})$ or functions thereof.

This is the general equation for an ellipse in two independent variables, as shown in Fig. 21. If a new coordinate system is defined that has its center at the point S and axes directed along X_1 and X_2 of Fig. 21, then Eq. (111) reduces to

$$r - r_s = B_{11} X_1^2 + B_{22} X_2^2 \tag{112}$$

Examination of the coefficients B, then, characterizes the rate surface. If, for example, B_{11} and B_{22} were significantly positive, any movement along the axes X_i would increase the rate r. Hence, the surface has a minimum at point S. Similarly, if the coefficients were negative, a maximum exists at point S.

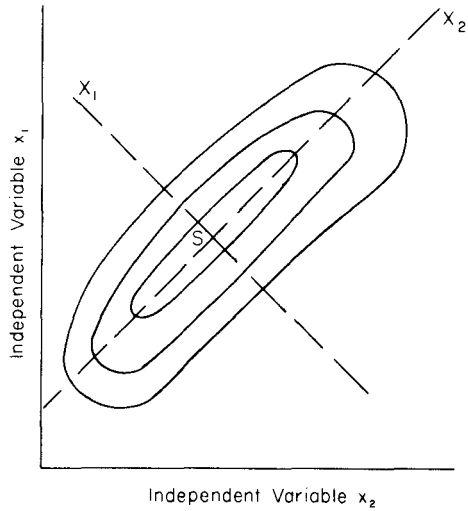

FIG. 21. Contours of constant reaction rate, Eqs. (111) and (112).

A saddle point exists if the coefficients are of unequal sign (B7, D2). Precise estimates of the parameters are of considerable importance.

The mathematical details of carrying out such a redefinition of coordinate system, termed a canonical transformation, have been presented elsewhere in detail (D2).

1. *Response-Surface Methods in Kinetic Modeling*

An augmented central composite design was used in obtaining reaction-rate data in a flow differential reactor; the reaction occurring was the isomerization of normal pentane to isopentane in the presence of hydrogen (C1). Using the subscripts 1, 2, and 3 for hydrogen, normal pentane, and isopentane respectively, an empirical rate equation can be written

$$r = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{11}x_1^2 + b_{22}x_2^2 + b_{33}x_3^2 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 \quad (113)$$

The parameter estimates obtained by a linear least-squares analysis of the original coded data (C1) are shown in Table XV. After a canonical analysis, this equation becomes

$$r - 2.11 = 0.155X_1^2 - 0.292X_2^2 + 0.419X_3^2 \quad (114)$$

The standard errors of the coefficients are, respectively, 0.094, 0.108, and 0.095. The center of the coordinate system is at $x_1 = 1.32$, $x_2 = 1.43$, and $x_3 = 2.08$.

TABLE XV

PARAMETER ESTIMATES FOR EQ. (113)

Parameter	Estimate (with standard error)
b_0	3.06 ± 0.20
b_1	-0.490 ± 0.12
b_2	1.26 ± 0.12
b_3	-1.50 ± 0.12
b_{11}	-0.077 ± 0.12
b_{22}	-0.208 ± 0.07
b_{33}	0.257 ± 0.12
b_{12}	-0.241 ± 0.12
b_{13}	0.505 ± 0.14
b_{23}	-0.158 ± 0.14

From a consideration of either Eqs. (113) or (114) (K3), it is evident that a saddle point is predicted from the fitted rate equation. This could eliminate from consideration any kinetic models not capable of exhibiting such a saddle point, such as the generalized power function model of Eq. (1) and the several Hougen-Watson models so denoted in Table XVI.

However, this particular experimental design only covered values of x_3 up to 1.68; consequently, the saddle point is only predicted by the model and not exhibited by the data. This is the reason the lack-of-fit tests of Section IV indicated neither model 3 nor model 4 of Table XVI could be rejected as inadequately representing the data. As is apparent, additional data must be taken in the vicinity of the stationary point to confirm this predicted nature of the surface and hence to allow rejection of certain models. This region of experimentation (or beyond) is also required by the parameter estimation and model discrimination designs of Section VII.

Response-surface methodology has also been used to gain theoretical insight into a reacting system (B19) and to determine the order of a reaction (P3).

2. *Response-Surface Methods in Process Exploitation*

Response-surface methodology has been used extensively for determining areas of process operation providing maximum profit. For example, the succinct representation of the rate surface of Eq. (114) indicates that increasing values of X_3 will increase the rate r. If some response other than reaction rate is considered to be more indicative of process performance (such as cost, yield, or selectivity), the canonical analysis would be performed on this response to indicate areas of improved process performance. This information

TABLE XVI

POSSIBLE RATE MODELS FOR PENTANE ISOMERIZATION

Model		Rate surface for model
1. Adsorption controlled single site	$r = \frac{k(x_2 - x_3/K)}{[1 + K_1x_1 + (K_2/K + K_3)x_3]}$	No stationary point except at equilibrium; first derivative with x_1 zero only at equilibrium; first derivative with x_2 always positive and with x_3 always negative except at equilibrium.
2. Adsorption controlled with dissociation	$r = \frac{k(x_2 - x_3/K)}{[1 + K_1x_1 + (K_2/Kx_3)^{\frac{1}{2}} + K_3x_3]^2}$	No stationary point except at equilibrium; first derivative with x_1 zero only at equilibrium; first derivative with x_2 always positive; first derivative with x_3 may be zero with second derivative positive.
3. Single site, surface reaction controlled	$r = \frac{kK_2(x_2 - x_3/K)}{(1 + K_1x_1 + K_2x_2 + K_3x_3)}$	No stationary point except at equilibrium; first derivative with x_1 zero only at equilibrium; first derivative with x_2 always positive and with x_3 always negative except at equilibrium.
4. Dual site, surface reaction controlled	$r = \frac{kK_2(x_2 - x_3/K)}{(1 + K_1x_1 + K_2x_2 + K_3x_3)^2}$	Can have saddle point; first derivative with x_1 zero at equilibrium; first derivative with x_2 may be zero with second derivative negative; first derivative with x_3 may be zero with second derivative positive.
5. Dual site, surface reaction controlled, hydrogen dissociated	$r = \frac{kK_2(x_2 - x_3/K)}{[1 + (K_1x_1)^{1/2} + K_2x_2 + K_3x_3]^2}$	Same as model 4.
6. Surface reaction controlled, dissociated *n*-pentane	$r = \frac{kK_2(x_2 - x_3/K)}{[1 + K_1x_1(K_2x_2)^{1/2} + K_3x_3]^2}$	Same as model 4.
7. Desorption controlled	$r = \frac{kK(x_2 - x_3/K)}{[1 + K_1x_1 + (K_2 + KK_3)x_2]}$	Same as model 1.

is obtained in addition to the modeling information described above. Hence, the response surface approach can yield dividends in both areas. The use of response surface methods for general process improvement (B7, D2) and for investigating the influence of process variables and catalysts on reactions (F2, F3) has been discussed.

B. Transformations of Variables

Particularly in empirical modeling, the transformation of a model can greatly assist in the modeling procedure. In some cases, a transformation can simplify the functional form of the model; in others, it can provide an improved fit of the data while requiring a previously specified function form (e.g., a power-function model). In still other cases, the transformations may ensure that certain assumptions are satisfied so that a simple and valid analysis may be performed. For example, it may be desirable to fit the logarithm of a dependent variable y by unweighted least squares rather than fitting y itself if the log function has constant variance. In this section, methods are discussed for determining transformations of the dependent variable, the independent variables, or both. Transformations of the parameters within a model were discussed in Section III.

1. *Transformations of Dependent Variable*

Theory for the transformation of the dependent variable has been presented (B11) and applied to reaction rate models (K4, K10, M8). In transforming the dependent variable of a model, we wish to obtain more perfectly (a) linearity of the model; (b) constancy of error variance, (c) normality of error distribution; and (d) independence of the observations to the extent that all are simultaneously possible. This transformation will also allow a simpler and more precise data analysis than would otherwise be possible.

Let us write an integrated ath-order rate equation as

$$\begin{aligned} C_{A0}^{1-a}[(1-x)^{1-a}-1] &= (a-1)kt \qquad a \neq 1 \\ -\ln(1-x) &= kt \qquad\qquad a = 1 \end{aligned} \tag{115}$$

If

$$\begin{aligned} \lambda &= a-1 \\ y &= (1-x)^{-1} \end{aligned} \tag{116}$$

then the rate equation may be written

$$y^{(\lambda)} = kt \tag{117}$$

where

$$y^{(\lambda)} = \begin{cases} (y^{\lambda}-1)/\lambda C_{A0}{}^{\lambda} & \lambda \neq 0 \\ \ln y & \lambda = 0 \end{cases} \tag{118}$$

Now, if the transformed variable is normalized to $Z^{(\lambda)}$ by the Jacobian of the transformation, we obtain

$$Z^{(\lambda)} = \begin{cases} (y^{\lambda} - 1)/\lambda \dot{y}^{\lambda - 1} & \lambda \neq 0 \\ \dot{y} \ln y & \lambda = 0 \end{cases} \tag{119}$$

where $\dot{y}$ is the geometric mean of the y's. Then it has been shown (B11, K10) that the steps required for effective reaction-order estimation are:

1. Estimate, by unweighted linear least squares, the parameter $\tilde{b} = k/J^{1/N}$, which minimizes the sum of squares

$$S(\lambda) = \sum_{i=1}^{N} (Z_i^{(\lambda)} - \tilde{b} t_i)^2 \tag{120}$$

for a given λ, and calculate the sum of squares $S(\lambda)$.

2. Plot this minimum sum of squares for several λ.
3. Read off the minimum of this plot to obtain the best λ, $\hat{\lambda}$.
4. Calculate the 99% confidence interval for this $\hat{\lambda}$ by

$$\ln S(\lambda) - \ln S_{\min} < \chi_1^2(0.01)/N = 6.63/N \tag{121}$$

If this procedure is followed, then a reaction order will be obtained which is not masked by the effects of the error distribution of the dependent variables If the transformation achieves the four qualities (a–d) listed at the first of this section, an unweighted linear least-squares analysis may be used rigorously. The reaction order, $a = \lambda + 1$, and the transformed forward rate constant, $\tilde{b}$, possess all of the desirable properties of maximum likelihood estimates. Finally, the equivalent of the likelihood function can be represented b: the plot of the transformed sum of squares versus the reaction order. This provides not only a reliable confidence interval on the reaction order, but also the entire sum-of-squares curve as a function of the reaction order. Then, for example, one could readily determine whether any previously postulated reaction order can be reconciled with the available data.

This method has been discussed in detail (K10) and extended to cover other important cases elsewhere (K4, M8). It can logically be asked why such a precise reaction order is wanted. This and several other points in the analysis can best be shown by the following example.

Pannetier and Davignon (P1) studied the solid–solid reaction

$$NiS_2(s) \rightarrow NiS(s) + \tfrac{1}{2}S_2(g) \tag{122}$$

They presented seven determinations of the mass of unreacted solid at each of four temperatures. By the method of differentiation, Pannetier and Davignon (P1) found the reaction to be two-thirds order.

Typical sum-of-squares contours are shown in Fig. 22, where $Z^{(\lambda)}$ represents the transformed dependent variable. The vertical broken lines

represent the 99% confidence interval. The results are shown below in complete form in Table XVII. Note that the two-thirds order is not compatible with the 415°C data. Although this order is adequate for the other data sets, a rather disturbing increase in the order with an increase of the temperature is present. In fact, the 395°C data alone would commonly be considered to be one-half order, with a relatively low probability of the two-thirds order being appropriate. This indication can be observed only for precise measures of the reaction order and its confidence interval. An analysis of variance for these data has been developed (K10).

TABLE XVII

ESTIMATES OF REACTION ORDER FOR EQ. (122)

Temperature (°C)	Reaction order (with 99% confidence interval)
395	0.53 ± 0.25
405	0.61 ± 0.05
415	0.91 ± 0.13
426	0.69 ± 0.09

If a single reaction order must be selected, an examination of the 95% confidence intervals (not shown) indicates that the two-thirds order is a reasonable choice. For this order, however, estimates of the forward rate constants deviate somewhat from an Arrhenius relationship. Finally, some trend of the residuals (Section IV) of the transformed dependent variable with time exists for this reaction order.

2. *Transformations of the Independent Variables*

In transforming the independent variables alone, it is assumed that the dependent variable already has all the properties desired of it. For example, if the y's are normally and independently distributed with constant variance, at least approximately, then any transformations such as described in Section VI,B,1 would be unnecessary. Under such assumptions, Box and Tidwell (B17) have shown how to transform the *independent* variables to reduce a fitted linear function to its simplest form. For example, a function that has been empirically fitted by

$$y = b_0 + b_1 x + b_2 x^2 \tag{123}$$

might be fitted more simply by

$$y = b \ln x \tag{124}$$

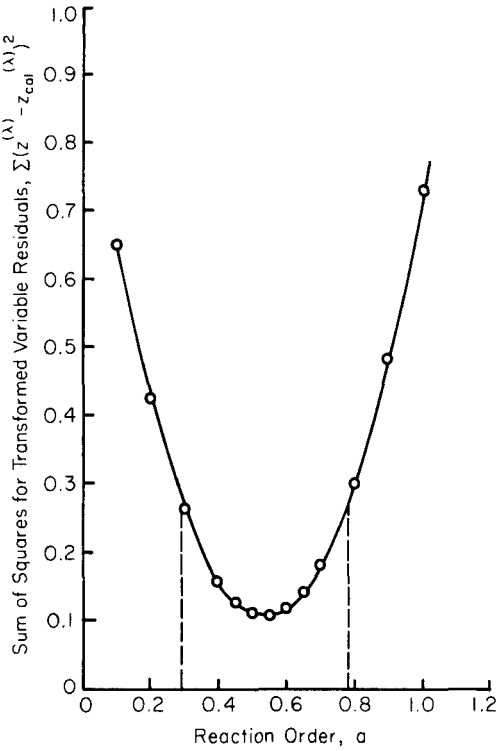

FIG. 22. Sum of squares curve for 395°C NiS_2 decomposition.

if the experimenter knew that it were preferable to transform x to $\ln x$.

A type of transformation explored in detail is

$$U_i \begin{cases} = x_i^{a_i} & a_i \neq 0 \\ = \ln x_i & a_i = 0 \end{cases} \tag{125}$$

This transformation thus includes the familiar reciprocal, square-root, and logarithmic transformations. Box and Tidwell have shown how the functions

$$y = \sum b_i U_i \tag{126}$$

or

$$y = \sum \sum b_{ij} U_i U_j \tag{127}$$

can be fitted iteratively. We examine the usefulness of this approach more fully in Section VI,B,3, immediately following.

3. *Transformations of Independent and Dependent Variables*

Combinations of the above two methods lead to equations of the form

(H2, W6):

$$y^{(\lambda)} = b_0 + \sum b_i U_i \tag{128}$$

where

$$y^{(\lambda)} \begin{cases} = (y^\lambda - 1)/\lambda & \lambda \neq 0 \\ = \ln y & \lambda = 0 \end{cases} \tag{129}$$

and U_i is given by Eq. (125). Here, if y is taken to be the reaction rate, then any rate expression from

$$r = k \prod_{i=1}^{N} p_i^{a_i} \tag{130}$$

to

$$r = \sum_{i}^{N} b_i p_i^{a_i} \tag{131}$$

may be determined. If, on the other hand,

$$y = \frac{\text{reaction driving force}}{r} \tag{132}$$

then, clearly, Eq. (128) can represent any of the hyperbolic models.

As an illustration, consider the hydrogenation of propylene over a platinum alumina catalyst discussed in Section II. These date were taken from 0 to 35°C, 1 to 4 atm total pressure, and 0 to 45% propylene. Equations (7) and (8) were obtained after considerable sifting and winnowing of rate equations. Both fit the observed data reasonably well.

An entirely empirical but very rapid technique is to fit

$$r^{(\lambda)} = b_0 + b_1 p_1^{a_1} + b_2 p_2^{a_2} + b_3 T^{a_3} \tag{133}$$

where p_1 and p_2 are partial pressures of hydrogen and propylene, respectively. By first using nonlinear least squares to estimate the parameters on the right-hand side of Eq. (133) at each of several values of λ, it was found that a_1 and a_2 should be zero, but that the values λ and a_3 were poorly defined. Consequently, the sums-of-squares contours similar to that of Fig. 22 were plotted as shown in Fig. 23 at the 95% confidence level. It is apparent that the effect of temperature on these data is not well defined, because a wide range of values of a_3 represents the data as well as the minimum sum-of-squares value. Choosing convenient values within the 95% confidence region, we obtain the rate equation

$$\ln r = 19.55 + 0.692 \ln p_1 - 0.226 \ln p_2 - 6.41 \times 10^3 T^{-1} \tag{134}$$

or

$$r = a_0 p_1^{0.692} p_2^{-0.226} \exp(-6.41 \times 10^3 T^{-1}) \tag{135}$$

A comparison of the predicted rates of Eqs. (7) and (135) is shown in Fig. 24. Since the bulk of the data were taken above 5% propylene, it is apparent that both models fit reasonably well. However, at lower concentrations of propylene, Eq. (135) will deviate widely from the data. Undoubtedly, deviations will also occur when extrapolating other directions from the data base as well. Checks on the adequacy of the transformation and calculations of the confidence regions for all parameters may also be carried out (B11, B17, H2).

The advantage of the approach is thus seen to be the speed and completeness with which a large number of empirical models can be tested and the best model of the group selected for closer examination. This empirical model may then be used as an end result, or it can serve as a starting point for more exhaustive analysis such as described in Sections III, V, and VII.

C. Empirical Model Tuning

In the correlation of kinetic data, one may spend considerable time and effort obtaining a theoretical model using the techniques presented in the accompanying sections. Alternatively, one may simply fit an empirical function to the data, using the several techniques already discussed in this section. Many cases between these extremes are met in practice however. This subsection discusses procedures for empirically modifying an approximate mechanistic model such that (a) the function form of the mechanistic

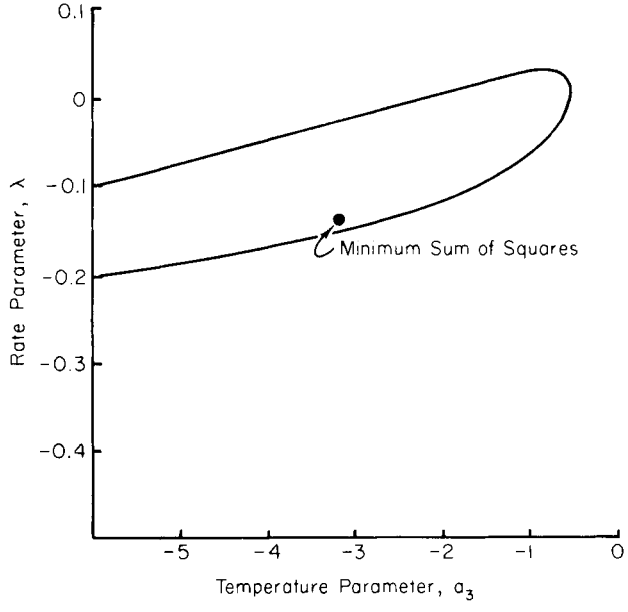

FIG. 23. Approximate 95% confidence region in $\lambda - a_3$ space for Eq. (133).

model can be retained; (b) the resulting model can describe the main features of the existing data; and (c) the model can possess a minimum number of parameters. This is done using the parametric residuals of Section IV,B,6, and already exploited in Section V.

Mickley, Sherwood, and Reed (M9) have discussed a useful method of empirically modifying a function to improve its fit of experimental data. The method would be applied when a particular rate equation

$$r_1 = f_1(\mathbf{x}; \mathbf{K}) \tag{136}$$

does not adequately describe the observed reaction rates, y. The procedure is to fit the residual of the dependent variable as an empirical function of the independent variables. If this empirical function were chosen to be linear, for example, we would estimate the empirical parameters $\mathbf{B}$ in

$$y - r_1 = \mathbf{xB} \tag{137}$$

After estimating the parameters $\mathbf{B}$, the resulting equation

$$r_1 = f_1(\mathbf{x}; \mathbf{K}) + \mathbf{xB} \tag{138}$$

may fit the data better than does Eq. (136). There are obvious disadvantages to this procedure for kinetic modeling. For example, we lose the advantage of the use of the functional form $f_1(\mathbf{x}; \mathbf{K})$, which would generally be selected

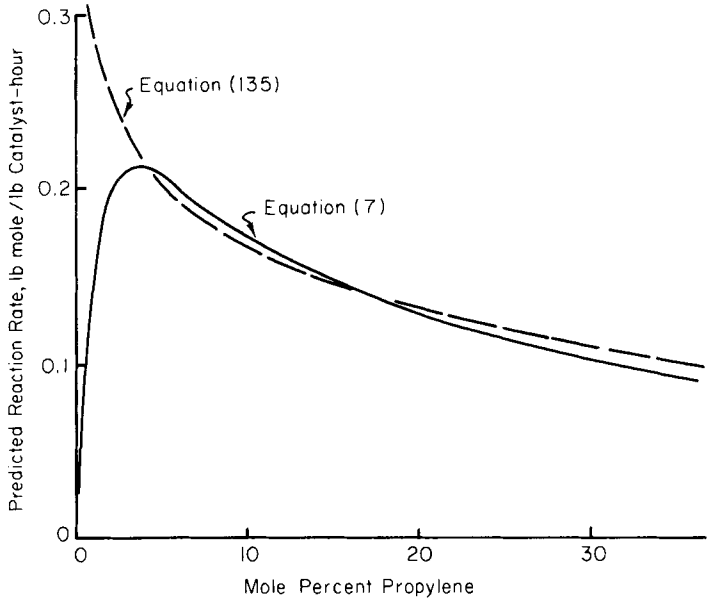

FIG. 24. Predicted reaction rates for Eqs. (7) and (135).

to provide reliable interpolation of kinetic data [e.g., Eqs. (1) or (2)]. Also, we add additional empirical constants **B** to Eq. (136), so that the residual mean square may actually be increased by following this procedure. Let us now describe how this general procedure may be utilized using the intrinsic parameters. In this way, the main behavior of, say, the Hougen–Watson model can be retained while making slight empirical modifications needed to fit the data. Also, the empirical modifications can be made without necessarily increasing the number of parameters in the model.

Let us consider the data taken by Laible (L1) on the dehydration of normal hexyl alcohol at 450°F over a silica alumina catalyst. The single- and dual-site surface reaction controlled models applying to alcohol dehydration were discussed in Section V,A,2. We now consider, however, the functional forms given, for example, by Eq. (84), as probably being capable of describing the data, but do not restrict the C_1 and C_2 plots to a linear pressure dependence as before. Rather, we obtain an empirical pressure dependence from the

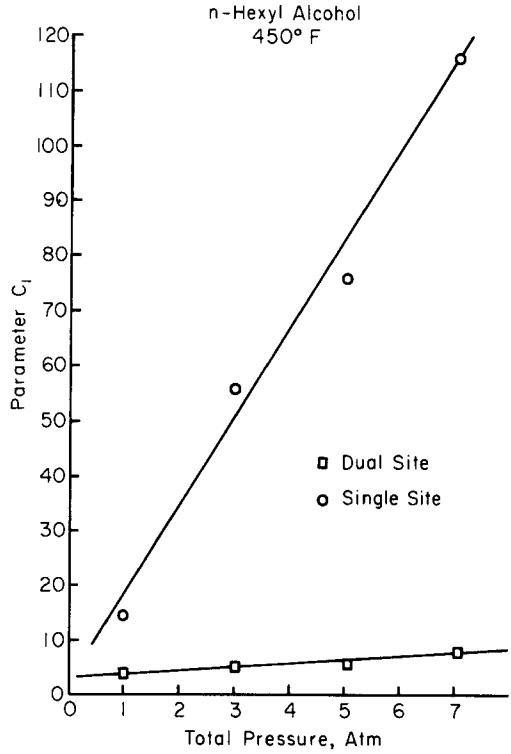

FIG. 25. Plot of parameter C_1 versus pressure for *n*-hexyl alcohol dehydration, Eqs. (80) and (81).

data. Figure 25 presents the single- and dual-site C_1 plots; it can be seen that either model is well described by

$$\hat{C}_1 = \bar{a} + B_1\pi \tag{139}$$

The C_2 plots of Fig. 26, however, indicate that a linear C_2 dependence upon pressure will not be adequate. The use of the dual-site model will require three additional parameters for C_2, in order that the model approximately describe the data. The single-site model, however, can be used if we take

$$\hat{C}_2 = \bar{a} + B_2\pi^2 \tag{140}$$

where the parameter $\bar{a}$ is contained in both Eqs. (139) and (140). Thus, a model with only three parameters, which describes the main features of the data, is (K4)

$$r = \frac{(1-x)\pi}{(3+15\pi)+(3+15.5\pi^2)x} \tag{141}$$

Although we have intentionally chosen a simple example here to illustrate the method, we note that the same result could be obtained by a regression of any of the parameters of Sections IV,B or V on the independent variables. This procedure is then analogous to that described by Eqs. (136)–(138) except with the parameters as dependent variables.

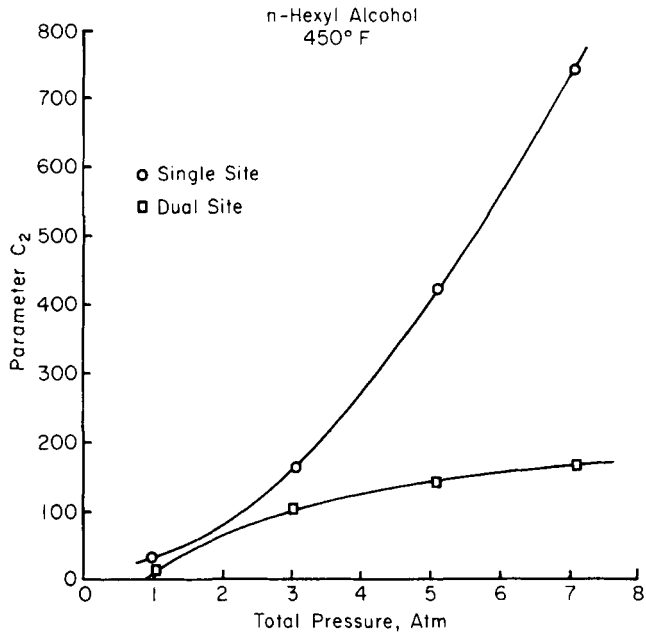

FIG. 26. Plot of parameter C_2 versus pressure for *n*-hexyl alcohol dehydration, Eqs. (80 and (81).

VII. Experimental Designs for Modeling

The importance of a sound experimental design in modeling has already been emphasized. Normally, however, one cannot construct a detailed plan for the conduct of an entire future experimental program. At the outset of many kinetics experiments, we are not only unaware of the *ranges* over which certain variables should be studied but also of *which* variables should be studied. Often, it is only after observing unexpectedly wide data scatter or unacceptably rapid catalyst deactivation that we begin to appreciate the importance of certain variables. Consequently, we generally take a limited set of data and use the results to guide our next experiments.

Box recognized that not only do most kinetics experiments proceed in such a sequential fashion, but indeed most scientific experiments proceed iteratively. In his descriptions of the iterative nature of experimentation (B9), he has identified the stages of *conjecture*, *design*, *experiment*, and *analysis*. In such experimentation, one designs experiments to test a certain conjecture, conducts the experiment and analyzes the data. From this analysis, one gains information leading to new conjectures. The sequence is then repeated, making full use of all the information available at any given stage in planning experiments for the next stage. This strategy is particularly valuable in kinetic modeling studies.

It is partly the fault of statistics that experimenters have misconstrued the value of the *number* and *precision* of data points relative to the value of the location of the points. The importance of the location of the data in the model specification stage can be seen from Fig. 1, which represents literature data (M3) on sulfur dioxide oxidation. The dashed and solid lines represent the predicted rates of two rival models, and the points are the results of two series of experimental runs. It can be seen that neither a greater number of experimental points nor data of greater precision will be of major assistance in discriminating between the two rival models, if data are restricted to the total pressure range from 2 to 10 atm. These data simply do not place the models in jeopardy, as would data below 2 atm and greater than 10 atm total pressure. This is presumably the problem in the water–gas shift reaction, which is classical in terms of the number of models proposed, each of which adequately represent given sets of data.

The proper location of data is also important in parameter-estimation situations. For the nitric oxide reduction reaction (K11), for example, the relative sizes of the three-dimensional confidence regions calculated after each observation are shown in Fig. 27. The size of the confidence region after 12 points taken according to a one factor at a time variation of hydrogen and nitric oxide partial pressures is seen to be equivalent the size of the region

after a 2^2 factorial followed by one point taken by a minimum-volume design. The size of the region would have been substantially smaller had the first three points been taken by a Box–Lucas design (B16), followed by the minimum-volume design for the next two points. Note, from Fig. 27, that 18 times more information is obtained from 12 minimum-volume design points than from the 12 one-factor design points. This is a very significant increase in information; it has been shown, for example, that an increase in the precision of an estimated rate constant by only a factor of 4 can reduce the reactor overdesign required for such uncertainties by 15% of the reactor volume (K6).

Experimental design procedures exist for discriminating among m rival models or estimating parameters in kinetic models. Figure 28 indicates three degrees of sophistication of such techniques that are available. At one extreme, the experimenter can simply use his intuition and experience, for example in suggesting that a tenfold variation from 1 to 10 atm total pressure should be adequate. Or, perhaps, high and low concentrations may be used. It is an unfortunate result of human nature that such experimental designs

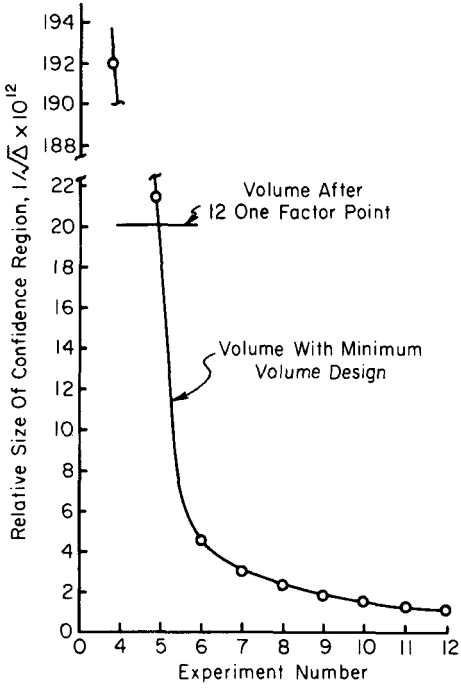

FIG. 27. Relative size of confidence region versus experiment number for nitric oxide reduction, Eq. (149).

may be influenced more by convenience than necessity; in any event, this method is best utilized by a gifted few. In very simple situations, such as the case in Fig. 1 with two models and one independent variable, a few hand calculations will suggest the proper regions of experimentation. For this example, given the experimental data of the figure and the scientific knowledge of two rival models, parameter estimates could be obtained and the predicted reaction rates sketched as shown in Fig. 1. These hand calculations, then, suggest that data at higher and lower pressure would be informative. The more usual situation, however, is where several complex models exist, each being influenced by several independent variables. In such cases, computer studies can best consider all of the available data and scientific knowledge, to evaluate the numerous possible alternatives.

The general scheme required here is shown in Fig. 29, a generalization of Fig. 28. With some ideas as to good areas of experimentation, the experimenter takes an initial set of data. These data are then analyzed to determine the best estimates of the parameters of the model or models under consideration. Since models that usually arise in these circumstances are nonlinear in the parameters, some version of nonlinear estimation will usually be employed in this analysis. Nonlinear estimation techniques, of course, almost always require the use of a computer.

Following this data analysis, the next step is the selection of a further set of runs to be performed; this also usually requires the use of a computer. For the two areas of modeling being discussed here, model discrimination and parameter estimation, this step involves maximization of the design criterion appropriate for each case. Settings of the independent variables, which extremize the criterion, are therefore determined, and these settings constitute the experimental design according to which the next set of runs

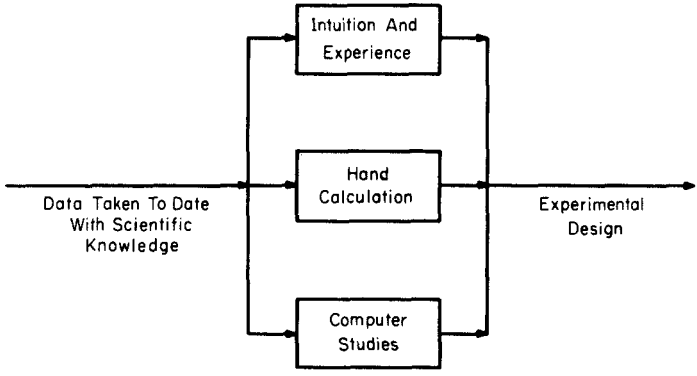

FIG. 28. General methods of selecting model discrimination or parameter estimation designs.

are performed. Once the data are collected, reestimation of the parameters follows. This cycle can be continued a sufficient number of times to accomplish the modeling objective (either model discrimination or parameter estimation). The strategy makes heavy use of the computer, there being a continual iteration between computer and laboratory.

A. Model-Discrimination Designs

Suppose that a given reacting system is about to be studied for which there exists a number of rival models; the object of the experimentation is here presumed to be the elimination of the inadequate models. Through this series of experiments, then, the experimenter is attempting to arrive at the best mathematical model for his system. The original paper (H13) in this area suggested sequential designs for discriminating between $m = 2$ models. Box and Hill (B13) later showed how to discriminate between an arbitrary number of rival models. In this later generalized approach, the

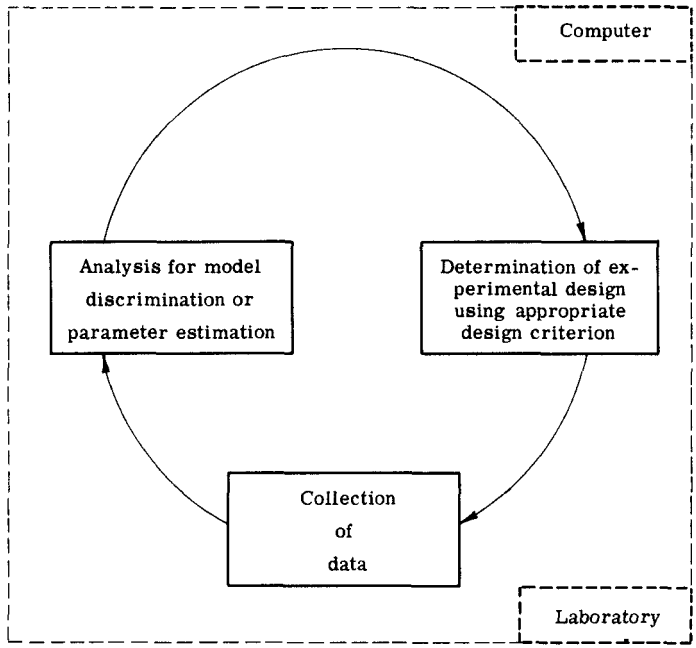

Fig. 29. Experimental strategy for mechanistic modeling.

experimenter iteratively selects those experimental conditions for the next stage of experimentation according to the following criterion:

$$D = \sum_{i=1}^{m} \sum_{j=1}^{m} \pi_i \pi_j D_{ij} \tag{142}$$

where

$$D_{ij} = \frac{1}{2}\left[\frac{(\hat{\sigma}_i^2 - \hat{\sigma}_j^2)^2}{(\sigma^2 + \hat{\sigma}_i^2)(\sigma^2 + \hat{\sigma}_j^2)} + (\hat{y}_i - \hat{y}_j)^2\left(\frac{1}{\sigma^2 + \hat{\sigma}_i^2} + \frac{1}{\sigma^2 + \hat{\sigma}_j^2}\right)\right]$$

The criterion D is a measure of divergence among the models, obtained from information theory. The quantity π_i is the prior probability associated with model i after the nth observation is obtained; σ^2 is the common variance of the n observations $y(1)$, $y(2)$, $\cdots$, $y(n-1)$, $y(n)$; $\hat{\sigma}_i^2$ is the variance for the predicted value of $y(n+1)$ by model i. When we have two models, D simplifies to

$$D = \frac{(\hat{\sigma}_2^2 - \hat{\sigma}_1^2)}{(\sigma^2 + \hat{\sigma}_1^2)(\sigma^2 + \hat{\sigma}_2^2)} + (\hat{y}_1 - \hat{y}_2)^2\left(\frac{1}{\sigma^2 + \hat{\sigma}_1^2} + \frac{1}{\sigma^2 + \hat{\sigma}_2^2}\right) \tag{143}$$

Note from Eq. (142) that this discriminant is large when $\hat{y}_i$ is far from $\hat{y}_j$, $\hat{\sigma}_i^2$ and $\hat{\sigma}_j^2$ are small, and π_i and π_j are jointly large. This means that in order to decide which of several models is adequate, we should seek out settings of the independent variables for which the responses of the various models are expected to be quite different [$(\hat{y}_i - \hat{y}_j)$ large] and for which the responses can be predicted relatively precisely ($\hat{\sigma}_i^2$ small). Also, we should not give much weight to models that we suspect are relatively poor ($\pi_i \pi_j$ small), even though the two models predict responses far from one another. This discrimination procedure, then, utilizes all of these concepts to provide effective model discrimination designs. In particular, after N observations become available, the quantity D will be maximized with respect to the settings for the next experimental run, $N+1$, and these maximizing values will be used for the $(N+1)$th experiment. This procedure could be continued until one model was found to be clearly superior to the others. This iterative discrimination procedure can thus be represented by Fig. 29, where the appropriate design criterion is taken to be, at each stage, the maximization of D as defined in Eq. (142).

As an illustration of the power of the method, two models known to fit the propylene hydrogenation data are shown in Eqs. (7) and (8). (See also Sections II,B and VI,B,3). Let us denote Eq. (7) by model 1 and Eq. (8) by model 2.

An initial nine data points were taken at 35°C in an adiabatic flow reactor. The initial prior probabilities were taken to be equal (equal probability of

adequacy) and σ^2 was estimated from experimental data (H12). Because of rapid catalyst deactivation and equipment limitations, total pressure was maintained at or below 3 atm.

Figure 30 portrays the grid of values of the independent variables over which values of D were calculated to choose experimental points after the initial nine. The additional five points chosen are also shown in Fig. 30. Note that points at high hydrogen and low propylene partial pressures are required. Figure 31 shows the posterior probabilities associated with each model. The acceptability of model 2 declines rapidly as data are taken according to the model-discrimination design. If, in addition, model 2 cannot pass standard lack-of-fit tests, residual plots, and other tests of model adequacy, then it should be rejected. Similarly, model 1 should be shown to remain adequate after these tests. Many more data points than these 14 have shown less conclusive results, when this procedure is not used for this experimental system.

For the multiresponse situation, several measurable responses are implicit in each model under consideration. For example, for the reaction

$$A \begin{matrix} \rightarrow B \rightarrow D \\ \rightarrow C \end{matrix}$$

several models could be postulated, one of which could be that each component undergoes a first-order decomposition. For each such model, several responses could be measured as a function of time, e.g., the concentrations of components A, B, C, and D. Just as there is substantial advantage in making use of each of these responses in parameter estimation (Section III,D), considerable information about model discrimination will be lost unless all of the possible responses are utilized. The model-discrimination procedure for several responses has been developed (H10) and exemplified (H14).

B. Parameter-Estimation Designs

Once model discrimination has been accomplished (that is, from a large group of rival models one best model has been selected as being adequate), further experimentation can be conducted to improve parameter estimates. For such parameter-estimation experiments, one design criterion suggested is to choose experiments providing the most desirable posterior distribution of the parameters. Under certain assumptions (B15), the procedure reduces to the maximization of the determinant

$$\Delta = |\mathbf{X}'\mathbf{X}| \tag{144}$$

where the (i, u)th element of this $N \times p$ $\mathbf{X}$ matrix is

$$f'_{iu} = \left. \frac{\partial f(\mathbf{x}_u ; \mathbf{K})}{\partial K_i} \right|_{\mathbf{K}=\mathbf{K}^\circ} \tag{145}$$

It is to be noticed that for linear models, such as

$$f(\mathbf{x}_u ; \mathbf{K}) = K_1 x_{1u} + K_2 x_{2u}^2 \tag{146}$$

then Eq. (145) becomes

$$\begin{aligned} f'_{1u} &= x_{1u} \\ f'_{2u} &= x_{2u}^2 \end{aligned} \tag{147}$$

The determinant is then, for N data points,

$$\Delta = \begin{vmatrix} \sum_{u=1}^{N} x_{1u}^2 & \sum_{u=1}^{N} x_{1u} x_{2u}^2 \\ \sum_{u=1}^{N} x_{1u} x_{2u}^2 & \sum_{u=1}^{N} x_{2u}^4 \end{vmatrix} \tag{148}$$

For this linear model, then, we simply find values of x_1 and x_2 that maximize Δ. We do not need estimates of the parameters to define this minimum variance design. For nonlinear models, such as Eq. (40), this is not true. The

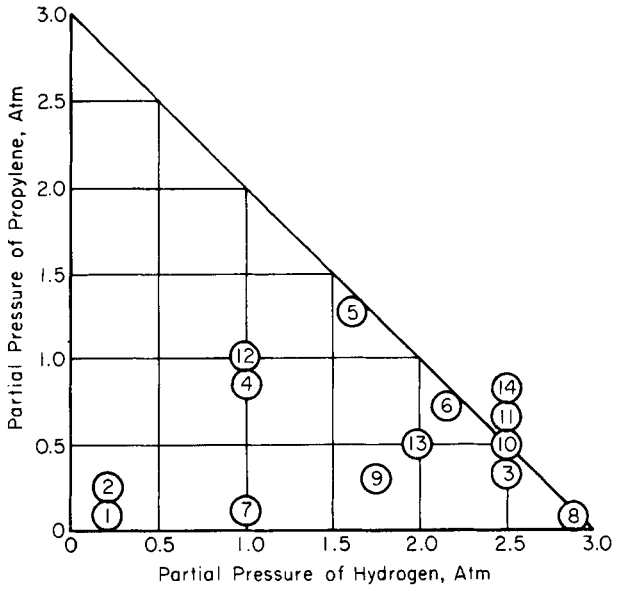

FIG. 30. Settings of independent variables for propylene hydrogenation for discriminating between Eqs. (7) and (8).

criterion Δ will be a function of the parameter values and, hence, the optimal design depends on these values; initial parameter estimates must be provided to use this criterion (B16).

By maximizing the determinant Δ, we can expect to have a smaller confidence region of the parameter estimates than that obtained with any other experiment in the possible region of experimentation. The larger the confidence region, of course, the more uncertain is our knowledge concerning the estimated parameters; the smaller the confidence region, the more precise is our knowledge.

The iterative parameter-estimation procedure can also be represented as in Fig. 29. An experimenter performs an experiment, obtains the data and estimates the parameters, for example, by nonlinear least squares. Then he checks the confidence region of the parameter estimates, perhaps through Δ as shown in Fig. 27. If it is too large, the experimenter finds those experimental settings that maximize the parameter estimation criterion of Eq. (144), that is, experimental settings that will reduce the size of confidence region as

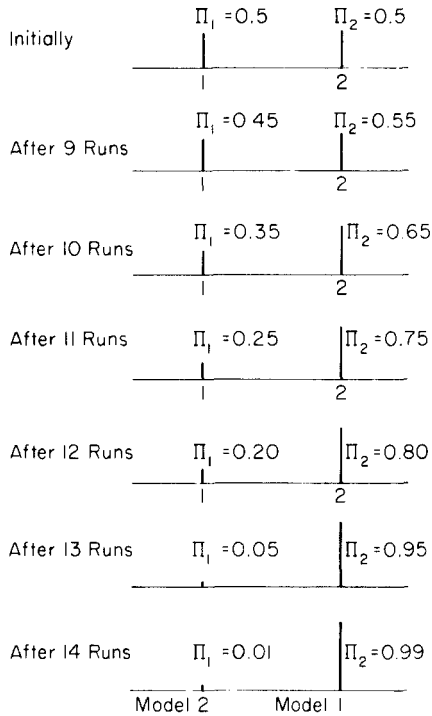

FIG. 31. Posterior probabilities in discrimination of propylene hydrogenation models of Eqs. (7) and (8).

much as possible. He then conducts an experiment as indicated by this experimental design, analyzes the data, and checks whether the parameters have now been estimated with sufficient precision. If they have, he terminates the experimentation; if not, he repeats the cycle.

To illustrate this criterion, consider a simulation problem chosen to be similar to the catalytic reduction of nitric oxide (A3, K11):

$$NO + H_2 \rightleftharpoons H_2O + \tfrac{1}{2}N_2$$

In an experimental study (A3), the reaction model was found to be the surface reaction between an adsorbed nitric oxide molecule and one adjacentlyl adsorbed hydrogen molecule:

$$r = \frac{kK_{NO}K_{H_2}p_{NO}\,p_{H_2}}{(1 + K_{NO}p_{NO} + K_{H_2}p_{H_2})^2} \tag{149}$$

The confidence region obtained using a simulated one variable at a time design was first examined, since this was the design used by the original experimenters.

The approximate 95% confidence region for the non-linear least-squares estimates using these data is shown in Fig. 32. The surface represented in this figure is the contour of the sums of squares surface, which has the value S_c given by Eq. (56).

In order to obtain more precise estimates of these parameters, the experimental design was also constructed using the above parameter estimation criterion. Using four preliminary observations taken according to a factorial design, the three parameters (k, K_{NO}, and K_{H_2}) were estimated. A fifth point was selected to minimize the joint confidence region of these parameters, using the settings of the partial pressures maximizing Δ. Following Fig. 29, the simulated reaction rate at this fifth point was used to reestimate the parameters by nonlinear least squares. Then the sixth point was chosen by again maximizing the determinant Δ, and the entire procedure was repeated.

Since the square root of the determinant Δ is inversely related to the size of the confidence region, the rate of decrease in volume obtained by this design procedure may be represented by Fig. 27. The solid horizontal line is the size of the region obtained by the twelve one variable at a time experimental points. In contrast, the estimates from the minimum-volume design are equally precise after the fifth point (i.e., the first point chosen by the minimum volume design). A visual measure of the size of the confidence region after twelve well-designed points can be obtained from Fig. 33, which is one-eighteenth the size of the one variable at a time confidence region of Fig. 32.

This parameter-estimation technique has also been extended to the multiple-response case (D3). Just as was seen in the multiple-response

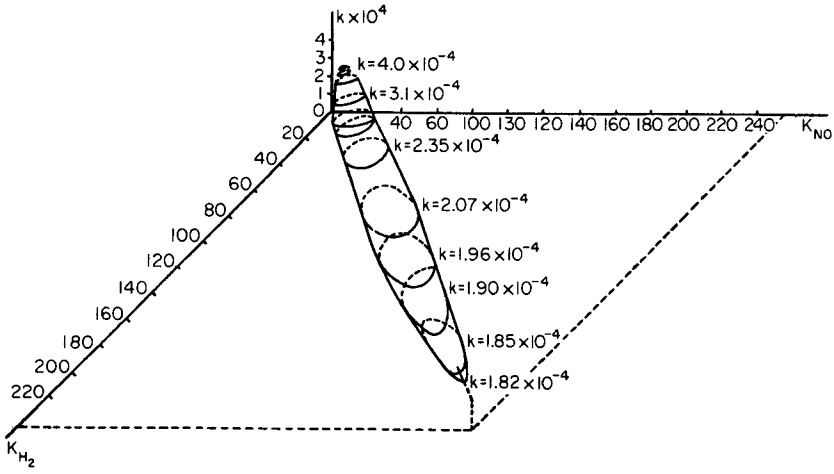

FIG. 32. Approximate 95% confidence region for Eq. (149) and one variable at a time design.

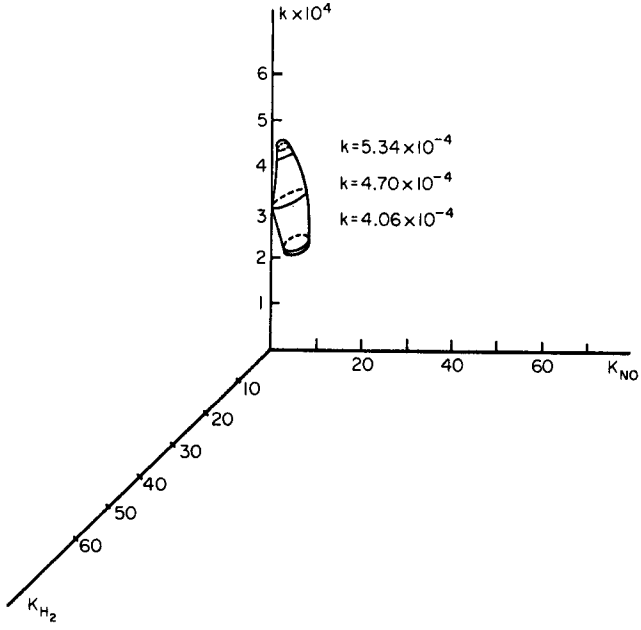

FIG. 33. Approximate 95% confidence region for Eq. (149) and sequential minimum volume design.

parameter-estimation example (B12), much greater power is available to the experimenter when all available responses are used. Also, a method has been suggested (H4) for combining the model-discrimination and parameter-estimation criteria, in order to increase the efficiency of the overall modeling program.

It will be noticed from Fig. 29 that there is a continual iteration between the laboratory and the computer. In the laboratory area, there could be a remote console linked directly to a time-shared digital computer, although immediate laboratory access is not essential. After obtaining his results, the experimenter would enter these into the digital computer, and the computer would perform routine analyses that have been programmed. For example, least-squares estimates would be calculated and the confidence region determined, for each model. Graphical output in the form of various plots of the residuals could also be obtained and examined. The computer could then be asked to determine those settings for a further run that would extremize the appropriate design criterion. The computer, having executed this instruction, would print out these optimal settings for consideration by the experimenter. The experimenter, in turn, could use these settings to proceed with further experimentation, or ask for other possible settings.

There is danger in describing this procedure in such an oversimplified way; it could be misconstrued that the investigator may only set variables and run experiments as the computer dictates. Experimentation in kinetics is far too complex for any such naive approach as this. In computer-assisted sequential experimentation, it is still necessary that the experimenter provide new ideas, intuition, and insights, which could possibly change the direction of the investigation in a fundamental manner. In short, the experimentation cannot be completely programmed ahead of time. Natural discoveries and surprises are the essence of research, and any experimental strategies for planning experiments must take them into account. However, the continual iteration between computer and laboratory, while not solving the problems completely, can contribute much in the way of solving modeling problems.

Nomenclature

A. A component of a reacting mixture

a A reaction order to be estimated for component A

a_i An empirical exponent of Eq. (133)

$\bar{a}$ An empirical parameter in Eqs. (139) and (140)

B A component of a reacting mixture

B_i An empirical parameter in Eqs. (139) and (140), $i = 1, 2$

B_{ii} The canonical parameters defining the nature of the response surface in transformed coordinates, $i = 1, 2, 3, \cdots$

$\mathbf{B}$ A vector of empirical parameters of Eqs. (137) and (138), obtained by linear regression

of residuals on x

b A reaction order to be estimated for component B

b_i An estimated value of a parameter

$\bar{b}$ A combination of rate and absorption constants, defined by Eqs. (94), (95), and (99)

$\tilde{b}$ A transformed rate constant for Eq. (115), obtained by dividing the rate constant by $J^{1/N}$

$\hat{b}$ Least-squares estimate of a parameter

$\mathbf{b}$ A vector of parameters to be estimated

C Designates a component of a reacting mixture

C_1 An observed value of an intrinsic parameter, which collects all denominator terms of hyperbolic models not multiplied by conversion

$\hat{C}_1$ The value of C_1 predicted by an assumed model

C_2 An observed value of an intrinsic parameter, which collects all denominator terms of hyperbolic models multiplied by conversion

$\hat{C}_2$ The value of C_2 predicted by an assumed model

C_A Concentration of any reactant A

C_{A0} Concentration of any reactant A at zero time

C_L Molar concentration of vacant active sites on catalyst

C_{Ai} Concentration of any component A at the ith data point, $i = 1, 2, \cdots, N$

$\hat{C}_{Ai}$ Concentration of component A at ith data point predicted by a particular model

$\mathbf{C}$ Variance–covariance matrix of observation errors, defined in Eq. (39)

$\text{cov}(b_1, b_2)$ Covariance of two parameters, b_1 and b_2

D Model discrimination criterion defined in Eq. (142)

D_{ij} Quantity related to the divergence D defined in Eq. (143)

E Activation energy

E' Ratio of activation energy to universal gas constant, E/R

$E(C_1-\hat{C}_1)$ Expected value of the residual of C_1

$F_{\alpha(\upsilon_1, \upsilon_2)}$ The $100(1-\alpha)\%$ point of the F statistic with (υ_1, υ_2) degrees of freedom

F Mass feed rate to reactor

f'_{iu} Partial derivative of nonlinear rate model with respect to the ith parameter and evaluated at the nth experimental point

$\mathbf{I}$ Matrix of unity diagonal terms and zero off-diagonal terms

J Jacobian of the transformation of the variable y to the variable $y^{(\lambda)}$

K Thermodynamic equilibrium constant for over-all reaction

K_i Adsorption equilibrium constant for species i, or product of absorption constant and initial mole fraction of species i

$\bar{K}_i$ Adsorption equilibrium constant for species i

$K^{(i)}$ Estimate of a parameter after ith iteration of nonlinear estimation program

$\mathbf{K}$ Vector of estimated rate and adsorption parameters

$\mathbf{K}^\circ$ Vector of initial estimates of K for use in nonlinear least squares

k Forward reaction-rate constant

k' Reparameterized forward reaction-rate constant

$\bar{k}$ Forward reaction-rate constant

k_0 Preexponential factor of forward rate constant

k_0' Reparameterized preexponential factor of forward rate constant

L Total concentration of active sites on catalyst

m Number of rival models
N Number of experimental observations
n Exponent of denominator of generalized hyperbolic rate equation, such as Eq. (94), or the reciprocal thereof
$\tilde{n}$ Number of replicated data sets
p Number of parameters in a model
p_i Partial pressure of species i in reaction mixture
q Number of replicated points, used to estimate experimental error
R Universal gas constant
r Reaction rate
r_i Reaction rate predicted from model i
r_0 Initial reaction rate
r_S Reaction rate at the stationary point of the reaction-rate surface
r_u^v Reaction rate predicted from response v at the uth experimental point
$r^{(\lambda)}$ Transformed reaction rate, where the transformation is defined generally in Eq. (129)
S Sum of squares of residuals, defined in Eq. (19)
S_c Critical sums of squares contour giving a $100(1-\alpha)\%$ confidence region
$S_{\min}$ Minimum sum of squares on a sum-of-squares surface, calculated by introducing least-squares parameter estimates into Eq. (19)
S_r Multiple-response sum of squares defined by Eq. (64)
S_Δ Multiple-response sum of squares defined by Eq. (65).
$S(\lambda)$ Minimum sum of squares of residuals for a given value of λ, defined by Eq. (120)
s^2 Estimate of experimental error variance
s_θ Standard error of parameter estimate $\hat{\theta}$
T Absolute temperature
$\bar{T}$ Mean value of absolute temperatures of a particular set of data
t Measure of reaction time
$t_{1/2}$ Half-life of a reaction
$t_{\upsilon,\alpha}$ The $100(1-\alpha/2)\%$ point of the t distribution with υ degrees of freedom
U_i Transformed dependent variable defined by Eq. (125)
$v(b_i)$ Variance of the parameter estimate b_i
$v(\mathbf{b})$ Variance–covariance matrix of the vector of parameter estimates $\mathbf{b}$
v Designates the response number in Eq. (65)
W Mass of catalyst in reactor
X_i The canonical axis i, $i = 1, 2, 3, \cdots$
$\mathbf{X}$ A matrix of independent variables, defined for a linear model in Eq. (29) and for nonlinear models in Eq. (43)
$\mathbf{X}^1$ Transpose of matrix $\mathbf{X}$
$\mathbf{X}^{-1}$ Inverse of the matrix $\mathbf{X}$
x Conversion in reactor
x_i A generalized independent variable
$\hat{x}_i$ Predicted value of the reactor conversion, for model i
$\mathbf{x}$ A matrix of the independent variables x_i
$\mathbf{y}$ A vector of the independent variables y_i
y Dependent variable related to conversion by Eq. (116)
y_i Generalized dependent variable
$\dot{y}$ Geometric mean of any set of variables y_i
$\bar{y}$ Arithmetic mean of any set of variables y_i
y_{A0} Initial mole fraction of component A
$\hat{y}_i$ Predicted value of the generalized dependent variable y_i
$y^{(\lambda)}$ Transformation of y, given by Eq. (118)

y_u^v Observed value of the response v at the uth experimental point
z Dependent variable defined by Eq. (77)
$\hat{z}$ Predicted value of z
$Z^{(\lambda)}$ Normalized transformed dependent variable given by Eq. (119)
α Confidence level of F or t distributions
β_i True value of an ith parameter
$\boldsymbol{\beta}$ Vector of parameter values β_i
γ_i Rate constants of Eq. (7), $i = 1, 2$
Δ Determinant defined in Eq. (65)
ε Experimental error
$\hat{\theta}$ Estimated value of any parameter
Λ A nonintrinsic parameter defined in Eqs. (77) and (78)
λ A transforming parameter related to the reaction order
υ Number of degrees of freedom
π Total pressure
π_i Posterior probability associated with model i
σ^2 Experimental error variance
σ_i^2 Experimental error variance for ith experimental point
σ_r^2 Variance of the reaction rate, r
$\hat{\sigma}_i^2$ Variance of the predicted rate from model i
$\chi_1^2(\alpha)$ The $100(1-\alpha)\%$ point of the chi-squared distribution with 1 degree of freedom
ψ A parameter setting the relative contributions of the linearization and steepest descent methods in determining the correction vector $\mathbf{b}$ of Eq. (45)

References

A1. Anscombe, F. J., *Proc. Berkeley Symp. Math. Statist. Probability*, 4*th*, **1**, 19 (1963).
A2. Anscombe, F. J., and Tukey, J. W., *Technometrics* **5**, 141 (1963).
A3. Ayen, R. J., and Peters, M. S., *Ind. Eng. Chem. Proc. Design Develop* **1**, 204 (1962).
B1. Ball, W. E., and Groenweghe, L. C. D., *Ind. Eng. Chem. Fundamentals* **5**, 181 (1966).
B2. Bartlett, M. S., *Biometrics* **3**, 39 (1947).
B3. Beale, E. M. L., *J. Roy. Statist. Soc., Ser. B* **22**, 41 (1960).
B4. Benson, S. W., "The Foundations of Chemical Kinetics." McGraw-Hill, New York, 1961.
B5. Booth, G. W., and Peterson, T. I., IBM SHARE Program "WLNLI" No. 687 (1958).
B6. Boudart, M., and Chembers, R. P., *J. Catalysis* **5**, 517 (1966).
B7. Box, G. E. P., *Biometrics* **10**, 16 (1954).
B8. Box. G. E. P., *Ann. N.Y. Acad. Sci.* **86**, 792 (1960).
B9. Box, G. E. P., Univ. of Wisconsin Dept. Statist. Tech. Rept. No. 111 (1967).
B10. Box, G. E. P., and Coutie, G. A., *Proc. Inst. Elect. Engrs. (London), Pt. B* **103**, No. 1, Suppl. 100 (1956).
B11. Box G. E. P., and Cox, D. R., *J. Roy. Statist. Soc., Ser. B* **26**, No. 2, 211 (1964).
B12. Box, G. E. P., and Draper, N. R., *Biometrika* **52**, 355 (1965).
B13. Box, G. E. P., and Hill, W. J., *Technometrics* **9**, 57 (1967).
B14. Box, G. E. P., and Hunter, W. G., *Technometrics* **4**, 301 (1962).
B15. Box, G. E. P., and Hunter, W. G., *Proc. IBM Sci. Computing Symp. Statist.*, 113 (1965).
B16. Box, G. E. P., and Lucas, H. L., *Biometrika* **46**, 77 (1959).

B17. Box, G. E. P., and Tidwell, P. W., *Technometrics* **4**, 531 (1964).
B18. Box, G. E. P., and Wilson, H. L., *Biometrika* **46**, 77 (1959).
B19. Box, G. E. P., and Youle, P. V., *Biometrics* **11**, 287 (1955).
C1. Carr, N. L., *Ind. Eng. Chem.* **52**, 391 (1960).
C2. Chou, Chan-Hui, *Ind. Eng. Chem.* **50**, 799 (1958).
C3. Cox, D. R., Proc. Berkeley Symp. Statist. **1**, 105 (1961).
C4. Cox, D. R., *J. Roy. Statist. Soc., Pt. B*, **24**, 406 (1962).
D1. Davies, O. L., ed., "Statistical Methods in Research and Production." Hafner, New York, 1958.
D2. Davies, O. L., ed., "The Design and Analysis of Industrial Experiments." Hafner, New York, 1960.
D3. Draper, N. R., and Hunter, W. G., *Biometrika* **54**, 376 (1967).
D4. Draper, N. R., and Smith, H., "Applied Regression Analysis." Wiley, New York, 1966.
F1. Franckaerts, J. F., and Froment, G. F., *Chem. Eng. Sci.* **19**, 807 (1964).
F2. Franklin, N. L., Pinchbeck, P. H., and Popper, F., *Trans. Inst. Chem. Eng.* **34**, 280 (1956).
F3. Franklin, N. L., Pinchbeck, P. H., and Popper, F., *Trans. Inst. Chem. Eng.* **36**, 259 (1958).
F4. Franks, R. G. E., "Mathematical Modeling in Chemical Engineering," Wiley, New York, 1967.
F5. Frost, A. A., and Pearson, R. G., "Kinetics and Mechanism." Wiley, New York, 1961.
G1. Guttman, I., and Meeter, D. A., Univ. of Wisconsin Dept. Statist. Tech. Rept. No. 37, (1964).
G2. Guttman, I., and Meeter, D. A., Univ. of Wisconsin Dept. Statist. Tech. Rept. No. 38 (1964).
H1. Hartley, H. O., *Technometrics* **3**, 269 (1961).
H2. Hill, W. J., Ph.D. Thesis, Univ. of Wisconsin, Madison, 1966.
H3. Hill, W. J., and Hunter, W. G., *Technometrics* **8**, 571 (1966).
H4. Hill, W. J., and Hunter, W. G., Univ. of Wisconsin Dept. Statist. Tech. Rept. No. 69, (1966).
H5. Hill, W. J., and Mezaki, R., *Am. Inst. Chem. Engrs. J.* (*A.I.Ch.E. J.*) **13**, 611 (1967).
H6. Himmelblau, D. M., Jones, C. R., and Bishoff, K. B., *Ind. Eng. Chem. Fundamentals* **6**, 539 (1967).
H7. Hinshelwood, C. N., and Burk, R. E., *J. Chem. Soc.* **127**, 1105 (1925).
H8. Hunter, W. G., *Ind. Eng. Chem. Fundamentals* **6**, 461 (1967).
H9. Hunter, W. G., and Atkinson, A. C., Univ. of Wisconsin Statist. Tech. Rept. No. 59 (1965).
H10. Hunter, W. G., and Hill, W. J., Univ. of Wisconsin Dept. Statist. Tech. Rept. No. 65 (1966).
H11. Hunter, W. G., and Mezaki, R., *A.I.Ch.E. J.*, **10**, 315 (1964).
H12. Hunter, W. G., and Mezaki, R., *Can. J. Chem. Eng.* **45**, 247 (1967).
H13. Hunter, W. G., and Reiner, A. M., *Technometrics* **7**, 307 (1965).
H14. Hunter, W. G., and Wichern, D. W., Univ. of Wisconsin Dept. Statist. Tech. Rept. No. 33 (1966).
J1. Johnson, R. A., Standel, N. A., and Mezaki, R., *Ind. Eng. Chem. Fundamentals* **7**, 181 (1968).
K1. Kabel, R. L., and Johanson, L. N., *A.I.Ch.E. J.* **8**, 621 (1962).
K2. Kittrell, J. R., Ph.D. Thesis, Univ. of Wisconsin, Madison, 1966.

K3. Kittrell, J. R., and Erjavec, John, *Ind. Eng. Chem. Proc. Design Develop.* **7**, No. 3 321 (1966).
K4. Kittrell, J. R., and Mezaki, R., *Ind. Eng. Chem.* **59**, No. 2, 28 (1967).
K5. Kittrell, J. R., and Mezaki, R., *A.I.Ch.E. J.* **13**, 389 (1967).
K6. Kittrell, J. R., and Watson, C. C., *Chem. Eng. Progr.* **62**, No. 4, 79 (1966).
K7. Kittrell, J. R., Hunter, W. G., and Watson, C. C., *A.I.Ch.E. J.* **11**, 1051 (1965).
K8. Kittrell, J. R. Mezaki, R., and Watson, C. C., *Ind. Eng. Chem.* **57**, No. 12, 19 (1965).
K9. Kittrell, J. R., Mezaki, R., and Watson, C. C., *Brit. Chem. Eng.* **11**, No. 1, 15 (1966).
K10. Kittrell, J. R., Mezaki, R., and Watson, C. C., *Ind. Eng. Chem.* **58**, No. 5, 51 (1966).
K11. Kittrell, J. R., Hunter, W. G., and Watson, C. C., *A.I.Ch.E.J.* **12**, 5 (1966).
K12. Kittrell, J. R., Hunter, W. G., and Mezaki, R., *A.I.Ch.E. J.* **12**, 1014 (1966).
L1. Laible, J. R., Ph.D. Thesis, Univ. of Wisconsin, Madison, 1959.
L2. Laidler, K. J., "Chemical Kinetics," 2nd ed. McGraw-Hill, New York, 1965.
L3. Lapidus, L., and Peterson, T. I., *A.I.Ch.E. J.* **11**, 891 (1965).
L4. Lapidus, L., and Peterson, T. I., *Chem. Eng. Sci.* **21**, 655 (1966).
L5. Levenburg, K., *Quart. Appl. Math.* **2**, 164 (1944).
M1. Marquardt, D. L., *J. Soc. Indust. Appl. Math.* **2**, 431 (1963).
M2. Marquardt, D. L., "Least Squares Estimation of Nonlinear Parameters," IBM SHARE Library Program No. 3094, Exhibit B.
M3. Mathur, G. P., and Thodos, G., *Chem. Eng. Sci.* **21**, 1191 (1966).
M4. Mezaki, R., and Butt, J. B., *Ind. Eng. Chem. Fundamentals* **7**, 120 (1968).
M5. Mezaki, R., and Kittrell, J. R., *Can. J. Chem. Eng.* **44**, No. 5, 285 (1966).
M6. Mezaki, R., and Kittrell, J. R., *A.I.Ch.E. J.* **13**, 176 (1967).
M7. Mezaki, R., and Kittrell, J. R., *Ind. Eng. Chem.* **59**, No. 5, 63 (1967).
M8. Mezaki, R., Kittrell, J. R., and Hill, W. J., *Ind. Eng. Chem.* **59**, No. 1, 93 (1967).
M9. Mickley, H. S., Sherwood, T.K., and Reed, C.E., "Applied Mathematics in Chemical Engineering," 2nd ed. McGraw-Hill, New York ,1957.
P1. Pannetier, G., and Davignon, L., *Bull. Soc. Chim. France* 2131 (1961).
P2. Peterson, T. I., *Chem. Eng. Sci.* **17**, 203 (1962).
P3. Pinchbeck, P. H., *Chem. Eng. Sci.* **6**, 105 (1957).
R1. Rogers, G. B., Lih, ,M.M. and Hougen, O.A., *A.I.Ch.E. J.* **12**, 369 (1966).
R2. Rosenbrock, H. H., and Storey, C., "Computational Techniques for Chemical Engineers." Macmillan (Pergamon), New York, 1966.
R3. Rudd, D. F., and Watson, C. C., "Strategy of Process Engineering," Wiley, New York, 1968.
S1. Shabaker, R. H., Ph.D. Thesis, Univ. of Wisconsin, Madison, 1965.
S2. Spang, H. A., *Soc. Indust. Appl. Math. Rev.* **4**, 343 (1962).
W1. Wei, J., and Prater, C. D., *Advan. Catalysis* **13**, 204 (1962).
W2. Weller, S., *A.I.Ch.E. J.* **2**, 59 (1956).
W3. White, R. R., and Churchill, S. W., *A.I.Ch.E. J.* **5**, 354 (1959).
W4. Williams, E. J., and Kloot, N. H., *Australian J. Appl. Sci.* **4**, 1 (1953).
W5. Williams, E. J., "Regression Analysis." Wiley, New York, 1959).
W6. Wu, S. M., Ermer, D. S., and Hill, W. J., *J. Eng. Ind. Ser. B* **88**, 81 (1966).
Y1. Yang, K. H., and Hougen, O. A., *Chem. Eng. Progr.* **46**, 146, (1950).