Statistical Techniques in Modeling of Complex Systems: Single and Multiresponse Models

S. SITHARAMA IYENGAR AND MUSTI S. RAO

Abstract—An exposition of statistical techniques in modeling complex systems (single and multiresponse models) that are representative of recent work on modeling systems is provided. The paper begins with several basic concepts related to linear and nonlinear models. We then examine four representative techniques of model discrimination which deal with use of nonintrinsic and intrinsic parameters, use of Bayesian methods, and likelihood discrimination. Next we examine multiresponse models with issues dealing with design of experiments for parameter estimation and model discrimination. A case study on sequential model discrimination in multiresponse models is also discussed. Finally an overview on estimating parameters in models of dynamical system is briefly discussed. The paper concludes with a summary of unresolved issues, and with suggestions on the future role of modeling in the complex situation.

I. Introduction

IN RECENT YEARS the importance of modeling of systems has attracted a considerable deal of interest from researchers with various interests. For example automatic control theory, biology, economics, and medicine are samples of fields where model building and analysis plays a central role. The basic aim of the study of a chemical, physical, or biological system is to investigate how it behaves so as to make recommendations for its future development. The most common means of doing this is to build a mathematical model for the system through which it becomes possible to predict, control, and optimize.

The modeling of any phenomenon is quite an involved task. It generally consists of assuming certain plausible models (which may be linear, nonlinear, or mechanistic), estimating the parameters of these models in the light of some experimental data, discriminating among the rival models, and ultimately checking the adequacy of the proposed model. The state of the art has been greatly advanced in the last decade, and is mainly due to the use of large digital computing facilities coupled with the advance of the state of mathematical system theory and statistical techniques. This paper provides an exposition of statistical techniques in modeling complex systems that are representative of recent work on modeling systems. The remainder of the paper is organized as follows.

Manuscript received November 23, 1981; revised June 24, 1982 and October 6, 1982.

S. S. Iyengar is with the Department of Computer Science, Louisiana State University, Baton Rouge, LA 70803.

M. S. Rao was with the Department of Chemical Engineering, University of Florida, Gainesville, FL 32611. He is now with the Indian Institute of Technology, Kanpur, India.

The general modeling strategy is considered in Section I. The problem of estimation of parameters in linear and nonlinear single response models is considered in Section II and III, respectively. Various model discrimination techniques for single-response models are discussed in Section IV. Once the most plausible model has been obtained from a set of competing models, it is necessary to validate the proposed model. Lack-of-fit F tests and residual analysis are used for this purpose. These tests of model adequacy are discussed in Section V. When discrimination cannot be achieved from the existing data, it becomes necessary for the experimenter to design further experiments that provide sharp discrimination. In order to achieve this goal using minimum amount of experimental effort, these experiments are conducted sequentially. The design strategy for model discrimination of single-response models is presented in Section VI. Once the final model has been obtained the goal of the experimenter is to obtain precise estimates of these parameters that will minimize their joint confidence region. The design strategy for the precise estimation of parameters in single-response models is presented in Section VII.

The analogous techniques of parameter estimation, model discrimination, and designs of experiments for multiresponse models are discussed in Sections VIII, IX, X, and XI, respectively. The paper concludes with a case study on sequential model discrimination in multiresponse systems in Section XII. Finally a very brief overview on dynamical systems is presented.

A. Mechanisms and Models

For any system it may be assumed that there exists a precise mathematical and physical representation of all phenomena that make up the system. In many situations, however, *a priori* it may not be known what this model is, and in fact the first goal of the experimenter is to obtain this relationship which is normally referred to as the model.

For example the biological oxygen demand (BOD), which is used a measure of the pollution produced by domestic and industrial wastes, may be given by an exponential model of the form [1]

$$\eta = \theta_1(1 - \exp - \theta_2 t), \tag{1}$$

where η is the BOD and t is the incubation period.

0018-9472/83/0300-0175$01.00

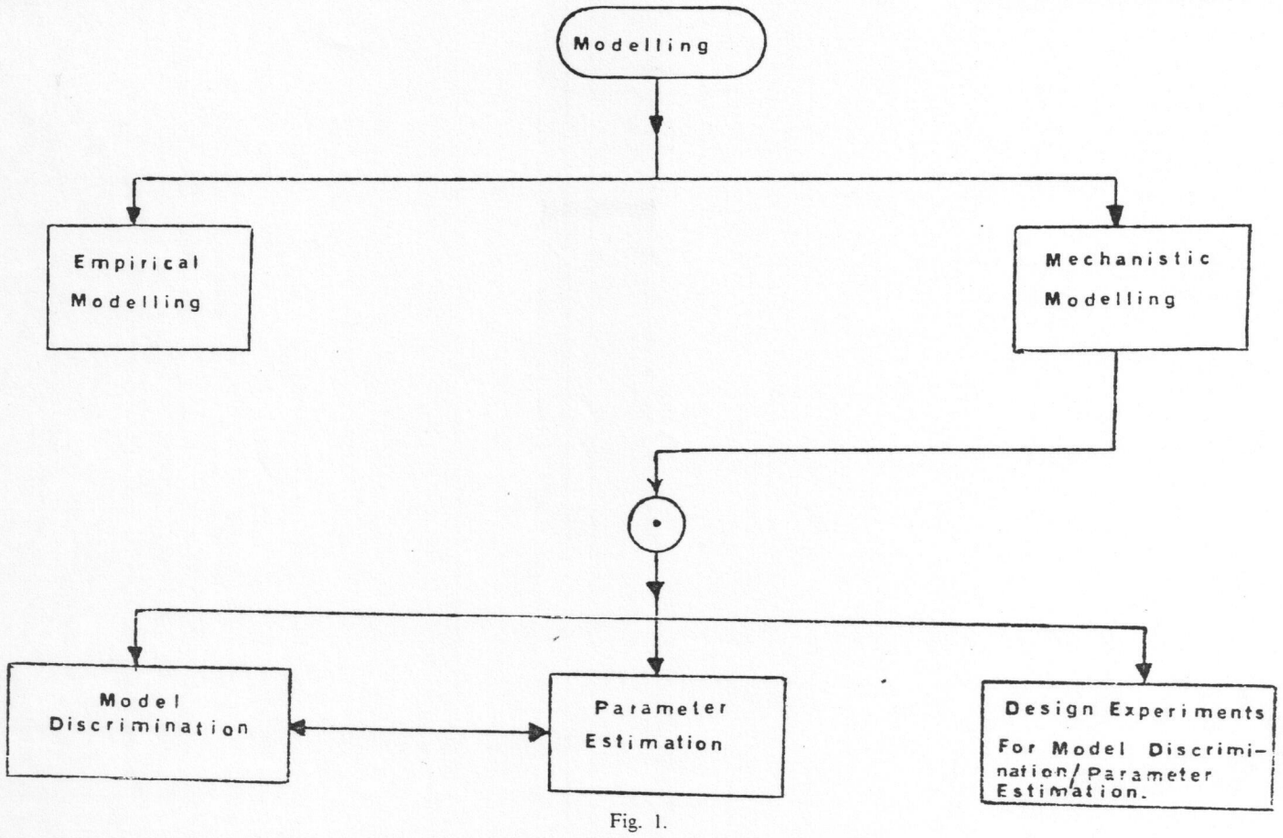

Fig. 1.

Another example of a model is the expression for the rate of a chemical reaction between two species A and B, which may be represented by

$$r_a = -dC_a/dt = kC_a^m C_b^n, \tag{2}$$

where r_a is the rate of consumption of the species A, k is the rate constant of the reaction, C_a and C_b are the concentrations of the species A and B, t is time, and m and n are the reaction orders.

Equation (2) is a power-law model in which the reaction rate r_a is the dependent variable and the concentrations C_a and C_b are the controllable quantities. However, in general, all phenomena can be theoretically represented by a mathematical model of the form

$$\eta = f(\theta, \xi), \tag{3}$$

where η is the expected value of the dependent variable y, θ is a (pxl) vector of parameters, and ξ is a (pxl) vector of independent variables.

It is frequently possible to postulate several physically meaningful models which can describe the same system. The investigator is then faced with the dual problem of choosing the best among the rival models and obtaining the best estimates (in some statistical sense) of the parameters involved in the selected model [2], [3]. The first stage, in which the precise mathematical relationship applicable to the system is identified, is known as the specification stage, and the second one, in which precise estimates of the parameters are obtained, is known as the estimation stage. Both these stages constitute the important goals of modeling.

B. Modeling Strategy

In some cases, when an experimenter starts with an object of modeling either the whole system or a part of it, he may have some knowledge about the possible mechanism. In others he may not know anything about the system. In most practical cases, normally he may have a partial knowledge about the system. If he does not have any knowledge about the system he may resort to purely empirical modeling, while if he has complete information about the system he may directly proceed with the estimation stage. In most cases, however, he may be in between these two broad aspects which leads to the so-called mechanistic modeling. The modeling strategy is summarized in Fig. 1 [1]. The various modeling strategies are further discussed below.

1) Empirical Models: Though one would like to find the "real world" applicable to a physical situation, invariably it may be an elusive goal. Under these conditions, one is left with two choices—viz., empirical modeling and mechanistic modeling. These empirical models may be polynomials in the independent variables. A typical empirical model may be represented by [4]

$$y_i = \theta_0 + \theta_1\xi_{1i} + \theta_2\xi_{2i} + \cdots + \theta_p\xi_{pi} + \theta_{11}\xi_{1i}^2 + \theta_{22}\xi_{2i}^2 + \cdots + \theta_{pp}\xi_{pi}^2 + \epsilon_i, \tag{4}$$

where ϵ is the error associated with a measurement, θ and ξ refer to the corresponding parameters and independent variables, respectively, and the subscript i refers to the ith measurement. Empirical models are arbitrarily chosen based on the apparent functional relationship of the response to the independent variables.

Generally, either polynomials of the type given by (4) or power-law models given by (2) are used to represent empirical models. These have no relation whatsoever, to the true mechanism of the system under consideration. Whenever the phenomenon under consideration is very complex, empirical models give useful guidance predicting the response within the range of experimentation.

2) Mechanistic Models: A mechanistic model is a mathematical relationship between the response (dependent variable) and the independent variables derived from a consideration of a plausible mechanism. For example consider a simple heterogeneously catalyzed surface reaction.

$$A.s + B.s \rightarrow C.s + s, \tag{5}$$

where $A.s$, $B.s$, and $C.s$ represent the adsorbed species A, B, and C, respectively, and s denotes a vacant site. If surface reaction is rate controlling, the model may be represented by

$$y = \frac{\theta_1 \xi_1 \xi_2}{(1 + \theta_2 \xi_1 + \theta_3 \xi_2)^2}, \tag{6}$$

where y is the rate of reaction (dependent variable) and ξ_1 and ξ_2 are the gas-phase concentrations of the species A and B, respectively.

Mechanistic modeling should be used judiciously. It is justified when the state of art is sufficiently advanced to formulate a useful mechanistic model. Judgment is needed in deciding when and when not to use mechanistic models [1]. Two important aspects of mechanistic modeling consist of parameter estimation and model discrimination. Consider a set of observations y and the corresponding values of an independent variable ξ available on a particular system. The experimental observations of a hypothetical situation are shown on the y versus ξ plot in Fig. 2. Also shown on the graph are two different models arising out of different mechanistic considerations. From the graph it is evident that either Model 1 or 2 describes the system under consideration. It is seen from the graph that as long as the experimental data are obtained in the region $0 \leqslant \xi \leqslant \xi_1$, it is difficult to say which of the two models governs the system. If the goal were to discriminate among these two models, the observations should be taken at higher values of the independent variable ξ. Thus the "design of experiments" plays a dominant role in the model of discrimination problem and its importance has been recognized for a long time [3]–[14]. Parameter estimation, model discrimination, and design of experiments for either model discrimination or parameter estimation constitute the important components of modeling. These topics are reviewed in the following sections.

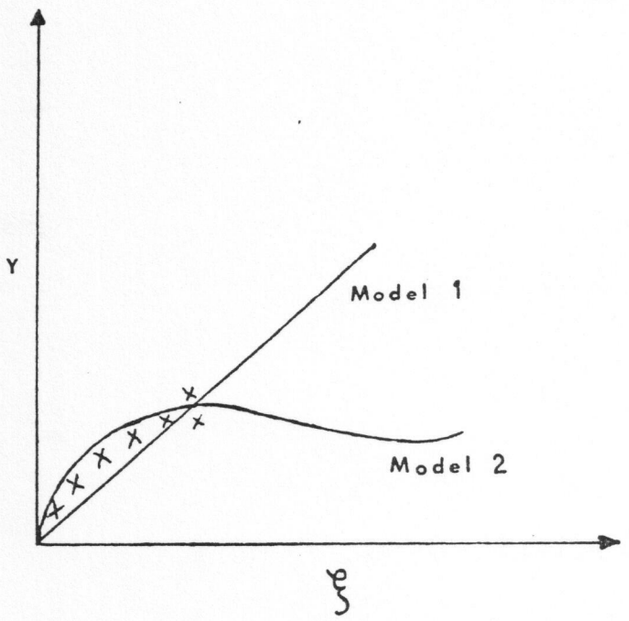

Fig. 2.

II. Estimation of Parameters in Linear Models

A. Definitions

A model can be either linear or nonlinear. If the partial derivative of the response function with respect to any parameter does not contain the parameters it is known as a linear model. For example the model

$$y = \theta_1 \xi_1^m + \theta_2 \xi_2^n + \epsilon \tag{7}$$

is a linear model in the parameters θ_1, θ_2. The constraints $m < 2$ and $n < 2$ must be recognized for linearity of the model described in (7). On the other hand the model of the following form is a nonlinear model.

$$y = \theta_1 \exp(-\theta_2 \xi_1 \xi_2) + \epsilon. \tag{8}$$

The technique to be employed for the estimation of parameters in any model depends on the nature of the model and it error structure. The parameter estimation techniques are reviewed briefly in the following sections. For details of the methods see [1]–[3], [15]–[17].

B. Linear Regression

1) Estimation of Parameters: It is assumed that there are n measurements y_i made on a system where the true value of the quantity which is measured by y_i is called η_i, and may be perfectly represented by

$$\eta_i = \theta_1 \xi_{i1} + \theta_2 \xi_{i2} + \cdots + \theta_p \xi_{ip}, \qquad p \leqslant n. \tag{9}$$

The quantities ξ_{ij}, $i = 1, \cdots, n$ and $j = 1, 2, \cdots, p$ are assumed to be known perfectly. From (9) it follows that

$$y_i = \theta_1 \xi_{i1} + \theta_2 \xi_{i2} + \cdots + \theta_p \xi_{ip} + \epsilon_i, \qquad i = 1, 2, \cdots, n, \tag{10}$$

where ξ_{ij} is the value of the jth independent variable

corresponding to the ith measurement. As a further assumption it is assumed that the errors are uniformly and independently distributed. Thus the covariance matrix of the errors may be given by

$$V(\epsilon) = I_{(n\times n)}\sigma^2, \tag{11}$$

where I is an $(n \times n)$ identity matrix and σ^2 is the unknown variance.

The least-squares estimates of the parameters ($\underline{\hat{\theta}}$) are those values of the parameters which minimize the sum of squares of the differences between the observations and the predictions. These estimators of the parameters, also known as BLUE (best linear unbiased estimators), can be given by

$$\underline{\theta} = (\underline{\xi}'\underline{\xi})^{-1}(\underline{\xi}'\underline{y}), \tag{12}$$

where $\underline{\xi}$ is given by

$$\underline{\xi} = \begin{bmatrix} \xi_{11} & \xi_{12} & \cdots\cdots & \xi_{1p} \\ \xi_{21} & \xi_{22} & \cdots\cdots & \xi_{2p} \\ \cdot & \cdot & \cdots\cdots & \\ \cdot & \cdot & \cdots\cdots & \\ \xi_{n1} & \xi_{n2} & \cdots\cdots & \xi_{np} \end{bmatrix} \tag{13}$$

and

$$\underline{y} = (y_1 y_2 \cdots, y_n)'. \tag{14}$$

2) Confidence Intervals and Confidence Regions: For a linear model of the type given by (10), the $(1 - \alpha/2)$ 100 percent confidence interval for the estimate of θ_i is given by [1], [2], [15]

$$\theta_i \pm t_{v,\alpha/2}[v(\theta_i)]^{1/2},$$

where $t_{v,\alpha/2}$ is the $(1 - \alpha/2)$ 100 percent point of the t distribution with v degrees of freedom.

If the errors in the response are assumed to be normally and independently distributed with constant variance σ^2, the quantity $v(\theta_i)$ is the ith diagonal element of the matrix $(\underline{\xi}'\underline{\xi})^{-1}\sigma^2$. If the parameter estimates are correlated, then the joint $(1 - \alpha)$ 100 percent confidence region is given by [1], [2], [15]

$$(\underline{\theta} - \underline{\hat{\theta}})'\underline{\xi}'\underline{\xi}(\theta - \underline{\hat{\theta}}) = s^2 pF_\alpha(p, v), \tag{15}$$

where $\underline{\theta}$ is a vector of parameter values which is being estimated with linear regression theory by $\underline{\hat{\theta}}$, s^2 is an unbiased independent estimate of experimental error variable σ^2, p is the number of parameters, and $F_\alpha(p, v)$ is the critical value of F at a significance level of α.

III. Estimation of Parameters—Nonlinear Models

Invariably, in many practical situations one encounters nonlinear models. For example, for the gaseous dehydration of ethanol on a resin catalyst, Kabel and Johanson [18] gave the rate expression as

$$r = \frac{kK_a\left(P_a^2 - P_e P_w / K_{\text{eq}}\right)}{\left(1 + K_a P_a + K_w P_w + K_e P_e\right)^2}. \tag{16}$$

In this expression r is the measured (dependent) variable, P_a, P_e, and P_w are the controllable quantities (independent variables), and k, K_a, K_e, K_w, and K_{eq} are the parameters to be estimated.

For scientific reasons the parameters in this nonlinear model should be estimated using nonlinear regression. In estimating the parameters in a linear model, one sets the partial derivatives of the residual sum of squares (RSS) with respect to each parameter to zero, and solves the resulting equations simultaneously. However in the case of nonlinear models, such equations cannot be easily solved. An iterative solution has been suggested [15], [19] for estimating the parameters in a nonlinear model. The procedure consists of first expanding the model by a Taylor's series expansion, retaining the linear terms, and then solving for the parameters by linear regression analysis. By retaining the linear terms in the Taylor's series expansion, the model may be structured as follows:

$$\eta_u = \eta_{u0} + \sum_{i=1}^{p} \left.\frac{\partial f(\underline{\xi}_u, \underline{\theta})}{\partial \theta_i}\right|_{\underline{\theta}=\underline{\theta}^0} \left(\theta_i - \theta_i^0\right). \tag{17}$$

Including experimental error (17) may be written as

$$z_u = \sum_{i=1}^{p} f'_{iu}\delta_i + \epsilon_u, \tag{18}$$

where

$$z_u = y_u - f(\underline{\xi}_u, \underline{\theta})|_{\underline{\theta}=\underline{\theta}_0} + \epsilon \tag{19}$$

$$f'_{iu} = \left.\frac{\partial f(\underline{\xi}_u, \underline{\theta})}{\partial \theta_i}\right|_{\underline{\theta}=\underline{\theta}^0} \tag{20}$$

and

$$\delta_i = \theta_i - \theta_i^0. \tag{21}$$

Since (18) is linear, the correction vector δ_i can be obtained by fundamentally minimizing $\delta(\theta)$.

The improved parameter estimates for the next trial are given by[1]

$$\theta_i^{(1)} = \theta_i^0 + \delta_i. \tag{22}$$

This procedure is iterated until the corrections δ_i become exceedingly small.

One can also apply the steepest descent procedures to determine parameter values [3]. This consists of setting up a first-order design in the parameter space about a set of parameter estimates, calculating the direction of steepest descent, obtaining the minimum, and using a second-order design for the precise location of the minimum. This approach has been found to converge for nearly any set of initial parameter estimates, but its convergence can be agonizingly slow. Marquardt [20] has suggested a compromise between these two primary methods which finds extensive application in estimating parameters in a nonlinear model.

[1]The Taylor series converges when the model solutions are sufficiently close to the true solution.

It is very important to note that a proper choice of the initial parameter estimates is important in the estimation procedure. Intrinsically nonlinear models may be linearized to obtain initial estimates of the parameters. In some situations one might have prior knowledge of the probable values of the parameters which may be used as initial estimates. Sequential simplex method has been successfully used for estimating the parameters in nonlinear models. This method is opportunistic and it converges even when the initial simplex straddles two or more ridges [21]–[23].

A derivative-free algorithm (DUD) has been developed by Ralston and Jennrich for fitting models defined by systems of nonlinear differential equations. On a number of test problems, DUD has been claimed to compete favorably with even the best derivative-based algorithms [24].

IV. Model Discrimination—Single Response Models

The problem of discriminating among a set of competing models is that of choosing the one, hopefully unique, that gives predictions which in some sense are better than those given by the others. Model discrimination is achieved by

1) use of nonintrinsic parameters, which are parameters introduced into a model for the purpose of discrimination from a set of rival models;
2) use of intrinsic parameters, i.e., parameters which are inherently present in the model;
3) use of Bayesian methods; and
4) use of non-Bayesian methods.

Several reviews [2], [4], [17], [25] have appeared on the model-discrimination aspect, the salient features of which are discussed briefly in the following sections.

A. Use of Nonintrinsic Parameters

Consider the discrimination between two rival Models 1 and 2 given by

$$\eta_1 = f(\underline{\theta}, \underline{\xi}) \tag{23}$$

and

$$\eta_2 = g(\underline{\phi}, \underline{\xi}).$$

The discrimination between these two models (described by (23)) can be achieved by considering an equation of the type

$$y - \frac{1}{2}(\eta_A + \eta_B) = \lambda(\eta_B - \eta_A). \tag{24}$$

Here λ is a nonintrinsic parameter since it was not originally present in the model. In order to discriminate between the two rival models (24) is regressed for λ. As seen from (24), Model 1 is the proper one if λ were equal to $-1/2$ and Model 2 is the proper one if λ were equal to $1/2$. In practice λ may not be exactly equal to either $+1/2$ or $-1/2$. In such cases the confidence limits of λ should be assessed to see whether it included $+1/2$ or $-1/2$, or both. If λ were equal to $-1/2$ in the desired confidence region, then Model 1 is the correct one. On the contrary, if λ were equal to $+1/2$ in the desired confidence region, then Model 2 is the correct one. The acceptable deviation depends on the desired degree of confidence. Normally 95 percent confidence limits should suffice in most engineering situations.

Mezaki and Kittrell [26] successfully used this model for discriminating between two models for the vapor-phase dehydration of secondary butyl alcohol to the olefin on a commercial cracking catalyst.

A somewhat analogous method consists of regressing the equation

$$y = (1 - \lambda)\eta_1 + \lambda\eta_2 \tag{25}$$

for λ. If λ were equal to 1, then Model 2 is the correct one. One the other hand, if λ were equal to 0, then Model 1 is the correct one. As before it is recommended that the value of λ should be considered in its confidence range.

B. Use of Intrinsic Parameters

Intrinsic parameters, which are inherently present in the model and of a simpler functional form than the entire model, facilitate the experimenter's ability to test the adequacy of a proposed model. Kittrell and Mezaki [2], [27] used this approach for proposing a model for the olefinic dehydration of a pure alcohol feed to a reactor. For details of this method see Kittrell and Mezaki's original article [27].

C. Use of Bayesian Methods

The Bayes' theorem provides a useful means of discriminating among rival models. Bayes' theorem states [28]

$$P(A_i/B) = \frac{P(A_i)P(B/A_i)}{\sum_{i=1}^{r} P(B/A_i)P(A_i)}, \tag{26}$$

where A_i, $(i = 1, \cdots, r)$ denotes the ith model, B denotes the data, $P(A_i)$ denotes the prior probability of the ith model, and $P(B/A_i)$ denotes the likelihood for the ith model.

In order to use the Bayes' theorem for model discrimination, one needs the values of prior probabilities of various models as seen in (26). When one does not have any strong preference for a particular model, he may assign equal probabilities to all the models.

The calculation of likelihood requires knowledge of the error structure. If one assumes that the errors are normally and independently distributed with a variance σ^2, the probability density function (pdf) of n observations $y_1, y_2, \cdots, y_n$ is given by [14]

$$\text{pdf} = \frac{1}{(\sqrt{2\pi}\,\sigma)^n} \exp - \frac{1}{2\sigma^2} \sum_{u=1}^{n} (y_u - f(\underline{\theta}, \underline{\xi}_u))^2. \tag{27}$$

TABLE I
NUMERICAL DATA FOR MODEL FITTING

u	ξ_u	y_u
1	0	-1.290
2	1	5.318
3	2	7.049
4	3	19.886

On the other hand, if one assumes that the errors are normally distributed but correlated with a variance–covariance matrix $\underline{V}\sigma^2$, then the pdf of n observations $y_1, y_2, \cdots, y_n$ is given by

$$\text{pdf} = \frac{1}{(\sqrt{2\pi}\,\sigma)^n \sqrt{|v|}} \cdot \exp - \frac{1}{2\sigma^2}\left[(\underline{y} - f(\underline{\theta}, \underline{\xi}))' \underline{V}^{-1} (\underline{y} - f(\underline{\theta}, \underline{\xi}))\right]. \quad (28)$$

Once the observations ($\underline{y}$) and the design matrix ($\underline{\xi}$) are substituted into one of the above expressions for the pdf, the resulting expression, which is a function of $\underline{\theta}$ and σ^2, is called the likelihood.

Example of Model Discrimination: Reilly [4] considered a simple numerical example of discriminating among rival models. Assuming ξ as a single independent variable, y_u as the dependent variable, the set of numerical data shown in Table I were used for discriminating among rival models.

The following three models were considered by Reilly for discrimination purposes

$$\begin{aligned} &\text{Model 1:} \quad y_u = \theta_{11}\xi_u + \epsilon_u, \\ &\text{Model 2:} \quad y_u = \theta_{21} + \theta_{22}\xi_u + \epsilon_u, \\ &\text{Model 3:} \quad y_u = \theta_{31} \exp \theta_{32}\xi_u + \epsilon_u. \end{aligned} \quad (29)$$

It is assumed that one of these three models represents the experimental data adequately. The errors were assumed to be normally and independently distributed. The probability density function, defined by (27), is applicable for this case. Using the least-squares criterion, the parameters were estimated and the maximum likelihoods were calculated. Assuming equal prior probabilities, the posterior probabilities were calculated for each model as shown in Table II. From the posterior probabilities it is seen that Model 3 is preferred. This simple example illustrates how the Bayesian approach may be used in model discrimination. Several investigators [7], [13], [29], [30] used Bayesian methods in model discrimination successfully.

D. Likelihood Discrimination

Likelihood discrimination shows the flexibility of Bayesian methods combined with an ability to "let the data speak for themselves" [4]. To illustrate this approach assume the ith model to be represented by the functional relationship

$$\eta_{iu} = f_i(\underline{\theta}, \underline{\xi}_u). \quad (30)$$

TABLE II
POSTERIOR PROBABILITIES OF MODELS

Model	Prior Probability	Likelihood	Posterior Probability
1	0.3333	0.05	0.0002
2	0.3333	1	0.0049
3	0.3333	202.2	0.9948

Let $\underline{y}$ denote the vector of observations. Also assume that the observations are correlated with a variance–covariance matrix $\underline{V}\sigma^2$. Assume that n observations are available. Under these assumptions the probability density function is given by (28).

The maximum of the likelihood function (maximum w.r. to σ^2) for the ith model can be shown to be given by [23]

$$L_i(\underline{\theta}, \sigma^2) = \frac{n^{n/2}}{(2\pi)^{n/2} M_i^{n/2} \sqrt{|\underline{V}|}} \exp(-n/2), \quad (31)$$

where M_i is the weighted sum of squares $(\underline{y} - \underline{\eta}_i)' \underline{V}^{-1} (\underline{y} - \underline{\eta}_i)$. The ratio of maximum likelihood for two models i and j, also known as the likelihood ratio, is a comparison of how well the two models can be made to fit the data.

For the above situation, then, one can show that the maximum likelihood ratio is given by

$$\lambda = (L_i)_{\max} / (L_j)_{\max} = (M_j / M_i)^{n/2}. \quad (32)$$

This ratio λ denotes the likelihood odds of Model i versus Model j.

The likelihood ratio method of discriminating among several rival models comprises of finding the likelihood ratio between the best model (having minimum weighted sum of squares) and the other models taken one at a time. Thus two-way comparisons are made by examining these ratios. This method (using the ratio of the maximum likelihood) is not the only way of discriminating nuisance parameters in likelihood inference.

Rao *et al.* [23] and Reilly [4] have clearly demonstrated the use of maximum likelihood ratio as a statistical criterion for discriminating among rival kinetic models. Even with a moderate number of experimental observations, the likelihood ratio appeared to be a powerful one for discriminating among a large number of competing models. There are inherent difficulties with any discrimination method when the models have different numbers of parameters. For more on this see Reilly [4] and Rao *et al.* [23].

V. TESTS OF MODEL ADEQUACY

Once the most plausible model has been obtained from a set of competing models it is necessary to test for the adequacy of the model. This is usually accomplished by lack-of-fit F tests and also by a residual analysis. Tests of model validation has been discussed by several investigators (e.g., [1], [3], [15], [31]–[33]) the basic principles of which will be discussed below.

A. Lack of fit F tests

The F statistic in model fitting may be defined by:

$$\frac{\text{lack of fit mean square}}{\text{pure error mean square}}.^{2}$$

This ratio is compared to the critical value of F at the required degree of confidence and the corresponding degrees of freedom. If the above ratio is greater than F_{crit}, then the model is inadequate. For a detailed treatment on lack-of-fit F tests, see [1], [15].

B. Residual Analysis

A residual is defined as the difference between the observed and the predicted values of a response (i.e., y-$\hat{\eta}$). The residual analysis can be applied either to the whole model as it is, or in some other cases (like in diagnostic checkup in model building) on some estimated parameters of a tentatively entertained theoretical model, so as to pinpoint the inadequacies.

As an illustration assume that the true model of a set of observations is given by

$$y_1 = \theta_0 + \theta_1\xi_1 + \theta_2\xi_2 + \epsilon. \tag{33}$$

Assume that the fit is given by

$$\hat{y}_1 = \hat{\theta}_0 + \hat{\theta}_1\xi_1. \tag{34}$$

It will be seen that the residual $y_1 - \hat{y}_1$ is correlated with the variable ξ_2 in a linear way. Thus if the residual is correlated with respect to ξ_2, it is indicative that the model should include a term consisting of ξ_2. Once this linear term is included and a residual analysis is performed, then the residuals should be random if the model were adequate. Time plots of residuals, as well as plots of residuals with respect to various controllable variables can detect possible model inadequacies, which can throw light on how to improve the model. This is the principle in adaptive model building. Kittrell *et al.* [34] have further demonstrated the use of diagnostic parameters for model building of chemical reaction rate models. For a more detailed treatment of residual analysis see [2], [15].

VI. Design of Experiments for Model Discrimination—Single Response Models

In the past, for kinetic modeling the "one-factor-at-a-time" method [35], in which the experimental factors are varied one at a time, with the remaining factors held constant, has been used. However this method of experimentation is found to be ineffective especially when there is interaction among the various factors under consideration.

When discrimination cannot be achieved from the existing data, it becomes necessary for the experimenter to design further experiments that provide maximum discrimination. It might be expected that all one has to do is to plan the future experiments carefully, run them, and then it may be expected that one of the models would emerge as the best one. But this is not the case in general. Different researchers claim different models for the same phenomenon. A classic example is the water–gas shift reaction [36] for which several models have been proposed. The reason for the discrepancy is that the "models are not put in jeopardy." In the past, investigators took the difference in response given by the models for choosing a measure of discrimination.

One of the very early methods of design of experiments for discriminating between two rival models is that of Hunter and Reiner [6], who suggested maximizing the design criterion

$$D = \left(\hat{f}_{N+1} - \hat{g}^{N}_{N+1}\right), \tag{35}$$

where f_{N+1} and g_{N+1} denote the predicted responses for the two rival models for the $(N+1)$th experimental trial (to be conducted) using the best estimates of parameters obtained after conducting N experimental trials.

When the number of rival models is greater than two, Roth [13] suggested a criterion that involved choosing the experimental points that maximized the product of absolute differences in the predicted values of the response. Roth's criterion is given by

$$C_{ji} = \prod_{\substack{k=1 \\ k \neq i}}^{m} \left(\eta_j^{(k)} - \eta_j^{(i)}\right). \tag{36}$$

The subscript j corresponds to the values of the independent variables ξ_j. The design criterion to be maximized is the weighted average of the spread among responses for the proposed models, weighted according to their probabilities, i.e.,[3]

$$D = \Sigma_i P_i^{(n)} C_{jy}. \tag{37}$$

Both Hunter and Reiner's method and Roth's method ignore the uncertainties associated with each model. To alleviate this difficulty, Box and Hill [7] proposed an excellent method for discriminating between m $(m > 2)$ rival models. Their method takes account of not only the difference in response given by the models to be discriminated among, but also the variance of the estimated response. The design criterion to be maximized is given by

[2] If $x_1, x_2, \cdots, x_m$ and $y_1, y_2, \cdots, y_m$ are two independent random samples of data from two different normal populations, then the F distribution is

$$F = \frac{\sum_{i=1}^{m} (x_i - \bar{x})^2 (m-1)\sigma^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2 (n-1)\sigma_2^2}$$

with the degrees of freedom = $(m-1, n-1)$.

[3] The probability $P_i^{(n)}$ of any model i denotes our degree of belief in that particular model after conducting n experiments.

Box and Hill as

$$D = \frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\pi_{in}\pi_{jn}\left\{\frac{\left(\sigma_i^2-\sigma_j^2\right)}{\left(\sigma^2+\sigma_i^2\right)\left(\sigma^2+\sigma_j^2\right)}+\left(\hat{y}_{n+1}^{(i)}-y_{n+1}^{(j)}\right)^2\left[\frac{1}{\left(\sigma^2+\sigma_i^2\right)}+\frac{1}{\left(\sigma^2+\sigma_j^2\right)}\right]\right\}. \quad (38)$$

Here σ^2 is the common variance of the n observations $y_1, y_2, \cdots, y_n$, σ_i^2 and σ_j^2 are the variance for the predicted values of y_{n+1} under the models i and j, respectively, and $\hat{y}_{n+1}^{(i)}$ and $\hat{y}_{n+1}^{(j)}$ are the predicted values of y_{n+1} under models i and j, respectively. Under certain assumptions, the pdf was assumed to be given by

$$\text{pdf} = \frac{1}{\sqrt{2\pi\left(\sigma^2+\sigma_i^2\right)}}\exp-\frac{1}{2\left(\sigma^2+\sigma_i^2\right)}\cdot\left[\left(y_{n+1}-y_{n+1}^{(i)}\right)^2\right]. \quad (39)$$

The above criterion is obtained by maximizing an upper bound on the expected change in the entropy.

Hill and Hunter [37] extended the criterion to the case where σ^2 is not known. In the procedure suggested by Box and Hill, the design and analysis are carried out simultaneously, and the stopping rule for experimentation is to stop when the posterior probabilities indicate that one model is clearly superior to the rest. Several studies [9], [14], [38] revealed that the model probabilities may oscillate considerably, and Hill [39] suggests that the Box and Hill's criterion must be applied very cautiously and the model should not be accepted too readily based on a small number of discriminating experiments.

Fedorov and his coworkers [40], [41] advocated a criterion for the design of experiments for model discrimination, which is essentially a formalization of the intuitive idea of Hunter and Reiner. Atkinson and Cox [42] proposed another criterion for discriminating between rival models which is similar to the one proposed by Fedorov and coworkers.

Hsiang and Reilly [10] adopted a Bayesian procedure for model discrimination which eliminated some of the stringent requirements for Box and Hill's procedure. However their method seems to require excessive computer storage space for problems involving many parameters. The existing methods of experimental designs for discriminating between rival regression models have been reviewed recently by Hill [39].

VII. Design of Experiments for Estimation of Parameters—Single Response Models

Suppose that after suitable model-building experiments have been carried out, a given model has been singled out as being adequate. Also assume that the space of the experimental variables is limited to some particular region of experimentation. Suppose it is desired to obtain estimates of the parameters θ_i evaluated at the uth set of experimental conditions and taken at some set of parameter values $\underline{\theta}_0$, i.e.,

$$f_{iu} = \partial f(\underline{\theta}, \underline{\xi}_u)/\partial\theta_i|_{\underline{\theta}=\underline{\theta}_0}. \quad (40)$$

The matrix of these derivatives, which will contain N rows (observations) and p columns (parameters), may be written

$$\boldsymbol{F} = \{f_{iu}\}. \quad (41)$$

Box and Lucas [11] have indicated that, under certain plausible assumptions, a choice of experimental points which will maximize $|\underline{F}'\underline{F}|$ will also be that choice of data points which will minimize the volume of the joint confidence region of parameters. $\sqrt{|\underline{F}'\underline{F}|}$ is inversely proportional to the volume of the joint confidence region.

Kittrell *et al.* [43] applied the above method for obtaining the precise estimates of parameters in the model for the reduction of nitric oxide, data for which were obtained by Ayen and Peters [44], and found that for the same number of data points the parameters in the model can be estimated 18 times more precisely than by another commonly used one-factor-at-a-time design.

Certain aspects of sequential design procedures for precise parameter estimation were discussed by Hosten and Emig [4], [45]. Box and Wilson [46] discussed certain aspects of design of experiments. The dual problem of model discrimination and parameter estimation has been discussed by Borth [47].

VIII. Parameters Estimation–Multiresponse Models

In some cases, for a given set of experimental conditions, not one but a number of responses can be measured in a process. There are numerous examples of such systems where several responses are obtained. The use of multiresponse techniques increases the precision and accuracy of the parameter estimates and decreases the volume of the joint confidence ellipsoid. Singh and Rao [48] reviewed parameter estimation and model discrimination in multiresponse models in a recent review.

A. Mathematical Formulation

A multiresponse model can be denoted by

$$\boldsymbol{Y} = \underline{\eta}(\underline{\theta}, \underline{\xi}) + \underline{\epsilon}, \quad (42)$$

where η represents the true value of the response Y, ϵ corresponds to the error associated with the measurement, $\underline{\theta}$ is a $(p \times 1)$ vector of parameters, and $\underline{\xi}_u$ is an $(s \times 1)$ vector of controllable variables.

More explicitly the ith response $(1 \leqslant i \leqslant r)$ for the uth experiment $(1 \leqslant u \leqslant n)$ may be denoted by

$$y_u^{(i)} = \eta_u^{(i)}(\underline{\theta}, \underline{\xi}_u) + \epsilon_u^{(i)}. \quad (43)$$

It is assumed that the errors are such that

$$\begin{aligned} E\left(\epsilon_u^{(i)}\right) &= 0, && \text{for all } i, u; \\ E\left(\epsilon_u^{(i)}\epsilon_v^{(j)}\right) &= 0, && \text{for all } i, j, u \neq v; \\ E\left(\epsilon_u^{(i)}\epsilon_u^{(j)}\right) &= \sigma_{ij}, && \text{for } i, j, u. \end{aligned} \quad (44)$$

The vector $\underline{y}_u = (y_u^{(1)}, y_u^{(2)}, \cdots, y_u^{(r)})'$ of r response for the uth experiment has a symmetric covariance matrix Σ given by

$$\Sigma = \{\sigma_{ij}\}, \qquad i = 1,2 \cdots, r; j = 1,2 \cdots, r. \quad (45)$$

Just as before, the problem before us is to discriminate among rival models of the type given by (42), and secondly to estimate the parameters once the most adequate model has been obtained.

B. Parameter Estimation

In a situation of r responses and p parameters which are common to all the r response equations, Box and Draper [49] showed that the point minimization of

$$z = \sum_{i=1}^{r} \sum_{j=1}^{r} \sigma^{ij} v_{ij} \quad (46)$$

gives the generalization of the method of least squares. Here σ^{ij} is the (i, j)th element of Σ^{-1}

and $$v_{ij} = \sum_{u=1}^{n} \left(y_u^{(i)} - \eta_u^{(i)}\right)\left(y_u^{(j)} - \eta_u^{(j)}\right). \quad (47)$$

When σ_{ij} is unknown, they suggest the minimization of

$$z_0 = \sum_{i=1}^{r} \sum_{j=1}^{r} v_{ij} V_{ij}/r \quad (48)$$

$$D_c = \begin{vmatrix} \sum_{u=1}^{n} \left(y_u^{(1)} - \eta_u^{(1)}\right)^2 & \cdots\cdots & \sum_{u=1}^{n} \left(y_u^{(1)} - \eta_u^{(1)}\right)\left(y_u^{(r)} - \eta_u^{(r)}\right) \\ - \quad - \quad - & & - \\ - \quad - \quad - \quad - \quad - & & \sum_{u=1}^{n} \left(y_u^{(r)} - \eta_u^{(u)}\right)^2 \end{vmatrix}. \quad (56)$$

for the estimation of parameters, where V_{ij} is the cofactor of v_{ij} in v_{ij}. However, they warn that the overall criterion is likely to be offset by the lack of fit in a particular response. In such situations it is safe to check for the adequacy of fit by a residual analysis, and also to consider the consistency of information from various responses by comparing the posterior distributions.

Beauchamp and Cornell [50] suggested an iterative procedure for the estimation of parameters in a multiresponse system. They suggest the minimization of

$$\phi(\underline{\theta}) = [\underline{Y} - \underline{\eta}(\underline{\theta}, \underline{\xi})]'\Omega^{-1}[\underline{Y} - \underline{\eta}(\underline{\theta}, \underline{\xi})] \quad (49)$$

Here Ω is defined by

$$\Omega = E(\underline{\epsilon}\underline{\epsilon}') = \Sigma \otimes I, \quad (50)$$

where $\otimes$ is the Kronecker product.

As a starting point the r response equations are expanded about a trial vector of parameters $\underline{\theta}^0$ to give

$$\underline{\eta}(\underline{\theta}, \underline{\xi}) = \underline{\eta}(\underline{\theta}^0, \underline{\xi}) + X\underline{\delta}^0, \quad (51)$$

where

$$X_i = \{x_{ul}^{(i)}\}, \qquad u = 1,2 \cdots, n; l = 1,2 \cdots, p, \quad (52)$$

for

$$x_{ul}^{(i)} = \partial\eta^{(i)}(\underline{\theta}, \underline{\xi}_u)/\delta\theta_i, \quad (53)$$

$$X = (x_1, x_2, \cdots, x_r)' \quad (54)$$

and

$$\delta^0 = \underline{\theta} - \underline{\theta}^0. \quad (55)$$

Modified Gauss–Newton method discussed by Hartley [51] can be used to estimate the least-squares estimates of $\underline{\theta}$.

Hunter [52] proposed the following criteria for various situations.

1) If σ^{ij} are known, the quantity $\Sigma_{i=1}^{r}\Sigma_{j=1}^{r}\sigma^{ij}v_{ij}$ should be minimized.
2) If $\sigma_{ij} = 0$ for $i \neq j$, $\Sigma_{i=1}^{r}\sigma^{ii}\Sigma_{u=1}^{n}(y_u^{(i)} - \eta_u^{(i)})^2)$ should be minimized.
3) For cases where the variances are assumed equal or when certain responses can be measured more precisely than others, then $\Sigma_{i=1}^{r}\Sigma_{u=1}^{n}(y_u^{(i)} - \eta_u^{(i)})^2$ should be minimized.

In all the above situations it is assumed that σ^{ij} are known. For the case when σ^{ij} are not known minimization of Box and Draper's determinant criterion [49] gives the best estimates of $\underline{\theta}$ in a multiresponse model. Their determinant criterion is given by

Mezaki and Butt [53] applied the above criterion to a complex reaction sequence and found the criterion to provide an effective means of estimation of parameters. With regards to convergence to final estimates, they observed that the determinant criterion to be much more rapid than the generalized nonlinear least-squares techniques. They found very little difference in these two procedures so far as computational effort is concerned. From a precision point of view, the determinant criterion puts the greatest weight on those responses which are measured most accurately, while the least-squares criterion places the greatest weight on the data known less precisely.

Erjavec [54] points out that the error variance of each of the responses must remain constant from run to run. Box *et al.* [55] discussed the possibility of one or the other kind of linear relationships which might exist among the responses. They suggest that if m-linear relationships are known to exist, then m-dependent responses must be deleted before analyzing the data.

Box *et al.* [56] proposed a method which minimizes v_{ij} with respect to $\underline{\theta}$ and $\underline{Y}_m$ for handling missing data $\underline{Y}_m$.

Reilly and Patinp-Leal [57] presented results of a study of the functional case of the problem of parameter estimation when there is error in all the variables. Their study leads to new and efficient algorithms for finding point estimates and their precisions.

IX. Model Discrimination—Multiresponse Models

Similar to single-response models, the task before an experimenter is to discriminate among various competing multiresponse models. Either the likelihood discrimination techniques described earlier for the single-response models or the Bayesian methods which make use of posterior probabilities can be used for modal discrimination.

The likelihood $L_{n,k}$ for the kth multiresponse model is given by [5], [30]

$$L(\underline{\theta}\Sigma, \underline{y}) = \frac{\Sigma_{n,k}^{1/2}}{(2\pi)^{r/2}} \cdot \exp\left[-\frac{1}{2}(\underline{y}_n - \underline{\eta}_{n,k})'\Sigma^{-1}(\underline{y}_n - \underline{\eta}_{n,k})\right], \quad (57)$$

where

$$\Sigma_{n,k} = \sum_{t=1}^{r} + X_{n,k}M_k^{-1}X'_{n,k}. \quad (58)$$

Here $X_{n,k}$ is an $(r \times p)$ matrix of partial derivatives whose (i, l)th element is the partial derivative of the ith response with respect to θ_l. M_k itself is given by

$$M_k = \sum_{i=1}^{r}\sum_{j=1}^{r} \sigma^{ij}X_k^{(i)\prime}X_k^{(j)}. \quad (59)$$

Likelihood discrimination is carried out by comparing maximum likelihoods for different models, taken one at a time.

In Bayesian methods discrimination is assumed to be achieved if one of the models attains a high posterior probability. At the end of n-experiments, if it is not possible to discriminate among the various competing models, further experimentation is necessary which is conducted by a sequential design of experiments as described below.

Bayesian methods, in general, use maximum likelihood for computing the posterior model probabilities. However the likelihood, being a function of the parameters which are themselves random variables, is in itself a random variable. Prasad and Rao [30] used expected likelihood in place of point or maximum likelihood in computing the posterior probabilities and demonstrated the utility of expected likelihood in efficient model discrimination.

X. Design of Experiments for Model Discrimination—Multiresponse Models

Hill and Hunter [58] extended the Box and Hill's criterion for application to multiresponse models. Their discrimination criterion, to be maximized for obtaining the experimental conditions for the $(n + 1)$th experimental run, is given by

$$D = \frac{1}{2}\sum_{h=1}^{m}\sum_{k=h+1}^{m} P_{n,h}P_{n,k}\{\mathrm{tr}\left[\Sigma_{n+1,h}\Sigma_{n+1,k}^{-1} + \Sigma_{n+1,k}\Sigma_{n+1,h}^{-1} - 2I_r\right] + (\hat{Y}_{n+1,h} - \hat{Y}_{n+1,k})'(\Sigma_{n+1,h}^{-1} + \Sigma_{n+1,k}^{-1})(\hat{Y}_{n+1,h} - \hat{Y}_{n+1,k})\}, \quad (60)$$

where I_r is an $(r \times r)$ identity matrix, $\hat{Y}_{n+1,k}$ is an $(r \times 1)$ vector of response for the $n + 1$th experimental conditions predicted using the kth model and the previous best estimates of parameters, and $P_{n,k}$ is the posterior probability of model k after the observations Y_n are obtained.

The design of experiments and analysis are conducted sequentially until discrimination is achieved. Less emphasis is placed on poorly fitting models. When the prior probabilities are in error, it is observed that the correct model emerged as the best one at the expense of more experimentation.

An important point to be noted about the Box–Hill procedure is related to the stopping rule. Box–Hill's criterion suggests that discrimination is achieved when the posterior probability of one of the models is rather high compared to those of other models. However, several investigators [9], [14], [38], [59] experience that model probabilities in certain situations may oscillate considerably from stage to stage. The rule must be applied cautiously and a model should not be accepted readily on the basis of a small number of discriminating experiments. Despite the limitations of Box–Hill's procedure, it has been successfully applied in various situations.

Another useful criterion for the design of experiments for model discrimination has been proposed by Roth [13]. For a multiresponse system, Roth's criterion is given by

$$D = \prod_{i=1}^{r}\left(\sum_{k=1}^{r}\left(P_{n,k}\prod_{\substack{h=1\\h\neq k}}^{m}(\hat{Y}_{n+1,h}^{(i)} - \hat{Y}_{n+1,k}^{(i)})\right)\right), \quad (61)$$

where $P_{n,k}\hat{Y}_{n+1,h}^{(i)}$ are defined earlier.

Roth's criterion consists of maximizing the divergence between the values of responses predicted by different models. If the parameters are not known to the same degree of accuracy in all the models a misleading set of experimental conditions may be predicted.

Reilly and Blau [60] observed that in the use of Roth's criterion the large divergences due to inaccurately known parameters outweigh the smaller divergences due to accurately known parameters, resulting in a probably wrong sequential design.

XI. Design of Experiments for Parameter Estimation—Multiresponse Models

Once the precise model applicable to a particular physicochemical situation has been determined, the goal of the experimenter is to estimate the parameters of the model

precisely, which needs additional experimentation. Draper and Hunter [61] proposed a method of conducting n^* additional experiments. Their approach consists of maximizing the posterior density of $\underline{\theta}$, after conducting $(n + n^*)$ runs, with respect to $\underline{\theta}$ and n^*—which is equivalent to maximizing the determinant

$$D = \sum_{i=1}^{r} \sum_{j=1}^{r} \sigma^{ij} \underline{X}_i' \underline{X}_j, \tag{62}$$

where $\underline{X}_i$ is an $(n + n^*) \times p$ matrix of elements defined by (52), the derivatives being evaluated at the current best estimates of $\underline{\theta}$.

Box [62], [63] discussed the problem of nonlinear model building in situations when constancy of covariance matrix cannot be assumed. He has also developed a computer program which undertakes the risk of formal computations of optimal experimental designs.

In the program he proposed the use of numerical differentiation for obtaining the derivatives of the model responses with respect to the unknown parameters. The program can be extended to handle errors in the input variables. Box [64] has described an experimental design criterion for precise estimation of a subset of the parameters in a nonlinear model.

This involves formulating an experimental design criterion for estimating only q parameters of interest out of a total p parameters. For a locally uniform prior distribution for $\underline{\theta}$, his criterion consists of maximizing the determinant

$$D = A_{11} - A_{12} A_{22}^{-1} A_{12}', \tag{63}$$

where A_{11} is a $(q \times q)$ matrix, and A_{12}, A_{22} are obtained by partitioning A, and is structured

$$A = \sum_{u=1}^{n} \sum_{i=1}^{r} \sum_{j=1}^{r} \sigma^{ij} X_{ij} X_{ju}' \tag{64}$$

for

$$X_{iu} = \left\{ \frac{\partial \eta^{(i)}(\underline{\theta}, \underline{\xi}_u)}{\partial \theta_1} \cdot \cdots \cdot \frac{\partial \eta^{(i)}(\underline{\hat{\theta}}, \underline{\xi}^u)}{\partial \theta_p} \right\}, \tag{65}$$

all derivatives being evaluated at $\underline{\hat{\theta}}$. Thus Box's [64] criterion for the design of experiments should be to maximize the determinant D described in (63). The derivatives are all evaluated of the current best estimate of θ.

XII. A Case Study of Sequential Model Discrimination in Multiresponse Systems

Prasad and Rao [30] advocated the use of expected likelihood in efficient model discrimination. From the basic definition, expected likelihood is given by

$$E(L_{n+1,u}(\underline{\theta}\Sigma, \underline{Y}_{n+1})) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} L_{n+1,u}(\underline{\theta}\Sigma, \underline{Y}_{n+1}) \cdot P_{n,u}(\underline{\theta})\, d\theta_1, d\theta_2, \cdots, d\theta_p. \tag{66}$$

See Prasad and Rao's article for details on evaluating the integral in (66).

Either the point likelihood ((57)) or the expected likelihood ((66)) can be used to update model probabilities using Bayes' theorem. Prasad and Rao presented computational results using multiresponse data to demonstrate the utility of the expected likelihood in efficient model discrimination in a catalyst fouling system. Two discriminatory criteria, viz., the Box–Hill criterion ((60)) and Roth's criterion ((61)), are compared.

The extensive multiresponse experimental data on the product distribution for various space velocities and decay times for the vapor phase reaction between tetrachloroethane and a large excess of chlorine on activated silica gel, collected by Prasad and Doraiswamy [65], are used in the present case study on model discrimination. The reaction proceeded according to the following scheme.

$$\underset{(A_1)}{C_2H_2Cl_4} \xrightarrow{Cl_2} \underset{(A_2)}{C_2HCl_5} \xrightarrow{Cl_2} \underset{(A_3)}{C_2Cl_6}.$$

Their experiments revealed a decline in the catalyst activity with time. Typical data used for illustrative purposes are given in Table III. The prior information consisted of a subset of six experimental data points, while the remaining 64 data points were used for sequential design purposes.

Under certain assumptions a general reaction rate model for the system under consideration may be represented by two simultaneous equations of the type

$$\eta = f_1(\underline{\xi}, \underline{\theta}) f_2(a), \tag{67}$$

$$da/dt = f_3(\underline{\xi}, \underline{\theta}) f_4(a), \tag{68}$$

where η is the rate of reaction, f_1, f_2, f_3, and f_4 are functions depending on the system, t is the decay time, and a is the activity of the catalyst $(a = \eta_{t=t}/\eta_{t=0})$.

For the chlorination reaction, 11 different combinations of fouling reactions were considered by Prasad and Rao, resulting in 11 different plausible models.

f_1 consists of two functions corresponding to the two independent responses—η_1, the rate of disappearance of A_1, and η_2, the rate of formation of A_2, i.e.,

$$\eta_1 = -\theta_1 \xi_1 a \tag{69}$$

and

$$\eta_2 = (\theta_1 \xi_1 - \theta_2 \xi_2) a. \tag{70}$$

A further assumption is given by

$$f_2(a) = f_4(a) = a. \tag{71}$$

The difference combinations of fouling reactions considered by the authors and the corresponding model equations are given in Table IV. The Box complex method [59], [66] was used for estimating the parameters in various models.

The prior information consisted of six random experiments (experiments numbered 8, 13, 18, 23, 41, and 46). Since no one model was preferred to start with, prior probability of each model was assumed to be 1/11. The following initial estimates were used for parameter estimation:

$$\theta_1 = 0.01\ \text{hr}^{-1}, \qquad \theta_2 = 0.001\ \text{hr}^{-1}, \qquad \theta_3 = 0.03\ \text{hr}^{-1}.$$

TABLE III
TYPICAL EXPERIMENTAL RATE DATA

S. No.	Expt. No.	Space time (g hr/ mol)	Mole % of A_1 ($100\xi_1$)	Mole % of A_2 ($100\xi_2$)	Decay time (hr) (ξ_3)	y_1 (mol/g hr x 10^2)	y_2
1	13	90	44.3	52.0	1	-0.460	0.242
2	58	30	82.1	17.4	8	-0.515	0.609
3	36	15	89.0	10.9	5	-0.715	0.668
4	46	60	63.2	35.2	6	-0.518	0.450
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
70	64	15	90.3	9.5	9	-0.585	0.524

TABLE IV
RIVAL MODELS

Model Number	Fouling Reactions	Model Equations
1	$A_1 \rightarrow P$	$\eta^1 = -\theta_1\xi_1 \exp(-\theta_3\xi_1\xi_3)$ $\eta^2 = (\theta_1\xi_1 - \theta_2\xi_2)\exp(-\theta_3\xi_1\xi_3)$
2	$A_2 \rightarrow P$	$\eta^1 = -\theta_1\xi_1 \exp(-\theta_3\xi_2\xi_3)$ $\eta^2 = (\theta_1\xi_1 - \theta_2\xi_2)\exp(-\theta_3\xi_2\xi_3)$
3	$A_3 \rightarrow P$	$\eta^1 = -\theta_1\xi_1 \exp(-\theta_3(1 - \xi_1 - \xi_2)\xi_3)$ $\eta^2 = (\theta_1\xi_1 - \theta_2\xi_2)\exp(-\theta_3(1 - \xi_1 - \xi_2)\xi_3)$
4	$A_1 + A_2 \rightarrow P$	$\eta^1 = -\theta_1\xi_1 \exp(-\theta_3\xi_1\xi_2\xi_3)$ $\eta^2 = (\theta_1\xi_1 - \theta_2\xi_2)\exp(-\theta_3\xi_1\xi_2\xi_3)$
5	$A_1 + A_3 \rightarrow P$	$\eta^1 = -\theta_1\xi_1 \exp[-\theta_3\xi_1(1 - \xi_1 - \xi_2)\xi_3]$ $\eta^2 = (\theta_1\xi_1 - \theta_2\xi_2)\exp[-\theta_3\xi_1(1 - \xi_1 - \xi_2)\xi_3]$
6	$A_2 + A_3 \rightarrow P$	$\eta^1 = -\theta_1\xi_1 \exp[-\theta_3\xi_2(1 - \xi_1 - \xi_2)\xi_3]$ $\eta^2 = (\theta_1\xi_1 - \theta_2\xi_2)\exp[-\theta_3\xi_2(1 - \xi_1 - \xi_2)\xi_3]$
7	$A_1 + A_2 + A_3 \rightarrow P$	$\eta^1 = -\theta_1\xi_1 \exp[-\theta_3\xi_1\xi_2(1 - \xi_1 - \xi_2)\xi_3]$ $\eta^2 = (\theta_1\xi_1 - \theta_2\xi_2)\exp[-\theta_3\xi_1\xi_2(1 - \xi_1 - \xi_2)\xi_3]$
8	$A_1 \rightarrow P$ $A_2 \rightarrow P$	$\eta^1 = -\theta_1\xi_1 \exp[-\theta_3(\xi_1 + \xi_2)\xi_3]$ $\eta^2 = (\theta_1\xi_1 - \theta_2\xi_2)\exp[-\theta_3(\xi_1 + \xi_2)\xi_3]$
9	$A_1 \rightarrow P$ $A_3 \rightarrow P$	$\eta^1 = -\theta_1\xi_1 \exp[-\theta_3(1 - \xi_2)\xi_3]$ $\eta^2 = (\theta_1\xi_1 - \theta_2\xi_2)\exp[-\theta_3(1 - \xi_2)\xi_3]$
10	$A_2 \rightarrow P$ $A_3 \rightarrow P$	$\eta^1 = -\theta_1\xi_1 \exp[-\theta_3(1 - \xi_1)\xi_3]$ $\eta^2 = (\theta_1\xi_1 - \theta_2\xi_2)\exp[-\theta_3(1 - \xi_1)\xi_3]$
11	$A_1 \rightarrow P$ $A_2 \rightarrow P$ $A_3 \rightarrow P$	$\eta^1 = -\theta_1\xi_1 \exp(-\theta_3\xi_3)$ $\eta^2 = (\theta_1\xi_1 - \theta_2\xi_2)\exp(-\theta_3\xi_3)$

The sequential discrimination procedure consisted of the following steps.

1) From the prior information, the parameters of each model were estimated.
2) Using the current best parameter values and either the Box–Hill criterion ((60)) or Roth's criterion ((61)), the next best discriminating experiment was designed.
3) Using the additional data of step 2) after conducting the experiment, the parameter of all the models were updated.
4) The probabilities of all the models were updated using either the point likelihood or the expected likelihood in the Bayes' theorem.

If the posterior probability of any one model is exceedingly high, discrimination is achieved. Otherwise the above sequence of steps were repeated.

Typical values of the posterior probabilities of the best model obtained by the above procedure are given in Table V.

The following salient features were observed by this case study:

1) An entirely different set of discriminating experiments was designed with the same prior information depending on the discriminatory criterion used.
2) The Box–Hill criterion proved to be more efficient in comparison to Roth's criterion.
3) Convergence towards the best model was faster with the expected likelihood than with the point likelihood.

Caution should be exercised with either the point-likelihood method or the expected-likelihood method. Sometimes the model probabilities oscillate from run to run. When the models have different number of parameters

TABLE V
POSTERIOR PROBABILITIES OF BEST MODEL USING BOX–HILL CRITERION AND EITHER POINT LIKELIHOOD OR EXPECTED LIKELIHOOD

Discrimination stage	Total number of experiments	Posterior probability of Model 9 using point likelihood $\pi_{n+1,9}$	Posterior probability of Model 9 using expected likelihood $\pi_{n+1,9}$
1	7	0.502	0.506
2	8	0.513	0.635
3	9	0.623	0.735
4	10	0.656	0.814
5	11	0.730	0.842

there are inherent difficulties with any discrimination method. Using the likelihood method, good discrimination requires that the likelihood ratio be higher than usual if the favored model is the one with the larger number of parameters [4].

A. Some General Comments on Dynamical System and Optimal Design

The process of modeling the dynamical properties of a system is an important step in the analysis and design of a control system [31]. Very often modeling results in a parametric model of the system which contains many unknown parameters. The problem of estimating parameters in models of dynamical system has been discussed in detail by Aström [31] in his classic paper on "maximum likelihood and prediction error methods." Furthermore Aström has shown that the maximum likelihood method and prediction error methods can be successfully applied for estimating parameters of a dynamical system. Estimation problems lead to maximization of the likelihood function or minimization of a function of the prediction error. A number of different ways of solving the optimization problem are also discussed by Aström [31]. It is further shown that the process of computations can be reduced significantly by organizing the measurements in a special way (equidistant sampling) or by choosing special model structures. For more see [31], [33]. The random process theory and state-space characterization of linear dynamic systems are essential to understanding optimal estimation theory. Gelb *et al.* [68] defines the concept of optimal estimation as follows.

> An optimal estimator is a computational algorithm that processes measurements to deduce a minimum error estimate of the state of a system by utilizing: knowledge of system and measurement dynamics, assumed statistics of system noises and measurement errors, and initial condition information.

The three types of estimation problems (estimate desired at time t), filtering, smoothing, and prediction, are discussed in detail by Gelb [68]. For a broader treatment on this see [68], [33]. The only theory of optimal design, at present, deals with linear models and parameter estimation. Furthermore the nonlinear models imply sequential methods and severe numerical difficulty. The problems of estimating parameters in models of dynamical systems is outside the scope of this paper. See [31], [33], [67]–[74], [76], [78].

XIII. CONCLUDING REMARKS

Parameter estimation, model discrimination, and design of experiments constitute the important components of modeling. Various computer-based techniques have been developed in the recent past for parameter estimation. Linear models are easy to handle while nonlinear models require initial estimates of parameters in the estimation problem. If the models are nonlinear but are intrinsically linear, then they can be linearized to obtain the initial estimates after which one of the several nonlinear estimation techniques can be applied for estimating the parameters. In recent years, derivative-free methods have been developed which can possibly save computation time.

In model discrimination, both the Bayesian and the non-Bayesian methods find extensive application. Also the sequential design techniques find extensive use both for model discrimination as well as for the parameter estimation. Experimental designs for simultaneous model discrimination and parameter estimation need to be developed in the future. Parameter estimation techniques are well-developed for multiresponse situations. While the various modeling techniques have been applied successfully so far on single response systems, their application to multiresponse systems needs to be investigated on experimental data.

ACKNOWLEDGMENT

One of the authors (MSR) gratefully acknowledges the Visiting Professorship at the Department of Chemical Engineering, University of Florida, Gainesville, FL 32611. We would like to thank the referees for their constructive comments which improved the presentation of this paper.

NOMENCLATURE

a	Activity.
A	Reaction species.
A_i	ith model; also used for the determinant defined by (64).
A_1, A_2, A_3	Reacting species.
B	Data; also used for reaction species as in (5).
C	Concentration as in (2); also used for spread among responses surfaces as in (36), and for reaction species as in (5).
D	Design criterion.
D_c	Determinant criterion as given in (56).
E	Expected likelihood as defined by (66).
f	Functional relationship.
F	Matrix of partial derivatives as defined by (41); also used for F in F-distribution.

g	Functional relationship.
I_r	$(r \times r)$ identity matrix.
k	Rate constant.
K_{eq}	Equilibrium constant.
K	Adsorption equilibrium constant.
L	Likelihood.
m	Order of reaction.
M_k	As defined by (59).
M_i	Weighted sum of squares as in (31).
n	Number of experimental observations; also used for the order of reaction as in (2).
p	Number of parameters; also used for partial pressure as in (16).
P	Probability.
$P(A_i)$	Prior probability of the ith model.
$P(B/A_i)$	Likelihood for the ith model.
r	Rate of reaction.
s	Vacant site on a catalyst surface.
s^2	Estimate of variance.
t	Time coordinate; also used for t in t-distribution.
v	Covariance.
V	Variance–covariance matrix.
V_{ij}	Cofactor of v_{ij} in v_{ij}.
X	Partial derivative matrix as defined by (52).
y	Observed value of response.
z	Residual as defined by (19); also used for the quantity defined by (46).
z_o	Quantity defined by (48).

Greek Symbols

Ω	$E(\underline{\varepsilon}\underline{\varepsilon}')$.
α	Significance level.
δ	Quantity defined by (21).
ε	Experimental error.
η	True or predicted value of response.
Σ	Covariance matrix as defined by (45).
$\underline{\theta}$	Vector of parameters.
ϕ	Vector of parameters.
$\underline{\xi}$	Vector of input variables.
σ^2	Variance.
σ	Element of variance–covariance matrix (if subscripts are used) or element of variance–covariance matrix (if superscripts are used).
ϕ	Functional relationship.
π	3.1426; also used for probability.
λ	Nonintrinsic parameter; also used for likelihood ratio as in (32).
υ	Degrees of freedom.
$\otimes$	Kronecker product.

Subscripts

1, 2	Parameters or models.
a, b	Species A, B, C, E, and W.
c, e, w	
o	Initial value.
i, j	Response or model.
N	Experimental trials.
r	Number of responses.

Superscripts

i, k	Models; also to measurements as in (4).
m, n	Reaction orders.
N	Experimental trials.
o	Initial value.
$\hat{}$	Best estimate.
$'$	Transpose.

References

[1] G. E. P. Box, W. G. Hunter, and J. S. Hunter, *Statistics for Experimenters*. New York: Wiley, 1978, ch. 16.

[2] J. R. Kittrell, "Mathematical modeling of chemical reactions," in *Advances in Chemical Engineering*, vol. 8, T. B. Drew, G. R. Cokelet, J. W. Hoopes Jr., and T. Vermeulen, Eds. New York: Academic, 1970, p. 97.

[3] J. R. Kittrell, R. Mezaki, and C. C. Watson, "Estimation of parameters for nonlinear least squares analysis," *Ind. Eng. Chem.*, vol. 57, no. 12, p. 19, 1965.

[4] P. M. Reilly, "Statistical methods in model discrimination," *Can. J. Chem. Eng.*, vol. 48, p. 168, 1970.

[5] W. J. Hill, "Statistical techniques for model-building," Ph.D. dissertation, Univ. Wisconsin, Madison, 1966.

[6] W. G. Hunter and A. M. Reiner, "Designs for discriminating between two rival models," *Technometrics*, vol. 7, p. 307, 1965.

[7] G. E. P. Box and W. J. Hill, "Discrimination among mechanistic models," *Technometrics*, vol. 9, p. 57, 1967.

[8] W. G. Hunter and R. Mezaki, "An experimental design strategy for discriminating among rival mechanistic models," *Can. J. Chem. Eng.*, vol. 45, p. 247, 1967.

[9] G. F. Froment and R. Mezaki, "Sequential discrimination and estimation procedures for rate modeling in heterogeneous catalysis," *Chem. Eng. Sci.*, vol. 25, p. 293, 1970.

[10] T. Hsiang and P. M. Reilly, "A practical method for discriminating among mechanistic models," *Can. J. Chem. Eng.*, vol. 49, p. 865, 1971.

[11] G. E. P. Box and H. L. Lucas, Design of experiments in nonlinear situations, *Biometrika*, vol. 46, p. 77, 1959.

[12] L. H. Hosten, "Non Bayesian sequential experimental design procedures for optimal discrimination between rival models," in *Proc. 4th Int. Symp. on Chemical Reaction Eng.*, Heidelberg, W. Germany, 1976.

[13] P. M. Roth, "Design of experiments for discrimination among rival models," Ph.D. dissertation, Princeton Univ., Princeton, NJ, 1965.

[14] P. D. H. Hill, "Optimal experimental designs for model discrimination," Ph.D. dissertation, Univ. Glasgow, Glasgow, Scotland, 1976.

[15] N. R. Draper and H. Smith, *Applied Regression Analysis*. New York: Wiley, 1966.

[16] R. L. Plackett, *Principles of Regression Analysis*. New York: Oxford, 1960.

[17] G. F. Froment, "Model discrimination and parameter estimation in heterogeneous catalysis," *Amer. Inst. Chem.*, vol. 21, p. 1041, 1975.

[18] R. L. Kabel and L. N. Johanson, "Reaction kinetics and adsorption equilibria in the vapor-phase dehydration of ethanol," *Amer. Inst. Chem.*, vol. 8, p. 621, 1962.

[19] Y. Bard, *Nonlinear Parameter Estimation*. New York: Academic, 1974.

[20] D. L. Marquardt, "An algorithm for least squares estimation of non-linear parameters," *J. Soc. Ind. Appl, Math.*, vol. 2, p. 431, 1963.

[21] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Comput. J.*, vol. 7, p. 308, 1965.

[22] M. S. Rao, "Identification of rate controlling steps in pentene-2 hydrogenation from selectivity—A maximum-likelihood ratio approach," Ph.D. dissertation, University of Waterloo, Waterloo, ON, Canada, 1969.

[23] M. S. Rao, R. R. Hudgins, P. M. Reilly, and P. L. Silveston, "A selectivity method for modelling the kinetics of pentene-2 hydrogenation," *Can. J. Chem. Eng.*, vol. 49, p. 354, 1971.

[24] M. L. Ralston and R. I. Jennrich, "Dud—A derivative-free algo-

rithm for nonlinear least squares," *Technometrics*, vol. 20, p. 7, 1978.
[25] K. M. Gaver and M. S. Geisel, "Discriminating among rival models: Bayesian and non-Bayesian methods," in *Frontiers of Econometrics*, P. Zarembka, Ed. New York: Academic, 1974, p. 49.
[26] R. Mezaki and J. R. Kittrell, "Discrimination between two rival models through non-intrinsic parameters," *Can. J. Chem. Eng.*, vol. 44, p. 285, 1966.
[27] J. R. Kittrell and R. Mexaki, "Discrimination among rival Hougen–Watson models through intrinsic parameters," *Amer. Inst. Chem.*, vol. 13, p. 389, 1967.
[28] W. Volk, *Applied Statistics for Chemical Engineers*. New York: McGraw-Hill, 1958.
[29] K. B. S. Prasad, "Use of expected likelihood in sequential model discrimination," M. Tech. Thesis, Indian Inst. Technology, Kanpur, India, 1976.
[30] K. B. S. Prasad and M. S. Rao, "Use of expected likelihood in sequential model discrimination in multiresponse systems," *Chem. Eng. Sci.*, vol. 32, p. 1411, 1977.
[31] K. J. Aström, "Maximum likelihood and prediction error methods," *Automatica*, vol. 16, p. 551, 1980.
[32] G. E. P. Box and G. M. Jenkins, *Time Series Analysis, Forecasting and Control*. San Francisco: Holden Day, 1970.
[33] R. Mehra, "On the identification of variances and adaptive Kalman filtering," *IEEE Trans. Automat.*, vol. 15, p. 175, 1970.
[34] J. R. Kittrell, W. G. Hunter, and R. Mezaki, "The use of diagnostic parameters for kinetic model building," *Amer. Inst. Chem.*, vol. 5, p. 1014, 1966.
[35] K. H. Yang and O. A. Hougen, "Determination of mechanism of catalyzed gaseous reactions," *Chem. Eng. Prog.*, vol. 46, p. 146, 1950.
[36] J. A. Cusumano, R. A. Dalla Betta, and R. B. Levy, *Catalysis in Coal Conversion*. New York: Academic, 1978, p. 65.
[37] W. J. Hill and W. G. Hunter, "A note on designs for model discrimination: Variance unknown case," *Technometrica*, vol. 11, p. 396, 1969.
[38] W. W. Wentzheimer, "Modelling of heterogeneous catalysed reactions using statistical experimental design and data analysis," Ph.D. dissertation, Univ. Pennsylvania, Philadelphia, 1970.
[39] P. D. H. Hill, "A review of experimental design procedures for regression model discrimination," *Technometrics*, vol. 20, no. 1, p. 15, 1978.
[40] A. C. Atkinson and V. V. Fedorov, "The design of experiments for discriminating between two rival models," *Biometrika*, vol. 62, p. 57, 1975.
[41] _____, "Optimal design: experiments for discriminating between several models," *Biometrika*, vol. 62, p. 289, 1975.
[42] A. C. Atkinson and D. R. Cox, "Planning experiments for discriminating between models," *J. Roy. Statist. Soc.*, vol. B, no. 36, p. 321, 1974.
[43] J. R. Kittrell, W. G. Hunter, and C. C. Watson, "Obtaining precise parameter estimates for nonlinear catalytic rate models," *Amer. Inst. Chem.*, vol. 12, no. 1, p. 5, 1966.
[44] R. J. Ayen and M. S. Peters, "Catalytic reduction of nitric oxide," *Ind. Eng. Chem. Process Des. Develop.*, vol. 1, p. 204, 1962.
[45] L. H. Hosten and G. Emig, "Sequential experimental design procedures for precise parameter estimation in ordinary differential equations," *Chem. Eng. Sci.*, vol. 30, p. 1357, 1975.
[46] G. E. P. Box and K. B. Wilson, "On the experimental attainment of optimum conditions," *J. Roy. Statist. Soc.*, vol. B, no. 31, p. 1, 1951.
[47] D. M. Borth, "A total entropy criterion for the dual problem of model discrimination and parameter estimation," *J. Roy. Statist. Soc.*, vol. B, no. 37, p. 77, 1975.
[48] S. Singh and M. S. Rao, "Parameter estimation and model discrimination in multiresponse models," *Indian Chem. Eng.*, vol. 23, no. 2, p. 19, 1981.
[49] G. E. P. Box and N. R. Draper, "The Bayesian estimation of common parameters for several responses," *Biometrika*, vol. 52, p. 355, 1965.
[50] J. J. Beauchamp and R. G. Cornell, "Simultaneous nonlinear estimation," *Technometrics*, vol. 8, p. 319, 1966.
[51] O. Hartley, "The modified Gauss–Newton method for the fitting of nonlinear regression functions by least squares," *Technometrics*, vol. 3, p. 269, 1961.
[52] W. G. Hunter, "Estimation of unknown constants from multiresponse data," *Ind. Eng. Chem. Fundam.*, vol. 6, p. 461, 1967.
[53] R. Mezaki and J. B. Butt, "Estimation of rate constants from multiresponse kinetic data," *Ind. Eng. Chem. Fundam.*, vol. 7, p. 120, 1968.
[54] J. Erjavec, "Strategy for estimation of rate constants from isothermal reaction data," *Ind. Eng. Chem. Fundam.*, vol. 9, p. 187, 1970.
[55] G. E. P. Box, W. G. Hunter, J. F. MacGregor, and J. Erjavec, "Some problems associated with the analysis of multiresponse data," *Technometrics*, vol. 15, p. 33, 1973.
[56] M. J. Box, N. R. Draper, and W. G. Hunter, "Missing values in multiresponse nonlinear model fitting," *Technometrics*, vol. 12, p. 613, 1970.
[57] P. M. Reilly and H. Patino-Leal, "A Bayesian study of the error-in-variables model," *Technometrics*, vol. 23, no. 3, p. 221, 1981.
[58] W. J. Hill and W. G. Hunter, Tech. Rep. 65, Dept. of Statistics, Univ. of Wisconsin, Madison, 1966.
[59] S. M. Siddik, "Kullback–Lieber information function and the sequential selection of experiments to discriminate among several linear models," Ph.D. dissertation, Princeton Univ., Princeton, NJ, 1972.
[60] P. M. Reilly and G. E. Blau, "The use of statistical methods to build mathematical models of chemical reacting systems," *Can. J. Chem. Eng.*, vol. 52, p. 289, 1974.
[61] N. R. Draper and W. G. Hunter, "Design of experiments for parameter estimation in multiresponse situations," *Biometrika*, vol. 53, p. 525, 1966.
[62] M. J. Box, "Improved parameter estimation," *Technometrics*, vol. 12, p. 219, 1980.
[63] _____, "Simplified experimental design," *Technometrics*, vol. 13, p. 19, 1971.
[64] _____, "An experimental design criterion for precise estimation of a subset of the parameters in a nonlinear model," *Biometrika*, vol. 58, p. 149, 1971.
[65] K. B. S. Prasad and L. K. Doraiswamy, "Effect of fouling in a fixed-bed reaction for a complex reaction: Test of proposed model and formulation of an optimal policy," *J. Cat.*, vol. 32, p. 384, 1974.
[66] M. J. Box, "A new method of constrained optimization and a comparison with other methods," *Comput. J.*, vol. 8, p. 42, 1965.
[67] B. O. Anderson and J. B. Moore, *Optimal Filtering*. Englewood Cliffs, NJ: Prentice-Hall, 1979.
[68] A. Gelb, Ed., *Applied Optimal Estimation*. Cambridge, MA: MIT, 1974.
[69] M. S. Maklad and S. T. Nichols, "A new approach to model structure discrimination," *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-10, no. 2, pp. 78–84, Feb. 1980.
[70] K. J. Aström, "On the achievable accuracy in identification problems," preprints of *1st IFAC Symp. on Identification in Aut. Contr. Systems*, Prague, Czechoslovakia, paper 1.8, 1967.
[71] R. K. Mehra, "Optimal input signals for parameters estimation in dynamic systems—Survey and new results," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 192–200, 1974.
[72] T. Söderström *et al.*, "On the accuracy of identification and design of identification experiment," Div. of Aut. Contr., Lund Institute of Technology, Lund, Sweden, Rep. 7428, 1974.
[73] K. J. Aström and T. Bohlin, "Numerical identification of linear dynamic systems from normal operating records," in *Proc. on IFAC Conference on Self-Adaptive Control Systems*, Teddington, England, pp. 96–111, 1965.
[74] K. J. Aström and T. Söderström, "Uniqueness of maximum likelihood estimates of parameters of ARMA models," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 769–773, 1974.
[75] M. W. Schwartz, "Identification of systems of unknown order," Ph.D. dissertation, Michigan State Univ., East Lansing, MI, 1972.
[76] R. L. Kashyap and R. L. Rao, *Dynamic Stochasti Models from Empirical Data*. New York: Academic, 1976.
[77] E. T. Janes, "Prior probabilities," *IEEE Trans. Syst., Sci., Cybern.*, vol. SSC-4, pp. 227–241, 1968.
[78] R. L. Kashyap, "A Bayesian comparison of different classes of dynamic models using empirical data," *IEEE Trans. Automat. Contr.*, vol. AC-22, pp 715–727, 1977.
[79] T. L. Fine and W. G. Hwang, "Consistent estimation of system order," *IEEE Trans. Automat. Contr.*, vol. AC-24, pp. 387–402, 1979.
[80] G. Schwartz, "Estimating the dimension of a model," *Ann. Statis.*, vol. 6, pp. 461–464, 1978.
[81] S. Sitharama Iyengar, "Computer modeling of complex biosystems," CRC Press, Inc., June 1983 (to appear).
[82] S. Sitharama Iyengar, M. R. Rao, and S. Guave, "A 4 level software engineering approach to model complex systems," submitted for publication.
[83] S. Sitharama Iyengar, "A systematic approach to modeling complex biological systems," in *Proc. 3rd Int. Congress on Med. Inform.*, Versailles, France, pp. 441–447.