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.

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]

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

(1)

where $\eta$ is the BOD and $t$ is the incubation period.

0018-9472/83/0300-0175$01.00 ©1983 IEEE
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

$$r_A = -\frac{dC_A}{dt} = kC_A^nC_B^m,$$  \hspace{1cm} (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

$$\eta = f(\theta, \xi),$$  \hspace{1cm} (3)

where $\eta$ is the expected value of the dependent variable $y$, $\theta$ is a (p×1) vector of parameters, and $\xi$ is a (p×1) 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]

$$y_i = \theta_0 + \theta_1 \xi_{i_1} + \theta_2 \xi_{i_2} + \cdots + \theta_p \xi_{ip} + \theta_{pp} \xi_{ip}^2 + \xi_i,$$ \hspace{1cm} (4)
where $\epsilon$ is the error associated with a measurement, $\theta$ and $\xi$ refer to the corresponding parameters and independent variables, respectively, and the subscript $i$ refers to the $i$th 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.

$$A.s + B.s \rightarrow C.s + s,$$

(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

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

(6)

where $y$ is the rate of reaction (dependent variable) and $\xi_1$ and $\xi_2$ 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 $\xi$ available on a particular system. The experimental observations of a hypothetical situation are shown on the $y$ versus $\xi$ 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 $0 < \xi < \xi_1$, 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 $\xi$. Thus the “design of experiments” plays a dominant role in the model 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.

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

$$y = \theta_1 \xi_1^m + \theta_2 \xi_2^n + \epsilon$$

(7)

is a linear model in the parameters $\theta_1$, $\theta_2$. 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.

$$y = \theta_1 \exp(-\theta_2 \xi_1 \xi_2) + \epsilon.$$  

(8)

The technique to be employed for the estimation of parameters in any model depends on the nature of the model and its 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 $\eta_i$, and may be perfectly represented by

$$\eta_i = \theta_1 \xi_{i1} + \theta_2 \xi_{i2} + \cdots + \theta_p \xi_{ip}, \quad p < n.$$  

(9)

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

$$y_i = \theta_1 \xi_{i1} + \theta_2 \xi_{i2} + \cdots + \theta_p \xi_{ip} + \epsilon_i,$$

(10)

where $\xi_{ij}$ is the value of the $j$th independent variable
corresponding to the $i$th 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

$$V(\varepsilon) = I_{n} \sigma^{2},$$

(11)

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

The least-squares estimates of the parameters $\hat{\theta}$ 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

$$\hat{\theta} = (\xi^T \xi)^{-1} \xi^T y,$$

(12)

where $\xi$ is given by

$$\xi = \begin{bmatrix}
\xi_{11} & \xi_{12} & \cdots & \xi_{1p} \\
\xi_{21} & \xi_{22} & \cdots & \xi_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
\xi_{n1} & \xi_{n2} & \cdots & \xi_{np}
\end{bmatrix}$$

(13)

and

$$\gamma = (y_{1}, y_{2}, \cdots, y_{n})^T.$$  

(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 $\theta$ is given by

$$\hat{\theta} \pm t_{n-\alpha/2} [v(\theta)]^{1/2},$$

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

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

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

(15)

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

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

$$r = \frac{k_{p} (P_{o}^{2} - P_{r}^{2} + K_{o}^{2})}{(1 + K_{p} P_{o} + K_{p} P_{u} + K_{p} P_{r})^{2}},$$

(16)

In this expression $r$ is the measured (dependent) variable, $P_{o}$, $P_{r}$, and $P_{u}$ are the controllable quantities (independent variables), and $k$, $k_{p}$, $k_{o}$, $k_{r}$, and $K_{o}$ 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:

$$\eta_{u} = \eta_{0} + \sum_{i=1}^{p} \frac{\partial f(\xi_{u}, \theta)}{\partial \theta_{i}} \delta_{i} + \epsilon,$$

(17)

Including experimental error (17) may be written as

$$z_{u} = \sum_{i=1}^{p} f_{i} \delta_{i} + \epsilon,$$

(18)

where

$$z_{u} = y_{u} - f(\xi_{u}, \theta)_{\theta = \theta_{0}} + \epsilon,$$

(19)

and

$$f_{i} = \frac{\partial f(\xi_{u}, \theta)}{\partial \theta_{i}} \bigg|_{\theta = \theta_{0}}$$

(20)

and

$$\delta_{i} = \theta_{i} - \theta_{0}.$$

(21)

Since (18) is linear, the correction vector $\delta$ can be obtained by fundamentally minimizing $\delta^T \delta$. The improved parameter estimates for the next trial are given by

$$\theta^{(1)} = \theta^{0} + \delta,$$

(22)

This procedure is iterated until the corrections $\delta$ 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.

1The 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 byRalston 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 nonintrinsinc parameters, which are parameters introduced into a model for the purpose of discrimination from a set of rival models;
2) use of intrinsinc 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 Nonintrinsinc Parameters

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

\[ \eta_1 = f(\theta, \xi) \]  

and

\[ \eta_2 = g(\phi, \xi). \]  

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

\[ y - \frac{1}{2} (\eta_d + \eta_a) = \lambda (\eta_b - \eta_d). \]  

Here \( \lambda \) is a nonintrinsin parameter since it was not originally present in the model. In order to discriminate between the two rival models (24) is regressed for \( \lambda \). As seen from (24), Model 1 is the proper one if \( \lambda \) were equal to \( -1/2 \) and Model 2 is the proper one if \( \lambda \) were equal to \( 1/2 \). In practice \( \lambda \) may not be exactly equal to either \( +1/2 \) or \( -1/2 \). In such cases the confidence limits of \( \lambda \) should be assessed to see whether it included \( +1/2 \) or \( -1/2 \), or both. If \( \lambda \) were equal to \( -1/2 \) in the desired confidence region, then Model 1 is the correct one. On the contrary, if \( \lambda \) 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

\[ y = (1 - \lambda)\eta_1 + \lambda \eta_2 \]  

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

B. Use of Intrinsinc Parameters

Intrinsinc 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]

\[ P(A_i/B) = \frac{P(A_i)P(B/A_i)}{\sum_{i=1}^{r} P(A_i)P(B/A_i)}, \]  

where \( A_i \), \( (i = 1, \cdots, r) \) denotes the \( i \)th model, \( B \) denotes the data, \( P(A_i) \) denotes the prior probability of the \( i \)th model, and \( P(B/A_i) \) denotes the likelihood for the \( i \)th 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 \( \sigma^2 \), the probability density function (pdf) of \( n \) observations \( y_1, y_2, \cdots, y_n \) is given by [14]

\[ \text{pdf} = \frac{1}{(\sqrt{2\pi}\sigma)^n} \exp\left[ -\frac{1}{2\sigma^2} \sum_{n=1}^{n} (y_n - f(\theta, \xi))^2 \right]. \]  

(27)
On the other hand, if one assumes that the errors are normally distributed but correlated with a variance–covariance matrix $\Sigma$, then the pdf of $n$ observations $y_1, y_2, \cdots, y_n$ is given by

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

(28)

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

**Example of Model Discrimination:** Reilly [4] considered a simple numerical example of discriminating among rival models. Assuming $\xi$ as a single independent variable, $y_i$ 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:

- **Model 1:** $y_i = \theta_{1i} \xi_i + \epsilon_i$
- **Model 2:** $y_i = \theta_{2i} + \theta_{22} \xi_i + \epsilon_i$
- **Model 3:** $y_i = \theta_{1i} \exp \theta_{2i} \xi_i + \epsilon_i$

(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 $i$th model to be represented by the functional relationship

$$y_i = f_i(\theta, \xi_i).$$

(30)

Let $y$ denote the vector of observations. Also assume that the observations are correlated with a variance–covariance matrix $\Sigma$. 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 $\sigma^2$) for the $i$th model can be shown to be given by [23]

$$L_i(\theta, \sigma^2) = \frac{n^{n/2}}{(2\pi)^{n/2}M^{n/2}} \exp \left( -\frac{n}{2\sigma^2} \right).$$

(31)

where $M$ is the weighted sum of squares $(y - \bar{y})^2 + \epsilon_i (y - \bar{y})$.

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

$$\Lambda = \left( \frac{L_i}{L_j} \right)_{\text{max}} = \frac{M_i}{M_j} n^{n/2}.$$

(32)

This ratio $\Lambda$ 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:

\[
\text{lack of fit mean square} = \frac{\sum (y_i - \hat{y}_i)^2}{\text{pure error mean square}}
\]

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., $y_i - \hat{y}_i$). 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

\[
y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + \epsilon_i.
\]

Assume that the fit is given by

\[
y_i = \hat{y}_0 + \hat{\beta}_1 x_i + \hat{\beta}_2 z_i.
\]

It will be seen that the residual $y_i - \hat{y}_i$ is correlated with the variable $z_i$ in a linear way. Thus if the residual is correlated with respect to $z$, it is indicative that the model should include a term consisting of $z$. 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 in 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 experi-

\[
F = \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2 (m - 1) \sigma^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2 (n - 1) \sigma^2}
\]

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

\[
D = \left( \hat{y}_{N+1} - \hat{y}_{N+1}^{*} \right)^2
\]

\[
C_j = \prod_{k=1}^{m} \left( \sigma_j^{(k)} - \sigma_j^{(k)} \right)\]

The subscript $j$ corresponds to the values of the independent variables $x_j$. 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.,

\[
D = \sum_{j=1}^{m} P_j^{(n)} C_j
\]

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

\[
\text{The probability $P_j^{(n)}$ of any model $j$ denotes our degree of belief in that particular model after conducting $n$ experiments.}
\]
Box and Hill as

\[
D = \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \pi_{ij} \rho_{ij} \left( \frac{(s_{ij} - \mu_{ij})}{(s_{ij} + \mu_{ij})} \right)^2 + \left( \frac{\mu_{ij} - \mu_{ij+1}}{\mu_{ij} + \mu_{ij+1}} \right)^2 \left( \frac{1}{s_{ij} + \mu_{ij}} + \frac{1}{s_{ij} + \mu_{ij+1}} \right) \right],
\]

(38)

Here \( s^2 \) is the common variance of the \( n \) observations \( y_1, y_2, \ldots, y_n \) and \( s^2 \) and \( \sigma^2 \) are the variance for the predicted values of \( y_{n+1} \) under the models \( i \) and \( j \), respectively, and \( \mu_{ij} \) and \( \mu_{ij+1} \) 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

\[
\text{pdf} = \frac{1}{(2\pi)^{m/2}} \exp \left( -\frac{1}{2} \sum_{i=1}^{m} \frac{(y_{n+1} - \mu_{ij})^2}{(s_{ij} + \mu_{ij})^2} \right).
\]

(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 \( k^2 \) 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 \( \theta \) evaluated at the \( u \) set of experimental conditions and taken at a set of parameter values \( \theta_i \), i.e.,

\[
f_{\theta} = \frac{\partial f(\theta, \xi_x)}{\partial \theta},
\]

(40)

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

\[
F = (f_{\theta}).
\]

(41)

Box and Lucas [11] have indicated that, under certain plausible assumptions, a choice of experimental points which will maximize \([FF]^{-1}\) will also be that choice of data points which will minimize the volume of the joint confidence region of parameters. \([FF]^{-1}\) 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 Both [47].

VIII. PARAMETERS ESTIMATION—MULTI RESPONSE 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 multiregion 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 multiregion model can be denoted by

\[
Y = \eta(\theta, \xi_x) + \epsilon,
\]

(42)

where \( \eta \) represents the true value of the response \( Y \), \( \epsilon \) corresponds to the error associated with the measurement, \( \theta \) is a \((p \times 1)\) vector of parameters, and \( \xi_x \) is an \((s \times 1)\) vector of controllable variables.

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

\[
y^{(u)}_{i} = \eta^{(u)}(\theta, \xi_{x}) + \epsilon^{(u)}_{i}.
\]

(43)

It is assumed that the errors are such that

\[
E(\epsilon^{(u)}_{i}) = 0, \quad \text{for all } i, u;
\]

\[
E(\epsilon^{(u)}_{i}\epsilon^{(u)}_{j}) = 0, \quad \text{for all } i, j, u = c;
\]

\[
E(\epsilon^{(u)}_{i}\epsilon^{(u)}_{j}) = \sigma_{ij}, \quad \text{for } i, j, u.
\]

(44)
The vector \( \mathbf{y}_i = (y_{ij1}, y_{ij2}, \ldots, y_{jir})' \) of \( r \) response for the \( i \)th experiment has a symmetric covariance matrix \( \Sigma \) given by

\[
\Sigma = (\sigma_{ij}), \quad i = 1, 2, \ldots, r; \quad j = 1, 2, \ldots, r. \tag{45}
\]

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

\[
z = \sum_{i=1}^{r} \sum_{j=1}^{r} a^{ij} v_{ij} \tag{46}
\]

gives the generalization of the method of least squares. Here \( a^{ij} \) is the \((i, j)\)th element of \( \Sigma^{-1} \) and

\[
v_{ij} = \sum_{u=1}^{n} (y_{iu} - \eta(u, \xi)) (y_{jtu} - \eta(u, \xi)). \tag{47}
\]

When \( \sigma_{ij} \) is unknown, they suggest the minimization of

\[
z_0 = \sum_{i=1}^{r} \sum_{j=1}^{r} v_{ij} V_{ij}/r \tag{48}
\]

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

\[
\phi(\theta) = \left[ Y - \eta(\theta, \xi) \right]' \Omega^{-1} \left[ Y - \eta(\theta, \xi) \right] \tag{49}
\]

Here \( \Omega \) is defined by

\[
\Omega = E(\epsilon \epsilon') = \Sigma \otimes I, \tag{50}
\]

where \( \otimes \) is the Kronecker product.

As a starting point the \( r \) response equations are expanded about a trial vector of parameters \( \theta^0 \) to give

\[
\eta(\theta, \xi) = \eta(\theta^0, \xi) + \chi \theta^0, \tag{51}
\]

where

\[
X_i = (x_{i1}, x_{i2}, \ldots, x_{ip}), \quad u = 1, 2, \ldots, n; \quad l = 1, 2, \ldots, p, \tag{52}
\]

for

\[
x_{il}^{(u)} = \frac{\partial \eta^{(i)}(\theta, \xi)}{\partial \theta}, \tag{53}
\]

and

\[
\delta^0 = \theta - \theta^0. \tag{55}
\]

Modified Gauss–Newton method discussed by Hartley [51] can be used to estimate the least-squares estimates of \( \theta \). Hunter [52] proposed the following criteria for various situations.

1) If \( a^{ij} \) are known, the quantity \( \sum_{i=1}^{r} \sum_{j=1}^{r} a^{ij} v_{ij} \) should be minimized.

2) If \( a_{ij} = 0 \) for \( i \neq j \), \( \sum_{i=1}^{r} a^{ii} (y_{iu}^{(i)} - \eta(u, \xi)) \) should be minimized.

3) For cases where the variances are assumed equal or when certain responses can be measured more precisely than others, then \( \sum_{i=1}^{r} \sum_{j=1}^{r} (y_{ij}^{(i)} - \eta(u, \xi)) \) should be minimized.

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

\[
D_c = \left| \sum_{u=1}^{n} (y_{iu}^{(i)} - \eta(u, \xi))' \sum_{u=1}^{n} (y_{iu}^{(i)} - \eta(u, \xi)) \right| - \left| \sum_{u=1}^{n} (y_{iu}^{(i)} - \eta(u, \xi))' \sum_{u=1}^{n} (y_{iu}^{(i)} - \eta(u, \xi)) \right|. \tag{56}
\]

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.

Ergenev [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 \( \theta \) and \( Y_m \) for handling missing data \( Y_m \).
Reilly and Patip-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 model discrimination.

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

$$L(\theta, y) = \frac{\Sigma_{n,k}^{1/2}}{(2\pi)^{n/2}} \exp \left[ -\frac{1}{2} (y_n - y_{n,k})^{\top} \Sigma_{n,k}^{-1} (y_n - y_{n,k}) \right],$$

where

$$\Sigma_{n,k} = \sum_{i=1}^r X_{n,k} M_k^{-1} X_{n,k}^{\top}.$$  

(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 $i$th response with respect to $\theta_l$. $M_k$ itself is given by

$$M_k = \sum_{i=1}^r \sum_{j=1}^r \sigma^{ij} X_{i,n}^{(i)} X_{i,n}^{(j)}.$$  

(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

$$D = \frac{1}{2} \sum_{k=1}^{m} \sum_{l=k+1}^{m} P_{n,k} P_{n,l} \left\{ \left[ \Sigma_{n+1,k} \Sigma_{n+1,l}^{-1} \right] + \left[ \Sigma_{n+1,k} \Sigma_{n+1,l}^{-1} - 2I \right] \right\}$$

$$+ \left[ (\hat{y}_{n+1,k} - \hat{y}_{n+1,l}) (\Sigma_{n+1,l}^{-1} + \Sigma_{n+1,k}^{-1}) + (\Sigma_{n+1,k}^{-1} + \Sigma_{n+1,l}^{-1}) \right] \right\}.$$  

(60)

where $I$ 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 $k$th 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

$$D = \prod_{i=1}^{r} \left\{ \sum_{k=1}^{m} P_{n,k} \prod_{k=1}^{m} \left( \hat{y}_{n+1,k} - \hat{y}_{n+1,k}^{(i)} \right) \right\},$$

(61)

where $P_{n,k}$ and $\hat{y}_{n+1,k}$ 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 maxi-
mizing the posterior density of \( \theta \). After conducting \( (n + n^*) \)
runs, with respect to \( \theta \) and \( n^* \)—which is equivalent to
maximizing the determinant

\[
D = \sum_{i=1}^{n} \sum_{j=1}^{n^*} a_{i}^* X_{ij} \tag{62}
\]

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

Box [62], [63] discussed the problem of nonlinear model
building in situations where 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 differen-
tiation for obtaining the derivatives of the model re-
ponses 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 para-
eters in a nonlinear model.

This involves formulating an experimental design crite-
orion for estimating only \( q \) parameters of interest out of a
total \( p \) parameters. For a locally uniform prior distribution
for \( \theta \), this criterion consists of maximizing the determinant

\[
D = A_{11} - A_{12} A_{22}^{-1} A_{12}^T \tag{63}
\]

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

\[
A = \sum_{i=1}^{q} \sum_{j=1}^{q} \sigma_{ij} X_{ij} X_{ij}^T \tag{64}
\]

for

\[
X_{ij} = \frac{\partial f(\theta, \xi)}{\partial \theta_i} \bigg|_{\theta = \theta_o} \frac{\partial f(\theta, \xi)}{\partial \theta_j} \bigg|_{\theta = \theta_o} \tag{65}
\]

all derivatives being evaluated at \( \theta \). Thus Box's [64]
criterion for the design of experiments should be to maxi-
mize the determinant \( D \) described in (63). The derivatives are all
evaluated of the current best estimate of \( \theta \).

**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

\[
E(L_{\theta}, \Sigma, Y_{+1}) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} L_{\theta}, \Sigma, Y_{+1} \theta, \Sigma, Y_{+1} \phi_{\theta, \Sigma, Y_{+1}}(\theta, \Sigma, Y_{+1}) \phi_{\theta, \Sigma, Y_{+1}}(\theta, \Sigma, Y_{+1}) d\theta, d\Sigma, dY. \tag{66}
\]

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

Either the point likelihood (57) or the expected likeli-
hood (66) can be used to update model probabilities
using Bayes' theorem. Prasad and Rao presented computa-
tional results using multiresponse data to demonstrate the
utility of the expected likelihood in efficient model dis-

crimination in a catalyst fouling system. Two discrimina-
tory 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 tetracloro-
ethane 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.

\[
\begin{align*}
C_2HCl_4 & \rightarrow C_2HCl_2 \rightarrow C_2Cl_6; \\
(A_1) & \rightarrow (A_2) \rightarrow (A_3) \\
\end{align*}
\]

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

\[
\eta = f_1(\xi_1, \theta) f_2(a) \tag{67}
\]

\[
da/dt = f_3(\xi_3, \theta) f_4(a) \tag{68}
\]

where \( \eta \) 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 (\( a = \eta_{+}/\eta_{-} \)).

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\( - \eta_1 \), the rate of disappearance
of \( A_1 \), and \( \eta_2 \), the rate of formation of \( A_2 \), i.e.,

\[
\eta_1 = -\theta_1 \xi_1 a \tag{69}
\]

and

\[
\eta_2 = (\theta_1 \xi_1 - \theta_3 \xi_2) a \tag{70}
\]

A further assumption is given by

\[
f_2(a) = f_3(a) = a \tag{71}
\]

The difference combinations of fouling reactions consid-
ered 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 experi-
ments (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 estima-
tion:

\[
\theta_1 = 0.01 \text{ hr}^{-1}, \quad \theta_2 = 0.001 \text{ hr}^{-1}, \quad \theta_3 = 0.03 \text{ hr}^{-1}.
\]
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 ((66)) 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
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

$A_i$ Activity.  
$A_i$ Reaction species.  
$A_1, A_2, A_3$ 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).  
$D_i$ Design criterion.  
$E_i$ Determinant criterion as given in (56).  
$F$ 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.
Functional relationship.

$(r \times r)$ identity matrix.

Rate constant.

Equilibrium constant.

Adsorption equilibrium constant.

Likelihood.

Order of reaction.

As defined by (59).

Weighted sum of squares as in (31).

Number of experimental observations; also used for the order of reaction as in (2).

Number of parameters; also used for partial pressure as in (16).

Probability.

Prior probability of the $i$th model.

Likelihood for the $i$th model.

Rate of reaction.

Vacant site on a catalyst surface.

Estimate of variance.

Time coordinate; also used for $t$ in $t$-distribution.

Covariance.

Variance–covariance matrix.

Partial derivative matrix as defined by (52).

Observed value of response.

Residual as defined by (19); also used for the quantity defined by (46).

Quantity defined by (48).

Greek Symbols

$\Omega$  

$E(\hat{g})$.  

Significance level.

$\delta$  

Quantity defined by (21).

$\varepsilon$  

Experimental error.

$\eta$  

True or predicted value of response.

$\Sigma$  

Covariance matrix as defined by (45).

$\psi$  

Vector of parameters.

$\phi$  

Vector of parameters.

$\xi$  

Vector of input variables.

$\sigma$  

Variance.

Element of variance–covariance matrix (if subscripts are used) or element of variance–covariance matrix (if superscripts are used).

$\phi$  

Functional relationship.

$3.1426$; also used for probability.

$\lambda$  

Nonintrinsc parameter; also used for likelihood ratio as in (32).

$\nu$  

Degrees of freedom.

$\otimes$  

Kronecker product.

Subscripts

1, 2  

Parameters or models.

$a, b$  

Species $A, B, C, E$, and $W$.

$c, e, w$  

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.

$q$  

Initial value.

Best estimate.

$\text{Transpose}$.  

REFERENCES


