Chemical Engineering Science, Vol. 45, No. 4, pp. 883–892, 1990.
Printed in Great Britain.

0009 2509/90 $3.00 + 0.00
© 1990 Pergamon Press plc

SEQUENTIAL EXPERIMENTAL DESIGN FOR PARAMETER ESTIMATION: A DIFFERENT APPROACH

J. C. PINTO, M. W. LOBÃO and J. L. MONTEIRO†
Programa de Engenharia Química/COPPE, Universidade Federal do Rio de Janeiro, CP 68502, 21945 Rio de Janeiro-RJ, Brasil

(*First received* 9 *November* 1988; *accepted in revised form* 21 *July* 1989)

Abstract—Six different strategies for sequential design for parameter estimation are analysed. Two of them have not been described before. According to the examples studied, the β-trace design criterion, based on the minimization of the trace of the posterior covariance matrix of parameters, seems to be the best. This criterion did not present most of the drawbacks presented by the other five criteria and provided the most accurate parameter estimates.

INTRODUCTION

Experiments are normally conceived and carried out to establish qualitative and quantitative relationships among the variables of a specific problem. However, as experiments are often expensive and time-consuming, their choice must be somehow optimized to reduce the number of experiments necessary to accomplish a specific task.

Generally, the qualitative analysis of a problem is related to the investigation of functional relationships among those variables. This analysis is often called model discrimination. Box and Hill (1967) were the first to establish a formal criterion for discriminating among mechanistic models. An interesting review on this subject is presented by Hill (1978). More recent results are presented by Ferraris and Forzatti (1983) and Ferraris *et al.* (1984).

When functional relationships among those variables are already known, experiments are usually performed for some model parameters to be evaluated. This is the so-called parameter evaluation (or parameter estimation) problem. Box and Lucas (1959) and Box and Hunter (1965) pioneered the formal development of optimal strategies for parameter estimation of mechanistic models. Their basic idea was to perform experiments sequentially, minimizing the volume of the confidence region formed by the estimated parameters in the parameter space. Afterwards, their original scheme was adapted to the evaluation of a subset from the original set of parameters by Hill and Hunter (1974).

Hill *et al.* (1968) presented a general criterion for sequential experimental design for both parameter estimation and model discrimination. However, according to Hill (1978), the method leads to worse parameter estimation and worse model discrimination.

Other design criteria for parameter evaluation were presented by Hosten (1974) and Pritchard and Bacon (1978). Hosten proposed choosing the experiment that would minimize the largest axis of the joint confidence hyperellipsoid, rather than minimizing its volume, as the next experiment to be performed. It is a shape design. Pritchard and Bacon proposed minimizing the correlation among the estimated parameters. However, according to Agarwal and Brisk (1985), these methods are not better than Box and coworkers' original one.

Practical applications of some of these methods to kinetic studies can be found in Froment (1975), Froment and Bischoff (1979) and Agarwal and Brisk (1985). Sequential experimental designs suited to dynamical systems are presented by Hosten and Emig (1975), Murray and Reiff (1984) and Kalogerakis and Luus (1984). Most of the basic theory can be found in Fedorov (1972).

In this work, two new criteria, based on the original scheme of Box and Lucas (1959), are developed and compared with existing ones. The various design criteria are applied to two simple examples and to a classical kinetic study of ethanol dehydrogenation by Franckaertz and Froment (1964). It is shown that difficulties can arise from the application of simplified criteria and that better results can be obtained with shape designs.

THEORY

Let

$$\mathbf{y} = \mathbf{f}(\mathbf{x}, \boldsymbol{\beta}) \tag{1}$$

be the functional relationship among the output variables, $y_i\,(i = 1, \ldots, n)$, the input variables, $x_u\,(u = 1, \ldots, m)$, and a set of parameters, $\beta_q\,(q = 1, \ldots, p)$, where $\mathbf{y}$, $\mathbf{x}$ and $\boldsymbol{\beta}$ are, respectively, n-dimensional, m-dimensional and p-dimensional real vectors, and $\mathbf{f}$ is an n-dimensional vector of real smooth functions.

Let $G_{\beta j}(\hat{\boldsymbol{\beta}}_j, \underline{\mathbf{V}}_j)$ be a Gaussian distribution that represents the present knowledge of $\boldsymbol{\beta}$, after j experiments. If l additional experiments are now performed and if $G_{\varepsilon k}(\mathbf{O}, \underline{\mathbf{V}}_{\varepsilon k})$ is a normal distribution that represents experimental deviations $\boldsymbol{\varepsilon}_k$, where $\boldsymbol{\varepsilon}_k$ is given by

$$\boldsymbol{\varepsilon}_k = \mathbf{y}_k - \mathbf{f}(\mathbf{x}_k, \hat{\boldsymbol{\beta}}_j) \tag{2}$$

†Author to whom correspondence should be addressed.

and $\mathbf{x}_k$ and $\mathbf{y}_k$ are measured in the kth experiment ($k = j+1, \ldots, j+l$), then the new Gaussian distribution G_β associated with the parameter vector $\boldsymbol{\beta}$ will be $G_{\beta j+l}(\hat{\boldsymbol{\beta}}_{j+l}, \hat{\underline{\mathbf{V}}}_{j+l})$, where $\hat{\underline{\mathbf{V}}}_{j+l}$ is the posterior covariance matrix of parameters, given by (Bard, 1974)

$$\hat{\underline{\mathbf{V}}}_{j+l} \cong \left(\sum_{k=j+1}^{j+l} \underline{\mathbf{B}}_k^t \, \underline{\mathbf{V}}_{\varepsilon k}^{-1} \, \underline{\mathbf{B}}_k + \underline{\mathbf{V}}_j^{-1} \right)^{-1} \tag{3}$$

where $\underline{\mathbf{B}}_k = (b_{iq})$ is the sensitivity matrix evaluated at $\mathbf{x} = \mathbf{x}_k$ and $\boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_j$, with

$$b_{iq} = \frac{\partial f_i}{\partial \beta_q}. \tag{4}$$

Therefore it may be written as

$$\hat{\underline{\mathbf{V}}}_{j+l} = \hat{\underline{\mathbf{V}}}_{j+l}(\mathbf{x}_{j+1}, \ldots, \mathbf{x}_{j+l}) \tag{5}$$

The posterior covariance matrix $\hat{\underline{\mathbf{V}}}_{j+l}$ is symmetric, positive-definite and can be associated with the hyperellipsoid in the parameter space:

$$|(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_{j+l})^t \, \hat{\underline{\mathbf{V}}}_{j+l}^{-1} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}_{j+l})| \leqslant z \tag{6}$$

where z depends on the number of degrees of freedom, the confidence level and the experimental error (Bard, 1974). This region in the parameter space is the joint confidence region. The greater the region, the smaller the confidence in the estimated parameters.

The basic idea of Box and Lucas (1959) and Box and Hunter (1965) was to perform the l experiments that would minimize the volume of the hyperellipsoid defined by eq. (6). As this volume is proportional to the square root of the determinant of $\hat{\underline{\mathbf{V}}}_{j+l}$, the chosen l experiments must be those which either minimize $\det(\hat{\underline{\mathbf{V}}}_{j+l})$ or maximize $\det(\hat{\underline{\mathbf{V}}}_{j+l}^{-1})$. According to Kiefer (1959) this criterion is a D-optimum criterion.

If one makes

$$\underline{\Pi} = \begin{pmatrix} \underline{\mathbf{V}}_{\varepsilon j+1} & \underline{\mathbf{0}} & \cdots & \underline{\mathbf{0}} \\ \underline{\mathbf{0}} & \underline{\mathbf{V}}_{\varepsilon j+2} & \cdots & \underline{\mathbf{0}} \\ \vdots & \vdots & & \vdots \\ \underline{\mathbf{0}} & \underline{\mathbf{0}} & \cdots & \underline{\mathbf{V}}_{\varepsilon j+l} \end{pmatrix} \tag{7}$$

and

$$\underline{\mathbf{B}}' = \begin{pmatrix} \underline{\mathbf{B}}_{j+1} \\ \underline{\mathbf{B}}_{j+2} \\ \vdots \\ \underline{\mathbf{B}}_{j+l} \end{pmatrix} \tag{8}$$

then eq. (3) can be written as

$$\hat{\underline{\mathbf{V}}}_{j+l} \cong [\underline{\mathbf{B}}'^t \underline{\Pi}^{-1} \underline{\mathbf{B}}' + \underline{\mathbf{V}}_j^{-1}]^{-1}. \tag{9}$$

After some rearrangement (Bard, 1974), it follows that

$$\det(\hat{\underline{\mathbf{V}}}_{j+l}^{-1}) \cong \det(\underline{\mathbf{V}}_j^{-1}) \det(\underline{\Pi}^{-1}) \det(\underline{\Pi} + \underline{\mathbf{B}}' \underline{\mathbf{V}}_j \underline{\mathbf{B}}'^t). \tag{10}$$

Thus, the l experiments that must be performed are those which either:

(a) maximize $\det(\underline{\mathbf{B}}'^t \underline{\Pi}^{-1} \underline{\mathbf{B}}' + \underline{\mathbf{V}}_j^{-1})$

or

(b) maximize $\det(\underline{\Pi} + \underline{\mathbf{B}}' \underline{\mathbf{V}}_j \underline{\mathbf{B}}'^t)$.

According to Bard (1974) the matrix $(\underline{\mathbf{B}}'^t \underline{\Pi}^{-1} \underline{\mathbf{B}}' + \underline{\mathbf{V}}_j^{-1})$ is related to the confidence in the parameter estimates, while the matrix $(\underline{\Pi} + \underline{\mathbf{B}}' \underline{\mathbf{V}}_j \underline{\mathbf{B}}'^t)$ is related to the confidence in the estimated output variables. From now on, both criteria (a) and (b) will be called the volume criterion (VC) without any distinction, since both of them lead to the same results.

Assuming that:

(1) $\mathbf{f}$ is linear in regard to the model parameters;

(2) $\underline{\mathbf{V}}_{\varepsilon k}$ is diagonal, homoscedastic and does not depend on the experimental conditions (so it is independent of k);

(3) the elements of $\underline{\mathbf{V}}_j$ are large enough;

then it can be concluded that the best set of experiments is that set that maximizes $\det(\underline{\mathbf{B}}'^t \underline{\mathbf{B}}')$. Although these conditions are rather restrictive, this is the criterion most often used in kinetic studies. From now on, it will be called the simple volume criterion (SVC).

As the axes of the joint confidence hyperellipsoid generated by eq. (6) are proportional to the square roots of the eigenvalues of $\hat{\underline{\mathbf{V}}}_{j+l}$, these eigenvalues give a good measure of the confidence in the parameters. Hosten (1974) proposed that the chosen set of experiments should minimize the largest eigenvalue of $\hat{\underline{\mathbf{V}}}_{j+l}$, so that the confidence region would be as round as possible. According to Hosten, the eigenvalues should be computed from the matrix $(\underline{\mathbf{B}}'^t \underline{\mathbf{B}}')$, in accordance with the simplifications already described. From now on, this criterion will be called the simple shape criterion (SSC). If eigenvalues are computed from eq. (9), it will be called the shape criterion (SC).

Two modifications are now introduced. First it is proposed that the best set of experiments is that set which minimizes the summation of the eigenvalues of $\hat{\underline{\mathbf{V}}}_{j+l}$. According to Kiefer (1959) it is an A-optimum criterion. If $\boldsymbol{\lambda}$ is the vector of eigenvalues of $\hat{\underline{\mathbf{V}}}_{j+l}$, then

$$\sum_{i=1}^{p} \lambda_i = \operatorname{tr}(\hat{\underline{\mathbf{V}}}_{j+l}). \tag{11}$$

Thus, the best set of experiments is the one that minimizes the trace of $\hat{\underline{\mathbf{V}}}_{j+l}$. From now on, it will be called the β-trace criterion (βTC). It can also be considered a kind of shape design.

As already said, $(\underline{\Pi} + \underline{\mathbf{B}}' \underline{\mathbf{V}}_j \underline{\mathbf{B}}'^t)$ is related to the confidence in the estimated output variables. It is proposed to choose the set of experiments that maximizes the summation of the eigenvalues of $(\underline{\mathbf{B}}' \underline{\mathbf{V}}_j \underline{\mathbf{B}}'^t)$, so that the set of output variables is placed in the most unknown region of the output space, where a good investigation is needed. From now on, it will be called the y-trace criterion (YTC).

PRELIMINARY ANALYSIS

The greatest failure of the VC is to give excessive importance to that parameter to which the model is most sensitive. Thus, although the volume of the confidence region can be reduced faster, as a consequence of faster reduction of the uncertainty of one specific model parameter, the uncertainties of all the other model parameters may remain very large. Therefore, volume criteria are good for prediction-oriented studies, but not well suited for the general parameter estimation problem.

As already said, the SVC is the most used in kinetic studies, in spite of not being very rigorous. Being a simple derivation of the more general VC, it has the same drawbacks described in the last paragraph and also other imperfections. From the point of view of sequential experimental design, the simple volume criterion presents the interesting property of allowing the design of all experiments at once, whenever good parameter estimates are available, since uncertainties do not appear in $\det(\underline{B}''\underline{B}')$. However, the method does not account for changes in the parameter covariance matrix and does not distinguish between experiments with different levels of experimental error, which is certainly a major drawback. Besides, let the mathematical model be

$$\begin{aligned} y_i &= f_i(\mathbf{x}, \boldsymbol{\beta}) \qquad (i = 1, \ldots, h) \\ y_i &= b_i + \sum_{k=1}^{n} a_{ik} y_k \qquad (i = h+1, \ldots, n) \end{aligned} \tag{12}$$

where b_i and a_{ik} are numerical constants. In such a case:

$$\frac{\partial y_i}{\partial \beta_q} = \sum_{k=1}^{n} a_{ik} \frac{\partial y_k}{\partial \beta_q} \qquad (i = h+1, \ldots, n) \tag{13}$$

so that the partial derivatives of y_i in respect to β_q are linear combinations of the other partial derivatives, when i is greater than h. When that is the case, $\det(\underline{B}''\underline{B}')$ can be singular everywhere, depending on the special arrangement of the number of parameters, number of equations and number of linear combinations in the model. If it really happens and the SVC is to be used, the linear equations must be discarded and a great amount of information is lost. It is worthwhile to notice that the model presented by eq. (12) is a generic kinetic model, where the linear equations are the stoichiometric relations and the mass balance equations.

The SSC has most of the imperfections described above, except that it is not a prediction-oriented criterion. The basic goal of Hosten's criterion is to reduce the uncertainty of the most uncertain parameter as much as possible. Nevertheless, the predictive capacity of the model is prejudiced to some extent. With SC, it is not necessary to discard information and uncertainties are also considered during the design of the next experiment. According to Agarwal and Brisk (1985) however, shape designs present slower convergence toward reasonable values than volume designs.

The βTC represents a transition from volume to shape design, as shown in the Appendix. This criterion is prediction-oriented when the uncertainties of the parameters are not very different, but it is parameter-oriented when the opposite happens, what seems to be a good property. Besides, if eq. (9) is used, it has none of the defects described before regarding linearity and uncertainties. It must also be stressed that, due to eq. (11), it is not necessary to calculate the individual eigenvalues of $\hat{\underline{V}}_{j+1}$ when the βTC is used.

The YTC is completely prediction-oriented, since experiments are designed to reduce the uncertainty in the prediction of output variables. It must also be noticed that the YTC does not require the computation of any determinants, eigenvalues or inverse matrixes. This may be particularly interesting when the model has a great number of parameters or when the time spent in model calculations is equivalent to or smaller than that spent in matrix manipulations.

EXPERIMENTS AND RESULTS

Three examples are presented now. The first and second experiments are simulations prepared to show the particular characteristics of each different design criterion analysed. The third example is a classical study on the dehydrogenation of ethanol. The general experimental procedure used was the following: an initial set of experiments was performed, parameters were evaluated, and, if results were not satisfactory, a new experiment was designed and performed.

A general version of the algorithm presented by Anderson *et al.* (1978) was used for parameter estimation. A slight modification of the algorithm presented by Law and Bailey (1963) was used to assure convergence.

Example 1

Let a mathematical model be represented by the following system of equations:

$$\begin{aligned} y_1 &= \beta_1 / x_1 \\ y_2 &= \beta_2 x_1^2 . \end{aligned} \tag{14}$$

Supposing that this system is homoscedastic and that errors are normally distributed with a standard deviation equal to 0.1 and with zero mean value, the set of experiments presented in Table 1 was generated. Both β_1 and β_2 were considered to be equal to 2.00.

Using the experimental data presented in Table 1, the following results can be obtained:

$$\begin{aligned} \hat{\beta}_1 &= 1.8620 \\ \hat{\beta}_2 &= 2.0400 \\ \underline{V}_3 &= \begin{pmatrix} 6.0515 \times 10^{-2} & 0 \\ 0 & 3.3722 \times 10^{-5} \end{pmatrix}. \end{aligned} \tag{15}$$

It is clear from the results above that β_2 is better evaluated than β_1.

If we are to use the SVC or SSC, it may be easily seen that the sequence of experiments depends neither on the parameters nor on the uncertainties. Supposing that the experiments that may be carried out are

Table 1. Initial set of experiments: example 1

x_1	y_1	y_2
0.441	4.201	0.400
2.116	0.960	8.571
9.902	0.230	200.00

Table 2. Experimental grid for parameter estimation: example 1

x_1			
0.1	0.6	2.0	7.0
0.2	0.7	3.0	8.0
0.3	0.8	4.0	9.0
0.4	0.9	5.0	10.0
0.5	1.0	6.0	

those presented in Table 2, the next experiment to be performed must be $x_1 = 10.0$ and $x_1 = 1.0$, according to the SVC and SSC, respectively.

In order to apply any of the other design criteria, some additional calculations are necessary. According to the VC, SC, βTC and YTC, the next experiment to be performed must be $x_1 = 0.1$.

After performing the fourth experiment, one can obtain the results presented in Table 3.

Example 2

Let a mathematical model be represented by the following system of equations:

$$y_1 = 0.1\frac{\exp(-\beta_1 x_1)}{x_1}$$
$$y_2 = \exp(\beta_2 x_1)/x_1. \tag{16}$$

Supposing that y_1 and y_2 are measured with different levels of experimental error, it can be written that

$$\underline{\hat{V}}_{j+1} = \begin{bmatrix} \dfrac{\sigma_{y1}^2}{0.01\exp(-2\hat{\beta}_1 x_1) + \alpha_1} & 0 \\ 0 & \dfrac{\sigma_{y2}^2}{\exp(2\hat{\beta}_2 x_1) + \alpha_2} \end{bmatrix} \tag{17}$$

where $\alpha_1 = \sigma_{y1}^2/\sigma_{\beta 1}^2$ and $\alpha_2 = \sigma_{y2}^2/\sigma_{\beta 2}^2$.

So the eigenvalues of $\underline{\hat{V}}_{j+1}$ are

$$\lambda_1 = \sigma_{y1}^2 \frac{100\exp(2\hat{\beta}_1 x_1)}{1 + 100\alpha_1\exp(2\hat{\beta}_1 x_1)} \tag{18}$$

$$\lambda_2 = \sigma_{y1}^2 \frac{K\exp(-2\hat{\beta}_2 x_1)}{1 + \alpha_2\exp(-2\hat{\beta}_2 x_1)} \tag{19}$$

where $K = \sigma_{y2}^2/\sigma_{y1}^2$.

According to the SC, it is easy to see that the best experiment of all is that one for which

$\lambda_1 = \lambda_2$

$$\frac{100\exp(2\hat{\beta}_1 x_1)}{1 + 100\alpha_1\exp(2\hat{\beta}_1 x_1)} = \frac{K\exp(-2\hat{\beta}_2 x_1)}{1 + \alpha_2\exp(-2\hat{\beta}_2 x_1)}. \tag{20}$$

As y_2 is a monotonic increasing function of x_1 and y_1 is a monotonic decreasing function of x_1, it is reasonable to suppose that

(1) $K > 1$,
(2) $\alpha_1 < 1$,
(3) $\alpha_2 > 1$.

Let K be equal to 2, α_2 be equal to 2, and α_1 be equal to 0.5. Then, according to the SC, if β_1 is equal to 2 and β_2 is equal to 0.25, the best experiment to be carried out is $x_1 = 1.1$.

According to the β-trace design, the best x_1 is that value that minimizes the function

$$\frac{100\exp(2\hat{\beta}_1 x_1)}{1 + 100\alpha_1\exp(2\hat{\beta}_1 x_1)} + \frac{K\exp(-2\hat{\beta}_2 x_1)}{1 + \alpha_2\exp(-2\hat{\beta}_2 x_1)} = \text{minimum}. \tag{21}$$

The solution is to use the largest x_1 available for experimentation.

It is easy to show that λ_1 is constrained between 1.96 and 2.00, and λ_2 between 0.67 and zero, when x_1 is greater than zero, so, when the SC is used, no significant improvement is obtained in the estimation of both β_1 and β_2. When the βTC is used, no significant improvement is obtained in the estimation of β_1; however, β_2 can be improved significantly. If K and α_2 are changed, no qualitative change is introduced in the analysis.

Table 3. Results after the new experiment: example 1

Criterion	x_1	y_1	y_2	β_1 (95%)	β_2 (95%)
SVC	10.0	0.214	200.05	1.8614 ± 0.7	2.0190 ± 0.01
SSC	1.0	1.965	2.087	1.8772 ± 0.6	2.0395 ± 0.01
VC, SC, βTC, YTC	0.1	19.63	0.037	1.9966 ± 0.09	2.0395 ± 0.01

Example 3

Experimental data on the dehydrogenation of ethanol in the presence of water vapour presented by Franckaertz (1963) and Franckaertz and Froment (1964) are analysed in this example.

The dehydrogenation of ethanol can be written as

$$CH_3CH_2OH \rightarrow CH_3CHO + H_2. \quad (22)$$

According to them, the unique significant parallel reaction was the ethanal condensation, given by

$$2CH_3CHO \rightarrow CH_3COOC_2H_5. \quad (23)$$

It was observed experimentally, however, that the amount of ethyl acetate produced could be calculated from

$$x_e = 0.0353 x_T \quad (24)$$

where x_T is the total conversion of ethanol, and x_e is the fraction of ethanol converted to ethyl acetate.

According to Franckaertz and Froment, the reactor model could be given by

$$x_T = \int_0^W \frac{r}{F_{AO}} d\omega \quad (25)$$

where

$$r = \frac{k[P_A - (P_R P_S/K)]}{(1 + k_A P_A + k_R P_R)^2} \quad (26)$$

and

$$K = \exp[(-14{,}159.9/T) + 25.2681] \quad (27)$$

T in Kelvin

where P_i are the partial pressures, k_i are the adsorption coefficients, and k is the rate constant.

Among more than 500 experimental data presented by Franckaertz (1963), we decided to analyse those data obtained at 225, 250 and 275°C. Following the same procedure adopted by Franckaertz and Froment (1964), the parameters k, k_A and k_R were estimated at each of these temperatures and afterwards they were expressed in accordance with the Arrhenius' equation. At each temperature, all of the design criteria were used to simulate a sequential experimental design for parameter estimation. The results obtained for sets of eight experiments are presented in Tables 4–10 and Figs 1–3. The sequences of experiments that resulted from the application of each design criterion at each different temperature are presented in Tables 4–6. The parameters evaluated after each sequence of experiments are presented in Tables 7 and 8. Tables 9 and 10 show the predictive capacity of each set of parameters, when all experiments caried out by Franckaertz were analysed. The evolution of the confidence in the parameters is presented in Figs 1–3.

It is important to notice that we do not present results for the simple shape design because it introduced some numerical difficulties to the problem (eigenvalues of $\underline{B}'\underline{B}$ were too small, because this matrix is singular everywhere). In spite of this difficulty, we decided to maintain the simple volume design.

The first four experiments of all different sequences are the same, to provide a common starting point of

Table 4. Sequence of experiments obtained with the VC and YTC: example 3

Temperature (°C)	W/F_{AO}	P	X_A	X_W	X_R	x_T
225	1.60	7	0.865	0.135	0.000	0.066
	0.80	4	0.865	0.135	0.000	0.083
	0.40	3	0.865	0.135	0.000	0.055
	1.00	1	0.865	0.135	0.000	0.118
	1.00	1	0.750	0.130	0.119	0.052
	0.40	1	0.865	0.135	0.000	0.060
	1.00	1	0.732	0.167	0.101	0.052
	0.40	10	0.865	0.135	0.000	0.038
250	1.60	7	0.865	0.135	0.000	0.149
	0.80	4	0.865	0.135	0.000	0.157
	0.40	3	0.865	0.135	0.000	0.108
	1.00	1	0.865	0.135	0.000	0.218
	1.00	1	0.672	0.145	0.183	0.123
	0.60	1	0.865	0.135	0.000	0.152
	0.80	1	0.672	0.145	0.183	0.106
	0.60	10	0.865	0.135	0.000	0.094
275	1.60	7	0.865	0.135	0.000	0.254
	0.80	4	0.865	0.135	0.000	0.262
	0.40	3	0.865	0.135	0.000	0.200
	1.00	1	0.865	0.135	0.000	0.362
	1.00	1	0.672	0.145	0.183	0.230
	0.20	1	0.865	0.135	0.000	0.118
	0.40	10	0.865	0.135	0.000	0.148
	0.40	1	0.865	0.135	0.000	0.196

Table 5. Sequence of experiments obtained with the SVC: example 3

Temperature (°C)	W/F_{AO}	P	X_A	X_W	X_R	x_T
225	1.60	7	0.865	0.135	0.000	0.066
	0.80	4	0.865	0.135	0.000	0.083
	0.40	3	0.865	0.135	0.000	0.055
	1.00	1	0.865	0.135	0.000	0.118
	1.53	1	0.865	0.135	0.000	0.153
	1.60	1	0.865	0.135	0.000	0.155
	1.40	1	0.865	0.135	0.000	0.144
	0.60	1	0.796	0.158	0.046	0.064
250	1.60	7	0.865	0.135	0.000	0.149
	0.80	4	0.865	0.135	0.000	0.157
	0.40	3	0.865	0.135	0.000	0.108
	1.00	1	0.865	0.135	0.000	0.218
	0.80	1	0.796	0.158	0.046	0.154
	0.85	1	0.796	0.158	0.046	0.162
	1.40	1	0.865	0.135	0.000	0.257
	0.80	3	0.865	0.135	0.000	0.154
275	1.60	7	0.865	0.135	0.000	0.254
	0.80	4	0.865	0.135	0.000	0.262
	0.40	3	0.865	0.135	0.000	0.200
	1.00	1	0.865	0.135	0.000	0.362
	0.60	3	0.865	0.135	0.000	0.250
	0.80	3	0.796	0.158	0.046	0.270
	0.80	1	0.796	0.158	0.046	0.290
	1.60	1	0.865	0.135	0.000	0.454

Table 6. Sequence of experiments obtained with the βTC and SC: example 3

Temperature (°C)	W/F_{AO}	P	X_A	X_W	X_R	x_T
225	1.60	7	0.865	0.135	0.000	0.066
	0.80	4	0.865	0.135	0.000	0.083
	0.40	3	0.865	0.135	0.000	0.055
	1.00	1	0.865	0.135	0.000	0.118
	0.20	1	0.865	0.135	0.000	0.035
	0.85	3	0.796	0.158	0.040	0.072
	1.00	1	0.672	0.145	0.183	0.048
	0.40	1	0.865	0.135	0.000	0.060
250	1.60	7	0.865	0.135	0.000	0.149
	0.80	4	0.865	0.135	0.000	0.157
	0.40	3	0.865	0.135	0.000	0.108
	1.00	1	0.865	0.135	0.000	0.218
	1.00	1	0.672	0.145	0.183	0.123
	0.40	1	0.865	0.135	0.000	0.111
	0.20	1	0.865	0.135	0.000	0.065
	0.85	3	0.796	0.158	0.046	0.147
275	1.60	7	0.865	0.135	0.000	0.254
	0.80	4	0.865	0.135	0.000	0.262
	0.40	3	0.865	0.135	0.000	0.200
	1.00	1	0.865	0.135	0.000	0.362
	0.20	1	0.865	0.135	0.000	0.118
	1.60	3	0.865	0.135	0.000	0.352
	0.40†	1	0.865	0.135	0.000	0.196
	1.60‡	4	0.865	0.135	0.000	0.320
	1.40†	3	0.865	0.135	0.000	0.338
	0.40‡	1	0.865	0.135	0.000	0.196

†βTC.
‡SC.

Table 7. Parameter estimates: example 3

Temperature (°C)	k (99%)	k_A (95%)	k_R (95%)	Criterion
225	0.63 ± 0.35	0.55 ± 0.30	11.04 ± 6.14	VC, YTC
	0.71 ± 0.65	0.75 ± 0.35	8.67 ± 12.96	SVC
	0.52 ± 0.23	0.64 ± 0.30	5.08 ± 2.96	SC, βTC
250	0.90 ± 0.14	0.53 ± 0.08	3.16 ± 0.76	VC, YTC
	1.08 ± 0.45	0.53 ± 0.17	5.00 ± 4.14	SVC
	0.92 ± 0.13	0.55 ± 0.09	3.12 ± 0.66	SC, βTC
275	1.74 ± 0.38	0.42 ± 0.13	3.03 ± 1.00	VC, YTC
	1.87 ± 0.51	0.54 ± 0.16	2.59 ± 1.54	SVC
	1.94 ± 0.42	0.51 ± 0.21	3.32 ± 0.89	SC
	1.93 ± 0.41	0.50 ± 0.20	3.31 ± 0.86	βTC

Table 8. Parameter estimates according to Arrhenius' equation: example 3

Parameter	ΔE	ΔS	Criterion
k	12,362	23.62	VC, YTC
	11,133	21.53	SVC
	15,379	29.35	SC, βTC
k_A†	− 2695	− 6.55	VC, YTC
	− 4047	− 8.76	SVC
	− 2727	− 6.37	SC, βTC
k_R†	− 19,737	− 34.88	VC, YTC
	− 12,419	− 20.63	SVC
	− 5571	− 8.04	SC, βTC

†$k = \exp(-\Delta E/RT + \Delta S/R)$.

analysis. They were carried out at different pressures and residence times, but with the same feed composition in order to simulate a common experimental procedure, since it is often much easier to vary pressures and flow rates than feed composition.

DISCUSSION

The main objective of example 1 is to show that simple design criteria are frequently inadequate for experimental design. That simple example shows that not taking all available information into consideration may lead to much poorer results. According to Table 3, it is clear that simple design criteria are not effective for the estimation of the parameter β_1 because they lead us to experiment in regions where output variables are not very sensitive to β_1. This happens only because these criteria do not take confidence in the parameters into consideration. If confidence in β_2 were considered, it would be concluded that β_2 had been much better evaluated than β_1 and that experimentation in regions where output variables were sensitive to β_1 would be necessary.

It can also be seen from Figs 1–3 that simple designs generally provide the less significant parameters, despite also providing good predictions of output variables, as shown in Tables 9 and 10. Remembering that numerical difficulties arise when applying the SSC in the third example, as has already been pointed out in the preliminary analysis, and that simple design criteria are computationally equivalent to the others, it can generally be said that using them to design experiments for parameter estimation is not the best choice.

The main objective of example 2 is to stress that the SC and βTC have fundamental differences, in spite of

Table 9. Predictive capacity of model parameters without using Arrhenius' equation: example 3

Temperature (°C)	$\overline{\Delta x_T}$	$\overline{\Delta x_T^2}$	Maximum Δx_T	Criterion
225	0.006	0.005	0.019	VC, YTC
	0.005	0.003	0.014	SVC
	0.005	0.004	0.020	SC, βTC
250	0.007	0.008	0.029	VC, YTC
	0.006	0.005	0.019	SVC
	0.008	0.008	0.029	SC, βTC
275	0.009	0.010	0.031	VC, YTC
	0.010	0.012	0.026	SVC
	0.010	0.012	0.031	SC, βTC

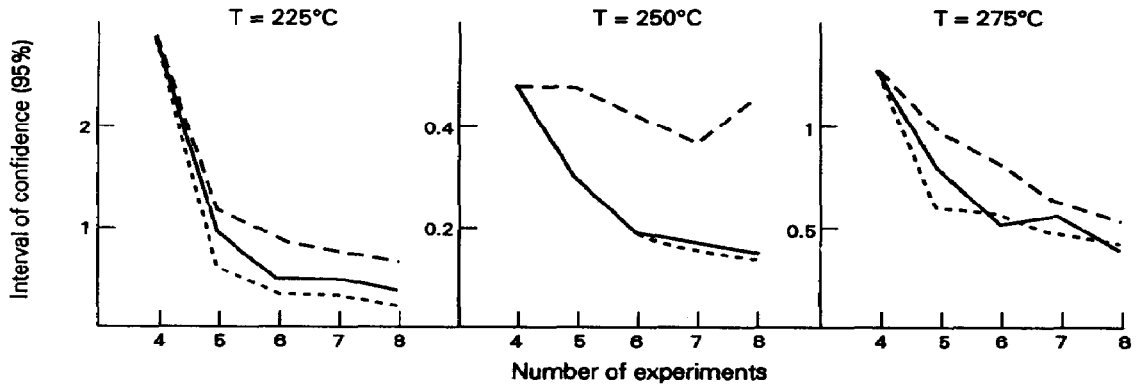

Fig. 1. Evolution of confidence limits of k. (———) y-trace design and volume design, (– – – –) simple volume design, (. . . . . .) β-trace design and shape design.

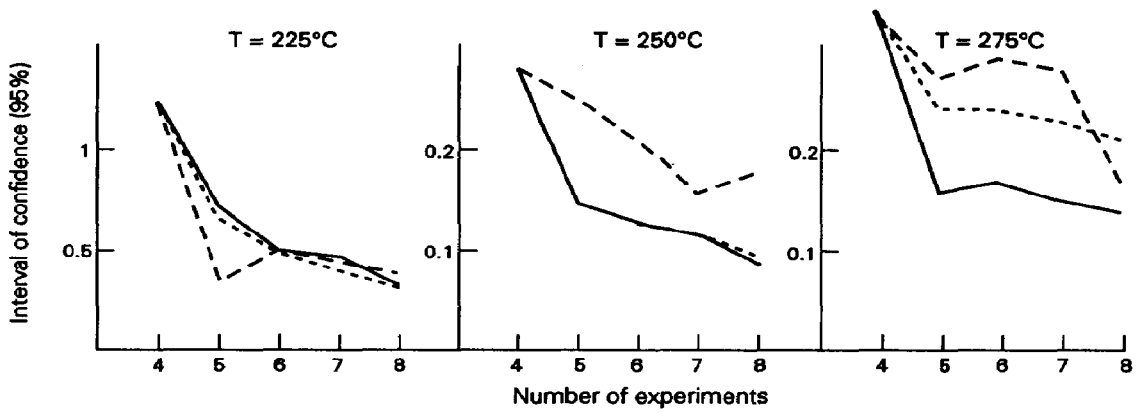

Fig. 2. Evolution of confidence limits of k_A. (———) y-trace design and volume design, (– – – –) simple volume design, (. . . . . .) β-trace design and shape design.

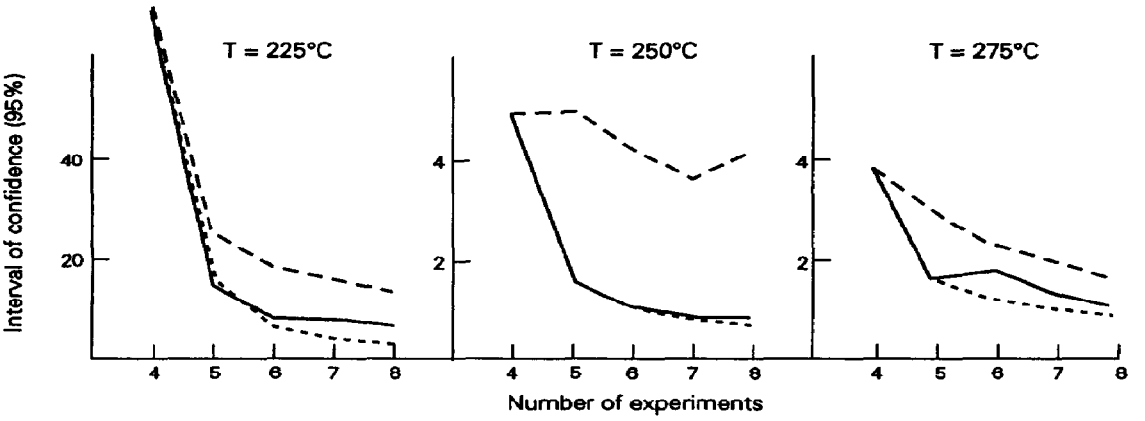

Fig. 3. Evolution of confidence limits of k_R. (———) y-trace design and volume design, (– – – –) simple volume design, (. . . . . .) β-trace design and shape design.

leading many times to similar sequences of experiments. When a general parameter estimation problem is somehow controlled by a parameter that cannot be well evaluated, and this is the case of example 2, the usage of the SC can prejudice the estimation of all other parameters considerably. In such a case, estimates will converge much slower to satisfactory results and a longer sequence of experiments will be necessary. If the βTC is used instead, the other parameters will be well evaluated and the convergence will

Table 10. Predictive capacity of model parameters with the usage of Arrhenius' equation: example 3

$\overline{\Delta x_T}$	$\overline{\Delta x_T^2}$	Maximum Δx_T	Criterion
0.014	0.020	0.057	VC, YTC
0.007	0.009	0.023	SVC
0.007	0.009	0.026	SC, βTC

be faster (see Appendix). In example 2. the VC and βTC provide the same results.

Example 3 is a practical application of sequential experimental design to a kinetic study. It can be seen from Tables 4–6 that all strategies generally provided good exploitation of the experimental grid. Experiments at different pressures, with different feed compositions and residence times were selected, no matter which design criterion was used. Nevertheless, it is important to notice that the interval of conversions analysed when the volume and y-trace design were used (Table 4) was much narrower than in the other cases.

It can be seen from Table 7 and Figs 1–3 that parameter estimates provided by the βTC and SSC are almost always better than the others. From Table 9, it can be seen that all sets of parameters can be considered to be equivalent to each other, if predictive capacity is confined to the temperature where they were evaluated. The same cannot be said when the Arrhenius' equation is used. In this case, results presented in Table 10 show that the VC and YTC provide parameters with worse predictive capacity. It seems to be a contradiction that prediction-oriented methods provide parameters with a worse predictive capacity. In reality, it is not only a contradiction, but it also happens very often.

If one assumes that the catalytic bed is long enough, so that the reaction reaches equilibrium, the total conversion will not depend on any of the three model parameters. Thus, high conversions do not contribute very much to the parameter estimation problem. Actually, the higher the value of k and the smaller the value of k_A, the smaller the catalytic bed must be for good parameter estimation. Besides, if it is assumed that the catalytic bed is very small, it can be written that

$$x_T = \frac{rW}{F_{AO}} \tag{28}$$

$$\frac{\partial x_T}{\partial k} = \frac{x_T}{k} \tag{29}$$

$$\frac{\partial x_T}{\partial k_A} = -\frac{x_T}{(1/2P_A) + (k_A/2)} \tag{30}$$

$$\frac{\partial x_T}{\partial k_R} \cong 0. \tag{31}$$

Therefore, it can be seen from Tables 4–7 and eqs (28)–(31) that k_A is the parameter to which x_T is most sensitive, specially at high temperatures and low conversions. For such a reason, trying to reduce the volume of the confidence region and improve the predictive capacity of the model, experiments are designed mostly to reduce the uncertainty of k_A. In this case, it means to carry out experiments at low conversions. However, it ends up biasing the results and reducing the predictive capacity of the model out of this region.

Although these three examples constitute a very small set of problems and conditions, other kinetic studies under way have been confirming all observations and facts presented in this article. Therefore, it can be said that all these phenomena are quite general.

CONCLUSIONS

Six different strategies for sequential experimental design for parameter estimation were analysed: the volume design criterion, simple volume design criterion, shape design criterion, simple shape design criterion, β-trace design criterion and y-trace design criterion. The last two criteria have not been described before. All design criteria but the y-trace and β-trace design criteria are computationally equivalent.

Simple design criteria are often not adequate for designing experiments for parameter estimation because they lead to less significant parameters and can introduce additional numerical difficulties into the original problem, such as singular determinants and ill-conditioned matrixes.

Giving excessive importance to the parameter to which the model is most sensitive, prediction-oriented criteria (the VC and YTC) can eventually generate biased results and poor parameter estimates.

When the estimation problem presents a parameter that cannot be well evaluated, the usage of the SC is inadequate because it causes poor estimation of all other parameters and also causes convergence to be too slow.

Being able to surmount all these problems in the examples presented, the βTC seems to be the best criterion for sequential experimental design for parameter estimation among those criteria analysed.

NOTATION

$\underline{\mathbf{B}}$	sensitivity matrix
$\underline{\mathbf{B}}'$	matrix defined by eq. (8)
det(.)	determinant of .
$\mathbf{f}$	vector of smooth real functions
F_{AO}	feed rate of ethanol, mol/h
G_ε	Gaussian distribution associated with experiments
G_β	Gaussian distribution associated with the parameters
k	kinetic constant, mol/h gcat
k	adsorption coefficient, atm^{-1}
K	equilibrium constant, atm
m	number of input variables

n	number of output variables
p	number of parameters
P	pressure, atm
P_i	partial pressure of i, atm
r	rate of reaction, mol/h gcat
T	absolute temperature, K
tr(.)	trace of .
$\underline{V}$	covariance matrix of parameters
$\hat{\underline{V}}$	posterior covariance matrix of parameters
$\underline{V}_e$	covariance matrix of output variables
$\mathbf{x}$	vector of input variables
x_e	fraction of ethanol converted to ethyl acetate
x_T	total conversion of ethanol
X_i	molar fraction of i
W	mass of catalyst, g
$\mathbf{y}$	vector of output variables

Greek letters

β	vector of parameters
Δx	absolute difference between measured and calculated values
ε	vector of experimental deviations
λ	vector of eigenvalues of $\hat{\underline{V}}$
$\underline{\Pi}$	matrix defined by eq. (7)
σ_i	standard deviation associated with variable i

Subscripts

A	ethanol
R	ethanal
S	hydrogen
W	water

Superscripts

t	transpose
$\hat{}$	estimate

REFERENCES

Agarwal, A. K. and Brisk, M. L., 1985, Sequential experimental design for precise parameter estimation. 2. Design criteria. *Ind. Engng Chem. Process Des. Dev.* **24**, 207–210.

Anderson, T. F., Abrams, D. S. and Grens II, E. A., 1978, Evaluation of parameters for nonlinear thermodynamic models. *A.I.Ch.E. J.* **24**, 20–29.

Bard, Y., 1974, *Nonlinear Parameter Estimation.* Academic Press, New York.

Box, G. E. P. and Hill, W. J., 1967, Discrimination among mechanistic models. *Technometrics* **9**, 57–71.

Box, G. E. P. and Hunter, W. G., 1965, *Proceedings—IBM Scientific Computing Symposium on Statistics*, p. 13.

Box, G. E. P. and Lucas, H. L., 1959, Design of experiments in non-linear situations. *Biometrika* **46**, 77–89.

Fedorov, V. V., 1972, *Theory of Optimal Experiment.* Academic Press, London.

Ferraris, G. B. and Forzatti, P., 1983, A new sequential experimental design procedure for discriminating among rival models. *Chem. Engng Sci.* **38**, 225–232.

Ferraris, G. B., Forzatti, P., Emig, G. and Hofmann, H., 1984, Sequential experimental design for model discrimination in the case of multiple responses. *Chem. Engng Sci.* **39**, 81–85.

Franckaertz, J., 1963, Kinetische studie van de katalytische dehydrogenatie van ethanol to acetaldehyde. PhD thesis, Rijksuniversiteit te Gent.

Franckaertz, J. and Froment, G. F., 1964, Kinetic study of the dehydrogenation of ethanol. *Chem. Engng Sci.* **19**, 807–818.

Froment, G. F., 1975, Model discrimination and parameter estimation in heterogeneous catalysis. *A.I.Ch.E. J.* **21**, 1041–1057.

Froment, G. F. and Bischoff, K. B., 1979, *Chemical Reactor Analysis and Design.* John Wiley, New York.

Hill, P. D. H., 1978, A review of experimental design procedure for regression model discrimination. *Technometrics* **20**, 15–21.

Hill, P. D. H. and Hunter, W. G., 1974, Design of experiments for subsets of parameters. *Technometrics* **16**, 425–434.

Hill, P. D. H., Hunter, W. G. and Wichern, D. W., 1968, A joint design criterion for the dual problem of model discrimination and parameter estimation. *Technometrics* **10**, 145–160.

Hosten, L. H., 1974, A sequential experimental design procedure for precise parameter estimation based on the shape of the joint confidence region. *Chem. Engng Sci.* **29**, 2247–2252.

Hosten, L. H. and Emig, G., 1975, Sequential experimental design procedures for precise parameter estimation in ordinary differential equations. *Chem. Engng Sci.* **30**, 1357–1364.

Kalogerakis, N. and Luus, R., 1984, Sequential experimental design of dynamic systems through the use of information index. *Can. J. chem. Engng* **62**, 730–737.

Kiefer, J., 1959, Optimum experimental designs. *J. R. statist. Soc. Ser. B* **21**, 272–319.

Law, V. J. and Bailey, R. V., 1963, A method for the determination of approximate system transfer functions. *Chem. Engng Sci.* **18**, 189–202.

Murray, L. E. and Reiff, E. K., 1984, Design of transient experiments for identification of fixed bed thermal transport properties. *Can. J. chem. Engng* **62**, 55–61.

Pritchard, D. J. and Bacon, D. W., 1978, Prospects for reducing correlations among parameter estimates in kinetic models. *Chem. Engng Sci.* **33**, 1539–1543.

APPENDIX

$$\det(\hat{\underline{V}}_{j+1}) = \prod_{i=1}^{p} \lambda_i \tag{A1}$$

$$\mathrm{tr}(\hat{\underline{V}}_{j+1}) = \sum_{i=1}^{p} \lambda_i. \tag{A2}$$

If $\lambda_1 \cong \lambda$ for any $i (1 \leqslant i \leqslant p)$, then

$$\det(\hat{\underline{V}}_{j+1}) \cong \lambda^p \tag{A3}$$

$$\mathrm{tr}(\hat{\underline{V}}_{j+1}) \cong p\lambda. \tag{A4}$$

Therefore the shape design, volume design and β-trace design criteria are similar when all λs are approximately equal throughout the experimental grid.

If $\lambda_1 \gg \lambda_i$ for any $i (2 \leqslant i \leqslant p)$, then

$$\mathrm{tr}(\hat{\underline{V}}_{j+1}) \cong \lambda_1. \tag{A5}$$

Therefore the shape design and β-trace design criteria are similar when any eigenvalue is much greater than the others (but not constant) throughout the experimental grid.

If any λ_1 is constant throughout the experimental grid, then

$$\mathrm{tr}(\hat{\underline{V}}_{j+1}) \cong \sum_{i=2}^{p} \lambda_i + \text{constant}. \tag{A6}$$

Therefore the parameter estimation will never be controlled by a poor parameter if the β-trace design criterion is used, because the minimum of eq. (A6) does not depend on λ_1. All parameters will be evaluated properly.