Automatica, Vol. 26, No. 2, pp. 195–213, 1990
Printed in Great Britain.

0005–1098/90 $3.00 + 0.00
Pergamon Press plc
© 1990 International Federation of Automatic Control

Qualitative and Quantitative Experiment Design for Phenomenological Models—A Survey*

E. WALTER† and L. PRONZATO†

The techniques reviewed can be used to decide, before collecting data for parameter estimation, if, where and how one will act on the system and where, how and when one will observe it.

Key Words—Identification; modeling; parameter estimation; identifiability; experiment design (not in the standard list).

Abstract—Designing an experiment for parameter estimation involves two steps. The first one is qualitative, and consists of selecting a suitable configuration of the input/output ports so as to make, if possible, all the parameters of interest identifiable. The second step is quantitative, and based on the optimization of a suitable criterion (with respect to the input shapes, sampling schedule, . . .) so as to get the maximum information from the data to be collected. When the model is nonlinear in the parameters, the typical situation for phenomenological models, both steps present specific difficulties which are discussed in this paper. The practical importance of qualitative experiment design is illustrated by a very simple biological model. Various policies presented in the literature for quantitative experiment design are reviewed. Special emphasis is given to methods allowing uncertainty on the prior information to be taken into account.

1. INTRODUCTION

THE CONFRONTATION of theories (or model structures) to data collected on the system to be described is essential to experimental sciences. This is necessary for selecting the model structure which explains the observations best, and for obtaining estimates of the unknown parameters involved in this structure.

Before collecting the data, the experiment must be designed, i.e. it must be decided if, where and how one will act on the system (location of input ports, type of actuators, shape of inputs . . .) and where, how and when one will observe it (location of output ports, type of sensors, sampling schedule . . .).

* Received 17 October 1988; revised 7 March 1989; received in final form 27 April 1989. The original version of this paper was presented at the 1st IFAC Symposium on Modelling and Control in Biomedical Systems, Venice, Italy, 6–8 April 1988. This paper was recommended for publication by Survey Editor K. J. Åström.

† Laboratoire des Signaux et Systèmes, CNRS/Ecole Supérieure d'Electricité, Plateau de Moulon, F-91192 Gif-sur-Yvette Cedex, France.

The *final purpose* of the experiment must be clearly defined. Possible situations are as follows. (i) One wants to get a better understanding of some physical phenomenon. One might then be interested in precisely estimating (some of) the model parameters. (ii) One wants to predict the process behaviour, or to control it in an optimal way. (iii) One wants to select the best description of a physical process among different model structures.

This paper is centred on the *parameter estimation* problem (i) for models the output of which is *nonlinear in the parameters*. Some tools concerning situation (ii) can be found in Brooks (1974, 1977), Goodwin and Payne (1977), Gevers and Ljung (1986) and O'Hagan (1978), with application to target attainment (Verdinelli and Wynn, 1988), and to quality control (see the brief survey (Pukelsheim, 1988)). Experiment design for model discrimination (iii) has received considerable attention, and one can refer to the survey papers by Atkinson and Cox (1974), and Hill (1978), together with the recent results of Fedorov and Khabarov (1986) and Fedorov (1987) that clarify the strong connection existing between experiment designs for parameter estimation and for model discrimination.

While experimenters and mathematicians agree on the fact that the quality of the model to be obtained depends heavily on that of the data collected, it must be admitted that they often disagree on the best way for designing an experiment. At the risk of being slightly provocative, one might say that most experimenters think of experiment design as an art owing much to intuition and very little to mathematics, while mathematicians view it as an essentially statistical problem. While we do not

question the obvious importance of the experimenter's intuition and prior knowledge, we shall try to show that mathematical treatment can also be helpful in designing an appropriate experiment, for at least two reasons.

First, a configuration of the input and output ports which seems reasonable intuitively, may nevertheless lead to the impossibility of obtaining a single optimal estimate for the parameters of interest, even for very simple models. It is therefore necessary to select the input/output port configuration to avoid such defects as far as possible. This can be done using the notions of structural identifiability and structural distinguishability, which allow one to know whether the right model structure and true value of its parameters could be recovered from noise-free data. This qualitative part of experiment design is addressed in Section 3.

Second, one must take into account the fact that measurements are often scarce and far from noise free, so that the quality of the parameter estimates will depend heavily on the input shape and measurement schedule. A quantitative design, based on the optimization of a suitable criterion, is then essential so as to return the maximum information from the data to be collected. Such a criterion depends on prior assumptions on the noise statistics, and, in general, on the value of the parameters to be estimated. It is therefore of special importance to take into account a suitable characterization of prior knowledge (or prior uncertainty) on the process under study. Depending on the reliability of prior information available, various policies can be considered and are reviewed in Section 4.

Section 2 briefly recalls the two types of nonlinearities that can affect a model and are of consequence in experiment design.

2. NONLINEARITIES

We shall concern ourselves with two types of nonlinearities, namely, in the parameters and in the inputs. Since each type of nonlinearity is associated with specific problems when designing an experiment, it seems useful to stress this point first. Here $\mathbf{y}_m(\boldsymbol{\theta}, \mathbf{u}, t)$ will be the outputs at time t of the model with parameters $\boldsymbol{\theta}$ when the inputs $\mathbf{u}(.)$ have been applied between times 0 and t, the model being in zero initial conditions.

A model is nonlinear in the inputs (non-LI) if its output does not satisfy the superposition principle concerning the inputs, i.e. if there exist some real scalars λ and μ, some input functions $\mathbf{u}_1(.)$ and $\mathbf{u}_2(.)$ and some positive time t such that

$$\mathbf{y}_m(\boldsymbol{\theta}, \lambda\mathbf{u}_1 + \mu\mathbf{u}_2, t) \neq \lambda\mathbf{y}_m(\boldsymbol{\theta}, \mathbf{u}_1, t) + \mu\mathbf{y}_m(\boldsymbol{\theta}, \mathbf{u}_2, t). \quad (1)$$

When control theoreticians speak of nonlinear models, they usually refer to this type of nonlinearity.

A model is nonlinear in the parameters (non-LP) if its output does not satisfy the superposition principle concerning the parameters, i.e. if there exist some real scalars λ and μ, some parameter vectors $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ and some positive time t such that

$$\mathbf{y}_m(\lambda\boldsymbol{\theta}_1 + \mu\boldsymbol{\theta}_2, \mathbf{u}, t) \neq \lambda\mathbf{y}_m(\boldsymbol{\theta}_1, \mathbf{u}, t) + \mu\mathbf{y}_m(\boldsymbol{\theta}_2, \mathbf{u}, t). \quad (2)$$

When statisticians speak of nonlinear estimation, they usually refer to this type of nonlinearity. Whether LI or non-LI, most phenomenological models are non-LP.

3. QUALITATIVE EXPERIMENT DESIGN

By qualitative experiment design, we mean the selection of input and output ports ensuring that the largest possible number of the parameters of interest are globally (or at least locally) identifiable. It can be performed in the absence of any prior information on the value of the parameters and even before collecting any data on the system to be studied. Section 3.1 recalls the definitions of identifiability to be used. In order to stress the need for such a qualitative design, we describe in Section 3.2 a very simple phenomenological model, with many possible output ports, but which nevertheless turns out to be rather deficient concerning identifiability. Section 3.3 reviews various tools that can be used for qualitative experiment design.

3.1. *Structural identifiability*

Anyone involved in estimating the parameters of a model $M(.)$ from measurements would like to know *beforehand* whether there is any chance of obtaining a single best value for the parameters, given the structure of the model considered and the type of measurements to be performed. Denote the equality of the model input/output maps obtained for two values $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$ of the parameter vector by

$$M(\boldsymbol{\theta}) \approx M(\boldsymbol{\theta}^*). \quad (3)$$

We shall use the definitions of Lecourtier and Walter (1981). The parameter θ_i is structurally globally identifiable (s.g.i.) if for almost any $\boldsymbol{\theta}^*$

$$M(\boldsymbol{\theta}) \approx M(\boldsymbol{\theta}^*) \Rightarrow \theta_i = \theta_i^*; \quad (4)$$

it is structurally locally identifiable (s.l.i.) if for almost any $\boldsymbol{\theta}^*$ there exists a neighbourhood $v(\boldsymbol{\theta}^*)$ such that if $\boldsymbol{\theta} \in v(\boldsymbol{\theta}^*)$, then

$$M(\boldsymbol{\theta}) \approx M(\boldsymbol{\theta}^*) \Rightarrow \theta_i = \theta_i^*. \quad (5)$$

Local identifiability is of course a necessary condition for global identifiability. A parameter that is not s.l.i. is structurally nonidentifiable (s.n.i.). A model $M(.)$ is s.g.i. (s.l.i.) if all its parameters are s.g.i. (s.l.i.). A model is s.n.i. if any of its parameters is s.n.i.

Remarks

(i) If a true value $\boldsymbol{\theta}^*$ for the parameters can be defined (i.e. if there is no error in the model structure), it cannot be uniquely recovered from noise-free input/output data unless the model is s.g.i.

(ii) The input is assumed here to be generated independently from the output. Feedback may result in a loss of identifiability (see, e.g. Gustavsson *et al.* (1981)).

The notion of identifiability is of importance for both linear and nonlinear models, but we would like to state some specific properties of non-LP and non-LI models concerning structural identifiability. Only non-LP models can have s.l.i. parameters which are not at the same time s.g.i. In fact, if a model is LP, the vector of *all* the model outputs $\mathbf{y}_m(\boldsymbol{\theta}, \mathbf{u}, t)$ can be written as

$$\mathbf{y}_m(\boldsymbol{\theta}, \boldsymbol{u}, \boldsymbol{t}) = A(\mathbf{u}, \mathbf{t})\boldsymbol{\theta} \tag{6}$$

and

$$M(\boldsymbol{\theta}) \approx M(\boldsymbol{\theta}^*) \Leftrightarrow A(\mathbf{u}, \mathbf{t})\boldsymbol{\theta} = A(\mathbf{u}, \mathbf{t})\boldsymbol{\theta}^*. \tag{7}$$

This set of linear equations in $\boldsymbol{\theta}$ (parametrized by $\boldsymbol{\theta}^*$) has either a unique solution $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ (if the columns of A are linearly independent) or an infinite number of solutions (located on a hyperplane of the parametric space). On the other hand, when dealing with non-LP models (the most frequent situation for phenomenological models), one obtains a set of nonlinear equations in $\boldsymbol{\theta}$, which can most often be put in the form of a set of polynomial equations in $\boldsymbol{\theta}$ (parametrized by $\boldsymbol{\theta}^*$). This set of equations may well have several distinct solutions for $\boldsymbol{\theta}$, which then implies that the model is s.l.i. without being s.g.i. When testing a non-LP model for identifiability, local studies give only partial results, and global approaches are needed.

The methods of test will differ depending on whether the model is LI or non-LI, as we shall see in Section 3.3. Moreover LI models are much more likely not to be s.g.i. than non-LI ones. Testing LI non-LP models for identifiability is thus of special importance.

3.2. *Example*

The three-compartment model to be considered, built with biological applications in mind, is taken from Venot *et al.* (1987). Other examples of biological models that turn out not to be s.g.i. can be found, for example, in Brown (1985), Carson *et al.* (1983), DiStefano and Mori (1977), and Milanese and Molino (1975). This model is used to describe the behaviour of a drug D such as glafenine (administered orally) and of its metabolite M, under the assumption that D can be transformed into M either by a first-pass effect after oral administration or after having entered the systemic circulation (Balant, 1984). It is described by the following state equation:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}x_1 &= -(\theta_1 + \theta_2)x_1 + u, \qquad x_1(0) = 0 \\
\frac{\mathrm{d}}{\mathrm{d}t}x_2 &= \theta_1 x_1 - (\theta_3 + \theta_5)x_2, \qquad x_2(0) = 0 \\
\frac{\mathrm{d}}{\mathrm{d}t}x_3 &= \theta_2 x_1 + \theta_3 x_2 - \theta_4 x_3, \qquad x_3(0) = 0.
\end{aligned} \tag{8}$$

Each of the state variables is associated with the content of a compartment; x_1 and x_2 respectively denote the quantity of drug in the gastrointestinal tract and in the systemic circulation, and x_3 denotes the quantity of metabolite in the systemic circulation.

We shall assume that there is only one input port (oral administration of the drug) but four possible output ports that we can measure, namely:

(i) the concentration of D in blood, $y_1 = \theta_6 x_2$, where θ_6 is the (unknown) inverse of the volume of Compartment 2,

(ii) the concentration of M in blood, $y_2 = \theta_7 x_3$, where θ_7 is the (unknown) inverse of the volume of Compartment 3,

(iii) the urinary excretion of D, $y_3 = \theta_5 x_2$,

(iv) the urinary excretion of M, $y_4 = \theta_4 x_3$. (9)

There are thus 15 feasible configurations of the selected output ports. Table 1 summarizes the structural identifiability properties of the model for each of them.

Remarks

(i) No feasible experiment can make the model s.g.i., but for each output-port configuration we are able to generate the set of all the models with the desired structure that will behave in exactly the same manner as a given generating model (i.e. a model with some given numerical value for each of the parameters), obtained, for example, by processing the data with a conventional package for data analysis.

(ii) Various output-port configurations are equivalent from a qualitative point of view. For example, the configurations 1, 2, 4 and 1, 2, 3, 4 yield the same structural identifiability properties. This does not mean that sampling y_3 is useless as far as reducing the uncertainty on the

TABLE 1. STRUCTURAL IDENTIFIABILITY PROPERTIES OF THE FIRST-PASS MODEL FOR THE 15 FEASIBLE CONFIGURATIONS OF THE OUTPUT PORTS; s.g.i., s.l.i. AND s.n.i., RESPECTIVELY STAND FOR STRUCTURALLY GLOBALLY IDENTIFIABLE, STRUCTURALLY LOCALLY IDENTIFIABLE AND STRUCTURALLY NON-IDENTIFIABLE

Output ports	Model	s.g.i. param.	s.l.i. param.	s.n.i. param.
1	s.n.i.	none	none	all
2	s.n.i.	none	θ_4 (3 sol.)	$\theta_1, \theta_2, \theta_3, \theta_5, \theta_6, \theta_7$
3	s.n.i.	none	none	all
4	s.n.i.	none	$\theta_1, \theta_2, \theta_3, \theta_4, \theta_5$ (6 sol.)	θ_6, θ_7
1, 2	s.n.i.	θ_4	none	$\theta_1, \theta_2, \theta_3, \theta_5, \theta_6, \theta_7$
1, 3	s.n.i.	none	none	all
1, 4	s.n.i.	θ_2, θ_4	$\theta_1, \theta_3, \theta_5, \theta_6$ (2 sol.)	θ_7
2, 3	s.n.i.	θ_4	$\theta_1, \theta_2, \theta_3, \theta_5, \theta_7$ (2 sol.)	θ_6
2, 4	s.n.i.	none	$\theta_1, \theta_2, \theta_3, \theta_4, \theta_5, \theta_7$ (6 sol.)	θ_6
3, 4	s.n.i.	θ_2, θ_4	$\theta_1, \theta_3, \theta_5$ (2 sol.)	θ_6, θ_7
1, 2, 3	s.l.i.	θ_4	$\theta_1, \theta_2, \theta_3, \theta_5, \theta_6, \theta_7$ (2 sol.)	none
1, 2, 4	s.l.i.	$\theta_2, \theta_4, \theta_7$	$\theta_1, \theta_3, \theta_5, \theta_6$ (2 sol.)	none
1, 3, 4	s.n.i.	θ_2, θ_4	$\theta_1, \theta_3, \theta_5, \theta_6$ (2 sol.)	θ_7
2, 3, 4	s.n.i.	$\theta_2, \theta_4, \theta_7$	$\theta_1, \theta_3, \theta_5$ (2 sol.)	θ_6
1, 2, 3, 4	s.l.i.	$\theta_2, \theta_4, \theta_7$	$\theta_1, \theta_3, \theta_5, \theta_6$ (2 sol.)	none

parameters is concerned, but the question of the optimal repartition of the measurements among the feasible outputs can only be addressed in the framework of quantitative experiment design.

(iii) Since this very simple model proves never to be s.g.i., one might think intuitively that for a more complex model the situation would be even worse. In fact this is not necessarily so. Consider, for example, the model obtained by adding a fourth compartment which can only exchange in both directions with Compartment 2. This model has also been used to describe some first-pass effects (Rowland *et al.* 1972; Hasegawa *et al.*, 1982). It proves to be s.g.i. under the output-port configuration 1, 2, 3, 4, although it contains two more parameters (Venot *et al.*, 1987). This apparently surprising fact can be explained as follows. Introducing a fourth compartment has increased the number of macro parameters (i.e. coefficients of exponentials and associated residuals) that can be estimated from the data in the noise-free context considered here. This increase was larger than that of the number of micro parameters (i.e. the θ_is).

3.3. *Tools for qualitative experiment design*

Most often, as was the case for the first-pass model previously described, the number of admissible input/output ports is very limited, so that one can readily test the model for identifiability under all the possible experimental configurations. What are therefore really needed for qualitative experiment design are good tools for the analysis of the identifiability properties of models. Since this analysis may have to be repeated a number of times, the procedure should be made as automatic as possible. Hence the interest in exploiting the facilities offered by computer algebra, both for obtaining the equations to be considered and for solving them (Lecourtier and Raksanyi, 1987; Raksanyi *et al.*, 1985). In what follows we shall very briefly mention some of the main methods that can be applied to LI and non-LI finite-dimensional state-space models. More detailed presentations and bibliographies can be found for example, in Godfrey (1983), Godfrey and DiStefano (1987), Jacquez (1985), and Walter (1982, 1987).

3.3.1. *LI case.* Of all the methods which can be used to test an LI model for identifiability with a global approach, we shall mention only three that we have found to be particularly useful.

The first one is to use *general results* that exist for given classes of models, like catenary or mamillary compartmental models or combinations of such structures, see for example Anderson (1982, 1983), Audoly and D'Angio (1983), and Cobelli *et al.* (1979a, b). It is thus

sometimes possible to prove that a model is (or is not) s.g.i. without any computation.

Example. Consider the three-compartment model of Section 3.2, under the output-port configuration 1. A mere inspection of the equations reveals that θ_4 and θ_7 have no influence on y_1. Another way of arriving at the same conclusion is to notice that y_1 is the output of a two-compartment model with zero initial condition. The impulse response of the model can thus be described in terms of three macro parameters as

$$y_1(t) = a(e^{-bt} - e^{-ct}) \tag{10}$$

when seven micro parameters have to be estimated.

However, it often happens that no conclusion can be reached by such methods, so that the test has then to be performed on a case-by-case basis. Moreover, when the model proves not to be s.g.i., it is often the only way to generate the set of all output-indistinguishable models and to know which parameters or parameter combinations are s.g.i.

A second possible approach, known as the *Laplace transform* (or *transfer function*) *approach*, and which can be traced back to Bellman and Åström (1970) is then to compute the transfer matrix associated with the experimental configuration to be tested

$$\begin{aligned}\mathbf{Y}(s, \boldsymbol{\theta}) &= C(\boldsymbol{\theta})[sI - A(\boldsymbol{\theta})]^{-1}B(\boldsymbol{\theta})\mathbf{U}(s) \\ &= H(s, \boldsymbol{\theta})\mathbf{U}(s)\end{aligned} \tag{11}$$

where s is the Laplace variable and I the identity matrix. One then has

$$M(\boldsymbol{\theta}) \approx M(\boldsymbol{\theta}^*) \Leftrightarrow H(s, \boldsymbol{\theta}) - H(s, \boldsymbol{\theta}^*) = 0 \tag{12}$$

and the identifiability properties of the model are directly related to the number of solutions for $\boldsymbol{\theta}$ of this equation.

Example. Consider the three-compartment model of Section 3.2, under the output-port configuration 4. The associated transfer function can be written as

$$\frac{Y_4(s, \boldsymbol{\theta})}{U(s)} = \frac{\theta_2\theta_4 s + \theta_2\theta_4(\theta_3 + \theta_5) + \theta_1\theta_3\theta_4}{(s + \theta_4)(s + \theta_1 + \theta_2)(s + \theta_3 + \theta_5)}. \tag{13}$$

Since the term in s^3 of the denominator of this transfer function is equal to 1 and there is no pole-zero cancellation, except for non-generic values of the parameters, $M(\boldsymbol{\theta}) \approx M(\boldsymbol{\theta}^*)$ is equivalent to the following set of equations:

$$\begin{aligned}&(s + \theta_4)(s + \theta_1 + \theta_2)(s + \theta_3 + \theta_5) \\ &\qquad = (s + \theta_4^*)(s + \theta_1^* + \theta_2^*)(s + \theta_3^* + \theta_5^*) \\ &\theta_2\theta_4 = \theta_2^*\theta_4^* \\ &\theta_2\theta_4(\theta_3 + \theta_5) + \theta_1\theta_3\theta_4 \\ &\qquad = \theta_2^*\theta_4^*(\theta_3^* + \theta_5^*) + \theta_1^*\theta_3^*\theta_4^*.\end{aligned} \tag{14}$$

Since θ_6 and θ_7 do not appear in these equations, they can take any value and the model is s.n.i.

To take advantage of the fact that, due to the special structure of the model, the denominator of the transfer function is available in fully factorized form, we have not expanded the polynomial corresponding to the denominator, contrary to what is usually done. As a result, instead of having to solve a single set of polynomial equations in $\boldsymbol{\theta}$, we have to solve six simpler sets of polynomial equations, obtained by considering each of the possible solutions for the first equation of (14) in turn. Each of these six sets of polynomial equations has a unique solution for θ_1 to θ_5. These parameters are thus s.l.i. Some of them take the same value in the solution of several of the six equation sets, so that they have less than six different values.

A third interesting method for testing an LI model for identifiability is the *similarity transformation approach*, which follows from the work of Berman and Schoenfeld (1956). In this method, one looks for the set of all the state-space similarity transformations that leave the model structure unchanged. Consider a model described by

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{x} = A(\boldsymbol{\theta}^*)\mathbf{x} + B(\boldsymbol{\theta}^*)\mathbf{u}, \quad \mathbf{x}(0) = \mathbf{0} \\ &\mathbf{y}_m = C(\boldsymbol{\theta}^*)\mathbf{x}.\end{aligned} \tag{15}$$

Clearly, if T is a non-singular square matrix such that $\mathbf{x} = T\mathbf{z}$, one has

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\mathbf{z} = T^{-1}A(\boldsymbol{\theta}^*)T\mathbf{z} + T^{-1}B(\boldsymbol{\theta}^*)\mathbf{u}, \quad \mathbf{z}(0) = \mathbf{0} \\ &\mathbf{y}_m = C(\boldsymbol{\theta}^*)T\mathbf{z}.\end{aligned} \tag{16}$$

Therefore, if two state-space models $(A(\boldsymbol{\theta}^*), B(\boldsymbol{\theta}^*), C(\boldsymbol{\theta}^*))$ and $(A(\boldsymbol{\theta}), B(\boldsymbol{\theta}), C(\boldsymbol{\theta}))$ are related by

$$\begin{aligned}&A(\boldsymbol{\theta}) = T^{-1}A(\boldsymbol{\theta}^*)T, \quad B(\boldsymbol{\theta}) = T^{-1}B(\boldsymbol{\theta}^*), \\ &C(\boldsymbol{\theta}) = C(\boldsymbol{\theta}^*)T\end{aligned} \tag{17}$$

then $M(\boldsymbol{\theta}) \approx M(\boldsymbol{\theta}^*)$. A less obvious result (Kalman, 1963) is that the existence of a transformation T such that (17) is satisfied is also *necessary* for the two models to have the same input/output behaviour, provided that both

models are controllable and observable. A controllable and observable LI state-space model $M(.)$ can thus be tested for identifiability by studying the number of solutions of (17) for T. For example, if there is only one solution $T = I$, then all models such that $M(\boldsymbol{\theta}) \approx M(\boldsymbol{\theta}^*)$ have the same matrices A, B and C, which, except for some pathological parametrization readily detected, implies that $\boldsymbol{\theta} = \boldsymbol{\theta}^*$ so that the model is s.g.i.

Example. Consider again the three-compartment model of Section 3.2, but now under the output-port configuration 3, 4. Its matrices A, B and C satisfy

$$A = \begin{bmatrix} -(\theta_1 + \theta_2) & 0 & 0 \\ \theta_1 & -(\theta_3 + \theta_5) & 0 \\ \theta_2 & \theta_3 & -\theta_4 \end{bmatrix},$$

$$B = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \qquad (18)$$

$$C = \begin{bmatrix} 0 & \theta_5 & 0 \\ 0 & 0 & \theta_4 \end{bmatrix}.$$

The parameters θ_6 and θ_7 do not appear in A, B or C and are therefore s.n.i. For notational simplicity, a reduced parameter vector $\boldsymbol{\theta}$ which contains only θ_1 to θ_5 is considered. This model is easily proved to be structurally controllable and observable (i.e. controllable and observable for almost any value of the parameters), so that the similarity transformation approach applies. Since, from (17), $B(\boldsymbol{\theta}^*) = TB(\boldsymbol{\theta})$ and $C(\boldsymbol{\theta}) = C(\boldsymbol{\theta}^*)T$, T can be written as

$$T(\boldsymbol{\alpha}) = \begin{bmatrix} 1 & \alpha_1 & \alpha_2 \\ 0 & \alpha_3 & 0 \\ 0 & 0 & \alpha_4 \end{bmatrix} \qquad (19)$$

where

$$\alpha_3 = \frac{\theta_5}{\theta_5^*} \quad \text{and} \quad \alpha_4 = \frac{\theta_4}{\theta_4^*}. \qquad (20)$$

Equation (17) can now be used to express $A(\boldsymbol{\theta})$ as a function of $\boldsymbol{\theta}^*$ and $\boldsymbol{\alpha}$:

$$A(\boldsymbol{\theta}) = T^{-1}(\boldsymbol{\alpha}) A(\boldsymbol{\theta}^*) T(\boldsymbol{\alpha}). \qquad (21)$$

Disregard sign constraints for the time being. A necessary and sufficient condition for $A(\boldsymbol{\theta})$ to have the right structure is then that it satisfies

$$\sum_{i=1}^{3} a_{i1} = 0, \quad a_{12} = 0, \quad a_{13} = 0, \quad a_{23} = 0. \qquad (22)$$

Together with (20), (22) admits two solutions for $\boldsymbol{\alpha}$, namely

$${}^1\boldsymbol{\alpha} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 1 \end{bmatrix}, \quad \text{and} \quad {}^2\boldsymbol{\alpha} = \begin{bmatrix} 1 - \theta_1^*/(\theta_3^* + \theta_5^* - \theta_2^*) \\ 0 \\ \theta_1^*/(\theta_3^* + \theta_5^* - \theta_2^*) \\ 1 \end{bmatrix}. \qquad (23)$$

$T({}^1\boldsymbol{\alpha}) = I \Rightarrow A({}^1\boldsymbol{\theta}) = A(\boldsymbol{\theta}^*) \Leftrightarrow {}^1\boldsymbol{\theta} = \boldsymbol{\theta}^*$, so that the first solution for $\boldsymbol{\alpha}$ corresponds to the generating model. The second solution is associated with $A({}^2\boldsymbol{\theta}) = T({}^2\boldsymbol{\alpha})^{-1} A(\boldsymbol{\theta}^*) T({}^2\boldsymbol{\alpha})$, from which the second possible value of the parameters ${}^2\boldsymbol{\theta}$ is easily obtained as a function of $\boldsymbol{\theta}^*$. $\boldsymbol{\theta}$ (restricted to its first five components) is therefore s.l.i. θ_2 and θ_4, which take the same value in $\boldsymbol{\theta}$ and $\boldsymbol{\theta}^*$, are s.g.i. and θ_1, θ_3 and θ_5 are s.l.i. Note that for ${}^2\boldsymbol{\theta}$ to be meaningful, all its components must be positive. Depending on the value of $\boldsymbol{\theta}^*$, one therefore has either two compartmental models with exactly the same input/output behaviour or only one. Neither of the two cases can be considered as atypical, for neither is associated with a subset of zero measure of the parametric space. There are structurally two solutions for $\boldsymbol{\theta}$ in $\mathbb{R}^5$, but when the sign constraints are taken into account the number of solutions for $\boldsymbol{\theta}$ is no longer a structural quantity.

Remark. The results presented in Table 1 could all be obtained either with the Laplace transform approach or with the similarity transformation approach. The only notable difference is in the complexity of the equations to be processed.

Several model structures $M_i(.)$ $(i = 1, \ldots, f)$ may be candidates to describe the same system. One would then like to design a configuration of the input and output ports such that there is at least some hope of selecting the best structure from experimental data. The notion of *structural distinguishability* can be used for that purpose. $M_1(.)$ is structurally distinguishable (s.d.) from $M_2(.)$ if for almost any (feasible) value $\boldsymbol{\theta}^*$ of the parameters of $M_2(.)$, there is no (feasible) value $\boldsymbol{\mu}$ of the parameters of $M_1(.)$ such that $M_1(\boldsymbol{\mu}) \approx M_2(\boldsymbol{\theta}^*)$. If $M_1(.)$ is s.d. from $M_2(.)$ and $M_2(.)$ is s.d. from $M_1(.)$, then $M_1(.)$ and $M_2(.)$ are said to be s.d. Although it can be shown that identifiability of $M_1(.)$ and $M_2(.)$ is neither necessary nor sufficient for the distinguishability of these two model structures (Walter *et al.*, 1984), the techniques to be used for testing a pair of model structures for distinguishability are similar to those used to test a given model structure for identifiability (Walter *et al.*, 1985). When one tests a model structure for identifiability, one hopes to be able to prove the uniqueness of the solution for $\boldsymbol{\theta}$ of

the relation $M(\boldsymbol{\theta}) \approx M(\boldsymbol{\theta}^*)$. When one tests a pair of model structures for distinguishability, one hopes to be able to prove the nonexistence of the solution for $\boldsymbol{\mu}$ of the relation $M_1(\boldsymbol{\mu}) \approx M_2(\boldsymbol{\theta}^*)$. In Walter *et al.* (1986), we investigated the identifiability and distinguishability properties of a pair of model structures as candidates for the description of the behaviour of ^{13}C in a chemical reactor producing methane from carbon monoxide and hydrogen. It turned out that these model structures were neither s.g.i. nor s.d. By using techniques of exhaustive modelling, it was nevertheless possible to generate all the models having the same input/output behaviour with either structure and to draw some conclusions on the possible values of the parameters (Walter *et al.*, 1986).

3.3.2. *Non-LI case.* There are basically three approaches for testing a non-LI model for structural identifiability. The first one (DiStefano, 1976; Walter, 1982) consists of *linearizing the model* (either mathematically around some equilibrium point or physically by introducing a tracer under appropriate experimental conditions), and then applying one of the methods available for LI models. The second approach uses a series expansion of the output, either in the time domain (Pohjanpalo, 1978) or in the time and input domain (Vajda, 1987; Walter, 1982). $M(\boldsymbol{\theta}) \approx M(\boldsymbol{\theta}^*)$ then implies that the coefficients of the series should be equal, which, as in the LI case, yields a set of equations in $\boldsymbol{\theta}$ parametrized by $\boldsymbol{\theta}^*$. A third approach has been recently developed (Vajda and Rabitz, 1989; Vajda *et al.*, 1989), which generalizes the similarity transformation approach to non-LI models.

It is well known that many LI compartmental models that are not s.g.i. become so if any of the flows involved is made nonlinear. For example, the non-LI model

$$\begin{aligned} \frac{d}{dt}x_1 &= -(\theta_1 + \theta_2)x_1 + \theta_3 x_1 x_2 + u, \quad x_1(0) = 0 \\ \frac{d}{dt}x_2 &= \theta_2 x_1 - (\theta_4 + \theta_3 x_1)x_2, \quad x_2(0) = 0 \\ y &= x_1 \end{aligned} \tag{24}$$

is s.g.i., whereas the LI model

$$\begin{aligned} \frac{d}{dt}\mathbf{x} &= \begin{bmatrix} -(\theta_1 + \theta_2) & \theta_3 \\ \theta_2 & -(\theta_3 + \theta_4) \end{bmatrix}\mathbf{x} + \begin{bmatrix} 1 \\ 0 \end{bmatrix}u, \quad \mathbf{x}(0) = \mathbf{0} \\ y &= [1 \quad 0]\mathbf{x} \end{aligned} \tag{25}$$

is s.n.i. This has led to the conjecture that all non-LI models might be s.g.i., barring some trivial exceptions such as models containing parts which are not connected to the output. This conjecture is false (Lecourtier *et al.*, 1987), but the fact remains that non-LI models built with a real application in mind are much more likely to be s.g.i. than LI models.

When input and output ports making all the parameters of interest identifiable have been selected, one has still to select the best feasible experiment to be performed to collect the data.

4. QUANTITATIVE EXPERIMENT DESIGN

Once the goal of the experiment has been fixed, one has to define the *criterion* to be used to compare experiments. An experiment that aims at estimating parameters must ensure a low uncertainty on the parameter estimates. The optimality criterion will therefore depend on the estimation method to be used. At this stage, the use of *prior knowledge* is *advisable,* and, somehow, *unavoidable.* It is advisable because it will improve the suitability of the experiment to the present situation (for instance, it is useless to stress estimating parameters which are already known precisely). It is unavoidable because any model structure is already considered as prior knowledge. When nothing is known *a priori,* there seems to be no alternative to performing experiments randomly spread throughout the admissible experimental domain. Reasonable specification of available prior knowledge, and reasonable use of this information are required for a good acceptance of the experiment designed. For instance, the lack of acceptance of experiments consisting of replications at a small number of different experimental conditions is generally due to some unreliability of the assumptions concerning the model structure or the noise distribution.

Once the optimality criterion $j(.)$ has been specified, designing an experiment reduces to a *constrained optimization problem.* The set e of admissible experiments is defined by qualitative constraints (for instance on the number and the nature of input and output ports on the process), and quantitative constraints (shape of the admissible inputs, duration of the experiment, physical limits to the excursion of some experimental conditions, minimum time between consecutive measurements . . .). Assume that N measurements are to be performed. The different variables that characterize the experimental conditions relative to the ith measurement will be summarized in a q_i-dimensional vector $\mathbf{e}^i$. All these vectors $\mathbf{e}^i$, $i = 1, \ldots, N$, are concatenated to form the experiment vector $\mathbf{e}$. Assuming that $j(.)$ is to be

minimized, an experiment $\mathbf{e}^*$ will be optimal if it satisfies

$$\mathbf{e}^* = \operatorname{Arg}\min_{\mathbf{e}\in\mathbb{E}} j(\mathbf{e}). \tag{26}$$

The availability of algorithmic procedures to solve this optimization problem fairly simply then appears essential to the practical use of experiment design.

Sections 4.1–3 recall the theoretical basis of (quantitative) experiment design. The specific problems raised by non-LP models are considered in Sections 4.4–9.

4.1. *Fisher information matrix*

Denote by $\mathbf{y}(\mathbf{e})$ the N-dimensional vector of all available measurements on the process and by $\boldsymbol{\theta}$ the p-dimensional vector of the parameters to be estimated. The measurement noise $\boldsymbol{\varepsilon}^*(\mathbf{e})$ is assumed to be additive, with zero mean, and with a distribution $f(\boldsymbol{\varepsilon}^*, \mathbf{e})$ independent of $\boldsymbol{\theta}$ (the situation where $f(.)$ also depends on $\boldsymbol{\theta}$ will be considered in Section 4.9). Denote by $\mathbf{y}_m(\boldsymbol{\theta}, \mathbf{e})$ the output vector of the model with parameters $\boldsymbol{\theta}$ associated with the observations $\mathbf{y}(\mathbf{e})$. Assume that there is no error in the model structure. A true value $\boldsymbol{\theta}^*$ then exists for $\boldsymbol{\theta}$ such that

$$\mathbf{y}(\mathbf{e}) = \mathbf{y}_m(\boldsymbol{\theta}^*, \mathbf{e}) + \boldsymbol{\varepsilon}^*(\mathbf{e}). \tag{27}$$

Under some regularity conditions, the maximum likelihood estimate of $\boldsymbol{\theta}$ based upon $\mathbf{y}$ is asymptotically normally distributed

$$\hat{\boldsymbol{\theta}}_{ml} \dashrightarrow N\left(\boldsymbol{\theta}^*, \frac{1}{N} M_F^{-1}(\boldsymbol{\theta}^*, \mathbf{e})\right) \tag{28}$$

where $M_F(\boldsymbol{\theta}, \mathbf{e})$ is the normalized Fisher information matrix, which can be written as

$$M_F(\boldsymbol{\theta}, \mathbf{e}) = \frac{1}{N} X'(\boldsymbol{\theta}, \mathbf{e})\, \Sigma^{-1}(\mathbf{e}) X(\boldsymbol{\theta}, \mathbf{e}) \tag{29}$$

with

$$X(\boldsymbol{\theta}, \mathbf{e}) = \frac{\partial \mathbf{y}_m(\boldsymbol{\theta}, \mathbf{e})}{\partial \boldsymbol{\theta}'} \tag{30}$$

and

$$\Sigma^{-1}(\mathbf{e}) = \int \frac{\partial \ln (f(\boldsymbol{\varepsilon}))}{\partial \boldsymbol{\varepsilon}} \frac{\partial \ln (f(\boldsymbol{\varepsilon}))}{\partial \boldsymbol{\varepsilon}'} f(\boldsymbol{\varepsilon})\, d\boldsymbol{\varepsilon} \tag{31}$$

where for notational convenience the dependence of $\boldsymbol{\varepsilon}$ in $\mathbf{e}$ has been omitted. From the Cramér–Rao inequality, $(1/N)M_F^{-1}(\boldsymbol{\theta}^*, \mathbf{e})$ is the lower bound for the asymptotic covariance matrix of any unbiased estimate of $\boldsymbol{\theta}$ (see, e.g. Fukunaga (1972), Goodwin and Payne (1977), Sorenson (1980), and Zacks (1981)). This property is the rationale for the use of the Fisher information matrix as a suitable characterization of the asymptotic parameter uncertainty. In most practical applications, the noise is assumed to be *white*, so that $\Sigma(\mathbf{e})$ is diagonal, and one obtains the well-known expression

$$M_F(\boldsymbol{\theta}, \mathbf{e}) = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{w_i(\mathbf{e}^i)} \frac{\partial y_{mi}(\boldsymbol{\theta}, \mathbf{e}^i)}{\partial \boldsymbol{\theta}} \frac{\partial y_{mi}(\boldsymbol{\theta}, \mathbf{e}^i)}{\partial \boldsymbol{\theta}'} \tag{32}$$

where $y_{mi}(\boldsymbol{\theta}, \mathbf{e}^i)$ is the ith component of $\mathbf{y}_m(\boldsymbol{\theta}, \mathbf{e})$, and where $w_i(\mathbf{e}^i)$ is the ith diagonal term of $\Sigma(\mathbf{e})$ (if ε_i is distributed $N(0, \sigma^2)$, then $w_i = \sigma^2$). When designing an experiment for parameter estimation, $\mathbf{e}$ should be chosen so as to minimize a measure of the uncertainty on the estimate of $\boldsymbol{\theta}$, and classical criteria of optimality are scalar functions Φ of $M_F(\boldsymbol{\theta}, \mathbf{e})$

$$j(\mathbf{e}) = \Phi(M_F(\boldsymbol{\theta}, \mathbf{e})). \tag{33}$$

Expression (32) is then the starting point for the mathematical theory of experiment design.

Remark. A sufficient condition for local identifiability is given by $\det(X'(\boldsymbol{\theta}, \mathbf{e})X(\boldsymbol{\theta}, \mathbf{e})) \neq 0$, where $X(\boldsymbol{\theta}, \mathbf{e})$ is defined by (30) (Cobelli and DiStefano, 1980; Jacquez and Greif, 1985; Walter, 1982). A *non-degenerated* or *informative* experiment (Goodwin, 1982) satisfies $\det(M_F(\boldsymbol{\theta}, \mathbf{e})) \neq 0$, where $M_F(\boldsymbol{\theta}, \mathbf{e})$ is given by (29), and ensures *local identifiability* for the model parameters. Designing an optimal experiment by minimizing a suitable criterion $\Phi(M_F(\boldsymbol{\theta}, \mathbf{e}))$ can thus be seen as maximizing a measure of identifiability, thereby transforming the initial qualitative problem into a quantitative one. Similarly, experiment design for model discrimination (Atkinson and Cox, 1974; Hill, 1978) can be seen as the quantitative counterpart to structural distinguishability (Walter *et al.*, 1984, 1985).

4.2. *Exact and approximate designs*

In the following, each measurement is assumed to depend on the same number q of experimental conditions. If $\mathbb{E}$ is the admissible domain for $\mathbf{e}^i$, then $\mathbf{e} = \mathbb{E}^N$. *Exact design* then corresponds to a constrained optimization problem in the (N . q)-dimensional space $\mathbb{E}^N$. Possible replications ($\mathbf{e}^i = \mathbf{e}^j$, $i \neq j$) lead to (32) being expressed as

$$M_F(\boldsymbol{\theta}, \mathbf{e}) = \sum_{i=1}^{n} \frac{1}{w_i(\mathbf{e}^i)} \frac{\partial y_{mi}(\boldsymbol{\theta}, \mathbf{e}^i)}{\partial \boldsymbol{\theta}} \frac{\partial y_{mi}(\boldsymbol{\theta}, \mathbf{e}^i)}{\partial \boldsymbol{\theta}'} \frac{r_i}{N} \tag{34}$$

where the integers r_i indicate the number of replications of measurements under the ith experimental condition and sum to N. The experiment $\mathbf{e}$ can then be interpreted as a discrete normalized measure on $\mathbb{E}$, with spectrum

$\mathbf{e}^i$, $i = 1, \ldots, n$, and frequencies r_i/N. In order to improve the tractability of the optimization of $j(\mathbf{e})$, Kiefer suggested that $\mathbb{E}^N$ be imbedded in the more general set of all discrete normalized measures ξ_d on $\mathbb{E}$, or if necessary into the set of all normalized measures on $\mathbb{E}$ (Kiefer and Wolfowitz, 1959; Kiefer, 1961). This generalization is associated with the concept of *approximate design*, which forms the basis for specific algorithms reviewed in Section 4.4.1. Once an optimal measure has been determined, it must be transformed into an exact experiment. From Caratheodory's theorem (see e.g. Fedorov (1972)), the optimal measure can be expressed as a discrete normalized measure with at most $(p(p+1)/2)+1$ points of support. The frequencies of this measure must then be approximated by n rational numbers r_i/N, where the r_is sum to N. An upper bound on the loss of optimality implied by such a procedure is given in Wynn (1972).

4.3. *Optimality criteria*

The theory of experiment design has been essentially concerned with approximate design for LP models. The definition of suitable optimality criteria has motivated many studies (see Pazman (1986) for a review of the desirable properties of these criteria). The essential properties of Φ, as given by (33), concern monotonicity and convexity on the set of positive semidefinite $p \times p$ matrices. Criteria of the general L_k-class (Galil and Kiefer, 1977; Kiefer, 1974; Pazman, 1986; Zarrop, 1979) satisfy these requirements. They are defined by

$$\Phi_k(M) = [p^{-1} \operatorname{trace} (HM^{-1}H^t)^k]^{1/k} \text{ if } \det(M) \neq 0$$
$$\Phi_k(M) = \infty \qquad \text{if } \det(M) = 0 \tag{35}$$

where H is a non-singular $p \times p$ matrix, and where $k \geq 0$. Approximate design is then a convex optimization problem which has received much attention. Since the pioncer work of Kiefer and Wolfowitz (1960), many results have been obtained concerning *equivalence theorems* that allow different formulations for the conditions of optimality of a design measure ξ^*, see, e.g. Kiefer (1974) and Silvey (1980) for results based on directional derivatives, Pukelsheim and Titterington (1983), Sibson (1972), Silvey (1972, 1980), and Silvey and Titterington (1973) for results using Lagrangian theory, and White (1973) for an extension of Kiefer's results to non-LP models. Algorithmic procedures for determining ξ^* have then been developed on the basis of these equivalence theorems (see Section 4.4 for references).

D-optimality. The D-optimality criterion is defined by

$$\Phi(M) = [\det(M)]^{-1}. \tag{36}$$

It can be deduced from (35) by setting $H = I_p$, and $k = 0$. A D-optimal experiment is thus obtained by maximizing the determinant of the Fisher information matrix. The large acceptance of this policy (Box and Lucas, 1959; Fedorov, 1972; Landaw, 1980; St John and Draper, 1975) can be partly explained by its appealing geometrical interpretation: the asymptotic confidence regions for the maximum likelihood estimate of $\boldsymbol{\theta}$ are ellipsoids (see Behnken (1964)), and a D-optimal experiment minimizes the volume of these ellipsoids (Bard, 1974; Box and Lucas, 1959; Galil and Kiefer, 1977; Silvey and Titterington, 1973). Moreover, D-optimal experiments possess the attractive property of being invariant with respect to any non-degenerated transformation applied on the model parameters such as a rescaling (Fedorov, 1972; Landaw, 1980). It is also well known that they often correspond to replications of a small number of different experimental conditions (Atkinson and Hunter, 1968; Box, 1968, 1970, 1972; Landaw, 1980; Pronzato, 1986; Vila, 1986, 1988).

L-optimality. Linear criteria are obtained from (35) by setting $k = 1$. The particular choices $H = I_p$ and $H = \operatorname{diag}(\theta_i^{-1}, i = 1, \ldots, p)$, repectively, correspond to *A- and C-optimality*. A-optimal experiments minimize the sum of variances of the estimate of $\boldsymbol{\theta}$ (Chernoff, 1953). C-optimality is generally preferred to A-optimality, as a C-optimal experiment is related to relative precision of the estimates and is independent from the choice of scale of the parameters (Landaw, 1980).

E-optimality. The E-optimality criterion is obtained from (35) by $H = I_p$ and $k = \infty$. E-optimal design corresponds to maximizing the minimum eigenvalue of $M_F(\boldsymbol{\theta}, \mathbf{e})$. E-optimal experiments thus minimize the maximum diameter of the asymptotic confidence ellipsoids for $\boldsymbol{\theta}$ (Galil and Kiefer, 1977).

C-optimality. Assume now that a scalar function $g(\boldsymbol{\theta})$ has to be estimated instead of the vector $\boldsymbol{\theta}$ itself. The asymptotic covariance of the maximum likelihood estimate of g is given by

$$\frac{\partial g}{\partial \boldsymbol{\theta}^t} M_F^{-1}(\boldsymbol{\theta}, \mathbf{e}) \frac{\partial g}{\partial \boldsymbol{\theta}}.$$

Minimizing this variance corresponds to a particular linear optimal design, and the c-optimality criterion is defined by (Silvey, 1980)

$$\Phi(M) = \operatorname{trace} (\mathbf{c} \cdot \mathbf{c}^t \cdot M^{-1}) \tag{37}$$

with

$$\mathbf{c} = \frac{\partial g}{\partial \boldsymbol{\theta}}. \tag{38}$$

D_s-*optimality*. When one is interested in a subset of s parameters, partial optimality criteria can be defined, such as D_s-optimality (Atwood, 1969; Gaffke, 1987; Karlin and Studden, 1966; Pazman, 1986; Silvey, 1980; Silvey and Titterington, 1973).

Turing's measure of conditioning. It is sometimes desirable to obtain a confidence region as spherical as possible for $\boldsymbol{\theta}$. Turing's measure of conditioning is then a suitable optimality criterion (Sutton and MacGregor, 1977), defined by

$$\Phi(M) = \frac{1}{p}\sqrt{(\text{trace}\,(M^{-1})\,\text{trace}\,(M))}. \tag{39}$$

Its minimum value is 1, obtained for spherical confidence regions. Note that this policy is rather similar in its purpose to E-optimal design (Hosten and Emig, 1975).

Efficiency criteria. Any optimality criterion $j(.)$ can be associated with an efficiency function, defined as a measure of the relative performance of any given experiment $\mathbf{e}$ compared to that of the optimal experiment $\mathbf{e}^*$. Galil and Kiefer (1977) define absolute inefficiency ratios for the L_k-class of criteria (35) by

$$\rho_k(\mathbf{e}) = \frac{\Phi_k(M_F(\boldsymbol{\theta}, \mathbf{e}))}{\Phi_k(M_F(\boldsymbol{\theta}, \mathbf{e}_k^*))} \tag{40}$$

where $\mathbf{e}_k^*$ is the optimal experiment for the criterion defined by Φ_k. Landaw (1980) defines the D-efficiency of $\mathbf{e}$ by

$$\zeta_d(\mathbf{e}) = \left[\frac{\det\,(M_F(\boldsymbol{\theta}, \mathbf{e}))}{\det\,(M_F(\boldsymbol{\theta}, \mathbf{e}_d^*))}\right]^{1/p} \tag{41}$$

where $\mathbf{e}_d^*$ is D-optimal.

The possible construction of compound criteria from basic ones (Kiefer, 1974) gives rise to a *boundless multiplicity of possible policies.* However, it seems reasonable to choose the functional Φ on the conjugate basis of *optimization tractability* and *physically appealing significance.* When the model is LP, the design criterion is completely defined by the functional Φ. When the model is *non-LP,* the Fisher information matrix (32), and consequently any criterion given by (33), depends on $\boldsymbol{\theta}$ which is of course unknown before experimentation. Several attempts have been made to overcome this difficulty. Sections 4.4–8 present various optimality criteria, and various optimal design policies, obtained by taking into account different formulations of prior knowledge, or prior uncertainty, on $\boldsymbol{\theta}$.

The most commonly used approach, known in the literature as *local optimal design* (Chernoff, 1953; Fedorov, 1972), consists in designing an optimal experiment for some given nominal value $\boldsymbol{\theta}^\circ$ of the model parameters.

4.4. *Local optimal design*

Optimizing a criterion $j(\boldsymbol{\theta}, \mathbf{e})$ for some given value $\boldsymbol{\theta}^\circ$ of $\boldsymbol{\theta}$ leads to an optimal experiment $\mathbf{e}^*$ that depends on $\boldsymbol{\theta}^\circ$. If in some special situations $\mathbf{e}^*(\boldsymbol{\theta}^\circ)$ can be obtained analytically (see, e.g. Endrenyi (1980), Landaw (1980), and Reilly *et al.* (1977)), the optimization problem generally has to be handled iteratively. The method to be used depends on the structure of the criterion function $j(.)$, on the admissible experimental domain $\mathbb{E}$ (see, e.g. Mori and DiStefano (1979) for an example of optimal design on a finite discrete set $\mathbb{E}$), and on the description chosen for the experiment (approximate or exact design).

4.4.1. *Approximate design.* The design measure approach needs not specify the number of measurements to be performed, but (32) must be valid. Two basic algorithms for constructing optimal measures ξ^* have been proposed, by Wynn (1972) for D-optimality, and Fedorov (1972) for L- and D-optimality. Several modifications have been suggested in order to accelerate their convergence (Atwood, 1973; St John and Draper, 1975), together with alternative procedures (Silvey and Titterington, 1973). Iterative procedures for general optimality criteria are presented in Wu (1978), and Wu and Wynn (1978). A review of these iterative procedures can be found in Pazman (1986), and a FORTRAN implementation for L- and D-optimality is given in Landaw (1980).

4.4.2. *Exact design.* Exact design requires the number of measurements N to be specified beforehand. Wynn (1972) and Fedorov (1972) have suggested exchange algorithms for D-optimal design (see also Atkinson and Donev (1988), and Yonchev (1988)). The convergence towards a D-optimal exact experiment is, however, not proved (see Silvey (1980)). A general non-linear programming approach in the $(N \cdot q)$-dimensional space $\mathbb{E}^N$ makes it possible to consider situations where (32) is not valid, and/or where the cost of performing N experiments is taken into account in the optimality criterion. First-order (e.g. steepest-descent) or second-order (e.g. Newton–Raphson) algorithms can then be used (see, e.g. Gill and Murray (1974), and Sadler (1975)). The possible multimodality of the criterion makes it advisable to use a global search method (see Dixon and Szëgo (1975, 1978)).

4.4.3. *Applications.* Phenomenological models are most often non-LP, and local experiment design is then a natural approach. Once a configuration of the input and output ports has been chosen so as to render the model identifiable (see, e.g. Brown (1982), Carson *et al.* (1983), DiStefano and Mori (1977), and Nathanson *et al.* (1984)), usual designs concern input and sampling times optimization.

Sampling times. The method to be used depends on the application. In electrical engineering for instance, long duration experiments involve the choice of *optimal sampling rates* (see, e.g. Goodwin *et al.* (1974)). In the case of uniform sampling, this design can be carried out in the frequency domain, thus leading to simple design procedures (see, e.g. Ng and Goodwin (1976), and Zarrop (1979)). On the other hand, if the experiment involves a small number of measurements each sampling time has to be optimized. Approximate design for compartmental models is used in Cobelli and Thomaseth (1986), and Landaw (1980, 1982), while an exact design approach is used in Cobelli *et al.* (1983), D'Argenio and Khakmahd (1983), DiStefano (1980), Mori and DiStefano (1979), and Nathanson *et al.* (1984).

Optimal inputs. Design of optimal inputs has received a considerable amount of attention (see the survey papers by Krolikowski and Eykhoff (1985), Mehra (1974), and the books by Zarrop (1979) and Kalaba and Spingarn (1982)). This problem can be expressed as an optimal control problem (see, e.g. Kalaba and Spingarn (1974, 1981)). The influence of feedback is investigated in Gevers and Ljung (1985). Sometimes, the input $u(t)$ can be parametrized (for instance as a combination of impulses and step functions (Mori and DiStefano, 1979; Pronzato, 1986)). Non-parametrized inputs have also been considered (Cobelli and Thomaseth, 1985, 1988a, b; Thomaseth and Cobelli, 1988).

Optimal input and sampling schedule. Simultaneous choice of input signal and sampling strategy is known to be essential for returning the maximum information from the experiment (see, e.g. Cobelli *et al.* (1985), Cobelli and Thomaseth (1988b), DiStefano (1979, 1980), Goodwin *et al.* (1974), and Weber and Rault (1979)).

Whatever the optimization procedure to be used, local optimal design does not take into account any prior uncertainty on the nominal value $\boldsymbol{\theta}^\circ$ chosen for $\boldsymbol{\theta}$. The experiment designed can thus prove to be far from optimal if $\boldsymbol{\theta}^\circ$ differs too much from the "true value" $\boldsymbol{\theta}^*$. This can hardly be considered as a robust approach, and has raised some doubts among experimenters about the practical use of optimal experiment design. Sections 4.5–8 review several approaches that have been suggested to overcome this difficulty.

4.5. *Sequential design*

The basic idea consists in improving the choice of $\boldsymbol{\theta}^\circ$ by alternating identification of the parameters and experiment design (see, e.g. Behnken (1964), and Box and Lucas (1959)). Each new experiment is designed using as nominal value $\boldsymbol{\theta}^\circ$ the estimate obtained from previous observations. Different policies can be used, which pertain to *exact design.*

4.5.1. *Full sequential and batch sequential designs.* Denote by $\hat{\boldsymbol{\theta}}_N$ the estimate of $\boldsymbol{\theta}$ based upon N measurements, and by $\mathbf{e}_1^N$ the concatenation of the experiment vectors $\mathbf{e}^i$, $i = 1, \ldots, N$, that characterize those N measurements. A *batch sequential design* (BSD) with size r concerns the choice of the next r vectors $\mathbf{e}^i$, $i = N+1, \ldots, N+r$, concatenated in $\mathbf{e}_{N+1}^{N+r}$. A *full sequential design* (FSD) is a BSD with size 1. Denote by $M_F(\hat{\boldsymbol{\theta}}_N, \mathbf{e}_1^{N+r})$ the Fisher information matrix corresponding to $N+r$ measurements evaluated at $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_N$. The usual practice is then to choose $\mathbf{e}_{N+1}^{N+r*}$ such that

$$\mathbf{e}_{N+1}^{N+r*} = \operatorname{Arg} \min_{\mathbf{e}_{N+1}^{N+r} \in \mathbb{E}^r} \Phi(M_F(\hat{\boldsymbol{\theta}}_N, \mathbf{e}_1^{N+r})). \quad (42)$$

Procedures for obtaining $\mathbf{e}_{N+1}^{N+r*}$, when Φ corresponds to L- or D-optimality, are described in Fedorov (1972). Note that the dependence of M_F in $\hat{\boldsymbol{\theta}}_N$ (due to the non-LP character of the model) implies that $\Phi(M_F(\hat{\boldsymbol{\theta}}_N, \mathbf{e}_1^{N+r}))$ cannot be determined recursively from $\Phi(M_F(\hat{\boldsymbol{\theta}}_{N-r}, \mathbf{e}_1^N))$. Problems concerning convergence properties are considered in Silvey (1980): $\mathbf{e}_{N+1}^{N+r*}$ depends on the previous observations, so that successive observations are not independent, and M_F is not the usual Fisher information matrix (see Ford *et al.* (1985), and Wu (1985)). However, a fast convergence of the sequential experiment is generally observed (see, e.g. Box and Hunter (1965), Hosten and Emig (1975) for FSD, and Draper and Hunter (1966), and Heineken *et al.* (1967) for BSD).

When a batch size $r \geq p$ is to be used, with $p = \dim(\boldsymbol{\theta})$, the following alternative approach, which we term incomplete batch design (IBD) can be considered.

4.5.2. *Incomplete batch design.* A batch size $r \geq p$ makes it possible to design non-degenerated experiments by the simpler procedure

$$\mathbf{e}_{N+1}^{N+r*} = \operatorname{Arg} \min_{\mathbf{e}_{N+1}^{N+r} \in \mathbb{E}^r} \Phi(M_F(\hat{\boldsymbol{\theta}}_N, \mathbf{e}_{N+1}^{N+r})). \quad (43)$$

The information resulting from previous measurements is simply summarized in $\hat{\boldsymbol{\theta}}_N$, and the design procedure does not take into account the fact that data have been collected with the previous experimental conditions $\mathbf{e}_1^N$. IBD is thus a repetition of classical local optimal designs (see, e.g. Behnken (1964), Burrows *et al.* (1982), and Johnson and Berthouex (1975)). The convergence of $\mathbf{e}_{N+1}^{N+r*}$ towards the optimal experiment for estimating $\boldsymbol{\theta}^*$ is simply related to the convergence fo $\hat{\boldsymbol{\theta}}_N$ towards $\boldsymbol{\theta}^*$, where $\boldsymbol{\theta}^*$ is the true parameter value given by (27) (see Box (1970) for a comparison between convergence properties of IBD and BSD on a catalytic chemical reaction model).

4.5.3. *Updating policy for the parameters.* When the designed experiment $\mathbf{e}_{N+1}^{N+r*}$ is nondegenerated (i.e. when $\det M_F(\boldsymbol{\theta}, \mathbf{e}_{N+1}^{N+r*}) \neq 0$)), the r associated measurements can be used to obtain a maximum-likelihood estimate $\hat{\boldsymbol{\theta}}_r$ of $\boldsymbol{\theta}$. The updated estimate of $\boldsymbol{\theta}$ can then be taken as follows (Fedorov, 1972):

$$\hat{\boldsymbol{\theta}}_{N+r} = [NM_F(\hat{\boldsymbol{\theta}}_N, \mathbf{e}_1^N) + rM_F(\hat{\boldsymbol{\theta}}_N, \mathbf{e}_{N+1}^{N+r})]^{-1} \times [NM_F(\hat{\boldsymbol{\theta}}_N, \mathbf{e}_1^N)\hat{\boldsymbol{\theta}}_N + rM_F(\hat{\boldsymbol{\theta}}_N, \mathbf{e}_{N+1}^{N+r})\hat{\boldsymbol{\theta}}_r]. \tag{44}$$

When BSD is used, with size $r < p$, the estimate $\hat{\boldsymbol{\theta}}_r$ cannot be obtained from the observations y_i, $i = N+1, \ldots, N+r$. If the noise distribution is white, so that (32) is valid, the following updating policy can be used (Fedorov, 1972):

$$\hat{\boldsymbol{\theta}}_{N+r} = [NM_F(\hat{\boldsymbol{\theta}}_N, \mathbf{e}_1^N) + rM_F(\hat{\boldsymbol{\theta}}_N, \mathbf{e}_{N+1}^{N+r})]^{-1} \times \left[NM_F(\hat{\boldsymbol{\theta}}_N, \mathbf{e}_1^N)\hat{\boldsymbol{\theta}}_N + \sum_{i=N+1}^{N+r} \frac{1}{w_i(\mathbf{e}^i)} \frac{\partial y_{mi}(\boldsymbol{\theta}, \mathbf{e}^i)}{\partial \boldsymbol{\theta}} \Big|_{\hat{\boldsymbol{\theta}}_N} \cdot y_i \right]. \tag{45}$$

If (32) is not valid, $\hat{\boldsymbol{\theta}}_{N+r}$ can be computed in a non-recursive way, according to the estimation method based upon the $N+r$ measurements.

4.5.4. *Population studies.* Sequential design on the same subject is sometimes impossible in practice, for example in biological studies, and an alternative approach is to make successive experiments on different subjects belonging to the same population (see, e.g. Cobelli *et al.* (1983), D'Argenio and Khakmahd (1983), and DiStefano (1981, 1982)). The usual approach consists of an IBD, with fixed size $r \geq p$, defined by

$$\mathbf{e}_{N+1}^{N+r*} = \operatorname{Arg} \min_{\mathbf{e}_{N+1}^{N+r} \in \mathbb{E}^r} \Phi(M_F(\langle \hat{\boldsymbol{\theta}}_N \rangle, \mathbf{e}_{N+1}^{N+r})) \tag{46}$$

where $\langle \hat{\boldsymbol{\theta}}_N \rangle$ denotes the estimate, after N measurements, of the mean parameter value $\langle \hat{\boldsymbol{\theta}} \rangle$ in the population. D'Argenio (1981) suggests choosing $\langle \hat{\boldsymbol{\theta}}_N \rangle$ as the mean of the available individual parameter estimates. One can refer to Prevost (1977), and Steimer *et al.* (1984) for alternative methods.

Conducting such multiple experiments is not always possible, and one is often interested in designing a single experiment. Moreover, when using sequential design, each experiment must be chosen as best given the available information on the parameters. For these reasons increasing attention has been devoted in recent years to *non-sequential* approaches aimed at determining a single optimal experiment while taking into account prior knowledge, or prior uncertainty, on the model parameters.

4.6. *Bayesian design*

Experiment design for Bayes estimators in LP models with normal measurement noise has motivated several studies (see, e.g. Chaloner (1984), Giovagnoli and Verdinelli (1983, 1985), and Verdinelli (1983), and the survey paper by Bandemer *et al.* 1987). Consider the classical linear Bayes estimator $\hat{\boldsymbol{\theta}}_B$ corresponding to N measurements, defined by

$$\hat{\boldsymbol{\theta}}_B = [NM_F(\mathbf{e}) + \Omega^{-1}]^{-1}[X^t(\mathbf{e})\Sigma^{-1}(\mathbf{e})\mathbf{y} + \Omega^{-1}\langle \boldsymbol{\theta} \rangle] \tag{47}$$

where $\langle \boldsymbol{\theta} \rangle$ and Ω are the prior mean and covariance matrix of $\boldsymbol{\theta}$, respectively, and where $X^t(\mathbf{e})$ and $\Sigma^{-1}(\mathbf{e})$ are given by (30) and (31) (M_F and X here do not depend on $\boldsymbol{\theta}$ because of the assumed linearity of the model in the parameters).

This estimator minimizes the Bayes risk $\underset{\boldsymbol{\theta}}{E}\{(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^t H^t H(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\}$ with respect to $\hat{\boldsymbol{\theta}}$. The minimal value of the risk is given by

$$\underset{\boldsymbol{\theta}}{E}\{(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_B)^t H^t H(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_B)\} = \frac{1}{N} \operatorname{trace}\left[H(M_F(\mathbf{e}) + \frac{1}{N}\Omega^{-1})^{-1} H^t \right] \tag{48}$$

where the choice of the matrix H depends on the relative interest in the parameters (see L- and c-optimality in Section 4.3). When the noise distribution and the prior distribution of $\boldsymbol{\theta}$ are both normal, $\hat{\boldsymbol{\theta}}_B$ is the maximum *a posteriori* estimator, and the minimum mean square estimator of $\boldsymbol{\theta}$ (Sorenson, 1980). Bayes L-optimal design is defined by the criterion function Φ_{BL}

$$\Phi_{BL}(M) = \operatorname{trace}\left[H\left(M_F(\mathbf{e}) + \frac{1}{N}\Omega^{-1} \right)^{-1} H^t \right] \tag{49}$$

and is strongly connected with the classical theory of approximate local optimal design (see, e.g. Bandemer *et al.* (1987), and Chaloner

(1984)). An algorithmic procedure for obtaining Bayes optimal experiments is described in Bandemer *et al.* (1987). Note that these experiments can be degenerated ($\det(M_F(\mathbf{e})) = 0$) if the prior distribution for $\boldsymbol{\theta}$ contains too much information (Landaw, 1980). The existence of one-point Bayes c-optimal experiments is proved in Chaloner (1984).

Bayes design for non-LP models has been somewhat ignored in the literature. The Cramér–Rao inequality for random parameters (see, e.g. Fukunaga (1972), and Sorenson (1980)) suggests optimality criteria based on scalar functions Φ of $N_F(\mathbf{e})$ (Pronzato, 1986), where $N_F(\mathbf{e})$ is given by

$$N_F(\mathbf{e}) = \underset{\boldsymbol{\theta}}{E}\{M_F(\boldsymbol{\theta}, \mathbf{e})\} + \underset{\boldsymbol{\theta}}{E}\left\{\frac{\partial \ln \pi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ln \pi(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}^t}\right\} \tag{50}$$

with $\pi(.)$ the prior distribution of $\boldsymbol{\theta}$. This approach is used by Mehra (1974) in the context of optimal input design, where the matrix $N_F(\mathbf{e})$ is approximated via a second-order Taylor-series development. Evaluating $\underset{\boldsymbol{\theta}}{E}\{M_F(\boldsymbol{\theta}, \mathbf{e})\}$ is generally a time consuming task, and thus an important obstacle to the iterative construction of optimal experiments.

BSD based on the criterion Φ_{BL} defined by (49) is considered in Draper and Hunter (1967a, b). The sequential procedure could be improved by updating the prior covariance matrix Ω after each estimation.

Remark. A similar procedure could be used for population studies (Section 4.5.4). Several methods suggested for estimating population characteristics such as $\langle\boldsymbol{\theta}\rangle$ and Ω (Mallet, 1983, 1986; Mentré 1984; Prevost, 1977; Sheiner and Beal, 1980; Steimer *et al.*, 1984) could be of use for setting up an updating policy for these characteristics.

4.7. *Optimal design in the average sense*

When the prior knowledge on $\boldsymbol{\theta}$ is too unreliable to use a Bayesian estimator of $\boldsymbol{\theta}$, one should allow the data to speak for themselves as much as possible. Bayesian design as defined in Section 4.6 is then inappropriate, and a probabilistic description of prior uncertainty on $\boldsymbol{\theta}$ leads to optimal design in the average sense. The criterion j_{AS} to be optimized is then the mathematical expectation of some local optimality criterion over all the possible values of $\boldsymbol{\theta}$

$$j_{AS}(\mathbf{e}) = \underset{\boldsymbol{\theta}}{E}\{\Phi(M_F(\boldsymbol{\theta}, \mathbf{e}))\}. \tag{51}$$

The set of available Φ corresponds to the functions used in local optimal design. The most commonly used are based on L- and D-optimality (see, e.g. Chaloner (1986, 1988), Chaloner and Larntz (1986, 1988), D'Argenio and Van Guilder (1988), Ermatov (1983), Fedorov (1980, 1981), Pronzato and Walter (1985), and Walter and Pronzato (1985, 1987a). It must be noted, however, that *one* local optimality criterion can lead to *several* average optimality criteria (see, e.g. Fedorov (1980, 1981), Fedorov and Atkinson (1988), and Pronzato (1986)). For instance, a D-optimal experiment can be obtained by maximizing $\det(M_F(\boldsymbol{\theta}, \mathbf{e}))$, or minimizing $[\det(M_F(\boldsymbol{\theta}, \mathbf{e}))]^{-1}$, whereas maximizing the ED-optimality criterion (Pronzato and Walter, 1985) defined by

$$j_{ED}(\mathbf{e}) = \underset{\boldsymbol{\theta}}{E}\{\det(M_F(\boldsymbol{\theta}, \mathbf{e}))\} \tag{52}$$

or minimizing the EID-optimality criterion (Walter and Pronzato, 1985, 1987a) defined by

$$j_{EID}(\mathbf{e}) = \underset{\boldsymbol{\theta}}{E}\{[\det(M_F(\boldsymbol{\theta}, \mathbf{e}))]^{-1}\} \tag{53}$$

leads to quite different experiments. The question of which criterion to use thus arises naturally. In Walter and Pronzato (1985, 1987b) EID-optimality is shown to have definite advantages over ED-optimality ($j_{EID}(\mathbf{e})$ is the average value of a scalar measure of the asymptotic uncertainty on $\boldsymbol{\theta}$). Note that average efficiency criteria could also be considered (D'Argenio and Van Guilder, 1988; Landaw, 1984; Pronzato, 1986), but the required computation of the local optimal experiment associated with each value of $\boldsymbol{\theta}$ (see Section 4.4) would considerably increase the amount of computation needed by any optimization procedure.

Except in very special situations (see, e.g. Zacks (1977)), the optimal experiment in the average sense cannot be obtained analytically, so that algorithmic procedures are needed. The results of Chaloner and Larntz (1986) and Dubov (1977) parallel those concerning L- and D-optimality for local approximate design (see Section 4.4.1), and a combination of Wynn's and Fedorov's algorithms is described in Chaloner and Larntz (1986). Their method converges slowly, and they suggest a faster nonlinear programming algorithm for discrete design with a fixed number of points of support (Chaloner and Larntz, 1988). For both procedures, mathematical expectations are evaluated by *numerical integration*. In order to avoid such evaluations, Ng and Goodwin (1976) suggest a Taylor-series expansion of j_{ED}, as given by (52), and D'Argenio and Van Guilder (1988) consider a

discrete approximation of the prior density $\pi(.)$. However, exact design based on any criteria such as (51) can be carried out *without any evaluation of mathematical expectation* (Pronzato and Walter, 1985). The method proposed uses a stochastic approximation algorithm (see, e.g. Dvoretsky (1956), Poliak and Tsypkin (1973), and Saridis (1974)), with a scaling policy that permits the avoidance of any internal tuning when treating a new problem (Pronzato, 1986; Pronzato and Walter, 1985; Walter and Pronzato, 1985). *EID*-optimal experiments that minimize (53) are then obtained almost as simply as classical local D-optimal experiments. *EID*-optimal design possesses some properties similar to those of D-optimal design concerning replications and independence from the probability distribution of the linear parameters (Pronzato, 1986; Walter and Pronzato, 1987b).

The sequential use of optimal design in the average sense is presented in Chaloner (1986), and Zacks (1977). Chaloner shows that batch sequential or non-sequential average optimal designs can be as efficient as (non-Bayesian) FSD. Finally, the Remark in Section 4.6 concerning population studies is also of interest for average optimality.

4.8. *Optimal design in the minimax sense*

The approach presented in Section 4.7 designs experiments that are good on average, but can perform poorly for some particular values of $\boldsymbol{\theta}$ associated with very low values of the prior distribution $\pi(.)$. One might prefer to optimize the worst possible performances of the experiment designed. This minimax approach relies on the knowledge of some prior admissible domain Θ for $\boldsymbol{\theta}$, without requiring any hypothesis on the prior distribution of $\boldsymbol{\theta}$ in Θ. Minimax criteria can be deduced from local optimality criteria by

$$j_{MM}(\mathbf{e}) = \max_{\boldsymbol{\theta} \in \Theta} \Phi(M_F(\boldsymbol{\theta}, \mathbf{e})) \tag{54}$$

where $j_{MM}(\mathbf{e})$ must be minimized with respect to $\mathbf{e}$. Classical functions Φ correspond to L- and D-optimality (see Ermatov (1983), Fedorov (1980, 1981), Fedorov and Atkinson (1988), and Pronzato (1986)). For instance, an experiment is said to be *MMD*-optimal (Pronzato and Walter, 1988a, b; Walter and Pronzato, 1987b) when it maximizes the criterion defined by

$$j_{MMD}(\mathbf{e}) = \det (M_F(\boldsymbol{\theta}_d(\mathbf{e}), \mathbf{e})) \tag{55}$$

where

$$\boldsymbol{\theta}_d(\mathbf{e}) = \operatorname{Arg} \min_{\boldsymbol{\theta} \in \Theta} \det (M_F(\boldsymbol{\theta}, \mathbf{e})). \tag{56}$$

Landaw (1984) suggests the use of a maximin D-efficiency criterion given by

$$j_{MMD}\zeta(\mathbf{e}) = \min_{\boldsymbol{\theta} \in \Theta} \left[\frac{\det [M_F(\boldsymbol{\theta}, \mathbf{e})]}{\det [M_F(\boldsymbol{\theta}, \mathbf{e}_d^*(\boldsymbol{\theta}))]} \right]^{1/p} \tag{57}$$

where $\mathbf{e}_d^*(\boldsymbol{\theta})$ is the D-optimal experiment associated with the value $\boldsymbol{\theta}$ for the parameters (see also D'Argenio and Van Guilder (1988)). However, each evaluation of $j_{MMD}\zeta(\mathbf{e})$ requires the computation of many D-optimal experiments, which thus makes its optimization quite a tremendous task. *MMD*-optimal experiments have properties similar to those of *EID*-optimal experiments, concerning replications and independence in the linear parameters (Pronzato and Walter, 1988a,b; Walter and Pronzato, 1987b).

MMD-optimal design can sometimes be reduced to local optimal design for some given value of $\boldsymbol{\theta}$. This frequently occurs for the important class of exponential regression models (Melas, 1978, 1981). When such a simplification is not possible, an algorithmic procedure for obtaining minimax optimal experiments (exact design) described in Pronzato (1986), and Pronzato and Walter (1988a,b) can be used. It is based on the conjugate use of a relaxation procedure (Shimizu and Aiyoshi, 1980), and a global optimizer (Pronzato *et al.*, 1984).

Minimax optimal design could also be used in a sequential approach. Each estimation should provide an updated estimate of the prior feasible domain for the parameters. Methods recently developed for membership set estimation could be used for that purpose (see, e.g. Belforte and Milanese (1981), and Walter and Piet-Lahanier (1986, 1988)).

4.9. *Unknown and heterogeneous error variance*

Design of experiments in the case of unknown and heterogeneous error variance has been somewhat neglected in the literature. Classical approaches using the Fisher information matrix given by (29) require that the noise distribution be either white and stationary, or known. However, practical situations may involve unknown and non-stationary variances. For instance, the measurement errors ε_i^* are frequently assumed to be independently normally distributed $N(0, \sigma_i^2(\rho, \nu, \mathbf{e}^i))$, with a non-homogeneous variance $\sigma_i^2(\rho, \nu, \mathbf{e}^i)$ given by

$$\sigma_i^2(\rho, \nu, \mathbf{e}^i) = \rho[y_{m_i}(\boldsymbol{\theta}, \mathbf{e}^i)]^\nu. \tag{58}$$

The parameter ν is assumed to belong to $[0, 2]$, $\nu = 0$ corresponds to a constant homogeneous variance, and $\nu = 2$ to a constant relative precision. The value assumed for ν is well known as having a great influence on the experiment designed (see, e.g. Box and Draper (1972), and

Landaw (1980)). It must first be noted that the normal distribution of the ε_is does not imply that the weighted least squares and the maximum likelihood estimators coincide. For weighted least squares estimation, the matrix to be used for designing the experiment no longer corresponds to the Fisher information matrix, but is still given by (29), where

$$\Sigma = \text{diag}\,\{\sigma_i^2(\rho, \nu, \mathbf{e}^i), i = 1, \ldots, N\}. \quad (59)$$

The optimal experiment is thus independent from ρ. The dependence on the unknown value of ν can be taken into account by an average optimal or a minimax optimal approach. For instance, when the prior uncertainty on ν is characterized by a distribution on $[0, 2]$ (e.g. uniform), a possible criterion is (Bezeau and Endrenyi, 1987; Endrenyi 1986; Schulz and Endrenyi, 1983) $\underset{\nu}{E}\{\zeta_d(\boldsymbol{\theta}, \mathbf{e}, \nu)\}$, where the D-efficiency criterion $\zeta_d(.)$ given by (41) here depends on ν. Another possible policy is to maximize the smallest efficiency (Endrenyi, 1986; Schulz and Endrenyi, 1983), which corresponds to maximizing the criterion $\min_{\nu}\{\zeta_d(\boldsymbol{\theta}, \mathbf{e}, \nu)\}$. The dependence on the parameter vector $\boldsymbol{\theta}$ could be taken simultaneously into account by a similar approach (see Sections 4.7 and 4.8). Note that replacing D-efficiency by other criteria based on D-optimality would lead to more tractable optimization procedures. An alternative policy, suggested in Pronzato (1986), considers the maximum likelihood estimator of $\boldsymbol{\lambda} = (\boldsymbol{\theta}', \rho, \nu)'$. The Fisher information matrix for this estimator can be constructed. If one is only interested in estimating $\boldsymbol{\theta}$, a partial optimality criterion, such as D_s-optimality (see, e.g. Gaffke (1987), Pazman (1986), Silvey (1980), Silvey and Titterington (1973)), could be used for designing the experiment. The dependence on the unknown value of $\boldsymbol{\lambda}$ could be taken into account by any of the approaches previously presented (sequential design, average or minimax optimal design . . .).

5. CONCLUSIONS

Even very simple phenomenological models may turn out not to be globally identifiable. There are then several values of the parameter vector that are associated to exactly the same input/output behaviour. Conventional softwares for data analysis will then converge (if at all!) to one of them, which may be completely erroneous, even if the model structure is correct and the data are noise-free. Similarly, if two models that are candidate for the description of the same data are not distinguishable, one may use the wrong model and nevertheless obtain a very satisfactory fit, even in a noise-free situation. It is therefore important in practice to analyse the structural properties of the model(s) considered under all the feasible input/output configurations. This qualitative experiment design can be done before data are collected. Computer algebra provides a great help, both for obtaining the equations expressing that the models have the same input/output behaviour and for solving them. This sometimes proves to be very complex and it may be necessary to try several methods, and even to combine them, to reach a conclusion. For large models, techniques making it possible to split the study into that of submodels can also prove very helpful (Audoly and D'Angio, 1983; Eisenfeld, 1982).

Taking into account prior information on the noise statistics and possible values of the parameters cannot be avoided for quantitative experiment design. A multiplicity of methods, which are associated with various formulations of the prior information, has been presented for that purpose. Local experiment design requires the specification of a nominal value $\boldsymbol{\theta}^\circ$, and should be restricted to situations where the efficiency function weakly depends on $\boldsymbol{\theta}$. When Bayesian estimation can be performed on the basis of reliable prior information $\pi(\boldsymbol{\theta})$, Bayesian experiment design should be preferred. However, in most situations, this prior information is too unreliable to be taken into account in the estimation procedure, and $\pi(\boldsymbol{\theta})$ can only be considered as a characterization of the prior uncertainty on $\boldsymbol{\theta}^\circ$. Average experiment design is particularly suited to this situation. The resulting experiment, although optimal in average, can prove rather poor for some particular values of $\boldsymbol{\theta}$. If this is unacceptable, minimax optimal design could be considered, provided the uncertainty on $\boldsymbol{\theta}^\circ$ is expressed as a prior admissible domain. When the experimentation phase can be repeated, each one of these approaches can form the basis of a sequential design. The existing algorithms should permit the application of any of these policies at a reasonable computational cost.

REFERENCES

Anderson, D. H. (1982). Structural properties of compartmental models. *Math. Biosci.*, **58**, 61–81.

Anderson, D. H. (1983). *Compartmental Modeling and Tracer Kinetics.*, Springer, Berlin.

Atkinson, A. C. and W. G. Hunter (1968). The design of experiments for parameter estimation. *Technometrics*, **10**(2), 271–289.

Atkinson, A. C. and D. R. Cox (1974). Planning experiments for discriminating between models (with discussion). *J. R. Statist. Soc.*, **B36**, 321–348.

Atkinson, A. C. and A. N. Donev (1988). Algorithms, exact design and blocking in response surface and mixture designs. In Y. Dodge, V. V. Fedorov and H. P. Wynn

(Eds), *Optimal Design and Analysis of Experiments*, pp. 61–69. North-Holland, Amsterdam.

Atwood, C. L. (1969). Optimal and efficient designs of experiments. *Ann. Math. Stat.*, **40**(5), 1570–1602.

Atwood, C. L. (1973). Sequences converging to *D*-optimal designs of experiments. *Ann. Stat.*, **1**(2), 342–352.

Audoly, S. and L. D'Angio (1983). On the identifiability of linear compartmental systems: a revisited transfer function approach based on topological properties. *Math. Biosci.*, **66**, 201–228.

Balant, L. (1984). Applicability of different types of models in health and disease. *Drug Metab. Rev.*, **15**, 75–102.

Bandemer, H., N. Wolfgang and J. Pilz (1987). Once more: optimal experimental design for regression models (with discussion). *Statistics*, **18**(2), 171–217.

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

Behnken, D. W. (1964). Estimation of copolymer reactivity ratios: an example of nonlinear estimation. *J. Polym. Sci.*, **A2**, 645–668.

Belforte, G. and M. Milanese (1981). Uncertainty interval evaluation in presence of unknown but bounded errors: nonlinear families of models. *Proc. IASTED Int. Symp. on Modelling Identification and Control*, Davos, pp. 75–79.

Bellman, R. and K. J. Åström (1970). On structural identifiability. *Math. Biosci.*, **7**, 329–339.

Berman, M. and R. Schoenfeld (1956). Invariants in experimental data on linear kinetics and the formulation of models. *J. Appl. Phys.*, **27**, 1361–1370.

Bezeau, M. and L. Endrenyi (1987). Design of experiments for the precise estimation of dose-response parameters: the Hill equation. *J. Theor Biol.*,

Box, G. E. P. and H. L. Lucas (1959). Design of experiments in nonlinear situations. *Biometrika*, **46**, 77–90.

Box, G. E. P. and W. G. Hunter (1965). The experimental study of physical mechanisms. *Technometrics*, **7**(1), 23–42.

Box, M. J. (1968). The occurrence of replications in optimal design of experiments to estimate parameters in nonlinear models. *J. R. Statist. Soc.*, **30**, 290–302.

Box, M. J. (1970). Some experiences with a nonlinear experimental design criterion. *Technometrics*, **12**(3), 569–589.

Box, M. J. (1972). Discussion of a paper by H. P. Wynn. *J. R. Statist. Soc.*, **B34**, 171–172.

Box, M. J. and N. R. Draper (1972). Estimation and design criteria for multiresponse nonlinear models with non-homogeneous variance. *Appl. Statist.*, **21**, 13–24.

Brooks, R. J. (1974). On the choice of an experiment for prediction in linear regression. *Biometrika*, **61**, 303–311.

Brooks, R. J. (1977). Optimal regression design for control in linear regression. *Biometrika*, **64**, 319–325.

Brown, R. F. (1982). Identifiability: its role in design of pharmacokinetic experiments. *IEEE Trans. Biomed. Engng*, **29**, 49–54.

Brown, R. F. (1985). *Biomedical Systems Analysis via Compartmental Concept*. Abacus Press, Tunbridge Wells.

Burrows, C. R., M. N. Sahinkaya and O. S. Turkay (1982). The use of *D*-optimality to improve the quality of squeeze-film parameter estimates. *Preprints 6th IFAC Symp. on Identification and System Parameter Estimation*, Washington, pp. 645–650.

Carson, E. R., C. Cobelli and L. Finkelstein (1983). *The Mathematical Modeling of Metabolic and Endocrine Systems*. Wiley, New York.

Chaloner, K. (1984). Optimal Bayesian experimental design for linear models. *Ann. Stat.*, **12**(1), 283–300.

Chaloner, K. (1986). Optimal Bayesian design for nonlinear estimation. Technical Report 468, University of Minnesota.

Chaloner, K. (1988). An approach to experimental design for generalized linear models. In Fedorov, V. V. and H. Läuter (Eds), *Model-Oriented Data Analysis, Proc. Eisenach GDR* 1987, Lecture Notes in Economics and Mathematical Systems, Vol. 297, pp. 3–12. Springer, Berlin.

Chaloner, K. and K. Larntz (1986). Optimal Bayesian design applied to logistic regression experiments. Technical Report 483, University of Minnesota.

Chaloner, K. and K. Larntz (1988). Software for logistic regression experiment design. In Y. Dodge, V. V. Fedorov and H. P. Wynn (Eds), *Optimal Design and Analysis of Experiments*, pp. 207–211. North-Holland, Amsterdam.

Chernoff, H. (1953). Locally optimal designs for estimating parameters. *Ann. Math. Stat.*, **24**, 586–602.

Cobelli, C., A. Lepschy and G. Romanin Jacur (1979a). Identifiability results on some constrained compartmental systems. *Math. Biosci.*, **47**, 173–196.

Cobelli, C., A. Lepschy and G. Romanin Jacur (1979b). Identifiability of compartmental systems and related structural properties. *Math. Biosci.*, **48**, 1–18.

Cobelli, C. and J. J. DiStefano, III (1980). Parameter and structural identifiability concepts and ambiguities: a critical review and analysis. *Am. J. Physiol.*, **239**, 7–24.

Cobelli, C., J. J. DiStefano, III and A. Ruggeri (1983). Minimal sampling schedules for identification of dynamic models of metabolic systems of clinical interest: case studies for two liver function tests. *Math. Biosci.*, **63**, 173–186.

Cobelli, C., A. Ruggeri and K. Thomaseth (1985). Optimal input and sampling schedule design for physiological system identification. Some theoretical and in vivo results on a compartmental model of glucose kinetics. In J. Eisenfeld and C. De Lisi (Eds), *Mathematics and Computers in Biomedical Applications*, pp. 161–172. Elsevier North-Holland, Amsterdam.

Cobelli, C. and K. Thomaseth (1985). Optimal input design for identification of compartmental models. Theory and applications to a model of glucose kinetics. *Math. Biosci.*, **77**, 267–286.

Cobelli, C. and K. Thomaseth (1986). Assessing output measurement optimality in physiological system identification by a design measure approach. *Proc. 25th Conf. on Decision and Control*, Athens, pp. 282–285.

Cobelli, C. and K. Thomaseth (1988a). On optimality of the impulse input for linear system identification. *Math. Biosci.*, **89**, 127–133.

Cobelli, C. and K. Thomaseth (1988b). Optimal equidose inputs and role of measurement error for estimating the parameters of a compartmental model of glucose kinetics from continuous- and discrete-time optimal samples. *Math. Biosci.*, **89**, 135–147.

D'Argenio, D. Z (1981). Optimal sampling times for pharmacokinetic experiments. *J. Pharmac. Biopharmac.*, **9**(6), 739–756.

D'Argenio, D. Z. and K. Khahmahd (1983). Adaptive control of theophylline therapy: importance of blood sampling times. *J. Pharmac. Biopharmac.*, **11**(5), 547–559.

D'Argenio, D. Z. and M. Van Guilder (1988). Design of experiments for parameter estimation involving uncertain systems, with application to pharmacokinetics. *Preprints of the 12th IMACS World Congress on Scientific Computation*, Paris, Vol. 2, pp. 511–513.

DiStefano, III, J. J. (1976). Tracer experiment design for unique identification of nonlinear physiological systems. *Am. J. Physiol.*, **230**, 476–485.

DiStefano, III, J. J. and F. Mori (1977). Parameter identifiability and experiment design: thyroid hormone metabolism parameters. *Am. J. Physiol.*, **233**, 134–144.

DiStefano, III, J. J. (1979). Matching the model and the experiment to the goals: data limitations, complexity and optimal experiment design for dynamic systems with biochemical signals. *J. Cybern. Inf. Sci.*, **2**, 6–20.

DiStefano, III, J. J. (1980). Design and optimization of tracer experiments in physiology and medicine. *Fed. Proc.*, **39**(1), 84–90.

DiStefano, III, J. J. (1981). Optimized blood sampling protocols and sequential design of kinetic experiments. *Am. J. Physiol.*, **9**, 259–265.

DiStefano, III, J. J. (1982). Algorithms, software and sequential optimal sampling schedule designs for pharmacokinetic and physiologic experiments. *Math. Comput. Simulat.*, **24**, 531–534.

Dixon, L. C. and G. P. Szëgo (1975). *Towards Global Optimization*. Vol. 1. North-Holland, Amsterdam.
Dixon, L. C. and G. P. Szëgo (1978). *Towards Global Optimization*. Vol. 2. North-Holland, Amsterdam.
Draper, N. R. and W. G. Hunter (1966). Design of experiments for parameter estimation in multiresponse situations. *Biometrika*, **53**, 525–533.
Draper, N. R. and W. G. Hunter (1967a). The use of prior distributions in the design of experiments for parameter estimation in nonlinear situations. *Biometrika*, **54**, 147–153.
Draper, N. R. and W. G. Hunter (1967b). The use of prior distributions in the design of experiments for parameter estimation in nonlinear situations: multiresponse case. *Biometrika*, **54**, 662–665.
Dubov, E. L. (1977). *D*-Optimal designs for nonlinear models under the Bayesian approach. In *Regression Experiments*. pp. 103–111 Moscow University Press, (in Russian).
Dvoretsky, A. (1956). On stochastic approximation. *Proc. 3rd Berkeley Symp. Math. Stat. and Probability*, Vol. 1, pp. 39–55.
Eisenfeld, J. (1982). New techniques for structural identifiability for large linear and nonlinear compartmental systems. *Math. Comput. Simulat.*, **24**, 494–501.
Endrenyi, L. (1980). Design of experiments for estimating enzyme and pharmacokinetic parameters. In *Kinetic Data Analysis*, pp. 137–167. Plenum Press, New York.
Endrenyi, L. (1986). Diagnosis, robust design and estimation in nonlinear regression. *Trans. 42nd Annual Quality Control Conf.*, Rochester, New York, pp. 21–41.
Ermatov, S. M. (Ed.) (1983). *Mathematical Theory of Experiment Design*. Nauka, Moscow (in Russian).
Fedorov, V. V. (1972). *Theory of Optimal Experiments*. Academic Press, New York.
Fedorov, V. V. (1980). Convex design theory. *Math. Operationsforsch. Statist., Ser. Statist.*, **11**(3), 403–413.
Fedorov, V. V. (1981). Active regression experiments. In V. V. Penenko (Ed.), *Mathematical Methods for Experiment Design*, pp. 19–73. Nauka, Novosibirsk, (in Russian).
Fedorov, V. V. and V. Khabarov (1986). Duality of optimal designs for model discrimination and parameter estimation. *Biometrika*, **73**, 183–190.
Fedorov, V. V. (1987). Various discrepancy measures in model testing (two competing regression models). In Y. Dodge (Ed.), *Statistical Data Analysis Based on the L_1-Norm and Related Methods*. pp. 357–366. North-Holland, Amsterdam.
Fedorov, V. V. and A. C. Atkinson (1988). The optimum design of experiments in the presence of uncontrolled variability and prior information. In Y. Dodge, V. V. Fedorov and H. P. Wynn (Eds), *Optimal Design and Analysis of Experiments*. pp. 327–344. North-Holland, Amsterdam.
Ford, I., D. M. Titterington and C. F. Wu (1985). Inference and sequential design. *Biometrika*, **72**(3), 545–551.
Fukunaga, K. (1972). *Introduction to Statistical Pattern Recognition*. Academic Press, New York.
Gaffke, N. (1987). Further characterizations of design optimality and admissibility for partial parameter estimation in linear regression *Ann. Stat.*, **15**(3), 942–957.
Galil, Z. and J. Kiefer (1977). Comparison of simplex designs for quadratic mixture models. *Technometrics*, **19**(4), 445–453.
Gevers, M. and L. Ljung (1985). Benefits of feedback in experiment design. *Preprints 7th IFAC/IFORS Symp. on Identification and System Parameter Estimation*, York, Vol. 1, pp. 909–914.
Gevers, M. and L. Ljung (1986). Optimal experiment design with respect to the intended model application. *Automatica*, **22**(5), 543–554.
Gill, P. E. and W. Murray (1974). *Numerical Methods for Constrained Optimization*. Academic Press, New York.
Giovagnoli, A. and I. Verdinelli (1983). Bayes *D*-optimal and *E*-optimal block designs. *Biometrika*, **70**(3), 695–706.
Giovagnoli, A. and I. Verdinelli (1985). Optimal block designs under a hierarchical linear model. In J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith (Eds.), *Bayesian Statistics*, Vol. 2, pp. 655–661. Elsevier North-Holland, Amsterdam.
Godfrey, K. R. (1983). *Compartmental Models and their Application*, Academic Press, London.
Godfrey, K. R. and J. J. DiStefano, III (1987). Identifiability of model parameters. In E. Walter (Ed.), *Identifiability of Parametric Models*, pp. 1–20. Pergamon Press, Oxford.
Goodwin, G. C., M. B. Zarrop and R. L. Payne (1974). Coupled design of test signals, sampling intervals, and filters for system identification. *IEEE Trans. Aut. Control*, **19**(6), 748–752.
Goodwin, G. C. and R. L. Payne (1977). *Dynamic System Identification: Experiment Design and Data Analysis*. Academic Press, New York.
Goodwin, G. C. (1982). An overview of the system identification problem, part 4: experiment design. *Proc. 6th IFAC Symp. on Identification and System Parameter Estimation*, pp. 19–25.
Gustavsson, I., L. Ljung and T. Söderström (1981). Choice and effect of different feedback configurations. In P. Eykhoff (Ed.), *Trends and Progress in System Identification*, pp. 367–388. Pergamon Press, Oxford.
Hasegawa, J., E. T. Lin, R. L. Williams, F. Sörgel and L. Z. Benet (1982). Pharmacokinetics of triamterene and its metabolite in man. *J. Pharmac. Biopharmac.*, **18**, 505–523.
Heineken, F. G., H. M. Tsuchiya and R. Aris (1967). On the accuracy of determining rate constants in enzymatic reactions. *Math. Biosci.*, **1**, 115–141.
Hill, P. D. H. (1978). A review of experimental design procedures for regression model discrimination. *Technometrics*, **20**(1), 15–21.
Hosten, L. H. and G. Emig (1975). Sequential experimental design for precise parameter estimation in ordinary differential equations. *Chem. Engng. Sci.*, **30**, 1357–1364.
Jacquez, J. A. (1985). *Compartmental Analysis in Biology and Medicine*. 2nd Edn. The University of Michigan Press, Ann Arbor.
Jacquez, J. A. and P. Greif (1985). Numerical parameter identifiability and estimability: integrating identifiability, estimability and optimal sampling design. *Math. Biosci.*, **77**, 201–227.
Johnson, D. B. and P. M. Berthouex (1975). Efficient biokinetic experimental designs. *Biotechnol. Bioengng.*, **17**, 557–570.
Kalaba, R. E. and K. Spingarn (1974). Optimal inputs for nonlinear process parameter estimation. *IEEE Trans. Aerospace Electronic Systems*, **10**(3), 339–345.
Kalaba, R. E. and K. Spingarn (1981). Optimal inputs for blood glucose regulation parameter estimation, part 3. *J. Optimiz. Theory Appl.*, **33**(2), 267–285.
Kalaba, R. E. and K. Spingarn (1982). *Control, Identification, and Input Optimization*. Plenum Press, New York.
Kalman, R. E. (1963). Mathematical description of linear dynamical systems. *SIAM J. Control, Ser. A*, **1**, 152–192.
Karlin, S. and W. J. Studden (1966). Optimal experimental designs. *Ann. Math. Stat.*, **37**, 783–815.
Kiefer, J. and J. Wolfowitz (1959). Optimum designs in regression problems. *Ann. Math. Stat.*, **30**, 271–294.
Kiefer, J. and J. Wolfowitz (1960). The equivalence of two extremum problems. *Can. J. Math.*, **12**, 363–366.
Kiefer, J. (1961). Optimum designs in regression problems II. *Ann. Math. Stat.*, **32**, 298–325.
Kiefer, J. (1974). General equivalence theory for optimum designs (approximate theory). *Ann. Stat.*, **2**(5), 849–879.
Krolikowski, A. and P. Eykhoff (1985). Input signal design for system identification: a comparative analysis. *Preprints 7th IFAC/IFORS Symp. on Identification and System Parameter Estimation*, York, Vol. 1, pp. 915–920.
Landaw, E. M. (1980). Optimal experimental design for biologic compartmental systems with application to pharmacokinetics. Ph.D. dissertation, UCLA.
Landaw, E. M. (1982). Optimal multicompartmental sampling designs for parameter estimation: practical aspects of the identification problem. *Math. Comp. Simulat.*, **24**, 525–530.

Landaw, E. M. (1984). Robust sampling designs for compartmental models under large prior eigenvalue uncertainties. In J. Eisenfeld and C. De Lisi (Eds), *Mathematics and Computers in Biomedical Applications*, pp. 181–187 Elsevier North-Holland, Amsterdam.

Lecourtier, Y. and E. Walter (1981). Comments on "On parameter and structural identifiability: nonunique observability/reconstructibility for identifiable systems, other ambiguities and new definitions. *IEEE Trans. Aut. Control*, **26**, 800–801.

Lecourtier, Y., F. Lamnabhi-Lagarrigue and E. Walter (1987). Volterra and generating power series approaches to identifiability testing. In E. Walter (Ed.), *Identifiability of Parametric Models*, pp. 50–66. Pergamon Press, Oxford.

Lecourtier, E. and A. Raksanyi (1987). The testing of structural properties through symbolic computation. In E. Walter (Ed.), *Identifiability of Parametric Models*, pp. 75–84. Pergamon Press, Oxford.

Mallet, A. (1983). Méthodes d'estimation de lois à partir d'observations indirectes d'un échantillon: application aux caractéristiques de population des modèles biologiques. Thèse de Doctorat d'Etat, Université Pierre & Marie Curie, Paris VI.

Mallet, A. (1986). A maximum likelihood estimation method for random coefficient regression models. *Biometrika*, **73**(3), 645–656.

Mehra, R. K. (1974). Optimal input signals for parameter estimation in dynamic systems, survey and new results. *IEEE Trans. Aut. Control*, **19**(6), 753–768.

Melas, V. B. (1978). Optimal designs for exponential regression. *Math. Operationsforsch. Statist., Ser. Statist.*, **9**(1), 45–59.

Melas, V. B. (1981). Design of optimal experiments for exponential regression. In V. V. Penenko (Ed.), *Mathematical Methods for Experiment Design*. pp. 174–198. Nauka, Novosibirsk (in Russian).

Mentré, F. (1984). Apprentissage de la loi de probabilité des paramètres d'un modèle par approximation stochastique. Thèse de 3ème cycle, Université Paris VII.

Milanese, M. and G. P. Molino (1975). Structural identifiability of compartmental models and pathophysiological information from the kinetics of drugs. *Math. Biosci.*, **26**, 175–190.

Mori, F. and J. J. DiStefano, III (1979). Optimal nonuniform sampling interval and test input design for identification of physiological systems from very limited data. *IEEE Trans. Aut. Control*, **24**, 893–900.

Nathanson, M. H., G. M. Saidel and G. D. McLaren (1984). Analysis of iron kinetics: identifiability, experiment design, and deterministic interpretations of a stochastic model. *Math. Biosci.*, **68**, 1–21.

Ng, T. S. and G. C. Goodwin (1976). On optimal choice of sampling strategies for linear system identification. *Int. J. Control*, **23**(4), 459–475.

O'Hagan, A. (1978). Curve fitting and optimal design for prediction (with discussion). *J. R. Statist. Soc.*, **B40**(1), 1–42.

Pazman, A. (1986). *Foundations of Optimum Experimental Design*, VEDA, Bratislava.

Pohjanpalo, H. (1978). System identifiability based on the power series expansion of the solution. *Math. Biosci.*, **41**, 21–33.

Poliak, B. T. and Ya. Z. Tsypkin (1973). Pseudogradient adaptation and training algorithms. *Aut. Remote Control*, **34**(3), 377–397.

Prevost, G. (1977). Estimation of a normal probability density function from samples measured with non-negligible and nonconstant dispersion. Internal Report, ADERSA-GERBIOS, Velizy-Villacoublay, France.

Pronzato, L., E. Walter, A. Venot and J. F. Lebruchec (1984). A general purpose global optimizer: implementation and applications. *Math. Comput. Simulat.*, **26**, 412–422.

Pronzato, L. and E. Walter (1985). Robust experiment design via stochastic approximation. *Math. Biosci.*, **75**, 103–120.

Pronzato, L. (1986). Synthèse d'expériences robustes pour modèles à paramètres incertains. Thèse de Doctorat, Université Paris XI.

Pronzato, L. and E. Walter (1988a). Robust experiment design for nonlinear regression models. In V. V. Fedorov and H. Läuter (Eds), *Model-Oriented Data Analysis, Proc. Eisenach GDR 1987*, Lecture Notes in Economics and Mathematical Systems, Vol. 297, pp. 77–86. Springer, Berlin.

Pronzato, L. and E. Walter (1988b). Robust experiment design via maximin optimization. *Math. Biosci.*, **89**, 161–176.

Pukelsheim, F. and D. M. Titterington (1983). General differential and Lagrangian theory for optimal experimental design. *Ann. Stat.*, **11**, 1060–1068.

Pukelsheim, F. (1988). Analysis of variability by analysis of variance. In Y. Dodge, V. V. Fedorov and H. P. Wynn (Eds), *Optimal Design and Analysis of Experiments*, pp. 289–292. North-Holland, Amsterdam.

Raksanyi, A., Y. Lecourtier, E. Walter and A. Venot (1985). Identifiability and distinguishability testing via computer algebra. *Math. Biosci.*, **77**, 245–266.

Reilly, P. M., R. Bajramovic, G. E. Blau, D. R. Branson and M. W. Sauerhoff (1977). Guidelines for the optimal design of experiments to estimate parameters in first order kinetic models. *Can. J. Chem. Engng*, **55**, 614–622.

Rowland, M., S. Riegelman, P. A. Harris and S. D. Sholkoff (1972). Absorption kinetics of aspirin in man following oral administration of an aqueous solution. *J. Pharm. Sci.*, **61**, 379–385.

Sadler, D. R. (1975). *Numerical Methods for Nonlinear Regression*. University of Queensland Press.

Saridis, G. N. (1974). Stochastic approximation methods for identification and control, a survey. *IEEE Trans. Aut. Control*, **19**(6), 798–809.

Schulz, M. and L. Endrenyi (1983). Design of experiments for estimating parameters with unknown heterogeneity of the error variance. *Proc. American Statistical Association, Computing Section*, pp. 177–181.

Sheiner, L. B. and S. L. Beal (1980). Evaluation of methods for estimating population pharmacokinetic parameters. I Michaelis Menten model. Routine clinical pharmacokinetic data. *J. Pharmac. Biopharmac.*, **8**(6), 553–571.

Shimizu, K. and E. Aiyoshi (1980). Necessary conditions for min-max problems and algorithm by a relaxation procedure. *IEEE Trans. Aut. Control*, **25**(1), 62–66.

Sibson, R. (1972). Discussion on a paper by H. P. Wynn. *J. R. Statist. Soc.*, **B34**, 181–183.

Silvey, S. D. (1972). Discussion on a paper by H. P. Wynn. *J. R. Statist. Soc.*, **B34**, 174–175.

Silvey, S. D. and D. M. Titterington (1973). A geometric approach to optimal design theory. *Biometrika*, **60**(1), 21–32.

Silvey, S. D. (1980). *Optimal Design*. Chapman & Hall, London.

Sorenson, H. W. (1980). *Parameter Estimation, Principles and Problems*. Marcel Dekker, New York.

Steimer, J. L., A. Mallet, J. L. Golmard and J. F. Boisvieux (1984). Alternative approaches to estimation of population pharmacokinetic parameters, comparison with NONMEM. *Drug. Metab. Rev.*, **15**, 265–292.

St John, R. C. and N. R. Draper (1975). *D*-optimality for regression designs: a review. *Technometrics*, **17**(1), 15–23.

Sutton, T. L. and J. F. MacGregor (1977). The analysis and design of binary vapour–liquid equilibrium experiments, part II: the design of experiments. *Can. J. Chem. Engng*, **55**, 609–613.

Thomaseth, K. and C. Cobelli (1988). Optimal input design for physiological systems identification: theory and practice. *Preprints of the 12th IMACS World Congress on Scientific Computation*, Paris, Vol. 2, pp. 514–516.

Vajda, S. (1987). Identifiability of polynomial systems: structural and numerical aspects. In E. Walter (Ed.), *Identifiability of Parametric Models*, pp. 42–49. Pergamon Press, Oxford.

Vajda, S. and H. Rabitz (1989). State isomorphism approach to global identifiability of nonlinear systems. *IEEE Trans. Aut. Control*, **34**, 220–223.

Vajda, S., H. Rabitz, E. Walter and Y. Lecourtier (1989). Qualitative and quantitative identifiability analysis of nonlinear chemical kinetic models. *Chem. Engng. Commun.*, **83**, 191–219.

Venot, A., E. Walter, Y. Lecourtier, A. Raksanyi and L. Chauvelot-Moachon (1987). Structural identifiability of "first-pass" models. *J. Pharmac. Biopharmac.*, **15**, 179–189.

Verdinelli, I. (1983). Computing Bayes *D*- and *A*-optimal block designs for a two-way model. *The Statistician*, **32**, 161–167.

Verdinelli, I. and H. P. Wynn (1988). Target attainment and experiment design, a bayesian approach. In Y. Dodge, V. V. Fedorov and H. P. Wynn (Eds), *Optimal Design and Analysis of Experiments*, pp. 319–324. North-Holland, Amsterdam.

Vila, J. P. (1986). Local optimality of replications from a minimal *D*-optimal design in nonlinear regression: a sufficient and quasi necessary condition. Internal Report No. 86–13, INRA, F78350 Jouy-en-Josas, France.

Vila, J. P. (1988). New algorithmic and software tools for *D*-optimal design computation in nonlinear regression. *COMPSTAT 1988*, pp. 409–414. Physica, Heidelberg.

Walter, E. (1982). *Identifiability of State Space Models*. Springer, Berlin.

Walter, E. (Ed.) (1987). *Identifiability of Parametric Models*. Pergamon Press, Oxford.

Walter, E., Y. Lecourtier and J. Happel (1984). On the structural output distinguishability of parametric models, and its relation with structural identifiability. *IEEE Trans. Aut. Control*, **29**, 56–57.

Walter, E., Y. Lecourtier, A. Raksanyi and J. Happel (1985). On the distinguishability of parametric models with different structures. In J. Eisenfeld and C. De Lisi (Eds), *Mathematics and Computers in Biomedical Applications*. pp. 145–160. Elsevier North-Holland, Amsterdam.

Walter, E. and L. Pronzato (1985). How to design experiments that are robust to parameter uncertainty. *Preprints 7th IFAC/IFORS Symp. Identification and System Parameter Estimation*, York, Vol. 1, pp. 921–926.

Walter, E., Y. Lecourtier, J. Happel and J.-Y. Kao (1986). Identifiability and distinguishability of fundamental parameters in catalytic methanation. *A.I.Ch.E.Jl*, **32**, 1360–1366.

Walter, E. and H. Piet-Lahanier (1986). Robust nonlinear parameter estimation in the bounded noise case. *Proc. 25th IEEE Conf. on Decision and Control*, Athens, pp. 1037–1042.

Walter, E., H. Piet-Lahanier and J. Happel (1986). Estimation of non-uniquely identifiable parameters via exhaustive modeling and membership set theory. *Math. Comput. Simulat.*, **28**, 479–490.

Walter, E. and L. Pronzato (1987a). Optimal experiment design for nonlinear models subject to large prior uncertainties. *Modeling Methodology Forum of the Am. J. Physiol.*, **253**, 530–534.

Walter, E. and L. Pronzato (1987b). Robust experiment design: between qualitative and quantitative identifiabilities. In E. Walter (Ed.), *Identifiability of Parametric Models* pp. 104–113. Pergamon Press, Oxford.

Walter, E. and H. Piet-Lahanier (1988). Estimation of the parameter uncertainty resulting from bounded-error data. *Math. Biosci.*, **92**, 55–74.

Weber, R. and A. Rault (1979). Sensitizing inputs and optimal sampling instants. *Preprints 5th IFAC Symp. on Identification and System Parameter Estimation*, Vol. 1, pp. 603–609.

White, L. V. (1973). An extension of the general equivalence theorem to nonlinear models. *Biometrika*, **60**(2), 345–348.

Wu, C. F. (1978). Some algorithmic aspects of the theory of optimal designs. *Ann. Stat.*, **6**(6), 1286–1301.

Wu, C. F. and H. P. Wynn (1978). The convergence of general step-length algorithms for regular optimum design criteria. *Ann. Stat.*, **6**(6), 1273–1285.

Wu, C. F. (1985). Asymptotic inference from sequential design in a nonlinear situation. *Biometrika*, **72**(3), 553–558.

Wynn, H. P. (1972). Results in the theory and construction of *D*-optimum experimental designs. *J. R. Statist. Soc.*, **B34**, 133–147.

Yonchev, H. (1988). New computer procedures for constructing *D*-optimal designs. In Y. Dodge, V. V. Fedorov and H. P. Wynn (Eds), *Optimal Design and Analysis of Experiments*, pp. 71–80. North-Holland, Amsterdam.

Zacks, S. (1977). Problems and approaches in design of experiments for estimation and testing in nonlinear models. In P. R. Krishnaiah (Ed.), *Multivariate Analysis IV*. pp. 209–223. North-Holland, Amsterdam.

Zacks, S. (1981). *Parametric Statistical Inference, Basic Theory and Modern Approaches*. Pergamon Press, London.

Zarrop, M. B. (1979). *Optimal Experiment Design for Dynamic System Identification*. Springer, Heidelberg.