Tatra Mt. Math. Publ. **26** (2003), 29–39

TATRA MOUNTAINS
Mathematical Publications

TRANSFORMING BOTH SIDES AND OPTIMUM EXPERIMENTAL DESIGN FOR A NONLINEAR MODEL ARISING FROM SECOND-ORDER CHEMICAL KINETICS

Anthony C. Atkinson

ABSTRACT. The paper develops methods for the design of experiments for mechanistic models when the response has to be transformed to achieve a symmetric error distribution with constant variance. Because of the nature of the relationship between response and the mechanistic model, it is necessary to transform both sides of the model. An example is given of optimum design for experiments in which multivariate responses are measured when the model is defined by a set of differential equations which cannot be solved analytically.

1. Introduction

This paper summarizes work on the effect of the assumed mean-variance relationship on the design of experiments for mechanistic models. Atkinson [2003] describes the background and gives simple examples. Here the emphasis is on a more complicated model arising in chemical kinetics in which the differential equations defining the model have to be solved numerically.

The experimental design problem considered is that of selecting a set of times t_i at which readings y_i should be taken in order to obtain good estimates of the parameter vector ψ in the nonlinear mechanistic model

$$y = \eta(t, \psi). \tag{1}$$

This becomes a statistical problem when error is introduced.

Let the expected value of Y be $E(Y) = \mu$ and the variance be $\operatorname{var} Y$. Then the variance of a function $g(Y)$ can be shown, by Taylor series expansion, to be approximately $\{g'(\mu)\}^2 \operatorname{var} Y$, where the dash denotes differentiation with

---

2000 Mathematics Subject Classification: Primary 62K05 Secondary 62J02.
Keywords: Box-Cox transformation, chemical kinetics, D-optimum design, direct method, mean variance relationship, parameter sensitivities.

respect to y. If the variance of the random variable Y increases as a power of the expected value so that

$$\operatorname{var} Y \propto \{E(Y)\}^{2(1-\lambda)}, \tag{2}$$

the Taylor expansion shows that power transformation of the response provides a random variable with variance approximately independent of the mean. For transformation of just the response y in a regression model, the normalized power transformation of Box and Cox [1964] is

$$z(\lambda) = \begin{cases} (y^\lambda - 1)/(\lambda \dot{y}^{\lambda-1}), & \lambda \neq 0, \\ \dot{y} \log y, & \lambda = 0, \end{cases} \tag{3}$$

where the geometric mean of the observations is written as $\dot{y} = \exp(\Sigma \log y_i / n)$. When $\lambda = 1$, there is no transformation. The model to be fitted is (1) with response $z(\lambda)$ and additive errors of constant variance.

However, if, as here, $\eta(t, \psi)$ is a mechanistic model based on chemical kinetics, the relationship between the response and the concentrations of the other reactants needs to be preserved after transformation. This is achieved by transformation of both sides of the model, as described in Chapter 4 of Caroll and Ruppert [1988]. For fixed $\lambda \neq 0$, estimation of the parameters ψ after transformation does not depend on whether the response is $z(\lambda)$ or the non-normalized y^λ. Simplification of the model and the introduction of observational error on this transformed scale leads to the statistical model

$$y^\lambda = \{\eta(t, \psi)\}^\lambda + \varepsilon. \tag{4}$$

We assume that the value of λ is known and compare designs for different λ. Designs for estimation of λ when the response of a regression model is to be transformed are described by Atkinson and Cook [1996].

A simple example of the effect of assumptions about the mean-variance relationship on experimental design comes from the model for first-order decay in which the concentration of chemical A at time t is given by the nonlinear function

$$[A] = \eta_A(t, \theta) = e^{-\theta t} \qquad (\theta, t \geq 0), \tag{5}$$

if it is assumed that the initial concentration of A is 1. If we put $\lambda = 1$ in (4) we have the usual model with additive errors of constant variance. Box and Lukas [1959] show that, for this error model, the locally D-optimum design consists of taking all measurements at a time $t^* = 1/\theta$.

However, if the log transformation is appropriate to obtain errors of constant variance and $[A]$ is measured, taking logarithms of both sides of (5), combined with additive errors, yields the statistical model

$$\log y_i = \log\{\eta_A(t_i, \theta)\} + \varepsilon_i = -\theta t_i + \varepsilon_i. \tag{6}$$

The optimum design now puts all observations at the maximum possible time, when the concentration is as small as possible, a clearly absurd answer. Thus, differing assumptions about the error structure lead to different designs for the mechanistic model (5). Data analysis shows that the value of λ should lie between 0 and 1.

Taking logarithms of the square root of both sides of the power law (2) yields

$$\log(\text{s.d. } Y) = \text{const.} + (1-\lambda)\log\{E(Y)\}\,, \tag{7}$$

where s.d. is the standard deviation of Y. If replicate observations are available, a plot of log standard deviation against log mean will indicate whether the power law holds. We check this for two examples.

Lindsey [2001, p. 113] introduces data on flosequinan, which give ten sets of replicate observations at a variety of times. The plot of log standard deviation against log mean shows a linear relationship, for which the value of λ is 0.35. The more extensive data set of Downing, Fedorov, and Leonov [2001] yields a value of 0.1, not far from the log. In comparison, Horwitz' rule for analytical chemistry (Lischer [1999]) yields a value around 0.14 for λ.

2. Optimum design: multivariate response

The ith experiment consists of running the reaction for a time t_i and taking a measurement of one or more responses. As in Box and Lucas [1959], each experiment allows measurements at only a single time point because, for example, the reaction has to be terminated before the measurements can be made. In this paper, only continuous designs are found, in which the design ξ is a continuous measure specifying both a set of k distinct points in a design region $\mathcal{T}$ and the proportions, w_i, of observations taken at these points. The times t_i are the points of support of the design ξ and w_i the design weights. In the examples the calculations are of locally D-optimum designs maximizing the logarithm of the determinant of the information matrix, $\log|M(\xi,\psi)|$, in which a best guess ψ^o is taken for the parameters.

If the concentration of more than one chemical is measured, there will be a model for each expected response giving a matrix F_i of parameter sensitivities for the ith response, $i = 1,\ldots,m$,

$$F_i = \{f_{ij}(t_u,\psi)\} = \left\{\frac{\partial\eta_i(t_u,\psi)}{\partial\psi_j}\right\},$$

where $u = 1,\ldots,k$ denote the design points and $j = 1,\ldots,p$ denote the parameters. The observations follow the model

$$y_{iu} = \eta_i(t_u,\psi) + \varepsilon_{iu}\,,$$

with

$$E(\varepsilon_{iu}) = 0\,, \qquad E(\varepsilon_{iu}\varepsilon_{lv}) = \begin{cases} 0 & \text{if } u \neq v\,, \\ \sigma_{il} & \text{if } u = v\,, \end{cases}$$

when the variance-covariance matrix of the responses is

$$\Sigma = \{\sigma_{il}\}_{i,l=1,\ldots,m}\,.$$

For normally distributed errors the information matrix is given by

$$M(\xi,\psi) = \sum_{i=1}^{m}\sum_{l=1}^{m} \sigma^{il} F_i^T W F_l\,, \tag{8}$$

where $\Sigma^{-1} = \{\sigma^{il}\}_{i,l=1,\ldots,m}$ and $W = \mathrm{diag}\{w_1,\ldots,w_k\}$.

3. Parameter sensitivities and transforming both sides

The parameter sensitivities for the model in §2 with $\lambda = 1$ will be written

$$f_j^1(t,\psi) = \frac{\partial\eta(t,\psi)}{\partial\psi_j}\,. \tag{9}$$

The parameter sensitivities in the transformed model (4) are then

$$f_j^\lambda(t,\psi) = \frac{\partial\{\eta(t,\psi)\}^\lambda}{\partial\psi_j} = \lambda\{\eta(t,\psi)\}^{\lambda-1}\frac{\partial\eta(t,\psi)}{\partial\psi_j} = \lambda\{\eta(t,\psi)\}^{\lambda-1} f_j^1(t,\psi)\,. \tag{10}$$

For fixed λ, multiplication by λ in (10) does not change the optimum design, so the sensitivities have the easily calculated form

$$f_j^\lambda(t,\psi) = \{\eta(t,\psi)\}^{\lambda-1} f_j^1(t,\psi) = f_j^1(t,\psi)/\{\eta(t,\psi)\}^{1-\lambda}\,. \tag{11}$$

If $\lambda < 1$, the variance of the observations increases with the value of $\eta(t,\psi)$. Thus transformation of both sides for such values of λ will increase the relative value of the sensitivities for times where the response is small. We can expect that designs for $\lambda < 1$ will include observations at lower concentrations than those when no transformation is needed.

4. Two consecutive second-order reactions: numerical derivatives

In the model for two consecutive reactions

$$A \xrightarrow{\theta_1} B \xrightarrow{\theta_2} C\,, \tag{12}$$

the kinetic differential equations for $[A]$, $[B]$ and $[C]$, the concentrations of the chemical compounds A, B and C as functions of time t are

$$\begin{aligned} \frac{d[A]}{dt} &= -\theta_1 [A]^{\nu_1}, \\ \frac{d[B]}{dt} &= \theta_1 [A]^{\nu_1} - \theta_2 [B]^{\nu_2}, \\ \frac{d[C]}{dt} &= \theta_2 [B]^{\nu_2}, \end{aligned} \tag{13}$$

where θ_1 and θ_2 are the rates of reaction and ν_1 and ν_2 are the orders. Atkinson and Bogacka [1997] discuss the consequences for experimental design of this distinction between the two parts of ψ.

If the orders ν_i in (13) are not both one, the differential equations for the responses have to be solved numerically. The numerical values of the parameter sensitivities are found using the "direct method" (Valko and Vajda [1984]). Examples in experimental design in chemical engineering are given by Bauer *et al.* [2000]. An introduction is given by Atkinson and Bogacka [2002]. A comparison of numerical methods for finding sensitivities is given by Uciński [1999].

The response functions $\eta_i(t, \psi)$, $i = 1, \ldots, m$, are solutions of the set of m differential equations

$$\frac{d\eta_i(t,\psi)}{dt} = g_i(\eta, t, \psi) \tag{14}$$

with given initial conditions. The differential equations satisfied by the parameter sensitivities are found by differentiation of (14) to be, in matrix notation,

$$\frac{df_j(t,\psi)}{dt} = \frac{\partial g(\eta,t,\psi)}{\partial \eta} f_j(t,\psi) + \frac{\partial g(\eta,t,\psi)}{\partial \psi_j}, \qquad j = 1, \ldots, p. \tag{15}$$

These have to be solved together with equations (14) for given prior values of the parameter vector ψ.

5. Two consecutive second-order reactions: Design

As an example of the use of numerical sensitivities, consider design for the consecutive reaction (13) when both reactions are second order, that is $\nu_1 = \nu_2 = 2$. The differential equations satisfied by the sensitivities are given by Atkinson and Bogacka [2002]. They were solved using the Maple procedure "`dsolve`", which produces answers at a grid of time points. Interpolation used the Fortran subroutines `locate` and `polint` from Press *et al.* [1992, Chapter 3].

Figure 1(a) shows the three responses as a function of time. We investigate the properties of three sets of designs as λ varies: those in which only [B] is measured, those in which [C] is measured and those in which both [B] and [C] are measured. The theory of §3 indicates that as λ decreases from one, the support points of the design move towards regions of lower concentration. From the figure we can expect that the effect will be strongest on designs in which only [C] is measured. But we first consider designs in which the single response is [B].

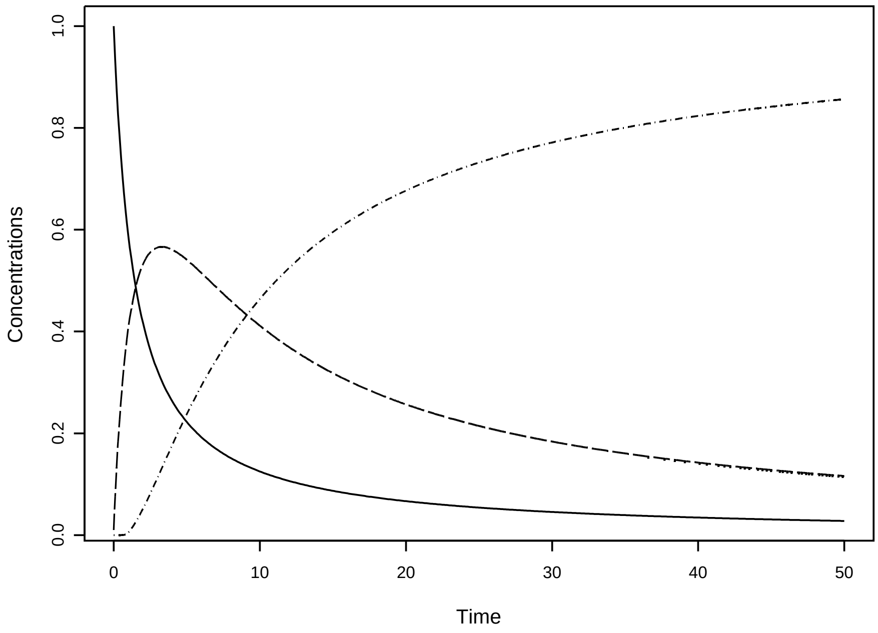
The model has four parameters: the two rate parameters θ_1 and θ_2 and the two orders of reaction ν_1 and ν_2. Figure 1(b) shows that, for all λ in the range zero to one, the D-optimum design when only [B] is measured has only four points of support. The design weights at the four points of support are therefore all equal to one quarter. The figure shows that the effect of λ on the design is not very strong. Throughout the uppermost design point is at the upper bound of the design region, $t = 50$. As λ decreases from one, the times for the first and second design points decrease steadily, whereas that for the third increases: all points are moving towards regions of lower response than their values when $\lambda = 1$.

The design when only [C] is measured is appreciably more complicated. Figure 2(a) shows that, for most values of λ, the design has four points of support. Initially the uppermost design point is at $t = 50$, but, as λ decreases, the times decrease steadily, including the largest time below $\lambda = 0.82$. When $\lambda = 0.75$ a five point design is optimum, as it is until $\lambda = 0.71$, when the four-point design includes a new design point and rejects one of the original four. Figure 2(b) shows the behaviour of the design weights, which are 0.25 for the periods of four-point designs, but change rapidly during the periods of five-point designs. The figure shows there are two further regions of five-point designs as λ decreases towards zero. These are formed by one of the design points dividing into two points close together. The general trend is that the times of the support points decrease steadily as λ does. The lowest design point has been constrained to be not less than 0.1. This design very clearly shows the move of the design points to lower values of the response as λ decreases.

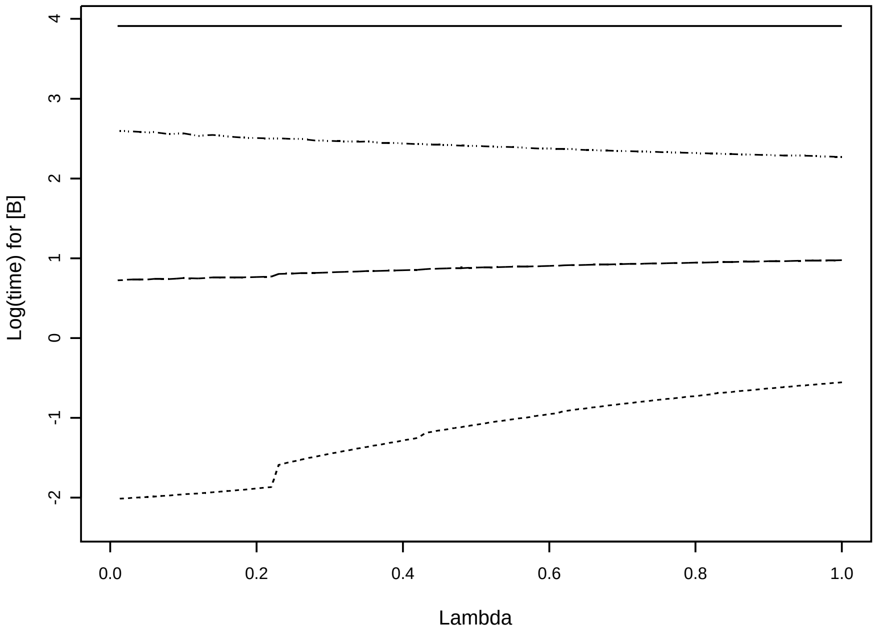
The design when both [B] and [C] are measured is rather different from these. Although there are four parameters in the model, three design points are optimum for many values of λ, as information is available from two responses. The support points of the design are in Figure 3(a) with the weights in Figure 3(b).

The times in general again trend downwards as λ decreases, although there is no weight on the upper boundary of the design region ($t = 50$) below $\lambda = 0.44$. The weights are however very different from those we have seen before. From $\lambda = 1$ down to $\lambda = 0.29$ most of the designs have three points of support with a weight near to one half at a time near 5 and weights near one quarter on

FIGURE 1. Two consecutive second-order reactions:

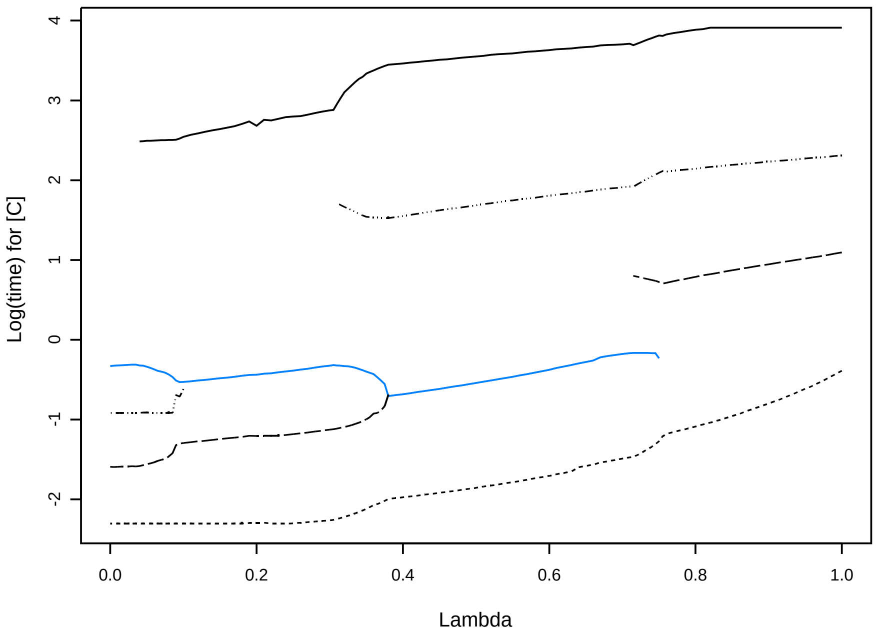


(a) responses. Reading up on the right: [A], [B] and [C];

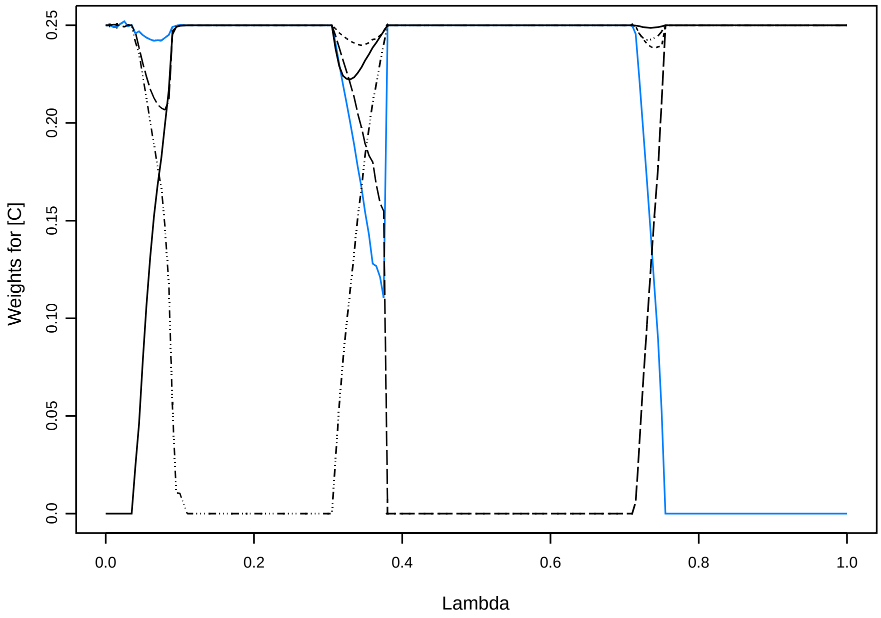


(b) logged times for the D-optimum design when only [B] is measured. All design weights are 0.25.

FIGURE 2. Two consecutive second-order reactions.
D-optimum design when only [C] is measured:

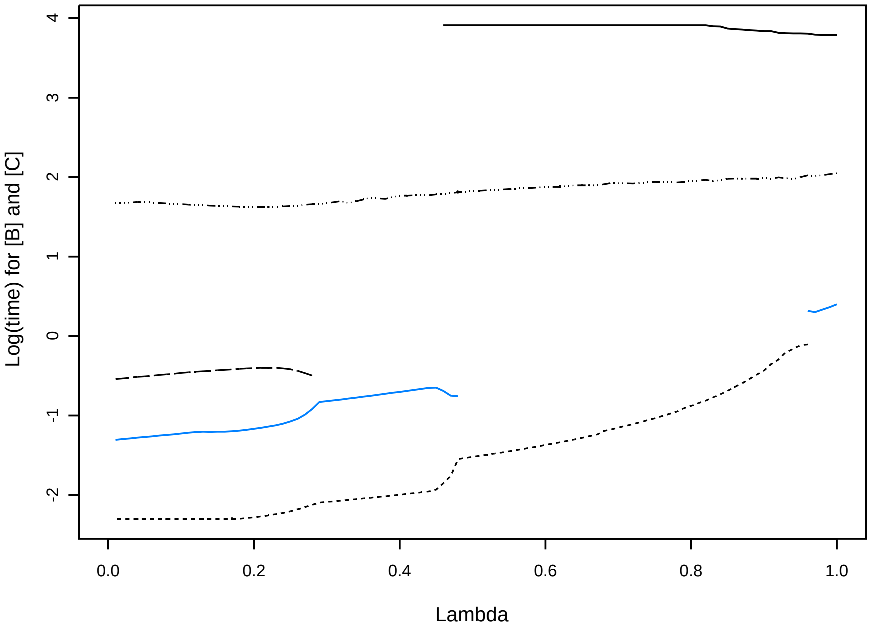


(a) logged times.

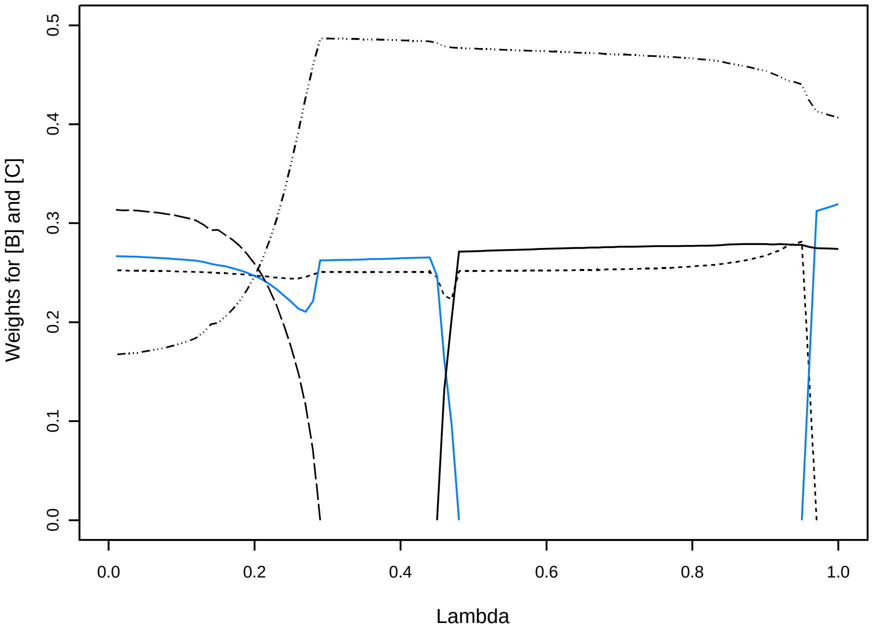


(b) design weights.

Figure 3. Two consecutive second-order reactions.
D-optimum design when both [B] and [C] are measured:



(a) logged times.



(b) design weights.

the other two points. For low values of λ the weights are nearer to one quarter on four points.

The concentration of the design points on lower values of time as λ decreases is more like the design for [C] only in Figure 2, than it is like that for [B] only. Efficiency calculations for these two designs relative to that when both responses are measured show that, as λ decreases, measuring [C] becomes much more informative than measuring [B], if measurements on only one response are possible. As λ approaches zero, the efficiency of the design rises to just over 80 %.

REFERENCES

[1] ATKINSON, A. C.: *Horwitz's rule, transforming both sides and the design of experiments for mechanistic models*, Appl. Statist. **52** (2003), 261–278.

[2] ATKINSON, A. C.—BOGACKA, B.: *Compound, D- and D_s-optimum designs for determining the order of a chemical reaction*, Technometrics **39** (1997), 347–356.

[3] ATKINSON, A. C.—BOGACKA, B.: *Compound and other optimum designs for systems of nonlinear differential equations arising in chemical kinetics*, Chemometrics and Intelligent Laboratory Systems **61** (2002), 17–33.

[4] ATKINSON, A. C.—COOK, R. D.: *Designing for a response transformation parameter*, J. Roy. Statist. Soc. B **59** (1996), 111–124.

[5] BAUER, I.—BOCK, H. G.—KÖRKEL, S.—SCHLÖDER, J. P.: *Numerical methods for optimum experimental design in* DAE *systems*, J. Comput. Appl. Math. **120** (2000), 1–25.

[6] BOX, G. E. P.—COX, D. R.: *An analysis of transformations (with discussion)*, J. Roy. Statist. Soc. Ser. B **26** (1964), 211–246.

[7] BOX, G. E. P.—LUCAS, H. L.: *Design of experiments in nonlinear situations*, Biometrika **46** (1959), 77–90.

[8] CARROLL, R. J.—RUPPERT, D.: *Transformation and Weighting in Regression*, Chapman and Hall, London, 1988.

[9] DOWNING, D.—FEDOROV, V. V.—LEONOV, S.: *Extracting information from the variance function: optimal design*, in: MODA 6–Advances in Model-Oriented Design and Analysis (A. C. Atkinson, P. Hackl, and W. G. Müller, eds.), Physica-Verlag, Heidelberg, 2001, pp. 45–52.

[10] LINDSEY, J. K.: *Nonlinear Models in Medical Statistics*, Oxford University Press, Oxford, 2001.

[11] LISCHER, P.: *Good statistical practice in analytical chemistry*, in: Probability Theory and Mathematical Statistics (B. Grigelionis, ed.), VSP, Dordrecht, 1999, pp. 1–12.

[12] PRESS, W. H.—TEUKOLSKY, S. A.—VETTERLING, W. T.—FLANNERY, B. P.: *Numerical Recipes in Fortran (2nd edition)*, Cambridge University Press, Cambridge, 1992.

[13] UCIŃSKI, D.: *Measurement Optimization for Parameter Estimation in Distributed Systems*, Technical University Press, Zielona Góra, 1999.

[14] VALKO, P.—VAJDA, S.: *An extended* ODE *solver for sensitivity calculations*, Comput. Chemistry **8** (1984), 255–271.

Received February 28, 2002

Department of Statistics
The London School of Economics
and Political Science
Houghton Street
London WC2A 2AE
GREAT BRITAIN
E-mail: a.c.atkinson@lse.ac.uk