ornl

OAK RIDGE
NATIONAL
LABORATORY

LOCKHEED MARTIN

ORNL/TM-13152

RECEIVED
APR 3 0 1996
OSTI

Design of Spatial Experiments:
Model Fitting and Prediction

Valerii V. Fedorov

MASTER

DESIGN OF SPATIAL EXPERIMENTS: MODEL FITTING AND PREDICTION

Valerii V. Fedorov

Mathematical Sciences Section
Computer Science and Mathematics Division
Oak Ridge National Laboratory
P. O. Box 2008
Oak Ridge, Tennessee 37831-6367

MASTER

DISCLAIMER

Portions of this document may be illegible
in electronic image products. Images are
produced from the best available original
document.

Contents

Abstract

The main objective of the paper is to describe and develop model oriented methods and algorithms for the design of spatial experiments. Unlike many other publications in this area, the approach proposed here is essentially based on the ideas of convex design theory.

1. Introduction

Since the earliest days of the experimental design theory, a number of concepts like split plots, strips, blocks, Latin squares, etc. (see Fisher (1947)), were strongly related to experiments with spatially distributed or allocated treatments and observations. In this survey we confine ourself to what can be considered as an intersection of ideas developed in the areas of response surface design of experiments and spatial statistics.

The results which we are going to consider are also related to the results developed by Cambanis (1985), Cambanis and Su (1993), Matern (1986), Micchelli and Wahba (1981), Sacks and Ylvisaker (1966, 1968, 1970) and Ylvisaker (1975,

1987). What differs in the approach of this paper from those cited? We intend to use the techniques which are based on the concept of regression models while the cited studies are based on the ideas developed in the theory of stochastic processes and the theory of integral approximation.

If this survey were to be written for a very applied audience, the title "optimal allocations of sensors" or "optimal allocation of observing stations" could be more appropriate. Environmental monitoring, meteorology, surveillance, some industrial experiments and seismology are the most typical areas in which the considered results may be applied. What are the most common features of the experiments to be discussed?

1. There are variables $x \in X \subset R^k$, which can be controlled. Usually $k = 2$, and in the observing station problem, x_1 and x_2 are coordinates of stations and X is a region where those stations may be allocated.

2. There exists a model describing the observed response(s) or dependent variable(s) y. More specifically y and x are linked together by a model, which may contain some stochastic components.

3. An experimenter or a practitioner can formulate the quantitative objective function.

4. Once a station or a sensor is allocated a response y can be observed either continuously or according to any given time schedule without any additional significant expense.

5. Observations made at different sites may be correlated.

Assumptions 1 - 5 are very loosely formulated and they will be justified when needed. In the subsequent sections the term "sensor" stands for what could be an observing station, meteorological station, radiosonde or well in the particular applied problem.

2. Standard design problem

In what follows we will mostly refer to experiments which are typical in environmental monitoring setting as a background for the exposition of the main results. We hope that the reader will be able to apply the ideas and techniques to other types of experiments.

When assumptions 4 and 5 are not considered we have what will be be called, the "standard design problem". The problem was extensively discussed (see for instance, Atkinson and Donev (1992), Fedorov (1972), Pazman (1986), Pukelsheim (1994) and Silvey (1980)), and it is difficult to add anything new in this area of experimental design theory. Theorem 1 which follows, is a generalized version of the Kiefer-Wolfowitz equivalence theorem, (see Kiefer (1959)) and stated here for the reader's convenience. It also serves as an opportunity to introduce the notation, which is sometimes different from that used in other articles of this volume.

Let

$$y_{ij} = \eta(x_i, \theta) + \varepsilon_{ij}, \tag{2.1}$$

$$i = 1, \ldots, n, \quad j = 1, \ldots, r_i \quad \sum r_i = N,$$

$$\eta(x, \theta) = \theta^T f(x),$$

where $\theta \in R^m$ are unknown parameters, $f^T(x) = (f_1(x) \ldots, f_m(x))$ are given functions, supporting points x_i are chosen from some set X, and the ε_{ij} are uncorrelated random errors with zero means and variances equal to one. We do not make distinctions in notation for random variables and their realizations when it is not confusing.

For the best linear unbiased estimator of unknown parameters the accumulated "precision" is described by the information matrix:

$$M(\xi) = \sum p_i f(x_i) f^T(x_i), \ p_i = r_i/N, \tag{2.2}$$

which is completely defined by the design $\xi = \{x_i, p_i\}_i^n$. In the context of the

standard design theory

$$M(\xi) = \int f(x)f^T(x)\xi(dx) = \int m(x)\xi(dx), \qquad (2.3)$$

where $\xi(dx)$ is a probability measure with the supporting set belonging to X : supp$\xi \subset X$, and

$$m(x) = f(x)f^T(x)$$

is the information matrix of an observation made at point x.

Regression model (2.1) and the subsequent comments do satisfy assumptions 1 and 2 from the previous section. To be consistent with assumption 3 let us introduce a function $\Psi(M)$, which is called the "criterion of optimality" in experimental design literature. A design

$$\xi^* = \arg\min_\xi \Psi\left[M(\xi)\right], \int \xi(dx) = 1, \qquad (2.4)$$

is called $(\Psi-)$ optimal.

Minimization must be over the set of all possible probability measures Ξ with supporting sets belonging to X. Now let us assume that:

(a) X is compact;

(b) $f(x)$ are continuous functions in X, $f \in R^m$;

(c) $\Psi(M)$ is a convex function and $\Psi(M) \leq \Psi(M + \Delta), M \geq 0, \Delta \geq 0$,

i.e. matrices M and Δ are nonnegative definite.

(d) there exists a real number q such that

$$\{\xi : \Psi[M(\xi)] \leq q < \infty\} = \Xi(q) \neq \emptyset;$$

(e) for any $\xi \in \Xi(q)$ and $\bar{\xi} \in \Xi(q)$:

$$\Psi\left[(1-\alpha)M(\xi) + \alpha M(\bar{\xi})\right] = \Psi\left[M(\xi)\right] + \alpha \int \psi(x,\xi)\bar{\xi}(dx) + \tau(\alpha,\xi,\bar{\xi}),$$

where $\tau(\alpha, \xi, \overline{\xi}) = o(\alpha)$.

Here and in what follows we use $\Psi(\xi)$ for $\Psi[M(\xi)]$, Ψ^* for $\Psi(\xi^*)$ and $\min_x, \min_\xi, \int$, and so on, instead of $\min_{x \in X}$, $\min_{\xi \in \Xi}$, $\int_X$, respectively, if it does not lead to ambiguity.

Theorem 1. . *If (a)–(e) hold, then*

1. *For any optimal design there exists a design with the same information matrix which contains no more than $n = m(m+1)/2$ supporting points.*

2. *A necessary and sufficient condition for a design ξ^* to be optimal is fulfillment of the inequality:*
$$\min_x \psi(x, \xi^*) \geq 0.$$

3. *The set of optimal designs is convex.*

4. $\psi(x, \xi^*)$ *achieves zero almost everywhere in $\text{supp}\,\xi^*$,*

where $\text{supp}\,\xi$, stands for supporting set of the design (measure) ξ.

Here and

Functions $\psi(x, \xi)$ for the most popular criteria of optimality may be found, for instance, in Atkinson and Fedorov (1984). Theorem 1 provides a starting point for analytical exercises with various relatively simple regression problems and makes possible the development of a number of simple numerical procedures for the optimal design construction in more complicated and more realistic situations. Most of these procedures are based on the following iterative scheme:

- (a) There is a design $\xi_s \in \Xi(q)$. Find

$$x_s = \arg\min\left\{\psi(x^+, \xi_s), \psi(x^-, \xi_s)\right\}, \tag{2.5}$$

$$x^+ = \arg\min_{x \in X} \psi(x, \xi_s), \quad x^- = \arg\max_{x \in X_s} \psi(x, \xi_s),$$

where $X_s = \text{supp}\,\xi_s$.

- (b) Choose $0 < \beta_s < 1$ and construct

$$\xi_{s+1} = (1 - \beta_s)\xi_s + \beta_s\xi(x_s), \qquad (2.6)$$

where $\xi(x_s)$ is a measure atomized at x_s.

The choice of a sequence β_s defines a variety of the algorithms; specific examples are given by Atkinson and Donev (1992), Cook and Nachtsheim (1989), Fedorov (1972, 1975) and Silvey (1980). The following sequences are most popular:

- $\beta_s = \alpha_s$, if $x_s = x^+$, and
  $\beta_s = \min\left\{\alpha_s, \xi_s(x^-)/(1 - \xi_s(x^-))\right\}$, if $x_s = x^-$,
  where $\lim \alpha_s = 0$, $\sum_{s=0}^{\infty} \alpha_s = \infty$;

- $x = \arg\min_x \psi\left[M(\xi_{s+1}(x))\right]$,
  where $\xi_{s+1}(x) = (1 - \beta)\xi_s + \beta\xi(x_s)$;

- $\alpha_s = \begin{cases} \beta_{s-1}, & \text{if } \psi\left[M(\xi_s(\beta_{s-1}))\right] < \psi[M(\xi_{s-1})]; \\ \beta_{s-1}/\gamma, & \gamma > 1 \quad \text{otherwise.} \end{cases}$

Theorem 1 together with iterative procedure (2.5), (2.6) provides quite powerful tools for constructing optimal design. The existing software products, see, for instance, Mitchell (1974), Nguen and Miller (1992), Nachtsheim (1987), SAS/QC Software (1995), Wheeler (1994), confirm this statement. Unfortunately, there are a few hurdles, which do not allow the direct use of the results reported above. The first one is that optimal designs defined by (2.4) may have unequal weights. What does this mean in the context of observing stations allocation? If we have N available stations or sensors, then $r_i^* = [p_i^* N]$ stations must be allocated at x_i^*, where $[p_i^* N]$ is some reasonable integer approximation of $p_i^* N$. It is obvious that in many cases (but not always) two or more stations sited in the immediate vicinity of each other will not give essentially more information than a single station. There are some arguments in favor of this statement, which can be expressed economically in colloquial statistical terminology: observations from these stations are strongly correlated. However, frequently weights p_i^* may be

considered as the desirable precision of measurements taken at the i-th station. The corresponding precision can be achieved through the proper technical steps or through controlling the longevity of the observational process.

Probably, Gribik et al (1976), were the first to use the optimal experimental design methods for environmental monitoring. They analyzed the problem of allocating measuring resources to aid in accurately estimating ground level pollution concentrations throughout a region X. The regression model was the linearized version of the diffusion model for four pollution sources and unknown background source. Since the diffusion model used in the study was a large scale model, measurements separated by distances smaller than a threshold value distance appeared to be correlated in the corresponding parameter estimation problem. At the same time the design method was a particular case of the method discussed in this section, where the independence of observational errors is essential. To avoid a contradiction the authors imposed the additional constraint: the distance between any two observing sites must be greater than the characteristic distance:

$$(x_i - x_j)^T (x_i - x_j) > d^2.$$

Imposing constraints of that type is one of the simplest way to handle possible correlation between the observed values at neighboring stations. Obviously the approach does not work for long-range correlation, when the widely separated observations are correlated.

It was assumed that the ground level pollution is of the prime interest. The authors proposed to use the weighted average variance of the best linear unbiased estimator of the ground level pollution:

$$\Psi[M(\xi)] = \int_X w(x) f^T(x) M^{-1}(\xi) f(x) dx$$

as the criterion of optimality. The weight function $w(x)$ was selected proportional to the population density in the considered region.

A rather detailed discussion of applicability of the standard design technique

for spatial experiments may be found in Fedorov et al(1988).

3. Optimal designs with bounded density

Gribik et al (1976), used a very simple and transparent idea to avoid clustering of sensors at particular points. This idea can be exploited in a more general and formal setting. Let the number of sensors N be sufficiently large and the density of stations per square unit be introduced into consideration:

$$\xi(dx) = \lim_{\Delta X \to 0} \frac{N(\Delta X)}{N}. \tag{3.1}$$

Introduction of (3.1) is very reasonable when the sensor allocation is considered in technological experiments. In the network allocation problem it is probably less realistic. Nevertheless, the results considered in this section help to explain why some intuitive approaches, similar to what was done by Gribik et al (1976), do work well in most cases.

If X is not uniform (as might be appropriate say, with different topography for different parts of X), then it is natural to assume that the sensor density has to be constrained:

$$\Phi_1(dx) \leq \xi(dx) \leq \Phi_2(dx).$$

With obvious redefining of the design measure $\xi(dx)$ and the upper bound $\Phi_2(dx)$ the latter may be reduced to a simpler statement:

$$0 \leq \xi(dx) \leq \Phi(dx).$$

Thus, the following optimization problem must be considered, (we skip the evident left hand side constraint):

$$\xi^* = \arg\min_{\xi} \Psi\left[M(\xi)\right], \tag{3.2}$$

$$\int \xi(dx) = 1, \ \xi(dx) \leq \Phi(dx) \tag{3.3}$$

This optimization problem was discussed by Wynn (1982), and Fedorov (1989). To avoid unnecessary technical complications let us assume additionally to (a)-(e) from section 2 that

(f) $\Phi(dx)$ is atomless, i.e.

$$\lim_{\Delta x \to 0} \int_{\Delta x} \Phi(dx) = 0$$

The following theorem summarizes the most important properties of designs with bounded density.

Theorem 2. . *Let Ξ_0 be a set of design ξ such that $\xi(dx) = \Phi(dx)$, when $\xi(dx) > 0$, and $\xi(dx) = 0$ otherwise, and let assumptions (a) - (f) hold. Then:*

- *There exists an optimal designs $\xi^* \in \Xi_0$.*

- *A necessary and sufficient condition for this design to be optimal is that $\psi(x, \xi^*)$ separates the two sets $X^* = supp\xi^*$ and its complement.*

In the above formulation "separate" means that there is a constant C such that $\psi(x, \xi^*) \leq C$ on X^* and $\psi(x, \xi*) > C$ on its complement.

Theorem 1 tells us that supporting sets of optimal designs must coincide with the points where $\psi(x, \xi^*)$ achieves its minimum. Therefore, in most cases the supporting set for the standard optimal design consists of a finite number of supporting points.

Theorem 2 forces $supp\xi^*$ to occupy the subsets of X. How is $\xi^*(dx)$ to be realized by a practitioner? One of the possibilities is to replace $\xi^*(\Delta X)$ for relatively small areas ΔX by $N^*(\Delta X) = [\xi^*(\Delta X)N]$. When $N^*(\Delta X)$ is defined then the corresponding number of sensors have to be allocated in ΔX. For instance, they can be sited at the nodes of some uniform grid. Generally, that allocation has to guarantee a reasonable approximation of the integral

$$\int_{\Delta X} \psi(x, \xi^*)\xi^*(dx)$$

by the sum

$$\sum_{x_i \in \Delta X} \psi(x_i, \xi^*) \Delta X_i, \ x_i \in \Delta X_i, \bigcup_i \Delta X_i = \Delta X, \Delta X_i \bigcap \Delta X_j = \emptyset, i \neq j.$$

The properties described by Theorem 2 allow us to formulate a simple numerical algorithm to construct optimal designs (see Fedorov (1989)). Let $\Phi(dx) = \phi(x)dx$ and

$$\lim_{s \to \infty} \alpha_s = 0, \ \lim_{s \to \infty} \sum_{s'=1}^{s} \alpha_{s'} = \infty \ \text{ and } \ \lim_{s \to \infty} \sum_{s'=1}^{s} \alpha_{s'}^2 < \infty.$$

(a) There is a design $\xi_s \in \Xi_0$. Let $X_{1s} = \text{supp}\xi_s$ and $X_{2s} = X \setminus X_{1s}$. Two sets $D_s \subset X_{1s}$ and $E_s \subset X_{2s}$ with equal measure,

$$\int_{D_s} \phi(x)dx = \int_{E_s} \phi(x)dx = \alpha_s,$$

and, correspondingly, including the points

$$x_{1s} = \arg \max_{x \in X_{1s}} \psi(x, \xi_s) \ \text{ and } \ x_{2s} = \arg \min_{x \in X_{2s}} \psi(x, \xi_s),$$

(b) The design ξ_{s+1} with the supporting set

$$\text{supp}\xi_{s+1} = X_{1(s+1)} = (X_{1s} \setminus D_s) \bigcup E_s$$

is constructed.

Usually $\phi(x)$ is assumed to be constant. All other cases may be converted to this one with the proper coordinate transformation. In the computerized version of the algorithm integrals in (a) are replaced with sums over some grid elements. If these elements and subsequently α_s are fixed and elements of both D_s and E_s coincide with the grid elements, then (a), (b) becomes an exchange type algorithm (see, for instance, Mitchell (1974)) with the simple constraint: every grid element cannot contain more than one supporting point and the weights of all supporting points are the same, i.e. N^{-1}. In practice it is sometimes convenient to consider

grids of varying density, which has to be proportional to $\phi(x)$. While it can be shown that the exchange algorithm (a), (b) converges to an optimal design for properly diminishing α_s , it is not generally true for finite α_s and, in particular, when $\alpha_s \equiv N^{-1}$. The accuracy of the limit designs is defined by the accuracy of the approximation (see assumption (e) from Section 2):

$$\Psi\left[(1-\alpha_s)M(\xi_s) + \alpha_s M(\xi)\right] \cong \Psi\left[M(\xi_s)\right] + \alpha_s \int \psi(x,\xi_s)\xi(dx)$$

$$\int_{\Delta X_i} \psi(x,\xi_s)\xi(dx) \simeq \psi(x_i,\xi_s)\Delta X_i.$$

When these approximations are reliable enough then we can hope that the limit designs do not deviate too much from the optimal ones. The term "limit design" must be used with some reservation when $\alpha_s \equiv N^{-1}$: instead of convergence some minor oscillations of $\Psi\left[M(\xi_s)\right]$ may be observed. Practical aspects of the iterative procedure (a),(b) were discussed by Fedorov and Müller(1989b) in the air pollution network design setting.

4. Correlated observational errors

Let us assume now that the random errors in model (2.1) are correlated and that the covariance structure is known, i.e. either the covariance matrix V or the covariance function $V(x, x')$ is given. There is no need to use the second subscript indicating the repeated observations and we consider

$$y_i = \eta(x_i, \theta) + \varepsilon_i, \tag{4.1}$$

where $i = 1, \ldots, N, E(\varepsilon_i) \equiv 0$ and

$$V(\xi_N) = \{E(\varepsilon_i \varepsilon_l), \}_{1,1}^N = \{V(x_i, x_l)\}_{1,1}^N.$$

For the obvious reason, in this section we will use the simplified notation:

$$\xi_N = \{x_1, \ldots, x_N\}.$$

In the case of correlated observations the best linear unbiased estimator is defined as (see, for instance, Rao (1973)):

$$\hat{\theta} = \underline{M}^{-1}(\xi_N)F(\xi_N)V^{-1}(\xi_N)Y, \tag{4.2}$$

and its dispersion matrix equals

$$Var(\hat{\theta}) = \underline{M}^{-1}(\xi_N), \tag{4.3}$$

where

$$Y^T = (y_1, \ldots, y_N), \quad F(\xi_N) = (f(x_1), \ldots, f(x_N)),$$

$$\underline{M}(\xi_N) = F(\xi_N)V^{-1}(\xi_N)F^T(\xi_N) = NM(\xi_N). \tag{4.4}$$

The best linear unbiased predictor at a point x is

$$\hat{y}(x) = f^T(x)\hat{\theta} + V^T(x, \xi_N)V^{-1}(\xi_N)(Y - F^T(\xi_N)\hat{\theta}), \tag{4.5}$$

$$E\left[(y(x) - \hat{y}(x))^2\right] = S^2(x, \xi_N) + \phi^T(x, \xi_N)\underline{M}^{-1}(\xi_N)\phi(x, \xi_N), \tag{4.6}$$

where

$$S^2(x, \xi_N,) = V(x, x) - V^T(x, \xi_N)V^{-1}(\xi_N)V(x, \xi_N),$$
$$\phi(x, \xi_N) = f(x) - F(\xi_N)V^{-1}(\xi_N)V(x, \xi_N),$$

Unlike (2.2), the information matrix (4.4) is not a sum of information matrices of single observations. Therefore we cannot use directly the results of the convex design theory, which is essentially based on the additivity of information matrices.

Actually, we have to consider the optimization problem

$$\xi_N^* = \arg\min_{\xi_N} \Psi\left[M(\xi)\right], \tag{4.7}$$

which does not have too much in common with (5) besides notation. For instance, the convexity of Ψ is not very helpful anymore.

In most studies authors try to imitate the iterative methods of optimal design construction considered in two previous sections. For instance, computations become similar to the standard (uncorrelated) case, if the following recursion formula is used (Brimkulov et al (1986)):

$$\underline{M}(\xi_{N+1}) = \underline{M}(\xi_N) + w(x, \xi_N)\phi(x, \xi_N)\phi^T(x, \xi_N), \tag{4.8}$$

where

$$\xi_{N+1} = \{\xi_N, x\} \text{ and } w^{-1}(x, \xi_N) = S^2(x, \xi_N).$$

We can easily derive, for instance, that

$$|\underline{M}(\xi_{N+1})| = |\underline{M}(\xi_N)|\left[1 + w(x, \xi_N)\phi^T(x, \xi_N)\underline{M}^{-1}(\xi_N)\phi(x, \xi_N)\right]. \tag{4.9}$$

Subsequently, for the D-criterion the point

$$x_{N+1} = \arg\max_{x \in X} w(x, \xi_N)\phi^T(x, \xi_N)\underline{M}^{-1}(\xi_N)\phi(x, \xi_N), \tag{4.10}$$

must be added to the design ξ_N. That is an imitation of step (a) from the iterative procedures considered in the two previous sections.

There exists a simple intuitive explanation why iterative procedures based on (4.10) provide "good" supporting points in the sense of the D-criterion. First, let us recollect that in the no-correlation case accordingly to stage (a) of the iterative procedure from Section 2 the additional observation(s) must be allocated

at point(s), where the ratio

$$\frac{\text{variance of prediction with the estimated } \theta}{\text{variance of prediction with the given } \theta}$$

$$= \frac{\sigma^2(x) + f^T(x)\underline{M}^{-1}(\xi)f(x)}{\sigma^2(x)} \qquad (4.11)$$

is maximal. This follows, for instance, from (2.5) when

$$\Psi(M) = -ln|M| \quad \text{and} \quad \psi(x,\xi) = m - f^T(x)M^{-1}(\xi)f(x) \text{ for } \sigma^2(x) \equiv 1$$

or

$$\psi(x,\xi) = m - \sigma^{-2}(x)f^T(x)M^{-1}(\xi)f(x),$$

in the more general case (see Fedorov (1972)) for details. In the case of correlated observations we are looking for a maximum of the same ratio

$$\frac{\text{variance of prediction with the estimated } \theta}{\text{variance of prediction with the given } \theta}$$

$$= \frac{S^2(x,\xi_N) + \psi^T(x,\xi_n)\underline{M}^{-1}(\xi_N)\psi(x,\xi_N)}{S^2(x,\xi_N)} \qquad (4.12)$$

When $x \to x_j \in \text{supp}\xi_N$, then

$$w(x,\xi_N)\psi^T(x,\xi_N)\underline{M}^{-1}(\xi_N)\psi(x,\xi_N)) \to 0 \qquad (4.13)$$

for $f(x)$ and $V(x,\xi_N)$ continuous in the vicinity of x. In other words the iterative procedure defined by (4.10) does not admit coinciding supporting points. The result follows from the definitions of $\psi(x,\xi_N)$ and $S^2(x,\xi_N)$, and the fact that

$$\left\{V^{-1}(\xi_N)V(x_i,\xi_N)\right\}_j = \delta_{ij},$$

where δ_{ij} is the Kronecker symbol. Formula (4.10) can be easily rewritten for the deleting procedure. ¿From (4.9) it follows that in the case of the D-criterion

candidates for deleting are defined by the equation:

$$x_{N+1}^- = \arg \min_\ell w(x_\ell, \xi_{N+1}) \phi^T(x_\ell, \xi_{N+1}) \underline{M}^{-1}(\xi_{N+1}) \phi(x_\ell, \xi_{N+1}). \qquad (4.14)$$

We are not aware of any results on the properties of the iterative procedures based on (4.10) and (4.14) for the D-criterion or similar procedures for other criteria. There are empirical confirmations that the exchange-type algorithms lead to a significant improvement of the starting design. For instance, Rabinowitz and Steinberg (1990) applied that type of algorithm to the problem of selecting sites for a seismographic network. They have shown that the computed designs are relatively efficient and are better than the standard D-optimal designs constructed for models with uncorrelated observations. It is reasonable to note that computationally (4.9) and (4.14) are much more demanding than their counterparts in the standard design theory. There exist a number of studies where the optimization problem (4.1) is considered for some special and relatively simple covariance functions, for instance, generated by autoregressive models. Various details and further references may be found in Bickel and Herzberg (1979), Bishoff (1992), Kunert (1988), Martin (1986), Müller and Pázman (1995) and Näther (1985).

In conclusion of this section let us emphasize again the significant difference between the case with correlated observations and the standard case. For uncorrelated observations the additiveness of the normalized information matrix $(\sigma(x) = 1)$:

$$M(\xi) = N^{-1} \sum_{i=1}^{N} f(x_i) f^T(x_i) = N^{-1} \underline{M}(\xi_N)$$

leads to many simple and elegant theoretical results initiated by Kiefer's pioneering findings. Very frequently normalized information matrices may be treated as a limit, i.e.:

$$M(\xi) = \lim_{N \to \infty} N^{-1} \underline{M}(\xi_N). \qquad (4.15)$$

In many cases for correlated observations the corresponding limit does not exist and the matrix $M(\xi)$ cannot be introduced. One of the most successful attempts

to replace (4.15) was due to by Sacks and Ylvisaker (1966, 1968); see more in Section 7.

5. Random coefficients regression models: Trend estimation

In what follows we intend to consider some simple models for the random component in (2.1). It is convenient to partition "intrinsic" or "process", and "observational" sources of randomness:

$$y_{ij} = \eta(x_i, \theta) + u_{ij} + \varepsilon_{ij}. \tag{5.1}$$

Values u_{ij} describes deviations of the observed response from $\eta(x_i, \theta)$ due to some causes which are independent of an observer. For instance, an average wind velocity may be disturbed by various local micro-eddies. The term ε_{ij} describes "observational" errors. Sometimes these errors are defined by the selected observational technique and, at least partly, they are controlled by an observer. The proposed partitioning is very conditional, and the reader may use a different one, which is more compatible with the corresponding experimental situation.

Let us assume that

$$y_{ij} = \theta_1^T f_1(x_i) + \theta_{2j}^T f_2(x_i) + \varepsilon_{ij}, \; i = 1,\ldots,N, \; j = 1,\ldots,k, \tag{5.2}$$

or

$$Y_j = F_1(\xi_N)^T \theta_1 + F_2(\xi_N)^T \theta_{2j} + \varepsilon_j, \tag{5.3}$$

where

$$\theta_1 \in R^t, \theta_{2j} \in R^l, \; t + l = m, \; F_\alpha(\xi_N) = (f_\alpha(x_1),\ldots,f_\alpha(x_N)), \; \alpha = 1,2,$$

$$\varepsilon_j^T = (\varepsilon_{1j},\ldots,\varepsilon_{Nj}).$$

Vector θ_2 is random with

$$E(\theta_2) = 0, \quad E(\theta_2 \theta_2^T) = Var(\theta_2) = \Lambda, \tag{5.4}$$

vector ε describes the observational errors, which are random and

$$E(\varepsilon) = 0 \quad E(\varepsilon \varepsilon^T) = \sigma^2 I.$$

We assume that θ_2 and ε are uncorrelated. In terms of (5.1) we have

$$E(u) = 0, E(uu^T) = F_2^T(\xi_N)\Lambda F_2(\xi_N).$$

If $e = u + \varepsilon$, then

$$E(e) = 0, \quad E(ee^T) = V(\xi_N) = \sigma^2 I + F_2^T(\xi_N)\Lambda F_2(\xi_N). \tag{5.5}$$

Thus we are going to consider a very special case of (4.1) with $V(\xi_N)$ defined by (5.5). It may be illuminating to associate index "j" with time (hour, day, ...) and "i" with location (x_i is a vector of coordinates of a particular site).

Model (5.5) gives an opportunity to introduce criteria of optimality which provide a very reasonable description of various experimental situations. Those criteria may be divided in two main groups. The first group is related to the "average over time" behavior of the observed response. The corresponding criteria depend upon the precision of estimators of θ_1. This means that we consider some functions of $Var(\tilde{\theta}_1)$, where $\tilde{\theta}_1$ is an estimator of θ_1.

The second group deals with "instant" responses and the corresponding criteria are based on $Var(\tilde{\theta}), \theta^T = (\theta_1^T, \theta_2^T)$.

Let us start with the first group, i.e. with estimating the subvector θ_1. The best linear unbiased estimator is (compare with (4.3):

$$\hat{\theta}_1 = \underline{M}_{11}^{-1}(\xi_N) F_1(\xi_N) V^{-1}(\xi_N) \sum_{j=1}^{k} k^{-1} Y_j \tag{5.6}$$

The dispersion matrix of $\hat{\theta}$ is

$$k \cdot Var(\hat{\theta}_1) = \underline{M}_{11}^{-1}(\xi_N), \tag{5.7}$$

where

$$\underline{M}_{11}(\xi_N) = F_1(\xi_N)V^{-1}(\xi_N)F_1^T(\xi_N). \tag{5.8}$$

An optimal design (or optimal observational network) is defined as

$$\xi_N^* = \arg\min_{\xi_N} \Psi\left[\underline{M}_{11}(\xi_N)\right], \tag{5.9}$$

which differs from (4.7) only by the more detailed information about $V(\xi_N)$. It may be expedient to note that unlike the situation described in comments accompanying (4.13) the covariance is not anymore a continuous function at the diagonal:

$$\lim_{x \to x'} E(e(x)e(x')) \neq \sigma^2 + f_2^T(x)\Lambda f_2(x'). \tag{5.10}$$

Therefore (5.9) may admit designs with repeated observations, i.e. it could be that $x \in \text{supp}\xi_N$ and $M(\xi_N + x)$ is better than $M(\xi_N)$.

Using the identity

$$(A + BDB^T)^{-1} = A^{-1} - A^{-1}B(B^TA^{-1}B + D^{-1})^{-1}B^TA^{-1} \tag{5.11}$$

and assuming the existence of the Λ^{-1}, one can find that

$$\underline{M}_{11}(\xi_N) = \overline{M}_{11}(\xi_N) - \overline{M}_{12}(\xi_N)\left[\overline{M}_{22}(\xi_N) + \Lambda^{-1}\right]^{-1}\overline{M}_{21}(\xi_N), \tag{5.12}$$

where $\overline{M}_{\alpha\beta}(\xi_N) = \sigma^{-2}F_\alpha(\xi_N)F_\beta^T(\xi_N)$

Now let us consider the matrix

$$D(\xi_N) = \begin{pmatrix} D_{11}(\xi_N) & D_{12}(\xi_N) \\ D_{21}(\xi_N) & D_{22}(\xi_N) \end{pmatrix} = \begin{pmatrix} \overline{M}_{11}(\xi_N) & \overline{M}_{12}(\xi_N) \\ \overline{M}_{21}(\xi_N) & \overline{M}_{22}(\xi_N) + \Lambda^{-1} \end{pmatrix}^{-1}.$$

¿From the Frobenius formula it follows that

$$D_{11}(\xi_N) = \underline{M}_{11}^{-1}(\xi_N).$$

The matrix $D(\xi_N)$ can be also considered as the dispersion matrix of the best linear unbiased estimator of parameter $\theta^T = (\theta_1^T, \theta_2^T)$ for the regression model.

$$y = \theta^T F(\xi_N) + \varepsilon,$$

where $F^T(\xi_N) = \left(F_1^T(\xi_N), F_2^T(\xi_N) \right), E(\varepsilon) = 0, E(\varepsilon\varepsilon^T) = \sigma^2 I$, with the prior information about parameters θ_2 described by a prior distribution $P(\theta_2)$ such that

$$E(\theta_2) = \int \theta_2 dP(\theta_2) = 0$$

and

$$\Lambda = E(\theta_2 \theta_2^T) = \int \theta_2 \theta_2^T dP(\theta_2),$$

See, for instance, Fedorov (1972), Pilz (1991), and Seber (1977). Thus, the optimization problem (4.1) may be embedded in the framework of convex design theory. For instance, for the D-criterion, when $|\underline{M}_{11}(\xi_N)|^{-1} = |D_{11}(\xi_N)|$ must be minimized, one can use any algorithm developed for the construction of "exact" or "discrete" optimal designs; see, for instance, Cook and Nachtsheim (1980), Fedorov (1972), Ermakov (1983), and Pukelsheim (1993) when only the subvector θ_1 has to be estimated. More generally we can now describe experimental design as the following optimization problem

$$\xi^* = \arg\min_{\xi} \Psi \left(M(\xi) + M_0 \right), \tag{5.13}$$

where

$$M_0 = \sigma^2 N^{-1} \begin{pmatrix} 0 & 0 \\ 0 & \Lambda^{-1} \end{pmatrix},$$

N is now the total number of possible observations, and

$$M(\xi) = \int f(x)f^T(x)\xi(dx),$$

$$f^T(x) = (f_1^T(x), f_2^T(x)).$$

Let us note that occasionally the total number of observations and the number of supporting points may coincide (like in (5.2). Then N stands for both. The results of Sections 2 and 3 may routinely be applied to (5.13) when Ψ is properly defined and σ^2 and Λ are known.

Subset D-optimality. According to (4.10) we have to minimize some function of the matrix $\underline{M_{11}^{-1}}$, when the parameters θ_1, are of prime interest. In terms of (5.13) it means that the objective function Ψ must depend upon elements of the matrix $\Delta_{11}(\xi)$, which may be defined as follows:

$$\Delta(\xi) = \begin{pmatrix} \Delta_{11}(\xi)\,\Delta_{12}(\xi) \\ \Delta_{21}(\xi)\,\Delta_{22}(\xi) \end{pmatrix} = \begin{pmatrix} M_{11}(\xi) & M_{12}(\xi) \\ M_{21}(\xi) & M_{22}(\xi) + M_{022} \end{pmatrix}^{-1} \tag{5.14}$$

where

$$M_{022} = \sigma^2 N^{-1}\Lambda^{-1}.$$

One of the possibilities is to select

$$\Psi\left(M(\xi) + M_0\right) = \ln|\Delta_{11}(\xi)|.$$

¿From Theorem 1 it immediately follows that a necessary and sufficient condition for ξ^* to be optimal is that (compare with Fedorov, 1972):

$$f^T(x)\Delta(\xi^*)f(x) - f_2^T(x)\left(M_{22}(\xi^*) + M_{022}\right)^{-1}f_2(x)$$
$$\leq \operatorname{tr}\Delta(\xi^*)M(\xi^*) - \operatorname{tr}\left(M_{22}(\xi^*) + M_{022}\right)^{-1}M_{22}(\xi^*). \tag{5.15}$$

Notice that

$$v_1(x,\xi) = f^T(x)\Delta(\xi)f(x)$$

is the normalized variance of $\hat{\theta}^T f(x)$, where $\hat{\theta}$ is the best linear unbiased estimator of θ from model (5.2), and

$$v_2(x, \xi) = f_2^T(x) \left(M_{22}(\xi^*) + M_{022} \right)^{-1} f_2(x)$$

may be considered as a normalized variance of the best linear estimator for the regression model with the same observational errors but with the response $\theta_2^T f_2(x)$. To get non-normalized values we have to multiply the normalized values by $\sigma^2 N^{-1}$. Simple, but rather long matrix calculations show that

$$v_1(x, \xi) = d(x, \xi) + R(x, x) - R^T(x, \xi) R(\xi) R(x, \xi)$$

$$+ \beta^T(x, \xi) \left(R^{-1}(\xi) + \rho^{-1}(\xi) \right)^{-1} \beta(x, \xi), \qquad (5.16)$$

where

$$
\begin{aligned}
d(x, \xi) &= f_1^T(x) M_{11}^{-1}(\xi) f_1(x), \quad R(x, x') = f_2^T(x) M_{022}^{-1} f_2(x'), \\[2mm]
R(x, \xi) &= F_2^T M_{022}^{-1} f_2(x), \qquad R(\xi) = F_2^T M_{022}^{-1} F_2, \\[2mm]
\rho^{-1}(\xi) &= W - W d(\xi) W, \ \beta(x, \xi) = W \underline{d}(x, \xi) - R^{-1}(\xi) R(x, \xi), \\[2mm]
d(\xi) &= F_1^T M_{11}^{-1} F_1, \quad \underline{d}(x, \xi) = F_1^T M_{11}^{-1} f_1(x), \\[2mm]
F_1 &= (f_1(x_1), \ldots, f_1(x_N)), \quad F_2 = (f_2(x_1), \ldots, f_2(x_N)), \\[2mm]
W_{ik} &= \delta_{ik} p_i, \quad p_i = \xi(x_i), \quad \{x_1 \ldots, x_n\} = \operatorname{supp}\xi.
\end{aligned}
$$

$$(5.17)$$

This collection of formulae looks much more complicated than the similar terms in (5.15). However, that presentation has one remarkable feature: it does not depend upon functions $f_2(x)$ and matrix Λ explicitly. All elements in (5.16) are

completely defined by the covariance function

$$\sigma^{-2} N R(x, x') = E\left(f_2(x)\theta_2^T \theta_2 f_2^T(x')\right) = V(x, x'), \quad x \neq x'.$$

Similar to (5.16)

$$v_2(x, \xi) = R(x, x) - R^T(x, \xi)\left(R(\xi) + W^{-1}\right)^{-1} R(x, \xi). \tag{5.18}$$

Thus, when the covariance function is known directly, i.e. we do not use (5.5) to get it, one can use (5.16) and (5.18) to construct optimal design. Moreover, the cases, in which $l = \dim \theta_2 \to \infty$ may be considered. The first attempt in this direction was done by Müller-Gronbach (1993).

6. Random coefficient regression models: Prediction

The presentation of the design problem for model (5.2)-(5.4) in the form (5.13) allows us to develop a rather simple technique for experimental design when the objective is the prediction of observed values. For the sake of simplicity, let $\eta(x, \theta) \equiv 0$ in (5.2) and

$$u_j(x_i) = u_{ij} = \theta_j^T f(x_i). \tag{6.1}$$

Then the corresponding optimal designs are defined as

$$\xi^* = \arg\min_\xi \Psi\left(M(\xi) + M_0\right), \tag{6.2}$$

where

$$M(\xi) = \int f(x) f^T(x) \xi(dx) \text{ and } M_0 = \sigma^2 N^{-1} \Lambda^{-1}.$$

It is expedient to note that

$$\min E\left[(\tilde\theta - \theta)(\tilde\theta - \theta)^T\right] = D(\xi) = \sigma^2 N^{-1}\left(M(\xi) + M_0\right)^{-1},$$

where the expectation operator E takes into account randomness of both observational errors and regression parameters. Minimization is taken with respect to all linear estimators, see Gladitz and Pilz (1982), Fedorov and Müller (1989), Pilz (1991) . The best linear estimator is

$$\hat{\theta}_j = N^{-1} \left(M(\xi) + M_0\right)^{-1} \sum_{i=1}^{N} y_{ij} f(x_i).$$

Similar to arguments in Section 5 we can apply the equivalence theorem to (6.2) to find, for instance, necessary and sufficient conditions for a design ξ^* to be optimal. Leaving to the reader the possibility to formulate them for the general case we focus only on three simple and very popular criteria.

Minimax and D-criterion. For the D-criterion, when

$$\Psi\left(M(\xi) + M_0\right) = -\ln|M(\xi) + M_0|, \tag{6.3}$$

one can easily derive from Theorem 1 that a necessary and sufficient condition for ξ^* to be optimal is that

$$f^T(x)\left(M(\xi^*) + M_0\right)^{-1} f(x) \leq \operatorname{tr}\left(M(\xi^*) + M_0\right)^{-1} M(\xi^*) \tag{6.4}$$

for all $x \in X$.

This inequality appears, especially when the dimension of f is large, more attractive and meaningful in notation described in comments accompanying (5.11):

$$R(x, x) - R^T(x, \xi)\left(W^{-1} + R(\xi)\right)^{-1} R(x, \xi)$$

$$\leq \operatorname{tr}\left(R(\xi) - R(\xi)\left(W^{-1} + R(\xi)\right)^{-1} R(\xi)\right). \tag{6.5}$$

The variance of the best linear unbiased predictor for u(x) equals

$$Var(u(x) - \hat{u}(x)|\xi),$$

$$= \sigma^2 + \sigma^2 N^{-1} \left[R(x,x) - R^T(x,\xi) \left(W^{-1} + R(\xi) \right)^{-1} R(x,\xi) \right], \qquad (6.6)$$

where

$$\hat{u}(x) = \hat{\theta}^T f(x) \; = \; R^t(x,\xi) \left(W^{-1} + R(\xi) \right)^{-1} \overline{Y} \qquad (6.7)$$

$$= \; V^T(x,\xi) \left(\sigma^2 W^{-1} N_{-1} + V(\xi) \right)^{-1} \overline{Y}, \qquad (6.8)$$

and the components of the vector $\overline{Y}$ are averages of observations at the corresponding points. For the continuously changing weights p_i the i-th component of $\overline{Y}$ may be considered as the observation made with a precision $\sigma / p_i N = \sigma_i^2$. If one introduce the covariance function

$$\tilde{V}(x, x\prime) = \begin{cases} V(x, x'), & x \neq x', \\ \sigma_i^2 + V(x, x) & \text{otherwise,} \end{cases} \qquad (6.9)$$

then (compare with (4.6) or coming later (7.1)) predictor $\hat{u}(x)$ coincides with the best linear unbiased predictor for $y(x) = u(x) + \varepsilon(x)$ everywhere except $x \in \text{supp}\xi$. At the design points x_i the realization(s) of $y(x)$ are measured directly and are not needed to be predicted, i.e. one may select $y(x) = \hat{y}(x)$ and $Var\left(y(x) - \hat{y}(x)|\xi \right) = 0$. Obviously

$$Var\left(y(x) - \hat{y}(x)|\xi \right) = \sigma^2 + Var\left(u(x) - \hat{u}(x)|\xi \right) \qquad (6.10)$$

otherwise.

Using Theorem 1 together with (6.5) and (6.10) we can formulate the analogue of the Kiefer-Wolfowitz equivalence theorem:

Theorem 3. *The following two design problems are equivalent:*

- $\min_\xi |D(\xi)|$,

- $\min_\xi \max_{x \in X} Var(y(x) - \hat{y}(x)|\xi)$.

There is one significant difference between this result and the original equivalence

theorem: an optimal design generally depends upon the number of observations N to be used.

Theorem 3 and formula (6.5) give another insight into numerical procedures from Sections 2 and 3: at every stage one has to relocate the design measure from the point(s) where $y(x)$ may be predicted easily (small $Var(y(x) - \hat{y}(x)|\xi_s)$) to the point(s), where the prediction is poor (large $Var(y(x) - \hat{y}(x)|\xi_s)$) .

Two linear criteria. Two objective functions which are very popular in spatial statistics are the weighted average variance of prediction:

$$Q_1(\xi) = E[\int_Z w^2(x) (y(x) - \hat{y}(x))^2 dx],$$

and the variance of the weighted average of prediction:

$$Q_2(\xi) = E\left[\int_Z w(x)y(x)dx - \int_Z w(x)\hat{y}(x)dx\right]^2,$$

where Z is the "prediction" set or the area of interest. Using (6.6) one can find that minimization of $Q_1(\xi)$ and $Q_2(\xi)$ is equivalent to minimization of

$$\Psi[M(\xi)] = \mathrm{tr}A(M(\xi) + M_0), \tag{6.11}$$

where in the first case

$$A = \int_Z w^2(x)f(x)f^T(x)dx,$$

and in the second one

$$A = aa^T, \quad a = \int_Z w(x)f^T(x)dx.$$

¿From part 2 of Theorem 1 it is easy to conclude that

Theorem 4. . *The design ξ^* is linear optimal if and only if*

$$\phi(x,\xi^*) \le \int_X \phi(x,\xi^*)\xi^*(dx), \quad \text{for all } x \in X, \tag{6.12}$$

where

$$\phi(x,\xi) = f^T(x)\left(M(\xi^*) + M_0\right)^{-1} A\left(M\xi^*\right) + M_0)^{-1} f(x).$$

Similar to the D-criterion we can show that for the average variance of prediction

$$\phi(x,\xi) = \phi_1(x,\xi) = \int_Z Cov^2(x,x'|\xi)w^2(x')dx', \tag{6.13}$$

while for the variance of the weighted average

$$\phi(x,\xi) = \phi_2(x,\xi) = w^2(x)\left(\int_Z Cov(x,x'|\xi)w(x')dx'\right)^2, \tag{6.14}$$

where

$$Cov(x,x'|\xi) = R(x,x') - R^T(x,\xi)\left(W^{-1} + R(\xi)\right)^{-1} R(x',\xi).$$

The counterparts of Theorem 3 and 4 may be formulated for optimal designs with bounded density. To do this function $\psi(x,\xi^*)$ in Theorem 2 should be replaced either with $Var\left(y(x) - \hat{y}(x)|\xi\right)$, or with $\phi_1(x,\xi)$, or with $\phi_2(x,\xi)$.

Remarks on applicability of the results. Let us note that the introduction of model (5.2) -(5.4) to generate correlated observations allows us to use the convex design theory for regression problems with correlated observations. Moreover, all results may be presented in a form which does not demand any direct knowledge of the functions $f(x)$. We can formulate results for a particular criterion using only information about the covariance function.

In this and in the previous section we have discussed only the properties of optimal designs. We hope, that having the sensitivity function $\psi(x,\xi)$ represented for various criteria in terms of the normalized covariance function $Cov(x,x'|\xi)$, the reader can easily construct numerical procedures similar to those discussed in sections 2 and 3.

7. Comparison with the methods based on the variance - covariance structure of observed random fields

Sacks-Ylvisaker approach. Let us suppose that in model (4.1) there is no trend,

i.e. $\eta(x, \theta) \equiv 0$, and the covariance function $V(x, x')$ is defined and known for all $x, x' \in X$. The objective of an experiment is to predict $y(x)$ at a given set of points Z, which can be either discrete or continuous.

The best linear unbiased predictor for $y(x)$ may be presented as follows (compare with (4.5 and (4.6)):

$$\hat{y}(x) = V^T(x, \xi_N) V^{-1}(\xi_N) Y_N, \qquad (7.1)$$

$$Var\left[(\hat{y}(x) - y(x))\right] = V(x, x) - V^T(x, \xi_N) V^{-1}(\xi_N) V(x, \xi_N). \qquad (7.2)$$

We again use notation $\xi_N = (x_1, \ldots, x_N)$ to emphasize that there is only one observation at every point x_i. Criteria $Q_1(\xi_N)$ and $Q_2(\xi_N)$ introduced in the previous section have been most intensively analyzed in the studies related to the design problem with correlated errors. Usually it has been assumed that $Z = X$.

A very good summary of the main results for the criterion $Q_1(\xi_N)$ may be found in Micchelli and Wahba (1981). The criterion $Q_2(\xi_N)$ was analyzed by Sacks and Ylvisaker (1970) and Ylvisaker (1987). Further references and comments may be found in Cambanis and Benhenni (1992), Cambanis (1985), Cambanis and Su (1993).

Noting (see (7.1)that

$$\int_X \hat{y}(x) dx = q^T(\xi_N) Y_N,$$

where

$$q^T(\xi_N) = \int_X V(x, \xi_N) V^{-1}(\xi_N) dx,$$

we can consider minimization of either $Q_1(\xi_N)$ or $Q_2(\xi_N)$ as a problem of finding an optimal basis for a quadrature formula in approximation theory, with a rather specific objective function; see Karlin and Studden (1966), Sacks and Ylvisaker (1970), Stroud (1975).

When $N \to \infty$, both $Q_1(\xi_N)$ and $Q_2(\xi_N)$ converge to zero for "smooth"

covariance functions $V(x, x')$ and for any atomless sequence ξ_N. As in Section 3 we may introduce the limit design measure that defines ξ_N. How a sequence ξ_N may be generated with a particular $\xi(dx)$ is discussed in details by Cambanis (1985) for $X \subset R^1$. For instance, the so called regular median sequence or design ξ_N is defined as

$$x_{Ni} = \arg \left(\int_a^{x_{Ni}} \xi(dx) = \frac{2i-1}{2N} \right), \ i = 1, 2, \ldots, n, X = [a, b]. \qquad (7.3)$$

When X is hypercube and $V(x, x')$ is separable with respect to all components of x, then design ξ_N may be defined as a direct product of univariate designs (see Ylvisaker (1975) for details). Thus the design problem is reduced to the search of the limiting measures providing the best convergence rate for the selected optimality criterion. The rather elaborate technique, a close sibling of the classical approximation theory, leads to a very special minimization problem. Introducing the design density $\xi(dx) = h(x)dx$ we may state this problem as follows

$$h^* = \arg \min_h \quad Q\left[B(h)\right], \qquad (7.4)$$

where

$$B_{\alpha\beta}(h) = \int_X \varphi_\alpha(x)\varphi_\beta(x)h^{-2k-2}(x)dx, \ \ \alpha, \beta = 1, \ldots, m, \xi(dx) = h(x)dx,$$

m is the number of estimated integrals (for instance, integrals of $Q_2(\xi_N)$-type with various weight functions), functions $\varphi_\alpha(x)$ and integer k are defined by a covariance function and by an optimality criterion.

Similar problems were considered in studies concerned with simultaneous calculation of m integrals by Monte-Carlo method; see Mikhailov and Zhigljavsky (1989), Zhigljavsky (1988), for details and further references. Actually, (7.4) being an optimization problem in a space of probability measures has many features in common with the standard design problem. Some interesting results including the analogue of the iterative numerical procedure from Section 2 are summarized

and discussed by Zhigljavsky (1988).

. Analytical solutions of (7.4) for the one-dimension case were proposed in the pioneering papers by Sacks and Ylvisaker (1966, 1968, 1970). Various generalizations may be found in Hajek and Kimeldorf (1976) and Wahba (1971, 1974). Following Cambanis (1985), the essence of those findings can be formulated as follows:

If there exist exactly k quadratic mean derivatives of the random process $y(x)$, then under certain regularity conditions (see details in the cited publications)

$$h^*(x) \sim \left[\alpha_k(x) w^2(x) \right]^{1/(2k+3)}, \tag{7.5}$$

where

$$\alpha_k(x) = V^{(k,k+1)}(x, x' - 0) - V^{(k,k+1)}(x, x' + 0)$$

and superscripts indicate the order of partial derivatives. For any design with density separated from zero the integral $Q(\xi_N)$ diminishes as $O(N^{-2k-2})$ and $h^*(x)$ minimizes

$$\lim_{N \to \infty} N^{2k+2} Q_2(\xi_N) \tag{7.6}$$

In fact, the immediate objective of Sacks and Ylvisaker (1966, 1968, 1970) was minimization of some function of the dispersion matrix of estimators of parameters θ describing a linear trend $\theta^T f(x)$ in model (4.1). They reduced the corresponding minimization problem to minimization of objective functions similar to $Q_2(\xi_N)$. For instance, when $\theta \in R^1$, then one has to minimize $Q_2(\xi_N)$ with a weight function which is a solution of the following integral equation:

$$f(x) = \int_X V(x, x') w(x') dx'.$$

Evidently, for stationary covariance functions $\alpha_k(x) \equiv$ constant. Subsequently the optimal limiting density $h^*(x)$ is completely defined by the weight function $w(x)$. In other words, only the behavior of $V(x, x')$ at its diagonal influences the solution! The Sacks-Ylvisaker approach (at least in its current form) cannot be

used in two cases of practical importance. First, it does not work for "infinitely" smooth covariance functions, when $\alpha_k(x) \equiv 0$ for any k. The covariance function

$$V(x, x') \sim exp\left[-(x - x')^2/2\right] \tag{7.7}$$

is a popular example (see Sacks and Ylvisaker (1966). The presence of the "white" noise (see model (5.1)) in observed variables gives another example, when the approach does not work. The latter case is of interest for many applications being a very reasonable model when a random process is observed with some instrumental error. In conclusion of this subsection let us note that the concept asymptotically optimal design ξ_N^* based on the existence of a continuous limit density $h^*(x)$ and assumption that

$$\lim_{N \to \infty} \max_i (x_{iN} - x_{(i-1)N}) = 0;$$

see Sacks and Ylvisaker (1966). The definition (7.3) of design ξ_N is one-dimension by its nature and that makes the approach difficult for spatial applications; see Ylvisaker (1975) for further details.

Random parameters approach. To understand the advantages and disadvantages of approach proposed in Sections 5 and 6 relative to to the Sacks-Ylvisaker approach let us introduce the following model:

$$y_{n\sigma}(x) = \sum_{\alpha=1}^{n} \theta_\alpha f_\alpha(x) + \sigma\varepsilon(x), \tag{7.8}$$

where $f_\alpha(x)$ are eigenfunctions of the covariance kernel $V(x, x')$

$$\lambda_\alpha f_\alpha(x) = \int_X V(x, x')f_\alpha(x')dx', \tag{7.9}$$

θ_α are random with zero means and diagonal covariance matrix, such that $Var(\theta_\alpha) = \lambda_\alpha$, $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_\alpha \geq \ldots$, $\varepsilon(x)$ is white noise with the variance 1, and σ^2 is a normalizing constant.

It is well known (Mercer's theorem; see for instance, Kanwal (1971)) that

under very mild conditions the series

$$V_n(x, x') = \sum_{\alpha=1}^{n} \lambda_\alpha f_\alpha(x) f_\alpha(x'), \quad x, x' \in X,$$

is uniformly and absolutely convergent and subsequently under very mild assumptions $\{\lambda_\alpha\}$ must diminish not slower then $O(1/\alpha)$. For many widely used stochastic processes or fields the rate is significantly faster; see Micchelli and Wahba (1981), Theorem 3.

Therefore, for sufficiently large n the kernel $V_n(x, x')$ may be a very reasonable approximation of $V(x, x')$. Allowing $\sigma \to 0$ we may hope that the process $y_{n,\sigma}(x)$ is "close" to $y(x)$ in the sense of their second moments. Subsequently, we might expect the closeness of the corresponding optimal designs. This probably holds for designs ξ_N with relatively small N. However, for $N \to \infty$ the diminishing σ does not guarantee closeness of optimal designs with $\sigma = 0$ and $\sigma > 0$. First, formally the Sacks-Ylvisaker approach does not work for any model with additive "white" noise, because it causes discontinuity of a covariance kernel at its diagonal:

$$E[y(x)y(x')] = \begin{cases} V(x, x'), & x \neq x', \\ V(x, x) + \sigma^2, & x = x'. \end{cases}$$

Secondly, for any $\sigma > 0$ and any ξ_N the rate of convergence for either $Q_1(\xi_N)$ or for $Q_2(\xi_N)$ will not be generally better than $O(N^{-1})$. This is slower than for any continuous covariance kernel.

Thus, for large N the Sacks-Ylvisaker approach and results from Sections 5 and 6 may lead to the different asymptotically optimal designs. If one believes that there is no instrument or any other observation error, then the Sacks-Ylvisaker approach leads to the better limit designs.

When the contribution of observation errors is significant then approximation (7.8) becomes very realistic and allows the use of methods from Sections 5 and 6, which usually produce optimal designs with very moderate numbers of supporting points. Usually these designs have about n supporting points.The existence

of well developed numerical procedures and software allows the construction of optimal designs for any reasonable covariance function $V(x, x')$ and various design regions X, including two and three dimension cases. Let us notice that the function $Cov(x, x'|\xi)$ used in Theorem 4 may be presented in the following form:

$$\sigma^2 N^{-1} Cov(x, x'|\xi) = V_n(x, x') - V_n^T(x, \xi) \left(\sigma^2 P + V_n(\xi)\right)^{-1} V_n(x, \xi), \quad (7.10)$$

$$P = \delta_{ij} r_i, \ r_i = p_i N.$$

This formula is convenient for some theoretical exercises. For more applied objectives and for development of numerical algorithms based on the iterative procedures from Sections 2 and 3, the direct use of eigenfunctions $f_\alpha(x)$ is more convenient.

Popular kernels. There are several show-case processes and design regions for which analytic expressions for the covariance kernel exist, and the corresponding eigenvalues and eigenfunctions are known:

For the Brownian motion the kernel is

$$V(x, x') = \min(x, x'), \ 0 \leq x, x' \leq 1,$$

and its eigenvalues and eigenfunctions are

$$\lambda_\alpha = (\alpha - 1/2)^{-2} \pi^{-2}, \ f_\alpha(x) = \sqrt{2} \sin(\alpha - 1/2)\pi x, \ \alpha = 1, 2, \ldots.$$

For the Brownian bridge,

$$V(x, x') = \min(x, x') [1 - \max(x, x')], \ 0 \leq x, x' \leq 1,$$

and

$$\lambda_\alpha = \alpha^{-2} \pi^{-2}, \ f_\alpha(x) = \sqrt{2} \sin \alpha \pi x, \ \alpha = 1, 2, \ldots.$$

Both kernels are not differentiable on the diagonal (see comments to (7.5)) and their "jump" functions $\alpha_k(x)$ are easy to calculate. These two kernels or some

simple functionals of them (compare with Wahba (1971)) are convenient candidates for the Sacks-Ylvisaker approach. For the Poisson kernel

$$V(x, x') = \frac{1 - \beta^2}{1 - 2\beta \cos 2\pi(x - x') + \beta^2}, \quad 0 \le x, x' \le 1, \quad 0 < \beta < 1,$$

and

$$\lambda_0 = 1, \quad \lambda_{2\alpha-1} = \lambda_{2\alpha} = \beta^\alpha,$$

$$f_0(x) = 1, \quad f_{2\alpha-1} = \sqrt{2} \cos 2\alpha\pi x, \quad f_{2\alpha} = \sqrt{2} \sin 2\alpha\pi x, \quad \alpha = 1, 2, \ldots.$$

The shape of the Poisson kernel may be controlled by the parameter β. It is "smooth" at the diagonal and the Sacks-Ylvisaker approach cannot be used.

For most real-world problems it is impossible to represent covariance kernels in a simple closed form. However, a representation in the form of an infinite series is standard. For instance, in many experiments related to either diffusion or heat conduction the covariance kernel may be expressed in the the two dimensional finite domain case (see Butkovskiy (1982)) as

$$V(x, x') = 4 \sum_{\alpha,\beta=1}^{\infty} \lambda_{\alpha\beta} \sin \alpha\pi x_1 \sin \beta\pi x_2 \sin \alpha\pi x_1' \sin \beta\pi x_2', \qquad (7.11)$$

where $X = \{0 \le x_1, x_2 \le 1\}$. Evidently,

$$f_{\alpha\beta}(x) = 2 \sin \alpha\pi x_1 \sin \beta\pi x_2, \quad \lambda_{\alpha\beta} = \exp\left[-a^2\pi^2(\alpha^2 + \beta^2)\right], \quad \alpha, \beta = 1, 2, \ldots,$$

and a is some constant. Representation (7.11) is very natural and convenient for the techniques considered in Sections 5 and 6. Note that the physical problems mentioned above lead to the Gaussian type kernels (compare with (7.7)) when X becomes infinite with respect to any or both coordinates.

The curious reader will find more covariance kernels in any serious book on integral equations containing a chapter on definite kernels or describing Green's functions (see, for instance, Kanwal (1971) and Butkovskiy (1982)).

Selecting the number of terms in (7.8) sufficiently large assures the closeness

of $V_n(x, x')$ and $V(x, x')$ may be assured. In many cases it is convenient to assume that coefficients $\{\theta_\alpha\}_1^\infty$ are normally distributed. This assumption does not help in the present problem and in fact can cause some theoretical difficulties. To avoid that we suppose that for any α distribution $P(\theta_\alpha)$ has a finite support set $[a_\alpha, b_\alpha]$ in R^1. For instance, we may select some simple symmetric distribution $P(\theta)$ defined on $[-a, a]$ with $E(\theta^2) = 1$ and use $P(\theta_\alpha \sqrt{\lambda_\alpha})$ as a distribution for θ_α. Selection of distributions with finite supporting sets assures not only closeness of an exact kernel and its approximation but proximity of $\sum_{\alpha=1}^\infty \theta_\alpha f_\alpha(x)$ and $\sum_{\alpha=1}^n \theta_\alpha f_\alpha(x)$ if the sequence $\{f_\alpha(x)\}_1^\infty$ satisfies some routine assumptions from the approximation theory. The basic idea of using model (7.8) is in deriving optimal designs for the approximate model $\sum_{\alpha=1}^n \theta_\alpha f_\alpha(x)$ and verifying the fact that these designs are optimal or close to optimal for $V_n(x, x')$ or $V(x, x')$. Furthermore, if the objective function is uniquely defined by a dispersion matrix of estimated parameters, then the constructed design is optimal for any model identical to the used one in terms of the first and second moments.

Using (7.8) with $\sigma = 0$ we may immediately conclude that minimization of $Q_2(\xi_N)$ is a rather standard problem from the approximate integration theory, see Davis and Rabinowitz (1985), Stroud (1975).

For instance, it is known (the Gauss-Jacobi Theorem) that for any polynomial $p(x)$ of degree $k \leq 2N - 1$ the exact equality

$$\int_a^b w(x)p(x)dx = \sum_{i=1}^N q_i p(x_i), \tag{7.12}$$

can be achieved the properly selected weights and supporting points. If $\{p_\alpha(x)\}_0^N$ are orthogonal polynomials with the weight function $w(x) > 0$, then $\xi_N = \{x_i\}_1^N$ are zeros of $p_N(x)$, and

$$q_i^{-1} = q_i^{-1}(\xi_N) = \sum_{\alpha=0}^{N-1} p_\alpha^2(x_i), \tag{7.13}$$

For example, let us consider the Brownian bridge kernel. Note that

$$- f_\alpha(x)dx = \frac{\sqrt{2}\sin\alpha\pi x}{\pi\sin\pi x}d\cos\pi x = \frac{\sqrt{2}}{\pi}U_{\alpha-1}(t)dt, \qquad (7.14)$$

where $t = \cos\pi x$ and $U_{\alpha-1}(t)$ is the second kind Tschebysheff polynomial. Now, with $x(t) = \pi^{-1}\arccos t$, we have for a proper $\xi_N = \{x_i\}_1^N$

$$\int_0^1 w(x)y_n(x)dx = \int_{-1}^1 w(x(t))\frac{y_n(x(t))}{\pi\sqrt{1-t^2}}dt = \sum_{i=1}^N q_i(\xi_N)\frac{y_n(x(t_i))}{\pi\sqrt{1-t_i^2}},$$

where $y_n(x) = \sum_{\alpha=1}^n \theta_\alpha f_\alpha(x)$, and it follows from (7.14) that $y_n(x(t))/\sqrt{1-t^2}$ is a polynomial of degree not higher than $n-1$. If $w(x(t)) = \sqrt{1-t^2}$, then ξ_N^* must coincide with zeros of $U_N(t)$, which are

$$x_i = x(t_i) = \frac{i}{N+1}, \quad i = 1,\ldots,N, \quad 2N \geq n;$$

compare with Müller-Gronbach (1993). Accordingly to (7.13) weights are $q_i(\xi_N) = \pi\sin^2\pi x_i/(N+1)$, and finally

$$\int_0^1 y_n(x)\sin\pi x\, dx = \frac{1}{N+1}\sum_{i=1}^N y_n(x_i)\sin\pi x_i.$$

The solution is extremely simple, but it could be more complicated for other weight functions, see Davis and Rabinowitz (1986). The value of $Q_2(\xi_N^*)$ is of order $O(\lambda_{2N})$. Various results about remainders in approximate integration theory may lead to the better estimates, but the corresponding technique is beyond the scope of this paper. Further details and related results may be found for instance, for the one dimension case in Davis and Rabinowitz (1986), Szegö (1959) and for the multi-dimension case in Stroud (1975). Similar exercises may be done for the criterion $Q_1(\xi_N)$ with $\sigma = 0$. The minimization of $Q_1(\xi_N)$ now becomes a problem from the function approximation theory. When $Z = X$ and $w(x) \equiv 1$,

then it follows from (7.1) that

$$Q_1(\xi_N) = E \int_X \left(y(x) - \hat{y}(x)\right)^2 dx = E \int_X \left(y(x) - \sum_{i=1}^{N} y(x_i)v_i(x)\right)^2 dx, \quad (7.15)$$

where $v^T(x) = V(x, \xi_N)V^{-1}(\xi_N)$.

It is known (see e.g. Micchelli and Wahba (1981)), that

$$Q(\xi_N) \geq E\left[\min_{\gamma} \int_X \left(y(x) - \sum_{\alpha=1}^{N} \gamma_\alpha f_\alpha(x)\right)^2 dx\right] = \sum_{\alpha=N+1}^{\infty} \lambda_\alpha. \quad (7.16)$$

This lower bound may be used to evaluate the efficiency of ξ_N and can be achieved for any singular kernel, i.e. when

$$V(x, x') = \sum_{\alpha=1}^{n} \lambda_\alpha f_\alpha(x) f_\alpha(x').$$

To verify the latter conjecture one has to select the design ξ_N coinciding with all zeros of $f_n(x)$; see some additional details in Fedorov and Hackl (1994). In cases when eigenfunctions cannot be found analytically the use of the remainder theory is probably one of the most reliable ways to construct satisfactory designs; see e.g. Davis and Rabinowitz (1984) or Achieser (1956). The ideas discussed in this subsection help to generate effective designs with very moderate number of observations N, obviously much less than we need using the Sacks-Ylvisaker approach based on the local approximation of $y(x)$. The author is not familiar with any studies where the connection between the classical approximation theory and the Sacks-Ylvisaker approach were analyzed systematically for models of type (7.8) with $n \to \infty$. Perhaps Micchelli and Wahba (1981) and Müller-Gronbach (1993) considered the closest ideas and models.

Again, we would like to note that in most cases measurement errors may contribute substantially to the randomness of observations. The rule of thumb in selection of the number of terms in (7.8) is that the least eigenvalue λ_N should be significantly less than σ.

8. Discrete case. Optimality criteria and the lower bounds

When the design region X and the set of interest Z are discrete and contain N_X and N_Z points correspondingly, then the covariance matrix of the vector $\{y(x_i) - \hat{y}(x_i)\}_1^{N_z}$, $x \in Z$, completely describes any objective function based on the second moments. We use the notation $D_Z(\xi_N)$, when the latter matrix consists of elements

$$D(\xi_N)_{ij} = V(x_i, x_j) - V^T(x_i, \xi_N)V^{-1}(\xi_N)V(x_j, \xi_N), x_i, x_j, \in Z.$$

The discrete versions of $Q_1(\xi_N)$ and $Q_2(\xi_N)$ are correspondingly

$$\mathrm{tr}D_Z(\xi_N) \quad \text{and} \quad L^T D_Z(\xi_N)L, \tag{8.1}$$

where $L^T = (1, \ldots, 1)$. We will introduce any weights as we did in the continuous case, to keep notations simple. In the discrete case we may introduce a very special version of D-optimality

$$Q_3(\xi_N) = \ln |D_Z(\xi_N)|. \tag{8.2}$$

It is assumed that there are no points in common for Z and $\mathrm{supp}\xi_N$. Otherwise the determinant equals zero, because

$$E\left[(\hat{y}(x_i) - y(x_i))^2\right] = 0,$$

when $x_i \in \mathrm{supp}\xi_N$.

Criterion (8.2) is very popular in the statistical literature related to the optimization of monitoring networks; see, for instance, Guttorp et al (1993), Carelton et al (1992), Schumacher and Zidek (1993), Shewry and Wynn (1987). In the cited papers the authors talk about either entropy or information. After the assumption of multivariate normality of the corresponding distributions is made, all approaches lead to various modifications of D-optimality; compare with Lindley

(1956).

In addition to (8.1) and (8.2) a number of other criteria were introduced for application in monitoring network improvement. A good collection of them can be found in Megreditchan (1979, 1989); see also Fedorov and Hackl (1994).

As soon as the criterion of optimality $Q(\xi_N)$ (we use this notation to emphasize that only the criteria of optimality related to the problem of interpolation or extrapolation are considered in this section) and the kernel $V(x, x')$ are defined, we have to find

$$\xi_N^* = \arg \min_{\xi_N} Q(\xi_N). \tag{8.3}$$

It is interestingly to note that optimization problem (8.3) was considered in a very different setting by Currin et al (1991) and by Morris et al (1993); see also Sacks et al (1989) for older references. They considered the Bayesian approach to design of computer experiments and introduced $Q(\xi_N)$ as a measure of discrepancy between a computer model and its approximation based on some prior knowledge expressed through the smoothness of the exact response. The latter was defined by a covariance function.

When N_Z is relatively "small" and N_X is not very "large" then exhaustive search may be a proper numerical procedure for a modern computer. With increase of N_Z and N_X one can use the exchange type algorithms discussed in Shewry and Wynn (1987) and in Fedorov and Hackl (1994), which are similar to those discussed in sections 2-4.

An alternative approach may be based on the introduction of a model similar to (7.8). For the sake of simplicity of notations, let $Z \subset X$, and let

$$\lambda_\alpha f_\alpha = V f_\alpha, \tag{8.4}$$

where $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{N_X}, V_{ij} = V(x_i, x_j)$ and $x_i, x_j \in X$. Then one may consider the following approximate model

$$Y \simeq Y_N = \sum_{\alpha=1}^{N} \theta_\alpha f_\alpha \tag{8.5}$$

where all vectors Y, Y_N and f_α have N_X components and

$$Y_i = y(x_i), \ x_i \in X, \ E(\theta_\alpha) = 0, \ E(\theta_\alpha^2) = \lambda_\alpha, \alpha = 1, \ldots, N.$$

Similar to (7.16)

$$Q_1(\xi_N) \geq \left[\min_\gamma (Y - F\gamma^T)^T (Y - F\gamma^T) \right]$$

$$= \left[(Y - Y_N)^T (Y - Y_N) \right] = \sum_{\alpha=N+1}^{N_X} \lambda_\alpha, \qquad (8.6)$$

where $F = (f_1, \ldots, f_N)$. There exists another result that can help to evaluate the closeness of Y_N and Y. Let

$$V_N = E(Y_n Y_n^T) = \sum_{\alpha=1}^{N} \lambda_\alpha f_\alpha f_\alpha^T.$$

Then (compare with Rao (1973), Ch. 8g)

$$V_N = \arg \min_A \|V - A\|,$$

where rank $A = N$ and symbol $\|B\|$ denotes the Frobenius norm of B defined by $(\mathrm{tr}B^2)^{1/2} = (\sum_{ij} B_{ij}^2)^{1/2}$. Moreover

$$\|V - V_N\|^2 = \sum_{\alpha=N+1}^{N_X} \lambda_\alpha^2. \qquad (8.7)$$

Thus, the vector Y_N is the best (maybe not unique) approximation of Y in the sense of two criteria (8.6) and (8.7). In fact, it is the best one for any strictly increasing function of $D = E\left[(Y - \tilde{Y}_N)(Y - \tilde{Y}_N)^T \right]$ which is invariant under orthogonal transformations, where $\tilde{Y}_N = BY$, rank $B \leq N$; see Seber (1984), Ch. 5.2. The vector

$$\hat{Y}_N = V(X, \xi_N) V^{-1}(\xi_N) Y,$$

where $V^T(X, \xi_N) = (V(x_1, \xi_N), \ldots, V(x_N, \xi_N))$, is one of the above linear es-

timates. Therefore (8.6) and (8.7) help to find the lower bounds for criteria depending upon

$$D(\xi_N) = V - V(X,\xi_N)V^{-1}(\xi_N)V(X,\xi_N) = E\left[(Y - \hat{Y}_N)(Y - \hat{Y}_N)^T\right].$$

Model (8.5) helps to understand some features of optimal designs and lead to some interesting numerical procedures (see next section). Adding the "white" noise, i.e. introducing the following model

$$Y_n = \sum_{\alpha=1}^{N} \theta_\alpha f_\alpha + \sigma\varepsilon, \tag{8.8}$$

where $E(\varepsilon) = 0$ and $E(\varepsilon\varepsilon^T) = I$, allows us to use all the tools discussed in Sections 2, 3, 5 to generate optimal designs.

9. Unknown covariance function

All the results discussed in the previous sections have essentially used the fact that either a covariance function $V(x,x')$ or a matrix Λ is known. That is possible but unfortunately uncommon in practice. In this section we explore two approaches to estimate the covariance structure.

Direct estimation of a covariance matrix. Let us start with a discrete design region X and assume there exist repeated observations at every point of X. Meteorological and environmental networks provide the most typical examples; see e.g. Megreditchan (1979, 1989) and Oehlert(1995a,b).

Let us define (compare with the previous section) the dispersion matrix of residuals $Y - \hat{Y}$ as

$$D(\xi_N) = \min_B E\left[(Y - BY(\xi_N))(Y - BY(\xi_N))^T\right], \tag{9.1}$$

where $\xi_N = (x_1,\ldots,x_N)$, $Y^T(\xi_N) = (y(x_1),\ldots,y(x_N))$, and B is an $N_X \times N$ matrix. For the sake of simplicity we assume that $E(Y) = 0$ and that this fact

is *a priori* known. Minimization is understood in the matrix ordering sense. A solution of (9.1) is

$$B^* = V(X, \xi_N) V^{-1}(\xi_N)$$

and

$$D(\xi_N) = V(X) - V(X, \xi_N) V^{-1}(\xi_N) V^T(X, \xi_N).$$

When sufficiently many observations are accumulated at every point of X the strong law of large numbers assures us that

$$\frac{1}{k} \sum_{j=1}^{k} (Y_j - BY_j(\xi_N)) (Y_j - BY_j(\xi_N))^T \simeq E\left[(Y - BY(\xi_N)) (Y - BY(\xi_N))^T\right] \tag{9.2}$$

and subsequently $\hat{B}^*$ and B^*, which minimize correspondingly the left and right hand-sides, are close to each other. Straightforward minimization gives

$$\hat{B}^* = \hat{V}(X, \xi_N) \hat{V}^{-1}(\xi_N), \tag{9.3}$$

where both matrices with caps are evident partitions of

$$\hat{V}(X) = k^{-1} \sum_{j=1}^{k} Y_j Y_j^T. \tag{9.4}$$

When there are missing observations, then it is better to use instead of (9.4) pairwise estimates

$$\hat{V}_{il}(X) = k_{il}^{-1} \sum_{j=1}^{k_{il}} Y_{ji} Y_{jl},$$

where k_{il} is the number of cases when the response variable was measured at x_i and x_l simultaneously.

Thus, (9.2) - (9.4) lead us to a very simple and widely used recipe: replace unknown parameters by their estimates and use methods developed for cases in which all parameters are known. Together with this simple recommendation (9.2) helps to generate other versions of numerical algorithms considered earlier. Let

us introduce matrix $I(\xi_N)$ with the following elements:

$$I_{il}(\xi_N) = \begin{cases} \delta_{il}, & \text{when } x_i \in \text{supp } \xi_N, \\ 0, & \text{otherwise.} \end{cases}$$

The left-hand side of (9.2) may be represented now as

$$\hat{D}(\xi_N, B) = k^{-1} \sum_{j=1}^{k} (Y_j - BI(\xi_N)Y_j)(Y_j - BI(\xi_N)Y_j)^T,$$

and the design problem may be viewed now as

$$\xi_N^* = \arg \min_{\xi_N} \min_{B} Q\left[\hat{D}(\xi_N, B)\right]. \tag{9.5}$$

¿From the numerical point of view (9.5) may be considered as a multi-dimension version of the best regression selection problem. Stepwise regression and best subset selection are the popular algorithms and can be easily adopted to solve (9.5). In fact the same methods may be used when the matrix $\hat{D}(\xi_N, B)$ is replaced by its true value; see comment in the conclusion of Section 8. Let

$$Q\left[\hat{D}(\xi_N, B)\right] = \hat{D}_{11}(\xi_N, B),$$

i.e. we want to minimize the variance of prediction at point x_1. In this case

$$\xi_N^* = \arg \min_{\xi_N} \min_{b} \sum_{j=1}^{k} \left(Y_{j1} - \sum_{i=2}^{N_X} b_i I_{1i}(\xi_N) Y_{ji}\right)^2, \tag{9.6}$$

and it is a very standard problem of selection of N predictors from $N_X - 1$ candidates and there exist a numerous number of the statistical packages which can be used to do that. The author is not familiar with multi-dimension versions of the corresponding software products, which are needed for more complicated criteria. The idea to use the least squares technique for selection of the most informative subset of sensors was probably initiated by Megreditchan (1979).

The search for an optimal design ξ_N may be viewed in this setting as a partitioning of X into two sets of given size $N_X - N$ and N. The latter must contain the most information about the whole set X; see, for instance, Shewry and Wynn (1987), who proposed using

$$\xi_N^* = \arg \min_{\xi_N} \ln |D_p(\xi_N)|, \tag{9.7}$$

where the subscript "p" indicates that the matrix contains only elements corresponding to the points (sites) with no observations. When (9.7) is replaced by its empirical version

$$\xi_N^* = \arg \min_{\xi_N} \min_{B} \ln \left| \hat{D}(\xi_N, B) \right|, \tag{9.8}$$

then the following simple and intuitively attractive exchange-type procedure may be used to construct ξ_N^*; see Fedorov and Hackl (1994):

(a) Given $\xi_{Ns} = \{x_{i_s}\}_1^N$ find

$$i^+ = \arg \max_i \min_b \sum_{j=1}^k \left(Y_{ji} - \sum_{i_s=1}^N b_{i_s} Y_{ji_s} \right)^2 .$$

Add the point x_{i^+} to the design: $\xi_{(N+1)s} = \xi_{Ns} + x_{i^+}$.

(b) Find

$$i^- = \arg \min_{i_s} \min_b \sum_{j=1}^k \left(Y_{ji_s} - \sum_{l \neq i_s} b_l Y_{jl} \right)^2 ,$$

where $x_l \in \xi_{(N+1)s}$ and delete the point x_{i^-} from the design, i.e. construct $\xi_{N(s+1)} = \xi_{(N+1)s} - x_{i^-}$. Return to (a).

Briefly, the exchange procedure (a), (b) may be spelled out in the following way: add to the design the worst explained sites and delete from it the best explained sites. Apparently, the approach may be called "model free": only existence of first two moments of observed Y is assumed. That may attract many practitioners. However in the search for an optimal network we are confined to sites where the measurements have been previously made. In other words, the

selection of the most informative subsets of sites (sensors, observing stations) may be discussed, but we cannot consider the problem of optimal extension.

Estimation of a parameterized covariance. In many practical cases the design region X is a continuous set and the covariance function has to be known everywhere at X. The most popular approach is based on the assumption that this function is homogeneous and isotropic, i.e.

$$V(x, x') = V(r), \quad r^2 = (x - x')^T (x - x')$$

with the subsequent parsimonious approximation of function $V(r)$; see e.g. Cressie (1991), Marshall and Mardia (1985), Matérn (1986), Ying (1995). The approach is frequently used in geostatistics, where a single realization of a random field is available, and in particular in the "kriging method" paradigm.

Methods from in Sections 5-8 are essentially based on approximation of the observed random fields by regression models with random coefficients. When prior to design of a network there exist some historical observations, then one may use the technique, which was developed for these models.It is expedient to note that accurate knowledge of Λ or Λ_n is useful but it is not as crucial as the knowledge of a covariance function in the Sacks-Ylvisaker approach. In fact, in basic optimization problems (5.13) and (6.2) the objective functions depend upon the sum $M(\xi) + M_0$, where M_0 is defined by Λ. For instance, in the case of (6.2)

$$M_0 = \sigma^2 N^{-1} \Lambda$$

and therefore the role of Λ diminishes when either $\sigma^2 \to 0$ or $N \to \infty$. Moreover, the simple dependence upon Λ allows to construct numerically optimal designs for different matrices Λ to learn about their sensitivity with respect to Λ.

In the simplest case, when the observational errors are negligible, the following estimators may be used:

$$\hat{\Lambda}_n = k^{-1} \sum_{j=1}^{k} \hat{\theta}_j \hat{\theta}_j^T, \tag{9.9}$$

$$\hat{\theta}_j = \arg \min_{\theta} \sum_{x_i \in X_j} \left(y_j(x_i) - \theta^T f(x_i) \right)^2 ,$$

where $\theta \in R^n$, X_j is the set of points with observations $y_j(x_i)$, and

$$\text{rank} \left(\sum_{x_i \in X_j} f(x_i) f^T(x_i) \right) = n$$

for all $j = 1, \ldots, k$. It is assumed that functions $f(x)$ are known and

$$V(x, x') \simeq V_n(x, x') = f^T(x) \Lambda_n f(x').$$

Subsequently,

$$\hat{V}(x, x') = f^T(x) \hat{\Lambda}_n f(x').$$

Actually, it is more convenient to use the matrix Λ_n directly than the function $\hat{V}(x, x')$ in all numerical procedures discussed in Sections 5, 6.

When the observational errors are comparable with the variations of θ, then (9.9) must be replaced with more sophisticated estimators, which are computationally much more demanding and complicate. Details and references may be found in Spjotvill (1977) and Fedorov et al. (1993).

10. Space and time

In most spatial experiments, after the sites are selected measurements are usually taken on some regular schedule, for instance, several times a day, or they are continuously recorded. Generally, the response function may depend upon time. Random errors can be correlated both in time and space. We consider only the simplest case, where there is no spatial correlation, following the ideas from Section 2. The generalization for more general models considered in Sections 3-6 is straightforward.

To adopt (2.1) for the time dependent response we assume that

$$\eta(x, t, \theta) = \theta^T f(x, t), \ 0 \leq t \leq T,$$

and

$$E\left[\varepsilon(x_i, t_j)\varepsilon(x_{i'}, t_{j'})\right] = \delta_{ii'}\rho_{jj'}(x).$$

When $\rho_{jj'} = \delta_{jj'}$, then the information matrix of observations made accordingly to the time schedule $\zeta(dt/x)$ may be presented in the following form

$$m(x) = \int_0^T f(x,t)f^T(x,t)\zeta(dt/x).$$

For measurements which are correlated in time,

$$m(x) = \sum_{jj'} f(x,t_j)R_{jj'}^{-1}(x)f^T(x,t_{j'}),$$

where, for the sake of simplicity, we assume that $\mathrm{supp}\zeta(dt/x)$ is a discrete set $t_1, t_2..., t_r$, and $R(x) = \rho_{jj'}(x)_1^r$.

When the measure $\zeta(dt/x)$ is fixed for each given x, then all the results from Section 2 may be used, with obvious replacement the function $\psi(x, \xi)$, which in the standard case has the form

$$\Phi(\xi) - f^T(x)A(\xi)f(x)$$

for all criteria satisfying assumptions (a) - (e), by the function

$$\psi(x, \xi) = \Phi(\xi) - \mathrm{tr}m(x)A(\xi).$$

For instance, for the D-criterion the sensitivity function $m - f^T(x)M^{-1}(\xi)f(x)$ must be replaced by $m - \mathrm{tr}m(x)M^{-1}(\xi)$. More details may be found in Atkinson and Fedorov (1988), Fedorov and Nachtsheim (1995), and Spruil and Studden (1979). Formally the time dependent observation may be treated as a vector-observation case (see, for instance, Fedorov (1972), Ch. 5).

Evidently, introducing the time variable does not change the basic theory, but makes all techniques, including computing of optimal designs, more time and effort consuming. However there exist models and optimality criteria for which

optimal designs are the same both for the static and for the time dependent cases. For instance, the latter is true for models with uncorrelated observations and with separable variables, when

$$f(x,t) = f_1(x) \otimes f_2(t)$$

or where

$$\eta(x,\theta) = \theta_0 + \theta_1^T f_1(x) + \theta_2^T f_2(t), \quad \theta^T = (\theta_0, \theta_1^T, \theta_2^T),$$

and the selected criterion satisfies assumptions $(a) - (e)$ from Section 2; see Cook and Thibodeau (1980), Hoel (1965), Huang and Hsu (1993), Schwabe (1994, 1995).

When time is included explicitly in model, then the concept of sensor allocation can be extended and "mobile" sensors may be introduced. In this case design consists of trajectories $x_i(t) \in X, 0 \leq t \leq T$. The topic is beyond the scope of this survey. A reader can find the results and references in Chang (1979), Fedorov and Nachtsheim (1995), Titterington (1980) and Zarrop (1979).

Acknowledgement

11. References

[1] Achieser, N. I. (1956). *Theory of Approximation*, Frederick Ungar, New York.

[2] Atkinson, A. C. and V. V. Fedorov (1988). Optimum Design of Experiments. In Kotz, S. and Johnson, N.I. *Encyclopedia of Statistics, Supplemental Volume.* Wiley, New York pp. 107-114.

[3] Atkinson, A.C. and A. N. Donev (1992). *Optimum Experimental Design.* Clarendon Press, Oxford.

[4] Benhenni, K. and S. Cambanis (1992). Sampling Designs for estimating Integrals of Stochastic Process. *Ann. Statist.*, 20, pp. 161-196.

[5] Bickel, P. J. and A. M. Herzberg (1979). Robustness of Design Against Autocorrelation in Time, I, *Ann. Statist.* 7, 77-95.

[6] Bishoff, W. (1992). On Exact D-optimal Designs for Models with Correlated Observations, *Ann. Inst. Statist. Math.* 44, 229-238.

[7] Bishoff, W. (1993). On D-optimal designs for linear models under correlated observations with applications to a linear model with multiple response. *JSPI.* 37, 69-80.

[8] Brimkulov, U., Krug G. and V. Savanov (1986). *Design of Experiments for Random Fields and Processes.* Nauka, Moscow.

[9] Butkovskiy, A. G. (1982). *Green's Functions and Transfer Functions Handbook,* Wiley, New York.

[10] Cambanis, S. (1985). Sampling Designs for Time Series. In *Handbook of Statistics.* Vol. 5, North-Holland, New York, pp. 337-362.

[11] Cambanis, S. and Y. Su (1993). Sampling Design for Estimation of a Random Process. *Stochastic Process Appl.* 46, 47-89.

[12] Chang D. (1979). Design of Optimal Control for a Regression Problem. *Ann. Statist.* 7, 1078-1085.

[13] Cook, R. D. and L. A. Thibodau (1980). Marginally Restricted D-optimal Designs. *JASA.* 75, 366-371.

[14] Cook, R.D. and C. J. Nachtsheim (1980). A Comparison of Algorithms for Constructing Exact D-optimal Designs. *Technometrics,* 24, 315-324.

[15] Cook, R.D. and V. V. Fedorov (1995). Constrained Optimization of Experimental Design. *Statistics.* <u>26</u>, 129-178.

[16] Cressie, N.A. (1991). *Statistics for Spatial Data.* Wiley, New York.

[17] Currin, C., Mitchell, T., Morris, M., and D. Ylvisaker (1991). Bayesian Prediction of Deterministic Functions, With Applications to the Design and Analysis of Computer Experiments. *JASA.* <u>31</u>, 953-963.

[18] Davis, P. J. and Rabinowitz, Ph. (1984). *Methods of Numerical Integration*, 2nd Edition. Academic Press, New York.

[19] Eaton, J.L., Giovagnoli A., and P. Sebastiani (1994). TR 598. School of Statistics, University of Minnesota.

[20] Ermakov, S.M. (Ed.) (1983). *Mathematical Theory of Experimental Design*, Nauka, Moscow.

[21] Ermakov, S.M. (1989). Random Interpolation in the Theory of Experimental Design, *Computational Statistics and Data Analysis*, <u>8</u>, 75-80.

[22] Fedorov, V.V. (1972). *Theory of Optimal Experiments.* Academic Press, New York.

[23] Fedorov, V.V. (1989). Optimum Design of Experiments with Bounded Density: Optimization Algorithms of the Exchange Type. *JSPI.* <u>24</u>, 1-13.

[24] Fedorov, V.V. and A.B. Uspensky (1975). *Numerical Aspects of Design and Analysis of Experiments.* Moscow State University Press, Moscow.

[25] Fedorov, V.V., Leonov, S.L. and S. A. Pitovranov (1988). Experimental Design Technique in the Optimization of a Monitoring Network. In Fedorov V.V. and Lauter M. (Eds). *Model-Oriented Data Analysis.* Springer-Verlag, New York, pp. 165-175

[26] Fedorov, V.V. and W. Müller (1989). Comparison of Two Approaches. In the Optimal Design of an Observation Network. *Statistics.* <u>19</u>, 339-351.

[27] Fedorov, V.V. and W. Müller (1989). Design of an Air-Pollution Monitoring Network. An Application of Experimental Design Theory. *Österreichische Zeitschrift für Statistik und Informatik.* 19, 5-17.

[28] Fedorov, V.V., Hackl, P. and W. Müller (1993), Estimation and Experimental Design for Second Kind Regression Models. *Informatik, Biometrie und Epidemiologie.* 24, 134-151.

[29] Fedorov, V.V. and P. Hackl (1994). Optimal Experimental Design: Spatial Sampling. *Calcutta Statistical Association Bulletin.*

[30] Fedorov, V.V. and C. J. Nachtsheim (1995). Optimal Designs for Time - Dependent Responses. In Kitsos L. P. and W. G. Mueller (Eds). *Advanced in Model-Oriented Data Analysis.* Springer-Verlag, New York. pp. 3-12. 44, 57-81.

[31] Fisher, R.A. (1947). *The Design of Experiments,* 4th Edition, Hafner-Publishing Co. Inc. New York.

[32] Gribik, P.R., Kortanek, K.O. and J. R. Sweigart (1976). Design a Regional Air Pollution Monitor Network: An Appraisal of a Regression Experimental Approach. In *"Proceedings of the Conference on Environmental Modeling and Simulation".* EPA 600/9-76-016. Research Triangle Park, North Carolina, pp.86-91.

[33] Guttorp, P., Le, N.D. Sampson, P.D. and J. V. Zidek (1993). Using Entropy in the Redesign of Environmental Monitoring Network. In Patil, G.P. and Rao, C.R. (Eds). *Multivariate Environmental Statistics.* Elsevier Science Publisher, New York, pp. 175-202.

[34] Hoel, P.G. (1965). Minimax designs in two dimensional regression. *Ann. Math. Stat.,* 29, 1134-1145.

[35] Huang, M.L. and M. C. Hsu, (1993). Marginally Restricted Linear-Optimal Designs. *JASA,* 35, 251-266.

[36] Kanwal R. P.(1971). *Linear Integral Equations, Theory and Technique.* Academic Press, New York.

[37] Karlin, S. and W. Studden (1966). *Tchebysheff Systems: with Applications in Analysis and Statistics.* Wiley, New York.

[38] Kiefer, J. (1959). Optimal Experimental Design. *JRSS (B)* , 21. 272-319.

[39] Kiefer, J. (1974). General Equivalence Theory for Optimum Designs (Approximate Theory). *Ann. Statist..* 2, 849-879.

[40] Kunert, J. (1988). Optimal Designs for Correlation in the Plane. In Dodge, Fedorov and Wynn (Eds). *Optimal Design and Analysis of Experiments.* North-Holland, New York. pp. 123-131.

[41] Lindley, D. (1956). On a Measure of Information provided by an experiment. *Ann. Math. Stat.* 27, 986-996.

[42] Marshall, R.J. and K. V. Mardia (1985). Minimum Norm Quadratic Estimation of Components of Spatial Covariance. *Mathematical Geology.* 17, 517-525.

[43] Martin, R.J. (1986). On the Design of Experiments under Spatial Correlation. *Biometrika.* 73, 247-277.

[44] Matèrn, B. (1986). *Spatial Variation, Lecture Notes in Statistics.* Springer-Verlag, Berlin.

[45] Megreditchan, G. (1979). L'optimization des reseaux d'observation des champs meterologiques. *La Meterologie.* 6, 51-66.

[46] Megreditchan, G. (1989). Statistical Redundancy as a Criterion for Meterological Network Optimization. *Ö sterreichische Zeitschrift für Statistik und Informatik.* 19, 18-29.

[47] Micchelli, C.A. and G. Wahba (1981). Design problems for optimal surface interpolation. In Ziegler Z. (Ed.). *Approximation Theory and Applications.* Academic Press New York. pp. 329-248.

[48] Mikhailov, G.A. and A. A. Zhigljavskii (1989). Uniform Optimization of Weighted Estimates of the Monte Carlo Method. *Soviet Math. Dokl.* 38, 523-526.

[49] Mitchell, T.J. (1974). (a) An Algorithm for Construction of D-optimal Experimental Design. (b) Computer construction of D-optimal "First-Order" Designs. *Technometrics.* 16, (a)203-210, (b)211-220.

[50] Morris M.D., Mitchell T.B. and D. Ylvisaker (1993). Bayesian Design and Analysis of Computer Experiments: Use of Derivatives in Surface Prediction. *Technometrics.* 35, 243-255.

[51] Müller W. and A. Pázman (1995). Design Measures and Extended Information Matrices for Optimal Designs when the Observations are Correlated, TR No.47, Dept. of Statistics, Univ. of Economics Vienna.

[52] Müller - Gronbach Th. (1993). Optimal Designs for Approximating the Path of Stochastic Process. Preprint Nr.. A-93-14. Freie Universität, Berlin.

[53] Nachtsheim, C.J. (1987). Tools for Computer-Aided Design of Experiments. *Journal of Quality Control.* 19, 132-160.

[54] Näther, W. (1985). *Effective Observation of Random Fields.* Teubner Texte Zur Mathematik-Band 72. Teubner Verlag, Leipzig.

[55] Nguen. N.K. and A. J. Miller (1992). A review of some exchange algorithms for constructing discrete D-optimal designs. *Comp. Statist. and Data Analysis.* 14, 489-498.

[56] Oehlert, G.W. (1995). The Ability of Wet Deposition Network to Detect Temporal Trends. *Environmetrics.* 6, 327-339.

[57] Oehlert, G.W. (1995). Shrinking a Wet Deposition Network. To appear: *Atmospheric Environment.*

[58] A. Pázman (1986). *Foundations of Optimum Experimental Design.* Reidel, Dordrecht.

[59] Pilz, T. (1991). *Bayesian Estimation and Experimental Design in Linear Regression Models.* Wiley, New York.

[60] Pukelsheim F. (1993). *Optimal Design of Experiments.* Wiley, New York.

[61] Rabinowitz, N. and D. M. Steinberg (1990). Optimal Configuration of a Seismographic Network: A Statistical Approach. *Bulletin of the Seismological Society of America.* 80, 187-196.

[62] Rao, C.R. (1973). *Linear Statistical Inference and its Applications*, 2nd Edition. Wiley, New York.

[63] Ripley, B.D. (1981). *Spatial Statistics*, Wiley, New York.

[64] Sacks, J., Welch, W.J., Mitchell, T.J., and H.P. Wynn (1989). Design and Analysis for Computer Experiments. *Statistical Science.* 4, 409-423.

[65] Sacks, J. and D. Ylvisaker D. (1966). Designs for regression problems with correlated errors. *Ann. Math. Statist.* 37, 66-89.

[66] Sacks, J. and D. Ylvisaker (1968). Designs for regression problems with correlated errors, many parameters.*Ann. Math. Statist.* 39, 49-69.

[67] Sacks, J. and D. Ylvisaker (1970). Designs for regression problems with correlated errors, III. *Ann. Math. Statist.* 41, 2057-2074.

[68] Sacks, J. and D. Ylvisaker (1970). Statistical designs and integral approximation. *Proc. 12tgh Bieu. Sem. Canad. Math. Cong.* pp.115-136. Canad. Math Cong., Montreal.

[69] *SAS/QC Software: Design of Experiments Tools.* (1995). SAS Institute, Inc.

[70] Seber G.A.F. (1977) *Linear Regression Analysis*, Wiley, New York.

[71] Seber G.A.F. (1984) *Multivariate Observations*, Wiley, New York.

[72] Shewry, M.C. and H. P. Wynn (1987). Maximum entropy sampling. *Journal of Applied Statistics.* 14, pp.165-170.

[73] Schumacher, P. and J. V. Zidek (1993). Using Prior Information in Designing Intervention Detection Experience. *Ann. Statist.* 21, 447–463.

[74] Schwabe, R. (1994). Optimal Designs for Additive Linear Models, Preprint No. 1-5-94, Freie Universität Berlin, Fachbereich Mathematik.

[75] Schwabe, R. (1995). Designing Experiments for Additive Nonlinear Models. In Kitsos, C.P. and W.G. Mueller. *Advances in Model-Oriented Data Analysis.* Springer-Verlag, New York. pp. 77-85.

[76] Silvey, S.D. (1980). *Optimal Design.* Chapman and Hall, London, New York.

[77] Spjotvill, E. (1977). Random Coefficient Regression Models - A Review. *Statistics.* 8, 69-93.

[78] Spruill, M.C. and W. J. Studden (1979). A Kiefer-Wolfowitz Theorem in Stochastic Process Setting. *Ann. Statist.* 7. 1329-1332.

[79] Stroud, A.H. (1975). *Approximate Calculation of Multiple Integrals.* Prentice-Hall, Inc., Englewood Cliffs, New Jersey.

[80] Szegö, G. (1959). *Ortogonal Polynomials.* Amer. Math. Soc., New York.

[81] Titterington, D.M. (1980). Aspects of Optimal Design in Dynamic Systems. *Technometrics.* 22, 287-300.

[82] Wahba, G. (1971). On the Regression Design Problem of Sacks and Ylvisaker, *Ann. Mtah. Statist.* 42, 1035-1053.

[83] Wahba, G. (1974). Regression Design for Some Equivalence Classes of Kernels. *Ann. Statist.* 2, 925-934.

[84] Wheeler, B. (1994). *ECHIP: Version 6.0 for Windows*. ECHIP, Inc., Hockessin.

[85] Wynn, H. (1982). Optimum Submeasures with Applications to Finite Population Sampling. In *Statistical Decision Theory and Related Topics III, Vol. 2*. Academic Press, New York. pp. 485-495.

[86] Ying, Z. (1993). Maximum Likelihood Estimation of Parameters under a Spatial Sampling Scheme. *Ann. Statist.* 21. 1567-1590.

[87] Ylvisaker, D. (1975). Design on Random Fields. In *A Survey of Statistical Design and Linear Models*. North-Holland, New York.

[88] Ylvisaker, D. (1987). Prediction and Design. *Ann. Statist* 15, 1-18.

[89] Ylvisaker, D. (1988). Bayesian Interpolation Schemes. In *Optimal Design and Analysis of Experiments*. Dodge V., Fedorov V. V., and Wynn, M. (Eds). North-Holland, New York. pp. 169-278.

[90] Zarrop, M.B. (1979). *Optimal Experimental Design for Dynamic System Identification*. Springer-Verlag, New York.

[91] Zhigljavsky, A.A. (1988). Optimal Designs for Estimating Several Integrals, in Dodge, Y., Fedorov, V. V. and Wynn, H.P. (Eds). *Optimal Design and Analysis of Experiments*. North-Holland, New York.

[92] Zimmerman, D.L. and M.B. Zimmerman (1991). A comparison of Spatial Semivariagram Estimators and Corresponding Ordinary Kriging Predictors. *Technometrics*, 33, 77-91.

ORNL/TM-13152

INTERNAL DISTRIBUTION

1.	D. J. Alexander	25.	G. R. Moline
2.	C. K. Bayne	26.	M. D. Morris
3.	W. R. Corwin	27.	R. K. Nanstad
4.	D. F. Craig	28.	C. E. Oliver
5.	T. S. Darland	29.	G. Ostrouchov
6.	C. S. Daw	30.	R. L. Schmoyer
7.	D. J. Downing	31.	R. F. Sincovec
8-12.	V. V. Fedorov	32.	P. T. Singley
13.	L. J. Gray	33.	D. A. Wolf
14.	H. W. Hayden, Jr.	34.	K-25 Applied Tech. Library
15.	S. K. Iskander	35.	Y-12 Technical Library
16.	P. Kanciruk	36.	Laboratory Records - RC
17-21.	M. R. Leuze	37-38.	Laboratory Records Dept.
22.	D. E. McCabe	39.	Central Research Library
23.	R. C. Mann	40.	M & C Records Office
24.	J. G. Merkle	41.	ORNL Patent Office

EXTERNAL DISTRIBUTION

42. Prof. A. C. Atkinson, London School of Economics, London WC2A 2AE, United Kingdom

43. Dr. Richard Beckman, Statistics Group A1, Los Alamos National Laboratory, MS F600, Los Alamos, NM 87545

44. C. Bolton, Nuclear Electric, Berkeley Nuclear Laboratories, Berkeley, Gloucestershire, GL139PB, United Kingdom

45. Prof. R. D. Cook, School of Statistics, University of Minnesota, 1994 Bufford Avenue, St. Paul, MN 55108-6042

46. Prof. D. F. Cox, Nuffield College, Oxford OX1 1NF, United Kingdom

47. Prof. Dennis Cox, Department of Statistics, Rice University, Houston, TX 77251-1892

48. E. Eason, Modeling and Computing Services, Suite 105, 6400 Lookout Road, Boulder, CO 80301

49. Prof. Sherwood Ebey, Department of Mathematics, University of the South, Sewanee, TN 37375

50. Prof. P. Hackl, Department of Statistics, Vienna University of Economics, Augasse 2-6, A-1090 WIEN, Austria

51. Dr. David Hall, Statistics, Systems Department, Pacific Northwest Laboratory, P.O. Box 999, Richland, WA 99352

52. Prof. J. A. Hartigan, Department of Statistics, Yale University, Box 2179 - Yale Station, New Haven, CT 06520

53. Dr. Dan Hitchcock, Office of Scientific Computing, ER-7, Applied Mathematical Sciences, Office of Energy Research, U. S. Department of Energy, Washington, DC 20585

54. Dr. Fred Howes, Office of Scientific COmputing, ER-7, Applied Mathematical Sciences, Office of Energy Research, U. S. Department of Energy, Washington, DC 20585

55. Prof. J. Stuart Hunter, 503 Lake Drive, Princeton, N.J. 08540

56. Prof. Mark Johnson, Department of Statistics, University of Central Flordia, Orlando, FL 32816-0370

57. Dr. A. M. Liebetrau, Computational Sciences Department, Battelle-Northwest, P. O. Box 999, Richland, WA 99352

58. Dr. J. Lijengren, Pacific Northwest Laboratories, P. O. Box 999, Richland, WA 99352

59. Dr. Michael McKay, Statistics Group Al, Los Alamos National Laboratory, MS F600, Los Alamos, NM 87545

60. Prof. Lisa Moore, Institute of Statistics and Decision Sciences, Duke University, Durham, NC 27706

61. Dr. David Nelson, Director of Scientific Computing, ER-7, Applied Mathematical Sciences, Office of Energy Research, U. S. Department of Energy, Washington, DC 20585

62. G. R. Odette, Department of Chemical and Nuclear Engineering, Ward Memorial Drive, University of California, Santa Barbara, A 93106

63. Dr. David J. Pack, CSR Institute, 8889 Bourgade AV, Lenaxa, KS 66219

64. Mr. Brent Pulsipher, Computational Sciences Department, Battelle Northwest, P. O. Box 999, K1-86, Richland, WA 99352

65. Dr. Jerome Sacks, NISS, P. O. Box 14162, Research Triangle Park, NC 27709-4162

66. Prof. A. F. Smith, Department of Mathematics, University of Nottingham, University Park, Nottingham NG7 2RD, England

67. Dr. Alan Solomon, P. O. Box 227, Omer 84965, Israel

68. Dr. Daniel L. Solomon, Department of Statistics, North Carolina State University, P. O. Box 5457, Raleigh, NC 27650

69. Prof. S. M. Stigler, Department of Statistics, University of Chicaago, 5734 University Avenue, Chicago, IL 60637-1546

70. A. Taboada, Mail Stop T0-E10, Division of Engineering, U.S. Nuclear Regulatory Commission, Washington, DC 20555

71. M. G. Vassilaros, MS T10-E10, Division of Engineering, U. S. Nuclear Regulatory Commission, Washington, DC 20555

72. Prof. William Welch, Department of Statistics and Acturial Science, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada

73. Prof. Kee Wong Weng, Biostatistics, UCLA School of Public Health, 10833 Le Conte Ave., Los Angeles, CA 90024-1722

74. Jeff Wu, Department of Statistics, Mason Hall, University of Michigan, Ann Arbor, MI 48109

75. Prof. Y. C. Yao, Department of Statistics, Colorado State University, Fort Collins, CO 80523

76. Prof. Don Ylvisaker, Department of Mathematics, University of California, Los Angeles, CA 90024

77. Prof. J. V. Zidek, Department of Statistics, University of British Columbia, 2021, West Mall, Vancouver, BC V6T 1Z2 Canada

78. Office of Assistant Manager for Energy Research and Development, Department of Energy, Oak Ridge Operations Office, P. O. Box 2001, Oak Ridge, TN 37831-8600

79-80. Office of Scientific and Technical Information, P. O. Box 62, Oak Ridge, TN 37830