Optimal Experimental Design: Spatial Sampling

Valerii V. Fedorov†
University of Minnesota.
Peter Hackl
Department of Statistics, University of Economics, Vienna.

†Currently at Computer Science and Mathematics Division, Oak Ridge National Laboratory, U.S.A..

Published in:

Calcutta Statistical Association Bulletin
Vol. 44, March & June, 1994, Nos. 173-174

Keywords: optimal experimental design, surface interpolation, optimal sampling

1 Introduction

This survey is mainly devoted to the problems related to the evaluation, estimation, and approximation of functions (or some functionals of them) which can be observed (measured) at a finite number of supporting points located on some (usually two-dimensional) surface.

Two cases can be distinguished. In the first one a function can be observed only once at every supporting point. Statistical geology provides the most typical examples for this case. In the second case observations (or measurements) can be done during some time interval. This type of experiment is mainly found in meteorology and environmetrics.

Both cases are considered in the paper with the main emphasis on the problem of optimal location of the supporting points (i.e., the points where observations have to be made). Inspite of the long history of the development of this problem the available results are rather sparse and are not systemized.

Nevertheless, it has to be pointed out that the following publications contributed a lot to the clarification of the theory: Cambanis (1985), Dalenius et al. (1961), Gribik et al. (1976), Katkovnik (1985), Matérn (1986), Ripley (1983), Sacks and Ylvisaker (1966), Thiebaux and Peddler (1987), Ylvisaker (1987). Among more applied work the following can be listed: Gandin et al. (1968), Kumar and Seinfeld (1978), Megreditchan (1979, 1985), Munn (1981). Of course, the choice of the cited publications is a subjective and small sample. In the subsequent sections the reader will find more extensive references.

2 Covariance Analysis Approach

2.1 Estimation and Optimality Criteria

Everywhere in this paper it is assumed that an observed surface (function, field) is a realization of some random process $y(x,t)$, where $x \in X$, $t \in T$, and X and T are compact. Measurement (observation, sampling) provides realizations y_{ij} at the points x_i and the times t_j, $x_i \in X$, $t_j \in T$.

Usually, x stands for the location of sensors or observation stations, and t stands for time. More generally, x corresponds to the controlled variables which can be chosen by the experimenter. Variables t comprise all those that cannot be controlled. E.g., in environmental network optimization problems t may represent time, wind direction, temperature, etc.

Let us start with the case when prior information (or model) is described by the mean trend $\mathrm{E}\{y(x,t)\} = m(x,t)$, and the covariance function $\mathrm{Cov}\{y(x,t), y(x',t)\} = C(x,x',t)$. It is assumed that $y(x,t)$ and $y(x,t')$ are independent from $t \neq t'$. If not stated beforehand we further assume that $m(x,t) = 0$ and $C(x,x',t) = C(x,x')$.

Given observations $y_l = (y_1, \ldots, y_l)^T$ with $y_i = y(x_i)$, $i = 1, \ldots, l$, we want to estimate the vector $y_p = (y_{l+1}, \ldots, y_n)^T$ with $y_i = y(x_i)$ for $i = l+1, \ldots, n$, where $n = l+p$. The best linear unbiased estimator is well-known and can be presented as follows [see, e.g., Ripley (1981, chp. 4)]:

$$\hat{y}_p = C_{pl} C_{ll}^{-1} y_l \tag{1}$$

and

$$D_{pp} = \mathrm{Cov}\{\hat{y}_p - y_p\} = C_{pp} - C_{pl} C_{ll}^{-1} C_{lp} \tag{2}$$

where $C_{pp} = \mathrm{Cov}\{y_p\}$, $C_{ll} = \mathrm{Cov}\{y_l\}$, and $C_{pl} = \mathrm{Cov}\{y_l, y_p\}$. It is clear that $D_{ll} = \mathrm{Cov}\{\hat{y}_l\} = 0$.

If in (2) all matrices are known then, similarly to the traditional experimental design theory, some monotonous function $Q(D_{pp})$ can be used to compare the efficiency of various locations of $x_1, \ldots, x_l$. For instance, if we want to estimate the average $\overline{y}_p = p^{-1} \sum_{i=l+1}^{n} y_i$, a suitable choice is

$$Q(D_{pp}) = \mathrm{Cov}\{\hat{\overline{y}}_p - \overline{y}_p\} = p^{-2} L^T D_{pp} L$$

with $L^T = (1, \ldots, 1)$. Criteria $\mathrm{tr}\, D_{pp}$, $\max_\alpha D_{pp,\alpha\alpha}$ and $\ln|D_{pp}|$ are discussed, e.g., by Micchelli and Wahba (1981), Ylvisaker (1987), and Shewry and Wynn (1987).

2.2 Most informative subsets

Let the n points $X_n = \{x_1, \ldots, x_n\}$ be fixed. Let us further assume that it is planned to make observations exactly at l points $X_l = \{x_1, \ldots, x_l\} \subset X_n$. We have the problem to choose a subset $X_l^* \subset X_n$ that provides the "best evaluation" of $y(x)$ over all $x \in X_n$. Similarly to the traditional experimental design theory, X_l will be called the "design (of experiment)".

It is clear that the estimator $\hat{y}$ given in (1) fulfils $\hat{y}(x) - y(x) = 0$ if $x \in X_l$. Therefore, $Q(D_{pp}) = Q(D_{nn})$, where D_{nn} corresponds to the whole set X_n. Consequently, the problem can be formulated in two equivalent forms:

$$X_l^* = \arg \min_{X_l \subset X_n} Q(D_{nn}),$$

or

$$X_l^* = \arg \min_{X_l \subset X_n} Q(D_{pp}), \quad X_p = X_n \setminus X_l. \tag{3}$$

Various cases of the optimization problem (3) were considered by Fedorov and Müller (1988), Megreditchan (1979, 1985), and Shewry and Wynn (1987).

It is convenient to illustrate the main features of (3) by considering the D-criterion

$$X^* = \arg \min_{X_l \subset X_n} |D_{pp}|. \tag{4}$$

Using

$$\begin{vmatrix} A & B \\ B^T & C \end{vmatrix} = |A| \cdot |C - BA^{-1}B^T| = |C| \cdot |A - B^T C^{-1} B|$$

we receive the identity

$$|C| = |C_{ll}||C_{pp} - C_{pl}C_{ll}^{-1}C_{lp}| = |C_{pp}||C_{ll} - C_{lp}C_{pp}^{-1}C_{pl}|. \tag{5}$$

It follows that $|C| = |C_{ll}||C_{pp} - C_{pl}C_{ll}^{-1}C_{lp}|$ and (4) is equivalent to

$$X_l^* = \arg \max_{X_l \subset X_n} |C_{ll}|. \tag{6}$$

Usually $l < p$ and, therefore, (6) is preferable to (4) from the computational point of view.

The exact solution of (6) or (4) generally requires the complete sorting out of all possible X_l. This, however, is unacceptable with respect to the computational effort for realistic values of l and n.

Shewry and Wynn (1987) utilized an exchange-type algorithm which is similar to that used in a number of optimization problems in statistics (e.g., the choice of the best subset of regressors, or the construction of discrete optimal designs).

For the D-criterion the exchange algorithm is based on the following simple idea. At the s-th iteration,

(a) delete the point x_s^- with the least contribution to $|C_{ll}(s)|$ from $X_l(s)$;

(b) add that point x_s^+ from $X_p(s)$ to $X_l^-(s) = X_l(s) - x_s^-$ that provides the largest increment of $|C_{ll}^-(s)| = |C[X_l^-(s)]|$.

The iterations are continued until $|C_{ll}(s)|$ increases or $|D_{pp}(s)|$ decreases.

¿From (5) and Frobenius' formula it follows that

$$x_s^- = \arg \max_{x_i \in X_l(s)} \{C_{ll}^{-1}(s)\}_{ii}.$$

Taking into account that $\{C_{ll}^{-1}(s)\}_{ii}^{-1}$ is the variance of the forecast $\hat{y}(x_i)$ [see (2)] that is based on the points that remain in $X_l(s)$ makes clear that at every step (a) the point at which the response is best explained by all others points of the set $X_l(s)$ has to be removed from this set.

Similarly

$$x_s^+ = \arg \min_{x \in X_p(s)} \{C^{-1}[X_l^-(s) + x]\}_{xx},$$

where the subscript xx stands for the element of the inverse matrix that corresponds to the added point. That point from $X_p(s)$ is added to $X_l^-(s)$ that is worst "explained" by the observations made at the points in $X_l^-(s)$.

Because of monotony the iterative procedure (a)–(b) converges, but not necessarily to the optimal separation into X_p^* and X_l^*. The procedure can give better results if, similar to Mitchell (1974), multiple forward and backward exchanges or "excursions" are used.

Probably one of the simplest versions of an iterative procedure was first proposed by Megreditchan (1979). He only used the deleting procedure for reducing X_l to the desired (prescribed) size.

Sometimes, $l \ll n$ (e.g., when l is the number of observing stations) and the optimization problem (6) can be solved by directly sorting out all possible X_l, especially when a formula similar to (5) is used for calculating $|C_{ll}|$ during subsequent replacement points in X_l.

For criteria different from D-criterion it is expedient to note that the matrix D_{pp} [see (2)] is a submatrix of

$$C^{-1} = \begin{pmatrix} C_{ll} & C_{lp} \\ C_{pl} & C_{pp} \end{pmatrix}^{-1} .$$

If n is not very large and C^{-1} can be computed, then the optimization problem (3) can be solved by the direct sorting out of all possible X_l. Otherwise one has to apply the exchange-type procedure similar to (a)–(b).

The reader can easily foresee the probable complication when instead of fixed l one has to find optimal X_l^* and l^* minimizing

$$w_1 Q(D_{pp}) + w_2 R(X_l) ,$$

where w_1 and w_2 are some weights and $R(X_l)$ are the costs of observations based on X_l.

2.3 Best supporting set

Let $\{x_1, \ldots, x_l\} = X_l \subset X$, where X is compact, and points $\{\overline{x}_1, \ldots, \overline{x}_p\} = X_p$ are fixed (they correspond to the points $x_{l+1}, \ldots, x_{l+p}$ from the previous sections).

One has to find $X_l^* \subset X$ providing the "best" forecast of $y_p = \{y(\overline{x}_1), ..., y(\overline{x}_p)\}^T$. According to (2), one has to search for

$$X_l^* = \arg \min_{X_l \subset X} Q[D_{pp}(X_l)] , \tag{7}$$

where $D_{pp}(X_l) = C_{pp} - C_p^T(X_l) C^{-1}(X_l) C_p(X_l)$, $C_p(X_l) = C_{lp}$, $C(X_l) = C_{ll}$.

When $p \le l$ and $X_p \subset X$ then (7) has the obvious solution: $X_l^* \supset X_p$, i.e., an optimal design has to include all points from X_p. The straightforward minimization in (7) can be cumbersome (usually $Q[D_{pp}(X_l)]$ is multimodal and the dimension of the problem is equal to $l \times \dim X$). The exchange-type algorithms (see, e.g., Fedorov and Müller (1988) and the previous section) can be a reasonable alternative.

2.4 Parametric Trend

Let

$$y(x) = f^T(x)\theta + u(x) , \tag{8}$$

where the random field $u(x)$ satisfies the same assumption which hold for $y(x)$ in the previous sections, $f(x)$ is a m-vector of known functions, and θ is a corresponding vector of parameters to be estimated. Then the best linear unbiased estimator is equal to [see, e.g., Ripley (1983, chp. 4.4), Fedorov (1989)]

$$\begin{aligned} \hat{\theta} &= M^{-1}(X_l)F(X_l)C^{-1}(X_l)y_l, \qquad (9)\\ \mathrm{Var}\{\hat{\theta}\} &= M^{-1}(X_l)\\ \hat{y}_p &= F^T(X_p)\hat{\theta} + C_p^T(X_l)C^{-1}(X_l)(y_l - F^T(X_l)\hat{\theta}), \qquad (10)\\ \mathrm{Var}\{\hat{y}_p\} &= \overline{D}_{pp} = D_{pp}(X_l) + \Delta_p^T(X_l)M^{-1}(X_l)\Delta_p(X_l)\,, \end{aligned}$$

where

$$\begin{aligned} M(X_l) &= F(X_l)C^{-1}(X_l)F^T(X_l)\\ \Delta_p(X_l) &= F(X_l) - F^T(X_l)C^{-1}(X_l)C_p(X_l)\\ F(X_p) &= \{f(x_i)\}_{i=1}^p, \quad x_i \in X_p\,. \end{aligned}$$

If the trend is of interest then

$$X_l^* = \arg\min_{X_l \subset X} \Psi[M^{-1}(X_l)]\,, \qquad (11)$$

and the design problem looks quite similar to the classical design theory problem, where the information matrix $M(X_l)$ is also the main object to cope with.

To optimize interpolation (extrapolation) one has to find

$$X_l^* = \arg\min_{X_l \subset X} Q[D_{pp}(X_l) + \Delta_p^T(X_l)M^{-1}(X_l)\Delta_p(X_l)]\,. \qquad (12)$$

Comparison with (3) or (4) shows that in (12) we have an additional term that is related to the trend: $\Delta_p^T(X_l)M^{-1}(X_l)\Delta_p(X_l)$. Frequently it is the dominative term.

The numerical construction of optimal designs for (11) and (12) is still a foggy problem. Similar to the previous sections, exchange-type algorithms can be used to construct reasonable good designs. Some useful formulas for that can be found in Brimkulov *et al.* (1986, chp. 4). For instance, the following recursion relation essentially simplifies the numerical algorithm:

$$M(X_{l+1}) = M(X_l) + \alpha(x)\varphi_l^T(x)\varphi_l(x)\,, \qquad (13)$$

where $X_{l+1} = X_l + x$,

$$\begin{aligned} \varphi_l(x) &= f(x) - F^T(X_l)C^{-1}(X_l)C_1(x)\\ \alpha^{-1}(x) &= C(x) - C_1^T(x)C^{-1}(X_l)C_1(x)\,; \end{aligned}$$

$C(x)$ is the variance of the observed value at x and $C_1(x)$ is the vector of its covariances with the observations at the points in X_l.

It is evident that (13) allows to use formally the results developed for the case with independent observations [see Fedorov (1972, chp. 3)]. For example, for the D-criterion one can use in the exchange algorithm based on the following recursion formula to include new points into the supporting set:

$$|M(X_{l+1})| = |M(X_l)|[1 + \alpha(x)\varphi_l^T(x)M^{-1}(X_l)\varphi_l(x)]^{-1}, \tag{14}$$

and subsequently

$$x^*_{l+1} = \arg\max_{x\in X} \alpha(x)\varphi_l^T(x)M^{-1}(X_l)\varphi_l(x). \tag{15}$$

A similar relation can be used for the deleting procedure.

Some matrix algebra [see also (1) and (2) and the comments given there] give

$$C^{-1}(X_l)C_1(x) = \begin{bmatrix} \delta(x,x_1) \\ \vdots \\ \delta(x,x_l) \end{bmatrix},$$

where $x \in X_l$, and $\delta(x,x_i) = 1$ if $x = x_i$ and 0 otherwise. Because of this fact, (15) does not admit any $x \in X_l$ as a candidate for being added again to X_l, e.g., $\alpha(x)\varphi_l^T(x)M^{-1}(X_l)\varphi_l(x) \equiv 0$ for all $x \in X_l$. Thus, only one measurement can be taken from any point (only one sensor can be located at any point). This corresponds to the fact that if a random field $u(x)$ has a continuous covariance structure and $x \to x'$, observations both in x and x' cannot contain more information than that in x alone.

Unfortunately, there are no results on the convergence to the optimal design of the iterative procedures based on (14) and (15) albeit some empirical confirmations that it usually leads to a significant improvement of the starting (initial) design.

3 Asymptotically Optimal Sampling Designs

When the number of available observations becomes reasonably large it is natural (compare the "continuous design" concept in traditional experimental design theory) to describe the location of observation points, i.e., the set X_l, by some distribution function $\xi(dx)$.

This idea was extensively used in a good cluster of publications initiated by Sacks and Ylvisaker (1966, 1968, 1970a, 1970b). The survey paper by Cambanis (1985) summarizes and illustrates various results and can be useful for readers who strive for details and bibliography.

3.1 Sampling designs, definitions

There are various sampling schemes which can correspond to the same distribution $\xi(dx)$, $x \in X$:

3.1.1 Deterministic sampling schemes

In these schemes it is assumed that the set X_l generated by $\xi(dx)$ is the same for measuring of all realizations of random field. For instance, $\xi(dx)$ can describe the location of observation stations in the optimal monitoring network problem [Fedorov (1989)].

To give an impression of possible schemes let us consider a one-dimensional case ($X \subset R$, $X = [a,b]$). A sequence $X_n = \{x_i\}_1^n$ with $\int_a^{x_i}\xi(dx) = \frac{i-1}{n-1}$ is called *regular deterministic sampling.*

It is clear that other schemes can easily be introduced, for instance, $X_n = \{x_i\}_1^n$ with $\int_a^{x_i} \xi(dx) = \frac{2i-1}{2n}$. This scheme is called *median sampling*.

Of course, multidimensional cases admit more varieties of deterministic sampling, fortunately usually leading to the same asymptotic results. In the observation network setup it is usually assumed that the integral

$$\int_{\Delta X} \xi(dx) = \Delta n \int_{\Delta X} dx$$

defines the number of observation points (stations, sensor locations) located at the area ΔX. It should be noted that the choice of ΔX may cause an additional problem: Δn has to be integer. A way to cope with this problem is first to fix Δn (e.g., to set $\Delta n \equiv 1$) and to choose then appropriate areas ΔX.

3.1.2 Simple random sampling

Here, every j-th realization of a set of n observations is taken at the points $\{x_{ij}\}_1^n$ which are chosen independently at random according to a distribution $\xi(dx)$. Usually, the simple random sampling scheme leads to the most simple asymptotical results. This scheme is also the most convenient in the multidimensional case.

3.1.3 Stratified sampling

In stratified sampling the set X is partitioned in a prescribed number of subsets. Usually the number of subsets is n and

$$\int_{\Delta_i^n} \xi(dx) \equiv n^{-1}\,.$$

The locations of observation point(s) is independently generated by some "local" distribution for every Δ_i^n . Usually it is a "normalized piece" of $\xi(dx)$.

3.1.4 Systematic sampling

It also assumes the stratification of X but only at Δ_1^n the location of the observation point is x_1 is generated by "local" distribution and subsequent points x_i $(i = 2, \ldots, n)$ are chosen from Δ_i^n according to some deterministic procedure depending upon x_1 (e.g., location x_i in Δ_i^n is homologous to x from Δ_1^n).

3.2 Parameter estimation, regular sampling

In this section the regression model (8) with one unknown parameter ($\theta \in R^1$) and with basic function $f(x)$ of the form

$$f(x) = \int_X C(x, x')\varphi(x')dx' \tag{16}$$

will be considered, where $\varphi(x)$ is a continuous function on X. Moreover, we will only deal with time series, i.e., X is a time interval.

Let $X_n = \{0 \le x_1 < \ldots < x_n \le 1\}$, then for the best linear unbiased estimator [see (9)]:

$$\begin{aligned} \hat{\theta}_n &= M^{-1}(X_n)Y_n && (17) \\ \mathrm{Var}\{\hat{\theta}_n\} &= M^{-1}(X_n), && (18) \end{aligned}$$

where $M(X_n) = \sum_{i,j=1}^n f(x_i)C_{ij}^{-1}(X_n)f(x_j)$ and $Y_n = \sum_{i,j=1}^n f(x_i)C_{ij}^{-1}(X_n)y(x_j)$. If observations are available over the entire interval X, then [see, e.g., Cambanis (1985)]

$$\mathrm{Var}\,\hat{\theta} = M^{-1}(X), \quad M(X) = \int_0^1 \int_0^1 \varphi(x)C(x,x')\varphi(x')dxdx'\,. \tag{19}$$

Let $\xi(dx)$ have a continuous derivative $p(x) > 0$, $x \in X$, and let X_n be the regular sampling generated by $p(x)$, then

$$\lim_{n\to\infty} M(X) - M(X_n) = 0\,.$$

When $C(X)$ is smooth and $C(x,x')$ has continuous (k,k) mixed partial derivatives, then

$$\varepsilon_n^2 = M(X) - M(X_n) = O(n^{-2k})\,. \tag{20}$$

If additionally these derivatives are smooth of the diagonal $X \times X$, then

$$\varepsilon_n^2 = O(n^{-2k-2})\,.$$

In other words, the smoother $C(x,x')$ is, the closer is $M(X_n)$ to $M(X)$. This assertion becomes clear, if one considers $M(X_n)$ as a quadrature formula for the approximation of the integral (19) [see also Sacks and Ylvisaker (1970b), Cambanis (1985)].

Under some further regularity conditions the rate of convergence is described by the following formula

$$n^{2k+2}\varepsilon_n^2 \to q_k \int_0^1 \frac{\alpha_k(x)\varphi^2(x)}{p^{2k+2}(x)}dx\,, \tag{21}$$

where

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

superscripts mean the corresponding derivatives, $q_k = |B_{2k+2}|/(2k+2)!$, and B_m is the m-th Bernoulli number.

¿From (21) follows that regular samplings generated by

$$p^*(x) \sim [\alpha_k(x)\varphi^2(x)]^{1/(2k+3)} \tag{22}$$

are asymptotically optimal and for them:

$$n^{2k+2}\varepsilon_n^2 \to q_k \left\{\int_0^1 [\alpha_k(x)\varphi^2(x)]^{1/(2k+3)}dx\right\}^{2k+3}\,. \tag{23}$$

3.3 Multiple regression

The case when $f(x)$, $\varphi(x)$, and θ are m-vectors was considered in details by Sacks and Ylvisaker (1968, 1970) for the regular sampling and one-dimensional X. Similar to the previous section the main problem consists of finding the discrepancy between $M(X)$ and $M(X_n)$ (now they are $m \times m$ matrices).

The complication comes from the fact that similar to the classical design problem various optimality criteria $\Psi(M)$ generate various optimal densities $p^*(x)$. The density $p_D^*(x)$ minimizing as $n \to \infty$

$$\frac{|M^{-1}(X_n^*)| - |M^{-1}(X)|}{\inf_{X_n}[|M^{-1}(X_n)| - |M^{-1}(X)|]}, \quad \Psi(M) = |M|,$$

is called asymptotically D-optimal. It was found that

$$p_D^*(x) \sim [\alpha_k(x)\varphi^T(x)M^{-1}(X)\varphi(x)]^{1/(2k+3)} \tag{24}$$

For linear criteria

$$\Psi(M) = \mathrm{tr}\, AM^{-1}, \quad A > 0\,,$$

the regular asymptotically optimal sampling generated by

$$p_A^*(x) \sim [\alpha_k(x)\varphi^T(x)M^{-1}(X)AM^{-1}(X)\varphi(x)]^{1/(2k+3)}. \tag{25}$$

For those who are familiar with the theory of optimal design for models with independent observations [Fedorov (1972)] formulas (24) and (25) should remind us the corresponding equivalence theorems.

4 Sampling Design with Approximated Covariance Function

4.1 Orthogonal expansion of the covariance kernel and the related regression model

Assume that the covariance function $C(x,x')$ is continuous on $X \times X$ and X is compact. Then this function can be presented in the following form:

$$C(x,x') = \sum_{\alpha=1}^{\infty} \lambda_\alpha \varphi_\alpha(x)\varphi_\alpha(x')\,, \tag{26}$$

where

$$\lambda_\alpha\varphi_\alpha(x) = \int_X C(x,x')\varphi_\alpha(x')dx'\,,$$
$$\lambda_1 \geq \lambda_2 \geq \ldots > 0,$$
$$\int_X \varphi_\alpha(x)\varphi_\beta(x)dx = \delta_{\alpha\beta}\,,$$

i.e., λ_α and $\varphi_\alpha(x)$ are eigenvalues and eigenfunctions, respectively, to the covariance kernel $C(x, x')$. Expansion (26) is well-known in the integral equation theory [see, e.g., Balakrishnan (1971, chp. 3)].

Formally, the random field with the covariance structure satisfying (26) may be generated (but not only) by the regression model with infinite number of random regression parameters with zero observation errors:

$$y_{ij} = \sum_{\alpha=1}^{\infty} \theta_{\alpha j}\varphi_\alpha(x_i), i = 1, \ldots, n, j = 1, \ldots, q, \tag{27}$$

with $\mathrm{E}\{\theta_{\alpha j}\} = 0$, $\mathrm{E}\{\theta_{\alpha j}\theta_{\beta j'}\} = \lambda_\alpha \delta_{\alpha\beta}\delta_{jj'}$. Model (27) can be considered as a special case of the intensively studied regression model with random coefficients—or second kind regression model [see, e.g., Spjotvoll (1977)]:

$$y_{ij} = \theta_j^T \varphi(x_i) + \varepsilon_{ij}, i = 1, \ldots, n, j = 1, \ldots, q, \tag{28}$$

where $\theta_j^T = \{\theta_1, \ldots, \theta_m\}$, $\varphi^T(x) = \{\varphi_1(x), \ldots, \varphi_m(x)\}$, $\mathrm{E}\{\theta_j\} = \theta_0$, $\mathrm{E}\{(\theta_j - \theta_0)(\theta_{j'} - \theta_0)^T\} = \Lambda\delta_{jj'}$, $\mathrm{E}\{\varepsilon_{ij}\} = 0$, $\mathrm{E}\{\varepsilon_{ij}\varepsilon_{i'j'}\} = \sigma^2\delta_{ii'}\delta_{jj'}$; θ_j and ε_{ij} are independent.

Let $X_n = \{x_1, \ldots, x_n\}$ and $X_n \subset X$. If in (26), starting from some $\alpha > m$, all $\lambda_\alpha = 0$, i.e., the covariance kernel is singular, then for any supporting set X_n ($n \geq m$) with a regular information matrix

$$M(X_n) = \sum_{i=1}^{n} \varphi(x_i)\varphi^T(x_i) > 0 \tag{29}$$

the exact parameters θ_j from (27) can be found:

$$\theta_j = \hat{\theta}_j = \arg\min_\theta \sum_{i=1}^{n} [y_{ij} - \theta^T\varphi(x_i)]^2, \tag{30}$$

where

$$\begin{aligned} \hat{\theta}_j &= M^{-1}(X_n)Y_j(X_n), \\ Y_j(X_n) &= \sum_{i=1}^{n} y_{ij}\varphi(x_i), \end{aligned}$$

and $y_j(x)$ can be predicted exactly at any $x \in X$.

(a) If $M(X_n)$ is singular (usually this occurs when $n < m$ and, of course, always when $m = \infty$), then (30) can be used with $\varphi_1(x)$ which is the subvector that comprises only the n first eigenfunctions from (26). The direct calculations lead to

$$y_j(x) - \hat{y}_j(x) = y_j(x) - \hat{\theta}_{1j}\varphi_1(x) = [\varphi_2^T(x) - \varphi_1^T(x)M_{11}^{-1}(X_n)M_{12}(X_n)]\theta_{2j}, \tag{31}$$

where $\varphi_2(x)$ corresponds to $(m - n)$ eigenfunctions not included in $\varphi_1(x)$, θ_{1j} and θ_{2j} are the corresponding parts of θ_j, $M_{11}(X_n) = \sum_{i=1}^n \varphi_1(x_i)\varphi_1^T(x_i)$ and $M_{12}(X_n) = \sum_{i=1}^n \varphi_1(x_i)\varphi_2^T(x_i)$.

(b) To make use of the information on λ_α and $\varphi_\alpha(x)$ with $\alpha > n$ a more complicated but usually more efficient estimator can be derived from (29) by taking into consideration the finite m and the vanishing σ^2. It is known that

$$\tilde{\theta}_j = [\Lambda^{-1} + \sigma^{-2}M(X_n)]^{-1}Y_j(X_n)$$

with

$$\mathrm{E}\{(\tilde{\theta}_j - \theta_j)(\tilde{\theta}_j - \theta_j)^T\} = [\Lambda^{-1} + \sigma^{-2} M(X_n)]^{-1}$$

is the best linear unbiased estimator; here, expectation is taken both over θ_j and ε_{ij}. Some matrix calculus leads to the following limit relations:

$$\begin{aligned} \tilde{y}_j(x) &= \lim_{\sigma^2 \to 0} \tilde{y}_j(x, \sigma^2) = \lim_{\sigma^2 \to 0} \varphi^T(x) \tilde{\theta}_j(\sigma^2) \\ &= C(x, X_n) C^{-1}(X_n) y_j(X_n)\,, \\ \lim_{\sigma^2 \to 0} \mathrm{E}\{[\tilde{y}_j(x, \sigma^2) - y_j(x)]^2\} &= C(x) - C(x, X_n) C^{-1}(X_n) C^T(x, X_n) \end{aligned} \tag{32}$$

where $C(X_n) = F^T(X_n) \Lambda F(X_n)$, $C(x, X_n) = \varphi^T(x) \Lambda F(X_n)$, and $C(x) = \varphi^T(x) \Lambda \varphi(x)$, with $F_{\alpha i}(X_n) = \varphi_\alpha(x_i)$ for $x_i \in X_n$.

Comparison of (32) with (1) and (2) demonstrates once more that the use of second kind regression models can help in understanding of various results for random fields.

It should be noted that the estimator that is described by (32) can also be obtained from $\tilde{y}_j(x) = \varphi^T(x) \tilde{\theta}_j$ to be

$$\tilde{\theta}_j = \arg\min_{\theta \in \Omega_j} \theta^T \Lambda^{-1} \theta\,, \tag{33}$$

with $\Omega_j = \{\theta : \theta_j^* = \arg\min_\theta \sum_{i=1}^n [y_{ij} - \varphi^T(x_i)\theta]^2\}$; cf., for instance, Section 4a.11 of Rao (1973).

4.2 Estimation of integrals

This problem was studied in many publications; the most recent ones are probably Micchelli and Wahba (1981), Schoenfelder and Cambanis (1982), and Cambanis (1985). In this section the problem is considered in the framework of the previous section.

Let

$$Q(X_n) = \mathrm{E}\{\int_X [y_j(x) - \hat{y}_j(x)]^2 dx\}\,. \tag{34}$$

¿From (27), (28), and (31) immediately follows that in case (a) of Section 4.1

$$Q_a(X_n) = \gamma_2^T \gamma_2 + \mathrm{tr}\, M_{11}^{-2}(X_n) M_{12}(X_n) \Lambda M_{21}(X_n) = \gamma_2^T \gamma_2 + \delta(\gamma_2, X_n)\,, \tag{35}$$

where $\gamma_2^T = (\lambda_{n+1}^{1/2}, \ldots, \lambda_m^{1/2})$, $\Lambda_{\alpha\beta} = \delta_{\alpha\beta}\gamma_\alpha^2$, $\alpha,\ \beta > n$, and it is assumed that the number of nonzero eigenvalues in (27) is finite. It is clear that

$$\min_{X_n} Q_a(X_n) \geq \gamma_2^T \gamma_2\,.$$

This inequality gives a new explanation of Theorem 1 by Micchelli and Wahba (1981), who prove that $\gamma_n^T \gamma_n$ is the lower bound of criterion (34) for any linear estimator $\hat{y}(x)$ and any design X_n, but "it is not known whether or not this lower bound can be achieved". Formula (35) helps to clarify the situation. If there exists a set

$$X_n^* = \arg\min_{X_n} \delta(\gamma_2, X_n) \tag{36}$$

such that

$$\delta(\gamma_2, X_n^*) = 0 \tag{37}$$

the cited theorem can naturally be justified. Condition (36) can be achieved if, for instance, a set X_n^* exists where the vector function $\varphi_1(x)$ is orthogonal to $\varphi_2(x)$, i.e., $M_{12}(X_n^*) = 0$.

Of course, (31) and (36) are closely related to the classical approximation theory.

In case (b) of Section 4.1

$$Q_b(X_n) = \mathrm{tr}[\Lambda - \Lambda F(X_n) C^{-1}(X_n) F^T(X_n) \Lambda] \,.$$

It can be shown that

$$Q_b(X_n) \leq Q_a(X_n)$$

with equality in all cases where $n = m$ and $F(X_n)$ of full rank.

4.3 Observations with random errors

The previous results can be generalized to the case when y_{ij} include observational errors [see (28)]:

$$y_{ij} = \sum_{\alpha=1}^{\infty} \theta_{\alpha j} \varphi_\alpha(x_i) + \varepsilon_{ij} \,, \tag{38}$$

where the errors ε_{ij} satisfy the assumptions stated in the context of (28). Calculations similar to (31)–(35) lead to

$$Q(X_n) = \gamma_2^T \gamma_2 + \delta(\gamma_2, X_n) + \sigma^2 \,\mathrm{tr}\, M_{11}^{-1}(X_n) \,. \tag{39}$$

The minimization of (39) reminds of the problems considered in the experimental design theory for nonadequate models [see, e.g., Box and Draper (1963), Karson, Manson, and Hader (1969)].

In cases when $\delta(\gamma_2, X_n)$ can be neglected the minimization of (39) becomes an optimization problem from experimental design theory. It was found by Fedorov and Müller (1988) that when the observed field is generated by (38) with a final number of $\theta_{\alpha j}$ (i.e., $\alpha = 1, \ldots, m$) then the exchange type algorithms known in the convex design theory will coincide with the algorithm discussed in Section 2.2. Of course, in the framework of sampling experiments repeated observations of the same realization $y(x)$ are not admitted and experimental designs with bounded density of supporting points have to be used [see, e.g., Fedorov (1989)].

5 Unknown Covariance Structure

All results presented in the previous sections have essentially used the fact that a covariance function $C(x, x')$ or a covariance matrix $C(X_n)$ is *a priori* known. There are two main approaches to evaluate the covariance structure.

5.1 Repeated observations

In many situations repeated observations on a set X_n are possible. Meteorology provides the most typical examples for that [see, e.g., Megreditchan (1979, 1985)].

Having repeated observations at every point of X_n one can directly estimate $C(X_n)$. If a process is stationary in time at every point $x \in X_n$ then one can use, for instance, the traditional estimator

$$\hat{C}(X_n) = k^{-1} \sum_{j=1}^{k} y_j y_j^T \,. \tag{40}$$

Similar to Section 2, $\mathrm{E}\{y_j\} = 0$ is assumed in (40).

A usual recommendation for the practitioner is to substitute $C(X_n)$ by $\hat{C}(X_n)$. Immediately, the problem of stability or robustness of optimization problems (3), (4), and (6) is arising. It is useful to emphasize again that one can choose optimal design X_l^* (optimal sites, sampler locations, etc.) only from points where the previous observations are available. Therefore, the optimization problem (7) is beyond the scope of the considered approach. Of course, all asymptotical results mentioned in Section 3 desperately demand more detailed information than that provided by (40).

Along with the recommendation to substitute true values by their estimates there exists a more natural approach leading to computational procedures based on the familiar algorithms developed for the best model choice in the traditional regression analysis.

¿From the definition it follows that the best linear unbiased estimator $\hat{y}_p$ [see (1)] minimizes the square risk matrix

$$V(b) = \mathrm{E}\{(y_p - \hat{y}_p)(y_p - \hat{y}_p)^T\} = \mathrm{E}\{(y_p - by_l)(y_p - by_l)^T\} \,,$$

where b is a $l \times p$ matrix. Direct minimization of any linear form $L^T V(b)$ gives [see (1)]

$$b^* = C_{ll}^{-1} C_{pl} \,.$$

Of course, $V(b^*) = D_{pp}$ and

$$V(b) - V(b^*) = v \,, \tag{41}$$

where v is nonnegatively defined matrix ($v \geq 0$).

If k realizations (observations) of y_l are available then

$$\hat{V}(b) = k^{-1} \sum_{j=1}^{k} (y_{pj} - b^T y_{lj})(y_{pj} - b^T y_{lj})^T \tag{42}$$

is a reasonable estimator for $V(b)$. Matrix b can formally be considered as a matrix of regression coefficients in the linear model

$$y_{pj} = b^T y_{lj} + \varepsilon_{pj} \,.$$

For any p-vector L, the quadratic form $L^T \hat{V}(b) L$ is minimal when

$$\hat{b} = \hat{C}_{ll}^{-1} \hat{C}_{pl} \,. \tag{43}$$

where $\hat{C}_{ll}$ and $\hat{C}_{pl}$ are the corresponding submatrices of matrix $\hat{C}(X_n)$ defined by (40).

Subsequently,

$$\hat{D}_{pp} = \hat{V}(\hat{b}) = \hat{C}_{pp} - \hat{C}_{pl} \hat{C}_{ll}^{-1} \hat{C}_{lp} \,. \tag{44}$$

Thus (42) and (43) confirm the previously expressed recommendations about substituting the "true" matrices by their direct estimates.

It is interesting to point out that, after substituting true matrices by their estimates, the iterative procedure (a)–(b) from Section 2.2 can be reformulated in the following way that probably will appeal to many practitioners [see Fedorov and Müller (1989)]:

(a) Remove that point x_s^- from the set $X_l(s)$ at which the observation is best explained by the observations at all other points of $X_l(s)$:

$$x_s^- = \arg \min_{x_i \in X_l(s)} v_i^2(s), \tag{45}$$

where

$$v_i^2(s) = \min_b \sum_{j=1}^k [y_{ij} - b^T y_{l(k),j}]^2,$$

y_{ij} are observations at point x_i, $j = 1, \ldots, k$, and $y_{l(k),j}$ are the observations at the points of $X_l(s)$ except point x_i, $j = 1, \ldots, k$.

(b) Add that point x_s^+ from $X_p(s)$ to the set $X_l(s) - x_s^-$ at which the observation is most poorly explained by the observations at all other points of $X_l(s) - x_s^-$:

$$x_s^+ = \arg \max_{x_i \in X_p(s)} v_i^2(s). \tag{46}$$

The deleting procedure (44) was used by Megreditchan (1985) to rank the "informativeness" of metereological stations. His procedure starts with $X_l(s = 0) = X_n$.

The application of (45)–(46) allows to use the well-developed methods of traditional regression analysis such as stepwise procedures or leap-and-bound algorithms [see, e.g., Seber (1977)] to avoid direct inversion of matrix $\hat{C}_{ll}(s)$ which can be large and ill-conditioned. Nevertheless, it has to be emphasized that these methods should be applied precautiously. They are not linear (do not provide linear estimators); therefore, the assertion that $\hat{b}$ minimizes any quadratic form $L^T \hat{V}(b) L$ [see the comment to (43)] is not generally true. Of course, this remark is of greater importance for theory than for a practitioner. One can easily predict further complications in tackling the optimization problem (4).

Concluding this section, it should be emphasized that the set X_n can only contain points (observing sites) where are series of previous observations. So no new "candidates" are admitted to the discussed approach.

5.2 Parametric estimation of a covariance function

For the design problem defined by (7) or for any design problem considered in Section 3 the covariance function $C(x, x')$ has to be known at every $x \in X$. And an experimenter is forced to approximate it using the finite number of available observations. To get the admissible approximation one needs to make various assumptions on the structure of $C(x, x')$ [see Ripley (1983, chp. 2)].

For instance, $C(x, x')$ is assumed to be homogeneous, i.e., it depends only on the vector $u = x - x'$. More frequently it is assumed that $C(x, x')$ is isotropic, i.e., it depends only on the distance between x and x': $C(x, x') = C(r)$, $r^2 = (x - x')^T(x - x')$. The last assumption converts the approximation to a one-dimensional problem.

One also has to use only those functions $\tilde{C}(x, x')$ for the approximation of the covariance function which are nonnegative definite and symmetric:

$$\tilde{C}(x, x') = \tilde{C}(x', x) \text{ and } \sum_{i,j=1}^{n} \gamma_i \gamma_j \tilde{C}(x_i, x_j) \geq 0 \tag{47}$$

for any γ_i, x_i, $i = 1, \ldots, n$, and any n.

The isotropy of a covariance function imposes some additional restrictions on the class of functions that can be used for the approximation. For instance, it is known that

$$C(r) \geq \inf_u \left[k! \frac{2}{u}^k J_k(u) \right] C(0), \tag{48}$$

where $k = (d-2)/2$, $d = \dim X$, and $J_k(u)$ is the Bessel function. In the two dimensional case (48) leads to the inequality

$$\frac{C(r)}{C(0)} \geq -0.403.$$

More details on the properties of the covariance functions and the parametric functions that can be used for the approximation can be found in Ripley (1983, chp. 2).

There exist experimental situations, where only the discussed approach can be used. Probably most of them are studied in geostatistics where only a single realization of the random field is available for observation (it could be a field of the concentration of some mineral). The methodology for these situations was mainly developed in the frame of "kriging methods" [see Matheron (1973), Journal and Huijbregts (1978)].

The idea of estimation of a covariance function is based on

(a) its parameterization $C(r, \gamma)$;

(b) selecting or collecting of site-pairs with the (appoximately) same distance r to have the "observed" values $\tilde{C}(r_i)$;

(c) estimation of parameters γ by the least square method or by a similar approach.

Comparative analysis of the linear interpolators and the kriging estimators are given by Ripley (1983, chp. 4) and by Fedorov (1989).

5.3 Fields generated by regression models with random parameters

The life of practitioner becomes easier if he believes that the observed field is generated (or can be appoximated) by a regression model (28).

Well-developed methods exist for estimating the covariance matrix Λ and the variance (if there are observational errors) σ^2. A comprehensive survey of them is given by Spjøtvoll (1977).

When the field is observed for various j at the same points X_n, then

$$\hat{\Lambda} = k^{-1} \sum_{j=1}^{k} \hat{\theta}_j \hat{\theta}_j^T - \hat{\sigma}^2 M^{-1}(X_n) , \tag{49}$$

where $\hat{\theta}_j$ is the least square estimates [see (30)] and

$$\hat{\sigma}^2 = k^{-1}(n-m)^{-1} \sum_j v_j^2(\hat{\theta}), \quad v_j^2(\hat{\theta}) = \sum_{i=1}^{k} [y_{ij} - \hat{\theta}_j^T \varphi(x_i)]^2 .$$

Subsequently,

$$\hat{C}(x,x') = \varphi^T(x)\hat{\Lambda}\varphi(x') + \hat{\sigma}^2 = k^{-1} \sum_{j=1}^{k} \eta(x,\hat{\theta}_j)\eta(x',\hat{\theta}_j) + \hat{\sigma}^2[1 - d(x,x')] , \tag{50}$$

where $\eta(x,\theta) = \theta^T \varphi(x)$ and $d(x,x') = \varphi^T(x) M^{-1}(X_n)\varphi(x')$.

In (49) and (50) it was assumed that $\mathrm{E}\{\theta_j\} = 0$. Both estimates admit various generalizations. If there is a trend ($\mathrm{E}\{\theta_j\} \neq 0$) and for every j the sample points X_n are different, then e.g.:

$$\hat{C}(x,x') = (k-1)^{-1} \sum_{j=1}^{k} [\eta(x,\hat{\theta}_j) - \overline{\eta}(x)][\eta(x',\hat{\theta}_j) - \overline{\eta}(x')] + \hat{\sigma}^2[1 - \overline{d}(x,x')] , \tag{51}$$

where $\overline{\eta}(x) = k^{-1} \sum_{j=1}^{k} \eta(x,\hat{\theta}_j)$ and $\overline{d}(x,x') = k^{-1} \sum_{j=1}^{k} \varphi^T(x) M^{-1}(X_{nj})\varphi(x')$.

It has to be emphasized that unlike Section 1, (49)–(51) provide estimates for the covariance function for every pair x, $x' \in X$ and for arbitrary locations X_{nj} that generate nonsingular information matrices $M(X_{nj})$ even with various numbers of sampling points for various realizations.

It is clear that similar to Sections 1 and 2 one can recommend to use $\hat{C}(x,x')$ as a substitution of $C(x,x')$ through all the results of Section 4.

5.4 Moving regression

The most sceptical practitioner may probably consider all estimation procedures presented above as sophisticated versions of the weighted average

$$\hat{y}(x) = \sum_{i=1}^{n} w(x,x_i) y_i , \tag{52}$$

or the moving regression predictors.

In the case of moving regression, in the vicinity of x the locally linear regression model is used for the surface approximation:

$$y_i \simeq \theta_0(x) + \theta^T(x) f(u_i,x) + \varepsilon(u_i,x), \quad i = 1,\ldots,k,$$

where $\theta_0(x)$ is the parameter to be estimated, $\theta(x)$ are nuissance parameters, $f(u,x)$ is a vector of given functions that vanish when $u = x_i - x \to 0$, and $\varepsilon(u,x)$ comprises the observation errors and the deviation of the model from the true values. The estimator is defined as a solution of the least square problem

$$[\hat{\theta}_0(x), \hat{\theta}(x)] = \arg\min_{\theta_0,\theta} \sum_{i=1}^{k} w(u_i, x)[y_i - \theta_0 - \theta^T f(u, x)]^2 , \tag{53}$$

$$\hat{y}(x) = \hat{\theta}_0.$$

Normally $w(u, x)$ is a unimodal function with the maximum at $u = 0$ and vanishing when $u \to \infty$. In many applications

$$w^{-1}(u, x) = \sigma^2 + \delta^2(u, x) ,$$

where σ^2 is the variance of the observation error and $\delta(u, x)$ is proportional to the assumed deviation of the model from the reality. The moving regression approach is evidently very close to the results reported in Sections 2.4 and 4.1.

The optimal location problem is now transformed into that of minimizing the average variance of the predicted values $\hat{y}(x_j)$, $j = 1, \ldots, p$:

$$\xi_n^* = \arg\min_{\xi_n} \sum_{j=1}^{p} d(x_j, \xi_n) , \tag{54}$$

where $\xi_n = \{x_1, \ldots, x_n\}$, $d(x_j, \xi_n) = l^T M^{-1}(x_j, \xi_n) l$, $M(x_j, \xi_n) = n^{-1} \sum_{i=1}^{n} w(x_i - x_j) f(x_i - x_j) f^T(x_i - x_j)$, and $l^T = (1, 0, \ldots, 0)$. The optimization problem (54) can be imbedded in the standard experimental design theory, and numerical algorithms considered by Fedorov (1989) can provide reasonable approximate solutions of (54).

6 Conclusions

In experiments where the dimension of the observation space is greater than one, we distinguish two cases:

- spatial sampling
- time series sampling

In the case of time series sampling the nature of time causes that we cannot sample in random order, but have to sample at times $t_1 < \ldots < t_n$.

However, in most practical situations – in meteorology, environmental studies, and others – observations are taken simultaneously. Also for the methods discussed in this paper, time is not essential. In situations where the dynamics of the field is crucial for a particular study, the optimal timing of observations can improve considerably the accuracy of the results [see Bensoussan (1972)]. Such experiments are based typically on "physical" models that are derived from behavioral differential equations so that the response function is not known explicitly; the models often are nonlinear in the parameters. Such problems are scarcely treated by studies known to the authors. Experiments of this kind also demand special methods that are beyond the scope of our survey.

In many technological experiments measurements are taken sequentially in time. To avoid systematic effects of time-trends randomization of observation points $\{x_1, \ldots, x_l\}$ can be recommended. People follow the same idea since the very beginning of experimental design history. This randomization can be combined with any type of sampling discussed in Section 3.1.

According to our opinion the approaches discussed in Sections 2 and 3 are mathematically most developed while methods treated in other sections are still in an early stage of research. We think that unlike the convex optimal experimental design theory the popular D-criterion is less important in optimal sampling than the linear criteria that are well suited for forecasting and evaluation (over space and/or time) averaged values.

It is clear that the quality of estimation of the variance-covariance structure is crucial both for prediction and for experimental design. The author failed to trace any well-promoted results on the evaluation of the effect of substituting $C(x, x')$ by $\hat{C}(x, x')$.

For data analysis some constructive comments can be found in Armstrong (1984). A sensitivity analysis examining how estimates will change for small perturbations in the variance-covariance structure was done by Warnes (1986). This, however, is a numerical analysis of the stability of the best linear unbiased estimator in the frame of the kriging method rather than a statistical analysis of the corresponding estimator which, for instance, must determine the increase of the estimator variance or the appearance of a nonzero bias.

Asymptotical properties of unbiased linear predictors of a random field with a misspecified covariance function were studied by Stein (1988). It is proved that kriging type estimators are asymptotically efficient under mild conditions as n tends to infinity; here, n is the number of observing points, but does not include points where the prediction is made. Unfortunately, the stability of optimal asymptotical locations (see Section 3) has not been investigated.

Acknowledgement. This paper was written during a stay of one of the authors (Valery V. Fedorov) at the Institut für Statistik of the University of Economics (Wirtschaftsuniversität) in Vienna, who wishes to express his appreciation to the whole staff of the Institute for their help and cooperation, and especially to Peter Hackl without whose activity and energy this paper would never have been published. Special thanks go to Christa Hackl for her patience in typing and editing the various versions of the manuscript.

References

[1] Armstrong, M. (1984) "Problems with universal kriging", *Mathematical Geology*, **16**, 101–108.

[2] Balakrishnan, A.V. (1971) *Introduction to Optimization Theory in a Hilbert Space*, Berlin: Springer-Verlag.

[3] Bensoussan, A. (1972) "Optimization of sensors' location in a distributed filtering problem", *Stability of Stochastic Dynamic Systems*, New York: Springer-Verlag.

[4] Box, G.E.P. and Draper, N.R. (1963) "The choice of a second order rotable design", *Biometrika*, **50**, 335–352.

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

[6] Cambanis, S. (1985) "Sampling designs for time series", *Handbook of Statistics,* New York: North-Holland, **5**, 337–362.

[7] Dalenius, T., Hajek, J., and Zubrzycki, S. (1961) "On plane sampling and related geometrical problems", *Proceeding of the 4th Berkeley Symposium on Mathematical Statistics*, 125–150.

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

[9] Fedorov, V.V. (1989a) "Kriging and other estimators of spatial field characteristics", *Atmospheric Environment*, **23**, 175–184.

[10] Fedorov, V.V. (1989b) "Optimum design with bounded density. Optimization algorithms of the exchange type", *Journal of Statistical Planning and Inference*, **22**, 1–13.

[11] Fedorov, V.V. and Müller, W. (1988) "Two approaches in optimization of observing networks", in Y. Dodge, V.V. Fedorov, and H.P. Wynn (Eds.). *Optimal Design and Analysis of Experiments.* New York: North-Holland, 239–255.

[12] Gandin, L.S., Kagan, R.L., and Tarakanova, V.P. (1968) "On the rational planning of the network for air temperature observations", *Trudy GGO*, **228**.

[13] Gribik, P.R., Kortanek, K.O., and Sweigart, J.R. (1976) "Designing a regional air pollution monitor network: An appraisal of a regression experimental design approach", *Proceedings of the Conference on Environment Modelling and Simulation*, EPA 600/9-76-016, Research Triangle Park, North Carolina, 86–91.

[14] Journel, A. and Huijbregts, C. (1978) *Mining Geostatistics*, New York: Academic Press.

[15] Karson, M.J., Manson, A.R., and Hader, R.J. (1969) "Minimum bias estimation and experimental design for response surfaces", *Technometrics*, **11**, 461–475.

[16] Katkovnik, V.Y. (1985) *On Parametric Identification and Data Smoothing*, Moscow: Nauka.

[17] Kumar, S. and Seinfeld, J.H. (1978) "Optimal location of measurements for distributed parameter estimation", *IEEE Transactions on Automatic Control*, **AC-22**, 690–698.

[18] Matérn, B. (1986) "Spatial variation", in D. Brillinger et al. (Eds.), *Lecture Notes in Statistics*, Berlin: Springer-Verlag, **36**.

[19] Matheron, C. (1973) "The intrinsic random functions and their approaches", *Advances in Applied Probability*, **5**, 434–468.

[20] Megreditchan, G. (1979) "L'optimization des reseaux d'observation des champs meteorologiques", *La Meteorologie*, **6**, 51–66.

[21] Megreditchan, G. (1985) *Methods statistiques d'analyse et d'interpolation des champs meteorologiques*, Geneve: Organisation Meteorologique Mondiale, 300 p.

[22] Micchelli, C.A. and Wahba, G. (1981) "Design problems for optimal surface interpolation", in Z. Ziegler (Ed.), *Approximation Theory and Applications*, New York: Academic Press, 329–348.

[23] Mitchell, T.J. (1974) "An algorithm for construction of D-optimal experimental design", *Technometrics*, **16**, 203–210.

[24] Munn, R.E. (1981) *The Design of Air Quality Monitoring Network*, London: MacMillan Publishers.

[25] Rao, C.R. (1973) *Linear Statistical Inference and Its Applications. 2nd Edition*, New York: John Wiley.

[26] Ripley, B.D. (1983) *Spatial Statistics*, New York: John Wiley.

[27] Sacks, J. and Ylvisaker, D. (1966) "Designs for regression problems with correlated errors", *Annals of Mathematical Statistics*, **37**, 66–89.

[28] Sacks, J. and Ylvisaker, D. (1968) "Designs for regression problems with correlated errors, many parameter", *Annals of Mathematical Statistics*, **39**, 49–69.

[29] Sacks, J. and Ylvisaker, D. (1970a) "Designs for regression problems with correlated errors, III", *Annals of Mathematical Statistics*, **41**, 2057–2074.

[30] Sacks, J. and Ylvisaker, D. (1970b) "Statistical designs and integral approximation", in *Proceedings of the Twelfth Biennial Seminar of the Canadian Mathematical Congress*, Montreal, 115–136.

[31] Schoenfelder, C. and Cambanis, S. (1982) "Random designs for estimating integrals of stochastic processes", *Annals of Statistics*, **10**, 526–538.

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

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

[34] Spjøtvoll, E. (1977) "Random coefficients regression models: A review", *Mathematische Operationsforschung und Statistik, Ser. Statistik*, **8**, 69–93.

[35] Stein, M.L. (1988) "Asymptotically efficient prediction of a random field with a misspecified covariance function", *Annals of Statistics*, **16**, 55–63.

[36] Thiebaux, H.J. and Peddler, M.A. (1987) *Spatial Objective with Applications in Atmospheric Science*, New York: Academic Press.

[37] Warnes, J.J. (1986) "Sensitivity analysis for universal kriging", *Mathematical Geology*, **18**, 653–676.

[38] Ylvisaker, D. (1975) "Design on random fields", in J.N. Srivastava (Ed.), *A Survey of Statistical Design and Linear Models*, New York: North-Holland, 593–607.

[39] Ylvisaker, D. (1987) "Prediction and design", *Annals of Statistics*, **15**, 1–19.