Working Paper

COMPARING THE PERFORMANCE OF TWO COMPETING MODELS

V. V. Fedorov

November 1986
WP-86-73

International Institute for Applied Systems Analysis
A-2361 Laxenburg, Austria

NOT FOR QUOTATION
WITHOUT THE PERMISSION
OF THE AUTHOR

COMPARING THE PERFORMANCE OF TWO COMPETING MODELS

V. V. Fedorov

November 1986
WP-86-73

INTERNATIONAL INSTITUTE FOR APPLIED SYSTEMS ANALYSIS
2361 Laxenburg, Austria

Foreword

Computerization of the environmental sciences is one of the most typical trends of the last two decades. A number of different models describing the same objects or phenomena are circulating in scientific media and interest in publications in which these models are compared has increased impressively.

Usually the comparison is based on intuitive ideas, and the object of this paper is to give some recommendations on how to use standard statistical techniques for model comparison. The key theme consists of an introduction to several statistically reasonable discrepancy measures for competing models and their subsequent maximization by varying the location of an experimental network.

In general, this paper covers the author's results in this sector of mathematical statistics.

Prof. M. Antonovsky

Acknowledgement

I am most grateful to Professor R.E. Munn for his constructive criticism and recommendations, which were extremely helpful in the preparation of this paper.

COMPARING THE PERFORMANCE OF TWO COMPETING MODELS

V. V. Fedorov

1. INTRODUCTION

Beginning with numerical weather prediction experiments in the 1920s, models of environmental processes have become more and more complex, keeping pace with advances in computer technology. Some of the current models can be run only on very large computers such as the CRAY on which, for example, the Navier-Stokes equations are solved using very short forward time steps and many points in space.

Investigators have long been interested in testing these big models with field data. In particular, when a new model is proposed (due to an advance in our physical understanding of the processes involved, or to advances in computer capabilities) it is important to determine whether the model is "better" before adopting it operationally for national weather forecasts, acid rain predictions, etc.

One problem is the definition of the word "better", which involves value judgements. For example, in an urban air pollution model the predictions could be widely different from observed values simply because the forecast wind direction was 30 degrees in error. By rotating the axis, a much improved match of observation and prediction could be obtained. Many similar examples could be given in which objective criteria must be established and promoters of competing models may sometimes disagree with one another because of the objective criteria they use.

Recent review articles on model performance in the air pollution field have been written, for instance, by Hayes and Moore (1986), Willmott (1982), Zwerver and Van Ham (1985), and a very interested paper by Fox (1981). Here we shall try to connect the ideas given in those papers with more formal results from mathematical statistics.

Probably a scientist who has worked with complex numerical models of physical processes would be very sceptical about the simplicity of the models considered in this paper. Nevertheless, simple diagnostic cases illuminate the main ideas and final results, and give some orientation which usually cannot be achieved in more complicated situations. Most certainly, the process of model comparison cannot be imbedded in a routine scheme (even one that is quite perfect). Usually there is need for some integration of standard mathematical techniques with the intuition of a researcher (for details, see Munn, 1981).

2. MAIN ASSUMPTIONS

Let a system "object under investigation - process of observation" be described by the following model

$$y_{ir} = \eta_i(x_i) + \varepsilon_{ir} (i = \overline{1,n}, \ r = \overline{1,r_1}). \tag{1}$$

A function $\eta_t(x)$ is a response function and x_i is a vector of conditions under which the i-th set of observations are made. Subscript "t" stands for "true" values which we try to observe or measure in direct experiments and to estimate in indirect experiments. Errors ε_{ir}, for instance, can reflect the imperfection of an observation process; stochastically of an object under observation; approximate character of used representation for $\eta_t(x)$; and so on.

One of the most crucial assumptions in the following is that ε_{ir} ***are random (stochastic) values***. This is a significant component of model (1) and one can say that a stochastic model is used for the description of ε_{ir}. It is necessary to emphasize that this assumption is essential to the whole concept of the paper. Details (to be supplied later) can technically change the final results, but they are adjustable to those details in the frame of the main idea of the paper.

Another significant assumption consists of the fact that *components of x_i (or at least part of them) can be chosen (or controlled)* by an experimenter. It will be assumed that $x_i \in X \subset R^k$, where X is compact.

The set of values

$$\xi_N = \{p_i, x_i\}, \quad p_i = r_i / N, \sum_{i=1}^{n} p_i = 1$$

specifies a design. The fractions p_i can be considered as measures prescribed to points x_i and variations of these measures must be proportional to N^{-1} in practice.

The major efforts of this paper will be directed to the case when it is ***a priori*** known that the function $\eta_t(x)$ (or true response) has to coincide with one of the two given functions, either $\eta_1(x, \vartheta_1$ or $\eta_2(x, \vartheta_2)$, where ϑ_j are parameters to be estimated, $\vartheta_j \in \Omega_j \subset R^{m_j}$. In general, there are no very special demands to these functions. For instance, they can be numerical solutions of some system of differential equation. However, for a number of presented results, their linearity ($\eta(x, \vartheta) = \vartheta^T f(x)$) will be important.

What is really essential in the last assumption is that *one of the comparing functions coincides with the true response*. In practice, this means that an experimenter believes in the closeness of (at least) one model to reality.

Cases where one needs to compare more than two models lead to certain mathematical difficulties but the corresponding techniques are more or less a straightforward generalization of the results presented here (compare with Atwood and Fedorov, 1975).

3. OPTIMALITY CRITERIA

The objective of an experiment is to choose the true model. To start the discussion of the experimental design problem one must apply to some criteria of optimality (Atkinson, Fedorov, 1975; Fedorov, 1980). The main idea behind these criteria is based on introduction of some measure of discrepancy between rival models depending upon the difference: $\eta_1(x, \vartheta_1) - \eta_2(x, \vartheta_2)$.

To be specific, suppose that the first model is true, i.e., there exists such ϑ_{1t} that $\eta_t(x) = \eta_1(x, \vartheta_{1t})$. If random errors ε_{ir} are independent and normally distributed with variances $\sigma^2 \equiv 1$ then it is reasonable to apply to the following measure of discrepancy

$$T_1^0(\xi_N, \vartheta_{1t}) = \inf_{\vartheta_2 \in \Omega_2} \sum_{i=1}^{n} p_i \{\eta_1(x_i, \vartheta_{1t}) - \eta_2(x_i, \vartheta_2)\}^2 . \tag{2}$$

The value $NT_1^\circ(\xi_N,\vartheta_{1t})$ coincides with the noncentrality parameter of χ^2-distribution (or F-distribution, when the variance of ε_{ir} is unknown) if $\eta_2(x_1\vartheta_2) = \vartheta_2^T f_2(x)$, where $f_2(x)$ is a vector of the given basic functions, and Ω_2 coincides with R^{m_2}. In other cases ($\eta_2(x_1\vartheta_2)$ is nonlinear or for arbitrary Ω) this fact has asymptotical ($N \to \infty$) character. More details see in Atkinson, Fedorov, 1975; Fedorov, 1981.

It will be useful to consider also a generalized version of (2):

$$T_1(\xi_N,\vartheta_{1t}) = \inf_{\vartheta_2\in\Omega_2} \sum_{i=1}^{N} p_i F\{\eta_1(x_i,\vartheta_{1t}) - \eta_2(x_i,\vartheta_2)\} . \tag{3}$$

For instance, robust M-estimators can lead to that kind of discrepancy measures (see, Huber, 1981).

The design

$$\xi_N^* = Arg \sup_{\xi_N} T_1(\xi_N,\vartheta_{1t}) \tag{4}$$

is called T_1-optimal design. To emphasize that the criterion of optimality is constructed under assumption that $\eta_1(x_1\vartheta_{1t}) = \eta_t(x)$ the design ξ_N^* will some times be called "locally T-optimal design".

Together with locally optimal designs we will consider maximin and Bayesian designs.

The design ξ_N^* is maximum if

$$\xi_{mN}^* = Arg \sup_{\xi_N} \inf_{\vartheta_1\in\Omega_1} \inf_{\vartheta_2\in\Omega_2} \sum_{i=1}^{n} p_i F\left\{\eta_1(x_i,\vartheta_1) - \eta_2(x_i,\vartheta_2)\right\} = \tag{5}$$

$$= Arg \sup_{\xi_N} \inf_{\vartheta_j\in\Omega_j} T_j(\xi_N,\vartheta_j) \ , \ j=1,2 .$$

If

$$\xi_{BN}^* = Arg \sup_{\xi_N} \sum_{j=1}^{2} \pi_j \int_{\Omega_j} T_j(\xi_N,\vartheta_j)\mu(d\vartheta_j) \tag{6}$$

where π_j is a prior probability of j-th model and $\mu(d\vartheta_j)$ is a corresponding prior distribution of $\vartheta_j\,(j=1,2)$, then ξ_{BN}^* is a Bayesian optimal design.

4. CONTINUOUS OPTIMAL DESIGNS

In what follows only the continuous versions of optimization problems (4)-(6) will be considered. In other words, the discreteness of p_i is neglected and

$$T_1(\xi_1\vartheta_1) = \inf_{\vartheta_2\in\Omega_2} \int_X F\{\eta_1(x,\vartheta_1) - \eta_2(x,\vartheta_2)\}\xi(dx) , \tag{7}$$

where $\xi(dx)$ can be any probabilistic measure with a supporting set belonging to X . It is clear that for the continuous case the subscript "N" does not bear any additional information and can be omitted.

Formally optimization problems (4) and (5) are similar. Both of them can be transformed to the following optimization problem:

$$\xi^* = Arg \sup_{\xi} T(\xi) = Arg \sup_{\xi} \inf_{\vartheta\in\Omega} \int_X F\{\eta(x,\vartheta)\}\xi(dx) . \tag{8}$$

For instance, in (5) one has to put $\vartheta^T = (\vartheta_1^T,\vartheta_2^T)$, $\eta(x,\vartheta) = \eta_1(x,\vartheta_1) - \eta_2(x,\vartheta_2)$

and $\Omega = \Omega_1 \times \Omega_2$. In the case of (4) when $\Omega = \Omega_2$ and $\eta(x,\vartheta) = \eta_1(x,\vartheta_{1t}) - \eta_2(x,\vartheta_2)$ it is crucial that the solutions of (8) will depend upon $\vartheta_{1t}: \xi^* = \xi^*(\vartheta_{1t})$.

Let us assume that

(a) the sets and Ω are compact and function $\eta(x,\vartheta)$ be continuous on $X \times \Omega$.

(b) the function $F(z)$ is monotonously increasing when $z \geq 0$ and monotonously decreasing when $z < 0$ and continuous on $Z = \{z : z = \eta(x,\vartheta), x \in X, \vartheta \in \Omega\}$.

Theorem 1.

(i) *There exists at least one solution of (8). The set of optimal designs is convex.*

(ii) *A necessary and sufficient condition for a design ξ^* to be optimal is the existence of a measure $\mu^*(d\vartheta)$ such that*

$$\tau(x,\xi^*) \leq T(\xi^*) ,$$

where

$$\tau(x,\xi^*) = \int_{\Omega^*} F\{\eta(x,\vartheta\}\mu^*(d\vartheta)$$

and the measure μ^* has the supporting set

$$\Omega^* = \{\vartheta^* : \vartheta^* = Arg \inf_{\vartheta \in \Omega} \int_X F\{\eta(x,\vartheta)\}\xi^*(dx)\} , \int_{\Omega^*} \mu^*(d\vartheta) = 1 .$$

(iii) *The function $\tau(x,\xi^*)$ achieves its upper bound on the supporting set of ξ^* .*

If in addition to (a) and (b):

(c) the function $F\{\eta(x,\vartheta)\}$ is a convex function of ϑ for all $x \in X$ and Ω is a convex compact, then

Theorem 2. There always exists an optimal design containing no more than $m+1$ supporting points, where m is the dimension of ϑ.

If, in addition to previous assumptions: (e) the function $F(z)$ is symmetrical, then:

Theorem 3. The supporting set of an optimal design for (8) belongs to Tchebysheff extremal basis:

$$(\vartheta^*, X^*) = Arg \inf_{\vartheta \in \Omega} \sup_{x \in X} |\eta(x,\vartheta)| .$$

Theorems 1-3 are helpful in the understanding of general structure of optimal designs and in some cases in this analytical construction (see, Fedorov, 1981; Denisov, Fedorov, Khabarov, 1981).

In cases when the definition of Ω includes at least k linearly independent constraints which are active for ϑ^* , i.e., $\psi(\vartheta^*) = O$, then the number of supporting points in Theorem 2 can be reduced until $m+1-k$. Moreover, if the location of these points is known then their measures can easily be calculated:

$$p_i^* = |\Delta_i| / \sum_{j=1}^{n} |\Delta_j| , \tag{9}$$

$$\Delta_i = \det \{\delta\eta(x,\vartheta)/\delta\vartheta\ ;\ \delta\psi/\delta\vartheta\}_{x=x_i^*,\vartheta=\vartheta^*},$$

where the existence of the corresponding derivatives and the regularity of the matrices are assumed.

Example 1. Let $\eta_1(x,\vartheta_1) = \sum_{\alpha=1}^{m} \vartheta_{1\alpha} x^{\alpha-1}$ and $\eta_2(x,\vartheta_2) = \sum_{\alpha=1}^{m+1} \vartheta_{2\alpha} x^{\alpha-1}$. Assume that $F(z)=z^2$, $X=[-1,1]$ and there are no other constraints except that $\vartheta_{2(m+1)} = \delta > O$. In terms of (8) it means that $\eta(x,\vartheta) = \sum_{\alpha=1}^{m+1} \vartheta_\alpha x^{\alpha-1}$ and $\Omega = \{\vartheta : \vartheta_{m+1} = \delta > O\}$.

The supporting set of the Tchebysheff problem

$$\inf_{\vartheta} \sup_{|x|\le 1} \left| \sum_{\alpha=1}^{m+1} \vartheta_\alpha x^{\alpha-1} \right|$$

is known (see, for instance, Karlin and Studden, 1966):

$$X^* = \{x_i^* = \cos \frac{m+1-i}{m}\pi\ ,\ i=\overline{1,m+1}\}.$$

The corresponding measures can be calculated with the help of (9): $p_1^* = p_{m+1}^* = 1/2m$, $p_2^* = \cdots = p_m^* = 1/m$.

5. DUALITY OF SOME MODEL TESTING AND PARAMETER ESTIMATION PROBLEMS

In this section only the linear case when $\eta(x,\vartheta) = \vartheta^T f(x)$ and $F(z) = z^2$ will be considered and all results will formulated in terms of (8).

Let us start with the most evident and simple case when one is interested in some linear combination $c^T\vartheta$ of unknown parameters. For interpolation or extrapolation, $c = f(x_0)$, where x_0 is the point of interest. Then if one wants to estimate $c^T\vartheta$, the following criterion (see, for instance, Fedorov, 1972; Silvey, 1980) can be used:

$$\Psi(\xi) = c^T M^-(\xi) c\ , \tag{10}$$

where the superscript means pseudo-inversion, and $M(\xi) = \int_X f(x) f^T(x) \xi(dx)$. If the model $\eta(x,\vartheta)$ is tested under the constraint $(c^T\vartheta)^2 \ge 1$ then:

$$T(\xi) = \inf_{(c^T\vartheta)^2 \ge 1} \int_X \eta^2(x,\vartheta)\xi(dx) = \inf_{(c^T\vartheta)^2 \ge 1} \vartheta^T M(\xi)\vartheta\ . \tag{11}$$

It is easy to check that in (11), instead of 1, any positive constant can be taken without influencing the optimal design if $\eta(x,\vartheta)$ depends linearly on ϑ. A similar result holds for the criteria considered below and it will be used without comment.

It is natural to suggest that

$$\{\xi : c^T M^-(\xi) c < \infty\} = \phi$$

for any type of pseudo-inverse matrix, or in other words, we assume that $c^T\vartheta$ is estimable in the experiments defined by ξ. The necessary and sufficient condition for the estimability of $c^T\vartheta$ is

$$c^T \{I - M^-(\xi) M(\xi)\} = 0 \tag{12}$$

Designs satisfying (12) will be called regular.

It is obvious that all optimal designs ξ^* for (11) coincide with the optimal designs for the simpler problem

$$\inf_{c^T\vartheta=1} \vartheta^T M(\xi)\vartheta . \tag{13}$$

Taking into account the condition (12) and using the standard Lagrangian technique, we get

$$\inf_{c^T\vartheta} \vartheta^T M(\xi)\vartheta = \{c^T M^-(\xi)c\}^{-1} \tag{14}$$

with $\vartheta^* = M^{-1}(\xi)c$. From the last equation, it immediately follows that regular optimal designs, are the same for both criteria (10) and (11), more details see in Fedorov, Khabarov, 1986. If there is some prior information on the parameters ϑ_t described by a prior distribution function, $\mu_0(d\vartheta)$, then it is reasonable to use the mean of the noncentrality parameter as a criterion of optimality

$$T_0(\xi) = \int_\Omega \int_x \eta^2(x,\vartheta)\xi(dx)\mu_0(d\vartheta) .$$

If the distribution $\mu_0(d\vartheta)$ has a dispersion matrix D_0 , then

$$T_0(\xi) = tr\{D_0 M(\xi)\} .$$

In practice, knowledge of D_0 is problematic and one can relax this demand and assume only that the determinant of the dispersion matrix has a value d greater than zero. In this case, the criterion

$$T_0(\xi) = \inf_{|D_0|\geq d} tr\{D_0 M(\xi)\}$$

can be the form of interest. If the matrix $M(\xi)$ is nonsingular, then

$$T_0(\xi) = md^{1/m} |M(\xi)|^{1/m} . \tag{15}$$

Evidently the maximization of (15) is equivalent to the maximization of $|M(\xi)|$. This criterion is one of the most widely used criteria in the estimation problem. Some properties of D-optimal designs connected with model testing were discussed by Kiefer (1958) and Stone (1958). The above result gives additional explanation of the relation between the D-criterion and the model testing problem.

Let us now consider a very natural criterion for the model testing problem,

$$\Phi(\xi) = \inf_{\sup_{x\in U}|\vartheta^T q(x)|^2\geq 1} \int_x \eta^2(x,\vartheta)\xi(dx) = \inf_{\sup_{x\in U}|\vartheta^T q(x)|^2\geq 1} \vartheta^T M(\xi)\vartheta , \tag{16}$$

for any function q defined on U .

It is not difficult to check the chain of equalities:

$$\inf_{\sup_{x\in U}|\vartheta^T q(x)|^2\geq 1} \vartheta^T M(\xi)\vartheta = \inf_{x\in U} \inf_{|\vartheta^T q(x)|^2\geq 1} \vartheta M(\xi)\vartheta$$

$$= \inf_{x\in U} \{q^T(x)M^-(\xi)q(x)\}^{-1} ,$$

where, of course, a design ξ has to be regular for any $c = q(x), x \in U$.

In most cases, the requirement of regularity causes the nonsingularity of $M(\xi)$. This happens, for instance, when $U = X$.

The criterion

$$\Psi(\xi) = \sup_{x \in U} q^T(x) M^-(\xi) q(x)$$

belongs to the family of g-criteria (Ermakov, 1983). When $U = X$ and $q(x) = f(x)$, one can bet an even stronger result because the criteria

$$|M(\xi)|^{-1}\ , \text{ and } \sup_{x \in U} f^T(x) M^{-1}(\xi) f(x)$$

are equivalent in the case of continuous designs, a result which follows from Kiefer & Wolfowitz's theorem. This leads immediately to the equivalence of (16) and D-criteria.

The equivalence of some criteria can be proved with the help of the well-known result on eigenvalues of matrices (compare with Jones, Mitchell, 1978). Let M be a symmetric matrix and let $C = BB^T$ be a positive-definite matrix. If $\lambda_1 \geq \cdots \geq \lambda_m$ are the roots of $|M - \lambda C| = 0$ then

$$\inf_{\vartheta} \frac{\vartheta^T M \vartheta}{\vartheta^T C \vartheta} = \lambda_m\ .$$

From this relation, the equivalence of the following two criteria immediately occurs

$$\Psi(\xi) = \lambda_m^{-1}(\xi)\ ,\quad T(\xi) = \inf_{\vartheta^T C \vartheta \geq 1} \int \eta^2(x, \vartheta) \xi(dx)\ .$$

When $C = I_m$, then $\Psi(\xi)$ is the popular E-criterion of design theory.

The results can be summarized in the following theorem.

Theorem 4. The following criteria are equivalent on the set of regular designs:

(i) $c^T M^-(\xi) c$ *and* $\inf_{(c^T\vartheta)^2 \geq \delta} \gamma(\xi, \vartheta)$;

(ii) $|M^{-1}(\xi)|$ *and* $\inf_{|D_0| \geq d} \int \gamma(\xi, \vartheta) \mu^0(d\vartheta)$;

(iii) $\sup_{x \in U} q^T(x) M^-(\xi) q(x)$ *and* $\inf_{\sup_{x \in U} |q^T(x)\vartheta|^2 \geq \delta} \gamma(\xi, \vartheta)$;

(iv) $\lambda_1\{B^T M^{-1}(\xi) B\}$ *and* $\inf_{\vartheta^T BB^T \vartheta \geq \delta} \lambda(\xi, \vartheta)$,

where $\delta > 0$ *and* $\gamma(\xi, \vartheta) = \int \{\vartheta^T f(x)\}^2 \xi(dx)$, *with the integral over the range* X .

The requirement of regularity is essential for (i) and (iii) of the theorem. In other cases for optimal designs the existence of the inverse matrix $M^{-1}(\xi)$ is evident.

The theorem is true both for discrete and continuous designs. But the equivalence of (ii) and (iii) is based on Kiefer & Wolfowitz's theorem which is true only for continuous designs.

6. NUMERICAL PROCEDURES

Theorem 2 gives possibility to construct T-optimal designs with the help of the algorithms developed for the parameter estimation problem. These algorithms were discussed repeatedly in the statistical literature (see, for instance, Fedorov, 1972; Ermakov, 1983). Therefore only the algorithms specially oriented to the model testing problem will be considered in this section. For the sake of simplicity they will be formulated for design problem (8) and we start with the algorithm, which is a generalized version of that proposed by Atkinson, Fedorov (1975).

This algorithm is based on the results of Theorem 1 and belongs to the family of steepest descent algorithms.

To avoid difficulties related to singularities in optimization problems

$$\inf_{\vartheta\in\Omega} \int F\{\eta(x,\vartheta)\}(dx) \tag{17}$$

we assume that for T-optimal design (17) has unique solution $\vartheta(\xi^*)$.

In practice, this assumption is not very restrictive because instead of (8) one can apply to the regularized version of it

$$\xi_0^* = Arg\ \sup_{\xi} T\ \{(1-\rho)\xi + \rho\xi_0\},\quad 1>\rho>0\ , \tag{18}$$

where ξ_0 is any design providing uniqueness of $\vartheta(\xi_0)$. Due to concavity of $T(\xi)$ (compare with Fedorov, Uspensky, 1975):

$$T(\xi^*) \geq T(\xi_0^*) \geq (1-\rho)T(\xi^*) + \rho T(\xi_0)\ .$$

(i) Let the design ξ_s was constructed at the previous iteration

$$x_{s+1} = Arg\ \max\{\sup_{x\in X} \varphi(x,\xi_1)\ ,\ -\inf_{x\in X_s} \varphi(x,\xi_s)\}\ ,$$

where $\varphi(x,\xi_s) = F\{\eta(x,\vartheta_s)\} - T(\xi_s)$, X_s is the supporting set of ξ_s and

$$\vartheta_s = Arg\ \inf_{\vartheta\in\Omega} \int_x F\{\eta(x,\vartheta)\}\xi_s(dx)$$

(ii) $\xi_{s+1} = (1-\gamma_s)\xi_s + \gamma_1\xi(x_s)$, where $\gamma_s = \alpha_s$ if $\sup_{x\in X} \varphi(x_{s+1},\xi_s) \geq -\inf_{x\in X_s} \varphi(x_{s+1},\xi_s)$, and $\gamma_s = -\max\{\alpha_s, p_{s+1}/(1-p_{s+1})\}$ otherwise, p_{s+1} is the measure of point x_{s+1} prescribed by ξ_s. The sequence $\{\alpha_s\}$ providing convergency of (i), (ii) can be chosen similarly to parameter estimation case (see, for instance, Ermakov, 1983; Denisov, Fedorov, Khabarov, 1981):

$$\alpha_s \to 0\ ,\ \sum \alpha_s = \infty\ ,\ \sum \alpha_s^2 < \infty\ .$$

If for given $\eta(x,\vartheta)$ and $\varphi(\vartheta)$ formula (9) is admissible for computing then instead of (ii) one can use this formula chosing from the sth-design $(m-k)$ supporting points with largest values of $\varphi(x,\xi_s)$. Together with x_{s+1} they will form a supporting set for ξ_{s+1}. This modification of the iterative procedure converges to an optimal design containing non more than $m+1-k$ supporting points and is very close to the Remez algorithm for the Tchebysheff best approximation problem (for details, see Demjanov, Malozemov, 1966; Denisov, Fedorov, Khabarov, 1981).

7. SEQUENTIAL DESIGN

Application to (4) and (5) makes it clear that iterative procedure (i), (ii) can be used in practice for the construction of maximin designs or locally optimal designs for given ϑ_{1t} . The latter design can be useful for the clarifying of general structure of T-optimal design. To be more specific, one can use some sequential design procedures which were repeatedly discussed by different authors (see, for instance, Atkinson, Fedorov, 1975; Atkinson, 1978).

The simplest sequential procedure is the following one:

(i) After N measurements one has to calculate

$$\hat{\vartheta}_{jn} = Arg \inf_{\vartheta_j \in \Omega_j} \sum_{i=1}^{N} F\{y_i - \eta_j(x_i, \vartheta_j)\} .$$

(ii) The $(N+1)$-th measurement has to be done at the point:

$$x_{N+1} = \sup_{x \in X} F\{\eta_1(x_1 \hat{\vartheta}_{1N}) - \eta_2(x_1 \hat{\vartheta}_{2N})\} .$$

This sequential procedure has its roots in iterative procedure (i), (ii) from the previous section. The similarity will be more evident if one put $\eta(x,\vartheta) = \eta_1(x,\vartheta_1) - \eta_2(x,\vartheta_2)$, $\gamma_N = (N_0+N)^{-1}$, where N_0 is a number of measurements in an initial experiment. Naturally the deletion of "bad" points permissible in the iterative procedure has no sense for the sequential design.

Some numerical examples illuminating the efficiency of sequential procedure (i), (ii) were discussed by Box, Hill, 1967; Fedorov, 1972; Atkinson, 1978. The weak convergency:

$$\lim_{N \to \infty} N^{-1} \sum_{i=1}^{N} F\{\eta_1(x_i, \hat{\vartheta}_{1N}) - \eta_2(x_i, \hat{\vartheta}_{2N})\} =$$

$$= \sup_{\xi} \inf_{\vartheta_{j^*}} \int_X F\{\eta_t(x,\vartheta) - \eta_{j^*}(x,\vartheta_{j^*})\} \xi(dx) ,$$

where $\eta_{j^*}(x,\vartheta)$ is a "wrong" model, follows from the convergency of iterative procedure (i), (ii) if one manages to prove that for the true model the parameter estimators are consistent for the sequence $\{\xi_N\}$. The consistency can be assured by application to regularization (18).

8. CONCLUSIONS

The results presented in this paper (based on formal, mathematical techniques) confirm the validity of the following simple, intuitive idea:

"Observing stations should be located at sites where the discrepancy between competing models is greatest".

Indeed, in case of two competing models $\eta_1(x,\vartheta_2)$ and $\eta_2(x,\vartheta_2)$. Theorems 1 and 3 lead to the recommendation that observing stations should be located at points where the function

$$F(x) = |\eta_1(x,\vartheta_1^*) - \eta_2(x_1 \vartheta_2^*)|$$

approaches its upper bound for the (in the model testing sense) worst values of parameters ϑ_1 and ϑ_2.

The same idea can be treated in numerical procedure (i), (ii) of Section 6 and the sequential methodology of experimental design.

In the first case, at every s-th step one has to relocate a possible point of observation from an area where the discrepancy $\eta_1(x, \vartheta_s) - \eta_2(x, \vartheta_s)$ is small, to an an area where it has its largest value.

In the sequential design, every new observation has to be located at a point where the current measure of discrepancy is largest (see (i), (ii), Section 7). It is evident that, to some extent, similar sequential procedures are used regularly in operational practice. Here, statistical theory provides a reasonable (from a statistical point of view) criteria of optimality, necessary formulae for calculations, and (this seems a most useful result) global optimality of the procedure: sequential designs generated by (i), (ii) converge to a design which is optimal in the sense of (3), (4).

Section 5 confirms the common feeling amongst practitioners that the problems of model testing and parameters estimation are essentially overlapping. If one can efficiently estimate the most characteristic parameters for competing models, then model discrimination can be performed appropriately.

REFERENCES

Atkinson, A.C. (1978). Posterior probabilities for choosing a regression model. *Biometrika 65*, 39-48.

Atkinson, A.C. & Fedorov, V.V. (1975). The design of experiments for discriminating between two rival models. *Biometrika 62*, 57-70.

Box, G.E.P & Hill, W.J. (1967). Discrimination among mechanistic models. *Technometrics 9*, 57-71.

Denisov, V., Fedorov, V. & Khabarov, V. (1981). Tchebysheff's approximation in the problem of asymptotical locally-optimal designs construction for discriminating experiments. In *Linear and Nonlinear Parameterization in Experimental Design Problems, Problems in Cybernetics*, Ed. V. Fedorov and V. Nalimov, pp. 3-10 (in Russian). Moscow.

Ermakov, S.M., Ed. (1983). *Mathematical theory of experimental design*. Nauka, Moscow (in Russian).

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

Fedorov, V.V. & Uspensky, A.B. (1975). *Numerical aspects of least squares methods*. Moscow State University, Moscow.

Fedorov, V.V. (1980). Convex design theory. *Math. Oper. Statist., Ser. Statist. 11*, 403-13.

Fedorov, V.V. (1981a). Design of model testing experiments. In *Symposia Mathematica*, pp. 171-80. Bologna: Monograph.

Fox, D.G. (1981). Judging air quality model performance. *Bulletin American Meteorological Society 62*, 599-608.

Hayes S.R. and Moore, G.E. (1986). Air quality model performance: a comparative analysis of 15 model evaluation studies. *Atmospheric Environment, 20*, 1897-1911.

Huber, P.J. (1981). *Robust statistics*. Wiley & Sons, N.Y.

Jones, E.R. & Mitchel, T.J. (1978). Design criteria for detecting model inadequacy. *Biometrika 65*, 541-51.

Karlin, S. & Studden, W.J. (1966). *Tchebysheff systems: with application in analysis and statistics*. Wiley & Sons, N.Y.

Kiefer, J. (1958). On the nonrandomized optimality and randomized nonoptimality of symmetrical designs. *Ann. Math. Statist.* *29*, 675-99.

Munn, R.E. (1981). The design of air quality monitoring networks, MacMillan Publishers Ltd., London, p.109.

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

Stone, M. (1958). Application of a measure of information to the design and comparison of regression experiments. *Ann. Math. Statist.* *29*, 55-70.

Willmott, C.J. (1982). Some comments on the evaluation of the model performance. *Bulletin American Meteorological Society*, *63*, 1309-1313.