

**OPTIMAL DESIGN OF EXPERIMENTS WHEN FACTORS
AFFECT DETECTION CAPABILITY**

Natalie Zayats¹ and David M. Steinberg²

¹ G-Stat, Tel Aviv, Israel

Email: natalie.zayats@g-stat.com

² Department of Statistics & Operations Research, Tel Aviv University, Israel

Tel Aviv 69978, Israel.

Email: dms@post.tau.ac.il

ABSTRACT

The configuration of a monitoring network affects the quality of information that is obtained. An important aspect is that some stations in a network may fail to provide data at all. In this work we study the important question of how to take account of detection capability in network configuration. Our approach is based on the theory of optimal design of experiments, extended to settings where detection must be considered. We illustrate our ideas by considering how to configure a seismic network that will provide precise estimates of the hypocenters of seismic events. An important consideration is that stations that are too far from the hypocenter may fail to detect the event. Our results provide a theoretical basis and practical guideline for configuring such a network.

KEYWORDS

D-optimal design; general equivalence theorem; network configuration; seismic monitoring.

2000 Mathematics Subject Classification: 62J02; 62K05.

1 Optimal Design and Detection Capability

In the five decades that have passed since the pioneering work of Kiefer and Wolfowitz (1960), the statistical theory of optimal design of experiments has been thoroughly explored and applied to many real experiments. The book by Silvey (1980) is an excellent short summary of the key ideas, that by Pukelsheim (1993) is both deep and general in coverage and the book by Atkinson, Donev and Tobias (2007) covers both theory and practical application.

Despite the large amount of research on the optimal design of experiments, the problem that we address here has attracted little previous attention. We are concerned with the design of monitoring networks in which the spatial configuration of the network affects the precision of the results in two different ways. First, it determines a regression matrix; and second, it influences detection capability, i.e. the probability that usable data will be obtained at all.

Our inquiry was motivated by the problem of configuring a seismic monitoring network, but the basic ideas that we develop are relevant to a wide variety of monitoring problems including sensor networks and environmental monitoring systems. Seismic networks are used to monitor events such as earthquakes and volcanoes. Data from the network are used to identify the occurrence of an event and to determine the source location (hypocenter). The precision of the location estimates is highly dependent on the network configuration. Moreover, a station that is too far from the hypocenter may not get a clear signal and could, therefore, provide no information at all. Thus practical design of the network must account for the detection capability.

We adopt the approach of Imhof, Song and Wong (2002, 2004), who proposed an extension of the statistical theory of optimal design to settings where detection capability is important. We exploit the theory of design measures, introduced by Kiefer and Wolfowitz (1960), to characterize D -optimal monitoring networks and thereby to obtain bounds on the efficiency of any proposed network. We then apply the theory to derive seismic network configurations that maximize the precision with which the location of earthquakes is determined by a seismic network, taking account of the fact that distant stations may not provide data.

A brief review of research on the optimal configuration of seismic networks will help to focus the goals of this paper. Kijko (1977) first proposed using the D -optimality criterion to design a seismic network and described an application using the conjugate gradient optimization algorithm. Rabinowitz and Steinberg (1990) used Mitchell's (1974) DETMAX

algorithm to find D -optimal networks for monitoring a specific source location. Later, Steinberg, Rabinowitz, Shimshoni, and Mizrachi (1995) proposed and studied a criterion for optimal network configuration for monitoring a system of faults. Steinberg and Rabinowitz (2003) gave a detailed theoretical characterization of D -optimal seismic networks for a single source.

The above studies assumed that all stations in the network will record primary phase arrivals from all events. In fact, though, most of the events detected by a local seismic network are in the micro-earthquake range (about magnitude 1.5-3 on the Richter scale) and data will typically be limited to stations that are close to the hypocenter. The same problem occurs for global networks, but with sparser station deployment and, consequently, a higher magnitude range. For example, the International Monitoring System uses a global seismic network comprised of about 50 primary stations to monitor compliance with the Comprehensive Nuclear-Test-Ban Treaty. That network has only limited detection capability for moderate earthquakes (magnitude 3-4.5 on the Richter scale) and for small-scale nuclear explosions that might be set off in violation of the Treaty.

Our focus in this work is on applying a suitable design criterion that reflects the detection capability of the network configuration. We begin with a review of optimal design theory and a summary of prior research on network configuration. Then we present the criterion, present a theoretical characterization of optimal network design accounting for possible non-detects and apply it to a configuration problem based on data from the Israel Seismic Network.

2 Review of Optimal Design Theory

We use this section to recount some important ideas from the theory of optimal experimental design. In the next section we describe the extension these ideas to settings where detection capability is also an important consideration. Our presentation in this section largely follows that in Silvey (1980).

Denote by \tilde{Y} an outcome variable whose distribution depends on controllable variables $u = (u_1, u_2, \dots, u_r)^T$; a vector of parameters $\theta = (\theta_1, \theta_2, \dots, \theta_k)^T$ of interest to the experimenter; and a vector $\tau = (\tau_1, \tau_2, \dots, \tau_l)^T$ of “nuisance” parameters, which are not of primary interest.

Vectors u can be chosen from a given set U in R^r and the true θ is known to belong to a set Θ in R^k . For given u , θ , and τ the distribution of \tilde{Y} is determined by a probability density function $p(y|u, \theta, \tau)$.

2.1 Fisher Information Matrix

The Fisher information matrix for θ is the standard tool for quantifying the efficiency of an experiment. For a single observation on \tilde{Y} at the vector u , the partitioned Fisher information matrix for θ and τ is

$$J(u; \theta, \tau) = \begin{pmatrix} J_{\theta\theta}(u) & J_{\theta\tau}(u) \\ J_{\theta\tau}^T(u) & J_{\tau\tau}(u) \end{pmatrix}$$

where $J_{\theta\theta}(u)$ is the $k \times k$ matrix whose (i,j) 'th element is

$$E \{ -\partial^2 \log p(y|u, \theta, \tau) / \partial\theta_i \partial\theta_j \}$$

The remaining matrices in $J(u; \theta, \tau)$ are defined similarly, using the second derivatives with respect to the nuisance parameters and the crossed derivatives with respect to one parameter of interest and one nuisance parameter. The information from N independent observations is

$$J(\underline{u}; \theta, \tau) = \sum_{i=1}^N J(u_i; \tau, \theta) = \begin{pmatrix} J_{\theta\theta}(\underline{u}) & J_{\theta\tau}(\underline{u}) \\ J_{\theta\tau}^T(\underline{u}) & J_{\tau\tau}(\underline{u}) \end{pmatrix}$$

Typically the experimental design will be chosen so that $J(u; \theta, \tau)$ is non-singular, in which case the inverse matrix $J^{-1}(\underline{u}; \theta, \tau)$ provides a lower bound for the variance matrix of an unbiased estimator of (θ, τ) and its leading $k \times k$ sub-matrix

$$V(\underline{u}; \theta, \tau) = \{ J_{\theta\theta}(\underline{u}) - J_{\theta\tau}(\underline{u}) J_{\tau\tau}^{-1}(\underline{u}) J_{\theta\tau}^T(\underline{u}) \}^{-1}$$

is a lower bound for the covariance matrix of an unbiased estimator of θ . If N is large, the variance matrix of the maximum likelihood estimator $\hat{\theta}$ is approximately $V(\underline{u}; \theta, \tau)$. The asymptotic variance matrix depends on \underline{u} , which implies that we can improve the inference by intelligent design of the experiment.

2.2 Linear and Nonlinear Model Theory

We review here results on how to use the ideas above to design an experiment for a linear or a nonlinear statistical model. We consider here problems where the following assumptions hold.

1. $E(\tilde{Y} | u, \theta, \tau) = f(u, \theta)$.
2. $\text{var}(\tilde{Y} | u, \tau, \theta) = \tau$.

3. The observations have a normal distribution and are independent.

We have a linear statistical model if the expected value has the form

$$E(\tilde{Y} | u, \theta, \tau) = \theta_1 f_1(u) + \theta_2 f_2(u) + \dots + \theta_k f_k(u) = f^T \theta$$

where $f_j(u)$ is a known function. Often $f_1(u) = 1$, corresponding to the intercept in the linear model. We have a nonlinear statistical model when $\partial f(u, \theta) / \partial \theta$ is a function of θ .

We assume that the variance is constant and equals τ . A more general description would allow the variance to depend on the design site, $\text{var}(\tilde{Y} | u, \theta, \tau) = \tau v(u)$ for v a known function of u . However this case can immediately be reduced to $\text{var}(\tilde{Y} | u, \theta, \tau) = \tau$ by considering, instead of \tilde{Y} , $\tilde{Y}' = \tilde{Y} / \{v(u)\}^{1/2}$ and instead of $f(u, \theta)$, $f'(u, \theta) = f(u, \theta) / \{v(u)\}^{1/2}$. So there is no essential loss of generality.

The Fisher information matrix for θ , with a linear model, is $(1/\tau) \sum f_i f_i^T$ and the covariance matrix of the least-squares estimator of θ is $V(\underline{u}; \tau) = \tau (\sum f_i f_i^T)^{-1}$. For nonlinear models, the Fisher information matrix has an approximate form that is analogous to the formula for linear models. Taking a first order Taylor series approximation to the true model leads to the formula $V^{-1}(\underline{u}; \theta, \tau) = (1/\tau) \sum f_i f_i^T$ where $f_i = \partial f(u_i, \theta) / \partial \theta$. See Gotwalt, Jones and Steinberg (2009) for details. For a nonlinear model, f_i depends on θ , so the information matrix depends not just on the design, but also on the true parameter value θ . Thus a design may be ideally suited to one value of θ but very inefficient for another one. This led Chernoff (1953) to suggest the notion of a ‘‘locally optimal’’ design, i.e. a design that is optimal for a particular θ . Most of our subsequent results will relate to the problem of maximizing Fisher information for θ in the local sense of Chernoff, where the special value of θ has been guessed in advance or involves a site that requires particular attention.

In the seismic monitoring setting that we will discuss later, the data are the arrival times of seismic phases at the stations and the controllable variables are the locations of the stations. The parameters of interest are the location of the source and the origin time of the event and τ is the variance associated with determining the seismic phase arrival times. A nonlinear model, which we will describe in Section 4.1, links the data to the source parameters.

2.3 Design Criteria and D-optimality

For linear statistical models, we can maximize information about θ if we choose N vectors $u_i \in U$, $i = 1 \dots N$ from the design space to make $M(\underline{u}) = \sum f_i f_i^T$ large in some sense. For nonlinear models, with a prescribed value of θ , the same basic problem is encountered,

with f_i the vector of partial derivatives. We now consider various ways in which we might wish to make $M(\underline{u})$ large.

It is known that there is little point in seeking a design \underline{u}_* which is optimal in the very strong sense that $M(\underline{u}_*) - M(\underline{u})$ is non-negative definite for all \underline{u} . Such a \underline{u} is likely to exist only in exceptional circumstances. Instead we attempt to find a design \underline{u}_* that maximizes some real-valued function $\phi\{M(\underline{u})\}$. Various functions have been suggested by practical considerations, leading to criteria such as D -optimality, G -optimality, E -optimality etc. In our research we will consider only the D -optimality criterion, $\phi(M(\underline{u})) = \log[\det\{M(\underline{u})\}]$, the most frequently used of all design criteria. Taking the logarithm of the determinant is a standard convention. As the log is a monotone function, the criterion ranks designs in the same way with or without the log, and $\log \det$ is concave, a desirable property for finding and characterizing optimal designs.

For nonlinear models, the dependence of $M(\underline{u})$ on θ can also be given a Bayesian solution. The idea is to assign a prior density $\pi(\theta)$ and then to use the Bayesian D -optimality criterion, $\int \log[\det M(\underline{u})] \pi(\theta) d\theta$. Averaging across a range of plausible θ values leads to designs that are robust with respect to uncertain knowledge of θ . See Chaloner and Verdinelli (1995) and Gotwalt, Jones and Steinberg (2009). For seismic monitoring, the multi-source criterion of Steinberg et al. (1990) has the same structure, but limits the prior to a finite discrete distribution with just a few support points.

2.4 Design Measure

Finding an optimal experimental design is by nature a discrete problem – what are the N design sites that give the best information matrix? However, Kiefer and Wolfowitz (1960) showed that theoretical results on optimal designs could sometimes be found more easily by replacing the discrete problem with a much more general problem, in which the design is regarded as a probability measure. Discrete designs in this formulation are a special class of measures with finite support and with mass equal to $1/N$ at each design point.

We assume here that design points u must be selected from a compact set U , which constitutes the design space. This assumption is practically realistic because typical design problems will involve a compact U . Let H be a class of probability distributions on the Borel sets of U . Any $\eta \in H$ will be called a design measure and we define its information matrix by

$$M(\eta) = E \{f(\tilde{u})f(\tilde{u})^T\}$$

where \tilde{u} is a random vector with distribution η .

Note that because U is compact, $M(\eta)$ exists for all $\eta \in H$. Let $\mathcal{M} = \{M(\eta) : \eta \in H\}$.

Suppose that ϕ is a real-valued function defined on the $k \times k$ symmetric matrices and bounded above on \mathcal{M} , though possibly not below. The continuous version of the design problem is to determine η_* to maximize $\phi\{M(\eta)\}$ over H . Any such η_* will be termed ϕ -optimal.

We now note some particular aspects of this problem.

2.4.1 The set \mathcal{M}

1. Each element of \mathcal{M} is a symmetric non-negative definite $k \times k$ matrix which can be represented by a point in $\mathfrak{R}^{(1/2)k(k+1)}$.
2. The set \mathcal{M} is convex. Indeed it is the closed convex hull of $\{f(u)f(u)^T : u \in U\}$.
3. By Caratheodory's Theorem, each element of \mathcal{M} can be expressed as a convex combination $\sum_{i=1}^I \lambda_i f_i f_i^T$, where $f_i = f(u_i)$, $u_i \in U$, $i = 1, \dots, I$ and $I \leq \frac{1}{2}k(k+1) + 1$. The important practical consequence is that if ϕ is maximal at M_* then M_* can always be expressed as $M(\eta_*)$, where η_* is a design measure supported on at most $\frac{1}{2}k(k+1) + 1$ points; that is, there always exists a discrete η_* which solves the continuous theory problem. Our aim is to approximate η_* by a design measure corresponding to an N -observation design and this is clearly going to be easier if η_* is discrete.

2.4.2 Fréchet Directional Derivative

The Fréchet directional derivative is a useful tool for exploring the D -optimality criterion. Given a design measure with information matrix M_1 , we consider what happens if we make a small modification in the direction of a design measure with information matrix M_2 . The Fréchet derivative quantifies the effect on the D -criterion.

The Fréchet derivative of ϕ at M_1 in the direction of M_2 is:

$$F_{\phi}(M_1, M_2) = \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} (\phi\{(1 - \varepsilon)M_1 + \varepsilon M_2\} - \phi(M_1))$$

Note that:

1. $M_1, M_2 \in \mathcal{M}$ implies that $(1 - \varepsilon)M_1 + \varepsilon M_2$ is an element of \mathcal{M} and so $\phi\{(1 - \varepsilon)M_1 + \varepsilon M_2\}$ is automatically defined.

2. Concavity of ϕ implies that $\frac{1}{\varepsilon}(\phi\{(1-\varepsilon)M_1 + \varepsilon M_2\} - \phi(M_1))$ is a non-increasing function of ε in $0 \leq \varepsilon \leq 1$. Hence when ϕ is concave, $F_\phi(M_1, M_2)$ exists if we allow the value $+\infty$.
3. Applying the last property with $\varepsilon = 1$, we have the result that $F_\phi(M_1, M_2) \geq \phi(M_2) - \phi(M_1)$.

Differentiability of ϕ implies that, if $\sum a_i = 1$,

$$F_\phi(M_1, \sum a_i M_i) = \sum a_i F_\phi(M_1, M_i)$$

If $\tilde{M} \in \mathcal{M}$ is a random matrix, ϕ is differentiable and E denotes expected value, we have by this linearity

$$E\{F_\phi(M_1, \tilde{M})\} = F_\phi(M_1, E(\tilde{M}))$$

2.5 The General Equivalence Theorem

With these preliminaries we are able to state the General Equivalence Theorem (GET) of Kiefer and Wolfowitz (1960), which both characterizes D -optimal design measures and gives useful information for constructing them. We state both theorems below in a general format that holds for any concave optimality function ϕ and, in particular, for the D -optimality criterion.

Theorem 1. *When ϕ is concave on \mathcal{M} , η_* is ϕ -optimal if and only if*

$$F_\phi\{M(\eta_*), M(\eta)\} \leq 0 \quad \text{for all } \eta \in H$$

This means that we are on top of a concave mountain when there is no direction in which we can look upward to another point of the mountain. However, the theorem has limited practical value as there are many directions in which we must look to determine whether or not the condition it contains is satisfied. The following theorem condenses these conditions to a much more useful one.

Theorem 2. *If ϕ is concave on \mathcal{M} and differentiable at $M(\eta_*)$ then η_* is ϕ -optimal if and only if $F_\phi\{M(\eta_*), f(u)f(u)^T\} \leq 0$ for all $u \in U$*

This is the key theorem of ϕ -optimal linear regression design theory because it provides a condition which can often be verified quite easily in practice. The GET (Kiefer & Wolfowitz, 1960) refers to the particular version of Theorem 2 obtained by taking $\phi = \log \det$. White (1973) extended the GET from linear to nonlinear models.

3 Optimal Design with Non-detects

We follow the approach of Imhof, Song and Wong (2002, 2004) and Baek, Zhu, Wu and Wong (2006) for extending the theory of D -optimal design (locally for nonlinear models) to problems where detection capability is an issue.

For a given experiment, let

$$I(u) = \begin{cases} 1 & \text{site } u \text{ provides data} \\ 0 & \text{otherwise} \end{cases}$$

be the indicator of getting data from a planned design point with the controllable factors set at u . We assume that $P(I_u = 1) = p(u)$ is a function of u , and possibly other relevant characteristics of the experiment. For example, in the seismic monitoring context, the detection probability depends on the hypocenter of the event θ , the station coordinates u and also on the event magnitude. Data from on going monitoring efforts are available and can be used to estimate $p(u)$.

The Fisher information for θ that is actually obtained in the experiment is

$$(1/\tau)M(\underline{u}) = (1/\tau) \sum f(u_i) f(u_i)^T I(u_i),$$

where $I(u_i)$ is the indicator for the i 'th design point, u_i .

There are a number of design criteria that might be used to reflect possible non-detects. For example, one might want to maximize the probability that enough data are collected to assure that all the parameters can be estimated. However, such a criterion may lead to data that provide very low precision for the parameter estimates. In the seismic monitoring context, stations close to the epicenter will have a much higher detection probability than those far from the origin; but the work of Steinberg and Rabinowitz (2003) shows that such networks provide little information about the location of the hypocenter.

Extensions of the D -optimality criterion to take account of possible non-detects include some balance between the goal of having enough data to estimate the parameters and getting precise estimates. Like Imhof, Song and Wong (2002), we advocate a version of D -optimality that first averages the Fisher information over the detection configurations and then applies the D -optimality functional. The alternative option is to compute $\det\{(M(\underline{u}))\}$ for each possible detection configuration and then to average with respect to the detection probabilities. Averaging the logs of the determinants is reminiscent of the Bayesian D -optimality criterion cited earlier for nonlinear models, but averaging over detection patterns rather than plausible parameter vectors. One common justification of the D -optimality

criterion is its relation to the size of a confidence ellipsoid for the parameters. With non-detects, the ellipsoid will be based on the data that are actually observed. The average size is then found by averaging the square root of the determinants. The drawback to computing a functional first and then averaging is that some detection patterns provide so few data that the observed matrix $M(\underline{u})$ is singular. Hence our preference for the first option.

Denoting by E_{ND} the expectation with respect to non-detection, the expected Fisher information of a design measure η , with respect to the detection distribution, is proportional to

$$M_{ND}(\eta) = E \{M(\eta)\} = E_{ND} E \{f(\tilde{u})f(\tilde{u})^T I_u\} = E \{f(\tilde{u})f(\tilde{u})^T P(I_u = 1)\} \quad (3.1)$$

where \tilde{u} is a random vector with distribution η . For a discrete N -point design \underline{u} , the corresponding expression is $M_{ND}(\underline{u}) = \sum f(u_i)f(u_i)^T p(u_i)$.

We can now use M_{ND} instead of M as the information matrix and can apply any optimality functional ϕ that was developed for use with full detection. In particular, the D -optimality criterion accounting for non-detects is

$$\log [\det \{M_{ND}(\eta)\}]. \quad (3.2)$$

Following common practice in optimal design theory, we define the relative efficiency of a design η_1 to a design η_2 by

$$RE(\eta_1, \eta_2) = \frac{\{\det [M_{ND}(\eta_1)]\}^{1/k}}{\{\det [M_{ND}(\eta_2)]\}^{1/k}}. \quad (3.3)$$

In some cases later in the paper, we will look at the same design, but use the full detection information matrix $M(\eta)$ in the denominator. This permits us to evaluate the reduction in assessed efficiency from taking account of incomplete detection.

We now extend Theorem 1 and Theorem 2 to settings with possible non-detection at each station. As in the earlier theorems, we need to consider the set of all possible expected information matrices $M_{ND}(\eta)$. We denote that set here by \mathcal{A} , to distinguish it from \mathcal{M} the set defined earlier assuming full detection. From the definition, it is immediate that \mathcal{A} is convex.

Theorem 3. *For a linear model, or a nonlinear model with interest on a fixed value of θ , and an optimality functional ϕ that is concave on \mathcal{A} , η_* is ϕ -optimal, accounting for non-detects, if and only if $F_\phi \{M_{ND}(\eta_*), M_{ND}(\eta)\} \leq 0$ for all $\eta \in H$ where H is a class of probability distributions on U .*

Proof. Necessity: ϕ maximal at $M_{ND}(\eta_*)$ implies

$$\phi\{(1-\varepsilon)M_{ND}(\eta_*) + \varepsilon M_{ND}(\eta)\} - \phi\{M_{ND}(\eta_*)\} \leq 0$$

for all ε in $[0, 1]$ and all $\eta \in H$. Since

$$(1-\varepsilon)M_{ND}(\eta_*) + \varepsilon M_{ND}(\eta) = M_{ND}\{(1-\varepsilon)\eta_* + \varepsilon\eta\},$$

this in turn implies, from the definition of F_ϕ , that

$$F_\phi\{M_{ND}(\eta_*), M_{ND}(\eta)\} \leq 0 \quad \text{for all } \eta \in H$$

Sufficiency: $F_\phi\{M_{ND}(\eta_*), M_{ND}(\eta)\} \leq 0$ for all $\eta \in H$, by Note 3.

$\phi\{M_{ND}(\eta)\} - \phi\{M_{ND}(\eta_*)\} \leq 0$ for all $\eta \in H$, that is η_* is ϕ -optimal.

Theorem 4 (the extension of Theorem 2). For a linear model, or a nonlinear model with interest on a fixed value of θ , and an optimality functional ϕ that is concave on \mathcal{A} and differentiable at $M_{ND}(\eta_*)$ then η_* is ϕ -optimal if and only if $F_\phi\{M_{ND}(\eta_*), gg^T\} \leq 0$ for all $g \in G$. Here U is a set of potential station sites and $G = \{g(u) : u \in U\}$ where $g(u) = f(u)\sqrt{p(u)}$, $f(u)$ is the vector of partial derivatives of the expected response at u with respect to θ and $p(u) = P(I(u) = 1)$ is the detection probability at u .

Proof. The necessity of the stated condition follows immediately from

Theorem 1: θ is fixed and ϕ is concave on \mathcal{A} implies

$$F_\phi\{M_{ND}(\eta_*), gg^T\} = F_\phi\{M_{ND}(\eta_*), J(\theta, u)\} \leq 0 \quad \text{for all } u \in U.$$

Sufficiency is proved as follows: Any $M_{ND}(\eta)$ can be expressed as a finite sum in the form

$$\begin{aligned} M_{ND}(\eta) &= \sum_{i=1}^N \lambda_i f_i f_i^T p(u_i) \\ &= \sum_{i=1}^N \lambda_i f_i \sqrt{p(u_i)} * f_i^T \sqrt{p(u_i)}^T \\ &= \sum_{i=1}^N \lambda_i g_i g_i^T \end{aligned} \tag{3.4}$$

where f_i is the vector of partial derivatives of the expected response at u_i and $g_i = f_i \sqrt{p(u_i)}$, $\lambda_i > 0$, $\sum_{i=1}^N \lambda_i = 1$. Then if ϕ is differentiable at $M_{ND}(\eta_*)$, by Note 4,

$$F_\phi\{M_{ND}(\eta_*), M_{ND}(\eta)\} = \sum_{i=1}^N \lambda_i F_\phi\{M_{ND}(\eta_*), g_i g_i^T\}$$

Hence

$$F_{\phi} \{M_{ND}(\eta_*), gg^T\} \leq 0 \quad \text{for all } g \in G$$

implies

$$F_{\phi} \{M_{ND}(\eta_*), M_{ND}(\eta)\} \leq 0 \quad \text{for all } \eta \in H$$

and the sufficiency now follows from Theorem 1.

Some further extensions are also possible. The non-detect adjusted information matrix $M_{ND}(\eta)$ for a design measure (or an N -point design) will typically depend on θ . This was already noted for nonlinear models due to the dependence of the partial derivatives on θ . Taking account of detection capability implies that this will also now be true of linear models if $p(u)$ depends on θ . Thus the Bayesian design criterion defined earlier, which averages across a prior distribution of plausible θ values, can also be applied here. The optimality results in Chaloner and Verdinelli (1995) for the Bayesian criterion will carry over to the problem here. As noted earlier, the detection probability may also depend on other aspects of the experiment and the Bayesian averaging can be extended to include those features as well.

4 Application to Seismic Monitoring

In this section we apply our ideas to the design of a seismic monitoring network that takes account of detection capability. The parameters of interest θ are the hypocenter (the spatial coordinates of the source) and the origin time. The controllable parameters u are the coordinates of the seismic stations. The estimation problem that we consider is to locate the hypocenter and determine the origin time. We make the common assumption that all stations are at the same height, so only the $X - Y$ coordinates need to be determined.

This setting provides a good framework for illustrating just how much a design may need to be modified as a result of limited detection. Steinberg and Rabinowitz (2003), assuming full detection, proved that locally optimal seismic networks for hypocenter location and origin time place one station right at the epicenter (the $X - Y$ coordinates of the source), and all the others on either one or two concentric circles centered at the epicenter. Moreover, one of those circles should be as far as possible from the epicenter. The latter conclusion is problematic, as detection should decrease with distance from the source.

4.1 Model for Seismic Location

In this section we provide a brief sketch of the physical model that relates the unknown hypocenter of a seismic event to the observed data and of the related questions of statistical inference. The interested reader can find detailed presentation in Steinberg and Rabinowitz (2003) or Aki and Richards (1980).

Earthquakes release energy that propagates through the earth in the form of seismic waves, which are recorded by seismographs. Based on the equations governing the propagation of elastic waves in a heterogeneous medium, the arrival time t_i at station i is modelled by

$$t_i = t_0 + g(P_i, P_0) + \varepsilon_i \quad (4.1)$$

where t_0 is the origin time, $P_0 = (X_0, Y_0, Z_0)$ is the hypocenter, $P_i = (X_i, Y_i, Z_i)$ is the location of the i th station, $g(P_i, P_0)$ gives the minimum travel time for a wave that originates at P_0 and arrives at P_i , and ε_i is an error term.

Throughout the remainder of this article, we will adopt the *layers above half-space* velocity model that is often used in routine analysis of local seismic data. In this model the earth's crust is divided into distinct vertical layers with velocity constant within each layer and increasing with depth. For such a model, waves can propagate from a hypocenter to a station along two types of ray paths: (1) direct waves travel up from the hypocenter and continue to rise until they reach the station; (2) refracted waves descend from the hypocenter to a deeper layer, travel along the top of the deeper layer, and then refract back up to the station. We follow the convention in seismology of making the positive Z -axis point downwards so that Z indicates depth. If the event occurs in the half-space (the lowest layer), only direct waves are possible. If the event occurs in a higher layer, it is necessary to compare travel times for a direct wave with those of all possible refraction waves.

The model derivatives play an important role in determining the precision of the estimates and thus also in design considerations. The derivatives are most easily expressed in terms of the initial ray angle and azimuth. The initial ray angle ϕ is the angle from the positive Z axis to the first segment of the ray path and the hypocenter. For a direct wave, ϕ is acute and is often called the takeoff angle; for a refracted wave, ϕ is obtuse. The azimuth α is the angle from the line $Y = Y_0$ to the ray from the epicenter (X_0, Y_0) through the station in the $X - Y$ plane. The partial derivatives of the travel times for an event that occurs in the j th crustal layer, where the velocity is v_j , are

$$\frac{\partial g}{\partial X_0} = -\frac{\cos \alpha \sin \phi}{v_j} \quad (4.2)$$

$$\frac{\partial g}{\partial Y_0} = -\frac{\sin \alpha \sin \phi}{v_j} \quad (4.3)$$

$$\frac{\partial g}{\partial Z_0} = \frac{\cos \phi}{v_j} \quad (4.4)$$

The arrival time derivatives with respect to the origin time t_0 are always equal to 1.

There are several types of waves associated with seismic events. The two types of waves most often detected by seismic networks are primary, or P -waves, which oscillate in the same direction as the ray path, and secondary, or S -waves, which oscillate in the wavefront plane that is orthogonal to the ray path. The wave equations imply that both types of waves propagate along the same path, but P -waves travel faster, and thus arrive sooner, than S -waves. As a result, it is often difficult to determine the onset time for the S -waves. Our subsequent results assume that only P -waves arrivals are available. Rabinowitz and Steinberg (2003) showed that essentially the same conclusions hold if both P and S -arrivals are available in the full detection setting.

4.2 Modeling Detection Probabilities

To apply the new design criterion one must have a reasonable model for the probability of detection. For the seismology problem of interest here, the detection probability depends primarily on the magnitude of the event and the distance from the hypocenter to the station. It can be modelled from existing seismic bulletin data with a generalized additive model (GAM) (Hastie and Tibshirani 1990). In this section we review briefly the GAM method and apply it to model the detection probability of seismic events in Israel.

GAM is a flexible statistical method that may be used to identify and characterize the effect of potential prognostic factors on an outcome variable. Generalized additive models extend the traditional linear statistical model. They can be applied in any setting where a linear or generalized linear model is typically used.

One of the most commonly used statistical models is the logistic regression model for binary data. Logistic regression models the effects of prognostic factors x_i via a linear predictor of the form $\sum x_i \beta_i$, where the β_i are parameters. The generalized additive model replaces $\sum x_i \beta_i$ with $\sum r_i(x_i)$ where r_i is an unspecified non-parametric function. This function is estimated in a flexible manner using a scatterplot smoother. The estimated function $\hat{r}_i(x_i)$ can reveal possible nonlinearities in the effects of the x_i .

In the logistic regression model the outcome y_i is 0 or 1, with 1 indicating a positive response (in our case detection of the earthquake by the station) and 0 indicating a negative

response (here failure to detect). We wish to model $p(y_i|x_1\dots x_p)$ the probability of a detection given prognostic factors $x_1\dots x_p$. The linear logistic model assumes that the log-odds are linear:

$$\log \frac{p(y_i|x_1\dots x_p)}{1-p(y_i|x_1\dots x_p)} = \beta_0 + x_1\beta_1 + \dots x_p\beta_p$$

The generalized additive logistic model assumes instead that

$$\log \frac{p(y_i|x_1\dots x_p)}{1-p(y_i|x_1\dots x_p)} = \beta_0 + r_1(x_1) + \dots r_p(x_p)$$

4.3 Detection by the Israel Seismic Network

Our analysis is based on data that was extracted from “The Seismological Bulletin of Israel (2005)”. The Bulletin contains the phase data and the processed information on source parameters of all earthquakes located by the Israeli seismic network during 2005; we used data from the time period January, 1, 2005 - July, 1, 2005. For each event located by the network we can find precise details and a list of the stations that successfully detected it.

Applying GAM theory for building a statistical model for the detection probability of a seismic signal:

$$\log \frac{p(I_{ij} = 1)}{1-p(I_{ij} = 1)} = \beta_0 + r_1(\text{Magnitude } j) + r_2(\text{Distance } (i, j))$$

where I_{ij} is the indicator that the j 'th event is detected by the i 'th station. Figure 1 shows the resulting estimate of detection probability as a function of distance for a magnitude 2.5 earthquake. Lower (higher) magnitudes had somewhat lower (higher) estimated detection curves; overall the effect of magnitude was not strong over the range of magnitudes present in our data set.

4.4 Optimal Design of a Seismic Network

In this subsection we illustrate the results of the theoretical research, considering design of a seismic network that is planned with a specific source. We assume that only P-wave arrivals will be available and we apply the detection model that we estimated from the Israel Seismic Bulletin. We examine magnitudes in the range of 2-4, which covers most of the seismic activity in Israel, and we adopt the crustal model in common use in Israel (Table 1).

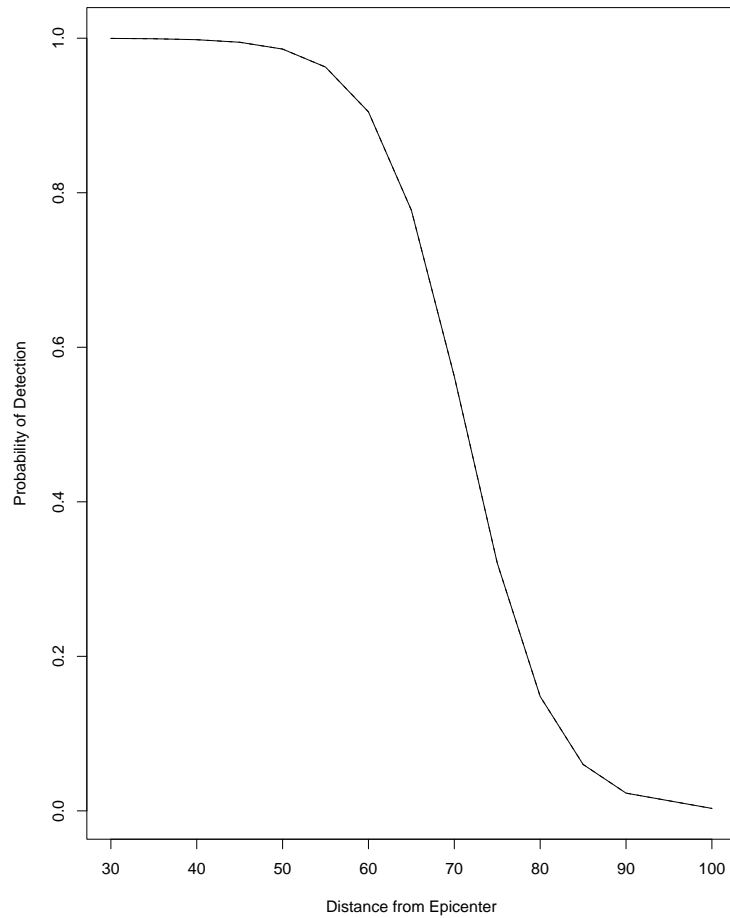


Figure 1: Estimated probability of detection vs. distance for a magnitude 2.5 earthquake.

Ignoring detectability, Steinberg and Rabinowitz (2003) characterized the structure of locally D -optimal seismic networks. These networks place one station at the epicenter and three equidistant stations on a circle about the epicenter of maximum possible radius. If the epicenter is above the half-space, so that refracted waves will be the first arrivals at some stations, the optimal network will also include three equidistant stations on a circle with a smaller radius. The proof that the optimal network can be limited to a station at

Table 1: Crustal Model

Layer	To Depth	P-wave Velocity
1	2.1 km	3.5 km/sec
2	12.7 km	5.7 km/sec
3	28.2 km	6.4 km/sec
4		7.9 km/sec

the epicenter and on at most two circles about the epicenter relied on symmetry conditions that remain true for the current problem. The impact of detectability comes to bear only in determining the optimal choice of radii for the circles. Henceforth, we examine networks with this structure.

We compare networks with 10 or 20 kilometers for the inner radius and 30 to 100 kilometers for the outer circle. The detection probability is modelled with the GAM that was described in the previous section and shown in Figure 1. The results in Figure 2 show that when the outer radius is less than 50 km the efficiency of both networks is almost identical. There is a sharp drop in the probability network efficiency at about 60 km from the epicenter. The loss of efficiency is especially severe for events with small magnitudes. From Figure 3 we see that the results are very similar for 10 km inner radius and for 20 km inner radius.

Figures 2 and 3 also show that the decreased efficiency, as a function of distance, is remarkably similar for all the magnitudes between 2 and 4. As a result, we can determine locations for the distant stations that have quite broad coverage, in the sense that they will be nearly optimal for all the magnitudes in the 2 - 4 range.

The results in Figure 4 show, for a fixed magnitude, the effect of considering detectability on the design criterion. We fix an inner radius and look at the value of the standard D -criterion and the value of the new criterion that includes the detection probability as functions of the outer radius. We found that the two criteria are almost identical when the outer radius is small, but they diverge as that radius becomes large. In particular, the standard criterion is monotone increasing. But for the new criterion, there is a clear optimum at an outer radius of approximately 50 km. This Figure effectively shows how adjusting for

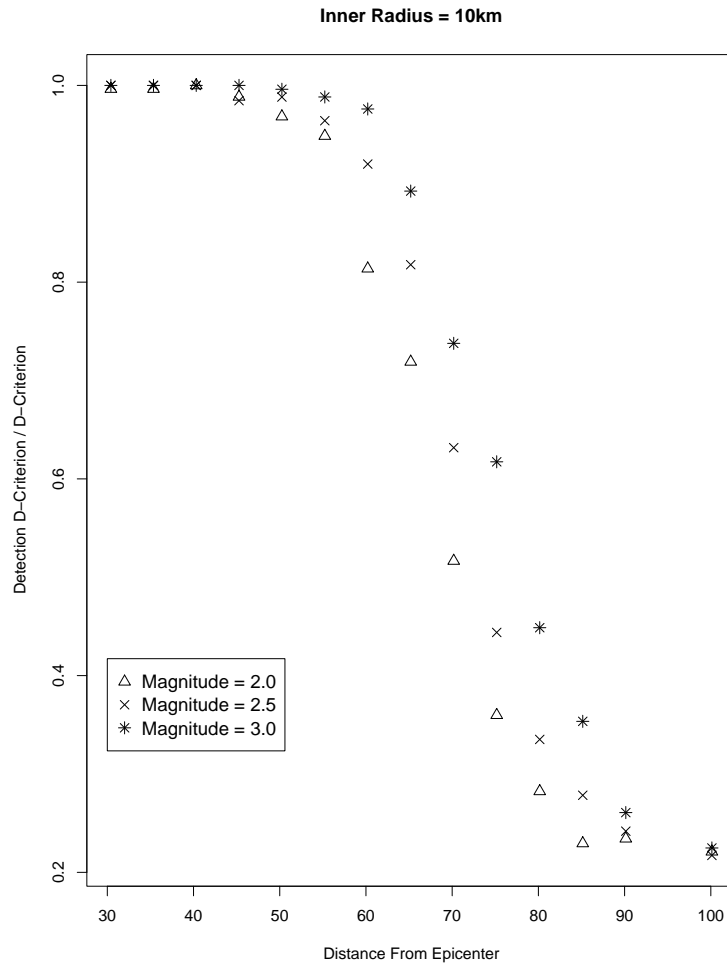


Figure 2: The ratio of detection-adjusted to full detection efficiency of a seismic network with 10 kilometers for the inner radius and 30 to 100 kilometers for the outer circle.

detectability modifies the choice of network configuration.

We can use Theorem 4 to determine optimal design. The implication of this theorem is that we can use a sufficient condition to verify whether or not a specific design is D -optimal. That is, if Fréchet directional derivative is equal to the number of parameters in the model, then the design is D -optimal. In Figure 5 we can see that the optimum is achieved at 40 km

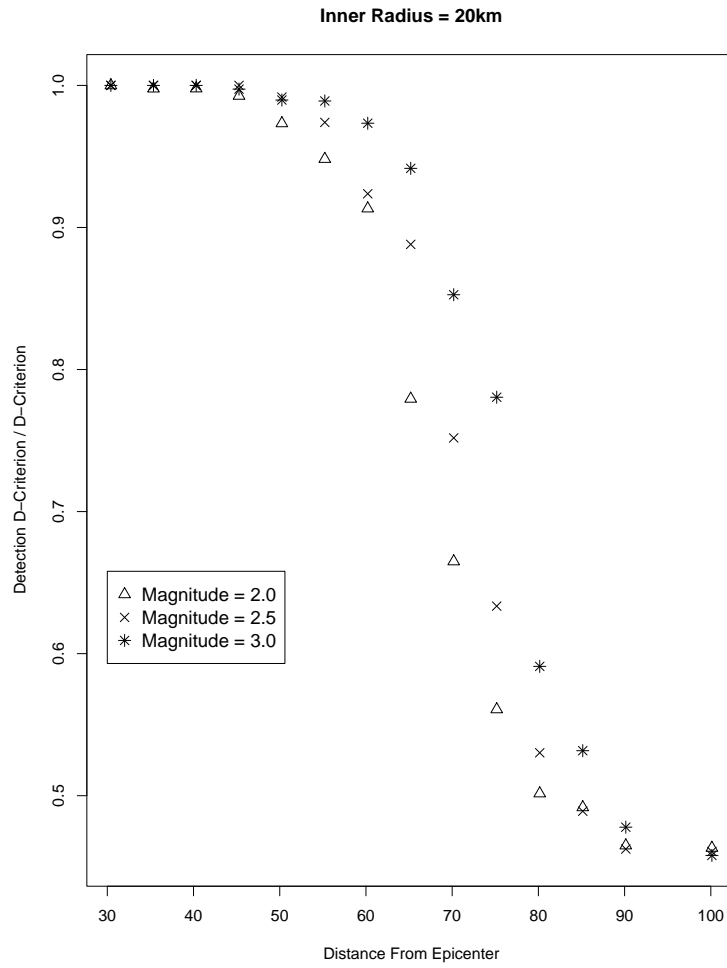


Figure 3: The ratio of detection-adjusted to full detection efficiency of probability seismic network with 20 kilometers for the inner radius and 30 to 100 kilometers for the outer circle.

outer radius, where the value of Fréchet derivative is maximized.

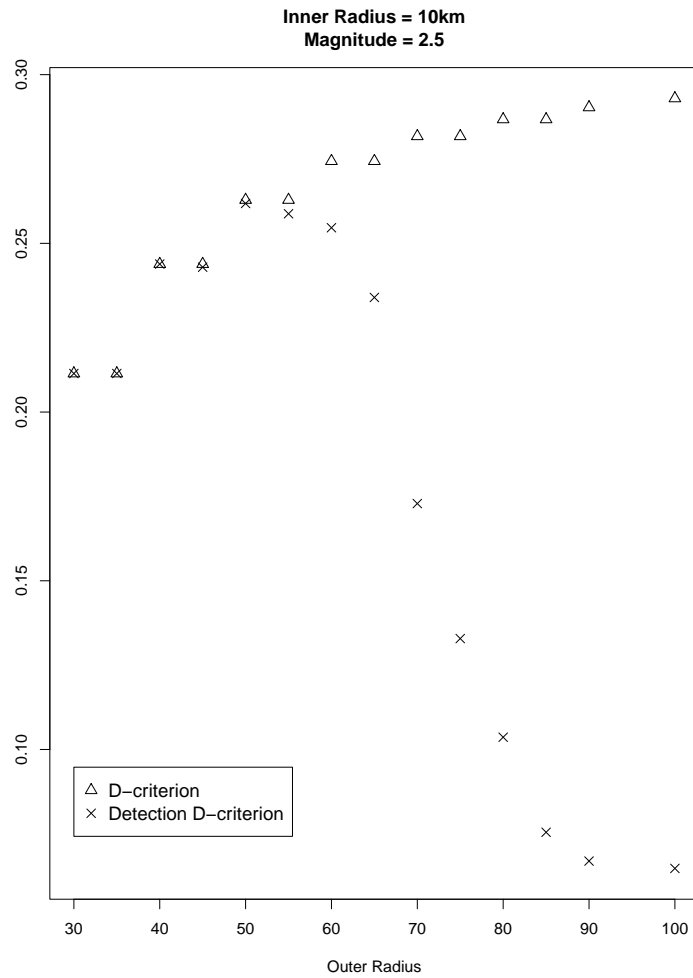


Figure 4: Simulation results that compare the value of the standard D -criterion and the value of the new criterion that includes the detection probability as functions of the outer radius for the following network settings: 10 kilometers for the inner radius, 2.5ML magnitude and 30 to 100 kilometers for the outer circle.

5 Conclusions

Station deployment is an important issue for monitoring networks. Not only do the station locations determine essential covariates used for parameter estimation, they also affect the

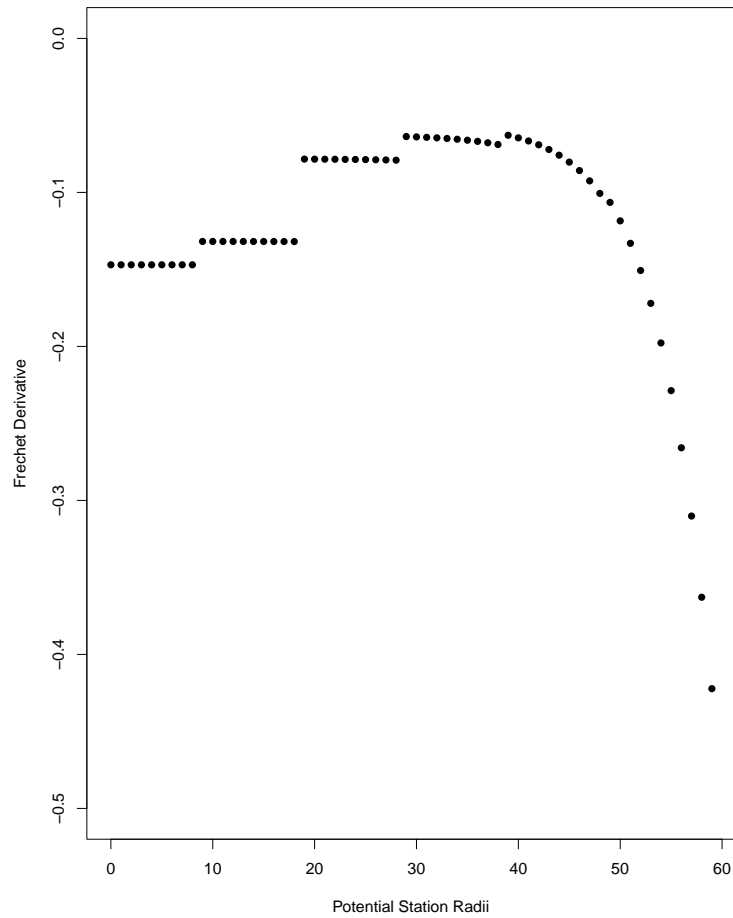


Figure 5: Fréchet directional derivative values as a function of the distance of the potential station from the epicenter for the following network settings: 10 kilometers for the inner radius, 2.5ML magnitude and 40 kilometers for the outer circle.

probability that usable data are collected at all. Network deployment must take account of the detectability. Our focus here has been on seismic networks, where previous research on configuration has ignored detectability.

We have shown here how to extend the D -optimality criterion to settings in which in-

complete detection is an issue. We have also shown that the general equivalence theorem can be extended to these problems. Finally, our empirical investigation illustrates that accounting for detectability can lead to radical changes in the structure of an optimal network.

6 Acknowledgments

We wish to thank Yochai Ben-Horin for helpful discussions on the seismological background and on the analysis of the bulletin data, and the two anonymous referees for constructive comments that improved the paper.

Bibliography

1. Aki K, Richards P (1980) *Quantitative seismology: Theory and methods*, vol I and II. New York: W.H. Freeman
2. Atkinson AC, Donev AN, Tobias R (2007) *Optimum experimental designs*. Oxford: Oxford University Press
3. Baek, I, Zhu, W, Wu, X, Wong WK (2006) “Bayesian optimal designs for a quantal dose-response study with potentially missing observations”, *Journal of Biopharmaceutical Statistics*, 16:679-693.
4. Chaloner K, Verdinelli I (1995), “Bayesian experimental design: a review”, *Statistical Science* 10:273-304
5. Chernoff H (1953) “Locally optimal designs for estimating parameters”, *Annals of Mathematical Statistics* 24:586-602
6. Gotwalt, C, Jones, BA, and Steinberg, DM (2009) “Fast computation of designs robust to parameter uncertainty for nonlinear settings”, *Technometrics* 51:88-95
7. Hastie, TJ, Tibshirani, RJ (1990) *Generalized Additive Models*, London: Chapman and Hall.
7. Kiefer J, Wolfowitz J (1960) “The equivalence of two extremum problems”, *Canadian Journal of Mathematics* 12:363-366
9. Lee WHK, Stewart SW (1981) *Principles and application of microearthquake networks*. New York: Academic Press

10. Imhof, LA, Song, D and Wong, WK (2002) "Optimal design of experiments with possibly failing trials", *Statistica Sinica* 12:1145-1155
11. Imhof, LA, Song, D and Wong, WK (2004) "Optimal design of experiments with anticipated pattern of missing observations", *Journal of Theoretical Biology* 228:251-260
12. Mitchell TJ (1974) "An algorithm for constructing of "D-optimal" experimental design", *Technometrics* 16:203-210
13. Pukelsheim F (1993) *Optimal Design of Experiments*. New York: John Wiley & Sons
14. Rabinowitz N, Steinberg DM (1990) "Optimal configuration of a seismographic network: a statistical approach", *Bulletin of the Seismological Society of America* 80:187-196
15. Silvey SD (1980) *Optimal Design, An Introduction To The Theory For Parameters Estimation*. New York: Chapman and Hall
16. Steinberg DM, Rabinowitz N (2003) "Optimal seismic monitoring for event location with application to on site inspection of the comprehensive nuclear test ban treaty", *Metrika* 58:31-57
17. Steinberg DM, Rabinowitz N, Shimshoni Y and Mizrachi D (1995) "Configuring a seismographic network for optimal monitoring of fault lines and multiple sources", *Bulletin of the Seismological Society of America* 85:1847-1857
18. White LV (1973) "An extension of the general equivalence theorem to nonlinear models", *Biometrika* 60:345-348