The Annals of Statistics
1993, Vol. 21, No. 1, 447–463

USING PRIOR INFORMATION IN DESIGNING INTERVENTION DETECTION EXPERIMENTS[1]

BY PETER SCHUMACHER AND JAMES V. ZIDEK

Statistical Sciences, Inc. and University of British Columbia

This paper investigates the effect of prior information on the design of experiments for detecting the potential impact of an event which is to occur at a specified time, in the knowledge of possible overall changes to the population as a whole. It is assumed that an F test of interaction is to be used to decide if an impact has occurred. Maximizing the power of this test, or rather the simpler, closely related goal of maximizing the noncentrality parameter is taken to be the designer's objective. Some of the results obtained are qualitative. For example, for certain fairly realistic general models of how a subregional impact might distribute itself, it is shown that it is never optimal to place more than 50% of the monitoring sites in any one of the homogeneous subregions in which the impact might occur. Another qualitative, more intuitively obvious result is that it is essential to monitor subregions where the impact is not likely to occur (''quasicontrols''); this would maximize the contrast created by the potential impact. A very general solution to the optimal design problem is given in a form which could be readily implemented in practice with the aid of a computer. Explicit solutions are also given for certain realistic impact models.

1. Introduction. Designing experiments invariably requires the use of prior information because the data which the experiment is designed to produce is not yet in hand. For example, in the first phase of a major study of surface water in the United States, Linthurst and his co-investigators chose the population of lakes to be sampled as ''lakes located within those regions expected to contain the most lakes in the U.S. characterized by alkalinity less than 400 μ eq/L (i.e., those areas where acidic deposition would potentially have the most effect)'' [Linthurst, Landers, Eilers, Brakke, Overton, Meier and Crowe (1986), page 4]. Because this prior information is usually introduced in an informal way, its influence on the selection of the ultimate design may well be obscured. In this paper we investigate the influence prior information can have on an experimental design. We do this in a very specific context to be described below, by introducing the prior information explicitly through intuitively natural parameters in an objective function which expresses the goal of

Received May 1990; revised April 1992.
[1]Research partially supported by the Societal Institute for the Mathematical Sciences from the United States Environmental Protection Agency. Support was also provided by the Natural Sciences and Engineering Research Council of Canada and by an Izaak Walton Killam Memorial Fellowship from the University of British Columbia.
AMS 1991 *subject classifications*. 62A15, 62C10, 62K05, 62P99.
Key words and phrases. Monitoring networks, optimal design, Bayesian designs, point impacts, environmental monitoring, interventions.

the design, taken here as the detection of a potential impact on a certain population of an ("intervention") event occurring at a specified time. A tractable objective function is developed in the next few paragraphs under what we believe to be reasonable simplifying assumptions. The optimization problem that emerges is to maximize the function H given in (2), as a function of $\mathbf{f}$ whose coordinates represent the fractions of the sample which are allocated to certain sampling zones defined below. Within zones, sampling will be random.

Our interest in the general problem addressed in this paper was originally stimulated by work related to various potential environmental impacts like those anticipated from the commencement of exploratory drilling in Harrison Bay on the North Slopes of Alaska [see Zidek (1984)]. In these examples, the population of interest is a spatial array of sites, some subset of which is to be sampled. The approach described here is an extensive generalization of that of Zidek (1984) and was stimulated in part by an anonymous referee who pointed out that our results might well apply outside the context of environmental monitoring and include human populations, for example.

Key features of the problem addressed in this paper are first, that it is not known, a priori, which if any items in the population will be changed by the intervention and second, that there may well be a pervasive or ambient change which affects the population as a whole. Thus interest focusses on the interaction, if any, between the event and the population, rather than on just the change in levels of the attributes of interest, since the latter might simply be due to the ambient change itself.

The item-event interaction, called the "space-event interaction" by Millard and Lettenmeier (1986) in the context of environmental monitoring, is quite commonly assessed by the F test in that context [cf. Green (1979)] and we assume it is to be used here. This implicitly assumes that the ambient change will be approximately constant. And clearly, to be effective, the design must not only incorporate a sample of items which do change as a result of the intervention, but also a sample of "pseudo controls" of those items which do not.

The designer's problem is made even more challenging, typically, by a paucity of background data which makes model parameters, like the interitem covariance matrix, inestimable. This forces subjective choices to be made on the basis of educated "hunches" about the size and likelihood of change. In this paper, these hunches are explicitly expressed by an assumption that the impact field is random and has a joint probability distribution which expresses the uncertainty in the prior opinions of the experimenters.

The power of the F test at the heart of this paper depends upon the noncentrality parameter δ^2, a quadratic functional of the differences between "before" and "after" expected values of measurements at the design stations. A marked departure of any one of the differences from their average over all design points would represent a change for that item and increase δ^2. But conversely $\delta^2 = 0$ would not necessarily imply no impact on the study population; this could happen simply because by bad luck or judgement, the design points did not include items affected by that impact. Clearly, to maximize the power of the test, the design points must maximize the unknown δ^2, pointing once again to the need, in practice, to resort to subjective strategies based on prior knowledge.

With uncertainty about the impact field expressed by a probability distribution, δ^2 becomes a random variable whose distribution depends on the choice of the subset of design points, say D. The power of the F test is therefore random, so the actual power of the test is the expected value of that of the F test. The optimal D, for testing at least, would maximize the expected power, but finding it with such a complex objective function is completely impractical. Instead we tackle the simpler problem of maximizing the expected noncentrality parameter. While our goal in doing so is primarily that of reaching a mathematically and numerically tractable problem and an objective which is at least consistent with that of maximizing the power, this simpler objective function has its own intuitive appeal: It is a natural index of the degree of change and might well be of interest even where testing is not the ultimate goal.

The simplified objective function thus becomes $E(\delta^2)$, which is just a quadratic function of the vector of 0's and 1's which represents the design. However, even this quadratic binary programming problem is well beyond the scope of modern programming methods (Professor M. Queyranne, personal communication). The problem is unrealistic as well in that it calls on the designer to provide input at a level of detail which would almost surely go well beyond the prior information at hand. So in this paper, an additional simplification is made by assuming the population can be stratified into relatively homogeneous clusters, which we will call "zones" in recognition of the context in which this problem originated. Our assumption is stated more precisely below in terms of *second order zonewise exchangeability* (SOZE).

As a lead-up to this assumption, suppose n replicate vector-valued measurements are to be made at each design point before and after the intervention. Assume these are conditionally independent given their expected values and common covariance matrix $\boldsymbol{\sigma}^2$. Denote the difference between the expected value vectors, "after" minus "before" say, by $\mathbf{Z}_{ij}$, for zone $i = 1, \dots, K$, and item $j = 1, \dots, n_i$ in zone i. A priori uncertainty translates into randomness of the various parameters of the measurement distribution and it is assumed in particular that the $\mathbf{Z}_{ij}$ are distributed independently of $\boldsymbol{\sigma}^2$.

It is easily shown that

$$E[\delta^2] = \sum_{(i,j)\in D} E\left[\left(\mathbf{Z}_{ij} - \overline{\mathbf{Z}}_D\right)^T Q\left(\mathbf{Z}_{ij} - \overline{\mathbf{Z}}_D\right)\right],$$

where $Q = E[(\boldsymbol{\sigma}^2/n)^{-1}]$ and $\overline{\mathbf{Z}}_D = (1/d)\Sigma_{(i,j)\in D}\mathbf{Z}_{ij}$, and d denotes the number of points in D. All expectations here and in the sequel except where otherwise indicated in specific contexts, are with respect to the joint prior distributions of the $\mathbf{Z}$'s and $\boldsymbol{\sigma}^2$.

This expression can be simplified using the notation $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T Q \mathbf{y}$ and $\|\mathbf{x}\|^2 = \langle \mathbf{x}, \mathbf{x} \rangle$ for arbitrary vectors $\mathbf{x}$ and $\mathbf{y}$. With this notation

$$(1) \qquad E[\delta^2] = \sum_{(i,j)\in D} E\|\mathbf{Z}_{ij} - \overline{\mathbf{Z}}_D\|^2.$$

Remembering $E[\delta^2]$'s role as an index of change, we would remark that the analysis of this paper will apply equally well whatever inner product is selected in obtaining (1), and $\mathbf{Z}_{ij}$ can even be infinite-dimensional as in the case where it is an analogue signal or time series.

We now impose the additional assumption of *second order zonewise exchangeability* (SOZE): $E(\mathbf{Z}_{ij}) = \mu_i$, $E(\|\mathbf{Z}_{ij}\|^2) = \gamma_i$, $E(\langle \mathbf{Z}_{ij}, \mathbf{Z}_{ij'} \rangle) = \beta_{ii}$, for certain constants $\boldsymbol{\mu}_i$, γ_i, β_{ii} and all i, $j \neq j'$ and at the same time $E(\langle \mathbf{Z}_{ij}, \mathbf{Z}_{kl} \rangle) = \beta_{ik}$ for all values of these subscripts, with $i \neq k$.

The assumption of SOZE implies

$$\begin{aligned} E(\delta^2) &= \sum_{(i,j)\in D} E\|\mathbf{Z}_{ij}\|^2 - dE\langle \overline{\mathbf{Z}}_D, \overline{\mathbf{Z}}_D \rangle \\ &= \sum n_i \gamma_i - d^{-1} \sum_{(i,j)\in D} \sum_{(i',j')\in D} E\langle \mathbf{Z}_{ij}, \mathbf{Z}_{i'j'} \rangle \\ &= \sum n_i \gamma_i - d^{-1} \sum n_i \gamma_i - d^{-1} \sum n_i (n_i - 1)\beta_{ii} - d^{-1} \sum_i \sum_{i' \neq i} n_i n_{i'} \beta_{ii'}, \end{aligned}$$

where n_i denotes the number of sampling points to be allocated to zone i. It thus follows that

$$(2) \qquad \begin{aligned} E[\delta^2] &= \sum f_i \{(d-1)\gamma_i + \beta_{ii}\} - d \sum \sum f_i f_{i'} \beta_{ii'} \\ &\triangleq H(\mathbf{f}), \end{aligned}$$

say, where $\mathbf{f} = (f_1, \dots, f_K)$ and f_i denotes the fraction of the total number d of points in zone i, for $i = 1, \dots, K$, that is, $f_i = n_i/d$.

It is assumed that the sample points within zones will be chosen at random. If no further randomization is admitted in the design, the problem of finding the optimal design now reduces to solving the quadratic integer programming problem of finding the optimal sampling fractions $\mathbf{f}$, those that maximize H in (2). The resulting design will be called the *optimal design* (OD).

In general, finding the OD remains out of reach of current techniques and technology. But our discretization of the original problem has an obvious approximate solution derived from treating the sampling fractions as continuous rather than discrete decision variables. The only constraints imposed in obtaining the approximate solution are $f_i \geq 0$ and $\sum f_i = 1$. In general, it may be necessary to impose the additional constraints that $L_i \leq f_i \leq U_i$, and this is permitted in the theory of Section 2. The resulting design will be called the *optimal approximate design* (OAD).

Randomization can be carried one step further and the n_i made random. One fairly simple scheme would endow them with a joint multinomial distribution with *zone randomization probabilities* π_i and $\boldsymbol{\pi} = (\pi_1, \dots, \pi_K)$; these cell probabilities satisfy exactly the same constraints as the sampling fractions f_i.

The expectation in (2) would need to be carried one step further with the result that

$$(3) \qquad E[\delta^2] = (d-1)\left[\sum \pi_i \gamma_i - \sum \sum \pi_i \pi_{i'} \beta_{ii'}\right],$$

for which we can write $H_R(\boldsymbol{\pi})$. An optimal design can now be found by maximizing H_R with respect to the π_i and the result will be referred to as the *optimal random design* (ORD). The OAD would be selected as an approximation to the OD purely for expediency and most of this paper is devoted to an investigation of the OAD. However, our results apply equally well to finding the ORD since, from a technical point of view finding these two designs are one and the same problem. Thus we are in fact investigating both the OAD and ORD in this paper.

Special features of the problem addressed in this paper place it beyond the current repertoire of optimal design theory which, broadly speaking, is based on either of two distinct paradigms [Federov and Mueller (1988), (1989)] which we will call the "regression" approach and the "random fields" approach for simplicity. The regression approach, deriving from the work of Smith (1918) has been developed by Elfving (1952), Kiefer (1959) and others [cf. Silvey (1980) for a review]. It seeks optimal designs for fixed effect regression experiments where estimation of functions of regression coefficients is the ultimate objective. There is a complementary literature on algorithms for finding the optimum or approximately optimum designs for this theory [see Pazman (1986) for a recent review]. While algorithms for finding optimal exact design are available, finding optimal designs can involve insurmountable computational difficulties and even the well known algorithm of Mitchell (1974) does not generally converge to an optimal solution, according to Fedorov, Leonov and Pitovranov (1987).

But the problem of this paper is not amenable to the regression approach. Our objective is testing and not estimation. And although the designer decides where to monitor for change, "nature" decides where the change occurs. Within the framework of fixed effects linear models the design matrix is thus unknown. This problem disappears in the random effects model. The design matrix can be specified and the randomness due to nature incorporated in the random effects. Our problem thus fits into the context of the second, or "random fields" approach to optimal design.

The random fields approach encompasses a number of design strategies, none of which applies to the problem of this paper. These strategies include designing to maximize the performance of the kriging interpolator, and designing to maximize some functional of entropy. The latter was proposed as a monitoring network design strategy by Caselton and Zidek [(1984); hereafter CZ] and again by Shewry and Wynn (1987). CZ formalize an idea applied by Caselton and Husain (1980); they advocate the use of an entropy based design criterion when there is a multiplicity of specified and unspecified objectives. This criterion has been implemented by Caselton, Kan and Zidek (1992), for the case of Gaussian fields where the spatial covariance matrix is unspecified, and refinements are added by Wu and Zidek (1992). The CZ approach is

Bayesian while that of this paper is only partially so for reasons discussed in the last section.

Federov and Mueller, in two very similar papers [Federov and Mueller (1988), (1989), hereafter FM], attempt to bridge the gap between the regression and random field approaches by making the regression effects random. The validity of their random effects model as a description of a spatial random field is unclear. Formally, the FM method does yield a design criterion which is related to, but more structured than, that of Caselton and Zidek (1984). The orientation of the resulting theory remains on estimation. And formally, their approach could be interpreted as Bayesian although FM do not exploit the rich existing Bayesian design theory.

In fact the CZ approach is really an implementation of a proposal of Lindley (1956) on the use of entropy in experimental design. There is a substantial body of work on optimal design in the Bayesian context although none as far as we are aware covers the application which is made here. For a recent review of Bayesian design see Verdinelli (1992).

While our work fits into the random fields context, our emphasis on testing and the resulting objective function (the expected noncentrality parameter) sets it apart. The pixelization scheme used in this paper and described above also gives our work a special feature which is vitally important to our analysis and its practical implementation. Finally, we are considering the case where the response variable is a vector rather than a scalar valued object, whereas the work we have been citing deals with scalar valued responses.

In summary, Section 2 of this paper gives a very general characterization of the OAD (ORD) and applies it in the particular case of $K = 2$ zones. The general characterization is then applied in Sections 3 and 4 to obtain the OAD (ORD) in certain special cases. A number of properties of the optimal sampling fractions (zone randomization probabilities) are presented. Except where otherwise indicated, the proofs of all of the results are given in the Appendix. The results derived in Sections 2–4 are discussed in Section 5. Some of the shortcomings of the present approach are noted and possible extensions are indicated.

2. Subregionally homogeneous impacts. The form of δ^2 allows us to assume without loss of generality that the uniform ambient change over the entire region under study is zero and we will now impose that assumption for convenience. Now to allow for the inclusion of a zone where it is certain that no impact occurs, that is, where $\mathbf{Z}_{ij} = \mathbf{0}$ for all j in zone i, under the assumption just made of a zero ambient change, we allow $\gamma_i = 0 = \beta_{ii'}$ for at most a single i, say $i = 1$, and for all i'. We assume that $\gamma_{i'} > 0$ for $i' > 1$ and that the $(K-1) \times (K-1)$ matrix $(\beta_{ii'})$ for which $i, i' > 1$ is positive definite. It follows that H as defined in (2) is a concave function of $\mathbf{f}$ with a unique global maximum $\mathbf{f}_a$ on the convex set $\mathscr{F} = \{\mathbf{f}\colon \mathbf{0} \le \mathbf{L} \le \mathbf{f} \le \mathbf{U} \le \mathbf{1}, \sum f_i = 1\}$, where $\mathbf{1}$ denotes the vector all of whose elements are 1, and $\mathbf{L}, \mathbf{U}$ are any feasible bounds on $\mathbf{f}$ imposed by the problem. The OAD sampling fractions are characterized in Theorem 1, where $\mathscr{I}_L$ and $\mathscr{I}_U$ denote the subscripts i for which $f_{ai} = L_i$ or $f_{ai} = U_i$, respectively, and the corresponding constraints are binding. Let $\bar{\mathscr{I}}$ denote the complement of $\mathscr{I}_L \cup \mathscr{I}_U$ in $\{1, \dots, K\}$.

THEOREM 1. *For $K \geq 2$, the sampling fractions which maximize the expected noncentrality parameter are, for a uniquely determined partition $\{1, \dots, K\} = \mathscr{I}_L \cup \bar{\mathscr{I}} \cup \mathscr{I}_U$, the solution of the equations*

$$\lambda = M_i(\mathbf{f}_a), \qquad i \in \bar{\mathscr{I}} \tag{4}$$

subject to $f_{ai} = L_i$, $i \in \mathscr{I}_L$, and $f_{ai} = U_i$, $i \in \mathscr{I}_U$, where $M_i(\mathbf{f}_a) = \gamma_i - (1/d)(\gamma_i - \beta_{ii}) - 2\sum_{j \geq 1} f_{aj}\beta_{ij}$, $M_i(\mathbf{f}_a) < \lambda$ or $> \lambda$ according as $i \in \mathscr{I}_L$ or $i \in \mathscr{I}_U$, and λ is a uniquely determined constant.

PROOF. See the Appendix.

The method of proof used in Theorem 1 uses the Kuhn–Tucker approach to optimization and since H is concave, $\mathscr{I}_L$, $\mathscr{I}_U$ and λ are unique. With $\mathscr{I}_L$ and $\mathscr{I}_U$ identified (binding constraints must be determined first), the optimization problem becomes Lagrangian in nature. In fact condition (4) simply asserts that the optimal $\mathbf{f}$ (on $\bar{\mathscr{I}}$) must be chosen so that the contours of the functions $H(\mathbf{f})$ and $\sum_{i \in \bar{\mathscr{I}}} f_i$ are tangent, that is, the gradient of $H(\mathbf{f})$ must be parallel to that of $\sum_{i \in \bar{\mathscr{I}}} f_i$, that is, must be a multiple (λ) of the vector all of whose coordinates are one and whose dimension is the number of points in $\bar{\mathscr{I}}$.

In the case of just two zones, the theorem yields an explicit result. The sampling fractions f_1, $f_2 = 1 - f_1$ must satisfy $L_i \leq f_i \leq U_i$, $i = 1, 2$ and a feasible solution exists if and only if the L's and U's satisfy $L_1 + L_2 \leq f_1 + f_2 \leq U_1 + U_2$, that is, $\min\{L_1, 1 - U_2\} < \max\{U_1, 1 - L_2\}$. In applying Theorem 1, we must first determine which if any of the constraints is binding. Now it is easily shown that subject only to $0 \leq f_i \leq 1$ and $f_2 = 1 - f_1$, the optimal solution is $f_{a1} = f \triangleq (1/2)(\alpha_1 - \alpha_2 + 2\beta_{11} - 2\beta_{12})(\beta_{11} + \beta_{22} - 2\beta_{12})^{-1}$, where $\alpha_i = \gamma_i + (1/d)(\beta_{ii} - \gamma_i)$ for $i = 1, 2$. One of the other constraints is binding when $f < \min\{1 - U_2, L_1\}$ or $f > \max\{U_1, 1 - L_2\}$, more explicitly when $f_{a1} = L_1$ if $1 - U_2 \leq f < L_1$, $f_{a1} = 1 - U_2$ if $L_1 \leq f < 1 - U_2$, $f_{a1} = U_1$ if $U_1 < f \leq 1 - L_2$, and $f_{a1} = 1 - L_2$ if $1 - L_2 < f \leq U_1$.

Observe that $f = 1/2$ if $\gamma_1 = \gamma_2$ and $\beta_{11} = \beta_{22}$ while if $f = 1/2$ is feasible, then it is optimal under these conditions regardless of the size of the "interclass correlation" β_{12}, roughly speaking. To interpret our results let us consider the univariate case with d large so that α_i is approximately equal to $\gamma_i = (\sigma_i^2 + \mu_i^2)(\sigma_i^2/n)^{-1}$, where $\sigma_i^2 = \mathrm{Var}(Z_{ij})$ for all j. Also $\beta_{ij} = (\sigma_i\sigma_j\rho_{ij} + \mu_i\mu_j)(\sigma_i^2/n)^{-1}$, where ρ_{ij} denotes the correlation between Z_{ik} and Z_{jm} (with $k \neq m$ if $i = j$). Then $f = (1/2)[1 - (\sigma_2^2\overline{\rho_{22}} - \sigma_1^2\overline{\rho_{11}})D^{-1}]$, where $\overline{\rho_{ij}} = 1 - \rho_{ij}$, $i = 1, 2$, $D = (\Delta\mu)^2 + (\Delta\sigma)^2 + 2\sigma_1\sigma_2\overline{\rho_{12}} - \sigma_1^2\overline{\rho_{11}} - \sigma_2^2\overline{\rho_{22}}$, $\Delta\mu = \mu_1 - \mu_2$ and $\Delta\sigma = \sigma_1 - \sigma_2$. It follows that if $\Delta\mu$ is large relative to the other parameters, $f_{a1} = 1/2$ is optimal if feasible, the same result as is obtained when $\sigma_1^2\overline{\rho_{11}} = \sigma_2^2\overline{\rho_{22}}$, a fact noted earlier. It is expected that within-zone covariances should exceed between-zone covariances and hence that typically $D > 0$. It then follows that $f < 1/2$ or $> 1/2$ according as $\overline{\sigma_2^2\rho_{22}} > \overline{\sigma_1^2\rho_{11}}$ or $< \overline{\sigma_1^2\rho_{11}}$. This seems intuitively natural since the designer's objective is to maximize the expected noncentrality parameter, that is, the contrasts among the Z_{ij}, and this is more likely to be achieved where σ_i^2 is large unless ρ_{ii} is unduly large at the same time. The optimal f takes account of these competing factors through $\overline{\sigma_i^2\rho_{ii}}$.

3. Particular covariance structures. A potentially applicable specialization of the model leading to Theorem 1, for which the OAD sampling fractions can be found explicitly, is analyzed in this section. Let

$$\begin{aligned} \beta_{ij} &= B_0 + B_1\phi_i\phi_j \quad \text{for all } i \neq j, \\ \gamma_i &= \beta_{ii} - C_0\phi_i^r = B_0 + B_1\phi_i^2 - C_1\phi_i^r \quad \text{for all } i, \end{aligned} \tag{5}$$

where $r = 0$ or 2, $0 \le \phi_1 \le \cdots \le \phi_K$, $B_1 > 0$, $C_0 \ge 0$, $C_1 < 0$ and B_0 are all specified constants and $\mathbf{f}$ is restricted only by $\mathbf{0} \le \mathbf{f} \le \mathbf{1}$ so that $\mathbf{0} = \mathbf{L}$, $\mathbf{U} = \mathbf{1}$. If $r = 2$, $\phi_1 > 0$ is assumed to avoid degeneracy. Example 1 shows how such a covariance structure can arise. With this added structure (4) becomes

$$\begin{aligned} \lambda = {} & B_0 + B_1\phi_i^2 - d^{-1}C_0\phi_i^r + 2C_1\phi_i^r f_{ai} \\ & - 2\sum_{j\ge 1} f_{aj}(B_0 + B_1\phi_i\phi_j), \qquad i \in \bar{\mathscr{I}}. \end{aligned} \tag{6}$$

Consider first the case where $r = 0$. Here (6) becomes

$$\lambda_1 = \phi_i^2 + 2C_2 f_{ai} - 2\phi_i\mu, \qquad i \in \bar{\mathscr{I}}, \tag{7}$$

where $\lambda_1 = (\lambda + B_0 + d^{-1}C_0)/B_1$, $C_2 = C_1/B_1$ and $\mu = \sum f_{aj}\phi_j$. Multiply both sides of (7) by ϕ_i and sum the result over all $i \in \bar{\mathscr{I}}$. The result is

$$\lambda_1 S_1 = S_3 + 2C_2\mu - 2S_2\mu, \tag{8}$$

where $S_t = \sum_{i\in\bar{\mathscr{I}}}\phi_i^t$, $-\infty < t < \infty$. Now sum both sides of (7) over $i \in \bar{\mathscr{I}}$ to get

$$\lambda_1 S_0 = S_2 + 2C_2 - 2S_1\mu. \tag{9}$$

Equations (7), (8) and (9) are readily solved for f_{ai}, λ_1 and μ. The results are

$$\begin{aligned} f_{ai} &= \tfrac{1}{2}C_2^{-1}[\lambda_1 - \phi_i^2 + 2\phi_i\mu], \qquad i \in \bar{\mathscr{I}}, \\ \mu &= \tfrac{1}{2}[c(\phi^2,\phi) - 2C_2S_1S_0^{-2}][c(\phi,\phi) - C_2S_0^{-1}]^{-1}, \\ \lambda_1 &= [S_2 + 2C_2 - 2S_1\mu]/S_0, \end{aligned} \tag{10}$$

where, in general, $c(x, y) = (1/S_0)\sum_{i\in\bar{\mathscr{I}}}(x_i - \bar{x})(y_i - \bar{y})$.

But $f_{ai} \ge 0$ entails $\lambda_1 \le \phi_i^2 - 2\phi_i\mu$ for $i \in \bar{\mathscr{I}}$ while Theorem 1 requires $\lambda_1 > \phi_i^2 - 2\phi_i\mu$ for $i \in \mathscr{I} \equiv \mathscr{I}_L$. Since $x \mapsto x^2 - 2x\mu$ is a quadratic function which attains its minimum at $x = \mu$, it follows that $\mathscr{I} = \{i\colon l \le i \le u\}$ for integers $l, u \in \{1, \dots, K\}$ unless $\bar{\mathscr{I}} = \{1, \dots, K\}$ where $\mathscr{I} = \varnothing$.

OAD sampling fractions are now readily found by trying successively smaller $\mathscr{I}$'s until a λ_1 and μ are found for which the conditions of Theorem 1 are satisfied. That a unique λ_1 and μ exist follows from that theorem as well. The procedure is illustrated in Example 1.

The case where $r = 2$ and $\phi_i > 0$ for all i in (5) is handled in a similar fashion. Instead of (7) we obtain

$$\lambda_2 = \phi_i^2 + 2C_3\phi_i^2 f_{ai} - 2B_3\phi_i\mu, \qquad i \in \bar{\mathscr{I}}, \tag{11}$$

where $\lambda_2 = (\lambda + B_0)/B_2$, $B_2 = (B_1 + C_0/d)$, $C_3 = C_1/B_2$, $B_3 = B_1/B_2$ and μ is as defined above. If $0 < \phi_i$ for all i, (11) may be solved in the same manner as (7) to yield

$$\begin{aligned} f_{ai} &= \tfrac{1}{2}C_3^{-1}\phi_i^{-2}\left[\lambda_2 - \phi_i^{-2} + 2B_3\phi_i\mu\right], \qquad i \in \bar{\mathscr{I}}, \\ \mu &= \tfrac{1}{2}\left[c(\phi, \phi^{-2}) + 2C_3S_{-1}\right]/\left[C_3S_{-2} - B_3c(\phi^{-1}, \phi^{-1})\right], \\ \lambda_2 &= \left[S_0 + 2C_3 - 2B_3\mu S_{-1}\right]/S_{-2} \end{aligned} \tag{12}$$

with $c(\cdot, \cdot)$ as defined just below (10). As before we may deduce that $\mathscr{I} = \varnothing$ or $\mathscr{I} = \{i: l \le i \le u\}$ so that $\bar{\mathscr{I}}$ can be found without resorting to a combinatorial search. The next theorem summarizes the results of this section.

THEOREM 2. *With the assumptions in* (5) *added to those of Theorem* 1, *the OAD sampling fractions are given in* (10) *or* (12) *according as* $r = 0$ *or* 2 *in* (5). *In any case, zones selected for monitoring by the OAD are the complement of* $\mathscr{I} = \varnothing$ *or* $\mathscr{I} = \{i: l \le i \le u\}$ *for integers* $l, u \in \{1, \dots, K\}$.

PROOF. The proof is contained in the discussion which precedes the statement of this result.

The following example involves a special case of some interest in its own right.

EXAMPLE 1. Suppose the data vectors are one-dimensional, $\beta_i = \sigma^2 + \mu_i^2$, $\gamma_i = \sigma^2\rho_w + \mu_i^2$ for all i and $\beta_{ij} = \sigma^2\rho_b + \mu_i\mu_j$ for all $i \ne j$ with $\rho_b < \rho_w$. Then assumptions (5) obtain with $B_0 = \sigma^2\rho_b$, $B_1 = 1$, $\phi_i = \mu_i$ for all i, $C_0 = \sigma^2(1 - \rho_w)$, $r = 0$ and $C_1 = \sigma^2(\rho_b - \rho_w) < 0$.

A natural alternative to the last covariance model would have $\mathrm{Var}(Z_i) \propto \mu_i^2$ provided $\mu_i > 0$ for all i. In this case assumptions (5) obtain with $r = 2$.

To illustrate how to get the optimal design corresponding to the first of the two models in this example, let $K = 5$, $\rho_b = 0.1$, $\rho_w = 0.6$, $\sigma^2 = 16$ and the $\{\mu_i\}$ be those given below:

$$\begin{aligned} i &= 1\,2\,3\,4\,5, \\ \mu_i &= 1\,2\,4\,5\,10. \end{aligned}$$

Both μ and λ_1 are computed successively for the various (but not all) possible choices of $\mathscr{I}$, each a subinterval of $\{1, \dots, 5\}$ to illustrate the process of finding the optimal subset. The results, shown in Table 1, reveal that

TABLE 1
Computations for selecting the optimal monitoring zones in Example 1, *using* (10)

			$\mu_i^2 - 2\mu_i\mu$ for given i				
Network zones	μ	λ_1	**1**	**2**	**3**	**4**	**5**
1	1	−17	−1	0	8	15	80
5	10	−116	−19	−36	−64	−75	−100
1, 5	5.5	−18	−10	−18	−28	−30	−10
1, 2, 5	5.5	−18	−10	−18	−28	−30	−10
1, 4, 5	5.540	−22.426	−10.081	−18.161	−28.322	−30.403	−10.805
1, 2, 3 5	4.896	−15.370	−8.793	−15.586	−23.172	−23.965	2.070
1, 2, 4 5	4.852	−15.164	−8.703	−15.406	−22.813	−23.516	2.969
1, 3, 4, 5	5.6	−24.5	−10.2	−18.4	−28.8	−31	−12
2, 3, 4, 5	5.472	−22.154	−9.944	−17.888	−27.776	−29.720	−9.440

$\bar{\mathscr{I}} = \{1, 2, 5\}$ is the optimal selection of zones since $\lambda_1 > \mu_i^2 - 2\mu_i\mu$ for $i = 3, 4$ while $\lambda_1 \le \mu_i^2 - 2\mu_i\mu$ for $i = 1, 2, 5$. Then using (10) it follows that $f_{a1} = f_{a5} = 1/2$ while $f_{a2} = 0$. Thus, equivalently, $\bar{\mathscr{I}} = \{1, 5\}$ is optimal. However, $f_2 \ge 0$ is a nonbinding constraint; $f_{a2} = 0$ cannot be improved on by choosing $f_{a2} < 0$ in violation of this constraint.

4. Uniform impacts. To the assumptions underlying Theorem 1 add: $\mathbf{Z}_{ij} = \mathbf{Z}_i$ for all j in zone i; $\mathbf{L} = \mathbf{0}$ and $\mathbf{U} = \mathbf{1}$. Consequently $\beta_{ii} = \gamma_i$. Assume $E\|\mathbf{Z}_i - \mathbf{Z}_j\|^2 > 0$ when $i \ne j$.

Theorem 1 simplifies substantially with these added assumptions and the result is the next theorem with $\mathscr{I}_U = \phi$ and $\mathscr{I}_L = \mathscr{I}$.

THEOREM 3. *The OAD sampling fractions*, $\{f_{ai}\}$, *are uniquely characterized by the conditions*

$$(13)\qquad \begin{aligned} E\|\mathbf{Z}_i - \bar{\mathbf{Z}}\|^2 &= \lambda, & i &\in \bar{\mathscr{I}}, \\ E\|\mathbf{Z}_i - \bar{\mathbf{Z}}\|^2 &< \lambda, & i &\in \mathscr{I} \end{aligned}$$

for some constant λ, *where* $\bar{\mathbf{Z}} = \sum_k f_{ak}\mathbf{Z}_k$.

It is easily shown that the constant λ of Theorem 3 is $\lambda = E\sum_i f_{ai}\|\mathbf{Z}_i - \bar{\mathbf{Z}}\|^2$. The proof of the following result is straightforward and omitted.

COROLLARY 4. *The OAD places sampling sites in at least two subregions.*

The next result is not intuitively obvious.

COROLLARY 5. *The following statements are equivalent*:

(i) *The OAD places all the monitoring stations in zones i and j.*
(ii) $f_{ai} = f_{aj} = 1/2$.
(iii) $E\langle \mathbf{Z}_l - \mathbf{Z}_i, \mathbf{Z}_l - \mathbf{Z}_j \rangle < 0$ *for all* $l \ne i, j$.

PROOF. See the Appendix

It is not obvious that if stations i, j satisfying (iii) exist, then they are unique. A simple direct proof is the following. For any two pairs of zone labels,

$$\begin{aligned} 0 &< E\|\mathbf{Z}_i + \mathbf{Z}_j - \mathbf{Z}_m - \mathbf{Z}_n\|^2 \\ &= E\langle \mathbf{Z}_i - \mathbf{Z}_m, \mathbf{Z}_i - \mathbf{Z}_n\rangle + E\langle \mathbf{Z}_j - \mathbf{Z}_m, \mathbf{Z}_j - \mathbf{Z}_n\rangle \\ &\quad + E\langle \mathbf{Z}_m - \mathbf{Z}_i, \mathbf{Z}_m - \mathbf{Z}_j\rangle + E\langle \mathbf{Z}_n - \mathbf{Z}_i, \mathbf{Z}_n - \mathbf{Z}_j\rangle. \end{aligned}$$

So it is impossible that (i, j) and (m, n) both satisfy (iii), proving the assertion of uniqueness.

COROLLARY 6. *$f_{ai} \le 1/2$ for all i.*

PROOF. See the Appendix.

Theorem 3 does not in general admit an explicit solution. Through further specialization, explicit solutions of certain particular cases of this theorem are given by Theorem 2. One of these cases will be investigated below. It is a distinguished special case in that it is the only one of reasonable generality for which an integer programming algorithm and hence exact solution is available. In this case it is possible then to compare the OAD solution with the exact solution, and this has been done in an example which for brevity will not be included here. The results are identical [Schumacher and Zidek (1988)].

To the other assumptions made above in this section, add $\beta_{ij} = 0$, $i \ne j$ which means that the case where at most one zone undergoes change is included, given the assumption we have made (without loss of generality) that the ambient change is zero. This case would arise in practice in situations where the affected zone was decided by a structural element like the direction of the wind at the time of the event.

Our assumptions imply

$$E(\delta^2) = d \sum_{k=1}^{K} f_i(1 - f_i)\gamma_i. \tag{14}$$

Relabel the $\{\gamma_i\}$ if necessary so that

$$0 \le \gamma_1 < \gamma_2 < \cdots < \gamma_K.$$

Observe that in our counterpart of (13),

$$E\|\mathbf{Z}_i - \overline{\mathbf{Z}}\|^2 = \gamma_i(1 - 2f_i) + \sum f_{ai}^2\gamma_i. \tag{15}$$

We now dispose of an easy special case to simplify our exposition.

THEOREM 7. *If $K = 2$, $f_{a1} = f_{a2} = 1/2$.*

PROOF. The result follows directly from (14). □

The following specialization of Theorem 1 gives the OAD (ORD) sampling fractions (zone sampling probabilities) in an explicit form.

THEOREM 8. *For $K > 2$, the sampling fractions which maximize the expected noncentrality parameter are*:

$$f_{ai} = \begin{cases} 0, & i = 1, \dots, m, \\ \frac{1}{2}[1 - \lambda\gamma_i^{-1}], & i = m+1, \dots, K, \end{cases}$$

where $\lambda = \lambda(m) = [K - m - 2][\sum_{i=m+1}^{K}\gamma_i^{-1}]^{-1}$, and $m = 0$ unless a unique positive $m < K - 2$ can be found which satisfies $\gamma_m < \lambda(m) \le \gamma_{m+1}$.

PROOF. This is essentially a consequence of Theorem 3. The proof appears in the Appendix.

There are a number of notable consequences of this theorem. These are given in the following corollaries whose proofs are immediate. The optimal approximate design (OAD) D_a is that having the sampling fractions stated in Theorem 8.

COROLLARY 9. *For any $K \ge 3$, the number of zones in D_a is $K - m \ge 3$.*

PROOF. The result is obvious since otherwise $\lambda(m) \le 0$ for all m and $\gamma_m < \lambda(m)$ cannot then be satisfied for any m, a contradiction. □

COROLLARY 10. *If $K = 4$, it is approximately optimal to monitor just $K - m = 3$ zones if and only if*

$$\gamma_1^{-1} > \gamma_2^{-1} + \gamma_3^{-1} + \gamma_4^{-1},$$

when zone 1 is excluded from D_a.

Thus far we have considered the design problem from the point of view of choosing continuous sampling fractions. An algorithm for finding the optimal solution in the discrete analogue is given by Schumacher and Zidek (1988). For brevity it is not presented here. In an example, the approximate solution given by rounding off the optimal continuous solution is exactly the same as that given by the integer programming algorithm, an encouraging result.

5. Discussion. This paper has investigated design problems involving subjective choices based on prior information, explicitly admitted and used in a fairly conventional setup where an experiment is to be designed to detect the potential impact of an event occurring at a known time. General optimal designs are derived and some interesting qualitative conclusions are reached.

It may well be objected that it is inconsistent to "break the Bayesian eggs" as we have done by expressing the prior information through probabilities and then not enjoy the "Bayesian omelette," opting instead to use the classical approach of hypothesis testing. However, our goal is not that of finding alternatives to the F test which as noted above is a conventional method in this context, but rather to discover how to incorporate prior information into the design *if* the F test is to be used, and to discover something about the character of the resulting design.

We recognize that in practice, "good" rather than "optimal" designs are needed and optimal designs like those in this paper must be considered as tentative proposals susceptible to modification depending on the circumstances prevailing in the context of their implementation. These "optimal" designs may well be valuable starting points, however, since they can be explicated in terms of their axiomatic underpinnings and proposed changes to these optimal designs can be interpreted in terms of the axioms. This can provide a degree of clarity in the typically complex situation confronting a designer.

Schumacher and Zidek (1988) discuss various issues, implications and interpretations of the theory developed in this paper and throw some additional light on the value of the theory. For brevity we will merely highlight some of these comments.

Our assumption that the unknown ambient change is approximately constant implies a need to minimize the breadth of the study population while preserving the pseudo controls to insure its validity.

Relying on the power as a design criterion is a concession to simplicity and has the shortcoming of ignoring substantive factors such as economic impact.

A surprising conclusion is that of Corollary 9, which says that when at most one of a group of three zones can potentially be impacted, with a uniform impact across the zone, monitoring must be done in all three, never in only two, regardless of the underlying parameters. This does not carry over to the more general model of Section 4 where monitoring just two zones may sometimes be optimal. Undoubtedly the most striking result of the paper is that of Corollary 6, which states that for the most general situation addressed in Section 4, no more than 50% of the network's sites should ever be placed in any one zone. This result seems very unintuitive. It would be interesting to know how general it really is.

APPENDIX

The proofs of most of the results in the paper are given in this Appendix.

PROOF OF THEOREM 1. Our assumptions imply $H(\mathbf{f}_a) > H(\mathbf{f}_a + \mathbf{h})$ when $\mathbf{f}_a + \mathbf{h}$ is feasible, that is, $\Sigma h_i = 0$, $h_i \leq 0$ for $i \in \mathscr{I}_U$, $h_i \geq 0$ for $i \in \mathscr{I}_L$ and $|\mathbf{h}|$ is small. But then $H(\mathbf{f}_a + \mathbf{h}) \simeq H(\mathbf{f}_a) + G(\mathbf{f}_a, \mathbf{h})$ where $G(\mathbf{f}_a, \mathbf{h}) = \Sigma h_i M_i(\mathbf{f}_a)$ and $M_i(\mathbf{f}_a)$ is defined in the statement of the theorem. Here " $\simeq$ " means equal up to the first two terms in the Taylor expansion in $\mathbf{h}$. It follows that $0 > G(\mathbf{f}_a, \mathbf{h})$ for all such $\mathbf{h}$. Suppose $\bar{\mathscr{I}}$ contains at least two elements, say $\{k, l\}$. Set $h_i = 0$ except for $i = k, l$ with $h_k = -h_l = h$. Then $H(\mathbf{f}_a + \mathbf{h}) \simeq H(\mathbf{f}_a) + h(M_k(\mathbf{f}_a) - M_l(\mathbf{f}_a))$. Now if $M_k(\mathbf{f}_a) - M_l(\mathbf{f}_a) \neq 0$, for sufficiently small $|\mathbf{h}|$, $H(\mathbf{f}_a) > H(\mathbf{f}_a) + h(M_k(\mathbf{f}_a) - M_l(\mathbf{f}_a))$. Then since h may be either negative or positive we have a contradiction. Thus, $M_k(\mathbf{f}_a) = M_l(\mathbf{f}_a)$ and, by extension, $M_i(\mathbf{f}_a)$ is a constant, say λ, on $i \in \bar{\mathscr{I}}$. Thus $G(\mathbf{f}_a, \mathbf{h}) = \Sigma h_i M_i(\mathbf{f}_a) = \Sigma_{i \in \bar{\mathscr{I}}} h_i(\lambda) + \Sigma_{i \in \mathscr{I}} h_i(M_i(\mathbf{f}_a)) = \Sigma_{i \in \mathscr{I}} h_i(M_i(\mathbf{f}_a) - \lambda)$ since

$\Sigma_{i \in \bar{\mathscr{I}}} h_i = -\Sigma_{i \in \mathscr{I}} h_i$. Now for any given $i \in \mathscr{I}_U$, take $h_i < 0$ and $h_j = 0$ for any other j except for some $j \in \bar{\mathscr{I}}$, to preserve the feasibility condition $\Sigma h_i = 0$. Then since $H(\mathbf{f}_a + \mathbf{h}) \simeq H(\mathbf{f}_a) + \Sigma_{i \in \mathscr{I}} h_i(M_i(\mathbf{f}_a) - \lambda)$, we obtain a contradiction unless $M_i(\mathbf{f}_a) - \lambda > 0$. Similarly we may choose, successively, $h_i > 0$ for any given $i \in \mathscr{I}_L$, to obtain the conclusion. The cases where $\bar{\mathscr{I}}$ consists of one or no elements are straightforward. □

PROOF OF COROLLARY 5. Suppose $f_{al} = 0$ except when $l = i$ or j. With $\alpha = f_{ai}$ and $(1 - \alpha) = f_{aj}$, Theorem 1 implies

$$0 = E\|\mathbf{Z}_i - \alpha\mathbf{Z}_i - (1-\alpha)\mathbf{Z}_j\|^2 - E\|\mathbf{Z}_j - \alpha\mathbf{Z}_i - (1-\alpha)\mathbf{Z}_j\|^2$$
$$= (1 - 2\alpha)E\|\mathbf{Z}_i - \mathbf{Z}_j\|^2.$$

Thus $\alpha = 1/2$ and this establishes the equivalence of (i) and (ii). Now suppose (iii) holds: $0 > E\langle \mathbf{Z}_l - \mathbf{Z}_i, \mathbf{Z}_l - \mathbf{Z}_j\rangle$ for all $l \neq i, j$. Then $-E\langle \mathbf{Z}_i, \mathbf{Z}_j\rangle > E\langle \mathbf{Z}_l, \mathbf{Z}_l\rangle - 2E\langle \mathbf{Z}_l, \overline{\mathbf{Z}}\rangle$ where $\overline{\mathbf{Z}} = (1/2)(\mathbf{Z}_i + \mathbf{Z}_j)$. Thus

$$E\langle \mathbf{Z}_l - \overline{\mathbf{Z}}, \mathbf{Z}_l - \overline{\mathbf{Z}}\rangle < E\langle \overline{\mathbf{Z}}, \overline{\mathbf{Z}}\rangle - E\langle \mathbf{Z}_i, \mathbf{Z}_j\rangle = (1/4)E\|\mathbf{Z}_i - \mathbf{Z}_j\|^2$$
$$= (1/2)E\|\mathbf{Z}_i - \overline{\mathbf{Z}}\|^2 + (1/2)E\|\mathbf{Z}_j - \overline{\mathbf{Z}}\|^2$$
$$= \lambda = E\|\mathbf{Z}_i - \overline{\mathbf{Z}}\|^2 = E\|\mathbf{Z}_j - \overline{\mathbf{Z}}\|^2,$$

which proves that (iii) implies (ii) by Theorem 3.

Conversely, suppose (ii) holds. Theorem 1 implies $E\|\mathbf{Z}_l - \overline{\mathbf{Z}}\|^2 < \lambda$ for every $l \neq i, j$, where $\lambda = (1/4)E\|\mathbf{Z}_i - \mathbf{Z}_j\|^2$ and $\overline{\mathbf{Z}} = (1/2)(\mathbf{Z}_i + \mathbf{Z}_j)$. A simple calculation shows that this is equivalent to the assertion that $0 > E\langle \mathbf{Z}_l - \mathbf{Z}_i, \mathbf{Z}_l - \mathbf{Z}_j\rangle$ for all $l \neq i, j$, so the proof of the converse and hence of the theorem is complete. □

PROOF OF COROLLARY 6. Suppose to the contrary $f_{ai} > 1/2$ for some i, say $i = 1$ for simplicity of exposition. By Theorem 1, $0 = E\|\mathbf{Z}_1 - \overline{\mathbf{Z}}\|^2 - E\|\mathbf{Z}_j - \overline{\mathbf{Z}}\|^2$ for every $j \in \bar{\mathscr{I}}$. But

$$\begin{aligned} E\|\mathbf{Z}_1 - \overline{\mathbf{Z}}\|^2 - E\|\mathbf{Z}_j - \overline{\mathbf{Z}}\|^2 &= E\|\mathbf{Z}_1\|^2(1 - 2f_{a1}) - 2f_{aj}E\langle \mathbf{Z}_1, \mathbf{Z}_j\rangle \\ &\quad - \sum_{k \neq 1, j} 2f_{ak}E\langle \mathbf{Z}_1, \mathbf{Z}_k\rangle \\ &\quad - E\|\mathbf{Z}_j\|^2(1 - 2f_{aj}) + 2f_{a1}E\langle \mathbf{Z}_1, \mathbf{Z}_j\rangle \\ &\quad + \sum_{k \neq 1, j} 2f_{ak}E\langle \mathbf{Z}_j, \mathbf{Z}_k\rangle \\ &= (1 - 2f_{a1})\big(E\|\mathbf{Z}_1\|^2 - E\langle \mathbf{Z}_1, \mathbf{Z}_j\rangle\big) \\ &\quad - (1 - 2f_{aj})\big(E\|\mathbf{Z}_j\|^2 - E\langle \mathbf{Z}_1, \mathbf{Z}_j\rangle\big) \\ &\quad - \sum_{k \neq 1, j} 2f_{ak}\big(E\langle \mathbf{Z}_1, \mathbf{Z}_k\rangle - E\langle \mathbf{Z}_j, \mathbf{Z}_k\rangle\big). \end{aligned}$$

But $-(1 - 2f_{aj}) = (1 - 2f_{a1}) - \Sigma_{k \neq 1, j} 2 f_{ak}$. Thus $0 = (1 - 2f_{a1}) E\|\mathbf{Z}_1 - \mathbf{Z}_j\|^2 - \Sigma_{k \neq 1} 2 f_{ak} E\langle \mathbf{Z}_j - \mathbf{Z}_1, \mathbf{Z}_j - \mathbf{Z}_k \rangle$. So $f_{a1} > 1/2$ implies $0 > \Sigma_{k \neq 1} f_{ak} E\langle \mathbf{Z}_j - \mathbf{Z}_1, \mathbf{Z}_j - \mathbf{Z}_k \rangle$ for every $j \neq 1$, $j \in \bar{\mathscr{I}}$. Thus

$$0 > \sum_{j \neq 1} \sum_{k \neq 1} f_{aj} f_{ak} E\langle \mathbf{Z}_j - \mathbf{Z}_1, \mathbf{Z}_j - \mathbf{Z}_k \rangle$$

or

$$0 > \sum_{k \neq 1} \sum_{j \neq 1} f_{ak} f_{aj} E\langle \mathbf{Z}_k - \mathbf{Z}_1, \mathbf{Z}_k - \mathbf{Z}_j \rangle,$$

where the last inequality is obtained from its predecessor simply by interchanging the indices of summation, j and k. But the sum of the right-hand sides of these last two inequalities is $\Sigma_{j \neq 1} \Sigma_{k \neq 1} f_{aj} f_{ak} E\|\mathbf{Z}_j - \mathbf{Z}_k\|^2 > 0$, and this is a contradiction. Thus $f_{a1} > 1/2$ is impossible and the conclusion of the corollary obtains. □

PROOF OF THEOREM 8. Theorem 3 and (15) imply that

$$\lambda = \gamma_i(1 - 2f_{ai}) + \sum f_{aj}^2 \gamma_j, \qquad i \in \bar{\mathscr{I}}. \tag{16}$$

If $\gamma_i = 0$ $i \in \mathscr{I}$ for otherwise, $\lambda = \Sigma f_{aj}^2 \gamma_j$, $\mathscr{I} = \phi$ (for $\gamma_j < 0$ is impossible), hence $f_{ai} = 1/2$, $i \in \bar{\mathscr{I}} = \{1, \dots, K]$ and $K = 2$ contrary to our assumptions. Thus $f_{ai} = 0$ in this case. Now (16) implies $f_{ai} = (1/2)(1 - \lambda\gamma_i^{-1})$ for all $i \in \bar{\mathscr{I}}$ where, with an abuse of notation, λ now replaces $\lambda - \Sigma f_{aj}^2 \gamma_j$. At the same time, Theorem 3 implies $\lambda > \gamma_i$ for all $i \in \mathscr{I}$. Thus $\mathscr{I} = \{1, \dots, m\}$ for some m and $\bar{\mathscr{I}} = \{m + 1, \dots, K\}$, therefore. Finally $\Sigma f_{aj} = 1$ implies $\lambda = (K - m - 2)(\Sigma_{i=m+1}^{K} \gamma_i^{-1})^{-1}$.

It remains to determine m explicitly by the requirement that $f_{ai} > 0$, $i \in \bar{\mathscr{I}}$. That such an integer m exists follows from considering

$$L(m) \triangleq (K - m - 1)\tau(m) - \tau(m) - \cdots - \tau(K), \qquad m = 1, \dots, K,$$

where $\tau(i) = \gamma_i^{-1}$ for all i. Observe that $L(K) = -2\tau(K) < 0$. Suppose $L(m) \geq 0$. Then

$$\begin{aligned} L(m - 1) - L(m) &= [K - (m - 1) - 1]\tau(m - 1) - \tau(m - 1) - \cdots - \tau(K) \\ &\quad - [K - m - 1]\tau(m) + \tau(m) + \cdots + \tau(K) \\ &= [K - m - 1][\tau(m - 1) - \tau(m)] \\ &\geq 0. \end{aligned}$$

Thus L has at most one sign change.

But

$$\begin{aligned} \lambda - \gamma_{m+1} &\propto \lambda\tau(m + 1) - 1 \\ &\propto [K - (m + 1) - 1]\tau(m + 1) - \tau(m + 1) - \cdots - \tau(K) \\ &= L(m + 1) \end{aligned}$$

for $m = 1, \dots, K - 1$. Thus either $\lambda < \gamma_i$ for all i and hence $m = 0$, or there is a unique m for which $\gamma_m < \lambda \leq \gamma_{m+1}$ when λ (a function of m) is as defined above. This completes the proof. □

Acknowledgments. The authors benefitted from stimulating discussions with Dr. Jonathan Berkowitz and Dr. Harry Joe about the problems treated here. Dr. Maurice Queyranne provided invaluable advice. Helpful comments were provided by the SIMS research group at the University of British Columbia, and by an anonymous referee. This manuscript was revised while the second author was on leave at the University of Bath, which generously provided the necessary facilities.

REFERENCES

CASELTON, W. F. and HUSAIN, T. (1980). Hydrologic networks: information transmission. *Water Resources Planning and Management Division, A.S.C.E.* **106** 503–520.

CASELTON, W. F., KAN, L. and ZIDEK, J. V. (1992). Quality data networks that minimize entropy. In *Statistics in the Environmental and Earth Sciences* (P. Guttorp and A. Walden, eds.). Griffin, London.

CASELTON, W. F. and ZIDEK, J. V. (1984). Optimal monitoring network designs. *Statist. Probab. Lett.* **2** 223–227.

ELFVING, G. (1952). Optimum allocation in linear regression theory. *Ann. Math. Statist.* **23** 255–262.

FEDOROV, V., LEONOV, S. and PITOVRANOV, S. (1987). Experimental design technique in the optimization of a monitoring network. In *Model-oriented Data Analysis: Proceedings Eisenach, GDR, 1987. Lecture Notes in Econom. and Math. Systems* (V. Fedorov and H. Lauter, eds.). Springer, New York.

FEDOROV, V. and MUELLER, W. (1988). Two approaches in optimization of observing networks. In *Optimal Design and Analysis of Experiments* (Y. Dodge, V. V. Fedorov and H. P. Wynn, eds.) 239–256. North-Holland, Amsterdam.

FEDOROV, V. and MUELLER, W. (1989). Comparison of two approaches in the optimal design of an observation network. *Statistics* **3** 339–351.

GREEN, R. H. (1979). *Sampling Design and Statistical Methods for Experimental Biologists.* Wiley, New York.

KIEFER, J. (1959). Optimum experimental design. *J. Roy. Statist. Soc. Ser. B* **21** 272–319.

LINDLEY, D. V. (1956). On the measure of the information provided by an experiment. *Ann. Math. Statist.* **27** 968–1005.

LINTHURST, R. A., LANDERS, D. H., EILERS, J. M., BRAKKE, D. F., OVERTON, W. S., MEIER, E. P. and CROWE, R. E. (1986). Characteristics of lakes in the eastern United States **1**. Population descriptions and physico-chemical relationships. Technical Report, EPA/600/4-86/007a, U.S. Environmental Protection Agency, Washington, DC.

MILLARD, S. P. and LETTENMEIER, D. P. (1986). Optimal design of biological sampling programs using the analysis of variance. *Estuarine, Coastal and Shelf Science* **22** 637–656.

MITCHELL, T. J. (1974). An algorithm for construction of D-optimal designs. *Technometrics* **16** 203–210.

PAZMAN, A. (1986). *Foundations of Optimum Experimental Design.* Reidel, Dordrecht.

SCHUMACHER, P. and ZIDEK, J. V. (1988). Using prior information in designing point impact detection networks. SIMS Technical Report 120, Dept. Statistics, Univ. British Columbia.

SHEWRY, M. and WYNN, H. (1987). Maximum entropy sampling. *Journal of Applied Statistics* **14** 165–170.

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

SMITH, K. (1918). On the standard deviations of adjusted and interpolated values of an observed polynomial function and its constants and guidance they give towards a proper choice of the distribution of observations. *Biometrika* **12** 1–85.

VERDINELLI, I. (1992). Advances in Bayesian experimental design. In *Bayesian Statistics* **4** (J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith, eds.). Oxford University Press, New York.

WU, S. and ZIDEK, J. V. (1992). An entropy based analysis of data from selected NADP/NTN network sites for 1983–86. *Atmospheric Environment.* **26A** 2089–2103.

ZIDEK, J. V. (1984). Detailed statistical approach to sediment chemistry monitoring. In *Beaufort Sea Monitoring Program Workshop Synthesis and Sampling Design Recommendations*. Technical Report, SEAMOcean, Inc. and Dept. Statistics, Univ. Washington.

STATISTICAL SCIENCES, INC.
SUITE 500, 1700 WESTLAKE N.
SEATTLE, WASHINGTON 98109

DEPARTMENT OF STATISTICS
2021 WEST MALL
UNIVERSITY OF BRITISH COLUMBIA
VANCOUVER, BRITISH COLUMBIA
CANADA V6T 1Z2