Statistics & Probability Letters 2 (1984) 223–227
North-Holland

OPTIMAL MONITORING NETWORK DESIGNS

W.F. CASELTON

Department of Civil Engineering, University of British Columbia, Vancouver, Canada

J.V. ZIDEK

Department of Statistics GN-22, University of Washington, Seattle, WA 98195, USA

Received October 1982
Revised February 1984

Abstract: The selection of a monitoring network is formulated as a decision problem whose solutions would then be optimal. The theory is applied where the underlying field has a multivariate normal probability structure.

Keywords: decision analysis, information theory, network design, proper local utility, monitoring networks.

1. Introduction

In environmental management and resource development, for example, it is often realized only retrospectively, that earlier and larger expenditures would have been justified to create a monitoring network of more adequate density and duration. This is always the case where long term monitoring is envisaged because the designer cannot fully foresee all of the future benefits derivable from a network by all possible users and uses. And even where he can see a benefit, he can still be ignorant of the procedures or models by which the overall regional information might be extrapolated from the possibly sparse network data. Thus, the designer may be obliged to assume, arbitrarily, both a utility function and a model before he can perform some kind of design analysis. This arbitrariness tends to undermine network design methodologies and as a consequence weaken the case for adequate networks. A methodology to alleviate these difficulties is the subject of this paper.

In the approach taken here, the severe problem of specifying probability distributions on uncountably infinite dimensional function spaces are avoided by adopting as a surrogate of the true environment in a region, the view provided by a (hypothetical) network of high station density. The performance of any prospective network is then judged by how well it conveys the information contained in this ultimate (but hypothetical) network view. Our analysis concentrates on this ultimate network of m locations, or more precisely, the state of nature $\psi = \{\psi^1, \ldots, \psi^m\}$ corresponding to the network. The column vector, ψ^j, consists of quantities determined at different times or for different attributes at location $j = 1, \ldots, m$.

The design problem is that of determining a subset of all the locations $\{1, \ldots, m\}$, say of size $n \ll m$, at which to locate monitoring stations. These stations would then yield measurements, $z = \{z^1, \ldots, z^n\}$, of the quantities of interest of these n sites, and the measurement can be used to make inferences about ψ.

Since the possible uses of the network's data are neighter unique nor entirely foreseen we recommend as the objective, an expression of the designer's uncertainty about ψ after observing z. We assume that the uncertainty about the ψ^i's at the

Research supported by the National Science and Engineering Research Council of Canada.

gauged sites will have been completely eliminated except for a residual, small component due to measurement error. The uncertainty about ψ will thus derive almost entirely from uncertainty about $\zeta = \{\zeta^1, \ldots, \zeta^{m-n}\}$, the component of ψ representing the ungauged sites. This uncertainty may be expressed as a conditional probability density function, $\hat{p}(\zeta \mid z)$ after z has been observed. This density, $\hat{p}$, need not equal the designer's actual posterior density, p. His optimal choice maximizes his posterior expected utility, $E[u(\hat{p}, \tilde{\zeta}) \mid z]$, after observing z. Here and below a tilde is placed over symbols such as ζ to indicate the quantities with respect to which the expectation or conditional expectation is being computed.

An objective of the theory of design is the determination of a broad reasonable class of prospective utilities from which a particular designer might choose his utility. This reduces the arbitrariness of his choice and simplifies his task if the class is not unduly large. A normative approach by which to achieve this objective starts with a statement of simple qualitative conditions and concludes with their translation into a restrictive class of utilities. If the designer finds these conditions reasonable, then his search for his particular utility is confined to this class. Our first condition is that the best subset of gauged locations should reduce as much as possible the uncertainty about ζ. A second condition is the widely accepted requirement that $\hat{p} = p$ be optimal (cf. Bernardo, 1979).

Bernardo (1979) shows how both of these requirements can be met. He shows that if u is *proper*, i.e. $\int u(g, \zeta) f(\zeta) \mathrm{d}\zeta$ is maximized over all densities g at $g = f$ (uniquely), and *local*, i.e. $u(g, \zeta) = u(g(\zeta), \zeta)$, then

$$u(g, \zeta) = A \log g(\zeta) + B(\zeta) \tag{1.1}$$

under appropriate regularity conditions.

With this choice for u it follows that

$$E[u(\hat{p}, \zeta) \mid z] = E[A \log \hat{p}(\tilde{\zeta} \mid z) + B(\tilde{\zeta}) \mid z]$$

after z is observed. The expectation of this quantity over z would give the expected utility before z has been observed.

If no monitoring network exists the maximum achievable utility is $E[A \log p_0(\tilde{\zeta}) + B(\tilde{\zeta})]$ where p_0 denotes the prior density of ζ. Thus the gain in expected utility provided by the network is $AE \log [\hat{p}(\tilde{\zeta} \mid \tilde{z}) / p_0(\tilde{\zeta})]$. The optimal choice, $\hat{p} = p$ yields the difference $AI(\zeta; z)$ where I is the measure of information originally proposed by Shannon (1948). Lindley (1956) adopted the same measure in the context of the design of experiment. Bernardo (1979) describes the coefficient A as the utility per unit of information.

We interpret the network design problem as that of choosing a set, k, of n monitoring stations where n is specified (see Section 3 for further discussion of this point). Let K denote the set of all possible n-location sets, k, and let $z(k)$ denote the prospective set of measurements which would be obtained from k and $\zeta(k)$, the corresponding ψ-values at ungauged sites. Then the optimal $k = k^*$ yields

$$z^* = z(k^*), \quad \zeta^* = \zeta(k^*)$$

and

$$I(\zeta^*, z^*) = \max\{ I[\zeta(k); z(k)] : k \in K \}.$$

The optimal design(s), k^*, will depend on parameters which would need to be specified. Caselton and Husain (1980) propose that these parameters be estimated from data provided by a temporary dense network. To make this proposal feasible in the context of particular interest to these authors, that of hydrology, Caselton initiated the development and production of a suitable low-cost microprocessor-based data logging instrument. This can be installed at widely distributed points on the grid and provide the preliminary data necessary for the determination of k^*, the points at which the much more expensive, permanent station can be established.

In the next section these ideas are developed further for a multivariate normal model and consequences are discussed in Section 3 where some background and related results are given.

2. Multivariate normal case

Order the m-locations in some way. Assume $\{\psi^j\}$ are column vectors of the same dimension so that the same is true of the n elements $z = z(k)$ and $\zeta(k)$ for any $k \in K$. Then (ζ, z) may be represented as an $m \times q$ matrix.

Assume $(\tilde{\xi}, \tilde{z}) \sim N(\mu, \Sigma)$. Here Σ, the $mq \times mq$ covariance, may be defined as that of the column vector $(\tilde{\xi}_1, \tilde{z}_1, \ldots, \tilde{\xi}_p, \tilde{z}_q)^{\mathrm{T}}$ where $(\tilde{\xi}_i, \tilde{z}_i)$ denotes the i-th row of $(\tilde{\xi}, \tilde{z})$. This would usually be highly structured, depending on the temporal and spatial patterns underlying the model.

It is easily shown using results of Anderson (1958, pp. 24, 29) and Kullback (1959) that

$$I(\zeta; z) = -\tfrac{1}{2}(\log|I - R|) \tag{2.1}$$

where $R = \operatorname{diag}\{\rho_1^2, \ldots, \rho_m^2\}$, and $\rho_1^2 \geqslant \cdots \geqslant \rho_m^2$ are obtained by diagonalizing

$$\eta \triangleq \Sigma_{11}^{-1/2}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}\Sigma_{11}^{-1/2}.$$

Here Σ_{11}, Σ_{12} and Σ_{22} are, respectively, $\operatorname{cov}(\tilde{\xi})$, $\operatorname{cov}(\tilde{\xi}, \tilde{z})$ and $\operatorname{cov}(\tilde{z})$. It is well known (cf. Anderson, 1958, p. 296) that these ρ's are the canonical correlation coefficients and $\rho_{n+1} = \cdots = \rho_m = 0$. Thus $I(\zeta; z) = -\frac{1}{2}\log\prod_{i=1}^{n}(1 - \rho_i^2)$. This expression is described by Osteyee and Good (1974) as the expected mutual information between two Gaussian processes.

To conclude this section two examples are considered.

Example 2.1. Suppose $q = 1$ and $\zeta^j = u^j + w$, $j = 1, \ldots, m - n$ and $z^j = v^j + w, j = 1, \ldots, n$. The $\{\tilde{u}^j\}$ and $\{\tilde{v}^j\}$ are independent within and between each group and they are all independent of $\tilde{w}$. Let $\sigma_u^2 = \operatorname{Var}(\tilde{u}^j)$, $\sigma_v^2 = \operatorname{Var}(\tilde{v}^j)$, for all j, and $\operatorname{Var}(\tilde{w}) = \sigma_w^2$.

Here, w may be regarded as the result of a common component event such as precipitation attributable to large (i.e., synoptic) scale storm structures while the u^j's might be thought of as representing local variations in precipitation due to topography and smaller scale storm structures such as convective cells.

Finally, the $\{v^j\}$ embody not only the local variations but measurement error as well, so that $\sigma_v^2 \geqslant \sigma_u^2$. The symmetry in our model implies that all n-station networks are equivalent, so our concern will focus on other effects.

It is easily shown that

$$\Sigma_{11} = \sigma_u^2 I_{m-n} + \sigma_w^2 j_{m-n}^{\mathrm{T}} j_{m-n}, \quad \Sigma_{12} = \sigma_w^2 j_{m-n}^{\mathrm{T}} j_n$$

and

$$\Sigma_{22} = \sigma_v^2 I_n + \sigma_w^2 j_n^{\mathrm{T}} j_n,$$

where I_r denotes the r-dimensional identity matrix and j_r is the r-dimensional row vector all of whose elements are 1. The matrix

$$\Sigma_{11}^{-1}\Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = n\mu\nu j_{m-n}^{\mathrm{T}} j_{m-n},$$

where

$$\mu = \sigma_w^2\left(\sigma_v^2 + n\sigma_w^2\right)^{-1}$$

and

$$\nu = \sigma_w^2\left(\sigma_u^2 + (m-n)\sigma_w^2\right)^{-1},$$

has the same nonzero eigenvalue as η, namely, $\rho_1^2 = n(m-n)\mu\nu$. Because m is large $\rho_1^2 = n\mu$, approximately, so $I(\zeta; z) = -\frac{1}{2}\log(\sigma_0^2[n + \sigma_0^2]^{-1})$, approximately, where $\sigma_0^2 = \sigma_v^2/\sigma_w^2$.

Observe that $I = 0$ as it must be when $n = 0$, that is, when no stations are gauged. When $\sigma_0^2 \to 0$, $I \to \infty$. A small value of σ_0^2 results when the combined instrument error and local variation are small compared to the variation due to the common event, the storm itself, for example. Any n-station monitoring network will then be an effective information transmitter. Of course, as $n \to \infty$, $I \to \infty$ as well.

If the cost of gauging sites can be expressed in the same utility unit as A in (1.1) and is C utility units per gauged site, then the number of stations, n^*, producing the maximum net benefit is given by $n^* = 2AC^{-1} - \sigma_0^2$.

This example can easily be extended to a case where the sites are not exchangeable as they are here. To do so involves the stratification of the overall region into homogeneous blocks. Additional components of variation, like the overall component w, are introduced, one for each block. This introduces an additional source of intra-block correlation beyond that already provided by w. The cross-correlation within the blocks may vary from block-to-block. However, no additional inter-block correlation is introduced.

The analysis of this very applicable extension of Example 2.1 is a numerical exercise once all the parameters have been specified. The best single station will tend to be found in a large block with a large intra-class correlation. Finding the optimal n-station network entails formidable computing costs and this invites a search for sub-optimal

approximate solutions. This subject will not be addressed here.

Example 2.2. A multivariate space-time system is envisaged with the same attributes being determined on q occasions. The spatial covariance matrix between sites i and j (whether or not either or both are gauged) is $\gamma_{ij}\tau$, $i, j \in \{1,\ldots,m\}$ where $\tau > 0$ and $\Gamma = (\gamma_{ij}) > 0$ are $q \times q$ and $m \times m$ matrices respectively. Then $|I - \eta| = |\Sigma_{11}^{-1}\Sigma_{11\cdot 2}|$, where $\Sigma_{11} = \Gamma_{11} \otimes \tau$, and $\Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} = \Gamma_{11\cdot 2} \otimes \tau$. Γ_{11} is essentially the spatial covariance matrix for ungauged sites, and so on. Thus

$$|I - \eta| = |\Gamma_{11}^{-1}\Gamma_{11\cdot 2} \otimes I_q| = |I|^{(m-n)}|\Gamma_{11}^{-1}\Gamma_{11\cdot 2}|^q$$

and

$$I(\zeta; z) = -\tfrac{1}{2}q \log|\Gamma_{11}^{-1}\Gamma_{11\cdot 2}|.$$

This shows that for this covariance structure an optimal design for one occasion will be optimal for measurements taken over several occasions.

3. Summary and discussion

This article presents a normative theory of network design which applies in cases where the possible use for the network's data are neither unique nor entirely foreseen. Our postulates imply, according to Bernardo, that the optimal network must maximize $I(\zeta, z) = E[\log p(\tilde{\zeta} \mid \tilde{z})/p_0(\tilde{\zeta})]$, where p and p_0 denote, respectively, the posterior and prior densities of $\tilde{\zeta}$, the unobserved quantities of interest at the ungauged sites. The measured quantities at the gauged sites constitute z.

Our approach differs from various others which have been proposed. All are concerned with a random field which is a family of vectors, $T = \{T(x); x \in X \subset \mathbb{R}^l\}$. In applications, $l = 2$ and x is a location vector; $T(x)$ will typically be a vector of observations recorded serially.

One major group of methods may be jointly labelled as LINEAR-QUADRATIC. These approaches essentially stipulate as the objective the estimation of $T(x)$ for a single x, or jointly for all x, with respect to quadratic estimation error used as a performance measure. The estimator is restricted to be linear in consequence of which only T's mean and covariance structure and not its joint distribution need be specified. Further gain in simplicity may be achieved by supposing that is wide sense stationary or that its covariance spherically symmetric. The optimal network of stations and estimator jointly minimize the ex pected quadratic estimation error measure of pe formance (see Cabannes, 1979). The difficulty wit this approach lies in the artificiality of its define objectives, the very specific and potentially subo timal estimator, and the estimator performan criterion.

A different approach by Caselton and Husa (1980), which was motivated by a desire to a leviate these difficulties, could be labelled INFOI MATION THEORETICAL. Their proposal i volves the use of Shannon's measure of inform tion to provide the network performance criterio This represents a departure from the statistic methods which have been previously brought bear upon network design. Problems of modelli complex spatial relationships are avoided adopting, as a representation of the true enviro ment in a region, the view provided by a hyp thetical network of extremely high station densi The performance of any prospective network then judged by how well it conveys the inform tion contained in this ultimate but hypothetic network view.

Our approach is related to, but different th that of Caselton and Husain (1980). They focus the discrete case, and Ψ represents the set of possible measurements over the dense grid inclu ing those measurements, z, which are actually ma at the subset of locations assigned to the netwo Measurement error is regarded as negligible and not explicitly accounted for, as it is in the mo introduced here. Because z is a subset of ψ, and these quantities are discrete, a well-known result information theory implies that $I(\psi; z) = I(z;$ $= H(z)$. the joint entropy of z. Thus Caselton a Husain (1980) advocate the choice of k wh maximizes $H(z)$. This criterion is intuitively na ral. The optimal n-station network is the one wh leads to the greatest reduction in uncertainty (e ropy) among n-station networks. A case where t criterion would seem unsatisfactory is one wher independent high entropy sites produce the larg joint entropy. At the same time, a certain l entropy site, which is representative of a large a

in the region to be served by the network, is excluded. Intuition would suggest that the best network should contain at least one of these low entropy sites. However such examples may seem too artificial to constitute a serious criticism of the criterion in discrete cases.

A serious difficulty does arise since $H \to \infty$ as the continuous case is approached through a series of discrete approximants (cf. Osteyee and Good, 1974). One could still adopt as a criterion the formal analogue of H, $-E \log f(\tilde{z})$, where f denotes $\tilde{z}$'s probability density function. However this choice would not, unlike the previous one, have axiomatic support.

We have side-stepped this issue by reinterpreting the objective as an expression of uncertainty about events only at ungauged sites as indicated in Section 1. This gives a methodology which includes both the continuous and discrete cases. It allows us to include instrument error explicitly. It does not require the statement of purposes to which the network's data will be put nor the evaluation of losses and gains which will accrue as a result of actions which might ultimately be taken. It does not even require as an input the procedures by which such actions might be chosen.

The criterion presented in Section 2 for the multivariate normal case could have been reached by a more conventional route. The natural predictor of ζ (when $q = 1$) is $\hat{\zeta} = \mu_1 + \beta(z - \mu_2)$ and the generalized variance of the normalized residuals is

$$|E(\zeta - \hat{\zeta})^{\mathrm{T}} \Sigma_{11}^{-1} (\zeta - \hat{\zeta})| = |\Sigma_{11}^{-1} \Sigma_{11 \cdot 2}| = \prod (1 - \rho_i^2).$$

This is minimized by a network which maximizes $\prod(1 - \rho_i^2)^{-1}$.

While the models used in the examples are not intended to be completely realistic, the results obtained are of practical interest. As a result of an approximation based on the fact that m is large, the expression for information obtained in Example 2.1 does not depend upon the magnitude of the independent variation at the ungaugued sites represented by u_i. The common variation at both gauged sites and the independent variation at gauged sites are the principal determinants of information transmitted by the network.

The linear relationship between the optimal network size and the ratio of the unit value of information to unit station cost points to a possible economic basis for sizing networks. The principal difficulty lies in establishing the value per unit of information in terms commensurate with station costs. The second nonnegative term in this expression might be construed as being the inverse of the signal to noise ratio at gauged sites. As this signal to noise ratio decreases it has the effect of reducing the size of the network which produces the maximum net benefit.

The implication of Example 2.2, that the optimal design is independent of the temporal covariance structure and number of sampling occassions, may well seem surprising. This result obtains because τ determines the temporal covariance at every site. If, for example, there were a marked degree of autocorrelation at the ungauged and gauged sites, the same would be true of the temporal cross-correlations. The information contained in all of the latter about all of the former would be the same as in the case of a common, low degree of temporal dependence.

References

Anderson, T.W. (1958), *An Introduction to Multivariate Statistical Analysis* (Wiley, New York).

Bernardo, J.M. (1979), Expected information as expected utility, *The Annals of Statistics*. **7** (3), 686–690.

Cabannes, A.F. (1979), Estimation of random fields from network observations, Technical Report No. 26, Department of Statistics, Stanford University, Stanford, California.

Caselton, W.F. and T. Hussain (1980), Hydrologic networks: Information transmission, *Journal of the Water Resources Planning and Management Division, A.S.C.E.*, **106** (WR2), 503–520.

Kullback, S. (1959), *Information Theory and Statistics* (Wiley, New York).

Lindley, D.V. (1956), On a measure of the information provided by an experiment, *Ann. Math. Statist.* **27**, 986–1005.

Osteyee, D.B. and I.J. Good (1974), *Information, Weight of Evidence, the Singularity between Probability Measures and Signal Detection* (Springer, New York).

Shannon, C.E. (1948), A mathematical theory of communication, *Bell System Technology Journal* **27**, 379–423, 623–656.