Bayesian Sequential Design of a Network of Sensors

by
Murray K. Clayton
James Shepherd
Chin–Fu Hsiao
and
Shun–Hwa Li
Abstract

This work is motivated by an interest in designing a pyranometer network for predicting
insolation for farms in Wisconsin. More generally, we discuss the construction of a network of
sensors when two-dimensional spatial information is present. Our approach is Bayesian, and is
based on a model developed by Handcock and Stein. We demonstrate how to use this model in
a sequential approach to network design and, by using existing satellite data, we illustrate the
approach for a number of optimality criteria.

Key words and phrases: kriging, variogram, spatial statistics
1 Introduction

The work in this paper is motivated by a problem that arises in agriculture, although it has the potential for applications in many other settings. Our particular interest is in predicting the “insolation” — the amount of solar radiation — for any one of a number of sites in the state of Wisconsin. Such information is useful to farmers, for example, for determining irrigation needs (Bland and Clayton, 1994).

One method for predicting insolation is to use satellite data, although such data are expensive to obtain. An alternative is to use a network of relatively inexpensive pyranometers — ground-based sensors designed to measure insolation. Statistical methods such as kriging (best linear unbiased prediction) can then be used to predict insolation at sites located away from pyranometers.

The general question that then arises is how to design the network of pyranometers. A number of more specific questions arise. For example, if a partial network is already in place, how should we add additional sensors to the network such that the quality of predictions between sensor sites is high? This is also an issue of economy: a dense network will usually offer precise predictions, but at a high cost. Instead we will more typically want to maximize the information available from each new sensor added to the network. This implies a sequential approach: add sensors one at a time, such that each sensor is added in an optimal manner.

Although perhaps less common, a parallel question also arises: if an existing network has numerous sensors in it, are there sensors that might be removed from the network without much sacrifice in information quality? This and the preceding question presume an existing network. The larger question might be: how can we efficiently design a network from scratch? Finally, in some situations, there might be a variable value to the information provided by the network. For example, in predicting insolation, there is limited interest in having high quality predictions in regions where there is little agricultural need for such predictions, especially if, at the same time, predictions are of low quality in areas of agricultural importance.

This paper is directed toward two of the above questions. We seek a method for expanding an existing, partial network, proceeding sequentially, and our focus will be on minimizing the variance of predictions over a region, noting that some regions might be of greater importance than others.

There have been several approaches proposed in the literature for addressing the above goals in network design. Notable in this are the several papers of Zidek and his colleagues (Caselton and Zidek, 1984, Wu and Zidek, 1992, Le and Zidek, 1992, Guttorp, Sampson and Newman, 1992, Guttorp, Le, Sampson and Zidek, 1993). Generally, these papers are based on Bayesian methods using an entropy criterion for assessing the quality of prediction. We vary from this approach first in that our model is closer to that of Handcock and Stein (1993), and is thus, as noted by Handcock and Stein, a more parametric approach than that of Zidek et al. Second, instead of using entropy for assessing the quality of a network, we base our evaluations on the variance of the resulting predictions: a measure more akin to squared error loss.

Other authors have used variance for assessing the quality of prediction, including, for example, Rouhani (1985), although the approach there is nonBayesian. A wide range of optimality criteria, many arising originally in classical experimental design, are discussed by Mueller (1998). This includes the work of Fedorov and Mueller (1988, 1989) who consider $D$-optimality for estimation of model parameters; and the work of Zimmerman and Homer (1991) and Mueller and Zimmerman (1999) who consider network design where the question of interest is estimation of the variogram (as opposed to prediction of observations). Although some of the design criteria are related, a considerably different modeling approach is taken by Nychka and Saltzman (1998), who use a random coefficient model for describing nonstationary ozone data.
Some authors apply multiple criteria for assessing network performance — Haas (1992), for example, considers both the relative error and the variance of the relative error of estimation, while Bogardi, Bardossy and Duckstein (1985) discuss multicriterion network design. The topic of combining design criteria is also covered by Mueller (1998).

Many of the above authors explicitly or implicitly suggest a sequential approach to network design. Nonsequential approaches can also be found in the literature; noteworthy is the work of Matérn (1986), Yfantis, Flatman and Behar (1987), and Johnson, Moore and Ylvisaker (1990). (See also Cressie, 1991, Section 5.6.)

Although not connected specifically to the problem of network design, we note in passing that Cressie (1994) gives a nice review of Bayesian approaches for modeling spatial data. More recently, Ecker and Gelfand (1997) discuss methods for the Bayesian modeling of variograms, although the larger prediction problem is not addressed. Finally, extensive reviews of Bayesian experimental design are provided by Pilz (1991) and Chaloner and Verdinelli (1995), although these authors do not address the special concerns that arise with spatial data.

As noted, a starting point for our approach is the model of Handcock and Stein (1993), which they proposed in developing a Bayesian approach to kriging. We briefly review their work in the next section and outline our approach to network design. In Section 3 we implement this approach for some specific examples. Some discussion and additional commentary are provided in Section 4.

2 Model

First we establish notation. Let $s_0, s_1, s_2, \ldots, s_n, \ldots$ represent locations in the region of interest and let $Z(s)$ denote a measurement (insolation, say) at a generic location $s$. For the network design problem we assume that the existing network consists of sites $s_1, \ldots, s_n$ and that we have a vector of observations $Z_n = (Z(s_1), Z(s_2), \ldots, Z(s_n))^T$ from those sites. For the moment we ignore time: we focus on observations taken on a single day, and assume that we only wish to make predictions for that day.

Consider placing a new sensor at the location $s_{n+1}$. We want $\text{Var}[Z(s_0)|Z_n, Z(s_{n+1})]$ to be small, where $s_0$ represents a farm, say, where it is of interest to predict insolation. We note that $Z(s_{n+1})$ is not actually observed until after the network has been augmented. Therefore, we will focus on the quantity

$$E_{Z(s_{n+1})|Z_n} \text{Var}[Z(s_0)|Z_n, Z(s_{n+1})]$$

where, as the notation implies, the expectation is over the distribution of $Z(s_{n+1})$ given $Z_n$. Finally, we hope that predictions will be of high quality over an entire region $S$, rather than just at a single location $s_0$. We therefore will examine two criteria — our goal being to minimize either of these. Specifically, we consider:

$$M(s_{n+1}) = \max_{s_0 \in S} E_{Z(s_{n+1})|Z_n} \text{Var}[Z(s_0)|Z_n, Z(s_{n+1})]$$

and

$$A(s_{n+1}) = E_{s_0} E_{Z(s_{n+1})|Z_n} \text{Var}[Z(s_0)|Z_n, Z(s_{n+1})].$$

The first criterion may be thought of as a minimax criterion, where we want to pick the new sensor location $s_{n+1}$ so that even the site $s_0$ with the largest expected variance will still yield good predictions (as measured by variance). For the second criterion we will take $E_{s_0}$ to mean Lebesgue measure over the region $S$. Thus, this criterion amounts to finding the new sensor location $s_{n+1}$ such that we minimize the average expected variance over the region of interest. Similar criteria
can be found in Cressie (1991, Sections 4.6.2 and 5.6.1) and are related to $G$ and $A$ optimality criteria. (See also Johnson et al., 1990, and Mueller, 1998.)

To apply either criterion, we must be able to compute the expected variance in expression (1). It will suffice if we can determine the probability density function $f(Z(s)\mid Z_m)$ for an arbitrary location $s$ and arbitrary vector of observations $Z_m$. We will use the model of Handcock and Stein (1993) to effect this; we now briefly review their model.

Suppose that $Z(s)$ is a stationary Gaussian random field with $E[Z(s)\mid \beta] = f(s)^T \beta$ where $f = (f_1, f_2, \ldots, f_q)^T$. We assume that $\text{cov}(Z(s + h), Z(s)\mid \alpha, \theta) = ak_\theta(|h|)$ where $\alpha$ represents a scalar variance parameter and $k_\theta$ represents a correlation function depending on a (vector) parameter $\theta$. We assume that the covariance structure is isotropic, and thus, from now on, will take $h$ to represent the scalar distance between two points. Then, given $\theta$,

$$
\hat{Z}_\theta(s_0) = \lambda_\theta^T Z_m
$$

is the best linear unbiased predictor of $Z(s_0)$ where

$$
\lambda_\theta^T = [k_\theta + F(F^TK_\theta^{-1}F)^{-1}b_\theta]^TK_\theta^{-1},
$$

and where $F(i, j) = f_j(s_i)$, $k_\theta = [k_\theta(s_0 - s_1), \ldots, k_\theta(s_0 - s_m)]^T$, $K_\theta(i, j) = k_\theta(s_i - s_j)$, and $b_\theta = f(s_0) - F^TK_\theta^{-1}k_\theta$.

At this point, following Handcock and Stein, we have parameterized the mean structure in terms of a parameter vector $\beta$, while the covariance structure is parameterized in terms of $\alpha$ and $\theta$. The usual geostatistical approach would be to estimate those, and plug them into the expression for $\hat{Z}_\theta(s_0)$ above. Indeed, this has already been done for $\beta$ (which is appropriate whether we simply plug in the maximum likelihood estimate for $\beta$, or whether we use a Bayesian approach with a flat prior on $\beta$). However, Handcock and Stein warn that using a plug-in estimate of $\theta$ can result in a false sense of the quality of the prediction. They propose a Bayesian approach to address this, and we follow their lead.

We suppose that $\alpha$, $\beta$, and $\theta$ have a joint prior of the form: $\pi(\alpha, \beta, \theta) = 1/\alpha$. (Handcock and Stein allow for a more general prior on $\theta$ than we will use.) In this case (Handcock and Stein, Equation 3.1):

$$
Z(s_0)\mid \theta, Z_m \sim t_{m-\alpha} \left( \hat{Z}_\theta(s_0), \frac{m}{m-\alpha} \hat{\alpha}(\theta) V_\theta \right)
$$

where

$$
V_\theta = 1 - k_\theta^T K_\theta^{-1} k_\theta + b_\theta^T (F^TK_\theta^{-1}F)^{-1} b_\theta.
$$

Moreover, (Handcock and Stein, Equation 3.2)

$$
p(\theta\mid Z_m) \propto \pi(\theta) |K_\theta|^{-1/2} |F^TK_\theta^{-1}F|^{-1/2} \hat{\alpha}(\theta)^{\alpha(m-\alpha)/2}.
$$

Of course, these results can be combined to determine

$$
p(Z(s_0)\mid Z_m) = \int p(Z(s_0)\mid \theta, Z_m) p(\theta\mid Z_m) d\theta.
$$

This gives us the predictive distribution for any generic site $s_0$ given observations on a network of size $m$. As noted, our particular interest is in $E_{Z(s_{n+1})\mid Z_n} \text{Var}(Z(s_0)\mid Z_n, Z(s_{n+1}))$. To calculate this, we take advantage of the standard result, for two random variables $X$ and $Y$, that $\text{Var}(Y) = E_X[\text{Var}(Y\mid X)] + \text{Var}_X[E(Y\mid X)]$. 

3
Thus,
\[
E_{Z(s_{n+1})} \frac{\text{Var}[Z(s_0)|Z_n, Z(s_{n+1})]}{Z_n} = E_{Z(s_{n+1})} \frac{Z_n}{E_{\theta}|Z_n \text{Var}[Z(s_0)|\theta, Z_n, Z(s_{n+1})]} \\
+ E_{Z(s_{n+1})} \frac{Z_n}{E_{\vartheta}|Z_n \text{Var}[Z(s_0)|\vartheta, Z_n, Z(s_{n+1})]} \\
= E_{Z(s_{n+1})} \frac{Z_n}{E_{\theta}|Z_n \left[ \frac{n+1}{n-q-1} \alpha_{n+1}(\theta) V_{\theta,n+1} \right]} \\
+ E_{Z(s_{n+1})} \frac{Z_n}{E_{\vartheta}|Z_n \text{Var}[Z_{\vartheta,n+1}(s_0)].}
\]

In the above, the subscript \(n + 1\) on \(\hat{\alpha}_{n+1}(\theta), V_{\theta,n+1}, \) and \(\hat{Z}_{\vartheta,n+1}(s_0) \) are used to indicate that they depend on the network of size \(n + 1\). For \(V_{\theta,n+1}\), this entails evaluating Equation (5) for the locations \(s_1, \ldots, s_{n+1}\); for \(\hat{Z}_{\vartheta,n+1}(s_0) \) we must evaluate Equation (4), taking into account both the set of \(n + 1\) sensor sites and the set of observations \(Z_n\) plus the (random) observation \(Z(s_{n+1})\).

In our application we have found that \(E_{Z(s_{n+1})} \text{Var}[\hat{Z}_{\vartheta,n+1}(s_0)]\) is small, and so we ignore it. Consequently, we may write
\[
E_{Z(s_{n+1})} \frac{\text{Var}[Z(s_0)|Z_n, Z(s_{n+1})]}{Z_n} \approx E_{Z(s_{n+1})} \frac{Z_n}{E_{\theta}|Z_n \left[ \frac{n+1}{n-q-1} \alpha_{n+1}(\theta) V_{\theta,n+1} \right]} \\
= \frac{n+1}{n-q-1} E_{\theta}|Z_n V_{\theta,n+1} E_{Z(s_{n+1})}|\theta, Z_n \left[ \alpha_{n+1}(\theta) \right] \tag{7}
\]

where the equality follows by changing the order of integration, and where \(\hat{\alpha}_{n+1}(\theta) = \frac{1}{n+1} W^T K_{\theta,n+1}^{-1} W\) and \(W = Z_{n+1} - F_{n+1} \hat{\beta}_{n+1}(\theta)\). To evaluate the expression in Equation (7), we use Markov chain Monte Carlo methods (as outlined in the following section), to handle the expectation \(E_{\theta}|Z_n\).

To deal with \(E_{Z(s_{n+1})}|\theta, Z_n \left[ \alpha_{n+1}(\theta) \right]\), we make a further approximation. Specifically, we take \(\hat{W} \approx Z_{n+1} - F_{n+1} \hat{\beta}_{n}(\theta)\). This has the effect of estimating the spatial trend surface based only on the data \(Z_n\), instead of the “complete” data set \(Z_{n+1}\). With this approximation in place, given \(Z_n\), the only random quantity in \(W\) is the last component of \(Z_{n+1}\), namely, \(Z(s_{n+1})\), which now enters into \(\hat{\alpha}\) in linear and quadratic terms.

To see this in a little more detail, write \(W = (W_1^T|W_2)\), where \(W_1\) has length \(n\), and \(W_2\) is scalar. In addition, partition \(K_{\theta,n+1}^{-1}\) as
\[
K_{\theta,n+1}^{-1} = \left( \begin{array}{cc} A & B \\ B^T & C \end{array} \right)
\]

where \(A\) is \(n \times n\) and the other components have the appropriate dimensions. In that case, we can write:
\[
(n+1)E_{Z(s_{n+1})}|\theta, Z_n \left[ \hat{\alpha}_{n+1}(\theta) \right] \approx W_1^T A W_1 + 2W_1^T B E_{Z(s_{n+1})}|\theta, Z_n \left[ W_2 \right] + CE_{Z(s_{n+1})}|\theta, Z_n \left[ W_2^2 \right] \\
= \left( \begin{array}{c} W_1 \\ W_2 \end{array} \right)^T \left( \begin{array}{cc} A & B \\ B^T & C \end{array} \right) \left( \begin{array}{c} W_1 \\ W_2 \end{array} \right) + C \text{Var}_{Z(s_{n+1})}|\theta, Z_n \left[ W_2 \right]
\]

where \(\hat{W}_2 = E_{Z(s_{n+1})}|\theta, Z_n \left[ W_2 \right] = k_2^T K_{\theta,n+1}^{-1} [Z_n - F_n \hat{\beta}(\theta)]\).

The upshot of this effort is that we have an approximation to \(E_{Z(s_{n+1})}|Z_n \text{Var}[Z(s_0)|Z_n, Z(s_{n+1})]\) that only involves integration over \(\theta|Z_n\). Integration over the distribution of the random \(Z(s_{n+1})\) has been handled algebraically (up to an approximation), and this simplifies the MCMC considerably.
3 Illustration

To illustrate these ideas, we return to the main motivation for this work, the problem of expanding a network of pyranometers for Wisconsin. Providing such an illustration is complicated by the fact that, at the time of this writing, no network of pyranometers currently exists in Wisconsin. We will therefore create an artificial starting point by pretending that a preliminary network of sensors exists corresponding to the stations in the (existing) University of Wisconsin Automated Weather Observation Network. (See Figure 1.) We then make use of a data set consisting of satellite observations of insolation in Wisconsin for 44 (nonconsecutive) days spanning the summer months of 1984. These observations constitute a grid on a spacing of roughly 20 km. (See Bland and Clayton, 1994, for more details regarding this data set.) For each of the 16 sensors in our artificial network, we take the nearest satellite observation to represent the observation at that sensor site. We then implement our approach to determine the next site at which to locate a new sensor.

To proceed, we need to fill in a few additional model details. First, we use $f(s)^T = (1, x, y)$ to model any first-order (mean) patterns. This model corresponds to modeling latitudinal effects plus longitudinal effects that might correspond, roughly, to effects of Lake Michigan along the eastern edge of our region of interest.

We also need to pick a model for $k_\theta(h)$. (In geostatistical terms, this is tantamount to picking a model for the variogram.) Handcock and Stein suggest a Matérn model: $k_{\theta_1, \theta_2}(h) = \frac{1}{2^{\theta_2-1} \Gamma(\theta_2)} \left( \frac{h}{\theta_1} \right)^{\theta_2} K_{\theta_2} \left( \frac{h}{\theta_1} \right)$ where $\theta_1' = \theta_1/(2\sqrt{\theta_2})$ and $K_{\theta_2}$ is a modified Bessel function of order $\theta_2$. This provides a flexible family for modeling the covariance structure. Indeed, this flexibility is important in our modeling efforts: for our data, a contour plot of the posterior for $(\theta_1, \theta_2)$ indicated that there is very little information in the data about $\theta_2$, but the suggestion is that $\theta_2$ should be large. As Handcock and Stein note, $\lim_{\theta_2 \to \infty} k_{\theta_1, \theta_2}(h) = \exp(-h^2/\theta_1^2)$, resulting in a “Gaussian” correlation model. Thus, in what follows, we reduce the vector parameter $\theta = (\theta_1, \theta_2)$ to the scalar $\theta_1$ and fit a Gaussian correlation model. We still evaluate the numerical integrals using MCMC. While this may represent overkill to some extent, it was straightforward to implement and gave results that seemed reasonable.

Finally, we must do a double search: for any potential new site $s_{n+1}$, we must either average or maximize the predicted variance (expression 1) over all possible $s_0$. In practice, we divide the region into a square grid of 121 points and conduct the search over this grid. We experimented with different grid sizes, and found that this worked well, providing relatively quick evaluations while sacrificing little in terms of accuracy.

Figure 2 shows the results of applying our approach to 3 sets of data, corresponding to days 146, 155, and 180 of the data set of Bland and Clayton (1994). In each case we show the results for the criteria $M$ and $A$. These results conform to the intuitive notion that the best place to put a sensor is in a region relatively vacant of existing sensors. For the criterion $A$, although the criterion values (and thus the prediction variance) differ from day to day, the optimal location of a new sensor is relatively constant. This is less true for the criterion $M$. As one colleague put it, using the minimax criterion requires us to try to “make the grumpiest person happy.” As we see, this can result in fairly substantial shifts in the optimal location of a sensor as we change from one day to another. Partly connected to this, we note that the optimal location using the criterion $M$ is not the same as that for $A$, nor would we expect it to be.

It is appealing that the optimal location for a new sensor does not vary much from day to day for criterion $A$. However, even though this variation is small, it is a practical fact that the sensor location cannot be changed daily. Therefore we must chose a single location that will serve for all
Figure 1: University of Wisconsin Automated Weather Observation Network station locations. (Axis scaling is with respect to an arbitrary origin; one unit equals approximately 100 km.)
Figure 2: Optimal location of sensor (X) based on simulated network of 16 locations (●) for Julian days 146, 155, and 180. Contours show values of optimization criterion; left column corresponds to minimax criterion $M$, right column corresponds to average loss criterion $A$. 


days. One way to do this is to simply average the locations obtained from the analyses of each day’s worth of data. Instead we look at a combined loss: Let \( d \) denote a subscript corresponding to a given day, and let \( D \) be the total number of days of interest. The quantity

\[
\frac{1}{D} \sum_{d=1}^{D} w(d) E_{Z_d(s_{n+1})} \{ \text{Var}[Z_d(s_0)|Z_n, Z_d(s_{n+1})] \}
\]

(8)

represents a weighted (through \( w(d) \)) combination of the prediction variances over all the days. (Unequal weighting is useful, for example, if predictions during certain parts of the season are more important than others.) The expression in (8) is then used as a substitute for expression (1) in Equation (2) and Equation (3). Figures 3 and 4 show the results of this. To save computational effort, we averaged over \( D = 10 \) days randomly sampled from the 44 available to us and used \( w(d) \equiv 1 \). We also pooled the daily estimates of the posterior for \( \theta \) to form a pooled estimate of the posterior.

Although we did not do so, we could take a more formal approach by constructing a spatial-temporal model for the data as done by Handcock and Wallis (1994) or Nychka and Saltzman (1998). The fact that the optimal \( s_{n+1} \) did not vary much from day to day for the optimality criterion \( A \) could be taken as an indication that explicit temporal modeling is not too important, or, perhaps, that the temporal correlation is high. (Although we did not see similar stability for the criterion \( M \), this may be more due to the resulting lack of robustness in optimal site location, as noted above, than due to temporal effects per se.)

Thus far we have only exhibited the use of our approach in locating a single new sensor. However, it is quite realistic that we would want to add in several sensors. We proposed doing this one step at a time, as illustrated in Figures 3 and 4 (and continuing to use an unweighted average of \( D = 10 \) days). Using, for example, the average variance criterion, we find that the best location for the next new sensor is \( s_{17} = (0.39, 1.08) \). We then pretend that a sensor was in fact installed at that location and impute data for that location by using the nearest satellite reading from the data of Bland and Clayton (1994). We then ask where the next sensor should be located, based on the 17 sensors now in the network. As we see in the top right plot of Figure 4, we should use \( s_{18} = (0.48, 2.47) \). The bottom plots of Figure 4 show the results of adding, sequentially, the 18th and 19th sensor sites in a similar manner.

We close this section with one further illustration. Just as we might use a weight function \( w(d) \) to emphasize the importance of predictions during particular parts of the season, we might also want to emphasize prediction for particular locations. That is, we might replace Equation (2) or Equation (3) by

\[
M_L(s_{n+1}) = \max_{s_0 \in \mathcal{S}} L(s_0) E_{Z(s_{n+1})} \{ \text{Var}[Z(s_0)|Z_n, Z(s_{n+1})] \}
\]

and

\[
A_L(s_{n+1}) = E_{s_0} L(s_0) E_{Z(s_{n+1})} \{ \text{Var}[Z(s_0)|Z_n, Z(s_{n+1})] \},
\]

where \( L(\cdot) \) is a multiplicative penalty. (A similar notion appears in Cressie, 1991 and Mueller, 1998, among others.) This allows us to put more emphasis on prediction in certain regions. In addition, we may again substitute in the quantity in (8) to combine data over a period of time. A simple example of this is shown in Figure 5. There, the penalty \( L(x, y) = 1 + (y - 1)^2 \) was chosen for illustrative purposes. This penalty has the effect of giving preference to locations in Southern Wisconsin near the line \( y = 1 \), and so it is not surprising that the optimal site is near that line (Figure 5). Although not done here, a more elaborate penalty function could be used to restrict the
Figure 3: Optimal location of sensor (X) based on simulated network of locations (●) pooled across 10 days. Contours values correspond to minimax optimization criterion $M$. Top Left: Optimal location of 17th sensor based on network of 16 sites. Top Right: Optimal location of 18th sensor based on original network of 16 sites, plus sequentially added 17th additional site (*). Bottom: Optimal locations of 19th (Left) and 20th (Right) sensors based on 16 original sensors, plus sequentially added sensors (*).
Figure 4: Optimal location of sensor (X) based on simulated network of locations (●) pooled across 10 days. Contours values correspond to average loss optimization criterion A. Top Left: Optimal location of 17th sensor based on network of 16 sites. Top Right: Optimal location of 18th sensor based on original network of 16 sites, plus sequentially added 17th additional site (*). Bottom: Optimal locations of 19th (Left) and 20th (Right) sensors based on 16 original sensors, plus sequentially added sensors (*).
Figure 5: Optimal location of sensor (X) based on simulated network of locations (●) pooled across 10 days. Contours correspond to minimax criterion $M_L$ (left plot) and average loss criterion $A_L$, where $L(x, y) = 1 + (y - 1)^2$.

network to lie exclusively in Wisconsin, say, by making $L$ extremely large outside the boundaries of the state.

4 Concluding Remarks

In this paper we have outlined an approach for expanding a network of sensors. Our approach is based on augmenting the network in a sequential fashion, and like many sequential design problems, it is convenient to take advantage of Bayesian methods. In this section we comment further on the approach that we have taken.

First, we took a fairly simple approach to modeling the trend surface, using only $f(s)^T = (1, x, y)$. A more sophisticated approach would be based on allowing for the presence of, for example, airborne particulate matter that might reduce insolation near cities. More sophisticated modeling of lake effects could also be added. Because these effects will likely have a temporal pattern, we chose not to model them directly, but to let them be absorbed into the model component accounting for spatial correlation.

Second, we used Markov chain Monte Carlo, specifically, the Hasting-Metropolis algorithm with a symmetric normal kernel to approximate the posterior for $\theta$. As noted, with so few parameters, this approach might be considered overkill. (See the discussion of (Handcock and Wallis, 1994) for some related comments.) On the other hand, from Equation (6) it is clear that some sort of numerical methods must be used and we found MCMC to be effective and efficient in this setting. More to the point, the computational challenge in this problem comes just as much from the optimization search that is necessary as it does from the estimation of the posterior of $\theta$. Because the criteria $M$ and $A$ have multiple local minima, no simple optimization algorithm is effective.
In our examples we have started with a fictitious network based on existing weather stations in Wisconsin, although in fact no such network of pyranometers exists. How, in practice, should we build a network from scratch? Our experience suggests that we cannot start with a single sensor and expand sequentially from there unless we change our priors considerably. This seems to arise because, in our approach, we have used uninformative priors with the intent that the data should have the greater influence on our decisions. However, the effect of that choice is that, when the number of sensor sites is small, then prediction variances grow too large and the criteria $M$ and $A$ become too unstable to reliably design a network.

To deal with this, suppose that a network with a maximum of 20 sensors is desired. We propose that, in the absence of a starting network, a small grid of say, 10, sensors be created. The work of Yfantis et al. (1987), for example, suggests an equilateral triangular grid could be appropriate. Given this grid, add sensors sequentially until 20 sensors have been added, or until the criterion $A$, say, is below a desired threshold. At that point, note that the methods of this paper can also be extended to ask which, if any, sensor should be removed, if there is to be minimal impact on the predictive properties of the network. Having removed that sensor, we might then ask whether, if it were added in any other location, there would be a benefit. If so, then move it. Thus proceeding somewhat analogously to a stepwise approach in regression, we reconfigure the network by moving, adding, or deleting sensors.

From the sequential perspective, the approach proposed in the preceding paragraph, and indeed in this entire paper, is a myopic one (Berger, 1980). For example, when expanding the network, each sensor is added as if it were the only sensor to be added. By contrast, a “two-step” approach would add sensors one at a time, but the first one would be added in recognition of the fact that a second one would be added later. From the standpoint of sequential decision theory, the myopic approach as used here is unlikely to be optimal. However, an optimal strategy for adding sensors would be extraordinarily difficult to find, since it would depend on performing a high dimensional backward induction (Berger, 1980). Conceivably we could find an optimal design through an indirect route: in a related setting Johnson et al. (1990) show that certain space-filling designs correspond to designs with good properties in terms of prediction error. It remains to be seen whether a similar relationship exists between optimal Bayesian sequential designs and designs chosen for their space-filling characteristics. In the interim, we propose the myopic approach outlined in this paper as a tool for network development.

5 Acknowledgements

The authors are grateful to Bill Bland, Dale Zimmerman, Doug Nychka, and Val Fedorov for their helpful comments.

References


