Groundwater Monitoring Network Optimization with Redundancy Reduction

L. M. Nunes1; M. C. Cunha2; and L. Ribeiro3

Abstract: Three optimization models are proposed to select the best subset of stations from a large groundwater monitoring network: (1) one that maximizes spatial accuracy; (2) one that minimizes temporal redundancy; and (3) a model that both maximizes spatial accuracy and minimizes temporal redundancy. The proposed optimization models are solved with simulated annealing, along with an algorithm parametrization using statistical entropy. A synthetic case-study with 32 stations is used to compare results of the proposed models when a subset of 17 stations are to be chosen. The first model tends to distribute the stations evenly in space; the second model clusters stations in areas of higher temporal variability; and results of the third model provide a compromise between the first two, i.e., spatial distributions that are less regular in space, but also less clustered. The inclusion of both temporal and spatial information in the optimization model, as embodied in the third model, contributes to selection of the most relevant stations.

DOI: 10.1061/(ASCE)0733-9496(2004)130:1(33)

CE Database subject headings: Monitoring; Optimization models; Ground-water management.

Introduction

Monitoring network dimension reduction is a need foreseen in Europe and other developed countries because of the already substantial financial effort put into maintaining the present environmental monitoring network. This amounts to many thousands of stations in the European Union (EU) space and the need to develop specific networks imposed by EU norms [see Nixon et al. (1996) for a discussion on this theme]. Rational allocation of financial resources means eliminating redundant stations from existing networks and placing new ones in unsampled locations, where necessary. Which stations from already existing national networks should be included in each of the specific EU networks? This is a question that needs not only expert judgment [as proposed by Nixon (1996)], but also optimization tools like those presented herein.

The optimization of environmental monitoring networks is a subject that has been studied for many years. Much effort has been put into the development of statistical methodologies to design new sampling campaigns and monitoring networks (e.g., Harmancioglu and Alspaslan 1992; Thompson 1992; Harmancioglu et al. 1999). Further developments considering the physical reality and its variability, particularly with groundwater stochastic simulation and modeling, were examined by Massmann and Freeze (1987a,b), Wagner and Gorelick (1989), and Ahlfeld and Pinder (1988) in the context of geostatistical simulation and modeling. These works concentrated either on the optimization of a new network or on expansion of an existing one. Network optimization when there is the need to reduce its dimension has been much less studied. Some of the rare exceptions are the works by Knoopman and Voss (1989), Meyer et al. (1994), Reed et al. (2000), and Grabow et al. (1993).

Monitoring networks for areal mean rainfall events was the subject of early works by Rodriguez-Iturbe and Mejia (1974a,b), where they considered the spatial and temporal variability of mean rainfall. The variance of mean rainfall was calculated as the product of point process variance, a reduction factor due to sampling in time (dependent only on the correlation in time and length of the time series), and a reduction factor due to sampling in space (dependent on the spatial correlation structure, the sampling geometry, and the number of stations). The writers studied random and stratified random sampling schemes, and obtained abacus for different correlation functions, number of stations, and area of the region. Lenton and Rodriguez-Iturbe (1974) further considered the density and location of the stations. Bras and Rodriguez-Iturbe (1975) compared sampling schemes for different point variances, covariance functions, and covariance functions parameter. Bras and Rodriguez-Iturbe (1976) included in the same formalism the cost associated with each station to help choose the best set (i.e., number and position) of stations with less cost and less mean rainfall variance. Delhomme (1978) applied the geostatistical fictitious point method (usually used to assess the quality of covariance models when estimating with kriging) to determine the optimal location of rain gauges. If the number of stations is large, then the dimension of the combinatorial problem may be exhaustively intractable. Pardo-Igüüzquiza (1998) solved this problem with a metaheuristic approach (simulated annealing). Rouhani (1985) also used variance reduction techniques to determine the number and position of groundwater monitoring stations, and also analyzed the robustness and resilience of the vari-
vance reduction analysis. Rouhani and Hall (1988) proposed the incorporation of risk, defined as the weighted sum of the expected value and the estimation variance, in order to correct the “blindness” of the estimation variance to extreme values. The method also considers temporal changes in the hydrologic variables. Loaiciga (1989) also proposed a variance reduction method using time-dependent spatial models (based on space-time means and covariances), with good results. This writer used the mixed-integer programming model of Hsu and Yeh (1989), originally developed for optimum design for parameter identification. Maximum periodicity was allowed in this study, but did not include trade-off analysis between sampling periodicity and further increasing the number of stations. Later developments in the solution of time-dependent models were proposed by Pardo-Igúzquiza (1998), with the inclusion of the climatological variogram (Bastin et al. 1984; Lebel et al. 1987). More extensive reviews on this topic may be found in Loaiciga et al. (1992), Dixon and Chiswell (1996), and Harmancioglu et al. (1999).

Variance reduction techniques use the variance of the estimation error ($\sigma^2_e$) as an indicator of the accuracy of the estimated values. In geostatistics, the mathematical definition of $\sigma^2_e$ means that its value does not depend on the actual values of the measured variables, but on the relative spatial distribution of the measuring locations. Therefore, one may use $\sigma^2_e$ as an indicator of which spatial distribution is best for a sampling network by testing all the combinations available and selecting the combination that minimizes $\sigma^2_e$.

Temporal redundancy is considered to be a function of the similarity between time series. By adopting this approach, stations that show larger differences between them over time and simultaneously have the best spatial distribution are retained. In principle, this would generate a network of monitoring stations smaller than the original, but which best reproduces the temporal and spatial behavior of the state variable. Temporal redundancy reduction had already been proposed by Amorocho and Espildora (1973), Caselton and Husain (1980), Harmonciglo and Yevjevich (1987), Husain (1989), and Harmonciglo and Alspaslan (1992). Harmonciglo et al. (1999) reviewed its applications in the context of information theory. Despite the elegance of this method, it is limited by the need to assume a probability distribution for the variables, which may be unknown or difficult to determine. Moreover the method is particularly well adapted to variables with equal probability distributions (usually normal or lognormal).

Space-time models have already been applied by, e.g., Bastin et al. (1984), Lebel et al. (1987), Buxton and Pate (1994), Dimitrakopoulos and Luo (1994), and Pardo-Igúzquiza (1998). These approaches require complex variogram fitting and are intended to reproduce the primary space-time nested variances. Pardo-Igúzquiza (1998) also used simulated annealing to solve the same problem with good results in a network augmentation problem.

Considered herein is the problem of reducing the dimension of an existing network of groundwater monitoring (GMN) stations from $\Pi$ to $\pi$, $\pi < \Pi$. Both the spatial covariance and available time series of a state variable are used. The problem is solved by replacing one station at a time and evaluating the effect on the cost function. Simulated annealing (SA) is used to approximate the solution. It should also be emphasized that the algorithms usually referenced as SA are in reality simulated quenching because the temperature schedule used is exponential instead of logarithmic [see Ingber (1993) for a discussion of this]. Faster cooling schedules are necessary to circumvent the slow convergence rate of SA, and have been proven to give good results, but without guaranteeing discovery of the global optimum. The exponential schedule is used herein and is referred to as SA in order to maintain the current nomenclature. Three different models for the optimization model are proposed and tested: (1) one considering only spatial information in the form of accuracy of spatial estimates; (2) one that considers only temporal information, by reducing redundancy in the time series; and (3) a model that includes both space and time information.

**Space and Time Models for Network Optimization**

Some groundwater monitoring programs have, in the past, been implemented on the basis of spatial and temporal intensive monitoring programs. These have been shown to require extensive manipulation or expensive and sensitive hardware. Cost rationalization and the need to orient monitoring programs to other specific objectives have made it necessary to reduce the size of the existing monitoring networks. The new size and spatial distribution must, however, be sufficient to ensure that both spatial and temporal variabilities are correctly included in the new design. In many instances some of the monitoring stations must be retained in the future design. This is the case for wells used to produce tap water, stations that detect specific contamination sources, or ones whose position is considered strategic for any other reason. These restrictions are easily implemented in the optimization algorithms if known a priori. A direct comparison method between all time series is used in this study that considers all the available raw data. A new definition for data temporal redundancy is proposed and used to obtain better spatial distribution of monitoring stations. Temporal redundancy and a measure of spatial coverage quality (i.e., variance of the error of estimation obtained by kriging, $\sigma^2_e$) are combined in a single model and subjected to minimization. Simulated annealing (SA) is used to find the solution, with entropy used to help parameterize the SA algorithm.

**Estimation Variance**

Field data within the geostatistical formalism is considered to be the result of random processes of regionalized variables, i.e., of random variables with space coordinates, with some spatial covariance. Regionalized variables are continuous in space and therefore not completely random, but at the same time it is not possible to model them by a deterministic function (or spatial process). They therefore lie between deterministic and stochastic processes (Matheron 1970), thus incorporating the notion of uncertainty in the conception of inference models or in the simulation of variables (Matheron 1970; David 1977; Journel and Huigbrechts 1978). A thorough review of the use of geostatistics for mapping and sampling design appears in Task (1990a,b).

The values of a state variable $z(x)$ at the sampled points in the field can be considered as realizations of a set of random variables $Z(x)$ in a field $\Gamma$. A set of random variables $Z(x)$, defined in a field $\Gamma$ is a random function $Z(x)$. Some restrictions with respect to stationarity are needed. The most common theory considers that the distribution function is invariant by translation and intrinsic stationarity. If increments are made at step $h$, then the resulting expression is the variogram:

$$\gamma(h) = \frac{1}{2}E[(Z(x+h) - Z(x))^2]$$

(1)

It is a necessary and sufficient condition that the variogram is a negative definite function.

If an estimate of the mean value of a state variable in an area $A$ from values at locations $x_i$, $Z(x_i)$, inside or outside the area is $V$, then a linear estimation of $V$ can be obtained from $p$ data points by
V = \sum_{i=1}^{p} \kappa_i \cdot Z(x_i) \tag{2}

which is unbiased if the sum of the weights \( \kappa \) is one. This is a common requirement in several methods, and also in kriging. The kriging system is (Journel and Huijbregts 1978)

\[ \sum_{i=1}^{n} \kappa_i \cdot \gamma(h) + \mu = \gamma(h_{iA}) \tag{3} \]

\[ \sum_{i=1}^{n} \kappa_i = 1 \]

where \( \mu = \) Lagrange parameter; and \( \gamma(h_{iA}) = \) average variogram between the point \( i \) and the area \( A \) when one extreme of the vector \( h \) is fixed in \( x_i \), and the other extreme describes the area \( A \) independently. The estimation variance is expressed by (Journel and Huijbregts 1978)

\[ \sigma_{E}^{2} = \sum_{i=1}^{n} \kappa_i \cdot (\gamma(h_{iA}) - \gamma(h_{AA})) + \mu \tag{4} \]

with \( \gamma(h_{AA}) \) being the average variogram inside \( A \). The estimation variance is a measure of the estimation accuracy of \( V \). Because \( \sigma_{E}^{2} \) only depends on the geometric configuration of the data points, and, once a variogram model is defined, it is possible to change data locations and calculate the estimation variance again. The spatial arrangement of points that minimizes \( \sigma_{E}^{2} \) has the lowest estimation error and therefore best reflects the spatial correlation introduced in the variogram model. By minimizing \( \sigma_{E}^{2} \) the stations that can best be predicted by the remaining stations are excluded. The estimation variance is incorporated in the OF model in a reduced form:

\[ \sigma_{E}^{2}(n) = \frac{\sigma_{E}^{2}(n)}{\max(\sigma_{E}^{2})} \tag{5} \]

where \( n \) = iteration number. In order to have a value that varies between fixed boundaries, \([0, 1]\), \( \sigma_{E}^{2} \) is divided at each iteration by the maximum estimation variance found so far. The estimation variance is calculated at each iteration (combination of stations) using \( p(= \pi) \) stations.

### Temporal Redundancy

Time series are considered to be synchronous, either because data are collected at the same time, or because the necessary interpolations are made to synchronize data. Time events can therefore be handled as realizations of random functions, \( Y_i(m), i = 1, ..., L_{\text{FIXED}} + L_{\text{EXP}}; m = 1, ..., D \), (with \( L_{\text{FIXED}} \) being the number of fixed stations, i.e., those that are to be included in all solutions; \( L_{\text{EXP}} \) the dimension of the subset of stations to be included in the new design, and \( D \) the dimension of the time series vectors). For the sake of simplicity, time indices are used only when needed; mathematical operations with time are made over all times. \( Y_i(m) \in X \in \chi \), with \( X \) being the current solution, \( \chi \) the set of possible solutions, and \( i = 1, ..., \pi \). The sum of the differences between time series, \( S \), is used to evaluate temporal redundancy: large values of \( S \) indicate that series are significantly different, while smaller values indicate the opposite. The sum \( S \) is an approximation to the sum of the integrals between time series, or to the sum of the areas defined between any two time functions. In order to calculate temporal redundancy, time series should have an equal mean value, with redundancy depending only on the time series variances and temporal fluctuations. Hence, a new variable is defined, \( Y_i^0(m) = Y_i(m) - \bar{Y}_i \), with \( \bar{Y}_i \) being the mean temporal value of the state variable in station \( i \). The summation of the difference between series is made for all time periods, with the possibility of shifting one time series in relation to the others by the time value \( \Delta \) so that the sum is the lowest, i.e., when summing the difference between series \( Y_i^0(m) \) and series \( Y_k^0(m) \neq i \), only the minimum summed values are used. Minimum temporal redundancy means maximum \( S \):

\[ S(n) = \max \left\{ \sum_{i=1}^{L_{\text{FIXED}}+L_{\text{EXP}}-1} \min_{k=i+1}^{D} \left\{ \sum_{m=1}^{D} \frac{Y_i(m) - Y_k(m+d)}{D-d}, \ 0 \leq d \leq +\Delta \right\} \right\} \tag{6} \]

\( \max(S(n)) \) is incorporated in the objective function model in a reduced form: \( \max(S(n)) \) is kept between fixed bounds, \([0, 1]\), by dividing it by the largest \( \max(S) \) found so far:

\[ S(n) = \frac{S(n)}{\max(S)} \tag{7} \]

\( n \) = iteration number.

This approach is believed to outperform statistical methods because (1) extreme values with low frequency tend to be masked by most statistical methods, (2) seasonality and trends are difficult to handle and preprocessing is usually required to filter these features, (3) statistical methods that handle seasonal phenomena well (i.e., either in correlation or frequency domains) are ill-suited to handling trends, and (4) nonparametric methods well suited to handling complex data (e.g., factorial analysis) require human intervention and are therefore not appropriate for automatic procedures.

### Resulting Model

Consider the following optimization model:

\[ \min \delta \cdot \sigma_{E}^{2}(n) + \eta \left[ -S(n) \right] \tag{8} \]

where \( \delta \) and \( \eta \) = weighting factors that allow the spatial component to be weighted differently from the temporal component in the objective function.

Three different optimization models are compared by (1) setting \( \eta \) equal to zero and therefore including only the spatial component in the objective function, (2) setting \( \delta \) equal to zero and therefore including only the time component, and (3) setting both...
δ and η equal to one and thereby building a space-time model with equal weighting. Different weightings may be used to compensate for different statistical reliability. The problem is single-objective as it depends only on the selected set of stations.

Solving the Optimization Problem

Iterative Improvement

The problem of reducing a monitoring network of dimension Π to a network of smaller dimension π requires testing all possible combinations of π in Π. The number of possible combinations is given by Π/(Π−π)! If the dimension of the network is large, the number of combinations may be high. The proposed solution to the combinatorial problem involves replacing one station at a time, evaluating the result, retaining the station if it reduces the objective function, or if the result fulfills a probabilistic criterion [the Metropolis criterion (Metropolis et al. 1953)], and rejecting the station otherwise. The iterative process of replacing the stations and analyzing the fulfillment of the Metropolis criterion is known as simulated annealing. The algorithm examines at each iteration the cost of a given set of stations (solution), and it may happen that a station that is rejected in one iteration, with a particular combination of stations, is accepted later if the combination is different (different solution). It is guaranteed that most of the solution space is searched. SA is one of the threshold algorithms included in the class of local search algorithms. The other two, as defined by Aarts and Korst (1990), are iterative improvement, where only objective function-reducing neighbors are accepted, and threshold accepting, where some deterministic nonincreasing threshold sequence is used. In the latter method, neighbor solutions with larger objective function values are accepted, but to a limited extent, because the threshold value is fixed and always decreasing, with a very rigid control on the size of the cost difference. Simulated annealing uses a more flexible control on the values of the threshold, allowing transitions out of a local minimum. SA was first introduced by Kirkpatrick et al. 1983 as an algorithm to solve well-known combinatorial optimization problems, and one that reduces the risk of falling into local minima (or metastable solutions) common to iterative improvement methods. This is a consequence of SA accepting only solutions that lower the objective function. These authors proposed the use of the Metropolis (Metropolis et al. 1953) procedure from statistical mechanics, which generalizes iterative improvement by incorporating controlled uphill steps (to worse solutions). The procedure states the following: Consider one small random change in the system at a certain temperature; the change in the objective function is ΔOF: if ΔOF<0, then the change in the system is accepted, and the new configuration is used as the starting point in the next step; if ΔOF>0 then the probability that the change is accepted is determined by P(ΔOF)=exp(−ΔOF/k_bT); a random number uniformly distributed in the interval (0,1) is taken and compared with the former probability; if this number is lower than P(ΔOF) then the change is accepted. The control parameters k_bT are replaced by the parameter t (also called temperature) to avoid using the Boltzmann constant, k_b, which would have no meaning in the present context.

The SA algorithm operates in the following way: (1) The system is melted at a high temperature (initial temperature, t_1); (2) the temperature is decreased gradually until the system freezes (i.e., no further objective function change occurs); and (3) at each iteration the Metropolis procedure is applied. The following describes how to establish the initial temperature, the rate of temperature decrease, the number of iterations at each temperature, and the stopping criteria to halt the process when the system is frozen. The generic SA algorithm (Fig. 1) incorporates a neighborhood function N( ), a heuristic h_n, a random function P( ), the iterations counter c, and the current temperature t; c←0; t←t_0. Choose a feasible solution X ∈ X, X_best ← X; do while c < C_MAX, or t > t_min.

In order to accelerate the process, several improvements have been proposed for limiting the number of iterations at each temperature. The dimension of the Markov chain (in a Markov chain the probability of the outcome of a given trial depends only on the outcome of the previous trial, as occurs in SA) has been proposed to be a function of the dimension of the problem (Kirkpatrick et al. 1983). Temperature is maintained until 100t_i solutions (iterations), or 100t_i successful solutions have been tested, whichever comes first, where Π is the number of variables (stations) in a problem. These authors propose that the annealing process cease if, after three consecutive temperatures, the number of acceptances is not achieved. The annealing can also be stopped if the average value of the objective function does not change after a preestablished number of temperature decreases (R_STOP). Along with these dynamic criteria, a static one may be used to halt the process when a minimum temperature, t_min, is attained. The former guarantees that the annealing will stop if none of the dynamic criteria are fulfilled, even before the total number of iterations is attained. In the algorithm used herein, both the dynamic and the static criteria are implemented.

Cunha and Sousa (1999) proposed the following expression to calculate the initial temperature, t_1:

\[ t_1 = \frac{b \cdot OF_0}{\ln a} \]

where OF_0 = cost of the initial configuration; a = elasticity of acceptance; and b = amount of dispersion around the cost of the initial solution. A prior model run determines the latter. The elasticity of acceptance represents the probability of accepting a solution worse than the initial one. The initial temperature determined by Eq. (9) is such that there is a probability a of accepting

\[
\begin{align*}
U & \leftarrow h_n(X) \\
\text{if } U \neq \text{Fail} & \\
\text{do} & \text{then if } P(U) > P(X) & X \leftarrow U \\
\text{else if } P(X) > P(X_{\text{best}}) & X_{\text{best}} \leftarrow X \\
\text{if } r < e^{[p(r)-p(x)]/t} & \text{then } X \leftarrow U \\
c & \leftarrow c + 1 \\
t & \leftarrow \alpha \cdot t \\
\text{Return } (X_{\text{best}})
\end{align*}
\]

Fig. 1. Simulated annealing algorithm; Random is a random real number taken from a uniform distribution in (0,1)
solutions that are b% worse than the initial solution. This approach is similar to the one proposed by Cunha and Sousa (1999), but here the amount of worsening is established. Temperature is usually decreased with a constant rate (cooling factor) \( \alpha \) such that after \( s \) temperature decreases, the temperature is \( t_s = t_1 \alpha^s \). The two stopping criteria—\( t_{\text{min}} \) and number of temperature decreases—are complementary because it is easy to calculate the minimum temperature attained if \( t_1 \) and \( \alpha \) are known. Criterion \( t_{\text{min}} \) may be useful if one wants to stop the annealing after a certain temperature is reached.

### Analysis of the Convergence

Due to the similarity between simulated annealing and the physical equivalent, a measure of disorder similar to entropy may be used to assess the evolution of the annealing process. Entropy is a classical statistical mechanics definition of a thermodynamic system used to evaluate the state of disorder. Marginal entropy is given by (if the Boltzmann constant is set equal to one and natural logarithms are used)

\[
H(t) = -\sum_{l \in \chi} p_l(t) \ln p_l(t) \tag{10}
\]

where \( p_l(t) \) = probability that the system is in state \( l \) at temperature \( t \); and \( \chi \) = solution space. The summation is carried out over the states allowed by the system. As temperatures decline, the number of states found that improve the best so far should also decrease, which is equivalent to the system reaching a more organized structure, or that entropy is decreasing. Marginal entropy should decrease monotonically to \( \ln |\chi| \) if equilibrium is attained at each temperature (Aarts and Korst 1990). If the best solution is unique, then marginal entropy converges to zero. It can be seen that the value of \( H(t) \) is maximum when all states have equal probability of occurrence, and that \( H_{\text{max}}(t) = \ln(l_{\text{MAX}}) \). \( l_{\text{MAX}} \) is the dimension of the Markov chain at each temperature (e.g., 100\( \Pi \) or 10\( \Pi \)).

Relative entropy is defined as

\[
\frac{H(t)}{H_{\text{max}}(t)} = \frac{H(t)}{\ln(l_{\text{MAX}})} \tag{11}
\]

Relative entropy should decrease monotonically from a sill (\( \approx 1 \)) as temperature decreases, attaining at the limit and the optimum a value close to zero. Hence, at sufficiently low temperatures the probability of finding better states tends to zero. If the algorithm is operating well, and if the tuning parameters are correct, then this should happen. This was used as a criterion to select the best set of annealing parameters. Entropy therefore seems to be an appropriate indicator of how well the search is made, and whether it is converging to viable solutions. Although it does not indicate whether the solution found is a global one, it can help to find good parameters for the simulated annealing algorithm. A good convergence schedule must fulfill the following criteria: (1) The time spent at high temperatures must be long enough to allow running through many of the candidate solutions, i.e., the variables must oscillate for the first high temperatures, with the exception of \( H/H_{\text{max}} \), which must be close to one; (2) as temperature decreases, \( S \) must converge to its maximum, \( \sigma^2 \), and the objective function must converge to its minima, and \( H/H_{\text{max}} \) must converge to \( \ln|\chi| \); and (3) for sufficiently low temperatures, no oscillation is expected. These criteria have been elegantly described in the theory of simulated annealing by Aarts and Korst (1990), and correspond to two well-defined regions in the convergence curve: Region 1 and Region 2. In Region 1 the objective function values are distributed uniformly (modeled with a normal density curve) and the structure of the combinatorial problem plays a minor role; in Region 2, close to the minimum, the structure of the combinatorial problem determines the solution density (as modeled with other density curves). Similar conclusions are also drawn from the entropy curves: at high temperatures the system is disordered and the entropy is maximum, as compatible with a normal density curve; at low temperatures entropy is minimum and should be modeled with densities that minimize entropy (e.g., exponential). Entropy analysis should help in the selection of a good set of SA parameters if the criteria mentioned previously are fulfilled, although it does not indicate the values directly.

### Application Example

To test the three models, a grid of 32 GMN stations randomly distributed in a square grid of eight by eight spatial units was created (Fig. 5). Monitoring stations numbers 1 and 2 represent wells for water supply, and are therefore to be included in all the solutions. Hypothetical time series for all stations are calculated using common mathematical functions, both purely deterministic and with normally distributed errors. The functions were selected based on empirical judgment and experience, and attempt to mimic the behavior of water quality variables. Some of the resulting variables are nonhomocedastic (i.e., variance is dependent on the magnitude of the value), thus reflecting a feature common to variables such as redox potential and electrical conductivity when the scale of measure has to be changed between consecutive observations. With the method proposed here, no statistical assumption on the data is made, which accelerates the preprocessing stage. The time series equations are presented in Table 1.

<table>
<thead>
<tr>
<th>Series number</th>
<th>Equation: ( x = )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>( \cos(\text{time}) + 0.54 \Delta \text{time} )</td>
</tr>
<tr>
<td>2</td>
<td>0.12 \text{time}</td>
</tr>
<tr>
<td>3</td>
<td>\sin(\text{time})</td>
</tr>
<tr>
<td>4</td>
<td>\cos(\Delta \text{time} + \text{time})</td>
</tr>
<tr>
<td>5</td>
<td>\cos(2 \Delta \text{time} + \text{time})</td>
</tr>
<tr>
<td>6</td>
<td>\log(\text{time}) - \log(\Delta \text{time})</td>
</tr>
<tr>
<td>7</td>
<td>\exp[-(\text{time}/(\text{end} \text{time} - \text{start} \text{time}))]</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
</tr>
<tr>
<td>9</td>
<td>\cos(0.3 \text{time} + 3 \Delta \text{time})</td>
</tr>
<tr>
<td>10</td>
<td>\text{atan}(\text{time})</td>
</tr>
<tr>
<td>11</td>
<td>\text{atan}(\text{time} + 1.5 \Delta \text{time})</td>
</tr>
<tr>
<td>12–22</td>
<td>Repeat Eqs. (1)–(11)+20% normally distributed error around mean</td>
</tr>
<tr>
<td>23–32</td>
<td>Repeat Eqs. (1)–(10)+40% normally distributed error around mean</td>
</tr>
</tbody>
</table>
single day. Therefore the cost of sampling is dependent only on transport costs. In this case II is equal to 30 and \( \pi \) is equal to 15, and the problem amounts to finding the minimum objective function value within a solution space of more than \( 1.55 \times 10^5 \) possible combinations.

### Results and Discussion

It should be noticed now that the example state variable of the case study is generic in that it represents the general behavior of natural parameters such as water levels, concentration of a chemical element, and temperature. Therefore the values of \( S \) and \( \sigma^2_F \) are given in unit (un) and square unit (un\(^2\)), respectively; objective function and \( H/H_{\text{max}} \) are dimensionless.

In order to guarantee stationarity of the Markov chains, a sufficient length of time must be spent at each temperature. Moreover, the amount of time spent at high temperatures must also be large enough to allow running through many of the candidate solutions, to avoid being trapped in local minima. The questions to be answered are these: (1) How high should the initial temperature be; (2) how many iterations are to be made at each temperature; and (3) how fast should the temperature fall. If the search is to take a feasible amount of time, some other aspects must also be considered: (1) The number of iterations at each temperature has to be the lowest possible; (2) the number of temperature reductions should be optimized; and (3) the search must stop if a crystallized system is attained.

Relative entropy may be used to complement the analysis of the change in the objective function, \( \sigma^2_F \), and \( S \) as temperature changes. What should be examined is the value of \( t_1 \) for which the objective function is minimized, and such that relative entropy is kept at high values for the first few iterations, decreasing to zero after that. This would indicate that the probability of accepting a new solution decreases as temperature decreases, starting with a value near 1, where most of the solutions are accepted and the probability of accepting worse solutions decreases as crystallization is attained.

The time spent by the search at high temperatures should be large enough to avoid being trapped in local optimal solutions, but small enough to minimize running time. Eq. (9) allows the time spent at the higher temperatures to be tuned, and also enables the amount of dispersion around the objective function value of the initial solution to be set. Both parameters were evaluated in prior batch runs, as described subsequently. The effect of adding temporal information to the optimization of the monitoring network is studied by comparing the resultant network when using only spatial information, only temporal information, and using both.

By construction, the objective function may decrease if a maximum of \( S \) or \( \sigma^2_F \) is found in any iteration. This is due to division by the maximum value whereby the objective function attains relatively small differences in equal solutions (since different maxima are used), but random jumps out of local minima are also allowed. These jumps are less frequent for lower temperatures, because a larger part of the solution space was searched. Even if they occur, the probability of missing the neighboring minimum decreases with \( r \) because most of the allowed transitions become more strictly descending.

An experimental spherical isotropic variogram for the first time period was fitted with the following parameters: nugget \( 8.1 \times 10^{-3} \text{un}^2 \); sill=\( 8.1 \times 10^{-2} \text{un}^2 \); range=2.428 m. A specific computer program was developed (OPTIVAR) and run on Pentium III (Intel 800 MHz) computers. Some prior batch runs using Eq. (9) were undertaken to obtain the proper initial temperature. The results indicated an initial temperature of 1.424, and this was used in all subsequent runs. Convergence was evaluated by analyzing the evolution in the variables \( S \), \( \sigma^2_F \), objective function, processing time, and \( H/H_{\text{max}} \), with temperature.

### Network Optimization with Only Spatial Information

Both \( \sigma^2_F \) and objective function decreased from high values to their minima [Figs. 2(a and b)]. At high temperatures the objective function and \( \sigma^2_F \) show high variability, indicating that worse solutions are still being accepted, but this variability decreases with temperature. No variability was detected at low temperatures, indicating that no further improvements were possible. The system was unable to jump to worse solutions because the probability of drawing a random number that fulfills the Metropolis criterion is low, as discussed previously. Marginal entropy also fulfills the criteria for convergence with a long sill for high temperatures. It decreases for lower temperatures, converging to zero as the best solution is found [Fig. 2(c)]. The minimum objective

![Fig. 2](image-url)
function value is attained at 0.3294, corresponding to a $\sigma^2_E$ of $4.244 \times 10^{-2}$ square units. The average processing time was 993 s.

When only spatial information is used, the objective function is only dependent on $\sigma^2_E$ and the sampling stations must therefore be distributed in space, thereby minimizing the variance of the estimation error. Table 2 shows the two sets that result for 50 runs, along with the frequency of occurrence of the resulting solution sets. The two solution sets have equal final objective function values, and are therefore equally good. Both solutions are shown in Fig. 3. The minimum objective function value is different from the minimum $\sigma^2_E$ because the cost is obtained after the division of $\sigma^2_E$ by its maximum value [Eq. (10)].

Considering only the spatial component, the stations are uniformly distributed in space. Similar results were obtained by Sacks and Schiller (1988) and Pardo-Igúzquiza (1998) for spatial models (without sampling costs in the case of the latter). The two solution sets (A1 and A2) have similar spatial distribution and correspond to geometric configurations with the same estimation variance. Had an anisotropic variogram been used, the spatial distribution would be different. One interesting feature of the data used in this case study is that variogram models are similar in all time periods. In cases where the variogram models vary with time, care is needed when identifying the period(s) when the spatial variability is best reproduced, and this/these is/are not necessarily when it is higher. High spatial variabilities may occur because of particular events, like extraordinary pumping regimes, and thus do not reflect the natural state of the system.

The frequency of occurrence of the two solutions is similar (55.6 and 44.4%, for A1 and A2, respectively). The difference between the two solutions are stations 6, 8, 13, and 30 (in bold in Table 2). The problem is therefore nonunique and some other criterion has to be used in choosing one over the other, such as travel costs and sampling costs.

Table 2. Solutions: Minimizing Only $\sigma^2_E$, Maximizing Only $S$, and Minimizing $\sigma^2_E$ and Maximizing $S$ (OF: Objective Function)

<table>
<thead>
<tr>
<th>Solution name</th>
<th>Solution set</th>
<th>Frequency (%)</th>
<th>$\sigma^2_E$ (un$^2$)</th>
<th>$S$ (un)</th>
<th>OF</th>
</tr>
</thead>
<tbody>
<tr>
<td>Only spatial information</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solution A1</td>
<td>{1,2,3,4,6,8,9,12,14,17,18,19,26,28,29,31,32}</td>
<td>55.6</td>
<td>$4.244 \times 10^{-2}$</td>
<td>—</td>
<td>0.3294</td>
</tr>
<tr>
<td>Solution A2</td>
<td>{1,2,3,4,9,12,13,14,17,18,19,26,28,30,31,32}</td>
<td>44.4</td>
<td>$4.244 \times 10^{-2}$</td>
<td>—</td>
<td>0.3294</td>
</tr>
<tr>
<td>Only temporal information</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solution B</td>
<td>{1,2,3,19,20,21,22,23,24,25,26,27,28,29,30,31,32}</td>
<td>100</td>
<td>—</td>
<td>438.2</td>
<td>$6.96 \times 10^{-8}$</td>
</tr>
<tr>
<td>Both spatial and temporal information</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Solution C1</td>
<td>{1,2,3,12,16,17,19,20,21,22,25,27,28,29,30,31,32}</td>
<td>87.5</td>
<td>$8.07 \times 10^{-2}$</td>
<td>335.7</td>
<td>0.557</td>
</tr>
<tr>
<td>Solution C2</td>
<td>{1,2,3,8,19,20,21,22,23,26,27,28,29,30,31,32}</td>
<td>12.5</td>
<td>$8.09 \times 10^{-2}$</td>
<td>335.8</td>
<td>0.596</td>
</tr>
</tbody>
</table>

Fig. 3. Minimizing only $\sigma^2_E$: optimal solution (Solution A1 in squares and Solution A2 in filled circles)

Fig. 4. Maximizing only $S$: (a) $S$ (un); (b) OF; (c) $H/H_{\text{max}}$
When only temporal information is used, the objective function depends solely on the value of $S$. The objective function should converge to zero because the spatial component of the objective function model was set constant and equal to zero; as $S$ is obtained by the division by the maximum $S$ value found so far, the temporal component of the objective function should converge to zero. The objective function value obtained was $6.96 \times 10^{-2}$ (Table 2). The discrepancy between this value and that expected may be justified by numerical oscillations and truncations during data storage and handling. The solution was attained at $S = 438.24$. It is clear from Figs. 4(a and b) that both $S$ and objective function converge to a maximum and a minimum, respectively. Relative entropy also shows a good convergence curve, fulfilling the convergence criteria stated in Section 3. The resulting solution is shown in Fig. 5. The average processing time was 886.5 s.

If only temporal information is used then the spatial distribution of the stations is dictated by local variability of the time series and by their mutual redundancy. The spatial coverage was more clustered than when using only spatial information, which may be explained by the need to choose stations that are close by but which have significantly different time series, as happens with stations 19 to 32 (with high random fluctuations around the mean). In practical problems this result may be a consequence of having stations near faults, pumping areas, contaminations sources, or any other source of strong local variability.

**Space-Time Network Optimization**

If both spatial and temporal information is considered, then the spatial distribution of stations is expected to be between the regular coverage proposed by the estimation variance and the more clustered distribution proposed by $S$. Again, the convergence curves fulfill the convergence criteria [Figs. 6(a and b)]. The minimum objective function value was attained at 0.557 for $\sigma_E^2 = 8.07 \times 10^{-2}$ squared units and $S = 335.7$. As expected, the solution for the complete model is between those for the single components: $\sigma_E^2$ is now higher and $S$ is now lower. Objective function value is now 40% higher than it would be if the components were independent (summing the objective function values for the first two models: objective function = 0.399), and the resulting solution sets (C1 and C2, see Table 2) are different from Solution Sets A and B, and could not be obtained by any simple combination of the latter. The frequency of the two solution sets is very different (87.5 and 12.5%, for C1 and C2, respectively) and does not correspond to equally good solutions—the costs are different. Rather, it indicates that a worse solution, very near to the best, was found one in every eight runs. Unlike the results of the first model, the best solution in this case is known, and no other criterion is needed to choose between them. The average processing time was 1,012 s.
The resulting monitoring network (Fig. 7) has the advantage of incorporating both spatial and temporal information, thereby maximizing the quality of the data collected. The space component of the model is reflected in a network with higher spatial accuracy, and the temporal component simultaneously selects the stations that are less redundant in time.

Conclusions

Three different models for the determination of the best set of monitoring stations of a monitoring network were evaluated. The models were designed to accommodate problems of network reduction when time series are available. This is usually the case in regional or subregional (e.g., aquifer scale) monitoring networks that have resulted from the aggregation of several local monitoring plans, and that must be transformed in a smaller network that still captures the regional variability of the state variable. A synthetic case study was used, and simulated annealing was selected as the algorithm to solve the combinatorial problem. The results can be summarized as follows:

A method based on the statistical entropy of simulated annealing chains (Markov chains) was proposed to help in choosing the SA parameters, and applied to the case study. The results showed that it is a good tool.

The first model considered only a measure of the quality of the spatial distribution, in the form of estimation variance obtained by kriging. The results indicate that with this model the best spatial distribution, in the form of estimation variance obtained by fixing the SA parameters, and applied to the case study. The results showed that it is a good tool.

The second model used a new variable, the temporal redundancy of the time series, to select the set of stations that had minimum common information, or that collectively assembled most of the variability of the time series possible with the reduced network. The temporal redundancy variable proposed in this article requires no assumption as to the probability distribution because it uses the sampled values directly, rather than a statistical transform. The results indicated that the spatial distribution may be more clustered than the one obtained with an optimization model with only a spatial component, if the time series less correlated in time are spatially clustered. This may be the case in areas of high heterogeneity, such as near important faults, at contacts between different flow regimes, or near pumping areas.

The third model included both the spatial and the temporal components. In the method proposed here only space variography is required, and therefore the preprocessing time is greatly reduced when compared to other approaches cited in “space-time network optimization.” The results of the space-time model show that the spatial distribution of monitoring stations with space-time information is less regular than that obtained with only spatial information, and less clustered than that resulting from time information only.

The space-time model proposed for selecting the subset of stations from a monitoring network best reflects the spatial variability of the state variable, and simultaneously includes the least redundant stations. The monitoring network thus obtained maximizes the relevance of the data collected, and contributes to a better cost-benefit ratio. This model seems a good alternative to other methods proposed in the literature.

Acknowledgments

The writers would like to acknowledge the financial support from the Portuguese Foundation for Science and Technology (Fundação Nacional para a Ciência e a Tecnologia), and also the anonymous referees for their contributions.

Notation

The following symbols were used in this paper:

\[ A = \text{area where mean value is to be estimated}; \]
\[ a = \text{elasticity of acceptance when calculating the initial temperature}; \]
\[ b = \text{amount of dispersion around cost of initial solution when calculating initial temperature}; \]
\[ D = \text{dimension of time series}; \]
\[ F_{\text{BETT}} = \text{frequency of solutions better than the best so far}; \]
\[ F_{\text{REP}} = \text{frequency of accepted transitions}; \]
\[ H(t) = \text{marginal entropy at temperature } t; \]
\[ I_{\text{MAX}} = \text{maximum number of iteration at each temperature}; \]
\[ I_{\text{TOTAL}} = \text{number of total iterations}; \]
\[ I(t) = \text{number of iterations at temperature } t; \]
\[ L_{\text{EXP}} = \text{number of stations to be experimented (others than } L_{\text{FIXED}}); \]
\[ L_{\text{FIXED}} = \text{number of fixed stations}; \]
\[ L_{\text{TOTAL}} = \text{number of total stations}; \]
\[ \text{OF} = \text{objective function}; \]
\[ M = \text{number of points used to discretize } A \text{ when kriging}; \]
\[ p_i(t) = \text{probability that system is in state } l \text{ at temperature } t; \]
\[ R_{\text{STOP}} = \text{number of temperature decreases without change in average value of cost function}; \]
\[ S = \text{summation of difference between time series}; \]
\[ T = \text{time spent until solution was found}; \]
\[ T(t) = \text{time spent at iteration } t; \]
\[ t = \text{temperature}; \]
\[ t_{\min} = \text{minimum temperature that can be attained during cooling}; \]
\[ t_i = \text{initial temperature}; \]
\[ V = \text{mean value of state variable in } A; \]
\[ X = \text{current solution}; \]
\[ x_a \] = value of state variable at location \( \alpha \) (known);
\[ Z(X) \] = state random variable;
\[ \gamma_X(z) \] = value of state random variable;
\[ \Delta \] = temperature decrease factor;
\[ \Gamma = \text{field where random state variable is defined}; \]
\[ \gamma(h) \] = variogram;
\[ \chi \] = solution space;
\[ \kappa_z \] = weights in estimation function;
\[ \mu \] = Lagrange parameter;
\[ \Pi = \text{Cartesian coordinate of set of stations in original design}; \]
\[ \pi = \text{Cartesian coordinate of subset of stations to be included in new design}; \]
\[ \sigma^2_k \] = estimation variance.

References


42 / JOURNAL OF WATER RESOURCES PLANNING AND MANAGEMENT © ASCE / JANUARY/FEBRUARY 2004

Downloaded 04 Nov 2008 to 129.116.232.152. Redistribution subject to ASCE license or copyright; see http://pubs.asce.org/copyright/

