ELSEVIER

Journal of Environmental Management 78 (2006) 294–304

Journal of Environmental Management

www.elsevier.com/locate/jenvman

# Optimal estuarine sediment monitoring network design with simulated annealing

L.M. Nunes[a,*], S. Caeiro[b], M.C. Cunha[c,1], L. Ribeiro[d,2]

[a]Faculty of Marine and Environmental Sciences, University of Algarve, Campus de Gambelas, 8000 Faro, Portugal
[b]Department of Exact and Technological Sciences of the Portuguese Distance Learning University, R. Escola Politecnica, 147, 1200 Lisbon, Portugal
[c]Civil Engineer Department, University of Coimbra, Pinhal de Marrocos, 3030 Coimbra, Portugal
[d]Instituto Superior Técnico, Lisbon Technical University, Av. Rovisco Pais, 1096 Lisboa Codex, Portugal

Received 29 April 2003; revised 7 March 2005; accepted 14 April 2005
Available online 19 August 2005

**Abstract**

An objective function based on geostatistical variance reduction, constrained to the reproduction of the probability distribution functions of selected physical and chemical sediment variables, is applied to the selection of the best set of compliance monitoring stations in the Sado river estuary in Portugal. These stations were to be selected from a large set of sampling stations from a prior field campaign. Simulated annealing was chosen to solve the optimisation function model. Both the combinatorial problem structure and the resulting candidate sediment monitoring networks are discussed, and the optimal dimension and spatial distribution are proposed. An optimal network of sixty stations was obtained from an original 153-station sampling campaign.


## 1. Introduction

A well designed, ongoing monitoring program is fundamental for the evaluation of environmental management of natural systems (Kay and Alder, 2000). The design of an effective monitoring program depends on the management objectives, resources (funding and staff) and available technology. Monitoring programmes should be designed to contribute to a synthesis of information or to evaluate impacts, or analyse the complex cross-linkages between environmental quality aspects, impacts and socio-economic driving forces (RIVM, 1994).

The technical design of monitoring networks is related to the determination of: (i) monitoring sites; (ii) monitoring frequencies; (iii) variables to be sampled; (iv) duration of sampling (the last two variables are not discussed here because they are case-specific). Most of the research results in this area have been obtained in the context of statistical procedures (Sanders et al., 1983; Moss, 1986; IAHS, 1986; Cochran, 1977). These rely in the principle that there are several sources of uncertainty, due to measuring errors, inherent heterogeneities of the involved variables, and in the cases where modelling is involved, also simplifications and errors in both the modelling and numerical/analysis solution phase. McBratney et al., 1981), as well as many other authors after them, indicated that uncertainties are the result of lack, in quality and quantity, of information concerning the systems under study, or as a result of spatial and temporal variations of parameters.

In many monitoring programs a first sampling stage with a large number of locations is undertaken, either because there is no prior information or it is considered necessary to collect more data. This stage is usually planned to give statistical information about the variables under study and to calculate their spatial covariance. A second stage is needed to transform the original set of sampling stations, with high cardinality, into a lower cardinality set of monitoring stations. Probably the methods used most to reduce cardinality are those based on the maximisation of spatial accuracy, or in other words, on the minimisation of the variance of the estimation error, also known as variance

* Corresponding author. Tel.: +351 289800900x7368; fax: +351 289818353.
*E-mail addresses:* lnunes@ualg.pt (L.M. Nunes), scaeiro@univ-ab.pt (S. Caeiro), mccunha@dec.uc.pt (M.C. Cunha), nlrib@alfa.ist.utl.pt (L. Ribeiro).
[1] Tel.: +351 239790200.
[2] Tel.: +351 218417834.

0301-4797/$ - see front matter © 2005 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jenvman.2005.04.024

reduction methods. This is usually carried out in the context of geostatistical theory (Matheron, 1963, 1965) and most frequently by interpolation with an unknown mean, i.e. by ordinary kriging. Other promising methods have been proposed for optimising the monitoring network design, in particular those based on information theory, as in articles such as those by Amorocho and Espildora (1973), Caselton and Husain (1980), Caselton and Zidek (1984), Harmancioglu and Yevjevich (1987), Husain (1989), and Harmancioglu and Alspaslan (1992). Despite the elegance of these methods, they are 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). When soft and other sources of information are available then the Bayesian Maximum Entropy geostatistical method, first developed by George Christakos (Christakos, 1990; Christakos, 1992), have proven to outperform ordinary kriging (D'Or et al., 2001), and also have the advantage over the latter that they do not require the specification of particular probability distributions.

Kriging variance has been extensively used for monitoring network design. Examples can be found in the work of Bras and Rodríguez-Iturbe (1976), Rouhani (1985), Loaiciga (1989), Rouhani and Hall (1988), Pardo-Igúzquiza (1998), van Groenigen et al. (1999), van Groenigen and Stein (1998), and Nunes et al. (2004a, b).

Two categories for monitoring optimisation with variance reduction have been proposed: (i) the local approach (e.g. Amorocho and Espildora, 1973); and (ii) the global approach (e.g. Ahmed et al., 1988). In the first the influence of each additional point is analysed separately. Total variance reduction after adding one point is easily computed by considering the individual values at each initial location or at the points in the vicinity of the point being estimated. In the global approach average estimation variances are used. Therefore, global approaches provide only average answers to monitoring designs. It is useful to analyse designs still on the drawing board or to perform extensive redesigns aimed at maintaining the efficiency of a monitoring network, which may require removal of poorly located sites. The local approach, on the other hand, is better suited to optimally expanding an existing network. The optimality in this case only relates to the additional points, which may not be acceptable if the original points are not optimal (Markus et al., 1999).

Minimisation of the average kriging variance approach was applied here to select the number and positions of sediment monitoring stations in the Sado river estuary located in the southwest coast oft Portugal (Fig. 1), such that different physically and chemically homogeneous areas identified in a prior sampling campaign were considered. This monitoring network will be further integrated into an environmental data management system for the Sado Estuary as a decision support tool for local authorities.

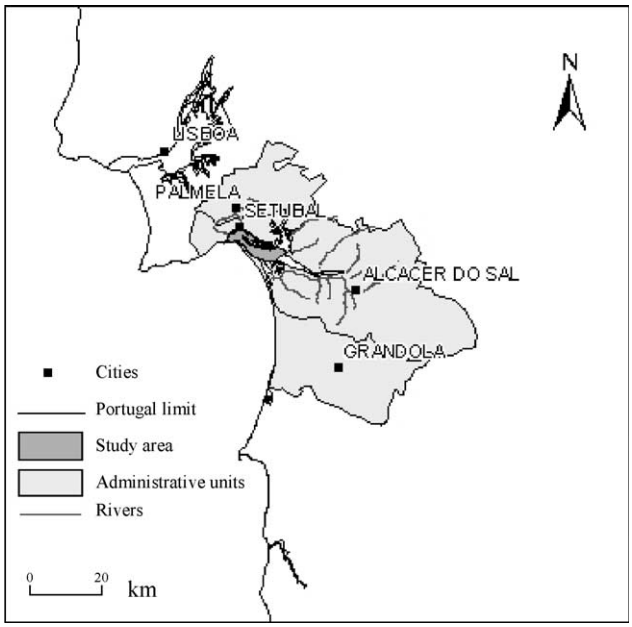


Fig. 1. Study area location (in dark grey).

The Sado Estuary in Portugal is an example where environmental problems are not well managed owing to the high natural values and diverse pressures for development and where the right tools to help evaluating the environmental quality status need to be developed. The objective here was on the development of a monitoring network that constitutes one the information sources of the Sado Estuary management system (physic-chemical data of sediment quality).

For practical and budgetary reasons the number of monitoring stations should be reduced to a minimum. The optimisation problem can be stated in a very simple way: maximising the spatial accuracy, constrained to a maximum number of stations, given the information collected in a prior sampling program (153 sampling sites). Maximisation of spatial accuracy is easily attained by minimising the variance of estimation error, though incorporating the patchiness of homogeneous areas is a more difficult problem. One alternative would be to fix several locations inside the different homogeneous areas, but then the choice of stations would be arbitrary. Another way is to use stratification, considering that a defined number of stations must be placed inside homogeneous areas. Stratification is a well-known statistical technique used for designing monitoring (or sampling) programs with denser networks in some areas than in others. The difference in probability density may be based, for example, on spatial autocovariances, statistical risk of contamination, plume detection probabilities or empirical judgement, among many others. Here we propose a statistically based stratification: homogeneous areas are monitored according to the frequency with which they appear in the prior sampling program. The inclusion of homogeneous areas was considered important by the manager because sediment granulometry and physical and

chemical characteristics have strong correlations with the amount of xenobiotics the sediment can retain and because these areas were planned to be geographic spatial units in an environmental management system. Hence, four types of sediments were established on the basis of three physical and chemical variables and the manager demanded that the proportion of stations in the four types of sediments in the monitoring network be similar to that of the sampling campaign (thus the constraint on the proportions).

Optimisation consists, then, of finding an optimal subset with a combination of stations taken from a larger set. Even for relatively small set cardinalities the number of combinations is too high to allow them all to be exhaustively evaluated in a reasonable amount of time. One of the most well known algorithms for solving combinatorial problems is simulated annealing, in particular in sampling/monitoring network optimisation (e.g. Meyer et al., 1994; Pardo-Igúzquiza, 1998; van Groenigen et al., 1999; Brus et al., 2000; Brus et al., 2002; Nunes et al., 2004a,b).

The article is divided in five sections. This Introduction is followed by a second section where the theoretical geostatistical and optimisation framework is presented. In this section the geostatistical parameter most frequently used to measure accuracy, the kriging estimation error variance, is explained and compared with another geostatistical measure of accuracy, the fictitious point estimation error variance. Also the simulated annealing heuristic used to solve the optimisation problem is introduced. In the third section a case-study is presented and data transformations are explained, while, in the fourth section, optimisation results are discussed. Finally, in the last section, the most important conclusions are drawn.

## 2. Theory

### 2.1. Estimation of probability distribution functions

Indicator coding implies transforming a continuous or discrete variable, $Z(x)$, into a discrete (0,1) one, the indicator $I(x)$. Considering a threshold value $z_c$ on $Z$, $I(x)$ is equal to 1 if $Z(x) \leq z_c$, and 0 otherwise. Therefore the variable at each location is transformed into a distribution function, i.e. the probability of exceeding the threshold is calculated within a region. With a sufficiently large number of thresholds the prior (and post) probability distribution of $Z$ is calculated at each location. Indicator transform is also at the core of nonparametric methods, which have some clear advantages over parametric methods: (i) the parametric hypothesis may not hold; (ii) there are no statistical tests to adequately investigate the validity of a multivariate distribution hypothesis (Alli et al., 1990); (iii) parametric methods are difficult for many practitioners to comprehend and apply due to their mathematical complexity (Sullivan, 1984). Several geostatistical methods can be used to estimate probability distributions, namely Multigaussian kriging (MK), disjunctive kriging (DK), lognormal kriging (LK), probability kriging (PK) and indicator kriging (IK). Both MK and DK are based on normality assumptions and LK on log-normality assumption. If these assumptions are not verified, e.g. in highly skewed distributions, variogram fitting tends to be very problematic and the estimation poor quality. IK is therefore a good alternative. It is actually one of the most frequently used nonparametric methods and will also be used here.

The theory and implementation of nonparametric estimators of spatial distributions is similar to that of nonparametric estimators of the local mean (Journel, 1987). Consider the indicator

$$i(x, z_c) = \begin{cases} 1, if z(x) \leq z_c \\ 0, otherwise \end{cases} \quad (1)$$

with $z_c$ representing some threshold values on $Z$.

The purpose of this indicator transformation is to estimate the posterior cumulative probability functions. These functions are linear combinations of the indicator function and represent the proportion of values less than the threshold,

$$\phi^*(z_c) = \sum_{\alpha=1}^{\omega} \lambda_\alpha i(x, z_c) \quad (2)$$

where $\lambda_\alpha$ represents weights, with the constraint $\sum_{\alpha=1}^{\omega} \lambda_\alpha = 1$ for unbiasedness and $\alpha$ the number of stations. Equation (2) can be solved by simple kriging using $i(x,z_c)$ and the indicator variogram:

$$\gamma_i(h, z_c) = \frac{1}{2N_h} \sum_{i=1}^{N_h} \left[ i(x+h, z_c) - i(x, z_c) \right]^2 \quad (3)$$

where $N_h$ is the number of pairs for lag $h$.

If one is interested in the proportions associated with each interval, $[z_{c-1}, z_c]$, then:

$$\Pr\left[ z_{c-1} < z(x_\alpha) \leq z_c \right] = \phi(z_c) - \phi(z_{c-1}) = \Psi(z_c) \quad (4)$$

and for the estimated proportions

$$\phi^*(z_c) - \phi^*(z_{c-1}) = \Psi^*(z_c)$$

### 2.2. Estimation error and estimation error variance

In this article the estimation error is sought at $x_\alpha$ locations because the monitoring stations to be included in the new design are taken from the initial locations. At the heart of geostatistical cross-validation lies the estimation error obtained by removing one of the $x_\alpha$ stations, estimating it by kriging with the remaining stations (those in the vicinity) and repeating this for all stations (Deutsch and Journel, 1992).Considering that the kriging mean estimation error is zero by construction, and there are $\omega$ stations, the estimation error variance is:

$$s^2 = \frac{1}{\omega}\sum_{\alpha=1}^{\omega}\left[i(x_\alpha, z_c) - i^*(x_\alpha, z_c)\right]^2 \quad (6)$$

where $i^*(x_\alpha,z_c)$ is the estimated IK value. Each station's value is obtained by removing it from the set and estimating it by IK using the remaining stations.

The estimation error variance might also be calculated by the combination of the individual estimation error variances, weighted by the relative frequency of the indicators. However this approach may prove prone to errors if the number of cutoffs is low, and still requires testing.

### 2.3. Model for the optimisation function

The testing of new candidate network designs is carried out by first choosing the number $\omega$ of stations to be included in the subset $S$'. Using the optimisation procedure the optimal combination of stations will be selected from the original set of stations $S$ (with cardinality $\Omega$).

The station locations that produce the lowest estimation error variance result in a spatial distribution with the highest accuracy (the ultimate objective). Therefore to optimise the spatial distribution of stations the estimation error variance has to be minimised. The resulting objective function model is Minimise

$$s^2 = \frac{1}{\omega}\sum_{\alpha=1}^{\omega}\left[i(x_\alpha, z_c) - i^*(x_\alpha, z_c)\right]^2, \omega \in S', S' \subset S \quad (7)$$

Subject to

$$\Psi_{S'}(z_c) \approx \Psi_S(z_c)$$

The constraint makes it necessary for a candidate solution set, $S$', to have the same proportion of stations with values in the intervals$]z_{c-1}, z_c]$ as the original set, $S$. The condition is not equality because for practical computation floating-point variable equality is machine-dependent and varies with the precision. Instead, $\Psi_{S'}(z_c)$ may be bounded and the constraint becomes:

$$\Psi_S(z_c)(1-\delta) \le \Psi_{S'}(z_c) \le \Psi_S(z_c)(1+\delta) \quad (8)$$

with $\delta$ the semi-amplitude of a]0,1] interval.

This condition is necessary to correct the bias introduced by variogram models fitting errors in fitting the variogram models (when adjusting the theoretical models to the experimental variogram). A practical example showing the bias is presented in this article.

The constraint was implemented in the algorithm by choosing, for the calculation of the objective function, only the solutions that fulfil the criterion. This is achieved by allowing random replacements of one element of the set S with one of S' and choosing only the sets S' for which the criterion is fulfilled. After kriging the order relations are corrected by post-processing with the GSLIB POSTIK routine (Deutsch and Journel, 1992). Gruijter et al. (1997) proposed an alternative method (Compositional Kriging) for guaranteeing the correct order relations and the constant sum of the proportions (i.e. 1). Though very promising, this method was not used here.

### 2.4. Solving the problem of the optimisation function model

The example studied here may be classified as a difficult combinatorial optimisation problem, for which an exhaustive search of all possible combinations is not possible in a reasonable amount of time. Solutions to these problems may however be sought in heuristic algorithms that iteratively look for better solutions by trial and error. One of such algorithms is the well-known simulated annealing (SA). It 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 OF-reducing neighbours are accepted, and threshold accepting, where some deterministic non-increasing threshold sequence is used, allowing neighbour solutions with larger OF to be accepted, but in a limited way because the threshold value is fixed and always decreasing, with a very rigid control on the size of the OF difference, $\Delta$OF. Simulated annealing uses a more flexible control of the values of the threshold, allowing transitions from a local minimum at nonzero temperatures.

SA was first introduced by Kirkpatrick et al. (1983) as an algorithm for solving well known combinatorial optimisation problems, reducing the risk of the search falling into local minima (or metastable solutions), that is common to iterative improvement methods. These authors proposed the use of the Metropolis procedure (Metropolis et al., 1953) from statistical mechanics. This procedure generalizes iterative improvement by incorporating controlled uphill steps (to worse solutions). The procedure states the following: consider that the change in the objective function is $\Delta$OF; if $\Delta\text{OF} \le 0$, then the change in the system is accepted and the new configuration is used as the starting point in the next step; if $\Delta\text{OF} > 0$ then the probability that the change is accepted is determined by $P(\Delta\text{OF}) = \exp(-\Delta\text{OF}/t)$ where $t$ is a control parameter called temperature; a random number uniformly distributed in the interval (0,1) is taken and compared with the former probability; if this number is lower than $P(\Delta\text{OF})$ then the change is accepted. The SA algorithm runs in the following way: (i) the system is *melted* at a high temperature (initial temperature, $t_1$); ii) the temperature is decreased gradually until the system *freezes* (because no better solutions are found and the probability of uphill steps is near zero); iii) at each iteration the Metropolis procedure is applied; iv) if any of the stopping criteria is satisfied the algorithm is stopped and the best solution found is presented.

The generic SA algorithm for a minimisation, considering a neighbourhood structure $N$, a solution space $\chi$, a constant temperature decrease rate $\alpha$ and an objective function OF, has the following pseudo-code.

```
Select an initial solution X_best;
Select an initial temperature t_1>0;
Select a temperature reduction factor, 0<α<1;
Set stopping condition
Repeat
Repeat
        Randomly select X ∈ N(X_best);
        ΔOF = OF(X) – OF(X_best);
        if ΔOF<0 then
                X_best = X
        else
                generate random z uniformly in (0,1);
        if z<exp(-ΔOF/t) then X_best = X;
Until iterations = max_iterations
Set t = α t;
Until stopping condition = true;
X_best is the optimal solution found.
```

In order to speed-up the process several improvements have been proposed, specifically by limiting the number of iterations at each temperature, i.e. defining the number *max_iterations*. It has been proposed that the dimension of the Markov chain should be a function of the dimension of the problem (Kirkpatrick et al., 1983): temperature is maintained until $100\,\Omega$ solutions (iterations), or $10\,\Omega$ successful solutions have been tested, whichever comes first. $\Omega$ stands for the number of variables (stations) in a problem. These authors also proposed that the annealing should be stopped (stopping criterion) if after three consecutive temperatures the number of acceptances is not achieved. It can also be considered that if the average value of the OF does not change after a pre-established number of temperature decreases ($R_{STOP}$), then the annealing should be stopped. These parameters control the time spent at each temperature and the total running time. Along with these dynamic criteria, a static one may be used to halt the process when a minimum temperature, $t_{min}$, is reached. The former will guarantee that the annealing will stop if none of the dynamic criteria is fulfilled, even before the total number of iterations is attained. In our algorithm both the dynamic and static criteria were implemented.

The initial temperature, $t_1$, is calculated by running a fast (rapid temperature decrease) schedule and picking up the temperature for which more than 95% of the iterations are accepted. Temperature is usually decreased at a constant rate, $\alpha$, usually close to one (e.g. 0.90 or higher). Aarts and Korst (1990) showed that SA can find optimal solutions if equilibrium is attained at each temperature (constant OF mean and variance) and proposed a temperature schedule dependent on OF variance that guarantees that. Despite this very attractive characteristic such a schedule tends to converge too slowly. Other $t$ schedules for optimality were also proposed by Geman and Geman (1984), Hajek (1988), and Siarry (1997). These however may not converge in an acceptable amount of time for many problems (Cohn and Fielding, 1999). The wealth of practical experience with the faster $t$ schedule used here indicates that the solutions found should be *good* local optimal ones. In practical terms: the local optimal solutions are a compromise between relatively good solutions in an amount of time significantly smaller than that necessary to guarantee the best quality solutions provided (in theory) by slower schedules.

A specific computer code in FORTRAN that incorporates both the estimation error variance and the SA algorithm was developed by the authors to optimise localisation problems and adapted to this specific problem.

## 3. Case study

### 3.1. Study area and source data

The Sado Estuary is the second largest estuary in Portugal with an area of approximately 24,000 hectares. It is located on the west coast of Portugal, 45 km south of Lisbon (Fig. 1). Most of the estuary is classified as a nature reserve. The Sado Estuary basin is subject to intensive land-use practices and plays an important role in the local and national economy. Most of the activities in the estuary (e.g. industry, shipping, intensive farming, tourism and urban development) have negative effects on the physical and chemical quality and biotic communities of water and sediment (Caeiro et al., 2003b).

During the year 2000/2001 sediment was sampled in 153 locations (stations) in an extensive estuarine sediment campaign (Caeiro et al., 2003b). Each sample was analysed for the fine fraction (FF), organic matter (OM) and redox potential (Eh). The data contributed to defining areas of similar physical and chemical characteristics (Caeiro et al., 2003a) and to designing a future sediment monitoring network.

### 3.2. Data processing

The definition of spatially homogenous physical and chemical areas was assumed as a necessary first methodological step. These areas resulted from data collected in a sampling campaign, after some statistical transformations (Fig. 2) (Caeiro et al., 2003a): (i) principal component analysis (PCA) using data on ln(FF), ln(OM), and Eh; (ii) variogram analysis on the first PCA component; (iii) computation of intervariable Euclidean distance (as opposed to geographic distance); (iv) computation of a dissimilarity matrix (Oliver and Webster, 1989), Eq. (9); (v) Euclidean distance computation on the dissimilarity matrix (cluster analysis); (vi) selection of four clusters and physical and chemical interpretation. A new variable, $Z$, was obtained by interpreting the cluster analysis of $d_{ij}$* projections. $Z$ is a discrete variable representing stations that share common physical and chemical characteristics and

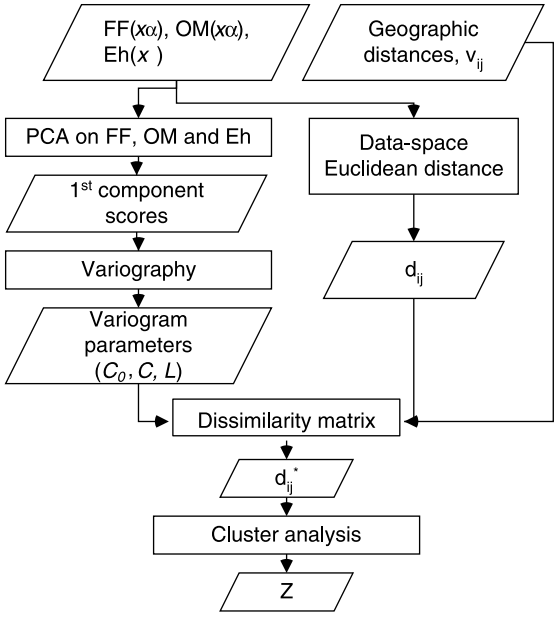


Fig. 2. Flowchart for the definition of the homogeneous sediment areas (adapted from Caeiro et al., 2003a).

similar spatial autocovariance.

$$d_{ij}^{*} = d_{ij} \times \frac{c}{c_0 + c} \times \left[ 1.5 \times \frac{|v_{ij}|}{L} - 0.5 \times \left( \frac{|v_{ij}|}{L} \right)^3 \right] + d_{ij} \times \frac{c_0}{c_0 + c} \qquad \text{for } 0 < v_{ij} \le L \; d_{ij}^{*}$$

$$= d_{ij} \quad \text{when} \quad v_{ij} > L \tag{9}$$

where: $d_{ij}$ is the Euclidean data-space dissimilarity of ln(OM), ln(FF) and Eh between sample sites; $c$ is the variogram sill; $c_0$ is the nugget variance; $L$ is the range; and $v_{ij}$ are the Euclidean geographic distances between locations $i$ and $j$.

The characteristics of $Z_i$, where $i = 1,\ldots,4$, are shown in Table 1 and reflect four separate physically and chemically homogeneous areas found in the sampling campaign. It was subsequently intended to estimate the probability distribution function of $Z$ considering four cut-offs: 1, 2, 3, and 4. The indicator transform is given by (1) with cut-off $z_c$ = 1,…4. For ease of identification, indicators have a subscript equal to the cut-off value.

The value of $\phi(z_c)$ and the proportion of stations with values in the interval$]z_{c-1}, z_c]$, $\psi(z_c)$, are shown in. As $Z$ can only take integer values, the proportions correspond to the stations for which $z(x_\alpha) = z_c$ (Table 2).

Table 1
Physical and chemical sediment parameters of each homogeneous area

| Parameter | Cut-offs | | | |
|---|---|---|---|---|
| | $z_1$ | $z_2$ | $z_3$ | $z_4$ |
| % OM | 8.6±2.4 | 4.2±1.4 | 1.9±0.7 | 0.9±0.3 |
| % FF | 60.4±27 | 21.7±11.8 | 9.1±7.8 | 1.5±1.3 |
| Eh | −278.9±68.6 | −178.8±72.6 | −137.4±50.9 | 74.4±49 |

Table 2
Indicator frequencies, $\phi(z_c)$ and $\Psi(z_c)$

| | Indicators | | | |
|---|---|---|---|---|
| | $i_1$ | $i_2$ | $i_3$ | $i_4$ |
| $\phi$ | 0.1176 | 0.4967 | 0.7320 | 1 |
| $\psi$ | 0.1176 | 0.3791 | 0.2353 | 0.2680 |

The maximum number of stations to be included in the sediment monitoring network had necessarily to be less than 100 for budgetary reasons, but if possible a much lower value was to be looked for ($\omega$). Moreover the new design had to reflect the physical and chemical variability of the sediment as detected in the prior sampling campaign and presented in the previous section. Accordingly, the proportion of monitoring stations in each of the identified homogeneous areas needed to be similar to that in the sampling campaign. This amounted to making the candidate solutions have $\Psi_{S'}(z_c) \approx \Psi_S(z_c)$.

Three different conditioning options for the objective function are presented: (i) no conditioning on the proportions is imposed; (ii) conditioning is imposed with $\delta = 0.5$; and (iii) conditioning is imposed with $\delta = 0.3$. In the first option the entire solution space is a feasible space, while in the others a solution is only feasible if it respects the condition. More stringent conditions were also tested ($\delta < 0.3$) but resulted in too long processing times and no solutions were obtained. It will be seen below that such conditioning may not be necessary. Hence, $\delta < 0.3$ corresponds to the lowest conditioning that produced results, $\delta < 0.5$ having been chosen because it corresponds to an interval with a range of 100% in respect of the proportions.

For each OF conditioning option several network dimensions were tested, according to the following scheme: (i) imposition of the maximum number of monitoring stations ($\omega$) to be included in the new design; (ii) detection of the optimal allocation solution with SA; and (iii) an increase in $\omega$ and a return to (i).

Eight different monitoring network dimensions (cardinality of $S'$: $\omega$) were tested, {30,40,50,60,70,80,90,100}. SA solutions were considered optimal when more than 70% of 20 consecutive runs with the same objective function conditions ($\omega$, $\delta$) and SA parameters had the lowest and equal $s^2$ value. Runs were made on Intel 2000 MHz PC's.

## 4. Optimisation results and discussion

### 4.1. Feasible space

The number of combinations of $\Omega$ sampling stations with $\omega$ possible monitoring stations is given by the well-known formula $W = \Omega! \,/\, ((\Omega - \omega)!\ \omega!)$. Now, if one wants to calculate the combinations conditioned to the reproduction

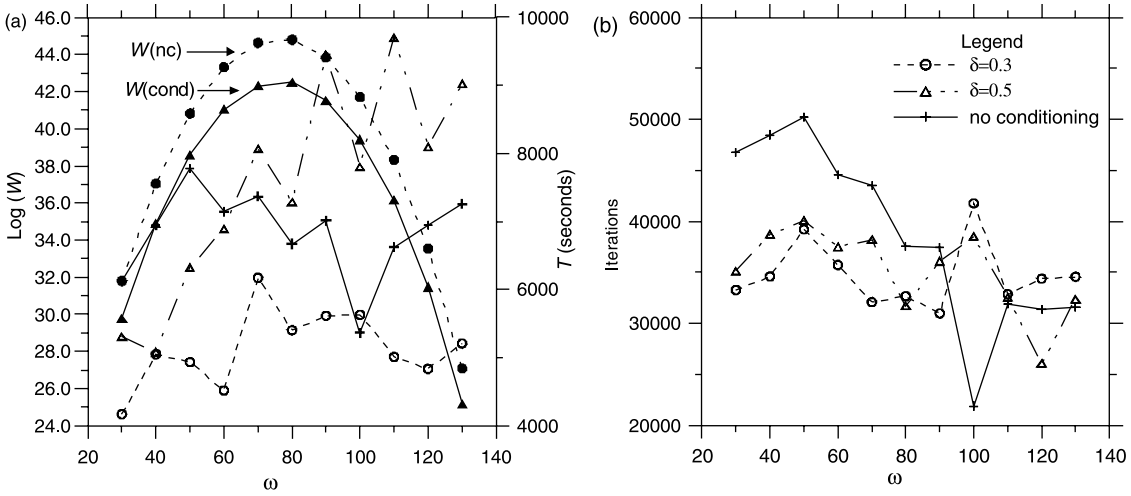


Fig. 3. Effect of conditioning and cardinality of S': (a) Dimension of the solution space (W) and time until optimal solution (T); (b) Number of iterations. In *W*(), nc and *cond* represent without and with conditioning, respectively.

of the proportions, the expression becomes

$$W = \prod_{i=1}^{k} \frac{\Omega_i!}{(\Omega_i - \omega_i)!\omega_i!} \tag{10}$$

where $i$ is the indicator number, $\Omega_i$ the number of sampling stations with the indicator $i$, and $\omega_i$ the number of monitoring stations with the indicator $i$ imposed by conditioning. The number of combinations in each case is represented in Fig. 3a, on a logarithmic scale, for different $\omega$ values. Conditioning reduces the dimension of the feasible space by more than two orders of magnitude when $\omega \approx 77$ (from $7.3\times10^{44}$ to $3.6\times10^{42}$). The lowest $W$ value is still higher than $10^{25}$, indicating that any attempt to solve even the lowest dimension combinatorial problem exhaustively would take (in the same machine) more than $10^6$ times the age of the universe! Despite this practical difficulty it is interesting to see how the optimisation problem structure (intrinsic to each specific problem) affects the time necessary to find an optimal solution (total running time), $T$, and the number of iterations. The time is dependent on the number of iterations and on the time necessary to compute the objective function. The number of iterations is however strongly dependent on the structure of the problem and not necessarily on the dimension of the feasible space, otherwise this parameter would have followed $W$ more closely. In reality it is observed that the problem structure changes for each $\omega$ value (Fig. 3b). However, when no conditioning is imposed the number of iterations decreases to $\omega=100$ (Fig. 3b and Table 3), stabilising after that. This may indicate that a higher proportion of indicators one and two, which have always had higher estimation errors (yet unpublished results), contributes to a more structured problem, possibly with fewer local minima. Such behaviour is not clear when conditioning is imposed due to a predetermined proportion of these indicators in all tested $\omega$ values. As a consequence, $T$ varies around a constant mean in the case of no conditioning (as $\omega$ increases the number of iterations decreases and the OF computing time increases) (Fig. 3b). When conditioning is imposed $T$ increases to $\omega=90$, with a tendency to stabilise after that (and is, therefore, dependent essentially on the OF computing time).

It is also interesting to see that strong conditioning ($\delta=0.3$) results in the lowest $T$, while weaker conditioning ($\delta=0.5$) results in the highest $T$. Once again the problem structure plays a fundamental part. This may not, however, be extrapolated to other problems because it is problem-dependent.

Table 3
Results for different $\omega$ and $\delta$ values: estimation error variance ($s^2$), number of iterations (Iter.) and time necessary to reach an optimal solution ($T$) in seconds

| $\omega$ | No conditioning | | | $\delta=0.5$ | | | $\delta=0.3$ | | |
|---|---|---|---|---|---|---|---|---|---|
| | $S^2$ | Iter. | $T$(s) | $s^2$ | Iter. | $T$(s) | $s^2$ | Iter. | $T$(s) |
| 30 | 0.12889 | 46760 | 6114 | 0.62667 | 35135 | 5306 | 1.00000 | 33260 | 4165 |
| 40 | 0.12000 | 48484 | 6946 | 0.62500 | 38729 | 5076 | 0.81938 | 34596 | 5046 |
| 50 | 0.11840 | 50209 | 7778 | 0.55000 | 40088 | 6325 | 0.68000 | 39215 | 4930 |
| 60 | 0.13639 | 44582 | 7142 | 0.54000 | 37523 | 6894 | 0.58750 | 35718 | 4515 |
| 70 | 0.18694 | 43560 | 7366 | 0.41429 | 38201 | 8071 | 0.58551 | 32078 | 6171 |
| 80 | 0.24938 | 37568 | 6669 | 0.41817 | 31705 | 7284 | 0.57000 | 32668 | 5400 |
| 90 | 0.33099 | 37441 | 7012 | 0.41421 | 36087 | 9456 | 0.56000 | 30972 | 5612 |
| 100 | 0.50910 | 21887 | 5361 | 0.41613 | 38523 | 7803 | 0.55700 | 41776 | 5626 |

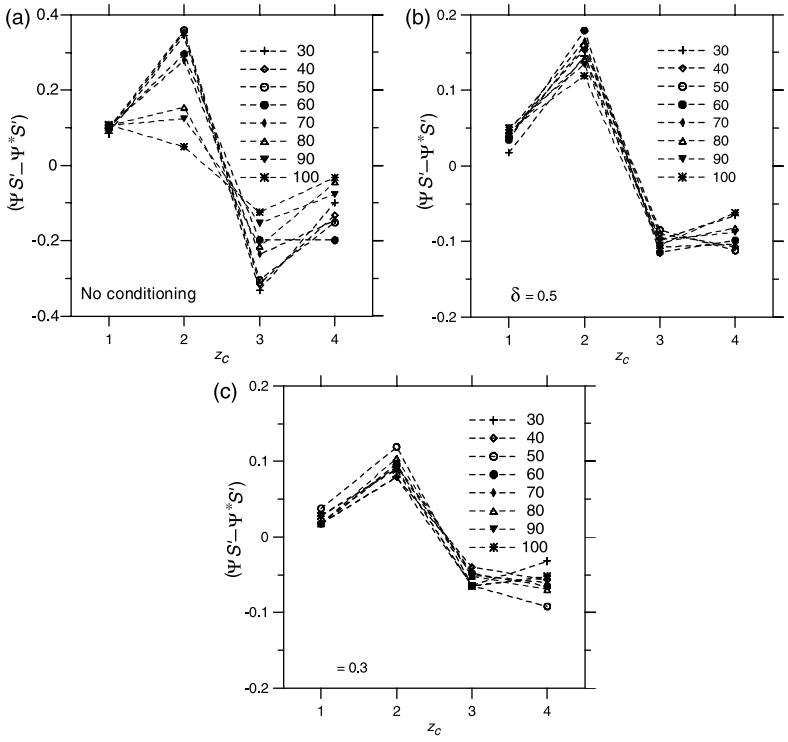


Fig. 4. Relative error when estimating $\Psi_{S'}(x_\omega, z_c)$: (a) no conditioning; (b) $\delta = 0.5$; (c) $\delta = 0.3$.

### 4.2. Optimal conditioning

Conditioning of the objective function has an interesting effect on the errors in the estimated indicator proportions. When no condition is imposed, the error in the estimated proportions $((\Psi_S(z_c) - \Psi_{S'}{}^*(z_c))$ is higher than 0.3 for low $\omega$ values and decreases with an increase in $\omega$ (Fig. 4a). Conditioning the proportions causes a reduction in the error of about 50% for $\delta = 0.5$ and about 25% for $\delta = 0.3$ (Fig. 4b and c) when compared with no conditioning. Moreover, $(\Psi_S(z_c) - \Psi_{S'}{}^*(z_c))$ is one third of the result of the imposed interval $(\Psi_S(z_c) - \Psi_{S'}(z_c))$, for both $\delta = 0.5$ and 0.3. This may indicate that imposing lower $\delta$ values would lead to similar results. If this is true then imposing $\delta < 0.3$ would also lead to very narrow intervals and eventually to a very limited number of neighbouring feasible solutions. Such behaviour is in line with the aforementioned practical difficulty—extremely long processing times. Furthermore, with $\delta = 0.3$, the $\Psi_{S'}{}^*(z_c)$ estimation error is lower than 10%, which is considered an acceptable error.

When the high or low values of a variable are clustered in small areas scattered about the study area, their relative frequencies are low or the data is too random, then variogram fitting becomes difficult and prone to error. The result is not only the fitting of theoretical variograms that only roughly approximate the real variability but also large estimation errors. This does not hinder the geostatistical method, but justifies the need to consider the conditioning of the proportions. An example of such a need is seen with indicators one and two, for which the estimation errors are higher: this leads the optimisation algorithm to select, preferentially, the two remaining indicators with lower estimation errors. As a result, in all the cases studied, the latter have higher proportions than in the original data set, as a way of compensating for the bias introduced with the first two indicators. However, conditioning significantly reduces the bias. Another even more important effect of estimation errors is reflected in Fig. 5: if no conditioning is used,

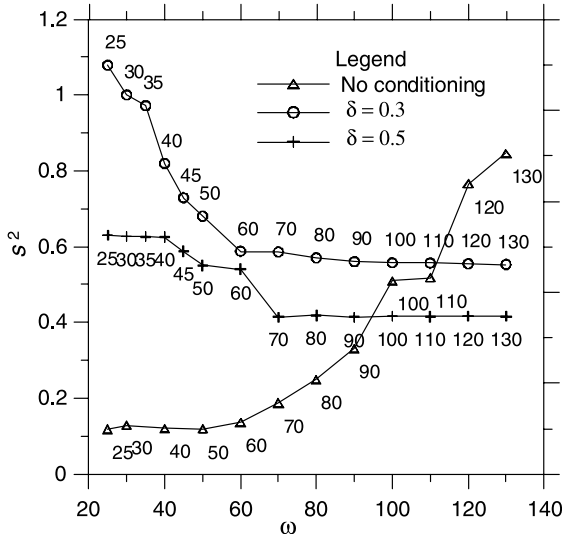


Fig. 5. Estimation error variance without conditioning and with conditioning ($\delta = 0.3$; $\delta = 0.5$).

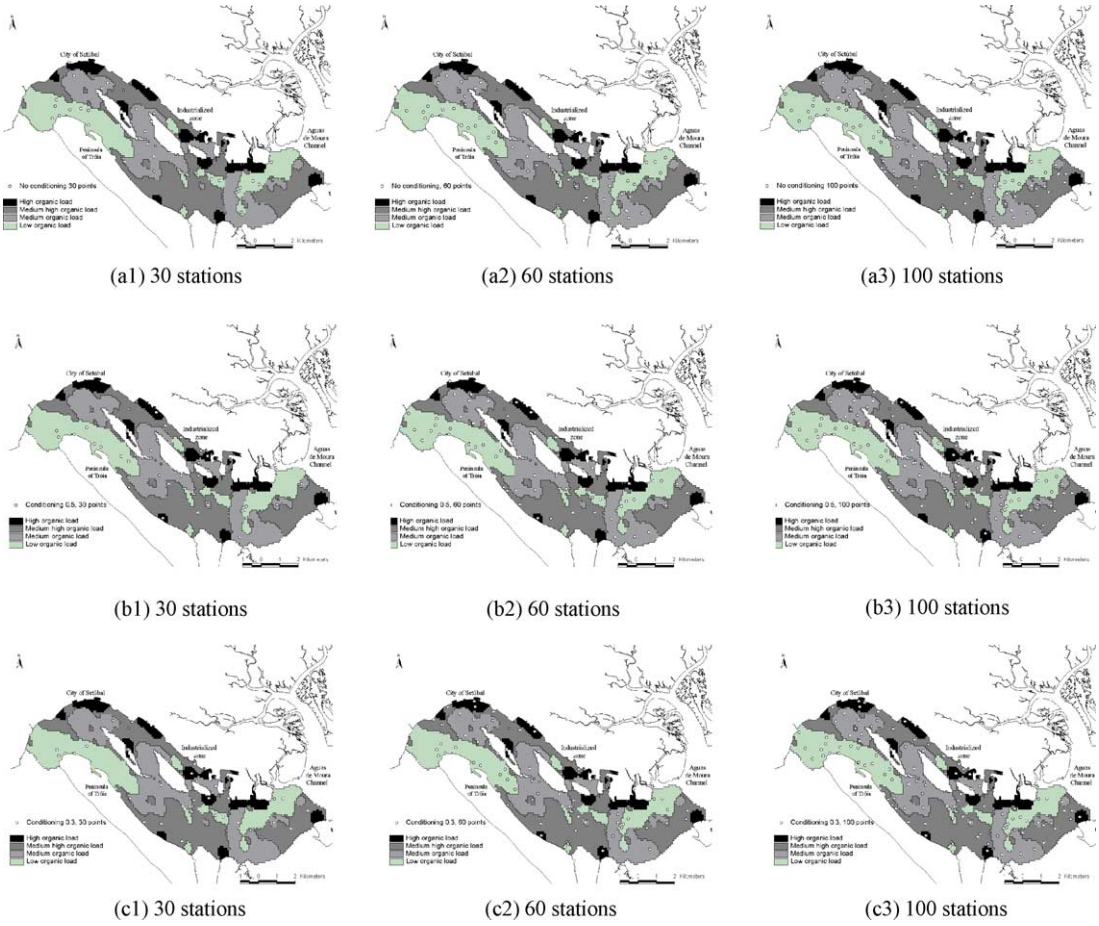


Fig. 6. Monitoring networks for different $\omega$ values: (a) no conditioning (nc); (b) $\delta=0.5$; (c) $\delta=0.3$.

increasing the number of stations will result in higher estimation error variances. This is opposite to what is expected because when increasing the number of points available for estimation the accuracy of the estimated value should decrease. This inversion may be explained by the consideration that, with a very low $\omega$, only stations with a low estimation error are included in the optimal solution; as $\omega$ increases, higher estimation error stations are included. Clearly, if no conditioning is imposed the monitoring network is dominated by the last two indicators (Fig. 6a).

When conditioning is used the expected increase in accuracy is observed (Fig. 5). Moreover, the stronger the conditioning the higher the $s^2$ because more high-error stations are imposed at lower $\omega$ values.

Fig. 6b and c show the resulting monitoring networks with different $\omega$ values. The proportions of the first two indicators are higher in these cases and, with $\delta=0.3$, a better reproduction of the probabilities is obtained, which is considered as an important decision-making criterion for monitoring the homogeneous areas.

### 4.3. Optimal monitoring network

A monitoring network dimension is considered optimal if each new station added to a $\omega$ value has little effect on the spatial accuracy of the monitoring, $s^2$, i.e. if the marginal gains are small. The gains are shown in Fig. 5. Gains in accuracy are high up to the 60th station, becoming much less important after that. Adding one new station produces an average increase in spatial accuracy of 1.24% up to the 60th station; after that the gains in accuracy reduce to an average of only 0.034%. Sixty is therefore considered as the optimal $\omega$ value. Fig. 7 shows the convergence results for this network, with the following simulated annealing parameters: temperature decrease coefficient, $\alpha=0.9$, initial temperature, $t_1=2.2$, $max\_iterations=2000$, $10\Omega=600$, $R_{STOP}=3$, $t_{min}=0.001$.

The resulting network is shown in Fig. 6. Had no conditioning or $\delta=0.5$ been chosen, the optimal number of stations would be similar, though the spatial distribution of stations quite different (cf. Figs. 5 and 6).

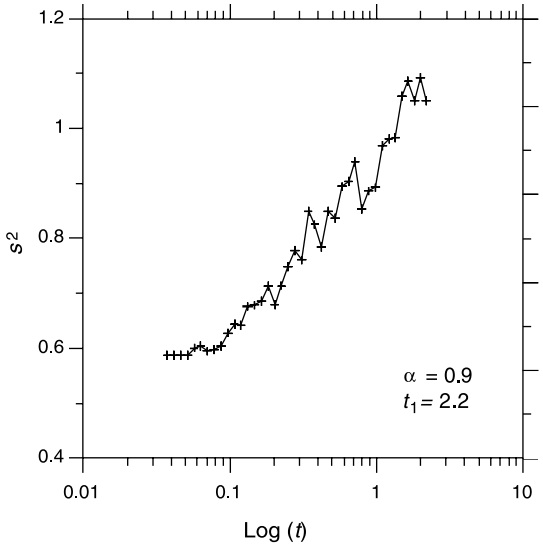


Fig. 7. Convergence results for the 60 station monitoring network with $\delta = 0.3$.

## 5. Conclusions

The following conclusions can be drawn: (i) Objective function conditioning is necessary to guarantee reproduction of the probability density functions of indicator variables; (ii) the higher the conditioning the closer the posterior (estimated) pdf is to the prior (data) pdf; (iii) conditioning with $\delta < 0.3$ leads to extremely long running times and has been shown to be unnecessary; (iv) if no conditioning is used the estimation error variance increases with the rise in the number of monitoring stations as a result of the bias introduced by variogram fitting errors; (v) the time necessary for SA to reach a solution is, in this particular case-study, more dependent on the structure of the problem than on its dimension; (vi) the optimisation problem studied here can not be solved exhaustively on account of the enormous number of possible combinations that would have to be tested; (vi) the solution attained may not be optimal globally but, locally, it should be optimal—a solution very close to the global minimum to be attained in an acceptable amount of time; (vii) a sediment monitoring network with sixty stations was obtained. In its construction this network has a proportion of stations inside each homogeneous sediment area similar to the proportions in the prior sampling program with 153 sampling stations.

## References

Aarts, E., Korst, J., 1990. Simulated Annealing and Boltzman Machines. Wiley, New York.

Ahmed, S., de Marsilly, G., Talbot, A., 1988. Combined use of hydraulic and electrical properties of an aquifer in a geostatistical estimation of transmissivity. Ground Water 26 (1), 78–85.

Alli, M.M., Nowatzki, E.A., Myers, D.E., 1990. Probabilistic analysis of collapsing soil by indicator kriging. Mathematical Geology 22 (1), 15–38.

Amorocho, J., Espildora, B., 1973. Entropy in the assessment of uncertainty of hydrologic systems and models. Water Resources Research 9 (6), 1511–1522.

Bras, R.L., Rodríguez-Iturbe, I., 1976. Network design for the estimation of areal mean of rainfall events. Water Resources Research 12 (6), 1185–1195.

Brus, D.J., Jansen, M.J.W., de Haan, W.F., 2000. Designing efficient sampling schemes for reconnaissance surveys of contaminated bed sediments in water courses. Geologie en Mijnbouw-Netherlands Journal of Geosciences 79 (4), 441–447.

Brus, D.J., Jansen, M.J.W., Gruijter, J.J., 2002. Optimising two- and three-stage designs for spatial inventories of natural resources by simulated annealing. Environmental and Ecological Statistics 9, 71–88.

Caeiro, S., Goovaerts, P., Painho, M., Costa, M.H., 2003a. Delineation of estuarine management areas using multivariate geostatistics: the case of Sado Estuary. Environmental Science and Technology 37 (18), 4052–4059.

Caeiro, S., Goovaerts, P., Painho, M., Costa, M.H., Sousa, S., 2003b. Spatial sampling design for sediment quality assessment in estuaries. Environmental Modelling and Software 18 (10), 853–859.

Caselton, W.F., Husain, T., 1980. Hydrologic networks: information transmission. Journal of the Water Resources Planning and Management Division 106 (WR2), 503–519.

Caselton, W.F., Zidek, J.V., 1984. Optimal monitoring network designs. Statistics and Probability Letters 2, 223–227.

Cochran, W.G., 1977. Sampling Techniques. Wiley, New York.

Cohn, H., Fielding, M., 1999. Simulated annealing: searching for an optimal temperature schedule. SIAM Journal on Optimization 9 (3), 779–802.

Christakos, G., 1990. A Bayesian/maximum-entropy view to the spatial estimation problem. Mathematical Geology 22 (7), 763–777.

Christakos, G., 1992. Random Field Models in Earth Sciences. Academic Press, San Diego, CA.

Deutsch, C.V., Journel, A.G., 1992. GSLIB, Geostatistical Software Library and User's Guide. Oxford University Press, New York.

D'Or, D., Bogaert, P., Christakos, G., 2001. Application of the BME approach to soil texture mapping. Stochastic Environmental Research and Risk Assessment 15, 87–100.

Geman, S., Geman, D., 1984. Stochastic relaxation, Gibbs distributions and the Baysian restoration of images. IEEE Transaction on Pattern Analysis and Machine Intelligence PAMI-6 (6), 721–741.

Gruijter, J.J., Walvoort, D.J.J., van Gaans, P.F.M., 1997. Continuous soil maps—a fuzzy set approach to bridge the gap between aggregation levels of process and distribution models. Geoderma 77, 169–195.

Hajek, B., 1988. Cooling schedules for optimal annealing. Mathematics of Operations Research 13 (2), 311–329.

Harmancioglu, N., Alspaslan, N., 1992. Water quality monitoring network design: a problem of multi-objective decision making. Water Resources Bulletin 28 (1), 179–192.

Harmancioglu, N., Yevjevich, V., 1987. Transfer of hydrologic information among river points. Journal of Hydrology 91 (1/2), 103–118.

Husain, T., 1989. Hydrologic uncertainty measure and network design. Water Resources Bulletin 25 (3), 527–534.

IAHS, 1986.Anon., 1986. Monitoring to Detect Changes in Water QualityAnon., 1986. Proceedings of the Budapest Symposium, July, 1986, Publication No. 157. IAHS Press, France.

Journel A., 1987. Geostatistics for the Environmental Sciences, An Introduction. Applied Earth Sciences Department, Stanford University, Project No. CR 811893, California, U.S.A.

Kay, R., Alder, J., 2000. Coastal Planning and Management. Spon Press, London and New York.

Kirkpatrick, S., Gellat Jr.., C.D., Vecchi, M.P., 1983. Optimization by simulated annealing. Science 220 (4598), 671–680.

Loaiciga, H.A., 1989. An optimization approach for groundwater quality monitoring network design. Water Resources Research 25 (8), 1771–1782.

Markus, A., Welch, W.J., Sacks, J., 1999. Design and analysis for modeling and predicting spatial contamination. Mathematical Geology 31 (1), 1–22.

Matheron, G., 1963. Principles of geostatistics. Economical Geology 58, 1246–1266.

Matheron, G., 1965. Les Variables Régionalisées et Leur Estimation. Masson et Cie, Paris.

McBratney, R.W., Webster, R., Burgess, T.M., 1981. The design of optimal sampling schemes for local estimation and mapping of regionalized variables—I. Computers and Geosciences 7 (4), 331–334.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations by fast computing machines. The Journal of Chemical Physics 21 (6), 1087–1092.

Meyer, P.D., Valocchi, A.J., Eheart, J.W., 1994. Monitoring network design to provide initial detection of groundwater contamination. Water Resources Research 30 (9), 2647–2659.

Moss, M., 1986. Integrated Design of Hydrological Networks IAHS Publication No. 158. Institute of Hydrology, Wallingford, Oxfordshire and IAHS Press, France.

Nunes, L.M., Cunha, M.C.C., Ribeiro, L., 2004a. Optimal space-time coverage and exploration costs in groundwater monitoring networks. Journal of Environmental Monitoring and Assessment 93 (1-3), 103–124.

Nunes, L.M., Paralta, E., Cunha, M.C.C., Ribeiro, L., 2004b. Groundwater nitrate monitoring network optimization with missing data. Water Resources Research 40 (2), 1–18.

Oliver, M.A., Webster, R., 1989. A geostatistical basis for spatial weighting in multivariate classification. Mathematical Geology 21 (1), 15–35.

Pardo-Igúzquiza, E., 1998. Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing. Journal of Hydrology 210 (1-4), 206–220.

RIVM, 1994. An Overview of Environmental Indicators: State of the Art and Perspectives. National Institute of Public Health and the Environment (RIVM) in co-operation with the University of Cambridge, Bilthoven, Netherlands.

Rouhani, S., 1985. Variance reduction analysis. Water Resources Research 21 (6), 837–846.

Rouhani, S., Hall, T., 1988. Geostatistical schemes for groundwater sampling. Journal of Hydrology 103 (1-2), 85–102.

Sanders, T.G., Ward, R.C., Loftis, J.C., Steel, T.D., Adrian, D.D., Yevjevich, V., 1983. Design of Networks for Monitoring Water Quality. Water Resources Publications, Littleton, CO.

Siarry, P., 1997. Enhanced simulated annealing for globally minimizing functions of many continuous variables. ACM Transactions on Mathematical Software 23 (2), 209–228.

Sullivan J.A., 1984. Non-parametric estimation of spatial distribution. PhD Thesis. Stanford University, Stanford.

van Groenigen, J.W., Stein, A., 1998. Constrained optimisation of spatial sampling using continuous simulated annealing. Journal of Environmental Quality 27 (5), 1078–1086.

van Groenigen, J.W., Siderius, W., Stein, A., 1999. Constrained optimisation of soil sampling for minimisation of the kriging variance. Geoderma 87, 239–259.