

# MetaHeuristics for a Non-Linear Spatial Sampling Problem

Eric M. Delmelle

Department of Geography and Earth Sciences

University of North Carolina at Charlotte

eric.delmelle@uncc.edu

## 1 Introduction

In spatial sampling, once samples of the primary variable have been collected, it is possible to augment the initial set by collecting additional measurements at other locations, a method known as second-phase sampling (Cressie 1991, Muller 1998, van Groenigen and Stein 1998 and recently de Gruitjer *et al.* 2006). Following a first sampling phase, the kriging variance is computed at each location using a covariogram function. Generally, additional observations are gathered away from existing points, that is where the kriging variance is large (see for instance Van Groenigen and Stein 1998). However, when the process under study is not stationary, sampling efforts should be directed in those strategic locations exhibiting strong spatial variation locally (Delmelle and Goovaerts 2009). In this paper, we formulate these two objectives into a single weighted-objective function -referred to as the weighted kriging variance-, where the weights reflect the roughness of the spatial process.

This objective function is highly non-linear (inversion of covariance matrices), and calls for robust heuristic methods. Additional samples can be collected *sequentially*, for instance by adding one sample at a time to the initial set. This procedure may be suboptimal but fast since it requires the inversion of a matrix augments by only one entry.

Practically, a covariogram summarizing the spatial variation in the observed variable with distance is determined following the collection of initial samples. Based on the covariance structure, the kriging variance is computed at each grid node, and weighted by the local variation at that node. The objective consists of locating those additional samples strategically to maximize the change in weighted kriging variance. Heuristic methods decide on the location of new samples. For instance a greedy algorithm will allocate additional observations on the peaks of the weighted kriging variance

surface but these local maxima may not be optimal to the objective function. In this paper, we propose a combination of heuristic methods: first, additional samples are determined using a sequential greedy algorithm and the objective function evaluated. Second, the points obtained using a greedy algorithm are used as a starting solution in simulated annealing. Through a swapping procedure, additional points are exchanged for other potential points, while the objective function is recomputed. This *metaheuristic* procedure combines the advantage of the greedy algorithm, that is its rapidity, with simulated annealing, which is recognized for its convergence towards optimal solutions.

## 2 Additional sampling methodology

A variable of interest  $Y$  has been measured at  $m$  locations within a study region,  $\mathfrak{D}$ . Measurements are denoted  $y(\mathbf{s}_i)$ ,  $\forall i = 1 \dots m$  (Goovaerts 1997). Using data values of the primary variable and a covariogram function, the kriging variance at a gridpoint  $\mathbf{s}_g$ :

$$\left(\sigma_k(\mathbf{s}_g)\right)^2 = \sigma^2 - \mathbf{c}^T(\mathbf{s}_g) \cdot \mathbf{C}^{-1} \cdot \mathbf{c}(\mathbf{s}_g), \quad (1)$$

where  $\mathbf{C}^{-1}$  is the inverse of the covariance matrix  $\mathbf{C}$  based on the covariogram function. The term  $\mathbf{c}$  is a column vector and  $\mathbf{c}^T$  its corresponding row vector. The Average Kriging Variance (AKV) is obtained by integrating Equation 1 over the area  $\mathfrak{D}$ . Computationally, discretizing  $\mathfrak{D}$  over a fine grid of points (set  $G$ ):

$$AKV = \int_{\mathfrak{D}} \left(\sigma_k(\mathbf{s}_g)\right)^2 \approx \frac{1}{|G|} \sum_{g \in G} \left(\sigma_k(\mathbf{s}_g)\right)^2 \quad (2)$$

Our first objective  $Z[S]$  is to select a set of  $n$  points to our existing set of  $m$  samples, which will maximize the change in kriging variance by as much as possible. This process can be thought as a simulation of what the change in kriging variance is expected to be, without having to collect additional points, assuming the covariogram structure would remain constant (Burgess, Webster and McBratney 1981 as well as Cressie 1993). Specifically:

$$\underbrace{\text{MAXIMIZE}}_{\{\mathbf{s}_{m+1}, \dots, \mathbf{s}_{m+n}\}} Z[S] = \frac{1}{|G|} \sum_{g \in G} \left(\sigma_k^{\text{old}}(\mathbf{s}_g)\right)^2 - \left(\sigma_k^{\text{new}}(\mathbf{s}_g)\right)^2, \quad (3)$$

where  $S$  denotes the sampling scheme. The set  $\mathbf{P}$  of  $p$  potential points is obtained by discretizing  $\mathfrak{D}$ , generating a total of  $\binom{p}{n}$  possible sampling combinations.

The kriging variance is unfortunately misused as a measure of reliability of the kriging estimate, as noted by several authors (Deutsch and Journel 1992; Armstrong 1994).

It is merely a function of the sample pattern, sample density, the numbers of samples and their covariance structure. The kriging variance assumes that the errors are independent of each other, which means that the process is stationary, an assumption violated in practice. Figure 1 illustrates the limitation of the kriging variance (Armstrong 1994), the objective being to interpolate the value of the inner grid point, highlighted with a question mark. The interpolation is a function of the values at the four surrounding observations. In scenario *b*, three very similar values and an extreme one. The scenario in *a* however shows four data values in a very narrow range. Assuming a similar spatial structure in both cases and given that the configuration of the data points is the same, the kriging variances are identical, and so are the kriged estimates. Nevertheless, since there is much less variation among its neighbors, the left-hand side scenario is a much safer option than the right hand-one when it comes to estimating the value of the primary variable.

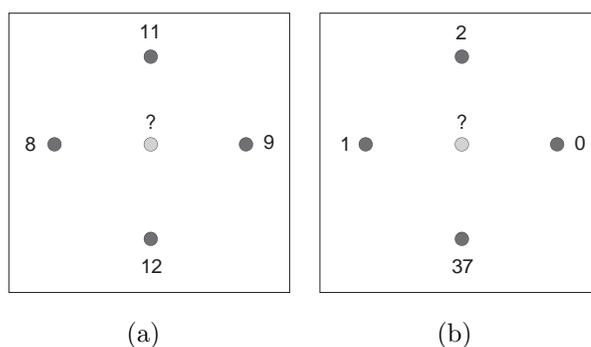


Figure 1: Example of two-dimensional nonstationarity. Dark points are used as data values to interpolate the center point (light gray). After Armstrong (1994).

This example illustrates the importance to account for local variations in the observed variable. Let  $\hat{y}(\mathbf{s}_g)$  be the interpolated value of the primary variable  $Y$  at a grid node  $\mathbf{s}_g$ . Estimating by how much that grid node is different in value from its surrounding points  $\mathbf{s}_j$  ( $j = 1, 2, \dots, J$ ) is possible through a filter process, specifically, a circular filter is constructed around each grid node  $\mathbf{s}_g$  that encompasses its neighbors. For illustration purposes, Figure 2 illustrates a 3 by 3 window, however the methodology can handle various neighborhood sizes. To determine an appropriate moving window size  $J$ , we compute the squared difference in interpolated value between the central grid node  $\hat{y}(\mathbf{s}_g)$  and the surrounding ones  $\hat{y}(\mathbf{s}_j)$ . We also introduce a distance factor  $d(\mathbf{s}_j, \mathbf{s}_g)$  and a parameter  $\beta$ , both regulating the importance given to nearby points. This is then summed over the set  $G$ . The weight  $\lambda(\mathbf{s}_g)$  becomes:

$$\lambda(\mathbf{s}_g) = \sum_{j=1, j \neq g}^J \frac{d(\mathbf{s}_j, \mathbf{s}_g)^{-\beta} \cdot (\hat{y}(\mathbf{s}_j) - \hat{y}(\mathbf{s}_g))^2}{\sum_{j=1, j \neq g}^J d(\mathbf{s}_j, \mathbf{s}_g)^{-\beta}} \quad (4)$$

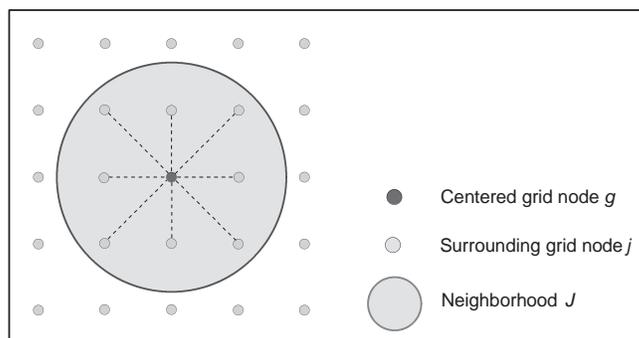


Figure 2: A  $3 \times 3$  moving window: a circle is passed around a grid node within a specific distance.

If the neighborhood  $J$  is kept constant,  $\lambda(\mathbf{s}_g)$  will exhibit great values when  $\beta < 1$ , because more weight is given to far away data points. As  $\beta$  increases,  $\lambda(\mathbf{s}_g)$  decreases and flattens out for high values of  $\beta$ . If  $J$  is too large, zones of rapid changes may go undetected. Equation 3 should be account for spatial variation of the primary variable. As such, a weighted second-phase sampling problem can be formulated as a single-weighted objective (Cressie 1991) where the kriging variance is weighted by Equation 4:

$$\underbrace{\text{MAXIMIZE}}_{\{\mathbf{s}_{m+1}, \dots, \mathbf{s}_{m+n}\}} Z[S] = \frac{1}{|G|} \sum_{g \in G} \lambda(\mathbf{s}_g) \cdot \left| \left( \sigma_k^{\text{old}}(\mathbf{s}_g) \right)^2 - \left( \sigma_k^{\text{new}}(\mathbf{s}_g) \right)^2 \right| \quad (5)$$

### 3 Application

In this paper, we use a sequential approach to strategically allocate new observations. To illustrate our methodology, we use primary data on soil concentration of Chromium (Cr) in a study area near La Chaux de Fonds, in the Swiss Jura (see, Goovaerts 1997 for the dataset). The Cr-concentration  $\frac{mg}{kg}$  represents the quantity of the heavy metal per kilogram of soil sampled.

Sequential addition assumes that one additional point has to be added to the initial set  $n$ -times. Once the first point has been selected and added to the initial set  $M$ ,  $n - 1$  additional locations are to be chosen in a similar, sequential fashion. The sequential addition approach is illustrated using algorithms such as random strategy, total enumeration, greedy, simulated annealing and simulated annealing with greedy start. The greedy approach has the drawback of getting stuck at local optima, while total enumeration is not time-efficient. Since simulated annealing has the inherent property of jumping out of a local optimum, we capitalize on this technique for finding the optimal solution  $S^*$  to the sequential addition, using a cooling factor  $\kappa$  at the end of a fixed number of iterations  $T_{it}$ . Similarly, the step size for determining new neighbors (for swapping purposes) was reduced by a factor  $\delta$ . A large initial step size

$\delta$  of 3 kilometers -corresponding to approximately half the size of the study area  $\mathcal{D}$ - was chosen to permit wide jump swaps.

Figure 3 to the left illustrates the performance of the greedy algorithm and total enumeration in maximizing the change in weighted kriging variance with the addition of new samples, against the changes obtained using naive addition. The total enumeration evaluates all possible solutions to the sequential addition, but may still be suboptimal. To check on global optimality (14.879%), we ran a simultaneous simulated annealing (see Van Groenigen and Stein 1998) and found that sequential results were very close to the optimal. The sequential total enumeration yielded a 14.8% improvement in the objective function. For the naive (random) addition, a total of 1500 simulations were performed, providing a good lower bound to evaluate other heuristics. In the best-case scenario, a reduction of 7.52% was obtained, in comparison with a change of 4.89% in the worst case. Table 1 reports on the computational time. When simulated annealing is used (Figure 3 to the right), the algorithm returns near optimal solutions, even more so when the algorithm uses a lower cooling schedule ( $\kappa$  closer to 1), and a greater number of iterations per temperature steps  $T_{it}$ .

Sequential heuristic	Time (min)	Reduction (%)	Optimality gap (%)
Total enumeration	229.72	14.768	.75
Naïve	8.56	[2.869; 7.521]	[80.72; 49.45]
Average naïve	8.56	4.892	67.12
Greedy	8.04	12.537	15.74
SA-Greedy( $\kappa = .875, \beta = .9$ )	106.76	14.768	.75
SA-Greedy( $\kappa = .35, \beta = .45$ )	33.35	14.649	.8
SA( $\kappa = .95, \delta = .965$ )	241.06	14.733	.98
SA( $\kappa = .35, \delta = .45$ )	33.82	14.420	3.08
SA( $\kappa = .05, \delta = .05$ )	26.50	13.95	6.24
Simultaneous heuristic SA	1500	14.879	0

Table 1: Average reduction (%), and optimality gap (%) for the sequential and simultaneous addition after the addition of  $n = 30$  points.

The combination of SA with a greedy start allows improvement upon a first very good solution. Since the starting solution is relatively good, SA may experience difficulties to improve upon that incumbent. Figure 4A shows the first 15 dynamic moves, with SA parameters  $\kappa = .875, \beta = .9$ , yielding the sequential optimal in 106.76 minutes. The location exhibiting the highest weighted kriging variance (point  $\mathbf{a} = \mathbf{s}_{m+1}^+$ ) is selected and serves as a starting point for SA, yet the latter is unable to locate a better point, hence  $\mathbf{a} = 1 = \mathbf{s}_{m+1}^+$ . That point is added to the set  $\mathbf{M}$  and the weighted kriging variance is re-computed accordingly. Location  $\mathbf{b} = \mathbf{s}_{m+2}^+$  is the point with the highest kriging variance and is selected as the starting point. SA finds a better sample at location **2**-symbolized by a white dot, and that point

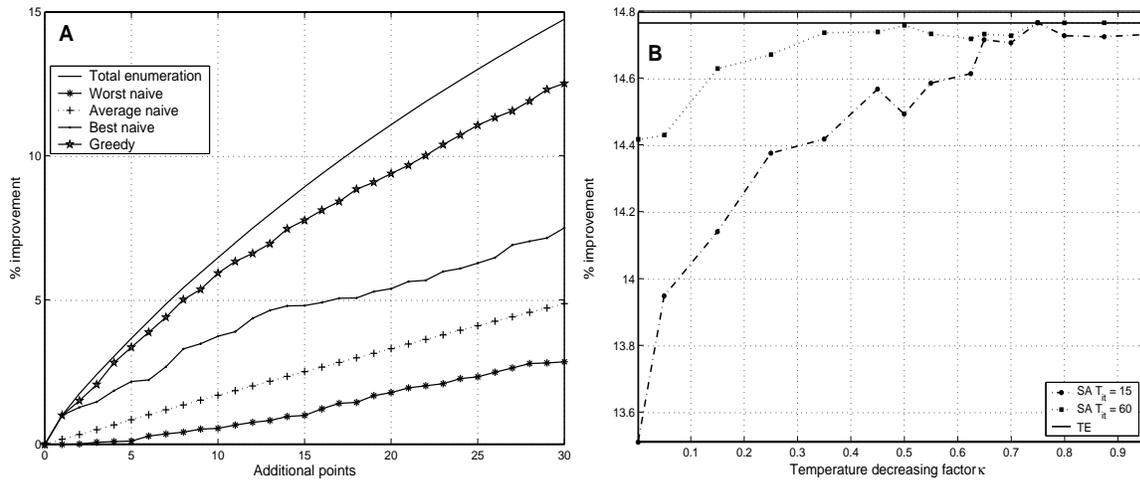


Figure 3: Percentage reduction in weighted kriging variance using a naïve approach versus total enumeration. The sensitivity of the sequential SA coupled with greedy to the cooling factor  $\kappa$  is illustrated in B. Notice for  $T_{it} = 60$  how near-optimal solutions are obtained even if the temperature drops quickly ( $\kappa = [.1; .25]$ ).

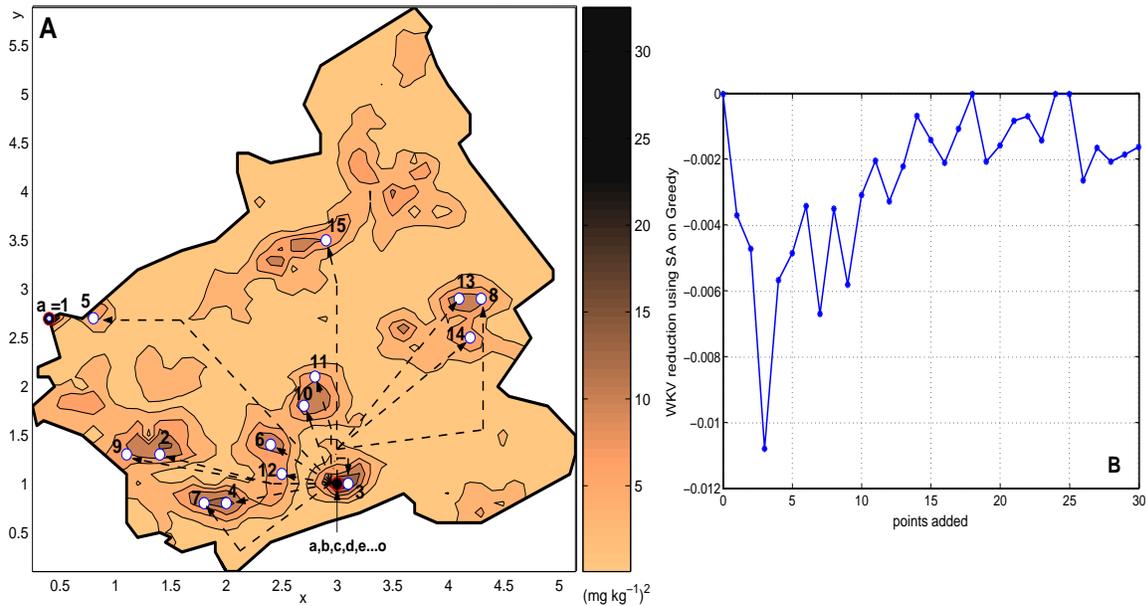


Figure 4: Illustration of the simulated annealing algorithm using a greedy start approach for the first 15 points (A), in the best-case scenario ( $\kappa = .875, \beta = .9$ ). Black dots denote initial points obtained using greedy. The arrows point to the locations obtained using SA. Graph B illustrates the reduction between successive steps using SA on greedy for  $n = 30$  points.

is added to  $\mathbf{M}$ . The weighted kriging variance is computed with the set  $\mathbf{M}$  that contains now two new samples, namely points **1** and **2**. In the following 17 additions, SA will ameliorate the incumbent greedy solution (see Figure 4B). Notice how often SA discovers a better solution from the initial greedy sample, yet the magnitude of that improvement decreases as new samples are being added.

## 4 Conclusions:

In this paper, we have addressed the second-phase spatial sampling problem based on two main criteria; the change in kriging variance, and the spatial variation of the primary variable. Results of our numerical testing showed that total enumeration outperformed all other heuristics in the sequential case, but at the cost of an extended running time. The greedy approach, which locates new samples points where the weighted kriging variance is the highest, returns near-optimal results in a short time-frame. Simulated annealing is very sensitive to the choice of the cooling factor, that governs the search procedure. The combination of simulated annealing with a greedy start performed remarkably well considering the optimality gap and the computational time.

### References

- Armstrong M. (1994). Is research in mining geostats as dead as dodo? *In*: Dimitrakopoulos R. (Ed.) *Geostatistics for the Next Century*. Kluwer Academic Publisher. Dordrecht: 303-312.
- Burgess T.M., Webster R. and A.B. McBratney (1981). Optimal interpolation and isarithmic mapping of soil properties: IV. Sampling strategy. *Journal of Soil Science*, vol. **32**: 643-659.
- Cressie, N., 1991. *Statistics for Spatial Data*. Wiley, New York, USA, 900p.
- De Gruijter, J., Brus, D.J., Bierkens, M.F.P. and Knotters M., 2006. *Sampling for Natural Resource Monitoring*. Springer, 332p
- Delmelle E. and P. Goovaerts (2009). Second-phase sampling designs for non-stationary spatial variables. *Geoderma* 153: 205-216
- Deutsch C.V. and A.G. Journel (1997) *Gslib: Geostatistical Software Library and User's Guide*. Oxford University Press, 2<sup>nd</sup> edition, 369p.
- Goovaerts P., 1997. *Geostatistics for natural resources evaluation*. 483p.
- Muller, W., 1998. *Collecting Spatial Data: Optimal Design of Experiments for Random Fields*. Heidelberg: Physica-Verlag.
- Van Groenigen, J.W. and Stein, A., 1998. Constrained optimization of spatial sampling using continuous simulated annealing. *Journal of Environmental Quality*, vol. 27: 1078-1086.