MetaHeuristics for a Non-Linear Spatial Sampling Problem

Eric Delmelle eric.delmelle@uncc.edu
Geography and Earth Sciences, University of North Carolina at Charlotte, U.S.A.

Summary

After samples of a primary variable are collected, it is possible to augment the initial set by collecting additional measurements at other locations, a method known as **second-phase sampling**.

Objectives are to gather new information:

- away from existing points
- where the phenomenon under study exhibits a strong spatial variation
- where covariates can not predict accurately the outcome of the primary variable.

⇓

single non-linear weighted-objective function

Scenarios for sample collection.

- *Sequential addition*, adding one sample at a time to initial set.
- *Simultaneous addition*, adding a new set of samples once.
- *Hybrid combination* (metaheuristic).

Which heuristics?

- Total enumeration
- Greedy
- Simulated annealing
- Simulated annealing with greedy start

Findings

- **Greedy**: good solution in a limited time-frame.
- **Simulated annealing**: fast convergence towards optimal solutions.
  − Cooling schedule impacts quality of solution
- **Metaheuristic**: Pairing *SA* with *greedy start* capitalizes on the rapidity of the greedy algorithm and the ability of SA to escape from a local optimum.

Kriging variance

With initial samples, the kriging variance is computed at each location using a covariogram function. Additional observations are gathered away from existing points, where the kriging variance is large (see for instance Van Groenigen and Stein 1998).
If process is not stationary, should sampling efforts should be directed in those strategic locations exhibiting strong spatial variation locally (Delmelle and Goovaerts 2009)?

Kriging variance: 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 $\mathbf{s}_g$:

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

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):

$$AKV = \int_{\mathfrak{D}} \Big(\sigma_k(\mathbf{s}_g)\Big)^2 \approx \frac{1}{\lfloor G \rfloor} \sum_{g \epsilon G} \Big(\sigma_k(\mathbf{s}_g)\Big)^2 \qquad (2)$$

⇓

I. Maximizing change in kriging variance

Our **first objective** $Z[S]$ is to select a set of n points to our exisiting 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:

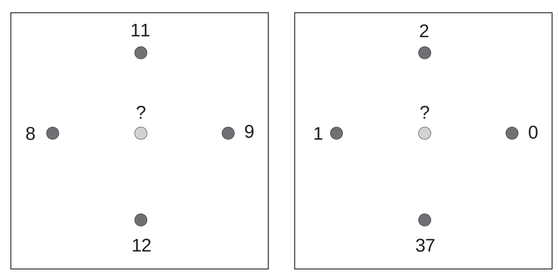
$$\underbrace{\text{Maximize}}_{\{\mathbf{s}_{m+1},\dots,\mathbf{s}_{m+n}\}} Z[S] = \frac{1}{\lfloor G \rfloor} \sum_{g \epsilon G} \Big(\sigma_k^{\text{old}}(\mathbf{s}_g)\Big)^2 - \Big(\sigma_k^{\text{new}}(\mathbf{s}_g)\Big)^2, \qquad (3)$$

where S denotes the sampling scheme.

Spatial roughness

The kriging variance is a function of the sampling pattern, sample density, the numbers of samples and their covariance structure, and assumes a stationary process, an assumption violated in practice (Deutsch and Journel 1992; Armstrong 1994).

Safest scenario for interpolation?



⇓

can attribute information be used in reformulating sampling objective?

Computing roughness with filter process:
Let $\widehat{y}(\mathbf{s}_g)$ be the interpolated value of the primary variable Y at a grid node $\mathbf{s}_g$. Estimate squared difference between interpolated value at grid node from its surrounding points $\mathbf{s}_j$ $(j = 1, 2, \dots J)$ using a **circular filter** with moving window size J.

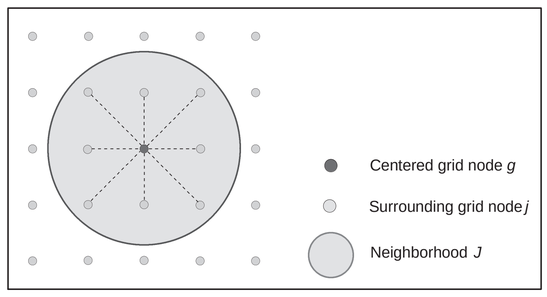


FIGURE 1: A 3×3 moving window filter process.

Distance factor $d(\mathbf{s}_j, \mathbf{s}_g)$ and parameter β regulate the importance given to nearby points. The weight $\lambda(\mathbf{s}_g)$ becomes:

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

$\lambda(\mathbf{s}_g)$ will exhibit great values when $\beta < 1$, because more weight is given to far away data points. As β increases, $\lambda(\mathbf{s}_g)$ decreases and flattens out for high values of β.

⇓

II. Maximizing change in weighted kriging variance: a non-linear problem

Our second objective is to **account for non-stationarity** under the form of weights → single-weighted objective:

$$\underbrace{\text{Maximize}}_{\{\mathbf{s}_{m+1},\dots,\mathbf{s}_{m+n}\}} Z[S] = \frac{1}{\lfloor G \rfloor} \sum_{g \epsilon G} \lambda(\mathbf{s}_g) \cdot \Big| \Big(\sigma_k^{\text{old}}(\mathbf{s}_g)\Big)^2 - \Big(\sigma_k^{\text{new}}(\mathbf{s}_g)\Big)^2 \Big| \qquad (5)$$

Heuristic techniques

Simulated annealing (SA) has the ability to escape local optima. The algorithm employs a random search that accepts changes improving the objective function, but also non-improving moves with probability δ_T (T= current temperature and cools down as the algorithm progresses).

$$\delta_T\big\{J(k) \rightarrow J(k+1)\big\} = 1 \quad \text{if} \quad Z\big[J(k+1)\big) \geq Z\big[J(k)\big]$$
$$\delta_T\big\{J(k) \rightarrow J(k+1)\big\} = \frac{1}{1 + e^{\left(\frac{\triangle Z}{T}\right)}} \quad \text{if} \quad Z\big[J(k+1)\big) < Z\big[J(k)\big]$$

To find the optimal point $\mathbf{s}^*$ to be added -or nearly optimal $\mathbf{s}^+$, use a high starting temperature T_{ini} and a cooling factor κ close to 1 → algorithm escapes local maximum.

Total enumeration evaluates all possible solutions to the sequential addition, but may still be suboptimal.

Greedy allocates new samples which corresponds to the highest peak of the surface.

Simulated annealing with greedy start allows improvement upon a first very good solution. When greedy solution is near optimal, SA may experience difficulties to improve upon that incumbent.

Application

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).

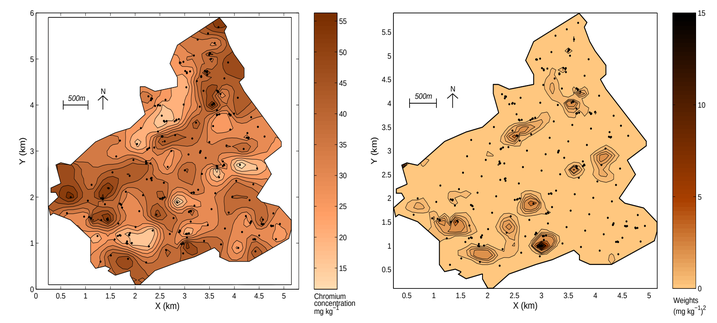


FIGURE 2: Kriged map of the Cr content and the weighted surface

In simulated annealing, at each temperature level, a fixed number of iterations T_{it} is run, then the temperature cools with factor κ, and so does the step size for determining new neighbors.

Illustration of SA with greedy start. First 15 dynamic moves, with SA parameters $\kappa = .875, \zeta = .9$. 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** (white dot). In the following 17 additions, SA will ameliorate the incumbent greedy solution

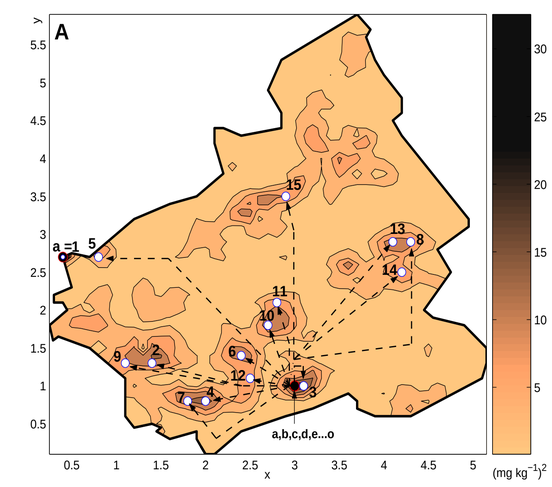


FIGURE 3: Simulated annealing algorithm using a greedy start approach for the first 15 points (A). (Black dots = initial points using greedy). The arrows point to the locations obtained using SA.

	Time (min)	ΔZ (%)	Opt. gap (%)
Total enumeration	**229.72**	**14.768**	.75
Naïve	8.56	[2.869; 7.521]	[80.72; 49.45]
Greedy	**8.04**	**12.537**	15.74
SA-Greedy($\kappa = .875, \zeta = .9$)	**106.76**	**14.768**	.75
SA-Greedy($\kappa = .35, \zeta = .45$)	**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**

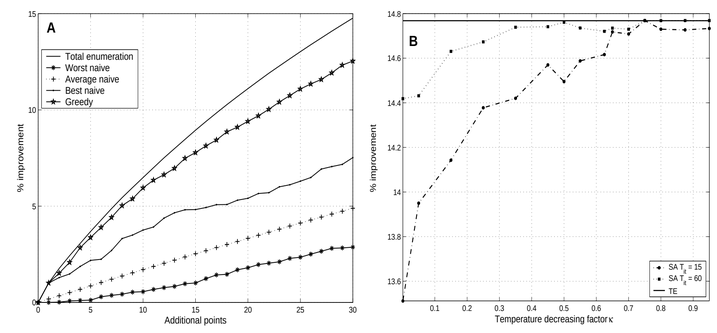


FIGURE 4: Percentage reduction in weighted kriging variance, with sensitivity of the sequential SA coupled with greedy to the cooling factor κ to the right.

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, 2nd edition, 369p.

Goovaerts P., 1997. *Geostatistics for natural resources evaluation*. 483p.

Michalewicz Z. and D. Fogel (2000). *How to Solve It: Modern Heuristics*. Springer. 467p.
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.