3rd *BYMAT Conference* (2020)

A multi-objective sampling design for spatial prediction problem

Elaheh Lotfian
Tarbiat Modares University
e.lotfian@modares.ac.ir

Mohsen Mohammadzadeh
Tarbiat Modares University
mohsen_m@modares.ac.ir

Abstract: In many fields such as image classification, geostatistical surveys, and air pollution, regarding the limitations of resources, time and technology, spatial sampling plays a crucial role. In spatial sampling, a set of sample locations are chosen such that the spatial prediction at unobserved locations be optimal. While many studies are focused on only one objective function as the predictor variance, and the predictor entropy, in applied problems we are interested in more than one objective. In this study, the optimization problem of spatial sampling with minimum cost is investigated from both the perspective of covariogram estimation and kriging variance. For this purpose, a bi-objective optimization problem of soil sampling has been considered. The first objective function is the mean total error and the second objective function is the cost of the distance travelled by the sampler. The mean total error is the sum of the ordinary kriging variance and uncertainties of the estimated covariogram parameters. The non-dominated sorting genetic algorithm-II and Taguchi method is applied to this problem. The results show the proper performance of this algorithm in multi-objective spatial sampling for spatial predictions.

Resumen: El muestreo espacial juega un papel crucial en muchos campos, como la clasificación de imágenes, los estudios geoestadísticos y la contaminación atmosférica, con respecto a las limitaciones de recursos, tiempo y tecnología. En el muestreo espacial, se elige un conjunto de ubicaciones de la muestra de forma que la predicción espacial en las ubicaciones no observadas sea óptima. Mientras que muchos estudios se centran en una sola función objetivo, como la varianza y la entropía del predictor, en los problemas aplicados interesa más de un objetivo. En este estudio, el problema de optimización del muestreo espacial con coste mínimo se investiga tanto desde la perspectiva de la estimación del covariograma como de la varianza de kriging. Para ello, se ha considerado un problema de optimización bi-objetivo de muestreo de suelos. La primera función objetivo es el error total medio y la segunda función objetivo es el coste de la distancia recorrida por el muestreador. El error total medio es la suma de la varianza ordinaria de kriging y las incertidumbres de los parámetros estimados del covariograma. Se aplica a este problema el algoritmo genético de ordenación no dominante-II y el método de Taguchi. Los resultados muestran el buen funcionamiento de este algoritmo en el muestreo espacial multiobjetivo para las predicciones espaciales.

Keywords: spatial sampling, spatial prediction, multi-objective optimization, uncertainty.

MSC2010: 62H11, 58E17.

Acknowledgements: Receiving support from the Center of Excellence in Analysis of Spatial and Spatio-Temporal Correlated Data at Tarbiat Modares University is acknowledged.

Reference: Lotfian, Elaheh, and Mohammadzadeh, Mohsen. “A multi-objective sampling design for spatial prediction problem”. In: *TEMat monográficos*, 2 (2021): *Proceedings of the 3rd BYMAT Conference*, pp. 163-166. ISSN: 2660-6003. URL: https://temat.es/monograficos/article/view/vol2-p163.

1. Introduction

Optimization of spatial sampling is a critical issue applied in many areas, including geostatistics, air pollution, and epidemiology. In spatial sampling, a set of sample locations is chosen. The spatial prediction at unobserved locations is optimal for the predictor variance and the predictor entropy. While there are several types of research on the single-objective spatial sampling, in many applied problems we are interested in more than one objective in sampling [4]. This paper aims to use Non-dominated Sorting Genetic Algorithm-II (NSGA-II) in spatial sampling subject to the spatial correlation of data. The rest of this paper is organized as follows. Section 2 is devoted to the problem statement. Materials and methods are introduced in Section 3. This section recalls some preliminaries on multi-objective optimization theory and customizes the NSGA-II for the introduced spatial sampling optimization model. Numerical results are provided in Section 4.

2. Problem statement

In this paper, the soil sampling of a field in Silsoe, Bedfordshire, UK, shown in Figure 2, is investigated [5]. It is assumed that the sampler enters and leaves the field at the corner $\{100, 100\}$ and collects 50 sample points across this domain. The first objective function is the spatial mean total error defined by [5, 7]

$$\bar{\sigma}_P^2 = \frac{1}{\mathcal{A}} \int_{s\in\mathcal{A}} (\sigma^2_{OK}(\mathbf{s}) + E[\tau^2(\mathbf{s})])\,\mathrm{d}s, \tag{1}$$

where $\sigma^2_{OK}(\mathbf{s})$ and $E[\tau^2(\mathbf{s})]$ are the squared prediction error (ordinary kriging variance) and the uncertainty in the estimated spatial model (covariogram) parameters, respectively, and

$$\sigma^2_{OK}(s_0) = Var(Z(s_0) - \tilde{Z}(s_0|\boldsymbol{\theta})) = \boldsymbol{C}(s_0 - s_0|\boldsymbol{\theta}) - \boldsymbol{\lambda}^T\boldsymbol{d},$$

$$E[\tau^2(s_0)] = \sum_{i=1}^{q}\sum_{j=1}^{q} Cov(\theta_i, \theta_j)\frac{\partial \boldsymbol{\lambda}^T}{\partial \theta_i}\boldsymbol{C}\frac{\partial \boldsymbol{\lambda}}{\partial \theta_j},$$

where $\boldsymbol{\lambda}^T$, $\frac{\partial \lambda}{\partial \theta_i}$, and $Cov(\theta_i, \theta_j)$ are the vector of kriging weights, the n-vector of partial derivatives of the kriging weights with respect to the ith variance parameter, and the covariance between the ith and jth parameters, respectively. Furthermore,

$$\begin{pmatrix} (\lambda_i)_{i=1}^n \\ \psi \end{pmatrix} = \underbrace{\begin{bmatrix} \left(C(s_i - s_j|\theta)\right)_{i,j=1}^n & \vec{\mathbf{1}} \\ \vec{\mathbf{1}} & 0 \end{bmatrix}^{-1}}_{A} \times \underbrace{\begin{pmatrix} (C(s_0 - s_i|\theta))_{i=1}^n \\ 1 \end{pmatrix}}_{d},$$

$$\frac{\partial \boldsymbol{\lambda}}{\partial \theta_i} = \boldsymbol{A}^{-1}\left(\frac{\partial \boldsymbol{d}}{\partial \theta_i} - \frac{\partial \boldsymbol{A}}{\partial \theta_i}\boldsymbol{A}^{-1}\boldsymbol{d}\right),$$

$$Cov(\theta_i, \theta_j) \approx \boldsymbol{F}^{-1}(\theta_i, \theta_j) = \left(\frac{1}{2}\,\mathrm{Tr}\left[\boldsymbol{C}^{-1}\frac{\partial \boldsymbol{C}}{\partial \theta_i}\boldsymbol{C}^{-1}\frac{\partial \boldsymbol{C}}{\partial \theta_j}\right]\right)^{-1}.$$

Note that C is the spherical covariogram [1]. Marchant and Lark [5] and Wadoux *et al.* [7] investigated some single-objective optimization problems with the same objective function (1). We consider the cost of the sampling defined by $Cost = d(\mathbf{s}) \times C_m$ as the other objective function where C_m and $d(\mathbf{s})$ are a fixed cost per each meter travelled by the sampler and the total distance walked to visit all points, respectively. We set, $C_m = 1$ and $d(\mathbf{s}) = \sum_{i=1}^{n-1} \|s_{i+1} - s_i\|$, $\mathbf{s} = (s_1, \dots, s_n)$.

3. Materials and methods

In this section, we provide some preliminaries on multi-objective optimization theory and Spatial NSGA-II Algorithm.

3.1. Multi-objective optimization

A general multi-objective optimization problem can be formulated as follows:

(2) $$\min f(x) = (f_1(x), \dots, f_m(x)), \qquad g_j(x) \le 0, \quad j \in J_\ell, \quad x \in X \subseteq \mathbb{R}^n,$$

where $f_i, g_j : \mathbb{R}^n \to \mathbb{R}$ ($i \in I_m := \{1, \dots, m\}$, $j \in J_\ell := \{1, \dots, \ell\}$), and $m > 1$. The nonempty set $S := \{x \in X \subseteq \mathbb{R}^n : g_j(x) \le 0, j \in J_\ell\}$ is called the feasible set.

Definition 1 ([3]). A feasible solution $\hat{x} \in S$ is called efficient or Pareto optimal solution of Problem (2), if there is no $x \in S$ such that $f_k(x) \le f_k(\hat{x})$ for each $k \in I_m$ and $f_i(x) < f_i(\hat{x})$ for some $i \in I_m$. The set of all Pareto optimal solutions of Problem (2) is called Pareto frontier for this problem. ◂

Now, the bi-objective spatial optimization problem can be represented as

$$\min_{\mathbf{s} \in \mathcal{A}} f(\mathbf{s}) = (\bar{\sigma}_P^2, d(\mathbf{s})),$$

$$\text{s.t. } \bar{\sigma}_P^2 - 1 \le 0, \quad d(\mathbf{s}) - 5000 \le 0, \|s_i - sj\| \ge 20 \quad \forall s_i, s_j \in \mathcal{A}, i \ne j,$$

where $\mathcal{A}$ is the interested area of study.

3.2. Spatial NSGA-II Algorithm

The NSGA-II algorithm is suggested by [2]. In the Figure 1, the diagram of the customized version of the NSGA-II for spatial sampling is shown. Note that the parameters of the algorithm are set by the Taguchi method [6].

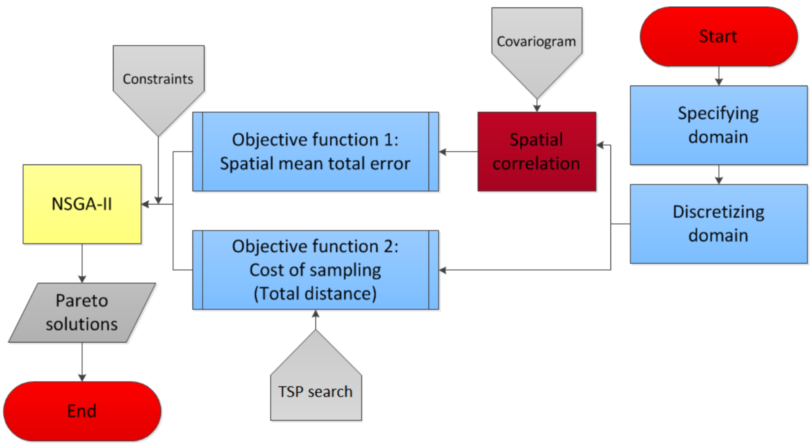

Figure 1: Diagram of the customized NSGA-II for the bi-objective spatial sampling optimization problem.

4. Numerical results

Note that, in Figure 2, each point on the Pareto frontier corresponds to a sample of 50 points in the field. Results show the suitable dispersion and coverage on the Pareto frontier obtained by the customized NSGA-II. It is clear from this frontier that there is a conflict between these two objective functions. For instance, sample 2 shows that minimizing the mean total error requires more cost.

As a sensitivity analysis, we change the parameters of the spherical covariogram. These values are chosen from the literature. In Table 1, in case 3, the mean of mean total errors is smallest, and in case 5, the mean of costs is smallest. Now, the choice of these parameters is also dependent on the decision-makers and their priorities. If they want less total mean error, they must choose parameters in case 3, and for less cost, they choose parameters in case 5.

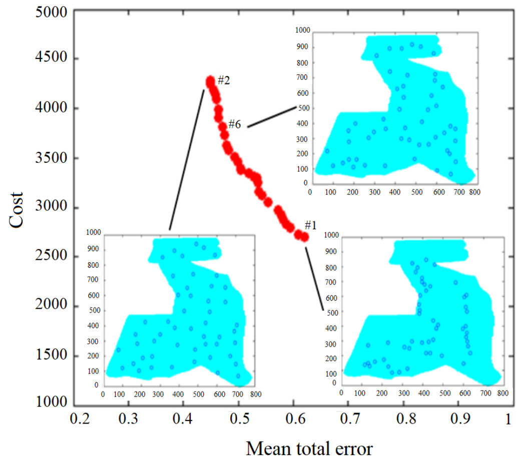

Figure 2: The region of interest and three sample designs from the estimated Pareto frontier obtained by NSGA-II with the parameters shown in the Table 1, case 1.

Table 1: Spherical covariogram parameters c_0, c_1, and a, and the mean of the objective functions values for each optimized sampling scheme.

Case	c_0	c_1	a	mean of $\bar{\sigma}_P^2$	mean of *Cost*
1	0.127	1.0	240.0	0.5173	3438
2	0.127	1.5	240.0	0.6610	3418
3	0.127	0.5	240.0	0.3395	3512
4	0.127	1.0	120.0	0.7675	3490
5	0.127	1.0	90.0	0.8941	3315

References

[1] Cressie, Noel. *Statistics for spatial data.* John Wiley & Sons, 2015.

[2] Deb, K.; Pratap, A.; Agarwal, S., and Meyarivan, T. “A fast and elitist multiobjective genetic algorithm: NSGA-II”. In: *IEEE Transactions on Evolutionary Computation* 6.2 (2002), pp. 182–197. https://doi.org/10.1109/4235.996017.

[3] Ehrgott, Matthias. *Multicriteria optimization.* Vol. 491. Springer Science & Business Media, 2005.

[4] Lark, R. M. “Multi-objective optimization of spatial sampling”. In: *Spatial Statistics* 18 (2016), pp. 412–430. https://doi.org/10.1016/j.spasta.2016.09.001.

[5] Marchant, B. P. and Lark, R. M. “Optimized Sample Schemes for Geostatistical Surveys”. In: *Mathematical Geology* 39 (2007), pp. 113–134. https://doi.org/10.1007/s11004-006-9069-1.

[6] Taguchi, Genichi. *Introduction to quality engineering: designing quality into products and processes.* Tech. rep. 1986.

[7] Wadoux, Alexandre M. J.-C.; Marchant, Benjamin P., and Lark, Richard M. “Efficient sampling for geostatistical surveys”. In: *European Journal of Soil Science* 70.5 (2019), pp. 975–989. https://doi.org/10.1111/ejss.12797.