

SPATIAL STATISTICS AND IMAGE MODELING

SPECIAL ISSUE: “III SEEMI’

RESEARCH PAPER

Optimization of spatial sample configurations using hybrid genetic algorithm and simulated annealing

LUCIANA PAGLIOSA CARVALHO GUEDES^{1,*}, PAULO JUSTINIANO RIBEIRO JR²,
SÔNIA MARIA DE STEFANO PIEDADE³ AND MIGUEL A. URIBE-OPAZO⁴

¹Centro de Ciências Exatas e Tecnológicas,

Universidade Estadual do Oeste do Paraná, Cascavel, Brazil

²L.E.G., Universidade Federal do Paraná, Curitiba, Brazil

³Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, Brazil

⁴Centro de Ciências Exatas e Tecnológicas,

Universidade Estadual do Oeste do Paraná, Cascavel, Brazil

(Received: 04 March 2011 · Accepted in final form: 20 June 2011)

Abstract

The accuracy of results obtained by geostatistical analysis, regarding spatial prediction, depends substantially on determining more efficient sampling schemes, with a reduced number of samples. This in the sense that the obtained results are similar to the actual results in the area, thus reducing operational costs. By using simulated data, the present work has laid plans for efficient spatial sampling in the prediction of variables with spatial dependence. The simulated annealing and hybrid genetic algorithm (GA) methods of optimization are used, considering the mean of the prediction variance as objective function. In addition, this work allows us to resize a sample configuration. This has already been applied to an experiment of precision agriculture in the cultivation of soybeans in Paraná, reducing by 50% its sample size and minimizing the efficiency losses of spatial prediction that this sample reduction may cause. The results for the simulations show that the optimized sample configurations produce lower estimates for the mean variance of the spatial prediction and better estimates for the characteristics related to spatial prediction in the studied area. Moreover, for the experiment considered in this study, the results show that the sample configuration reduce by hybrid GA shows a greater similarity with the initial sample configuration. Thus, the 50% reduction in the sample size by using the hybrid GA produces effective results for the classification of potassium (K) fertilizer in the area. Therefore, this reduced sample configuration may be used in future experiments for this area, reducing by 50% costs of chemical analysis of soil, without great loss of efficiency in the conclusions drawn by the spatial prediction.

Keywords: Geostatistics · Sampling spatial · Spatial variability.

Mathematics Subject Classification: Primary 62M30 · Secondary 68T20.

*Corresponding author. Luciana Pagliosa Carvalho Guedes. Rua Universitária 2069, Sala 65, CEP 85819-110, Cascavel, Paraná, Brasil. Email: luciana.pagliosa@hotmail.com

1. INTRODUCTION

The study of spatial variability structure of geo-referenced variables can be carried out using techniques of geostatistics, which allow, from a set of sample elements, understand the continuity of variables of interest across the area, thus demonstrating the spatial variation of the phenomenon by means of thematic maps of variability (see Diggle and Ribeiro Junior, 2007), which provide support information in the decision-making process for better management. However, restricted financial resources can cause great conflicts when having research objectives such as defining a sampling scheme used in the analysis of spatial variability, with the smallest size possible, to reduce operational costs when acquiring it, and maximize the results obtained for the spatial prediction; see Stingel (2007).

In one area, the entire population of points is not enumerable. Therefore, determining a sample set, that maximizes the efficiency of spatial prediction, from this population of points, makes it even more complex. A choice of methodology is to consider an initial sampling grid, with a large number of points, as a discretization of the area. Thus, defining the best sample configuration becomes a problem of choosing from an initial sampling grid, a reduced sample setting, minimizing the loss of information regarding spatial prediction, and resulting on cost reduction of the process of sample collection. This methodology is used for problems of avoiding expenses in environmental monitoring networks stations (see Bueso et al., 1999; Banjevic and Switzer, 2001; Nunes et al., 2005) and re-configuration of electric power networks; see Miasaki and Romero (2007).

Methods in the literature that regard resizing of sample configurations consider, as search criteria, a measure called objective function, which is minimized or maximized, and which summarizes the efficiency of the optimization of sample configuration for spatial prediction. The pursuit for the best sample configuration can be made by complete enumeration of all possible solutions (see Wu and Zidek, 1992; Le et al., 2003) or by sequential search methods (see Cressie, 1993; Boer et al., 2002; Royle, 2002), in which a sample is added or deleted in each step of the sequential process. However, when sample size is large, these methods are computationally exhaustive and impractical. Therefore, metaheuristics strategies, used in artificial intelligence, are an alternative in the pursuit for better sample configurations, since they easily adapt to different structures, are directed to the global optimization of the problem and promote search procedures that prevent a premature stop in excellent locations; see Ferri and Piccioni (1992), Groenigen and Stein (1998), Costa and Poppi (1999) and Zhu and Stein (2006).

In the past decade, several procedures categorized to be meta heuristics emerged, such as the simulated annealing algorithm and genetic algorithms (GA), including its hybrid version, which is an iterative method that pursuits optimal solution; see Banjevic and Switzer (2001); Nunes et al. (2005); Zhu and Stein (2006), Costa and Poppi (1999); Ruiz-Cárdenas et al. (2010), and Medeiros and Guimarães (2004); Chakrapani and Rajan (2008); Ruiz-Cárdenas et al. (2010).

The goal of this paper is to define optimal sample configurations for ten simulated data sets, using optimization algorithms called simulated annealing and hybrid GA. Furthermore, using this same methodology, we propose a hybrid GA to resize a sample set of 256 points obtained by stratified systematic unaligned sampling (see Wollenhaupt and Wolcowski, 1994; Souza et al., 1999), considering potassium content as variable. This new sample configuration, with a previously established 50% reduction in size, reduces losses that may be caused by its resizing in the accuracy of results regarding spatial prediction.

The paper is organized as follows. Section 2 describes in detail the simulation study. Section 3 presents the SA and GA algorithms for determining the best sample configuration. Section 4 compares the performance of these algorithms using simulated data. Section 5 applies these algorithms to the reduction of original sample configuration of the potassium values. Finally, a short discussion and concluding remarks are presented in Section 6.

2. SIMULATION STUDY

The discretization of the studied area is represented by an initial regular lattice, 20×20 , consisting of 400 points. The simulated results for the regionalized variable, located at this sample grid, are obtained through simulations. These represent the achievements of multivariate stochastic processes, assuming a set of stationary random variables, with no directional trend, isotropic and ceiling in the x and y coordinates equal to 1. This set of variables is represented by the spatial linear model

$$Z(x_i) = \mu(x_i) + S(x_i), \quad (1)$$

where $\mu(x_i) = \mu$ is a constant, x_i , for $i = 1, \dots, 400$, are the locations of the sampling grid points. In addition, $S(x_i)$ is the Gaussian process, spatially referenced, with variance equal to C_1 and semivariance function represented by the exponential model

$$\gamma(h) = \begin{cases} C_0 + C_1[1 - \exp(-\frac{3h}{a})], & \text{if } 0 < h < a; \\ C_0 + C_1, & \text{if } h \geq a, \end{cases} \quad (2)$$

where C_0 is called nugget, $C_0 + C_1$ is the value sill and a is the range.

Considering the spatial linear model described in Equation (1), the set of random variables have normal distribution given by $\mathbf{Z} \sim N_n(\mu\mathbf{1}, \mathbf{\Sigma})$, where $\mathbf{1}$ is a vector of ones and $\mathbf{\Sigma} = [(\sigma_{ij})]$, $\sigma_{ij} = \text{Cov}(x_i, x_j) = C(h)$, with $|x_i - x_j| = h$, for $i = j = 1, \dots, 400$, is the symmetric and positive covariance matrix, established by the following relationship with the model described in Equation (2) and given by $\gamma(h) = C(0) - C(h)$. Thus, according to Cressie (1993), each simulated data set is obtained by pre-establishing a mean vector $\mu\mathbf{1}$ and a covariance matrix $\mathbf{\Sigma}$ with parameters of its respective semivariance function model described in Equation (2), previously determined with: range equal to 0.6, nugget effect (NE) equal to 0, and value sill equal to 1. Then, the matrix \mathbf{L} is obtained using the Cholesky decomposition, given by $\mathbf{\Sigma} = \mathbf{L}\mathbf{L}^\top$. Thus, each simulated data set is obtained using the following relationship:

$$\mathbf{Z} = \mu\mathbf{1} + \mathbf{L}\boldsymbol{\epsilon}, \quad (3)$$

where $\boldsymbol{\epsilon}$ is a vector of uncorrelated random variables following a normal distribution with zero mean and variance equal to one.

3. STOCHASTIC SEARCH ALGORITHMS

Determining the best sample configuration consists on reducing the sample size of the initial grid. However, this reduction is obtained by iterative methods called simulated annealing and hybrid GA, which in this study consist of a combination of GA with simulated annealing method; see Jackson and Norgard (2008) and Zhou et al. (2010). The simulated annealing algorithm (see Kirkpatrick et al., 1983) makes an analogy to a natural phenomenon of annealing, which main objective is to increase the resistance of metals or glasses, by heating at elevated temperatures and subsequently cooling them slowly and gradually. The implementation of simulated annealing (SA) is considered for this work following the steps mentioned next.

STEP 0. The pursuit for the optimal solution is made from the iteration $i = 0$, considering an initial temperature value, for example $t_i = 8$, and a stopping criterion of 1000 iterations for the data from each simulation and 500 iterations for the potassium variable under study.

The choice for such stopping criterion is motivated by the observation that with this many iterations, the parameter of temperature had become small enough, changing very little current sample configuration and its corresponding objective function value;

STEP 1. An initial random sample S_i is selected, belonging to the original grid, reduced to a set with the predetermined size of 196 points for the simulated data sets and 128 points for potassium content, which respectively represent 49% and 50% of the initial grid;

STEP 2. The objective function for S_i is calculated as follows: for this sample configuration, an exponential model is adjusted using the method of maximum likelihood and the spatial prediction of the variable values in the study is performed, using the geostatistical interpolation called kriging. For the simulation, the prediction is made in relation to the values of the variable under this study, in the locations of the original grid. For the application example, the prediction performed referred to the points that are not sampled in a very thin grid in the area, with 17100 points. Next step is to calculate the objective function $f(\cdot)$ determined by the prediction variance mean ($\bar{\sigma}_0^2$), given the prediction previously described and expressed by

$$\bar{\sigma}_0^2 = \sum_{i=1}^n \lambda_i \gamma_{i0} + \alpha - \gamma_{ii}, \quad (4)$$

where γ_{i0} is the value of the semivariance function within the sampled point x_i and the non-sampled point x_0 , γ_{ii} is the value of the semivariance function within the sampled point x_i and itself and α is the Lagrange multiplier;

STEP 3. A neighbor sample configuration is obtained S_{i+1} , considering a disturbance in a random point of the previous sample configuration S_i . This disturbance is made by exchanging this point by another randomly chosen in its neighborhood;

STEP 4. The objective function for the sample configuration S_{i+1} , the same way as Step 2;

STEP 5. The variation that occurred from the variance mean values of the spatial prediction within both sample configurations is given by $\Delta = f(S_{i+1}) - f(S_i)$. The new solution S_{i+1} is accepted with probability and described as

$$P[\text{ accept } S_{i+1}] = \begin{cases} 1, & \text{if } \Delta \leq 0; \\ \exp\left(\frac{-\Delta}{t_i}\right), & \text{if } \Delta > 0, \end{cases}$$

where t_i is called current temperature, and its purpose is to control the quantity of solutions that is accepted in the optimization process. The process of simulated annealing is initiated with a high value for t_i , to allow a greater probability of acceptance of any new solution, including inefficient solutions, thus avoiding local minima;

STEP 6. The optimization process is ended when and if the stop criterion is met. If not, the value of t_i , is reduced through the relation $t_{i+1} = 0.85 t_i$. Thus, the quantity of inefficient solutions to be accept are reduced, admitting only S_{i+1} solutions with low value of the objective function, causing engagement with global optimum solution; see Nunes et al. (2005);

STEP 7. A $i = i + 1$ equation is determined and resumed to Step 3.

The structure of the hybrid GA, used in this work and for each data set, is described below.

STEP 0. Considering 60 iterations from $i = 0$, a initial population is determined with 100 sample configurations (called individuals), randomly chosen, all with the same previously established size;

STEP 1. For each individual is made estimation and spatial prediction of non-sampled localizations, as described in the process of simulated annealing, and the objective function is calculated $f(\cdot)$, represented by the variance mean of spatial prediction, shown in Equation (4);

STEP 2. The first selection is made, in which 50% of individuals with the highest values of objective function are eliminated from the population;

STEP 3. Two sample configurations (called parents) are randomly selected from the population of individuals, and then he crossing is made. These settings are coded into vectors, and its elements represent respectively the sample points of the original grid that either belong or not to the initial sample configuration. Then, another position in these two vectors is randomly selected, called the cutoff, so that both sample configurations are partitioned, and its shares are recombined, generating two new sample configurations; see Chakrapani and Rajan (2008);

STEP 4. The new sample configurations are evaluated by the objective function described in Equation (4), and the best sample configuration evaluated compared to the best of its parents, is chosen or not to the step of mutation, with the same probability used in simulated annealing;

STEP 5. The step of mutation is carried out with the individual selected in the previous step. This mutation consists in modifying the resulting configuration from the cross, by the process of simulated annealing;

STEP 6. The modified sample configuration is added to the population of individuals, replacing the individual with the highest value of mean variance of spatial prediction;

STEP 7. The optimization process is ended when and if the stop criterion is met. Otherwise, a $i = i + 1$ equation is determined and resumed to Step 3.

The optimized sampling grids are compared to samples found in the literature, such as simple random sampling (Ale), regular lattice (Sis), lattice plus close pairs (Pro) and lattice plus in-fill (Fill). For the comparison of these sampling schemes, measures of comparison associated with the prediction of points belonging to the original grid are considered. However, in simulated data sets, sampling schemes are compared by the mean variance of prediction, percentage and total sum of greater variable than the third quartile, and estimates of model parameters adjusted to the semivariance.

The obtained reduced sample configuration for potassium content is compared in relation to the mean variance of spatial prediction, percentage of predicted values that are inserted in each interval of potassium classification for the fertilization of potash fertilizers in soybean crops in Paraná; see Oleynik et al. (1998). For each reduced data set and for the real data set, maps that express the spatial variability of potassium are designed. In order for these maps to be compared it is necessary to quantify the areas and determine the correlation among them. This information can be obtained from the error matrix, considered as reference map, the map obtained by sample configuration with the totality of points and as a base map, the map that is obtained by reduced sample configuration. Each element in the matrix represents the area that belongs to the class i from the base map and class j from the reference map. The main diagonal (when $i = j$) represents cases when the area presented the same classification on both maps while the elements off the main diagonal represent the mistaken classifications.

From the error matrix, measures of accuracy called overall accuracy (OA) and Kappa statistics (\hat{K}) are calculated as (see Couto, 2003)

$$\text{OA} = \frac{\sum_{i=1}^r x_{ii}}{N} \quad \text{and} \quad \hat{K} = \frac{N \sum_{i=1}^r x_{ii} - \sum_{i=1}^r x_{i+} x_{+i}}{N^2 - \sum_{i=1}^r x_{i+} x_{+i}},$$

where x_{ii} is the total area that presented the same classification in both maps, r is the total number of classes, $x_{i+} = \sum_{j=1}^r x_{ij}$ is the amount of area of class i of the reference map, $x_{+i} = \sum_{u=1}^r x_{ui}$ is the amount of area of class i of the base map and N is the total area of the maps (pixel's). This allows us to measure the similarity between two maps.

To identify whether the results obtained by the optimized mesh sample are similar, we determinate the best configuration reduced sample from the optimization process performs on ten simulated data sets. The attainment of the simulated data sets, the implementation of algorithms and geostatistical analysis are carried out using the software R (www.R-project.org); see R Development Core Team (2009).

4. COMPARISON OF ALGORITHMS ON SIMULATED DATA

Through the plots of simulated values at their respective locations and from the reduced sample configuration points, shown in Figure 1, one may observe that the sampling schemes obtained by the optimization processes in Figure 1(e) and Figure 1(f), when compared with other, provide better coverage of the sampling points in the area.

Figure 2 shows, for the simulations, the scatter plots obtained, of the order of iterations versus the mean variance of spatial prediction, by the optimization processes of SA and hybrid GA. On the plots it is observed that the optimization processes sought the lowest value for the mean variance of spatial prediction, with about 600 iterations for the simulated annealing and 50 for the hybrid GA, and the search time for the best optimized sample configuration is one hour.

For the results regarding the estimates of the exponential model of semivariance function, for sampling schemes under study, shown in Table 1, we have that in all reduced samples the estimate values, on average, are very close to actual values, and the sampling grid optimized by simulated annealing showed, on average, a closer result to the value sill ($C_0 + C_1$), nugget (C_0) and NE coefficient ($\text{NE} = [C_0 / (C_0 + C_1)] \times 100$) of simulated data.

Table 1. Estimate mean of the exponential model parameters.

Parameter	Ale	Sis	Pro	Fill	SA	GA
Range (a)	0.77	0.77	0.82	0.83	0.82	0.80
Nugget (C_0)	0.010	0.017	0.016	0.009	0.006	0.010
NE coefficient	1.17	2.14	1.90	1.08	0.72	1.15
Value sill ($C_0 + C_1$)	0.82	0.83	0.86	0.86	0.87	0.85

Table 2 presents the mean results related to measures regarding spatial prediction. The lowest values for the mean variance of spatial prediction σ_0^2 are observed in the sampling schemes obtained by the optimization methods (SA and GA). The same results are shown in the boxplots in Figure 3(a), which highlight a low dispersion of results, especially in sampling schemes obtained by SA.

When analyzing Table 2 and the boxplots from Figures 3(b) and 3(c), it appears that only the sampling schemes obtained by the GA showed values for the predict value percentage above the 75th percentile (PVPP75), closer to the actual value, with lower dispersion and lower square sum of errors (SQEPVPP75). Moreover, although in general the mean values

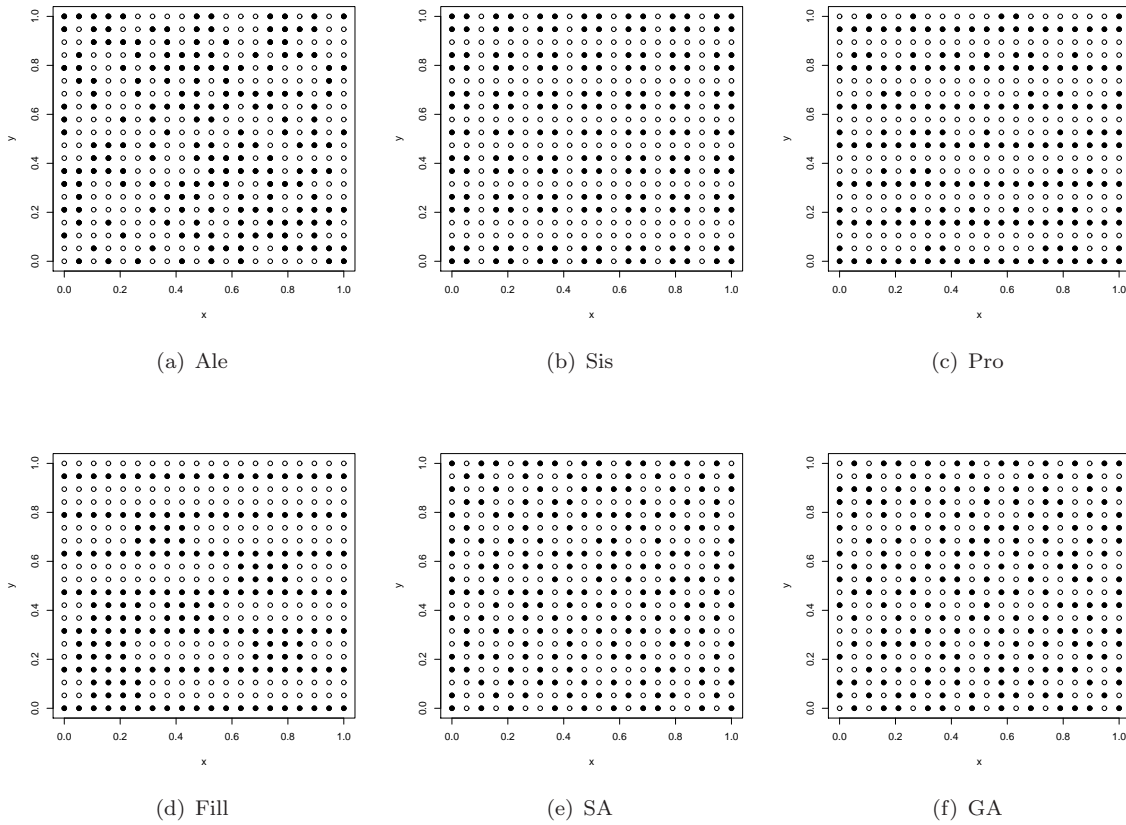


Figure 1. Location of the 400 sampling points (of one simulation) in the simulated area (o) and 196 points (•).

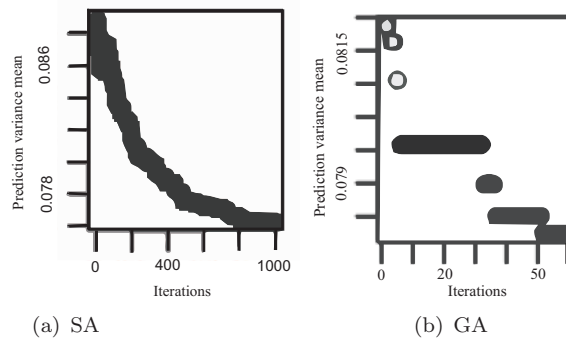


Figure 2. Scatter plots of the order of iterations versus the mean variance of the spatial prediction, of one simulation.

of the sum of predicted values above the 75th percentile (SVPP75) are different from true mean value (108.54), the lowest square sum of errors (SQE_{SVPP75}) are submitted by the sample configurations obtained by optimization methods (GA and GA), mainly by hybrid GA, as shown in Figures 3(d) and 3(e).

Table 2. Estimate mean of measures related to spatial prediction, considering the mean value that represents the sum of simulated values of the original grid equal to 108.54.

Parameter	Ale	Sis	Pro	Fill	SA	GA
σ_0^2	0.091	0.085	0.092	0.104	0.078	0.076
PVPP75	24.61	24.35	24.56	25.08	24.41	25.16
SQE_{PVPP75}	0.15	0.18	0.26	0.24	0.23	0.10
SVPP75	108.50	104.20	106.40	109.04	108.20	101.53
SQE_{SVPP75}	32×10^3	43×10^3	32×10^3	60×10^3	23×10^3	19×10^3

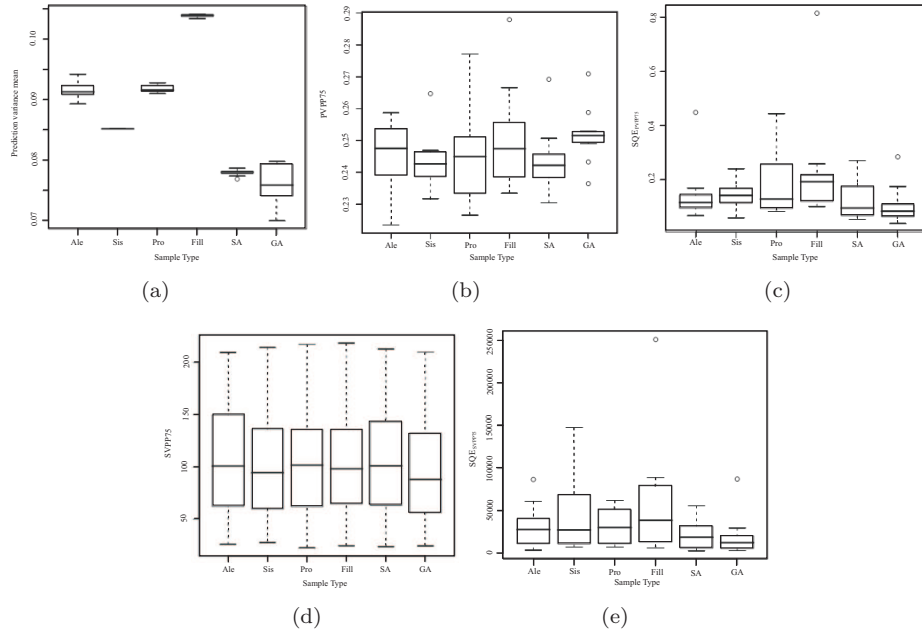


Figure 3. Boxplot of estimates: (a) variance mean of prediction (σ_0^2), (b) percentage of predicted values above the 75th percentile (PVPP75), (c) squared sum of errors of the percentage of predicted values above the 75th percentile (SQE_{PVPP75}), (d) sum of predicted values above the 75th percentile (SVPP75) and (e) squared sum of errors of the sum of predicted values above the 75th percentile (SQE_{SVPP75}).

5. ANALYSIS OF POTASSIUM CONTENT DATA RESULTS

Figure 4 shows plots of potassium values in their respective locations, where the area of each point is proportional to its value for the original sample configuration (see Figure 4a) and for sampling schemes with reduced number of sample elements: simple random (see Figure 4b) regular lattice (see Figure 4c), optimized by SA (see Figure 4d) and optimized by the GA (see Figure 4e). Regularity is observed to exist in the arrangement of points on the systematic sampling and on the optimized sampling (SA and GA). In which one can identify the formation of regions with similar values and excellent coverage of the samples in the area, because there are no regions without sample points.

Parameters of the exponential model of semivariance function and the mean variance of spatial prediction (σ_0^2) are estimated for each of the reduced sampled sets and for the actual sample set, and such parameters are shown in Table 3. Comparing these results, it is noted that in all reduced samples, the values of the NE coefficient are not similar to the estimate of the initial sample. Except for the sample optimized by simulated SA, estimates with range (a) and value sill ($C_0 + C_1$) are similar to the value obtained in the initial sampling. It is also noted that the optimized sampling yielded a lower value for the mean variance of spatial prediction, even in relation to the value obtained by initial sampling.

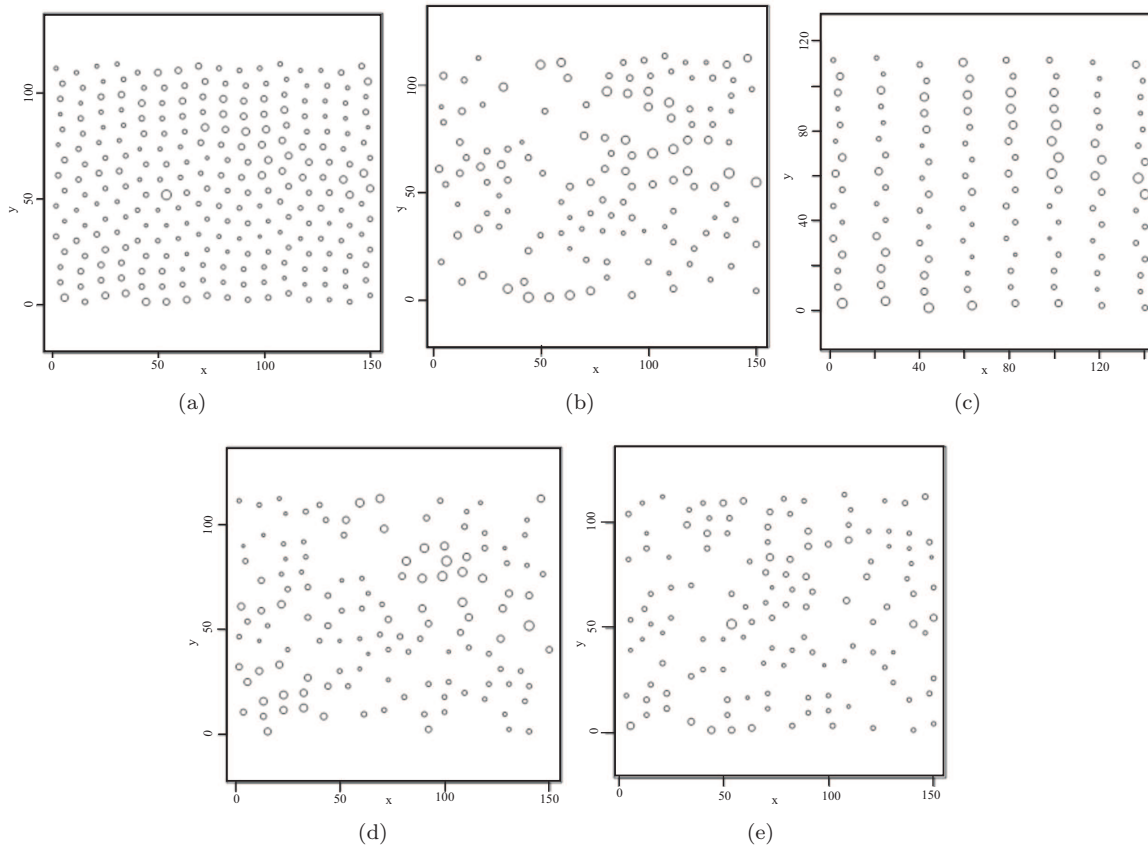


Figure 4. (a) Plot of potassium content values distributed on their respective locations where the area of each point is proportional to its value for the initial sample set and of samples: (b) Random simple, (c) regular lattice, (d) sampling optimized by simulated annealing algorithm and (e) sampling optimized by hybrid GA.

Table 3. Estimates of semivariance function model parameters for potassium content in the samples under study and mean variance of spatial prediction.

Statistic	Initial set	Ale	Sis	SA	GA
Range (a)	56.41	47.90	52.36	101.35	45.42
Nugget (C_0)	0.0007	0.0000	0.0006	0.0000	0.0037
NE coefficient	21.52	0.00	6.96	0.00	48.68
Value sill ($C_0 + C_1$)	0.0034	0.0065	0.0080	0.0047	0.0076
σ_0^2	0.0034	0.024	0.032	0.0008	0.0005

On the same spatial prediction grid, predicted values of reduced samples and of the initial grid are categorized, in accordance with the recommendation of the levels of potassium content for the fertilization of soybean crops in the State of Paraná (see Oleynik et al., 1998). Table 4 shows the amounts and percentages of predicted values, in the reduced sampling and in the initial sampling, which are in each of the intervals of classification of manure for potassium content. In general, it is observed that the sample reduced by the optimization method GA shows percentage and values of the amount of predicted values in each classification more similar to the values obtained by initial sampling. The accuracy indexes called overall accuracy (OA) and kappa statistics (\hat{K}), described in Table 5, show that all sampling schemes have high accuracy, since $OA \geq 0.85$ (see Anderson et al., 1976) and $\hat{K} \geq 0.80$ (see Krippendorf, 1980), and the highest values of these indexes are submitted by sample configuration optimized by the hybrid GA.

Table 4. Estimates of number and percentage of predicted values inserted in each interval of the classification of potassium for fertilization in the State of Paraná (see Oleynik et al., 1998).

Statistic	Classification of Emater	Initial set	Ale	Sis	SA	GA
Number of predicted values	$K \leq 0.1$	0	0	0	0	0
	$0.1 < K \leq 0.2$	0	0	21	0	0
	$0.2 < K \leq 0.3$	6830	6011	7555	5541	6916
	$K > 0.3$	10270	11089	9524	11559	10184
	Sum	17100	17100	17100	17100	17100
Percentage of predicted values	$K \leq 0.1$	0.00	0.00	0.00	0.00	0.00
	$0.1 < K \leq 0.2$	0.00	0.00	0.12	0.00	0.00
	$0.2 < K \leq 0.3$	39.94	35.15	44.18	32.40	40.44
	$K > 0.3$	60.06	64.85	55.70	67.60	59.56
	Sum	100.00	100.00	100.00	100.00	100.00

Table 5. Measures of accuracy considering the variability map which is designed from the sample of 256 points, used as base map; And the variability map which is designed from the sampling configurations with 128 points, used as the reference map.

Accuracy measures	Ale	Sis	SA	GA
Overall accuracy	0.952	0.956	0.925	0.995
Kappa	0.900	0.917	0.838	0.900

Figure 5 shows maps of spatial variability of potassium content, taking into account the rating for fertilization in Paraná, designed by using the grid with all data (see Figure 5a) and the reduced sampling schemes from Figure 5(b) to Figure 5(e). It is observed that the samples reduced by random simple (see Figure 5b) and optimized by simulated annealing (see Figure 5d) overestimate some regions predicted by the map of spatial variability of potassium content obtained using data of sampling that contain the entire data (see Figure 5a) not indicating the suggested classification. Variability maps which are designed through the regular lattice (see Figure 5c) and sampling scheme optimized by hybrid GA (see Figure 5e) had the highest similarity and an underestimation of some predicted regions, in relation to the variability map designed by using the grid containing all data, especially with closer approximation of each region, so understated.

6. CONCLUSIONS

The analysis that we have been carried out for the simulations has shown that the samples obtained by the optimization methods (SA and GA), mainly by the hybrid genetic algorithm, provided estimates that were closer in similarity to the actual value, regarding the measures related to spatial prediction in non-sampled locations, and the lowest values of mean variance of spatial prediction. Give so that samples obtained by the optimization methods proved to be more reliable for spatial prediction. Analyses on reduced samples with 128 samples, for the potassium content variable, showed that the sampling optimized by the hybrid genetic algorithm had lower estimates for the mean variance of spatial prediction, even in relation to the value of this estimate obtained by means of the initial data set with 256 sampling units. In addition, in sampling optimized by hybrid genetic algorithm, estimates of the number and of the percentage of predicted values belonging to each interval of classification, used to determine the amount of product to be applied at each sampling location, were similar to the values obtained in the initial data set. These results indicate that sampling optimized by hybrid genetic algorithm, with 50% of samples, can be used in place of the sample set that contains all data in future experiments in this area, reducing half of the expenses spent with soil chemical analysis with little influence on the final results.

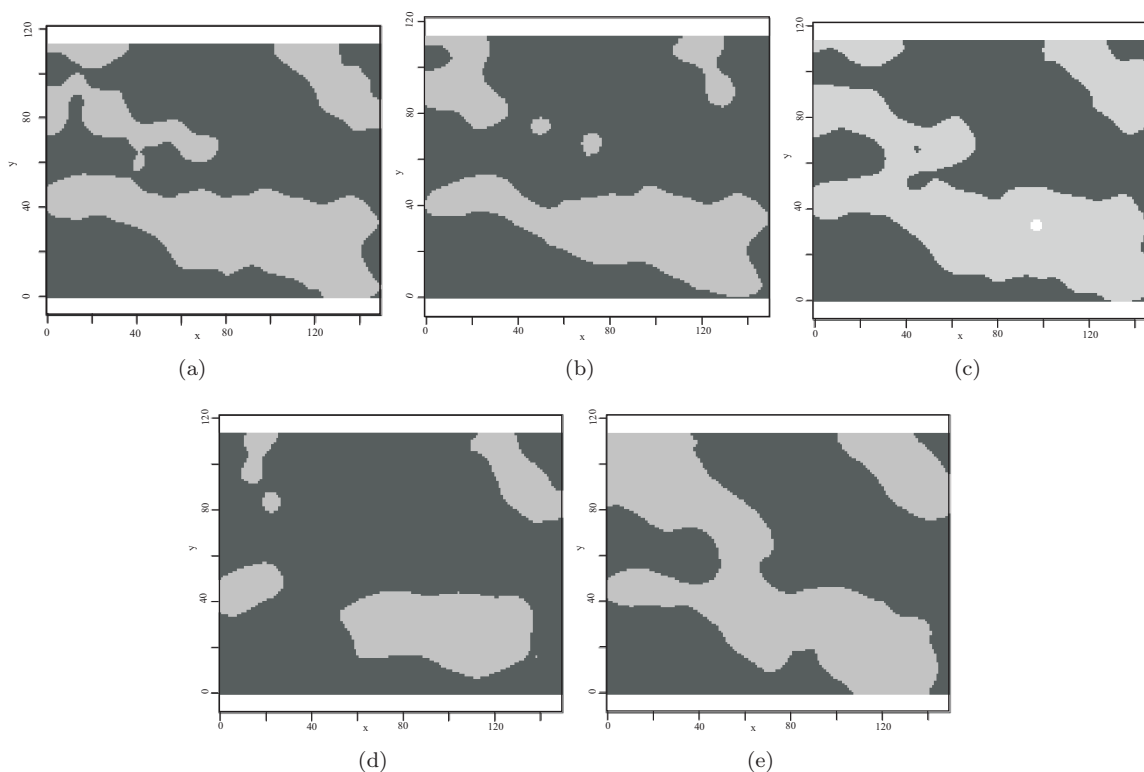


Figure 5. Map which represents the spatial variability of K ($\text{cmol}_c/\text{dm}^3$) in the area, according to the classification for fertilization in soybean crops in Paraná, obtained from the following samplings: (a) initial set, (b) random simple, (c) regular lattice, (d) optimized by SA and (e) optimized by hybrid GA.

7. ACKNOWLEDGEMENTS

The authors to thank to CAPES, Coordination of Improvement of Higher Education Personnel, through the Program of Social Demand and the National Council for Scientific and Technological Development (CNPq) for financial support.

REFERENCES

- Anderson, J.F., Hardy, E.E., Roach, J.T., Witmer, R.E., 1976. A Land Use and Land Cover Classification System for use with Remote Sensor Data. Geologic Survey, Washington.
- Banjevic, M., Switzer, P., 2001. Optimal network designs in spatial statistics. In Royal Statistical Society Conference on Spatial Modelling, Glasgow, pp. 1-15.
- Boer, E.P.J., Dekkers, A.L.M., Stein, A., 2002. Optimization of a monitoring network for sulfur dioxide. *Journal of Environmental Quality*, 31, 121-128.
- Bueso, M.C., Angulo, J.M., Cruz-Sanjulián, J., García-Aróstegui, J.L., 1999. Optimal spatial sampling design in a multivariate framework. *Mathematical Geology*, 31, 507-525.
- Chakrapani, Y., Rajan, K.S., 2008. Hybrid genetic-simulated annealing approach for fractal image compression. *International Journal of Computational Intelligence*, 4, 308-313.
- Costa Filho, P.A.C., Poppi, R.J., 1999. Algoritmo genético em química. *Química Nova*, 27, 405-411.
- Couto, P., 2003. Assessing the accuracy of spatial simulation models. *Ecological Modelling*, 167, 181-198.
- Cressie, N.A.C., 1993. *Statistics for Spatial Data*. Wiley, New York.
- Diggle, P.J., Ribeiro Junior, P.J., 2007. *Model-based Geostatistics*. Springer, New York.

- Ferri, M., Piccioni, M., 1992. Optimal selection of statistical units: An approach via simulated annealing. *Computational Statistics and Data Analysis*, 13, 47-61.
- Groenigen, J.W.V., Stein, A., 1998. Constrained optimization of spatial sampling using continuous simulated annealing. *Journal of Environmental Quality*, 27, 1078-1086.
- Jackson, W.C., Norgard, J.D., 2008. A hybrid genetic algorithm with Boltzmann convergence properties. *Journal of Optimization Theory and Applications*, 136, 431-443.
- Kirpatrick, S.C., Gelatt, C. D., Vecchi, J.M.P., 1983. Optimization by simulated annealing. *Science*, 220, 671-680.
- Kriependorf, K., 1980. *Content Analysis: An Introduction to its Methodology*. Sage Publications, Beverly Hills.
- Le, N.D., Sun, L., Zidek, J.V., 2003. Designing networks for monitoring multivariate environmental fields using data with monotone pattern. Technical report. Statistical and Applied Mathematical Sciences Institute.
- Medeiros, F.L.L., Guimarães, L.N.F., 2004. Hybrid genetic algorithm applied to the steady state determination problem of nonlinear dynamic systems. *IEEE Computer Society Press*, 1, 1-6.
- Miasaki, C.T., Romero, R., 2007. Um algoritmo genético especializado aplicado ao planejamento da expansão do sistema de transmissão com alocação de dispositivos de compensação série. *Sba Controle e Automação*, 18, 210-222.
- Nunes, L.M.A.F.F., Caeiro, S., Cunha, M.C., Ribeiro, L., 2005. Optimal estuarine sediment monitoring network design with simulated annealing. *Journal of Environmental Management*, 78, 294-304.
- Oleynik, J., Braganolo, N., Bublitz, U., Silva, J.C.C., 1998. Tabela para transformação de resultados analíticos e interpretação dos resultados, Curitiba, Brazil.
- R Development Core Team, 2009. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
- Royle, J.A., 2002. Exchange algorithms for constructing large spatial designs. *Journal of Statistical Planning and Inference*, 100, 121-134.
- Ruiz-Cárdenas, R., Ferreira, M.A.R., Schmidt, A.M., 2010. Stochastic search algorithms for optimal design of monitoring networks. *Environmetrics*, 21, 102-112.
- Souza, E.G., Ribeiro, S.R.A., Uribe-Opazo, M.A., Vilas Boas, M., Silva, M.S., Johann, J.A., Molin, J.P., Nóbrega, L.H.P., Oliveira, E.F., Carraro, I.A., 1999. Metodologia para análise da variabilidade espacial dos atributos do solo e da produtividade em uma área piloto de agricultura de precisão. *Congresso Brasileiro de Engenharia Agrícola, Poço de Caldas*, pp. 109-111.
- Stingel, E., 2007. Uso da geoestatística na avaliação da distribuição de Mahanarva fimbriolata em cana-de-açúcar. *Bragantia*, 66, 449-455.
- Wu, S., Zidek, J.V., 1992. An entropy-based analysis of data from selected NADP/NTN network sites for 1983-1986. *Atmospheric Environment*, 26A, 2089-2103.
- Zhou, Y.Y., Rao, Y.Q., Chang, C.Y., Gao, L., 2010. Hybrid genetic algorithm with simulated annealing based on best-fit strategy for rectangular packing problem. *Advanced Materials Research*, 379, 118-120.
- Zhu, Z., Stein, M.L., 2006. Spatial sampling design for prediction with estimated parameters. *Journal of Agricultural, Biological and Environmental Statistics*, 11, 24-44.
- Wollenhaupt, N.C., Wolcowski, R.P., 1994. Grid soil sampling. *Better Crops with Plant Food*, 78, 6-9.