Universidade de São Paulo
Escola Superior de Agricultura “Luiz de Queiroz”

Otimização de amostragem espacial

Luciana Pagliosa Carvalho Guedes

Tese apresentada para obtenção do título de Doutor em
Agronomia. Área de concentração: Estatística e Experi-
mentação Agronômica

Piracicaba
2008

Luciana Pagliosa Carvalho Guedes

Licenciada em Matemática

Otimização de amostragem espacial

Orientadora:

Prof^ª Dr^ª **SÔNIA MARIA DE STEFANO PIEDADE**

Tese apresentada para obtenção do título de Doutor
em Agronomia. Área de concentração: Estatística e
Experimentação Agronômica

Piracicaba

2008

Dados Internacionais de Catalogação na Publicação (CIP)
DIVISÃO DE BIBLIOTECA E DOCUMENTAÇÃO - ESALQ/USP

Guedes, Luciana Pagliosa Carvalho

Otimização de amostragem espacial / Luciana Pagliosa Carvalho. - - Piracicaba, 2008.
143 p. : il.

Tese (Doutorado) - - Escola Superior de Agricultura Luiz de Queiroz, 2008.
Bibliografia.

1. Algoritmos genéticos 2. Amostragem 3. Estatística de processos estocásticos
I. Título

CDD 519.5

“Permitida a cópia total ou parcial deste documento, desde que citada a fonte – O autor”

Dedicatória

A

Deus

Força maior de todo ser humano, que guia os meus passos na busca pelos meus sonhos.

Ao meu marido **Marcelo Guedes**, o amor, a compreensão, o apoio e, principalmente, por se manter sempre forte e zelar pela saúde e bem-estar da nossa família, nos momentos mais difíceis.

Ao meu filho **Pedro**, que alegra minha vida dando-me motivos e estímulos para progredir.

A minha família pelo constante apoio, incentivo e auxílio.

AGRADECIMENTOS

Ao professor Dr. Paulo Justiniano Ribeiro Junior, pela confiança depositada, pelas brilhantes idéias e pela sua constante orientação à melhoria desse trabalho.

A minha orientadora Dra. Sônia Maria De Stefano Piedade pelo apoio incondicional na elaboração desse trabalho.

Aos professores Dr. Décio Barbin, Dra. Clarice Demétrio e Dra. Roseli Aparecida Leandro pelos ensinamentos repassados durante o curso de doutorado e pelas inúmeras vezes que fizeram excelentes sugestões.

Ao professor Dr. Miguel Angel Uribe Opazo pelo incentivo, motivação e proveitosas discussões.

Aos meus colegas do mestrado e doutorado em Estatística e Experimentação Agronômica da ESALQ pela amizade e troca de idéias.

As secretárias Solange e Luciane pela ajuda incondicional e constante auxílio.

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) e à Coordenadoria para o Aperfeiçoamento de Pessoal de Nível Superior (CAPES) o apoio financeiro em forma de bolsa.

SUMÁRIO

RESUMO	7
ABSTRACT	9
LISTA DE FIGURAS	10
LISTA DE TABELAS	17
LISTA DE QUADROS	18
1 INTRODUÇÃO	19
2 DESENVOLVIMENTO	21
2.1 Revisão de Literatura	21
2.1.1 Agricultura de precisão e variabilidade espacial	21
2.1.2 Geoestatística	22
2.1.2.1 Variável regionalizada	22
2.1.2.2 Semivariância	23
2.1.3 Modelos teóricos para o ajuste ao semivariograma	25
2.1.3.1 Método de máxima verossimilhança e máxima verossimilhança restrita	28
2.1.4 <i>Krigagem</i>	30
2.1.4.1 <i>Krigagem</i> ordinária	30
2.1.4.2 <i>Krigagem</i> universal	33
2.1.4.3 <i>Krigagem</i> indicatriz	34
2.1.5 Esquemas de amostragem espacial	35
2.1.6 Otimização de amostras espaciais	40
2.1.6.1 Critério de otimização	40
2.1.6.2 Seleção de algoritmos no processo de otimização: método de seleção sequencial	42
2.1.6.3 Seleção de algoritmos no processo de otimização: recozimento simulado ou têmpera simulada	43
2.1.6.3.1 Funcionamento do recozimento simulado	44
2.1.6.4 Seleção de algoritmos no processo de otimização: algoritmo genético	45
2.1.6.4.1 Introdução aos algoritmos genéticos	45
2.1.6.4.2 Codificação dos indivíduos da população	47
2.1.6.4.3 Funcionamento de um algoritmo genético	47
2.1.6.4.4 Cruzamento ou recombinação genética	48
2.1.6.4.5 Mutação	51
2.1.6.5 Critério de otimização na predição espacial	51
2.1.6.6 Critério baseado na variância da <i>krigagem</i>	51
2.1.6.6.1 Critério baseado na máxima entropia	54

2.1.6.7 Critério de otimização baseado em medidas geométricas	56
2.1.6.8 Critério de otimização na estimação dos parâmetros do modelo ajustado à função semivariância	57
2.2 Material e métodos	58
2.2.1 Simulações	58
2.2.2 Área experimental e variáveis em estudo	63
3 RESULTADOS E DISCUSSÕES	64
3.1 Resultados das simulações	64
3.2 Resultados da análise do teor de potássio	72
3.2.1 Análise espacial	72
3.2.2 Comparação entre as amostragens com tamanho reduzido	74
4 CONCLUSÕES	78
REFERÊNCIAS	79
ANEXOS	83

RESUMO

Otimização de amostragem espacial

O objetivo desse trabalho foi estabelecer planos de amostragem com redução no tamanho amostral, a partir de conjuntos de dados com dependência espacial, que fossem eficientes na predição de localizações não amostradas e que gerassem estimativas eficientes de características relacionadas à predição espacial. Esses planos amostrais reduzidos foram obtidos por processos de otimização denominados recozimento simulado e algoritmo genético híbrido, considerando a média da variância da predição espacial, obtida pelo método de interpolação chamado *krigagem*, como função objetivo minimizada. Para isso, utilizaram-se conjuntos de dados simulados, com diferentes valores de alcance e efeito pepita, cujo intuito foi identificar a influência que esses parâmetros exercem na escolha da configuração amostral otimizada. Para cada conjunto de dados simulados, foram obtidas amostras pelos processos de otimização e seus resultados foram comparados aos esquemas de amostragem: aleatório, sistemático, sistemático centrado adicionado de delineamentos menores e sistemático centrado adicionado de pontos próximos. Os resultados mostraram que os planos de amostragem otimizados, principalmente os planos obtidos pelo algoritmo genético híbrido, produziram menores estimativas para a média da variância da *krigagem* e melhores estimativas para a porcentagem e soma de valores preditos acima do terceiro quartil e do percentil 90, que são características relacionadas à predição espacial. Observou-se, também, que o aumento do tamanho amostral produziu melhores estimativas para todos os resultados analisados e, independente do valor de alcance e efeito pepita, a amostragem otimizada pelo algoritmo genético híbrido produziu melhores resultados. Além disso, obtiveram-se conjuntos amostrais reduzidos de 128 parcelas pelo algoritmo genético híbrido, pelo processo de recozimento simulado e pelas amostragens aleatória e sistemática, para a propriedade química teor de potássio pertencente ao conjunto de dados, com 256 parcelas, de um experimento de agricultura de precisão em uma área experimental. Por intermédio dos dados resultantes dessas amostragens, realizou-se uma análise geoestatística para identificar o comportamento de dependência espacial da variável potássio na área e foram feitas predições espaciais do potássio em localizações não amostradas nessa mesma área. Em todos os esquemas de amostragem utilizados, os valores preditos foram classificados segundo o critério de adubação do potássio no Paraná em culturas de soja (EMATER - Empresa Paranaense de Assistência Técnica e Extensão Rural, 1998). Esses resultados foram comparados com a análise realizada no conjunto de dados inicial e observou-se uma maior similaridade desses resultados com os obtidos pela análise realizada através dos dados da amostragem obtida pelo algoritmo genético híbrido. Assim, tiveram-se evidências de que a redução em 50% do tamanho amostral do conjunto de dados da variável potássio, utilizando nessa redução uma amostragem obtida pelo algoritmo genético híbrido, produziu resultados eficientes para a classificação de adubação de potássio na área em estudo, reduzindo em 50% os custos

com análise química do solo, sem grande perda de eficiência nas conclusões obtidas pela predição espacial.

Palavras-chave: Geoestatística; Amostragem Espacial; Algoritmo Genético.

ABSTRACT

Optimization of sampling space

The aim of this work was to establish plans for sampling with reduced in the sample size, from sets of dependent spatial data, and they are efficient in terms of prediction of the nonsampled observations and prediction of linear targets. These plans were obtained by sampling reduced processes optimization of the algorithm called simulated annealing and hybrid genetic algorithm, considering the average kriging variance as objective function to be minimised. Therefore, it was used simulated data sets, With different values of range and nugget effect, whose aim was to identify the influence that these exercise parameters in choosing the sample configuration optimized. For each set of data simulated samples were obtained through the optimization process and its results were compared to sampling schemes: random, systematic, lattice plus in fill and lattice plus close pairs. The results show that the sampling plans optimized, especially the plans obtained by hybrid genetic algorithm, produced lower estimates for the average kriging variance and best estimates for the percentage and amount of predicted values above the third quartile and the 90 percentile, which are characteristics related to spatial prediction. It was also observed that the increase of sample size produces best estimate for all results analyzed and independent of the value range and nugget effect, sampling optimized by the hybrid genetic algorithm produced better results. In addition, sets up sampling reduced of 128 samples by sampling schemes: hybrid genetic algorithm, simulated annealing, random and systematic, for the property belonging to the chemical potassium set data, with 256 samples, an experiment of precision agriculture, in an experimental area. Through these sampling data, an analysis was carried out geostatistics to identify the behavior of spatial dependence of the variable potassium in the area under study and predictions were made of potassium space in locations not sampled that same area. In all sampling schemes used, the predicted values were classified at the discretion of the potassium fertilization in the cultivation of soybeans in Parana (EMATER-PARANA, 1998). These results were compared with the analysis in the initial set of data and there was a greater similarity of these results with those obtained by the analysis performed by data obtained by the sampling hybrid genetic algorithm. So there has been evidence that the reduction by 50% of the sample size of the data set of variable potassium, using this reduction a sample obtained by the hybrid genetic algorithm, produced effective results for classification of potassium fertilizer in the area under study, reducing by 50% the costs with chemical analysis of soil, without much loss of efficiency in the conclusions obtained by predicting space.

Keywords: Geostatistics; Design geostatistical; Genetic algorithm.

LISTA DE FIGURAS

Figura 1 - Gráfico semivariograma experimental	24
Figura 2 - Modelo teórico exponencial ajustado à função semivariância, em que a representa o alcance, C_0 é o efeito pepita e $C_0 + C_1$ é o patamar	25
Figura 3 - Modelo teórico gaussiano	27
Figura 4 - Modelo teórico esférico	27
Figura 5 - Amostragem Aleatória: (a) desalinhada, (b) alinhada em uma direção e (c) alinhada em ambas as direções	37
Figura 6 - Amostragem Sistemática Centrada	38
Figura 7 - Amostragem Sistemática: (a) desalinhada em ambas as direções, (b) alinhada em ambas as direções e (c) alinhada em uma direção	39
Figura 8 - (a) <i>Lattice plus close pairs</i> e (b) <i>Lattice plus in-fill</i>	40
Figura 9 - Boxplot das estimativas da média da variância de predição para as amostragens com tamanho amostral igual a 40, com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1)	67
Figura 10 - Boxplot das estimativas da média da variância de predição para as amostragens com tamanho amostral igual a 196, com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1)	68
Figura 11 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens de tamanho 40, obtidas pelo algoritmo de recozimento simulado, nas simulações dos modelos com alcance igual a 0,9, efeito pepita igual a 0 e patamar igual a 1	69
Figura 12 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens de tamanho 40, obtidas pelo algoritmo genético híbrido nas simulações dos modelos com alcance igual a 0,9, efeito pepita igual a 0 e patamar igual a 1	70
Figura 13 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens de tamanho 196, obtidas pelo algoritmo de recozimento simulado nas simulações dos modelos com alcance igual a 0,9, efeito pepita igual a 0 e patamar igual a 1	71

Figura 14 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens de tamanho 196, obtidas pelo algoritmo genético híbrido nas simulações dos modelos com alcance igual a 0,9, efeito pepita igual a 0 e patamar igual a 1	72
Figura 15 - (a) Boxplot do teor de potássio, (b) Histograma do potássio e (c) Gráfico para estimação do parâmetro da transformação Box Cox	73
Figura 16 - (a) Postplot de K, (b) Envelope do semivariograma omnidirecional de K, (c) Semivariogramas Direcionais de K nas direções 0°, 45°, 90° e 135° e (d) Mapa da variabilidade espacial do teor de potássio, segundo a classificação para adubação de áreas com cultivo de soja no Paraná	75
Figura 17 - Postplot de K nas amostragens: (a) Aleatória, (b) Sistemática, (c) Amostragem otimizada pelo algoritmo de recozimento simulado e (d) Amostragem otimizada pelo algoritmo genético híbrido	76
Figura 18 - Mapa que representa a variabilidade espacial de K na área, segundo a classificação para adubação de áreas com cultivo de soja no Paraná, obtidos pelas amostragens: (a) Aleatória, (b) Sistemática, (c) Amostragem otimizada pelo algoritmo de recozimento simulado e (d) Amostragem otimizada pelo algoritmo genético híbrido	77
Figura 19 - Gráfico da localização dos 400 pontos amostrais na área simulada (○) e localizações dos 40 pontos (●) para as amostragens: (a) aleatória, (b) sistemática, (c) sistemática centrada adicionada de pontos próximos e (d) sistemática centrada adicionada de amostragens sistemáticas centradas menores	88
Figura 20 - Gráfico da localização dos 400 pontos amostrais na área simulada (○) e localizações dos 196 pontos (●) para as amostragens: (a) aleatória, (b) sistemática, (c) sistemática centrada adicionada de pontos próximos e (d) sistemática centrada adicionada de amostragens sistemáticas centradas menores	89
Figura 21 - Gráfico da localização dos 400 pontos amostrais na área simulada (○) e localizações dos 40 pontos (●) para uma das amostragens obtidas nas diferentes simulações pelo algoritmo de recozimento simulado com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1)	90
Figura 22 - Gráfico da localização dos 400 pontos amostrais na área simulada (○) e localizações dos 40 pontos (●) para uma das amostragens obtidas nas diferentes simulações pelo algoritmo genético híbrido com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1)	91

- Figura 23 - Gráfico da localização dos 400 pontos amostrais na área simulada ($\circ$) e localizações dos 196 pontos ($\bullet$) para uma das amostragens obtidas nas diferentes simulações pelo algoritmo de recozimento simulado com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1) 92
- Figura 24 - Gráfico da localização dos 400 pontos amostrais na área simulada ($\circ$) e localizações dos 196 pontos ($\bullet$) para uma das amostragens obtidas nas diferentes simulações pelo pelo algoritmo genético híbrido com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1) 93
- Figura 25 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 40 pontos amostrais pelo algoritmo de recozimento simulado nas simulações dos modelos com alcance igual a 0,6, efeito pepita igual a 0 e patamar igual a 1 94
- Figura 26 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 40 pontos amostrais pelo algoritmo de recozimento simulado nas simulações dos modelos com alcance igual a 0,3, efeito pepita igual a 0 e patamar igual a 1 95
- Figura 27 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 40 pontos amostrais pelo algoritmo de recozimento simulado nas simulações dos modelos com alcance igual a 0,6, efeito pepita igual a 0,2 e patamar igual a 1 96
- Figura 28 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 40 pontos amostrais pelo algoritmo de recozimento simulado nas simulações dos modelos com alcance igual a 0,6, efeito pepita igual a 0,5 e patamar igual a 1 97
- Figura 29 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 40 pontos amostrais pelo algoritmo genético híbrido nas simulações dos modelos com alcance igual a 0,6, efeito pepita igual a 0 e patamar igual a 1 98
- Figura 30 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 40 pontos amostrais pelo algoritmo algoritmo genético híbrido nas simulações dos modelos com alcance igual a 0,3, efeito pepita igual a 0 e patamar igual a 1 99

Figura 31 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 40 pontos amostrais pelo algoritmo genético híbrido nas simulações dos modelos com alcance igual a 0,6, efeito pepita igual a 0,2 e patamar igual a 1	100
Figura 32 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 40 pontos amostrais pelo algoritmo genético híbrido nas simulações dos modelos com alcance igual a 0,6, efeito pepita igual a 0,5 e patamar igual a 1	101
Figura 33 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 196 pontos amostrais pelo algoritmo de recozimento simulado nas simulações dos modelos com alcance igual a 0,6, efeito pepita igual a 0 e patamar igual a 1	102
Figura 34 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 196 pontos amostrais pelo algoritmo de recozimento simulado nas simulações dos modelos com alcance igual a 0,3, efeito pepita igual a 0 e patamar igual a 1	103
Figura 35 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 196 pontos amostrais pelo algoritmo de recozimento simulado nas simulações dos modelos com alcance igual a 0,6, efeito pepita igual a 0,2 e patamar igual a 1	104
Figura 36 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 196 pontos amostrais pelo algoritmo de recozimento simulado nas simulações dos modelos com alcance igual a 0,6, efeito pepita igual a 0,5 e patamar igual a 1	105
Figura 37 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 196 pontos amostrais pelo algoritmo genético híbrido nas simulações dos modelos com alcance igual a 0,6, efeito pepita igual a 0 e patamar igual a 1	106
Figura 38 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 196 pontos amostrais pelo algoritmo genético híbrido nas simulações dos modelos com alcance igual a 0,3, efeito pepita igual a 0 e patamar igual a 1	107

- Figura 39 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 196 pontos amostrais pelo algoritmo genético híbrido nas simulações dos modelos com alcance igual a 0,6, efeito pepita igual a 0,2 e patamar igual a 1 108
- Figura 40 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 196 pontos amostrais pelo algoritmo genético híbrido nas simulações dos modelos com alcance igual a 0,6, efeito pepita igual a 0,5 e patamar igual a 1 109
- Figura 41 - Boxplot das estimativas da porcentagem de valores preditos, que estão acima do terceiro quartil para as amostragens com tamanho amostral igual a 40 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1) 110
- Figura 42 - Boxplot das estimativas da porcentagem de valores preditos, que estão acima do terceiro quartil para as amostragens com tamanho amostral igual a 196 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1) 111
- Figura 43 - Boxplot da soma quadrada dos erros das estimativas de porcentagem de valores preditos, que estão acima do terceiro quartil para as amostragens com tamanho amostral igual a 40 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1) 112
- Figura 44 - Boxplot da soma quadrada dos erros das estimativas de porcentagem de valores preditos, que estão acima do terceiro quartil para as amostragens com tamanho amostral igual a 196 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1) 113
- Figura 45 - Boxplot das estimativas da porcentagem de valores preditos, que estão acima do percentil 90 para as amostragens com tamanho amostral igual a 40 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1) 114
- Figura 46 - Boxplot das estimativas da porcentagem de valores preditos, que estão acima do percentil 90 para as amostragens com tamanho amostral igual a 196 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1) 115

- Figura 47 - Boxplot da soma quadrada dos erros das estimativas de porcentagem de valores preditos, que estão acima do percentil 90 para as amostragens com tamanho amostral igual a 40 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1) 116
- Figura 48 - Boxplot da soma quadrada dos erros das estimativas de porcentagem de valores preditos, que estão acima do percentil 90 para as amostragens com tamanho amostral igual a 196 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1) 117
- Figura 49 - Boxplot das estimativas da soma dos valores preditos, que estão acima do terceiro quartil para as amostragens com tamanho amostral igual a 40 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1) 118
- Figura 50 - Boxplot das estimativas da soma dos valores preditos, que estão acima do terceiro quartil para as amostragens com tamanho amostral igual a 196 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1) 119
- Figura 51 - Boxplot da soma quadrada dos erros das estimativas da soma dos valores preditos, que estão acima do terceiro quartil para as amostragens com tamanho amostral igual a 40 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1) 120
- Figura 52 - Boxplot da soma quadrada dos erros das estimativas da soma dos valores preditos, que estão acima do terceiro quartil para as amostragens com tamanho amostral igual a 196 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1) 121
- Figura 53 - Boxplot das estimativas da soma dos valores preditos, que estão acima do percentil 90 para as amostragens, com tamanho amostral igual a 40 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1) 122
- Figura 54 - Boxplot das estimativas da soma dos valores preditos, que estão acima do percentil 90, para as amostragens com tamanho amostral igual a 196 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1) 123

Figura 55 - Boxplot da soma quadrada dos erros das estimativas da soma dos valores preditos, que estão acima do percentil 90, para as amostragens com tamanho amostral igual a 40 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1) 124

Figura 56 - Boxplot da soma quadrada dos erros das estimativas da soma dos valores preditos, que estão acima do percentil 90, para as amostragens com tamanho amostral igual a 196 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1) 125

LISTA DE TABELAS

Tabela 1 - Parâmetros dos modelos simulados	58
Tabela 2 - Médias das estimativas dos parâmetros do modelo e de algumas medidas de comparação relacionadas à predição espacial para as amostragens em estudo, considerando um modelo com alcance igual a 0,90 e efeito pepita igual a 0, em que: ($\% >$) é a porcentagem de valores preditos acima de; (SQE) é a soma quadrada dos erros e ($\sum >$) é a soma dos valores preditos que estão acima de	65
Tabela 3 - Estatísticas Descritivas e estimativas dos parâmetros do modelo da função semivariância para a variável teor de potássio	73
Tabela 4 - Estimativas dos parâmetros do modelo da função semivariância para o teor de potássio nas amostragens e média da variância de predição	75
Tabela 5 - Estimativas do número e porcentagem dos valores preditos inseridos em cada intervalo da classificação do potássio para a adubação no Estado do Paraná (EMATER, 1998) .	77
Tabela 6 - Médias das estimativas dos parâmetros do modelo e de algumas medidas de comparação relacionadas à predição espacial para as amostragens, considerando um modelo com alcance igual a 0,60 e efeito pepita igual a 0, em que: ($\% >$) é a porcentagem de valores preditos acima de; (SQE) é a soma quadrada dos erros e ($\sum >$) é a soma dos valores preditos que estão acima de	84
Tabela 7 - Médias das estimativas dos parâmetros do modelo e de algumas medidas de comparação relacionadas à predição espacial para as amostragens, otimizada pelo recozimento simulado (RS) e algoritmo genético híbrido (AGH), considerando um modelo com alcance igual a 0,30 e efeito pepita igual a 0, em que: ($\% >$) é a porcentagem de valores preditos acima de; (SQE) é a soma quadrada dos erros e ($\sum >$) é a soma dos valores preditos que estão acima de	85
Tabela 8 - Médias das estimativas dos parâmetros do modelo e de algumas medidas de comparação relacionadas à predição espacial para as amostragens, considerando um modelo com alcance igual a 0,60 e efeito pepita igual a 0,2, em que: ($\% >$) é a porcentagem de valores preditos acima de; (SQE) é a soma quadrada dos erros e ($\sum >$) é a soma dos valores preditos que estão acima de	86
Tabela 9 - Médias das estimativas dos parâmetros do modelo e de algumas medidas de comparação relacionadas à predição espacial para as amostragens, considerando um modelo com alcance igual a 0,60 e efeito pepita igual a 0,5, em que: ($\% >$) é a porcentagem de valores preditos acima de; (SQE) é a soma quadrada dos erros e ($\sum >$) é a soma dos valores preditos que estão acima de	87

LISTA DE QUADROS

Quadro 1 - Etapas de um algoritmo genético	49
Quadro 2 - Algoritmo do processo de recozimento simulado	61
Quadro 3 - Etapas de um algoritmo genético híbrido	62

1 INTRODUÇÃO

Na agricultura tradicional, com o advento da mecanização e a produção em larga escala, a aplicação de fertilizantes é feita de forma uniforme em toda a área, com base em índices médios de fertilidade. Como consequência, áreas com maior nível de fertilidade são adubadas em excesso, causando riscos de contaminação ambiental e um aumento nos custos da produção agrícola, enquanto que áreas com um menor nível de fertilidade são adubadas abaixo da real necessidade do local, impedindo um ótimo desenvolvimento do cultivo.

O recente processo de globalização da economia vem impondo ao setor agrícola nacional a obtenção de uma maior sustentabilidade em mercados competitivos, impulsionando o aprimoramento agrícola, que possa gerar a prática da produção de culturas em larga escala. Nesse contexto, pesquisadores e agricultores estão adotando um novo sistema de agricultura sustentável, chamado de Agricultura de Precisão, que permite a aplicação no local correto e no momento adequado das quantidades de insumos necessários à produção agrícola, proporcionando um aumento racional da produtividade de culturas, uma redução nos custos de produção e uma minimização do impacto negativo ambiental.

Assim, para identificar a necessidade de insumos agrícolas para a aplicação localizada, utiliza-se o mapeamento da fertilidade do solo por meio da análise da variabilidade espacial, possível pelas técnicas de *Geoestatística*, que permitem, a partir de um conjunto aleatório de elementos amostrais, conhecer a continuidade das variáveis de interesse em toda a área de estudo, demonstrando assim, a variação espacial do fenômeno por intermédio de mapas de variabilidade.

Logo, a escolha do tamanho amostral n e das localizações amostrais $(x_1, x_2, \dots, x_n)$ são uma preocupação de pesquisadores e agricultores, pois um planejamento amostral deve considerar o risco do plano de amostragem para o mapeamento da população e o custo desse plano, isto é, deve-se ter uma boa estimativa dos parâmetros do modelo ajustado à função semivariância e uma eficiência na precisão na predição espacial, de tal forma que os resultados obtidos pelo método de interpolação denominado *krigagem* sejam confiáveis, considerando a quantidade necessária de recursos financeiros disponíveis para a execução do plano de amostragem.

Especificamente na análise da variabilidade espacial de propriedades químicas do solo, fator que exerce influência direta no nível de produtividade das culturas, a configuração e tamanho amostral devem ser cuidadosamente definidos, pois, quanto maior a quantidade de amostras do solo na realização da análise química, maior será o custo para o agricultor.

Nesse sentido, o objetivo desse trabalho consiste em definir planos de amostragem, utilizando processos de otimização de inteligência artificial, denominados algoritmo de recozimento simulado e algoritmo genético híbrido. Esses planos de amostragem otimizados devem gerar predições eficientes da variável regionalizada e de algumas de suas características, considerando diferentes tamanhos amostrais

e diferentes características do processo estocástico espacial, tal como alcance e efeito pepita.

2 DESENVOLVIMENTO

2.1 Revisão de Literatura

2.1.1 Agricultura de precisão e variabilidade espacial

Com o aumento da produção agrícola mundial e o advento da mecanização, ocorreu uma mudança no processo de produção agrícola, sendo seu principal objetivo a produção e o lucro em larga escala. No entanto, para maximizar a produção e eliminar as pragas existentes nas plantações houve a necessidade da utilização excessiva de produtos químicos nos campos, gerando um aumento de fatores com efeitos nocivos ao meio ambiente e um aumento de custos na produção agrícola.

Logo, passou a ser objeto de atenção de diversos pesquisadores a necessidade de otimizar a utilização dos recursos de produção no processo agrícola, diminuindo a utilização de insumos agrícolas nos campos, contribuindo para a redução da contaminação do meio ambiente e promovendo um aumento na margem de lucro do agricultor.

Dentro desse cenário, buscou-se um sistema de agricultura sustentável chamado Agricultura de Precisão. Segundo Molin (2002), a agricultura de precisão é definida como um conjunto de procedimentos e tecnologias da informação que possibilitam um gerenciamento da produção agrícola, considerando a variabilidade espacial da produtividade e os fatores a ela relacionados. Seu desenvolvimento ocorreu na Europa, na década de 80, com os avanços tecnológicos, como computadores, satélites, equipamentos e máquinas necessárias para a aplicação de insumos à taxas variadas; Sistema de Posicionamento Global (GPS), Sistema de Informações Geográficas (SIG) e sensoriamento remoto, que possibilitaram a prática da agricultura de precisão.

Silva (2000) cita que o segredo da agricultura de precisão é o uso racional das informações obtidas sobre a variabilidade espacial, que permite a tomada de decisão sobre o melhor manejo para cada local do campo. Enquanto o conceito convencional da aplicação de insumos é feito por taxas médias, onde extensas áreas são gerenciadas de forma idêntica, na agricultura de precisão a aplicação de fertilizantes e corretivos é feita de forma diferenciada entre as zonas de manejo, ou seja, a aplicação é feita no local correto, momento adequado e quantidade necessária à produção agrícola, promovendo uma produtividade máxima do cultivo e uma melhor uniformização da produtividade. Além disso, pela agricultura de precisão é possível uma maior eficiência e utilização dos recursos disponíveis, gerando uma redução dos custos da produção. Baio (2001) comparou a eficiência e a economia da aplicação de herbicidas pela utilização da aplicação localizada de defensivos agrícolas e a aplicação convencional em duas áreas vizinhas localizadas no município de Campos Novos Paulista. A aplicação localizada de insumos agrícolas permitiu uma economia de 31,6% de herbicidas.

Uma das ferramentas importantes na implantação de um programa de agricultura de pre-

cisão é a construção de mapas que expressem a variabilidade de características importantes da produção agrícola. Por exemplo, mapas de produtividade permitem identificar as localizações de áreas críticas, em termos de rentabilidade, enquanto que mapas dos atributos do solo e do relevo podem auxiliar na identificação dos fatores que estão limitando a produtividade.

A geração desses mapas, que expressam a variabilidade, é possível pelo estudo da variabilidade espacial da produtividade e das propriedades do solo. Morgan (1997) define variabilidade espacial como uma variação acerca da distância, profundidade ou tempo que ocorre na produção de grãos, características físicas e químicas do solo, ou população de pragas. Ortiz (2002) ressalta que identificação da estrutura da variabilidade espacial fornece um melhor conhecimento do comportamento da variável na área em estudo e possibilita um melhor planejamento das futuras amostragens.

Para estudar a variabilidade espacial da produtividade e de características do solo existentes em áreas de produção agrícola, há a necessidade de que as informações coletadas sejam referenciadas geograficamente, sendo que uma das técnicas utilizadas para se estudar a variabilidade espacial é conhecida como Geoestatística ou Teoria das Variáveis Regionalizadas.

2.1.2 Geoestatística

2.1.2.1 Variável regionalizada

De acordo com Garcia (1997), as técnicas de estatística clássica utilizam medidas como média e desvio padrão para representarem um conjunto de dados e pressupõem que as variações de um local para outro sejam aleatórias e independentes. Contudo, segundo Johann (2001), pesquisas têm demonstrado que muitos fenômenos naturais possuem variabilidade espacial, existindo portanto, um raio de dependência espacial entre as amostras. Assim, métodos de estatística clássica como análise de variância e testes de hipóteses, que consideram as amostras independentes, não podem ser aplicados, sendo mais indicada a utilização de técnicas que considerem a posição espacial dos valores amostrados.

A Geoestatística, também chamada de *Teoria das Variáveis Regionalizadas*, foi idealizada por George Matheron e entende-se, como tal, a um conjunto de ferramentas utilizadas para analisar e descrever a variabilidade espacial de variáveis, cujas amostras são dependentes da distância. São duas características que diferenciam a geoestatística de outras técnicas de estatística espacial: a localização espacial da variável é considerada fixa e distribuída de forma contínua no espaço.

As variáveis estudadas pela geoestatística são chamadas de Variáveis Regionalizadas. Cada valor do atributo em estudo, amostrado na localização x_i , isto é, $z(x_i)$, representa uma realização de uma variável aleatória $Z(x_i)$ e, considerando as n observações mensuradas nas localizações $x_1, x_2, \dots, x_n$ na região D , tem-se uma realização de um conjunto de variáveis aleatórias $Z(\mathbf{x}) = \{Z(x_1), Z(x_2), \dots, Z(x_n)\}$ chamado de Função Aleatória ou Processo Estocástico (SCHABENBERGER; PIERCE, 2002). As-

sim, cada conjunto de valores amostrados $z(\mathbf{x}) = \{z(x_1), z(x_2), \dots, z(x_n)\}$ é uma realização da função aleatória e representa na prática a variável regionalizada. Essa, possui tanto o caráter aleatório, no sentido de que seus valores são decorrentes de variáveis aleatórias, quanto o caráter estruturado, seguindo uma lei de distribuição espacial que considera a variação de um local para o outro com uma certa continuidade, mas que não pode ser representada por uma simples função matemática. A variável regionalizada tem três características:

- i) Localização geográfica, isto é, no ponto x_i , $Z(x_i)$ é uma variável aleatória;
- ii) A estrutura espacial do fenômeno é expressa pela correlação espacial entre as variáveis aleatórias;
- iii) Anisotropia, isto é, diferentes padrões de continuidade e de variabilidade nas diversas direções da região D .

2.1.2.2 Semivariância

Se a variável é considerada estacionária intrínseca, a função semivariância é definida por:

$$\gamma(h) = \frac{1}{2} \text{Var}[Z(x+h) - Z(x)] \quad (1)$$

e pode ser utilizada para caracterizar a autocorrelação entre $Z(x+h)$ e $Z(x)$ separados pela distância h .

Na prática, para descrever quantitativamente a continuidade espacial da variável de interesse na área em estudo, utiliza-se o *semivariograma experimental*, que consiste num gráfico de dispersão dos valores da semivariância das respectivas distâncias h versus os valores das distâncias h , como exemplificado na Figura 1.

O valor da função semivariância pode ser calculado pelo estimador de Matheron, obtido pelo método dos momentos, que corresponde a uma média da dissimilaridade entre $Z(x)$ e $Z(x+h)$ expresso por:

$$\gamma(h_j) = \frac{1}{2N(h_j)} \sum_{i=1}^n [Z(x_i + h_j) - Z(x_i)]^2,$$

sendo que $N(h_j)$ representa a quantidade de pares de pontos separados pela distância euclidiana h_j que separa dois pontos na malha amostral em estudo e n expressa a quantidade de observações.

Se a amostragem é regular, uma escolha natural para os valores de h_j são as próprias distâncias entre os pontos amostrais. Mas, se a amostragem é realizada em áreas irregulares, então existem diversas distâncias h , sendo provável ter um número muito pequeno de pares de pontos separados pela mesma distância h e, conseqüentemente, o gráfico semivariograma experimental tem muitos pontos, dificultando sua interpretação. Logo, há necessidade de se agruparem as distâncias h em intervalos,

definindo previamente as distâncias e sua tolerância. Geralmente, é considerada como a média das distâncias de cada intervalo.

A semivariância possui as seguintes propriedades:

- i) $\gamma(0) = 0$;
- ii) $\gamma(h) = \gamma(-h) \geq 0$; isto é, a função semivariância é par e sempre produz valores positivos;
- iii) $\lim_{|h| \rightarrow \infty} \frac{\gamma(h)}{|h|^2} = 0$, ou seja, $\gamma(h)$ cresce mais lentamente que h^2 .

Logo, se as semivariâncias $\gamma(h)$ crescerem mais rapidamente do que h^2 há um indício de que a variável não é estacionária intrínseca (SCHABENBERGER; PIERCE, 2002).

Figura 1 - Gráfico semivariograma experimental

Além de ser usado para analisar o grau de dependência espacial entre amostras de um campo experimental, o semivariograma permite definir parâmetros, apresentados na Figura 2, necessários para a estimativa de valores em locais não amostrados. Segundo Wackernagel (1995), nesse gráfico pode-se observar, geralmente, que a média da dissimilaridade entre os valores aumenta quando se aumenta a distância h que separa os pontos amostrais. Assim, para grandes valores de h os valores da semivariância convergem para um determinado valor constante chamado de patamar (*sill*), denotado por $C_0 + C_1$ que, teoricamente, corresponde à variância dos dados. O valor da distância h em que a semivariância atinge o patamar é chamado de alcance (*range*), expresso por a , consistindo na máxima distância onde os pontos amostrais podem ser considerados dependentes. Nesse sentido, a partir do alcance, os dados não possuem uma estrutura espacial e, sendo assim, define-se o tamanho da zona de influência em uma amostra. Para valores pequenos de h , próximos à origem do semivariograma, pode-se ter uma idéia do tipo de continuidade da variável em estudo: diferenciável, contínua mas não diferenciável ou descontínua (DIGGLE; RIBEIRO JUNIOR, 2006). Quando o semivariograma é descontínuo na origem, tem-se o efeito pepita (nugget), expresso por C_0 , mostrando que para pequenas distâncias o valor da variável muda abruptamente. Essa mudança abrupta reflete uma variabilidade não explicada, que pode ser devido a erros de medida ou a uma microvariação não detectada para pequenas distâncias.

Figura 2 - Modelo teórico exponencial ajustado à função semivariância, em que a representa o alcance, C_0 é o efeito pepita e $C_0 + C_1$ é o patamar

2.1.3 Modelos teóricos para o ajuste ao semivariograma

O gráfico do semivariograma experimental sugere que pode ser ajustado um modelo à semivariância como função da distância. O modelo a ser ajustado deve ser o mais simples possível, que represente de forma correta a tendência dos valores de $\gamma(h)$ em relação a h , pois será utilizado no processo de predição dos valores da variável regionalizada em locais não amostrados, permitindo construir um mapa da variabilidade da variável em estudo. Assim, se o modelo ajustado estiver errado, todos os cálculos seguintes estarão também errados e o mapa obtido não representará corretamente a variável em estudo.

Dependendo dos valores de $\gamma(h)$ para valores grandes de h , os modelos podem ser denominados como modelos com patamar ou modelos sem patamar. Os modelos que atingem um patamar representam a existência da estacionaridade de segunda ordem, enquanto os modelos sem patamar satisfazem apenas a hipótese intrínseca.

Utilizando propriedades da variância na função semivariância descrita na eq. (1), para um processo estacionário, tem-se que, de forma geral, o modelo ajustado à semivariância como função da distância é expresso por:

$$\gamma(h) = \frac{1}{2}[2\sigma^2 - 2COV(h)] = \frac{1}{2}[2\sigma^2 - 2\sigma^2\rho(h)],$$

pois $\rho(h) = CORR(h) = \frac{COV(h)}{\sigma^2}$.

Considerando que $\sigma^2 = C_0 + C_1$, tem-se que o modelo ajustado à função semivariância pode ser representado por:

$$\gamma(h) = C_0 + C_1[1 - \rho(h)]. \quad (2)$$

Segundo Diggle e Ribeiro Junior (2006), o modelo ajustado à função semivariância depende dos parâmetros C_0 , $C_0 + C_1$ e da função de correlação $\rho(h)$, que incorpora as seguintes características:

1. A função $\rho(h)$ é monótona não crescente, isto é, a correlação entre duas observações decresce ou permanece constante com o aumento da sua respectiva distância;
2. A função $\rho(h) \rightarrow 0$ quando $h \rightarrow \infty$, isto é, a correlação tende a zero para distâncias h grandes;
3. Pelo menos um parâmetro controla a razão de decaimento de $\rho(h)$ para 0.

A seguir, serão descritas algumas famílias de funções de correlação que satisfazem essas características:

1) Família Exponencial: Esta família com dois parâmetros é definida por:

$$\rho(h) = \exp[-(h/a)^k],$$

com $a > 0$ e $0 \leq k \leq 2$, sendo que h representa a distância entre dois pontos amostrais e a expressa o alcance.

Se $k = 1$, tem-se a função de correlação exponencial e considerando o valor $3a$ como o alcance real, tem-se o seguinte modelo exponencial:

$$\begin{cases} \gamma(h) = C_0 + C_1[1 - \exp(-3\frac{h}{a})], & \text{se } 0 < h < a \\ \gamma(h) = C_0 + C_1, & \text{se } h > a. \end{cases}$$

Nesse modelo, quando $h \rightarrow \infty$, o valor da semivariância se aproxima do valor do patamar assintoticamente, como apresentado na Figura 2, não havendo alcance. Contudo, na prática, considera-se como sendo um alcance efetivo, o valor de h para o qual $\gamma(h)$ é aproximadamente $0,95(C_0 + C_1)$ (MCBRATNEY; WEBSTER, 1986). Além disso, o modelo exponencial é contínuo mas não diferenciável na origem.

Se $k = 2$, tem-se a função de correlação gaussiana e, considerando o valor $3a$ como o alcance real, tem-se o seguinte modelo gaussiano:

$$\gamma(h) = C_0 + C_1[1 - \exp(-3\frac{h^2}{a^2})].$$

O modelo gaussiano, apresentado na Figura 3, possui um patamar assintótico, como o modelo exponencial e, quando $h \rightarrow 0$, esse modelo é infinitamente diferenciável e contínuo, sendo o modelo mais contínuo na origem, em relação aos modelos apresentados neste trabalho.

2) Família Esférica com um parâmetro é definida por:

$$\begin{cases} \rho(h) = 1 - \frac{3}{2}(\frac{h}{a}) + \frac{1}{2}(\frac{h}{a})^3, & \text{se } 0 < h < a \\ \rho(h) = 0, & \text{se } h > a \end{cases}$$

em que h representa a distância entre dois pontos amostrais, a representa o alcance, tal que $a > 0$.

Utilizando a eq. (2), tem-se que o modelo esférico ajustado à função semivariância é dado por:

$$\begin{cases} \gamma(h) = C_0 + C_1 \left[\frac{3}{2} \cdot \left(\frac{h}{a} \right) - \frac{1}{2} \cdot \left(\frac{h}{a} \right)^3 \right], & \text{se } 0 < h \leq a \\ \gamma(h) = C_0 + C_1, & \text{se } h > a. \end{cases}$$

Figura 3 - Modelo teórico gaussiano

Um exemplo do modelo esférico é apresentado na Figura 4 e suas características principais são: existência de alcance, patamar finito e próximo à origem e esse modelo é uma vez diferenciável.

Todos os modelos descritos são considerados com patamar, sendo que C_0 é o efeito pepita, C_1 é chamado de contribuição, $C_0 + C_1$ é o patamar e a é o alcance.

Figura 4 - Modelo teórico esférico

Os modelos sem patamar podem ser escritos da seguinte forma:

$$\gamma(h) = C_0 + Ah^b, \quad 0 < b < 2.$$

2.1.3.1 Método de máxima verossimilhança e máxima verossimilhança restrita

Se a variável em estudo segue a distribuição Gaussiana, pode-se utilizar o Método de Máxima Verossimilhança na estimação dos parâmetros do modelo ajustado à função semivariância. Esse método baseia-se no fato de que a partir das observações de uma variável Z , contínua ou discreta, pode-se obter alguma informação sobre a distribuição de Z ou, equivalentemente, sobre os parâmetros $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_m)'$ da distribuição de Z .

Por meio de experimentos de simulação com variáveis regionalizadas, Mardia e Marshall (1984) concluíram que a estimação dos parâmetros pela Máxima Verossimilhança para amostras pequenas, produz estimativas viesadas, principalmente, para o patamar, mas quando utilizada em grandes amostras, produz estimativas não viesados, consistentes e eficientes. Sua principal desvantagem é a necessidade de uma especificação geral do modelo probabilístico e a inversão de matrizes complexas, podendo haver dificuldades computacionais (CRESSIE, 1993 e DIGGLE; RIBEIRO JUNIOR, 2006).

Seja $\mathbf{Z} = (Z(x_1), Z(x_2), \dots, Z(x_n))$, representado por um modelo linear Gaussiano, expresso por (DIGGLE; RIBEIRO JUNIOR, 2006):

$$Z_i = \mu(x_i) + S(x_i) + \epsilon_i,$$

sendo que $S(x_i)$ é o processo gaussiano referenciado espacialmente com variância igual a σ^2 e função de correlação $\rho(\mathbf{h}, a)$, em que a representa o alcance; $\mu(x_i)$ é a média representada por um modelo de regressão linear $\sum_{j=1}^{p+1} f_{j-1}(\mathbf{x})\beta_{j-1}$ tal que f_{j-1} são variáveis explanatórias espacialmente referenciadas, e ϵ_i é o ruído ou erro, mutuamente independente e identicamente distribuído tal que $\epsilon_i \sim N(0, \tau^2)$.

Considerando $\mu(x_i) = \mu$ uma constante, tem-se que $\mathbf{Z} \sim N(\mu\mathbf{1}, C(\boldsymbol{\phi}))$, em que $\mathbf{1}$ é um vetor de uns e $C(\boldsymbol{\phi})$ é a transposta da matriz de covariâncias definida pela função covariograma, cujo elemento posicionado na i -ésima linha e j -ésima coluna é expresso por:

$$C_{ij}(\boldsymbol{\phi}) = \text{cov}(Z(\mathbf{x}_i), Z(\mathbf{x}_j)) = \text{var}(Z(\mathbf{x})) - \gamma(h; \boldsymbol{\phi}). \quad (3)$$

Assim, a função de verossimilhança é dada por:

$$L(\boldsymbol{\phi}, \mu; \mathbf{Z}) = f(\mathbf{Z}; \boldsymbol{\phi}, \mu) = \frac{1}{(2\pi)^{\frac{n}{2}} |C(\boldsymbol{\phi})|^{\frac{1}{2}}} \exp -\frac{1}{2}(\mathbf{Z} - \mu\mathbf{1})' C(\boldsymbol{\phi})^{-1} (\mathbf{Z} - \mu\mathbf{1}).$$

Como a função logarítmica é monótona, maximizar $L(\boldsymbol{\phi}, \mu; \mathbf{Z})$ é equivalente a maximizar o logaritmo de $L(\boldsymbol{\phi}, \mu; \mathbf{Z})$. Então, a estimativa de $\boldsymbol{\phi} = (C_0, C_0 + C_1 = \sigma^2, a)$, por máxima verossimilhança (ML), é obtida pela maximização de:

$$l(\boldsymbol{\phi}) = \ln f(\mathbf{Z}; \boldsymbol{\phi}, \mu) = l(\boldsymbol{\phi}, \mu; \mathbf{Z}) = -\frac{n}{2} \ln(2\pi) - \frac{1}{2} \ln |C(\boldsymbol{\phi})| - \frac{1}{2}(\mathbf{Z} - \mu\mathbf{1})' C(\boldsymbol{\phi})^{-1} (\mathbf{Z} - \mu\mathbf{1}).$$

Quando $m = 1$, sendo m a quantidade de parâmetros do modelo, o gráfico de $l(\phi)$ versus ϕ , chamado de superfície suporte é bastante informativo. Mas, quando $m > 3$, a superfície suporte não pode ser obtida e tem-se que recorrer às técnicas iterativas.

É aconselhável eliminar o parâmetro *nuisance* μ antes de estimar ϕ , o que é obtido, considerando a verossimilhança dos contrastes:

$$\mathbf{Z}_c = Z(\mathbf{x}_i) - Z(\mathbf{x}_1) \quad \text{com } i = 2, \dots, n.$$

Assim:

$$\mathbf{Z}_c = \mathbf{Q}\mathbf{Z} \sim N(\mathbf{0}, \mathbf{Q}C(\phi)\mathbf{Q}'),$$

com:

$$\mathbf{Z}_c = \begin{bmatrix} Z(x_2) - Z(x_1) \\ Z(x_3) - Z(x_1) \\ \dots \\ Z(x_n) - Z(x_1) \end{bmatrix}, \quad \mathbf{Q} = \begin{bmatrix} -1 & 1 & 0 & \dots & 0 \\ -1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & 0 & 0 & \dots & 1 \end{bmatrix}, \quad \mathbf{Z} = \begin{bmatrix} Z(x_1) \\ Z(x_2) \\ \dots \\ Z(x_n) \end{bmatrix} \quad \text{e}$$

$$C_{ij}(\phi) = \begin{bmatrix} C_{11}(\phi) & C_{21}(\phi) & \dots & C_{n1}(\phi) \\ C_{12}(\phi) & C_{22}(\phi) & \dots & C_{n2}(\phi) \\ \dots & \dots & \dots & \dots \\ C_{1n}(\phi) & C_{2n}(\phi) & \dots & C_{nn}(\phi) \end{bmatrix},$$

sendo que:

$$C_{ii} = C_{ii}(\phi) = \text{cov}[Z(x_i), Z(x_i)] = \text{var}[Z(x_i)]$$

e:

$$\mathbf{Q}C(\phi)\mathbf{Q}' = \begin{bmatrix} C_{22} - C_{21} - C_{12} + C_{11} & \dots & C_{n2} - C_{n1} - C_{12} + C_{11} \\ C_{23} - C_{21} - C_{13} + C_{11} & \dots & C_{n3} - C_{n1} - C_{13} + C_{11} \\ \dots & \dots & \dots \\ C_{2n} - C_{21} - C_{1n} + C_{11} & \dots & C_{nn} - C_{n1} - C_{1n} + C_{11} \end{bmatrix}.$$

Como:

$$C_{nn} - C_{n1} - C_{1n} + C_{11} = 2\gamma(h) = 2\gamma(x_n, x_1; \phi),$$

então, a diagonal da matriz $\mathbf{Q}C(\phi)\mathbf{Q}'$ tem elementos dados por:

$$\psi_{ii}(\boldsymbol{\phi}) = 2\gamma(x_i, x_1; \boldsymbol{\phi}).$$

Os elementos fora da diagonal da matriz $\mathbf{Q}C(\boldsymbol{\phi})\mathbf{Q}'$ são dados por:

$$\psi_{ij}(\boldsymbol{\phi}) = \gamma(x_i, x_1; \boldsymbol{\phi}) + \gamma(x_j, x_1; \boldsymbol{\phi}) - \gamma(x_i, x_j; \boldsymbol{\phi}).$$

Assim,

$$L(\boldsymbol{\phi}; \mathbf{Z}_c) = f(\mathbf{Z}_c; \boldsymbol{\phi}) = \frac{1}{(2\pi)^{\frac{n}{2}} |C(\boldsymbol{\phi})|^{\frac{1}{2}}} \exp -\frac{1}{2} \mathbf{Z}_c' \boldsymbol{\psi}^{-1}(\boldsymbol{\phi}) \mathbf{Z}_c. \quad (4)$$

Maximizando a eq. (4) em relação a $\boldsymbol{\phi}$, têm-se as estimativas obtidas pelo método de *Máxima Verossimilhança Restrita* (REML). Pettitt e Mcbratney (1993) afirmam que o método de Máxima Verossimilhança Restrita produz estimativas mais próximas de seu valor real, especialmente se existem muitos efeitos fixos no modelo.

2.1.4 *Krigagem*

Uma etapa importante na Geoestatística para a descrição e análise da variabilidade espacial de uma variável de interesse é a construção de mapas de variabilidade espacial. Sua construção é possível por um método de interpolação geoestatístico chamado de *krigagem*. Segundo Thompson (1992), como $Z(x_0)$ é considerada uma variável aleatória, então o problema de inferência é conhecido como predição. Sendo assim, pela *krigagem*, é possível prever um valor desconhecido $Z(x_0)$ em qualquer posição não amostrada x_0 e também de toda função de distribuição acumulada da variável em cada ponto, possibilitando a construção de mapas da média, medianas, quantis e probabilidades. Existem diversos tipos de *krigagem*, tais como: ordinária, indicadora e universal.

2.1.4.1 *Krigagem ordinária*

Segundo Câmara e Medeiros (1998), a *krigagem* ordinária é baseada na idéia da regressão linear, em que a predição de um valor desconhecido $Z(x_0)$ é obtida pela combinação linear de n valores conhecidos $Z(x_i)$ adicionada a um parâmetro λ_0 , pelo preditor:

$$Z^*(x_0) = \lambda_0 + \sum_{i=1}^n \lambda_i Z(x_i), \quad (5)$$

sendo que $Z^*(x_0)$ é o valor predito na posição não amostrada x_0 , λ_i são os pesos e $Z(x_i)$ são os valores conhecidos da variável nos locais amostrados.

Os pesos são baseados nas distâncias entre os pontos vizinhos ao ponto predito e também na distância entre estes. Segundo Schabenberger e Pierce (2002), como se deseja um preditor não tendencioso, no sentido de que:

$$E[Z(x_0) - Z^*(x_0)] = 0 \implies E[Z(x_0)] = E[Z^*(x_0)],$$

obtém-se:

$$E[Z(x_0)] = E[\lambda_0 + \sum_{i=1}^n \lambda_i Z(x_i)] \implies m = \lambda_0 + \sum_{i=1}^n \lambda_i m,$$

sendo satisfeita se e somente se:

$$\lambda_0 = 0 \quad \text{e} \quad \sum_{i=1}^n \lambda_i = 1.$$

Logo, a eq. (5) pode ser expressa como:

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i Z(x_i). \quad (6)$$

Assim, considerando que $\sum \lambda_i = 1$, espera-se que o erro quadrático médio de predição, também chamado de variância do erro, descrita na eq. (7) seja mínimo:

$$Var[Z^*(x_0) - Z(x_0)] = E\{[Z^*(x_0) - Z(x_0)]^2\} - \{E[Z^*(x_0) - Z(x_0)]\}^2. \quad (7)$$

No segundo lado da igualdade da eq. (7), o primeiro termo corresponde ao erro quadrático médio e o segundo termo corresponde ao viés. Como o preditor da *krigagem* é não tendencioso, tem-se que o segundo termo é igual a zero, logo a variância do erro pode ser escrita como:

$$Var[Z^*(x_0) - Z(x_0)] = E\{[Z^*(x_0) - Z(x_0)]^2\}.$$

A variância do erro também pode ser escrita pelas covariâncias, isto é:

$$\begin{aligned} Var[Z^*(x_0) - Z(x_0)] &= COV[Z^*(x_0), Z^*(x_0)] - 2COV[Z^*(x_0), Z(x_0)] + COV[Z(x_0), Z(x_0)] \\ &= VAR[Z^*(x_0)] - 2COV[Z^*(x_0), Z(x_0)] + VAR[Z(x_0)]. \end{aligned} \quad (8)$$

Como a variância de $Z^*(x_0)$ é uma combinação de variáveis aleatórias, então pode ser escrita como:

$$Var[Z^*(x_0)] = VAR\left[\sum_{i=1}^n \lambda_i Z(x_i)\right] = \sum_{i=1}^n \sum_{j=1}^n C_{ij}(\phi), \quad (9)$$

em que $C_{ij}(\phi)$ é a covariância entre $Z(x_i)$ e $Z(x_j)$, definida na eq. (3).

Tem-se ainda que:

$$\begin{aligned} 2COV[Z^*(x_0), Z(x_0)] &= 2COV\left[\left(\sum_{i=1}^n \lambda_i Z(x_i)\right), Z(x_0)\right] \\ &= 2 \sum_{i=1}^n \lambda_i Z(x_i) COV[Z(x_i), Z(x_0)] \\ &= 2 \sum_{i=1}^n \lambda_i Z(x_i) C_{i0}(\phi), \end{aligned} \quad (10)$$

em que $C_{i0}(\boldsymbol{\phi})$ é a covariância entre $Z(x_i)$ e $Z(x_0)$.

Substituindo na eq. (8) os resultados obtidos pelas eq. (9) e (10), tem-se que:

$$Var[Z^*(x_0) - Z(x_0)] = \sum_{i=1}^n \sum_{j=1}^n C_{ij}(\boldsymbol{\phi}) - 2 \sum_{i=1}^n \lambda_i Z(x_i) C_{i0}(\boldsymbol{\phi}) + \sigma^2. \quad (11)$$

Desse modo, tem-se que encontrar o conjunto de pesos λ_i , para $i = 1, 2, \dots, n$ tal que a variância do erro como definida na eq. (11) seja mínima, considerando também que $\sum \lambda_i = 1$. Como essa minimização é extremamente difícil (ISAAKS; SRIVASTAVA, 1989), deve-se usar o multiplicador de Lagrange, para transformar o problema de minimização com restrição em um problema de minimização sem restrição. Logo,

$$Var[Z^*(x_0) - Z(x_0)] = \sum_{i=1}^n \sum_{j=1}^n C_{ij} - 2 \sum_{i=1}^n \lambda_i Z(x_i) C_{i0} + \sigma^2 + 2\alpha(\sum_{i=1}^n \lambda_i - 1), \quad (12)$$

sendo que α é o multiplicador de Lagrange e o termo $2\alpha(\sum_{i=1}^n \lambda_i - 1)$ é igual a zero.

Minimizando-se a eq. (12) em relação a λ_i , tem-se que esses pesos são obtidos a partir do sistema de equações chamado de sistema de krigagem ordinária:

$$\begin{cases} \sum_{i=1}^n \lambda_i C_{ij} + \alpha = C_{i0} \\ \sum_{i=1}^n \lambda_i = 1. \end{cases}$$

Como $\gamma_{ij} = \sigma^2 - C_{ij} \Rightarrow C_{ij} = \sigma^2 - \gamma_{ij}$, tem-se que:

$$\begin{cases} \sum_{i=1}^n \lambda_i \gamma_{ij} - \alpha = \gamma_{i0} \\ \sum_{i=1}^n \lambda_i = 1 \end{cases}$$

o que pode ser escrito na forma matricial como:

$$\boldsymbol{\gamma} \boldsymbol{\lambda} = \boldsymbol{b}$$

cujas soluções é dada por:

$$\boldsymbol{\lambda} = \boldsymbol{\gamma}^{-1} \boldsymbol{b}$$

sendo que:

$$\boldsymbol{\lambda} = \begin{bmatrix} \gamma(x_1, x_1) & \gamma(x_1, x_2) & \cdots & \gamma(x_1, x_n) & 1 \\ \gamma(x_2, x_1) & \gamma(x_2, x_2) & \cdots & \gamma(x_2, x_n) & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \gamma(x_n, x_1) & \gamma(x_n, x_2) & \cdots & \gamma(x_n, x_n) & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} \gamma(x_1, x_0) \\ \gamma(x_2, x_0) \\ \vdots \\ \gamma(x_n, x_0) \\ 1 \end{bmatrix} \quad \text{e} \quad \boldsymbol{\lambda} = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \alpha \end{bmatrix}.$$

De acordo com Ribeiro Junior (1995), a *krigagem* ordinária é o melhor preditor linear não viesado e com variância mínima (BLUP - *best linear unbiased predictor*), o que assegura que este interpolador seja exato, fornecendo valores iguais aos observados quando a localização do ponto estimado coincidir com uma localização observada (ANDRIOTTI, 2004). Mas como as equações da *krigagem* não dependem dos valores individuais, mas sim das localizações dos pontos amostrados e das semivariâncias, suas estimativas e estatísticas derivadas são sensíveis às pequenas mudanças no variograma. Assim, os resultados obtidos podem ser viesados quando o variograma não for representativo na área de interesse (LLOYD; ATKINSON, 1999).

Uma qualidade importante da *krigagem* é que além de se fazerem previsões do valor da variável regionalizada em um ponto, é possível associar a essa previsão, uma medida de precisão ou confiabilidade da previsão, denominada de variância da *krigagem*, dada por:

$$\sigma_0^2 = \sum_{i=1}^n \lambda_i \gamma_{i0} + \alpha - \gamma_{ii}. \quad (13)$$

Por meio da eq. (13), vê-se que a variância da *krigagem*, também como seu interpolador, depende apenas das localizações dos pontos amostrados e das semivariâncias. Dessa forma, a variância será máxima em locais mais distantes de valores medidos em relação ao local de previsão. Tem-se ainda que, quanto maior for a diferença entre o efeito pepita e o patamar no semivariograma, maior será a continuidade espacial e a confiança que se pode ter na estimativa, conseqüentemente, menor será a variância da estimativa (VIEIRA, 1998).

2.1.4.2 *Krigagem* universal

Uma generalização da *krigagem* ordinária é a *krigagem* universal, que também é um interpolador linear uniformemente não viesado, utilizado para predizer o valor de Z no ponto x_0 , isto é, $Z(x_0)$. Considerando nesse modelo, um mesmo termo, o processo estacionário e o ruído, tem-se que:

$$Z(\mathbf{x}) = \sum_{j=1}^{p+1} f_{j-1}(\mathbf{x}) \beta_{j-1} + \delta(\mathbf{x}), \quad (14)$$

ou escrevendo a eq. (14) em notação matricial:

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\delta},$$

sendo que $\mathbf{Z} = [Z(x_1), Z(x_2), \dots, Z(x_n)]'$ é o conjunto de dados da variável aleatória Z ; $\boldsymbol{\beta} = [\beta_0, \beta_1, \dots, \beta_p]'$ é o vetor de parâmetros desconhecidos; $\boldsymbol{\delta} = [\delta(x_1), \delta(x_2), \dots, \delta(x_n)]'$ é o vetor do processo aleatório tal que $E[\delta(x_i)] = 0$ e com variograma $2\gamma(\cdot)$ e $\mathbf{X}$ é a matriz do esquema amostral ou matriz de variáveis explanatórias dada por:

$$\mathbf{X} = \begin{bmatrix} f_0(x_1) & f_1(x_1) & \cdots & f_p(x_1) \\ f_0(x_2) & f_1(x_2) & \cdots & f_p(x_2) \\ \vdots & \vdots & \vdots & \vdots \\ f_0(x_n) & f_1(x_n) & \cdots & f_p(x_n) \end{bmatrix}.$$

O preditor de $Z(x_0)$ linear em relação aos dados $\mathbf{Z}$, uniformemente não viesado, é expresso pela eq. (15):

$$Z^*(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (15)$$

tal que: $\boldsymbol{\lambda}'\mathbf{X} = \mathbf{x}'$, em que $\mathbf{x}' = [f_0(x_0), f_1(x_0), \dots, f_p(x_0)]$ e $\boldsymbol{\lambda}' = [\lambda_1, \lambda_2, \dots, \lambda_n]$. Se se considerar $f_0(\mathbf{x}) = 1$ e $p = 0$, tem-se a *krigagem* ordinária e a restrição $\boldsymbol{\lambda}'\mathbf{X} = \mathbf{x}'$ é igual a $\sum \lambda_i = 1$.

Ainda espera-se que o erro quadrático médio de predição ou variância do erro,

$$\sigma_e^2 = E[Z(x_0) - Z^*(x_0)]^2 = E[Z(x_0) - \sum_{i=1}^n \lambda_i Z(x_i)]^2 \quad (16)$$

seja mínimo, considerando $\boldsymbol{\lambda}'\mathbf{X} = \mathbf{x}'$.

Conforme Cressie (1993), introduzindo os multiplicadores de Lagrange $\mathbf{m}$ que garantem que $\boldsymbol{\lambda}'\mathbf{X} = \mathbf{x}'$, os valores de $\boldsymbol{\lambda}$ e $\mathbf{m}$ são obtidos por:

$$\boldsymbol{\lambda}' = [\boldsymbol{\gamma} + \mathbf{X}(\mathbf{X}'\boldsymbol{\Gamma}^{-1}\mathbf{X})^{-1}(\mathbf{x} - \mathbf{X}'\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma})]'\boldsymbol{\Gamma}^{-1} \text{ e}$$

$$\mathbf{m}' = -(\mathbf{x} - \mathbf{X}'\boldsymbol{\Gamma}^{-1}\boldsymbol{\gamma})'(\mathbf{X}'\boldsymbol{\Gamma}^{-1}\mathbf{X})^{-1},$$

sendo que $\boldsymbol{\gamma} = [\gamma_{10}, \gamma_{20}, \dots, \gamma_{n0}]'$ e $\boldsymbol{\Gamma}$ é uma matriz $n \times n$ tal que o (i, j) -ésimo elemento é γ_{ij} .

2.1.4.3 *Krigagem* indicatriz

Segundo Vieira (1998), a utilização da *krigagem* ordinária produz resultados sub-estimados ou super-estimados da variância da predição, que representa uma medida de precisão na predição. Uma alternativa é a utilização da *krigagem* indicatriz, que possibilita além da estimação da esperança em cada ponto, também a estimação de toda função de distribuição acumulada da variável em cada ponto não

amostrado. A *krigagem* indicatriz consiste em transformar o conjunto de dados da variável de interesse num conjunto de dados binários, de tal forma que a variável se torne uma variável indicadora. Nessa transformação, são designados pontos de corte z_c , chamados de *cutoff*, selecionados segundo algum critério do pesquisador, como, por exemplo, os decis da distribuição acumulativa amostral (LLOYD; ATKINSON, 1999).

Assim, a transformação do conjunto de dados é dada por:

$$I(x_i, z_c) = \begin{cases} 1, & \text{se } Z(x_i) \leq z_c \\ 0, & \text{caso contrário.} \end{cases}$$

Depois da transformação do conjunto de dados, mantêm-se os procedimentos utilizados na krigagem ordinária, isto é, calculam-se os semivariogramas, ajusta-se um modelo e pelas estimativas obtidas para os parâmetros deste modelo, calculam-se as estimativas da variável em locais não amostrados, em que o estimador BLUE expresso na eq. (6), segundo Almeida e Ribeiro Junior (1996), é transformado no preditor linear expresso pela eq. (17):

$$\hat{I}(x_0, z_c) = \sum_{i=1}^n \hat{\lambda}_i i(x_i). \quad (17)$$

2.1.5 Esquemas de amostragem espacial

Geralmente, o pesquisador desenvolve sua pesquisa com o intuito de compreender determinadas características de todo o grupo de interesse, seja ele um grupo de pessoas, animais, ou uma área agrícola. Esse grupo, denominado população, consiste em um conjunto de indivíduos que possuem pelo menos uma característica em comum. Como existem limitações, tal como tempo disponível e recursos econômicos, torna-se inviável e impraticável o estudo de todos os indivíduos da população, utilizando-se apenas uma parte da população denominada amostra, cujos elementos são chamados de unidades amostrais. A forma de se escolher as unidades amostrais na população de interesse, para a variável em estudo, é chamado de método de amostragem. Segundo Machado et. al (2005), o processo de escolha da amostra permite um controle inicial da pesquisa possibilitando que características estranhas à amostra sejam irrelevantes ou constantes.

Para o caso bi-dimensional, os métodos de amostragem espacial são uma forma racional de utilizar as ferramentas estatísticas, para determinar o tamanho amostral n e as localizações amostrais $\{x_1, x_2, \dots, x_n\}$, na região de interesse $D \subset \mathbb{R}^d$ ($d \geq 1$), definindo assim, a configuração da amostra. De acordo com Stingel (2005), é de suma importância o desenvolvimento de esquemas de amostragem mais eficientes, no sentido de que os resultados obtidos se aproximem dos resultados reais presentes na área em estudo. O objetivo dessa eficiência pode ser a predição dos valores de Z em locais não amostrados, a predição de alguma função da variável $g(Z(.))$, ou ainda, a estimação dos parâmetros do modelo ajustado

à semivariância como função da distância. Existem diversos planos de amostragem espacial e sua escolha depende, além de tempo e recursos econômicos, também do propósito do experimento, das características da área em estudo e das características da distribuição espacial da variável na área em estudo. Alguns planos de amostragem espacial clássicos são os que se seguem.

Na amostragem aleatória simples, os pontos são escolhidos aleatoriamente, por meio de uma distribuição uniforme sobre a região D . Na amostragem sistemática, escolhe-se a localização de um ponto amostral de forma aleatória e, em seguida, especificam-se as $(n - 1)$ localizações dos pontos amostrais, de tal forma que todos os pontos estejam localizados segundo algum padrão regular. Quando a localização do ponto inicial não é escolhida de forma aleatória, o plano de amostragem é denominado determinístico ou regular. Sua principal característica é a estrutura geométrica da localização das amostras, sendo a área de interesse dividida em polígonos com todos os ângulos internos iguais e todos os lados iguais e os elementos da amostras estão posicionados no vértice dos polígonos. Exemplos de planos de amostragem regular são: grade quadrada, grade triangular eqüilateral, grade retangular, grade hexagonal. Segundo Cressie (1993), a amostragem regular tem fácil implementação e produz melhores resultados de eficiência em relação à média e o máximo da variância da *krigagem*.

A amostragem estratificada consiste na divisão da região D em estratos não sobrepostos e dentro de cada estrato as localizações dos pontos amostrais são escolhidas por uma amostra aleatória uniforme. Esse tipo de amostragem é utilizado quando existem áreas, dentro de uma região, heterogêneas entre si, mas dentro de cada estrato a área é homogênea. O caso mais simples, denominado estrato quadrado, é quando a área é dividida em quadrados iguais, definindo os estratos (RIPLEY, 1981). Na prática, o conhecimento prévio da região D pode sugerir a forma e o tamanho dos estratos.

Quenoulli (1949) generaliza e amplia as formas de se definir a configuração amostral, empregando combinações nas duas direções da área de interesse das amostragens: aleatória, sistemática e estratificada, surgindo assim, diversos tipos de amostragem. Logo, pode-se obter, por exemplo, na direção x a amostragem aleatória e na direção y a amostragem sistemática. Considera-se, ainda, se as amostras são alinhadas ou independentes (desalinhadas) em uma ou em ambas as direções.

Krishnaiah e Rao (1988) descrevem o método de Quenoulli (1949) para dados distribuídos no $\mathbf{R}^2$ de forma discreta, assumindo que a população consiste em LM unidades, distribuídos em L linhas na direção x e M colunas na direção y das quais se desejam obter lm amostras, pela amostragem aleatória, considerando que $L = lk$ e $M = mq$. Se as amostras são alinhadas apenas na direção x , então os elementos amostrados em cada linha da direção x terão 0 ou l elementos, sendo que as linhas dos l elementos serão escolhidas de forma aleatória e, na direção y , a posição dos elementos relativa a essa direção será escolhida de forma aleatória.

Se as amostras são alinhadas apenas na direção y , então cada coluna da direção y terá 0

ou m elementos e as colunas que terão os m elementos serão escolhidas de forma aleatória. Na direção x , a posição dos elementos relativa a essa direção será escolhida de forma aleatória. Se as amostras são alinhadas em ambas as direções, então cada linha da direção x terá 0 ou l elementos e cada coluna da direção y terá 0 ou m elementos. Se as amostras são desalinhadas, a posição dos elementos amostrados será obtida de forma aleatória e cada linha da direção x terá no máximo l elementos amostrados e cada coluna da direção y terá no máximo m elementos amostrados. Exemplos de planos de amostragem aleatórios alinhados e desalinhados são apresentados na Figura 5.

Figura 5 - Amostragem Aleatória: (a) desalinhada, (b) alinhada em uma direção e (c) alinhada em ambas as direções

Na amostragem sistemática, o conjunto de pontos para o caso com desalinhamento em ambas as direções pode ser obtido por meio de:

$$S = \{(r_j + (i - 1)k, s_i + (j - 1)q) : j = 1, \dots, m; i = 1, \dots, l\}, \quad (18)$$

sendo que r_j , com $j = 1, \dots, m$ são m valores inteiros sorteados com reposição entre 1 e k e s_i , com $i = 1, \dots, l$ são l valores inteiros sorteados com reposição entre 1 e q . Quando a amostra sistemática é obtida com alinhamento em ambas as direções, tem-se que $r_1 = \dots = r_m = r$ e $s_1 = \dots = s_l = s$, com r e s escolhidos de forma aleatória entre os inteiros 1 e k e 1 e q respectivamente. O conjunto de pontos pode ser obtido por:

$$S = \{(r + (i - 1)k, s + (j - 1)q) : j = 1, \dots, m; i = 1, \dots, l\}. \quad (19)$$

Quando a amostra sistemática é obtida considerando o alinhamento em apenas uma direção, y ou x , tem-se que o conjunto de pontos pode ser obtido respectivamente por:

$$S = \{(r + (i - 1)k, s_i + (j - 1)q) : j = 1, \dots, m; i = 1, \dots, l\}$$

e

$$S = \{(r_j + (i - 1)k, s + (j - 1)q) : j = 1, \dots, m; i = 1, \dots, l\},$$

com $s_1, \dots, s_l$ escolhidos de forma aleatória com reposição quando o alinhamento é atribuído à direção y e $r_1, \dots, r_m$ são escolhidos de forma aleatória com reposição quando o alinhamento é atribuído à direção x .

Se a amostra sistemática é alinhada em ambas as direções, como apresentada na Figura 6, então é chamada de amostra sistemática centrada (*lattice*). Nessa, a área populacional é dividida em quadrados iguais e a unidade amostral estará no centro de cada quadrado. Assim, as unidades da amostra são equidistantes, definindo apenas uma única amostra sistemática centrada, com determinado tamanho em cada área de interesse. Além de sua fácil implementação na área de interesse, essa forma de amostragem sistemática é preferível, pois não favorece nenhum lado da área e devido à localização central das amostras em cada quadrado, não leva a resultados errôneos por causa do aumento consistente da variável de interesse em determinada direção ou variação periódica (RIPLEY, 1981).

A amostragem sistemática centrada comparada com as amostragens aleatória e estratificada aleatória produz uma melhor estimativa para a média, menor erro padrão, intervalo de confiança para a média populacional mais eficiente (MILNE, 1959) e menor valor empírico da variância da amostragem (BELLHOUSE, 1977). Contudo, os valores empíricos apresentados para a variância de amostragem são altamente variáveis. Duas possíveis explicações para essa grande variabilidade é a pequena quantidade de amostras e a existência de uma dependência espacial complexa e variável, resultados que se podem confundir com a eficiência da amostragem sistemática (DUNN; HARRISON, 1993).

Figura 6 - Amostragem Sistemática Centrada

Como exemplo, suponha que a população de interesse consiste em 100 quadrados distribuídos em 10 linhas e 10 colunas ($M = N = 10$). Deseja-se obter três tipos de amostras com 10 quadrados, em que $L = l \cdot k = 5 \cdot 2 \implies l = 5, k = 2$ e $M = m \cdot q = 2 \cdot 5 \implies m = 2, q = 5$ e com os seguintes esquemas amostrais: sistemática desalinhada em ambas as direções, sistemática alinhada em

ambas as direções e sistemática alinhada apenas na direção y .

Para se obter uma amostra com 10 unidades amostrais, para a amostragem sistemática desalinhada em ambas as direções, sorteia-se com reposição dois ($m = 2$) números inteiros entre um e dois ($k = 2$) para serem os valores de r , e cinco ($l = 5$) números inteiros entre 1 e cinco ($q = 5$) para serem os valores de s , obtendo-se, por exemplo: $[r_1, r_2] = [1, 2]$ e $[s_1, s_2, s_3, s_4, s_5] = [1, 5, 3, 2, 4]$. Utilizando a eq. (18) para a obtenção do conjunto de pontos amostrais, tem-se que: $S = \{(1, 1), (3, 5), (5, 3), (7, 2), (9, 4), (2, 6), (4, 10), (6, 8), (8, 7), (10, 9)\}$, sendo esse conjunto de pontos amostrais apresentado na Figura 7-(a).

Uma amostra de tamanho 10 pode ser obtida pela amostragem sistemática alinhada em ambas as direções, fazendo um sorteio para os valores de r e s , conseguindo-se, por exemplo, $r = 2$ e $s = 3$. Utilizando a eq. (19), chega-se ao seguinte conjunto de pontos amostrais: $S = \{(2, 3), (2, 8), (4, 3), (4, 8), (6, 3), (6, 8), (8, 3), (8, 8), (10, 3), (10, 8)\}$, representado na área da população pela Figura 7-(b).

Pela amostragem sistemática alinhada apenas na direção y , sorteia-se um valor para r , obtendo $r = 1$ e sorteam-se, com reposição, 10 valores para s , tendo-se o seguinte conjunto de pontos amostrais: $S = \{(1, 2), (1, 8), (3, 5), (3, 9), (5, 1), (5, 10), (7, 3), (7, 8), (9, 2), (9, 9)\}$, representado na área da população pela Figura 7-(c).

Figura 7 - Amostragem Sistemática: (a) desalinhada em ambas as direções, (b) alinhada em ambas as direções e (c) alinhada em uma direção

Na amostragem estratificada alinhada em uma determinada direção, os estratos são definidos na direção de interesse e todos os elementos em cada estrato estão na mesma linha naquela direção. Se a amostragem é estratificada desalinhada, não há um posicionamento alinhado dos elementos amostrais dentro de cada estrato.

Os métodos de amostragem espacial: aleatório, sistemático e estratificado, ignoram considerações importantes na prática como, por exemplo, efeito pepita relevante e amostras mensuradas muito próximas, em localizações praticamente idênticas, que podem influenciar na estimação dos parâmetros desconhecidos do modelo.

Logo, dois métodos de amostragem são propostos: amostragem sistemática centrada com pares de pontos próximos (*lattice plus close pairs*) e amostragem sistemática centrada com amostragens sistemáticas centradas menores (*lattice plus in-fill*) que, além de considerar a regularidade espacial, cobre toda a área de interesse D, inclusive pares de pontos próximos. O método de amostragem *lattice plus close pairs*, apresentado na Figura 8-a, consiste numa amostragem regular $k \times k$ com espaçamento Δ acrescentando z pontos, os quais estão localizados de forma uniformemente aleatória dentro de uma circunferência de raio $\delta = \alpha\Delta$, cujo centro será sempre um ponto da amostragem regular, escolhido de forma aleatória.

O método de amostragem *lattice plus in-fill*, apresentado na Figura 8-b, consiste também numa amostragem regular $k \times k$ com espaçamento Δ , acrescentando t esquemas amostrais regulares menores $r \times r$ neste, adicionando assim, $t(r^2 - 4)$ pontos.

Figura 8 - (a) *Lattice plus close pairs* e (b) *Lattice plus in-fill*

2.1.6 Otimização de amostras espaciais

2.1.6.1 Critério de otimização

Quando a amostragem é rápida de ser realizada e gera baixos custos, pode-se obter facilmente um grande número de amostras e assim, há pouco interesse num algoritmo de otimização da seleção de amostras. Mas, como, na prática, o número de localizações amostrais é um fator limitante, pois a amostragem, geralmente, tem um custo alto ou é demorada, há necessidade de um critério de otimização da configuração amostral que considere os custos da realização do experimento e a confiabilidade dos resultados.

A utilização de um plano de amostragem introduz um erro amostral, que pode ser a diferença entre a estatística $\hat{\theta}$ (usada como estimador do parâmetro θ de interesse da população) e o parâmetro θ de interesse na população, isto é, $\hat{\theta} - \theta$. Se a esperança matemática dessa diferença for igual a zero, isto é,

$$E[\hat{\theta} - \theta] = 0 \longrightarrow E[\hat{\theta}] - E[\theta] = 0 \longrightarrow E[\hat{\theta}] = \theta,$$

tem-se que o estimador é não viesado. Se, entretanto, $E[\hat{\theta} - \theta] \neq 0$, tem-se que o estimador erra, em média, o valor correto do parâmetro populacional e o estimador é viesado. Logo:

$$E[\hat{\theta}] = \theta + b(\theta),$$

sendo $b(\theta)$ o viés.

Como o critério de otimização é representado por uma medida, que será minimizado ou maximizado e resume o quão eficiente é o esquema amostral, então o critério de otimização da configuração amostral será aquele que especifique uma função objetivo que deverá ser minimizada ou maximizada; isto é, uma medida estatística de proximidade do estimador (ou preditor) para o parâmetro desconhecido (ou valor desconhecido). Essa medida de proximidade pode ser o Erro Quadrático Médio (EQM), que fornece informações sobre a precisão do estimador, dado por:

$$EQM(\hat{\theta}) = E[\hat{\theta} - \theta]^2 = E[\hat{\theta}^2 - 2\theta\hat{\theta} + \theta^2].$$

Utilizando as propriedades da esperança matemática e somando e subtraindo $(E[\hat{\theta}])^2$, tem-se:

$$EQM(\hat{\theta}) = E[\hat{\theta}^2] - (E[\hat{\theta}])^2 + (E[\hat{\theta}])^2 - 2\theta \cdot E[\hat{\theta}] + \theta^2,$$

sendo que os dois primeiros termos do segundo lado da igualdade correspondem à variância de $\hat{\theta}$ e os três últimos termos do segundo lado da igualdade formam um quadrado perfeito. Assim:

$$EQM(\hat{\theta}) = VAR[\hat{\theta}] + (E[\hat{\theta} - \theta])^2,$$

corresponde ao viés do estimador $\hat{\theta}$. Logo, tem-se que:

$$EQM(\hat{\theta}) = Var[\hat{\theta}] + (\text{viés}[\hat{\theta}])^2.$$

Portanto, o erro quadrático médio do estimador $\hat{\theta}$ possui dois componentes, o que significa que o estimador erra o valor do parâmetro em função da sua variância e ainda pelo fato de ser viesado, ou seja, de não acertar na média. Geralmente, procuram-se estimadores que possuem baixos valores do Erro Quadrático Médio. Quando o estimador é não viesado, o Erro Quadrático Médio é simplesmente a variância do estimador, ou seja:

$$EQM(\hat{\theta}) = Var[\hat{\theta}].$$

Outra característica desejável de um estimador é que este seja o *Melhor Estimador Linear Não Viesado*, isto é, o estimador deve ser não viesado, ser obtido por uma combinação linear das

observações da amostra e, ainda, entre todos os estimadores lineares não viesados, apresentar a menor variância.

De acordo com Cressie (1993), na amostragem espacial a função objetivo a ser minimizada dependerá do propósito do experimento, sendo os mais comuns: a estimação dos parâmetros do modelo ajustado à semivariância, a predição pontual espacial, a predição total da variável na área em estudo ou a predição média da variável em estudo. Assim, a base do método de amostragem espacial é:

- i) Escolher o parâmetro que será estimado ou os valores que serão preditos;
- ii) Escolher o estimador ou o preditor;
- iii) Escolher a medida de proximidade que será minimizada.

Além da utilização de critérios estatísticos para se encontrar um esquema de amostragem espacial, a função objetivo pode ser implementada com componentes econômicos ou de custo, cujo intuito é introduzir no processo de escolha da melhor configuração amostral espacial, um fator que considere os custos envolvidos no processo de amostragem (GROENIGEN; STEIN, 1998). Para Bras e Rodriguez-Iturbe (1976), o melhor conjunto de amostras espaciais será aquele encontrado pela minimização de uma função objetivo composta por uma medida de precisão denotada por $\delta(n, x_i)$, um parâmetro de custo $C(n, x_i)$ e um outro parâmetro C_Δ , que corresponde a uma medida de precisão equivalente a cada mudança de unidade no custo, ou seja,

$$\delta(n, x_i) + C_\Delta \cdot C(n, \mathbf{x}_i).$$

O parâmetro de mudança influencia no grau de importância do custo no modelo, isto é, se o modelo possui um parâmetro de mudança alto, então o custo é o fator dominante no modelo; mas se o parâmetro de mudança for pequeno, então o custo não é relevante no modelo. Para Stingel (2005), deve-se buscar um equilíbrio entre o custo e o nível de precisão do plano de amostragem, para que a amostra utilizada seja economicamente viável e produza resultados com um melhor nível de precisão.

2.1.6.2 Seleção de algoritmos no processo de otimização: método de seleção sequencial

Quando o objetivo na amostragem for escolher um subconjunto ótimo de n pontos amostrais localizados na região D , o método de redução da variância pode ser usado no processo de otimização, o qual visa minimizar uma função objetivo, sujeita a restrições. Esse método é dividido em duas categorias: redução local da variância e redução global da variância.

A principal característica da técnica local é que a influência de cada ponto da região D é analisada separadamente. Logo, pode ser usado quando se tem um esquema amostral já existente na

área em estudo e deseja-se eliminar ou adicionar amostras nesse esquema amostral, uma de cada vez, de tal forma a minimizar a função objetivo. A técnica de redução local também é chamada de Método de Seleção Seqüencial, cuja implementação computacional é fácil e rápida, contudo pode não gerar um esquema amostral ótimo, se as localizações iniciais da configuração amostral não forem eficientes (BANJEVIC, 2004).

2.1.6.3 Seleção de algoritmos no processo de otimização: recozimento simulado ou têmpera simulada

A técnica de redução global da variância é utilizada para as situações denominadas *design prospective* (LOPHAVEN, 2004), cujo objetivo é redefinir um esquema amostral de tamanho Λ num esquema menor de tamanho w ($\Lambda > w$), testando todas as possíveis combinações para o novo esquema amostral e calculando seus valores da função objetivo. O esquema amostral escolhido será aquele que produzir o menor valor para a função objetivo. Como serão testadas todas as possíveis combinações de amostras, sendo que o número de possíveis combinações é dado por $\frac{\Lambda!}{w!(\Lambda-w)!}$, esse método é computacionalmente exaustivo e não se pode analisar a influência de cada ponto do novo esquema amostral separadamente como no método local, obtendo apenas uma resposta geral do novo esquema amostral.

Uma técnica de inteligência artificial de busca global de soluções ótimas proposto por Kirkpatrick et al. (1983) é chamada Recozimento Simulado (SA - *Simulated Annealing*) ou Têmpera Simulada. Esse processo é baseado nos princípios da minimização numérica, introduzido por Metropolis em 1953 e faz uma analogia ao fenômeno físico da têmpera, cujo objetivo é tornar metais e vidros mais fortes, por intermédio de uma alteração em suas propriedades físicas.

No processo da têmpera, o metal ou vidro é aquecido a temperaturas elevadas, o que corresponde a um nível máximo de energia, permitindo que os átomos se movam livre e rapidamente num estado muito desordenado. Em seguida, resfria-se o material de maneira lenta e gradual, dando aos átomos uma oportunidade maior de se rearranjarem, de tal forma que ao chegar na temperatura ambiente, o material irá adquirir uma estrutura cristalina estável. O processo da têmpera termina quando o material atingir o seu ponto de solidificação, tornando-o mais forte, atingindo o estado mínimo de energia (BUSTAMANTE, 2002).

Esse algoritmo de otimização consiste num método iterativo de busca de uma solução ótima, em que a função objetivo a ser minimizada corresponde ao nível de energia do processo de recozimento.

2.1.6.3.1 Funcionamento do recozimento simulado

Em problemas de busca por um melhor esquema de amostragem espacial, inicia-se a busca de otimização considerando como solução inicial um esquema de amostragem aleatório ou um esquema de amostragem específico, levando em conta um conhecimento prévio do problema a ser minimizado (RIBAS, 2003).

A partir de $i = 0$, e em cada iteração, esse método gera aleatoriamente um único esquema de amostragem vizinho S_{i+1} da solução anterior S_i , obtido por uma estratégia de movimentação. Essa estratégia de movimentação depende da natureza do problema estudado e pode ser realizada pela perturbação aleatória, que consiste em substituir um ponto, escolhido aleatoriamente, do esquema de amostragem inicial. O novo ponto do esquema de amostragem pode ser escolhido aleatoriamente ou considerando alguma característica relevante do processo espacial.

Em seguida, calcula-se a variação do valor da função objetivo f expressa por $\Delta = f(S_{i+1}) - f(S_i)$. A nova solução S_{i+1} será aceita com probabilidade:

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

em que t_i é chamada de temperatura corrente.

Se o valor da probabilidade de aceitação da nova solução for diferente de 1, um número $r \in (0, 1)$ é escolhido aleatoriamente utilizando uma distribuição uniforme e comparado ao valor da probabilidade. Se $r < P[\text{aceitar } S_{i+1}]$, então o esquema de amostragem S_{i+1} torna-se a solução corrente do problema.

Esse processo é repetido até que um dos seguintes critérios de parada seja satisfeito: t_i seja suficientemente pequeno, o método para atingir um determinado número fixo de iterações, ou a função objetivo é pouco modificada em sucessivas iterações.

O parâmetro temperatura é quem controla a quantidade de soluções que serão aceitas no processo de otimização. Quanto maior for o valor de t_i , mais próximo de zero será o valor de $\frac{-\Delta}{t_i}$ e conseqüentemente, mais próximo de 1 será $\exp(\frac{-\Delta}{t_i})$. Isso possibilita uma maior probabilidade de se aceitar qualquer nova solução, inclusive soluções ineficientes, escapando assim, de mínimos locais. Quanto menor for o valor de t_i , a quantidade de soluções ineficientes, que serão aceitas, diminui, admitindo apenas soluções S_{i+1} com baixo valor da função objetivo, gerando o encontro com a solução ótima global (NUNES, 2003).

Segundo Bustamante (2002), o algoritmo de recozimento simulado é mais eficiente pela busca de uma solução ótima global, quando se consegue evitar soluções ótimas locais. Como é o parâmetro temperatura quem controla a quantidade de soluções que serão aceitas no processo de otimização, o

processo de recozimento simulado é iniciado com um valor alto para t_i , pois quanto maior for seu valor, mais próximo de zero será o valor de $\frac{-\Delta}{t_i}$ e consequentemente, mais próximo de 1 será $\exp(\frac{-\Delta}{t_i})$. Isso possibilita uma maior probabilidade de se aceitar qualquer nova solução, inclusive soluções ineficientes, escapando assim, de mínimos locais.

Mas, ao longo do processo o valor de t_i deverá diminuir, de modo que a quantidade de soluções ineficientes que serão aceitas seja reduzida, admitindo apenas soluções S_{i+1} com baixo valor da função objetivo, gerando o encontro com a solução ótima global (NUNES, 2003). Para isso, é implementado no processo de recozimento simulado uma agenda de resfriamento, que consiste em um mecanismo que reduza lentamente, a cada iteração, a temperatura corrente (FERRI; PICCIONI, 1992). Algumas das regras de redução da temperatura corrente que podem ser utilizadas são:

i) Geométrica:

$$t_i = \alpha \cdot t_{i-1},$$

em que $0 < \alpha < 1$ é o parâmetro de proporção da temperatura, devendo ser um número constante e próximo de um ($0,8 \leq \alpha \leq 0,95$).

ii) Aritmética:

$$t_i = t_{i-1} - \alpha,$$

em que α é o parâmetro de redução da temperatura, devendo ser um número constante e próximo de 1.

iii) Logarítmica:

$$t_i = \frac{\alpha}{\ln(1+t_{i-1})},$$

em que α é uma constante.

2.1.6.4 Seleção de algoritmos no processo de otimização: algoritmo genético

2.1.6.4.1 Introdução aos algoritmos genéticos

Os algoritmos genéticos são métodos probabilísticos de busca global, para encontrar a solução ótima para um determinado problema pela simulação da evolução de uma população de indivíduos. Baseados na teoria de Charles Darwin, a evolução natural das espécies originou-se das observações de sua jornada científica ao redor do mundo. Durante essa expedição ele verificou, na ilha de

Galápagos, que cada ilha possuía uma fauna diferenciada, principalmente em relação a pássaros, lagartos e tartarugas. Com esse estudo, sua teoria especifica que todos os seres vivos são resultados da evolução biológica, a qual provoca mudanças nas características dos indivíduos, herdáveis via material genético.

Assim, ao longo das gerações há um processo de renovação da população. Indivíduos menos adaptados são eliminados e indivíduos mais adaptados ao meio ambiente têm maior probabilidade de sobreviverem e, conseqüentemente, teriam um maior índice de reprodução, aumentando sua quantidade de descendentes a cada geração em relação aos menos adaptados.

Os algoritmos genéticos são versões computacionais simplificadas da evolução natural, cuja idéia básica é trabalhar com uma população ou conjunto de indivíduos representando soluções para um determinado problema e por meio de transformações sucessivas nesse conjunto, baseadas na seleção e evolução natural, gerar novas e melhores soluções. Ao término desse processo, a melhor solução encontrada no conjunto de indivíduos será a solução ótima para o problema.

Os algoritmos genéticos surgiram no campo da inteligência artificial na década de 60, introduzidos por John Holland em seu livro *Adaptatton in Natural and Artificial Systems*, cujo objetivo era abstrair e explicar os processos adaptativos de sistemas naturais e projetar sistemas artificiais capazes de simular mecanismos de sistemas naturais (MEDEIROS, 2002). Atualmente, sua aplicação se estende por diversas áreas do conhecimento, como por exemplo, na elaboração de modelos matemáticos para explicar certas características de doenças na epidemiologia, na avaliação de jogos de computador, ou na área da engenharia para a otimização nos tamanhos de redes de comunicação. Segundo Medeiros (2002), suas principais vantagens em relação a outros métodos de busca por soluções ótimas são as seguintes:

- i) Esse método não restringe sua busca pela solução ótima a uma única aproximação para a solução a cada iteração, mas trabalha com um conjunto de soluções aproximadas que sofrem transformações a cada iteração;
- ii) Combina regras e operadores aleatórios, como por exemplo, a escolha aleatória dos indivíduos da população inicial, que permite criar uma certa diversidade na população, com regras e operadores determinísticos, como a função de aptidão, que possibilita a convergência da busca, pois delimita as regiões mais promissoras do espaço de busca;
- iii) Não utilizam, na sua implementação, cálculos de derivadas, pois esses algoritmos usam apenas as informações contidas na função de otimização. São fáceis de serem implementados e são aplicados a uma grande quantidade de problemas aparentemente sem soluções;
- iv) São flexíveis, podendo ser combinados com outros métodos de busca por soluções ótimas (algoritmo híbrido).

2.1.6.4.2 Codificação dos indivíduos da população

A maneira como os indivíduos são representados influencia no sucesso do algoritmo genético e essa escolha depende do problema abordado. Qualquer que seja a representação empregada, ela deve ser capaz de representar todos os indivíduos no espaço de busca do problema. Os principais tipos de representação são por meio de: números binários, números reais, permutações de símbolos e símbolos repetidos.

Em problemas que envolvem, por exemplo, variáveis que assumem valores lógicos, que portanto só podem assumir dois valores, uma escolha natural para representar os indivíduos é uma sequência binária, ou seja, uma sequência de zeros e uns. Essa sequência será composta por r elementos (sub-sequências), em que cada um desses elementos representa uma variável. Esse tipo de codificação tem diversas vantagens como facilidade na implementação computacional, na execução das operações genéticas e na transformação de variáveis com valores contínuos em sequências binárias (COSTA FILHO e POPPI, 1999).

Outra forma de representação dos indivíduos é por um vetor de números reais (ponto flutuante), em que cada número real representa uma variável do problema.

2.1.6.4.3 Funcionamento de um algoritmo genético

O funcionamento do algoritmo genético inicia-se gerando uma população inicial de indivíduos. Essa população pode ser gerada aleatoriamente ou por intermédio de algum método específico, que possibilite gerar indivíduos com qualidade melhor. A função objetivo a ser otimizada é chamada de função de aptidão e calcula a qualidade de cada indivíduo na população inicial. Logo, para cada indivíduo, associa-se um valor de aptidão que indica qual a sua proximidade com a solução ótima.

Se o objetivo é minimizar, por exemplo, a média da variância da predição espacial, os indivíduos da população com os menores valores para a média da variância da predição serão os mais próximos da solução ótima. Mas se, por exemplo, o objetivo é maximizar o determinante da matriz de covariância, os indivíduos da população com os maiores valores para o determinante da matriz de covariância serão os mais próximos da solução ótima.

Depois da avaliação de todos os indivíduos, ocorre a etapa da seleção. O processo de seleção natural é a ligação dos indivíduos e seu desempenho de sobrevivência. Nessa etapa, determina-se, pela função de aptidão, quais são os melhores indivíduos que serão escolhidos para a reprodução. Assim, indivíduos com maiores valores da função de aptidão têm uma maior chance de serem selecionados para compor a nova população e os indivíduos com os menores valores da função de aptidão possuem uma pequena probabilidade de serem selecionados, sendo, geralmente, eliminados da população.

Existem diversos mecanismos de seleção. Na *seleção por roleta*, todos os indivíduos são

dispostos num círculo, em que o ângulo da área que esse indivíduo irá ocupar nesse círculo é proporcional a sua probabilidade de seleção, calculada pelo seu valor de aptidão dividido pela soma da adaptação de todos os indivíduos. Um ponto da extremidade do círculo é escolhido aleatoriamente e em seguida, é selecionado para compor a nova população, o indivíduo correspondente a região do círculo na qual esse ponto pertence. Na *seleção por torneio*, dois indivíduos são escolhidos aleatoriamente na população, e de acordo com seu valor de aptidão, o indivíduo mais adaptado será selecionado para compor a nova população. Na *seleção por ordenação*, os indivíduos são ordenados segundo seu valor de aptidão e serão selecionados para compor a nova população uma quantidade pré definida de indivíduos que obtiveram os maiores valores de adaptação.

Selecionados os indivíduos que irão compor a nova população, é na etapa da reprodução que acontece a evolução dos indivíduos, permitindo que a cada geração a população esteja diversificada, mas mantendo as características principais de adaptação de seus antecedentes. A reprodução dos indivíduos ocorre por meio de dois operadores aleatórios: cruzamento e mutação. Um esquema de um algoritmo genético é apresentado na Quadro 1.

2.1.6.4.4 Cruzamento ou recombinação genética

O cruzamento (*crossover*) ou recombinação genética é o principal mecanismo na reprodução de novos indivíduos no espaço de busca da solução ótima, o qual proporciona troca de informação genética entre indivíduos. No processo de cruzamento, escolhe-se dois indivíduos denominados *pais* e por uma recombinação genética são gerados dois novos indivíduos chamados de *filhos ou proles*. Existem diversas técnicas de recombinação genética, que dependem da forma de codificação dos indivíduos. Para indivíduos codificados em número binário, os principais tipos de recombinação genética são: cruzamento com um ponto de corte, cruzamento com dois ou mais pontos de corte, cruzamento uniforme e cruzamento por variável (SOARES, 1997).

No cruzamento com um ponto de corte, escolhe-se aleatoriamente um ponto em dois indivíduos selecionados, onde haverá troca de material genético. São então recombinadas as características dos dois indivíduos, definidas pelo ponto de corte e são gerados dois novos indivíduos. Supondo que escolhidos dois indivíduos e um ponto de corte na posição 4, tem-se:

Pai 1: 1 1 0 1 || 0 1

Pai 2: 1 0 0 1 || 1 0

Resultando:

Filho 1: 1 1 0 1 1 0

Filho 2: 1 0 0 1 0 1

Quadro 1 - Etapas de um algoritmo genético

Passo 1) Seja $i = 0$, faça:

S_0 = população inicial com N esquemas de amostragens de tamanho n

p_c = taxa de cruzamento

p_m = taxa de mutação

ns = número de indivíduos a serem selecionados

nc = número de cruzamentos realizados

Passo 2) Calcule o valor de f para cada indivíduo da população

Passo 3) Seleção:

Gere S_{i+1} e selecione os ns melhores indivíduos de S_i

Passo 4) Cruzamento:

Calcule $prob$ = valor aleatório entre 0 e 1

Se $prob \leq p_c$, faça:

Selecione de S_{i+1} : pai_1 e pai_2

Gere: $(filho_1, filho_2) = \text{Cruzamento}(pai_1, pai_2)$

Calcule $f(filho_1)$ e $f(filho_2)$

Adicione em S_{i+1} : $(filho_1, filho_2)$

Passo 5) Mutação:

Calcule $prob$ = valor aleatório entre 0 e 1

Se $prob \leq p_m$, faça:

Selecione de S_{i+1} : indivíduo $_j$

Mutação(indivíduo $_j$)

Calcule $f(indivíduo_j)$

Adicione em S_{i+1} : indivíduo $_j$

Passo 6) Faça $i = i + 1$ e retorne ao passo (3)

O cruzamento com n pontos de corte é similar ao método anterior, diferindo apenas na escolha aleatória e sem reposição dos n pontos de corte. Supondo que serão escolhidos dois indivíduos com 4 pontos de corte nas posições 4, 6, 9 e 12, tem-se que:

Pai 1: 1 0 1 0 || 0 0 || 0 0 1 || 0 0 1 || 1 1 0

Pai 2: 1 1 0 1 || 1 1 || 0 1 1 || 0 1 0 || 0 0 0

Resultando:

Filho 1: 1 0 1 0 1 1 0 0 1 0 1 0 1 1 0

Filho 2: 1 1 0 1 0 0 0 1 1 0 0 1 0 0 0

O cruzamento uniforme é capaz de escolher qualquer posição nos indivíduos para a re-

combinação genética. Esse operador utiliza uma palavra binária padrão escolhida aleatoriamente e um dos valores 0 ou 1 para designar a posição em que serão recombinaos os indivíduos. No cruzamento uniforme, é trocado apenas o bit daquela posição e não um segmento de bits. Supondo que escolhidos dois indivíduos e a palavra padrão escrita abaixo, considerando como ponto de corte o número binário 0, tem-se:

Padrão: 1 1 1 1 0 1

Pai 1: 1 1 0 1 0 1

Pai 2: 1 0 0 1 1 0

Resultando:

Filho 1: 1 1 0 1 1 1

Filho 2: 1 0 0 1 0 0

Diferenciando-se dos métodos anteriores, no cruzamento por variável escolhe-se aleatoriamente um ponto de corte para cada variável e dentro de cada variável os indivíduos pais são recombinaos. Considerando dois indivíduos escolhidos aleatoriamente com 2 variáveis e ponto de corte nas posições 2 e 3 para cada uma das variáveis respectivamente, tem-se:

Padrão: Variável 1 || Variável 2

Pai 1: 1 1 || 0 1 0 1 0 0 || 1 0

Pai 2: 1 0 || 0 1 1 0 1 0 || 0 1

Resultando:

Padrão: Variável 1 || Variável 2

Pai 1: 1 1 0 1 1 1 0 0 0 1

Pai 2: 1 0 0 1 0 0 1 0 1 0

Para indivíduos codificados em ponto flutuante, os principais tipos de recombinação genética são: cruzamento aritmético e cruzamento geométrico. Sejam dois pais representados por $\mathbf{x}_1$ e $\mathbf{x}_2$, o cruzamento aritmético é definido como uma combinação linear dos dois pais, resultando nos filhos $a\mathbf{x}_1 + (1 - a)\mathbf{x}_2$ e $a\mathbf{x}_2 + (1 - a)\mathbf{x}_1$, em que a é um número aleatório pertencente ao intervalo $[0, 1]$. No cruzamento geométrico, os filhos resultantes serão da forma $a\mathbf{x}_1 \cdot (1 - a)\mathbf{x}_2$ e $a\mathbf{x}_2 \cdot (1 - a)\mathbf{x}_1$, em que a é um número aleatório pertencente ao intervalo $[0, 1]$.

Um parâmetro importante na fase de cruzamento é a taxa de cruzamento (p_c), sendo essa a probabilidade de um indivíduo ser recombinao com outro. Uma alta taxa de cruzamento fornece uma chance maior de surgirem na população novos indivíduos com características das gerações anteriores. Porém, segundo Ribeiro Filho (2001), para uma alta taxa de cruzamento há uma elevada probabilidade

de que com o passar das gerações, a população fique estagnada em uma região de ótimo local, convergindo prematuramente para uma solução ótima local.

2.1.6.4.5 Mutação

A mutação é um operador aleatório que tem a função de introduzir características novas aos indivíduos, diversificando a população de soluções para o problema, evitando assim, a homogeneidade da população ao longo das gerações, impedindo que o espaço de busca pela melhor solução não fique estagnado a uma região com um ótimo local. Diversas formas de mutação podem ser implementadas, dependendo da codificação dos indivíduos e do problema abordado. Se a sequência que representa os indivíduos for binária, uma forma de mutação é escolher aleatoriamente uma posição na sequência e inverter seu valor. Assim, se na posição sorteada o indivíduo tem valor 1, passará a ser 0 após a operação da mutação e vice-versa.

Nos problemas que envolvem a representação dos indivíduos em ponto flutuante, os operadores de mutação mais comum são a mutação uniforme e a mutação gaussiana. Na mutação uniforme, escolhe-se aleatoriamente uma posição na sequência de um indivíduo e troca seu valor naquela posição por um número aleatório, com distribuição uniforme num intervalo pré definido $[a, b]$. Na mutação gaussiana, todos os componentes do indivíduo selecionado $\mathbf{x}$ serão modificados por meio de $\mathbf{x}' = \mathbf{x} + \mathbf{n}$, em que $\mathbf{n}$ é um vetor de números escolhidos por uma distribuição gaussiana com média 0 e desvio padrão igual a σ (RIBEIRO FILHO, 2001).

A operação de mutação é aplicada segundo uma probabilidade chamada de taxa de mutação (p_m). Não é recomendável utilizar uma elevada taxa de mutação, pois essa proporciona um considerável aumento na diversidade da população a cada geração, tornando a busca pela solução ótima altamente aleatória, impedindo a convergência do algoritmo genético.

2.1.6.5 Critério de otimização na predição espacial

2.1.6.6 Critério baseado na variância da *krigagem*

De acordo com Cressie (1993) e Groenigen e Stein (1998), nas duas técnicas citadas, anteriormente, para o processo de otimização de uma malha amostral, se o objetivo for a predição espacial, então as funções objetivos a serem minimizadas utilizadas como critério de otimização são a média da variância da *krigagem* definida pela eq. (20), dependendo se D é uma região ou um número finito de pontos.

$$\frac{1}{|D|} \int_D \sigma_e^2(x) dx \quad \text{ou} \quad \frac{1}{|D|} \sum_{i \in D} \sigma_e^2(x_i) \quad (20)$$

e o pior erro de predição definido pela eq. (21), que corresponde à máxima variância da *krigagem* definido por:

$$\max_{\mathbf{x} \in D} \sigma_e^2(x), \quad (21)$$

em que σ_e^2 é a variância da predição, descrita nas eq. (13) e (16) para a *krigagem* ordinária e *krigagem* universal, respectivamente.

Nas duas funções objetivos, é utilizado o Erro Quadrático Médio de predição definido pela variância da *krigagem* ou da predição, pois além de ser uma medida de confiabilidade das estimativas da predição, também é capaz de avaliar a adequação de uma configuração particular dos dados, sendo que se pode considerar uma precisão requerida numa situação de interesse.

Como essas funções objetivos dependem da variância da predição, que por sua vez é uma função da semivariância, tem-se então que as funções objetivos descritas são sensíveis a pequenas mudanças no semivariograma ou em outra função a ser utilizada. Assim, podem-se obter resultados viesados, quando o semivariograma não for representativo na área de interesse, resultando em estimativas de confiabilidade superestimadas ou subestimadas. Uma alternativa é a utilização da *krigagem* indicatriz como meio de predição, com Erro Quadrático Médio a ser minimizado, representado pela variância da *krigagem* indicatriz, chamada também de variância condicional, pois é uma medida condicional aos valores do conjunto de dados. A seleção das amostras espaciais pela *krigagem* indicatriz não fornece uma melhor precisão na predição espacial, quando ocorre a presença de tendência direcional dos dados (LLOYD; ATKINSON, 1999).

Como um bom esquema amostral espacial deve considerar a heterogeneidade espacial existente na área em estudo, então para áreas com sub-regiões distintas entre si, pode-se segmentar a área em estudo classificando-a em sub-regiões separadas, as quais analisadas, individualmente, podem ser consideradas estacionárias em termos de seus respectivos variogramas. Essa forma de análise permite identificar melhor a forma de variação espacial da variável de interesse, na área em estudo, para assim, construir estratégias de amostragens ótimas (LLOYD; ATKINSON, 1998).

Observando a variância da predição, tem-se que a melhor predição, isto é, a que resultará num menor valor da variância da predição, será aquela obtida pelas n amostras espaciais que possuem a menor covariância entre si e a maior covariância com o valor a ser predito. No entanto, covariância grande com o valor a ser predito é encontrada em amostras localizadas próximas ao local do valor a ser predito. Quando o valor médio de toda a área em estudo for predito, maior covariância será encontrada pelas amostras do interior ou próximas do centro da área em estudo. Por outro lado, baixa covariância entre as amostras é encontrada considerando um espaçamento das amostras de forma sistemática ou pela divisão da área em estudo em diversos pequenos estratos. Além disso, existem outros fatores que podem

influenciar na escolha do melhor esquema de amostragem para a predição espacial.

De acordo com Groenigen e Stein (1998), os fatores que podem influenciar no esquema de amostragem otimizado em relação a predição espacial são: o alcance, o modelo ajustado à função semivariância e a vizinhança da *krigagem*. Segundo esses mesmos autores, a escolha de um esquema de amostragem para um caso anisotrópico resultará em uma maior concentração de pontos na direção de maior continuidade espacial.

Lark (2002) simulou nove conjuntos de dados dentro de uma mesma região regular, obtidos por meio de variogramas com diferentes parâmetros. Em cada um dos conjuntos de dados, selecionou 49 pontos amostrais otimizados, utilizando como função objetivo a variância da *krigagem*. Concluiu que pequenos valores do alcance implicam numa melhor amostragem por agrupamentos de pontos, e para valores grandes de alcance obtém-se uma melhor amostragem regular.

No entanto, Zimmerman e Cressie (1990), por meio de estudos com simulação, concluíram que o aumento da correlação espacial e do efeito pepita têm pouca influência na escolha do esquema de amostragem ótimo, mas a escolha da função média gera uma grande influência na escolha do esquema de amostragem ótimo. Quando a média é constante, o esquema amostral ótimo tende a ser regular, mas quando a média não é constante, o esquema amostral ótimo tende a ter os pontos mais dispostos na periferia da área em estudo.

Yfantis et al. (1987) compararam a eficiência da predição espacial pela variância da *krigagem* para os esquemas amostrais sistemático quadrado, triangular eqüilateral e hexagonal e concluíram que se o variograma tem efeito pepita pequeno, então o esquema de amostragem triangular produz melhor predição da *krigagem* em relação aos esquemas amostrais quadrado e hexagonal. Quando o efeito pepita é grande e a densidade amostral é pequena, isto é, existe uma pequena quantidade de elementos amostrais, a malha hexagonal produz menor variância da *krigagem*, produzindo então, uma melhor predição espacial.

Segundo Lophaven (2004), um dos principais problemas da otimização da amostragem espacial, considerando a predição espacial como principal objetivo, é que são assumidos como conhecidos, ou bem estimados, os parâmetros do modelo ajustado à semivariância como função da distância, sem considerar a incerteza das estimativas dos parâmetros. Logo, a estimação incorreta dos parâmetros do modelo ajustado à semivariância como função da distância produz esquemas amostrais que não são ótimos. Portanto, além de considerar a eficiência da predição espacial na escolha de melhores planos de amostragem espacial, é necessário também considerar a eficiência na estimação dos parâmetros do modelo ajustado à semivariância como função da distância.

2.1.6.6.1 Critério baseado na máxima entropia

A informação é um agente dissipador de incertezas cujo objetivo é proporcionar alterações no comportamento das pessoas, reduzindo a incerteza. A mudança no comportamento do receptor de uma dada mensagem depende do quanto de novo tem a mensagem. Assim, entende-se que, quanto maior for a taxa de novidade, maior será a mudança no comportamento das pessoas, e a mensagem de maior valor será aquela que tende para o máximo de informação. Isso leva ao estabelecimento da interação entre taxa de informação e originalidade: mais informação resulta de maior originalidade, ou menor previsibilidade.

Shannon (1948) foi o primeiro a associar os conceitos de entropia e informação. Em seu modelo de comunicação, a quantidade de informação transmitida em uma mensagem é um valor matemático bem definido e mensurável, que pode ser expresso como uma função da probabilidade de que uma mensagem, de expressão verbal ou não, seja recebida. Suponha um evento E com probabilidade de ocorrência $p(E)$. Shannon (1948) definiu uma medida de informação com base na sua incerteza ou dificuldade de previsão contida no evento E . Essa medida é definida por:

$$Inf(E) = -\log(p(E)).$$

Assim, desde que $0 < p(E) \leq 1$, o logaritmo de uma fração é negativo, logo essa medida que quantifica a incerteza da informação contida no evento E é positiva. Se $p(E) = 1$, tem-se certeza da ocorrência do evento E , e como era de se esperar a quantidade de incerteza é nula. Quanto menor a probabilidade de ocorrência do evento, maior será a medida de incerteza da ocorrência do evento E .

Seja a ocorrência de N eventos $i : 1, 2, \dots, N$, com probabilidades p_i , Shannon considerou como medida de incerteza a esperança matemática desses eventos, chamada de entropia, expressa por:

$$H(X) = E[-\log f(X)].$$

Shannon (1948), considerando um sistema que envolve um sinal de entrada X que produz um sinal Y na presença de ruído, em que as variáveis X e Y podem assumir valores discretos x_i e y_j , respectivamente com probabilidades $p(x_i)$ e $p(y_j)$ e probabilidade conjunta $p(x_i, y_j)$, caracterizou uma medida de capacidade de transmitir uma informação na presença de ruído, ou seja, uma medida de redução da incerteza média concentrada em X depois de receber o sinal Y , chamada de Índice de Informação de Shannon, definida por:

$$I(X, Y) = H(X) - H(X|Y), \quad (22)$$

em que $H(X) = E[-\log f(X)]$, é chamada de entropia de X , representa uma medida de incerteza ou de variabilidade associada a X e $H(X|Y) = E[-\log f(X|Y)]$ é a entropia condicional de X dado Y .

Se a incerteza total sobre X é expressa pela entropia de sua distribuição, a informação sobre X é dada por:

$$I(X) = -H(X). \quad (23)$$

O critério baseado na máxima entropia tem como objetivo escolher esquemas amostrais de forma a maximizar o ganho de informação para a predição de locais não amostrados. Assim, considerando um processo aleatório para a variável regionalizada $Z(\mathbf{x})$, de tal forma que sua área em estudo possa ser discretizada em m localizações, a escolha da melhor configuração amostral consiste em escolher n localizações amostrais ($n < m$), de forma que a predição de toda a área, por meio dessa amostra, produzirá a máxima informação possível em relação aos $(m - n)$ pontos não amostrados. Dividindo a variável regionalizada:

$$Z(\mathbf{x}) = \begin{bmatrix} Z_1(\mathbf{x}) \\ Z_2(\mathbf{x}) \end{bmatrix},$$

em que $Z_1(\mathbf{x})$ é o conjunto de valores da variável regionalizada nas $(m - n)$ localizações não amostradas e $Z_2(\mathbf{x})$ é o conjunto de valores da variável regionalizada nas n localizações selecionadas pelo processo de otimização.

Considerando a eq. (22) e (23), a medida de ganho de informação sobre $Z_1(\mathbf{x})$ como resultado da mensuração de $Z_2(\mathbf{x})$ é dada por:

$$I(Z_1, Z_2) = I(Z_1|Z_2) - I(Z_1).$$

Logo, a melhor informação possível sobre Z_1 é encontrada pela maximização de $I(Z_1, Z_2)$. Segundo Bueso et. al (2005), se $Z \sim N_m(\boldsymbol{\mu}, \Sigma)$, utilizando a eq. (22) e considerando $H(Z_1)$ como uma constante, maximizar $I(Z_1, Z_2)$ se reduz a minimizar $H(Z_1|Z_2)$, que é equivalente a minimizar o determinante da matriz de covariância condicional de Z_1 dado Z_2 , expresso por $\Sigma_{Z_1|Z_2}$. Ou ainda:

$$\text{mín} \left[-\frac{1}{2} \log |\Sigma_{Z_1|Z_2}| \right] = \text{mín} \left[-\frac{1}{2} \log |I - R| \right],$$

em que I é a matriz identidade e R é a matriz diagonal, sendo que a diagonal é composta pela correlação canônica quadrada entre os elementos de Z_1 e Z_2 (BUESO et. al, 1999).

No entanto, segundo Bueso et. al (1999), como $H(Z_1, Z_2) = H(Z_1|Z_2) + H(Z_2)$, uma forma mais simples considerando $H(Z_1, Z_2)$ como uma constante, minimizar $H(Z_1|Z_2)$ como proposto

é equivalente a maximizar $H(Z_2)$. Considerando $Z \sim N_m(\boldsymbol{\mu}, \Sigma)$, maximizar $H(Z_2)$ é equivalente a maximizar o determinante da matriz de covariância de Z_2 , expresso por Σ_{Z_2} , ou ainda maximizar o logaritmo do determinante da matriz de covariância de Z_2 , apresentada na eq. (24):

$$\text{máx } \log|\Sigma_{Z_2}|. \quad (24)$$

2.1.6.7 Critério de otimização baseado em medidas geométricas

Independente da função de covariância assumida, um critério de otimalidade pode ser baseado em medidas geométricas. Chen e Sudjianto (2002) propuseram a Técnica da Máxima Distância que consiste em adicionar amostras a uma malha já existente, de tal forma que maximizem a menor distância entre os pontos candidatos e os pontos existentes na configuração inicial. Logo, considerando $\mathbf{X} = \{X_i\}$, em que $i = 1, \dots, n$, como o conjunto de pontos já existentes na configuração inicial e $\mathbf{D} = \{D_j\}$, em que $j = 1, \dots, m$, o conjunto de pontos candidatos, um ponto $D_k \in \mathbf{D}$ será selecionado se esse maximizar:

$$\text{mín } [d(D_j, X_i)].$$

Outra técnica, chamada de (*Space Filling Designs*), tem como objetivo avaliar a eficiência da configuração amostral na cobertura da região de interesse. Por esse método de escolha do melhor plano de amostragem, a função objetivo a ser utilizada estará apenas em função da distância entre as localizações amostrais e os pontos candidatos ao esquema de amostragem.

Nessa técnica, calcula-se a distância entre cada ponto $\mathbf{x}$ de uma região de interesse R e o conjunto de pontos candidatos D do esquema de amostragem, expressa por:

$$d_p(\mathbf{x}, D) = \left[\sum_{\mathbf{u} \in D} \|\mathbf{x} - \mathbf{u}\|^p \right]^{(1/p)},$$

em que p é uma constante previamente definida e considerando pontos bi-dimensionais $\mathbf{x} = (x_1, x_2)$, $\|\mathbf{x}\| = \sqrt{x_1^2 + x_2^2}$ representa a distância euclidiana.

Essa medida informa o quão eficiente é o esquema de amostragem na cobertura do ponto $\mathbf{x}$. Assim, um esquema de amostragem ótimo na cobertura de uma região de interesse, será aquele que minimizar a seguinte função objetivo:

$$C_{p,q}(D) = \left[\sum_{\mathbf{x} \in R} d_p(\mathbf{x}, D)^q \right]^{(1/q)}, \quad (25)$$

em que $q > 0$ é uma constante previamente definida.

A medida geométrica $d_p(\mathbf{x}, D)$ é uma medida similar à variância de predição e quando $p \rightarrow -\infty$ e $q \rightarrow \infty$ a função objetivo $C_{p,q}(D)$ converge à escolha do esquema amostral que minimiza a distância máxima entre os pontos candidatos em D .

Considerando $p = -1$ e $q = 1$ na eq. (25), tem-se que a função objetivo a ser minimizada é dada por:

$$C_{p,q}(D) = \sum_{\mathbf{x} \in R} \left[\sum_{\mathbf{u} \in D} \frac{1}{\|\mathbf{x} - \mathbf{u}\|} \right]. \quad (26)$$

Refletindo o parâmetro de correlação ϕ e o patamar σ^2 , associado a um processo espacial, por exemplo, um modelo exponencial de covariância isotrópico, a função objetivo descrita na eq. (26) a ser minimizada, baseada na geometria da configuração amostral, será:

$$C_{p,q}(D) = \left(\frac{1}{\sigma^2}\right) \sum_{\mathbf{x} \in R} \left[\sum_{\mathbf{u} \in D} \frac{1}{1 - e^{-\|\mathbf{x} - \mathbf{u}\|/\phi}} \right].$$

2.1.6.8 Critério de otimização na estimação dos parâmetros do modelo ajustado à função semivariância

Como a qualidade do ajuste do modelo do semivariograma depende, principalmente, da configuração amostral e do tamanho amostral, então o objetivo na amostragem pode ser escolher configurações amostrais que possibilitem uma maior precisão nas estimativas dos parâmetros $\phi = (C_0, C_0 + C_1 = \sigma^2, a)$ do modelo ajustado à função semivariância. Tal escolha depende do estimador utilizado e de uma medida de qualidade do estimador. O vetor de parâmetro ϕ do modelo ajustado à função semivariância pode ser estimado por máxima verossimilhança (LARK, 2002), mínimos quadrados generalizados não linear (MULLER; ZIMMERMAN, 1999) e máxima verossimilhança restrita.

De acordo com Lark (2002), uma medida utilizada para quantificar a eficiência do estimador do vetor de parâmetros ϕ é a matriz de covariância, que sob condições de regularidade e para grandes amostras, é uma aproximação do determinante do EQM do estimador. Assim, a incerteza das estimativas de ϕ é quantificada pela matriz de covariância, obtida pela inversão da matriz de informação de Fisher ($\mathbf{I}_\phi$) (KITANIDIS, 1987), cujo ij -ésimo elemento é dado por:

$$(\mathbf{I}_\phi)_{ij} = \frac{1}{2} \text{Tr}[A_{ij}^{-1}(\phi) \cdot A_i \cdot A_{ij}^{-1}(\phi) \cdot A_j],$$

sendo que $A_{ij}^{-1}(\phi)$ é a inversa da matriz de correlação e $A_i = \frac{\partial A_{ij}}{\partial \phi_i}$.

Esse mesmo autor, baseado no critério de *D-otimalidade* (que utiliza como função objetivo a ser minimizada ou maximizada, o determinante da função de proximidade do estimador para o parâmetro desconhecido), considera como função objetivo o determinante do inverso da matriz de informação de Fisher, descrito na eq. (22), utilizada para otimizar uma configuração amostral cujo propósito é obter

melhores estimativas dos parâmetros do modelo ajustado à função semivariância. Minimizar o determinante do inverso da matriz de informação de Fisher é similar a maximizar o determinante da matriz de informação de Fisher.

$$\delta(n, \mathbf{x}_i) = |(\mathbf{I}_\phi)^{-1}|.$$

Zimmerman (2004), também baseado no critério de *D-otimalidade*, utilizou, como critério de otimização, maximizar o determinante da matriz de informação associada com o estimador de mínimos quadrados generalizados não linear dos parâmetros. Concluiu que, considerando apenas o caso em que a função média é constante e efeito pepita igual a zero, tem-se que quando o parâmetro de correlação espacial é pequeno, os pontos na configuração amostral ficam muito próximos, gerando um conglomerado compacto, mas conforme se aumenta o parâmetro de correlação espacial, os pontos tendem a se aproximar de uma linha reta. Quando se tem um moderado efeito pepita, os pontos na configuração amostral não formam conglomerados e se afastam um pouco da linha reta.

2.2 Material e métodos

2.2.1 Simulações

A primeira etapa do trabalho foi realizada por meio de simulações, em que os conjuntos de dados utilizados foram simulados por um experimento de Monte Carlo, ou seja, a partir de simulações que representam realizações de processos estocásticos multivariados, assumindo variáveis estacionárias Gaussianas, sem tendência direcional, modelo exponencial e isotrópico, com limite máximo nas coordenadas x e y igual a 1. Foram utilizados 5 conjuntos de dados resultantes da combinação dos seguintes fatores determinados previamente:

Tabela 1 - Parâmetros dos modelos simulados

Alcance (a)	Efeito Pepita (τ^2)	Patamar ($\tau^2 + \sigma^2$)
0,9	0,0	1,0
0,6	0,0	1,0
0,3	0,0	1,0
0,6	0,2	1,0
0,6	0,5	1,0

Conforme Cressie (1993), cada conjunto de dados simulados $Z(\mathbf{x}) = Z(x_1), Z(x_2), \dots, Z(x_n)$ foi obtido pela relação:

$$\mathbf{Z} = \boldsymbol{\mu} + \mathbf{L}\boldsymbol{\varepsilon}$$

sendo que $\boldsymbol{\mu}$ é o vetor de médias, $\boldsymbol{\varepsilon}$ é o vetor de variáveis aleatórias e $\mathbf{L}$ é uma matriz triangular inferior obtida pela decomposição de Cholesky tal que:

$$\boldsymbol{\Sigma} = \mathbf{L}\mathbf{L}'$$

em que $\boldsymbol{\Sigma}$ é a matriz de covariância simétrica e definida positiva.

Com o objetivo de escolher uma melhor configuração amostral reduzida para um conjunto amostral já existente, considerou-se que essas simulações, realizadas em uma malha amostral regular 20×20 composta por 400 pontos amostrais representaram o conjunto de dados inicial. Essa malha inicial pode ser tida como uma discretização da área, já que a quantidade de pontos nessa área é infinita, ou pode ser considerada como um esquema amostral já existente em que o pesquisador, por algum motivo específico deseja reduzir o tamanho amostral. A configuração inicial foi reduzida para um conjunto de 40 e 196 pontos, que representam respectivamente 10% e 49%.

A redução do tamanho amostral realizou-se pelos seguintes métodos de otimização:

i) **Recozimento Simulado - RS:**

O esquema do processo de otimização pelo método de recozimento simulado utilizado nesse trabalho é descrito na Quadro 2. Tomou-se como amostra inicial S_0 , uma malha regular, com tamanhos iguais a 40 e 196, respectivamente e, considerando 500 iterações, iniciou-se o processo do recozimento simulado com temperatura inicial igual a 8 e agenda de resfriamento geométrica com parâmetro de proporção igual a 0,85. Em cada iteração, foi obtido um esquema de amostragem vizinho S_{i+1} , substituindo um ponto do esquema de amostragem anterior, escolhido aleatoriamente, por outro ponto escolhido de forma aleatória. Para os dois esquemas de amostragem, a função objetivo definida pela média da variância de predição foi calculada. Se o esquema S_{i+1} obtivesse a menor média da variância da predição, era considerado como o melhor esquema de amostragem, e era o esquema de amostragem inicial da próxima iteração. Caso contrário, esse esquema era considerado como o melhor, segundo uma probabilidade de aceitação.

ii) **Algoritmo Genético Híbrido - AGH:**

Considerou-se uma população com 100 conjuntos amostrais iniciais S_0 (chamados de indivíduos), escolhidos aleatoriamente, com tamanhos iguais a 40 e 196, respectivamente. Em cada conjunto amostral, as amostras foram escolhidas aleatoriamente. Esses conjuntos amostrais foram avaliados pela função objetivo definida pela média da variância da predição espacial e os 50 conjuntos amostrais que apresentaram os menores valores para a função objetivo foram selecionados para compor a nova população. Para facilitar o processo de cruzamento, os indivíduos que compõem a população foram codificados da seguinte forma:

Enumerou-se, de um a cem, os 400 pontos amostrais da malha simulada regular inicial 20×20 . Cada amostra da população foi comparada com a malha regular inicial e construído um vetor de tamanho 400 com 360 zeros, nas posições em que há a ausência de ponto amostral e 40 uns (para o tamanho amostral igual a 40), nas posições em que há presença de ponto amostral. Para o tamanho amostral igual a 196, o vetor de tamanho 400 construído foi composto por 204 zeros e 196 uns.

Para os indivíduos selecionados, escolheu-se aleatoriamente dois indivíduos (pais). Pelo cruzamento com um ponto de corte, escolhido de forma aleatória, foram gerados dois novos indivíduos (filhos) e para estes foi calculada a média da variância da predição espacial. O filho que obteve o menor valor para a média da variância de predição foi comparado com o melhor pai. Se o filho apresentou o melhor resultado para a média da variância de predição, ele foi selecionado para ser otimizado pelo algoritmo de recozimento simulado, descrito, anteriormente, usando 300 iterações. Se o pai apresentou o melhor valor da média da variância da predição, então o filho foi escolhido ou não de acordo com uma probabilidade de aceitação. Nos cruzamentos iniciais, a probabilidade de aceitação é grande, mas a cada geração, a probabilidade de aceitação diminui, de tal forma que apenas filhos melhores serão otimizados no processo de recozimento simulado.

Pelo algoritmo de recozimento simulado, gerou-se um novo indivíduo com características de seus antecedentes (pais), mas com algumas modificações que possibilitaram seu aprimoramento em relação à média da variância de predição. Esse filho melhorado pelo processo de recozimento simulado foi adicionado na população e em seguida, retirou-se da população o indivíduo com o maior valor da média da variância de predição.

Esse processo de reprodução foi repetido 60 vezes, gerando ao final, uma população com 50 indivíduos, sendo que desses, o que obteve o menor valor da média da variância de predição, foi considerado o esquema de amostragem otimizado. O esquema do algoritmo genético híbrido utilizado nesse trabalho é apresentado na Quadro 3.

Para cada um dos 5 conjuntos de dados propostos na Tabela 1, foram realizadas 10 simulações cujo objetivo foi identificar se ocorreram resultados consistentes, isto é, se os resultados obtidos através da malha amostral otimizada e o valor final da média da variância de predição produziram

resultados similares.

As malhas amostrais otimizadas selecionadas pelos dois métodos de otimização foram comparadas com esquemas de amostragens encontrados na literatura, tais como: amostragem aleatória (Alea), sistemática (Sis), amostragem sistemática centrada adicionadas de pontos próximos (*lattice plus close pairs* - Prox) e amostragem sistemática centrada adicionada de esquemas amostrais sistemáticos centrados menores (*lattice plus in-fill* - Infill).

Além da média da variância de predição, as comparações entre os esquemas de amostragens e os dois métodos de otimização foram feitas pelas estimativas dos parâmetros do modelo ajustado à semivariância como função da distância e de características de interesse da variável regionalizada na área em estudo: porcentagem da área onde a variável é maior que o terceiro quartil, porcentagem da área, soma total da área onde a variável é maior que o terceiro quartil e soma total da área onde a variável é maior que o percentil 90. Em cada simulação, as estimativas das características de interesse relacionadas a predição espacial foram calculadas por 500 simulações condicionais dos valores preditos em cada ponto da malha inicial. Em cada uma das simulações condicionais, calculou-se, também, a soma quadrada dos erros, medida que possibilita identificar a proximidade da estimativa com seu valor real.

Quadro 2 - Algoritmo do processo de recozimento simulado

Passo 1) Seja $i = 0$, faça:

S_i = esquema de amostragem inicial de tamanho n

$t_0 = 2$ é a temperatura inicial

$\alpha = 0,85$ é o fator de redução da temperatura

$ni = 300$ é o número de iterações

Passo 2) Gere aleatoriamente S_{i+1}

Passo 3) Calcule $\Delta = f(S_{i+1}) - f(S_i)$

Passo 4) Se $\Delta \leq 0$, escolha S_{i+1}

Passo 5) Se $\Delta > 0$:

Calcule $P = P[\text{aceitar } S_{i+1}] = \exp(\frac{-\Delta}{t_i})$

Escolha aleatoriamente $r \in U(0, 1)$

Escolha S_{i+1} se $r < P[\text{aceitar } S_{i+1}]$

Caso contrário, escolha S_i

Passo 6) Faça $t_{i+1} = \alpha \cdot t_i$

Passo 7) Faça $i = i + 1$ até $i = ni$ e retorne ao passo (2)

Quadro 3 - Etapas de um algoritmo genético híbrido

Passo 1) Seja $j = 0$, faça:

P_j = população inicial com 100 esquemas de amostragens de tamanho 40 (ou 25)

$ns = 50$ é o número de indivíduos a serem selecionados

$nc = 60$ é o número de cruzamentos realizados

$t_0 = 8$ é a temperatura inicial

$\alpha = 0,85$ é o fator de redução da temperatura

$ni = 300$ é o número de iterações do recozimento simulado

Passo 2) Calcule o valor de f para cada indivíduo da população

Passo 3) Seleção:

Selecione os ns melhores indivíduos de P_j

Passo 4) Cruzamento: faça:

Selecione de P_j aleatoriamente: pai_1 e pai_2

Gere: $(filho_1, filho_2) = \text{Cruzamento}(pai_1, pai_2)$

Calcule $f(filho_1), f(filho_2), f(pai_1), f(pai_2)$

Calcule $\Delta = \min[f(filho_1), f(filho_2)] - \min[f(pai_1), f(pai_2)]$, se $\Delta > 0$:

Calcule $P = P[\text{aceitar } filho] = \exp(\frac{-\Delta}{t_i})$

Escolha aleatoriamente $r \in U(0, 1)$

Escolha $filho$ se $r < P[\text{aceitar } filho]$

Caso contrário, faça $filho = pai$

Passo 5) Recozimento Simulado: para $i = 0$ faça:

Selecione do passo (4) $filho = S_i$ com melhor valor da função objetivo

Gere S_{i+1} e calcule $\Delta = f(S_{i+1}) - f(S_i)$

Se $\Delta \leq 0$, escolha S_{i+1}

Se $\Delta > 0$:

Calcule $P = P[\text{aceitar } S_{i+1}] = \exp(\frac{-\Delta}{t_i})$

Escolha aleatoriamente $r \in U(0, 1)$

Escolha S_{i+1} se $r < P[\text{aceitar } S_{i+1}]$

Caso contrário, escolha S_i

Faça $t_{i+1} = \alpha \cdot t_i$

Faça $i = i + 1$ até $i = ni$ e retorne ao passo (*)

Passo 6) Adicione S_{i+1} na população P_j

Passo 7) Elimine de P_j o indivíduo com pior valor da função objetivo

Passo 8) Faça $j = j + 1$ até nc e volte ao passo (4)

2.2.2 Área experimental e variáveis em estudo

A segunda etapa do trabalho foi realizada por meio de um conjunto de dados pertencente ao banco de dados dos projetos desenvolvidos pelo Núcleo de Inovações Tecnológicas (NIT) da Universidade Estadual do Oeste do Paraná (UNIOESTE), proveniente de uma área experimental, localizada no Centro de Pesquisa Eloy Gomes da Cooperativa Central Agropecuária de Desenvolvimento Tecnológico e Econômico Ltda (COODETEC), com 1,33 ha, onde o solo é classificado como Latossolo Vermelho Distoférico. Nessa área, foi cultivada soja e os dados coletados foram produtividade e às características físico-químicas do solo, coletados no ano agrícola 1997/1998. A malha é irregular com 256 pontos amostrais, onde 128 parcelas têm um sistema de manejo localizado e 128 parcelas sem manejo localizado. O esquema de amostragem utilizado nesta área é denominado de amostragem desalinhada sistemática estratificada (WOLLENHAUPT; WOLICOWSKI, 1994 e SOUZA et al., 1999).

Selecionou-se desse conjunto de dados como aplicação do método a variável teor de potássio (K) como variável de interesse na eficiência da predição espacial com a redução do número de amostras. A escolha dessa variável para análise se baseia no fato de que ela representa o segundo nutriente mais exigido pela soja, influenciando diretamente o nível de produtividade e, conseqüentemente, o lucro do agricultor e que além disso, pela análise geoestatística, o comportamento dessa variável apresentou-se o mais similar às condições propostas nas simulações realizadas.

Com base nestes resultados, usaram-se os algoritmos de recozimento simulado e genético híbrido citados anteriormente, considerando como função objetivo a média da variância de predição espacial. Além disso, também foram obtidas amostragens aleatória e sistemática. Esses esquemas de amostragem foram comparados pelas estimativas obtidas para o modelo ajustado, média da variância de predição e quantidade e porcentagem de valores preditos que estão inseridos em cada intervalo de classificação do teor de potássio para a adubação de fertilizantes potássicos de áreas com cultivo de soja no Paraná (EMATER, 1998).

3 RESULTADOS E DISCUSSÕES

3.1 Resultados das simulações

Nas Figuras 19 a 22 são apresentados todos os esquemas de amostragem resultantes dos processos de otimização por recozimento simulado e algoritmo genético híbrido, com 40 e 196 pontos amostrais, obtidos dos 400 pontos simulados utilizando os modelos descritos na Tabela 1. Observa-se, por esses esquemas de amostragem, que nos dois processos de otimização, os pontos escolhidos para compor a amostra estão distantes ao máximo entre si, cobrindo da melhor forma possível a área.

Como se pode observar pelos gráficos de dispersão da ordem das iterações do algoritmo de recozimento simulado e algoritmo genético híbrido, versus a média da variância da predição espacial, descritos nas Figuras 11 a 14 e, 27 a 38, em todas as simulações esses processos de otimização buscaram o menor valor possível para a média da variância da predição espacial, sendo que na maioria das simulações, o algoritmo de recozimento simulado encontra o menor valor global da média da variância da predição com 300 iterações nas amostragens de tamanho 40 e com 600 iterações nas amostragens com tamanho 196. Em ambos os tamanhos amostrais, na maioria das simulações, o algoritmo genético híbrido encontra o menor valor global com 20 ou 40 iterações (cruzamentos).

Considerando os conjuntos de dados simulados com o modelo exponencial e efeito pepita igual a 0, patamar igual a 1 e alcance igual a 0,90, são apresentadas, na Tabela 5, as médias das estimativas do coeficiente de efeito pepita, alcance, patamar, além das seguintes medidas de comparação da eficiência da predição espacial: média da variância da *krigagem* ($\overline{\sigma_0^2}$), porcentagem média e a média da soma dos valores preditos que estão acima do terceiro quartil e do percentil 90. Todas essas medidas de comparação da eficiência da predição espacial foram calculadas, utilizando na predição espacial as estimativas dos parâmetros do modelo.

Por meio dessas estatísticas, verifica-se que as estimativas médias do coeficiente de efeito pepita foram próximas de seu valor real (0,00%) na maioria das amostragens estudadas, com exceção das amostragens obtidas pelo algoritmo de recozimento simulado. Com o aumento do tamanho amostral, de 40 para 196 pontos amostrais, as estimativas médias do coeficiente de efeito pepita são ainda mais próximas do valor real em todos os esquemas de amostragem.

Comparando os resultados do coeficiente pepita desse modelo, com os resultados das simulações dos modelos descritos na Tabela 2, observa-se que com o decréscimo do alcance, as estimativas do coeficiente de efeito pepita aumentam, se afastando do seu valor real.

Para as estimativas do patamar e alcance, observa-se que em todas as amostragens e em todos os tamanhos amostrais, as estimativas médias do patamar e do alcance foram similares e próximas de seu valor real.

Tabela 2 - Médias das estimativas dos parâmetros do modelo e de algumas medidas de comparação relacionadas à predição espacial para as amostragens em estudo, considerando um modelo com alcance igual a 0,90 e efeito pepita igual a 0, em que: (% >) é a porcentagem de valores preditos acima de; (SQE) é a soma quadrada dos erros e ($\sum$ >) é a soma dos valores preditos que estão acima de

Medidas	n	Alea	Sist	Próx	Infill	RS	AGH
Coef. de efeito pepita %	40	3,37	11,94	5,40	2,63	17,85	4,60
	196	1,17	2,14	1,90	1,08	0,72	1,15
Alcance	40	0,75	0,70	0,63	0,82	0,82	0,89
	196	0,77	0,77	0,82	0,83	0,82	0,80
Patamar	40	0,73	0,83	0,75	0,76	0,78	0,80
	196	0,82	0,83	0,86	0,86	0,87	0,85
$\overline{\sigma_0^2}$	40	0,318	0,310	0,292	0,313	0,258	0,250
	196	0,091	0,085	0,092	0,104	0,078	0,076
% > do percentil 75 SQE	40	25,18	25,92	23,56	24,04	24,07	24,25
		1,86	1,68	1,53	1,35	1,19	0,84
	196	24,61	24,35	24,56	25,08	24,41	25,16
		0,15	0,18	0,26	0,24	0,23	0,10
% > do percentil 90 SQE	40	11,04	11,73	10,40	10,87	10,07	10,02
		1,06	0,89	0,65	0,74	0,39	0,39
	196	10,38	9,61	9,84	10,04	10,15	9,99
		0,13	0,18	0,10	0,14	0,06	0,06
$\sum$ > percentil 75 (108,54) SQE	40	107,82	116,15	106,40	109,44	105,82	107,70
		$58 \cdot 10^4$	$52 \cdot 10^4$	$47 \cdot 10^4$	$45 \cdot 10^4$	$28 \cdot 10^4$	$31 \cdot 10^4$
	196	108,50	104,20	106,40	109,04	108,20	101,53
		$32 \cdot 10^3$	$43 \cdot 10^3$	$32 \cdot 10^3$	$60 \cdot 10^3$	$23 \cdot 10^3$	$19 \cdot 10^3$
$\sum$ > percentil 90 (58,04) SQE	40	61,35	68,69	60,82	64,09	58,01	58,19
		$33 \cdot 10^4$	$13 \cdot 10^4$	$36 \cdot 10^4$	$41 \cdot 10^4$	$33 \cdot 10^4$	$23 \cdot 10^4$
	196	60,64	55,27	56,99	58,19	60,18	54,48
		$42 \cdot 10^3$	$49 \cdot 10^3$	$28 \cdot 10^3$	$40 \cdot 10^3$	$26 \cdot 10^3$	$22 \cdot 10^3$

As amostragens otimizadas utilizando o algoritmo de recozimento simulado e algoritmo genético híbrido, apresentaram as menores estimativas para a média da variância da predição espacial. Como era de se esperar, com o aumento do tamanho amostral, as estimativas médias para a média da variância de predição espacial diminuíram; mas as melhores estimativas para a média da variância da predição espacial ainda foram obtidas com as amostragens otimizadas.

Em todas as amostragens, com o decréscimo do alcance e/ou com o aumento do efeito pepita, as estimativas médias da média da variância da *krigagem* aumentam consideravelmente. Esses

resultados podem ser confirmados pelos gráficos boxplot da média da variância de predição espacial descritos nas Figuras 9 e 10 para as simulações com todos os modelos descritos na Tabela 2.

Contudo, mesmo com a mudança dos valores do alcance e efeito pepita nos modelos simulados, as estimativas médias da média da variância da *krigagem* continuaram menores nas amostragens obtidas pelo algoritmo genético híbrido e recozimento simulado.

As amostragens obtidas pelo algoritmo genético também apresentaram melhores resultados nas estimativas médias das porcentagens de valores preditos, que estão acima do terceiro quartil e do percentil 90, em todos os tamanhos amostrais e na maioria dos modelos simulados, com exceção das simulações no modelo com o menor valor do alcance, igual a 0,30. Pois, essas estimativas obtiveram valores médios próximos do valor real e menores valores para a soma quadrada dos erros (SQE), resultados que podem ser confirmados pelos gráficos boxplot dessas estimativas, descritos nas Figuras 39 a 46. Com o aumento do tamanho amostral, podemos observar estimativas médias mais próximas ao valor real e uma queda na SQE.

Nas estimativas da soma dos valores preditos acima do terceiro quartil e do percentil 90, todas as amostragens obtiveram estimativas médias similares à média de valores reais. Mas, as amostragens obtidas por recozimento simulado e pelo algoritmo genético apresentaram menores valores para a soma quadrada dos erros, indicando que nesses métodos de amostragens, ocorreu um maior número de simulações condicionais com o valor da soma dos valores preditos acima do terceiro quartil e do percentil 90 próximos do valor real.

No entanto, nas estimativas da soma dos valores preditos acima do percentil 90, nas amostragens com tamanho amostral igual a 196, não houve diferenças relevantes nos resultados apresentados das amostragens otimizadas com as demais amostragens.

Figura 9 - Boxplot das estimativas da média da variância de predição para as amostragens com tamanho amostral igual a 40, com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1)

Figura 10 - Boxplot das estimativas da média da variância de predição para as amostragens com tamanho amostral igual a 196, com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1)

Figura 11 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens de tamanho 40, obtidas pelo algoritmo de recozimento simulado, nas simulações dos modelos com alcance igual a 0,9, efeito pepita igual a 0 e patamar igual a 1

Figura 12 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens de tamanho 40, obtidas pelo algoritmo genético híbrido nas simulações dos modelos com alcance igual a 0,9, efeito pepita igual a 0 e patamar igual a 1

Figura 13 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens de tamanho 196, obtidas pelo algoritmo de recozimento simulado nas simulações dos modelos com alcance igual a 0,9, efeito pepita igual a 0 e patamar igual a 1

Figura 14 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens de tamanho 196, obtidas pelo algoritmo genético híbrido nas simulações dos modelos com alcance igual a 0,9, efeito pepita igual a 0 e patamar igual a 1

3.2 Resultados da análise do teor de potássio

3.2.1 Análise espacial

Pelas estatísticas descritivas apresentadas na Tabela 6, que resumem o comportamento da variável teor de potássio, tem-se que o valor da média e mediana são muito próximos, indicando uma similaridade do conjunto de dados. Observa-se, ainda, pelos valores do desvio padrão e coeficiente de variação, uma baixa dispersão e uma homogeneidade na distribuição dos valores do teor de potássio (K), sendo essas conclusões confirmadas pelos gráficos boxplot e histograma apresentados na Figuras 15-(a)

e 15-(b). Observando-se o gráfico que descreve o valor estimado para o parâmetro da transformação Box Cox, descrito na Figura 15-(c), tem-se que com 95% de confiança podemos considerar que não há a necessidade de transformação da variável teor de potássio, considerando então, que a variável teor de potássio possui distribuição normal de probabilidade.

Tabela 3 - Estatísticas Descritivas e estimativas dos parâmetros do modelo da função semi-variância para a variável teor de potássio

Estatísticas	K
Média	0,3301
Mediana	0,3150
Desvio padrão	0,0865
Coefficiente de Variação	26,21%
Alcance	56,406
Efeito Pepita	21,52%
Patamar	0.0062
$\overline{\sigma_0^2}$	0,0034

Figura 15 - (a) Boxplot do teor de potássio, (b) Histograma do potássio e (c) Gráfico para estimação do parâmetro da transformação Box Cox

Na Figura 16-(a), é apresentado o gráfico Postplot da variável teor de potássio. Nesse gráfico, podemos observar que valores mais baixos do teor de potássio estão localizados em pontos mais próximos, ocorrendo o mesmo para os valores mais altos do teor de potássio, indicando uma similaridade entre amostras mais próximas. Observa-se, ainda, que não há uma variação gradativa dos valores do teor de potássio em qualquer direção, mostrando que não há tendência direcional.

Para identificar a presença de continuidade espacial da variável teor de potássio, construiu-se o gráfico do semivariograma experimental omnidirecional juntamente com o gráfico de envelope, descrito na Figura 16-(b). Nele, nota-se a existência de continuidade espacial pois, conforme há um aumento da distância, o valor da semivariância aumenta até se aproximar de uma constante. Além disso, a região do envelope representa a região sem dependência espacial e como alguns pontos do semivariograma estão fora do envelope, têm-se evidências de que existe continuidade espacial. Na Figura

16-(c), estão descritos os gráficos dos semivariogramas direcionais nas direções 0°, 45°, 90° e 135°. Nesses gráficos observa-se que não há uma diferença significativa dos valores do alcance e patamar entre os semivariogramas, indicando ausência de anisotropia.

Identificada a presença de continuidade espacial, ajustou-se pelo método de Máxima Verossimilhança um modelo exponencial à variável teor de potássio, em que as estimativas dos parâmetros do modelo são apresentadas na Tabela 6. Por essas estimativas, tem-se que a variável teor de potássio apresenta um raio de dependência espacial igual a 56,406 metros, uma moderada dependência espacial, pois seu valor de coeficiente pepita está entre 25% e 75%. Assim, temos que 73,70% da variação total dos dados do teor de potássio é explicada pela dependência espacial.

Por meio dessas estimativas do modelo da função semivariância ajustado aos dados, foi possível construir um mapa da área em estudo que expressa o comportamento da variabilidade espacial do teor de potássio, descrito na Figura 16-(d), segundo a classificação para a adubação de fertilizantes a base de potássio no Paraná para áreas com cultivo de soja (EMATER, 1998). Nesse, nota-se que os níveis do teor de potássio na área são classificados como médio (0,21 à 0,30) e alto (acima de 0,30), necessitando, necessitando nessas regiões fazer uma adubação corretiva de K_2O^4 com 50 e 40 kg/ha respectivamente.

3.2.2 Comparação entre as amostragens com tamanho reduzido

Reduzindo as 256 amostras do teor de potássio na área para 50% que corresponde a 128 amostras, pelas amostragens aleatória, sistemática, amostragem otimizada pelo algoritmo de reconhecimento simulado e amostragem otimizada pelo algoritmo genético híbrido, construíram-se seus respectivos gráficos Postplot, apresentados na Figura 17. Pode-se observar, nesses gráficos, a disposição dos pontos na área em cada amostragem reduzida. Assim, além da ausência visual de um padrão da configuração da amostragem aleatória e uma regularidade na disposição dos pontos na amostragem sistemática. Nota-se, também, uma regularidade visual na disposição dos pontos nas amostragens otimizadas, nas quais se pode identificar uma formação de regiões com valores mais próximos e uma excelente cobertura das amostras na área, pois não há regiões da área sem pontos amostrados.

Ainda se estimaram para cada um dos conjuntos amostrados reduzidos os parâmetros do modelo exponencial da função semivariância, que são apresentados na Tabela 7. Comparando esses resultados com as estimativas do conjunto de dados inicial, temos que em todas as amostragens reduzidas obtidas, os valores do efeito pepita não foram próximos da estimativa obtida na amostra inicial. Com exceção da amostragem otimizada pelo método de reconhecimento simulado, as estimativas do alcance e patamar foram próximas do valor obtido na amostragem inicial.

Considerando, também, os dados iniciais e os dados das amostras reduzidas, realizou-se

Figura 16 - (a) Postplot de K, (b) Envelope do semivariograma omnidirecional de K, (c) Semivariogramas Direcionais de K nas direções 0°, 45°, 90° e 135° e (d) Mapa da variabilidade espacial do teor de potássio, segundo a classificação para adubação de áreas com cultivo de soja no Paraná

a predição espacial em uma mesma malha amostral, sendo que os resultados das estimativas da média da variância da predição estão descritos na Tabela 7 para as amostragens reduzidas e na Tabela 6 para a amostragem inicial. Com esses resultados, observa-se que as amostragens otimizadas produziram um menor valor da média da variância da predição espacial, inclusive em relação ao valor obtido pela amostragem inicial.

Tabela 4 - Estimativas dos parâmetros do modelo da função semivariância para o teor de potássio nas amostragens e média da variância de predição

Estatísticas	Aleatória	Sistemática	RS	AGH
Alcance	47,90	52,36	101,35	45,42
Efeito Pepita	0,00%	6,96%	0,00%	48,68%
Patamar	0,0065	0,0080	0,0047	0,0076
$\overline{\sigma_0^2}$	0,0024	0,0032	0,0008	0,0005

Ainda nessa mesma malha de predição espacial, foram classificados os valores preditos das amostragens reduzidas de acordo com a recomendação dos níveis do teor de potássio para a adubação

Figura 17 - Postplot de K nas amostragens: (a) Aleatória, (b) Sistemática, (c) Amostragem otimizada pelo algoritmo de recozimento simulado e (d) Amostragem otimizada pelo algoritmo genético híbrido

de áreas produtivas de soja no Estado do Paraná (EMATER, 1998), sendo que a recomendação de aplicação de K_2O^4 em (kg/ha) é de 90, 70, 50 e 40 para os respectivos intervalos de classificação descritos na Tabela 8. Nessa mesma tabela, são apresentadas as quantidades e porcentagens de valores preditos, nas amostragens reduzidas e na amostragem inicial, que estão em cada um dos intervalos da classificação de adubação a base de potássio. Em geral, percebe-se que as amostragens reduzidas pelos métodos de otimização, principalmente pelo processo chamado algoritmo genético, apresentaram valores da porcentagem e da quantidade de valores preditos em cada classificação mais próximos aos valores obtidos pela amostragem inicial.

Esses resultados podem ser confirmados pelos mapas da variabilidade espacial do teor de potássio, levando em conta a classificação para adubação do Paraná, apresentados nas Figuras 16-(c e d), que se mostram mais similares ao mapa de variabilidade espacial do teor de potássio obtido pelos dados da amostragem inicial, descrito na Figura 14-(d).

Tabela 5 - Estimativas do número e porcentagem dos valores preditos inseridos em cada intervalo da classificação do potássio para a adubação no Estado do Paraná (EMATER, 1998)

Estatísticas	Classificação	Conjunto inicial	Alea	Sist	RS	AGH
Número de valores preditos	$K < 0,10$	0	0	0	0	0
	$0,10 < K \leq 0,20$	0	33	21	0	0
	$0,20 < K \leq 0,30$	6830	6011	7514	7541	6916
	$K > 0,30$	10270	11015	9524	9559	10184
Porcentagem de valores preditos	$K < 0,10$	0,00	0,00	0,00	0,00	0,00
	$0,10 < K \leq 0,20$	0,00	0,19	0,12	0,00	0,00
	$0,20 < K \leq 0,30$	39,94	35,15	43,94	43,10	40,44
	$K > 0,30$	60,06	64,42	55,70	56,90	59,56

Figura 18 - Mapa que representa a variabilidade espacial de K na área, segundo a classificação para adubação de áreas com cultivo de soja no Paraná, obtidos pelas amostragens: (a) Aleatória, (b) Sistemática, (c) Amostragem otimizada pelo algoritmo de recozimento simulado e (d) Amostragem otimizada pelo algoritmo genético híbrido

4 CONCLUSÕES

As análises obtidas para as simulações mostraram que, com excessão da amostragem obtida pelo algoritmo de recozimento simulado, as amostragens apresentaram excelentes estimativas para o coeficiente de efeito pepita, em ambos os tamanhos amostrais analisados, produzindo em média, estimativas muito próximas de seu valor real e uma maior porcentagem de simulações que estimaram corretamente o coeficiente de efeito pepita. No entanto, com o aumento do alcance, as estimativas do coeficiente pepita aumentam, se afastando do seu valor real. Todas as amostragens e todos os tamanhos amostrais apresentaram estimativas de alcance e patamar similares e próximas ao seu valor real.

Considerando a predição espacial, utilizando as estimativas dos parâmetros do modelo, as amostragens otimizadas obtidas pelo recozimento simulado e algoritmo genético híbrido demonstraram as menores estimativas para a média da variância de predição e resultados mais próximos aos reais da porcentagem e soma dos valores preditos acima do terceiro quartil e do percentil 90, em ambos os tamanhos amostrais. Com o aumento do tamanho amostral, as estimativas médias da média da variância de predição diminuíram em todas as amostragens. Comparando os dois métodos de amostragem otimizada, a amostragem obtida pelo algoritmo genético híbrido apresentou os menores valores das estimativas da média da variância de predição e resultados mais próximos ao real, em relação as estimativas da porcentagem e soma dos valores preditos acima do terceiro quartil e do percentil 90.

Em relação as amostragens reduzidas, obtidas do conjunto de dados da variável potássio, observa-se que as amostragens otimizadas, principalmente a amostragem otimizada pelo algoritmo genético híbrido, apresentaram os menores valores para as estimativas da média da variância de predição, inclusive em relação ao valor dessa estimativa obtida através do conjunto inicial com 256 amostras.

Além disso, na amostragem otimizada pelo algoritmo genético híbrido, as estimativas do número e da porcentagem de valores preditos pertencentes a cada intervalo da classificação, para a determinação da quantidade de produto a ser aplicado em cada localização amostral, foram mais próximas aos valores obtidos no conjunto de dados inicial com 256 amostras. Esses resultados evidenciam que a amostragem otimizada pelo algoritmo genético híbrido, com 50% das amostras, pode ser utilizada no lugar do conjunto de amostras inicial, reduzindo à metade o custo com análise de solo e refletindo pouca influência nos resultados, quanto à adubação potássica.

No entanto, o algoritmo genético híbrido despende mais tempo computacional que o processo de recozimento simulado, sendo inviável sua utilização em conjuntos de dados com grande tamanho amostral. Estudos complementares são ainda necessários para que se possa atender de forma mais eficaz as exigências e restrições da implementação.

REFERÊNCIAS

- ALMEIDA, C. F. P.; RIBEIRO JUNIOR, P. J. **Estimativa da distribuição espacial de retenção de água em um solo utilizando krigagem indicatriz**. Curitiba: Laboratório de Estatística - Universidade Federal do Paraná, 1996. 36 p. Relatório Técnico.
- ANDRIOTTI, J. L. S. **Fundamentos de Estatística e Geoestatística**. Rio Grande do Sul: Unisinos, 2004. 165 p.
- BAIO, F. H. R. **Aplicação localizada de defensivos baseada na variabilidade espacial das plantas daninhas**. 2001. 113 p. Dissertação (Mestrado em Entomologia) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2001.
- BANJEVIC, M. **Optimal network designs in spatial statistics**. 2004. 119 p. Tese (Doutorado em Filosofia) - Stanford University, California, 2004.
- BELHOUSE, D. R. Some optimal designs for sampling in two dimensions. **Biometrika**, Oxford, v. 64, n. 3, p. 605-611, dec. 1977.
- BRAS, R. L.; RODRIGUEZ-ITURBE, I. Network design for the estimation of areal mean of Rainfall Events. **Water Resources Research**, Washington, v. 12, n. 6, p. 1185-1195, dec. 1976.
- BUESO, M. C.; ANGULO, J. M.; CRUZ-SANJULIÁN, J.; GARCÍA-ARÓSTEGUI, J. L. Optimal Spatial Sampling Design in a multivariate framework. **Mathematical Geology**, Montreal, v. 31, n. 5, p. 507-525, july. 1999.
- BUESO, M. C.; ANGULO, J. M.; ALONSO, F. J.; RUIZ-MEDINA, M. D. A study on sensitivity of spatial sampling designs to a priori discretization schemes. **Environmental Modelling e Software**, Washington, v. 20, p. 891-902, july. 2005.
- BUSTAMANTE, R. C. R. **Otimização de painéis sanduíches utilizando o método de reconhecimento simulado**. 2002. 103 p. Dissertação (Mestrado em Engenharia e tecnologias espaciais/Mecânica espacial e controle) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2002.
- CÂMARA, G.; MEDEIROS, J. S. GIS para meio ambiente. In: SISTEMA DE INFORMAÇÕES GEOGRÁFICAS, 1998, Curitiba. **Anais...** Curitiba: Sagres, 1998. p. 140-180.
- CHEN, R. J. W.; SUDJANTO, A. On sequential sampling for global metamodeling in engineering design. In: DESIGN ENGINEERING TECHNICAL CONFERENCES AND COMPUTERS AND INFORMATION IN ENGINEERING CONFERENCE, 2002, Montreal. **ASME 02: trabalhos...** Montreal: ASME, 2002, p. 1-10.
- COSTA FILHO, P. A. C.; POPPI, R. J. Algoritmo genético em química. **Química Nova**, São Paulo, v. 27, n. 5, p. 405-411, fev. 1999.
- CRESSIE, N. A. C. **Statistics for Spatial Data**. New York: A Wiley-Interscience, 1993. 900 p.
- DIGGLE, P. J.; RIBEIRO JUNIOR, P. J. **Model-based Geostatistics**. New York: Springer, 2006. 230 p.
- DUNN, R.; HARRISON, A. R. Two dimensional systematic sampling of land use. **Applied Statistics**, New York, v. 42, n. 4, p. 585-601, dec. 1993.
- EMATER. **Tabelas para transformação de resultados analíticos e interpretação de resultados**. Curitiba, 1998, 64 p.
- FERRI, M.; PICCIONI, M. Optimal selection of statistical units: An approach via simulated

annealing. **Computational Statistics e Data Analysis**, Connecticut, v. 13, n. 1, p. 47-61, jan. 1992.

GARCIA, A. G. **Variabilidade espacial de atributos físicos do solo e resposta espectral da cultura de feijão irrigado, em imagens aéreas digitais**. 1997. 78 p. Dissertação (Mestrado em Agronomia) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 1997.

GROENIGEN, J. W. V.; STEIN, A. Constrained Optimization of Spatial Sampling using Continuous Simulated Annealing. **Journal of Environmental Quality**, Madison, v. 27, n. 5, p. 1078-1086, oct. 1998.

ISAAKS, E.; SRIVASTAVA, R. M. **An Introduction to Applied Geostatistics**, New York: Oxford University Press, 1989. 553 p.

JOHANN, A. J. **Variabilidade espacial dos atributos do solo e da produtividade em uma área piloto com e sem manejo localizado de agricultura de precisão**. 2001. 89 p. Dissertação (Mestrado em Engenharia Agrícola) - Universidade Estadual do Oeste do Paraná, Cascavel, 2001.

KIRKPATRICK, S. C.; GELATT, C. D.; VECCHI, J. M. P. Optimization by simulated annealing. **Science**, Oxford, v. 220, n. 4598, p. 671-680, may. 1983.

KITANIDIS, P. K. Parametric Estimation of covariances of regionalized variables. **Water Resources Bulletin**, Middleburg, v. 23, n. 4, p. 557-567, aug. 1987.

KRISHNAIAH, P. R.; RAO, C. R. **Sampling**. Amsterdam: Elsevier Science Publishers, 1988. 450 p. (Handbook of Statistics, 6).

LARK, R. M. Optimized spatial sampling of soil for estimation of the variogram by maximum likelihood. **Geoderma**, Orlando, v. 105, n. 1, p. 49-80, jan. 2002.

LLOYD, C. D.; ATKINSON, P. M. Scale and the spatial struture of landform: optimising sampling strategies with geostatistics. In: GEOCOMPUTATION, 1998, Montreal. **Anais...** Montreal: AMCE, 1998. p. 95-112.

LLOYD, C. D.; ATKINSON, P. M. Designing optimal sampling configurations with ordinary and indicator kriging. In: GEOCOMPUTATION, 1999, Montreal. **Anais...** Montreal: AMCE, 1999. p. 115-132.

LOPHAVEN, S. **Design and analysis of environmental monitoring programs**. 2004. 190p. Tese (Ph.D. em Engenharia) - Technical University of Denmark, Kongens Lyngby, 2004.

McBRATNEY, A. B; WEBSTER, R. Chosing functions for semivariograms of soil properties and fitting then to samples estimates. **Journal of Soil Science**, Oxford, v. 37, n. 4, p. 617-639, dec. 1986.

MACHADO, A. A.; DEMÉTRIO, C. G. B.; FERREIRA, D. F.; SILVA, J. G. C. Estatística Experimental: uma abordagem no planejamento e no uso de recursos computacionais. In: REUNIÃO ANUAL DA REGIÃO BRASILEIRA DA SOCIEDADE INTERNACIONAL DE BIOMETRIA, 50; SIMPÓSIO DE ESTATÍSTICA APLICADA À EXPERIMENTAÇÃO AGRONÔMICA, 11, 2005, Londrina. **Minicurso...** Londrina: UEL, 2005, p. 25-50.

MARDIA, K. V.; MARSHALL, R. J. Maximum likelihood estimation of models for residual covariance in spatial regression. **Biometrika**, Cambridge, v. 71, n. 1, p. 135-146, apr. 1984.

MEDEIROS, F. L. L. **Algoritmo Genético Híbrido como um método de busca de estados estacionários de sistemas dinâmicos**. 2002. 191 p. Dissertação (Mestrado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2002.

- MILNE, A. The Centric Systematic Area-Sample Treated as a Random Sample. **Biometrics**, Arlington, v. 15, n. 2, p. 270-297, jun. 1959.
- MOLIN, J. P. Definição de unidades de manejo a partir de mapas de produtividade. **Revista da Sociedade Brasileira de Engenharia Agrícola**, Jaboticabal, v. 22, n. 1, p. 83-92, jan. 2002.
- MORGAN, M. **The precision-farming guide for agriculturists: An agricultural primer**. Litho: Deere e Company, 1997. 117 p.
- MULLER, W; ZIMMERMAN, D. Optimal designs for variogram estimation. **Environmetrics**, West Sussex, v. 10, n. 1, p. 23-27, jan. 1999.
- NUNES, L. M. A. F. F. **Monitoring networks optimization with simulated annealing**. 2003. 205 p. Tese (Doutorado em Ciências de Engenharia) - Instituto Superior Técnico, Universidade Técnica de Lisboa, Lisboa, 2003.
- ORTIZ, G. C. **Aplicação de Métodos Geoestatísticos para identificar a magnitude e a estrutura da variabilidade espacial de variáveis físicas do solo**. 2002. 91 p. Dissertação (Mestrado em Agronomia) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2002.
- PETTIT, A. N.; MCBRATNEY, A. B. Sampling Designs for estimating spatial variance-components. **Applied Statistics**, New York, v. 42, p. 185-209, jan. 1993.
- QUENOUILLE, M. H. Problems in plane sampling. **The Annals of Mathematical Statistics**, Ann Arbor, v. 20, n. 3, p. 335-375, sept. 1949.
- RIBAS, P. C. **Análise do uso de têmpera simulada na otimização do planejamento mestre da produção**. 2003. 132 p. Dissertação (Mestrado em Engenharia de produção) - Pontifícia Universidade Católica do Paraná, Curitiba, 2003.
- RIBEIRO FILHO, G. R. **Melhoramentos no algoritmo genético construtivo e novas aplicações em problemas de agrupamento**. 2001. 120 p. Tese (Doutorado em Computação Aplicada) - Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 2001.
- RIBEIRO JUNIOR, P. J. **Métodos Geoestatísticos no Estudo da Variabilidade Espacial de Parâmetros do solo**. 1995. 91 p. Dissertação (Mestrado em Agronomia) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 1995.
- RIPLEY, B. D. **Spatial Statistics**. New York: John Wiley and Sons, 1981. 658 p.
- SCHABENBERGER, O; PIERCE, F. J. **Contemporary Statistical Models: for the plant and soil sciences**. Ed. CRC Press, Florida, 2002. 725 p.
- SHANNON, E. C. A Mathematical Theory of Communication. **The Bell System Technical Journal**, Oxford, v.27, p. 379 - 423, oct. 1948.
- SILVA, E. A. A. **Aplicação de um estimador robusto na análise de variabilidade espacial de um latossolo roxo**. 2000. 109 p. Dissertação (Mestrado em Engenharia Agrícola) - Universidade Estadual do oeste do Paraná, Cascavel, 2000.
- SOARES, G. L. **Algoritmos Genéticos: Estudo, novas técnicas e aplicações**. 1997. 145 p. Dissertação (Mestrado em Engenharia Elétrica) - Escola de Engenharia da Universidade Federal de Minas Gerais, Belo Horizonte, 1997.

SOUZA, E. G.; RIBEIRO, S. R. A.; URIBE-OPAZO, M. A.; VILAS BOAS, M.; SILVA, M. S.; JOHANN, J. A.; MOLIN, NÓBREGA, L. H. P.; OLIVEIRA, E. F.; CARRARO, I. A. Metodologia para análise da variabilidade espacial dos atributos do solo e da produtividade em uma área piloto de agricultura de precisão. In: CONGRESSO BRASILEIRO DE ENGENHARIA AGRÍCOLA, n. 3, 1999, Poços de Caldas. **Anais...** Poços de Caldas: ASNN, 1999. p. 109-111.

STINGEL, E. **Distribuição espacial e plano de amostragem para a cigarrinha das raízes, *Mahanarva fimbriolata* (Stal, 1854), em cana-de-açúcar.** 2005. 74 p. Dissertação (Mestrado em Entomologia) - Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, 2005.

THOMPSON, S. K. **Sampling.** New York: John Wiley, 1992. 430 p.

VIEIRA, S. R. Geoestatística Aplicada á Agricultura de Precisão. In: SISTEMA DE INFORMAÇÕES GEOGRÁFICAS, 1998, Curitiba. **Anais...** Curitiba: Sagres, 1998, p. 1-48.

WACKERNAGEL, H. **Multivariate Geostatistics: an introduction with applications.** Germany: Springer, 1995. 387 p.

WOLLENHAUPT, N. C.; WOLICOWSKI, R. P. Grid soil sampling. **Norcross: Better Crops with Plant Food**, Washington, v.78, n. 4, p. 6-9, 1994.

YFANTIS, E. A.; FLATMAN, G. T.; BEHAR, J. V. Efficiency of kriging estimation for square, triangular and hexagonal grids. **Mathematical Geology**, Montreal, v. 19, n.º. 3, p. 183-205, apr. 1987.

ZIMMERMAN, D. L.; CRESSIE, N.. Mean squared prediction error in the spatial linear model with estimated covariance parameters. **Annals of the Institute for Statistical Mathematics**, Boston, v. 44, n. 1, p. 27-43, mar. 1992.

ZIMMERMAN, D. L. **Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction.** Iowa: The University of Iowa, 2004. 32 p. Relatório Técnico.

ANEXOS

ANEXO-A: Tabelas e Figuras

Tabela 6 - Médias das estimativas dos parâmetros do modelo e de algumas medidas de comparação relacionadas à predição espacial para as amostragens, considerando um modelo com alcance igual a 0,60 e efeito pepita igual a 0, em que: ($\% >$) é a porcentagem de valores preditos acima de; (SQE) é a soma quadrada dos erros e ($\sum >$) é a soma dos valores preditos que estão acima de

Medidas	n	Alea	Sist	Próx	Infill	RS	AGH
Coef. de efeito pepita %	40	4,42	17,32	6,78	4,68	10,22	7,02
	196	1,71	2,87	2,73	1,59	0,43	2,65
Alcance	40	0,54	0,61	0,55	0,50	0,71	0,60
	196	0,55	0,55	0,59	0,60	0,56	0,59
Patamar	40	0,80	0,97	0,97	0,97	0,94	0,81
	196	0,88	0,88	0,93	0,92	0,90	0,90
$\overline{\sigma_0^2}$	40	0,442	0,429	0,411	0,436	0,368	0,360
	196	0,135	0,127	0,136	0,153	0,116	0,114
% > acima do percentil 75 SQE	40	25,23	26,23	23,69	24,40	24,71	23,73
		1,92	1,90	1,67	1,59	1,02	1,08
	196	24,93	24,17	24,58	25,08	25,36	24,45
		0,16	0,18	0,22	0,30	0,16	0,12
% > acima do percentil 90 SQE	40	11,00	11,71	10,27	10,93	10,79	10,00
		1,08	0,97	0,70	0,83	0,57	0,56
	196	10,85	9,83	9,99	10,30	10,50	9,88
		0,09	0,14	0,08	0,12	0,08	0,08
$\sum >$ percentil 75 (116,53) SQE	40	117,27	126,15	113,66	118,19	118,43	111,77
		$68 \cdot 10^4$	$67 \cdot 10^4$	$59 \cdot 10^4$	$59 \cdot 10^4$	$35 \cdot 10^4$	$39 \cdot 10^4$
	196	117,79	111,68	114,44	117,25	119,38	113,64
		$39 \cdot 10^3$	$61 \cdot 10^3$	$49 \cdot 10^3$	$77 \cdot 10^3$	$37 \cdot 10^3$	$29 \cdot 10^3$
$\sum >$ percentil 90 (62,51) SQE	40	66,43	73,85	68,74	64,44	67,35	60,90
		$39 \cdot 10^4$	$29 \cdot 10^4$	$44 \cdot 10^4$	$51 \cdot 10^4$	$36 \cdot 10^4$	$29 \cdot 10^4$
	196	65,79	60,15	62,15	63,96	65,93	60,92
		$38 \cdot 10^3$	$64 \cdot 10^3$	$34 \cdot 10^3$	$55 \cdot 10^3$	$38 \cdot 10^3$	$35 \cdot 10^3$

Tabela 7 - Médias das estimativas dos parâmetros do modelo e de algumas medidas de comparação relacionadas à predição espacial para as amostragens, otimizada pelo reconhecimento simulado (RS) e algoritmo genético híbrido (AGH), considerando um modelo com alcance igual a 0,30 e efeito pepita igual a 0, em que: ($\% >$) é a porcentagem de valores preditos acima de; (SQE) é a soma quadrada dos erros e ($\sum >$) é a soma dos valores preditos que estão acima de

Medidas	n	Alea	Sist	Próx	Infill	RS	AGH
Coef. de efeito pepita %	40	18,92	30,22	12,20	16,58	22,69	20,29
	196	20,40	24,45	23,41	3,36	20,46	3,34
Alcance	40	0,37	0,34	0,31	0,32	0,38	0,35
	196	0,51	0,53	0,57	0,32	0,53	0,31
Patamar	40	0,80	0,97	0,97	0,97	0,94	0,94
	196	0,89	0,89	0,91	0,96	0,90	0,92
$\overline{\sigma_0^2}$	40	0,682	0,667	0,658	0,686	0,620	0,61
	196	0,233	0,227	0,232	0,28	0,218	0,209
% acima do percentil 75 SQE	40	25,69	27,21	23,61	24,77	24,88	25,94
		2,23	2,33	2,08	2,00	1,36	1,21
	196	24,83	24,08	24,26	24,76	24,77	25,24
		0,20	0,34	0,24	0,25	0,29	0,19
% acima do percentil 90 SQE	40	11,13	12,08	10,22	11,01	10,49	11,00
		1,19	1,32	0,85	1,10	0,64	0,60
	196	10,36	9,56	9,81	10,14	10,02	10,46
		0,10	0,13	0,10	0,11	0,16	0,14
$\sum >$ percentil 75 (123,78) SQE	40	127,13	137,62	118,09	125,28	123,03	123,92
		$86 \cdot 10^4$	$89 \cdot 10^4$	$74 \cdot 10^4$	$80 \cdot 10^4$	$53 \cdot 10^4$	$39 \cdot 10^4$
	196	119,70	113,09	115,55	123,15	118,29	126,80
		$56 \cdot 10^3$	$13 \cdot 10^4$	$73 \cdot 10^3$	$78 \cdot 10^3$	$92 \cdot 10^3$	$73 \cdot 10^3$
$\sum >$ percentil 90 (66,81) SQE	40	73,19	81,20	74,05	68,53	69,87	64,47
		$46 \cdot 10^4$	$36 \cdot 10^4$	$60 \cdot 10^4$	$68 \cdot 10^4$	$41 \cdot 10^4$	$27 \cdot 10^4$
	196	67,04	60,61	62,99	68,23	64,60	70,84
		$47 \cdot 10^3$	$71 \cdot 10^3$	$39 \cdot 10^3$	$58 \cdot 10^3$	$71 \cdot 10^3$	$71 \cdot 10^3$

Tabela 8 - Médias das estimativas dos parâmetros do modelo e de algumas medidas de comparação relacionadas à predição espacial para as amostragens, considerando um modelo com alcance igual a 0,60 e efeito pepita igual a 0,2, em que: ($\%$ $>$) é a porcentagem de valores preditos acima de; (SQE) é a soma quadrada dos erros e ($\sum >$) é a soma dos valores preditos que estão acima de

Medidas	n	Alea	Sist	Próx	Infill	RS	AGH
Coef. de efeito pepita %	40	20,51	28,88	16,76	19,44	28,58	12,55
	196	20,40	24,45	23,40	18,92	20,47	21,65
Alcance	40	0,48	0,57	0,43	0,44	0,52	0,39
	196	0,51	0,53	0,57	0,52	0,54	0,50
Patamar	40	0,74	0,93	0,90	0,91	1,00	0,83
	196	0,89	0,89	0,91	0,91	0,90	0,85
$\overline{\sigma_0^2}$	40	0,580	0,566	0,558	0,580	0,528	0,524
	196	0,233	0,227	0,232	0,244	0,218	0,216
% acima do percentil 75 SQE	40	24,82	26,34	23,57	24,45	27,42	24,86
		2,13	1,92	2,13	2,04	1,98	1,05
	196	24,83	24,08	24,26	24,46	24,73	24,76
		0,20	0,34	0,24	0,31	0,29	0,20
% acima do percentil 90 SQE	40	10,63	11,72	10,19	10,94	12,54	10,39
		1,00	1,23	0,81	1,01	1,34	0,54
	196	10,36	9,56	9,81	10,10	10,01	10,53
		0,11	0,13	0,09	0,12	0,15	0,08
$\sum >$ percentil 75 (118,52) SQE	40	117,76	129,31	115,27	121,45	137,18	122,05
		$76 \cdot 10^4$	$75 \cdot 10^4$	$74 \cdot 10^4$	$78 \cdot 10^4$	$65 \cdot 10^4$	$39 \cdot 10^4$
	196	119,70	113,09	115,55	117,42	118,29	132,01
		$57 \cdot 10^3$	$13 \cdot 10^3$	$72 \cdot 10^3$	$87 \cdot 10^3$	$93 \cdot 10^3$	$42 \cdot 10^3$
$\sum >$ percentil 90 (64,18) SQE	40	66,43	76,07	71,38	66,16	82,53	67,74
		$43 \cdot 10^4$	$52 \cdot 10^4$	$55 \cdot 10^4$	$50 \cdot 10^4$	$36 \cdot 10^4$	$29 \cdot 10^4$
	196	67,03	60,62	63,00	65,12	64,60	73,02
		$47 \cdot 10^3$	$70 \cdot 10^3$	$40 \cdot 10^3$	$58 \cdot 10^3$	$71 \cdot 10^3$	$30 \cdot 10^3$

Tabela 9 - Médias das estimativas dos parâmetros do modelo e de algumas medidas de comparação relacionadas à predição espacial para as amostragens, considerando um modelo com alcance igual a 0,60 e efeito pepita igual a 0,5, em que: ($\%$ $>$) é a porcentagem de valores preditos acima de; (SQE) é a soma quadrada dos erros e ($\sum >$) é a soma dos valores preditos que estão acima de

Medidas	n	Alea	Sist	Próx	Infill	RS	AGH
Coef. de efeito pepita %	40	68,45	63,31	30,77	54,71	43,43	45,60
	196	56,01	59,24	53,67	47,45	50,83	58,43
Alcance	40	0,21	0,41	0,35	0,35	0,50	0,20
	196	0,50	0,52	0,64	0,47	0,55	0,46
Patamar	40	0,75	0,98	0,95	0,95	1,00	0,93
	196	0,92	0,93	0,94	0,93	0,92	0,89
$\overline{\sigma_0^2}$	40	0,749	0,740	0,739	0,753	0,724	0,713
	196	0,360	0,355	0,358	0,366	0,351	0,348
% acima do percentil 75 SQE	40	24,85	26,48	23,34	24,63	25,75	25,64
		2,17	2,08	2,39	2,29	2,34	1,55
	196	25,37	24,05	24,25	24,07	25,34	24,23
		0,26	0,44	0,30	0,29	0,28	0,25
% acima do percentil 90 SQE	40	11,10	12,41	10,64	11,42	12,23	11,99
		1,04	1,15	0,94	1,23	1,58	0,86
	196	10,61	9,74	9,90	10,05	11,13	10,40
		0,11	0,16	0,18	0,20	0,19	0,10
$\sum >$ percentil 75 (122,15) SQE	40	122,24	134,10	117,31	125,73	128,39	130,97
		$76 \cdot 10^4$	$83 \cdot 10^4$	$78 \cdot 10^4$	$89 \cdot 10^4$	$83 \cdot 10^4$	$53 \cdot 10^4$
	196	125,72	116,32	118,62	119,43	124,51	123,17
		$80 \cdot 10^3$	$16 \cdot 10^4$	$10 \cdot 10^4$	$87 \cdot 10^3$	$98 \cdot 10^3$	$78 \cdot 10^3$
$\sum >$ percentil 90 (66,44) SQE	40	71,74	82,43	76,39	70,33	78,86	79,49
		$46 \cdot 10^4$	$63 \cdot 10^4$	$64 \cdot 10^4$	$58 \cdot 10^4$	$40 \cdot 10^4$	$32 \cdot 10^4$
	196	70,62	63,68	65,66	67,37	72,09	69,42
		$54 \cdot 10^3$	$92 \cdot 10^3$	$63 \cdot 10^3$	$70 \cdot 10^3$	$80 \cdot 10^3$	$59 \cdot 10^3$

Figura 19 - Gráfico da localização dos 400 pontos amostrais na área simulada ($\circ$) e localizações dos 40 pontos ($\bullet$) para as amostragens: (a) aleatória, (b) sistemática, (c) sistemática centrada adicionada de pontos próximos e (d) sistemática centrada adicionada de amostragens sistemáticas centradas menores

Figura 20 - Gráfico da localização dos 400 pontos amostrais na área simulada ($\circ$) e localizações dos 196 pontos ($\bullet$) para as amostragens: (a) aleatória, (b) sistemática, (c) sistemática centrada adicionada de pontos próximos e (d) sistemática centrada adicionada de amostragens sistemáticas centradas menores

Figura 21 - Gráfico da localização dos 400 pontos amostrais na área simulada (o) e localizações dos 40 pontos (●) para uma das amostragens obtidas nas diferentes simulações pelo algoritmo de recozimento simulado com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1)

Figura 22 - Gráfico da localização dos 400 pontos amostrais na área simulada (o) e localizações dos 40 pontos (●) para uma das amostragens obtidas nas diferentes simulações pelo algoritmo genético híbrido com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1)

Figura 23 - Gráfico da localização dos 400 pontos amostrais na área simulada (o) e localizações dos 196 pontos (●) para uma das amostragens obtidas nas diferentes simulações pelo algoritmo de recozimento simulado com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1)

Figura 24 - Gráfico da localização dos 400 pontos amostrais na área simulada ($\circ$) e localizações dos 196 pontos ($\bullet$) para uma das amostragens obtidas nas diferentes simulações pelo pelo algoritmo genético híbrido com os seguintes valores de (alcance, efeito pepita, patamar): (a) $(0,9; 0; 1)$, (b) $(0,6; 0; 1)$, (c) $(0,3; 0; 1)$, (d) $(0,6; 0,2; 1,0)$ e (e) $(0,6; 0,5; 1)$

Figura 25 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 40 pontos amostrais pelo algoritmo de recozimento simulado nas simulações dos modelos com alcance igual a 0,6, efeito pepita igual a 0 e patamar igual a 1

Figura 26 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 40 pontos amostrais pelo algoritmo de recozimento simulado nas simulações dos modelos com alcance igual a 0,3, efeito pepita igual a 0 e patamar igual a 1

Figura 27 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 40 pontos amostrais pelo algoritmo de recozimento simulado nas simulações dos modelos com alcance igual a 0,6, efeito pepita igual a 0,2 e patamar igual a 1

Figura 28 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 40 pontos amostrais pelo algoritmo de recozimento simulado nas simulações dos modelos com alcance igual a 0,6, efeito pepita igual a 0,5 e patamar igual a 1

Figura 29 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 40 pontos amostrais pelo algoritmo genético híbrido nas simulações dos modelos com alcance igual a 0,6, efeito pepita igual a 0 e patamar igual a 1

Figura 30 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 40 pontos amostrais pelo algoritmo algoritmo genético híbrido nas simulações dos modelos com alcance igual a 0,3, efeito pepita igual a 0 e patamar igual a 1

Figura 31 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 40 pontos amostrais pelo algoritmo genético híbrido nas simulações dos modelos com alcance igual a 0,6, efeito pepita igual a 0,2 e patamar igual a 1

Figura 32 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 40 pontos amostrais pelo algoritmo genético híbrido nas simulações dos modelos com alcance igual a 0,6, efeito pepita igual a 0,5 e patamar igual a 1

Figura 33 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 196 pontos amostrais pelo algoritmo de recozimento simulado nas simulações dos modelos com alcance igual a 0,6, efeito pepita igual a 0 e patamar igual a 1

Figura 34 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 196 pontos amostrais pelo algoritmo de recozimento simulado nas simulações dos modelos com alcance igual a 0,3, efeito pepita igual a 0 e patamar igual a 1

Figura 35 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 196 pontos amostrais pelo algoritmo de recozimento simulado nas simulações dos modelos com alcance igual a 0,6, efeito pepita igual a 0,2 e patamar igual a 1

Figura 36 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 196 pontos amostrais pelo algoritmo de recozimento simulado nas simulações dos modelos com alcance igual a 0,6, efeito pepita igual a 0,5 e patamar igual a 1

Figura 37 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 196 pontos amostrais pelo algoritmo genético híbrido nas simulações dos modelos com alcance igual a 0,6, efeito pepita igual a 0 e patamar igual a 1

Figura 38 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 196 pontos amostrais pelo algoritmo algoritmo genético híbrido nas simulações dos modelos com alcance igual a 0,3, efeito pepita igual a 0 e patamar igual a 1

Figura 39 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 196 pontos amostrais pelo algoritmo genético híbrido nas simulações dos modelos com alcance igual a 0,6, efeito pepita igual a 0,2 e patamar igual a 1

Figura 40 - Gráfico de dispersão da ordem das iterações versus as estimativas da média da variância de predição para as amostragens obtidas com 196 pontos amostrais pelo algoritmo genético híbrido nas simulações dos modelos com alcance igual a 0,6, efeito pepita igual a 0,5 e patamar igual a 1

Figura 41 - Boxplot das estimativas da porcentagem de valores preditos, que estão acima do terceiro quartil para as amostragens com tamanho amostral igual a 40 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1)

Figura 42 - Boxplot das estimativas da porcentagem de valores preditos, que estão acima do terceiro quartil para as amostragens com tamanho amostral igual a 196 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1)

Figura 43 - Boxplot da soma quadrada dos erros das estimativas de porcentagem de valores preditos, que estão acima do terceiro quartil para as amostragens com tamanho amostral igual a 40 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1)

Figura 44 - Boxplot da soma quadrada dos erros das estimativas de porcentagem de valores preditos, que estão acima do terceiro quartil para as amostragens com tamanho amostral igual a 196 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1)

Figura 45 - Boxplot das estimativas da porcentagem de valores preditos, que estão acima do percentil 90 para as amostragens com tamanho amostral igual a 40 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1)

Figura 46 - Boxplot das estimativas da porcentagem de valores preditos, que estão acima do percentil 90 para as amostragens com tamanho amostral igual a 196 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1)

Figura 47 - Boxplot da soma quadrada dos erros das estimativas de porcentagem de valores preditos, que estão acima do percentil 90 para as amostragens com tamanho amostral igual a 40 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1)

Figura 48 - Boxplot da soma quadrada dos erros das estimativas de porcentagem de valores preditos, que estão acima do percentil 90 para as amostragens com tamanho amostral igual a 196 e com os seguintes valores de (alcance, efeito pepita, pata-mar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1)

Figura 49 - Boxplot das estimativas da soma dos valores preditos, que estão acima do terceiro quartil para as amostragens com tamanho amostral igual a 40 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1)

Figura 50 - Boxplot das estimativas da soma dos valores preditos, que estão acima do terceiro quartil para as amostragens com tamanho amostral igual a 196 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1)

Figura 51 - Boxplot da soma quadrada dos erros das estimativas da soma dos valores preditos, que estão acima do terceiro quartil para as amostragens com tamanho amostral igual a 40 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1)

Figura 52 - Boxplot da soma quadrada dos erros das estimativas da soma dos valores preditos, que estão acima do terceiro quartil para as amostragens com tamanho amostral igual a 196 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1)

Figura 53 - Boxplot das estimativas da soma dos valores preditos, que estão acima do percentil 90 para as amostragens, com tamanho amostral igual a 40 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1)

Figura 54 - Boxplot das estimativas da soma dos valores preditos, que estão acima do percentil 90, para as amostragens com tamanho amostral igual a 196 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1)

Figura 55 - Boxplot da soma quadrada dos erros das estimativas da soma dos valores preditos, que estão acima do percentil 90, para as amostragens com tamanho amostral igual a 40 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1)

Figura 56 - Boxplot da soma quadrada dos erros das estimativas da soma dos valores preditos, que estão acima do percentil 90, para as amostragens com tamanho amostral igual a 196 e com os seguintes valores de (alcance, efeito pepita, patamar): (a) (0,9; 0; 1), (b) (0,6; 0; 1), (c) (0,3; 0; 1), (d) (0,6; 0,2; 1,0) e (e) (0,6; 0,5; 1)

ANEXO-B: Programa R utilizado na geração das amostragens em conjuntos de dados simulados

```

require(geoR)
n <- 400
al <- 0.3
nu <- 0
sil <- 1
sss <- 10
resample <- function(x, size, ...) if(length(x) <= 1) {
  if(!missing(size) && size == 0) x[FALSE] else x} else
  sample(x, size, ...)
for (jjj in 1:sss) {
  set.seed(0+jjj)
  sim <- grf(n, grid= reg, xlims = c(0, 1), ylims = c(0,1),
  nx = (n^0.5), ny = (n^0.5), cov.pars=c(sil, al), nu = nu,
  cov.model=exp)
  grid1 <- sim$coords
  t_0 <- 8          ## temperatura inicial ##
  alfa <- 0.85      ## fator de ponderação de t ##
  d <- 40           ## tamanho do conjunto de dados reduzido ##
  w <- d
  ## AMOSTRA ALEATÓRIA ##
  x <- 1:n
  a1 <- sample(x, w, replace = FALSE, prob = NULL)
  o1 <- sort(a1)
  v <- as.vector(sim$data)
  v1 <- v[o1]
  c1 <- as.matrix(sim$coords)
  co1 <- c1[,1]
  co2 <- c1[,2]
  sco1 <- co1[o1]
  sco2 <- co2[o1]
  dados <- cbind(sco1,sco2,v1)
  amo1 <- as.geodata(dados)
  amo11exp <- likfit(amo1, ini=c(0.5,0.2), nug=0)
  OC <- output.control(n.pred=500)
  kb1 <- krige.conv(amo1,loc=grid1,krige=krige.control(obj.m = amo11exp,out=OC)
  mvar1 <- mean(kb1$krige.var)
  ## porcentagem dos dados simulados estarem acima do 3° quartil ##
  c75 <- boxplot.stats(sim$data)$stats[4]
  rep <- rep(NA,dim(kb1$simulations)[2])
  repsum <- rep(NA,dim(kb1$simulations)[2])
  for(i in 1:dim(kb1$simulations)[2]){
    sim1 <- kb1$simulations[,i]
    t4175 <- table(sim1 > c75)
    p4175 <- t4175/(length(sim1)[1])
    rep[i] <- as.vector(p4175[2])
    sum75 <- sum(kb1$simul[,i][(kb1$simulations[,i] > c75) == T])
    repsum[i] <- sum75
  }
  rep[!is.na(rep) == F] <- 0
  mp751 <- mean(rep)
  sq751 <- sum((rep-0.25)^2)
  sum751 <- mean(repsum)
  sumreal75 <- sum(sim$data[(sim$data > c75) == T])
  sqsum751 <- sum((repsum - sumreal75)^2)
  ## porcentagem dos dados simulados estarem acima do percentil 90 ##
  c90 <- as.vector(quantile(sim$data, prob = seq(0,1,0.1)))[10]
  rep1 <- rep(NA,dim(kb1$simulations)[2])
  rep1sum <- rep(NA,dim(kb1$simulations)[2])

```

```

for(i in 1:dim(kb1$simulations)[2]){
  sim1 <- kb1$simulations[,i]
  t4190 <- table(sim1 > c90)
  p4190 <- t4190/(length(sim1)[1]) rep1[i] <- as.vector(p4190[2])
  sum90 <- sum(kb1$simul[,i][(kb1$simulations[,i] > c90) == T])
  replsum[i] <- sum90
}
rep1[!is.na(rep1) == F] <- 0
mp901 <- mean(rep1)
sq901 <- sum((rep1-0.10)^2)
sum901 <- mean(replsum)
sumreal90 <- sum(sim$data[(sim$data > c90) == T])
sqsum901 <- sum((replsum - sumreal90)^2)
## AMOSTRA PARES DE PONTOS PRÓXIMOS ##
m <- 20 ## valores a mais de pontos próximos ##
mm <- 15 ## nº de pontos próximos ##
alpha <- 0.12 ## raio ##
nlat <- 25 ## nº de pontos no látice 4 x 4 ##
a2 <- seq(1,(n^0.5),by=4)
reg.grid <- as.matrix(expand.grid(a2, a2)) ## grid 5 x 5 ##
matriz <- matrix(sim$data[1:n],(n^0.5),(n^0.5),byrow = TRUE)
dalat <- rep(0,nlat)
for (g in 1:nlat) {
  result <- matriz[reg.grid[g, 1], reg.grid[g, 2]]
  dalat[g] = result
} ## dados da variável no látice ##
seq <- 1:n
matrix <- matrix(seq,(n^0.5),(n^0.5),byrow = TRUE)
vecoord <- rep(0,nlat)
for (j in 1:nlat) {
  resul <- matrix[reg.grid[j, 1], reg.grid[j, 2]]
  vecoord[j] = resul
}
coordlat <- sim$coord[vecoord,]
dadoslat <- cbind(coordlat, dalat) ## matriz com coordenadas e dados do látice ##
num <- 1:nlat
mpontos <- sample(num, m, replace = FALSE, prob = NULL)
mps <- dadoslat[mpontos,] ## m pontos selecionados ##
md <- matrix(0, nrow = m, ncol = n) ## cálculo das distâncias dos m pontos
a todos os pontos do grid ##
for (r in 1:m){
  d <- ((mps[r,1] - sim$coords[,1])^2 + (mps[r,2] - sim$coords[,2])^2)^0.5
  md[r,] = d
}
md <- t(md)
tf <- matrix(0, nrow = n, ncol = m) ## quais distâncias estão entre 0 e alpha ##
numeros <- 1:n
posi <- rep(0,m)
for (z in 1:m){
  e <- md[,z] <= alpha & md[,z] > 0
  tf[,z] = e
  sa <- sample(numeros[tf[,z] == 1],1)
  posi[z] = sa ## posição dos elementos sorteados ##
}
b <- table(posi)
c <- as.vector(names(b))
dad <- cbind(sim$coords,sim$data)
dado <- dad[c,]
sort <- sample(c,mm, rep = F)
dados <- rbind(dadoslat,dado[sort,]) ## dados do látice e dos pontos próximos ##
amo2 <- as.geodata(unique(dados))
amo22exp <- likfit(amo2, ini=c(0.5, 0.2), nug=0, cov.model=exp)

```



```

kb2 <- krige.conv(amo2, loc=grid1,krige=krige.control(obj.m = amo22exp, out=0C)
mvar2 <- mean(kb2$krige.var)
## porcentagem dos dados simulados estarem acima do 3° quartil ##
c75 <- boxplot.stats(sim$data)$stats[4]
rep <- rep(NA,dim(kb2$simulations)[2])
repsum <- rep(NA,dim(kb2$simulations)[2])
for(i in 1:dim(kb2$simulations)[2]){
  sim1 <- kb2$simulations[,i]
  t4175 <- table(sim1 > c75)
  p4175 <- t4175/(length(sim1)[1])
  rep[i] <- as.vector(p4175[2])
  sum75 <- sum(kb2$simul[,i][(kb2$simulations[,i] > c75) == T])
  repsum[i] <- sum75
}
rep[!is.na(rep) == F] <- 0
mp752 <- mean(rep)
sq752 <- sum((rep-0.25)^2)
sum752 <- mean(repsum)
sumreal75 <- sum(sim$data[(sim$data > c75) == T])
sqsum752 <- sum((repsum - sumreal75)^2)
## porcentagem dos dados simulados estarem acima do percentil 90 ##
c90 <- as.vector(quantile(sim$data, prob = seq(0,1,0.1)))[10]
rep1 <- rep(NA,dim(kb2$simulations)[2])
rep1sum <- rep(NA,dim(kb2$simulations)[2])
for(i in 1:dim(kb2$simulations)[2]){
  sim1 <- kb2$simulations[,i]
  t4190 <- table(sim1 > c90)
  p4190 <- t4190/(length(sim1)[1])
  rep1[i] <- as.vector(p4190[2])
  sum90 <- sum(kb2$simul[,i][(kb2$simulations[,i] > c90) == T])
  rep1sum[i] <- sum90
} rep1[!is.na(rep1) == F] <- 0
mp902 <- mean(rep1)
sq902 <- sum((rep1-0.10)^2)
sum902 <- mean(rep1sum)
sumreal90 <- sum(sim$data[(sim$data > c90) == T])
sqsum902 <- sum((rep1sum - sumreal90)^2)
## AMOSTRA LATTICE IN FILL ##
##### lattice maior 5 x 5 #####
nlat <- 25
a2 <- seq(1,(n^0.5),by=4)
reg.grid <- as.matrix(expand.grid(a2, a2))
matriz <- matrix(sim$data[1:n],(n^0.5),(n^0.5),byrow = TRUE)
dalat <- rep(0,nlat)
for (g in 1:nlat) {
  result <- matriz[reg.grid[g, 1], reg.grid[g, 2]]
  dalat[g] = result
}
## dados da variável no látice ## seq <- 1:n
matrix <- matrix(seq,(n^0.5),(n^0.5),byrow = TRUE)
vecoord <- rep(0,nlat)
for (j in 1:nlat) {
  resul <- matrix[reg.grid[j, 1], reg.grid[j, 2]]
  vecoord[j] = resul
}
coordlat <- sim$coord[vecoord,] ## dados das coordenadas no látice ##
dadoslat <- cbind(coordlat, dalat) ## matriz com coordenadas e dados do látice ##
## Domínio do sorteio ##
dadosdomi <- dadoslat
##### sort 1 #####
for(u in 1:5){
  sample <- sample(as.numeric(row.names(dadosdomi)),1)
  ele <- c(sample+1,sample+20, sample+21)
}

```

```

simi <- cbind(sim$coords,sim$data) elesim <- simi[ele,]
dadoslat <- rbind(dadoslat,elesim)
}
amo3 <- as.geodata(dadoslat)
amo33exp <- likfit(amo3, ini=c(0.5, 0.2), nug=0, cov.model=exp)
kb3 <- krige.conv(amo3, loc=grid1,krige=krige.control(obj.m = amo33exp, out = 0C)
mvar3 <- mean(kb3$krige.var)
## porcentagem dos dados simulados estarem acima do 3° quartil ##
c75 <- boxplot.stats(sim$data)$stats[4]
rep <- rep(NA,dim(kb3$simulations)[2])
repsum <- rep(NA,dim(kb3$simulations)[2])
for(i in 1:dim(kb3$simulations)[2]){
  sim1 <- kb3$simulations[,i]
  t3175 <- table(sim1 > c75)
  p3175 <- t3175/(length(sim1)[1])
  rep[i] <- as.vector(p3175[2])
  sum75 <- sum(kb3$simul[,i][(kb3$simulations[,i] > c75) == T])
  repsum[i] <- sum75
}
rep[!is.na(rep) == F] <- 0
mp753 <- mean(rep)
sq753 <- sum((rep-0.25)^2)
sum753 <- mean(repsum)
sumreal75 <- sum(sim$data[(sim$data > c75) == T])
sqsum753 <- sum((repsum - sumreal75)^2)
## porcentagem dos dados simulados estarem acima do percentil 90 ##
c90 <- as.vector(quantile(sim$data, prob = seq(0,1,0.1)))[10]
rep1 <- rep(NA,dim(kb3$simulations)[2])
rep1sum <- rep(NA,dim(kb3$simulations)[2])
for(i in 1:dim(kb3$simulations)[2]){
  sim1 <- kb3$simulations[,i]
  t3190 <- table(sim1 > c90)
  p3190 <- t3190/(length(sim1)[1])
  rep1[i] <- as.vector(p3190[2])
  sum90 <- sum(kb3$simul[,i][(kb3$simulations[,i] > c90) == T])
  rep1sum[i] <- sum90
}
rep1[!is.na(rep1) == F] <- 0
mp903 <- mean(rep1)
sq903 <- sum((rep1-0.10)^2)
sum903 <- mean(rep1sum)
sumreal90 <- sum(sim$data[(sim$data > c90) == T])
sqsum903 <- sum((rep1sum - sumreal90)^2)
## AMOSTRA SISTEMÁTICA ##
a2 <- seq(3,(n^0.5),by=5)
a3 <- seq(1,(n^0.5),by=2)
reg.grid <- as.matrix(expand.grid(a2, a3)) ## grid 14 x 14 ##
matriz <- matrix(sim$data[1:n],(n^0.5),(n^0.5),byrow = TRUE)
dalat <- rep(0,w)
for (g in 1:w) {
  result <- matriz[reg.grid[g, 1], reg.grid[g, 2]]
  dalat[g] = result
}
## dados da variável no látice ##
seq <- 1:n
matrix <- matrix(seq,(n^0.5),(n^0.5),byrow = TRUE)
vecoord <- rep(0,w)
for (j in 1:w) {
  resul <- matrix[reg.grid[j, 1], reg.grid[j, 2]]
  vecoord[j] = resul
}
coordlat <- sim$coord[vecoord,] ## dados das coordenadas no látice ##
dados <- cbind(coordlat, dalat) ## matriz com coordenadas e dados do látice ##

```

```

amo4 <- as.geodata(dados)
amo44exp <- likfit(amo4, ini=c(0.5, 0.2), nug=0, cov.model=exp)")
kb4 <- krige.conv(amo4, loc=grid1,krige=krige.control(obj.m = amo44exp, out = 0C)
mvar4 <- mean(kb4$krige.var)
## porcentagem dos dados simulados estarem acima do 3° quartil ##
c75 <- boxplot.stats(sim$data)$stats[4]
rep <- rep(NA,dim(kb4$simulations)[2])
repsum <- rep(NA,dim(kb4$simulations)[2])
for(i in 1:dim(kb4$simulations)[2]){
  sim1 <- kb4$simulations[,i]
  t4175 <- table(sim1 > c75)
  p4175 <- t4175/(length(sim1)[1])
  rep[i] <- as.vector(p4175[2])
  sum75 <- sum(kb4$simul[,i][(kb4$simulations[,i] > c75) == T])
  repsum[i] <- sum75
}
rep[!is.na(rep) == F] <- 0
mp754 <- mean(rep)
sq754 <- sum((rep-0.25)^2)
sum754 <- mean(repsum)
sumreal75 <- sum(sim$data[(sim$data > c75) == T])
sqsum754 <- sum((repsum - sumreal75)^2)
## porcentagem dos dados simulados estarem acima do percentil 90 ##
c90 <- as.vector(quantile(sim$data, prob = seq(0,1,0.1)))[10]
rep1 <- rep(NA,dim(kb4$simulations)[2])
rep1sum <- rep(NA,dim(kb4$simulations)[2])
for(i in 1:dim(kb4$simulations)[2]){
  sim1 <- kb4$simulations[,i]
  t4190 <- table(sim1 > c90)
  p4190 <- t4190/(length(sim1)[1])
  rep1[i] <- as.vector(p4190[2])
  sum90 <- sum(kb4$simul[,i][(kb4$simulations[,i] > c90) == T])
  rep1sum[i] <- sum90
}
rep1[!is.na(rep1) == F] <- 0
mp904 <- mean(rep1)
sq904 <- sum((rep1-0.10)^2)
sum904 <- mean(rep1sum)
sumreal90 <- sum(sim$data[(sim$data > c90) == T])
sqsum904 <- sum((rep1sum - sumreal90)^2)
## RECOZIMENTO SIMULADO ##
alpha <- 0.30      ## raio ##
d <- 40
for (i in 1:500){
  z <- 1:d
  a <- sample(z, 1, replace = FALSE, prob = NULL)
  psub <- dados[a,]      ## ponto escolhido p/ ser substituído ##
  npsub <- as.integer(row.names(dados)[a])      ## n° do ponto a ser substituído ##
  o1 <- as.numeric(row.names(dados))
  simul <- cbind(sim$coords,sim$data)
  simul1 <- simul[-o1,]      ## conjunto de pontos que podem ser sorteados sem S_0 ##
  ## cálculo das distâncias ##
  di <- as.vector((((psub[1] - simul1[,1])^2) + ((psub[2] - simul1[,2])^2))^0.5)
  e <- di <= alpha & di > 0
  tf <- 1: length(di)
  qq <- tf[e == T]
  sa <- resample(qq,1)
  rr <- as.integer(row.names(simul1)[sa])
  qqqq <- row.names(dados)
  ponto <- simul1[sa,]      ## ponto sorteado a substituir em S_0 ##
  row.names(dados) <- as.integer(qqqq)
  amo11 <- as.geodata(dados)

```

```

amo11exp <- likfit(amo11, ini=c(0.5, 0.2), nug=0, cov.model=exp)")
kb11 <- krige.conv(amo11, loc=grid1,krige=krige.control(obj.m = amo11exp))
m11 <- mean(kb11$krige.var)
z <- 1:d
q <- as.integer(row.names(dados))
as <- z[q == npsub]
dado <- dados[-as,]
S_{1} <- rbind(dado, ponto) ## conjunto S_1 modificado ##
qqqq <- row.names(S_1)
qqqq <- as.integer(replace(qqqq,d,rr))
row.names(S_1) <- qqqq
amo22 <- as.geodata(S_1)
amo22exp <- likfit(amo22, ini=c(0.5, 0.2), nug=0, cov.model=exp)")
kb22 <- krige.conv(amo22, loc=grid1,krige=krige.control(obj.m = amo22exp))
m22 <- mean(kb22$krige.var)
deltam <- m22 - m11
S_0 <- dados P <- exp(-(deltam)/t_0)
runif <- runif(1000,0,1)
r <- sample(runif,1, replace = FALSE, prob = NULL)
ifelse(deltam <= 0, dados <- S_1, ifelse(r <= P, dados <- S_1, dados <- S_0))
t_0 <- alfa * t_0
write.table(t(c(m11,m22,min(m11,m22),t_0)), C:\\recozimento400nu05.txt, dec = .,
col.names = F,row.names = F, append = T)
}
amo5 <- as.geodata(dados)
amo55exp <- likfit(amo5, ini=c(0.5, 0.2), nug=0, cov.model=exp)")
kb5 <- krige.conv(amo5, loc=grid1,krige=krige.control(obj.m = amo55exp, out = 0C)
mvar5 <- mean(kb5$krige.var)
## porcentagem dos dados simulados estarem acima do 3° quartil ##
c75 <- boxplot.stats(sim$data)$stats[4]
rep <- rep(NA,dim(kb5$simulations)[2])
repsum <- rep(NA,dim(kb5$simulations)[2])
for(i in 1:dim(kb5$simulations)[2]){
  sim1 <- kb5$simulations[,i]
  t4175 <- table(sim1 > c75)
  p4175 <- t4175/(length(sim1)[1])
  rep[i] <- as.vector(p4175[2])
  sum75 <- sum(kb5$simul[,i][(kb5$simulations[,i] > c75) == T])
  repsum[i] <- sum75
}
rep[!is.na(rep) == F] <- 0
mp755 <- mean(rep)
sq755 <- sum((rep-0.25)^2)
sum755 <- mean(repsum)
sumreal75 <- sum(sim$data[(sim$data > c75) == T])
sqsum755 <- sum((repsum - sumreal75)^2)
## porcentagem dos dados simulados estarem acima do percentil 90 ##
c90 <- as.vector(quantile(sim$data, prob = seq(0,1,0.1)))[10]
rep1 <- rep(NA,dim(kb5$simulations)[2])
rep1sum <- rep(NA,dim(kb5$simulations)[2])
for(i in 1:dim(kb5$simulations)[2]){
  sim1 <- kb5$simulations[,i]
  t4190 <- table(sim1 > c90)
  p4190 <- t4190/(length(sim1)[1])
  rep1[i] <- as.vector(p4190[2])
  sum90 <- sum(kb5$simul[,i][(kb5$simulations[,i] > c90) == T])
  rep1sum[i] <- sum90
}
rep1[!is.na(rep1) == F] <- 0
mp905 <- mean(rep1)
sq905 <- sum((rep1-0.10)^2)
sum905 <- mean(rep1sum)

```

```

sumreal90 <- sum(sim$data[(sim$data > c90) == T])
sqsum905 <- sum((rep1sum - sumreal90)^2)
## PROGRAMA ALGORITMO GENÉTICO HÍBRIDO ##
## passo1: Inserir o conjunto total Z de tamanho D ##
grid1 <- sim$coords
simu <- cbind(sim$coords,sim$data)
alfa <- 0.85
t_0 <- 8
alpha <- 0.30
w <- 40      ## tamanho do conjunto de dados reduzido ##
we <- 60     ## número de iterações para o cruzamento ##
## criando os indivíduos da população com amostragem aleatória ##
## selecionando os indivíduos codificados ##
mvar11 <- matrix(rep(NA,(bb)) , ncol = 1, nrow = bb)
matrix01 <- matrix(rep(NA,(bb*n)) , ncol = bb, nrow = n)
matrixso <- matrix(rep(NA,(bb*w)) , ncol = bb, nrow = w)
for (cc in 1:bb) {
  seq0 <- rep(0,n)
  so <- sample(1:n, w, replace = FALSE, prob = NULL)
  sort <- sort(so)
  matrixso[,cc] = sort
  indi <- replace(seq0, sort, 1)
  matrix01[,cc] = indi
}
## Decodificando para a avaliação dos indivíduos e avaliação ##
decodi <- matrix(rep(NA,(bb*w*3)) , ncol = 3, nrow = bb*w)
for (cc in 1:bb) {
  indiv <- simu[matrixso[,cc],]
  decodi[(((cc-1)*w)+1):(cc*w),] <- indiv
amo11 <- as.geodata(indiv)
amo11exp <- likfit(amo11, ini=c(0.5, 0.2), nug=0, cov.model="exp")
kc11 <- krige.conv(amo11, loc=grid1,krige=krige.control(obj.m = amo11exp))
m11 <- mean(kc11$krige.var)
mvar11[cc,] <- m11
}
## seleção inicial de 50% dos indivíduos da população ##
select <- matrix(rep(NA,(bb*w/2)*3) , ncol = 3, nrow = (bb*w/2))
tablee <- decodi
vm11 <- as.vector(mvar11)
median <- median(vm11)
vm1 <- as.vector(vm11)
vm1[vm1 <= median] <- 1
z <- 1:bb
esco <- z[vm1 == 1]  ## número das amostras selecionadas ##
for (cc in 1:length(esco)) {
  as <- tablee[(1+((esco[cc]-1)*w)):(w*esco[cc]),]
  select[(((cc-1)*w)+1):(cc*w),] <- indiv
}
vm11 <- vm11[esco]
## select são os 50 dos indivíduos selecionados ##
selecao <- select      ## conjunto de dados selecionados ##
mselect <- matrix01[,esco]  ## conjunto selecionado codificado ##
## recombinação (crossover) dos indivíduos codificados ##
ti <- 2
for (ee in 1:we){
  sequencia <- 1:w
  pcorte <- sample(sequencia,1)  ## ponto de corte ##
## recombinação de pais aleatórios ##
sample <- sample(1:(dim(mselect)[2]),2)
pai1sam <- mselect[,sample[1]]    ## pai1 codificado ##
pai2sam <- mselect[,sample[2]]    ## pai2 codificado ##
filho1sam <- c(pai1sam[1:pcorte],pai2sam[(pcorte+1):n])

```

```

filho2sam <- c(pai2sam[1:pcorte],pai1sam[(pcorte+1):n])
## reparo dos filhos ##
nf1 <- (w-table(filho1sam)[2])
nf2 <- (w-table(filho2sam)[2])
qq <- 1:n
f1 <- qq[filho1sam == 0]
f2 <- qq[filho2sam == 0]
f11 <- qq[filho1sam == 1]
f21 <- qq[filho2sam == 1]
if(nf1 >= 0) filho1sam <- replace(filho1sam,sample(f1,abs(nf1)),rep(1,abs(nf1)))
filho1sam <- replace(filho1sam,sample(f11,abs(nf1)),rep(0,abs(nf1)))
else filhodeco1sam <- filho1sam
if(nf2 >= 0) filho2sam <- replace(filho2sam,sample(f2,abs(nf2)),rep(1,abs(nf2)))
else filho2sam <- replace(filho2sam,sample(f21,abs(nf2)),rep(0,abs(nf2)))
filhodeco2sam <- filho2sam
## conjunto de filhos decodificados ##
p11 <- qq[filho1sam == 1]
p21 <- qq[filho2sam == 1]
filhod1sam <- simu[p11,]
filhod2sam <- simu[p21,]
pai11 <- qq[pai1sam == 1]
pai21 <- qq[pai2sam == 1]
paid1sam <- simu[pai11,]
paid2sam <- simu[pai21,]
## avaliação dos filhos e dos pais ##
filho1sam <- as.geodata(filhod1sam)
filho2sam <- as.geodata(filhod2sam)
pai1 <- as.geodata(paid1sam)
pai2 <- as.geodata(paid2sam)
amo11exp <- likfit(filho1sam1, ini=c(0.5, 0.2), nug=0, cov.model=exp)
kb11 <- krige.conv(filho1sam, loc=grid1,krige=krige.control(obj.m = amo11exp))
mfilho1sam <- mean(kb11$krige.var)
amo22exp <- likfit(filho2sam2, ini=c(0.5, 0.2), nug=0, cov.model=exp)
kb22 <- krige.conv(filho2sam, loc=grid1,krige=krige.control(obj.m = amo22exp))
mfilho2sam <- mean(kb22$krige.var)
amo11exp <- likfit(pai1, ini=c(0.5, 0.2), nug=0, cov.model=exp)
kbp11 <- krige.conv(pai1, loc=grid1,krige=krige.control(obj.m = amo11exp))
mpai1sam <- mean(kbp11$krige.var)
amo22exp <- likfit(pai2, ini=c(0.5, 0.2), nug=0, cov.model=exp)
kbp22 <- krige.conv(pai2, loc=grid1,krige=krige.control(obj.m = amo22exp))
mpai2sam <- mean(kbp22$krige.var)
## escolha do melhor filho ##
dif <- mfilho1sam - mfilho2sam
m <- ifelse(dif <=0, filhodeco <- filhod1sam, filhodeco <- filhod2sam)
mdeco <- ifelse(dif <=0, filhoco <- filhodeco1sam, filhoco <- filhodeco2sam)
vmfilho <- ifelse(dif <=0, vmfilho <- mfilho1sam, vmfilho <- mfilho2sam)
## escolha do melhor pai ##
dif <- mpai1sam - mpai2sam
m <- ifelse(dif <=0, paico <- pai1sam, paico <- pai2sam)
mdeco <- ifelse(dif <=0, paideco <- paid1sam, paideco <- paid2sam)
vmpai <- ifelse(dif <=0, vmpai <- mpai1sam, vmpai <- mpai2sam)
delta <- vmfilho - vmpai
P <- exp(-(delta)/ti)
runif <- runif(1000,0,1)
r <- sample(runif,1, replace = FALSE, prob = NULL)
ifelse(delta <= 0, filho <- filhodeco, ifelse(r <= P, filho <- filhodeco,
filho <- paideco))
ti <- alfa * ti
t_0 <- 8 for (i in 1:300){
z <- 1:w
a <- sample(z, 1, replace = FALSE, prob = NULL)
psub <- filho[a,] ## ponto escolhido p/ ser substituído ##

```



```

npsub <- as.integer(row.names(filho))[a]
ol <- as.numeric(row.names(filho))
simul <- cbind(sim$coords,sim$data)
simul1 <- simul[-ol,] ## conjunto de pontos que podem ser sorteados sem S_0 ##
## cálculo das distâncias ##
di <- as.vector((((psub[1] - simul1[,1])^2) + ((psub[2] - simul1[,2])^2))^0.5)
e <- di <= alpha & di > 0
tf <- 1:length(di)
qqq <- tf[e == T]
sa <- resample(qqq,1)
rr <- as.integer(row.names(simul1)[sa])
qqqq <- row.names(filho)
ponto <- simul1[sa,] ## ponto sorteado a substituir em S_0 ##
row.names(filho) <- as.integer(qqqq)
amo1 <- as.geodata(filho)
## krigagem com as estimativas dos parâmetros na malha sim$coords ##
amo11exp <- likfit(amo1, ini=c(0.5, 0.2), nug=0, cov.model=exp)
kc11 <- krige.conv(amo1, loc=grid1,krige=krige.control(obj.m = amo11exp))
m11 <- mean(kc11$krige.var)
z <- 1:w
q <- as.integer(row.names(filho))
as <- z[q == npsub]
dado <- filho[-as,]
S_1 <- rbind(dado, ponto) ## conjunto S_1 modificado ##
qqqq <- row.names(S_1)
qqqq <- as.integer(replace(qqqq,d,rr))
row.names(S_1) <- qqqq
amo2 <- as.geodata(S_1)
## krigagem com as estimativas dos parâmetros na malha sim$coords ##
amo22exp <- likfit(amo2, ini=c(0.5, 0.2), nug=0, cov.model=exp)
kc22 <- krige.conv(amo2, loc=grid1,krige=krige.control(obj.m = amo22exp))
m22 <- mean(kc22$krige.var)
deltam <- m22 - m11
S_0 <- filho
P <- exp(-(deltam)/t_0)
runif <- runif(1000,0,1)
r <- sample(runif,1, replace = FALSE, prob = NULL)
ifelse(deltam <= 0, filho <- S_1, ifelse(r <= P, filho <- S_1, filho <- S_0))
t_0 <- alfa * t_0
vm <- min(m22,m11)
plot(filho[,1],filho[,2])
}
selecao <- as.matrix(selecao)
selecao <- rbind(selecao, filho)
sorter <- sort(as.integer(row.names(filho)))
filhodeco <- replace(seq0,sorter,1)
mselect <- cbind(mselect,filhodeco)
vm11 <- c(vm11,vm)
## seleção dos melhores: eliminando 1 indivíduo pior ##
mselect <- mselect[,-(rank(vm11, ties.method = first) == length(vm11))]
vm11 <- vm11[ -(rank(vm11, ties.method = first) == length(vm11))]
t <- 1:length(vm11)
nn <- t[ifelse((rank(vm11, ties.method = first) == length(vm11)) == T,1,0)==1]
selecao <- selecao[-((1+(nn-1)*w):(w*nn)),]
write.table(mselect, C:\\selectfinal.txt, dec = ., col.names = F,row.names = F)
write.table(selecao, C:\\selecaofinal.txt, dec = ., col.names = F,row.names = F)
write.table(vm11, C:\\vm11final.txt, dec = ., col.names = F,row.names = F)
write.table(t(c(ee,min(vm11))), C:\\graficobest.txt,col.names = F,row.names = F,
append = T)
}
vmfinal <- read.table(C:\\vm11final.txt)
ssss <- 1: dim(vmfinal)[1]

```

```

posicao <- ssss[vmfinal == min(vmfinal)]
se <- read.table(C:\\selectfinal.txt)
bestcodi <- se[,posicao] ## best codificado ##
sed <- read.table(C:\\selecaofinal.txt)
best <- sed[((1+(posicao-1)*w):(w*posicao)),]
plot(best[,1], best[,2])
amo <- as.geodata(best)
amoexp <- likfit(amo, ini=c(0.5, 0.2), nug=0, cov.model=exp)
OC <- output.control(n.pred=500)
kb1 <- krige.conv(amo, loc=grid1, krige=krige.control( obj.m = amoexp, out=OC)
mvar6 <- mean(kb1$krige.var)
## porcentagem dos dados simulados estarem acima do 3° quartil ##
c75 <- boxplot.stats(sim$data)$stats[4]
rep <- rep(NA, dim(kb1$simulations)[2])
repsum <- rep(NA, dim(kb1$simulations)[2])
for(i in 1:dim(kb1$simulations)[2]){
  sim1 <- kb1$simulations[,i]
  t4175 <- table(sim1 > c75)
  p4175 <- t4175/(length(sim1)[1])
  rep[i] <- as.vector(p4175[2])
  sum75 <- sum(kb1$simul[,i][(kb1$simulations[,i] > c75) == T])
  repsum[i] <- sum75
}
rep[!is.na(rep) == F] <- 0
mp756 <- mean(rep)
sq756 <- sum((rep-0.25)^2)
sum756 <- mean(repsum)
sumreal756 <- sum(sim$data[(sim$data > c75) == T])
sqsum756 <- sum((repsum - sumreal756)^2)
## porcentagem dos dados simulados estarem acima do percentil 90 ##
c90 <- as.vector(quantile(sim$data, prob = seq(0,1,0.1)))[10]
rep1 <- rep(NA, dim(kb1$simulations)[2])
rep1sum <- rep(NA, dim(kb1$simulations)[2])
for(i in 1:dim(kb1$simulations)[2]){
  sim1 <- kb1$simulations[,i]
  t4190 <- table(sim1 > c90)
  p4190 <- t4190/(length(sim1)[1])
  rep1[i] <- as.vector(p4190[2])
  sum90 <- sum(kb1$simul[,i][(kb1$simulations[,i] > c90) == T])
  rep1sum[i] <- sum90
}
rep1[!is.na(rep1) == F] <- 0
mp906 <- mean(rep1)
sq906 <- sum((rep1-0.10)^2)
sum906 <- mean(rep1sum)
sumreal906 <- sum(sim$data[(sim$data > c90) == T])
sqsum906 <- sum((rep1sum - sumreal906)^2)
write.table(t(round(c(amo1exp$nu, amo2exp$nu, amo3exp$nu, amo4exp$nu,
amo5exp$nu, amo6exp$nu, amo1exp$phi, amo2exp$phi, amo3exp$phi, amo4exp$phi,
amo5exp$phi, amo6exp$phi, amo1exp$si, amo2exp$si, amo3exp$si, amo4exp$si,
amo5exp$si, amo6exp$si), 4)), C:\\novo400a03estimados.txt, dec = ., col.names = F,
row.names = F, append = T)
write.table(t(round(c(mvar1, mvar2, mvar3, mvar4, mvar5, mvar6), 4)),
C:\\novo400a03mvar.txt, dec = ., col.names = F, row.names = F, append = T)
write.table(t(round(c(mp751, mp752, mp753, mp754, mp755, mp756, sq751, sq752, sq753,
sq754, sq755, sq766), 4)), C:\\novo400a03p75.txt, dec = ., col.names = F,
row.names = F, append = T)
write.table(t(round(c(mp901, mp902, mp903, mp904, mp905, mp906, sq901, sq902, sq903,
sq904, sq905, sq906), 4)), C:\\novo400a03p90.txt, dec = ., col.names = F,
row.names = F, append = T)
write.table(t(round(c(sqsum751, sqsum752, sqsum753, sqsum754, sqsum755, sqsum756), 4)),
C:\\novo400a03sum75.txt, dec = ., col.names = F, row.names = F, append = T)

```



```
write.table(t(round(c(sqsum901,sqsum902,sqsum903,sqsum904,sqsum905,sqsum906),4)),  
C:\\novo400a03sum90.txt, dec = ., col.names = F, row.names = F, append = T)  
}
```

ANEXO-C: Programa R utilizado na geração das amostragens reduzidas para o conjuntos de dados da variável potássio

```
## ANÁLISE DO CONJUNTO DE DADOS INICIAL ##
require(geoR)
K <- read.geodata(C:\\dados1998.dat, head=T, coords.col=2:3, data.col=5)
summary(K$data) hist(K$data, main = , xlab = K, ylab = Frequência)
range(dist(K$coords))
boxplot(K$data)
plot(K)
points(K, xlab = Coordenada X, ylab = Coordenada Y)
require(MASS)
boxcox(K)
boxcox(K, lam=seq(-1,1,l=21), ylab = log Verossimilhança)
K.vario <- variog(K, max.dist = 90)
plot(K.vario)
variog4 <- variog4(K, max.dist = 90) plot(variog4, xlab = Distância,
ylab = Semivariância)
K.env <- variog.mc.env(K, obj.variog = K.vario) plot(K.vario, env = K.env)
K.ml <- likfit(K, ini=c(0.05,40), nugget=0)
summary(K.ml)
K.grid <- expand.grid(seq(0,149.60,by=1), seq(0,113.6,by=1))
K.kc <- krige.conv(K, loc=K.grid, krige=krige.control(obj=K.ml))
image(K.kc, loc = K.grid, col=gray(seq(.9,0,l=21)), xlab=Coordenada X,
ylab=Coordenada Y,levels=seq(0, 2, 10))
par.ori <- par(no.readonly=T)
as <- ifelse(K.kc$predi > 0.30,3,FALSE)
dg <- 1:length(K.kc$pre)
dg030 <- dg[as == 3]
as <- ifelse(K.kc$predi >= 0.21 & K.kc$predi <= 0.30,2,FALSE)
dg <- 1:length(K.kc$pre) dg021 <- dg[as == 2]
as <- ifelse(K.kc$predi > 0.30,3, ifelse(K.kc$predi >= 0.21 & K.kc$predi <= 0.30,
2, ifelse(K.kc$predi >= 0.11 & K.kc$predi <= 0.20,1, ifelse(K.kc$predi <= 0.10,
0,-1))))
image(K.kc, loc = K.grid, col=gray(seq(0.9,0.3,l=4)), val = as,
xlab=Coordenada X, ylab=Coordenada Y)
## porcentagem dos valores preditos estarem na classificação do Paraná ##
t11 <- table(K.kc$predict < 0.11)
p11 <- t11/(length(K.kc$predict))
t1120 <- table(K.kc$predict >= 0.11 & K.kc$predict <= 0.20)
p1120 <- t1120/(length(K.kc$predict))
t2130 <- table(K.kc$predict >= 0.21 & K.kc$predict <= 0.30)
p2130 <- t2130/(length(K.kc$predict))
t30 <- table(K.kc$predict > 0.30)
p30 <- t30/(length(K.kc$predict))
t <- cbind(t11[2],t1120[2],t2130[2],t30[2])
p <- cbind(p11[2],p1120[2],p2130[2],p30[2])
## AMOSTRA ALEATÓRIA ##
n <- dim(K$coords)[1]
w <- n/2
x <- 1:n
a1 <- sample(x, w, replace = FALSE, prob = NULL)
o1 <- sort(a1)
v <- as.vector(K$data)
v1 <- v[o1]
c1 <- as.matrix(K$coords)
co1 <- c1[,1]
co2 <- c1[,2]
sco1 <- co1[o1]
sco2 <- co2[o1]
dados <- cbind(sco1,sco2,v1)
```

```

amo1 <- as.geodata(dados)
points(amo1, xlab = Coordenada X, ylab = Coordenada Y)
amol1exp <- likfit(amo1, ini=c(0.05,20), nug=0)
K11.kc <- krige.conv(amo1, loc=K.grid, krige=krige.control(obj=amol1exp))
m11 <- mean(K11.kc$krige.var)
image(K11.kc, loc=K.grid, col=gray(seq(.9,0,l=21)), xlab=Coordenada X,
ylab=Coordenada Y)
par.ori <- par(no.readonly=T)
as <- ifelse(K11.kc$predi > 0.30,1,0)
image(K11.kc, loc = K.grid, col=gray(seq(0.9,0.3,l=21)), val = as,
xlab=Coordenada X, ylab=Coordenada Y)
as <- ifelse(K11.kc$predi > 0.30,3, ifelse(K11.kc$predi >= 0.21 &
K11.kc$predi <= 0.30,2, ifelse(K11.kc$predi >= 0.11 & K11.kc$predi <= 0.20,1,
ifelse(K11.kc$predi <= 0.10,0,-1))))
image(K11.kc, loc = K.grid, col=gray(seq(0.9,0.3,l=4)), val = as,
xlab=Coordenada X, ylab=Coordenada Y)
## porcentagem dos valores preditos estarem na classificação do Paraná ##
t111 <- table(K11.kc$predict < 0.11)
p111 <- t111/(length(K11.kc$predict))
t11201 <- table(K11.kc$predict >= 0.11 & K11.kc$predict <= 0.20)
p11201 <- t11201/(length(K11.kc$predict))
t21301 <- table(K11.kc$predict >= 0.21 & K11.kc$predict <= 0.30)
p21301 <- t21301/(length(K11.kc$predict))
t310 <- table(K11.kc$predict > 0.30)
p310 <- t310/(length(K11.kc$predict))
t1 <- cbind(t111[2],t11201[2],t21301[2],t310[2])
p1 <- cbind(p111[2],p11201[2],p21301[2],p310[2])
## AMOSTRA SISTEMÁTICA ##
a22 <- seq(1,n, by = 2)
Kdados <- cbind(K$coords,K$data)
dados <- Kdados[a22,]
amo2 <- as.geodata(dados)
resample <- function(x, size, ...) if(length(x) <= 1) {
if(!missing(size) && size == 0) x[FALSE] else x } else
sample(x, size, ...)
points(amo2, xlab = Coordenada X, ylab = Coordenada Y)
amo22exp <- likfit(amo2, ini=c(0.05,20), nug=0)
K22.kc <- krige.conv(amo2, loc=K.grid, krige=krige.control(obj=amo22exp))
m22 <- mean(K22.kc$krige.var)
image(K22.kc, loc=K.grid, col=gray(seq(.9,0,l=21)), xlab=Coordenada X,
ylab=Coordenada Y)
par.ori <- par(no.readonly=T)
as <- ifelse(K22.kc$predi > 0.30,1,0)
image(K22.kc, loc = K.grid, col=gray(seq(0.9,0.3,l=21)), val = as,
xlab=Coordenada X, ylab=Coordenada Y)
as <- ifelse(K11.kc$predi > 0.30,1,0)
image(K11.kc, loc = K.grid, col=gray(seq(0.9,0.3,l=21)), val = as,
xlab=Coordenada X, ylab=Coordenada Y)
as <- ifelse(K22.kc$predi > 0.30,3, ifelse(K22.kc$predi > 0.20 &
K22.kc$predi <= 0.30,2, ifelse(K22.kc$predi > 0.10 & K22.kc$predi <= 0.20,1,
ifelse(K22.kc$predi <= 0.10,0,-1))))
image(K22.kc, loc = K.grid, col=gray(seq(1,0.3,l=4)), val = as,
xlab=Coordenada X, ylab=Coordenada Y)
## porcentagem dos valores preditos estarem na classificação do Paraná ##
t112 <- table(K22.kc$predict < 0.11)
p112 <- t112/(length(K22.kc$predict))
t11202 <- table(K22.kc$predict >= 0.11 & K22.kc$predict <= 0.20)
p11202 <- t11202/(length(K22.kc$predict))
t21302 <- table(K22.kc$predict >= 0.21 & K22.kc$predict <= 0.30)
p21302 <- t21302/(length(K22.kc$predict))
t320 <- table(K22.kc$predict > 0.30)
p320 <- t320/(length(K22.kc$predict))

```

```

t2 <- cbind(t112[2],t11202[2],t21302[2],t320[2])
p2 <- cbind(p112[2],p11202[2],p21302[2],p320[2])
## RECOZIMENTO SIMULADO ##
t_0 <- 8                ## temperatura inicial ##
alfa <- 0.85            ## fator de ponderação de t ##
alpha <- 20             ## raio ##
d <- w grid1 <- expand.grid(seq(0,149.60,by=6), seq(0,113.6,by=6))
for (i in 1:500){
  z <- 1:d
  a <- sample(z, 1, replace = FALSE, prob = NULL)
  psub <- dados[a,] ## ponto escolhido p/ ser substituído ##
  npsub <- as.integer(row.names(dados))[a] ## nº do ponto a ser substituído ##
  o1 <- as.numeric(row.names(dados))
  simul <- cbind(K$coords,K$data)
  simul1 <- simul[-o1,] ## conjunto de pontos que podem ser sorteados sem S_0 ##
  ## cálculo das distâncias ##
  di <- as.vector((((psub[1] - simul1[,1])^2) + ((psub[2] - simul1[,2])^2))^0.5)
  e <- di <= alpha & di > 0
  tf <- 1:length(di)
  qq <- tf[e == T]
  sa <- resample(qq,1)
  rr <- as.integer(row.names(simul1)[sa])
  qqqq <- row.names(dados)
  ponto <- simul1[sa,] ## ponto sorteado a substituir em S_0 ##
  row.names(dados) <- as.integer(qqqq)
  amo11 <- as.geodata(dados)
  amo11exp <- likfit(amo11, ini=c(0.05,20), nug=0)
  kb11 <- krige.conv(amo11, loc=grid1,krige=krige.control(obj=amo11exp))
  m11 <- mean(kb11$krige.var)
  z <- 1:d
  q <- as.integer(row.names(dados))
  as <- z[q == npsub]
  dado <- dados[-as,]
  S_1 <- rbind(dado, ponto) ## conjunto S_1 modificado ##
  qqqq <- row.names(S_1)
  qqqq <- as.integer(replace(qqqq,W,rr))
  row.names(S_1) <- qqqq
  amo22 <- as.geodata(S_1)
  amo22exp <- likfit(amo22, ini=c(0.05,20), nug=0)
  kb22 <- krige.conv(amo22, loc=grid1,krige=krige.control(obj=amo22exp))
  m22 <- mean(kb22$krige.var)
  deltam <- m22 - m11
  S_0 <- dados
  P <- exp(-(deltam)/t_0)
  runif <- runif(1000,0,1)
  r <- sample(runif,1, replace = FALSE,prob = NULL)
  ifelse(deltam <= 0, dados <- S_1, ifelse(r <= P, dados <- S_1, dados <- S_0))
  t_0 <- alfa * t_0
  plot(dados[,1],dados[,2])
  write.table(t(c(m11,m22,min(m11,m22),t_0)), C:\\recozimentodados1998.txt,
  dec = ., col.names = F,row.names = F, append = T)
}
write.table(dados, C:\\amostrareco1998.txt, dec = ., col.names = F,
row.names = F, append = T)
dados <- read.table(C:\\amostrareco1998.txt)
amo3 <- as.geodata(dados[1:w,])
points(amo3, xlab = Coordenada X, ylab = Coordenada Y)
amo33exp <- likfit(amo3, ini=c(0.05,20), nug=0)
K33.kc <- krige.conv(amo3, loc=K.grid, krige=krige.control(obj=amo33exp))
m33 <- mean(K33.kc$krige.var)
image(K33.kc, loc=K.grid, col=gray(seq(.9,0,l=21)),xlab=Coordenada X,
ylab=Coordenada Y)

```

```

par.ori <- par(no.readonly=T)
as <- ifelse(K33.kc$predi > 0.30,3, ifelse(K33.kc$predi >= 0.21 &
K33.kc$predi <= 0.30,2, ifelse(K33.kc$predi >= 0.11 & K33.kc$predi <= 0.20,1,
ifelse(K33.kc$predi <= 0.10,0,-1)))
image(K33.kc, loc = K.grid, col=gray(seq(0.7,0.3,l=21)), val = as,
xlab=Coordenada X, ylab=Coordenada Y)
as <- ifelse(K33.kc$predi > 0.30,1,0)
image(K33.kc, loc = K.grid, col=gray(seq(0.9,0.3,l=21)), val = as,
xlab=Coordenada X, ylab=Coordenada Y)
## porcentagem dos valores preditos estarem na classificação do Paraná ##
t113 <- table(K33.kc$predict < 0.11)
p113 <- t113/(length(K33.kc$predict))
t11203 <- table(K33.kc$predict >= 0.11 & K33.kc$predict <= 0.20)
p11203 <- t11203/(length(K33.kc$predict))
t21303 <- table(K33.kc$predict >= 0.21 & K33.kc$predict <= 0.30)
p21303 <- t21303/(length(K33.kc$predict))
t330 <- table(K33.kc$predict > 0.30)
p330 <- t330/(length(K33.kc$predict))
t3 <- cbind(t113[2],t11203[2],t21303[2],t330[2])
p3 <- cbind(p113[2],p11203[2],p21303[2],p330[2])
## ALGORITMO GENÉTICO HÍBRIDO ##
n <- dim(K$coords)[1]
w <- n/2
t_0 <- 8          ## temperatura inicial ##
alfa <- 0.85      ## fator de ponderação de t ##
alpha <- 30       ## raio ##
d <- w
grid1 <- expand.grid(seq(0,149.60,by=6), seq(0,113.6,by=6))
K.grid <- expand.grid(seq(0,149.60,by=6), seq(0,113.6,by=6))
simu <- cbind(K$coords,K$data)
bb <- 100          ## tamanho da população inicial ##
we <- 100          ## número de iterações para o cruzamento ##
resample <- function(x, size, ...) if(length(x) <= 1) {
if(!missing(size) && size == 0) x[FALSE] else x} else
sample(x,size, ...)
## criando os indivíduos da população com amostragem aleatória ##
## selecionando os indivíduos codificados ##
mvar11 <- matrix(rep(NA,(bb)), ncol = 1, nrow = bb)
matrix01 <- matrix(rep(NA,(bb*n)), ncol = bb, nrow = n)
matrixso <- matrix(rep(NA,(bb*w)), ncol = bb, nrow = w)
for(cc in 1:bb) {
seq0 <- rep(0,n)
so <- sample(1:n, w, replace = FALSE, prob = NULL)
sort <- sort(so)
matrixso[,cc] = sort
indi <- replace(seq0, sort, 1)
matrix01[,cc] = indi
}
## Decodificando para a avaliação dos indivíduos e avaliação ##
decodi <- matrix(rep(NA,(bb*w*3)), ncol = 3, nrow = bb*w)
for (cc in 1:bb) {
indiv <- simu[matrixso[,cc],]
decodi[(((cc-1)*w)+1):(cc*w),] <- indiv
amo11 <- as.geodata(indiv)
amo11exp <- likfit(amo11, ini=c(0.05,20), nug=0)
kc11 <- krige.conv(amo11, loc=K.grid,krige=krige.control(obj=amo11exp))
m11 <- mean(kc11$krige.var)
mvar11[cc,] <- m11
}
## seleção inicial de 50% dos indivíduos da população ##
select <- matrix(rep(NA,(bb*w/2)*3), ncol = 3, nrow = (bb*w/2))
tablee <- decodi

```

```

vm11 <- as.vector(mvar11)
median <- median(vm11)
vm1 <- as.vector(vm11)
vm1[vm1 <= median] <- 1
z <- 1:bb
esco <- z[vm1 == 1] ## número das amostras selecionadas ##
for (cc in 1:length(esco)) {
  as <- tablee[(1+((esco[cc]-1)*w)):(w*esco[cc]),]
  select[((cc-1)*w)+1):(cc*w),] <- indiv
}
vm11 <- vm11[esco]
## select40a03 são os 50% dos indivíduos selecionados ##
selecao <- select ## conjunto de dados selecionados ##
mselect <- matrix01[,esco] ## conjunto selecionado codificado ##
## recombinação (crossover) dos indivíduos codificados ##
ti <- 1.5
for (ee in 1:100){
  sequencia <- 1:w
  pcorte <- sample(sequencia,1) ## ponto de corte ##
  ## recombinação de pais aleatórios ##
  sample <- sample(1:(dim(mselect)[2]),2)
  pai1sam <- mselect[,sample[1]] ## pai1 codificado ##
  pai2sam <- mselect[,sample[2]] ## pai2 codificado ##
  filho1sam <- c(pai1sam[1:pcorte],pai2sam[(pcorte+1):n])
  filho2sam <- c(pai2sam[1:pcorte],pai1sam[(pcorte+1):n])
  ## reparo dos filhos ##
  nf1 <- (w-table(filho1sam)[2])
  nf2 <- (w-table(filho2sam)[2])
  qq <- 1:n
  f1 <- qq[filho1sam == 0]
  f2 <- qq[filho2sam == 0]
  f11 <- qq[filho1sam == 1]
  f21 <- qq[filho2sam == 1]
  if(nf1 >= 0) filho1sam <- replace(filho1sam,sample(f1,abs(nf1)),rep(1,abs(nf1)))
  else filho1sam <- replace(filho1sam,sample(f11,abs(nf1)),rep(0,abs(nf1)))
  filhodeco1sam <- filho1sam
  if(nf2 >= 0) filho2sam <- replace(filho2sam,sample(f2,abs(nf2)),rep(1,abs(nf2)))
  else filho2sam <- replace(filho2sam,sample(f21,abs(nf2)),rep(0,abs(nf2)))
  filhodeco2sam <- filho2sam
  ## conjunto de filhos decodificados ##
  p11 <- qq[filho1sam == 1]
  p21 <- qq[filho2sam == 1]
  filhod1sam <- simu[p11,]
  filhod2sam <- simu[p21,]
  pai11 <- qq[pai1sam == 1]
  pai21 <- qq[pai2sam == 1]
  paid1sam <- simu[pai11,]
  paid2sam <- simu[pai21,]
  ## avaliação dos filhos e dos pais ##
  filho1sam <- as.geodata(filhod1sam)
  filho2sam <- as.geodata(filhod2sam)
  pai1 <- as.geodata(paid1sam)
  pai2 <- as.geodata(paid2sam)
  amo1filhoexp <- likfit(filho1sam, ini=c(0.05,20), nug=0)
  kb11 <- krige.conv(filho1sam, loc=K.grid,krige=krige.control
    (obj.m = amo1filhoexp))
  mfilho1sam <- mean(kb11$krige.var)
  amo2filhoexp <- likfit(filho2sam, ini=c(0.05,20), nug=0)
  kb22 <- krige.conv(filho2sam, loc=K.grid,krige=krige.control
    (obj.m = amo2filhoexp))
  mfilho2sam <- mean(kb22$krige.var)
  amo1paiexp <- likfit(pai1, ini=c(0.05,20), nug=0)

```

```

kbp11 <- krige.conv(pai1, loc=K.grid,krige=krige.control(obj.m = amo1paiexp))
mpailsam <- mean(kbp11$krige.var)
amo2paiexp <- likfit(pai2, ini=c(0.05,20), nug=0)
kbp22 <- krige.conv(pai2, loc=grid1,krige=krige.control(obj.m = amo2paiexp))
mpai2sam <- mean(kbp22$krige.var)
## escolha do melhor filho ##
dif <- mfilho1sam - mfilho2sam
m <- ifelse(dif <=0, filhodeco <- filhod1sam, filhodeco <- filhod2sam)
mdeco <- ifelse(dif <=0, filhoco <- filhodeco1sam, filhoco <- filhodeco2sam)
vmfilho <- ifelse(dif <=0, vmfilho <- mfilho1sam, vmfilho <- mfilho2sam)
## escolha do melhor pai ##
dif <- mpailsam - mpai2sam
m <- ifelse(dif <=0, paico <- pailsam, paico <- pai2sam)
mdeco <- ifelse(dif <=0, paideco <- paid1sam, paideco <- paid2sam)
vmpai <- ifelse(dif <=0, vmpai <- mpailsam, vmpai <- mpai2sam)
delta <- vmfilho - vmpai
P <- exp(-(delta)/ti)
runif <- runif(1000,0,1)
r <- sample(runif,1, replace = FALSE, prob = NULL)
ifelse(delta <= 0, filho <- filhodeco, ifelse(r <= P, filho <- filhodeco,
filho <- paideco))
ti <- alfa * ti
t_0 <- 1.5
for (i in 1:500){
z <- 1:w
a <- sample(z, 1, replace = FALSE, prob = NULL)
psub <- filho[a,] ## ponto escolhido p/ ser substituído ##
npsub <- as.integer(row.names(filho))[a] ## nº do ponto a ser substituído ##
o1 <- as.numeric(row.names(filho))
simul <- cbind(K$coords,K$data)
simul1 <- simul[-o1,] ## conjunto de pontos que podem ser sorteados sem S_0 ##
# cálculo das distâncias ##
di <- as.vector((((psub[1] - simul1[,1])^2) + ((psub[2] - simul1[,2])^2))^0.5)
e <- di <= alpha & di > 0
tf <- 1:length(di)
qqq <- tf[e == T]
sa <- resample(qqq,1)
rr <- as.integer(row.names(simul1)[sa])
qqqq <- row.names(filho)
ponto <- simul1[sa,] ## ponto sorteado a substituir em S_0 ##
row.names(filho) <- as.integer(qqqq)
amo1 <- as.geodata(filho)
## krigagem com as estimativas dos parâmetros na malha sim$coords ##
amo1exp <- likfit(amo1, ini=c(0.05,20), nug=0)
kc11 <- krige.conv(amo1, loc=K.grid,krige=krige.control(obj.m = amo1exp))
m11 <- mean(kc11$krige.var)
z <- 1:w
q <- as.integer(row.names(filho))
as <- z[q == npsub]
dado <- filho[-as,]
S_1 <- rbind(dado, ponto) ## conjunto S_1 modificado ##
qqqq <- row.names(S_1)
qqqq <- as.integer(replace(qqqq,128,rr))
row.names(S_1) <- qqqq
amo2 <- as.geodata(S_1)
## krigagem com as estimativas dos parâmetros na malha sim$coords ##
amo2exp <- likfit(amo2, ini=c(0.05,20), nug=0)
kc22 <- krige.conv(amo2, loc=K.grid,krige=krige.control(obj.m = amo2exp))
m22 <- mean(kc22$krige.var)
deltam <- m22 - m11
S_0 <- filho
P <- exp(-(deltam)/t_0)

```



```

runif <- runif(1000,0,1)
r <- sample(runif,1, replace = FALSE, prob = NULL)
ifelse(deltam <= 0, filho <- S_1, ifelse(r <= P, filho <- S_1, filho <- S_0))
t_0 <- alfa * t_0
vm <- min(m22,m11)
plot(filho[,1],filho[,2])
}
selecao <- as.matrix(selecao)
selecao <- rbind(selecao, filho)
seq0 <- rep(0,n)
sorter <- sort(as.integer(row.names(filho)))
filhodeco <- replace(seq0,sorter,1)
mselect <- cbind(mselect,filhodeco)
vm11 <- rbind(vm11,vm)
## selecao dos melhores: eliminando 1 indivíduo pior ##
mselect <- mselect[,-(rank(vm11, ties.method = first) == length(vm11))]
vm11 <- vm11[,-(rank(vm11, ties.method = first) == length(vm11))]
t <- 1:length(vm11)
nn <- t[ifelse((rank(vm11, ties.method = first) == length(vm11)) == T,1,0)==1]
selecao <- selecao[-((1+(nn-1)*w):(w*nn)),]
write.table(mselect, C:\\selectfinal1998.txt, dec = ., col.names = F,
row.names = F)
write.table(selecao, C:\\selecaofinal1998.txt, dec = ., col.names = F,
row.names = F)
write.table(vm11, C:\\vm11final1998.txt, dec = ., col.names = F,row.names = F)
write.table(min(vm11), C:\\graficobest1998.txt,col.names = F,row.names = F,
append = T)
}
vmfinal <- read.table(C:\\vm11final1998.txt)
ssss <- 1: dim(vmfinal)[1]
posicao <- ssss[vmfinal == min(vmfinal)]
vm11 <- as.vector(vmfinal)
se <- read.table(C:
selectfinal1998.txt")"
bestcodi <- se[,posicao]      ## best codificado ##
mselect <- se
sed <- read.table(C:\\selecaofinal1998.txt)
best <- sed[((1+(posicao-1)*w):(w*posicao)),]
selecao <- sed
plot(best[,1], best[,2])
amo4 <- as.geodata(best)
amo44exp <- likfit(amo4, ini=c(0.05,20), nug=0)
K44.kc <- krige.conv(amo4, loc=K.grid, krige=krige.control(obj=amo44exp))
m44 <- mean(K44.kc$krige.var)
as <- ifelse(K44.kc$predi > 0.30,3, ifelse(K44.kc$predi >= 0.21 &
K44.kc$predi <= 0.30, 2, ifelse(K44.kc$predi >= 0.11 & K44.kc$predi <= 0.20,
1, ifelse(K44.kc$predi <= 0.10,0,-1))))
image(K44.kc, loc = K.grid, col=gray(seq(0.7,0.3,l=21))), val = as,
xlab=Coordenada X, ylab=Coordenada Y)
## porcentagem dos valores preditos estarem na classificação do Paraná ##
t114 <- table(K44.kc$predict < 0.11)
p114 <- t114/(length(K44.kc$predict))
t11204 <- table(K44.kc$predict >= 0.11 & K44.kc$predict <= 0.20)
p11204 <- t11204/(length(K44.kc$predict))
t21304 <- table(K44.kc$predict >= 0.21 & K44.kc$predict <= 0.30)
p21304 <- t21304/(length(K44.kc$predict))
t340 <- table(K44.kc$predict > 0.30)
p340 <- t340/(length(K44.kc$predict))
t4 <- cbind(t114[2],t11204[2],t21304[2],t340[2])
p4 <- cbind(p114[2],p11204[2],p21304[2],p340[2])

```