

UNIVERSIDADE ESTADUAL DO OESTE DO PARANÁ
CAMPUS DE CASCAVEL
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA AGRÍCOLA

**ALGORITMO GENÉTICO APLICADO À DETERMINAÇÃO DA MELHOR
CONFIGURAÇÃO E DO MENOR TAMANHO AMOSTRAL NA ANÁLISE DA
VARIABILIDADE ESPACIAL DE ATRIBUTOS QUÍMICOS DO SOLO**

Tamara Cantú Maltauro

Cascavel – Paraná – Brasil

Fevereiro 2018

Tamara Cantú Maltauro

**ALGORITMO GENÉTICO APLICADO À DETERMINAÇÃO DA MELHOR
CONFIGURAÇÃO E DO MENOR TAMANHO AMOSTRAL NA ANÁLISE DA
VARIABILIDADE ESPACIAL DE ATRIBUTOS QUÍMICOS DO SOLO**

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Agrícola em cumprimento parcial aos requisitos à obtenção do título de mestre em Engenharia Agrícola, área de concentração em Engenharia de Sistemas Agroindustriais.

Orientadora: Professora Dra. Luciana Pagliosa Carvalho Guedes.

Coorientadora: Professora Dra. Rosângela Villwock.

Cascavel – Paraná – Brasil

Fevereiro de 2018

Ficha de identificação da obra elaborada através do Formulário de Geração Automática do Sistema de Bibliotecas da Unioeste.

Maltauro, Tamara Cantú
ALGORITMO GENÉTICO APLICADO À DETERMINAÇÃO DA
MELHOR CONFIGURAÇÃO E DO MENOR TAMANHO AMOSTRAL NA
ANÁLISE DA VARIABILIDADE ESPACIAL DE ATRIBUTOS
QUÍMICOS DO SOLO / Tamara Cantú Maltauro;
orientador(a), Luciana Pagliosa Carvalho Guedes;
coorientador(a), Rosangela Villwock, 2018.
103 f.

Dissertação (mestrado), Universidade Estadual do
Oeste do Paraná, Campus de Cascavel, Centro de
Ciências Exatas e Tecnológicas, Programa de Pós-
Graduação em Engenharia Agrícola, 2018.

1. Agricultura de precisão. 2. Geoestatística. 3.
Exatidão Global. 4. Matriz de informação de Fisher.
I. Guedes, Luciana Pagliosa Carvalho. II. Villwock,
Rosangela. III. Título.

Revisão de português, inglês e normas realizada por Ana Maria Martins Alves Vasconcelos em 09 de maio de 2018.



unioeste

Universidade Estadual do Oeste do Paraná

Campus de Cascavel CNPJ 78680337/0002-65
Rua Universitária, 2069 - Jardim Universitário - Cx. P. 000711 - CEP 85819-110
Fone:(45) 3220-3000 - Fax:(45) 3324-4566 - Cascavel - Paraná



PARANÁ
GOVERNO DO ESTADO

TAMARA CANTÚ MALTAURO

Algoritmo genético aplicado à determinação da melhor configuração e do menor tamanho amostral na análise da variabilidade espacial de atributos químicos do solo

Dissertação apresentada ao Programa de Pós-Graduação em Engenharia Agrícola em cumprimento parcial aos requisitos para obtenção do título de Mestra em Engenharia Agrícola, área de concentração Sistemas Biológicos e Agroindustriais, linha de pesquisa Geoprocessamento, Estatística Espacial e Agricultura de Precisão, APROVADO(A) pela seguinte banca examinadora:

Orientador(a) - Luciana Pagliosa Carvalho Guedes

Universidade Estadual do Oeste do Paraná - Campus de Cascavel (UNIOESTE)

Alan Gavioli

Universidade Tecnológica Federal do Paraná (UTFPR)

Miguel Ángel Uribe Opazo

Universidade Estadual do Oeste do Paraná - Campus de Cascavel (UNIOESTE)

Cascavel, 21 de fevereiro de 2018

BIOGRAFIA

Tamara Cantú Maltauro, nascida em Palotina – Paraná – Brasil em maio de 1992, graduada em Licenciatura em Matemática pela Universidade Estadual do Oeste do Paraná (UNIOESTE) em 2015. Experiência profissional como professora de Matemática do Integral do Colégio Anglo. Em fevereiro de 2016, ingressou no Mestrado em Engenharia Agrícola, área de concentração Sistemas Biológicos e Agroindustriais, linha de pesquisa Geoprocessamento, Estatística Espacial e Agricultura de Precisão. Atualmente é bolsista CAPES.

AGRADECIMENTOS

Meus sinceros agradecimentos...

Em primeiro lugar a Deus, por ter me iluminado nesta caminhada, me concedendo saúde, capacidade e força para concluir este trabalho.

A minha mãe, pela minha educação, pelo apoio, incentivo, carinho e motivação durante o trabalho.

A minha irmã pelo carinho e companheirismo.

Ao meu namorado, pacientemente sempre me dando apoio, carinho e motivação para nunca desistir. E por entender o tempo necessário para a realização deste trabalho.

Em especial agradeço a minha orientadora, Luciana Pagliosa Carvalho Guedes, por todo o conhecimento e sabedoria repassados, por ser uma orientadora extraordinária, tendo muita paciência, competência, confiança, disponibilidade e dedicação nesse trabalho.

A minha coorientadora professora Rosângela Villwock, por estar sempre disponível para discussões do trabalho.

À banca, pelas sugestões e elogios.

Aos colegas do grupo, em especial à Letícia e ao Weverton, pelo companheirismo durante o trabalho.

A esta Universidade, ao Programa de Pós-Graduação em Engenharia Agrícola - PGEAGRI, seu corpo docente, direção e administração.

À CAPES, pelo apoio e amparo financeiro.

Enfim, a todos que, de alguma forma, contribuíram para a realização deste trabalho.

ALGORITMO GENÉTICO APLICADO À DETERMINAÇÃO DA MELHOR CONFIGURAÇÃO E DO MENOR TAMANHO AMOSTRAL NA ANÁLISE DA VARIABILIDADE ESPACIAL DE ATRIBUTOS QUÍMICOS DO SOLO

RESUMO

É necessário determinar um esquema de amostragem com um tamanho que minimize os custos operacionais e maximize a qualidade dos resultados durante a montagem de um experimento que envolva o estudo da variabilidade espacial de atributos químicos do solo. Assim, o objetivo deste trabalho foi redimensionar uma configuração amostral com o menor número de pontos possíveis para uma área comercial composta por 102 pontos, considerando a informação sobre a variabilidade espacial de atributos químicos do solo no processo de otimização. Inicialmente, realizaram-se simulações de Monte Carlo, assumindo as variáveis estacionárias Gaussiana, isotrópicas, modelo exponencial para a função semivariância e três configurações amostrais iniciais: sistemática, aleatória simples e *lattice plus close pairs*. O Algoritmo Genético (AG) foi utilizado para a obtenção dos dados simulados e dos atributos químicos do solo, a fim de se redimensionar a amostra otimizada, considerando duas funções-objetivo. Essas estão baseadas na eficiência quanto à predição espacial e à estimação do modelo geoestatístico, as quais são respectivamente: a maximização da medida de acurácia exatidão global e a minimização de funções baseadas na matriz de informação de Fisher. Observou-se pelos dados simulados que, para ambas as funções-objetivo, quando o efeito pepita e o alcance variaram, em geral, as amostragens apresentaram os menores valores da função-objetivo, com efeito pepita igual a 0 e alcance prático igual a 0,9. O aumento do alcance prático gerou uma leve redução do número de pontos amostrais otimizados para a maioria dos casos. Em relação aos atributos químicos do solo, o AG, com ambas as funções-objetivo, foi eficiente quanto à redução do tamanho amostral. Para a maximização da exatidão global, tem-se que o tamanho amostral da nova amostra reduzida variou entre 30 e 35 pontos que corresponde respectivamente a 29,41% e a 34,31% da malha inicial, com uma similaridade mínima de predição espacial, em relação à configuração original, igual ou superior a 85%. Vale ressaltar que tais dados refletem no processo de otimização, os quais apresentam similaridade entre os mapas construídos com as configurações amostrais: original e otimizada. Todavia, o tamanho amostral da amostra otimizada variou entre 30 e 40 pontos para minimizar a função baseada na matriz de informação de Fisher, a qual corresponde respectivamente a 29,41% e 39,22% da malha original. Mas, não houve similaridade entre os mapas elaborados quando se considerou a configuração amostral inicial e a otimizada. Para ambas as funções-objetivo, os atributos químicos do solo apresentaram moderada dependência espacial para a configuração amostral original. E, a maioria dos atributos apresentaram moderada ou forte dependência espacial para a configuração amostral otimizada. Assim, o processo de otimização foi eficiente quando aplicados tanto nos dados simulados como nos atributos químicos do solo.

PALAVRAS-CHAVE: Agricultura de precisão, Geoestatística, exatidão global, matriz de informação de Fisher.

GENETIC ALGORITHM APPLIED TO DETERMINE THE BEST CONFIGURATION AND THE LOWEST SAMPLE SIZE IN THE ANALYSIS OF SPACE VARIABILITY OF CHEMICAL ATTRIBUTES OF SOIL

ABSTRACT

It is essential to determine a sampling design with a size that minimizes operating costs and maximizes the results quality throughout a trial setting that involves the study of spatial variability of chemical attributes on soil. Thus, this trial aimed at resizing a sample configuration with the least possible number of points for a commercial area composed of 102 points, regarding the information on spatial variability of soil chemical attributes to optimize the process. Initially, Monte Carlo simulations were carried out, assuming Gaussian, isotropic, and exponential model for semi-variance function and three initial sampling configurations: systematic, simple random and lattice plus close pairs. The Genetic Algorithm (GA) was used to obtain simulated data and chemical attributes of soil, in order to resize the optimized sample, considering two objective-functions. They are based on the efficiency of spatial prediction and geostatistical model estimation, which are respectively: maximization of global accuracy precision and minimization of functions based on Fisher information matrix. It was observed by the simulated data that for both objective functions, when the nugget effect and range varied, samplings usually showed the lowest values of objective-function, whose nugget effect was 0 and practical range was 0.9. And the increase in practical range has generated a slight reduction in the number of optimized sampling points for most cases. In relation to the soil chemical attributes, GA was efficient in reducing the sample size with both objective functions. Thus, sample size varied from 30 to 35 points in order to maximize global accuracy precision, which corresponded to 29.41% to 34.31% of the initial mesh, with a minimum spatial prediction similarity to the original configuration, equal to or greater than 85%. It is noteworthy that such data have reflected on the optimization process, which have similarity between the maps constructed with sample configurations: original and optimized. Nevertheless, the sample size of the optimized sample varied from 30 to 40 points to minimize the function based on Fisher information matrix, which corresponds to 29.41% and 39.22% of the original mesh, respectively. However, there was no similarity between the constructed maps when considering the initial and optimum sample configuration. For both objective functions, the soil chemical attributes showed mild spatial dependence for the original sample configuration. And, most of the attributes showed mild or strong spatial dependence for optimum sample configuration. Thus, the optimization process was efficient when applied to both simulated data and soil chemical attributes.

Keywords: Fisher Information Matrix, Geostatistics, Global Accuracy, Precision Agriculture.

SUMÁRIO

LISTA DE FIGURAS	ix
LISTA DE TABELAS.....	xi
1. INTRODUÇÃO	12
2. OBJETIVOS	14
2.1. Objetivo geral	14
2.2. Objetivos específicos	14
3. REVISÃO BIBLIOGRÁFICA	15
3.1. Agricultura de precisão e variabilidade espacial	15
3.2. Geoestatística	16
3.2.1. Modelo espacial linear Gaussiano	17
3.2.2. Semivariograma.....	19
3.2.3. Modelos teóricos.....	22
3.2.3.1. Modelo esférico.....	22
3.2.3.2. Modelo exponencial	23
3.2.3.3. Modelo gaussiano	24
3.2.3.4. Modelos da família Matérn.....	25
3.2.4. Métodos de estimação de parâmetros por máxima verossimilhança - ML	26
3.2.5. Desvios-padrões das estimativas	27
3.2.5.1. Função score.....	27
3.2.5.2. Matriz de informação observada para modelos espaciais Gaussianos.....	27
3.2.5.3. Matriz de informação.....	28
3.2.6. Critério de validação dos ajustes de modelos	29
3.2.6.1. Validação cruzada	29
3.3. Krigagem ordinária	31
3.4. Comparação de mapas temáticos	35
3.4.1. Matriz de erros	35
3.4.2. Exatidão Global.....	36
3.4.3. Índice Kappa.....	36
3.4.4. Índice de concordância Tau.....	37
3.5. Esquema de amostragem espacial	37
3.5.1. Amostragem aleatória simples	38
3.5.2. Amostragem sistemática	38
3.5.3. Amostragem <i>lattice plus close pairs</i>	39

3.6. Otimização de amostras espaciais	40
3.6.1. Critério de otimização	40
3.6.2. Algoritmos selecionados no processo de otimização: algoritmo genético.	40
3.6.2.1. Histórico	40
3.6.2.2. Funcionamento	41
3.6.2.3. Vantagens.....	42
3.6.2.4. Representação genética e codificação	43
3.6.2.5. Geração da população inicial	44
3.6.2.6. Função de avaliação	44
3.6.2.7. Métodos de seleção	44
3.6.2.8. Operadores genéticos.....	46
3.6.2.8.1. Cruzamento	47
3.6.2.8.2. Mutação	49
3.6.2.9. Critérios de parada.....	49
3.6.2.10. Parâmetros dos algoritmos genéticos.....	50
4. REFERÊNCIAS BIBLIOGRÁFICAS	51
5. ARTIGOS	57
5.1. ARTIGO 1	57
RESUMO	57
5.1.1. Introdução.....	58
5.1.2. Material e Métodos.....	60
5.1.2.1. Estudo de simulações	60
5.1.2.2. Estudo prático.....	63
5.1.3. Resultados.....	65
5.1.3.1. Estudos de simulação	65
5.1.3.2. Redimensionamento amostral considerando os atributos químicos do solo.....	67
5.1.4. Conclusão.....	72
5.1.6. Referências Bibliográficas.....	73
APÊNDICE	76
5.2. ARTIGO 2	79
RESUMO	79
5.2.1. Introdução.....	80
5.2.2. Material e Métodos.....	81
5.2.2.1. Estudo de simulações	81

5.2.2.2. Estudo prático.....	84
5.2.3. Resultados.....	86
5.2.3.1. Estudos de simulação	86
5.2.3.2. Otimização do tamanho e da configuração amostral considerando os atributos químicos do solo.....	88
5.2.4. Conclusão.....	93
5.2.6. Referências Bibliográficas.....	93
6. CONSIDERAÇÕES FINAIS	100

LISTA DE FIGURAS

Figura 1 Exemplo de um semivariograma omnidirecional.....	21
Figura 2 Representação gráfica do modelo teórico da função semivariância, em que α representa o alcance, φ_1 é o efeito pepita e $\varphi_1 + \varphi_2$ o patamar.	21
Figura 3 Representação gráfica do modelo esférico.....	23
Figura 4 Representação gráfica do modelo exponencial.....	24
Figura 5 Representação gráfica do modelo gaussiano.....	25
Figura 6 Representação gráfica do modelo da família Matérn, com $k = 0,5; k = 1; k = 1; k = 2; k = 2,5$	26
Figura 7 Exemplo da amostragem aleatória simples.....	38
Figura 8 Exemplo da amostragem sistemática ou regular.	39
Figura 9 Exemplo da amostragem <i>lattice plus close pairs</i>	39
Figura 10 Fluxograma do funcionamento do Algoritmo Genético.	42
Figura 11 Exemplo de representação dos indivíduos do método da seleção da roleta, em que a porcentagem (%) representa a probabilidade de seleção do indivíduo.	45
Figura 12 Exemplo de aplicação do método da seleção por torneio, em que os indivíduos e seus respectivos valores do <i>fitness</i> são dados em (a) e, em (b) apresentam-se os sete torneios realizados. Em cinza, destaca-se o indivíduo que venceu cada torneio.....	46
Figura 13 Exemplo de um cruzamento com um ponto de corte, em que P1 e P2 são os pais e F1 e F2 são os filhos gerados.....	47
Figura 14 Exemplo de um cruzamento com dois pontos de corte, em que P1 e P2 são os pais e F1 e F2 são os filhos gerados.	48
Figura 15 Exemplo de um cruzamento com quatro pontos de corte, em que P1 e P2 são os pais e F1 e F2 são os filhos gerados.	48
Figura 16 Exemplo de cruzamento uniforme, em que P1 e P2 são os pais e F1 e F2 são os filhos gerados.....	49
Figura 17 Exemplo de uma mutação binária simples.....	49
Figura 18 Gráfico da localização dos 100 pontos amostrais iniciais (\circ), para um exemplo de uma simulação da amostragem sistemática (a), aleatória (b) e <i>lattice plus close pairs</i> (c). ...	62
Figura 19 Fluxograma do funcionamento para a busca da melhor configuração amostral de tamanho reduzido.....	63
Figura 20 Mapa da área em estudo e da configuração amostral.	64
Figura 21 Mapas temáticos dos atributos químicos do solo elaborados considerando a malha amostral original e otimizada, em que \bullet representa a localização dos pontos selecionados no processo de otimização e valores estimados das medidas de similaridade EG, Tau e Kappa. X: Coordenadas do eixo X e Y: Coordenadas do eixo Y.	71
Figura 22 Gráfico da localização dos 100 pontos iniciais (\circ) e a localização dos pontos selecionados (\bullet), para um exemplo de cada simulação da amostragem aleatória simples, (a) alcance 0,45 e efeito pepita 0; (b) alcance 0,60 e efeito pepita 0; (c) alcance 0,90 e efeito pepita 0; (d) alcance 0,45 e efeito pepita 0,50; (e) alcance 0,60 e efeito pepita 0,50; (f) alcance 0,90 e efeito pepita 0,50; (g) alcance 0,45 e efeito pepita 0,80; (h) alcance 0,60 e efeito pepita 0,80; (i) alcance 0,90 e efeito pepita 0,80.....	76
Figura 23 Gráfico da localização dos 100 pontos iniciais (\circ) e a localização dos pontos selecionados (\bullet), para um exemplo de cada simulação da amostragem <i>lattice plus close pairs</i> , (a) alcance 0,45 e efeito pepita 0; (b) alcance 0,60 e efeito pepita 0; (c) alcance 0,90 e efeito pepita 0; (d) alcance 0,45 e efeito pepita 0,50; (e) alcance 0,60 e efeito pepita 0,50; (f)	

alcance 0,90 e efeito pepita 0,50; (g) alcance 0,45 e efeito pepita 0,80; (h) alcance 0,60 e efeito pepita 0,80; (i) alcance 0,90 e efeito pepita 0,80.....	77
Figura 24 Gráfico da localização dos 100 pontos iniciais (○) e a localização dos pontos selecionados (●) para um exemplo de cada simulação da amostragem sistemática, (a) alcance 0,45 e efeito pepita 0; (b) alcance 0,60 e efeito pepita 0; (c) alcance 0,90 e efeito pepita 0; (d) alcance 0,45 e efeito pepita 0,50; (e) alcance 0,60 e efeito pepita 0,50; (f) alcance 0,90 e efeito pepita 0,50; (g) alcance 0,45 e efeito pepita 0,80; (h) alcance 0,60 e efeito pepita 0,80; (i) alcance 0,90 e efeito pepita 0,80.....	78
Figura 25 Mapa da área em estudo e da configuração amostral.	85
Figura 26 Mapas temáticos dos atributos químicos do solo elaborados considerando a malha amostral original e otimizada, em que ● representa a localização dos pontos selecionados no processo de otimização e valores estimados das medidas de similaridade EG, Tau e Kappa.....	92
Figura 27 Gráfico da localização dos 100 pontos iniciais (○) e a localização dos pontos selecionados (●), para um exemplo de cada simulação da amostragem aleatória simples, (a) alcance 0,45 e efeito pepita 0; (b) alcance 0,60 e efeito pepita 0; (c) alcance 0,90 e efeito pepita 0; (d) alcance 0,45 e efeito pepita 0,50; (e) alcance 0,60 e efeito pepita 0,50; (f) alcance 0,90 e efeito pepita 0,50; (g) alcance 0,45 e efeito pepita 0,80; (h) alcance 0,60 e efeito pepita 0,80; (i) alcance 0,90 e efeito pepita 0,80.....	97
Figura 28 Gráfico da localização dos 100 pontos iniciais (○) e a localização dos pontos selecionados (●), para um exemplo de cada simulação da amostragem <i>latticeplus close pairs</i> , (a) alcance 0,45 e efeito pepita 0; (b) alcance 0,60 e efeito pepita 0; (c) alcance 0,90 e efeito pepita 0; (d) alcance 0,45 e efeito pepita 0,50; (e) alcance 0,60 e efeito pepita 0,50; (f) alcance 0,90 e efeito pepita 0,50; (g) alcance 0,45 e efeito pepita 0,80; (h) alcance 0,60 e efeito pepita 0,80; (i) alcance 0,90 e efeito pepita 0,80.....	98
Figura 29 Gráfico da localização dos 100 pontos iniciais (○) e a localização dos pontos selecionados (●), para um exemplo de cada simulação da amostragem sistemática, (a) alcance 0,45 e efeito pepita 0; (b) alcance 0,60 e efeito pepita 0; (c) alcance 0,90 e efeito pepita 0; (d) alcance 0,45 e efeito pepita 0,50; (e) alcance 0,60 e efeito pepita 0,50; (f) alcance 0,90 e efeito pepita 0,50; (g) alcance 0,45 e efeito pepita 0,80; (h) alcance 0,60 e efeito pepita 0,80; (i) alcance 0,90 e efeito pepita 0,80.....	99

LISTA DE TABELAS

Tabela 1 Matriz de erros de ordem $k \times k$	36
Tabela 2 Valores médios e desvio padrão (entre parênteses) do mínimo e o máximo da EG, calculada com a amostra de tamanho reduzido, considerando-se todas as simulações, obtidos no início e no final do processo de otimização, para as amostragens aleatória, <i>lattice plus close pairs</i> e sistemática.....	66
Tabela 3 Estatística descritiva e valor do coeficiente de correlação linear de Pearson dos atributos químicos do solo Ca [cmolc dm ⁻³], C [g dm ⁻³], Cu [mg dm ⁻³], Mn [cmolc dm ⁻³] e pH, considerando a configuração amostral original e a de tamanho reduzido	68
Tabela 4 Valores estimados dos parâmetros dos melhores modelos ajustados, com seus respectivos desvios padrões, para os atributos químicos do solo Ca, C, Cu, Mn e pH.	69
Tabela 5 Valores médios e desvio padrão (entre parênteses) do mínimo e máximo do valor estimado do logaritmo do determinante da matriz inversa de Fisher, calculado na amostra de tamanho reduzido, obtidos no início e no final do processo de otimização (o valor em negrito representa o valor mínimo obtido ao final do processo de otimização), de acordo com a amostragem inicial e os valores dos parâmetros efeito pepita e alcance, considerados em todas as simulações.....	87
Tabela 6 Estatística descritiva dos seguintes atributos químicos do solo Ca [cmolc dm ⁻³], C [g dm ⁻³], Cu [mg dm ⁻³], Mn [cmolc dm ⁻³] e pH, e valor do coeficiente de correlação linear de Pearson dos atributos químicos do solo com as coordenadas dos eixos X e Y, considerando a configuração amostral original e a configuração amostral de tamanho reduzido.....	89
Tabela 7 Valores estimados dos parâmetros dos melhores modelos geoestatísticos ajustados, com seus respectivos desvios-padrões estimados, e valor da função-objetivo obtida pelo AG, para os atributos químicos do solo Ca [cmolc dm ⁻³], C [g dm ⁻³], Cu [mg dm ⁻³], Mn [cmolc dm ⁻³] e pH, considerando a configuração amostral original e a configuração amostral de tamanho reduzido.....	90

1. INTRODUÇÃO

É de notável importância a busca por soluções ótimas para os mais variados problemas no setor agrícola, pois a completa imersão em conceitos como desempenho, eficiência e custos faz com que o setor agrícola almeje maior sustentabilidade em mercados competitivos.

Na agricultura tradicional, a aplicação de fertilizantes nas áreas agrícolas é feita de forma uniforme em toda a área. Porém, essa maneira de aplicação não é considerada otimizada, pois o solo e a cultura de uma mesma área nem sempre são uniformes (MOLIN; AMARAL; COLOÇO, 2015). Desta forma, partes da área são adubadas em excesso e causam riscos ambientais, além de elevados custos. Todavia, áreas com quantidade inferior de fertilizante não recebem o suficiente para o ótimo desenvolvimento do cultivo (PHILLIPS, 2014). Deste modo, um sistema chamado de agricultura de precisão está em processo de implantação, pois possibilita o gerenciamento localizado da cultura, com a aplicação das quantidades adequadas de insumos a cada local, a fim de que se reduzam os gastos e riscos ambientais (SCHULZ, 2015).

A agricultura de precisão permite a identificação da variabilidade espacial da produtividade e de atributos químicos e físicos do solo, bem como o gerenciamento localizado de culturas e alternativas de manejo diferenciado de acordo com a necessidade do solo (HAUSCHILD *et al.*, 2014).

A configuração e o tamanho amostral devem ser cuidadosamente definidos, especificamente na análise da variabilidade espacial de atributos químicos do solo (GUEDES, 2008). Pois, quanto maior for a quantidade de amostras do solo usadas na realização da análise química, maior será o custo para descrever a variabilidade espacial de atributos do solo de acordo com os mapas temáticos.

Assim, é necessário determinar um esquema de amostragem com o menor tamanho amostral possível em um experimento que envolva a análise da variabilidade espacial, e que minimize os custos operacionais com a sua obtenção. Todavia, a descrição da dependência espacial depende substancialmente da confiabilidade dos resultados gerados pelas análises geoestatísticas. Desta forma, a determinação da configuração amostral e do tamanho reduzido deve considerar quão bons são os resultados gerados no estudo da variabilidade espacial (GUEDES *et al.*, 2011).

Logo, a escolha do tamanho e da configuração amostral eficientes na análise da variabilidade espacial pode ser definida como um problema de otimização, que consiste em, a partir de uma malha amostral inicial, representar uma discretização da área sob estudo, escolher a melhor configuração amostral com o menor tamanho possível e que minimize as perdas quanto à acurácia dos resultados ou de um resultado específico gerado pela análise geoestatística (GUEDES *et al.*, 2016).

Existem diversos métodos de busca pela melhor configuração amostral, a qual pode ser feita pelo estudo de todas as possíveis combinações de soluções ou por métodos de busca sequencial (CÁRDENAS *et al.*, 2010; GUEDES *et al.*, 2014). Estes métodos são eficientes computacionalmente somente quando o tamanho amostral for pequeno. Caso contrário, tais métodos tornam-se exaustivos e inviáveis computacionalmente. Por conseguinte, existem as metodologias Meta-Heurísticas utilizadas em inteligência artificial, tais como o Recozimento Simulado (*Simulated Annealing*) (KRIPKA e KRIPKA, 2010) e o algoritmo genético (AG) (CÁRDENAS *et al.*, 2010; MOORI *et al.*, 2010; KONOWALENKO *et al.*, 2012).

As configurações amostrais otimizadas devem gerar previsões eficientes da variável regionalizada e de algumas de suas características, e assim levam em consideração os distintos tamanhos amostrais e as características do processo estocástico espacial.

2. OBJETIVOS

2.1. Objetivo geral

Determinar a melhor configuração como melhor tamanho amostral reduzido, para uma área comercial de cultivo de soja, onde se desenvolve uma atividade de aplicação localizada de insumos, com a utilização de algoritmo genético (AG).

2.2. Objetivos específicos

- A partir de dados simulados, avaliar a influência que o efeito pepita, o alcance e a configuração amostral inicial exercem no processo de otimização;
- Aplicar o AG em dados simulados e nos atributos químicos do solo, considerando-se duas funções objetivos, as quais descrevem a qualidade da estimação do modelo geoestatístico ou a qualidade da estimação espacial da variável georreferenciada em localizações não amostradas.

3. REVISÃO BIBLIOGRÁFICA

3.1. Agricultura de precisão e variabilidade espacial

O emprego das técnicas da agricultura de precisão (AP) vem sendo amplamente utilizado no setor agropecuário (BOTTEGA *et al.*, 2013). As técnicas fundamentam-se no fato de que as áreas agrícolas não são uniformes no espaço. A AP possui diversas formas de ser abordada e ainda diferentes modos de ser praticada, mas seu objetivo é sempre o mesmo, resolver problemas relacionados à desuniformidade do solo (JUNTOLLI *et al.*, 2013; MOLIN *et al.*, 2013). Porém, no Brasil, a prática que predomina é a gestão da adubação das lavouras de acordo com a amostragem georreferenciada de solo e aplicação de adubos de forma localizada e em dosagens variadas, de acordo com a necessidade (MOLIN; AMARAL; COLOÇO, 2015).

Segundo Soares Filho e Cunha (2015), a AP está relacionada à análise da variabilidade espacial e é caracterizada pelas seguintes fases: coletar dados, gerenciar as informações, aplicar insumos com teor variado e, por fim, avaliar econômica e ambientalmente os resultados. Sua origem deu-se pela pressão por uma agricultura menos destrutiva ao meio ambiente nos países europeus e logo após nos Estados Unidos, pela necessidade de dosagem de aplicações de insumos que possibilitam maximizar a produção sem exceder na quantidade de insumos que comprometam a qualidade do meio (MOLIN, 2002). O surgimento dos GPS (*Global Positioning System*) possibilitou a expansão das práticas da AP.

A utilização da AP é uma possibilidade para a agricultura tradicional na qual a aplicação de fertilizante é realizada com taxa fixa no solo, a fim de provocar impactos ambientais e custos desnecessários ao produtor (BERNARDI *et al.*, 2014). O objetivo é melhorar o rendimento econômico da atividade agrícola, aumentara produtividade e diminuir os custos e riscos ambientais (BRAGA; PINTO, 2011). Pode ainda ser definida como um sistema que permite a identificação da variabilidade espacial da produtividade e dos atributos químicos e físicos do solo, pois, segundo Guedes (2008), tais fatores exercem influência direta no nível de produtividade das culturas e, assim, permitem o gerenciamento localizado de culturas e alternativas de manejo diferenciado de acordo com a necessidade (HAUSCHILD *et al.*, 2014).

A AP engloba um conjunto de ferramentas que combina sistemas de informação, sensores, máquinas adaptadas e conhecimento para melhorar a produtividade.

A construção de mapas temáticos é de notável importância na utilização da AP, pois eles representam a variabilidade espacial de características da produção (MOLIN, 2002), logo, podem indicar localizações críticas na área. A variabilidade espacial é uma variação

em relação à profundidade, à distância ou ao tempo que acontece na produtividade, características químicas e físicas do solo, entre outras (MORGAN, 1997).

A obrigatoriedade das informações coletadas é geograficamente referenciada para o estudo desta variabilidade existente em áreas de produção agrícola. Assim, a Geoestatística ou teoria das variáveis regionalizadas é uma técnica empregada para estudar a variabilidade espacial (CÂMARA; MEDEIROS, 1998).

3.2. Geoestatística

A geoestatística teve início em 1951 na África do Sul com o matemático Daniel G. Krige, a partir de estudos realizados com dados de concentração de ouro. Ele concluiu que somente a variância dos dados seria insuficiente para explicar o fenômeno em estudo. Para isso, seria necessário levar em conta a distância entre as amostras. O conceito de geoestatística surge após as observações de Krige, o qual nos permite, a partir de alguns pontos amostrais, conhecer a continuidade espacial de uma variável em estudo, em toda a área de interesse, levando em consideração a localização geográfica e a dependência espacial (CÂMARA; MEDEIROS, 1998). Ou seja, a geoestatística permite identificar a existência, ou não, de dependência espacial entre observações (TEIXEIRA; SCALON, 2013).

Matheron, baseado nos trabalhos dos pioneiros Daniel G. Krige, e H. J. de Wijs, apresentou uma série de publicações que, por sua importante contribuição para o estudo e formalização da Teoria das Variáveis Regionalizadas, ficou conhecido como o criador da Geoestatística (LANDIM; YAMAMOTO, 2013).

As variáveis regionalizadas são constituídas por um duplo aspecto contraditório. Pela sua característica “aleatória” que apresenta irregularidades e variação imprevisível de um ponto para outro e pela sua característica “estrutural” que apresenta relações existentes entre os pontos no espaço motivadas pela sua gênese (GUEDES, 2008). É uma função definida em um domínio no espaço; além disto, os valores podem ser tomados em suportes geométricos (BRAGA, 1990).

Qualquer estudo geoestatístico inicia-se com uma análise estrutural, que consiste em modelar a variabilidade espacial dos dados (JOURNEL; HUIJBREGTS, 1978). O estudo desta variabilidade espacial de uma variável georreferenciada trata de uma abordagem probabilística, a qual envolve três passos fundamentais: a configuração da amostragem no campo, o delineamento dos ensaios de laboratório (quando esse for necessário) e a seleção de um método apropriado para prever o valor da variável em estudo em locais não medidos (não amostrados).

A previsão geoestatística difere da estimativa clássica em um aspecto importante: ela se baseia em modelos espaciais, todavia, o mesmo não ocorre com os métodos clássicos. A

geoestatística, ao contrário, requer o pressuposto de que a variável é aleatória, que os fatores e processos da formação no solo, no mar ou no ar são o resultado de um ou mais processos aleatórios. Os modelos nos quais as predições geoestatísticas são baseadas nestes processos aleatórios. Eles não são dos dados, nem mesmo da realidade que poderia se observar completamente, se houvesse tempo e paciência infinitos (WEBSTER; OLIVER, 2007). Desta forma, com a geoestatística, é possível conhecer a maneira que qualquer variável contínua varia no espaço, com um nível de detalhes que quantifica a variação espacial desta variável em diferentes direções espaciais (GALLARDO, 2006).

É importante considerar a variabilidade dos atributos químicos do solo quando são estudados, pois os fatores e processos da sua formação, que atuam ao longo do tempo, causam variabilidades naturais que, acrescentadas ao manejo, acentuam a variabilidade dos atributos, logo, elas exibem estruturas espaciais (BARBIERI *et al.*, 2013). Ou seja, tais estruturas são caracterizadas por terem apenas um dado em cada locação, porém, dados em locações diferentes são mais semelhantes em amostras mais próximas. Além disso, observa-se que quando se aumenta a distância entre as amostras, a similaridade entre a variável em estudo diminui. Essas características nos fornecem indícios da existência de dependência espacial da variável georreferenciada em uma área em estudo. Quando fenômenos naturais possuem variabilidade espacial, existe, portanto, um raio de dependência espacial entre as amostras, que define a maior distância na qual as amostras são consideradas dependentes.

A predição de uma variável georreferenciada em localizações não amostrada sem uma área em estudo a partir das técnicas de geoestatística é possível de ser realizada, depois da identificação da existência de dependência espacial dessa variável e da determinação de qual é o melhor modelo geoestatístico, que descreve sua estrutura de dependência espacial.

Aplicações da geoestatística podem ser encontradas em áreas bem distintas que vão desde os campos clássicos de mineração e geologia para a ciência do solo, hidrologia, meteorologia, ciências ambientais, agricultura, até mesmo engenharia estrutural (BÁRDOSSY, 2008).

3.2.1. Modelo espacial linear Gaussiano

A variável georreferenciada simulada Z é representada pelo modelo linear espacial gaussiano $\{Z(\mathbf{s}_i), \mathbf{s}_i \in S\}$, em que $S \subset \mathcal{R}^2$, sendo que \mathcal{R}^2 é o espaço euclidiano bidimensional ($d \geq 1$) (MARDIA; MARSHALL, 1984), ou ainda, o processo $\mathbf{Z} = (Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n))^T$, em que $\mathbf{s}_i (i = 1, \dots, n)$ é uma localização espacial conhecida, com distribuição gaussiana n -variada com vetor de médias $\mu \mathbf{1}$ para processos estacionários e a matriz de covariância Σ , isto é, $\mathbf{Z} \sim N_n(\mu \mathbf{1}, \Sigma)$, em que μ é uma constante; $\mathbf{1}$ é um vetor de uns

de dimensão $n \times 1$ e Σ é uma matriz chamada de covariância definida positiva simétrica, com dimensão $n \times n$, dada por $\Sigma = [Cov(Z(s_i), Z(s_j))]_{i=j-1, \dots, n}$. Supondo-se que os dados são descritos pelo modelo da Equação (1):

$$Z(s_i) = \mu(s_i) + \epsilon(s_i) \quad (1)$$

em que, $s_i (i = 1, \dots, n)$ representa as localizações dos pontos da malha amostral; n é o número de localizações, e $\epsilon(s_i)$ será o processo gaussiano, referenciado espacialmente, com função semivariância $\gamma(h)$, representada por um modelo geoestatístico a ser definido, que poderá ser um processo isotrópico (quando a função semivariância for função apenas da distância) ou um processo anisotrópico (quando a função semivariância for uma função da distância e de uma direção que apresente uma maior continuidade espacial). Além disso, $\mu(s_i)$ é o termo determinístico, de forma que, $\mu(s_i) = \mu$ é uma constante, quando o processo considerado for estacionário, e caso contrário, ou seja, é uma tendência linear quando o processo não for estacionário e representar uma tendência direcional da variável georreferenciada.

Considere um processo estocástico gaussiano dado pela Equação (1), para algumas funções conhecidas de s_i se assume que $x_1(s_i), \dots, x_p(s_i)$, a média do processo estocástico é dada pela Equação (2):

$$\mu(s_i) = \sum_{j=1}^p x_j(s_i) \beta_j, \quad i = 1, \dots, n \quad (2)$$

em que, β_1, \dots, β_p são parâmetros desconhecidos e precisam ser estimados.

A função de covariância $C(s_i, s_j) = Cov[\epsilon(s_i), \epsilon(s_j)]$, utilizada no estudo da dependência espacial do processo intrinsecamente estacionário, é especificada por um vetor tridimensional $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \varphi_3)^T$. Desta forma, pode-se utilizar a seguinte notação (URIBE-OPAZO; BORSOI; GALEA, 2012):

$Z(s_i) = z_i$, $\mathbf{Z} = (z_1, \dots, z_n)^T$, $x_{ij} = x_j(s_i)$, $\mathbf{x}_i^T = (x_{i1}, \dots, x_{ip})$, donde $i = 1, \dots, n$; $j = 1, \dots, p$, e \mathbf{X} é uma matriz $n \times p$, com as posto completo e as linhas \mathbf{x}_i^T , $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$, $\epsilon_i = \epsilon(s_i)$ e $\boldsymbol{\epsilon} = (\epsilon_1, \dots, \epsilon_n)^T$, com $i = 1, \dots, n$; $j = 1, \dots, p$. Portanto, a média do processo é dada pela Equação (3):

$$\mu(s_i) = \mathbf{x}_i^T \boldsymbol{\beta}, \quad (3)$$

Assim,

$$Z_i = \mathbf{x}_i^T \boldsymbol{\beta} + \epsilon_i, \text{ com } i = 1, \dots, n.$$

Da mesma forma, pode-se escrever da forma matricial, como a Equação (4):

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (4)$$

em que, o vetor de erros aleatórios $\boldsymbol{\epsilon}$ tem $E(\boldsymbol{\epsilon}) = \mathbf{0}$ (vetor nulo) e a matriz de covariância de $\boldsymbol{\epsilon}$, sendo $\Sigma[(\sigma_{ij})]$, com $Cov[\epsilon(s_i), \epsilon(s_j)]$. Assume-se que Σ é não singular e \mathbf{X} é de posto completo.

Uma forma paramétrica que especifica para a matriz de covariância (MARDIA; MASHALL, 1984) é dada pela Equação (5):

$$\Sigma = \varphi_1 \mathbf{I}_n + \varphi_2 \mathbf{R}, \quad (5)$$

em que φ_1 é o efeito pepita, sendo ideal que, $\gamma(0) = 0$, porém, na prática, conforme h tende a zero, $\gamma(h)$ se aproxima de um valor positivo, conhecido como efeito pepita, porém, ele revela a descontinuidade do semivariograma para distâncias menores do que a menor distância entre amostras (CÂMARA; MEDEIROS, 1998); \mathbf{I}_n é a matriz identidade $n \times n$; φ_2 é a contribuição, conhecida como a diferença entre o patamar e o efeito pepita (CÂMARA; MEDEIROS, 1998). O segmento gráfico neste intervalo que, de fato, caracteriza a dependência espacial dos processos estocásticos contínuos, pode ser entendido como uma variância estruturada, correspondente à estrutura do fenômeno em estudo (SEIDEL; OLIVEIRA, 2013); $\mathbf{R} = \mathbf{R}(\varphi_3) = [r_{ij}]$, com $r_{ij} = \frac{a - \gamma(\|s_i - s_j\|)}{a - \gamma(0)}$ (sendo a distância Euclidiana entre s_i e s_j) uma matriz simétrica $n \times n$, com elementos $r_{ij} = 1$, para $i = j$ com $i, j = 1, \dots, n$; φ_3 é a função do alcance do modelo. O alcance ($a = g(\varphi_3)$) é a distância limite de dependência espacial, ou seja, medições localizadas a distância maior que a são consideradas independentes entre si, enquanto amostras separadas por distâncias menores do que a são correlacionadas umas com as outras (VIEIRA, 2000; SEIDEL; OLIVEIRA, 2013). Ainda tem-se o patamar ($\varphi_1 + \varphi_2$) que, conforme a distância h aumenta, $\gamma(h)$ também aumenta, até determinado valor em que se estabiliza. O valor em que $\gamma(h)$ se estabiliza é conhecido como patamar, o qual é normalmente igual à variância dos dados (VIEIRA, 2000). Tal forma paramétrica ocorre para vários processos isotrópicos, onde a covariância $Cov[\epsilon(s_i), \epsilon(s_j)]$ é definida pela função de covariâncias $Cov[\epsilon(s_i), \epsilon(s_j)] = \varphi_2 r_{ij}$ (De BASTIANI; URIBE-OPAZO; DALPOSSO, 2012).

A função semivariância $\gamma(h)$ é representada por um modelo geoestatístico, porém, para o caso isotrópico, ela está descrita na Equação (6).

$$\gamma(h) = C(0) - C(h), \quad (6)$$

em que h é a distância entre dois pontos, $C(0)$ é a variância dos dados e $C(h)$ é a função covariância.

3.2.2. Semivariograma

A semivariância é uma medida do nível de dependência espacial entre duas amostras $Z(s_i)$ e $Z(s_i + h)$, em dois pontos s_i e $s_i + h$ separados por uma distância h (TEIXEIRA; SCALON, 2013). A variabilidade entre estas quantidades é caracterizada pela função de semivariância $\gamma(s_i, h)$, definida como a expectativa da variável aleatória $[Z(s_i) - Z(s_i + h)]^2$ (JOURNEL; HUIJBREGTS, 1978), ou simplesmente definida como a variância do incremento (LANDIM; YAMAMOTO, 2013), dada pela Equação (7).

$$\gamma(h) = \frac{1}{2} E[Z(s) - Z(s+h)]^2, \quad (7)$$

em que, $Z(s)$ e $Z(s+h)$ são valores amostrados nas posições s e $s+h$, respectivamente.

Utiliza-se o semivariograma para se descrever quantitativamente a continuidade espacial da variável de interesse e se identificar a existência de dependência espacial da mesma, o qual se constitui em um gráfico de dispersão dos valores da semivariância, calculados para amostras separadas, respectivamente, pela distância h versus os valores h (GUEDES, 2008; SEIDEL; OLIVEIRA, 2013).

O estimador mais conhecido da função semivariância é o estimador de Matheron (MATHERON, 1963), definido pela Equação (8).

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(s_i) - Z(s_i+h)]^2, \quad (8)$$

em que: $N(h)$ é o número de pares de pontos amostrais separados pelo vetor de norma h ; $Z(s_i)$ e $Z(s_i+h)$ são valores amostrados nas posições s_i e s_i+h , respectivamente, $i = 1, \dots, N$.

O semivariograma é a pedra fundamental da geoestatística, e seu comportamento próximo à origem é muito informativo sobre as propriedades de continuidade do processo aleatório $Z(s)$ (CRESSIE, 2015).

O cálculo dos valores da semivariância não é algo simples e direto, na verdade, o semivariograma é muito sensível à disposição dos pontos amostrais, a qual pode ser regular ou irregular (LANDIN; YAMAMOTO, 2013).

O cálculo pode ser realizado diretamente com base nos pontos amostrais para a determinação dos valores de semivariância a partir de amostras regularmente espaçadas. Os pares de pontos obtidos para determinada distância h , levando-se em conta uma direção, são utilizados para calcular as diferenças ao quadrado, os quais são acumulados para o cálculo da média conforme a Equação (8) (ISAAKS; SRIVASTAVA, 1989).

Já para o caso de amostras irregularmente espaçadas, há a necessidade de que se definam parâmetros novos e se introduzam limites de tolerância para a direção e distância (CÂMARA; MEDEIROS, 1998). Leva-se em consideração como referência o lag (refere-se a uma distância pré-definida, a qual é utilizada para o cálculo do gráfico semivariograma) (ISAAKS; SRIVASTAVA, 1989; LANDIM; YAMAMOTO, 2013).

Faz-se necessária a escolha de um ponto de corte para a construção do semivariograma, ou seja, a determinação da distância até onde será estudado o semivariograma (*cutoff*), isto é, a distância máxima utilizada no cálculo da semivariância. Em geral, deve-se ter bom senso para essa escolha. Porém, segundo Clark (1979), a distância de corte (*cutoff*) seria de 50% da distância máxima e possuir no mínimo 30 pares para o cálculo das semivariâncias (JOURNEL; HUIJBREGTS, 1978; GUERRA, 1988).

Um exemplo de um semivariograma é dado na Figura 1.

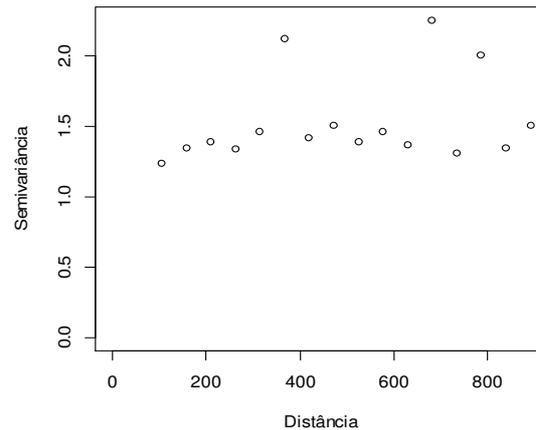


Figura 1 Exemplo de um semivariograma omnidirecional.

O semivariograma (Figura 1) é utilizado também para definir alguns parâmetros que definem o modelo que expressa a função semivariância, os quais são necessários para a estimativa de valores em locais não amostrados. Tais parâmetros definem a estrutura de variabilidade espacial $\varphi = (\varphi_1, \varphi_2, \varphi_3)^T$ e foram descritos anteriormente (Figura 2).

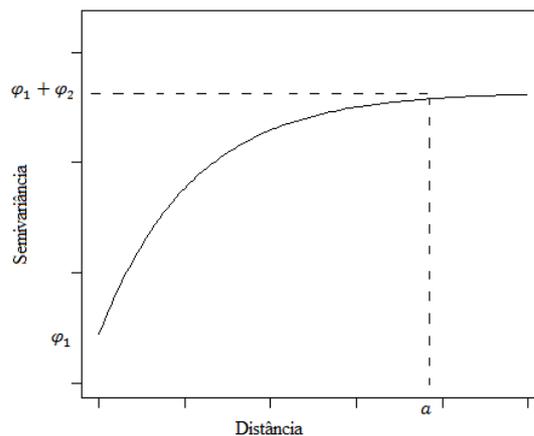


Figura 2 Representação gráfica do modelo teórico da função semivariância, em que a representa o alcance, φ_1 é o efeito pepita e $\varphi_1 + \varphi_2$ o patamar.

Espera-se que as diferenças $Z(s_i) - Z(s_i + h)$ decresçam assim que h , a distância que os separa, decresce e que as observações geograficamente mais próximas apresentem comportamento mais parecido entre si do que aquelas separadas por maiores distâncias. Desta forma, espera-se que $\gamma(h)$ aumente com a distância h (Figura 2) (CÂMARA; MEDEIROS, 1998).

O semivariograma mostra a estrutura da dependência espacial entre as amostras no decorrer de um suporte específico (CÂMARA; MEDEIROS, 1998), levando em consideração a autocorrelação em função da distância e da direção delas (MELLO; OLIVEIRA, 2016). O grau desta dependência espacial pode ser determinado pelo coeficiente de efeito pepita relativo (EPR), definido pela Equação (9):

$$EPR = \frac{\varphi_1}{\varphi_1 + \varphi_2} \cdot 100. \quad (9)$$

A dependência é considerada forte quando o valor do coeficiente EPR apresentar valor inferior a 25%, moderada dependência espacial quando apresentar valores entre 25% a 75% e forte dependência espacial para valores acima de 75% (CAMBARDELLA *et al.*, 1994).

Segundo Landin (2006), a modelagem, ou seja, o ajuste de um semivariograma experimental a uma função é um passo fundamental na análise variográfica bem como um processo que envolve várias tentativas e na qual a experiência pesa muito. Alguns modelos básicos, usuais em geoestatística são: modelo esférico, modelo exponencial, modelo gaussiano e o modelo da família Matérn. A escolha de um dos modelos citados requer um bom julgamento, baseado na experiência e uma compreensão das limitações matemáticas da função.

3.2.3. Modelos teóricos

Um dos aspectos mais relevantes da aplicação da Teoria das variáveis regionalizadas é o ajuste de um modelo teórico à função semivariância que representa o semivariograma experimental, e pode ser uma das maiores fontes de ambiguidade e polêmica nesta aplicação (VIEIRA, 2000).

Os modelos teóricos da função semivariância podem ser convenientemente divididos nos que atingem o patamar (transitivos) e os modelos que não atingem o patamar (não transitivo) (ISAAKS; SRIVASTAVA, 1989).

Os modelos teóricos com patamar mais utilizados na literatura são: esférico, exponencial, gaussiano e família Matérn (ISAAKS; SRIVASTAVA, 1989; VIEIRA, 2000; DIGGLE; RIBEIRO JR., 2007; CRESSIE, 2015). Tais modelos levam em consideração o fato de a variável apresentar o mesmo padrão de continuidade espacial em todas as direções, ou seja, apresentam dados isotrópicos.

3.2.3.1. Modelo esférico

Trata-se de um dos modelos mais utilizados na geoestatística (ISAAKS; SRIVASTAVA, 1989; SOARES, 2014). É dado em função de dois parâmetros: o patamar $\varphi_1 + \varphi_2$, limite superior para onde os valores da semivariância tendem com o aumento dos valores de (h) e a amplitude $h = a$, distância a partir da qual os valores de $\gamma(h)$ param de crescer e são iguais a um patamar que é normalmente coincidente com a variância de $Z(x)$ (SOARES, 2014).

A Equação 10 representa o modelo esférico de função semivariância (Figura 3) com efeito pepita φ_1 , patamar $\varphi_1 + \varphi_2$ e φ_3 que é uma função do alcance prático $a = g(\varphi_3)$.

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ \varphi_1 + \varphi_2 \left[1 - \frac{3}{2} \left(\frac{h}{\varphi_3} \right) - \frac{1}{2} \left(\frac{h}{\varphi_3} \right)^3 \right], & 0 < h \leq \varphi_3 \\ \varphi_1 + \varphi_2, & h > \varphi_3 \end{cases} \quad (10)$$

A função de covariância para o modelo esférico é dada pela Equação (11).

$$C(h) = \begin{cases} \varphi_1 + \varphi_2, & h = 0 \\ \varphi_2 \left[1 - \frac{3}{2} \left(\frac{h}{\varphi_3} \right) - \frac{1}{2} \left(\frac{h}{\varphi_3} \right)^3 \right], & 0 < h \leq \varphi_3 \\ 0, & h > \varphi_3 \end{cases} \quad (11)$$

Para o modelo esférico, tem-se que a função de correlação espacial é dada pela Equação (12).

$$\rho(h) = \begin{cases} 1, & h = 0 \\ 1 - \frac{3}{2} \left(\frac{h}{\varphi_3} \right) - \frac{1}{2} \left(\frac{h}{\varphi_3} \right)^3, & 0 < h \leq \varphi_3 \end{cases} \quad (12)$$

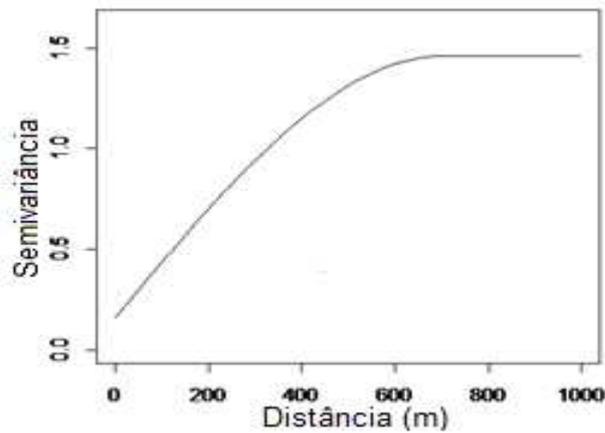


Figura 3 Representação gráfica do modelo esférico.

Este modelo é utilizado em \mathbb{R} , \mathbb{R}^2 e \mathbb{R}^3 (WEBSTER; OLIVER, 2007).

O modelo esférico é obtido pela escolha dos valores do efeito pepita φ_1 e do patamar $\varphi_1 + \varphi_2$. Assim, traça-se uma reta que intercepte o eixo Y em φ_1 e seja tangente aos primeiros pontos de $h=0$. A reta cruzará o patamar à distância, $a' = \frac{2}{3}a$. Desta forma, o alcance prático será $a = \frac{3a'}{2}$. Este modelo é linear até aproximadamente $\frac{1}{3}a$ (VIEIRA, 2000).

3.2.3.2. Modelo exponencial

A Equação 13 representa o modelo exponencial da função semivariância (Figura 4), com efeito pepita φ_1 , patamar $\varphi_1 + \varphi_2$ e φ_3 que é uma função do alcance prático $a = g(\varphi_3)$.

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ \varphi_1 + \varphi_2 \left[1 - \exp\left(-\frac{h}{\varphi_3}\right) \right], & 0 < h \leq d \end{cases} \quad (13)$$

Em que d é a máxima distância na qual o semivariograma é definido (VIEIRA, 2000).

A função de covariância para o modelo exponencial é dada pela Equação (14).

$$C(h) = \begin{cases} \varphi_1 + \varphi_2, & h = 0 \\ \varphi_2 \left[\exp\left(-\frac{h}{\varphi_3}\right) \right], & h > 0 \end{cases} \quad (14)$$

Tem-se que a função de correlação espacial do modelo exponencial é dada pela Equação (15).

$$\rho(h) = \begin{cases} 1, & h = 0 \\ \left[\exp\left(-\frac{h}{\varphi_3}\right) \right], & h > 0 \end{cases} \quad (15)$$

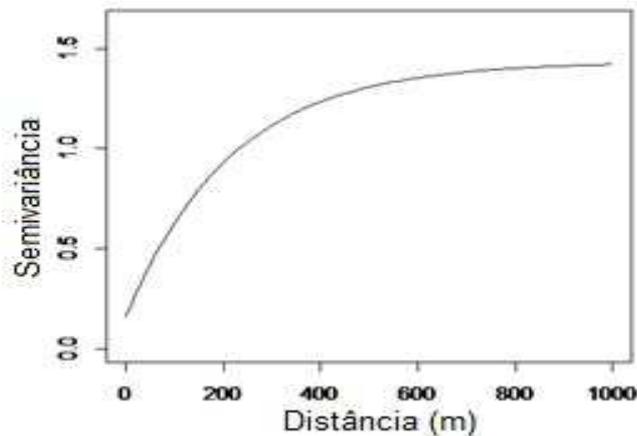


Figura 4 Representação gráfica do modelo exponencial.

Para Vieira (2000), um diferencial entre o modelo esférico e o exponencial é que o modelo esférico atinge o patamar no valor do alcance, enquanto o exponencial atinge apenas assintoticamente. O parâmetro a é determinado visualmente como a distância na qual o semivariograma se estabiliza.

O modelo exponencial atinge o patamar assintoticamente, com o alcance prático definido como a distância em que o valor do semivariograma é 95% de φ_2 , aproximadamente (ISAAKS; SRIVASTAVA, 1989; WEBSTER; OLIVER, 2007). Desta forma, seu alcance prático é $a = 3\varphi_3$.

Mesmo que a função semivariância do modelo exponencial possua um crescimento junto à origem de forma mais rápida do que o modelo esférico, as estruturas têm maior continuidade espacial, ou seja, maiores prolongamentos resultantes das maiores correlações espaciais e grandes distâncias h (SOARES, 2014). O modelo exponencial é válido em \mathbb{R}^d , com $d \geq 1$ (CRESSIE, 2015).

3.2.3.3. Modelo gaussiano

O modelo gaussiano é um modelo de transição constantemente utilizado para modelar fenômenos extremamente contínuos (ISAAKS; SRIVASTAVA, 1989). A Equação 16 representa o modelo gaussiano da função semivariância (Figura 5), com efeito pepita φ_1 , patamar $\varphi_1 + \varphi_2$ e φ_3 que é uma função do alcance prático $a = g(\varphi_3)$.

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ \varphi_1 + \varphi_2 \left\{ 1 - \exp \left[\left(-\frac{h}{\varphi_3} \right)^2 \right] \right\}, & h > 0 \end{cases} \quad (16)$$

Tem-se que a função de covariância do modelo gaussiano é dada pela Equação (17).

$$C(h) = \begin{cases} \varphi_1 + \varphi_2, & h = 0 \\ \varphi_2 \left\{ \exp \left[\left(-\frac{h}{\varphi_3} \right)^2 \right] \right\}, & h > 0 \end{cases} \quad (17)$$

A função da correlação espacial para o modelo gaussiano é dada pela Equação (18).

$$\rho(h) = \begin{cases} 1, & h = 0 \\ \exp \left[\left(-\frac{h}{\varphi_3} \right)^2 \right], & h > 0 \end{cases} \quad (18)$$

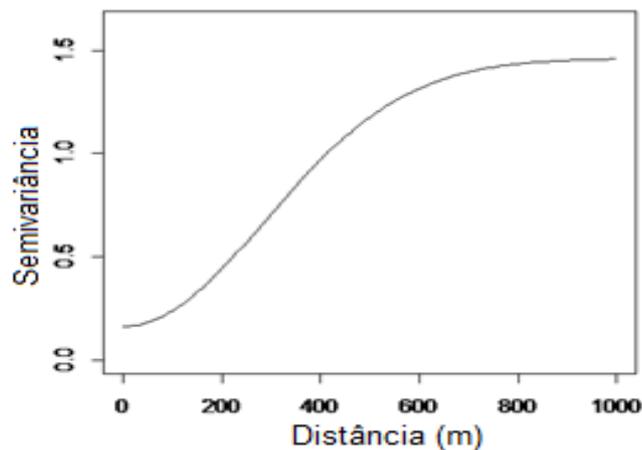


Figura 5 Representação gráfica do modelo gaussiano.

O modelo gaussiano atinge o patamar assintoticamente e pode ser considerado como tendo um alcance prático de aproximadamente de $a = \sqrt{3}\varphi_3$ (WEBSTER; OLIVER, 2007), onde atinge 95% da variância do patamar.

A característica distinta do modelo gaussiano é que ele possui um comportamento parabólico próximo à origem (ISAAKS; SRIVASTAVA, 1989; SOARES, 2014). Ainda segundo ISAAKS e SRIVASTAVA(1989), é o único modelo de transição apresentado cuja forma possui um ponto de inflexão.

3.2.3.4. Modelos da família Matérn

A função da família Matérn é apresentada por Matérn (1986). A Equação 19 apresenta a modelo da família Matérn, com efeito pepita φ_1 , patamar $\varphi_1 + \varphi_2$ e φ_3 que é uma função do alcance prático $a = g(\varphi_3)$.

$$\gamma(h) = \begin{cases} 0, & h = 0 \\ \varphi_1 + \varphi_2 \left[1 - (2^{k-1}\Gamma(k))^{-1} \left(\frac{h}{\varphi_3} \right)^k K_k \left(\frac{h}{\varphi_3} \right) \right], & h > 0 \end{cases} \quad (19)$$

em que, k é um parâmetro;

K_k é a função de Bessel de terceiro tipo, de ordem k , sendo $K_k(u) = \frac{1}{2} \int_0^\infty x^{k-1} e^{-\frac{1}{2}u(x+\frac{1}{x})} dx$, e esta função é válida para φ_3 e k maiores do que zero. Nesta família, o parâmetro k , chamado de ordem, consiste em um parâmetro de forma que determina a suavização analítica do processo subjacente $Z(s)$. Especificamente $Z(s)$ é $[k - 1]$ vezes diferenciáveis, em que $[k]$ denota o menor inteiro maior ou igual a k (DE BASTIANI, 2012).

A função da covariância do modelo da família Matérn é dada pela Equação (20).

$$C(h) = \begin{cases} \varphi_1 + \varphi_2, & h = 0 \\ \varphi_2 \left[(2^{k-1} \Gamma(k))^{-1} \left(\frac{h}{\varphi_3} \right)^k K_k \left(\frac{h}{\varphi_3} \right) \right], & h > 0 \end{cases} \quad (20)$$

A Equação 21 apresenta a função da correlação espacial do modelo da família Matérn.

$$\rho(h) = \{ (2^{k-1} \Gamma(k)) \}^{-1} \left(\frac{h}{\varphi_3} \right)^k K_k \left(\frac{h}{\varphi_3} \right), \quad h > 0 \quad (21)$$

A família Matérn é válida para a e $k > 0$ até o infinito, isso inclui os casos especiais, corresponde à função de correlação Exponencial quando $k = 0,5$ (WEBSTER; OLIVER, 2007); e à função de correlação Gaussiana quando $k \rightarrow \infty$. A Figura 6 apresenta o gráfico da família Matérn para $k = 0,5$; $k = 1$; $k = 1,5$; $k = 2$; $k = 2,5$.

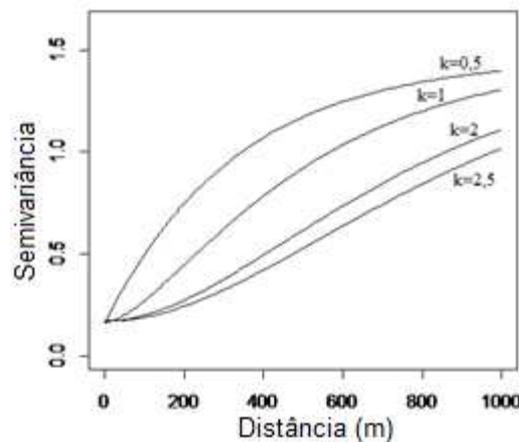


Figura 6 Representação gráfica do modelo da família Matérn, com $k = 0,5$; $k = 1$; $k = 1,5$; $k = 2$; $k = 2,5$.

3.2.4. Métodos de estimação de parâmetros por máxima verossimilhança - ML

Considere um processo gaussiano $\{Z(s), s \in S\}$, com distribuição gaussiana n -variada com vetor de médias $\mu \mathbf{1}$ e a matriz de covariância Σ , isto é, $\mathbf{Z} \sim N_n(\mu \mathbf{1}, \Sigma)$, em que μ é uma constante; $\mathbf{1}$ é um vetor de uns de dimensão $n \times 1$ e Σ é uma matriz definida positiva, com dimensão $n \times n$, dada por $\Sigma = [\text{Cov}(Z(s_i), Z(s_j))]$.

Os parâmetros desconhecidos $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\varphi}^T)^T$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ e $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \varphi_3)^T$, $i = j = 1, \dots, n$, do modelo linear espacial definido na Equação 1 podem ser estimados

maximizando a função de verossimilhança do logaritmo desta função Equação (22) (BORSSOI *et al.*, 2009),

$$l(\theta) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\Sigma| - \frac{1}{2}(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})^T \Sigma^{-1}(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta}). \quad (22).$$

A estimativa da máxima verossimilhança é um método estatístico bastante aceito e tem propriedades de otimização em grandes amostras (DIGGLE; RIBEIRO JR., 2007), além disso, consiste em maximizar o logaritmo da função verossimilhança, expresso por:

$$l(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\varphi}}) = \max(\boldsymbol{\beta}, \boldsymbol{\varphi}).$$

O estimador de máxima verossimilhança tem uma distribuição assintoticamente normal, imparcial e totalmente eficiente, dentro do cenário da geoestatística. Ademais, a implementação da estimativa da máxima verossimilhança é apenas direta quando os dados são gerados por um modelo gaussiano (DIGGLE; RIBEIRO JR., 2007). Ainda, este estimador é a combinação de parâmetros que produz o maior produto dessas probabilidades (BÁRDOSSY, 2008).

3.2.5. Desvios-padrões das estimativas

3.2.5.1. Função score

A função score do vetor $\theta = (\boldsymbol{\beta}^T, \boldsymbol{\varphi}^T)^T$ é dada por (URIBE-OPAZO; BORSOI; GALEA, 2012; De BASTIANI *et al.*, 2015) com o intuito de obter estimadores de máxima verossimilhança de $\theta = (\boldsymbol{\beta}^T, \boldsymbol{\varphi}^T)^T$, $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$ e $\boldsymbol{\varphi} = (\varphi_1, \varphi_2, \varphi_3)^T$ para o modelo espacial linear Gaussiano (4):

$$U(\boldsymbol{\beta}) = \frac{\partial l(\theta)}{\partial \boldsymbol{\beta}} = \mathbf{X}^T \Sigma^{-1}(\mathbf{Z} - \mathbf{X}\boldsymbol{\beta}),$$

$$U(\boldsymbol{\varphi}) = \frac{\partial l(\theta)}{\partial \boldsymbol{\varphi}} = (U(\varphi_1), \dots, U(\varphi_z))^T,$$

que possui como o j -ésimo elemento expresso por:

$$U(\varphi_j) = -\frac{1}{2} \left\{ \text{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_j} \right) + \frac{1}{2} \boldsymbol{\varepsilon}^T \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_j} \Sigma^{-1} \boldsymbol{\varepsilon} \right\}, \quad j = 1, 2, 3$$

Com $\boldsymbol{\varepsilon} = (\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})$.

3.2.5.2. Matriz de informação observada para modelos espaciais Gaussianos

A matriz de informação observada para o modelo espacial linear Gaussiano (4) é definida por $\mathbf{I}(\boldsymbol{\theta}) = -\mathbf{L}(\boldsymbol{\theta})$. Similarmente, a segunda matriz de derivadas é dada pela Equação (23) (URIBE-OPAZO; BORSOI; GALEA, 2012).

$$\mathbf{L}(\boldsymbol{\theta}) = \begin{pmatrix} L_{\beta\beta} & L_{\beta\varphi} \\ L_{\varphi\beta} & L_{\varphi\varphi} \end{pmatrix} \quad (23)$$

com,

$$\mathbf{L}_{\beta\beta} = \frac{\partial^2 l(\theta)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} = \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X},$$

$$\mathbf{L}_{\beta\boldsymbol{\varphi}} = \mathbf{L}_{\boldsymbol{\varphi}\beta}^T = \frac{\partial^2 l(\theta)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\varphi}^T},$$

com elementos $\frac{\partial^2 l(\theta)}{\partial \boldsymbol{\beta} \partial \varphi_j} = -\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}$, $j = 1, 2, 3$, sendo $\boldsymbol{\varepsilon} = (\mathbf{Z} - \mathbf{X}\boldsymbol{\beta})$.

$$\mathbf{L}_{\boldsymbol{\varphi}\boldsymbol{\varphi}} = \frac{\partial^2 l(\theta)}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}^T},$$

com elementos,

$$\begin{aligned} \frac{\partial^2 l(\theta)}{\partial \varphi_i \partial \varphi_j} &= \frac{1}{2} \text{tr} \left\{ \boldsymbol{\Sigma}^{-1} \left(\frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} - \frac{\partial^2 \boldsymbol{\Sigma}}{\partial \varphi_i \partial \varphi_j} \right) \right\} \\ &+ \frac{1}{2} \boldsymbol{\varepsilon}^T \boldsymbol{\Sigma}^{-1} \left\{ \frac{\partial^2 \boldsymbol{\Sigma}}{\partial \varphi_i \partial \varphi_j} - \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} - \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i} \right\} \boldsymbol{\Sigma}^{-1} \boldsymbol{\varepsilon}, \end{aligned}$$

para $i, j = 1, 2, 3$.

3.2.5.3. Matriz de informação

Os erros padrões assintóticos podem ser calculados ao se inverter a matriz de informação de Fisher (URIBE-OPAZO; BORSOI; GALEA, 2012). A matriz para o modelo espacial linear gaussiano (4) é dada por (LANGE; LITTLE; TAYLOR, 1989; MITCHELL, 1989):

$$\mathbf{I}_F(\boldsymbol{\theta}) = \mathbf{K}(\boldsymbol{\theta}) = \begin{pmatrix} \mathbf{K}(\boldsymbol{\beta}) & \mathbf{0} \\ \mathbf{0} & \mathbf{K}(\boldsymbol{\varphi}) \end{pmatrix}.$$

A qual é uma matriz diagonal bloco, donde:

$$\mathbf{K}(\boldsymbol{\beta}) = \mathbf{E}[-\mathbf{L}_{\beta\beta}] = \mathbf{E} \left[-\frac{\partial^2 l(\theta)}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}^T} \right] = \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X}$$

$$\mathbf{K}(\boldsymbol{\varphi}) = \mathbf{E}[-\mathbf{L}_{\boldsymbol{\varphi}\boldsymbol{\varphi}}] = \mathbf{E} \left[-\frac{\partial^2 l(\theta)}{\partial \boldsymbol{\varphi} \partial \boldsymbol{\varphi}^T} \right],$$

que tem como elementos:

$$k_{ij}(\boldsymbol{\varphi}) = \frac{1}{2} \text{tr} \left(\boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_i} \boldsymbol{\Sigma}^{-1} \frac{\partial \boldsymbol{\Sigma}}{\partial \varphi_j} \right), \quad i, j = 1, 2, 3.$$

Assim, ao se utilizar ML para estimar o parâmetro $\boldsymbol{\theta}$, pode-se utilizar o inverso da matriz de informação de Fisher para aproximar a matriz de covariância, que é uma medida natural do desempenho para a estimação de parâmetros (ZHU; STEIN, 2005).

A razão para isso é que o inverso da matriz de informação é, em condições de regularidade, a matriz de covariância assintótica do correspondente estimador de $\boldsymbol{\theta}$ (ZIMMERMAN, 2006).

Segundo Muller (2007), quando o vetor de parâmetros $\boldsymbol{\beta}^T$ é conhecido e fixa em $\boldsymbol{\beta} = \boldsymbol{\mu}_1$ resta apenas o vetor de parâmetros $\boldsymbol{\varphi}^T$ para ser estimado, logo:

$$I_F(\boldsymbol{\theta}) = I_F(\boldsymbol{\varphi}) = K(\boldsymbol{\varphi})$$

Assim, quando se maximiza o determinante da matriz de Fischer ($I_F(\boldsymbol{\theta})$) é o mesmo que:

$$\max |I_F(\boldsymbol{\theta})| = \max |K(\boldsymbol{\varphi})|.$$

Projetos de maximização do determinante de $K(\boldsymbol{\varphi})$ são sugeridos por Zhu e Stein (2005). Eles ainda propõem o uso do critério de *minimax* ou bayesiano para que se evitem efeitos indesejáveis devido à linearização, para a otimização de amostras.

Zhu e Stein (2005) utilizam o log do determinante da matriz inversa de informação de Fisher como critério para encontrar amostragem ótima.

$$V_0(S, \theta) = -\log |I_F(\theta, S)| \quad (24)$$

A Equação 24 depende do parâmetro desconhecido que se objetiva estimar. Logo, segundo Zhu e Stein (2005), a proposta de $S(\theta) = \arg \min_{S \subset D} V_0(S; \theta)$ chama-se o design otimizado localmente em θ . Desta forma, para se utilizar a Equação (24) diretamente como critério de projeto, precisa-se assumir a existência de uma estimativa preliminar para $\hat{\theta}_0$, e assim pode-se encontrar um projeto localmente ótimo, ao se avaliar a Equação 24 em $\hat{\theta}_0$:

$$S(\hat{\theta}_0) = \arg \min_{S \subset D} V_0(S, \hat{\theta}_0).$$

De outra forma:

$$S(\hat{\theta}_0) = \arg \min_{S \subset D} (-\log |I_F(\theta, S)|).$$

Todavia, quando os parâmetros $\boldsymbol{\beta}^T$ e $\boldsymbol{\varphi}^T$ não são conhecidos, é preciso maximizar o determinante da matriz de Fischer ($K(\boldsymbol{\theta})$), ou seja:

$$\max |I_F(\boldsymbol{\theta})| = \max |K(\boldsymbol{\theta})|.$$

3.2.6. Critério de validação dos ajustes de modelos

A escolha dos modelos teóricos ocorre de acordo com os métodos de validação, quando são realizadas comparações entre os valores teóricos do modelo geoestatístico e os valores empíricos adquiridos por amostragem. Baseado na análise desses erros de estimação, o melhor modelo de semivariância poderá ser selecionado (FARACO *et al.*, 2008). Sendo assim, na sequência, apresenta-se o critério de validação cruzada.

3.2.6.1. Validação cruzada

A validação cruzada é uma técnica de erros de estimativas, que possibilita a comparação entre valores estimados e amostrados (ISAAKS; SRIVASTAVA, 1989). As comparações são importantes e ajudam na escolha do melhor modelo da função semivariância. Em certa localização, $s_i Z(s_i)$, que é o valor da amostra, será ocultado do conjunto de dados, e assim, a partir do conjunto de dados restante, será modelada uma função para a semivariância e o valor de $Z(s_i)$ será predito por krigagem no local, obtendo-

se $\hat{Z}(s_i)$, de acordo com o modelo proposto (OLIVER; WEBSTER, 2014; DIAS; ALVES; SANCHES, 2015).

Logo, depois de calculada a estimativa, pode-se compará-la ao valor real da amostra que inicialmente tinha sido removido do conjunto de dados. Este processo deve ser repetido para todas as amostras do conjunto de dados (ISAAKS; SRIVASTAVA, 1989). Portanto, o erro médio para a validação cruzada (EM) é dado pela seguinte Equação.

$$EM = \frac{1}{n} \sum_{i=1}^n (Z(s_i) - \hat{Z}(s_i)), \quad (25)$$

em que, n é o número de valores amostrados da variável georreferenciada; $Z(s_i)$ é o valor da amostra observada no ponto s_i e $\hat{Z}(s_i)$ é o predito por krigagem no ponto s_i sem considerar na amostra $Z(s_i)$.

Mcbratney e Webster (1986); Cressie (2015) apresentam o erro médio reduzido (ER), o desvio padrão dos erros médios (S_E), o desvio padrão dos erros reduzidos (S_{ER}) e o erro absoluto (EA), como medidas para avaliar os modelos. O erro médio reduzido (ER) é definido pela Equação (26).

$$\overline{ER} = \frac{1}{n} \sum_{i=1}^n \frac{(Z(s_i) - \hat{Z}(s_i))}{\sigma(\hat{Z}(s_i))}, \quad (26)$$

em que $\sigma(\hat{Z}(s_i))$ é o desvio padrão da krigagem no ponto s_i , sem considerar a observação $Z(s_i)$.

O desvio padrão dos erros reduzidos é definido pela Equação (27).

$$S_{ER} = \sqrt{\frac{1}{n} \sum_{i=1}^n \frac{|Z(s_i) - \hat{Z}(s_i)|}{\sigma(\hat{Z}(s_i))}}. \quad (27)$$

Ainda segundo os autores, quando se aplica a condição de não tendenciosidade, o valor para o erro médio reduzido deve ser zero e o valor do desvio padrão do erro reduzido igual a um. Logo, os valores para EM e \overline{ER} devem ser mais próximos de zero, o valor de S_E o menor possível e o valor de S_{ER} mais próximo de um (FARACO *et al.*, 2008), os quais são os critérios para a escolha do melhor modelo ajustado.

O erro absoluto (EA) é uma medida da magnitude dos erros na unidade variável, e deve ser o menor possível (FARACO *et al.*, 2008). Assim, se o conjunto de valores medidos e preditos por krigagem $Z(s_i)$ e $\hat{Z}(s_i)$ forem conhecidos, é possível determinar o erro absoluto na unidade variável estudada, dado pela Equação (28).

$$EA = \sum_{i=1}^n |Z(s_i) - \hat{Z}(s_i)|, \quad (28)$$

em que, n é o número de valores amostrados da variável georreferenciada; $Z(s_i)$ é o valor da amostra observada no ponto s_i ;

$\hat{Z}(s_i)$ é o predito por krigagem no ponto s_i , sem considerar na amostra $Z(s_i)$.

Vieira (2000) leva em consideração para a qualidade da estimação somente duas estatísticas de diagnósticos o ER e EA.

3.3. Krigagem ordinária

Uma fase importante para a descrição e análise da variabilidade espacial de uma variável é a construção de mapas. A construção é possível a partir de um método de interpolação geoestatístico chamado krigagem. O termo krigagem é derivado da homenagem dada por Matheron (1963) ao matemático sul-africano, Daniel G. Krige (VIEIRA, 2000), pois seus trabalhos inspiraram Matheron.

Krigagem é um nome dado a um grupo de técnicas de estimação, as quais são baseadas na minimização da variância do erro de estimação (BRAGA, 1990). A krigagem utiliza informações a partir da função semivariância para encontrar os pesos ótimos que são atribuídos no estimador de krigagem aos valores observados de variáveis georreferenciadas nas amostras. O uso dos valores da função semivariância para a estimativa de valores da variável georreferenciada, em locais não amostrados por krigagem, não exige que o conjunto de amostras tenha distribuição normal. Mas, a presença de distribuição assimétrica, com muitos valores anômalos, deve ser considerada, pois a krigagem é um estimador linear.

A diferença entre krigagem e outros métodos de interpolação é a forma como os pesos são atribuídos às distintas amostras (CÂMARA; MEDEIROS, 1998). A krigagem em AP é uma técnica de interpolação muito utilizada para estimativa dos valores de propriedade de solo e produtividade em locais não amostrados, considerando a dependência espacial entre valores amostrais vizinhos destas variáveis, expresso no semivariograma (DALLACORT *et al.*, 2012).

É possível estimar valores desconhecidos $Z(s_0)$, a partir da técnica de krigagem, para qualquer localização s_0 , onde não existam valores medidos, e que a estimativa deva ser uma combinação linear dos valores amostrais observados (VIEIRA, 2000).

Existem diversos tipos de krigagem, dentre elas a krigagem ordinária. A krigagem ordinária é o método mais utilizado pela sua simplicidade e resultados que proporciona. É um método local de estimativa e, desta maneira, a estimativa em um ponto não conhecido resulta na combinação linear de n valores encontrados na vizinhança próxima (LANDIM; YAMAMOTO, 2013).

O estimador da krigagem ordinária é definido pela Equação (29) (ISAAKS; SRIVASTAVA, 1989; SOARES, 2014).

$$Z^*(\mathbf{s}_0) = \sum_{i=1}^n \lambda_i Z(\mathbf{s}_i), \quad (29)$$

em que:

$Z^*(\mathbf{s}_0)$ é o valor estimado para a localização \mathbf{s}_0 não amostrado;

n é o número de valores medidos $Z(\mathbf{s}_i)$ das variáveis em pontos amostrados e

λ_i são os pesos associados a cada valor $Z(\mathbf{s}_i)$ medido.

Assim, o que torna a krigagem um ótimo interpolador, é a forma como os pesos são distribuídos. Desta forma, o estimador não pode ser tendencioso e deve ter variância mínima (VIEIRA, 2000).

A Equação (30) representa o erro que se comete ao estimar o valor de $Z(\mathbf{s})$ (JOURNEL; HUIJBREGTS, 1978).

$$E[Z^*(\mathbf{s}_0) - Z(\mathbf{s}_0)] = 0. \quad (30)$$

Logo, quando se substitui a Equação (29) na Equação (30), tem-se que:

$$E[Z^*(\mathbf{s}_0) - Z(\mathbf{s}_0)] = E \left[\sum_{i=1}^n \lambda_i Z(\mathbf{s}_i) - Z(\mathbf{s}_0) \right] = 0.$$

E, quando se aplica a linearidade do operador do valor esperado, tem-se:

$$E[Z^*(\mathbf{s}_0) - Z(\mathbf{s}_0)] = \sum_{i=1}^n \lambda_i E[Z(\mathbf{s}_i)] - E[Z(\mathbf{s}_0)] = 0.$$

Levando-se em consideração a condição de estacionaridade ($E[Z(\mathbf{s}_i)] = m$) $\forall i = 1, \dots, n$, tem-se que:

$$E[Z(\mathbf{s}_i)] = E[Z(\mathbf{s}_0)] = m.$$

Logo,

$$E[Z^*(\mathbf{s}_0) - Z(\mathbf{s}_0)] = \sum_{i=1}^n \lambda_i m - m = m \left(\sum_{i=1}^n \lambda_i - 1 \right) = 0. \quad (31)$$

Desta forma, para que a Equação (28) seja verdadeira para qualquer m , é necessário que

$$\sum_{i=1}^n \lambda_i - 1 = 0,$$

Isto é,

$$\sum_{i=1}^n \lambda_i = 1.$$

Desta forma, para que a estimativa não seja tendenciosa, é necessário que a soma dos pesos tenha valor igual a um, qualquer que seja a distribuição de seus valores (VIEIRA, 2000).

Outro critério para a qualidade do estimador é a variância do erro, denotada por $\varepsilon(s_0)$. Dois estimadores podem apresentar a média de $\varepsilon(s_0)$ nula, porém, a dispersão da média determina a diferença da qualidade dos estimadores (SOARES, 2014).

A variância de estimativa ou a variância do erro de estimativa é definida pela Equação (32) (VIEIRA, 2000).

$$\varepsilon(s_0) = \text{var} [Z^*(s_0) - Z(s_0)]. \quad (32)$$

Logo,

$$\varepsilon(s_0) = \text{var} [Z^*(s_0) - Z(s_0)] = E\{[Z^*(s_0) - Z(s_0)]^2\}.$$

Decompondo os termos:

$$\begin{aligned} \varepsilon(s_0) &= E[[Z^*(s_0) - Z(s_0)] \cdot [Z^*(s_0) - Z(s_0)]] \\ &= E[Z^{*2}(s_0) - Z^*(s_0)Z(s_0) - Z(s_0)Z^*(s_0) + Z^2(s_0)] \end{aligned}$$

$$\varepsilon(s_0) = E[Z^{*2}(s_0) - 2Z^*(s_0)Z(s_0) + Z^2(s_0)]. \quad (33)$$

Assim, quando se substitui a Equação (29) na Equação (33), tem-se que:

$$\varepsilon(s_0) = E \left\{ \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} Z(s_{\alpha}) Z(s_{\beta}) \right\} - 2E \left\{ \sum_{\alpha=1}^n \lambda_{\alpha} Z(s_{\alpha}) Z(s_0) \right\} + E\{Z^2(s_0)\}.$$

Após definido o modelo de covariância (ou semivariograma), válido para todo o campo de análise de $Z(s)$, pode-se escrever a expressão da variância de estimação de qualquer estimador em função das covariáveis entre as amostras e das covariáveis entre as amostras e o ponto a se estimar, a expressão é definida na Equação (34) (SOARES, 2014).

$$\varepsilon(s_0) = C(0) + \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C(s_{\alpha}, s_{\beta}) - 2 \sum_{\alpha=1}^n \lambda_{\alpha} C(s_{\alpha}, s_0). \quad (34)$$

A Equação (34) deve ser minimizada sob a restrição de $\sum_{i=1}^n \lambda_i = 1$, assim a minimização pode ser dada pelas técnicas de Lagrange, que implica em adicionar mais uma equação $\sum_{\alpha=1}^n \lambda_{\alpha} = 1$ e, conseqüentemente, mais uma incógnita (o parâmetro de Lagrange μ) à Equação (34). Logo tem-se que (SOARES, 2014):

$$\varepsilon(s_0) = C(0) + \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C(s_{\alpha}, s_{\beta}) - 2 \sum_{\alpha=1}^n \lambda_{\alpha} C(s_{\alpha}, s_0) + 2\mu \left[\sum_{\alpha=1}^n \lambda_{\alpha} - 1 \right]. \quad (35)$$

Sendo $2\mu [\sum_{\alpha=1}^n \lambda_{\alpha} - 1] = 0$ o termo adicional nulo e $\alpha = \beta = 1, \dots, n$.

A minimização da Equação (32) equivale, assim, no cálculo das $n + 1$ derivadas parciais em relação à λ_{α} e μ , e, quando são igualadas a zero, tem-se um sistema de $n + 1$ equações a $n + 1$ incógnitas, da qual a solução resulta em ponderadores λ_{α} , que desempenham a condição de não enviesamento $\sum_{\alpha=1}^n \lambda_{\alpha} = 1$ ($\alpha = 1, \dots, n$) e, ao mesmo tempo, minimizam a variância de estimação (SOARES, 2014):

$$\frac{\partial [E\{[[Z(s_0)]^* - Z(s_0)]^2\} + 2\mu(\sum_{\alpha} \lambda_{\alpha} - 1)]}{\partial \lambda_{\alpha}} = 0, \quad \alpha = 1, \dots, n$$

$$\frac{\partial [E\{[[Z(s_0)]^* - Z(s_0)]^2\} + 2\mu(\sum_{\alpha} \lambda_{\alpha} - 1)]}{\partial \mu_{\alpha}} = 0.$$

O desenvolvimento das n primeiras equações conduz a:

$$\frac{\partial [C(0) + \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} C(s_{\alpha}, s_{\beta}) - 2 \sum_{\alpha=1}^n \lambda_{\alpha} C(s_{\alpha}, s_0) + 2\mu (\sum_{\alpha=1}^n \lambda_{\alpha} - 1)]}{\partial \lambda_{\alpha}} = 0,$$

$$\alpha = \beta = 1, \dots, n.$$

$$2 \sum_{\beta=1}^n \lambda_{\beta} C(s_{\alpha}, s_{\beta}) - 2C(s_{\alpha}, s_0) + 2\mu = 0, \quad \alpha = 1, \dots, n \quad (36)$$

A última derivada parcial em relação a μ conduz à equação:

$$\sum_{\alpha=1}^n \lambda_{\alpha} = 1, \quad \alpha = 1, \dots, n.$$

Desta forma, o sistema de krigagem de $n + 1$ equações, cuja solução fornece os n ponderados de λ_{α} , é dado por:

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta} C(s_{\alpha}, s_{\beta}) + \mu = C(s_{\alpha}, s_0), & \alpha = 1, \dots, n, \\ \sum_{\alpha=1}^n \lambda_{\alpha} = 1 \end{cases} \quad (37)$$

O valor mínimo da variância de estimação é obtido pela substituição da Equação (37) na equação da variância de estimação Equação (34):

$$\sigma_E^2(s_0) = C(0) + \sum_{\alpha=1}^n \lambda_{\alpha} C(s_{\alpha}, s_0) - \mu - 2 \sum_{\alpha=1}^n \sum_{\alpha} \lambda_{\alpha} C(s_{\alpha}, s_0),$$

$$\sigma_E^2(s_0) = C(0) - \sum_{\alpha=1}^n \lambda_{\alpha} C(s_{\alpha}, s_0) - \mu.$$

O sistema de krigagem também pode ser descrito em função do semivariograma $\gamma(h)$, tendo em conta que $\gamma(h) = C(0) - C(h)$.

$$\begin{cases} \sum_{\alpha=1}^n \lambda_{\beta} \gamma(s_{\alpha}, s_{\beta}) + \mu = \gamma(s_{\alpha}, s_0), & \alpha = 1, \dots, n \\ \sum_{\alpha=1}^n \lambda_{\alpha} = 1. \end{cases}, \quad (38)$$

Logo, a variância de estimação é dada pela Equação (39),

$$\sigma_E^2(s_0) = \sum_{\alpha=1}^n \lambda_{\alpha} \gamma(s_{\alpha}, s_0) + \mu \quad \alpha = 1, \dots, n. \quad (39)$$

O sistema de $n + 1$ equações também pode ser descrito da forma matricial, quando se designam a matriz de covariâncias entre amostras \mathbf{K} , \mathbf{M} , a matriz do segundo membro – covariância entre amostras, o ponto a se estimar e λ a matriz dos pesos ponderados:

$$\mathbf{K} = \begin{bmatrix} C(s_1, s_1) & C(s_1, s_2) & \dots & C(s_1, s_N) \\ C(s_2, s_1) & C(s_2, s_2) & \dots & C(s_2, s_N) \\ \vdots & \vdots & \vdots & \vdots \\ C(s_N, s_1) & C(s_N, s_2) & \dots & C(s_N, s_N) \\ 1 & 1 & \dots & 1 \end{bmatrix},$$

$$\mathbf{M} = \begin{bmatrix} C(s_1, s_0) \\ C(s_2, s_0) \\ \vdots \\ C(s_N, s_0) \\ 1 \end{bmatrix},$$

$$\lambda = \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \\ \mu \end{bmatrix}.$$

Assim, o sistema de krigagem fica igual a

$$\mathbf{K} \cdot \lambda = \mathbf{M}$$

Na qual a solução resulta, após a inversão de \mathbf{K} , em:

$$\lambda = \mathbf{K}^{-1} \cdot \mathbf{M}$$

e

$$\sigma_E^2(s_0) = C(0) - \lambda^T \cdot \mathbf{M}.$$

Quando se define $[\mathbf{Z}]$ como vetor dos valores $Z(s_\alpha)$, $[\mathbf{Z}] = [Z(s_1), \dots, Z(s_n)]$, então o estimador $[Z(s_0)]^*$ é igual a:

$$[Z(s_0)]^* = \lambda^T \cdot \mathbf{Z} = \mathbf{M}^T \cdot \mathbf{K}^{-1} \cdot \mathbf{Z}.$$

3.4. Comparação de mapas temáticos

3.4.1. Matriz de erros

A Matriz de erros é uma técnica utilizada para comparar mapas temáticos, na qual um deles é denominado mapa de referência e o outro mapa modelo, a qual utiliza indicadores de acurácia, derivados da matriz de erros, matriz $k \times k$ (Tabela 1) e tem como unidade de medida o *pixel* (De BASTIANI; URIBE-OPAZO; DALPOSSO, 2012).

Nesta matriz, os *pixels* no mapa de referência são quantificados nas colunas, enquanto os *pixels* do mapa modelo são quantificados nas linhas. Cada elemento da matriz reflete a área que pertence à classe m do mapa modelo e à j do mapa de referência (DALPOSSO *et al.*, 2012). É importante que, para comparar o mapa modelo com o mapa de referência, ambos devem possuir as mesmas classes.

Tabela 1 Matriz de erros de ordem $k \times k$.

		Mapa de referência				
		C_1	C_2	...	C_k	TOTAL
Mapa modelo	C_1	n_{11}	n_{12}	...	n_{1k}	$\sum_{j=1}^k n_{1j}$
	\vdots	\vdots	\vdots	\ddots	\vdots	\vdots
	C_k	n_{k1}	n_{k2}	...	n_{kk}	$\sum_{j=1}^k n_{kj}$
TOTAL		$\sum_{n=1}^k n_{m1}$	$\sum_{n=1}^k n_{m2}$...	$\sum_{n=1}^k n_{mk}$	N

N é o número total de pixels; k é o número de classes; C_k é a classe k; $\sum_{j=1}^k n_{kj}$ é a soma dos pixels na classe C_k do mapa modelo; $\sum_{n=1}^k n_{mk}$ é a soma dos pixels na classe C_k do mapa de referência.

A diagonal principal representa as situações em que as áreas tiveram a mesma classificação para os dois mapas. Já os elementos fora da diagonal principal representam as classificações distintas dos mapas, ou seja, que não coincidem (DE BASTIANI; URIBE-OPAZO; DALPOSSO, 2012).

A partir dos elementos da matriz de erros são obtidos os índices de similaridade, exatidão global (EG) (CONGALTON; GREEN, 1999; ANDERSON *et al.* 2001), o índice Kappa (K) (CONGALTON; GREEN, 1999; KRIPENDORFF, 2013) e o índice Tau (T) (MA; REDMOND (1995); KRIPENDORFF, 2013).

3.4.2. Exatidão Global

A exatidão global é uma das medidas mais simples para se mensurar a similaridade do mapa de referência com o mapa modelo (DEBASTIANI; URIBE-OPAZO; DALPASSO, 2012). É calculada pela soma total dos pixels classificados corretamente, dividida pelo número total de *pixels* da matriz de erros (CONGALTON; GREEN, 1999), conforme a Equação (40).

$$EG = \frac{\sum_{i=1}^k n_{ii}}{N}, \quad (40)$$

em que n_{ii} são os elementos da diagonal principal da matriz de informação, os quais representam as áreas que tiveram a mesma classificação para os dois mapas; k é o número de classes e N é o total de *pixels*.

De acordo com Anderson *et al.*, (2001), o nível mínimo de precisão de interpretação deve ser 85%, para que os mapas apresentem similaridade.

3.4.3. Índice Kappa

O índice Kappa é utilizado para mensurar a exatidão das classificações temáticas e é indicado como uma medida adequada da exatidão por levar em consideração todos os elementos da matriz de erros. Apresenta uma medida de concordância entre os valores do

mapa de referências e os valores do mapa modelo, dados pela Equação (41) (CONGALTON; GREEN, 1999).

$$K = \frac{N \sum_{i=1}^k n_{ii} - \sum_{i=1}^k n_{i \bullet} \cdot n_{\bullet i}}{N^2 - \sum_{i=1}^k n_{i \bullet} \cdot n_{\bullet i}}, \quad (41)$$

em que, $n_{i \bullet} \cdot n_{\bullet i}$ é a soma total da linha i multiplicado pela soma total da coluna i ; N é o total de pixels e n_{ii} são os elementos da diagonal principal da matriz de erros, na respectiva linha e coluna i .

O índice Kappa é classificado como de baixa exatidão se $K < 0,67$, como média exatidão se $0,67 \leq K < 0,80$ e alta exatidão se $K \geq 0,80$ (KRIPENDORFF, 2013).

3.4.4. Índice de concordância Tau

Ma e Redmond (1995) introduziram o coeficiente de Tau, sendo esse menos subjetivo e de mais fácil compreensão e utilização, que mede a melhoria de uma classificação sobre uma atribuição aleatória de pixels a grupos. Também é conhecido como Kappa modificado por ser parecido com o índice Kappa, diferenciando-se desse por utilizar em sua formulação as probabilidades a priori para cada classe p_k . Quando essas probabilidades a priori para as classes forem iguais, tem-se $p_k = 1/k$, onde k é o número de classes. Este índice é calculado pela Equação (42).

$$T = \frac{\frac{\sum_{i=1}^k n_{ii}}{N} - p_k}{1 - p_k}, \quad (42)$$

em que n_{ii} são os elementos da diagonal principal da matriz de erros, na respectiva linha e coluna i e p_k é a probabilidade a priori quando as classes são iguais.

A classificação do índice Tau segue a mesma classificação do índice Kappa, segundo Kripendorff (2013).

3.5. Esquema de amostragem espacial

Uma amostra é um subconjunto de valores do fenômeno espacial. Se representativa, ela deve apresentar a variabilidade espacial, tanto em tamanho, ou seja, número de pontos, como em termos de distribuição dos pontos no domínio a ser estudado. Assim, qualquer estimativa que se baseia nestes pontos amostrais, está, porém, a uma incerteza, e nessa perspectiva, a metodologia geoestatística se destaca por oferecer a incerteza associada à estimativa (LANDIM; YAMAMOTO, 2013).

A ideia básica de um método de amostragem espacial é inferir as propriedades de uma população usando uma amostra geograficamente distribuída ao longo do espaço, segundo uma estatística adequada (WANG *et al.*, 2013).

3.5.1. Amostragem aleatória simples

Em estatística, na amostragem aleatória simples, a população constituída por n unidades é numerada sequencialmente e, desta forma, n unidades serão sorteadas sem reposição. Assim, a componente aleatória é o número sequencial escolhido entre 1 e n . Todavia, em estudos que envolvem geoestatística, as observações são realizadas em pontos de amostragem com localização dentro da região em estudo e, assim, as coordenadas geográficas a serem escolhidas casualmente são a componente aleatória (LANDIM; YAMAMOTO, 2013).

Um exemplo da amostragem aleatória simples é apresentado na Figura 7.

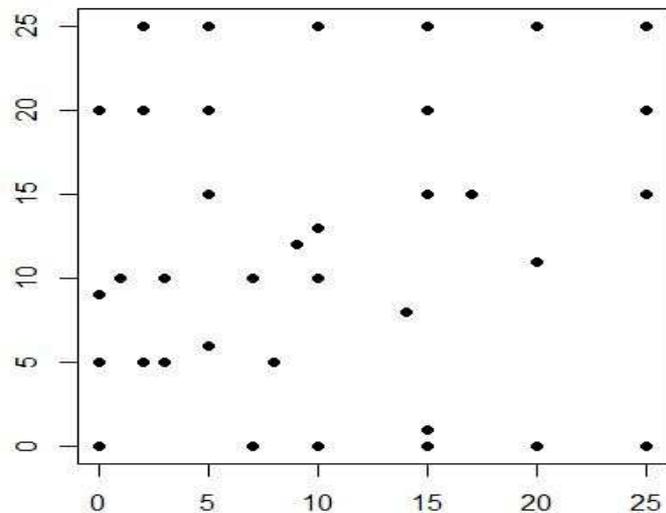


Figura 7 Exemplo da amostragem aleatória simples.

A amostragem aleatória simples garante que o projeto seja independente do fenômeno espacial subjacente de interesse. Entretanto, a partir de uma perspectiva espacial, este projeto é potencialmente ineficiente, pois pode levar a uma cobertura muito desigual na área em estudo (DIGGLE; RIBEIRO JR., 2007).

3.5.2. Amostragem sistemática

A amostragem sistemática é realizada com base em uma malha amostral regular já definida, de forma aleatória. Na teoria, a componente aleatória seria dada pela escolha do ponto de origem, mas, na prática, isso não ocorre, pois, a malha regular é definida inicialmente pelo responsável pela amostragem para otimizar a coleta das unidades da região de estudo (LANDIM e YAMAMOTO, 2013).

Um exemplo da amostragem sistemática ou regular é apresentado na Figura 8.

A cobertura desigual na área de estudo é dada como um problema que não ocorre na amostragem regular, haja vista existir uma cobertura uniforme na área (DIGGLE; RIBEIRO JR., 2007; CHIPETA *et al.*, 2016). Ainda tem-se que, do ponto de vista da predição

espacial, um projeto de estrutura é geralmente mais eficiente do que um projeto completamente aleatório (DIGGLE; RIBEIRO JR., 2007).

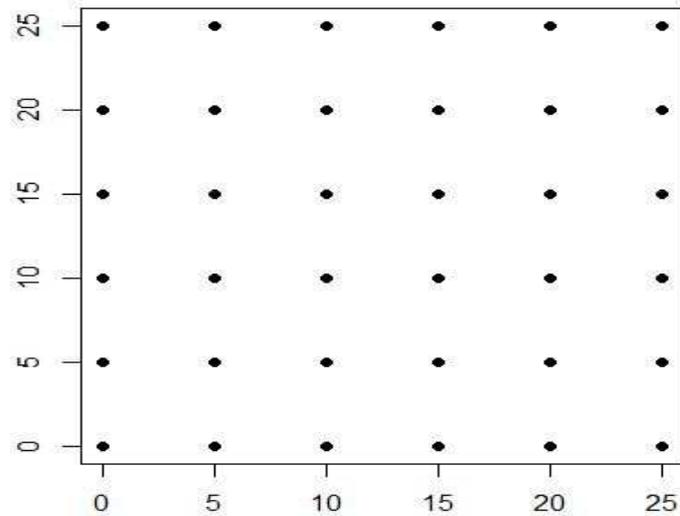


Figura 8 Exemplo da amostragem sistemática ou regular.

A amostragem regular pode utilizar grades amostrais triangulares, quadradas ou hexagonais (CHIPETA *et al.*, 2016). Uma malha triangular é um conjunto de poliedros cujas faces são triângulos, as faces da malha quadrada são quadrados e as faces da malha hexagonal são hexágonos.

3.5.3. Amostragem *lattice plus close pairs*

A amostragem *lattice plus close pairs* ($k \times k$, m , δ) (Figura 9) consiste em uma amostragem regular $k \times k$ com espaçamento Δ adicionando outros m pontos, e cada um deles está localizado de forma uniformemente aleatória dentro de uma circunferência de raio δ , cujo centro sempre está em um ponto da grade selecionado aleatoriamente (DIGGLE; RIBEIRO JR., 2007).

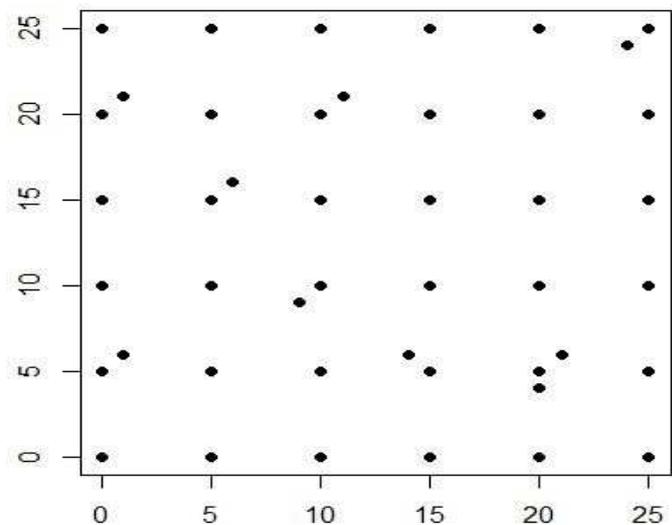


Figura 9 Exemplo da amostragem *lattice plus close pairs*.

3.6. Otimização de amostras espaciais

Os problemas de otimização são de maximização ou minimização de uma função, isto é, o maior ou o menor valor que uma função pode assumir em um intervalo. Estes problemas baseiam-se principalmente em três pontos:

1. Codificação do problema: matematicamente, a escolha da codificação para um problema de otimização é a função ou regra que associa os elementos do espaço codificado, com aqueles do espaço de busca.
2. Função objetivo que se deseja maximizar ou minimizar: é a representação matemática do critério de eficiência do problema de otimização, ou seja, é a função que se quer otimizar, minimizando-a ou maximizando-a.
3. Espaço de soluções: é o conjunto, espaço ou região que contém as possíveis ou viáveis soluções do problema a ser otimizado. Também pode ser chamado de domínio viável do problema.

3.6.1. Critério de otimização

Quando a amostragem é realizada de maneira rápida e gera baixos custos, pode-se obter assim um grande número de amostras. Desta forma, há pouco interesse em um algoritmo de otimização da seleção de amostras. Entretanto, na prática, o número de localizações amostrais é um fator limitante, pois a amostragem, normalmente, tem um custo alto ou é demorada, assim, é necessário um critério de otimização da configuração amostral que leve em conta os custos da realização do experimento e a eficiência dos resultados pelo experimento.

3.6.2. Algoritmos selecionados no processo de otimização: algoritmo genético.

3.6.2.1. Histórico

O algoritmo genético (AG) é um ramo dos algoritmos evolucionários e como tal pode ser definido como uma técnica de busca que se baseia no processo da evolução natural das espécies (PESSOA *et al.*, 2015). Eles fazem parte de uma classe de paradigmas e técnicas computacionais, denominada computação evolucionista.

Segundo Filitto (2008), os algoritmos genéticos são utilizados para resolver problemas de busca e otimização, encontrados no mundo real. Eles buscam a melhor solução para os problemas de otimização, por meio de um processo iterativo. Segundo Zini *et al.* (2014), os algoritmos genéticos estão se expandindo por toda a comunidade científica, assim, são obtidas diversas aplicações dos algoritmos que podem ajudar a resolver problemas importantes, como no roteamento de veículos (HEINEN; OSÓRIO, 2006), no setor de petróleo e gás, no setor musical, em telecomunicações, na medicina (REDUSINO, 2010), estimação de componentes harmônicas (PESSOA *et al.*, 2015) e planejamento de

geração distribuída em sistemas elétricos de distribuição, que envolvem energia térmica baseada em biomassa e energia eólica (OLIVEIRA *et al.*, 2016).

Segundo Linden (2012), a história dos algoritmos genéticos tem início na década de 1940, quando cientistas se inspiraram na natureza para criar o ramo da inteligência artificial. Até o final da década de 1950, a pesquisa se desenvolveu mais nos ramos da pesquisa cognitiva e na compreensão dos processos de raciocínio e aprendizagem. A partir daí, a pesquisa começou a buscar modelos de sistemas genéticos que pudessem gerar soluções candidatas para problemas com difícil desenvolvimento computacional.

Em 1858, Charles Darwin desenvolveu sua teoria de evolução das espécies. Para ele, quanto mais adaptado o indivíduo for com o ambiente, maior será a sua chance de sobrevivência. Ou seja, sua teoria explica como o ambiente atua sobre os indivíduos, forçando a seleção dos mais aptos (GOLDBARG; LUNA, 2005). Assim, em 1900, com a teoria combinada com a genética e as ideias de Darwin e Alfred Wallace (co-inventor do conceito), foi criado o princípio de Genética Populacional (LUCAS, 2002; LINDEN, 2012). Porém, somente nas décadas de 1950 e 1960, os biólogos começaram a desenvolver simulações computacionais de sistemas genéticos. Apesar de não ser o primeiro a desenvolver atividades na área, John Holland é considerado o pai dos algoritmos genéticos, pois foi seu inventor (HOLLAND, 1975).

Holland estudou a evolução das espécies e desenvolveu um modelo heurístico computacional que, quando implementado, gerava boas soluções para problemas com grau de dificuldade elevado, que eram insolúveis computacionalmente até aquela época (LINDEN, 2012). O autor apresentou os algoritmos genéticos como uma metáfora para os processos evolutivos, de modo que seja possível estudar a evolução e adaptação de espécies, a partir de simulações dentro de um computador. Porém, os algoritmos genéticos transcenderam o papel imaginado originalmente por Holland e transformou-se em uma ferramenta de uso dos cientistas da computação.

Um fator que merece destaque no trabalho de Holland é que ele usou originalmente cromossomos binários, em que o indivíduo é representado por um vetor composto de zeros e uns. Esta limitação mais tarde foi eliminada por outros pesquisadores, mas é comum hoje cientistas utilizarem a representação binária. Contudo, mesmo quando há outras representações possíveis, como a inteira ou a real, ainda a representação binária é mais simples do que estas outras representações (LINDEN, 2012). A essa representação se dá o nome de alfabeto do AG.

3.6.2.2. Funcionamento

Segundo Rocha Jr. e Tostes (2013), seu funcionamento consiste em um processo iterativo, iniciado com uma população inicial aleatória de cromossomos (indivíduos), que são as possíveis soluções do problema. Os indivíduos são avaliados com o objetivo de

selecionar os mais aptos. Os indivíduos selecionados são recombinados a partir de operadores genéticos e, assim, gera-se uma nova população. Realiza-se este processo até encontrar a solução ótima, ou até atingir um critério de parada pré-estabelecido pelo pesquisador.

Um fluxograma genérico é capaz de englobar a maioria dos Algoritmos Genéticos (Figura 10). A figura mostra que, em um primeiro momento, uma população inicial de indivíduos que serão as possíveis soluções para o problema é gerada de forma aleatória. No momento seguinte, esta população é avaliada por uma função de avaliação (chamada também de função objetivo), no sentido de identificar boas soluções. O passo seguinte é verificar se um critério de parada pré-definido foi atingido. Se não, aplica-se o processo de seleção e reprodução nos indivíduos desta população. Os indivíduos selecionados passam pelos operadores genéticos, como o cruzamento e a mutação. Desta forma, gera-se uma nova população de filhos, a partir da geração anterior, que será avaliada novamente. Este processo é repetido até que o critério de parada seja atingido.

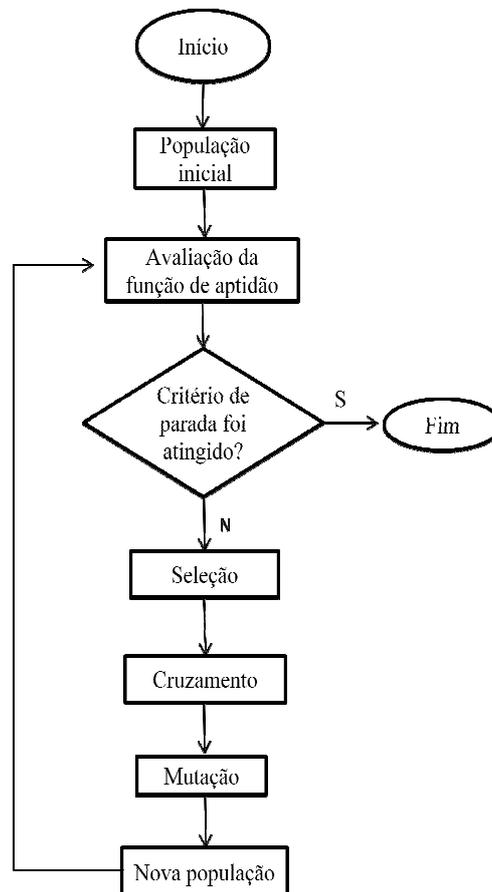


Figura 10 Fluxograma do funcionamento do Algoritmo Genético.

3.6.2.3. Vantagens

Algumas vantagens dos AGs sugeridas por Linden (2012) quando comparadas a outras técnicas de otimização:

- Não utilizam métodos de busca totalmente aleatórios;

- Necessitam somente de informação sobre o valor de uma função objetivo para cada membro da população;
- Não são afetados por descontinuidade na função ou na sua derivada, quando comparado a outras técnicas de otimização. Isto faz com que os AGs sejam muito adequados para funções com descontinuidade ou para os quais não temos como calcular uma derivada;
- São capazes de lidar com funções discretas e contínuas, podendo inclusive lidar com funções reais, discretas, categóricas e com funções mistas;
- São boas técnicas para atacar problemas com espaços de busca extremamente grandes, os quais não podem ser resolvidos por técnicas tradicionais;
- Operam em um espaço de soluções codificadas, e não necessariamente no espaço de busca diretamente;
- Apresentam resultados bastante aceitáveis, com relação à precisão e aos recursos empregados para ampla gama de problemas de difícil resolução por outros métodos.

Ainda, o AG possui uma grande vantagem pelo fato de trabalhar com o conceito de população, diferente de outros métodos que trabalham somente com um ponto. Assim, tendo uma população de pontos adaptados, diminui a possibilidade de alcançar um falso ótimo.

3.6.2.4. Representação genética e codificação

Segundo Linden (2012), a representação dos indivíduos é fundamental para o algoritmo genético. É a maneira como são traduzidas as informações do problema para uma forma viável a ser tratada pelo computador, pois ela é a principal responsável para o desempenho e execução do programa (LUCAS, 2002). Quanto mais adequada ela for ao problema, maior será a qualidade dos resultados obtidos. Qualquer representação empregada deve ser simples e sempre representar todos os indivíduos da população e esta é uma representação arbitrária. Se houver soluções proibidas então, preferencialmente elas não devem ter representação no algoritmo e se houver condições no problema, devem estar implícitas na representação.

Segundo Zini (2009), a maioria dos trabalhos desenvolvidos e apresentados na literatura com o AG utiliza a codificação binária, em que o indivíduo é representado por um vetor composto por 0's e 1's. Assim, a aplicação dos operadores genéticos é mais simples de ser utilizada com este tipo de codificação (ZINI *et al.*, 2014).

Para Bueno (2009), os principais tipos de representação ocorrem por: números binários, números reais e categóricas. E, na última representação, os conjuntos de valores preestabelecidos são codificados, como as direções e bases de DNA. As mais utilizadas são a binária, a inteira e a de ponto flutuante.

3.6.2.5. Geração da população inicial

O processo da escolha da população inicial pode ser realizado de diferentes formas. Normalmente depende do problema, sendo que os indivíduos são selecionados de acordo com uma regra probabilística.

A literatura na área sugere utilizar a estratégia mais simples possível de inicialização. Escolhem-se n indivíduos de forma aleatória e utiliza-se um gerador de números pseudo-aleatórios com distribuição uniforme em uma faixa previamente definida pelo usuário. Tal faixa é definida levando-se em consideração algum conhecimento prévio do problema a ser otimizado. Outra maneira é se dividir o espaço de busca em k espaços iguais e gerar n/k indivíduos em cada um destes espaços e assim garante-se a inclusão de indivíduos que possuem as principais características de todo o espaço (LINDEN, 2012).

O número de elementos que comporá a população depende do problema. O tamanho da população inicial é um parâmetro importante dos AGs. Barbosa (2014) diz que se deve encontrar um ponto de equilíbrio para o tamanho da população inicial, o qual não pode ser muito pequeno, pois assim o espaço de exploração das soluções fica restrito, reduzindo assim as possibilidades de se atingir o ótimo global. Mas, se a população for muito grande, isso acarretará em maiores esforços computacionais e um maior tempo de execução para chegar à solução ótima.

3.6.2.6. Função de avaliação

Depois de gerada a população inicial, os indivíduos precisam ser avaliados, para assim alguns darem continuidade à criação de outros que possam evoluir quanto às suas características (ZINI, 2009). A avaliação é denominada função de aptidão (ou função objetivo, ou ainda, *fitness*), e serve para medir o desempenho de cada indivíduo, com base no que se quer otimizar. Por exemplo, se o processo for de maximização, a função deve apresentar um valor mais elevado, diferentemente do processo de minimização (MEDEIROS; KRIPKA, 2012). A função de avaliação dos indivíduos é a componente mais importante em qualquer algoritmo genético, pois é por ela que se mede quão próximo um indivíduo está da solução desejada ou quão boa é a solução. Deve ser escolhida com cuidado, pois deve refletir os objetivos a serem alcançados na resolução do problema e deriva das condições impostas pelo problema (LINDEN, 2012).

3.6.2.7. Métodos de seleção

Para Guedes (2008), o processo de seleção é a ligação dos indivíduos e seu desempenho de sobrevivência, ou seja, os indivíduos mais aptos ao ambiente devem ser selecionados, para assim poderem passar suas características para as próximas gerações. Os métodos de seleção têm a função de selecionar os indivíduos que participarão do

processo de reprodução, ou seja, se obtêm indivíduos (pais) capazes de gerar novos indivíduos (filhos) (RODRIGUES *et al.*, 2004).

A escolha deve ser feita de tal forma que os membros da população mais adaptados ao meio ambiente tenham maior chance de reprodução, isto é, aqueles que apresentam um valor da função *fitness* mais elevado se o processo de otimização for de maximização.

Segundo Linden (2012), os métodos de seleção são usados no sentido de direcionar o processo de seleção para melhores regiões no espaço de busca. Existem diversas formas de seleção de indivíduos sobre os quais serão aplicados os operadores genéticos. A seguir, apresentam-se algumas:

- **Roleta**

Para Linden (2012), o método Roleta foi proposto inicialmente por Holland. Consiste em cada indivíduo ter seu valor de aptidão representado proporcionalmente em relação ao valor total das aptidões.

Os indivíduos são dispostos em setores circulares em que o ângulo do setor circular que o indivíduo ocupará no círculo é proporcional à sua probabilidade de seleção (Figura 11). A probabilidade de seleção é calculada pelo seu valor de aptidão dividido pela soma da adaptação de todos os indivíduos.

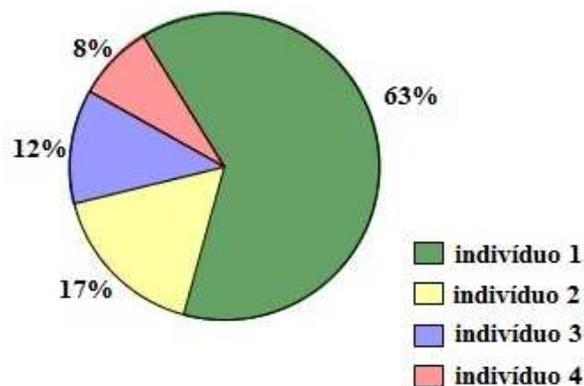


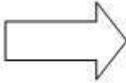
Figura 11 Exemplo de representação dos indivíduos do método da seleção da roleta, em que a porcentagem (%) representa a probabilidade de seleção do indivíduo.

O sorteio dos elementos é feito por um "jogo de roleta". Os indivíduos são selecionados por sorteio, quando a roleta é girada. Segundo Filitto (2008) e Zini (2009), os indivíduos que possuem elevada aptidão ocuparão uma porção maior do que os indivíduos com baixa aptidão. Assim, fica claro que os indivíduos que apresentarem maior representação terão maiores chances para serem selecionados. Um novo indivíduo é selecionado a cada vez que a roleta é girada. Para Lopes (2013), uma desvantagem do método é que a roleta pode parar várias vezes em um mesmo indivíduo com elevado valor de aptidão, o que resultaria em uma população com pouca diversidade.

- Seleção por torneio

Para Zini (2009), a ideia básica do método de seleção por torneio é a promover um torneio entre um grupo de n indivíduos aleatoriamente tomados na população ($n \geq 2$). O indivíduo que vencer o torneio (aquele que tiver o melhor valor do *fitness*) fará parte na nova população. Um exemplo com sete torneios está apresentado na Figura 12, e existem três competidores em cada torneio. O número de competidores não pode ser menor do que dois, caso contrário não há competição.

Indivíduo	Fitness
X1	100
X2	200
X3	800
X4	50
X5	50
X6	1000
X7	30



X1 100	X4 50	X7 30	Torneio 1
X2 200	X3 800	X5 50	Torneio 2
X6 1000	X2 200	X7 30	Torneio 3
X4 50	X6 1000	X5 50	Torneio 4
X7 30	X4 50	X1 100	Torneio 5
X3 800	X5 50	X7 30	Torneio 6
X1 100	X2 200	X5 50	Torneio 7

(a) (b)

Figura 12 Exemplo de aplicação do método da seleção por torneio, em que os indivíduos e seus respectivos valores do *fitness* são dados em (a) e, em (b) apresentam-se os sete torneios realizados. Em cinza, destaca-se o indivíduo que venceu cada torneio.

Uma vantagem desse método é que os indivíduos são selecionados de forma aleatória para participar de cada torneio, não favorecendo assim os indivíduos com maior aptidão, como no caso da roleta. Portanto, quando os melhores indivíduos forem selecionados, eles vencem o torneio. Ainda há a impossibilidade de uns indivíduos dominarem a população, o que resultaria em uma rápida convergência para um ótimo local (BUENO, 2009). Neste método, também não há necessidade de um cálculo da probabilidade de seleção. Assim, este procedimento torna-se mais simples e rápido.

3.6.2.8. Operadores genéticos

Os operadores genéticos são transformações realizadas nos indivíduos que restaram no processo de seleção. Isto ocorre para que haja uma recombinação e produção de novos indivíduos, ou seja, gera-se um novo indivíduo com as informações genéticas de outros dois indivíduos selecionados de forma aleatória (MEDEIRO; KRIPTA, 2012).

Eles são os responsáveis para que se garanta a evolução dos indivíduos a partir de uma população inicial; e assim novas populações são criadas. Espera-se que cada nova população obtenha melhoria no desempenho dos indivíduos (RODRIGUES *et al.*, 2004).

Os operadores que merecem destaque são os de recombinação (cruzamento) e os de mutação. Logo, uma nova população terá sido criada ao final desta etapa.

3.6.2.8.1. Cruzamento

O cruzamento (ou crossover) é o principal mecanismo na reprodução dos novos indivíduos, pois proporciona a troca de informações genéticas entre os indivíduos, e ainda tem o importante papel de garantir a diversidade e a evolução dos indivíduos (GUEDES, 2008; BUENO, 2009).

A troca de informações é responsável pela combinação das características genéticas dos pais durante a reprodução e permite que sejam herdadas pelos próximos descendentes.

A reprodução mais simples realizada nos indivíduos ocorre com a seleção de somente dois indivíduos (FRANCO; SILVA, 2014). Para a realização do cruzamento, dois indivíduos (pais) são selecionados e o material genético é combinado para gerar um novo indivíduo, o qual pode ser mais bem adaptado ao ambiente (ter melhor valor da função objetivo) do que os genitores (LINDEN, 2012).

Existem diversos tipos de operadores de cruzamento, porém, os mais tradicionais são: cruzamento com um ponto de corte, cruzamento com dois pontos de corte; cruzamento com n pontos de corte e cruzamento uniforme, descritos a seguir.

- Cruzamento com um Ponto de Corte

Segundo Linden (2012), o cruzamento de um ponto é considerado o mais simples, e consiste em escolher aleatoriamente um ponto de corte em dois indivíduos selecionados, onde haverá troca de material genético. Em outras palavras, dividem-se os pais em duas partes e recombina-se as características dos indivíduos. Assim, são gerados dois novos indivíduos (filhos), como ilustrado na Figura 13.

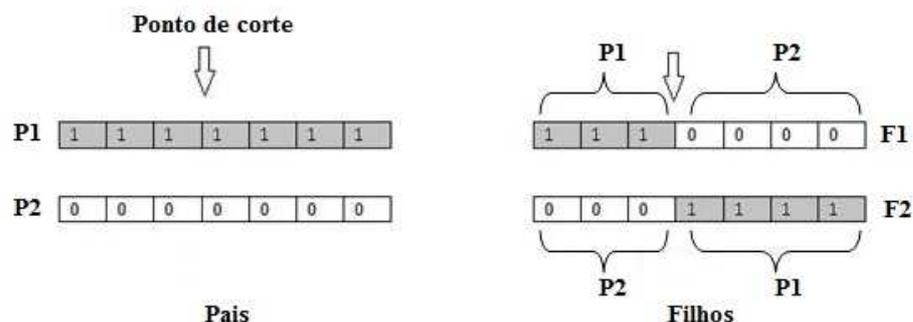


Figura 13 Exemplo de um cruzamento com um ponto de corte, em que P1 e P2 são os pais e F1 e F2 são os filhos gerados.

- Cruzamento com dois Pontos de Corte

O cruzamento de dois pontos é semelhante ao cruzamento com um ponto. Nele, são escolhidos dois pontos de corte aleatoriamente e sem reposição, assim são gerados três segmentos em cada indivíduo pai (LINDEN, 2012). Na sequência, ocorre a troca do segmento do meio de cada indivíduo pai e os dois novos indivíduos (filhos) são gerados, como ilustrado na Figura (14).

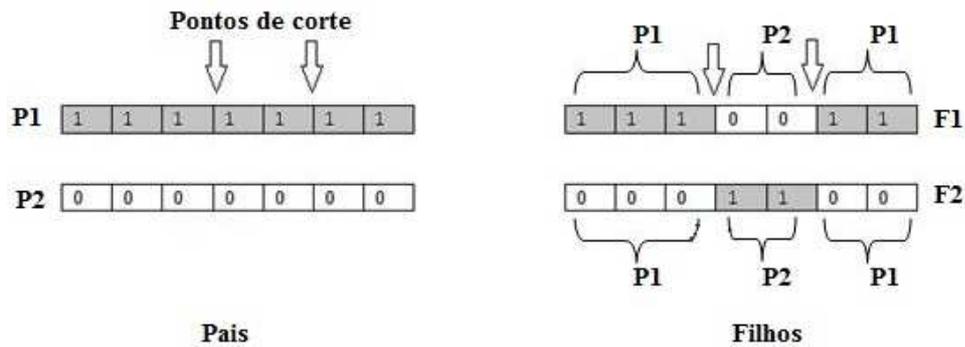


Figura 14 Exemplo de um cruzamento com dois pontos de corte, em que P1 e P2 são os pais e F1 e F2 são os filhos gerados.

- Cruzamento com n pontos de corte

Para Lucas (2002), o cruzamento com n pontos de corte segue a mesma ideia do cruzamento com um ponto de corte. Difere apenas na escolha aleatória e sem reposição dos n pontos de corte. A Figura (15) ilustra um exemplo deste cruzamento em que são escolhidos dois indivíduos com quatro pontos de corte.

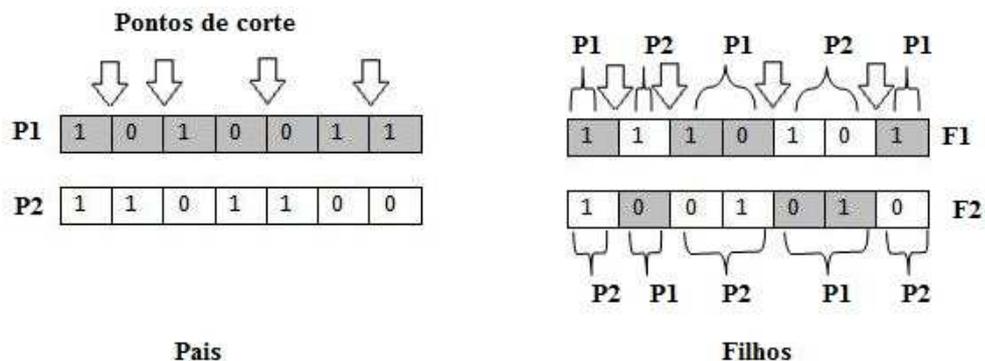


Figura 15 Exemplo de um cruzamento com quatro pontos de corte, em que P1 e P2 são os pais e F1 e F2 são os filhos gerados.

- Cruzamento Uniforme

O cruzamento uniforme é capaz de escolher qualquer posição nos indivíduos para a recombinação genética. Utiliza uma palavra binária-padrão escolhida aleatoriamente e um dos valores 0 ou 1 para designar a posição em que serão recombinados os indivíduos (LUCAS, 2002). No cruzamento uniforme, é trocado apenas o *bit* (valor) daquela posição e não um segmento de *bits*. A Figura 16 ilustra o funcionamento do cruzamento uniforme, enquanto a segunda, quinta e sétima posições foram sorteadas para a troca do *bit* entre os pais, para a geração dos filhos.

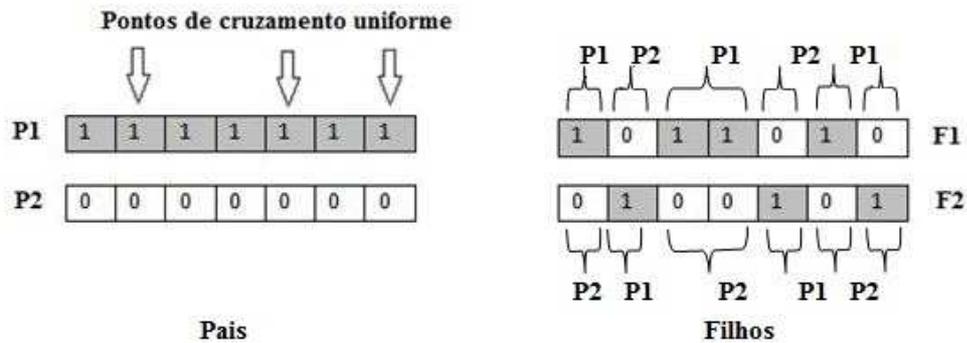


Figura 16 Exemplo de cruzamento uniforme, em que P1 e P2 são os pais e F1 e F2 são os filhos gerados.

3.6.2.8.2. Mutação

Os operadores de mutação são aplicados depois do cruzamento. Os operadores de mutação têm como funções introduzir novas características aos indivíduos, diversificar a população de soluções e evitar que ocorra homogeneidade da população ao longo das gerações. A mutação pode ser benéfica, deixando o indivíduo mais apto para permanecer na população, ou maléfica, deixando o indivíduo menos apto e ele pode não sobreviver à seleção natural (BUENO, 2009). Este operador auxilia o algoritmo a fugir de máximos ou mínimos locais.

Entre os operadores de mutação para codificação binária existe a mutação binária simples, na qual as posições são sorteadas e os genes correspondentes são invertidos, ou seja, se o valor do gene é 1, muda-se para zero e vice-versa, gerando-se assim um novo indivíduo que fará parte da população. A Figura (17) ilustra o funcionamento do operador de mutação.

Estes operadores são aplicados segundo uma probabilidade chamada taxa de mutação (MOORI; KIMURA; ASAKURA, 2010).

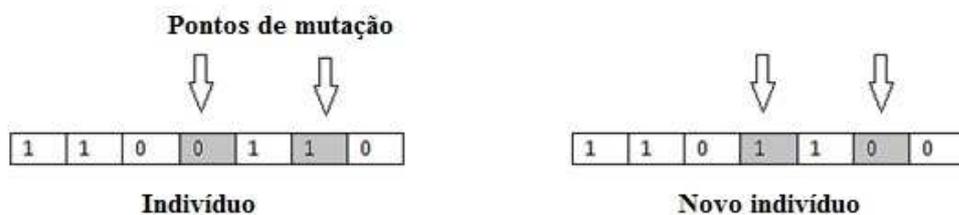


Figura 17 Exemplo de uma mutação binária simples.

3.6.2.9. Critérios de parada

Segundo Youssef *et al.* (2001), um problema enfrentado na confecção de um AG é determinar o critério de parada. O fator de complicação se deve à dificuldade desses algoritmos em avaliar a qualidade da solução em um instante da busca, pois, em geral, não há informação de quão a solução se aproxima da solução ótima. Geralmente, os critérios utilizados são:

- Número máximo de iterações (gerações);
- Solução ótima encontrada (quando esta é conhecida);
- Perda de diversidade, ou seja, a cada iteração não há mais alteração nas características dos indivíduos;
- Nas últimas k gerações não houve melhoria na função de aptidão.

3.6.2.10. Parâmetros dos algoritmos genéticos

Os principais parâmetros de configuração dos algoritmos genéticos são: o tamanho da população, o número de gerações, a taxa de cruzamento e a taxa de mutação. Zini (2009) defende que a utilização correta dos parâmetros é, sem dúvidas, um dos aspectos mais relevantes dentro da estrutura de um algoritmo genético.

Para o autor, o tamanho da população pode afetar o desempenho e a eficiência do AG, por isso a população não pode ser nem muito grande nem muito pequena. Experimentos realizados por De Jong em 1975 indicam que o melhor tamanho de uma população está entre 50 e 100 indivíduos, considerada a faixa mais comum em aplicações de AG (LINDEN, 2012).

Diversos autores divergem quanto à escolha do número de gerações. Goldberg (1989) considera como critério o desvio padrão dos valores de aptidão de cada indivíduo, em substituição a este parâmetro. Assim, tem-se uma comparação do desempenho da atual geração (dos indivíduos) com a geração de indivíduos anterior. Se o desvio padrão for menor ou igual ao estabelecido como aproximação aceitável, o processo então termina. Neste caso, o número de gerações não é fornecido no início do processo, e é conhecido somente quando o critério de parada for aceito.

A taxa de cruzamento indica a probabilidade em que ocorrerá o cruzamento entre os indivíduos a cada iteração. Quanto maior a probabilidade for, mais rapidamente novas estruturas serão colocadas na população. Mas, também existe o problema de se retirar indivíduos com boa qualidade, quando a probabilidade é muito elevada. Já com um valor muito baixo, a evolução pode se tornar muito lenta. Na literatura, de acordo com Rezende (2003), normalmente utilizam-se valores entre 0,6 e 0,99. Porém, Tanomaru (1995) sugere uma taxa de cruzamento superior a 0,7.

A taxa de mutação indica qual a probabilidade em que ocorrerá o processo de mutação nos indivíduos a cada iteração. A mutação tem o trabalho de fornecer novas informações à população e aumentar a diversidade. A taxa de mutação possibilita maior varredura do espaço de busca. Deve-se tomar cuidado em se utilizar taxa de mutação muito elevada, pois pode tornar a busca essencialmente aleatória. Na literatura, de acordo com Rezende (2003), encontram-se valores entre 0,001 e 0,1 para esta taxa. Enquanto Tanomaru (1995) sugere uma taxa de mutação menor que 0,01.

4. REFERÊNCIAS BIBLIOGRÁFICAS

- ANDERSON, J. F.; HARDY, E. E.; ROACH J. T.; WITMER R. E. **A land use and land cover classification system for use with remote sensor data**. U.S. Washington, DC: U.S. Geologic Survey, 2001. 41 p.
- BARBIERI, D. M.; MARQUES JÚNIOR, J.; PEREIRA, G. T.; LA SCALA JÚNIOR, N.; SIQUEIRA, D. S.; PANOSSO, A. R. Comportamento dos óxidos de ferro da fração argila e do fósforo adsorvido, em diferentes sistemas de colheita de cana-de-açúcar. **Revista Brasileira de Ciência do Solo**, v. 37, n.6, p.1557-1568, Nov./dec. 2013.
- BARBOSA, R. C. **Aplicação da meta-heurística Algoritmo Genético na Otimização das rotas de entregas da distribuição física de produtos no município de Fortaleza**. 2014. 91 p. Dissertação de Mestrado, Programa de Mestrado em Logística e Pesquisa Operacional da Universidade Federal do Ceará, Fortaleza-CE.
- BÁRDOSSY, A. **Introduction to Geostatistics**. Institute of Hydraulic Engineering University of Stuttgart, 2008. 135 p.
- BERNARDI, A. C. C.; NAIDE, J. M.; RESENDE, A. V.; BOSSOI, L. H.; INAMASU, R. Y. **Agricultura de precisão: Resultados de um Novo olhar**. Embrapa, Brasília- DF, 2014.600p.
- BORSSOI, J. A.; URIBE-OPAZO, A. M; GALEA, M. Diagnostics techniques applied in geostatistics for agricultural data analysis. **Revista Brasileira de Ciência do Solo**, v.6, n. 6, p.1-16, 2009.
- BOTTEGA, E. L.; QUEIROZ, D. M.; PINTO, F. A. C.; SOUZA, C. M. A. Variabilidade espacial de atributos do solo em sistema de semeadura direta com rotação de culturas no cerrado brasileiro. **Revista Ciência Agrônômica**, Fortaleza – CE, v.44, n.1, p.1-9, jan./mar. 2013.
- BRAGA, L. P. V. **Geoestatística e aplicações**. 9º Simpósio Brasileiro de Probabilidade e Estatística, IME – USP, 1990. 36 p.
- BRAGA, R.; PINTO, P. A. Agricultura de precisão: adoção & principais obstáculos. **AGROTEC - Revista técnico-científica agrícola**, Coimbra – PT, n.1, p.84-89, dez. 2011.
- BUENO, F. **Métodos Heurísticos Teoria e Implementação**. Tutorial, Instituto Federal de Santa Catarina – IFSC, Araranguá- SC, 2009. 37 p.
- CÂMARA, G.; MEDEIROS, J. S. GIS para meio ambiente. In: Sistema de informações geográficas, 1998, Curitiba. **Anais...** Curitiba: Sagres, 1998. p. 140-180.
- CAMBARDELLA, C.A.; MOORMAN, T. B; PARKIN, T.B.; NOVACK, J. M; KARLEN, D. L; TURCO, R. F; KNOPKA, A.E. Field-scale variability of soil properties in Central Iowa Soils. **Soil Science Society America Journal**, Medison, v.58, n.5, p. 1501-1511, 1994.
- CÁRDENAS, R. R.; FERREIRA, M. A. R.; SCHMIDT, A. M. Stochastic search algorithms for optimal design of monitoring networks. *Environmetrics*.v. 21, p. 102-112, 2010. Published online on March, 8th, 2009 in **Wiley InterScience** - (www.interscience.wiley.com) DOI: 10.1002/env.989.
- CHIPETA, M. G.; TERLOUW, D. J.; PHIRI, K. S.; DIGGLE, P. J. **Adaptive geostatistical design and analysis for prevalence surveys**. v.15, p.70-84, 2016.
- CLARK, I. **Practical geostatistics**. London: Applied science publishers Ltda, 1979, 123 p.

CONGALTON, R. G.; GREEN, K. **Assessing the accuracy of remotely sensed data: principles and practices**. New York: Lewis Publisher, 1999, 130p.

CRESSIE, N. A. C. **Statistics for Spatial Data** (Rev. Ed.). John Wiley & Sons, p.928, 2015.

DALLACORT, R.; MORAES, V. R.; ALVES, E. D. L.; COMUNELLO, E.; FIETZ, C. R. Modelos de semivariogramas para análise da variabilidade espacial da precipitação pluvial no Pantanal de Mato Grosso. **Anais...** 4^o Simpósio de Geotecnologias no Pantanal, Bonito – MS, p.467-477, out., 2012.

DALPOSSO, G. H.; URIBE-OPAZO, M. A.; MERCANTE, E.; JOHANN, J. A.; BORSSOI, J. A. Comparison measures of maps generated by geostatistical methods. **Engenharia Agrícola**, Jaboticabal, v.32, n.1, p.174-183, jan./fev. 2012.

DE BASTIANI, F. **Influência local em modelos espaciais lineares com distribuição da família de contornos elípticos**. 137 f. Dissertação (Mestrado) – Programa de Pós-Graduação em Estatística, Universidade Federal de Pernambuco, 2012.

DE BASTIANI, F.; URIBE-OPAZO, M. A.; DALPOSSO, G. H. Comparison of maps of spatial variability of soil resistance to penetration constructed with and without covariables using a spatial linear model. **Engenharia Agrícola**, Jaboticabal, v.32, n.2, p.393-404, mar./abr. 2012.

DE BASTIANI, F.; CYSNEIROS, A. H. M. A.; URIBE-OPAZO, M. A.; GALEA, M. Influence diagnostics in elliptical spatial linear models. **Sociedad de Estadística e Investigación Operativa**, TEST, Berlin, v.24, n.2, p. 322-340, 2015.

DIAS, V. R. M.; ALVES, M. C.; SANCHES, L. Métodos de ajuste de semivariogramas para modelagem espacial de íons de precipitação pluvial em Cuiabá, Brasil. **Ciência e Natura – Revista do Centro de Ciências Naturais e Exatas** – UFSM, Santa Maria – RS, v.37, n.2, p. 312-320, mai./ago. 2015.

DIGGLE, P.J.; RIBEIRO JR., P.J. **Model-based geostatistics**. New York: Springer, 2007. 230 p.

FARACO, M. A.; URIBE-OPAZO, M. A. SILVA, E. A.; JOHANN, J. A.; BORSSOI, J. A. Seleção de modelos de variabilidade espacial para elaboração de mapas temáticos de atributos físicos do solo e produtividade da soja. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 32, n. 2, p. 463-476, 2008.

FILITTO, D. Algoritmos Genéticos: uma visão explanatória. **Revista Multidisciplinar da UNIESP- Saber acadêmico** – n.6, p. 136-143, 2008/ ISSN 1980-5950.

FRANCO, R. A. P.; SILVA, K. R. G. Templates de algoritmos genéticos para a geração de estímulos aplicados à verificação funcional de dispositivos. **XII Conferência de Estudos em Engenharia Elétrica**- CEEL-Universidade Federal de Uberlândia – UFA, Uberlândia, MG, out. 2014. ISSN 2178-8308.

GOLDBARG, M. C.; LUNA, H. P. L. Otimização Combinatória e Programação Linear. **Elsevier**, Rio de Janeiro, 2^aed., 2005. 536 p.

GOLDBERG, D. E. **Genetic Algorithm in search, Optimization and Machine Learning**. Addison - Wesley, Reading – USA, 1989.

GALLARDO, A. Geoestatística. **Ecosistemas Revista científica y técnica de ecología y medio ambiente**. v.15, n.3, p. 48-58, 2006.

GUEDES, L. P. C. **Otimização de amostragem espacial**. 2008. 143f. Tese de doutorado, Programa de Pós-Graduação em Estatística e Experimentação Agrônômica, Universidade de São Paulo, Escola Superior de Agricultura “Luiz de Queiroz”, São Paulo, 2008.

GUEDES, L. P. C.; URIBE-OPAZO, M. A.; JOHANN, J. A.; SOUZA, E. G. Anisotropia no estudo da variabilidade espacial de algumas variáveis químicas do solo. **Revista Brasileira de Ciência do Solo**, Viçosa, v.32, n.6, p. 2217-2226, Nov./dec. 2008.

GUEDES, L.P.C.; RIBEIRO JR. P.J.; PIEDADE, S.M.D.S.; URIBE-OPAZO, M.A. Optimization of spatial sample configurations using hybrid genetic algorithm and simulated annealing. **Chilean Journal of Statistics**, Chile, v. 2, n.2, p. 39-50, set., 2011.

GUEDES, L. P. C.; URIBE-OPAZO, M. A., RIBEIRO JR. P. J. Optimization of sample design size and shapes for regionalized variables using simulated annealing. **Ciência e Investigación Agraria**, v. 41, n.1, p. 33-48, jan./abr. 2014.

GUEDES, L. P. C.; RIBEIRO JR. P. J.; URIBE-OPAZO, M. A.; BASTIANI, F. Soybean yield maps using regular and optimized sample with different configurations by simulated annealing. **Engenharia Agrícola**, Jaboticabal, v.36, n.1, p.114-125, jan./fev. 2016. ISSN: 1809-4430.

GUERRA, P. A. G. **Geoestatística operacional**. Brasília: Ministério das Minas e Energia, Departamento Nacional da Produção Mineral, 1988. 145p.

HAUSCHILD, F. E. G.; FIORIN, J. E.; AMADO, T. J. C.; SANTI, A. L. Técnicas de agricultura de precisão para definição de zonas de manejo. **Congresso Brasileiro de Agricultura de Precisão – ConBAP**, São Pedro – SP, set. 2014.

HEINEN, M. R.; OSÓRIO, F. S. Algoritmos Genéticos Aplicados ao Problema de Roteamento de Veículos. **Hífen**, Uruguiana, v.30, n.58, p. 89-95, 2006.

HOLLAND, J. H. **Adaptation in natural and artificial systems**. The University of Michigan Press, Ann Arbor, MI, 1975.

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

JOURNEL, A. G.; HUIJBREGTS, C. J. **Mining geostatistics**. London: Academic Press, 1978. 600 p.

JUNTOLLI, F. V.; SANTOS, R. L. B.; INAMASU, R. Y.; MOLIN, J. P. Agricultura de precisão. **Boletim técnico**. Ministério da agricultura, pecuária e abastecimento. Secretaria de desenvolvimento agropecuário e cooperativismo. Brasília – DF, 2013. 40 p.

KONOWALENKO, F.; BENEVIDES, P. F.; COSTA, D. M. B.; BARBOZA, A. O.; NUNES, L. F. Aplicação de um algoritmo genético para o problema do carteiro chinês em uma situação real de cobertura de arcos. **Revista Ingeniería Industrial**, n.1, p.27-36, mai. 2012.

KRIPPENDORFF, K. **Content Analysis an Introduction to its Methodology**. Sage Publications Ltda, California, 2ª ed., 2013. 422 p.

KRIPTA, R. M. L.; KRIPTA, M. Simulated Annealing aplicado na otimização da alocação de salas em instituição de ensino superior. **Asociación Argentina de Mecánica Computacional**. Buenos Aires, Argentina, v.29, p. 9317-9325, nov. 2010.

- LANDIM, P. M. B. Sobre Geoestatística e mapas. **Terra e Didática**, v. 2, n. 1, p. 19-33, 2006.
- LANDIM, P. M. B.; YAMAMOTO, J. K. **Geoestatística: conceitos e aplicações**. São Paulo-SP: Oficina de Textos, 2013. 215 p.
- LANGE, K. L.; LITTLE, R. J. A.; TAYLOR, J. M. G. Robust statistical modeling using the t distribution. **Journal of the American Statistics Association**, v.84, n.408, p. 881-896, 1989.
- LINDEN, R. **Algoritmos Genéticos**. Ciência Moderna Ltda, Rio de Janeiro, 2012. 496 p.
- LOPES, H. S.; RODRIGUES, L. C. A.; STEINER, M. T. A. **Meta-heurística em pesquisa operacional**. Ominipax, Curitiba, 1ª edição, 2013.
- LUCAS, D. C. **Algoritmos Genéticos: uma Introdução**. Apostila elaborada para a disciplina de Ferramentas de Inteligência Artificial, 2002. Disponível em: <http://www.inf.ufrgs.br/~alvares/INF01048IA/ApostilaAlgoritmosGeneticos.pdf>. Acesso em: 15 jun. 2016.
- MA, Z.; REDMOND, R. L. Tau coefficients for accuracy assessment of classification of remote sensing data. **Photogrammetric Engineering and Remote Sensing**. Bethesda, v. 61, n.4, p. 453-439, 1995.
- MATÉRN, B., 1986. **Spatial Variation. Lecture notes in Statistics**, 2ªed. New York: Springer, 151p.
- MATHERON, G. Principles of Geostatistics. **Economic Geology**, v. 58, p. 1246-1266, 1963.
- 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.
- MCBRATNEY A. B., WEBSTER R. Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. **J. soilsci.** v. 37, p. 617-639, 1986.
- MEDEIROS, G. F., KRIPKA, M. Algumas Aplicações de Métodos Heurísticos na Otimização de Estruturas. **Revista de Ciências Exatas Aplicadas e Tecnológicas da Universidade de Passo Fundo – CIATEC UPF**, Passo Fundo, v. 4, n. 1, p. 19-32, 2012.
- MELLO, Y. R.; OLIVEIRA, T. M. N. Análise Estatística e Geoestatística da Precipitação Média para o município de Joinville (SC). **Revista Brasileira de Meteorologia – RBMET**. Rio de Janeiro – RJ, v.31, n.2, p. 229-239, 2016.
- MITCHELL, A. F. S. The information matrix, skewness tensor and α -connections for the general multivariate elliptic distribution. **Annals of the Institute of Statistical Mathematics**. Tokyo - JP, v.41, n.2, p. 289-304, 1989.
- 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.
- MOLIN, J. P.; INAMASU, R. Y.; JUNTOLLI, F. V.; SANTOS, R. L. B. Agricultura de Precisão. **Boletim técnico**, Ministério da Agricultura, Pecuária e Abastecimento Secretaria de Desenvolvimento Agropecuário e Cooperativismo, 3ed. – Brasília - DF, 2013.

MOLIN, J. P.; AMARAL, L. R.; COLAÇO, A. F. **Agricultura de Precisão**. Oficina de Textos, 1ª ed. - São Paulo, 2015.

MOORI, R. G.; KIMURA, H.; ASAKURA, O. K. Aplicação do algoritmo genético na gestão de suprimentos. **Revista de Administração e Inovação – RAI**. São Paulo, v.7, n.2, p.171-192, abr./jun. 2010.

MORGAN, M. **The precision farming**: Guide for agriculturists. 1ª ed. Moline: Deere & Company, 1997. 117p.

MULLER, W. G. **Collecting Spatial Data**. Springer-Verlag Berlin Heidelberg. 3ª ed., 2007. 242 p.

OLIVEIRA, L. W.; OLIVEIRA, A. R.; GOMES, F. V.; DIAS, B. H.; SILVA JR., I. C. Aplicação de algoritmos genéticos para o planejamento de geração distribuída em sistemas de distribuição. **Revista LSP- científica interdisciplinar**. v. 3, n.1, p. 79-97, jan./mar. 2016.

PESSOA, A. L. S.; ULISSES, P. H. C.; BRANCO, H. M. G. C.; RABÊLO, R. A. L. Uma aplicação de algoritmos genéticos simples e compacto para estimação de componentes harmônicas. **Revista Brasileira de Computação Aplicada**, Passo Fundo, v.7, n.2, p. 77-91, mai. 2015.

PHILLIPS, S. Indústria de fertilizantes e agricultura de precisão: apoio à segurança alimentar mundial. **Jornal Informações Agronômicas**, n.146, 2014.

REDUINO, A. C. E. **Aplicações de Algoritmos Genéticos**. Faculdade Salesiana Maria Auxiliadora: Macaé-RJ, 2010. Disponível em http://www.fsma.edu.br/si/edicao3/aplicacoes_de_alg_geneticos.pdf. Acesso: 15 jun. 2016.

REZENDE, S. O. **Sistemas inteligentes**: fundamentos e aplicações. Editora Manole, Barueri, SP, 2003.

ROCHA JR., P. A. S.; TOSTES, M. E. L. Algoritmo Genético aplicado a otimização da furação de placas de circuito impresso. **Anais... XI Simpósio Brasileiro de Automação Inteligente – SBAI**, Fortaleza – CE, out. 2013.

RODRIGUES, F. L.; LEITES, H. G.; SANTOS, H. N.; SOUZA, A. L.; SILVA, G. F. **Meta-heurística Algoritmo Genético para solução de problemas de planejamento florestal com restrições de integridade**. Sociedade de Investigações Florestais- SIF. R. Árvore, Viçosa- MG, v.28, n.2, p.233-245, 2004.

SCHULZ, M. **Incorporação de técnicas de agricultura de precisão em uma propriedade rural**: uma abordagem teórica. 5ª Semana Internacional de Engenharia e Economia FAHOR, Horizontina – RS, OUT. 2015.

SEIDEL, E. J.; OLIVEIRA, M. S. Proposta de uma generalização para os modelos de semivariogramas Exponencial e Gaussiano. **Semina: ciências Exatas e Tecnológicas**. Londrina – PR, v.34, n.1, p.125-132, jan./jul. 2013.

SOARES, A. **Geoestatística para as ciências da terra e do ambiente**. 3ª ed. Lisboa: Press, 2014. 214p.

SOARES FILHO, R.; CUNHA, J. P. A. R. Agricultura de Precisão: particularidades de sua adoção no Sudoeste de Goiás – Brasil. **Engenharia Agrícola**, Jaboticabal, v.35, n.4, p. 689-698, jul./ago. 2015. ISSN: 1809-4430.

TANOMARO, J. Motivação, Fundamentos e Aplicações de Algoritmos Genéticos. **II Congresso Brasileiro de redes neurais**, III Escola de redes neurais, Curitiba – PR, 1995.

TEIXEIRA, M. B. R.; SCALON, J. D. Comparação entre estimadores de semivariância. **Revista Brasileira Biom.**, São Paulo – SP, v.31, n.2, p.248-269, 2013.

URIBE-OPAZO, M. A.; BORSSOI, J. A.; GALEA, M. Influence diagnostics in Gaussian spatial linear models. **Journal of Applied Statistics**, v. 39, n. 3, p. 615-630, 2012.

VIEIRA, S. R. Tópicos em ciência do solo. **Sociedade Brasileira de Ciência do Solo**, v. 1. Viçosa – MG, 2000. 53p.

WANG, J. F.; JIANG, C. S.; HU, M. G.; CAO, Z. D.; GUO, Y. S.; LI, L. F.; LIU, T. J.; MENG, B. **Design based spatial sampling**: Theory and implementation, Environmental Modeling and Software, v. 40, p. 280–288, 2013.

WEBSTER, R.; OLIVER, M. **Geostatistics for environmental scientists**. Chichester: John Wiley & Sons, 2007. 315 p.

YOUSSEF, H.; SAIT, S. M.; ADICHE, H. Evolutionary algorithms, simulated annealing end tabu search: a comparative study. **Engineering Applications of Artificial Intelligence**. v. 14, p.167-181, 2001.

ZIMMERMAN, D. L. Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. John Wiley & Sons, Ltd. **Environmetrics**, v.17, p.635-652, 2006.

ZINI, E. O. C. **Algoritmo Genético Especializado na Resolução de Problemas com Variáveis Contínuas e Altamente Restritos**. Dissertação de mestrado, Programa de Pós-Graduação em Engenharia Elétrica, Universidade Estadual Paulista “Júlio de Mesquita Filho” – Campus de Ilha Solteira, São Paulo, 2009.

ZINI, E. O. C.; BONINI NETO, A.; GARBELINI, E. Algoritmo Multi objetivo para otimização de problemas restritos aplicados a indústria. **Congresso Nacional de Matemática Aplicada à Indústria** – CNMAI, Caldas Novas – Goiás, Nov. 2014.

ZHU, Z.; STEIN, M. L. Spatial sampling for design for parameter estimation of the covariance function. **Journal of Statistical Planning and Inference**.v.134, p.583-603, 2005.

5. ARTIGOS

5.1. ARTIGO 1

Algoritmo genético aplicado ao redimensionamento amostral de atributos químicos do solo otimizando a predição espacial

RESUMO

As técnicas geoestatísticas permitem realizar estudos de dependência espacial de variáveis georreferenciadas. Na realização de um experimento que envolva a variabilidade espacial, é essencial definir um plano de amostragem que tenha um tamanho amostral adequado, que permita a redução dos custos operacionais e aumente a qualidade dos resultados da predição espacial. O objetivo desse trabalho foi redimensionar uma configuração amostral, em uma área comercial composta por 102 pontos, reduzir seu tamanho amostral, para realizações de experimentos futuros, considerando-se a variabilidade espacial de atributos químicos do solo. Para isso, primeiramente, avaliou-se a influência da configuração inicial e dos parâmetros do modelo geoestatístico na busca por um tamanho e uma configuração amostral otimizados a partir de dados simulados. Realizaram-se simulações, assumindo-se variáveis estacionárias Gaussiana, isotrópicas, modelo exponencial para a função semivariância e três configurações amostrais iniciais a serem reduzidas: sistemática, aleatória e *lattice plus close pairs*. Tanto para os dados simulados quanto para os atributos químicos do solo, utilizou-se o Algoritmo Genético (AG) para o redimensionamento amostral otimizado e maximização da medida de acurácia Exatidão Global. A partir dos dados simulados, observou-se que o efeito pepita influenciou negativamente a redução do tamanho amostral, enquanto o alcance não influenciou de forma relevante o redimensionamento amostral. Em relação aos atributos químicos do solo, o AG foi eficiente quanto à redução do tamanho amostral, de tal forma que o tamanho amostral da nova amostra reduzida variou entre 30 e 35 pontos que corresponde respectivamente a 29,41% e 34,31% da malha inicial, com similaridade mínima de predição espacial, em relação à configuração original, igual a 85%. Para todos os atributos químicos do solo, houve similaridade quanto à estrutura de dependência espacial e à construção dos mapas temáticos, entre as configurações original e otimizada de pontos amostrais, pois a maioria das medidas de similaridade indica elevada exatidão.

PALAVRAS-CHAVE: Agricultura de precisão, geoestatística, simulação, tamanho da amostra.

Genetic algorithm applied to sample resizing of soil chemical attributes to optimize spatial prediction

ABSTRACT

Geostatistical techniques have allowed the spatial dependence study of geo-referenced variables. It is essential to define a sampling plan that has an adequate sample size during a trial related to spatial variability, which allows reducing operational costs and increasing the quality of spatial prediction results. Thus, this trial aimed at resizing a sample configuration, in a commercial area composed of 102 points, as well as at reducing its sample size for future experiments, considering the spatial variability of soil chemical attributes. Firstly, the influence of initial configuration and geostatistical model parameters were evaluated in order to find out an optimized sample size and configuration from simulated data. Simulations were carried out, assuming stationary Gaussian and isotropic variables, exponential model for semivariance function and three initial sampling configurations to be reduced: systematic, random and lattice plus close pairs. For both simulated data and soil chemical attributes,

Genetic Algorithm (GA) was optimized to maximize the Global Accuracy measurement. It was observed from the simulated data that nugget effect showed a negative effect on sample size reduction, while the range did not significantly influence the sample resize. In relation to soil chemical attributes, GA was efficient in reducing the sample size, in such way that the sample size of the new reduced sample varied from 30 to 35 points, corresponding respectively to 29.41% and 34.31% of the initial mesh, with a minimum similarity of spatial prediction, in relation to the original configuration (85%). There was similarity on spatial dependence structure and thematic maps drawing for all chemical attributes of soil, between the original and optimized sample point configurations, since most similarity measures have suggested high accuracy.

KEYWORDS: Geostatistics, precision agriculture, sample size, simulation.

5.1.1. Introdução

Atualmente, conhecer a variabilidade dos atributos do solo e das culturas, tanto no espaço quanto no tempo, é apontado como o princípio básico para um manejo de forma precisa nas áreas agrícolas (GREGO; VIEIRA, 2005; DE BASTIANI *et al.*, 2017). Assim, o estudo da variabilidade espacial de variáveis georreferenciadas contínuas em uma área de cultivo é muito importante para que se possa realizar o manejo diferenciado de acordo com a necessidade do solo, para melhorar o rendimento da produção (BERNARDI *et al.*, 2014).

Os atributos químicos do solo são fatores que exercem influência direta no nível de produtividade das culturas. Portanto, quando são estudados, em especial, a análise da variabilidade espacial de atributos químicos do solo, a configuração amostral e o tamanho amostral, os fatores devem ser definidos com cuidado, pois o número de amostras do solo utilizadas para a análise da variabilidade espacial, obtida pelas técnicas geoestatísticas, está diretamente ligado aos custos para a realização do experimento. Ou seja, quanto maior for a quantidade de amostras utilizadas, maior será o custo para descrever a variabilidade espacial de atributos químicos do solo a partir de mapas temáticos (DIGGLE; RIBEIRO JÚNIOR, 2007).

Logo, em um experimento que envolva a análise da variabilidade espacial, é essencial se definir um plano de amostragem que: tenha um tamanho amostral adequado, permita a minimização dos custos operacionais e maximize a qualidade dos resultados da predição espacial. Desta forma, a determinação da configuração amostral e do tamanho amostral reduzido deve considerar quão bons são os resultados gerados no estudo da variabilidade espacial (GUEDES *et al.*, 2011).

A escolha de uma configuração e de um tamanho amostral eficiente para a predição espacial pode ser definida como um problema de otimização que, a partir de uma malha amostral inicial, se escolhe uma configuração amostral que minimize as perdas quanto à acurácia dos resultados da predição espacial (GUEDES *et al.*, 2016).

Um desses processos de otimização é chamado de Algoritmo Genético (AG), definido como uma técnica computacional inspirada na evolução natural das espécies

(LINDEN, 2012; PESSOA *et al.*, 2015; GOLDBARG; GOLDBARG; LUNA, 2016). Os processos de redimensionamento de configurações amostrais consideram como critério de busca uma função-objetivo do problema, que será minimizada ou maximizada e resume a eficiência da otimização.

Trabalhos sobre resultados e soluções para as configurações amostrais otimizadas em variáveis que apresentam dependência espacial são recentes. A ideia de otimização é encontrada, por exemplo, em problemas de racionalização de estações de rede de monitoramento ambiental (CÁRDENAS *et al.*, 2010); em problemas para determinar uma configuração amostral, com o tamanho amostral fixado previamente, para a classificação do fertilizante potássio (K) na área em estudo (GUEDES *et al.*, 2011); em problemas para otimizar o tamanho e as configurações amostrais para análise da produtividade de soja, cuja produtividade foi obtida pelo monitor de colheita (GUEDES *et al.*, 2014); para determinar a localização ótima para a instalação de fontes de geração distribuída (GD) em sistemas elétricos de distribuição, que envolvem energia térmica baseada em biomassa e eólica visando minimizar perdas de energia, emissões e os custos de investimento e operação, além da melhoria da regulação de tensão e qualidade de energia (OLIVEIRA *et al.*, 2016).

Tais processos de redimensionamento de configurações amostrais consideram como critério de busca uma função-objetivo do problema, a qual pode ser minimizada ou maximizada, de acordo com o problema. Existem funções-objetivo que levam em consideração a eficiência quanto à predição, tais como a variância média da predição espacial (GUEDES *et al.*, 2011), a soma do erro quadrático da predição espacial (GUEDES *et al.*, 2014) e a medida de acurácia Exatidão Global (GUEDES *et al.*, 2014; GUEDES *et al.*, 2016). Entretanto, durante o processo de otimização, ainda se pode levar em consideração a eficiência quanto à estimação do modelo geoestatístico, como é o caso da função-objetivo logaritmo do determinante da matriz inversa de Fisher (ZHU; STEIN, 2005).

O presente trabalho traz contribuições em relação ao redimensionamento de uma configuração amostral, além de otimizar simultaneamente o tamanho e a configuração amostral com o uso de algoritmo genético. Desta forma, o objetivo do trabalho foi determinar a melhor configuração com melhor tamanho amostral reduzido, para o estudo da variabilidade espacial de atributos químicos do solo, provenientes de uma área comercial de cultivo de soja, onde se desenvolve uma atividade de aplicação localizada de insumos, usando-se algoritmo genético e as informações destes atributos. Para isto, avaliaram-se também, a partir de dados simulados, a influência da configuração amostral inicial e a influência dos parâmetros do modelo geoestatístico para a obtenção de um tamanho e uma configuração amostral otimizada.

5.1.2. Material e Métodos

5.1.2.1. Estudo de simulações

Os conjuntos de dados simulados foram obtidos a partir de simulações de Monte Carlo, de forma que estes conjuntos representam realizações de processos estocásticos $\{Z(\mathbf{s}_i), \mathbf{s}_i \in S\}$, com $\mathbf{s}_i = (x, y)^T$ o vetor que representa determinada localização na área de estudo, com $S \subset \mathcal{R}^2$, e \mathcal{R}^2 é o espaço euclidiano bidimensional (MARDIA; MARSHALL, 1984). Suponha que os dados $Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)$ deste processo são observações da variável aleatória sob estudo nas \mathbf{s}_i ($i = 1, \dots, n$) localizações espaciais conhecidas e geradas pelo modelo linear espacial gaussiano, dado Pela Equação 43 (URIBE-OPAZO; BORSSOI; GALEA, 2012):

$$Z(\mathbf{s}_i) = \mu(\mathbf{s}_i) + \epsilon(\mathbf{s}_i) \quad (43)$$

em que $\mu(\mathbf{s}_i)$ é o termo determinístico e $\epsilon(\mathbf{s}_i)$ é o termo estocástico e ambos dependem da localização espacial na qual $Z(\mathbf{s}_i)$ é observada. Assume-se que o erro estacionário $\epsilon(\mathbf{s}_i)$ tem média zero, $E[\epsilon(\mathbf{s}_i)] = 0$, e a variação entre pontos no espaço é determinada por uma função covariância $C(\mathbf{s}_i, \mathbf{s}_j) = cov[\epsilon(\mathbf{s}_i), \epsilon(\mathbf{s}_j)]$, e $\mu(\mathbf{s}_i) = \mu \mathbf{1}$, em que μ é uma constante; $\mathbf{1}$ é um vetor de uns de dimensão $n \times 1$, quando o processo considerado for estacionário. Caso contrário, representa uma tendência direcional da variável georreferenciada (GRESSIE, 2015). Neste caso, a média do processo é dada para algumas funções conhecidas de \mathbf{s}_i como $x_1(\mathbf{s}_i), \dots, x_p(\mathbf{s}_i)$, por,

$$\mu(\mathbf{s}_i) = \sum_{j=1}^p x_j(\mathbf{s}_i) \beta_j, \quad i = 1, \dots, n$$

em que β_1, \dots, β_p são parâmetros desconhecidos e precisam ser estimados.

Da mesma forma, na notação matricial, temos o modelo espacial linear dado pela Equação 44:

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (44)$$

em que, o vetor de erros aleatórios $\boldsymbol{\epsilon}$ tem $E(\boldsymbol{\epsilon}) = \mathbf{0}$ (vetor nulo) e a matriz de covariância $\boldsymbol{\Sigma}[(\sigma_{ij})]$, com $\sigma_{ij} = C(\mathbf{s}_i, \mathbf{s}_j)$. Assume-se que $\boldsymbol{\Sigma}$ é não singular e \mathbf{X} é de posto completo; $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$; e \mathbf{Z} segue uma distribuição gaussiana n -variada com vetor de médias $\mu \mathbf{1}$ e matriz de covariância $\boldsymbol{\Sigma}$, isto é, $\mathbf{Z} \sim N_n(\mu \mathbf{1}, \boldsymbol{\Sigma})$ (DE BASTIANI *et al.*, 2015).

Uma forma paramétrica que especifica para a matriz de covariância (MARDIA; MASHALL, 1984; URIBE-OPAZO; BORSSOI; GALEA, 2012) é dada por:

$$\boldsymbol{\Sigma} = \varphi_1 \mathbf{I}_n + \varphi_2 \mathbf{R}(\varphi_3),$$

em que φ_1 é o efeito pepita, sendo ideal que $\gamma(0) = 0$ (γ é a semivariância). Porém, na prática, conforme h tende a zero, $\gamma(h)$ se aproxima de um valor positivo, conhecido como efeito pepita, e este revela a descontinuidade do semivariograma para distâncias menores

do que a menor distância entre amostras; I_n é a matriz identidade $n \times n$; φ_2 é a contribuição, conhecida como a diferença entre o patamar e o efeito pepita (CÂMARA; MEDEIROS, 1998). Assim, o segmento gráfico neste intervalo, de fato, caracteriza a dependência espacial dos processos estocásticos contínuos, a qual pode ser entendida como uma variância estruturada correspondente à estrutura do fenômeno em estudo (SEIDEL; OLIVEIRA, 2013); φ_3 é a função do alcance do modelo, e o alcance ($a = g(\varphi_3)$) é a distância limite de dependência espacial, ou seja, medições localizadas à distância maior que a são consideradas independentes entre si, enquanto amostras separadas por distâncias menores do que a são correlacionadas umas com as outras (VIEIRA, 2000; SEIDEL; OLIVEIRA, 2013); e $\mathbf{R}(\varphi_3)$ é uma matriz $n \times n$ que é função de φ_3 .

$\mathbf{R} = \mathbf{R}(\varphi_3) = [r_{ij}]$, em que r_{ij} depende de $h_{ij} = \|s_i - s_j\|$ (distância Euclidiana entre s_i e s_j) é uma matriz simétrica $n \times n$, com elementos da diagonal $r_{ij} = 1$, para $i = j$ com $i, j = 1, \dots, n$; $r_{ij} = \varphi_2^{-1}C(S_i, S_j)$ para $\varphi_2 \neq 0$ e $i \neq j$; e $r_{ij} = 0$ para $\varphi_2 = 0$, $i \neq j$. Tal forma paramétrica da matriz de covariância Σ , ocorre para vários processos isotrópicos (GUEDES *et al.*, 2008), onde a covariância $Cov[\epsilon(s_i), \epsilon(s_j)]$ é definida pela função de covariâncias $Cov[\epsilon(s_i), \epsilon(s_j)] = \varphi_2 r_{ij}$, para $i \neq j$ (De BASTIANI *et al.*, 2015).

As seguintes três configurações amostrais iniciais foram simuladas em uma área quadrada com limite máximo para as coordenadas x e y igual a 1, com 100 pontos amostrais cada: sistemática 10×10 (Figura 18a), aleatória (Figura 18b) e sistemática 9×9 com adição de 19 pontos próximos, posicionados dentro de uma circunferência de raio 0,10, cujo centro consiste em um ponto da amostragem regular, escolhido de forma aleatória (Figura 18c) (*lattice plus close pairs*, DIGGLE; RIBEIRO JUNIOR, 2007). Nove ensaios foram considerados para cada configuração amostral. E, em cada ensaio, foram construídas 100 simulações, totalizando-se 900 simulações para cada amostragem. Em cada ensaio, as simulações foram realizadas considerando-se o modelo linear espacial gaussiano (Equação 43) e um modelo exponencial para a função semivariância, com contribuição igual a 1, bem como todas as combinações dos seguintes valores para os parâmetros alcance prático (0,45; 0,60; e 0,90) e efeito pepita (0; 0,5; e 0,8). Os valores foram escolhidos com o objetivo de contemplar diferentes intensidades de dependência espacial (fraca, moderada e forte, CAMBARDELLA *et al.*, 1994).

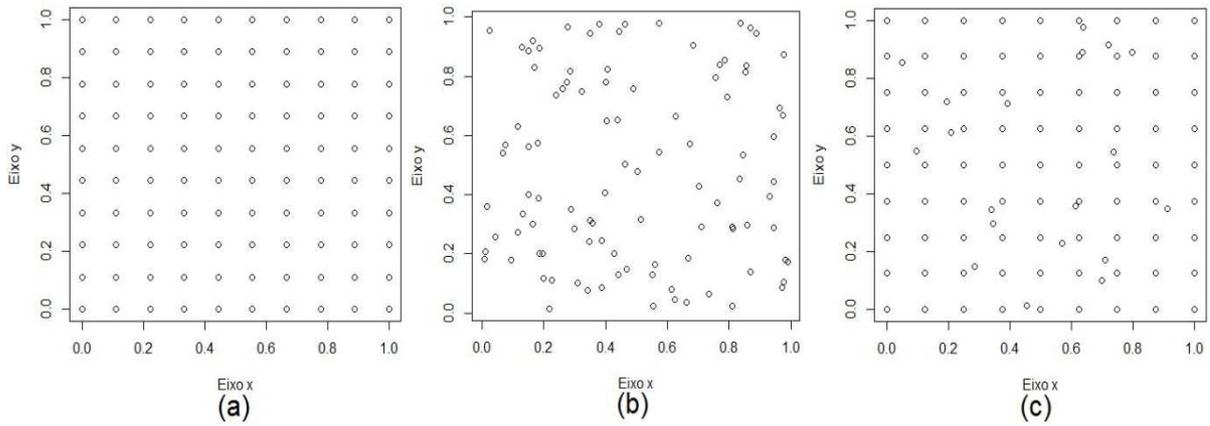


Figura 18 Gráfico da localização dos 100 pontos amostrais iniciais (\circ), para um exemplo de uma simulação da amostragem sistemática (a), aleatória (b) e *lattice plus close pairs*(c).

Aplicou-se o AG para cada simulação em cada ensaio. Para isso, considerou-se um número de indivíduos da população inicial igual a 100 indivíduos. Cada indivíduo era composto por uma amostra inicial reduzida, composta por 20 pontos da malha amostral original. Para cada indivíduo, calculou-se a EG (CONGALTON; GREEN, 1999), conforme a Equação 45. Tal medida compara os valores preditos obtidos pelas configurações inicial e otimizada a partir da matriz de erros, que tem como unidade de medida o pixel. Cada elemento desta matriz representa o número total de pixels ou a área total pertencente à classe i do mapa modelo e à classe j do mapa de referência. Neste caso, o mapa modelo é construído com a configuração amostral inicial e o mapa de referência com a configuração amostral otimizada (DE BASTIANI; URIBE-OPAZO; DALPOSSO, 2012).

$$EG = \frac{\sum_{i=1}^k n_{ii}}{N}, \quad (45)$$

em que n_{ii} são os elementos da diagonal principal da matriz de erros, que representam os pixels (ou áreas) que tiveram a mesma classificação para os dois mapas comparados; k é o número de classes e N é o total de pixels ou a área total (DE BASTIANI; URIBE-OPAZO; DALPOSSO, 2012).

No passo seguinte, 75% dos indivíduos que apresentaram os maiores valores para a função objetivo EG foram selecionados para o processo de reprodução (cruzamento). Esta porcentagem (75%) foi definida a partir de testes para que os indivíduos selecionados para o processo de reprodução pudessem garantir a representatividade da população (LINDEN, 2012).

Realizaram-se, no AG, dois mil (2000) cruzamentos entre os indivíduos considerando-se um ponto de corte (LINDEN, 2012). Em cada cruzamento, dois filhos (indivíduos obtidos pelo cruzamento) foram criados e, para cada filho, foi calculado o valor de EG. Posteriormente, o melhor filho foi acrescentado à população do AG, ao mesmo tempo em que se eliminava desta população o indivíduo que apresentou o menor valor de EG.

Ao final do AG avaliou-se se o melhor indivíduo atingiu valor da EG maior ou igual a 0,85 e, caso tenha sido satisfeito, o processo termina. Caso contrário, o tamanho amostral de cada indivíduo é aumentado em 5% do total da amostra inicial. E a nova população de filhos foi composta por 50% dos melhores indivíduos da etapa anterior. Desta forma, o AG foi aplicado sucessivamente, até que a amostra otimizada pelo algoritmo atingisse o valor mínimo da EG igual a 0,85, pois valores de $EG \geq 0,85$ indicam similaridade (ANDERSON *et al.*, 2001). Ao final deste processo, obteve-se a configuração amostral otimizada de tamanho reduzido. O fluxograma apresentado na Figura 19 exemplifica o funcionamento do AG para a busca da melhor configuração amostral de tamanho reduzido.

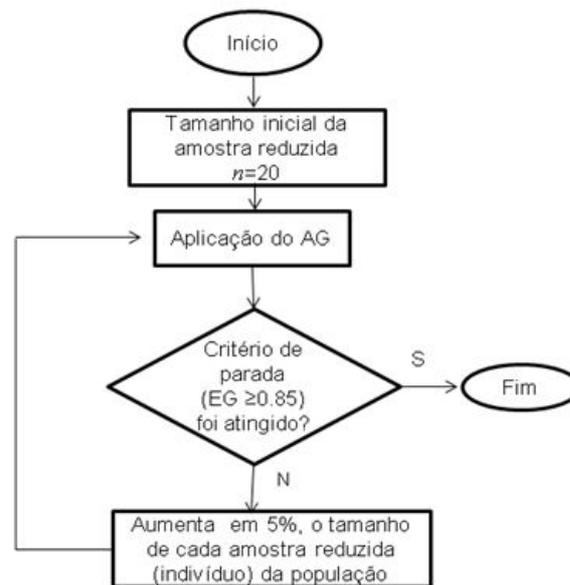


Figura 19 Fluxograma do funcionamento para a busca da melhor configuração amostral de tamanho reduzido.

5.1.2.2. Estudo prático

O conjunto de dados utilizado nesta pesquisa refere-se a uma área comercial de produção de grãos com 167,35 ha, composta por 102 pontos amostrais e que está localizada na Fazenda Agassiz em Cascavel/PR, cuja localização geográfica é, aproximadamente, 24,95° Sul de latitude, 53,57° Oeste de longitude e altitude média de 650 metros.

A configuração amostral definida nesta área foi sistemática centrada com pares de pontos próximos (*lattice plus close pairs*), com distância mínima entre os pontos da grade regular de 141 metros. A amostragem foi realizada com distâncias menores (75 metros e 50 metros entre pares de pontos) em alguns locais escolhidos aleatoriamente (Figura 20). As amostras foram localizadas e georreferenciadas por um aparelho receptor de sinal com o sistema de posicionamento global (GPS) GEOEXPLORE 3, em um sistema espacial de coordenadas UTM.

Os seguintes atributos químicos do solo foram estudados: teor de Cálcio (Ca) [cmol dm^{-3}], de Carbono (C) [g dm^{-3}], de Cobre (Cu) [cmol dm^{-3}] e de Manganês (Mn) [cmol dm^{-3}]; e pH, observados no ano safra 2010/2011.

A amostragem do solo para a determinação dos níveis dos atributos químicos do solo foi realizada em cada ponto demarcado (Figura 20). Nesses, foram coletadas quatro subamostras de solo, de 0,0 a 0,2 m de profundidade nas proximidades dos pontos, misturadas e colocadas em sacos plásticos, com amostras de aproximadamente 500 g. Assim, as amostras representativas para cada parcela foram compostas e encaminhadas ao Laboratório da Cooperativa Central de Pesquisas Agrícolas – COODETEC, onde se procedeu a análise química, utilizando-se o Método Walkley-Black.

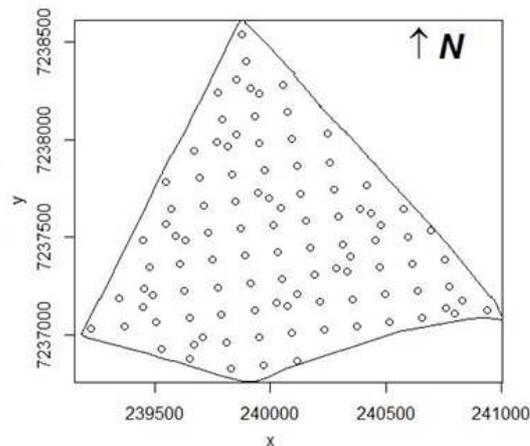


Figura 20 Mapa da área em estudo e da configuração amostral.

Realizaram-se as análises exploratória e geoestatística para cada atributo químico do solo. A existência de anisotropia foi avaliada pelo teste não paramétrico de isotropia de Maity e Sherman (2012) (MS). Em seguida, estimaram-se modelos para a função semivariância pelo método máxima verossimilhança (DIGGLE; RIBEIRO JR., 2007), considerando-se os modelos: exponencial, gaussiano e da família Matérn com $k=1$; 1,5 e 2. O melhor modelo foi escolhido pelo método da validação cruzada (FARACO *et al.*, 2008; ASSUMPÇÃO; HADLICH, 2017). Posteriormente, por krigagem ordinária, realizou-se a predição espacial de cada atributo químico do solo em localizações não amostradas na área agrícola em estudo, e confeccionaram-se os mapas temáticos dos valores estimados de cada atributo (CRESSIE, 2015).

Em seguida, aplicou-se o AG para cada atributo químico do solo. Ele foi realizado da mesma forma e com os mesmos critérios que foram aplicados nas simulações. Ao final do processo de otimização, obteve-se uma configuração amostral de tamanho reduzido para cada atributo químico do solo, na qual, novamente, foram realizadas as análises exploratória e geoestatística. E a comparação dos mapas temáticos da configuração amostral original com a configuração amostral otimizada, para cada um dos atributos, foi obtida a partir do

cálculo da medida de similaridade EG e dos índices de concordância Kappa e Tau (De BASTIANI; URIBE-OPAZO; DALPOSSO, 2012; KESTRING *et al.*, 2015).

As simulações, as rotinas computacionais do AG, as análises estatísticas e geoestatísticas foram realizadas no software R (R DEVELOPMENT CORE TEAM, 2017) utilizando os pacotes geoR (RIBEIRO JR.; DIGGLE, 2001) e sm (BOWMAN; AZZALINI, 2015).

5.1.3. Resultados

5.1.3.1. Estudos de simulação

Observa-se, que em todas as simulações realizadas para as três configurações amostrais (Tabela 2), os valores estimados dos mínimos e máximos da EG calculados para os indivíduos que foram obtidos após o processo de otimização estão muito próximos, como também, obteve-se um baixo valor do desvio padrão para o valor estimado de EG ao final do processo. Os resultados indicam que houve eficiência e estabilidade do processo de otimização, mostrando que o espaço de busca pela solução do processo está reduzido em poucos valores.

Ainda, em média, observou-se aumento relevante da exatidão global (de 58,1% para 93,9%), quando se compararam os valores máximos da EG obtidos no início e ao final do processo de otimização, indicando uma eficiência do processo (Tabela 2).

Considerando-se todos os ensaios e todas as configurações amostrais simuladas, as melhores configurações amostrais de tamanho reduzido obtidas pelo AG tiveram, em média, entre 41 e 57 pontos. Houve, portanto, uma redução de 59% e 43% quanto ao número de pontos em relação à amostra original (Tabela 2). O menor tamanho amostral foi obtido com o menor valor para o efeito pepita e o maior alcance prático. Enquanto o maior tamanho amostral foi obtido com o maior valor para o efeito pepita e o menor valor do alcance prático.

Em todas as amostragens simuladas, para cada alcance prático considerado, observou-se que o aumento do parâmetro efeito pepita gerou aumento do tamanho amostral reduzido. O aumento do alcance prático, para cada valor do efeito pepita, gerou redução do número de pontos amostrais otimizados, principalmente quando foi considerado o menor valor de efeito pepita. Os resultados foram semelhantes aos obtidos por Guedes *et al.* (2011), mesmo trabalhando somente com a otimização da configuração amostral e tamanho amostral previamente fixado. Foram encontrados resultados semelhantes sobre a diminuição do tamanho amostral com o aumento do valor do alcance prático, sob o cálculo do tamanho amostral em GRIFFITH (2005) e VALLEJOS e OSORIO (2014). Os autores utilizaram como metodologia o cálculo do tamanho amostral efetivo (ESS) para a redução deste tamanho amostral. E, com a variação do efeito pepita, também houve maior variação

(em %) do valor máximo estimado da EG, quando se comparou seu valor obtido no início e ao final do processo de otimização.

Tabela 2 Valores médios e desvio padrão (entre parênteses) do mínimo e o máximo da EG, calculada com a amostra de tamanho reduzido, considerando-se todas as simulações, obtidos no início e no final do processo de otimização, para as amostragens aleatória, *lattice plus close pairs* e sistemática.

Amostragem inicial / Parâmetros		$\alpha = 0,45$		$\alpha = 0,60$		$\alpha = 0,90$	
		Mín.	Max.	Mín.	Max.	Mín.	Max.
Aleatória	Inicial	0,3886 (0,0301)	0,5009 (0,0374)	0,4023 (0,0317)	0,5172 (0,0426)	0,4225 (0,0364)	0,5468 (0,0436)
	Final	0,8630 (0,0129)	0,8638 (0,0129)	0,8636 (0,0130)	0,8640 (0,0129)	0,8632 (0,0364)	0,8644 (0,0136)
	Aumento (%) T.amostr	46,5		44,6		41,05	
$\varphi_1 = 0$	Inicial	0,3668 (0,0274)	0,4580 (0,0336)	0,369 (0,0269)	0,4653 (0,0361)	0,3729 (0,0264)	0,4717 (0,0365)
	Final	0,8621 (0,0129)	0,8635 (0,0130)	0,8618 (0,0112)	0,8621 (0,0115)	0,8630 (0,0116)	0,8634 (0,0115)
	Aumento (%) T.amostr	55,4		55,25		54,35	
$\varphi_1 = 0,5$	Inicial	0,3595 (0,0268)	0,4468 (0,0355)	0,3638 (0,0264)	0,4525 (0,0320)	0,3658 (0,0244)	0,4573 (0,0328)
	Final	0,8627 (0,0120)	0,8632 (0,0114)	0,8607 (0,0130)	0,8613 (0,0132)	0,8615 (0,1290)	0,8620 (0,0130)
	Aumento (%) T.amostr	57,25		55,25		55,55	
$\varphi_1 = 0,8$	Inicial	0,3856 (0,0238)	0,4841 (0,0307)	0,3933 (0,0275)	0,5010 (0,0374)	0,4076 (0,0320)	0,5221 (0,0413)
	Final	0,8634 (0,0118)	0,8641 (0,0117)	0,8615 (0,0115)	0,8624 (0,0111)	0,8622 (0,0112)	0,8625 (0,0110)
	Aumento (%) T.amostr	49,5		47		44,35	
<i>Lattice plus close pairs</i>	Inicial	0,3712 (0,0238)	0,4605 (0,0320)	0,3728 (0,0245)	0,4643 (0,0304)	0,3744 (0,0260)	0,4738 (0,0335)
	Final	0,8614 (0,0119)	0,8626 (0,0119)	0,8629 (0,0103)	0,8634 (0,0107)	0,8624 (0,0125)	0,8629 (0,0127)
	Aumento (%) T.amostr	54,25		53,95		53,20	
$\varphi_1 = 0,5$	Inicial	0,3668 (0,0226)	0,4547 (0,0291)	0,3671 (0,0230)	0,4551 (0,0298)	0,3687 (0,0245)	0,4619 (0,0315)
	Final	0,8635 (0,0147)	0,8638 (0,0146)	0,8613 (0,0107)	0,8618 (0,0111)	0,8629 (0,0130)	0,8635 (0,0131)
	Aumento (%) T.amostr	56,2		55,3		55,35	
$\varphi_1 = 0,8$	Inicial	0,3783 (0,0240)	0,4732 (0,318)	0,3850 (0,0277)	0,4860 (0,0341)	0,3975 (0,0303)	0,5118 (0,0394)
	Final	0,8619 (0,0118)	0,8621 (0,0117)	0,8627 (0,0129)	0,8634 (0,0124)	0,8626 (0,0110)	0,8630 (0,0112)
	Aumento (%) T.amostr	50,65		48,9		45,5	
Sistemática	Inicial	0,3659 (0,0248)	0,4515 (0,0334)	0,3661 (0,0255)	0,4526 (0,0317)	0,3703 (0,0242)	0,4608 (0,0348)
	Final	0,8631 (0,0124)	0,8635 (0,0123)	0,8627 (0,0123)	0,8632 (0,0126)	0,8638 (0,0126)	0,8641 (0,0124)
	Aumento (%) T.amostr	55,25		54,75		54	
$\varphi_1 = 0,5$	Inicial	0,3636 (0,0253)	0,4450 (0,0338)	0,3640 (0,0256)	0,4469 (0,0322)	0,3631 (0,0253)	0,4507 (0,0316)
	Final	0,8627 (0,0129)	0,8629 (0,0129)	0,8629 (0,0122)	0,8632 (0,0121)	0,8638 (0,0135)	0,8647 (0,0135)
	Aumento (%) T.amostr	56,25		56,30		55,85	
$\varphi_1 = 0,8$	Inicial	0,3636 (0,0253)	0,4450 (0,0338)	0,3640 (0,0256)	0,4469 (0,0322)	0,3631 (0,0253)	0,4507 (0,0316)
	Final	0,8627 (0,0129)	0,8629 (0,0129)	0,8629 (0,0122)	0,8632 (0,0121)	0,8638 (0,0135)	0,8647 (0,0135)
	Aumento (%) T.amostr	56,25		56,30		55,85	

α : é o valor do alcance simulado; φ_1 : é o valor do efeito pepita simulado; Min.: é o valor mínimo da exatidão global (EG); Max.: é o valor máximo da exatidão global (EG); % de aumento médio do valor máximo final da exatidão global em relação ao valor máximo inicial; T.: tamanho.

O valor estimado da EG obtido ao final da otimização não sofreu tanta influência com a variação dos parâmetros efeito pepita e alcance prático, os quais variaram de 0,8613 a 0,8647. Este fato foi consequência do critério de parada do algoritmo (valor estimado de EG ser maior ou igual a 0,85) (Tabela 2).

Comparando-se os resultados obtidos em todos os ensaios, quando variou-se a configuração amostral original, a amostragem sistemática apresentou em média os maiores tamanhos amostrais otimizados para quase todos os ensaios.

Comparando-se a disposição dos pontos da malha amostral otimizada para cada configuração amostral, não se identificou um padrão quanto à disposição dos pontos escolhidos, para nenhum dos casos das simulações. É possível visualizar a localização dos 100 pontos amostrais iniciais e dos pontos selecionados pelo AG, pelas Figuras 7, 8 e 9 (APÊNDICE), para um exemplo de cada simulação de cada configuração amostral.

5.1.3.2. Redimensionamento amostral considerando os atributos químicos do solo

Inicialmente a área comercial possuía 102 pontos. Após a utilização do algoritmo genético, a malha amostral dos atributos químicos do solo Ca e Mn foi composta por 35 pontos (Tabela 3 e Figuras 21a e 21d). Entretanto, a malha amostral otimizada foi composta por 30 pontos para os atributos C, Cu e pH (Tabela 3 e Figuras 21b, 21c e 21e). Tal redução corresponde respectivamente em 29,41% e 34,31% da malha amostral inicial. No artigo proposto por Guedes *et al.* (2011), quando otimizaram uma malha amostral com o tamanho amostral reduzido fixo em 50% da malha inicial. Os referidos autores obtiveram bons resultados quanto à eficiência da predição espacial para a classificação de Potássio na área em estudo. Guedes *et al.* (2014) obtiveram redução do tamanho amostral em 40% da malha amostral, porém, havia uma quantidade maior de pontos. O estudo foi realizado por recozimento simulado, considerando-se a EG. Com isso, pode-se perceber que neste trabalho encontrou-se maior redução amostral do que naqueles trabalhos.

O algoritmo AG demorou aproximadamente 50 minutos para otimizar o tamanho e a configuração amostral, com relação ao tempo computacional, para cada atributo químico do solo, haja vista a máquina usada possuir Processador Intel(R) Core(TM) i7-4790 CPU @ 3.60GHz, 3601 Mhz, 4 Núcleos, 8 Processadores Lógicos e memória física (RAM) instalada 4,00 GB.

Os resultados da análise exploratória dos atributos químicos do solo bem como o valor do coeficiente de correlação linear de Pearson dos atributos químicos do solo com as coordenadas dos eixos X e Y são apresentados na Tabela 3, considerando-se a configuração amostral original e a otimizada.

Observa-se, pela Tabela 3, que mesmo com a redução amostral, as estatísticas descritivas dos atributos químicos do solo, da malha amostral reduzida foram semelhantes

aos resultados da malha original. O mesmo fato também ocorreu em Guedes *et al.* (2014), os quais utilizaram o recozimento simulado para otimizar o tamanho e as configurações amostrais para análise da produtividade de soja. Em particular, a similaridade dos valores da média e mediana entre os diferentes tamanhos amostrais, de cada atributo químico do solo, indica que tais configurações representam uma configuração amostral semelhante. Este fato também ocorreu em Souza *et al.*, (2014), os quais analisaram diferentes intensidades de amostragens de solo com relação à precisão na análise geoestatística e na interpolação de mapas, para fins de agricultura de precisão em áreas de cana-de-açúcar.

Tabela 3 Estatística descritiva e valor do coeficiente de correlação linear de Pearson dos atributos químicos do solo Ca [cmolc dm⁻³], C [g dm⁻³], Cu [mg dm⁻³], Mn [cmolc dm⁻³] e pH, considerando a configuração amostral original e a de tamanho reduzido

Estatística	Configuração amostral	Ca	C	Cu	Mn	pH
Média	Original	5,20	26,93	2,95	49,32	5,10
Mínimo		2,37	19,87	1,10	17,00	4,40
Q1		4,21	24,55	2,30	36,25	4,80
Mediana		5,08	26,88	2,80	43,00	5,10
Q3		6,10	29,61	3,68	59,50	5,30
Máximo		11,76	34,29	4,90	107,00	6,70
Variância		1,88	10,55	0,68	379,51	0,15
C.V (%)		26,41	12,06	27,86	39,50	7,58
Coef. X		0,14	-0,08	-0,08	-0,08	0,14
Coef. Y		0,40	0,36	-0,56	0,59	0,35
<i>n</i>		35	30	30	35	30
Média	De tamanho reduzido	5,17	26,88	3,01	50,94	5,06
Mínimo		2,51	20,65	1,10	23,00	4,40
Q1		3,96	24,74	2,33	34,00	4,80
Mediana		4,93	26,88	2,90	47,00	5,05
Q3		6,15	28,83	3,78	64,00	5,30
Máximo		11,76	33,90	4,70	93,00	5,90
Variância		3,11	11,52	0,87	414,11	0,13
C.V (%)		34,09	12,62	31,07	39,95	7,25
Coef. X		0,13	0,06	-0,01	-0,10	0,17
Coef. Y		0,50	0,39	-0,57	0,62	0,45

Q1: representa o primeiro quartil, Q3: representa o terceiro quartil, C.V: é o coeficiente de variação, coef.: coeficiente de correlação linear de Pearson entre os atributos químicos do solo e as coordenadas X e Y e *n*: representa o número de pontos obtidos pela otimização.

Inclusive observa-se que os teores de cálcio (Ca), carbono (C), cobre (Cu) e o atributo pH na amostra reduzida apresentam os mesmos comportamentos de tendência direcional da configuração amostral original (Tabela 3), com valores próximos de coeficiente de correlação linear de Pearson. Há também moderada associação linear dos respectivos valores com as coordenadas do Eixo Y (CALLEGARI-JACQUES, 2003). Já o teor de manganês (Mn) apresentou moderada associação linear para a configuração original e forte associação linear para a configuração amostral reduzida (CALLEGARI-JACQUES, 2003). Desta forma, nos atributos químicos do solo que apresentaram tendência direcional, houve a necessidade de se trabalhar com um modelo linear espacial gaussiano (Equação 43) com média expressa a partir de covariáveis (GRESSIE, 2015).

Além disso, o teor de cálcio (Ca) apresentou pontos discrepantes nas configurações amostrais original e otimizada. Enquanto os atributos químicos do solo Mn e pH apresentaram pontos discrepantes somente na configuração amostral original.

Considerou-se que a estrutura de dependência espacial é isotrópica para todos os atributos químicos do solo, pois pelo teste não paramétrico de isotropia MS, não se rejeitou a hipótese nula de isotropia (p -valor $>0,05$).

Para ambas as configurações amostrais (original e otimizada), o melhor modelo da função semivariância, obtido pelo método da máxima verossimilhança considerando-se a validação cruzada foi: o modelo da família Matérn com $k=2$ para os teores de cálcio (Ca) e carbono (C); o modelo exponencial para os teores de cobre (Cu), de manganês (Mn) e o atributo pH (Tabela 4).

Tabela 4 Valores estimados dos parâmetros dos melhores modelos ajustados, com seus respectivos desvios padrões, para os atributos químicos do solo Ca, C, Cu, Mn e pH.

	Ca	C	Cu	Mn	pH
Modelos	Matérn $k=2$	Matérn $k=2$	Exponencial	Exponencial	Exponencial
$\hat{\beta}_0$	-9916,96	-21325,50	6283,82	$-1,7 \cdot 10^5$	-2455,07
$\hat{\beta}_1$	0,001	0,003	-0,001	0,023	0,0003
$\hat{\varphi}_1$	1,14	4,50	0,09	95,25	0,10
$\hat{\varphi}_1 + \hat{\varphi}_2$	1,56	9,08	0,45	241,4	0,13
\hat{a}	249,14	383,47	805,21	453,49	300,96
\overline{EPR} (%)	72,43	49,53	20,71	39,47	73,75
DP ($\hat{\beta}_0$)	$3,0 \cdot 10^{-9}$	$1,1 \cdot 10^8$	$4,1 \cdot 10^{-9}$	$5,9 \cdot 10^{-8}$	$8,7 \cdot 10^{-10}$
DP ($\hat{\beta}_1$)	$2,1 \cdot 10^{-5}$	$7,7 \cdot 10^{-5}$	$2,9 \cdot 10^{-5}$	0,00043	$6,3 \cdot 10^{-6}$
DP ($\hat{\varphi}_1$)	0,11223	0,00027	0,06244	$8,8 \cdot 10^{-5}$	0,02528
DP ($\hat{\varphi}_2$)	0,11195	$3,9 \cdot 10^{-5}$	0,12878	$2,8 \cdot 10^{-5}$	0,02616
DP ($\hat{\varphi}_3$)	0,02242	0,01961	0,14615	0,05603	0,01952
Modelo	Matérn $k=2$	Matérn $k=2$	Exponencial	Exponencial	Exponencial
$\hat{\beta}_0$	-16370,00	-28110,00	8299,05	$-1,8 \cdot 10^5$	-3346,60
$\hat{\beta}_1$	0,0002	0,004	-0,001	0,025	0,0005
$\hat{\varphi}_1$	1,71	3,43	0,09	74,41	0,06
$\hat{\varphi}_1 + \hat{\varphi}_2$	2,27	9,16	0,56	247,81	0,10
\hat{a}	311,81	435,15	625,84	579,09	237,42
\overline{EPR} (%)	71,09	37,46	14,30	30,02	54,76
DP ($\hat{\beta}_0$)	$5,5 \cdot 10^{-9}$	$1,5 \cdot 10^8$	$4,6 \cdot 10^{-9}$	$8,2 \cdot 10^{-8}$	$1,2 \cdot 10^{-9}$
DP ($\hat{\beta}_1$)	$4,0 \cdot 10^{-5}$	0,00011	$3,3 \cdot 10^{-5}$	0,00059	$8,9 \cdot 10^{-6}$
DP ($\hat{\varphi}_1$)	0,00169	0,00053	0,07800	0,00019	0,01365
DP ($\hat{\varphi}_2$)	0,00113	$8,6 \cdot 10^{-5}$	0,10143	$6,0 \cdot 10^{-5}$	0,01325
DP ($\hat{\varphi}_3$)	0,05742	0,03690	0,12970	0,11130	0,00037

$\hat{\beta}_0, \hat{\beta}_1$: são os valores estimados dos parâmetros do modelo de regressão, que explicam a média; $\hat{\varphi}_1$: valor estimado do efeito pepita, $\hat{\varphi}_2$: valor estimado da contribuição, \hat{a} : valor estimado do alcance prático (m), $\overline{EPR} = \hat{\varphi}_1 / (\hat{\varphi}_1 + \hat{\varphi}_2)$: valor estimado do efeito pepita relativo (%), $\hat{\varphi}_1 + \hat{\varphi}_2$: valor estimado do patamar e DP: desvio padrão.

Observou-se um valor estimado para o raio de dependência espacial para os atributos químicos do solo Ca, C, Cu, Mn e pH, considerando a configuração amostral original (alcance), cuja variação foi de 249,14m a 805,21m (Tabela 4). Todavia, o alcance variou de 237,42 a 625,84m para a configuração amostral otimizada.

Assim, ambas as configurações amostrais, apresentaram valor estimado do raio de dependência espacial similar para cada atributo químico do solo, com exceção de Cu e Mn, que apresentam maior diferença quando comparado ao raio de dependência espacial, considerando-se a malha inicial, com o obtido considerando a otimizada. A comparação dos alcances dos atributos químicos Ca, C e Mn obtidos, de acordo com a malha amostral original com a de tamanho reduzido, revelou que o raio de dependência espacial teve um leve aumento de 25%, 13,5% e 27,7%, respectivamente. E para os atributos químicos Cu e pH, o raio de dependência espacial teve uma leve queda de 22,3% e 21%, respectivamente (Tabela 4).

Segundo os critérios usados por Cambardella *et al.* (1994) no que tange à intensidade de dependência espacial, tem-se que os atributos Ca, C, Mn e pH apresentaram moderada dependência espacial para as configurações amostrais original e otimizada (Tabela 4). Enquanto o teor de cobre (Cu) no solo apresentou forte dependência espacial nas configurações amostrais original e otimizada (Tabela 4).

Os valores estimados dos desvios padrões dos parâmetros foram calculados para ambas as configurações amostrais. Observou-se um aumento nos valores dos desvios padrão da média estimada, quando se comparou a configuração amostral reduzida com a configuração original (Tabela 4).

Os mapas construídos a partir da krigagem ordinária (Figura 21) para a configuração otimizada reduzida apresentaram para todos os atributos químicos do solo semelhanças visuais, quando comparados com os mesmos mapas construídos considerando-se a configuração inicial.

Além disso, observa-se que os maiores valores dos teores de cálcio (Ca) e manganês (Mn) concentraram-se na região norte da área de estudo, enquanto o teor de cobre (Cu) concentra seus maiores valores na região sul (Figura 21). Os atributos químicos do solo teor de carbono (Ca) e pH apresentam concentrações mais espalhadas pela área de estudo (Figura 21).

Quando se compara a disposição dos pontos obtidos pelo processo de otimização, verifica-se que o algoritmo de otimização escolheu os pontos amostrais da configuração otimizada de forma a selecionar pontos de cada subárea, ou ponto próximo a cada subárea descrita pelo mapa temático da malha original. Sabe-se que o número de pontos correspondentes de cada subárea está relacionado à proporção que tal subárea representa na área agrícola. Observou-se ainda que os atributos químicos do solo carbono (C) e pH apresentaram formação de alguns pequenos aglomerados de pontos escolhidos, mas a posição dos aglomerados não segue uma tendência (Figura 21).

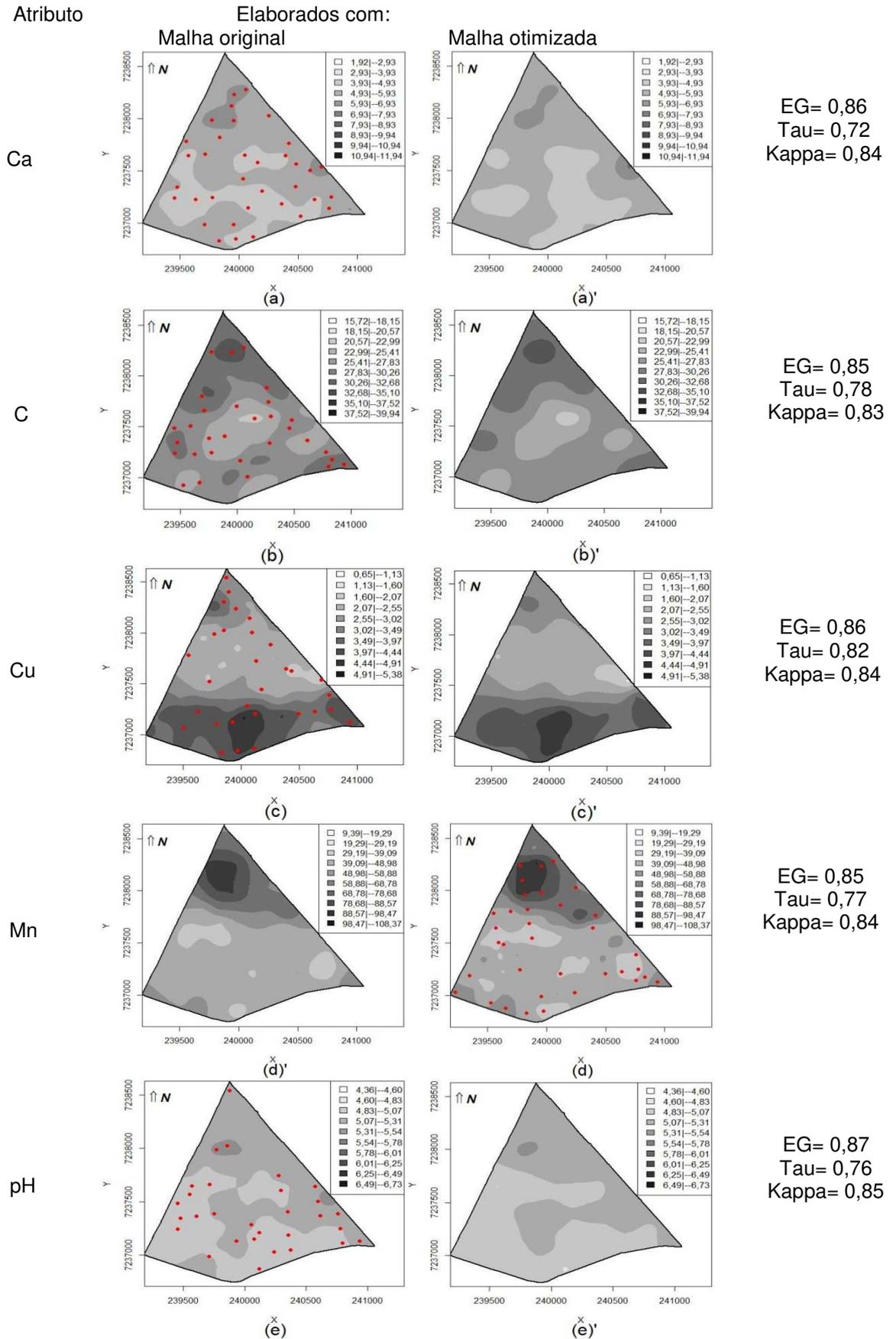


Figura 21 Mapas temáticos dos atributos químicos do solo elaborados considerando a malha amostral original e otimizada, em que ● representa a localização dos pontos seleccionados no processo de otimização e valores estimados das medidas de similaridade EG, Tau e Kappa. X: Coordenadas do eixo X e Y: Coordenadas do eixo Y.

Quando comparados os mapas temáticos dos atributos químicos do solo Ca, C, Cu, Mn e pH gerados a partir das malhas original e otimizada, observou-se que os atributos químicos do solo apresentaram valores estimados de exatidão global (EG) entre 0,85 a 0,87. Isso indica que os mapas são similares, ou seja, são elaborados considerando com ambas as configurações são semelhantes quanto à distribuição dos teores dos atributos na área em estudo $EG \geq 0,85$ (ANDERSON *et al.*, 2001) (Figura 21). Além disso, de acordo com os valores estimados dos índices de concordância Kappa e Tau, todos os atributos químicos do solo apresentaram média ou alta exatidão, com valores estimados entre 0,72 a 0,82 (média exatidão se $0,67 \leq K < 0,80$ e alta exatidão se $K \geq 0,80$) (KRIPENDORFF, 2013).

5.1.4. Conclusão

O processo de otimização foi eficiente e estável quando aplicado nos dados simulados e nos atributos químicos do solo. Para todos os atributos químicos do solo, recomenda-se que o número de pontos a serem utilizados em experimentos futuros varie de 30 a 35 pontos. Recomenda-se também que a escolha de configurações amostrais com tamanho reduzido seja obtida pelo AG, pois considerou-se como similaridade mínima de predição espacial, em relação à configuração original, igual a 85%.

Assim, conclui-se que é possível redimensionar a grade amostral que envolve este experimento ao se reduzir o tamanho amostral entre 29,41% e 34,31% do tamanho amostral da malha original, sem perda de informação quanto à descrição da estrutura de dependência espacial e à descrição da variabilidade espacial apresentada pelos mapas.

Observou-se que o parâmetro efeito pepita exerceu relevante influência negativa no tamanho amostral para os dados simulados pois, quanto maior o valor do efeito pepita, maior o tamanho amostral. Os menores tamanhos amostrais ocorreram com o menor valor de efeito pepita e o maior valor do alcance, para ambas as configurações amostrais.

Agradecimentos

Os autores agradecem à CAPES, ao CNPq e à Fundação Araucária pelo apoio financeiro.

5.1.6. Referências Bibliográficas

- ANDERSON, J. F.; HARDY, E. E.; ROACH J. T.; WITMER R. E. **A land use and land cover classification system for use with remote sensor data**. U.S. Washington, DC: U.S. Geologic Survey, 2001. 41 p.
- ASSUMPÇÃO, H. C. P.; HADLICH, G. M. Estatística descritiva e estacionaridade em variáveis geoquímicas ambientais. **Engenharia Sanitária e Ambiental**, v.22, n.4, p. 671-677, 2017.
- BERNARDI, A. C. C.; NAIDE, J. M.; RESENDE, A. V.; BOSSOI, L. H.; INAMASU, R. Y. **Agricultura de precisão: resultados de um Novo olhar**. EMBRAPA, Brasília- DF, 600p, 2014.
- BOWMAN A. W.; AZZALINI A. **R package sm: nonparametric smoothing methods** (version 2.2-5.4). University of Glasgow, UK and Università di Padova, Itália. Accessed on Dec. 12, 2015.
- CALLEGARI-JACQUES, S. M. **Bioestatística: princípios e aplicações**. Porto Alegre: Artemed, 2003. 255 p.
- CÂMARA, G.; MEDEIROS, J. S. GIS para meio ambiente. In: Sistema de informações geográficas, 1998, Curitiba. **Anais...** Curitiba: Sagres, p.140-180, 1998.
- CAMBARDELLA, C.A.; MOORMAN, T.B.; PARKIN, T.B.; NOVACK, J.M.; KARLEN, D.L.; TURCO, R.F.; KNOPKA, A.E. Field-scale variability of soil properties in Central Iowa Soils. **Soil Science Society American Journal**, Medison, v.58, n.5, p. 1501-1511, 1994.
- CÁRDENAS, R. R.; FERREIRA, M. A. R.; SCHMIDT, A. M. Stochastic search algorithms for optimal design of monitoring networks. **Environmetrics**, London, v. 21, n.1, p.102–112, 2010.
- CONGALTON, R. G.; GREEN, K. **Assessing the accuracy of remotely sensed data: principles and practices**. New York: Lewis Publisher, 1999, 130p.
- CRESSIE, N. A. C. **Statistics for Spatial Data** (Rev. Ed.). Canada, John Wiley & Sons, 2015, 928p.
- DE BASTIANI, F.; URIBE-OPAZO, M. A.; DALPOSSO, G. H. Comparison of maps of spatial variability of soil resistance to penetration constructed with and without covariables using a spatial linear model. **Engenharia Agrícola**, v.32, n.2, p.393-404, 2012.
- DE BASTIANI, F.; CYSNEIROS, A. H. M. A.; URIBE-OPAZO, M. A.; GALEA, M. Influence diagnostics in elliptical spatial linear models. **Sociedad de Estadística e Investigación Operativa**, TEST, Berlin, v.24, n.2, p. 322-340, 2015.
- DE BASTIANI, F.; GALEA, M.; CYSNEIROS, A. H. M. A.; URIBE-OPAZO, M. A. Gaussian spatial linear models with repetitions: an application to soybean productivity. **Spatial Statistics**, v.1, p. 319-335, 2017.
- DIGGLE, P.J.; RIBEIRO JR., P.J. **Model-based geostatistics**. New York: Springer, 2007. 230 p.
- FARACO, M. A.; URIBE-OPAZO, M. A. SILVA, E. A.; JOHANN, J. A.; BORSSOI, J. A. Seleção de modelos de variabilidade espacial para elaboração de mapas temáticos de atributos físicos do solo e produtividade da soja. **Revista Brasileira de Ciência do Solo**,

Viçosa, v. 32, n. 2, p. 463-476, 2008.

GOLDBARG, M. C.; GOLDBARG, E. G.; LUNA, H. P. L. **Otimização Combinatória e Meta-heurísticas**: algoritmos e aplicações. Elsevier, Rio de Janeiro, 2016.

GREGO, C. R.; VIEIRA, S. R. Variabilidade espacial de propriedades físicas do solo em uma parcela experimental. **Revista Brasileira de Ciência do Solo**, v. 29, n.2, p. 169-177, 2005.

GRIFFITH, D. Effective Geographic Sample Size in the Presence of Spatial Autocorrelation. **Annals of the Association Geographers**, v.95, n.4, p. 740-760, 2005.

GUEDES, L. P. C.; URIBE-OPAZO, M. A.; JOHANN, J. A.; SOUZA, E. G. Anisotropia no estudo da variabilidade espacial de algumas variáveis químicas do solo. **Revista Brasileira de Ciência do Solo**, Viçosa, v.32, n.6, p. 2217-2226, Nov./dec. 2008.

GUEDES, L.P.C.; RIBEIRO JR. P.J.; PIEDADE, S.M.D.S.; URIBE-OPAZO, M.A. Optimization of spatial sample configurations using hybrid genetic algorithm and simulated annealing. **Chilean Journal of Statistics**, Chile, v. 2, n.2, p. 39-50, 2011.

GUEDES, L. P. C.; URIBE-OPAZO, M. A., RIBEIRO JR. P. J. Optimization of sample design size and shapes for regionalized variables using simulated annealing. **Ciência e Investigación Agraria**, v. 41, n.1, p. 33-48, 2014.

GUEDES, L. P. C.; RIBEIRO JR. P. J.; URIBE-OPAZO, M. A.; De BASTIANI, F. Soybean yield maps using regular and optimized sample with different configurations by simulated annealing. **Engenharia Agrícola**, v.36, n.1, p.114-125, 2016.

KESTRING, F. B. F.; GUEDES, L. P. G.; DE BASTINI, F.; URIBE-OPAZO, M. A. Comparação de mapas temáticos de diferentes grades amostrais para a produtividade da soja. **Revista Engenharia Agrícola**, Jaboticabal, v. 35, n. 4, p.733-743, 2015.

KRIPPENDORFF, K. **Content Analysis an Introduction to its Methodology**. California, Sage Publications Ltd, 3rded., 2013, 441p.

LINDEN, R. **Algoritmos Genéticos**. Rio de Janeiro: Ciência Moderna Ltda, 2012. 496p.

MAITY A.; SHERMAN M. Testing for spatial isotropy under general designs. **Journal of Statistical Planning and Inference**, v.142, n.5, p. 1081-1091, 2012.

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

OLIVEIRA, L. W.; OLIVEIRA, A. R.; GOMES, F. V.; DIAS, B. H.; SILVA JR., I. C. Aplicação de algoritmos genéticos para o planejamento de geração distribuída em sistemas de distribuição. **Revista LSP- científica interdisciplinar**. v. 3, n.1, p. 79-97, jan./mar. 2016.

PESSOA, A. L. S.; ULISSES, P. H. C.; BRANCO, H. M. G. C.; RABÊLO, R. A. L. Uma aplicação de algoritmos genéticos simples e compacto para estimação de componentes harmônicas. **Revista Brasileira de Computação Aplicada**, v.7, n.2, p. 77-91, 2015.

R DEVELOPMENT CORE TEAM. R: A language and environment for statistical computing. **R Foundation for Statistical Computing**, Vienna, Austria. ISBN 3-900051-07-0. Disponível em: <<http://www.R-project.org>> Acesso em: 14 out. 2017.

RIBEIRO JR., P. J.; DIGGLE, P. J. geoR: A package for geostatistical analysis. **R-NEWS**, v. 1, n. 2, ISSN 1609-3631, 2001.

SEIDEL, E. J.; OLIVEIRA, M. S. Proposta de uma generalização para os modelos de semivariogramas Exponencial e Gaussiano. **Semina: ciências Exatas e Tecnológicas**. Londrina/PR, v.34, n.1, p.125-132, jan./jul. 2013.

SOUZA, Z. M.; SOUZA, G. S.; MARQUES JÚNIOR., J.; PEREIRA, G. T. Number of samples in geostatistical and kriging maps of soil properties. **Ciência Rural**, v.44, n.2, p. 261-268, 2014.

URIBE-OPAZO, M. A.; BORSSOI, J. A.; GALEA, M. Influence diagnostics in Gaussian spatial linear models. **Journal of Applied Statistics**, v. 39, n. 3, p. 615-630, 2012.

VALLEJOS, R.; OSORIO, F. Effective sample size of spatial process models. **Spatial Statistics**, v.9, n.1, p. 66-92, 2014.

VIEIRA, S. R. Tópicos em ciência do solo. **Sociedade Brasileira de Ciência do Solo**, v.1. Viçosa/MG, 2000. 53 p.

ZHU, Z.; STEIN, M. L. Spatial sampling for design for parameter estimation of the covariance function. **Journal of Statistical Planning and Inference**.v.134, p.583-603, 2005.

APÊNDICE

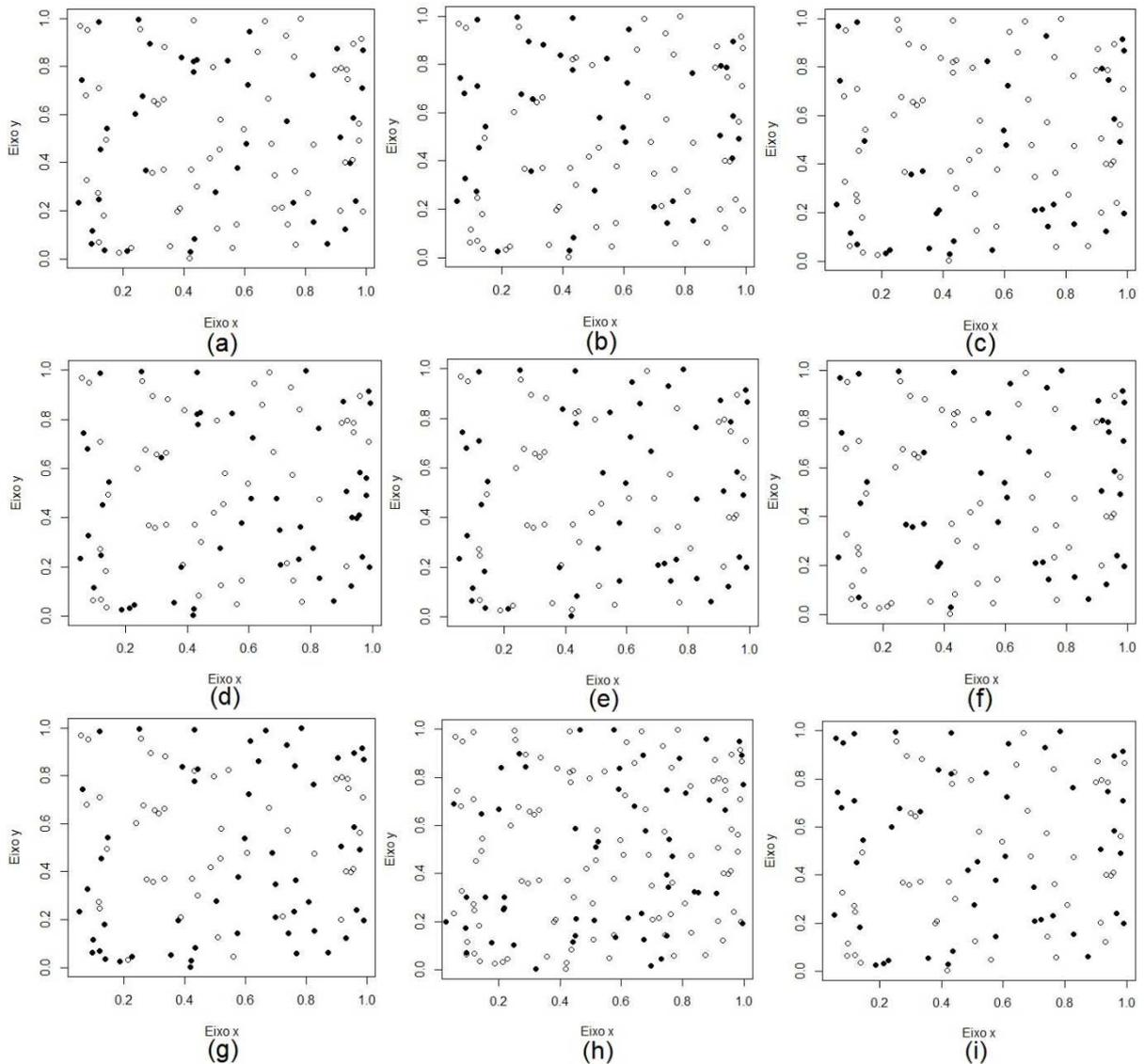


Figura 22 Gráfico da localização dos 100 pontos iniciais (\circ) e a localização dos pontos selecionados (\bullet), para um exemplo de cada simulação da amostragem aleatória simples, (a) alcance 0,45 e efeito pepita 0; (b) alcance 0,60 e efeito pepita 0; (c) alcance 0,90 e efeito pepita 0; (d) alcance 0,45 e efeito pepita 0,50; (e) alcance 0,60 e efeito pepita 0,50; (f) alcance 0,90 e efeito pepita 0,50; (g) alcance 0,45 e efeito pepita 0,80; (h) alcance 0,60 e efeito pepita 0,80; (i) alcance 0,90 e efeito pepita 0,80.

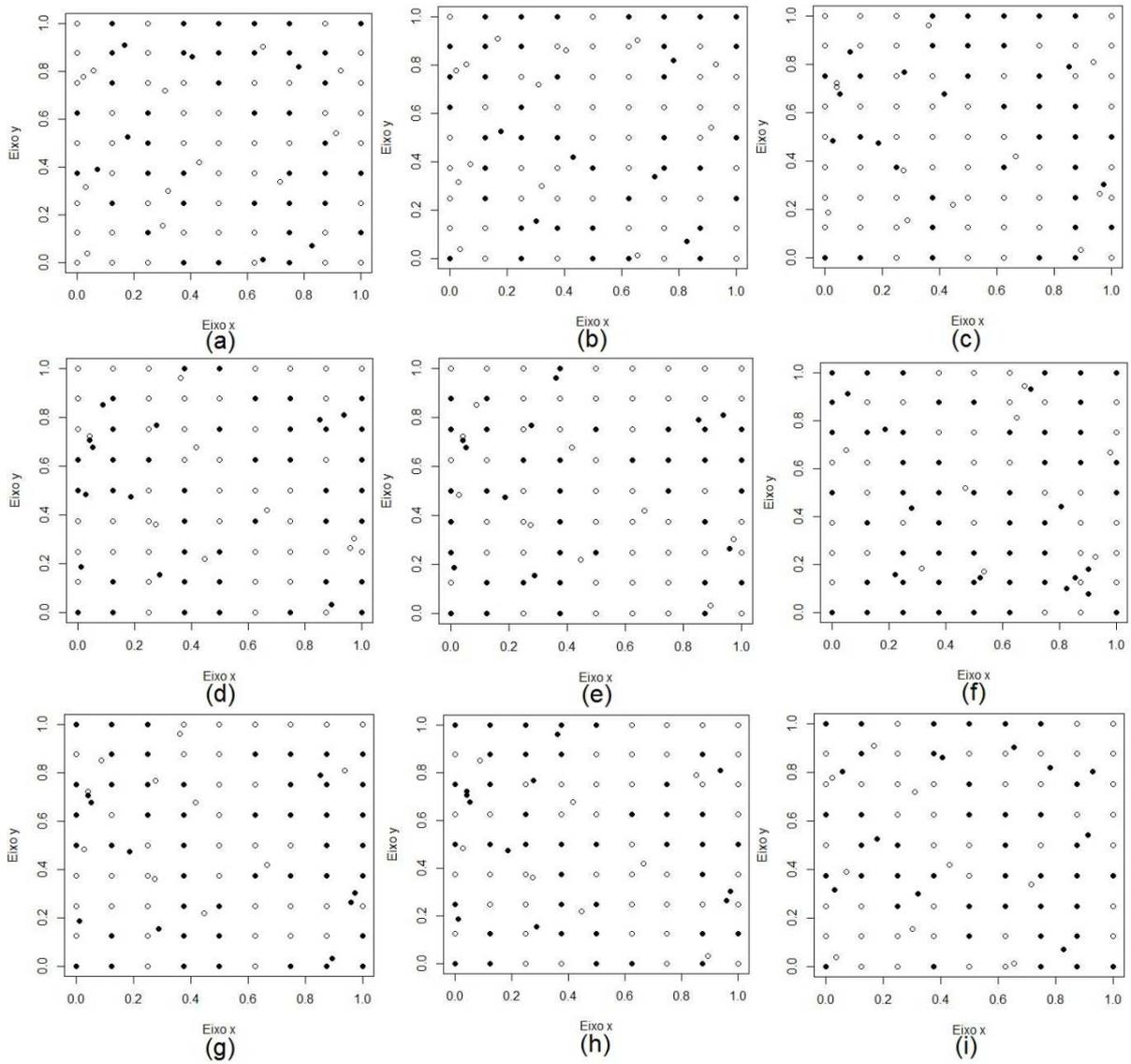


Figura 23 Gráfico da localização dos 100 pontos iniciais (\circ) e a localização dos pontos selecionados (\bullet), para um exemplo de cada simulação da amostragem *latticeplus close pairs*, (a) alcance 0,45 e efeito pepita 0; (b) alcance 0,60 e efeito pepita 0; (c) alcance 0,90 e efeito pepita 0; (d) alcance 0,45 e efeito pepita 0,50; (e) alcance 0,60 e efeito pepita 0,50; (f) alcance 0,90 e efeito pepita 0,50; (g) alcance 0,45 e efeito pepita 0,80; (h) alcance 0,60 e efeito pepita 0,80; (i) alcance 0,90 e efeito pepita 0,80.

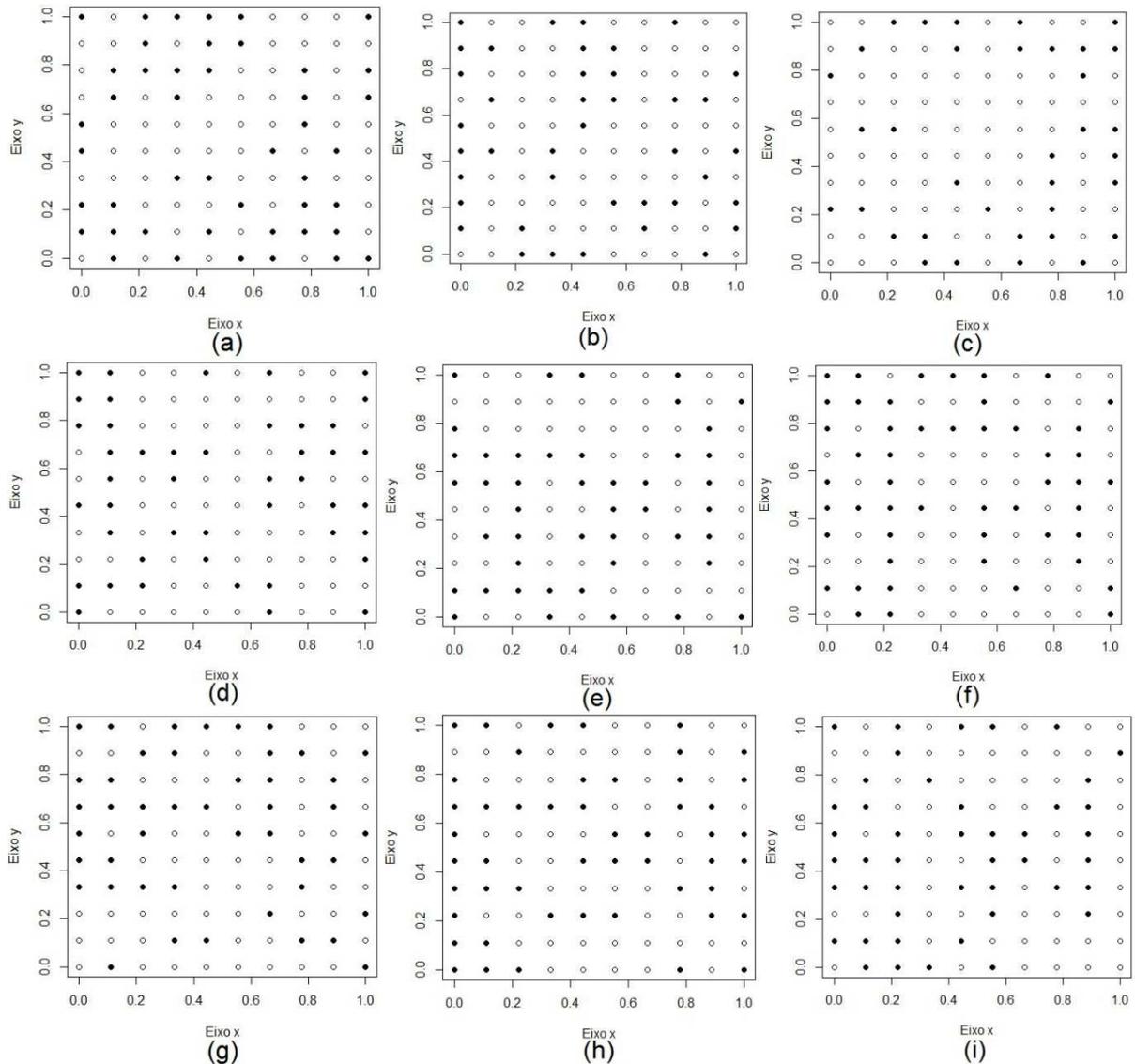


Figura 24 Gráfico da localização dos 100 pontos iniciais (\circ) e a localização dos pontos selecionados (\bullet) para um exemplo de cada simulação da amostragem sistemática, (a) alcance 0,45 e efeito pepita 0; (b) alcance 0,60 e efeito pepita 0; (c) alcance 0,90 e efeito pepita 0; (d) alcance 0,45 e efeito pepita 0,50; (e) alcance 0,60 e efeito pepita 0,50; (f) alcance 0,90 e efeito pepita 0,50; (g) alcance 0,45 e efeito pepita 0,80; (h) alcance 0,60 e efeito pepita 0,80; (i) alcance 0,90 e efeito pepita 0,80.

5.2. ARTIGO 2

Redução otimizada do tamanho amostral na análise da variabilidade espacial de atributos químicos do solo otimizando a eficiência da estimação do modelo geoestatístico

RESUMO

Este trabalho teve por objetivo determinar a melhor configuração amostral com o menor tamanho amostral reduzido, para uma área comercial de cultivo de soja composta por 102 pontos amostrais, onde se desenvolve uma atividade de aplicação localizada de insumos. A redução amostral foi realizada por algoritmo genético (AG) e considerou-se a variabilidade espacial de atributos químicos do solo. O processo AG minimizou uma função-objetivo relacionada à eficiência da estimação do modelo geoestatístico, baseada na matriz de informação de Fischer. Ainda foi avaliada, em estudos de simulação, a influência dos parâmetros do modelo geoestatístico na busca por um tamanho e uma configuração amostral otimizada. Observou-se que, de modo geral, o efeito pepita influenciou negativamente a redução do tamanho amostral de acordo com os dados simulados, enquanto o alcance prático influenciou de forma positiva a redução do tamanho amostral. Ainda, o menor tamanho amostral foi obtido quando foi considerada a amostragem aleatória simples nas simulações na malha inicial. O AG foi eficiente na redução do tamanho amostral para os atributos químicos do solo. Isso ocorreu de tal forma que o tamanho da amostra otimizada variou entre 30 e 45 pontos, o que corresponde respectivamente a 29,41% e 44,12% da malha original. Quanto à tendência direcional, todos os atributos químicos do solo apresentaram aumento da intensidade da tendência direcional dos atributos químicos do solo na direção norte, comparando-se as configurações inicial e otimizada. Quanto aos desvios-padrões das estimativas dos parâmetros, observa-se que, de modo geral, a configuração amostral otimizada apresentou os menores valores para os desvios. Logo, é possível que a configuração amostral otimizada apresente melhor qualidade em relação à estimação dos parâmetros. Todos os atributos químicos apresentaram moderada ou forte dependência espacial para a estrutura de dependência espacial. Em relação à construção dos mapas temáticos, não houve similaridade entre os mapas obtidos para todos os atributos químicos do solo, com o uso na predição espacial das configurações amostrais inicial e otimizada.

Palavras-chave: Algoritmo genético, configurações amostrais, matriz de informação de Fischer, simulação.

Optimized reduction of the sample size during spatial variability analysis of soil chemical attributes to optimize the estimation efficiency of the geostatistical model

ABSTRACT

This trial aimed at determining the best sample configuration with the smallest sample size, for a commercial area of soybean crop. It was composed of 102 sampling points, where an activity of specific inputs application has been developed. This sample reduction was carried out by genetic algorithm (GA) and spatial variability of soil chemical attributes was considered. The GA process minimized an objective function related to the efficiency estimation of geostatistical model, based on Fischer information matrix. The parameters influence of geostatistical model in order to find out an optimized sample size and configuration was also evaluated during simulation studies. It was observed that, in general, the nugget effect influenced negative on sample size reduction according to the simulated data. While, the practical range influenced positively the sample size reduction. The smallest sample size was also obtained when simple random sampling was considered in simulations

in the initial mesh. GA was efficient in reducing the sample size for soil chemical attributes. This occurred in such a way that the optimized sample size range from 30 to 45 points. This corresponds respectively to 29.41% and 44.12% of the original mesh. Regarding the directional trend, all soil chemical attributes showed an increase of the directional trend intensity concerning soil chemical attributes towards the northern area, when compared to the initial and optimized configurations. While for the standard deviations of parameters estimates, it has been observed that the optimum sample configuration showed the lowest values for deviations. Therefore, the optimized sample configuration might present the best quality in relation to the parameters estimation. All chemical attributes showed some mild or strong spatial dependence for the spatial dependence structure. But, in relation to the thematic maps drawing, there was no similarity between the obtained maps for all soil chemical attributes, with the use in the spatial prediction of initial and optimized sampling configurations.

Keywords: Fischer information matrix, genetic algorithm, sample configurations, simulation.

5.2.1. Introdução

A qualidade do solo é essencial ao desenvolvimento sustentável e à preservação dos ecossistemas e da biodiversidade (MANUEL *et al.*, 2017). Além da variabilidade natural que os solos apresentam, as práticas agrícolas de manejo e uso das terras são fatores que também provocam variabilidade (ARTUR *et al.*, 2014). O estudo e o conhecimento desta variabilidade espacial das propriedades do solo e das culturas são de extrema importância para o sistema de cultivo, para que se possa melhorar o manejo e a rentabilidade da produção agrícola (BERNARDI *et al.*, 2014).

Vale ressaltar que as informações sobre a variabilidade espacial de atributos químicos do solo servirão para otimizar as aplicações de insumos na lavoura em cada ponto, fazendo com que a produtividade aumente e os problemas ambientais diminuam (MACHADO *et al.*, 2007). As técnicas geoestatísticas são ferramentas que permitem realizar o estudo da variabilidade espacial dos atributos químicos e físicos do solo (CAVALCANTE *et al.*, 2011).

Desta forma, quando se conhece a variabilidade dos atributos químicos do solo e da cultura, pode-se analisar a variabilidade da produtividade, melhorar o manejo em áreas agrícolas e conduzir o gerenciamento agrícola de forma pontual (AMADO *et al.*, 2009). Neste tipo de experimento, é importante um plano de amostragem reduzido, que permita a redução dos custos operacionais e aumente a qualidade dos resultados obtidos (GUEDES *et al.*, 2014).

A escolha de uma configuração e um tamanho amostral adequado pode ser definida como um problema de otimização que envolve a discretização da área em estudo, ou seja, a partir de uma malha amostral inicial, escolhe-se uma configuração amostral que minimize as perdas quanto aos resultados das análises (GUEDES *et al.*, 2016).

Um desses processos de otimização é chamado de Algoritmo Genético (AG), o qual

é um ramo dos algoritmos evolucionários e como tal pode ser definido como uma técnica de busca que se baseia no processo da evolução natural das espécies (LINDEN, 2012; PESSOA *et al.*, 2015). Os processos de redimensionamento de configurações amostrais consideram como critério de busca uma função-objetivo do problema, a qual será minimizada ou maximizada e resume a eficiência da otimização.

Trabalhos com o objetivo de encontrar configurações amostrais otimizadas em variáveis que apresentam dependência espacial são recentes. A ideia de otimização é encontrada, por exemplo, em Guedes *et al.* (2014), que utilizaram o recozimento simulado para otimizar o tamanho e as configurações amostrais para análise da produtividade de soja, obtida pelo monitor de colheita, utilizando funções-objetivo baseadas na eficiência da predição espacial da variável georreferenciada em localizações não amostradas.

Ainda, pode-se levar em conta a eficiência quanto à estimação do modelo geoestatístico no processo de otimização. Desta forma, o objetivo do trabalho foi redimensionar uma configuração amostral, para uma área comercial de cultivo de soja onde se desenvolve uma atividade de aplicação localizada de insumos. O tamanho amostral foi reduzido a partir do algoritmo genético e foi minimizada a função-objetivo logaritmo do determinante da matriz inversa de Fischer. Para isto, avaliou-se ainda por meio de dados simulados, a influência dos parâmetros do modelo geoestatístico e da configuração amostral inicial utilizada no processo de otimização, em busca por um tamanho e uma configuração amostral otimizada.

5.2.2. Material e Métodos

5.2.2.1. Estudo de simulações

Os valores das variáveis foram obtidos por um experimento de Monte Carlo para cada conjunto de dados simulados, que representam realizações de processos estocásticos, com variáveis gaussianas, sem tendência direcional, e isotrópica.

Assim, a simulação de uma variável georreferenciada que apresenta dependência espacial demandou um processo estocástico $\{Z(\mathbf{s}_i), \mathbf{s}_i \in S\}$, com $\mathbf{s}_i = (x, y)^T$ como o vetor que representa determinada localização na área de estudo, em que $S \subset \mathcal{R}^2$, e \mathcal{R}^2 é o espaço euclidiano bidimensional (MARDIA; MARSHALL, 1984), em que $Z(\mathbf{s}_1), \dots, Z(\mathbf{s}_n)$ são observações da variável georreferenciada sob estudo nas \mathbf{s}_i ($i = 1, \dots, n$) localizações espaciais amostradas, que pode ser expressa pelo modelo linear espacial gaussiano (Equação 46) (URIBE-OPAZO; BORSSOI; GALEA, 2012):

$$Z(\mathbf{s}_i) = \mu(\mathbf{s}_i) + \epsilon(\mathbf{s}_i) \quad (46)$$

em que $\mu(\mathbf{s}_i)$ é o termo determinístico e $\epsilon(\mathbf{s}_i)$ é o termo estocástico; ambos dependem da localização espacial na qual $Z(\mathbf{s}_i)$ é observada. Assume-se que o erro estacionário $\epsilon(\mathbf{s}_i)$ tem

média zero, $E[\epsilon(\mathbf{s}_i)] = 0$, e a variação entre pontos no espaço é determinada por uma função covariância $C(\mathbf{s}_i, \mathbf{s}_j) = cov[\epsilon(\mathbf{s}_i), \epsilon(\mathbf{s}_j)]$, e $\mu(\mathbf{s}_i) = \mu\mathbf{1}$, em que μ é uma constante; $\mathbf{1}$ é um vetor de uns de dimensão $n \times 1$, quando o processo considerado for estacionário, e caso contrário, representa uma tendência direcional da variável georreferenciada (GRESSIE, 2015). Neste caso, para algumas funções conhecidas de \mathbf{s}_i como $x_1(\mathbf{s}_i), \dots, x_p(\mathbf{s}_i)$, a média do processo é dada por

$$\mu(\mathbf{s}_i) = \sum_{j=1}^p x_j(\mathbf{s}_i)\beta_j, \quad i = 1, \dots, n$$

em que β_1, \dots, β_p são parâmetros desconhecidos e precisam ser estimados.

Da mesma forma, na notação matricial, temos o modelo espacial linear dado pela Equação (47):

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (47)$$

em que, o vetor de erros aleatórios $\boldsymbol{\epsilon}$ tem $E(\boldsymbol{\epsilon}) = \mathbf{0}$ (vetor nulo) e a matriz de covariância $\Sigma[(\sigma_{ij})]$, com $\sigma_{ij} = C(\mathbf{s}_i, \mathbf{s}_j)$. Assume-se que Σ é não singular e \mathbf{X} é de posto completo; $\boldsymbol{\beta} = (\beta_1, \dots, \beta_p)^T$; e \mathbf{Z} segue uma distribuição gaussiana n -variada com vetor de médias $\mu\mathbf{1}$ e matriz de covariância Σ , isto é, $\mathbf{Z} \sim N_n(\mu\mathbf{1}, \Sigma)$ (DE BASTIANI *et al.*, 2015).

Uma forma paramétrica que especifica para a matriz de covariância (MARDIA; MASHALL, 1984; URIBE-OPAZO; BORSSOI; GALEA, 2012) é dada por:

$$\Sigma = \varphi_1 \mathbf{I}_n + \varphi_2 \mathbf{R}(\varphi_3),$$

em que φ_1 é o efeito pepita, sendo ideal que $\gamma(0) = 0$ (γ é a semivariância), porém, na prática, conforme h tende a zero, $\gamma(h)$ se aproxima de um valor positivo, conhecido como efeito pepita, o qual revela a descontinuidade do semivariograma para distâncias menores do que a menor distância entre amostras; \mathbf{I}_n é a matriz identidade $n \times n$; φ_2 é a contribuição, conhecida como a diferença entre o patamar e o efeito pepita (CÂMARA; MEDEIROS, 1998). Portanto, φ_2 é o segmento gráfico neste intervalo que, de fato, caracteriza a dependência espacial dos processos estocásticos contínuos e pode ser entendida como uma variância estruturada que é correspondente à estrutura do fenômeno em estudo (SEIDEL; OLIVEIRA, 2013); φ_3 é a função do alcance do modelo, e o alcance ($a = g(\varphi_3)$) é a distância limite de dependência espacial, ou seja, medições localizadas à distância maior que a são consideradas independentes entre si, enquanto amostras separadas por distâncias menores do que a são correlacionadas umas com as outras (VIEIRA, 2000; SEIDEL; OLIVEIRA, 2013); e $\mathbf{R}(\varphi_3)$ é uma matriz $n \times n$ que é função de φ_3 .

Sendo que $\mathbf{R} = \mathbf{R}(\varphi_3) = [r_{ij}]$ e r_{ij} depende de $h_{ij} = \|\mathbf{s}_i - \mathbf{s}_j\|$ (a distância Euclidiana entre \mathbf{s}_i e \mathbf{s}_j) é uma matriz simétrica $n \times n$, com elementos da diagonal $r_{ij} = 1$, para $i = j$ com $i, j = 1, \dots, n$; $r_{ij} = \varphi_2^{-1}C(\mathbf{s}_i, \mathbf{s}_j)$ para $\varphi_2 \neq 0$ e $i \neq j$; e $r_{ij} = 0$ para $\varphi_2 = 0$, $i \neq j$. Esta

forma paramétrica da matriz de covariância Σ , ocorre para vários processos isotrópicos (GUEDES *et al.*, 2008), onde a covariância $Cov[\epsilon(\mathbf{s}_i), \epsilon(\mathbf{s}_j)]$ é definida pela função de covariâncias $Cov[\epsilon(\mathbf{s}_i), \epsilon(\mathbf{s}_j)] = \varphi_2 r_{ij}$, para $i \neq j$ (De BASTIANI *et al.*, 2015).

As simulações foram realizadas em uma área quadrada com limite máximo para as coordenadas x e y igual a 1. Foram utilizadas três configurações amostrais, compostas por 100 pontos amostrais: sistemática 10×10 , aleatória e sistemática 9×9 com adição de 19 pontos próximos, posicionados dentro de uma circunferência de raio 0,10, cujo centro será sempre um ponto da amostragem regular, escolhido de forma aleatória (*lattice plus close pairs*, DIGGLE; RIBEIRO JR., 2007).

Foram considerados nove ensaios em cada configuração amostral e realizadas com simulações para cada ensaio. Em cada ensaio, as simulações foram realizadas considerando média igual a cinco (5), o modelo linear espacial gaussiano (Equação 46) e um modelo exponencial para a função semivariância, com parâmetro contribuição igual a um e todas as combinações dos seguintes valores para os seguintes parâmetros: alcance prático (0,45; 0,60; e 0,90) e efeito pepita (0; 0,5; e 0,8).

Para cada simulação de cada ensaio aplicou-se o AG, o qual foi compreendido pelas seguintes etapas:

Etapa 0: Com base em testes iniciais, em busca de melhores resultados no processo de otimização, foi pré-definido que o critério de parada seja igual a dois mil (2000) cruzamentos, e cada cruzamento foi realizado com um ponto de corte (LINDEN, 2012).

Etapa 1: Definiu-se uma população inicial de forma aleatória, com cem (100) indivíduos. Cada indivíduo foi composto por uma amostra inicial reduzida composta por vinte (20) pontos da malha original.

Etapa 2: Para cada indivíduo, calculou-se o valor da função-objetivo, definida pelo negativo do logaritmo de determinante da matriz de informação de Fischer, ou de forma equivalente, pelo logaritmo do determinante da matriz inversa de Fisher (ZHU; STEIN, 2005). As funções são descritas na Equação (48).

$$V_0(\theta) = -\log|\mathbf{I}_F(\theta)| = \log|(\mathbf{I}_F(\theta))^{-1}| \quad (48)$$

em que $\mathbf{I}_F(\theta)$ é a matriz de informação de Fisher, $\theta = (\beta^T, \varphi^T)^T$ (URIBE-OPAZO; BORSOI; GALEA, 2012). Esta matriz é dada para o modelo espacial linear gaussiano por (LANGE; LITTLE; TAYLOR; 1989; MITCHELL, 1989):

$$\mathbf{I}_F(\theta) = \mathbf{K}(\theta) = \begin{pmatrix} \mathbf{K}(\beta) & 0 \\ 0 & \mathbf{K}(\varphi) \end{pmatrix}.$$

Sendo uma matriz diagonal bloco, donde:

$$\mathbf{K}(\beta) = E[-L_{\beta\beta}] = E\left[-\frac{\partial^2 l(\theta)}{\partial \beta \partial \beta^T}\right] = \mathbf{X}^T \Sigma^{-1} \mathbf{X},$$

$$\mathbf{K}(\varphi) = E[-L_{\varphi\varphi}] = E\left[-\frac{\partial^2 l(\theta)}{\partial \varphi \partial \varphi^T}\right],$$

que tem como elementos:

$$k_{lm}(\varphi) = \frac{1}{2} \text{tr} \left(\Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_l} \Sigma^{-1} \frac{\partial \Sigma}{\partial \varphi_m} \right), \quad l, m = 1, 2, 3.$$

Detalhes sobre a matriz de informação podem ser encontrados em URIBE-OPAZO; BORSOI; GALEA (2012).

Assim, ao se utilizar ML para se estimar o parâmetro θ , pode-se utilizar o inverso da matriz de informação de Fisher para aproximar a matriz de covariância, que é uma medida natural do desempenho para a estimação de parâmetros (ZHU; STEIN, 2005).

A razão para isso é que o inverso desta matriz é, em condições de regularidade, a matriz de covariância assintótica do correspondente estimador de θ (ZIMMERMAN, 2006).

Segundo Muller (2007), quando o vetor de parâmetros β^T (vetor de médias) é conhecido e fixo, isto é, $\beta = \mu$ resta na matriz de informação de Fisher apenas o vetor de parâmetros φ^T para ser estimado. Porém, neste trabalho, não se considerou a média fixa.

Etapa 3: Foram selecionados 75% dos indivíduos da população para o processo de reprodução (cruzamento), os quais apresentaram os menores valores para função-objetivo.

Etapa 4: Em cada cruzamento, dois filhos (indivíduos obtidos pelo cruzamento) foram criados e, para cada filho, foi calculado o valor da função objetivo (Equação 48). Porém, apenas o melhor filho (com menor valor para a função objetivo) foi acrescentado na população do AG, ao mesmo tempo em que se eliminava desta população o indivíduo que havia apresentado o maior valor para a função-objetivo.

Etapa 5: Calculou-se o critério de parada do processo de otimização, conforme a Equação (49):

$$C = \min_m - \min_{m-1} \quad (49)$$

em que \min_m e \min_{m-1} são os mínimos obtidos com o tamanho amostral atual e o anterior respectivamente.

Se $|C| < 0,5$, então o processo termina. Caso contrário, o tamanho amostral de indivíduo da população foi aumentado em 5% do total da amostra inicial, ou seja, houve acréscimo de indivíduos novos nesta população, e voltou para a etapa 2. O processo foi repetido até que a amostra otimizada pelo algoritmo atingisse o critério de parada (Equação 49). Ao final deste processo, obteve-se a configuração amostral otimizada.

5.2.2.2. Estudo prático

O conjunto de dados utilizado nesta pesquisa refere-se a uma área comercial de produção de grãos com 167,35 ha, localizada na Fazenda Agassiz em Cascavel, PR. A área possui 102 pontos amostrais dispostos em uma configuração amostral denominada sistemática centrada com pares de pontos próximos (*lattice plus close pairs*), com distância mínima entre os pontos da grade regular de 141 metros e em alguns locais, escolhidos

aleatoriamente. A amostragem foi realizada com distâncias menores (75 metros e 50 metros entre pares de pontos) (Figura 25). As amostras foram localizadas e georreferenciadas com um aparelho receptor de sinal e o sistema de posicionamento global (GPS) GEOEXPLORE 3, em um sistema espacial de coordenadas UTM.

Estudaram-se os seguintes atributos químicos do solo: teor de Cálcio (Ca) [cmol dm^{-3}], de Carbono (C) [g dm^{-3}], de Cobre (Cu) [cmol dm^{-3}] e de Manganês (Mn) [cmol dm^{-3}]; e pH, observados no ano safra 2010/2011. As análises químicas do solo foram realizadas com amostras de aproximadamente 500 g, e encaminhadas ao Laboratório da Cooperativa Central de Pesquisas Agrícolas – COODETEC pelo Método Walkley-Black.

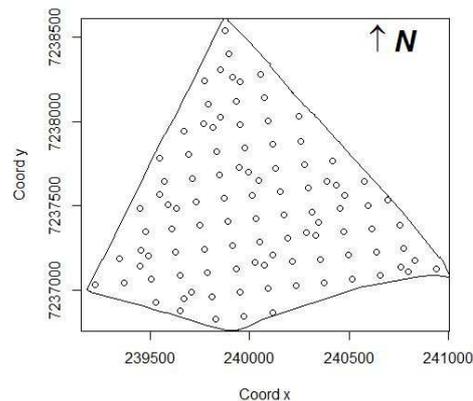


Figura 25 Mapa da área em estudo e da configuração amostral.

Cada atributo químico do solo foi submetido às análises descritivas e geoestatística. A existência de anisotropia foi avaliada pelo teste não paramétrico de isotropia de Maity e Sherman (2012) (MS), considerando-se 5% de significância. Na sequência, estimaram-se os seguintes modelos à função semivariância pelo método máxima verossimilhança (DIGGLE; RIBEIRO JR., 2007): exponencial, gaussiano e da família Matérn com $k=1$; 1,5 e 2 (CRESSIE, 2015). A escolha do melhor modelo foi feita pelo método da validação cruzada (FARACO *et al.*, 2008; ASSUMPÇÃO; HADLICH, 2017). E, por krigagem ordinária, realizou-se a predição espacial de cada atributo químico do solo em localizações não amostradas na área agrícola em estudo (Figura 25). Em seguida, confeccionaram-se os mapas temáticos dos valores estimados de cada atributo (CRESSIE, 2015).

O próximo passo foi a aplicação do AG em cada atributo químico do solo. O AG foi realizado da mesma forma e com os mesmos critérios aplicados nas simulações. Ao final do processo de otimização, uma configuração amostral de tamanho reduzido foi obtida para cada atributo químico do solo, na qual realizaram-se novamente as análises exploratórias e geoestatísticas.

As simulações, as rotinas computacionais do AG, as análises estatísticas e geoestatísticas foram realizadas no software R (R DEVELOPMENT CORE TEAM, 2017) utilizando os pacotes geoR (RIBEIRO JR. e DIGGLE, 2001) e sm (BOWMAN and AZZALINI, 2015).

5.2.3. Resultados

5.2.3.1. Estudos de simulação

Em todas as simulações realizadas para as três configurações amostrais, observou-se que os valores estimados dos mínimos e máximos para a função logaritmo do determinante da matriz inversa de Fischer, obtido após o processo de otimização, estão muito próximos (Tabela 5). Isso indica que o espaço de busca pela solução do processo está reduzido em poucos valores.

Ainda, em média, observou-se decréscimo relevante no valor estimado do logaritmo do determinante da matriz inversa de Fisher (54,5% a 109%), quando se compararam os valores estimados de mínimos obtidos no início e ao final do processo de otimização, indicando eficiência na minimização do valor da função objetivo (Tabela 5).

Considerando-se todas as simulações, as amostragens *lattice plus close pairs* e aleatória simples apresentaram respectivamente a menor e a maior variabilidade dos valores estimados mínimos obtidos ao final do processo de otimização para a função objetivo (Tabela 5).

Além disso, nas simulações realizadas com a amostragem *lattice plus close pairs*, foram observados os menores valores estimados da função-objetivo ao final do processo de otimização quando comparada com as demais amostragens utilizadas nas simulações (Tabela 5). Logo, tal amostragem inicial permitiu a determinação de uma configuração amostral de tamanho reduzido com uma maior minimização da função objetivo pelo processo de otimização.

Considerando-se todas as simulações realizadas, independente da amostragem inicial no processo de otimização, observaram-se os menores valores estimados da função objetivo obtidos ao final do processo de otimização nas simulações realizadas com efeito pepita igual a 0 e alcance prático igual a 0,9 (Tabela 5).

Considerando-se todas as configurações amostrais iniciais, as melhores configurações amostrais de tamanho reduzido, obtidas pelo processo de otimização, tiveram em média 32 a 40 pontos. Houve, portanto, uma redução de 60% a 68% do número de pontos amostrais, em relação à malha inicial (Tabela 5). O menor tamanho amostral foi obtido quando foi considerada a amostragem aleatória simples como malha amostral inicial nas simulações, e com o menor valor para o efeito pepita (igual a 0) e o maior alcance prático (igual a 0,9). Já o maior tamanho amostral foi obtido quando foi considerada a amostragem *lattice plus close pairs* como malha amostral inicial nas simulações, e o segundo maior valor tanto para o efeito pepita (0,50) quanto para o alcance prático (0,60).

Tabela 5 Valores médios e desvio padrão (entre parênteses) do mínimo e máximo do valor estimado do logaritmo do determinante da matriz inversa de Fisher, calculado na amostra de tamanho reduzido, obtidos no início e no final do processo de otimização (o valor em negrito representa o valor mínimo obtido ao final do processo de otimização), de acordo com a amostragem inicial e os valores dos parâmetros efeito pepita e alcance, considerados em todas as simulações

Amostragem inicial/Efeito pepita (ϕ_1)		Alcance (a)						
		$a = 0,45$		$a = 0,60$		$a = 0,90$		
		Mín.	Max.	Mín.	Max.	Mín.	Max.	
Irregular	Inicial	-8,21	-4,85	-8,47	-5,16	-18,85	-12,44	
		(9,31)	(5,91)	(9,57)	(6,23)	(1,78)	(1,22)	
	$\phi_1 = 0$	Final	-13,01	-12,88	-13,18	-13,07	-29,12	-28,87
			(14,14)	(14,01)	(14,33)	(14,22)	(2,34)	(2,33)
	Decréscimo (%)	58,5%		55,5%		54,5%		
	T. amostral	37,30		37,60		31,90		
$\phi_1 = 0,5$	Inicial	-8,21	-4,87	-14,63	-7,92	-7,01	-3,71	
		(9,30)	(5,925)	(1,56)	(0,84)	(8,11)	(4,77)	
	Final	-13,01	-12,84	-23,94	-23,67	-11,66	-11,49	
		(14,14)	(14,01)	(2,23)	(2,20)	(12,79)	(12,62)	
	Decréscimo (%)	58,5%		63,5%		65,5%		
	T. amostral	37,30		37,40		36,75		
$\phi_1 = 0,8$	Inicial	-6,17	-2,73	-6,27	-2,84	-6,45	-3,03	
		(7,26)	(3,78)	(7,36)	(3,89)	(7,56)	(4,09)	
	Final	-10,97	-10,81	-11,15	-10,97	-11,19	-11,03	
		(12,13)	(11,96)	(12,35)	(12,15)	(12,35)	(12,17)	
	Decréscimo (%)	77,5%		78%		73,5%		
	T. amostral	36,90		37,20		36,80		
<i>Lattice plus close pairs</i>	Inicial	-15,04	-9,21	-15,59	-9,77	-16,45	-11,04	
		(1,61)	(1,06)	(1,54)	(1,13)	(1,63)	(1,66)	
	$\phi_1 = 0$	Final	-25,26	-24,97	-26,00	-25,75	-26,82	-26,59
		(2,11)	(2,00)	(2,11)	(2,05)	(2,36)	(2,25)	
	Decréscimo (%)	68%		66,5%		63,5%		
	T. amostral	36,50		36,35		36,10		
$\phi_1 = 0,5$	Inicial	-12,58	-6,62	-12,68	-6,84	-12,99	-7,48	
		(1,44)	(0,96)	(1,35)	(0,97)	(1,39)	(1,12)	
	Final	-23,12	-22,80	-22,74	-22,47	-23,22	-22,89	
		(2,32)	(2,23)	(2,24)	(2,18)	(2,28)	(2,26)	
	Decréscimo (%)	83,5%		79,5%		78,5%		
	T. amostral	36,10		39,50		39,20		
$\phi_1 = 0,8$	Inicial	-11,59	-5,57	-11,67	-5,70	-11,94	-6,22	
		(1,36)	(1,07)	(1,29)	(0,98)	(1,28)	(1,11)	
	Final	-21,95	-21,64	-22,08	-21,80	-22,19	-21,93	
		(2,20)	(2,13)	(2,36)	(2,28)	(2,13)	(2,04)	
	Decréscimo (%)	89,5%		89,5%		85,5%		
	T. amostral	37,50		36,60		36,10		
Regular	Inicial	-13,61	-8,54	-14,02	-9,11	-14,98	-10,16	
		(1,44)	(0,94)	(1,47)	(0,99)	(1,55)	(1,11)	
	$\phi_1 = 0$	Final	-24,02	-23,72	-24,49	-24,21	-25,17	-24,93
		(1,98)	(1,93)	(2,27)	(2,25)	(2,40)	(2,31)	
	Decréscimo (%)	76,5%		74,5%		68%		
	T. amostral	37,90		37,20		37,10		
$\phi_1 = 0,5$	Inicial	-5,14	-2,51	-5,29	-2,66	-5,48	-2,94	
		(6,24)	(3,58)	(6,39)	(3,73)	(6,58)	(4,02)	
	Final	-10,53	-10,37	-10,46	-10,30	-10,55	-10,41	
		(11,65)	(11,49)	(11,59)	(11,44)	(11,68)	(11,53)	
	Decréscimo (%)	100%		97%		92%		
	T. amostral	35,90		38,50		37,70		
$\phi_1 = 0,8$	Inicial	-10,38	-4,99	-4,79	-2,10	-10,95	-10,16	
		(1,47)	(0,92)	(5,89)	(3,18)	(1,41)	(1,11)	
	Final	-20,83	-20,49	-10,01	-9,87	-21,30	-21,02	
		(2,24)	(2,19)	(11,15)	(11,01)	(2,22)	(2,14)	
	Decréscimo (%)	100%		109%		94,5		
	T. amostral	37,50		37,10		36,90		

a é o valor do alcance simulado; ϕ_1 é o valor do efeito pepita simulado; Min. é o valor mínimo do logaritmo do determinante da matriz inversa de Fischer; Max. é o valor máximo do logaritmo do determinante da matriz inversa de Fischer; % de decréscimo do valor máximo inicial do logaritmo do determinante da matriz inversa de Fischer em relação ao valor máximo final; T.: tamanho.

O aumento do alcance prático nas simulações, para cada valor do efeito pepita considerado nas simulações, gerou uma leve redução (em média 1 ponto) do número de pontos amostrais otimizados para a maioria dos processos de otimização. Porém, apenas para a amostragem aleatória simples com efeito pepita 0, houve maior variação da quantidade de pontos amostrais obtidos ao final de processo de otimização, igual a 5 pontos amostrais. Apesar de Guedes *et al.* (2011) terem trabalhado com o tamanho amostral previamente fixado, a fim de otimizar apenas a configuração amostral inicial, os resultados encontrados pelos pesquisadores foram semelhantes aos desta pesquisa.

A disposição dos pontos da malha amostral otimizada foi comparada com cada configuração amostral, mas não se identificou qualquer padrão quanto à disposição dos pontos escolhidos, para todos os casos simulados. Isso pode ser observado em um exemplo de simulação, tal como descrito nas Figuras 22, 23 e 24 (APÊNDICE), nas quais é possível visualizar a localização dos cem pontos amostrais iniciais e dos pontos selecionados no processo de otimização.

5.2.3.2. Otimização do tamanho e da configuração amostral considerando os atributos químicos do solo

A área comercial era composta inicialmente por 102 pontos amostrais. Após a aplicação do algoritmo genético para cada atributo químico do solo, tem-se que a configuração amostral otimizada e reduzida variou de 30 a 40 pontos (Tabela 6). A maior quantidade de pontos na amostra reduzida foi encontrada para os teores de cálcio (Ca), cobre (Cu) e de manganês (Mn), e a menor quantidade para o teor de carbono (C). Tal redução corresponde respectivamente a 29,41% e 39,22% dos pontos da malha amostral inicial. Observa-se que o processo de otimização ora proposto, que simultaneamente otimizou a configuração e o tamanho amostral, determinou tamanhos amostrais menores, quando comparados às malhas amostrais otimizadas por Guedes *et al.* (2011), que fixaram o tamanho amostral reduzido em 50% da malha inicial, usando um algoritmo genético híbrido. Considerou-se também a média da variância da predição espacial como função-objetivo, associada à eficiência da predição espacial da variável georreferenciada em localizações não amostradas.

Em relação ao tempo computacional do algoritmo genético, tem-se que para se otimizar cada tamanho e configuração amostral para cada atributo químico do solo, foram necessários quinze minutos. A máquina em estudo possui Processador Intel(R) Core (TM) i7-4790 CPU @ 3.60GHz, 3601 Mhz, 4 Núcleos, 8 Processadores Lógicos e memória física (RAM) instalada 4,00 GB.

Os resultados da análise exploratória dos atributos químicos do solo, assim como o valor do coeficiente de correlação linear de Pearson dos atributos químicos do solo com as

coordenadas dos eixos X e Y, considerando as configurações amostrais inicial e otimizada, também são evidenciados na Tabela 6.

Tabela 6 Estatística descritiva dos seguintes atributos químicos do solo Ca [cmolc dm⁻³], C [g dm⁻³], Cu [mg dm⁻³], Mn [cmolc dm⁻³] e pH, e valor do coeficiente de correlação linear de Pearson dos atributos químicos do solo com as coordenadas dos eixos X e Y, considerando a configuração amostral original e a configuração amostral de tamanho reduzido.

Atributos	Configuração	Ca	C	Cu	Mn	pH
Média	Original	5,20	26,93	2,95	49,32	5,10
Mínimo		2,37	19,87	1,10	17,00	4,40
Q1		4,21	24,55	2,30	36,25	4,80
Mediana		5,08	26,88	2,80	43,00	5,10
Q3		6,10	29,61	3,68	59,50	5,30
Máximo		11,76	34,29	4,90	107,00	6,70
Variância		1,88	10,55	0,68	379,51	0,15
C.V (%)		26,41	12,06	27,86	39,50	7,58
Coef. X		0,14	-0,08	-0,08	-0,08	0,14
Coef. Y		0,40	0,36	-0,56	0,59	0,35
<i>n</i>		40	30	40	40	35
Média	Otimizada com tamanho reduzido	5,22	27,11	2,75	44,60	5,12
Mínimo		3,80	24,55	1,80	21,00	4,60
Q1		4,59	26,10	2,30	36,00	4,80
Mediana		5,03	26,88	2,75	42,00	5,20
Q3		5,90	28,05	3,53	53,25	5,35
Máximo		7,52	30,39	4,30	82,00	5,60
Variância		0,74	2,84	0,50	202,50	0,08
C.V (%)		16,48	6,20	24,60	31,91	5,44
Coef. X		-0,10	0,20	-0,06	-0,07	0,26
Coef. Y		0,86	0,85	-0,75	0,65	0,63

Q1 é o primeiro quartil, Q3 é o terceiro quartil, C.V é o coeficiente de variação, coef.: coeficiente de correlação linear de Pearson entre os atributos químicos do solo e as coordenadas X e Y e *n* é o número de pontos obtidos no processo de otimização.

De acordo com a Tabela 6, nota-se que, mesmo com a redução amostral, as estatísticas descritivas dos atributos químicos do solo, obtidos para a configuração amostral reduzida e otimizada pelo AG, foram semelhantes aos resultados da configuração amostral original. Resultados semelhantes foram obtidos em Guedes *et al.* (2014), mesmo trabalhando com um algoritmo diferente (recozimento simulado) e uma função a ser otimizada com objetivos diferentes (otimizar a eficiência da predição espacial).

No entanto, observa-se ainda que, em geral, nas configurações amostrais otimizadas, foram eliminados os valores máximos e mínimos dos atributos químicos do solo, observados na configuração amostral inicial (Tabela 6).

Nota-se que, para cada atributo químico do solo, a presença de tendência na direção Y foi intensificada, pois se obteve um valor coeficiente de correlação linear de Pearson maior na configuração amostral otimizada (Tabela 6).

A partir do teste não paramétrico de isotropia de Maity e Sherman (2012) (MS), aplicado para cada atributo químico do solo, não se rejeitou a hipótese nula de isotropia (p -valor > 0,05), logo, considerou-se que a estrutura de dependência espacial de todos os atributos químicos do solo é isotrópica.

Para ambas as configurações amostrais, o melhor modelo da função semivariância, obtido pelo método da máxima verossimilhança e considerando a validação cruzada, foi: o modelo da família Matérn com $k=2$ para os teores de cálcio (Ca) e de carbono (C) e o modelo exponencial para os teores de cobre (Cu) e de manganês (Mn) e o atributo químico do solo pH (Tabela 7).

Tabela 7 Valores estimados dos parâmetros dos melhores modelos geoestatísticos ajustados, com seus respectivos desvios-padrões estimados, e valor da função-objetivo obtida pelo AG, para os atributos químicos do solo Ca [cmolc dm⁻³], C [g dm⁻³], Cu [mg dm⁻³], Mn [cmolc dm⁻³] e pH, considerando a configuração amostral original e a configuração amostral de tamanho reduzido.

Atributo s	Configuração	Modelos	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\varphi}_1$	$\hat{\varphi}_2$	\hat{a}	\overline{EPR} (%)	$\widehat{V}_0(\theta)$
Ca	Original	Matérn k=2	-9916,31 (3,0*10 ⁻⁹)	1,37 (2,1*10 ⁻⁵)	1,14 (0,11)	0,43 (0,11)	0,25 (0,02)	72,43	-30,03
C		Matérn k=2	-2,1*10 ⁴ (1,1*10 ⁻⁸)	3,00 (7,7*10 ⁻⁵)	4,50 (1,0*10 ⁻³)	9,08 (3,9*10 ⁻⁵)	0,38 (0,02)	49,53	5,99
Cu	Exponencial	Exponencial	6283,82 (4,1*10 ⁻⁹)	-0,87 (2,9*10 ⁻⁵)	0,09 (0,06)	0,45 (0,13)	0,81 (0,15)	20,71	-20,04
Mn		Exponencial	-1,7*10 ⁵ (5,9*10 ⁻⁸)	23,74 (0,001)	95,25 (8,8*10 ⁻⁵)	241,35 (2,8*10 ⁻⁵)	0,45 (0,06)	39,47	16,13
Ph	Exponencial	Exponencial	-2455,07 (8,7*10 ⁻¹⁰)	0,34 (6,3*10 ⁻⁶)	0,10 (0,03)	0,13 (0,03)	0,30 (0,02)	73,75	-29,37
Ca		Otimizada com tamanho reduzido	Matérn k=2	-1,1*10 ⁴ (1,7*10 ⁻⁹)	1,52 (1,2*10 ⁻⁵)	0,00 (0,02)	1,95 (0,02)	0,22 (0,01)	0,00
C	Matérn k=2		-2,0*10 ⁴ (4,8*10 ⁻⁹)	2,73 (3,4*10 ⁻⁵)	0,51 (0,11)	0,78 (0,11)	0,75 (0,11)	39,53	-3,08
Cu	Exponencial	Exponencial	6694,32 (4,7*10 ⁻⁹)	-0,92 (3,4*10 ⁻⁵)	0,00 (0,03)	0,24 (0,10)	1,50 (0,02)	0,00	-36,76
Mn		Exponencial	-1,6*10 ⁵ (8,2*10 ⁻⁸)	21,99 (0,001)	0,00 (1,0*10 ⁻³)	133,50 (1,0*10 ⁻³)	0,81 (0,08)	0,00	-1,49
Ph	Exponencial	Exponencial	-2951,69 (9,7*10 ⁻¹⁰)	0,41 (7,1*10 ⁻⁶)	0,04 (0,01)	0,05 (0,01)	0,96 (1,0*10 ⁻³)	44,44	-16,96

$\hat{\beta}_0$, $\hat{\beta}_1$: são os valores estimados dos parâmetros do modelo de regressão, que explicam a média; $\hat{\varphi}_1$: efeito pepita estimado, $\hat{\varphi}_1 + \hat{\varphi}_2$: patamar estimado, \hat{a} : alcance prático estimado (m), EPR: efeito pepita relativo estimado ($\overline{EPR} = \hat{\varphi}_1 / (\hat{\varphi}_1 + \hat{\varphi}_2)$) (%), k: parâmetro de ordem (CRESSIE, 2015), e f. obj: função objetivo.

Observou-se, pela Tabela 7, que todos os atributos químicos do solo, com ambas as configurações amostrais, apresentaram dependência espacial, quando avaliados o EPR (CAMBARDELLA *et al.*, 1994) e o alcance prático. Além disso, observou-se que a maioria dos atributos químicos do solo apresentaram diferenças nos valores estimados dos parâmetros do modelo geoestatístico, principalmente para o alcance prático e efeito pepita, quando se comparou a estimação realizada usando as configurações amostrais inicial e a otimizada.

Houve redução do valor estimado do efeito pepita para a maioria dos atributos químicos do solo e aumento do raio de dependência espacial quando se utilizou a configuração amostral otimizada reduzida na estimação no modelo geoestatístico. Pois, para os atributos químicos Ca, C, Cu, Mn e pH, considerando-se a configuração amostral original, observou-se um raio de dependência espacial (valor estimado do alcance) com variação entre 0,25km e 0,88km (Tabela 7). E, ao se considerar a configuração amostral reduzida, o valor estimado do alcance variou de 0,22km a 1,50km.

Os valores estimados dos desvios padrões dos parâmetros foram calculados para ambas as configurações amostrais. Observa-se que, de modo geral, a configuração amostral otimizada apresenta melhor qualidade na estimação dos parâmetros do modelo, pois apresentaram menores valores para os desvios padrões dos parâmetros, comparando-se ambas configurações amostrais.

Considerando-se a configuração amostral otimizada para todos os atributos químicos do solo, observou-se redução dos valores estimados da função-objetivo (de 34% a 151%), quando comparados com os valores obtidos com amostra original.

Os mapas construídos para todos os atributos químicos do solo, considerando-se a configuração amostral otimizada (Figura 26), não apresentaram semelhanças visuais, quando comparados aos mesmos mapas construídos considerando-se a configuração inicial (Figura 26). No entanto, para ambos os mapas (Figura 26), observou-se que os maiores valores de Ca, C e Mn estão concentrados na região norte da área de estudo. O atributo químico do solo Cu possui concentração de seus maiores valores na região sul; e o atributo pH apresentou valores mais espalhados pela área de estudo.

Quando se comparou a disposição dos pontos (Figura 26), verificou-se que o algoritmo de otimização selecionou os pontos amostrais da configuração otimizada, a fim de que fossem selecionados os pontos de sub-regiões heterogêneas da área em estudo. Assim, de forma geral, ocorreu um espalhamento dos pontos escolhidos em toda a área, principalmente para os teores de carbono (C) e manganês (Mn) e do atributo pH, os quais apresentaram maior espalhamento da configuração otimizada.

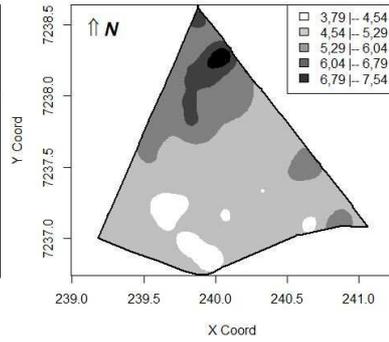
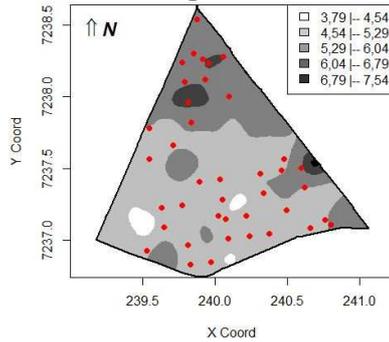
Quando comparados os mapas construídos de cada atributo, considerando a malha inicial e a otimizada, pelas medidas de similaridade, exatidão global e índices de concordância Kappa e Tau (Figura 26), observou-se que todos os atributos químicos apresentaram valores de exatidão global (EG) abaixo de 0,85 e valores inferiores a 0,80 para os índices de concordância Kappa e Tau. Há, portanto, indícios de dissimilaridade entre os mapas analisados (ANDERSON *et al.*, 2001; KRIPENDORFF, 2013).

Atributo

Elaborados com:
Malha original

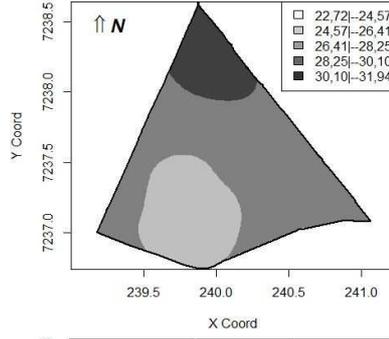
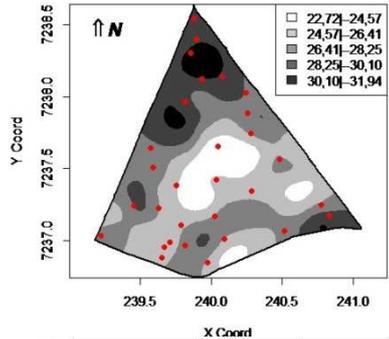
Malha otimizada

Ca



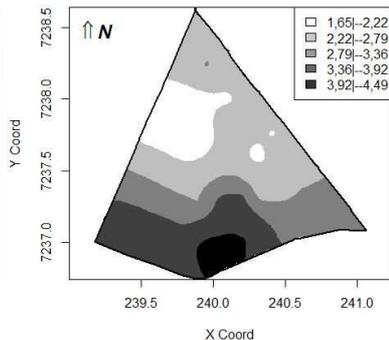
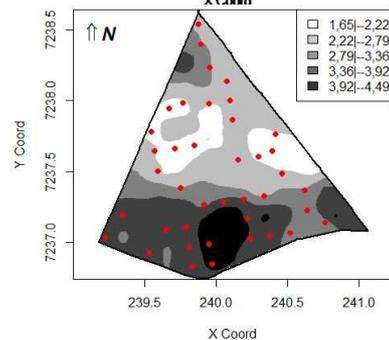
EG= 0,67
Tau= 0,39
Kappa= 0,59

C



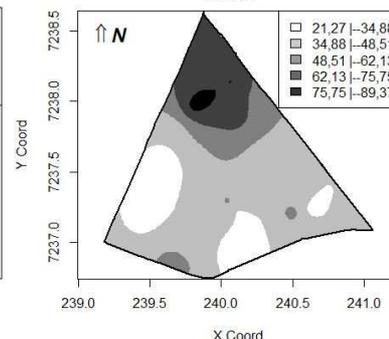
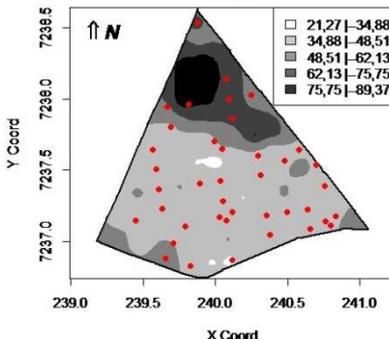
EG= 0,43
Tau= 0,20
Kappa= 0,29

Cu



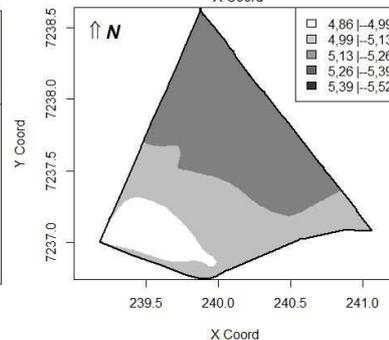
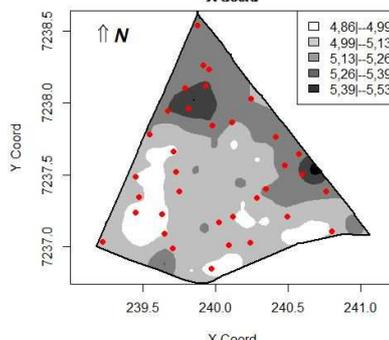
EG= 0,63
Tau= 0,51
Kappa= 0,54

Mn



EG= 0,61
Tau= 0,37
Kappa= 0,51

pH



EG= 0,59
Tau= 0,35
Kappa= 0,48

Figura 26 Mapas temáticos dos atributos químicos do solo elaborados considerando a malha amostral original e otimizada, em que ● representa a localização dos pontos selecionados no processo de otimização e valores estimados das medidas de similaridade EG, Tau e Kappa.

5.2.4. Conclusão

O processo de otimização foi eficiente e estável quando aplicado nos dados simulados e dados reais.

Conclui-se que foi possível: redimensionar a grade amostral que envolve o experimento e reduzir seu tamanho amostral, por AG, a partir da minimização da função-objetivo.

A nova configuração amostral otimizada para todos os atributos químicos do solo variou de 30 a 40 pontos, o que corresponde respectivamente de 29,41% a 39,22% da malha original. Quanto à estimativa da estrutura de dependência espacial, todos os atributos químicos do solo, com ambas as configurações amostrais, apresentaram moderada ou forte dependência espacial, quando avaliados o EPR e o alcance prático.

Houve diferenças relevantes entre os mapas temáticos elaborados para todos os atributos químicos do solo, considerando-se a configuração de pontos amostrais da malha original e de tamanho reduzido.

Quanto à disposição dos pontos da configuração amostral otimizada, verificou-se que o AG selecionou os pontos a fim de que fossem selecionados os pontos de sub-regiões heterogêneas da área em estudo.

A partir dos dados simulados, observou-se que o processo de otimização se mostrou mais eficiente quanto à minimização da função-objetivo quando se considerou a amostragem inicial *lattice plus close pairs*, quando comparada às demais amostragens. Além disso, em geral, as amostragens otimizadas e reduzidas apresentaram os menores valores da função-objetivo, com o menor valor simulado de efeito pepita e o maior valor simulado de alcance.

Agradecimentos

Os autores agradecem à CAPES pelo apoio financeiro e aos laboratórios de estatística aplicada (LEA) e estatística espacial (LEE) da UNIOESTE.

5.2.6. Referências Bibliográficas

ANDERSON, J. F.; HARDY, E. E.; ROACH J. T.; WITMER R. E. **A land use and land cover classification system for use with remote sensor data**, U.S. Washington, DC: U.S. Geologic Survey, 2001. 41p.

AMADO, T. J. C.; PES, L. Z.; LEMAINSKI, C. L.; SCHENATO, R. B. Atributos químicos e físicos de Latossolos e sua relação com os rendimentos de milho e feijão irrigados. **Revista Brasileira de Ciência do solo**, v.33, n.4, p.831-843, 2009.

ARTUR, A. G.; OLIVEIRA, D. P.; COSTA, M. C. G.; ROMERO, R. E.; SILVA, M. V. C.; FERREIRA, T. O. Variabilidade espacial dos atributos químicos do solo, associada ao

microrrelevo. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.18, n.2, p.141-149, 2014.

ASSUMPÇÃO, H. C. P.; HADLICH, G. M. Estatística descritiva e estacionaridade em variáveis geoquímicas ambientais. **Engenharia Sanitária e Ambiental**, v.22, n.4, p. 671-677, 2017.

BERNARDI, A. C. C.; NAIDE, J. M.; RESENDE, A. V.; BOSSOI, L. H.; INAMASU, R. Y. **Agricultura de precisão: Resultados de um novo olhar**. EMBRAPA, Brasília- DF, 600p, 2014.

BOWMAN A. W.; AZZALINI A. **R packagesm: non parametric smoothing methods** (version 2.2-5.4). University of Glasgow, UK and Università di Padova. Itália. Accessed on Dec. 12, 2015.

CÂMARA, G.; MEDEIROS, J. S. GIS para meio ambiente. In: Sistema de informações geográficas, 1998, Curitiba. **Anais...** Curitiba: Sagres, p. 140-180, 1998.

CAMBARDELLA, C.A.; MOORMAN, T.B.; PARKIN, T.B.; NOVACK, J.M.; KARLEN, D.L.; TURCO, R.F.; KNOPKA, A.E. Field-scale variability of soil properties in central Iowa soils. **Soil Science Society American Journal**, Medison, v.58, n.5, p. 1501-1511, 1994.

CAVALCANTE, E. G. S.; ALVES, M. C.; SOUZA, Z. M.; PEREIRA, G. T. Variabilidade espacial de atributos químicos do solo sob diferentes usos e manejos. **Revista Brasileira de Engenharia Agrícola e Ambiental**, v.15, n.3, p.237-243, 2011.

CRESSIE, N. A. C. **Statistics for Spatial Data** (Rev. Ed.). John Wiley & Sons, 2015, 928 p.

DE BASTIANI, F.; CYSNEIROS, A. H. M. A.; URIBE-OPAZO, M. A.; GALEA, M. Influence diagnostics in elliptical spatial linear models. **Sociedad de Estadística e Investigación Operativa**, TEST, Berlin, v.24, n.2, p. 322-340, 2015.

DIGGLE, P.J.; RIBEIRO JR., P.J. **Model-based geostatistics**. New York: Springer, 2007. 230 p.

FARACO, M. A.; URIBE-OPAZO, M. A. SILVA, E. A.; JOHANN, J. A.; BORSSOI, J. A. Seleção de modelos de variabilidade espacial para elaboração de mapas temáticos de atributos físicos do solo e produtividade da soja. **Revista Brasileira de Ciência do Solo**, Viçosa, v. 32, n. 2, p. 463-476, 2008.

GUEDES, L. P. C.; URIBE-OPAZO, M. A.; JOHANN, J. A.; SOUZA, E. G. Anisotropia no estudo da variabilidade espacial de algumas variáveis químicas do solo. **Revista Brasileira de Ciência do Solo**, Viçosa, v.32, n.6, p. 2217-2226, Nov./dec. 2008.

GUEDES, L.P.C.; RIBEIRO JR. P.J.; PIEDADE, S.M.D.S.; URIBE-OPAZO, M.A. Optimization of spatial sample configurations using hybrid genetic algorithm and simulated annealing. **Chilean Journal of Statistics**, Chile, v. 2, n.2, p. 39-50, set. 2011.

GUEDES, L. P. C.; URIBE-OPAZO, M. A., RIBEIRO JR. P. J. Optimization of sample design size and shapes for regionalized variables using simulated annealing. **Ciência e Investigación Agraria**, v. 41, n.1, p. 33-48, jan./abr. 2014.

GUEDES, L.P.C.; RIBEIRO JR. P.J.; URIBE-OPAZO, M.A.; BASTIANI, F. Soybean yield maps using regular and optimized sample with different configurations by simulated annealing. **Engenharia Agrícola**, v.36, n.1, p.114-125, jan./fev. 2016. ISSN: 1809-4430

KRIPPENDORFF, K. **Content Analysis an Introduction to its Methodology**. California, Sage Publications Ltd, 3rded., 2013.441 p.

LANGE, K. L.; LITTLE, R. J. A.; TAYLOR, J. M. G. Robust statistical modeling using the t distribution. **Journal of the American Statistics Association**, v.84, n.408, p. 881-896.dec. 1989.

LINDEN, R. **Algoritmos Genéticos**. Ciência Moderna Ltda, Rio de Janeiro, 2012. 496 p.

MACHADO, L. O.; LANA, A. M. Q.; LANA, R. M. Q.; GUIMARÃES, E. C.; FERREIRA, C. V. Variabilidade espacial de atributos químicos do solo em áreas sob sistema plantio convencional. **Revista Brasileira de Ciência do Solo**, v.31, n.3, p.591-599, Viçosa, 2007.

MAITY A, SHERMAN M. Testing for spatial isotrop yunder general designs. **Journal of Statistical Planning and Inference**, v.142, n.5, p.1081-1091, 2012.

MANUEL, L.; SCALON, J. D.; GONZALEZ, G. G. H.; MOREIRA, J. M.; RIBEIRO, M. C.; MELO, A. F. Estudo da variabilidade espacial da concentração de cobre e cádmio ao longo da margem do rio Meuse. **Pubvet – Medicina veterinária e zootecnia**, v.11, n.8, p.802-807, ago., 2017.

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.

MITCHELL, A. F. S. The information matrix, skewness tensor and α -connections for the general multivariate elliptic distribution. **Annals of the Institute of Statistical Mathematics**.Tokyo– JP, v.41, n.2, p. 289-304, 1989.

MULLER, W. G. **Collecting Spatial Data**. Springer-Verlag Berlin Heidelberg.3rded., 2007.

PESSOA, A. L. S.; ULISSES, P. H. C.; BRANCO, H. M. G. C.; RABÊLO, R. A. L. Uma aplicação de algoritmos genéticos simples e compacto para estimação de componentes harmônicas. **Revista Brasileira de Computação Aplicada**, Passo Fundo, v.7, n.2, p. 77-91, mai. 2015.

R DEVELOPMENT CORE TEAM. R: A language and environment for statistical computing. **R Foundation for Statistical Computing**, Vienna, Austria. ISBN 3-900051-07-0.Disponívelem: <<http://www.R-project.org>>Acessoem: 14 out. 2017.

RIBEIRO JR., P. J.; DIGGLE, P. J.geoR: a package for geostatistical analysis. **R-NEWS**, v. 1, n. 2, ISSN 1609-3631, 2001.

SEIDEL, E. J.; OLIVEIRA, M. S. Proposta de uma generalização para os modelos de semivariogramas Exponencial e Gaussiano. **Semina: ciências Exatas e Tecnológicas**. Londrina – PR, v.34, n.1, p.125-132, jan./jul. 2013.

URIBE-OPAZO, M. A.; BORSSOI, J. A.; GALEA, M. Influence diagnostics in Gaussian spatial linear models. **Journal of Applied Statistics**, v. 39, n. 3, p. 615-630, 2012.

VIEIRA, S. R. Tópicos em ciência do solo. **Sociedade Brasileira de Ciência do Solo**, v.1. Viçosa – MG, 2000. 53 p.

ZHU, Z.; STEIN, M. L. Spatial sampling for design for parameter estimation of the covariance function. **Journal of Statistical Planning and Inference**, v.134, p.583-603, 2005.

ZIMMERMAN, D. L. Optimal network design for spatial prediction, covariance parameter estimation, and empirical prediction. John Wiley& Sons, Ltd. **Environmetrics**, v.17, p.635-652, 2006.

APÊNDICE

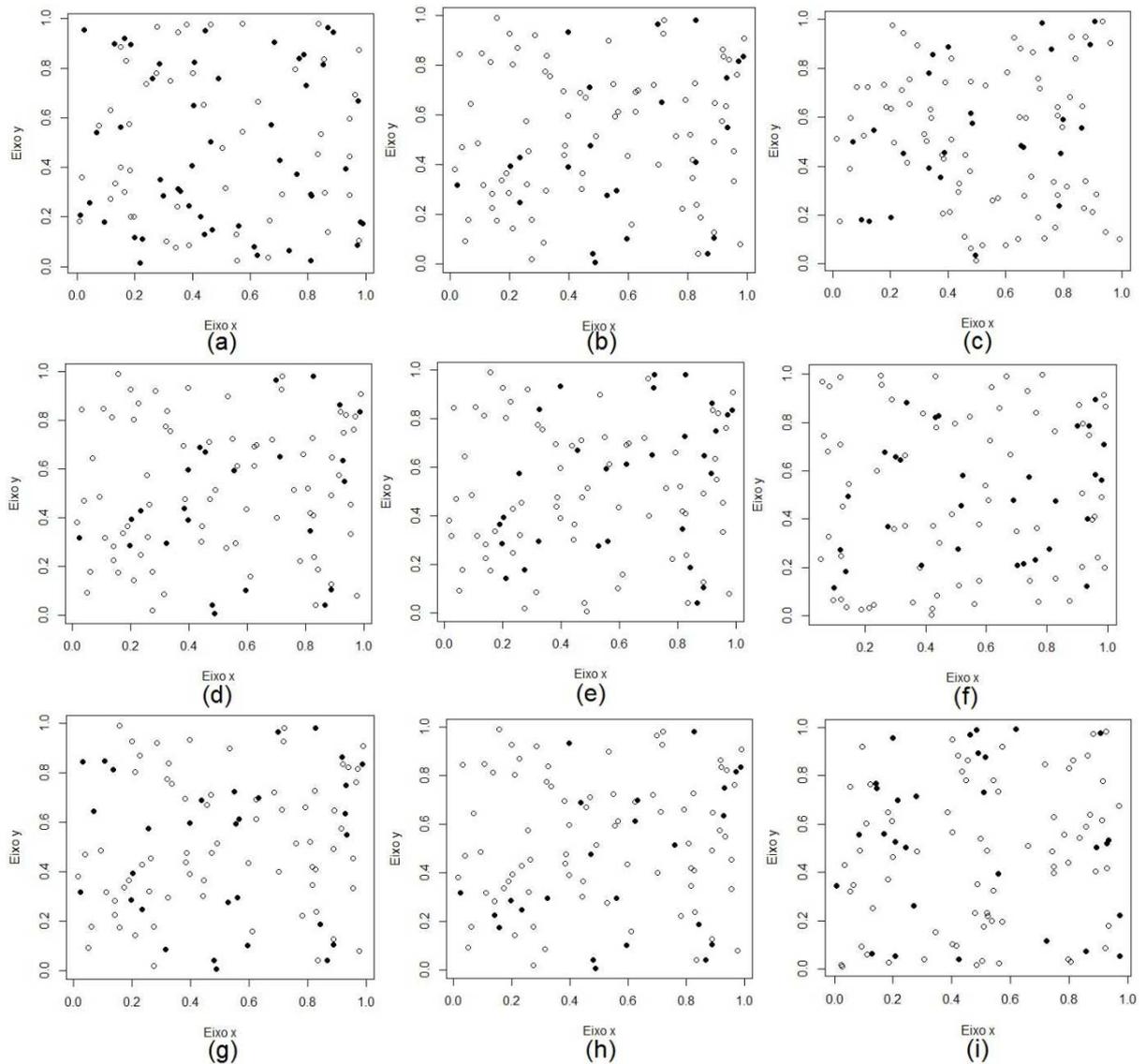


Figura 27 Gráfico da localização dos 100 pontos iniciais (\circ) e a localização dos pontos selecionados (\bullet), para um exemplo de cada simulação da amostragem aleatória simples, (a) alcance 0,45 e efeito pepita 0; (b) alcance 0,60 e efeito pepita 0; (c) alcance 0,90 e efeito pepita 0; (d) alcance 0,45 e efeito pepita 0,50; (e) alcance 0,60 e efeito pepita 0,50; (f) alcance 0,90 e efeito pepita 0,50; (g) alcance 0,45 e efeito pepita 0,80; (h) alcance 0,60 e efeito pepita 0,80; (i) alcance 0,90 e efeito pepita 0,80.

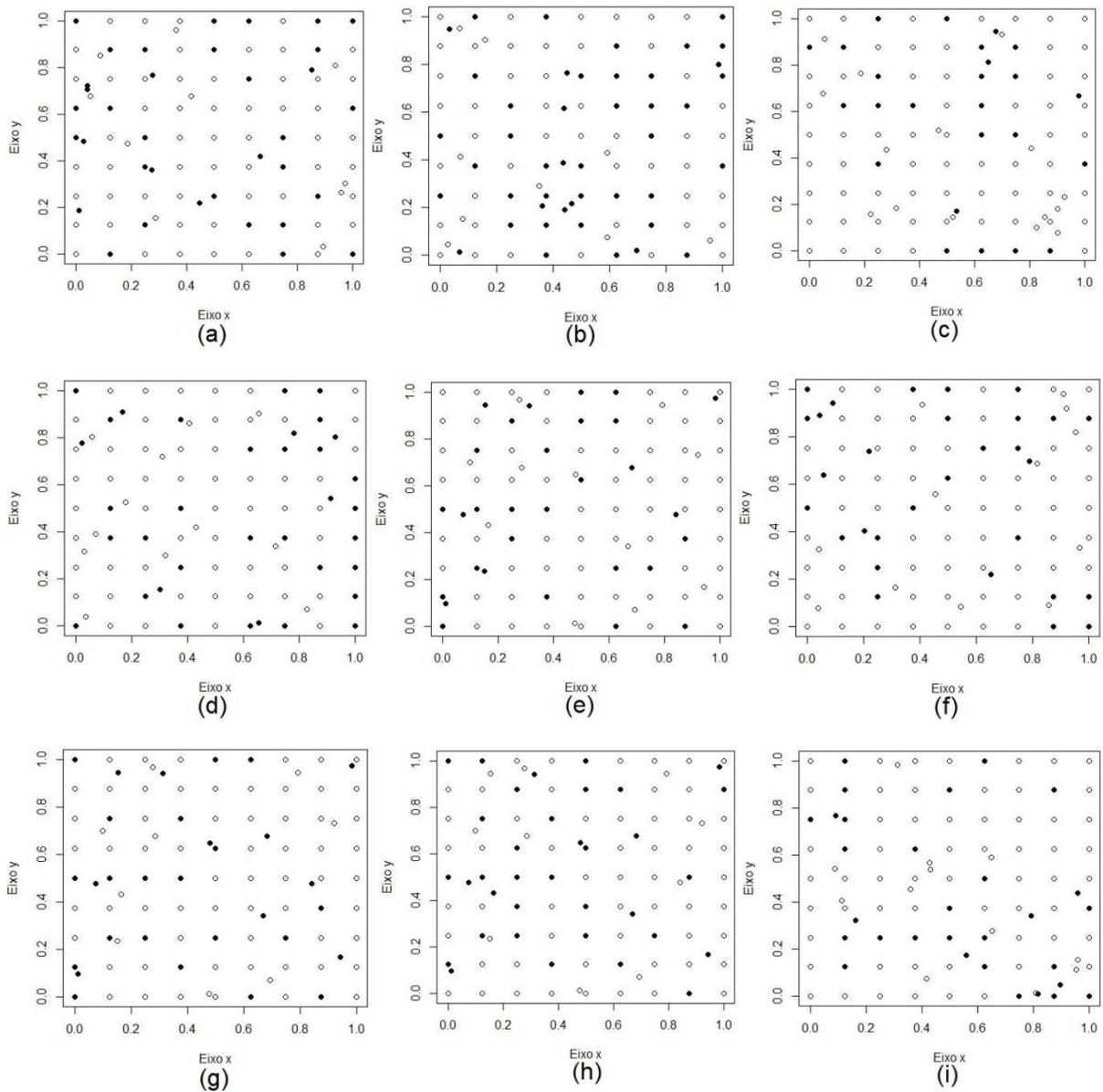


Figura 28 Gráfico da localização dos 100 pontos iniciais (○) e a localização dos pontos selecionados (●), para um exemplo de cada simulação da amostragem *latticeplus close pairs*, (a) alcance 0,45 e efeito pepita 0; (b) alcance 0,60 e efeito pepita 0; (c) alcance 0,90 e efeito pepita 0; (d) alcance 0,45 e efeito pepita 0,50; (e) alcance 0,60 e efeito pepita 0,50; (f) alcance 0,90 e efeito pepita 0,50; (g) alcance 0,45 e efeito pepita 0,80; (h) alcance 0,60 e efeito pepita 0,80; (i) alcance 0,90 e efeito pepita 0,80.

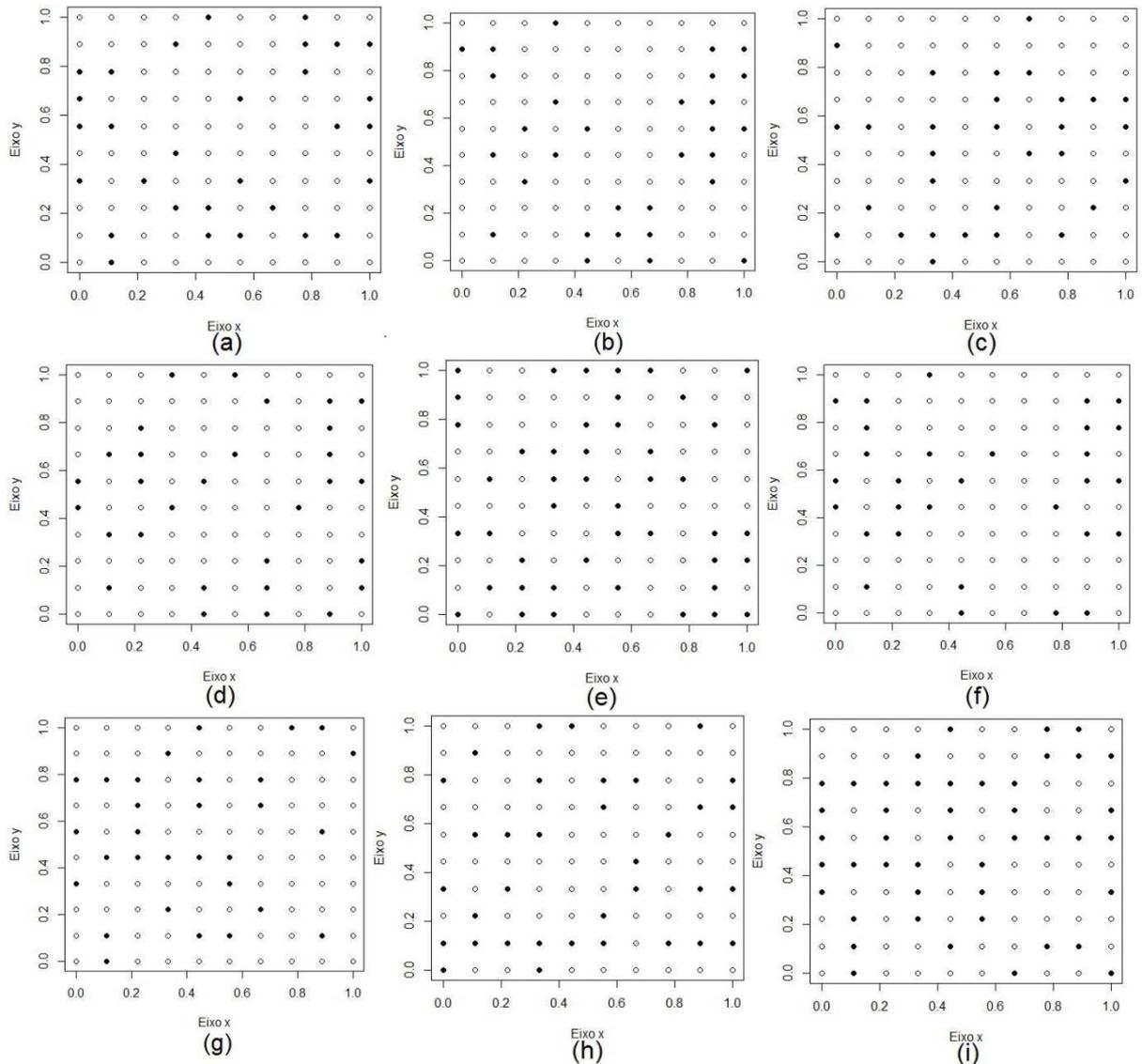


Figura 29 Gráfico da localização dos 100 pontos iniciais (\circ) e a localização dos pontos selecionados (\bullet), para um exemplo de cada simulação da amostragem sistemática, (a) alcance 0,45 e efeito pepita 0; (b) alcance 0,60 e efeito pepita 0; (c) alcance 0,90 e efeito pepita 0; (d) alcance 0,45 e efeito pepita 0,50; (e) alcance 0,60 e efeito pepita 0,50; (f) alcance 0,90 e efeito pepita 0,50; (g) alcance 0,45 e efeito pepita 0,80; (h) alcance 0,60 e efeito pepita 0,80; (i) alcance 0,90 e efeito pepita 0,80.

6. CONSIDERAÇÕES FINAIS

- Conclui-se que os processos de otimização AG, considerando: (1º) maximização da medida de acurácia EG e, (2º) minimização da função-objetivo, definida pelo negativo do logaritmo de determinante da matriz de informação de Fischer, ou de forma equivalente, o logaritmo do determinante da matriz inversa de Fisher foram eficientes e estáveis quando aplicados tanto nos dados simulados quanto nos atributos químicos do solo;
- O algoritmo de otimização possibilitou o redimensionamento de uma grade amostral composta por 102 pontos referente a uma área comercial e reduziu seu tamanho amostral;
- Para a maximização da EG, tem-se que o tamanho amostral da nova amostra reduzida variou entre 30 e 35 pontos que corresponde respectivamente de 29,41% a 34,31% da malha inicial. Mas, para a minimização da função, definida pelo negativo do logaritmo de determinante da matriz de informação de Fischer, o tamanho da amostra otimizada variou entre 30 e 40 pontos e correspondeu respectivamente em 29,41% e 39,22% da malha original.
- Houve similaridade entre os mapas elaborados com a configuração de pontos amostrais original e a otimizada para todos os atributos químicos do solo quanto à construção dos mapas temáticos, para a maximização da medida de acurácia EG. Entretanto, para a minimização do negativo do logaritmo do determinante da matriz de informação de Fischer, para todos os atributos químicos do solo não houve similaridade entre os mapas elaborados, com a configuração amostral inicial e otimizada. Isso pode ser consequência do que cada função-objetivo considera como prioridade no seu processo de otimização.
- Em relação aos dados simulados, quando o efeito pepita e o alcance variaram, em geral, as amostragens apresentaram os menores valores da função-objetivo, com efeito pepita igual a 0 e alcance prático igual a 0,9.
- Mais estudos podem ser realizados: 1º) levando-se em consideração, simultaneamente, as diversas variáveis, ou seja, um processo de otimização multivariado; (2º) outros processos de otimização que possibilitem obter resultados eficazes com menor tempo computacional; (3º) considerando-se a presença de anisotropia na estrutura de dependência espacial; (4º) considerando-se a otimização de medidas associadas ao custo do experimento; (5º) considerando-se outras distribuições de probabilidade para a variável georreferenciada; e (6º) reproduzir experimentos em outras áreas.