

**Unioeste - Universidade Estadual do Oeste do Paraná**  
CENTRO DE CIÊNCIAS EXATAS E TECNOLÓGICAS  
Colegiado de Ciência da Computação  
*Curso de Bacharelado em Ciência da Computação*

**Estimação dos Parâmetros de Semivariogramas Utilizando Algoritmos Genéticos**

*Hemerson Leandro Siqueira Carlin*

**CASCADEL**  
**2010**

**Hemerson Leandro Siqueira Carlin**

**Estimação dos Parâmetros de Semivariogramas Utilizando Algoritmos Genéticos**

Monografia apresentada como requisito parcial para obtenção do grau de Bacharel em Ciência da Computação, do Centro de Ciências Exatas e Tecnológicas da Universidade Estadual do Oeste do Paraná - Campus de Cascavel

Orientador: Prof. André Luiz Brun

CASCADEL  
2010

**Hemerson Leandro Siqueira Carlin**

**Estimação dos Parâmetros de Semivariogramas Utilizando Algoritmos Genéticos**

Monografia apresentada como requisito parcial para obtenção do Título de Bacharel em Ciência da Computação, pela Universidade Estadual do Oeste do Paraná, Campus de Cascavel, aprovada pela Comissão formada pelos professores:

---

Prof. André Luiz Brun (Orientador)  
Colegiado de Ciência da Computação,  
UNIOESTE

---

Prof. Adair Santa Catarina  
Colegiado de Ciência da Computação,  
UNIOESTE

---

Prof. Adriana Postal  
Colegiado de Ciência da Computação,  
UNIOESTE

Cascavel, 17 de novembro de 2010

*Durante muito tempo eu fiquei preocupado com o que os outros achavam ao meu respeito. Mas hoje, o que os outros acham de mim muito pouco me importa (a não ser que sejam pessoas que me amam), porque a minha salvação não depende do que os outros acham de mim, mas do que Deus sabe ao meu respeito.*

Pe. Fábio de Melo

## **AGRADECIMENTOS**

Agradeço à Deus pela vida, conquistas e vitórias a mim concedidas.

À toda minha família de um modo geral, e em especial aos meus pais Hemerson e Nelci e à minha namorada Gabriela, que me ajudou, apoiou e soube entender o fato de não podermos estar juntos nos momentos de desenvolvimento deste trabalho.

Ao professor André Brun por acreditar na minha capacidade e por todo apoio por ele fornecido.

À UNIOESTE, todos os funcionários e professores, com os quais muito aprendi durante minha graduação.

Aos meus amigos que contribuíram direta ou indiretamente para a conclusão deste trabalho.

À todas as pessoas que, de uma forma ou de outra, deram auxílio e incentivo para que conseguisse concluir esta pesquisa.

# Lista de Figuras

2.1	Estrutura Clássica de um Semivariograma . . . . .	10
2.2	Modelo Linear . . . . .	11
2.3	Modelo Circular . . . . .	11
2.4	Modelo Esférico . . . . .	12
2.5	Modelo Exponencial . . . . .	13
2.6	Modelo Gaussiano . . . . .	13
3.1	Fluxograma que descreve um AG . . . . .	22
3.2	Método da roleta com base nos indivíduos das Tabelas 3.1 e 3.2 . . . . .	26
3.3	Troca de segmentos com um ponto de cruzamento . . . . .	28
3.4	Troca de segmentos com cruzamento multiponto . . . . .	28
3.5	Exemplo de mutação utilizando codificação binária . . . . .	31
4.1	Leitura dos dados de entrada . . . . .	35
4.2	Biblioteca <i>JFreeChart</i> utilizada para montagem dos gráficos . . . . .	36
4.3	Exemplo do cálculo das distâncias $h$ e $N(h)$ . . . . .	36
4.4	Estrutura do cromossomo implementado no AG . . . . .	39
4.5	Cálculo da aptidão dos cromossomos utilizando o método dos mínimos quadrados	39
4.6	Tela de configuração do AG . . . . .	40
4.7	Resultado mostrado após execução do AG . . . . .	41
4.8	Fluxograma do AG Implementado. . . . .	41
4.9	Resultado Final . . . . .	42
4.10	Configuração dos testes . . . . .	43
4.11	Resultados do AG salvos durante execução do programa . . . . .	43

A.1	Leitura de um arquivo .xls . . . . .	48
A.2	Criação de planilha . . . . .	49
A.3	Planilha criada . . . . .	49
B.1	Configuração inicial do desenho do gráfico . . . . .	50
B.2	Criação das séries . . . . .	51
B.3	Customização do gráfico . . . . .	51
B.4	Desenho do gráfico . . . . .	51

# Lista de Tabelas

3.1	Representação dos indivíduos pela codificação binária para seleção do método da roleta . . . . .	25
3.2	Representação dos indivíduos pela codificação real para seleção do método da roleta . . . . .	25
3.3	Seleção por Torneio . . . . .	26
5.1	Configuração do computador utilizado para os testes . . . . .	44
5.2	Resultado para mutação <i>creep</i> e estimador Cressie . . . . .	45
5.3	Resultado para mutação <i>creep</i> e estimador Matheron . . . . .	45
5.4	Resultado para mutação <i>creep</i> e estimador Pairwise . . . . .	45
5.5	Resultado para mutação uniforme e estimador Cressie . . . . .	46
5.6	Resultado para mutação uniforme e estimador Matheron . . . . .	46
5.7	Resultado para mutação uniforme e estimador Pairwise . . . . .	46



# Lista de Abreviaturas e Siglas

a	alcance
AG	Algoritmo Genético
AIC	Akaike's Information Criterion
AP	Agricultura de Precisão
C	Patamar
C0	Efeito Pepita
C1	Contribuição
EA	Erro Absoluto
EM	Erro Médio
ER	Erro Médio Reduzido
IDE	Integrated Development Environment
LP	Linguagem de Programação
MMQ	Método dos Mínimos Quadrados
OO	Orientada a Objetos

# Sumário

<b>Lista de Figuras</b>	<b>vi</b>
<b>Lista de Tabelas</b>	<b>viii</b>
<b>Lista de Abreviaturas e Siglas</b>	<b>ix</b>
<b>Sumário</b>	<b>x</b>
<b>Resumo</b>	<b>xiii</b>
<b>1 Introdução</b>	<b>1</b>
<b>2 Agricultura de Precisão</b>	<b>3</b>
2.1 Introdução . . . . .	3
2.2 Vantagens . . . . .	4
2.3 Funcionamento . . . . .	4
2.4 Geoestatística . . . . .	5
2.5 Semivariogramas . . . . .	6
2.6 Ajuste de Semivariogramas . . . . .	10
2.6.1 Modelo Linear . . . . .	10
2.6.2 Modelo Circular . . . . .	11
2.6.3 Modelo Esférico . . . . .	12
2.6.4 Modelo Exponencial . . . . .	12
2.6.5 Modelo Gaussiano . . . . .	13
2.7 Krigagem . . . . .	14
2.8 Indicadores de Melhor Ajuste . . . . .	14
2.8.1 Validação Cruzada . . . . .	14
2.8.2 Critério de Akaike . . . . .	16
2.8.3 Critério de Filliben . . . . .	17

<b>3</b>	<b>Algoritmos Genéticos</b>	<b>19</b>
3.1	Introdução . . . . .	19
3.2	Histórico . . . . .	20
3.3	Estrutura de um Algoritmo Genético . . . . .	21
3.4	Representação dos Indivíduos . . . . .	22
3.4.1	Representação Binária . . . . .	23
3.4.2	Representação Real . . . . .	23
3.5	Seleção . . . . .	24
3.5.1	Método da Roleta ( <i>Roulette Wheel</i> ) . . . . .	24
3.5.2	Seleção por Torneio . . . . .	25
3.5.3	Seleção por Rank . . . . .	26
3.6	Cruzamento . . . . .	27
3.6.1	Representação Binária . . . . .	27
3.6.2	Representação Real . . . . .	28
3.7	Mutação . . . . .	30
3.7.1	Representação Binária . . . . .	30
3.7.2	Representação Real . . . . .	30
3.8	Elitismo . . . . .	31
3.9	Parâmetros Genéticos . . . . .	32
3.9.1	Tamanho da População . . . . .	32
3.9.2	Taxa de Cruzamento . . . . .	32
3.9.3	Taxa de Mutação . . . . .	33
3.9.4	Intervalo de Cruzamento . . . . .	33
<b>4</b>	<b>Metodologia</b>	<b>34</b>
4.1	Materiais e Métodos . . . . .	34
4.2	Configuração dos Testes . . . . .	42
<b>5</b>	<b>Resultados e Discussões</b>	<b>44</b>
5.1	Resultados . . . . .	44
<b>6</b>	<b>Conclusão</b>	<b>47</b>
<b>A</b>	<b>Tutorial de utilização da biblioteca <i>JExcelAPI</i></b>	<b>48</b>

<b>B Tutorial de utilização da biblioteca <i>JFreeChart</i></b>	<b>50</b>
<b>Referências Bibliográficas</b>	<b>52</b>

# Resumo

Agricultura de Precisão e Semivariogramas trazem estudos focados principalmente no solo, como análises de densidade, microporosidade e variabilidade de atributos físico-químicos. Estes questionam quais fatores influenciam no sucesso e/ou insucesso dos parâmetros do semivariograma.

Com isso, a redução de custos para o agricultor é uma consequência, fazendo com que os insumos agrícolas sejam aplicados somente onde de fato se faz necessário e não indiscriminadamente em toda a extensão da área cultivada como tradicionalmente é feito.

Técnicas de estimação de semivariogramas buscam encontrar uma aproximação para as curvas geradas pelos modelos matemáticos. Este trabalho aplica Algoritmos Genéticos (AGs) para encontrar os valores dos parâmetros de dependência espacial efeito pepita, contribuição e alcance, de forma automática procurando tornar o processo de estimação mais eficiente e seguro.

A função do software proposto é encontrar a melhor combinação de cruzamento  $\times$  mutação do Algoritmo Genético que traga os melhores resultados em um menor intervalo de tempo. Para isto, os testes foram executados sobre três estimadores, cinco modelos matemáticos, dois métodos de mutação, um de seleção e mais cinco de cruzamento, sendo que os melhores resultados continham, sem exceção, o cruzamento Aritimético.

**Palavras-chave:** Agricultura de precisão, semivariograma, geoestatística, estimação, algoritmo genético.

# Capítulo 1

## Introdução

A produção agrícola brasileira vem apresentando um grande crescimento. Isto é possível graças à enorme quantidade de áreas agricultáveis, que vão desde terras já exploradas e utilizadas para o plantio à terras inexploradas. Outros dois principais fatores que contribuem com esse crescimento são o clima favorável e o alto nível tecnológico dos produtores e agroindústrias.

A Agricultura de Precisão (AP) teve início na pressão por uma agricultura que afetasse menos o meio ambiente, primeiramente nos países europeus e, logo após, nos Estados Unidos. O incentivador deste desenvolvimento se deu na busca da maximização da produção pela dosagem precisa dos insumos, dada a variabilidade espacial existente.

Esta variabilidade espacial é o mais importante foco de aplicação da Geoestatística, que é um conjunto de procedimentos estatísticos que permite identificar a existência ou não de dependência espacial entre as observações.

A principal contribuição da Geoestatística, neste processo, é a gerência adequada do solo, através de um estudo detalhado das áreas de plantio, construindo-se mapas temáticos que auxiliam na aplicação correta dos insumos, diminuindo os impactos e otimizando os custos.

Entre os instrumentos geoestatísticos destaca-se o semivariograma, que é uma das principais abordagens empregadas para estudo espacial de propriedades do solo. É um artifício estatístico usado na verificação da presença de dependência espacial entre pontos amostrais georreferenciados espacialmente.

A estrutura clássica de um semivariograma é composta por quatro parâmetros: alcance ( $a$ ), efeito pepita ( $C_0$ ), contribuição ( $C_1$ ) e patamar, os quais são de grande importância para a construção de mapas de produtividade e para o processo estimativo. Conseqüentemente, a má es-

colha do valor desses parâmetros pode comprometer a eficiência da estimação. A determinação dos parâmetros é feita com base em um modelo matemático conhecido.

Algoritmos Genéticos (AGs) são métodos de busca de propósito geral com seus princípios inspirados pela genética natural das populações que envolvem soluções para os problemas. A ideia básica constitui-se em manter uma população de cromossomos gerada aleatoriamente, representando supostos candidatos para a solução do problema, e que evoluem ao longo do tempo através de um processo de competição e variação controlada.

Durante sucessivas iterações (ou gerações), os cromossomos da população são avaliados como soluções e, com base nessas avaliações, uma nova população é formada. Cada cromossomo da população possui uma função *fitness* associada que fornece uma medida da qualidade da solução do problema e determina quais cromossomos são utilizados para formar novos indivíduos no processo de competição, que consiste em oferecer os melhores indivíduos para a fase de reprodução.

Com o auxílio de Algoritmos Genéticos, seria mais eficiente encontrar os valores aproximados entre semivariograma e modelo. A importância deste trabalho é auxiliar na pesquisa de forma a deixar este processo mais apropriado. Para tanto, desenvolveu-se uma ferramenta computacional capaz de determinar, através de um Algoritmo Genético, o melhor ajuste do semivariograma, encontrando assim os valores ideais para os parâmetros deste, trazendo mais eficiência na aplicação dos insumos, no rendimento da colheita, maior produtividade da lavoura e menor impacto ambiental, contribuindo também para o estudo de técnicas geoestatísticas.

# Capítulo 2

## Agricultura de Precisão

### 2.1 Introdução

Em todos os setores da economia globalizada há a necessidade do aumento da eficiência para manter a competitividade. Para a agricultura isso não é diferente. Um dos fatores que mais contribuíram foi a evolução da informática, com o surgimento de novas tecnologias e ferramentas de apoio. Essa evolução proporcionou uma nova forma de enxergar a propriedade que antes era vista como apenas uma e, após esse processo de evolução, passou a ser vista como várias propriedades dentro de uma só, cada uma com suas características específicas.

Segundo [Werner 2007], o tema Agricultura de Precisão é bastante atual no meio agrícola brasileiro. Para alguns produtores, sua utilização é uma realidade como ferramenta para a busca da máxima eficiência gerencial da propriedade. Enquanto o sistema de agricultura tradicional (que ainda é amplamente praticado no Brasil) desconsidera a variabilidade existente nas áreas de produção, onde os recursos de produção são aplicados da maneira mais uniforme possível, a agricultura de precisão busca otimizar a aplicação dos recursos de produção de acordo com a variabilidade existente. A proposta da agricultura de precisão é permitir que se faça aquilo que o pequeno agricultor sempre fez, porém em larga escala associando todo o conhecimento acumulado pelas ciências agrárias até hoje [Molin 2001].

Conforme [Balastreire 1998], a agricultura de precisão envolve um conjunto de técnicas que permite um tratamento mais eficiente de plantio, por meio do gerenciamento localizado de culturas. A AP tem seus fundamentos calcados na percepção da variabilidade espacial da produtividade e de fatores físicos, químicos e hídricos do solo que nela tem direta influência.

A AP permite, pelo uso de demarcação de lavouras por coordenadas georreferenciadas, um



planejamento mais racional do manejo de nutrientes, pragas, umidade do solo, plantas daninhas, além da seleção de cultivares em função de sua adaptabilidade às diferentes condições identificadas nas áreas de cultivo. A expectativa de redução de custos está, principalmente, relacionada ao fato de que os insumos agrícolas somente serão aplicados onde de fato se faz necessário e não indiscriminadamente, como tradicionalmente é feito [Silva, Moretto e Rodrigues 2004].

## **2.2 Vantagens**

Segundo [Gentil e Ferreira 1999] e [Batchelor, Whigham e Dewitt 1997], a agricultura de precisão traz grandes benefícios para os usuários, tais como:

- a) redução do grave problema do risco da atividade agrícola, como impacto ambiental;
- b) redução dos custos da produção, dada a aplicação diferenciada dos insumos;
- c) tomada de decisão rápida e certa;
- d) controle de toda situação, pelo uso da informação;
- e) maior produtividade da lavoura;
- f) mais tempo livre para o administrador;
- g) melhoria do meio ambiente pelo menor uso de defensivo.
- h) fornecer informações para tomar decisões de manejo mais embasadas;
- i) prover registros de fazenda mais detalhados e úteis.

## **2.3 Funcionamento**

As etapas básicas do sistema de agricultura de precisão são:

- a) a coleta de dados;
- b) o planejamento do gerenciamento;
- c) a aplicação localizada dos insumos.

Na etapa de coleta de dados o principal objetivo é identificar a variabilidade existente em campo dos diversos fatores de produção, tais como: solo, pragas, ervas daninhas e da própria produção da cultura. O primeiro passo é fazer o mapa de produtividade da colheita através da instalação, em colheitadeiras, de equipamentos como GPS e sensores, utilizados para medir o volume e as condições físicas dos grãos colhidos em cada trecho mapeado.

Essas informações são passadas para um computador que constrói os mapas de acordo com a quantidade produzida em cada trecho, através de técnicas de estimação como a Krigagem (seção 2.7).

A segunda fase processa os dados coletados na primeira etapa trabalhando com a avaliação e quantificação da variabilidade medida. A partir daí, são analisados os fatores de produção, as características da terra e, então, as estratégias de gerenciamento agrícola são propostas.

Após a análise das amostras, o agricultor terá a fertilidade da terra traduzida, bem como a ocupação das plantas daninhas e outros mapas que podem mostrar o nível de  $pH$  existente, estrutura e drenagem do solo, densidade das plantas, estágio de crescimento, etc.

Conhecidas todas as características da propriedade, o próximo estudo é feito para a aplicação dos insumos, ou seja, quais insumos devem ser utilizados, suas quantidades e, principalmente, a posição exata de aplicação. O maior ganho de desempenho nesse processo quando comparado com a agricultura tradicional é que o agricultor aplica apenas a quantidade necessária em cada região distinta do terreno.

Na terceira e última etapa, máquinas agrícolas são utilizadas para a aplicação desses insumos, que variam desde sementes, pesticidas e fertilizantes até defensivos e corretivos.

Variabilidade espacial de solos sempre existiu e deve ser considerada toda vez que a amostragem de campo for efetuada [Vieira 1998] e, recentemente, uma ferramenta compreensível surgiu, capaz de utilizar técnicas para estudar o solo onde apenas a estatística clássica não consegue bons resultados. Essa ferramenta chama-se geoestatística.

## **2.4 Geoestatística**

Em 1951 [Krige 1951] notou que era necessário considerar a distância entre as amostras de dados de concentração de ouro e verificou que estas apresentavam algum tipo de dependência espacial.

[Júnior 1995] afirma que a Geoestatística é definida como um conjunto de procedimentos estatísticos aplicados a problemas, onde os dados são referenciados espacialmente. Assim, as observações apresentam características de acordo com a posição e não somente com relação aos valores que assumem. Portanto, observações mais próximas geograficamente apresentam a tendência de valores mais parecidos.

[Freitas 2000] cita que a variabilidade espacial de propriedades do solo é ainda a grande área de aplicação da geoestatística e que esta vem crescendo atualmente, pois tem sido aplicada em vários campos do conhecimento como análise espacial da distribuição das cigarras [Soares et al. 2008], plantas daninhas [Shiratsuchi 2001], precipitação pluviométrica [Carvalho e Assad 2005] e sucesso da influência de condição e predação da casca de besouros [Gilbert e Grégoire 2003].

A utilização da geoestatística permite identificar a existência ou não de dependência espacial entre as observações, podendo ser aplicada em mapeamentos, orientação de futuras amostragens e modelagens, permitindo, assim, estimar o valor do atributo em locais não amostrados, facilitando a gestão dos recursos naturais, especialmente em bacias hidrográficas representativas [Gomes et al. 2007].

A base de estudo da Geoestatística é a variabilidade espacial de fenômenos naturais. Para realizar este estudo, deve ser considerada a dependência espacial entre as amostras e a localização geográfica das mesmas [Câmara e Medeiros 1998].

Entre os instrumentos geoestatísticos, o semivariograma é uma das principais abordagens empregadas para estudo espacial de propriedades do solo [Burgess e Webster 1980].

## 2.5 Semivariogramas

O semivariograma pode ser interpretado como uma técnica estatística usada na verificação da presença de dependência espacial entre pontos amostrais georreferenciados, representado por um gráfico da função semivariância *versus* a sua distância. A distância máxima para qual a semivariância é estimada, é chamada *Cut-off*. Os pontos que estão posicionados, além desta distância máxima, são considerados não influentes [Isaaks e Srivastava 1989].

Segundo [Guerra 1988], há três tipos de semivariograma:

1. semivariograma verdadeiro: apresenta a dependência real que é desconhecida;

2. semivariograma experimental: é obtido por meio dos pontos amostrais;
3. semivariograma teórico: é o semivariograma ajustado mediante um modelo teórico ao semivariograma experimental.

A função semivariância  $\gamma(h)$  é uma importante estatística utilizada para a construção do semivariograma experimental e é definida pela Equação 2.1.

$$\gamma(h) = \frac{1}{2}E[Z(s) - Z(s+h)]^2 \quad (2.1)$$

onde:

$h$  representa cada distância considerada entre quaisquer dois pontos;

$Z(s)$  é o valor da variável  $Z$  medida em um local representado por  $s$ ;

$Z(s+h)$  é o valor da variável  $Z$  medida em um local distanciado de  $s$  por uma distância  $h$ ;

$s$  é a posição amostral.

Estimar uma função de semivariância significa estimar convenientemente os parâmetros de um modelo teórico. Os estimadores mais conhecidos para a função semivariância  $\gamma(h)$  são Estimador Clássico de Matheron [Matheron 1963], o Estimador de Cressie e Hawkins [Cressie e Hawkins 1980] e o Estimador de Semivariância Relativa Pairwise [Srivastava e Parker 1989].

Para [Júnior 1995], o estimador de Matheron é um estimador não viciado para  $\gamma(h)$  quando a função  $Z()$  é um processo estocástico intrinsecamente estacionário, mas é pouco resistente e muito afetado por observações atípicas (*outliers*). Nestes casos surge a necessidade de um estimador robusto, que não seja afetado por valores discrepantes que não podem ser eliminados ou substituídos como é o caso do Estimador de Cressie e Hawkins. [Júnior 1995] afirma ainda que o Estimador de Cressie e Hawkins é adequado em casos onde são simuladas distribuições assimétricas, pois fornece uma estimativa robusta e estável do semivariograma.

Já o Estimador de Semivariância Relativa Pairwise [Srivastava e Parker 1989], caracteriza-se por diminuir o efeito de dados discrepantes, porém apresenta a desvantagem da soma dos

valores da variável  $Z$  nos pontos de coordenadas  $s$  e  $(s + h)$  não poder ser nula. Além disso, quando a média se aproxima de zero, podem ocorrer semivariogramas com “saltos” nas semi-variâncias [Isaaks e Srivastava 1989].

O Estimador Clássico de Matheron [Matheron 1963], o Estimador de Cressie e Hawkins [Cressie e Hawkins 1980] e o Estimador de Semivariância Relativa Pairwise [Srivastava e Parker 1989] são representados pelas Equações 2.2, 2.3 e 2.4 respectivamente.

$$\hat{\gamma}_M(h) = \frac{1}{2N(h)} \sum^{N(h)} E[Z(s) - Z(s+h)]^2 \quad (2.2)$$

$$\hat{\gamma}_C(h) = \frac{\left\{ \frac{1}{N(h)} \sum^{N(h)} E|Z(s) - Z(s+h)|^{\frac{1}{4}} \right\}^4}{0,914 + \frac{0,988}{N(h)}} \quad (2.3)$$

$$\hat{\gamma}_P(h) = \frac{2}{N(h)} \sum^{N(h)} \left[ \frac{Z(s) - Z(s+h)}{Z(s) + Z(s+h)} \right]^2 \quad (2.4)$$

onde:

$N(h)$  é o número de pares de dados utilizados a uma distância  $h$ ;

$Z$  é a variável regionalizada ou função aleatória;

$s = (x_i; y_i) \in S \subset \mathfrak{R}^2$  são as posições amostrais.

[Li e Lake 1994] propuseram dois novos estimadores de semivariância: os estimadores New 1 e New 2, a fim de tentar reduzir a imprecisão do estimador quando a distância entre as amostras aumenta. Os estimadores de Semivariância New 1 e New 2 são dados pelas Equações 2.5 e 2.6, respectivamente.

$$\hat{\gamma}_{N1}(h) = \frac{1}{N} \sum_{i=1}^N \left\{ \frac{1}{2m} \sum_{j \in D_{i,h}} [Z(s_i) - Z(s_j)]^2 \right\} \quad (2.5)$$

$$\hat{\gamma}_{N2}(h) = \hat{\gamma}_{N1}(h) + \frac{h}{d} \hat{\gamma}_{N1}(h) \quad (2.6)$$

onde:

$n$  é o número de pontos amostrais na área toda;

$D_{i,h}$  é o conjunto indexado dos valores da variável regionalizada  $Z$ , em pontos de uma janela móvel bidimensional de área  $\Delta_{i,h}$  com tamanho  $h$  centrada no bloco  $i$ , excluindo o ponto  $s_i$ ;

$m$  é o número de dados em  $D_{i,h}$ ;

$\hat{\gamma}'_{N1}(h)$  é o valor da derivada da função  $\hat{\gamma}_{N1}$  na distância  $h$ ; sendo  $\hat{\gamma}_{N1}$  o modelo ajustado a  $\gamma(h)$ ;

$d$  é o momento da semivariância ( $d = X$  ou  $d = S^2 + \bar{X}^2$ ).

Além do modelo de semivariograma experimental escolhido para ser usado no processo de estimação, é necessário determinar o valor do *Cut-off*, ou ponto de corte. O valor determinado refere-se à distância máxima utilizada no cálculo da semivariância.

A estrutura clássica de um semivariograma é composta por quatro parâmetros: alcance ( $a$ ), efeito pepita ( $C_0$ ), contribuição ( $C_1$ ) e patamar ( $C = C_0 + C_1$ ), os quais são de grande importância para a construção de mapas de contorno confrontando com o modelo matemático [Isaaks e Srivastava 1989].

Conforme [Cressie 1993], o alcance ( $a$ ) representa a distância máxima onde as amostras se apresentam correlacionadas espacialmente. Quando as amostras não apresentam dependência, significa que as distâncias são superiores ao valor do alcance. [Journel e Huijbregts 1978] relatam que o alcance fornece uma ideia da região de influência da variável regionalizada.

[Câmara e Medeiros 1998] afirmam que o efeito pepita ( $C_0$ ) revela a descontinuidade do semivariograma para distâncias maiores que a menor distância entre as amostras. Esta descontinuidade pode ocorrer em casos onde houve erros de análise ou em que a variabilidade de pequena escala não foi detectada pela amostragem.

Segundo [Isaaks e Srivastava 1989], o patamar ( $C$ ) é o valor na qual a função semivariância  $\gamma(h)$  se estabiliza. O valor de  $\gamma(h)$  aumenta à medida que  $h$  cresce, até atingir um ponto com valor máximo estabilizado, ou seja, quando  $h$  é igual ao alcance.

A contribuição ( $C_1$ ) é a diferença entre o patamar e o efeito pepita.

A Figura 2.1 apresenta a estrutura clássica de um semivariograma, bem como seus parâmetros: alcance ( $a$ ), efeito pepita ( $C_0$ ), contribuição ( $C_1$ ) e patamar ( $C = C_0 + C_1$ ).

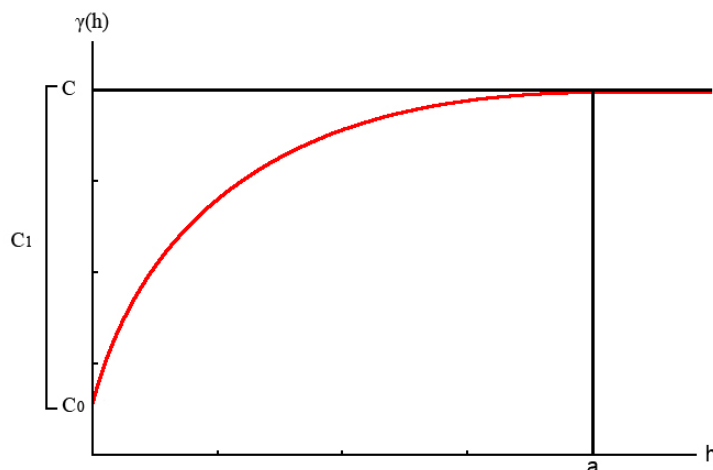


Figura 2.1: Estrutura Clássica de um Semivariograma

## 2.6 Ajuste de Semivariogramas

O padrão normal de um semivariograma se apresenta de forma intuitiva ao que esperamos de dados de campos, isto é, normalmente as observações mais próximas geograficamente apresentam comportamento mais semelhante entre si do que observações separadas por distâncias maiores [Matos 2003].

Existem diferentes modelos de semivariogramas que podem ser utilizados convenientemente, conforme o melhor ajuste aos valores observados no semivariograma experimental. Os principais modelos são: Modelo Linear, Circular, Esférico, Exponencial e Gaussiano.

### 2.6.1 Modelo Linear

O Modelo Linear com patamar é um modelo válido apenas em  $\mathbb{R}$  (espaço unidimensional), não sendo válido para dimensões mais abrangentes [Cressie 1993]. A equação 2.7 e o gráfico da Figura 2.2 representam este modelo.

$$\gamma(h) = \begin{cases} C_0 + C_1 \left(\frac{h}{a}\right), & \text{se } 0 \leq h \leq a; \\ C_0 + C_1, & \text{se } h > a. \end{cases} \quad (2.7)$$

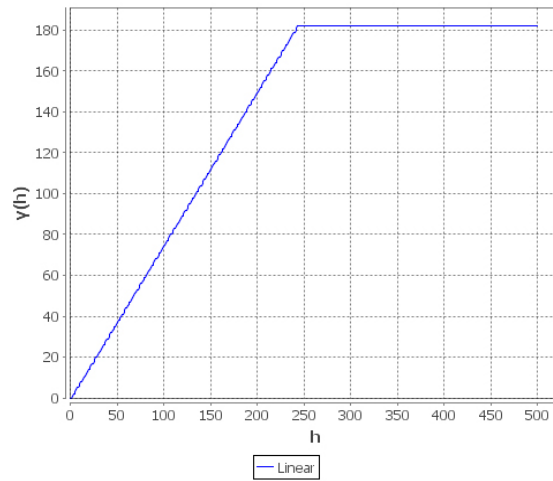


Figura 2.2: Modelo Linear

### 2.6.2 Modelo Circular

O Modelo Circular apresenta validade apenas nos planos unidimensionais e bidimensionais, não podendo ser aplicado a planos tridimensionais, onde é aplicado o modelo esférico. Este modelo é representado pela equação 2.8 e o gráfico da Figura 2.3.

$$\gamma(h) = \begin{cases} \frac{2(C_0+C_1)}{\Pi} \left[ \left(\frac{h}{a}\right) \sqrt{1 - \left(\frac{h}{a}\right)^2} + \arcsin\left(\frac{h}{a}\right) \right], & \text{se } 0 \leq h \leq a; \\ C_0 + C_1, & \text{se } h > a. \end{cases} \quad (2.8)$$

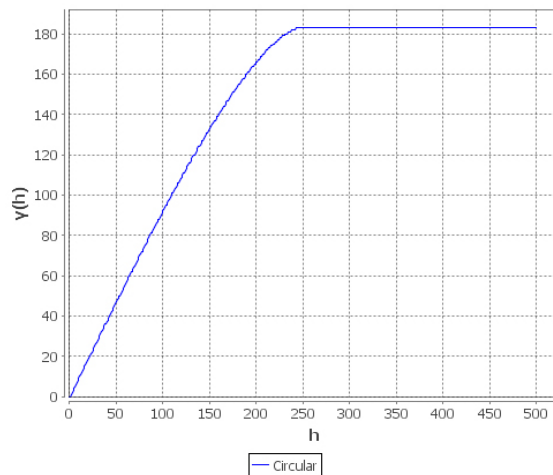


Figura 2.3: Modelo Circular



### 2.6.3 Modelo Esférico

O Modelo Esférico apresenta um alcance que vai sendo aumentado até determinado instante, a partir do qual a semivariância se torna constante, ou seja, limitando a sua área de influência para cada uma das amostras. A equação 2.9 e o gráfico da Figura 2.4 representam este modelo.

$$\gamma(h) = \begin{cases} C_0 + C_1 \left[ \frac{3}{2} \left( \frac{h}{a} \right) - \frac{1}{2} \left( \frac{h}{a} \right)^3 \right], & \text{se } 0 \leq h \leq a; \\ C_0 + C_1, & \text{se } h > a. \end{cases} \quad (2.9)$$

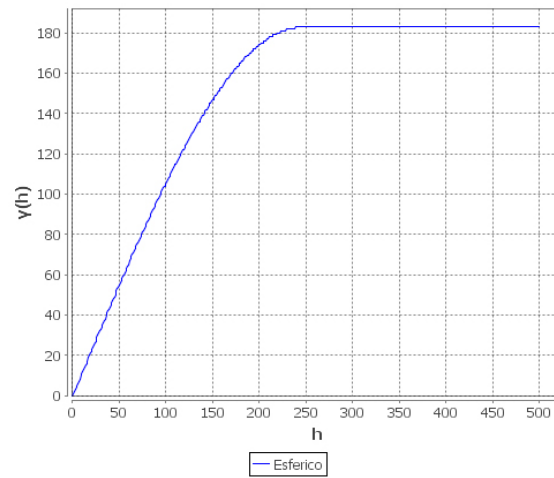


Figura 2.4: Modelo Esférico

### 2.6.4 Modelo Exponencial

O Modelo Exponencial é diferente do modelo esférico pois apresenta uma continuidade nos primeiros valores de  $h$ . Neste modelo o patamar é atingido apenas assintoticamente, pois a curva não chega a tocar  $C$ . Este modelo é representado pela equação 2.10 e o gráfico da Figura 2.5.

$$\gamma(h) = \begin{cases} C_0 + C_1 \left[ 1 - e^{-3\frac{h}{a}} \right], & \text{se } 0 \leq h \leq a; \\ C_0 + C_1, & \text{se } h > a. \end{cases} \quad (2.10)$$

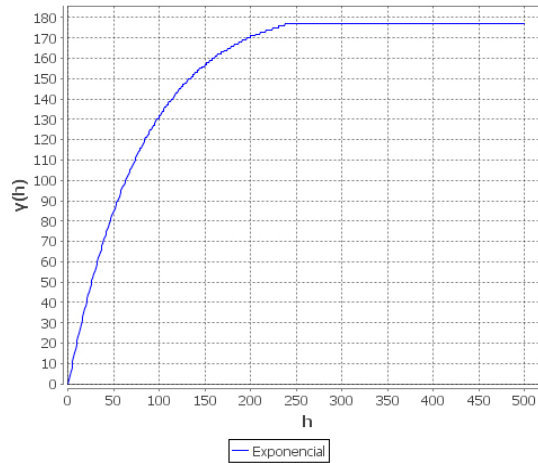


Figura 2.5: Modelo Exponencial

### 2.6.5 Modelo Gaussiano

O Modelo Gaussiano traz um alcance longo, já seu patamar possui comportamento semelhante ao modelo exponencial. Este modelo apresenta um comportamento parabólico próximo à origem e, dentre os cinco modelos mostrados neste trabalho, é o único que tem um ponto de inflexão na sua curva. A equação 2.11 e o gráfico da Figura 2.6 representam este modelo.

$$\gamma(h) = \begin{cases} C_0 + C_1 \left[ 1 - e^{-3\left(\frac{h}{a}\right)^2} \right], & \text{se } 0 \leq h \leq a; \\ C_0 + C_1, & \text{se } h > a. \end{cases} \quad (2.11)$$

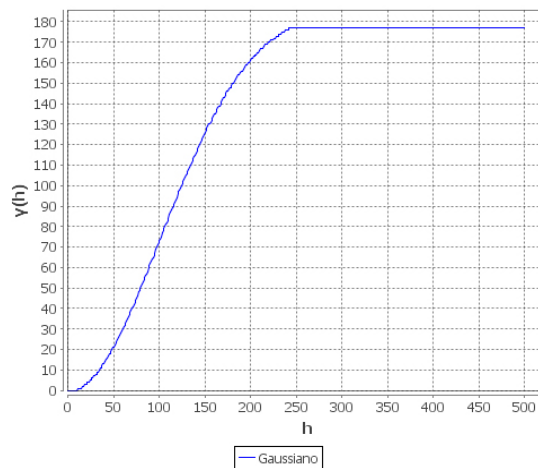


Figura 2.6: Modelo Gaussiano

## 2.7 Krigagem

A krigagem é uma técnica eficiente de interpolação que permite estimar valores em locais onde observações não foram extraídas, fazendo com que a estimação seja realizada com base no conjunto de pontos onde os atributos foram medidos [Matos 2003].

Este método considera que os pontos amostrados mais próximos possuem uma influência maior que os pontos mais distantes, atribuindo pesos baseados na distância entre os vizinhos, o ponto estimado e também na distância entre estes pontos [Cressie 1993].

O termo krigagem é derivado do nome Daniel G. Krige [Krige 1951], que foi o pioneiro a introduzir o uso de médias móveis para evitar a superestimação sistemática de reservas de mineração. Inicialmente o método de krigagem foi desenvolvido para solucionar problemas de mapeamentos geológicos, mas seu uso expandiu-se com sucesso no mapeamento de solos [Burgess e Webster 1980], mapeamento hidrológico [Kitanidis e Vomvoris 1983] e mapeamento atmosférico [Lajaunie 1984].

A diferença entre a krigagem e os outros métodos de interpolação é a forma como os pesos são atribuídos às diferentes amostras. No caso de interpolação linear simples, por exemplo, os pesos são todos iguais à  $1/N$ , enquanto que na interpolação baseada no inverso do quadrado das distâncias, os pesos são definidos como o inverso do quadrado da distância que separa o valor interpolado dos valores observados.

## 2.8 Indicadores de Melhor Ajuste

Os métodos de validação comparam valores teóricos do modelo geoestatístico escolhido e os valores empíricos obtidos na amostragem. Com base na análise desses erros de estimação, poderá ser selecionado o melhor modelo de semivariância. Os principais critérios para validação são: a Validação Cruzada, o Critério de Akaike, e o Critério de Filliben.

### 2.8.1 Validação Cruzada

A validação cruzada é uma técnica de avaliação de erros de estimativas que permite comparar os valores previstos com os amostrados [Isaaks e Srivastava 1989].

[Vieira 1998] afirma que a técnica consiste em ignorar, um a um, cada valor amostrado

e prever um valor interpolado pela krigagem, utilizando o modelo proposto. Comparando os valores preditos e os valores amostrados, pode-se escolher o melhor modelo dentre um conjunto de modelos propostos.

[Júnior 1995] e [GONÇALVES 1997] citam que após conhecidos os valores obtidos da amostragem bem como estes estiverem estimados para cada ponto, a variância total da estimativa é conhecida, permitindo realizar a avaliação de qualidade e o erro de estimação pode ser calculado (Erro Médio). [Brun 2007] cita que estes erros provenientes do processo de krigagem deveriam ter média nula e variância unitária, fazendo com que as características apresentadas pelo erro de estimação indiquem a eficácia do ajuste dos modelos de semivariogramas utilizados, ou seja, a Validação Cruzada é utilizada na tarefa de identificar qual modelo melhor representa o fenômeno em estudo [Davis 1987].

A Equação do Erro Médio (EM) (ver equação 2.12) é a equação que calcula os resíduos na validação cruzada.

$$EM = \frac{1}{n} \sum_{i=1}^n (Z_i(s) - \hat{Z}_i(s)) \quad (2.12)$$

onde:

$n$  é o número de dados;

$Z_i(s)$  é o valor observado no ponto  $s$ ;

$\hat{Z}_i(s)$  é o valor predito por krigagem ordinária no ponto  $s$ , sem considerar a observação  $Z_i(s)$ .

Esse procedimento pode ser visto como um experimento no qual se imita o processo de estimação, ao supor que nunca se toma uma amostra naquela localização. Uma vez que a estimação é feita, pode-se compará-la ao valor da amostra que foi inicialmente removida do conjunto de dados amostrais.

O erro médio reduzido (ER) (ver equação 2.13) é uma outra forma de calcular os resíduos da validação cruzada, segundo [McBratney e Webster 1986] e [Cressie 1993]:

$$ER = \frac{1}{n} \sum_{i=1}^n \frac{(Z_i(s) - \hat{Z}_i(s))}{\sigma_i(s)} \quad (2.13)$$

onde:

$\sigma_i(s)$  é o desvio padrão da krigagem no ponto  $s$ , sem considerar a observação  $Z_i(s)$ .

O desvio padrão dos erros reduzidos é obtido pela Equação 2.14

$$S_{ER} = \sqrt{\left(\frac{1}{n} \sum_{i=1}^n \frac{|Z_i(s) - \hat{Z}_i(s)|}{\sigma_i(s)}\right)} \quad (2.14)$$

Segundo [McBratney e Webster 1986], [Cressie 1993] e [Gomes et al. 2007], a escolha do melhor modelo ajustado ocorre quando o valor de  $EM$  e  $ER$  estão mais próximos de zero, o valor do desvio padrão menor e o valor de  $S_{ER}$  mais próximo de 1. O erro absoluto ( $EA$ ) é uma medida da magnitude dos erros na unidade da variável. Conhecendo-se o conjunto de valores medidos e preditos por krigagem ordinária  $Z_s(i)$  e  $\hat{Z}_s(i)$ , respectivamente, é possível definir o erro absoluto na unidade da variável estudada, segundo a Equação 2.15

$$EA = \sum_{i=1}^n |Z_i(s) - \hat{Z}_i(s)| \quad (2.15)$$

## 2.8.2 Critério de Akaike

O Critério de Informação de Akaike - AIC (*Akaike's Information Criterion*) procura uma solução satisfatória entre o bom ajuste e o princípio da parcimônia [Akaike 1973]. [Sakamoto, Ishiguro e Kitagawa 1986] desenvolveram estudos visando conhecer como os modelos são usados para fazer estimação. O AIC propõe verificar se dois modelos representam dados igualmente satisfatórios, sendo que a opção mais óbvia é optar pelo modelo mais simples, esperando melhor desempenho para a estimação de novos dados.

Modelos de avaliação de desempenho, como o método dos mínimos quadrados ordinários e mínimos quadrados ponderados, são intensamente utilizados para a escolha do melhor modelo, mas eles não consideram a complexidade utilizada no processo. A Equação 2.16 estima a parte variável de AIC para um conjunto de dados.

$$\hat{A} = n * \ln(R) + 2p \quad (2.16)$$

onde:

$n$  é o número de pontos experimentais do semivariograma;

$R$  é a soma dos quadrados dos resíduos  $\sum_{i=1}^n (Z_i(s) - \hat{Z}_i(s))^2$ ;

$\hat{Z}_s(i)$  é o valor predito no ponto  $i$ ;

$p$  é o número de parâmetros do modelo.

O modelo que apresentar o menor valor para  $\hat{A}$  será o escolhido.

### 2.8.3 Critério de Filliben

O Critério de [Filliben 1975] para resíduos ortonormais derivados do processo de validação cruzada também pode ajudar na determinação do melhor modelo geoestatístico para o ajuste.

Estes resíduos são obtidos através da diferença entre os valores amostrados e estimados divididos pelo desvio padrão da estimativa da krigagem, obtendo-se, assim, um erro para cada ponto amostrado [Brun 2007].

Esse erro é definido pela equação 2.17:

$$\varepsilon_k = \frac{Z(S_k) - \hat{Z}_\varepsilon(S_k)}{\sigma_k}, k = p + 1, \dots, n \quad (2.17)$$

onde:

$\varepsilon_k$  são os resíduos amostrados;

$Z(S_k)$  são os valores amostrados;

$\hat{Z}_\varepsilon(S_k)$  são os valores estimados pela krigagem;

$\sigma_k$  é o desvio-padrão da  $k$ -ésima estimativa da krigagem.

No teste de [Filliben 1975], calcula-se o coeficiente de correlação linear  $r^*$  entre as observações ordenadas e a estatística das medianas ordenadas de distribuição normal  $N(0, 1)$ . Quanto mais próximo  $r^*$  estiver de 1, mais normal é a distribuição dos dados [Faraco et al. 2008].

Se o valor tabelado de  $r$  para determinado tamanho de amostra é maior do que  $r^*$ , a hipótese de que os dados seguem uma distribuição normal é rejeitada para um certo nível percentual de

significância. Praticamente, o que se testa é se os resíduos ortonormais seguem uma distribuição normal em determinado nível de significância.

Para calcular o valor de  $r^*$ , utiliza-se a Equação 2.18, conhecida como cálculo do coeficiente de correlação linear:

$$r^* = \frac{\sum_{i=1}^n (W_i - \bar{W})(M_i - \bar{M})}{\sqrt{\left[ \sum_{i=1}^n (W_i - \bar{W})^2 \right] \left[ \sum_{i=1}^n (M_i - \bar{M})^2 \right]}} \quad (2.18)$$

onde:

$W_i$  são os resíduos ordenados;

$\bar{W}$  é a média dos resíduos;

$M_i$  é a estatística de uma distribuição Normal (0, 1);

$\bar{M}$  é a média da estatística de uma distribuição Normal (0, 1).

# Capítulo 3

## Algoritmos Genéticos

### 3.1 Introdução

Algoritmos Genéticos (AGs) são métodos de busca de propósito geral com seus princípios inspirados pela genética natural das populações que envolve soluções para os problemas [Holland 1975].

Os AGs são uma família de modelos computacionais baseados na evolução, que incorporam uma solução potencial para um problema específico numa estrutura semelhante a de um cromossomo e aplicam operadores de seleção, mutação e cruzamento a essas estruturas de forma a preservar informações críticas relativas à solução do problema [Miranda 2010].

A ideia básica constitui-se em manter uma população de cromossomos gerada aleatoriamente que representam supostos candidatos para a solução do problema e evoluem ao longo do tempo através de um processo de competição e variação controlada. Durante sucessivas interações (ou gerações), os cromossomos da população são classificados pela sua adaptação como soluções e, com base nessas avaliações, uma nova população é formada. Cada cromossomo da população possui uma função *fitness* associada que fornece uma medida da proximidade da solução do problema e determina quais cromossomos são utilizados para formar novos indivíduos no processo de competição, o qual é chamado de seleção [Goldberg 1989].

Os AGs possuem uma grande medida de sucesso na busca e otimização de problemas, e a razão para a grande parte deste sucesso está na capacidade de explorar a informação acumulada sobre um espaço de busca desconhecido, tendendo a realizar as próximas buscas em subespaços úteis. Esta é a sua característica fundamental, principalmente nos espaços de busca grandes, complexos e mal compreendidos onde os métodos tradicionais de busca são impróprios



[Herrera, Lozano e Verdegay 1998].

Segundo [Carvalho 2010], os Algoritmos Genéticos diferem dos métodos tradicionais de busca em quatro aspectos:

- a) AGs trabalham com uma codificação do conjunto de parâmetros e não com os próprios parâmetros;
- b) AGs trabalham com uma população e não com um único ponto;
- c) AGs utilizam informações de custo ou recompensa e não derivadas ou outro conhecimento auxiliar;
- d) AGs utilizam regras de transição probabilísticas e não determinísticas.

## **3.2 Histórico**

Até meados do século 19, os naturalistas acreditavam que cada espécie havia sido criada separadamente por um ser supremo ou através de geração espontânea. O trabalho do naturalista Carolus Linnaeus [Linneu 1758] e [Jr 2002] sobre a classificação biológica de organismos despertou o interesse pela similaridade entre certas espécies, levando a acreditar na existência de uma certa relação entre elas. Outros trabalhos influenciaram os naturalistas em direção à teoria da seleção natural, tais como os de Jean Baptiste Lamarck [Ferreira 2010], que sugeriu uma teoria evolucionária no “uso e desuso” de órgãos e de Thomas Robert Malthus [Henriques 2007] e [Landry 2010], que propôs que fatores ambientais tais como doenças e carência de alimentos, limitavam o crescimento de uma população.

Após mais de 20 anos de observações e experimentos, Charles Darwin [Darwin 1859] apresentou sua teoria de evolução através de seleção natural, provocando uma revolução no mundo inteiro sobre evolução natural. Ele defendia que o homem, assim como os outros seres vivos, era resultado da evolução.

Conforme [Brun e Opazo 2009], Darwin concluiu que nem todos os organismos nascem, sobrevivem e reproduzem-se. Os indivíduos mais propensos à sobrevivência são aqueles mais adaptados para enfrentar determinadas condições ambientais. Logo estes indivíduos teriam

maior chance de se reproduzir e assim deixar seus descendentes. Dessa forma, com o passar dos anos, as variações favoráveis tendem a permanecer e as desfavoráveis tendem a serem suprimidas.

O trabalho de Darwin influenciou muito nos anos seguintes não apenas a Biologia, Botânica e Zoologia, mas também teve grande influência sobre o pensamento religioso, filosófico, político e econômico da época. A teoria da evolução e a computação nasceram praticamente na mesma época: Charles Babbage, um dos fundadores da computação moderna e amigo pessoal de Darwin desenvolveu sua máquina analítica em 1837. Ambos provavelmente estariam surpresos e orgulhosos com a ligação entre estas duas áreas [Carvalho 2010].

Em meados do século 19, Gregor Mendel desenvolveu seus primeiros estudos sobre herança genética, o qual mais tarde foi redescoberto por cientistas e teve grande influência sobre os futuros trabalhos que estariam relacionados à evolução.

Após simulações e importantes teorias criadas à respeito da genética e evolução, [Holland 1975] começou a desenvolver as primeiras pesquisas no assunto, sendo o criador da “Bíblia dos Algoritmos Genéticos”. Este trabalho veio a ser corroborado mais tarde por [Goldberg 1989].

### **3.3 Estrutura de um Algoritmo Genético**

Embora existam muitas variantes possíveis do algoritmo genético básico, o mecanismo fundamental opera em uma população de cromossomos ou indivíduos que representam possíveis soluções para o problema. Uma implementação de um algoritmo genético começa com uma população aleatória de cromossomos. Esses elementos são, então, avaliados e associados a uma probabilidade de reprodução de tal forma que as maiores probabilidades são associadas aos cromossomos que representam uma melhor solução para o problema de otimização do que àqueles que representam uma solução pior. A aptidão da solução é tipicamente definida com relação à população corrente [Miranda 2010].

O fluxograma da Figura 3.1 apresenta brevemente o funcionamento de um AG [Cortes 1999].

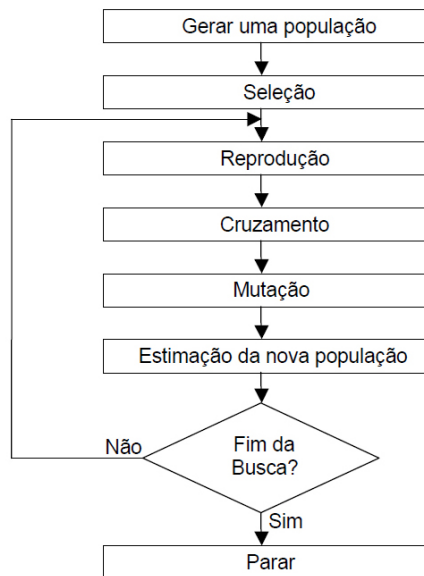


Figura 3.1: Fluxograma que descreve um AG

### 3.4 Representa o dos Indiv duos

A representa o (ou codifica o) dos cromossomos   a forma como os indiv duos ser o representados.   a primeira preocupa o a ser resolvida ao desenvolver uma solu o envolvendo AGs que dependem fortemente do problema a ser resolvido.

Existem 3 abordagens distintas utilizadas para realizar o processo de codifica o dos AGs. A codifica o cl ssica (ou bin ria) que trabalha com cadeias de bits, a codifica o real que trabalha diretamente com valores de ponto flutuante e a codifica o inteira, que trabalha apenas com valores inteiros, por isso menos aplicada.

A fun o *fitness*, fun o objetivo ou aptid o dos cromossomos   a fun o que representa a capacidade do indiv duo como poss vel solu o para resolver o problema. A aptid o do indiv duo   determinada atrav s do c lculo da fun o de acordo com as especifica es de projeto e do problema a ser resolvido, onde os indiv duos com valor de *fitness* possuem maiores chances de se reproduzir [Carvalho 2010].

Um cromossomo representa um conjunto de par metros da fun o objetivo. Por exemplo: se a fun o objetivo fosse  $f(x) = x \sin(10\pi x) + 1$ , um cromossomo codificaria um valor do par metro  $x$ .

[Catarina e Bach 2003] utilizaram AGs para encontrar a melhor solu o para uma determi-

nada função. Os autores utilizaram tanto a representação clássica quanto a codificação real para resolver o problema. O resultado obtido foi a solução ótima para ambas as representações, porém os algoritmos que estavam aplicando codificação real apresentaram a solução ótima em menos tempo. Isto se deve ao consumo de tempo utilizado pela codificação binária para calcular o valor numérico da cadeia de bits que representa tal valor.

### **3.4.1 Representação Binária**

É a codificação mais comum devido à sua simplicidade. Cada cromossomo é representado por uma *string* de bits formada por 0's e 1's. Porém devido à necessidade de maior precisão numérica, as cadeias de bits tornam-se longas, aumentando o esforço computacional e o consumo de tempo para convergência do algoritmo, além da necessidade de conversão das cadeias binárias para um valor numérico.

[Herrera, Lozano e Verdegay 1998] cita que devido a este problema, novas técnicas de codificação dos cromossomos surgiram, diferentes da técnica clássica apresentada por [Holland 1975]. Uma das soluções propostas para essa melhora de desempenho foi a utilização de valores reais para representar os cromossomos, técnica esta que será apresentada na próxima seção deste trabalho e é conhecida como codificação real.

### **3.4.2 Representação Real**

A codificação real é uma abordagem encontrada para suprir as desvantagens encontradas na codificação clássica, como precisão numérica, tempo de convergência do algoritmo, esforço computacional, etc.

[Herrera, Lozano e Verdegay 1998] afirmam que o uso de parâmetros reais torna possível cobrir um espaço de busca bastante abrangente, mesmo para domínios desconhecidos das variáveis, o que é difícil de se conseguir trabalhando com as cadeias binárias, onde conforme há aumento do domínio do problema há perda de precisão.

Outra vantagem é que a utilização da codificação real traz uma solução mais próxima da ótima, sendo que a codificação binária pode trazer uma convergência prematura do algoritmo, principalmente quando o problema é de alta dimensionalidade.

## 3.5 Seleção

Segundo [Schneider 1998] a ideia principal do operador de seleção, em um algoritmo genético, é oferecer aos melhores indivíduos da população corrente preferência para o processo de reprodução, permitindo que estes indivíduos possam passar as suas características às próximas gerações. Isto funciona como na natureza, onde os indivíduos mais adaptados ao seu ambiente possuem mais oportunidades para reproduzir do que aqueles indivíduos considerados mais fracos ou inaptos.

Nesta fase, os indivíduos que se mostraram mais aptos da geração atual são utilizados para gerar uma nova população através do cruzamento, ao menos teoricamente, visto que alguns métodos não necessariamente selecionam apenas os melhores elementos de cada geração. O processo de seleção é realizado, geralmente, através de três mecanismos bastante conhecidos:

- a) amostragem direta: o conjunto de indivíduos representantes da geração é selecionado baseados em um critério fixo. Pode ser visto como escolha “a dedo” ou “os  $n$  melhores”;
- b) amostragem aleatória simples ou equiprovável: todos os elementos possuem a mesma chance de serem escolhidos para compor o conjunto que gerará a prole seguinte;
- c) amostragem estocástica: a chance de cada indivíduo ser selecionado é diretamente proporcional ao valor de seu *fitness*.

### 3.5.1 Método da Roleta (*Roulette Wheel*)

Esta abordagem, que também é conhecida como método de Monte Carlo é a estratégia mais aplicada. Segundo [Goldberg 1989], cada indivíduo é representado em uma roleta de acordo com sua aptidão. A roleta representa a soma das aptidões de todos os indivíduos da população e é girada o número de vezes necessárias à seleção a fim de que seja composta a população auxiliar que será utilizada na propagação da espécie. Os indivíduos com *fitness* maior possuem uma área maior na roleta, acarretando maior probabilidade de seleção.

O funcionamento do método da Roleta é descrito a seguir:

1. sorteia-se um número aleatório entre zero e a soma das aptidões;
2. o indivíduo equivalente a este número na roleta é selecionado;

3. os indivíduos com maior aptidão (maior *fitness*) possuem maior probabilidade de serem selecionados.

Para exemplificar este método, as Tabelas 3.1 e 3.2 mostram seis indivíduos gerados aleatoriamente, cada um com seu valor de *fitness* e *fitness* relativo. Este último é calculado pela divisão entre o *fitness* do indivíduo pela soma total de todas as aptidões. A Tabela 3.1 mostra um exemplo com codificação binária e a Tabela 3.2 mostra um exemplo com codificação real.

Com os valores de *fitness* relativo obtidos na quarta coluna da Tabela 3.1 e terceira coluna da Tabela 3.2, é possível construir a roleta como ilustra a Figura 3.2.

Tabela 3.1: Representação dos indivíduos pela codificação binária para seleção do método da roleta

Indivíduo	String	Fitness	Fitness Relativo (%)
$S_1$	11110	30	8,57
$S_2$	1001100	76	21,71
$S_3$	10101	21	6,0
$S_4$	10010100	148	42,29
$S_5$	1000011	67	19,14
$S_6$	1000	8	2,29
Total		350	100

Tabela 3.2: Representação dos indivíduos pela codificação real para seleção do método da roleta

Indivíduo	Fitness	Fitness Relativo (%)
$S_1$	30	8,57
$S_2$	76	21,71
$S_3$	21	6,0
$S_4$	148	42,29
$S_5$	67	19,14
$S_6$	8	2,29
Total	350	100

### 3.5.2 Seleção por Torneio

Na seleção por torneio uma série de indivíduos da população é selecionada. Logo após estes entram em competição, comparando seus valores de aptidão onde o melhor deles é selecionado.

A Tabela 3.3 mostra um exemplo de funcionamento da seleção por torneio: se o número de indivíduos escolhidos fosse 3 e esses indivíduos fossem  $S_1$ ,  $S_4$  e  $S_5$ , o cromossomo a ser escolhido seria o indivíduo  $S_4$ , pois este apresenta-se mais apto à solução do problema.

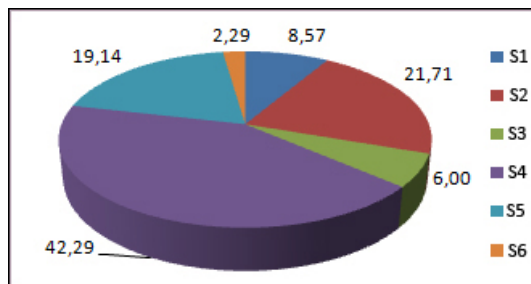


Figura 3.2: Método da roleta com base nos indivíduos das Tabelas 3.1 e 3.2

Tabela 3.3: Seleção por Torneio

Invidíduo	Fitness
$S_1$	2,35
$S_2$	6,98
$S_3$	14,2
$S_4$	20,43
$S_5$	0,71
$S_6$	10,5

Este método traz a possibilidade do pior indivíduo ser selecionado para participar do processo de mutação ou cruzamento e ser o único competidor de um torneio, porém, quanto maior o tamanho do torneio, maior a probabilidade do melhor indivíduo ser escolhido [Lacerda 2009].

### 3.5.3 Seleção por Rank

O método da seleção por Rank evita a convergência prematura ou a dominância de um “super” indivíduo. Geralmente ocorre quando surgem indivíduos de alta aptidão, mas os indivíduos realmente ótimos ainda não estão na população [Lacerda 2009].

O princípio básico deste método é ordenar todos os elementos de acordo com a sua função *fitness*, utilizar este ranking como base da seleção e o número de vezes que cada indivíduo será selecionado depende de sua posição em relação aos outros.

Desvantagem: ser muito lento para ordenar os indivíduos quando a população é muito grande.

## 3.6 Cruzamento

O processo de cruzamento (ou *crossover*) é a etapa onde ocorre a troca de material genético (propagando as características) entre os “pares” de cromossomos que irão gerar novos indivíduos para as próximas populações [Brun 2005].

O operador de cruzamento é geralmente aplicado somente com uma dada probabilidade, também chamada de taxa de cruzamento. Essa taxa de cruzamento é normalmente alta (entre 60% e 90%) e durante a aplicação do operador, é gerado um número aleatório  $r$  entre 0 e 1 para aplicar o seguinte teste:

- a) se o valor sorteado for menor que a taxa de *crossover*, então operador é aplicado aos cromossomos pais para gerar um novo filho;
- b) senão, os filhos tornam-se iguais aos pais para permitir que algumas boas soluções sejam preservadas.

### 3.6.1 Representação Binária

Na codificação binária o cruzamento é feito com base na troca de segmentos dos cromossomos. Esta troca pode ser feita de duas formas: dividindo-se em um ponto ou em multiponto.

#### Cruzamento com um Ponto

Também chamado de *single-point crossover*, ocorre quando o ponto onde é realizada a quebra é escolhido de forma aleatória, ou a critério do projetista, dentro da string que o representa. Baseado neste ponto realiza-se a troca de material cromossômico entre os dois indivíduos.

Estes indivíduos são recombinações de dois cromossomos pais, como por exemplo: Supondo-se que o tamanho dos cromossomos seja de 10 bits e que o ponto de cruzamento escolhido seja a posição 4 dos cromossomos pais. Ambos os pais são divididos em duas partes e cada cromossomo filho fica com uma das partes de cada pai: o filho 1 fica com a primeira parte do pai 1 e a segunda parte do pai 2 enquanto o segundo filho fica com a segunda parte do pai 1 e a primeira parte do pai 2. A Figura 3.4 mostra esse exemplo.



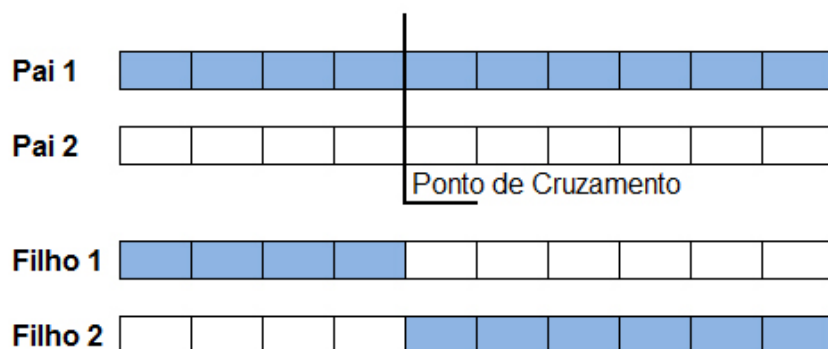


Figura 3.3: Troca de segmentos com um ponto de cruzamento

### Cruzamento Multiponto

Cruzamento Multiponto ou *multi-point crossover* é realizado similarmente ao cruzamento de um ponto, porém a troca de segmentos é realizada com um número maior de intercalações.

Considere-se que os pontos de cruzamento escolhidos são as posições 2 e 7 dos cromossomos pais. Ambos são divididos em 3 partes e, de forma intercalada, cada filho fica com uma sequência dos cromossomos pais, como mostra a Figura 3.4.

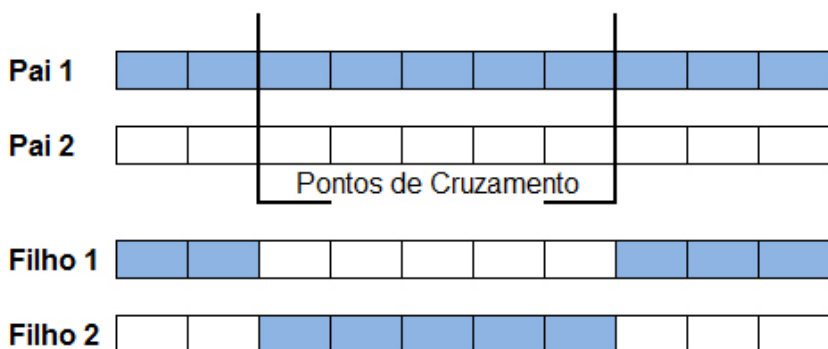


Figura 3.4: Troca de segmentos com cruzamento multiponto

### 3.6.2 Representação Real

O cruzamento com codificação real dos cromossomos trata o valor do gene de forma direta, sem nenhuma forma de codificação. A técnica mais conhecida para realizar cruzamento

entre cromossomos na codificação real é através de operadores aritméticos da média ou média geométrica [Brun 2005].

Os cruzamentos que utilizam média e média geométrica geram um novo cromossomo usando a média simples e a média geométrica de dois cromossomos pais através das Equações 3.1 e 3.2.

$$C_1 = \frac{P_1 + P_2}{2} \quad (3.1)$$

$$C_1 = \sqrt{P_1^2 * P_2^2} \quad (3.2)$$

Operadores de média tendem a diminuir muito a diversidade dos filhos pois estes estarão sempre no meio do intervalo dos pais. Para evitar esse possível problema outras soluções foram propostas:  $BLX - \alpha$ , aritmético e heurístico.

Segundo [Catarina 2005], o cruzamento  $BLX - \alpha$  consiste em gerar um novo cromossomo a partir da Equação 3.3.

$$C = P_1 + \beta(P_2 - P_1) \quad (3.3)$$

onde  $C$  é o cromossomo gerado,  $P_1$  e  $P_2$  são os cromossomos pais,  $\beta$  é um valor sorteado aleatoriamente que deve estar compreendido entre  $-\alpha$  e  $1 + \alpha$  e  $\alpha$  é um pequeno valor que estende os limites para a definição de  $C$ . Caso o cromossomo seja formado por múltiplos genes, a Equação 3.3 é aplicada a cada par de genes de  $P_1$  e  $P_2$ .

O cruzamento aritmético é realizado através da aplicação das Equações 3.4 e 3.5.

$$C_1 = \beta * P_1 + (1 - \beta) * P_2 \quad (3.4)$$

$$C_2 = (1 - \beta) * P_1 + \beta * P_2 \quad (3.5)$$

onde  $C_1$  e  $C_2$  são os cromossomos filhos,  $P_1$  e  $P_2$  são os cromossomos pais e  $\beta \in U(0, 1)$ .

No cruzamento heurístico é necessário ter conhecimento do valor da função *fitness* dos pais. Este método consiste em gerar um cromossomo filho a partir de uma interpolação linear entre os pais usando a informação da aptidão. O cruzamento heurístico favorece o pai com maior

*fitness*. Dados dois cromossomos pais  $P_1$  e  $P_2$ , onde o cromossomo  $P_1$  possui o valor de *fitness* superior ao de  $P_2$ . O novo cromossomo é gerado a partir da Equação 3.6.

$$C = P_1 + r(P_1 - P_2) \quad (3.6)$$

onde o valor de  $r$  varia de 0 a 1 e  $f(P_1) > f(P_2)$ .

## 3.7 Mutaç o

A principal funç o do operador de mutaç o   permitir que todo o espaço de busca tenha chance de ser atingido. A mutaç o altera arbitrariamente um ou mais genes de um cromossomo selecionado procurando aumentar a variabilidade estrutural da populaç o.

O papel da mutaç o nos Algoritmos Gen ticos   restaurar material gen tico perdido ou inexplorado da populaç o para evitar converg ncia prematura do AG para soluç es que n o s o o resultado esperado. A posiç o de cada cromossomo na populaç o pode sofrer uma altera o aleat ria de acordo com uma probabilidade definida por uma taxa de mutaç o ou probabilidade de mutaç o.

Deve-se ressaltar que ap s os processos de cruzamento e mutaç o, uma estrat gia adicional de seleç o pode ser adotada, chamada de estrat gia elitista ou elitismo [Jong 1975]. Esta estrat gia garante que o melhor cromossomo sempre permanece inalterado de uma geraç o para outra para evitar que esse cromossomo desapareça (seç o 3.8).

Taxas de mutaç o devem ser pequenas, geralmente menores que 5%, pois devem ser apenas o suficiente para aumentar a diversidade da populaç o.

### 3.7.1 Representa o Bin ria

Seu objetivo   manter a variabilidade da populaç o de indiv duos. Analisa-se cada gene e, atrav s de uma probabilidade pr -determinada, altera-se o valor deste gene. A Figura 3.5 mostra um exemplo de mutaç o que utiliza codifica o bin ria, onde o 4  e o 7  genes foram alterados.

### 3.7.2 Representa o Real

Segundo [Catarina 2005], existem diversas formas de se realizar a mutaç o quando a codifica o real est  sendo utilizada, tais como mutaç o uniforme, mutaç o gaussiana, mutaç o

<b>Antes</b>	1	0	1	0	0	1	1	1	0	0
<b>Depois</b>	1	0	1	1	0	1	0	1	0	0

Figura 3.5: Exemplo de mutação utilizando codificação binária

*creep*, mutação não uniforme e não-uniforme múltipla. Cada uma dessas formas trabalha de acordo com uma regra específica e previamente elaborada.

A mutação uniforme (ou randômica) trabalha com valores obtidos aleatoriamente de uma distribuição uniforme através de um intervalo determinado de acordo com o máximo ou mínimo permitidos. Ela realiza teste de mutação para cada parâmetro codificado e substitui esses parâmetros por outros escolhidos uniformemente em um intervalo  $[a; b]$ .

A mutação gaussiana realiza teste de mutação para cada parâmetro codificado e substitui o parâmetro por um número aleatório escolhido usando uma distribuição normal  $N(\mu; \sigma^2)$  onde  $\mu$  é igual ao valor do gene a ser substituído e a variância é definida pelo próprio projetista.

A mutação *creep* acrescenta ou subtrai um pequeno número aleatório obtido pela distribuição  $N(0; \sigma^2)$ . Pode também multiplicar o valor atual por número próximo de 1. Busca uma solução próxima através de ajustes aleatórios em ambas as direções. Este método é comumente utilizado para explorar o espaço de busca localmente [Brun 2005].

A mutação não-uniforme consiste na simples substituição de um gene por um número extraído de uma distribuição não-uniforme, onde os valores possuem uma probabilidade diferente de ocorrerem [Herrera, Lozano e Verdegay 1998].

A mutação não-uniforme múltipla aplica a mutação não-uniforme em todos os genes do cromossomo selecionado.

### 3.8 Elitismo

Elitismo é uma técnica usada para melhorar o desempenho e convergência do algoritmo, fazendo com que este venha a convergir mais cedo. Ela preserva os indivíduos considerados ótimos ou próximos da solução esperada sem permitir que estes sofram cruzamento ou mutação. Para tal feito, esses indivíduos são passados para a próxima geração garantindo que não sejam

perdidos ou que sofram algum tipo de alteração.

[Miranda 2010] ressalta que na prática isto resulta numa busca mais agressiva e, geralmente, é bastante efetiva. No entanto existe o perigo de uma convergência prematura para mínimos locais, isto é, cada indivíduo selecionado e cruzado com seu parceiro é colocado no lugar do pior indivíduo da população anterior.

Vale lembrar que o uso do elitismo é indicado com poucos elementos, pois se o número de elementos for muito grande, o algoritmo pode convergir para um máximo local.

## **3.9 Parâmetros Genéticos**

A configuração dos parâmetros genéticos utilizados no AG é um dos fatores mais importantes na implementação, pois podem influenciar muito o resultado final e interagem um com o outro de maneira não-linear, ou seja, não se pode otimizar um independentemente de outro. São eles: tamanho da população, taxa de cruzamento, taxa de mutação e intervalo de cruzamento.

### **3.9.1 Tamanho da População**

Determina o número de cromossomos de cada população, afeta o desempenho global e a eficiência do AG [Catarina e Bach 2003]. Uma população menor pode cobrir uma pequena parcela do espaço de busca, tem baixo custo computacional e pode gerar convergência prematura. Populações grandes oferecem uma cobertura representativa do domínio do problema e ainda evitam o problema de uma convergência prematura para soluções locais. Utilizar grandes populações, porém, pode aumentar o custo computacional e o tempo para chegar ao resultado.

### **3.9.2 Taxa de Cruzamento**

Indica a probabilidade de ocorrência de um cruzamento. Quanto maior, maior a diversidade em cada geração. Quando a taxa de cruzamento é muito baixa a convergência para uma solução torna-se mais demorada, por outro lado, quando a taxa de cruzamento é muito alta pode ocorrer a perda de bons indivíduos.

### **3.9.3 Taxa de Mutação**

Determina a probabilidade de ocorrência de mutação nos cromossomos. Valor baixo para taxa de mutação aumenta o espaço de busca, além de prevenir a convergência prematura, porém, se for muito pequeno, pode entrar em estagnação. Valor alto para taxa de mutação torna a busca aleatória além de aumentar muito a chance de que uma boa solução do problema seja destruída.

### **3.9.4 Intervalo de Cruzamento**

Controla a porcentagem de indivíduos que será substituída em cada geração. Para valores baixos a busca torna-se muito lenta, necessitando de um número maior de gerações e para valores muito altos acarretará na perda de indivíduos de alta aptidão.

# Capítulo 4

## Metodologia

### 4.1 Materiais e Métodos

O software proposto, nomeado *SEMAG* (**SEM**ivariogramas e **Algoritmos G**enéticos), foi desenvolvido na linguagem de programação (LP) Java, que é uma linguagem orientada a objeto (OO) e multiplataforma criada pela empresa *Sun Microsystems*. Aliada à linguagem, a IDE (Integrated Development Environment - Ambiente de Desenvolvimento Integrado) *NetBeans* também foi utilizada, pois facilita uma boa programação OO, auxilia na forma abstrata de se construir as regras, possibilita a criatividade por parte do desenvolvedor e, principalmente, atende às necessidades requeridas para o desenvolvimento deste trabalho.

A ferramenta construída permite a construção de semivariogramas experimentais com base em dados de entrada e, posteriormente, a escolha dos parâmetros de Alcance, Efeito Pepita e Contribuição de forma a permitir um processo estimativo eficiente e confiável.

A primeira etapa do processo é a leitura dos dados de entrada, os quais possuem as coordenadas geográficas  $X$ ,  $Y$  e a propriedade observada  $Z$  neste ponto. Para isso, a biblioteca *Java Excel API* foi incorporada à IDE *NetBeans*, a qual possibilita a criação, leitura, escrita e edição de planilhas do *Microsoft Excel*. Explicações de uso desta biblioteca estão descritas no Apêndice A deste documento. A opção de escolha de arquivo com dados de entrada pode ser vista na Figura 4.1.

Outra biblioteca adicional foi utilizada para criação dos gráficos mostrados durante a execução do programa. Essa biblioteca é a *JFreeChart* e, assim como a biblioteca citada anteriormente, também é livre para uso (ver Apêndice B). Essa é uma ferramenta muito poderosa para construção de gráficos em Java. Basicamente, utiliza as classes *DataSet* que armazena os pontos

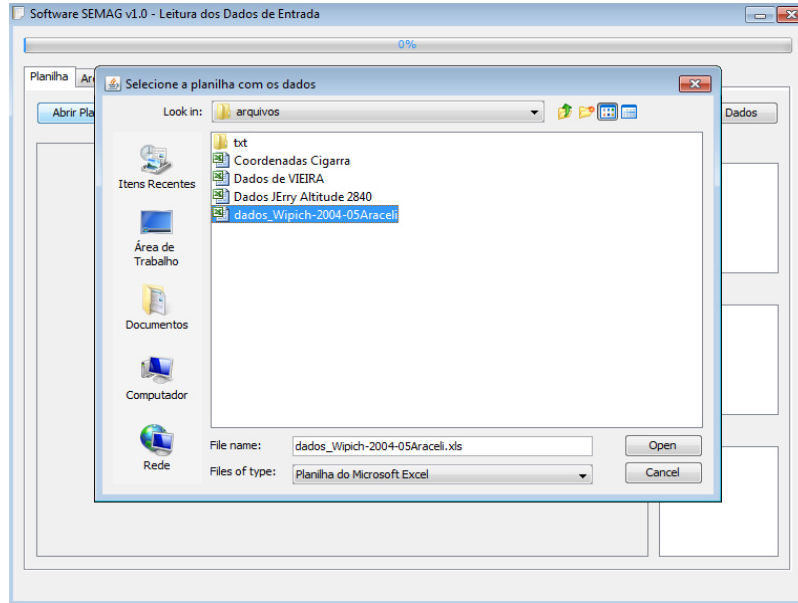


Figura 4.1: Leitura dos dados de entrada

que serão amostrados na tela e *Plot* que define as propriedades do gráfico (cores, imagens de fundo, efeitos...). A *JFreeChart* gerencia o desenho do gráfico (ver Figura 4.2).

Após a leitura dos pontos de entrada, são calculadas as distâncias  $h$  entre cada par de coordenadas, o número de ocorrências de cada distância ( $N(h)$ ) (Figura 4.3) e as funções de semivariâncias para os estimadores Matheron, Cressie e Hawkins e Pairwise, de acordo com as Equações 2.2, 2.3 e 2.4.

O usuário tem a liberdade de escolher o valor da distância máxima de estimação da semi-variância (*Cut-off*) na tela. Pontos posicionados além desta distância são considerados não influentes. Este valor é dado em porcentagem, assim, se o usuário escolher 80% (como na Figura 4.2), a distância máxima considerada será 80% da maior distância encontrada entre os pares de coordenadas.

O número de comparações para encontrar  $N(h)$  é uma combinação do total de pontos amostrais tomados 2 a 2, conforme a equação 4.1.

$$C_N^S = \frac{N!}{S!(N-S)!} \quad (4.1)$$

onde:

$N$  representa o total de pontos amostrais;



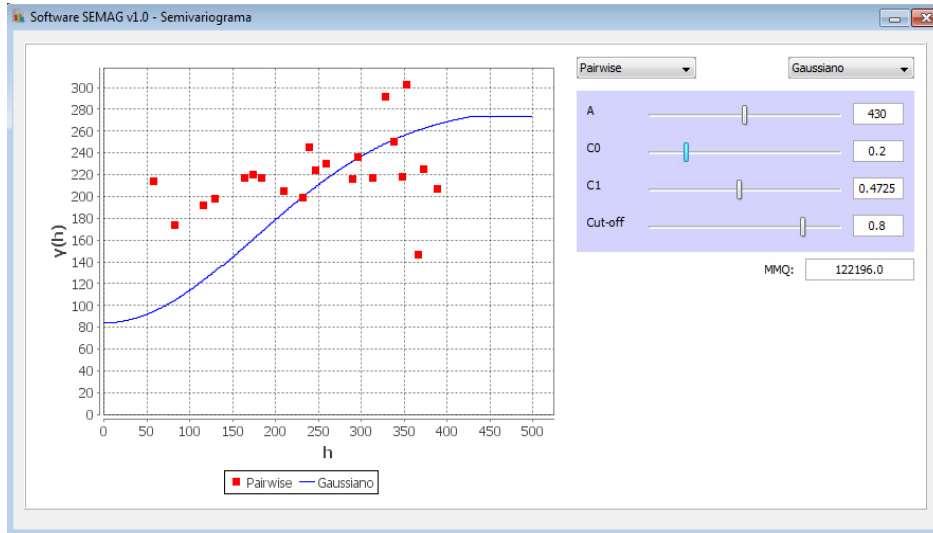


Figura 4.2: Biblioteca *JFreeChart* utilizada para montagem dos gráficos

$$S = 2;$$

logo:

$$\begin{aligned}
 C_N^S &= \frac{N!}{2!(N-S)!} \\
 &= \frac{N * (N-1) * (N-2)!}{2!(N-2)!} \\
 &= \frac{N * (N-1)}{2!} \\
 &= \frac{N^2 - N}{2} \\
 &= \frac{N^2}{2} - \frac{N}{2}
 \end{aligned}$$



Figura 4.3: Exemplo do cálculo das distâncias  $h$  e  $N(h)$

Os modelos matemáticos implementados neste trabalho foram: Linear, Circular, Esférico, Exponencial e Gaussiano (como mostram as Figuras 2.2, 2.3, 2.4, 2.5 e 2.6, respectivamente). A

aplicação desses modelos visa realizar o ajuste da função de semivariância (pontos amostrados) com a curva do modelo escolhido.

O cálculo dos modelos é feito como segue:

- a) Percorrem-se todos os pontos horizontais da tela de 0 (zero) até o máximo valor no eixo  $X$ ;
- b) Para cada valor  $X$ , um correspondente  $Y$  é armazenado de acordo com a escolha dos parâmetros e a equação do modelo matemático;
- c) As tuplas  $(X, Y)$  são plotadas no gráfico e armazenadas, formando a curva do modelo escolhido a ser aplicado posteriormente no ajuste.

A escolha do modelo e o estimador da função não é feita aleatoriamente, haja vista que os resultados provenientes desta escolha podem não ser os mais indicados e esperados. Para isto, o modelo deve ser o mais próximo possível com a distribuição dos pontos.

Para todos os testes realizados, os valores de  $\gamma(h)$  foram muito pequenos, fazendo com que a visualização destes pontos no gráfico se concentrasse apenas nas regiões inferiores do eixo  $y$  (próximos de zero) quando comparadas à disposição dos modelos matemáticos na tela do programa, impossibilitando assim, visualização e interpretação da função de variância construída. Visando contornar este percalço, aplicou-se uma normalização para obter uma melhor representação visual das informações.

O ajuste desta distribuição foi feito através da proporção e redistribuição destes pontos nos eixos  $x$  e  $y$  para coordenadas de tela. Para isso, durante o cálculo das semivariâncias e dos modelos matemáticos, os maiores valores de  $x$  e  $y$  foram armazenados e o cálculo das novas coordenadas ficaram de acordo com as equações 4.2 e 4.3.

$$Coord_X = \left( \frac{U_{max}}{X_{max}} \right) * P_x \quad (4.2)$$

$$Coord_Y = \left( \frac{V_{max}}{Y_{max}} \right) * P_y \quad (4.3)$$

onde:

$U_{max}$  é o maior valor de  $X$  da tela;

$X_{max}$  é o maior valor de X do método de estimação ou modelo matemático;

$V_{max}$  é o maior valor de Y da tela;

$Y_{max}$  é o maior valor de Y do método de estimação ou modelo matemático;

$P_x$  e  $P_y$  são os pontos a serem mapeados para coordenada de tela.

O comportamento da curva de ajuste é influenciado diretamente pelos valores dos parâmetros do semivariograma. Dessa forma, desenvolveu-se um Algoritmo Genético capaz de determinar automaticamente tais parâmetros, procurando encontrar o melhor ajuste entre a função de semivariância obtido através do estimador e a curva construída com base no modelo matemático.

No contexto do trabalho, os genes que compõem os indivíduos que formam as possíveis soluções para o ajuste do semivariograma são os parâmetros buscados, fazendo com que cada cromossomo possua 4 genes: Alcance ( $a$ ), Efeito Pepita ( $C_0$ ), Contribuição ( $C_1$ ) e o valor de *fitness* ( $F$ ) associado. Dadas as características das informações (valores com ponto flutuante e com necessidade de grande precisão), a abordagem adotada na construção do AG foi a codificação real.

Durante cada operação de cruzamento ou mutação, independente do método escolhido, as operações entre cromossomos são feitas nos mesmos parâmetros, ou seja, o valor de alcance de um cromossomo pai com o valor de alcance do outro cromossomo pai, assim como os demais parâmetros. A Figura 4.4 ilustra a estrutura dos indivíduos que compõem o AG, onde o campo demarcado por “A” armazena o valor referente ao Alcance, os campos “ $C_0$ ” e “ $C_1$ ” armazenam os valores de Efeito Pepita e Contribuição, respectivamente, e o campo “F” armazena o valor de *fitness* de cada um dos cromossomos.

Para o cálculo da aptidão dos indivíduos, optou-se por utilizar o método dos mínimos quadrados (MMQ), pois minimiza o erro resultante, é comumente utilizado em ajuste de curvas e consiste em comparar a diferença entre dois conjuntos de dados ([Fonte 1994]). Este método está descrito na Equação 4.4.

$$S = \sum_{i=1}^n (y_i - f(x_i))^2 \quad (4.4)$$

O *fitness* de cada cromossomo é calculado percorrendo todas as distâncias  $h$  armazenadas na etapa da leitura dos dados de entrada, atribuindo  $\gamma(h)$  à variável  $y_i$  e  $Y(h)$  à variável  $f(x_i)$  da

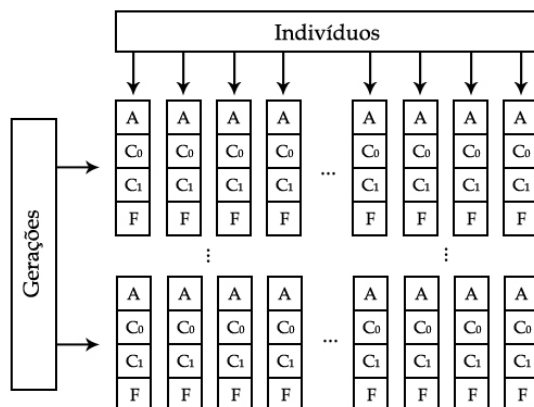


Figura 4.4: Estrutura do cromossomo implementado no AG

Equação 4.4, onde  $Y(h)$  é o valor armazenado no momento do cálculo do modelo matemático (Figura 4.5). Assim o *fitness* de cada elemento depende diretamente da qualidade do ajuste do modelo matemático ao semivariograma. Quanto melhor o ajuste, menor o MMQ e maior o *fitness*.

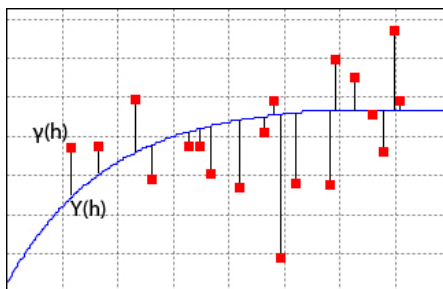


Figura 4.5: Cálculo da aptidão dos cromossomos utilizando o método dos mínimos quadrados

Após este processo, o valor de *fitness* armazenado é  $(100/\text{fitness})$ , pois aumenta a possibilidade de elementos com aptidão pequena serem escolhidos no processo de seleção do AG por serem possíveis soluções do problema, ou seja, quanto menor for o valor calculado, maior será o *fitness*.

O usuário tem total liberdade para configuração do AG, podendo determinar o tamanho da população, quantidade de gerações, taxa de mutação, taxa de cruzamento, porcentagem do elitismo e quais métodos de cruzamento, mutação e seleção deseja utilizar (ver Figura 4.6).

A primeiro procedimento executado pelo AG é a construção da população inicial. Para isso,

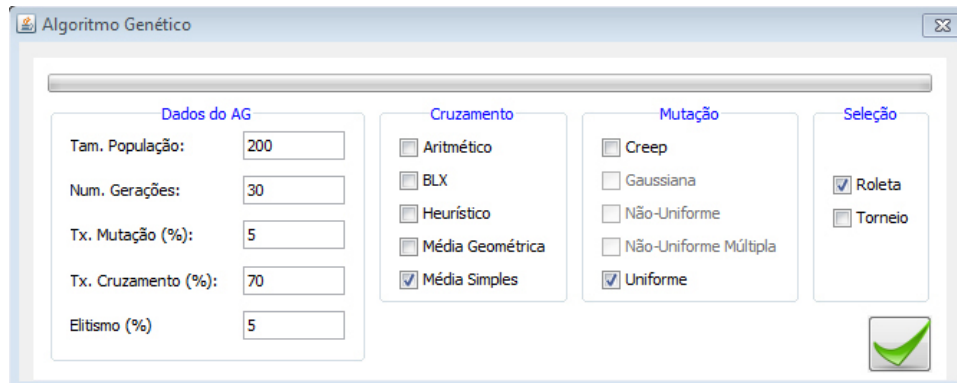


Figura 4.6: Tela de configuração do AG

os valores de Alcance, Efeito Pepita e Contribuição são atribuídos randomicamente e, após testes realizados, concluiu-se que estes valores trazem resultados mais precisos variando de 0 a 500 para o primeiro parâmetro e 0 a 1 para os demais.

Após este processo, uma quantidade de indivíduos determinada pela taxa de elitismo é automaticamente passada para a próxima geração sem sofrer nenhum tipo de alteração em seus genes. Estes indivíduos não são retirados da população e podem ser escolhidos pelo método de seleção.

Enquanto a nova população não estiver completa, o algoritmo utiliza o método de seleção (roleta ou torneio) escolhido pelo usuário para selecionar dois cromossomos pais e geram novo(s) filho(s) pelo cruzamento (também escolhido pelo usuário). Na maioria das vezes, o AG tem comportamento melhor com taxas de cruzamento entre 60% e 70%, fazendo com que o AG não fique muito lento ou que elementos com características consideradas boas fossem perdidos, no caso de uma taxa muito alta. O cruzamento só é realizado quando o valor sorteado for menor do que a taxa de cruzamento, caso contrário, o pai com maior aptidão é clonado e enviado à próxima geração.

Essa próxima geração é então submetida à etapa de mutação. Similarmente ao cruzamento, um valor é sorteado e a mutação é realizada quando este valor for inferior à taxa de mutação e, caso contrário, o cromossomo não sofre nenhuma alteração em seu gene. É indicado que taxas de mutação sejam relativamente baixas (geralmente menores que 8%), pois são mais conservadoras e evitam modificação na maioria dos elementos ou possível geração de um algoritmo aleatório. Após a etapa de mutação, os indivíduos tem seu *fitness* calculado através da Equação

#### 4.4

Ao fim da execução das  $N$  gerações, o elemento com melhor valor de *fitness* é escolhido como resposta para o problema. Clicando no botão em destaque (ver Figura 4.7), os valores de Alcance, Efeito Pepita e Contribuição são enviados à tela do Semivariograma e a curva é atualizada (Figura 4.9) mostrando o resultado final. O fluxograma da Figura 4.8 mostra o funcionamento do AG implementado.

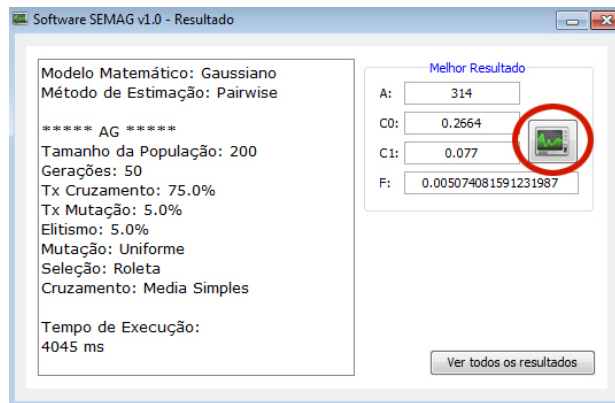


Figura 4.7: Resultado mostrado após execução do AG

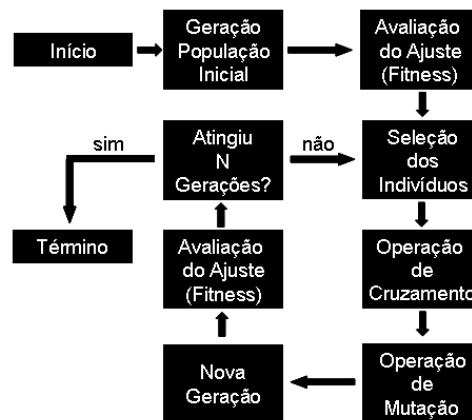


Figura 4.8: Fluxograma do AG Implementado.

Depois de obtido o melhor ajuste, o pesquisador pode aplicar os parâmetros encontrados para processos estimativos, como a Krigagem ordinária.

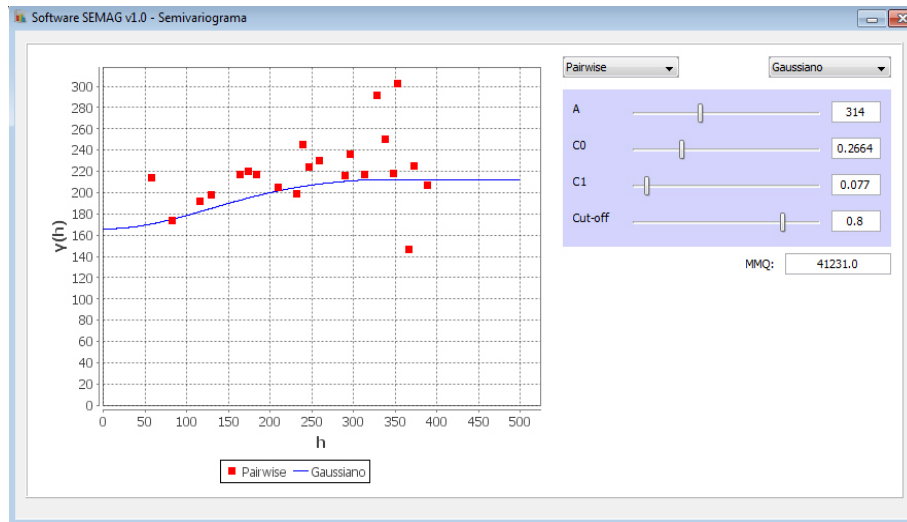


Figura 4.9: Resultado Final

## 4.2 Configuração dos Testes

Muitas são as combinações fornecidas pelo software implementado. Visando melhores resultados, foram criados ambientes de teste, mantidos parcialmente fixos durante cada iteração executada. Cada um desses ambientes é composto por um método de estimação, um modelo matemático e um método de mutação.

Após formados, esses ambientes foram associados à todos os operadores de cruzamento, um a um, para depois iniciar o teste. Cada teste foi executado doze vezes garantindo boa confiança estatística nos resultados e conclusões, sendo considerados melhores aqueles que tem melhor eficiência no ajuste ao modelo teórico em um intervalo menor de tempo.

A Figura 4.10 esboça como é dado o funcionamento dos testes.

Toda execução do AG é armazenada pelo software e posteriormente salva em planilha eletrônica para análise final dos resultados como mostra a Figura 4.11.

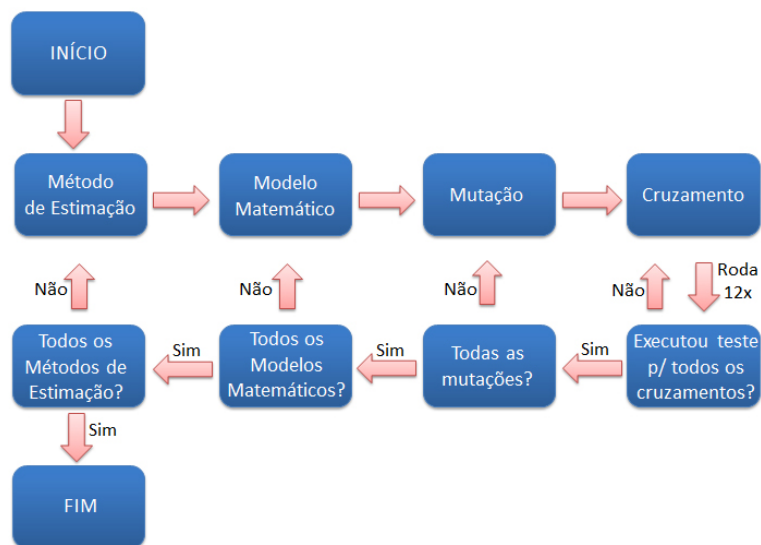


Figura 4.10: Configuração dos testes

Software SEMAG v1.0 - Resultado Sessão

	Estimador	Modelo	Cruzamento	Mutaçao	Tempo (ms)	Eficiência
1	Pairwise	Gaussiano	Media Simples	Creep	3822	5,049204497831367E-4
2	Pairwise	Gaussiano	Media Simples	Creep	3597	4,641469674957879E-4
3	Pairwise	Gaussiano	Media Simples	Creep	3457	4,144304689280756E-4
4	Pairwise	Gaussiano	Media Simples	Creep	3371	4,4606813244654986E-4
5	Pairwise	Gaussiano	Media Simples	Creep	3232	4,989820765638098E-4
6	Pairwise	Gaussiano	Media Geometrica	Creep	3485	3,6275914606497016E-4
7	Pairwise	Gaussiano	Media Geometrica	Creep	3455	5,087142755400002E-4
8	Pairwise	Gaussiano	Media Geometrica	Creep	3582	5,279217831086146E-4
9	Pairwise	Gaussiano	Media Geometrica	Creep	3551	3,184236754371161E-4
10	Pairwise	Gaussiano	Media Geometrica	Creep	3499	4,0416777812805654E-4
11	Pairwise	Gaussiano	Heuristico	Creep	3654	1,609932640418325E-4
12	Pairwise	Gaussiano	Heuristico	Creep	3344	5,712686162160309E-4
13	Pairwise	Gaussiano	Heuristico	Creep	3364	5,326685629667508E-4
14	Pairwise	Gaussiano	Heuristico	Creep	3326	3,9251859556846505E-4
15	Pairwise	Gaussiano	BLX	Creep	3711	5,507790770044227E-4
16	Pairwise	Gaussiano	BLX	Creep	3871	5,49004106550717E-4
17	Pairwise	Gaussiano	BLX	Creep	3901	5,594906397215975E-4
18	Pairwise	Gaussiano	BLX	Creep	3876	5,258093520451355E-4
19	Pairwise	Gaussiano	BLX	Creep	3924	5,463376257259461E-4
20	Pairwise	Gaussiano	Aritmetico	Creep	3033	4,846557974526491E-4
21	Pairwise	Gaussiano	Aritmetico	Creep	2772	4,8348184283939215E-4
22	Pairwise	Gaussiano	Aritmetico	Creep	2747	5,618892965707896E-4
23	Pairwise	Gaussiano	Aritmetico	Creep	2807	4,860550797616386E-4
24	Pairwise	Gaussiano	Aritmetico	Creep	2474	3,301528607745386E-4
25	Pairwise	Gaussiano	Aritmetico	Creep	2727	4,7292056826135485E-4
26	Pairwise	Gaussiano	Aritmetico	Creep	3189	4,7270595798589446E-4

Salvar Tabela

Figura 4.11: Resultados do AG salvos durante execuçao do programa



# Capítulo 5

## Resultados e Discussões

### 5.1 Resultados

O software construído possui uma interface amigável e de fácil operação, permitindo ao usuário escolher os vários parâmetros para configuração do processo a ser executado (semivariograma, modelo matemático, taxa de mutação e cruzamento, tamanho da população e número de gerações).

Durante a realização dos testes, verificou-se que o sistema mostrou-se eficiente para determinação dos parâmetros utilizando quaisquer combinações de modelos matemáticos e de estimativas, bem como combinações distintas de métodos de seleção, cruzamento e mutação, sendo o método de cruzamento Aritmético o mais rápido entre todos os demais métodos.

O computador utilizado para obtenção de tais resultados possui a configuração mostrada pela Tabela 5.1.

Tabela 5.1: Configuração do computador utilizado para os testes

<b>Processador</b>	AMD Athlon (tm) 64 X2 <i>Dual Core Processor</i> 5000+ 2.60 GHz
<b>Memória</b>	2 GB DDR2
<b>HD</b>	120 GB
<b>cache L1</b>	64 KB
<b>cache L2</b>	512 KB
<b>Sistema Operacional</b>	Windows 7 Professional

Após realizada a construção do software e executados os testes, verificou-se que a ferramenta obtida permite a obtenção dos parâmetros ( $C_0$ ,  $C_1$  e  $A$ ) de forma satisfatória, rápida e eficiente. O uso de um AG pode auxiliar na aplicação e desenvolvimento de métodos geostatísticos nos estudos da estrutura de dependência espacial.

Em todos os testes realizados, observou-se proximidade nos resultados quanto à eficiência do algoritmo, tornando-o apto e certamente eficiente apenas quanto ao tempo de execução do mesmo. As Tabelas 5.2, 5.3, 5.4, 5.5, 5.6 e 5.7 esboçam os resultados finais sendo o cruzamento aritmético predominantemente o melhor entre os cinco métodos de cruzamento estudados neste trabalho.

Tabela 5.2: Resultado para mutação *creep* e estimador Cressie

<b>Estimador</b>	<b>Modelo</b>	<b>Cruzamento</b>	<b>Tempo (ms)</b>
Cressie	Circular	Média Simples	3291,92
	Esférico	Aritmético	2763,67
	Exponencial	Aritmético	2131,83
	Gaussiano	Aritmético	2033,67
	Linear	Aritmético	1853,33

Tabela 5.3: Resultado para mutação *creep* e estimador Matheron

<b>Estimador</b>	<b>Modelo</b>	<b>Cruzamento</b>	<b>Tempo (ms)</b>
Matheron	Circular	Aritmético	3297,08
	Esférico	Aritmético	2684,00
	Exponencial	Aritmético	2107,50
	Gaussiano	Aritmético	2240,58
	Linear	Aritmético	1803,75

Tabela 5.4: Resultado para mutação *creep* e estimador Pairwise

<b>Estimador</b>	<b>Modelo</b>	<b>Cruzamento</b>	<b>Tempo (ms)</b>
Pairwise	Circular	Média Simples	2980,00
	Esférico	Média Simples	2453,92
	Exponencial	Aritmético	2147,25
	Gaussiano	Aritmético	2215,83
	Linear	Aritmético	1681,50

Tabela 5.5: Resultado para mutação uniforme e estimador Cressie

<b>Estimador</b>	<b>Modelo</b>	<b>Cruzamento</b>	<b>Tempo (ms)</b>
Cressie	Circular	Média Simples	2915,50
	Esférico	Média Simples	2522,00
	Exponencial	Aritmético	2112,92
	Gaussiano	Aritmético	2345,33
	Linear	Aritmético	1785,42

Tabela 5.6: Resultado para mutação uniforme e estimador Matheron

<b>Estimador</b>	<b>Modelo</b>	<b>Cruzamento</b>	<b>Tempo (ms)</b>
Matheron	Circular	Média Simples	3415,25
	Esférico	Aritmético	2828,42
	Exponencial	Aritmético	2360,67
	Gaussiano	Média Simples	2656,67
	Linear	Aritmético	2203,67

Tabela 5.7: Resultado para mutação uniforme e estimador Pairwise

<b>Estimador</b>	<b>Modelo</b>	<b>Cruzamento</b>	<b>Tempo (ms)</b>
Pairwise	Circular	Média Simples	2898,75
	Esférico	Aritmético	2553,08
	Exponencial	Aritmético	2007,58
	Gaussiano	Aritmético	2342,67
	Linear	Aritmético	1985,58

# Capítulo 6

## Conclusão

A qualidade de um software pode ser avaliada em parte pela funcionalidade oferecida por ele. Um software que possua uma gama considerável de opções e funções, tende a um escopo de aplicações maior e, conseqüentemente, mais procurado por usuários potenciais.

O software SEMAG mostrou-se bastante confiável e apto aos resultados, confirmando o fato de que a computação pode auxiliar na aplicação e desenvolvimento de ferramentas geostatísticas, possibilitando também a realização de estudos de dependência espacial e processos estimativos.

Espera-se, futuramente, realizar testes para verificação de qual configuração do Algoritmo Genético (método de seleção, cruzamento e mutação) é mais apropriada (em termos de consumo de tempo e eficiência de resultado) para a obtenção dos parâmetros do semivariograma.

Para novos trabalhos, seria interessante a implementação de outros métodos de estimação como o New 1 e o New 2 que reduzem a imprecisão do estimador quando a distância entre as amostras aumenta, bem como a implementação do mesmo software em outras linguagens de programação tais como *Delphi* e *C#*.

# Apêndice A

## Tutorial de utilização da biblioteca *JExcelAPI*

Nesta biblioteca, os arquivos de planilhas são interpretados como livros ou *Workbooks* e cada um destes livros é composto por uma ou mais planilhas (*Sheets*).

A leitura começa com a criação de um *Workbook* a partir do diretório do arquivo *xls*. As planilhas contidas nesse *Workbook* são acessadas a partir da posição 0 e o conteúdo do arquivo é acessado pela varredura da planilha em colunas e linhas, como mostra a Figura A.1.

```
private void leituraPlanilha() {
    try {
        Workbook workbook = Workbook.getWorkbook(new File("diretorio/planilha.xls"));
        Sheet sheet = workbook.getSheet(0);

        Object nomeColunas[] = new Object[sheet.getColumns()];
        for (int i = 0; i < sheet.getRows(); i++) {
            Object linha[] = new Object[sheet.getColumns()];
            for (int j = 0; j < sheet.getColumns(); j++) {
                if (i == 0) {
                    /* Pega o nome de todas as colunas */
                    nomeColunas[j] = sheet.getCell(j, i).getContents();
                } else {
                    /* Acessa o conteúdo da planilha */
                    linha[j] = sheet.getCell(j, i).getContents();
                }
            }
        }
    } catch (IOException ex) {
        System.out.println("Erro: " + ex);
    } catch (BiffException ex) {
        System.out.println("Erro: " + ex);
    }
}
```

Figura A.1: Leitura de um arquivo .xls

O processo de criação de uma planilha é similar ao de leitura. Primeiro cria-se um *Workbook* e uma Planilha para escrita (*WritableWorkbook* e *WritableSheet*). A inserção do conteúdo na planilha é feita pelo método *addCell(coluna, linha, dado)* enviando o dado a ser inserido e a localização do mesmo no arquivo (Figura A.2).

```
public void criaPlanilha() {
    try {
        /* Cria um Workbook para escrita */
        WritableWorkbook workbook = Workbook.createWorkbook(new File("planilha.xls"));
        /* Cria uma Planilha para escrita */
        WritableSheet sheet = workbook.createSheet("Resultado", 0);

        /* Adiciona os dados */
        for (int i = 0; i < 10; i++) {
            for (int j = 0; j < 10; j++) {
                /* Cria uma célula na posição [j,i] */
                Label label = new Label(j, i, "Célula: "+j+", "+i);
                sheet.addCell(label);
            }
        }
        workbook.write();
        workbook.close();
    } catch (WriteException ex) {
        System.out.println("Erro: " + ex);
    } catch (IOException ex) {
        System.out.println("Erro: " + ex);
    }
}
```

Figura A.2: Criação de planilha

O resultado deste exemplo é mostrado na Figura A.3.

	A	B	C	D	E	F	G	H	I	J
1	Célula: 0,0	Célula: 1,0	Célula: 2,0	Célula: 3,0	Célula: 4,0	Célula: 5,0	Célula: 6,0	Célula: 7,0	Célula: 8,0	Célula: 9,0
2	Célula: 0,1	Célula: 1,1	Célula: 2,1	Célula: 3,1	Célula: 4,1	Célula: 5,1	Célula: 6,1	Célula: 7,1	Célula: 8,1	Célula: 9,1
3	Célula: 0,2	Célula: 1,2	Célula: 2,2	Célula: 3,2	Célula: 4,2	Célula: 5,2	Célula: 6,2	Célula: 7,2	Célula: 8,2	Célula: 9,2
4	Célula: 0,3	Célula: 1,3	Célula: 2,3	Célula: 3,3	Célula: 4,3	Célula: 5,3	Célula: 6,3	Célula: 7,3	Célula: 8,3	Célula: 9,3
5	Célula: 0,4	Célula: 1,4	Célula: 2,4	Célula: 3,4	Célula: 4,4	Célula: 5,4	Célula: 6,4	Célula: 7,4	Célula: 8,4	Célula: 9,4
6	Célula: 0,5	Célula: 1,5	Célula: 2,5	Célula: 3,5	Célula: 4,5	Célula: 5,5	Célula: 6,5	Célula: 7,5	Célula: 8,5	Célula: 9,5
7	Célula: 0,6	Célula: 1,6	Célula: 2,6	Célula: 3,6	Célula: 4,6	Célula: 5,6	Célula: 6,6	Célula: 7,6	Célula: 8,6	Célula: 9,6
8	Célula: 0,7	Célula: 1,7	Célula: 2,7	Célula: 3,7	Célula: 4,7	Célula: 5,7	Célula: 6,7	Célula: 7,7	Célula: 8,7	Célula: 9,7
9	Célula: 0,8	Célula: 1,8	Célula: 2,8	Célula: 3,8	Célula: 4,8	Célula: 5,8	Célula: 6,8	Célula: 7,8	Célula: 8,8	Célula: 9,8
10	Célula: 0,9	Célula: 1,9	Célula: 2,9	Célula: 3,9	Célula: 4,9	Célula: 5,9	Célula: 6,9	Célula: 7,9	Célula: 8,9	Célula: 9,9

Figura A.3: Planilha criada

# Apêndice B

## Tutorial de utilização da biblioteca *JFreeChart*

O primeiro passo é criar uma coleção de pontos, que funciona como uma lista dos pontos utilizados para o desenho do gráfico. Nesta coleção estarão contidas as séries que possuem as coordenadas  $x$  e  $y$ , sendo que cada série é interpretada como um subgráfico independente. A Figura B.1 exemplifica essa primeira etapa.

```
@Override
public void paint(Graphics g) {
    super.paint(g);
    /* Cria coleção de pontos */
    XYSeriesCollection dataset = new XYSeriesCollection();
    /* Cria os pontos a serem plotados */
    XYSeries xy = this.geraGraficoLinhaXY();
    /* Adiciona na coleção */
    dataset.addSeries(xy);
    /* Desenha o gráfico */
    this.desenhaGrafico(g, dataset);
}
```

Figura B.1: Configuração inicial do desenho do gráfico

Para a criação das séries basta apenas dar um nome à mesma (geralmente uma variável do tipo *String*), e depois adicionar as coordenadas  $x$  e  $y$  de acordo com a necessidade do programador (ver Figura B.2).

Adicionadas as séries à coleção, o último passo é criar o gráfico. Este possui alguns parâmetros de customização tais como: título, nome dos eixos  $x$  e  $y$ , legenda, tooltips e URLs, como mostra a Figura B.3. O resultado final é mostrado na Figura B.4.

```

public XYSeries geraGraficoLinhaXY() {
    try {
        XYSeries series = new XYSeries("Grafico");
        for (int i = 1; i <= 10; i++) {
            series.add(i, i);
        }
        return series;
    } catch (Exception e) {
        return null;
    }
}

```

Figura B.2: Criação das séries

```

public void desenhaGrafico(Graphics g, XYSeriesCollection dataset) {
    JFreeChart chart = ChartFactory.createXYLineChart("Titulo do Gráfico", "Nome eixo X",
        "Nome eixo Y", dataset, PlotOrientation.VERTICAL, true, true, false);

    /* Customizacao das cores gráfico - cor de fundo e cor das linhas do grid */
    chart.setBackgroundPaint(Color.WHITE);
    XYPlot plot = chart.getXYPlot();
    plot.setBackgroundPaint(Color.WHITE);
    plot.setDomainGridlinePaint(Color.BLACK);
    plot.setRangeGridlinePaint(Color.BLACK);

    /* Desenha gráfico com dimensões 500x500 */
    g.drawImage(chart.createBufferedImage(500, 500), 0, 0, 500, 500, this);
}

```

Figura B.3: Customização do gráfico

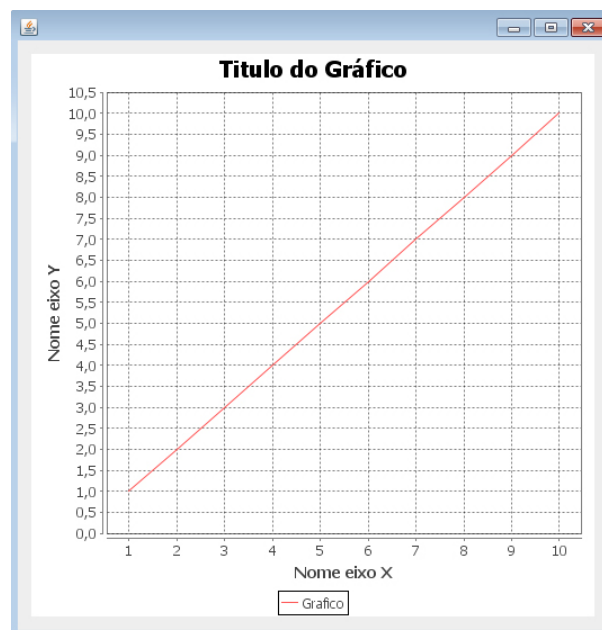


Figura B.4: Desenho do gráfico



# Referências Bibliográficas

- [Akaike 1973]AKAIKE, H. Information theory as an extension of the maximum likelihood principle. In: *International Symposium On Information Theory*. Budapest: [s.n.], 1973. p. 267–281.
- [Balastreire 1998]BALASTREIRE, L. A. Estudo de caso: Uma pesquisa brasileira em agricultura de precisão. In: *Congresso Brasileiro de Engenharia Agrícola*. Poços de Caldas, MG: [s.n.], 1998. p. 203–231.
- [Batchelor, Whigham e Dewitt 1997]BATCHELOR, B.; WHIGHAM, K.; DEWITT, J. *Precision Agriculture: Introduction to Precision Agriculture*. 1997. Iowa Cooperative Extension, 4p.
- [Brun 2005]BRUN, A. L. Algoritmos genéticos. In: *Encontro Paranaense de Computação – EPAC*. Cascavel, PR: UNIOESTE – Universidade Estadual do Oeste do Paraná: [s.n.], 2005.
- [Brun 2007]BRUN, A. L. *Indicadores de qualidade de ajuste de semivariograma experimental no software Geogac 2.1*. Dissertação (Dissertação de Mestrado) — UNIOESTE – Universidade Estadual do Oeste do Paraná, Cascavel, PR, 2007.
- [Brun e Opazo 2009]BRUN, A. L.; OPAZO, M. A. U. Estimação dos parâmetros de semivariogramas através de um algoritmo genético. In: *III Encontro de Paranaense de Computação – EPAC*. Cascavel, PR: UNIOESTE – Universidade Estadual do Oeste do Paraná: [s.n.], 2009. p. 267–281.
- [Burgess e Webster 1980]BURGESS, T. M.; WEBSTER, R. Optimal interpolation and isarithmic mapping of soil properties. i the semi-variogram and punctual kriging. *Journal of Soil Science*, n. 31, p. 315–331, 1980.

- [Carvalho 2010]CARVALHO, A. P. de Leon F. de. *Algoritmos Genéticos*. 2010. [Http://www.icmc.usp.br/ andre/research/genetic/](http://www.icmc.usp.br/andre/research/genetic/) Consultado na Internet em: jan/2010.
- [Carvalho e Assad 2005]CARVALHO, J. R. P.; ASSAD, E. D. Análise espacial da precipitação pluviométrica no estado de são paulo: comparação de métodos de interpolação. *Engenharia Agrícola*, v. 25, n. 2, p. 377–384, maio/agosto 2005.
- [Catarina 2005]CATARINA, A. S. *Algoritmos evolutivos aplicados ao processo de análise de dados geográficos*. São José dos Campos, SP, 2005.
- [Catarina e Bach 2003]CATARINA, A. S.; BACH, S. L. Estudo do efeito dos parâmetros genéticos na solução otimizada e no tempo de convergência em algoritmos genéticos com codificações binária e real. *Acta Scientiarum: Technology*, v. 25, n. 2, p. 147–152, 2003.
- [Câmara e Medeiros 1998]CÂMARA, G.; MEDEIROS, J. Gis para meio ambiente. In: *Sistemas de Informações Geográficas*. Curitiba, PR: [s.n.], 1998. p. 185.
- [Cortes 1999]CORTES, M. B. S. Introdução à otimização. In: *II Jornada de Estatística de Maringá*. Maringá, PR: UEM – Universidade Estadual de Maringá: [s.n.], 1999.
- [Cressie 1993]CRESSIE, N. A. C. *Statistics for spatial data*. New York: John Wiley, 1993.
- [Cressie e Hawkins 1980]CRESSIE, N. R.; HAWKINS, D. M. Robust estimation of the variogram. In: *Mathematical Geology*. New York: [s.n.], 1980. p. 115–125.
- [Darwin 1859]DARWIN, C. *On The Origin of Species by Means of Natural Selection*. London: John Murray, 1859.
- [Davis 1987]DAVIS, B. M. Uses and abuses of cross-validation in geostatistics. *Journal of Mathematical Geology*, New York, v. 19, n. 3, p. 405–421, 1987.
- [Faraco et al. 2008]FARACO, M. A. et al. 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, MG, v. 32, n. 2, 2008.

- [Ferreira 2010]FERREIRA, M. A. *Transformismo e Extinção de Lamarck à Darwin*. 2010. [Http://www.teses.usp.br/teses/disponiveis/8/8133/tde-24102007-150401/](http://www.teses.usp.br/teses/disponiveis/8/8133/tde-24102007-150401/) Consultado na Internet em: 22/março/2010.
- [Filliben 1975]FILLIBEN, J. J. *The probability plot correlation coefficient test for normality*. 1975. *Technometrics*, p111-117.
- [Fonte 1994]FONTE, C. M. P. da C. *Ajustamento de observações utilizando o método dos mínimos quadrados*. Dissertação (Dissertação de Mestrado) — Universidade de Coimbra, 1994.
- [Freitas 2000]FREITAS, V. A. *Análise de Dados Espaciais por Meio de Semivariogramas*. Dissertação (Dissertação de Monografia) — Universidade Federal de Uberlândia, Uberlândia, MG, 2000.
- [Gentil e Ferreira 1999]GENTIL, L. V.; FERREIRA, S. M. Agricultura de precisão: Prepare-se para o futuro, mas com os pés no chão. *Revista A Granja*, Porto Alegre, RS, n. 610, p. 12–17, 1999.
- [Gilbert e Grégoire 2003]GILBERT, M.; GRÉGOIRE, J. C. Site condition and predation influence a bark beetle's success: a spatially realistic approach. *Agricultural and Forest Entomology*, v. 5, n. 2, p. 87, 2003.
- [Goldberg 1989]GOLDBERG, D. E. *Genetic Algorithms in Search, Optimization, and Machine Learning*. New York: Addison-Wesley, 1989.
- [Gomes et al. 2007]GOMES, N. M. et al. *Métodos de Ajuste de Modelos de Semivariograma Aplicados ao Estudo da Variabilidade Espacial de Atributos Físico-Hídricos do Solo*. Dissertação (Dissertação de Mestrado) — UFLA – Universidade Federal de Lavras, Lavras, MG, 2007.
- [GONÇALVES 1997]GONÇALVES, A. C. A. *Variabilidade espacial de propriedades físicas do solo para fins de manejo da irrigação*. Tese (Tese de Doutorado) — Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, SP, 1997.
- [Guerra 1988]GUERRA, P. A. G. *Geoestatística operacional*. São Paulo: Editora Polígono, 1988.

- [Henriques 2007]HENRIQUES, A. *Thomas Robert Malthus: A Teoria Malthusiana*. Coimbra, Portugal, 2007.
- [Herrera, Lozano e Verdegay 1998]HERRERA, F.; LOZANO, M.; VERDEGAY, J. L. Tackling real-coded genetic algorithms: Operators and tools for behavioural analysis. *Artificial Intelligence Review*, v. 12, 1998.
- [Holland 1975]HOLLAND, J. H. *Adaptation in Natural and Artificial Systems*. Michigan: The University of Michigan Press, 1975.
- [Isaaks e Srivastava 1989]ISAACS, E. H.; SRIVASTAVA, R. M. *An introduction to Applied Geostatistics*. New York: Oxford University Press, 1989.
- [Júnior 1995]JÚNIOR, P. J. R. *Abordagem Geoestatística em um estudo de variabilidade espacial*. Dissertação (Dissertação de Mestrado) — Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, SP, 1995.
- [Jong 1975]JONG, K. A. D. *An analysis of the behavior of a class of genetic adaptive systems*. Tese (Tese de Doutorado) — University of Michigan, Michigan, 1975.
- [Journel e Huijbregts 1978]JOURNEL, A. G.; HUIJBREGTS, C. J. *Mining Geoestistics*. London: Academic Press, 1978.
- [Jr 2002]JR, T. R. H. Chasing tyrannosaurus and deinonychus around the tree of life: Classifying dinosaurs. In: *Fiorillo, Dinosaurs: The Science Behind the Stories, Society of Vertebrate Paleontology, Paleontological Society e American Geological Institute*. [S.l.: s.n.], 2002. p. 31–38.
- [Kitanidis e Vomvoris 1983]KITANIDIS, P. L.; VOMVORIS, E. G. A geoestatical approach to the inverse problem in groundwater modelling and one dimensional simulations. *Water Resources Research*, p. 677–690, 1983.
- [Krige 1951]KRIGE, D. G. A. A statistical approach to some basic mine evaluation problems on the witwatersrand. *Journal of South African Institution of Minning and Metallurgy*, Johannesburg, v. 52, p. 119–139, 1951.

- [Lacerda 2009]LACERDA, E. G. M. *Algoritmos Genéticos: Aspectos Práticos*. Natal, RN, 2009.
- [Lajaunie 1984]LAJAUNIE, C. A geostatistical approach to air pollution modelling. In: *Geostatistics for Natural Resources Characterization*. California: [s.n.], 1984. p. 877–891.
- [Landry 2010]LANDRY, P. *Thomas Robert Malthus*. 2010. [Http://www.blupete.com/Literature/Biographies/Philosophy/Malthus.htm](http://www.blupete.com/Literature/Biographies/Philosophy/Malthus.htm) Consultado na Internet em: 22/março/2010.
- [Li e Lake 1994]LI, D.; LAKE, L. W. A moving window semivariance estimator. *Water Resources Research*, v. 30, n. 5, p. 1479–1489, 1994.
- [Linneu 1758]LINNEU, C. de. *Systema Naturæ per regna tria naturæ, secundum classes, ordines, species, cum characteribus differentiis, synonymis, locis*. [S.l.]: Holmiae, 1758.
- [Matheron 1963]MATHERON, G. Principles of geostatistics. *Economic Geology*, v. 58, p. 1246–1266, 1963.
- [Matos 2003]MATOS, J. D. de. *Utilização de Semivariogramas como Redutor de Dimensionalidade no Reconhecimento de Padrões em Imagens Digitais Hiperespectrais*. Dissertação (Dissertação de Mestrado) — UFRGS – Universidade Federal do Rio Grande do Sul, Porto Alegre, RS, 2003.
- [McBratney e Webster 1986]MCBRATNEY, A. B.; WEBSTER, R. Choosing functions for semi-variograms of soil properties and fitting them to sampling estimates. *Journal Soil Science*, Oxford, v. 37, n. 3, p. 617–639, 1986.
- [Miranda 2010]MIRANDA, M. N. de. *Algoritmos Genéticos: Fundamentos e Aplicações*. 2010. [Http://www.gta.ufrj.br/marcio/genetic.html](http://www.gta.ufrj.br/marcio/genetic.html) Consultado na Internet em: fev/2010.
- [Molin 2001]MOLIN, J. P. *Agricultura de precisão: O Gerenciamento da Variabilidade*. 2001. Piracicaba, SP, 83p.
- [Sakamoto, Ishiguro e Kitagawa 1986]SAKAMOTO, Y.; ISHIGURO, M.; KITAGAWA, G. *Akaike information criterion statistics*. New York: Springer, 1986.

- [Schneider 1998]SCHNEIDER, A. M. *Algoritmo Adaptativo Genético para Acompanhamento da Trajetória de Alvos Móveis*. Dissertação (Dissertação de Mestrado) — UFRGS – Universidade Federal do Rio Grande do Sul, Instituto de Informática, Porto Alegre, RS, 1998.
- [Shiratsuchi 2001]SHIRATSUCHI, L. S. *Mapeamento da variabilidade espacial das plantas daninhas com a utilização de ferramentas da agricultura de precisão*. Dissertação (Dissertação de Mestrado) — Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, SP, 2001.
- [Silva, Moretto e Rodrigues 2004]SILVA, C. B.; MORETTO, A. C.; RODRIGUES, R. L. Viabilidade econômica do uso da agricultura de precisão: o caso do paraná. In: *XLII Congresso da Sociedade Brasileira de Economia e Sociologia Rural*. Cuiabá, MT: [s.n.], 2004.
- [Soares et al. 2008]SOARES, V. P. et al. Análise espacial da distribuição de cigarras (quesada gigas oliver) em povoamentos de paricá (schizolobium amazonicum huber ex ducke) na região de dom eliseu, pa. In: *Sociedade de Investigações Florestais*. Viçosa, MG: [s.n.], 2008. p. 251–258.
- [Srivastava e Parker 1989]SRIVASTAVA, R. M.; PARKER, H. M. Robust measures of spatial continuity. In: *Geostatistics, on the 3rd. Geostatistical Congress*. Holland, Armstrong: [s.n.], 1989. p. 295–308.
- [Vieira 1998]VIEIRA, S. R. Geoestatística aplicada à agricultura de precisão. In: *Sistema de Informações Geográficas*. Curitiba, PR: [s.n.], 1998. p. 1–48.
- [Werner 2007]WERNER, V. *Análise Econômica e Experiência Comparativa Entre a Agricultura de Precisão e Tradicional*. Tese (Tese de Doutorado) — Centro de Ciências Rurais, Universidade Federal de Santa Maria, Santa Maria, RS, 2007.