

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

**INDICADORES DE QUALIDADE DE AJUSTE DE
SEMIVARIOGRAMA EXPERIMENTAL NO SOFTWARE
GEOGAC 2.1**

ANDRÉ LUIZ BRUN

**CASCADEL - PR
Junho - 2007**

Livros Grátis

<http://www.livrosgratis.com.br>

Milhares de livros grátis para download.

ANDRÉ LUIZ BRUN

**INDICADORES DE QUALIDADE DE AJUSTE DE
SEMIVARIOGRAMA EXPERIMENTAL NO SOFTWARE
GEOGAC 2.1**

Projeto apresentado ao Programa de Pós-Graduação em Engenharia Agrícola em cumprimento parcial aos requisitos para obtenção do título de Mestre em Engenharia Agrícola, área de concentração em **Engenharia de Sistemas Agroindustriais**.

Orientador: Prof. Dr. Miguel Angel Uribe
Opazo.

**CASCVEL - PR
Julho – 2007**

Ficha catalográfica
Elaborada pela Biblioteca Central do Campus de Cascavel - Unioeste

B919i Brun, André Luiz
 Indicadores de qualidade de ajuste de semivariograma experimental
 no software GEOCAC 2.1 / André Luiz Brun— Cascavel, PR: UNIOESTE,
 2007.
 85 f. ; 30 cm.

 Orientador: Prof. Dr. Miguel Angel Uribe Opazo
 Dissertação (Mestrado) – Universidade Estadual do Oeste do Paraná.
 Programa de Pós-Graduação Stricto Sensu em Engenharia Agrícola,
 Centro de Ciências Exatas e Tecnológicas.
 Bibliografia.

 1. Geoestatística. 2. GEOCAC (Software). I. Opazo, Miguel Angel
 Uribe. II. Universidade Estadual do Oeste do Paraná. III. Título.

CDD 21ed. 620.001171
 519.53

Bibliotecária: Jeanine da Silva Barros CRB-9/1362

“Quero que a estrada venha sempre até você e que o vento esteja sempre a seu favor! Quero que haja sempre uma cerveja em sua mão e que esteja sempre ao seu lado, seu grande amor”

Jimmy London

AGRADECIMENTOS

À minha esposa Greicy, que durante todos estes anos me apoiou, incentivou, foi amiga, exemplo, companheira e paciente quando não pude estar presente pela necessidade de desenvolver o presente trabalho.

À minha família, por todo carinho, suporte e incentivo.

Ao Professor Dr. Miguel Angel Uribe Opazo, por acreditar em mim para o desenvolvimento deste trabalho e pelo apoio durante a execução do mesmo.

A todos os meus amigos que, de uma forma ou outra contruíram para o cumprimento desta etapa em minha vida.

À todas as pessoas que, no decorrer deste período, deram auxílio e incentivo para que conseguisse concluir esta pesquisa.

À todos os funcionários e amigos da UNIOESTE que auxiliaram o desenvolvimento desta tarefa.

À todas as bandas de rock que ajudaram esta árdua tarefa e que lutam para manter acesa a chama do Rock n' Roll mas que não serão citadas nas referências do trabalho.

SUMÁRIO

| | Página |
|--------------------------------------------------------------------|--------|
| RESUMO | 09 |
| ABSTRACT | 10 |
| 1 INTRODUÇÃO..... | 11 |
| 2 REVISÃO BIBLIOGRÁFICA | 16 |
| 2.1 Agricultura de Precisão | 16 |
| 2.2 Geoestatística | 18 |
| 2.2.1 Hipóteses usadas na Geoestatística | 22 |
| 2.2.2 Semivariogramas | 24 |
| 2.2.2.1 Modelagem e Ajuste de Semivariogramas | 29 |
| 2.2.3 Anisotropia e Isotropia | 32 |
| 2.2.4 Krigagem | 33 |
| 2.2.5 Indicadores de Melhor Ajuste | 34 |
| 2.2.5.1 Validação Cruzada | 34 |
| 2.2.5.2 Critério de Akaike | 37 |
| 2.2.5.3 Critério de Filliben | 38 |
| 2.3 Principais Softwares Geoestatísticos | 39 |
| 3 MATERIAL E MÉTODOS..... | 42 |
| 3.1 Módulo 1 (Entrada de Dados) | 44 |
| 3.2 Módulo 2 (Saída de Dados) | 46 |
| 3.3 Módulo 3 (Estimadores da Função Semivariância) | 46 |
| 3.3.1 Estimador de Pairwise | 47 |
| 3.3.2 Estimador New 1 | 50 |
| 3.3.3 Estimador New 2 | 52 |
| 3.4 Módulo 4 (Krigagem) | 53 |
| 3.5 Módulo 5 (Indicadores para escolha do melhor modelo) | 56 |
| 3.5.1 Validação Cruzada | 56 |
| 3.5.2 Critério de Akaike | 58 |
| 3.5.3 Teste de Filliben | 58 |
| 3.6 Módulo 6 (Previsão de Pontos não amostrados) | 59 |
| 4 RESULTADOS E DISCUSSÕES | 61 |
| 4.1 Estimador de Paiwise | 65 |
| 4.2 Estimador New 1 | 66 |
| 4.3 Estimador New 2 | 67 |
| 4.4 Krigagem, Validação Cruzada, Critério de Filliben e Akaike ... | 69 |
| 4.5 Estimadores de Matheron e Cressie & Hawkins | 74 |
| 5 CONCLUSÕES | 78 |
| 6 BIBLIOGRAFIA | 79 |

LISTA DE TABELAS

| | | Página |
|-----------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|
| Tabela 1 | Parâmetros encontrados para a construção do semivariograma que apresentaram casamento aproximado entre o semivariograma experimental e a curva do modelo utilizado..... | 77 |
| Tabela 2 | Valores encontrados para os métodos de avaliação de ajuste..... | 77 |

LISTA DE FIGURAS

| | | Página |
|------------------|-------------------------------------------------------------------------------------|--------|
| Figura 1 | Ciclo da Agricultura de Precisão | 17 |
| Figura 2 | Área com Concentração Heterogênea | 18 |
| Figura 3 | Grid Regular de Amostragem | 20 |
| Figura 4 | Exemplo de Janela Móvel Bi-dimensional..... | 27 |
| Figura 5 | Gráfico de um Semivariograma Experimental | 28 |
| Figura 6 | Gráfico do modelo linear. | 29 |
| Figura 7 | Gráfico do modelo circular..... | 30 |
| Figura 8 | Gráfico do modelo esférico..... | 30 |
| Figura 9 | Gráfico do modelo exponencial | 31 |
| Figura 10 | Gráfico do modelo gaussiano..... | 32 |
| Figura 11 | Processo de desenvolvimento através do modelo espiral... | 40 |
| Figura 12 | Tela principal do GEOCAC | 42 |
| Figura 13 | Estrutura do Vetor de Pairwise..... | 45 |
| Figura 14 | Exemplo de gráfico $u \times \gamma(u)$ | 46 |
| Figura 15 | Exemplo de gráfico com busca de ajuste a modelo matemático | 46 |
| Figura 16 | Exemplo de janela deslizante de forma circular..... | 49 |
| Figura 17 | Estrutura do processo da krigagem..... | 51 |
| Figura 18 | Fluxograma da Validação Cruzada | 52 |
| Figura 19 | Mapa de pontos amostrados e estimados | 55 |
| Figura 20 | Tela principal do GEOCAC com novas inclusões..... | 58 |
| Figura 21 | Tela Semivariograma por Pairwise | 60 |
| Figura 22 | Tela Semivariograma por New 1..... | 61 |
| Figura 23 | Tela para escolha dos parâmetros para a construção do semivariograma de New 2 | 63 |
| Figura 24 | Tela Semivariograma por New 2 | 63 |
| Figura 25 | Módulo de Krigagem | 65 |
| Figura 26 | Módulo de Validação Cruzada | 67 |
| Figura 27 | Critério de Akaike | 68 |
| Figura 28 | Teste de Filliben | 68 |
| Figura 29 | Relatório sobre a variável em estudo | 69 |
| Figura 30 | Tela Semivariograma por Matheron | 70 |
| Figura 31 | Tela Semivariograma por Cressie & Hawkins | 71 |
| Figura 32 | | |
| Figura 33 | | |

RESUMO

A geoestatística vem sendo utilizada com sucesso dentro do estudo de variáveis com dependência espacial, tornando mais eficiente o processo estimativo e a produção agrícola, visto que permite um estudo detalhado das propriedades da área em utilização. Com a evolução das técnicas geoestatísticas, surgiram também softwares para auxiliar no processo estimativo e decisivo. Eles vêm se mostrando grandes aliados no estudo das propriedades do solo. Este trabalho visou implementar novas técnicas geoestatísticas no software GEOCAC com o intuito de torná-lo não só mais completo e abrangente, mas também mais eficiente, servindo de grande valia aos pesquisadores da área. O software obtido ao término do trabalho mostrou-se mais completo e eficiente, pois as novas técnicas incluídas possibilitam análises mais completas e confiáveis, permitindo aos usuários a construção de novos modelos teóricos a semivariogramas experimentais, a estimação de pontos não amostrados pela técnica da krigagem e a avaliação dos modelos ajustados, por meio dos métodos de validação cruzada, teste de Filliben e critério de Akaike.

PALAVRAS CHAVE: Krigagem, Validação de Módulos, Geoestatística.

Pointers of Adjustment's Quality of Experimental Semivariograms is GEOCAC 2.1 Software

ABSTRACT

The geostatistics has been successfully used inside of the study of variables with spatial dependence, becoming more efficient the esteem process and agricultural production, since it allows a detailed study of the properties of the area in use. With the evolution of the geostatistics techniques, they had also appeared softwares to assist in the estimative and decisive process. These come if showing great allies in the study of the ground properties. This work aimed at to implement new geostatistics techniques in software GEOCAC with intention to become it more complete and including, but not only also more efficient, serving of great value the researchers of the area. The software gotten to the ending of the work revealed more complete and efficient, therefore the new enclosed techniques allow analyze more complete and trustworthy, allowing to the users the construction of new theoretical models at experimental semivariograms, the esteem of points not showed for the technique of the krigagem and the evaluation of the adjusted models, by means of the methods of cross validation, Filliben's test and Akaike's criterion.

KEYWORDS: Kriging, Model Validation, Geostatistics.

1 INTRODUÇÃO

Atualmente, existem softwares comerciais que realizam estudos geoestatísticos, porém estes possuem alto custo de aquisição e são implementados em ambiente fechado, ou seja, o usuário não pode realizar alterações ou mesmo acompanhar passo a passo o processo, visto que muitos deles geralmente funcionam como uma “caixa preta”. Porém, como há a necessidade, em alguns casos, de um acompanhamento mais detalhado do processo, é interessante existir a interação entre usuário e programa que permita acompanhar gradativamente o processo. Outro fator importante ao processo é a definição dos indicadores utilizados para a escolha do melhor modelo teórico ao modelo experimental. Nesse sentido, é interessante a utilização de um software que disponibilize diversos novos indicadores para realizar a escolha do melhor modelo, visto que o processo se tornaria mais rápido e eficiente.

Existem vários softwares livres intensamente utilizados por pesquisadores da área estatística. Entre eles, o software chamado R que, além de ser livre, é um ambiente que permite construções estatísticas e gráficas com boa qualidade. Com ele, o usuário criar suas próprias rotinas, de acordo com sua necessidade específica, ou seja, o software R pode ser compreendido também como uma linguagem de programação.

No ano de 2000, foram realizadas pesquisas pelo Grupo de Geoestatística Aplicada (GGEA) da UNIOESTE, visando contornar os problemas seguintes: softwares com alto custo e fechados, falta de suporte técnico, ou mesmo, softwares livres que dependem de softwares comerciais. Como resultado destas pesquisas, surgiu o software GEOCAC, que é um software desenvolvido pelo grupo. Este software veio a auxiliar as pesquisas voltadas à geoestatística, porém ainda foi necessária a implementação de novas teorias desenvolvidas nos últimos anos, fato que confirma a importância de continuar incrementando novos módulos ao software GEOCAC, com a intenção de torná-lo mais eficiente, mais abrangente e, sobretudo, uma comunicação amigável entre usuário e programa.

SILVA et al. (2002) apresentaram a versão 1.1 do software GEOCAC para o ajuste de modelos teóricos de semivariogramas experimentais. BRUN et al. (2004) implementaram um módulo que realizasse o processo de krigagem e validação cruzada que, posteriormente, veio a ser integrado ao Software GEOCAC, precisando, porém, este módulo, de ajustes e correções. Tais melhorias foram contornadas pela presente pesquisa.

A interação entre a informática e demais áreas é cada vez mais presente e eficiente. A extensão do conhecimento de profissionais em áreas específicas com o conhecimento proveniente de especialistas da área da computação vêm obtendo resultados mais expressivos e interessantes com o passar do tempo. Um exemplo desta interação é o trabalho de DAINESE et al. (2004) que buscou a combinação do sensoriamento remoto, computação (por meio da utilização de redes neurais) e estatística espacial, para realizar a análise da incidência de plantas invasoras em determinada área com o intuito de obter uma representação realística da distribuição da produção dentro desta área. Dessa forma, pode direcionar, de maneira mais eficiente, o manejo do solo.

A Geoestatística está cada vez mais presente no trabalho de pesquisadores na área agrícola, pois vem se tornando importante recurso de apoio à Agricultura de Precisão, seja fornecendo métodos mais confiáveis e eficazes de estimação ou por meio de técnicas mais robustas (valores discrepantes não influenciar no resultado) para a análise dos dados.

Dentro da Geoestatística, existem vários métodos muito utilizados, como os semivariogramas e correlogramas que visam determinar uma estrutura de dependência espacial, caso ela exista. Nos últimos anos, diferentes autores utilizaram semivariogramas para realizar a análise da variabilidade espacial de características quantitativas com certa continuidade espacial.

O estudo da Geoestatística se mostra cada vez mais eficaz e, conseqüentemente, abrangente para o estudo de dados de superfícies. Essa observação deve-se ao fato de que esta técnica, que considera o

posicionamento geográfico das amostras como fator de influência, permite a construção de mapas temáticos que possibilitem não só uma representação do fenômeno em estudo, mas também fornecem informações para realização de um estudo de zoneamento de áreas geográficas.

Outro método eficiente utilizado na Geoestatística, o semivariograma, apresenta vários modelos nos quais é baseada a sua construção. Tais modelos podem apresentar valores distintos. Este trabalho visa desenvolver técnicas de avaliação para a verificação de qual modelo melhor se ajusta aos dados observados (dados de entrada), pois tendo em mãos tal indicador (de melhor ajuste) é possível ao usuário escolher o modelo que melhor se adapta à distribuição dos dados e, desta forma, evitar possíveis influências, como pontos discrepantes, por exemplo.

Assim, o semivariograma constitui-se numa importante representação quantitativa da variabilidade espacial e temporal de determinadas variáveis regionalizadas. Segundo NOVAIS et al. (2000), estas técnicas geoestatísticas estão entre as mais adequadas para medir dependência espacial, sendo que a escolha do melhor modelo repercute diretamente na confiabilidade dos resultados obtidos no processo de interpolação. Dessa forma, como a escolha do modelo de semivariograma, a escolha dos indicadores utilizados para verificação da qualidade dos resultados também é crucial.

Além dos estimadores clássicos de semivariogramas existentes, Matheron e Cressie & Hawkins, foram propostos outros estimadores que apresentam maior robustez, eficiência e confiabilidade nos resultados, aumentando assim a eficiência dos processos já existentes.

Nesse sentido, utilizando um software que permita trabalhar com estas estimações, seria possível obter uma estimacão da estrutura de dependência espacial mais confiável. Com a construção deste software, que além de possuir código aberto, será gratuito, tem-se um poderoso instrumento de estimacão, permitindo o acompanhamento detalhado do processo.

CARVALHO & ASSAD (2005) realizaram estudos de estimativas e interpolação de valores, onde mostraram que a krigagem é uma técnica bastante eficaz e confiável. Nesta pesquisa, verifico-se que o processo de krigagem foi mais eficiente, inclusive, que outros métodos de estimação conhecidos, como métodos de interpolação do quadrado do inverso das distâncias, média móvel e da curvatura mínima. A krigagem permite que seja realizada a construção de mapas temáticos, para explicar ou estudar determinado fenômeno, baseados na sua estrutura de variabilidade espacial ou temporal encontrada no ajuste do semivariograma experimental.

Assim, o desenvolvimento deste trabalho visa, justamente, contornar estes desafios e fornecer aos pesquisadores da área maior praticidade e confiança em realizar suas pesquisas, através de um software gratuito e seguro, implementando novos módulos geoestatísticos no software GEOCAC que ajudará aos usuários realizarem pesquisas confiáveis e eficientes, utilizando novas técnicas científicas desenvolvidas nos últimos anos.

Com o intuito de demonstrar mais compreensivelmente os objetivos propostos, foi adotada uma estrutura de módulos para analisar cada etapa separadamente:

Módulo 1: alterar a entrada de dados do software, permitindo melhor interação e eficiência, provendo ao usuário novas possibilidades de operação, como utilizar um arquivo de texto para entrada dos dados ou a possibilidade de analisar um conjunto de dados com tamanho considerável (8000 amostras, por exemplo), já que a versão anterior trabalhava apenas com um número restrito de pontos;

Módulo 2: realizar modificações no software, visando melhorar a etapa de saída de dados, tanto na parte de formatação dos dados quanto nas possibilidades de exibição dos dados finais, como gerar relatórios, salvar os gráficos gerados pelos dados, ou mesmo nova formatação dos gráficos permitindo melhor interpretação destes;

Módulo 3: implementar novos estimadores de semivariogramas experimentais aos já existentes no programa (Matheron e Cressie & Hawkins), que são: Pairwise, New 1 e New 2;

Módulo 4: implementar um algoritmo que realize o processo de krigagem ordinária, considerando áreas com diferentes contornos;

Módulo 5: implementar indicadores de validação para a escolha do melhor modelo ajustado ao semivariograma experimental. Estes indicadores são: validação cruzada, teste de Filliben e critério de Akaike;

Módulo 6: previsão de pontos não amostrados em mapas temáticos.

Assim, este software irá prover aos pesquisadores, professores, estudantes e usuários, o uso de um programa que permita um melhor acompanhamento do processo de análise, representação e escolha das melhores técnicas e modelos geoestatísticos de acordo com os dados em estudo.

2 REVISÃO BIBLIOGRÁFICA

2.1 Agricultura de Precisão

Segundo (BALASTREIRE, 1998), a agricultura de precisão (AP) é um conjunto de técnicas que permite um tratamento mais eficiente de culturas, 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. São considerados os níveis de necessidade de insumos a serem aplicados em função do que realmente a área precisa. Esta forma de tratamento das áreas cultivadas, conforme DAINESE et al. (2004), ocorre principalmente devido ao processo de mecanização da agricultura aliado à necessidade da produção em larga escala.

Os principais objetivos buscados com a implantação da AP, segundo BALESTREIRE (1998) e MOLIN (2001), são o aumento da rentabilidade da cultura e a redução dos impactos ambientais ocorridos pela aplicação deste. Esses objetivos podem ser alcançados pela aplicação de insumos a taxas variadas, de acordo com as características do solo, ou seja, cada região receberá apenas a quantidade que lhe é necessária originando um zoneamento das áreas.

Segundo MOLIN (2001), quando se depara com uma área bastante heterogênea é interessante dividi-la em regiões menores, mas que apresentem características semelhantes com o intuito de possibilitar a aplicação mais eficiente de insumos, evitando desperdício e diminuindo o impacto ambiental causado por eles.

A AP apresenta, conforme MOLIN (2001), um ciclo que deve ser seguido com o intuito de obter resultados mais efetivos e vantajosos. Tal ciclo está representado na Figura 1.

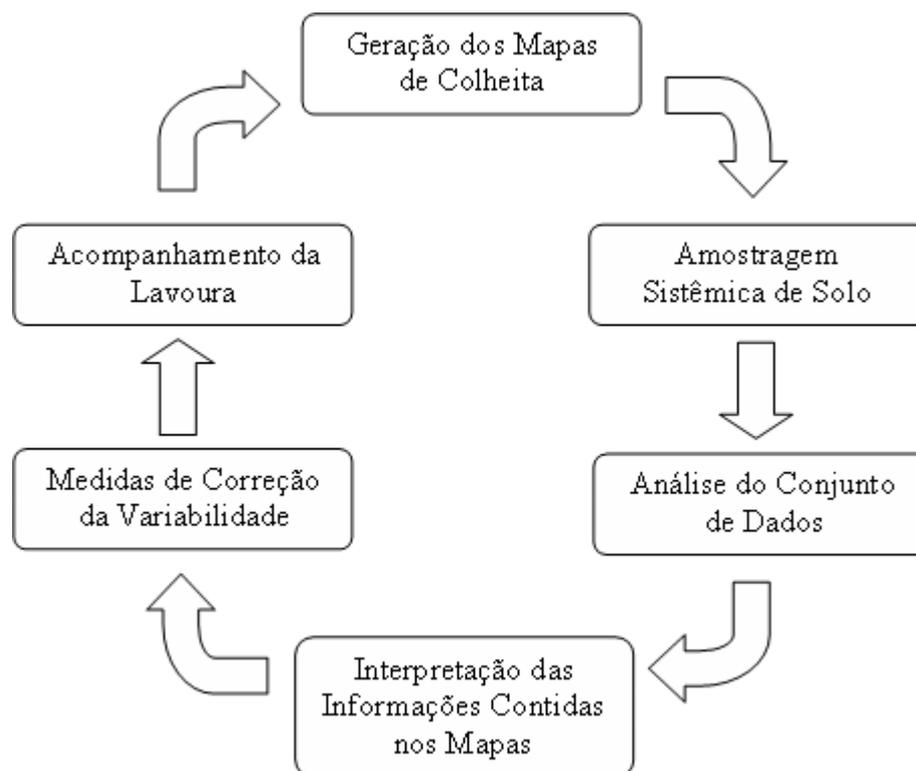


Figura 1 Ciclo da Agricultura de Precisão. Adaptado de MOLIN (2001).

A primeira etapa consiste na geração dos mapas de colheita, em seguida são realizadas amostragens sistêmicas do solo e posterior análise do conjunto de dados. O passo seguinte, então, é a interpretação das informações contidas nos mapas e aplicação das medidas de correção da variabilidade (de acordo com a distribuição de nutrientes) e, por fim, o acompanhamento da lavoura. Após realizado todo o ciclo, é necessário retornar ao passo inicial da geração dos mapas de colheita e realizar todo o processo novamente. Este ciclo visa não somente acompanhar em detalhes e conhecer o desempenho da área, mas, com o tempo, poder aplicar quantidades distintas de insumos em cada parte da área, de acordo com sua composição (MOLIN, 2001).

A Figura 2 é um exemplo de como uma área pode apresentar características heterogêneas. Neste caso, segundo a AP, cada área deveria receber uma quantidade distinta de insumos e não uma quantidade média para toda a área, como acontece na agricultura convencional, visto que apresentam concentrações diferentes de nutrientes.

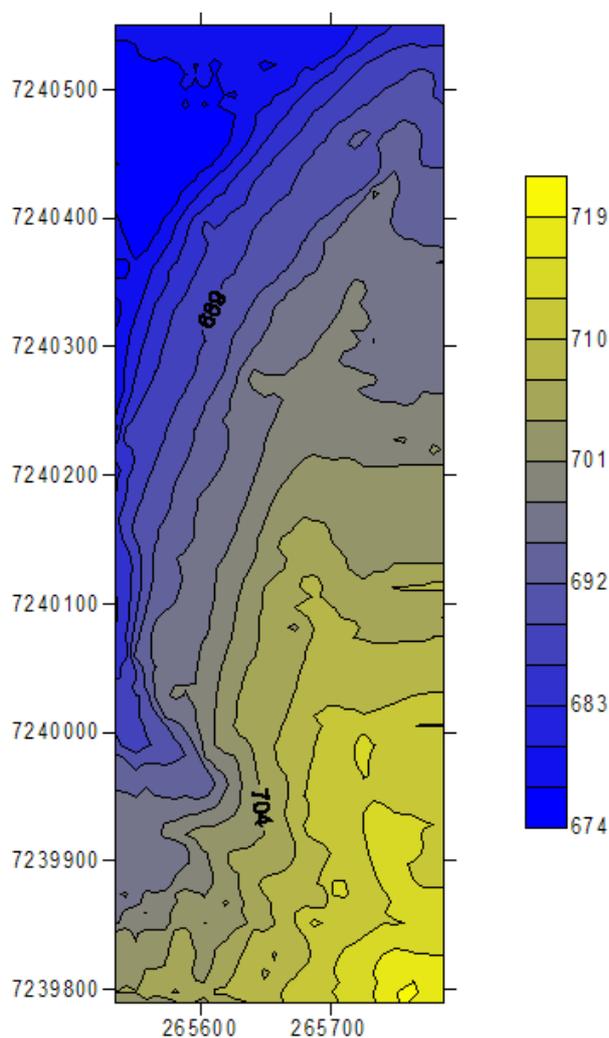


Figura 2 Área com Concentração Heterogênea.

2.2 Geoestatística

A Geoestatística teve início, quando KRIGE, em 1951, na África do Sul, concluiu que era necessário considerar a distância entre as amostras de dados de concentração de ouro, pois verificou que apresentavam algum tipo de dependência espacial.

RIBEIRO JUNIOR (1995) define a Geoestatística como um conjunto de procedimentos de caráter estatísticos, que são aplicados a problemas, onde os dados são referenciados espacialmente. Segundo o autor, na Geoestatística, a observação de propriedades do solo apresenta suas características de acordo

com sua posição e não somente com relação ao seu valor. Portanto, observações mais próximas geograficamente apresentam a tendência de valores mais parecidos.

Segundo CÂMARA & MEDEIROS (1998), 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.

GUERRA (1988) afirma que o nome “Geoestatística” surgiu para fazer referência à Aplicação da Teoria das Variáveis Regionalizadas, apresentada por MATHERON (1963), na solução de problemas de Geologia e Mineração.

A Geoestatística sofreu grande influência das pesquisas e trabalhos estatísticos desenvolvidos na escola sul-africana. GUERRA (1988) ressalta, ainda, que a Teoria das Variáveis Regionalizadas pode ser aplicada em outras áreas além da mineira. Por exemplo: na área florestal, visando a distribuição de humos no solo; na meteorologia, estudando a distribuição de chuvas em determinadas regiões, como apresentado em CARVALHO & ASSAD (2005); na engenharia agrícola, estudando a produtividade de soja e seus atributos físicos e químicos do solo (SOUZA et al., 1999; JOHANN et al., 2002, 2004).

Nesse sentido, a teoria das variáveis regionalizadas pode ser entendida como uma função que varia de acordo com sua posição geográfica, mas com certa tendência de continuidade. Ou seja, os valores amostrados em determinadas posições dependem das amostras encontradas nas proximidades, pois estas amostras possuem influências sobre a variável em estudo. Assim, quanto mais próximas estiverem estas amostras, maior será sua influência (GUERRA, 1988).

Considerando uma área $S \subset R^2$, no qual se tem um conjunto de valores medidos $\{z(s_{kl}), s_{kl} = (x_k, y_l) = s \in S; k = 1, 2, \dots, n_1, l = 1, 2, \dots, n_2\}$ onde $n = n_1 * n_2$ são os pontos amostrais. Para uma dada posição fixa s_{kl} , cada valor medido da variável em estudo, $z(s_{kl})$, pode ser considerado como uma realização de uma certa variável aleatória $Z(s_{kl})$, que por sua vez, pode ser

considerada uma realização do conjunto de variáveis aleatórias $\{Z(s_{ij}), \forall s_{ij} \in S\}$. Esse conjunto é chamado uma função aleatória e é denotado por $Z(s)$. Na literatura, $Z(s)$ é conhecida como variável regionalizada, função aleatória ou processo estocástico. Denota-se, neste trabalho, $Z(s)$ ou Z como uma variável regionalizada (GUERRA, 1988).

A Figura 3 apresenta um exemplo de um *grid* regular (\mathbb{R}^2) que representa um campo de coleta de amostras em determinadas características a estudar, onde cada amostra possuirá uma posição (x,y) definida no plano cartesiano.

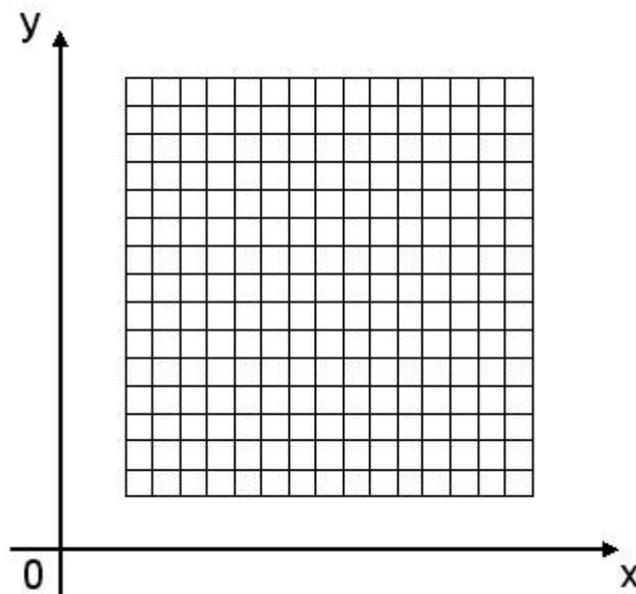


Figura 3 Grid Regular de Amostragem.

Para GUERRA (1988), as variáveis regionalizadas apresentam duas características interessantes:

- Aleatória: pois os valores numéricos de determinada propriedade podem variar consideravelmente de um ponto para outro;
- Espacial: os valores numéricos que são observados não são inteiramente independentes (dependem das posições vizinhas).

Existem, segundo GUERRA (1988), três características qualitativas das variáveis regionalizadas que influenciam, diretamente, nos valores numéricos obtidos em uma análise. São elas a localização da amostra, a continuidade do fenômeno na região e a anisotropia.

A amostra deve considerar a homogeneidade da localização onde foi realizada, pois pode, por exemplo, encontrar um local de alta concentração de determinado fenômeno, como um acúmulo de minério. Então, é interessante observar as semelhanças nas posições circundantes para verificar o campo de influência na amostra realizada.

A continuidade serve para se verificar o grau de dependência entre as amostras, pois, de acordo com a variável em estudo e da característica do local, pode-se obter uma dependência espacial alta, baixa ou inexistente.

Porém, como os fenômenos não podem ser tratados simplesmente como eventos aleatórios, surgiu a necessidade de técnicas de caráter matemático que permitam estudar estas características de um modo sintético, considerando o aspecto aleatório e espacial. A Geoestatística se propõe a estudar dois objetivos principais:

- Buscar uma imagem da variabilidade dos dados e uma medida da correlação entre os valores tomados em dois pontos do espaço. Para solucionar estas buscas, foi desenvolvido o semivariograma;
- Obter estimativas e medir a precisão das mesmas. Este objetivo vem sendo alcançado através da técnica conhecida como krigagem.

Por meio da utilização de técnicas geoestatísticas, é possível analisar a variabilidade espacial de determinado fenômeno, com continuidade espacial, ou analisar a distribuição de pragas dentro de uma área de cultivo, como realizado por AVENDAÑO et al. (2005).

SOUZA et al. (2005) realizaram um estudo da variabilidade espacial dos fatores de erosão, onde características do solo foram analisadas a fim de determinar a chance de ocorrerem erosões na área, através de técnicas

geoestatísticas. Já, NICOLELLA et al. (2005) apresentaram o comportamento da derivação de agrotóxicos em determinada área, devido à ocorrência de chuvas, através de análises espaciais.

2.2.1 Hipóteses usadas na Geoestatística

Utilizando uma amostragem única, tudo que se sabe sobre a variável regionalizada $Z(s)$ é uma única realização, onde $s \in S$ é um espaço paramétrico. Então, para que seja possível estimar valores para locais que não foram amostrados, é necessário assumir alguma hipótese de estacionaridade. As hipóteses de estacionaridade de segunda ordem ou a hipótese intrínseca são as hipóteses necessárias no estudo Geoestatístico da variável analisada. Para aplicarem-se as técnicas geoestatísticas pelo menos uma das hipóteses acima deve ser satisfeita, caso alguma não seja cumprida, os resultados não são considerados confiáveis e não poderiam ser realizadas inferências sobre a variável regionalizada ou processo estocástico.

Um processo estocástico, segundo HOEL et al. (1972), é estacionário de segunda ordem, se o primeiro e o segundo momentos estatísticos existam, e são constantes em relação à origem, isto é,

$$E[Z(s)] = m_1(s) = m_1$$

e

$$E[Z^2(s)] = m_2(s) = m_2, \quad (1)$$

em que:

$E[Z(s)]$ é a esperança matemática da variável regionalizada Z ;

$E[Z^2(s)]$ é o segundo momento ordinário da variável regionalizada Z ;

$$m_i(s) = \frac{\sum_{j=1}^n Z_j^i(s)}{n} \text{ é o momento de ordem } i = 1, 2, 3, \dots, r, \text{ sendo } n$$

pontos amostrais e r a ordem do momento;

m_1 e m_2 são constantes;

Logo, a variância da variável regionalizada Z , se existir, é dada por:

$$VAR[Z(s)] = E[Z^2(s)] - [E(Z(s))]^2 = m_2(s) - [m_1(s)]^2 = m_2 - m_1^2 = \sigma^2. \quad (2)$$

Conforme CARVALHO (2004), um exemplo de estacionaridade são as oscilações de tensão em uma rede elétrica. Esta tensão não pode ter mudanças bruscas, pois danificaria diversos aparelhos elétricos a ela ligados.

Conforme VIEIRA (1998), satisfazendo a hipóteses de estacionaridade de segunda ordem, já é suficiente para o emprego da geoestatística. Porém, alguns fenômenos físicos, como a concentração de ouro, apresentam uma variância infinita de dispersão, então se deve aplicar a hipótese intrínseca, pois é menos restritiva, necessitando apenas da estacionaridade do semivariograma.

VIEIRA et al. (2000) afirma que a existência de estacionaridade permite a repetição de um experimento, mesmo realizando a amostragem, processo no qual são coletadas as amostras, em pontos diferentes dos pontos iniciais, pois todas as posições são consideradas pertencentes às populações com os mesmos momentos estatísticos.

Um processo estocástico, segundo HOEL et al. (1972), é intrinsecamente estacionário, se sua média é constante e a variância de $Z(s+h) - Z(s)$ depende apenas da distância h , conforme as Equações (3) e (4) apresentadas respectivamente:

$$E[Z(s)] = m_1(s) = m_1; \quad (3)$$

$$Var[Z(s+h) - Z(s)] = E\{[Z(s+h) - Z(s)]^2\} = 2\gamma(h). \quad (4)$$

Sendo $\gamma(h)$ uma função que depende apenas de h , conhecida como semivariância.

2.2.2 Semivariogramas

Conforme ISAACS & SRIVASTAVA (1989), o semivariograma é uma técnica estatística usada na verificação da presença de dependência espacial entre os pontos amostrais georreferenciados espacialmente, representado por um gráfico da função semivariância contra 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.

Segundo GUERRA (1988) há três tipos de semivariograma: o semivariograma verdadeiro, o semivariograma experimental e o semivariograma teórico. O semivariograma verdadeiro apresenta a dependência real que é desconhecida; o semivariograma experimental é obtido por meio dos pontos amostrais e o semivariograma teórico é o semivariograma ajustado mediante um modelo teórico ao semivariograma experimental.

A função semivariância $\gamma(h)$ é uma importante estatística que é utilizada para a construção do semivariograma experimental e é definida na Equação (5).

$$\gamma(h) = \frac{1}{2} E[Z(s) - Z(s+h)]^2 \quad (5)$$

Os estimadores da função semivariância $\gamma(h)$ definida em (5) mais conhecidos são: o Estimador Clássico de Matheron (MATHERON, 1963) e o Estimador de Cressie & Hawkins (CRESSIE & HAWKINS, 1980) que são apresentados nas Equações (6) e (7), respectivamente.

$$\hat{\gamma}_M(h) = \frac{1}{2N(h)} \sum^{N(h)} [Z(s) - Z(s+h)]^2 \quad (6)$$

$$\hat{\gamma}_c(h) = \frac{\left\{ \frac{1}{N(h)} \sum |Z(s) - Z(s+h)|^2 \right\}^4}{0,914 + \frac{0,988}{N(h)}} \quad (7)$$

em que:

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

Z : é a variável regionalizada ou função aleatória;

$s = (x_i, y_j) \in S \subset R^2$; são as posições amostrais.

Segundo RIBEIRO JUNIOR (1995), o estimador de Matheron é pouco resistente e muito afetado por observações atípicas, os *outliers*. Quando ocorrem estes casos, é necessário aplicar um estimador mais robusto que não seja afetado por pontos discrepantes e que não possam ser eliminados ou substituídos.

ROMAN (2005) afirma que alguns autores, em alguns casos, retiram do conjunto dos dados observados valores discrepantes (valores atípicos quando comparados aos demais) ou aplicam transformações, visando obter uma distribuição normal de probabilidade, isentando assim o conjunto total de dados discrepantes, que possam afetar e influenciar a distribuição dos dados.

Segundo SRIVASTAVA & PARKER (1989), existem estimadores que podem ser usados para reduzir o efeito sofrido pelo estimador de Matheron, e um deles é o estimador de Cressie & Hawkins (outros são New 1 e New 2), pois fornece uma estimativa mais estável do semivariograma. O estimador de semivariância relativa Pairwise (SRIVASTAVA & PARKER, 1989; ISAAKS & SRIVASTAVA, 1989) é apresentado na Equação (8):

$$\hat{\gamma}_p(h) = \frac{2}{N(h)} \sum \left[\frac{Z(s) - Z(s+h)}{Z(s) + Z(s+h)} \right]^2 \quad (8)$$

LI & LAKE (1994) propõem 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 (9) e (10), respectivamente:

$$\hat{\gamma}_{N1}(h) = \frac{1}{n} \sum_{i=1}^n \left\{ \frac{1}{2m} \sum_{j \in D_{i,h}} [Z(s_i) - Z(s_j)]^2 \right\} \quad (9)$$

$$\hat{\gamma}_{N2}(h) = \hat{\gamma}_{N1}(h) + \frac{h}{d} \hat{\gamma}'_{N1}(h) \quad (10)$$

em que:

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

$D_{i,h}$: é um conjunto indexado dos valores da variável regionalizada Z, em pontos de uma janela móvel bi-dimensional 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 segundo momento ordinário da semivariância.

Na Figura 4, é representada a janela bi-dimensional $D_{i,h}$ na área hachurada, $D_{i,h}$ é o conjunto de pontos s_j e o elemento s_i é o que se encontra no centro.

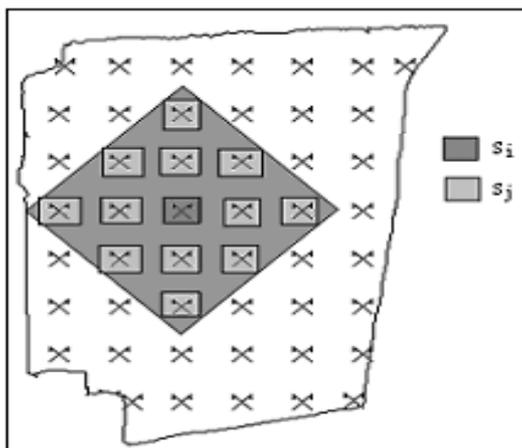


Figura 4 Exemplo de janela Móvel Bi-Dimensional.

Além do modelo de semivariograma experimental escolhido para ser usado na estimação, é necessário determinar o valor ideal do *Cut-off*, ou ponto de corte. O valor determinado refere-se à distância máxima utilizada no cálculo da semivariância. Segundo CLARCK (1979), geralmente, usa-se valores proporcionais à distância máxima, como 1/4, 1/3 ou 1/2, porém podem ser utilizados valores diferentes dos citados, mas é necessário que os valores adotados sejam escolhidos com bom senso.

Ainda não existe um consenso na literatura com relação ao número mínimo exato para que os dados apresentados pelo semivariograma sejam confiáveis. Segundo JOURNEL & HUIJBREGTS (1978) e WOLLENHAUPT et al. (1997), deve utilizar-se no mínimo 30 pares, enquanto WEBSTER & OLIVER (1990) recomendam um número mínimo de pares superior a cem.

A Figura 5 apresenta a estrutura clássica de um semivariograma, composto de seus parâmetros alcance (a), efeito pepita (C_0), contribuição (C_1) e patamar ($C = C_0 + C_1$), que são de grande importância para a construção de mapas de contorno confrontando com o modelo, conforme ISAAKS & SRIVASTAVA (1989).

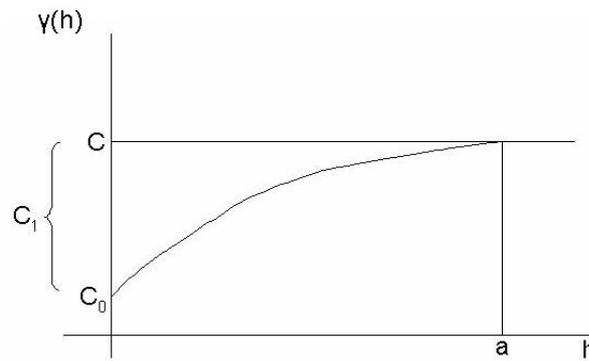


Figura 5 Gráfico de um Semivariograma Experimental.

Segundo CRESSIE (1993), o alcance (a) representa a distância máxima onde as amostras são correlacionadas, onde as distâncias são superiores ao valor do alcance, diz-se que as amostras não apresentam dependência. Então, o valor do alcance delimita a região máxima com dependência espacial. JOURNEL & HUIJBREGTS (1978) relatam que o alcance fornece uma idéia da região de influência da variável regionalizada. CAMARA & 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 que houve erros de análise ou onde a variabilidade de pequena escala não foi detectada pela amostragem.

O patamar, (C , em que $C = C_0 + C_1$), 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 onde um ponto com valor máximo estabilizado, afirmam ISAACS & SRIVASTAVA (1989).

Conforme KAVANAGH (2001), os parâmetros alcance, efeito pepita e patamar, são os meios nos quais se procura quantificar a dependência espacial da propriedade em estudo.

2.2.2.1 Modelagem e Ajuste de Semivariogramas

O Modelo Linear com patamar é um modelo válido apenas em R (espaço unidimensional), não sendo válido para dimensões mais abrangentes (CRESSIE, 1993). Este modelo é definido pela Equação 11. A Figura 6 apresenta o gráfico correspondente ao modelo linear:

$$\gamma(h) = \begin{cases} C_0 + C_1 \left(\frac{h}{a}\right) & , \text{se } 0 \leq h \leq a \\ C_0 + C_1 & , \text{se } h > a. \end{cases} \quad (11)$$

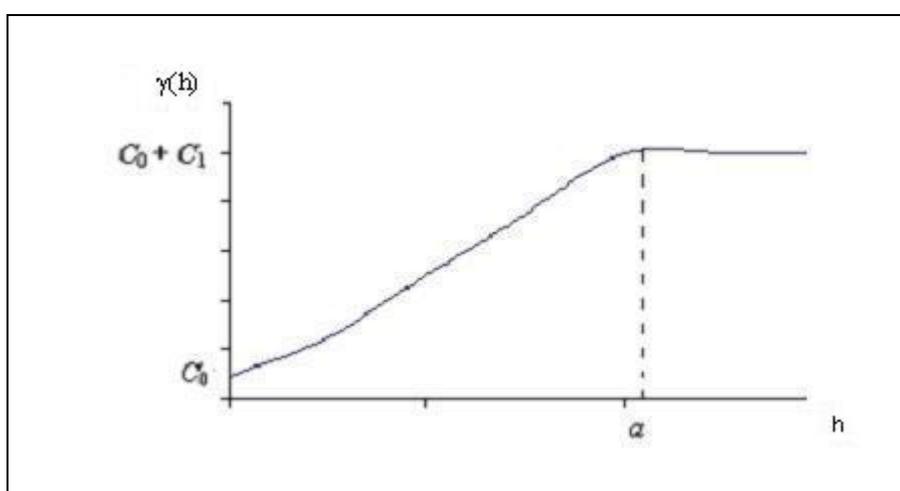


Figura 6 Gráfico do modelo linear.

O Modelo Circular apresenta validade apenas nos planos unidimensionais e bidimensionais, não podendo ser aplicado a planos tridimensionais, onde é aplicado o modelo esférico. O modelo circular é representado pela Equação 12, abaixo e seu gráfico está presente na Figura 7:

$$\gamma(h) = \begin{cases} C_0 + C_1 \left[1 - \frac{\pi}{2} \cos^{-1} \left(\frac{h}{a} \right) + \frac{2h}{\pi a} \sqrt{1 - \frac{h^2}{a^2}} \right] & , \text{se } 0 \leq h \leq a \\ C_0 + C_1 & \text{se } h > a \end{cases} \quad (12)$$

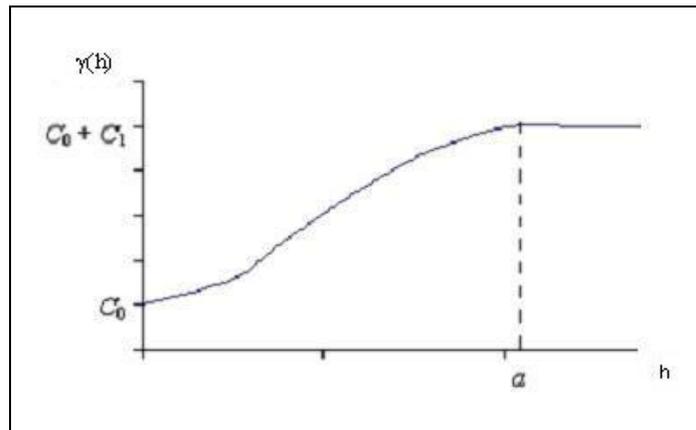


Figura 7 Gráfico do modelo circular.

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. Este modelo é dado pela Equação 13:

$$\gamma(h) = \begin{cases} C_0 + C_1 \left[\frac{3}{2} \left(\frac{h}{a} \right) - \frac{1}{2} \left(\frac{h}{a} \right)^3 \right] & , \text{se } 0 \leq h \leq a \\ C_0 + C_1 & \text{se } h > a \end{cases} \quad (13)$$

A Figura 8 apresenta o gráfico do modelo esférico.

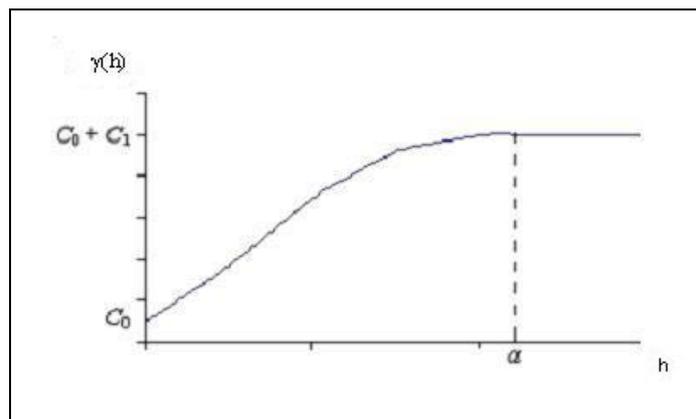


Figura 8 Gráfico do modelo esférico.

O Modelo Exponencial é diferente do modelo esférico pois apresenta uma continuidade nos primeiros valores de h. Neste modelo, que é definido pela Equação 14, o patamar é atingido apenas assintoticamente, pois a curva

não chega a tocar C_1 . O gráfico para este modelo está apresentado na Figura 9:

$$\gamma(h) = \begin{cases} C_0 + C_1 \left[1 - e^{-3\frac{h}{a}} \right] & \text{Se } 0 < u \leq a \\ C_0 + C_1 & \text{Se } u > a \end{cases} \quad (14)$$

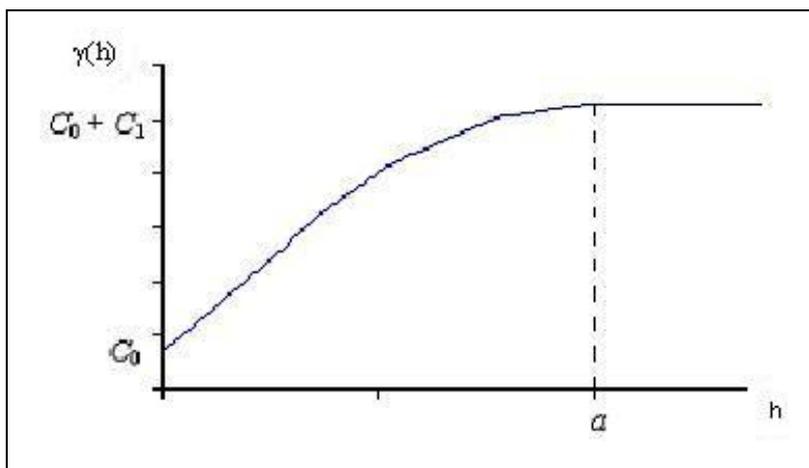


Figura 9 Gráfico do modelo exponencial.

O Modelo Gaussiano apresenta um alcance longo, já seu patamar apresenta comportamento semelhante ao modelo exponencial. Este modelo apresenta um comportamento parabólico próximo à origem e, dentre os vários modelos apresentados, é o único que apresenta um ponto de inflexão na sua curva. É definido pela Equação 15. O gráfico do modelo gaussiano pode ser visto na Figura 10.

$$\gamma(h) = \begin{cases} C_0 + C_1 \left[1 - e^{-3\left(\frac{h}{a}\right)^2} \right] & \text{Se } 0 < u \leq a \\ C_0 + C_1 & \text{Se } u > a \end{cases} \quad (15)$$

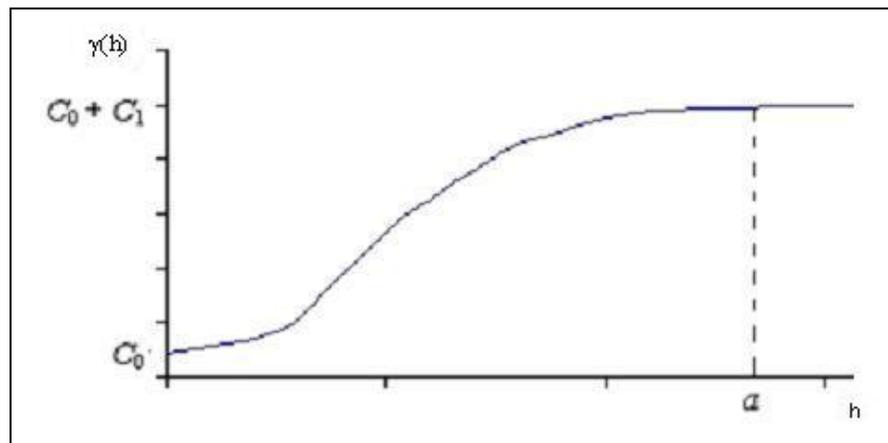


Figura 10 Gráfico do modelo gaussiano.

2.2.3 Anisotropia e Isotropia

De acordo com CRESSIE (1993), quando semivariogramas construídos para direções diferentes apresentam características semelhantes, o que sugere um mesmo padrão de continuidade espacial, o evento em questão é considerado isotrópico. Quando o fenômeno é isotrópico, o estudo da dependência pode ser feito através do semivariograma omnidirecional, onde as semivariâncias são calculadas baseadas em informações de todas as direções, considerando apenas as distâncias.

Porém, caso semivariogramas construídos para diferentes direções apresentem diferentes padrões de dependência espacial, o fenômeno é considerado anisotrópico. A literatura recomenda trabalhar com as direções 0° , 45° , 90° e 135° . Três abordagens particulares são apontadas por ISAAKS & SRIVASTAVA (1989): a anisotropia geométrica e a anisotropia zonal, a mista e combinada.

Segundo RENDU (1978), a anisotropia geométrica, um ou mais semivariogramas direcionais possuem alcances diferentes, porém um mesmo valor para o patamar. Já no caso da anisotropia zonal, o inverso ocorre, já que os valores dos patamares são distintos e os valores dos alcances são semelhantes.

CARVALHO (2004) realizou um estudo da anisotropia em uma área agrícola espacialmente referenciada, verificando que, quanto maior a intensidade da anisotropia melhor será representada a continuidade espacial nos mapas construídos.

2.2.4 Krigagem

O processo de interpolação pode ser aplicado em diversas áreas com características distintas. Entre estas, podem ser destacadas (GUERRA, 1988):

- Fornecer mapas de isolinhas utilizados para visualização gráfica, permitindo a análise e tomada de decisões baseadas nesta representação;
- Calcular atributos da superfície numa determinada posição, ou em várias posições;
- Alterar a unidade espacial que permite posterior comparação quando utilizadas estruturas diferentes de dados em diferentes layers;
- Servir de auxílio no processo de decisão espacial, como correção do solo, prospecção mineral ou de petróleo, entre outros.

O processo de estimação e interpolação adotado neste trabalho será o método da krigagem ordinária, visto que é um estimador considerado ótimo. É reconhecido como ótimo, por produzir estimativas não viciadas e com variância mínima, referida, segundo RIBEIRO JUNIOR (1995) como *best linear unbiased predictor* (BLUE). A krigagem, como técnica eficiente de interpolação, permite a construção de mapas para explicar-se ou estudar determinado fenômeno. 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 e o ponto estimado e também na distância entre estes pontos (CRESSIE, 1993).

A krigagem é dita linear uma vez que o valor predito pelo método ($Z^*(s_0)$) de determinada variável em uma posição qualquer é a combinação linear dos valores desta mesma variável em pontos amostrados:

$$Z^*(s_0) = \sum_{i=1}^n \lambda_i Z(s_i) \quad (16)$$

Segundo CARVALHO & ASSAD (2005), a krigagem, como método geoestatístico de interpolação, é muito utilizada no estudo da distribuição espacial de precipitação pluviométrica. O aspecto fundamental que diferencia os interpoladores geoestatísticos dos demais é a utilização da correlação espacial entre as observações vizinhas para prever valores em locais não amostrados.

Em seu trabalho, CARVALHO & ASSAD (2005) realizaram uma comparação entre estes três métodos de interpolação. O resultado verificado apontou a krigagem como mais eficiente, enquanto os demais métodos mostraram equivalência, quanto à eficiência demonstrada, porém o método do inverso do quadrado das distâncias possui implementação muito mais simples que o método da curvatura mínima.

2.2.5 Indicadores de Melhor Ajuste

A seguir, serão apresentadas técnicas para obter o melhor modelo ajustado ao semivariograma experimental.

2.2.5.1 Validação Cruzada

VIEIRA (1998) descreve alguns procedimentos que podem ser utilizados para validar o ajuste de um modelo teórico e seus respectivos parâmetros. Tais procedimentos têm como base a técnica conhecida como Validação Cruzada (VC), a qual consiste em ignorar, um a um, cada valor

amostrado e predizer 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.

Assim, a Validação Cruzada é utilizada na tarefa de identificar qual modelo melhor representa o fenômeno em estudo (DAVIS, 1987).

RIBEIRO JUNIOR (1995) afirma que, após conhecidos os valores obtidos da amostragem e os valores estimados, pode ser conhecida, também, a variância total da estimativa, o que permite realizar a avaliação da qualidade ou precisão do processo.

Já, segundo VIEIRA (1996), a validação cruzada se apresenta como uma técnica adequada para a avaliação dos procedimentos de modelagem, pois se o fenômeno em estudo puder ser modelado de maneira satisfatória, é possível reproduzir, com boas aproximações, informações da amostra.

Conforme GONÇALVES (1997), depois de obtidos os valores amostrados e estimados para cada ponto, pode-se então calcular o erro de estimação. Espera-se então, que estes erros provenientes do processo de krigagem tenham média nula e variância unitária. As características apresentadas pelo erro de estimação indicam a eficácia do ajuste dos modelos de semivariogramas utilizados, a modelagem do processo em questão, o peso e importância da presença de outliers e, ainda, auxilia na avaliação da estacionaridade.

Apesar de eficiente, a VC apresenta, segundo ISAAKS & SRIVASTAVA (1989), algumas limitações. O processo de validação pode gerar pares de valores amostrados e estimados somente nos locais onde ocorreram as amostras. Conforme os mesmos autores, A VC encontra resultados que normalmente não refletem a real performance de um método de estimação com exatidão, pois a estimação no local amostrado é considerada não-representativa.

A Equação 17 apresenta como são calculados os resíduos na validação cruzada, utilizando o Erro Médio (EM) como métrica de avaliação:

$$VC = \sum_{i=1}^n (Z_i(s) - \hat{Z}_i^*(s)) \quad (17)$$

em que:

$Z_i(s)$ é o valor observado do ponto i em s ;

$\hat{Z}_i^*(s)$ é o valor estimado no ponto i em s , sem considerar a observação $Z_i(s)$.

Outra forma de calcular os resíduos da validação cruzada é calculando o erro médio reduzido (\overline{ER}), conforme a Equação 18, Segundo MCBRATNEY e WEBSTER (1986) e CRESSIE (1993):

$$\overline{ER} = \frac{1}{n} \sum_{i=1}^n \frac{Z_i(s) - \hat{Z}_i(s)}{\sigma_i(s)} \quad (18)$$

em que:

$\sigma_i(s)$: desvio padrão da krigagem no ponto s , sem considerara a observação $Z_i(s)$.

O desvio padrão dos erros reduzidos foi obtido a partir da Equação 19:

$$S_{ER} = \sqrt{\frac{1}{n} \sum_{i=1}^n \frac{|Z_i(s) - \hat{Z}_i(s)|}{\sigma_i(s)}} \quad (19)$$

Segundo MCBRATNEY e WEBSTER (1986), CRESSIE (1993) e MELLO et al. (2005), o valor de EM e \overline{ER} mais próximo de zero, o valor do desvio padrão menor e o valor de S_{ER} mais próximo de um são os critérios para escolha do melhor modelo ajustado.

O erro absoluto (EA) é usado para medir a amplitude 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, foi possível definir o erro absoluto na unidade da variável em estudo através da Equação 20:

$$EA = \sum_{i=1}^n |\hat{z}(s_i) - z(s_{(i)})| \quad (20)$$

2.2.5.2. Critério de Akaike

O teorema de Akaike busca a otimização e também a menor complexidade, pois ele visa demonstrar que dois modelos distintos representam igualmente dados de maneira satisfatória. Uma vez que os dois modelos apresentam fielmente os dados, a opção mais óbvia é optar pelo mais simples. Este estudo visa conhecer como os modelos são utilizados para fazer previsões, a partir de AKAIKE (1973) e SAKAMOTO, ISHIGURO & KITAGAWA (1986).

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. O Critério de Informação de Akaike (AIC – Akaike's Information Criterion) é utilizado para procurar uma solução satisfatória entre um bom ajuste e a parcimônia. A Equação 18 estima a parte variável de AIC par um conjunto de dados.

$$\hat{A} = n \ln(R) + 2p \quad (21)$$

em que:

n: é o número de pontos experimentais do semivariograma;

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

$\hat{Z}_i(s)$: 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.

2.2.5.3. Critério de Filliben

O teste de Filliben (FILLIBEN, 1975) é realizado na verificação dos resíduos ortogonais provenientes do processo de validação cruzada e contribui, positivamente, na determinação de qual o melhor modelo geoestatístico a ajustar. Estes resíduos ortogonais 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.

Depois de obtidos um valor amostrado e um estimado para cada ponto, é possível calcular a regressão linear entre estes pares de dados e calcular o coeficiente linear a , o coeficiente angular b (visto que o processo de regressão linear pode ser representado por uma reta que possui coeficiente angular e linear) a correlação r^* entre os pares e o erro reduzido com sua média e variância. O melhor ajuste será obtido, quando os valores encontrados se aproximarem dos valores ideais: $b = 1$, $a = 0$, $r^* = 1$, média do erro zero e variância um.

A Equação 19 apresenta como são calculados os resíduos ortonormais:

$$\varepsilon_k = \frac{Z(s_k) - Z_e(s_k)}{\sigma_k}, \quad k = p+1, \dots, n \quad (22)$$

em que:

ε_k : são os resíduos normais;

$Z(s_k)$: são os valores amostrados;

$\hat{Z}_e(s_k)$: são os valores estimados pela krigagem;

σ_k : é o desvio padrão da k-ésima estimativa da krigagem.

Para calcular o valor de r^* , utiliza-se a Equação 20, conhecida como cálculo do coeficiente de correlação linear.

$$r^* = \frac{\sum_{i=1}^n (W_i - \bar{W})(M_i - \bar{M})}{\sqrt{\left[\sum_{i=1}^n (W_i - \bar{W})^2 \right] \left[\sum_{i=1}^n (M_i - \bar{M})^2 \right]}} \quad (23)$$

em que:

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).

2.3 Principais Softwares Geoestatísticos

Atualmente, existem diversos softwares Geoestatísticos em uso por pesquisadores, dentre os quais se podem destacar os softwares GSLIB, GSTAT, CRIMESTAT, SPACESTAT, SPRING e ARCGIS.

O GSLIB (Geoestatistical Software Library) é um pacote de programas desenvolvido na Universidade de Stanford (EUA), em FORTRAN. Este software é voltado, principalmente, para alunos de pós-graduação e pesquisadores que possuam alguma experiência no tratamento geoestatístico de dados. Uma desvantagem do GSLIB é a sua interface pouco amigável, visto

que não é gráfica e exige a utilização de utilitários para visualização dos resultados. Os código-fonte, arquivos executáveis, arquivos de dados e links de utilitários podem ser encontrados e obtidos no endereço <http://www.gslib.com>.

O GSTAT é um software de código aberto para modelagem e predição geoestatística multivariável e simulação de variáveis em 1, 2 ou 3 dimensões. Este software foi desenvolvido em 1997 e desde então vem sendo aplicado a vários GIS (Geographic Information Systems). Informações, manuais, versões e o código fonte do GSTAT podem ser encontrados no endereço <http://www.gstat.org>.

O CRIMESTAT é um software desenvolvido com o intuito de auxiliar no combate ao crime. É uma aplicação autônoma que analisa estatísticas espaciais da localização de incidência de crimes. Este software é gratuito para analistas, pesquisadores, educadores e estudantes. O código fonte é escrito em C++ e permite multi-threading (processamento paralelo). O programa, bem como seus manuais podem ser adquiridos pelo endereço <http://www.oip.usdoj.gov/cmrc>.

O SPACESTAT fornece muitas funções para análise espacial. Apesar de eficiente, apresenta uma interface pobre, visto que é desenvolvido em ambiente DOS. Uma abordagem mais completa sobre o software pode ser encontrada em ANSELIN (1992). Vários manuais e exemplos podem ser obtidos através do endereço www.spacestat.com.

Um dos softwares de maior sucesso e mais utilizados atualmente é o SPRING, visto sua capacidade gráfica de resultados e a gama de funções que possibilita aos usuários. Esta ferramenta permite construção de gráficos, processamento de imagens, análise espacial, modelagem numérica de terreno e consulta a banco de dados espaciais. O SPRING foi desenvolvido pelo INPE (Instituto Nacional de Pesquisas Espaciais), através da Divisão de Processamento de Imagens. A obtenção do software, bem como manuais do mesmo, pode ser feita através do endereço <http://www.dpi.inpe.br/spring/>.

Já, o ARCGIS oferece um conjunto integrado de softwares de Sistemas de Informações geográficas (SIG) para a construção de um SIG mais completo. Este software apresenta uma série de características interessantes como facilidade de uso, uma vez que possui uma interface amigável ao usuário, é bastante funcional, escalável e permite customização. Porém, um fator indesejado neste software é o fato de que é pago. O usuário pode escolher dentre vários modelos existentes, qual se encaixa no perfil requisitado por ele. A aquisição do ARCGIS (através de pagamento) pode ser feita pelo endereço <http://www.esri.com/software/arcgis/>.

R é uma linguagem e um ambiente voltado para a computação estatística e gráfica. Foi desenvolvido no Bell Laboratories (formalmente conhecido como AT&T, hoje chamado Lucent Technologies) por John Chambers e seus colegas. É uma linguagem livre, permitindo aquisição gratuita e também alterações em seu código. R fornece uma grande variedade de funções estatísticas (modelos linear e não-linear, testes estatísticos clássicos, análise de série temporais, classificação...) e as técnicas gráficas. Além disso, como ela tem código aberto, é altamente extensível. O R pode ser adquirido pelo endereço <http://www.r-project.org/>.

Quanto ao GEOCAC, é um software geoestatístico desenvolvido por pesquisadores do Laboratório de Estatística Aplicada (LEA) da Universidade Estadual do Oeste do Paraná (UNIOESTE). Este software possui interface bastante amigável ao usuário e permite uma extensa gama de técnicas estatísticas para análise de dados. Não possui custo para aquisição e pode ser adquirido através do laboratório supracitado.

3 MATERIAL E MÉTODOS

O desenvolvimento do software foi realizado na linguagem Delphi 6 ®, pois esta oferece diversas funcionalidades interessantes ao desenvolvimento, como orientação a objetos, interface gráfica bastante ergonômica ao usuário e, principalmente, porque as versões iniciais do GEOCAC (SILVA et al., 2002) já haviam sido implementadas nesta linguagem.

Como visto, o objetivo deste trabalho consistiu em tornar o software GEOCAC mais eficiente e apropriado ao desenvolvimento de pesquisas na área. Para tanto, foram realizadas correções na versão anterior, visando adaptar e corrigir o projeto possuído e, então, desenvolver novas funções que, posteriormente, foram inseridas no software, buscando torná-lo mais completo e abrangente.

O método utilizado para a construção do software segue o modelo espiral (conforme a Figura 11), apresentado por PRESSMAN (2006). Este modelo, que recebe o nome de espiral devido à sua similaridade com uma forma eliquoidal, é dividido em quatro etapas subseqüentes, abordadas a seguir.

O planejamento (primeira etapa) foi realizado no início do projeto. Foram delimitados os objetivos e estratégias que foram adotados durante a execução do trabalho.

A segunda etapa, a análise de riscos, tratou-se também ao início do desenvolvimento, pois os riscos devem ser tratados seriamente e eficientemente. A terceira etapa consiste basicamente na implementação propriamente dita, onde o projeto foi construído.

A quarta e última etapa é a avaliação do software. Esta ficou incumbida aos pesquisadores da área, que utilizarão o software em suas pesquisas, quando ele estiver terminado.

Esta metodologia (espiral) mostrou-se vantajosa devido ao contato entre desenvolvedor e pesquisadores, o que permitiu uma interação eficiente e satisfatória. Outro fator favorável ao desenvolvimento do trabalho foi a possibilidade da implementação em módulos individuais. Este fato permitiu que os módulos pudessem ser avaliados e testados separadamente, possibilitando, assim, maior eficiência e confiabilidade nos resultados obtidos. Não foram utilizadas ferramentas com o intuito de verificar a qualidade do software. Este foi modelado de acordo com as necessidades e aplicações solicitadas pelos usuários do sistema. Desta forma, acredita-se ter alcançado um estado satisfatório com relação à qualidade, pois, durante o processo de desenvolvimento, o software foi submetido a avaliações e, conseqüentemente, alterações, visando melhorar não só a eficiência, mas também torná-lo mais adequado aos pesquisadores.

A Figura 11 apresenta o diagrama do processo de desenvolvimento seguindo o modelo espiral. É interessante observar que a interação entre o processo de implementação e avaliação é bidirecional. Este fato ocorre devido à possibilidade do software não satisfazer plenamente e haver a necessidade de alterações e correções.



Figura 11 Processo de desenvolvimento através do modelo espiral.

Para uma melhor compreensão das modificações realizadas neste projeto, acreditou-se que uma estrutura modular seria a mais indicada. Logo, a metodologia foi dividida de acordo com a característica das funções.

3.1 Módulo 1 (Entrada de Dados)

As alterações realizadas sobre a entrada de dados tiveram por objetivo melhorar a interação e a eficiência do software, permitindo desta forma, novas possibilidades de operação e principalmente o aumento na capacidade de trabalhar com um conjunto maior de dados.

Em um primeiro momento foi corrigido o funcionamento do botão importar, que anteriormente não realizava o processo de abrir e carregar um arquivo de dados corretamente.

Através de reuniões com pesquisadores que utilizam o GEOCAC, verificou-se que os arquivos com dados não deveriam, apenas, ficar restritos às planilhas do Excel®, mas serem carregados, também, de um arquivo texto. Esta funcionalidade foi implementada para oferecer mais opções de interação e, conseqüentemente, maior eficiência do software. Esta função é apresentada na Figura 12 (no destaque em vermelho) representada pelo caractere W.

Outra alteração de grande importância foi com relação ao tamanho do conjunto de dados de entrada. Na versão inicial do GEOCAC, eram permitidos apenas arquivos contendo cerca de trezentos pontos. Após a alteração, o programa trabalha agora com uma gama consideravelmente maior, com cerca de oito mil pontos amostrais, permitindo analisar dez variáveis distintas para cada ponto.

A última função construída, abordando a entrada de dados, foi a tarefa de carregar um arquivo com coordenadas de borda para a construção de mapas de fronteira, que são mapas construídos para representar a linha fronteira da região em estudo. Para tanto, são necessários os pontos que constituem a linha externa da área. A versão inicial do software permitia apenas a entrada de dados em extensões do tipo xls. Já a nova versão possibilita a operação com dados no formato txt.

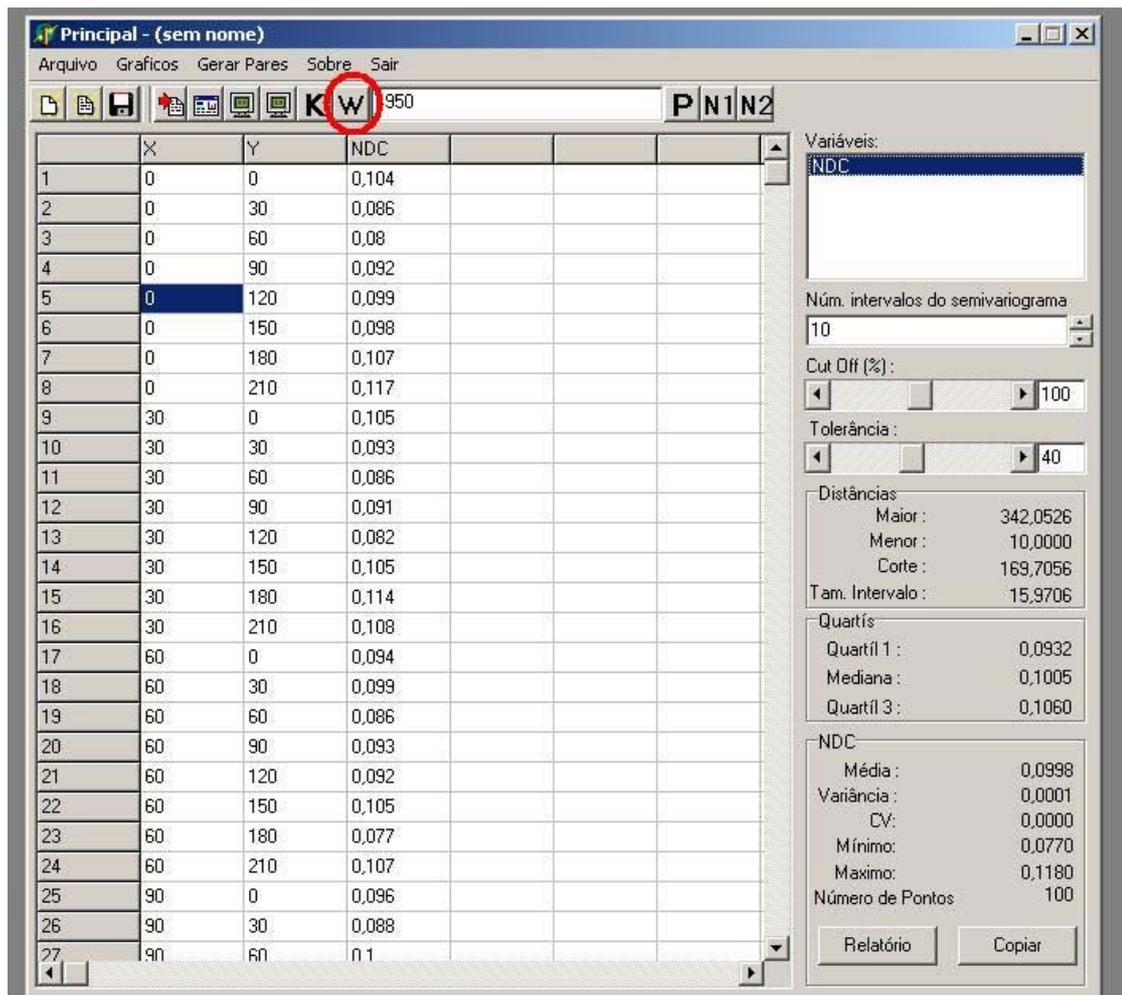


Figura 12 Tela principal do GEOCAC.

A Figura 13 apresenta um diagrama do funcionamento da entrada de dados para posterior análise.

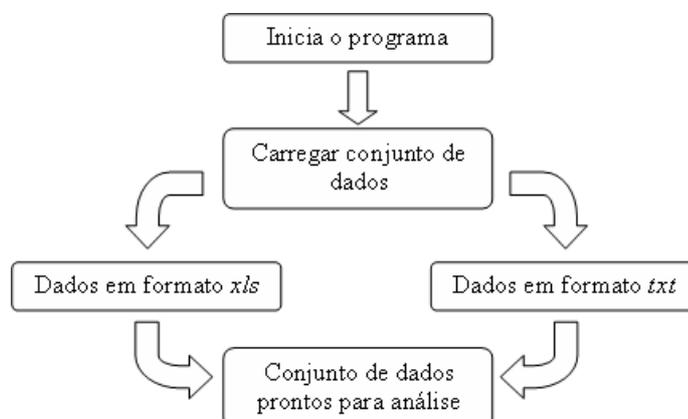


Figura 13 Entrada de Dados do GEOCAC.

3.2 Módulo 2 (Saída de Dados)

Esta etapa teve como objetivo melhorar a formatação dos dados fornecidos ao usuário, com o intuito de torná-los mais facilmente compreensíveis e ilustrativos.

Foi implementada a opção de salvar todos os gráficos gerados pelo software, permitindo, dessa forma, que os mesmos possam ser utilizados posteriormente, independentes do GEOCAC. Os gráficos também sofreram alterações em sua formatação, visando torná-los mais claros e sucintos. Entre as mudanças, pode-se citar a inclusão de mais casas decimais nos valores presentes nos gráficos, a mudança na opção das cores para obter um melhor contraste na visualização, a atualização das variáveis em estudo.

Relatórios com a análise descritiva do conjunto de dados foram criados, permitindo que o usuário possa imprimir ou exportar os dados estatísticos para outras aplicações.

3.3 Módulo 3 (Estimadores da Função de Semivariância)

Esta etapa teve como objetivo acrescer à versão anterior do GEOCAC os estimadores de semivariograma de Pairwise, New 1 e New 2 propostos por SRIVASTAVA & PARKER (1989) e por LI & LAKE (1994), respectivamente.

Para tornar estes novos estimadores independentes do restante do software, optou-se pela sua implementação de maneira separada, facilitando, dessa forma, correções necessárias e a opção de utilizar um conjunto de dados consideravelmente grande (algo que a versão anterior não possibilitava).

Foram construídas, também, opções que permitem ao usuário determinar Alcance (a), Efeito Pepita (C_0) e Contribuição (ou sill, C_1) ao

construir os semivariogramas e variar estes parâmetros de forma a ajustar o melhor conjunto de valores para obter uma melhor representação da distribuição.

Para melhor interpretação dos valores utilizados no processo de construção dos semivariogramas, foram implementados gráficos que ilustram a curva do modelo geoestatístico escolhido, bem como a curva representada pelos pares provenientes do conjunto de dados.

3.3.1 Estimador de Pairwise

A construção dos estimadores para ajuste aos semivariogramas mostrou-se tarefa desafiadora perante a implementação, visto que além da confiabilidade dos resultados buscou-se a eficiência, de maneira que o software pudesse trabalhar com um número maior (cerca de 8000) de pontos amostrais.

Para construir este estimador (Pairwise), foi utilizada a equação 8 apresentada na Seção 2.2.2 em que o valor encontrado é chamado gama. Esta equação armazena o quadrado das diferenças entre os valores amostrados e, em seguida, divide este valor pelo número de pares para cada distância diferente. O processo será melhor compreendido na seqüência.

O primeiro passo para a elaboração do estimador de Pairwise foi construir um vetor dinâmico com tamanho inicial suficientemente grande para armazenar todos os pares possíveis.

Em seguida, foram calculadas todas as distâncias possíveis entre os pares com o intuito de encontrar a maior distância possível entre dois pontos amostrais distintos. Para então, de acordo com o parâmetro *Cut-off* (Ponto de corte, determinado pelo usuário no início do processo), determinar a distância máxima aceita, dentre todas as distâncias obtidas. Este corte faz com que as distâncias maiores que o *Cut-off* sejam ignoradas.

Então, são calculados os valores de $\gamma(h)$ para cada distância possível (para todas as distâncias inferiores ao ponto de corte). Este processo é realizado comparando-se todos os pares entre si. Para cada distância distinta, são armazenados o valor de gama e também o número de pares que estão separados por esta medida. Dessa forma, ao fim do processo, para obter o valor final de $\gamma(h)$, basta realizar a divisão do valor de gama temporário armazenado pelo número de pares para aquela respectiva distância. Este processo de comparação entre todos os pontos amostrados torna o método consideravelmente lento, variando exponencialmente de acordo com o número de pontos amostrados. Os valores encontrados (h , $\gamma(h)$ e número de pares) são armazenados no vetor construído ao início do processo. A Figura 14 apresenta um esboço da estrutura do vetor citado, onde o campo “Distância” armazena a distância entre os pontos, o campo “Pares” contém o número de pares separados pela distância e $\gamma(h)$ possui o valor de Gama para a distância. “N” representa o número de distâncias possíveis entre os pontos (dentro do ponto de corte). Os valores, bem como a ordenação do vetor, são fictícios, não sendo baseados em um conjunto de dados real.

| | | | | |
|-------------|-----------|-----------|-----|-----------|
| Distância | 30 m | 60 m | ... | 142,302 m |
| Pares | 179 | 124 | | 16 |
| $\gamma(h)$ | 0,0014235 | 0,0014771 | | 0,002592 |
| | 1 | 2 | | N |

Figura 14 Estrutura do Vetor de Pairwise.

Após obtido o valor de gama para todas as distâncias dentro do valor de corte, os dados são plotados em um gráfico bidimensional onde o eixo X é representado pelos valores das distâncias e os valores de Y variam de acordo com os valores encontrados para gama (um valor para cada distância). A

Figura 15 apresenta um exemplo de gráfico com os valores já plotados (neste gráfico o valor de da distância é representado por h).

Para verificar o ajuste entre os valores plotados pelo semivariograma e um modelo conhecido, foram implementadas funções para a construção das curvas destes modelos matemáticos conhecidos. Os modelos adotados foram o Linear, Circular, Esférico, Exponencial e Gaussiano, representados pelas Equações 11, 12, 13, 14 e 15, respectivamente.

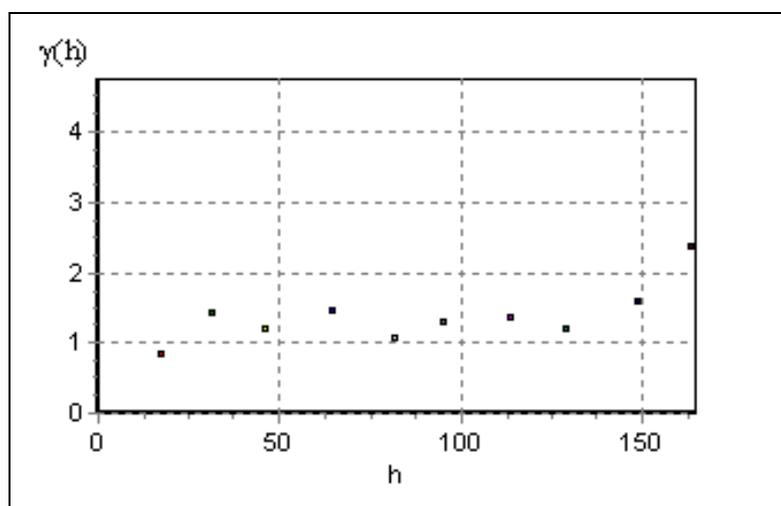


Figura 15 Exemplo de gráfico h versus $\gamma(h)$.

A Figura 16 exibe um gráfico onde se buscou ajustar um modelo matemático ao conjunto de dados do semivariograma. Neste exemplo, foi utilizada a curva do modelo matemático Esférico.

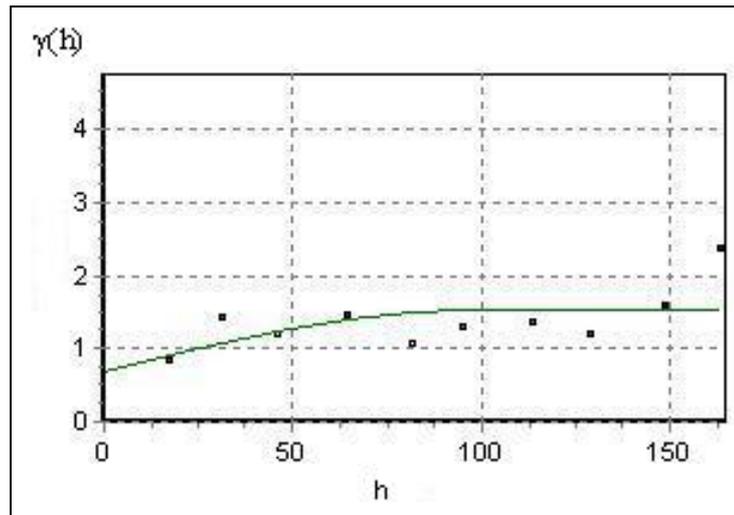


Figura 16 Exemplo de gráfico com busca de ajuste ao modelo esférico.

Esta tentativa de ajuste entre os valores de gama com os valores de algum modelo matemático é utilizada para verificar qual modelo melhor representa os dados e também para encontrar os valores do Efeito Pepita (C_0), Contribuição (C_1) e Alcance (a) que permitam o melhor ajuste entre a curva e os dados.

3.3.2 Estimador New 1

A aplicação do estimador de New 1 possui o mesmo objetivo que os outros estimadores de Matheron, Cressie & Hawkins e Pairwise, porém este estimador apresentado por LI & LAKE (1994) possui a vantagem de ser mais robusto, ou seja, menos suscetível a valores discrepantes (*outliers*) e também por tentar reduzir a imprecisão do estimador quando à distância entre as amostras aumenta. A Equação 9 define o estimador New 1.

Outra diferença encontrada entre os estimadores clássicos e o New 1 é de que o segundo utiliza um mecanismo de janela móvel, onde esta percorre a área amostrada, considerando apenas os valores presentes dentro da janela.

Como pode ser visto na parte interna (região mais escura) da Figura 4, presente na Seção 2.2.2.

Inicialmente, é construído um vetor dinâmico de tamanho suficiente para armazenar todos os valores necessários no processo. Então, é calculada a maior distância entre os pontos amostrados, para, a partir dela e do valor de *Cut-off*, determinar a distância de corte do método.

Após obtida a distância (ou ponto) de corte, são comparados todos os pontos amostrais entre si para calcular todas as distâncias possíveis dentro do limite imposto pelo corte. Estas distâncias serão utilizadas para delimitar o tamanho da janela deslizante utilizada.

O passo seguinte foi calcular o valor de $\gamma(h)$ para cada distância diferente. Neste estimador, são calculados dois somatórios para obtenção do valor de $\gamma(h)$. O primeiro considera todos os pontos dentro da janela em uma determinada posição sobre a área de amostragem. O segundo somatório refere-se a todas as janelas de mesmo tamanho, posicionada em todas as posições da área de amostragem. Ou seja, para cada posição da janela, é realizada combinação dos pontos presentes dentro desta e um somatório. Observando este importante detalhe, pode-se tirar a conclusão de que este método desprende um poder computacional maior.

Após obtidos os valores de $\gamma(h)$ para cada uma das distâncias, os passos seqüentes são similares aos do Estimador de Pairwise apresentado na Seção 3.3.1, construindo o gráfico bidimensional (chamado semivariograma) e a busca pelo melhor ajuste.

A Figura 17 apresenta um exemplo fictício de área amostral, onde o círculo corresponde à janela deslizante. O valor h representa a distância que determina o tamanho da janela. Para cada valor de h , a janela é deslocada sobre toda a área amostral (no exemplo representado pelo retângulo mais externo) e em cada iteração são considerados apenas os pontos amostrados dentro do raio da janela (os pontos amostrais são representados pelos círculos duplos menores).

No contexto da Figura 17, podem-se determinar dois parâmetros utilizados pelo Estimador New 1: o valor de N (que é o número de pontos amostrados em toda a área) que é igual a 26 pontos e o valor de m (que é o número de pontos amostrais posicionados dentro da janela deslizante) que é de nove pontos.

3.3.3 Estimador New 2

A estrutura de construção do estimador New 2 é bastante similar à utilizada pelo estimador New 1. A única diferença está no momento do cálculo do valor de $\gamma(h)$, onde no New 2 são utilizadas as derivadas de um modelo matemático. A opção pelo modelo utilizado no cálculo de $\gamma(h)$ é feita pelo usuário ao início do processo.

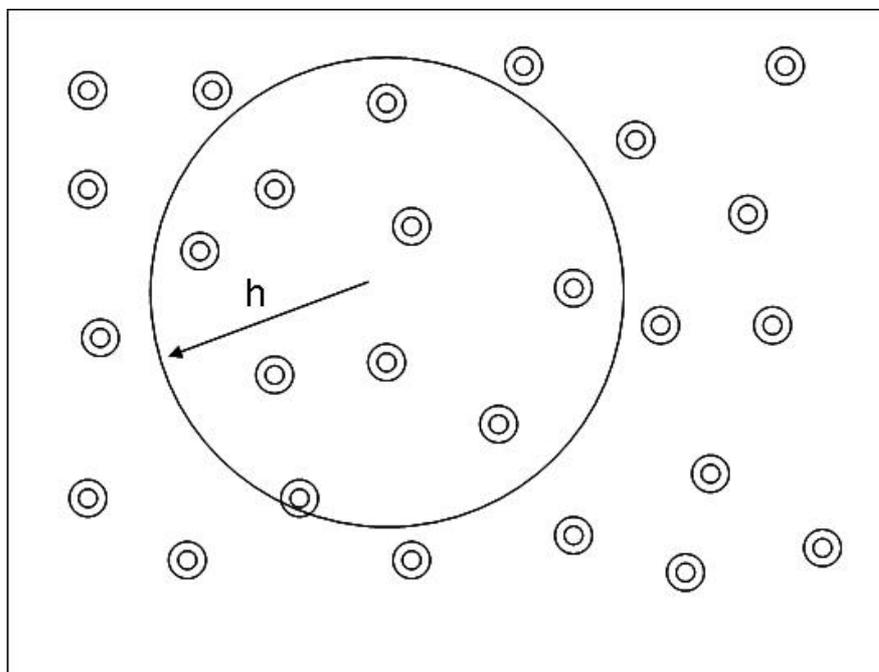


Figura 17 Exemplo de janela deslizante de forma circular.

3.4 Módulo 4 (krigagem)

O processo de krigagem (KRIGE, 1951) tem por objetivo obter estimações em pontos não amostrados, baseado em um conjunto de dados de entrada. Porém, para que seja possível realizar esta etapa, deve-se anteriormente determinar o modelo de semivariograma adotado. Essa necessidade ocorre, pois a Krigagem necessita dos valores dos parâmetros efeito pepita (C_0), patamar ($C_0 + C_1$) e alcance (α), que são obtidos através do ajuste do modelo teórico no semivariograma experimental. Para que estes parâmetros possam ser determinados pela aplicação dos semivariogramas, deve-se, inicialmente, carregar um conjunto de dados com estrutura definida: coordenadas (X,Y), bem como os valores amostrados nestas posições.

Dentro do processo de krigagem, o primeiro passo a ser realizado foi a construção do mapa, contendo os pontos amostrados (de acordo com os dados de entrada). Baseado nesta representação da região de interesse, fica a encargo do usuário determinar qual a posição, dentro da área de amostragem, que será estimada.

Depois de determinada a posição a ser estimada, são encontrados os pontos que se situam dentro do alcance, tomando por referência o ponto escolhido pelo usuário para amostragem. O valor do alcance foi obtido através da modelagem do semivariograma experimental em processo anterior. Então, é construída uma estrutura vetorial armazenando as coordenadas (X,Y) e o valor amostrado da variável de interesse de cada ponto situado dentro do alcance, este vetor foi chamado de Z.

Na seqüência, com este vetor em mãos, é construída a Matriz **H** de ordem $n \times n$, que é uma estrutura bidimensional que armazena as distâncias entre todos os pontos presentes dentro do alcance (não é incluído aqui o ponto a ser estimado).

A etapa seguinte é a construção do vetor $h = (h_1, \dots, h_u)^T$, que é uma estrutura unidimensional (de ordem $n \times 1$) que armazena as distância entre os pontos que se situam dentro do alcance do ponto a ser amostrado.

Possuindo então a matriz bidimensional \mathbf{H} e o vetor h , é possível obter, utilizando o modelo de variabilidade espacial definido na modelagem do semivariograma, a matriz bidimensional de covariância Σ . E, posteriormente, a obtenção do vetor de pesos que permitirá o alcance do valor estimado. A estrutura montada, nesta etapa, pode ser entendida como um sistema linear de equações aritméticas, onde a matriz de covariância “ Σ ” é multiplicada por um vetor de variáveis (pesos) resultando em uma igualdade, que é o vetor obtido através de h' . É interessante ressaltar, que este vetor, utilizado como resultado da multiplicação, é obtido através da aplicação do vetor h' na equação do modelo de variabilidade espacial. A matriz de covariância é obtida aplicando-se esta equação à matriz de distâncias \mathbf{H} .

Uma vez construída a estrutura do SELA (sistema de equações lineares) pôde-se utilizar técnicas de solução de sistemas para encontrar o vetor de pesos. A opção, neste trabalho, foi pelo método de Eliminação de Gauss sem pivoteamento (STEINBRUCH & WINTERLE, 1987). Esta técnica consiste em obter a matriz triangular superior da matriz de covariância e em seguida fazer a substituição com retrocesso, obtendo ao fim do processo os valores dos pesos, armazenados em uma estrutura vetorial.

Então, com os pesos obtidos, o passo subsequente é a aplicação da Equação 16, utilizando como parâmetros o vetor de pesos e o vetor que guarda os valores amostrados nos pontos situados dentro do alcance (vetor Z). Dessa forma, chega-se ao valor estimado para a posição apontada pelo usuário. A Figura 18 apresenta um esboço do processo de krigagem

Para avaliar se o método empregado chegou a resultados confiáveis e se eficientes indicadores podem ser utilizados, estes, neste trabalho, são representados pela validação cruzada, critério de Akaike e Filliben, representados pelas Equações 17, 18 e 19, respectivamente.

O primeiro método implementado foi o estimador de Pairwise, devido à sua semelhança com os estimadores tradicionais de Matheron e Cressie & Hawkins. Em seguida, foram implementados os estimadores New 1 e New 2, respectivamente.

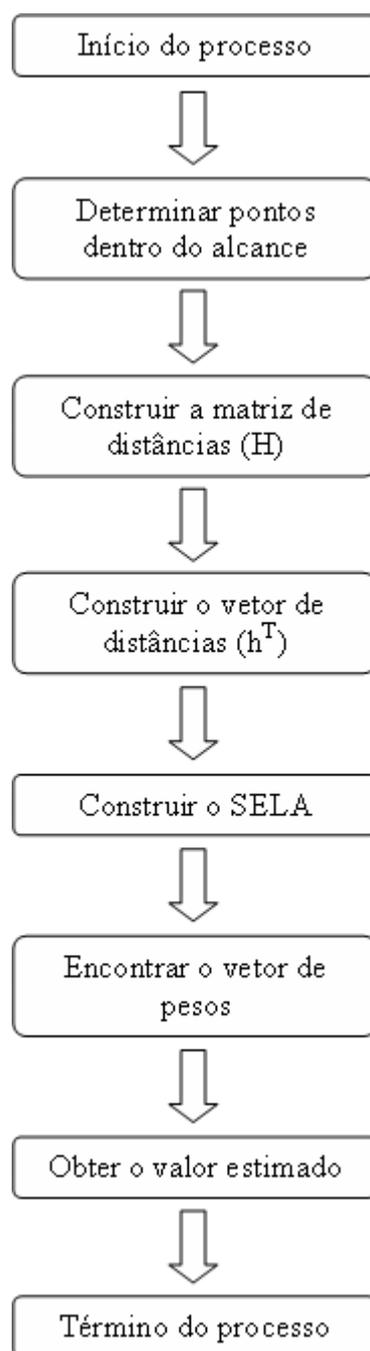


Figura 18 Estrutura do processo da krigagem.

3.5 Módulo 5 (Indicadores para a escolha do melhor modelo)

3.5.1 Validação Cruzada

O processo de validação cruzada é um procedimento para verificar a precisão do processo de estimação adotado. Para tanto, é realizada uma comparação entre os valores estimados e amostrados (VIEIRA, 1998).

Este processo consiste em realizar estimações exatamente sobre os pontos amostrados, que são os pontos onde os valores da variável em estudo são conhecidos. Para realizar a estimação em cada ponto deve-se, inicialmente, excluir o ponto (posição) onde o valor amostrado é procurado e em seguida utilizar o método de estimação para encontrar o valor estimado no ponto em questão, através da Equação 17. Neste trabalho, o método de estimação utilizado foi a krigagem ordinária.

O processo de validação consiste em uma seqüência de passos que devem ser seguidos para verificar a precisão do método de estimação. Esta seqüência é apresentada na Figura 19.

Em um primeiro momento, é realizada a estimação para cada um dos pontos amostrados na área, como descrito anteriormente. O passo seguinte é o cálculo da média simples dos valores amostrados e estimados.

Então, na seqüência, são calculados os desvios padrão dos valores estimados e amostrados. Possuindo estes dados, pode-se encontrar a covariância e o coeficiente de correlação entre os dois conjuntos de dados. Munidos deste conjunto estatístico descritivo dos dados, pode-se verificar a eficiência do método de estimação adotado. Quanto mais próximos forem os conjuntos de dados amostrados e estimados, mais preciso é considerado o interpolador.

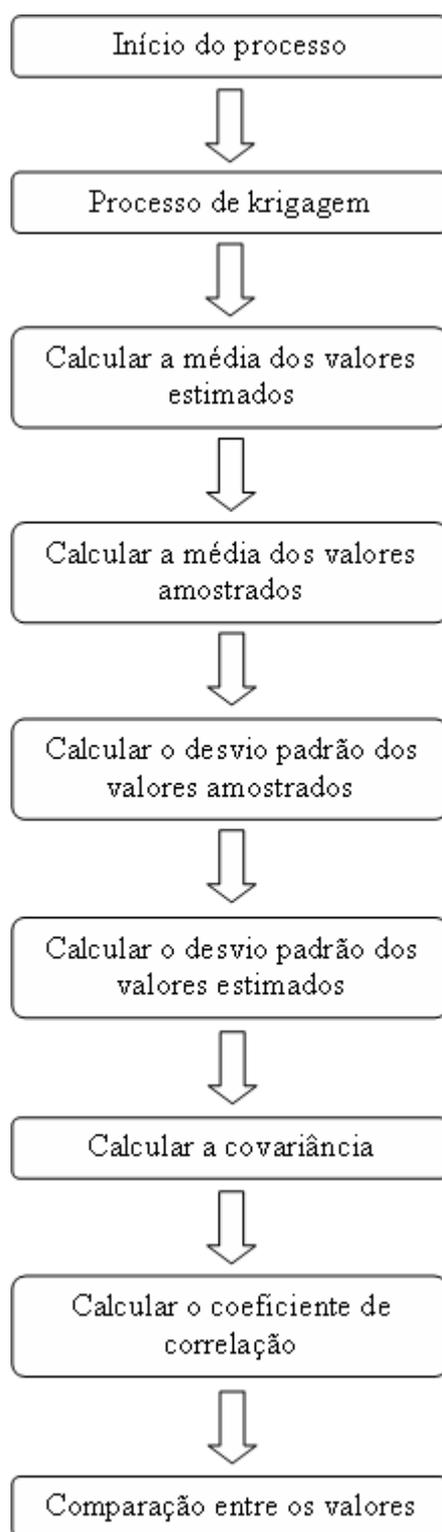


Figura 19 Fluxograma da Validação Cruzada.

3.5.2 Critério de Akaike

Assim como a validação cruzada, o critério de Akaike (AKAIKE, 1973) é utilizado para verificar a eficiência da técnica estimativa utilizada. O presente método é definido pela Equação 18, apresentada na Seção 2.2.5.2.

A obtenção do critério de Akaike é realizada através de três passos. O primeiro consiste em calcular do valor da soma dos quadrados dos resíduos. Para obter este valor, são somadas as diferenças entre os valores amostrados e estimados elevados à segunda potência, para cada ponto da região de interesse.

O segundo passo consiste na determinação do número de parâmetros utilizados pelo processo estimativo. Neste trabalho, os parâmetros aplicados foram três: efeito pepita, alcance e contribuição.

A última etapa é obtida substituindo-se os valores encontrados anteriormente na Equação 17. Onde R corresponde à soma dos quadrados dos resíduos, p define o número de parâmetros aplicados e n o número de pontos em estudo.

Para análise da precisão do modelo utilizado, o Critério de Akaike tem a seguinte definição: quanto menor for o valor encontrado, melhor é considerado o método de estimação envolvido no processo.

3.5.3 Critério de Filliben

Este método, bem como a validação cruzada, é uma forma de verificar a qualidade do processo estimativo, ou seja, um indicador de melhor ajuste.

O Critério de Filliben é similar ao processo de regressão linear, onde são calculados o coeficiente linear e angular e também o erro reduzido, utilizando para tanto a Equação 19 definida na Seção 2.2.4.3.

Dentro do processo, são calculados, novamente, as médias dos valores amostrados e estimados. Então, calculam-se as diferenças entre os valores amostrados e sua média correspondente e, também, a diferença entre os valores estimados e sua média. Estas diferenças são armazenadas em vetores, pois são utilizadas, posteriormente, no processo (FILLIBEN, 1975).

Para atingir o resultado do critério em questão foram necessárias três variáveis temporárias, com o intuito de tornar o processo mais visível ao usuário. A primeira recebe o resultado da multiplicação dos valores dos dois vetores, ou seja, a multiplicação da diferença armazenada nos vetores (estimados e amostrados). A segunda recebe o valor armazenado no vetor de diferenças dos pontos amostrais multiplicados por eles próprios e a terceira variável recebe os valores armazenados no vetor das diferenças dos pontos estimados. O critério de Filliben é obtido, então, pela divisão da primeira variável pela soma da segunda e terceira variáveis.

3.6 Módulo 6 (Previsão de Pontos não amostrados)

A previsão de pontos não amostrados permite, aos pesquisadores e usuários da informação verificada na área em estudo, ter uma estimativa de valores em posições não amostradas, o que permite um estudo mais eficiente do solo bem como um custo menor, visto que o número de pontos amostrados pode ser menor.

Ao início do processo, o usuário escolheu o conjunto de dados com os quais vai trabalhar. Então, baseado nestes, é construído um mapa bidimensional, contendo todos os pontos amostrais. Em seguida, na estimativa, cabe ao usuário escolher o modelo de ajuste de semivariograma que deseja, aquele que determinará os valores para o Alcance (α), Efeito Pepita (C_0) e Contribuição (C_1), pois estes já foram escolhidos, anteriormente, no momento da escolha do melhor ajuste ao modelo matemático). Depois de determinadas as variáveis utilizadas (parâmetros), o usuário informa ao software uma

posição, contendo coordenada X e Y dentro da área em estudo, que será estimada.

Em seguida, realizada a estimação na posição informada, o software apresenta ao usuário o valor estimado obtido para a localização e permite a visualização do novo ponto, dentro do mapa da área toda. A Figura 20 apresenta um exemplo de mapa de pontos, onde no detalhe em vermelho, pode ser visto um ponto estimado pelo software escolhido aleatoriamente.

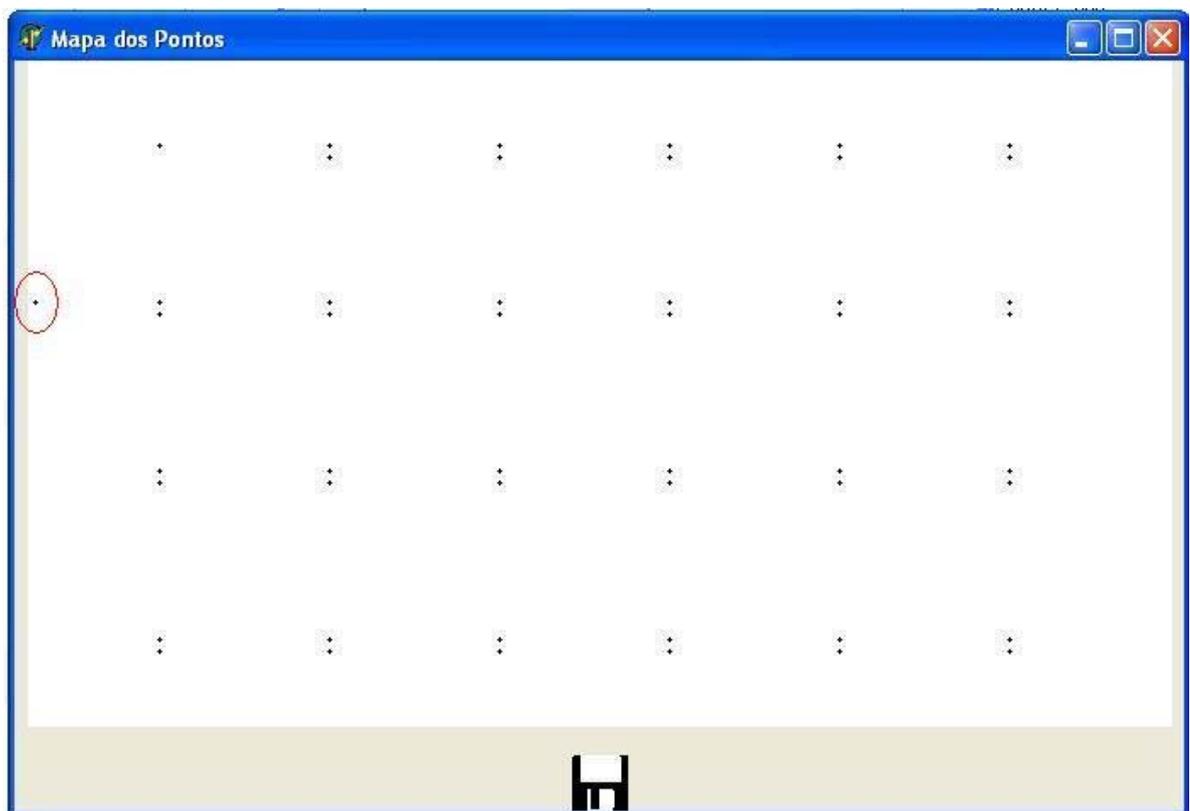


Figura 20 Mapa de pontos amostrados e estimados (detalhe).

4 RESULTADOS E DISCUSSÕES

Para ilustrar os resultados obtidos por meio da implementação do software, foram utilizados conjuntos de dados reais, onde foi amostrada a densidade do solo com profundidade entre 20 e 30 cm, por pesquisadores do Laboratório de Estatística Aplicada (LEA) da UNIOESTE, Campus de Cascavel. A aplicação deste exemplo visa uma melhor interpretação do resultado com o software. Os dados utilizados são apresentados no ANEXO I.

A Figura 21 exibe a tela principal do GEOCAC, após as atualizações realizadas neste projeto. As inclusões realizadas estão realçadas por marcadores em vermelho e serão explicadas a seguir:

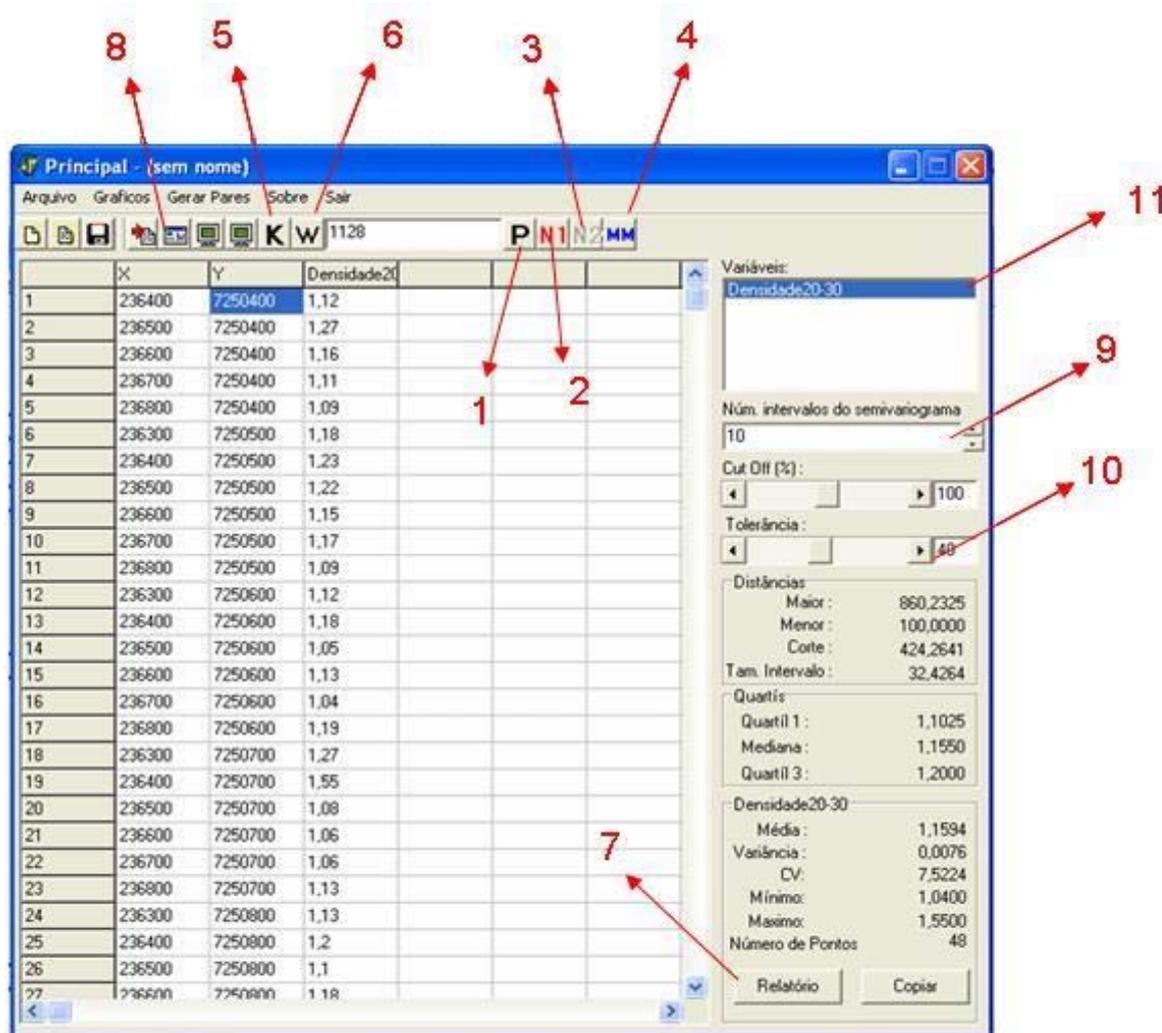


Figura 21 Tela principal do GEOCAC com novas inclusões.

1) Estimador de Pairwise

Acionando o botão apontado pelo marcador número 1 (representado pela letra P), é realizado o processo de cálculo do estimador de semivariância de Pairwise. A Figura 23 apresenta o gráfico do estimador de Pairwise implementado.

2) Estimador de New 1

O estimador de semivariância de New1 é acionado através do clique no objeto apontado pelo marcador 2. Este botão é representado pelo símbolo N1 (referente a New 1). O estimador de New 1 obtido pode ser visto na Figura 24.

3) Estimador de New 2

O apontador número 3 indica o botão que aciona o processo de construção da semivariância do estimador New 2. Este botão é representado na Figura pelo símbolo N2 (referente a New 2). A Figura 26 apresenta o estimador de New 2 implementado.

4) Estimadores de Matheron e Cressie & Hawkins

Através do botão número 8, pode-se realizar a construção e exibição dos semivariogramas de Matheron e Cressie & Hawkins, que foram implementados na versão 1.1 do GEOCAC. Tais semivariogramas são vistos nas Figuras 29 e 30, respectivamente.

5) Modelo Matemático (MM)

Este botão, representado pelo apontador número 4, abre uma janela que permite ao usuário escolher qual modelo matemático (esférico, exponencial, circular e gaussiano) deve ser escolhido para a construção do semivariograma por meio do estimador New 2. O estimador de New 2 fica bloqueado até que este botão seja clicado e um modelo matemático seja escolhido. Este botão é representado pelo símbolo MM (modelo matemático). Os estimadores de Pairwise e New 1 não dependem desta escolha do modelo matemático.

6) Cut-off

A barra apontada pela seta número 9 permite ao usuário a possibilidade de corte ou delimitação da distância para a construção do estimador.

7) Tolerância

A seta 10 aponta para a barra que permite ao usuário determinar um valor de tolerância para a angulação, no momento em que são construídos os semivariogramas direcionais.

8) Variável

Ao clicar no objeto apontado pela seta número 11, o usuário pode escolher qual variável, dentre as amostradas, ele deseja estudar.

9) Krigagem (K)

Acionando o botão indicado pelo marcador 5, o usuário abre a janela responsável pela estimação por krigagem ordinária.

10) Arquivo texto (W)

O botão apontado pelo marcador número 6 permite que arquivos com dados em formato texto possam ser carregados e analisados pelo software.

11) Relatório

O botão número 7 gera um relatório que contém a estatística descritiva da variável escolhida para análise. Este relatório permite impressão dos dados.



Figura 22 Relatório sobre a análise descritiva obtida.

4.1 Estimador de Pairwise

Clicando no botão 1, presente na Figura 21, o usuário acessa o semivariograma experimental de Pairwise.

Neste ambiente, é realizada a construção do semivariograma através do estimador de Pairwise, baseado nas posições carregadas na tela inicial e na variável escolhida.

Uma vez acessada esta função do GEOCAC, é permitido, ao usuário, a escolha do modelo matemático para a construção da curva, a fim de ajustar os dados amostrados. Podem ser escolhidos, também, os parâmetros Alcance, Efeito Pepita e Contribuição que influenciam na curva construída. Estes valores podem ser alterados até o momento em que seja alcançado um ajuste considerado bom para os dados.

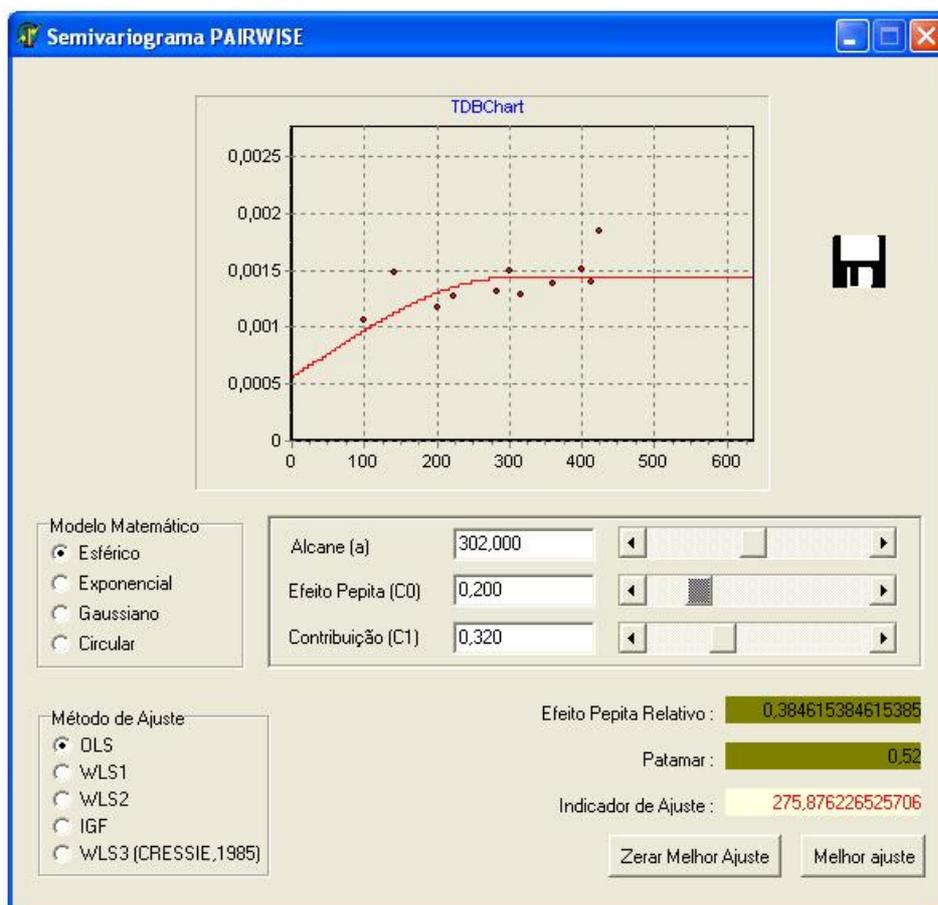


Figura 23 Tela Semivariograma por Pairwise.

Os valores de Alcance (a), Efeito Pepita (C_0) e Contribuição (C_1) são utilizados no método de estimação de Krigagem, portanto é interessante que sejam escolhidos ou determinados com cuidado.

A Figura 21 apresenta, ainda, um botão que permite que o gráfico construído seja salvo em arquivo para posterior utilização. Este botão está representado pelo esboço de um disquete dentro da Figura.

4.2 Estimador New 1

Clicando no botão 2 da Figura 21, o usuário estará construindo o semivariograma experimental, por meio do método de NEW 1, proposto por LI & LAKE (1994). A Figura 24 apresenta o gráfico obtido por esta opção.

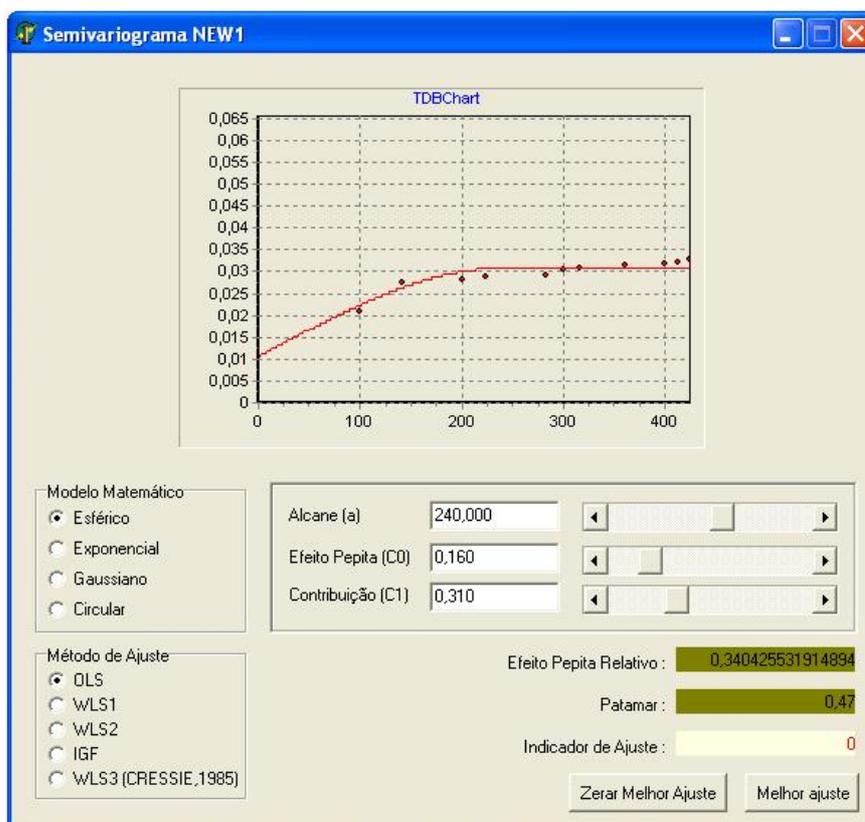


Figura 24 Tela Semivariograma por New 1.

As funcionalidades oferecidas ao usuário, nesta opção do software, são bastante similares àquela presente no método de Pairwise. A diferença entre as duas opções está, justamente, no momento da construção dos gráficos.

Esta diferença na construção do gráfico pode ser vista, comparando-se os valores plotados nos dois gráficos, das Figuras 23 e 24, respectivamente, onde se verifica grande discrepância na distribuição entre os dois modelos.

4.3 Estimador New 2

Quando o usuário quiser construir o semivariograma, por meio do estimador de New 2, ele deverá, inicialmente, acionar o botão 4 da tela principal (Figura 21), eleger o modelo matemático que deseja e escolher os parâmetros de Alcance (α), Efeito Pepita (C_0) e Contribuição (C_1) para a construção do semivariograma. Estas opções podem ser feitas através do menu apresentado na Figura 23.

Como podem ser observadas na Figura 25, as opções de modelos matemáticos oferecidos ao usuário para a construção do semivariograma são os modelos esférico, gaussiano, exponencial e circular.

Depois de ter escolhido o modelo matemático, devem ser determinados os valores dos parâmetros utilizados no método.

Assim que o modelo for escolhido e os parâmetros determinados, o usuário clica em “OK” para prosseguir com o método de New 2, pois o botão 3 presente na Figura 21, estará habilitado, uma vez que os dados necessários foram escolhidos. Então, o usuário pode clicar no botão referente ao New 2 e obter o semivariograma de New 2 apresentado pela Figura 26.

The 'New2' dialog box contains the following elements:

- Radio buttons for model selection:
 - Modelo Esférico
 - Modelo Exponencial
 - Modelo Gaussiano
 - Modelo Circular
- Input fields for parameters:
 - Alcance: 400
 - Efeito Pepita (C0): 0,001
 - Contribuição (C0): 0,33
- An 'OK' button at the bottom.

Figura 25 Tela para escolha dos parâmetros para a construção do semivariograma de New 2.

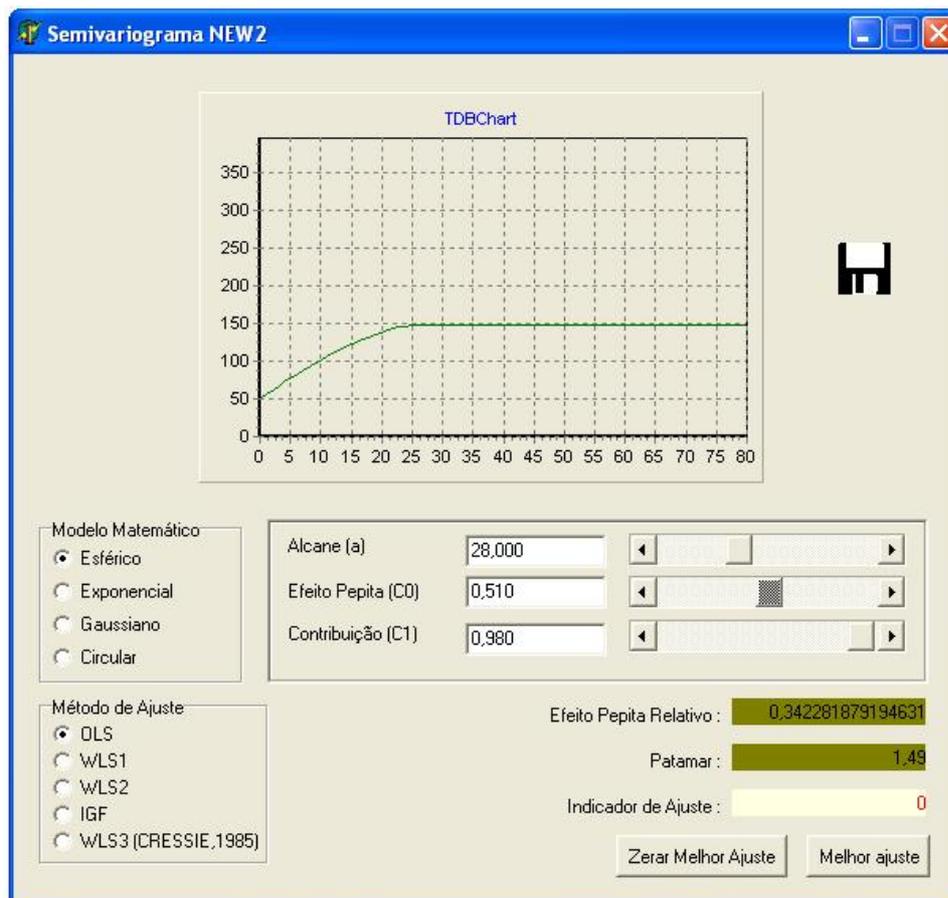


Figura 26 Tela Semivariograma por New 2.

Depois de construído o semivariograma através de New 2, as opções de interação são semelhantes às apresentadas aos métodos de Pairwise e New 1. O usuário pode escolher o modelo matemático para a construção da curva de ajuste, pode eleger os parâmetros para verificar o melhor ajuste e por fim, tem a opção de salvar o gráfico do semivariograma em arquivo.

4.4 Krigagem, Validação Cruzada, Teste de Filliben e Critério de Akaike

Ao clicar no botão 5 da tela principal (presente na Figura 20), o usuário aciona o módulo do software responsável pela estimação por krigagem. Porém, antes de acessar esta função, deve-se construir o semivariograma dos dados, utilizando um dos estimadores disponíveis com o intuito de determinar os valores dos parâmetros Alcance (α), Efeito Pepita (C_0) e Contribuição (C_1), que serão utilizados aqui.

Este módulo do GEOCAC permite que se estimem valores para a variável em estudo em qualquer lugar dentro da área amostrada. A Figura 27 apresenta o módulo de krigagem desenvolvido no software.

Como pode ser visto na Figura 27, o módulo de krigagem possibilita ao usuário escolher uma posição qualquer, através de uma coordenada (X,Y), dentro da área para estimação de um valor.

Após informada a posição a ser estimada, deve ser escolhido o método de construção de semivariograma. A escolha de cada método acarreta em valores distintos, para as variáveis: Efeito Pepita, Alcance e Contribuição. O software permite, também, que o se possa determinar, manualmente, independente dos semivariogramas, os valores dos parâmetros.

Processo de Krigagem

Krigagem | Validação Cruzada | Akaike | Filiben

Realizar a krigagem de um ponto dentro do Grid

Coordenada X: 1 Coordenada Y: 2,5

Método Escolhido:

- Matheron
- Cressie & Hawkins
- Pairwise
- New1
- New2

Alcance: 35

Efeito Pepita (C0): 4

Contribuição (C1): 3

Krigar Ponto

Valor Estimado: _____

Dados Retirados da Amostragem

| | |
|-----------------------|---------|
| Número de Amostrados: | 48 |
| Menor Posição em X | 236300 |
| Menor Posição em Y | 7250400 |
| Maior Posição em X | 236900 |
| Maior Posição em Y | 7251100 |

Ver mapa Contorno

Figura 27 Módulo de Krigagem.

Ao clicar em “Krigar Ponto”, o usuário realiza o processo de krigagem, baseado nos valores fornecidos por ele. O resultado obtido, o valor estimado, é então fornecido pelo objeto logo abaixo do botão krigar ponto.

É possível verificar, ainda, os menores e maiores valores para os eixos X e Y (do plano cartesiano), bem como o número de pontos amostrais da área.

O botão “Ver mapa” permite visualizar o mapa de pontos amostrados e estimados ao longo do processo, possibilitando ainda que este seja salvo. Já, clicando no botão “Contorno”, o usuário pode carregar um arquivo que constrói o mapa de contorno da área em estudo.

Clicando-se nas abas ao lado da aba chamada “Krigagem” da tela, apresentada na figura 27, é possível obter acesso à novas funções: Validação Cruzada, Akaike e Filliben.

A Figura 28 apresenta a estrutura do módulo de validação cruzada. O botão “realizar processo” é responsável por executar o processo da validação cruzada. Os campos “valor estimado” e “valor amostrado” servem para realizar uma comparação visual entre os dois conjuntos de dados.

Os campos presentes na parte inferior da Figura 26 apresentam alguns dados estatísticos para análise dos dados, tanto amostrados quanto estimados.

Porém existem formas mais confiáveis de verificação da qualidade do processo de estimação. Para tanto, o software oferece os critérios de Akaike e Filliben para realizar este processo de validação do estimador.

As Figuras 29 e 30 representam, respectivamente, os módulos de Akaike e Filliben.

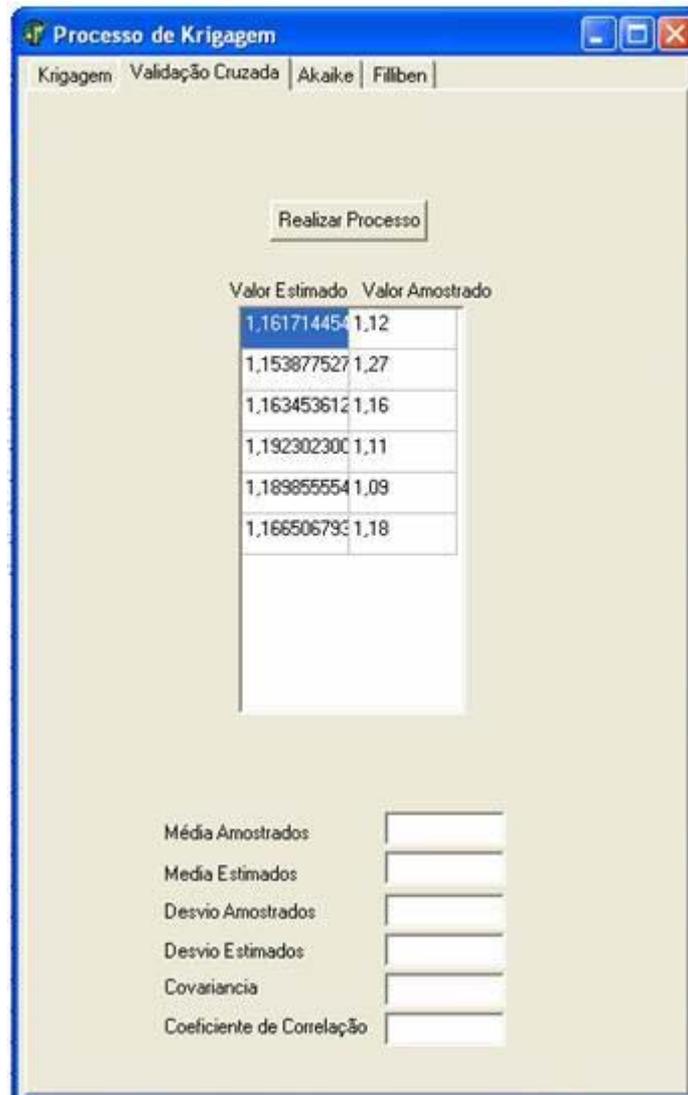


Figura 28 Módulo de Validação Cruzada.

Os módulos de Filliben e Akaike são bastante simples. São realizados junto com o processo da validação cruzada e, simplesmente, oferecem, ao usuário, valores que servem de apoio para verificação da qualidade do método de estimação.

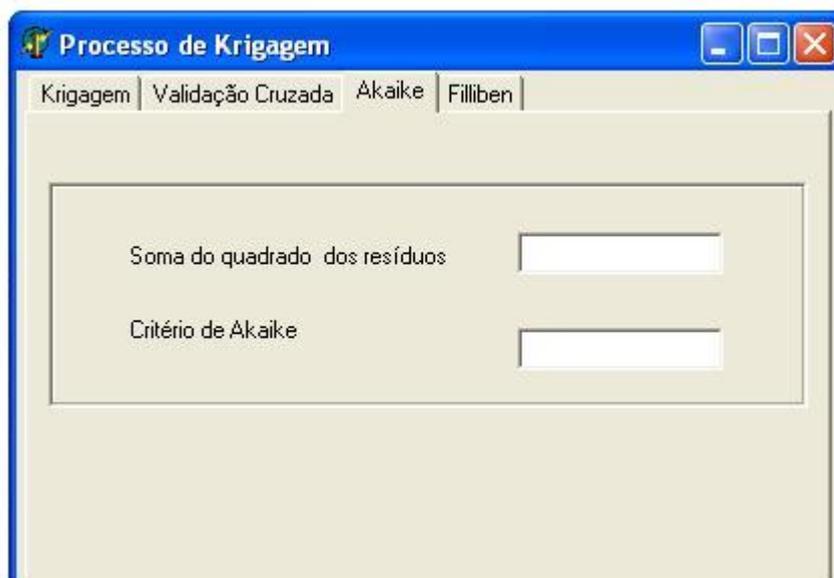


Figura 29 Critério de Akaike.

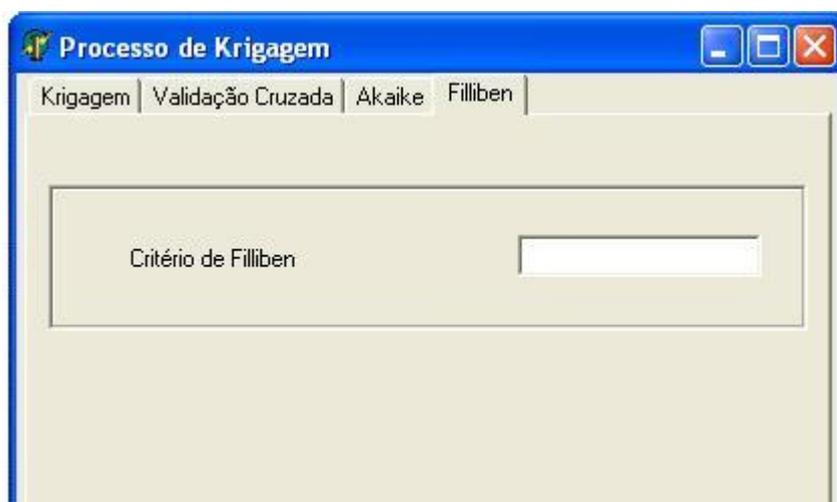


Figura 30 Teste de Filliben.

Ao clicar no botão 7 da tela principal (Figura 21), o usuário obterá um rápido relatório, contendo a estatística descritiva da variável em estudo. Este

relatório pode ser observado na Figura 31. O botão “imprime” do relatório permite a impressão dos dados.

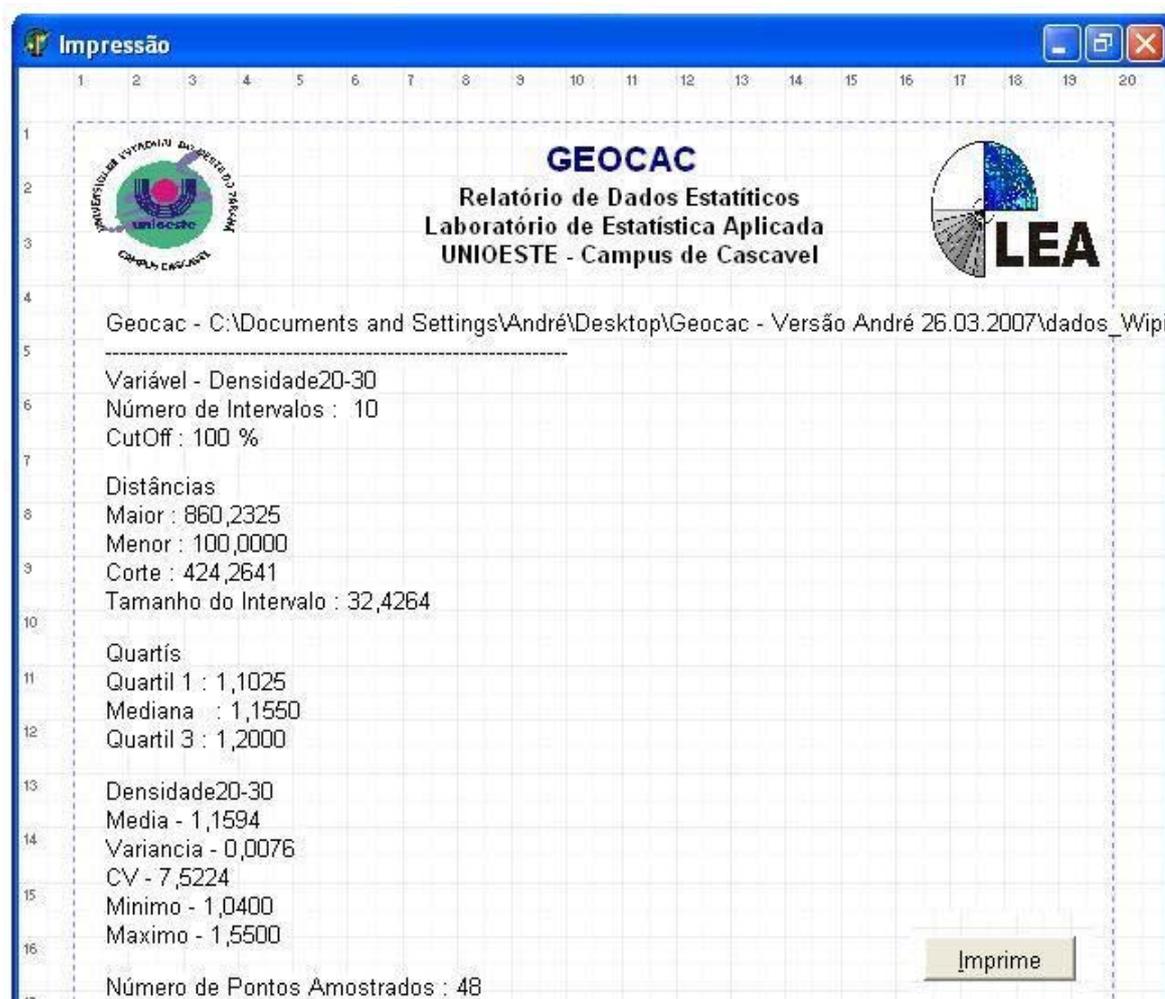


Figura 31 Relatório sobre a variável em estudo.

4.5 Estimadores de Matheron e Cressie & Hawkins

Clicando no botão 8 da Figura 20, o usuário terá acesso às telas dos semivariogramas de Matheron e Cressie. Nestas telas, existem as opções de interação, fornecendo os valores dos parâmetros (Alcance, Efeito Pepita e

Contribuição), a escolha do modelo matemático para verificação de “casamento” e o modelo matemático para verificação da qualidade do “casamento” entre a curva do modelo matemático e a curva construída pelo usuário. As Figuras 32 e 33 apresentam as telas do Semivariograma de Matheron e Cressie & Hawkins respectivamente.

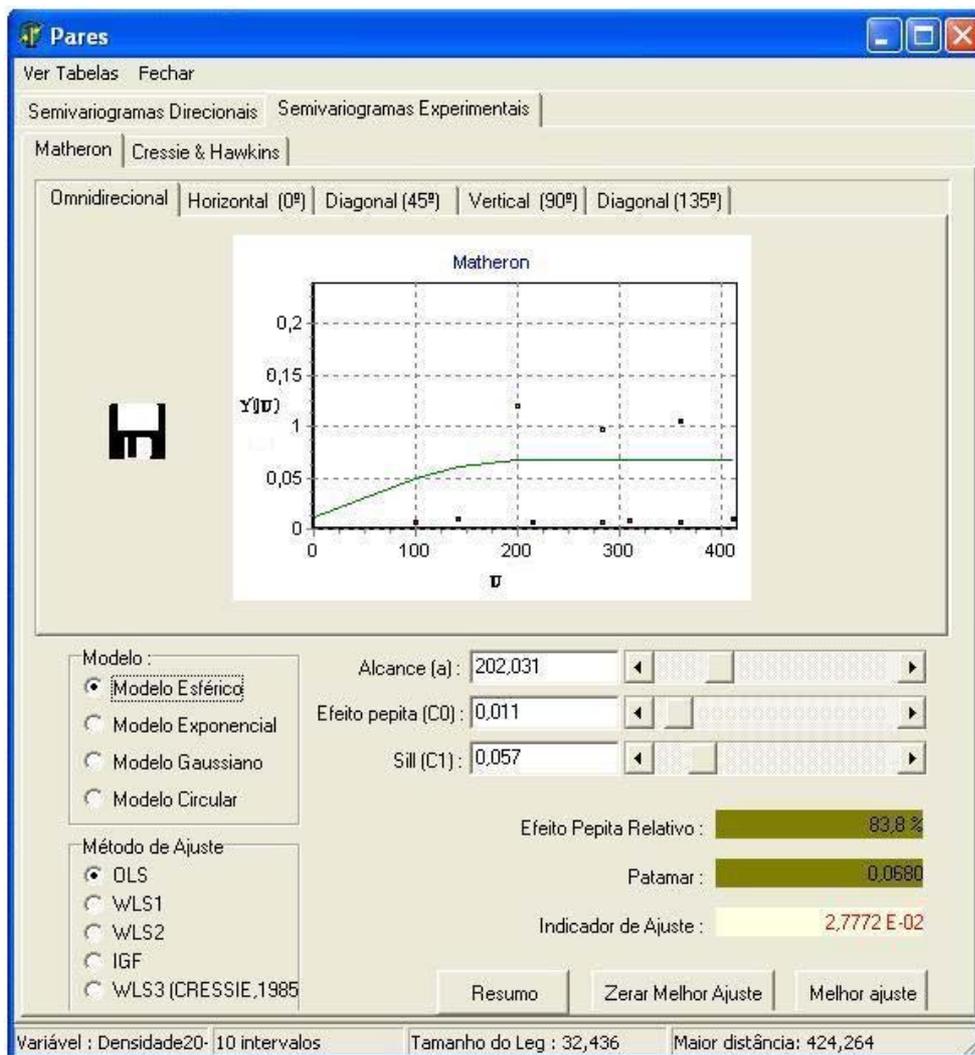


Figura 32 Semivariograma de Matheron.

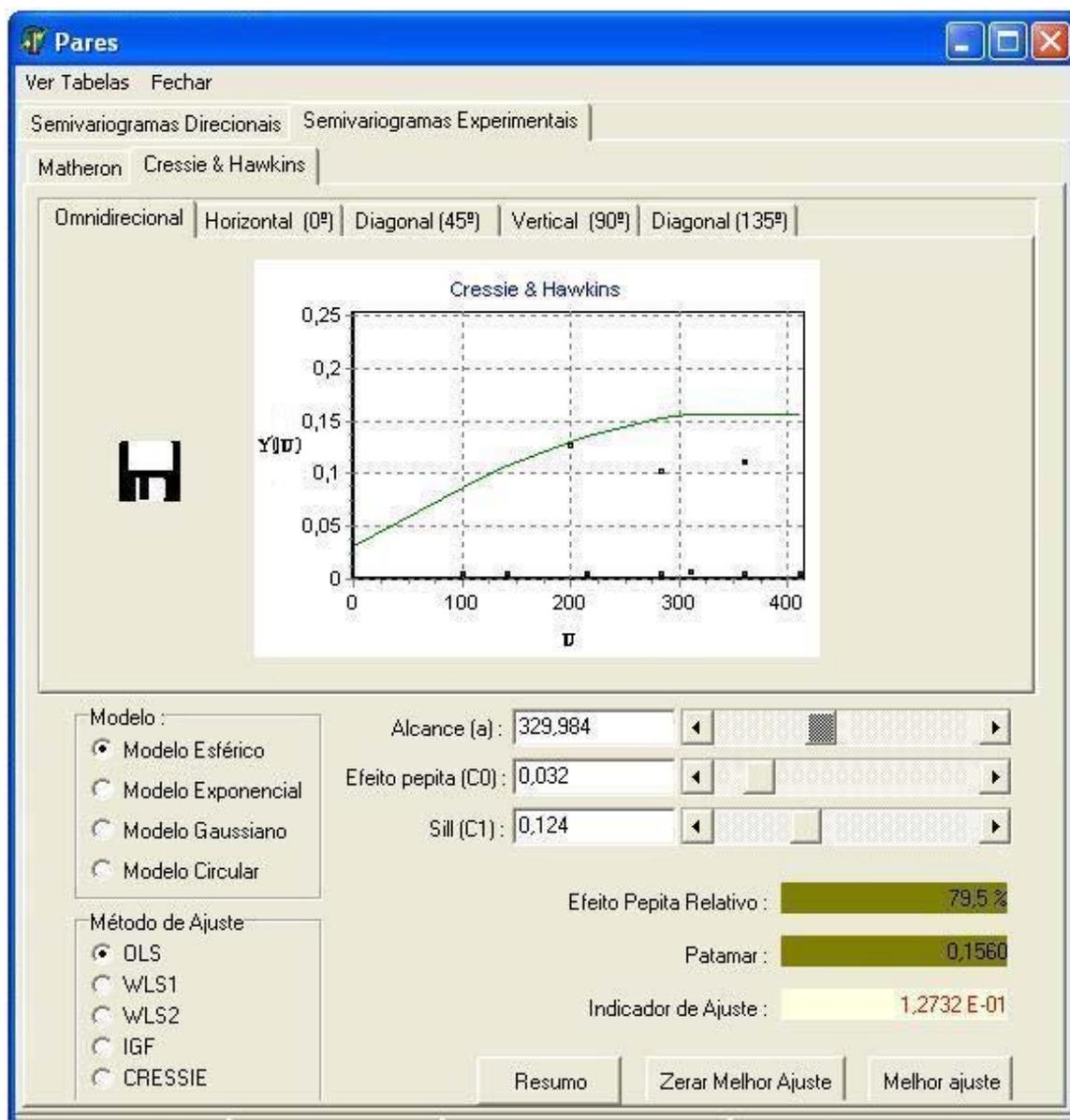


Figura 33 Semivariograma de Cressie & Hawkins.

Utilizando os modelos construídos, foram criadas as Tabelas 1 e 2 com os valores encontrados para os parâmetros envolvidos na construção dos semivariogramas e estimação dos pontos.

Tabela 1 Parâmetros encontrados para a construção do semivariograma que apresentaram casamento aproximado entre o semivariograma experimental e a curva do modelo esférico.

| Método | Estimativas | | | |
|-------------------|--------------------|----------------------|----------------------|------------------------------------|
| | a | C₀ | C₁ | C₀+C₁ |
| Matheron | 356,92 | 0,023 | 0,019 | 0,042 |
| Cressie & Hawkins | 356,92 | 0,022 | 0,020 | 0,042 |
| Pairwise | 342,00 | 0,029 | 0,021 | 0,050 |
| NEW 1 | 413,00 | 0,026 | 0,023 | 0,049 |
| NEW 2 | 398,00 | 0,026 | 0,024 | 0,050 |

Tabela 2 Valores encontrados para os métodos de avaliação de ajuste ao modelo.

| Dados | Avaliação de Ajuste | | |
|--------------------|----------------------------|---------------|-----------------|
| | Validação Cruzada | Akaike | Filliben |
| Densidade 20-30 cm | 0,6665362 | -13,4717 | 0,145095 |

5 CONCLUSÕES

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. Baseado neste pensamento, o GEOCAC foi incrementado com novos módulos desenvolvidos durante este trabalho, com o intuito de torná-lo não só mais aplicável como também mais eficiente.

A Versão 2.1 do software GEOCAC se mostrou (pois permite ao usuário a análise de um conjunto de dados consideravelmente maior) abrangente (pois possui novos métodos de análise que possibilitam ao usuário não só analisar os dados, mas também verificar se o mecanismo de análise mostrou-se eficiente) e mais ergonômica, pois foram realizadas também alterações com o intuito de melhorar a interação homem-software, perante a versão anterior (1.0).

Por meio deste software, a Geoestatística e outras áreas que dela fazem uso, como a Agricultura de Precisão, se beneficiam, pois contam com uma nova opção de análise, gratuita e eficaz.

Para novos trabalhos, seria interessante a implementação de novos métodos de estimação, como o método da máxima verossimilhança, máxima verossimilhança restrita e também funções para o estudo de diagnósticos, fazendo, dessa forma, que o GEOCAC ofereça uma gama de aplicações ainda maior e permita aos pesquisadores, que dele se utilizam, o acesso a uma ferramenta robusta, confiável e ao alcance no momento de analisar dados.

Outro objetivo que poderia ser alcançado, em longo prazo, seria a implementação do software em uma linguagem multiplataforma, tal como Java.

6 BIBLIOGRAFIA

AKAIKE, H. Information theory and an extension of the maximum likelihood principle. In **2nd International Symposium on Information Theory** (eds B.N. Petrov and F. Csáki), p. 267–281. Akadémiai Kiadó, Budapest, 1973.

ANSELIN, L. **SpaceStat tutorial**: a workbook for using SpaceStat in the analysis of spatial data. University of Illinois, Urbana-Champaign, 1992.

AVENDAÑO, F.; SCHANBENRGER O.; PIERCE, F. J.; MALAKERBEHAN H. Geostatistical Analysis of Field Spatial Distribution Patterns of Soybean Cyst Nematode. **Agronomy Journal**, Michigan, v.95, Disponível on-line em <http://www.scielo.br>. Acesso em 16/Dez/2005.

BALASTREIRE, L. A. Estudo de caso: uma pesquisa brasileira em agricultura de precisão. In: **Congresso Brasileiro de Engenharia Agrícola**, v.27, 1998, Poços de Caldas. Anais. Lavras: Universidade Federal de Lavras, p. 203-231, 1998.

BRUN, A. L.; SILVA, E. A. A; VILAS BOAS; OPAZO, M. A. U. Algoritmo de Validação Cruzado Usando Krigagem Ordinária. **Congresso Brasileiro de Agricultura de Precisão**. Piracicaba, 2004, 6p.

CÂMARA, G.; MEDEIROS, J. S. GIS para meio ambiente. In: **SISTEMA DE INFORMAÇÕES GEOGRÁFICAS – GIS 98**, Curitiba, 1998, Curso D, Curitiba: Sagres, 1998, 185p.

CARVALHO, J. R. P.; ASSAD, E. D. Análise espacial da precipitação pluviométrica no Estado de São Paulo: comparação de métodos de interpolação. **Engenharia Agrícola**, Piracicaba, v.25, n.2, p.377-384. Disponível on-line em <http://www.scielo.br>. Acesso em 29/Nov/2005.

CARVALHO, L. P. **Estudo da Anisotropia em uma Área Agrícola Espacialmente Referenciada**. Cascavel, 2004. 80f. Dissertação (Mestrado em

Engenharia Agrícola). Centro de Ciências Exatas e Tecnológicas. Universidade Estadual do Oeste do Paraná.

CLARCK, I. **Practical Geoestatistics**. London, Applied Science Publishers LTD, 1979, 129p.

CRESSIE, N. R.; HAWKINS, D. M. Robust Estimation of the variogram. In: **Mathematical Geology**, New York, v. 12, n. 2, p. 115-125, 1980.

CRESSIE, N. **Statistics for spacial data**. New York> John Wiley, 1993. 900p.

DAINESE, R. C.; PUGLIESE, P. B.; JORGE, L. A. C.; TRINDADE JR, O.; CEREDA JR, A. Avaliação de Incidência de Plantas Invasoras por meio de Segmentação de Imagens e Redes Neurais. **Congresso Brasileiro de Agricultura de Precisão**. Piracicaba, 2004.

DAVIS, B. M. Uses and abuses of cross-validation in geostatistics. **Mathematical Geology**, New York, v. 19, n. 3, p. 405-421, 1987.

FILLIBEN, J.J. The Probability plot correlation coefficient test of normality. **Technometrics**, v. 17, n. 1, p. 111-117, 1975.

GONÇALVES, A.C.A. **Variabilidade espacial de propriedades físicas do solo para fins de manejo da irrigação**. Piracicaba, 1997. 118f. Tese (Doutorado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.

GUERRA, P. A. G. **Geoestatística Operacional**. Brasília: Ministério das Minas e Energia, 1988, 150p.

HOEL, P.G.; PORT, S.C.; STONE, C.J. **Introduction to Stochastic Processes**. Ed. Houghton Mifflin Company Boston. 1972, 203p.

ISAAKS, E. H.; SRIVASTAVA, R. M. **Na introduction to Applied Geoestatics**. New York: Oxford, Oxford University Press, 1989, 561p.

JOURNEL, A. G.; HUIJBREGTS, CH. J. Mining **Geoestatics**. London, Academic Press, 1978, 600p.

KAVANAGH, C. R. **Estimação de Parâmetros no ajuste de Modelos Teóricos a Semivariogramas Experimentais**. Cascavel, 2001. 85f. Dissertação (Mestrado em Engenharia Agrícola). Centro de Ciências Exatas e Tecnológicas. Universidade Estadual do Oeste do Paraná.

KRIGE, D. G. A statistical approach to some basic mine evaluation problems on the witwatersrand. **Journal of South African Institution of Mining and Metallurgy**, Johannesburg, v. 52, p. 119-139, 1951.

LI, D.; LAKE, L. W. A moving window semivariance estimator. **Water Resources Research**. v. 30, n.5, p. 1479-1489. 1994.

MATHERON, G. Principles of Geoestatics. **Economic Geology**, v. 58, p. 1246-1266, 1963.

MOLIN, J. P. Agricultura de Precisão – **O gerenciamento da variabilidade**. Piracicaba 2001. 83p.

NICOLELLA, G.; PEREZ FILHO, A.; SOUZA, M. D.; FERRACINI, V. L. Geoestatics as a Basic to the CMLS Pesticide Simulation Model with Validation in Soil Columns. **Scientia Agrícola**, Piracicaba, v.62, n.1, p. 50-56. Disponível on-line em <http://www.scielo.br>. Acesso em 18/Dez/2005.

NOVAIS, R. F.; ALVAREZ V.; SCHAEFER, C. E. G. R Editora. **Tópicos em Ciência do Solo**. Viçosa, SBCS, 2000, 352p. PRESSMAN, R. S. Engenharia de software. McGraW Hill, São Paulo, 2006, 6ª ed. 720p.

RENDU, J. M. **An Introduction to Geoestatistical methods of mineral evaluation**, South África, IMM, Monograph Geoestatistics, 1978, 100p.

RIBEIRO JÚNIOR, P. J. **Abordagem Geoestatística em um estudo de variabilidade espacial**. Piracicaba, 1995, 99f. Dissertação (Mestrado). Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo.

ROMAN, M. **Variabilidade espacial do desenvolvimento e da produtividade do trigo (*Triticum aestivum*)**. Cascavel, 2005, 128f. Dissertação (Mestrado em Engenharia Agrícola). Centro de Ciências Exatas e Tecnológicas. Universidade Estadual do Oeste do Paraná, UNIOESTE, Brasil.

SAKAMOTO, Y.; ISHIGURO, M.; KITAGAWA, G. **Akaike information criterion statistics**. Dordrecht: Kluwer Academic Publishers, 1986. 240p.

SILVA, E. A. A. ; OPAZO, M. A. U. ; VILAS BOAS, M. A. ; LAMPARELLI, R. . Programa computacional para produção e ajuste de modelos teóricos a semivariogramas experimentais em variáveis regionalizadas - Geocac. In: **XXXI Congresso Brasileiro de Engenharia Agrícola**, 2002, Salvador. XXXI Congresso Brasileiro de Engenharia Agrícola, 2002.

SOUZA, E. G.; JOHANN, J. A.; ROCHA, J. V.; RIBEIRO, S. R. A.; SILVA, M. S.; URIBE-OPAZO, M. A.; MOLIN, J. P.; OLIVEIRA, E. F.; NÓBREGA, L. H. P. Variabilidade espacial dos atributos químicos do solo em um latossolo roxo distrófico na região de Cascavel, PR. **Engenharia Agrícola**, Jaboticabal, v. 8, n. 3, p. 80-92, 1999.

SOUZA, Z. M.; MARTINS FILHO, M. V.; MARQUES JR, J.; PEREIRA, G. T. Variabilidade Espacial de Fatores de Erosão em Latossolo Vermelho Eutroférico sob Cultivo de Cana-de-Açúcar. **Engenharia Agrícola**, Campinas, v.25, n.1, p.105-114. Disponível on-line em <http://www.scielo.br>. Acesso em 18/Dez/2005.

SRIVASTAVA, R.M; PARKER, H.M. Robust measures of spatial continuity, in **Geoestatistics**, v. 1, p. 295-308, 1989.

STEINBRUCH, A.; WINTERLE, P. **Álgebra Linear**. Editora Makron Brooks, 2ª Edição, volume único, 1987, 594p.

VIEIRA, S.R. **Geoestatística**: Curso de extensão universitária. Botucatu: UNESP, 1996, 166p.

VIEIRA, S. R. **Geoestatística Aplicada à Agricultura de Precisão**. In: SISTEMA DE INFORMAÇÕES GEOGRÁFICAS – GIS 98. Curitiba, Curso P, Curitiba: Sagres, p. 1-48, 1998.

VIEIRA, S. R. et al. **Mapeamento de propriedades de solo e plantas usando geoestatística**. In: BALASTREIRE, L. A. O estado da arte da agricultura de precisão no Brasil. Piracicaba: L. A. Balastreire, p. 93-104, 2000.

WEBSTER, R.; BURROUGH, P.A. Multiple discriminant analysis in soil survey. **Journal of Soil Science**, v.25, p.120-134, 1974.

WEBSTER, R.; OLIVER, M.A. **Statistical methods in soil and land resource survey**. Oxford: Oxford University Press, 1990, 316p.

WOLLENHAUPT, N. C.; MULLA, D. J.; CRAWFORD, G. Soil sampling and interpolation techniques for mapping spatial variability of soil properties. In: PIERCE, F. J.; SAEDLER, E. J. **The State of site-specific management for agriculture**. MADison, ASA Misc. Publ., ASA, CSSA and SSA, p. 19-53, 1997.

ANEXO I

| PARCELA | X (leste) | Y (norte) | Densidade 20-30 (cm) |
|---------|-----------|-----------|----------------------|
| 1 | 236400 | 7250400 | 1,12 |
| 2 | 236500 | 7250400 | 1,27 |
| 3 | 236600 | 7250400 | 1,16 |
| 4 | 236700 | 7250400 | 1,11 |
| 5 | 236800 | 7250400 | 1,09 |
| 6 | 236300 | 7250500 | 1,18 |
| 7 | 236400 | 7250500 | 1,23 |
| 8 | 236500 | 7250500 | 1,22 |
| 9 | 236600 | 7250500 | 1,15 |
| 10 | 236700 | 7250500 | 1,17 |
| 11 | 236800 | 7250500 | 1,09 |
| 12 | 236300 | 7250600 | 1,12 |
| 13 | 236400 | 7250600 | 1,18 |
| 14 | 236500 | 7250600 | 1,05 |
| 15 | 236600 | 7250600 | 1,13 |
| 16 | 236700 | 7250600 | 1,04 |
| 17 | 236800 | 7250600 | 1,19 |
| 18 | 236300 | 7250700 | 1,27 |
| 19 | 236400 | 7250700 | 1,55 |
| 20 | 236500 | 7250700 | 1,08 |
| 21 | 236600 | 7250700 | 1,06 |
| 22 | 236700 | 7250700 | 1,06 |
| 23 | 236800 | 7250700 | 1,13 |
| 24 | 236300 | 7250800 | 1,13 |
| 25 | 236400 | 7250800 | 1,2 |
| 26 | 236500 | 7250800 | 1,1 |
| 27 | 236600 | 7250800 | 1,18 |
| 28 | 236700 | 7250800 | 1,16 |
| 29 | 236800 | 7250800 | 1,04 |
| 30 | 236300 | 7250900 | 1,14 |
| 31 | 236400 | 7250900 | 1,22 |
| 32 | 236500 | 7250900 | 1,11 |
| 33 | 236600 | 7250900 | 1,11 |
| 34 | 236700 | 7250900 | 1,14 |
| 35 | 236800 | 7250900 | 1,16 |
| 36 | 236300 | 7251000 | 1,21 |
| 37 | 236400 | 7251000 | 1,2 |
| 38 | 236500 | 7251000 | 1,17 |
| 39 | 236600 | 7251000 | 1,09 |
| 40 | 236700 | 7251000 | 1,1 |
| 41 | 236800 | 7251000 | 1,05 |
| 42 | 236300 | 7251100 | 1,25 |
| 43 | 236400 | 7251100 | 1,16 |
| 44 | 236500 | 7251100 | 1,25 |
| 45 | 236600 | 7251100 | 1,3 |
| 46 | 236700 | 7251100 | 1,25 |
| 47 | 236800 | 7251100 | 1,16 |

| | | | |
|----|--------|---------|------|
| 48 | 236900 | 7251100 | 1,12 |
|----|--------|---------|------|

Livros Grátis

(<http://www.livrosgratis.com.br>)

Milhares de Livros para Download:

[Baixar livros de Administração](#)

[Baixar livros de Agronomia](#)

[Baixar livros de Arquitetura](#)

[Baixar livros de Artes](#)

[Baixar livros de Astronomia](#)

[Baixar livros de Biologia Geral](#)

[Baixar livros de Ciência da Computação](#)

[Baixar livros de Ciência da Informação](#)

[Baixar livros de Ciência Política](#)

[Baixar livros de Ciências da Saúde](#)

[Baixar livros de Comunicação](#)

[Baixar livros do Conselho Nacional de Educação - CNE](#)

[Baixar livros de Defesa civil](#)

[Baixar livros de Direito](#)

[Baixar livros de Direitos humanos](#)

[Baixar livros de Economia](#)

[Baixar livros de Economia Doméstica](#)

[Baixar livros de Educação](#)

[Baixar livros de Educação - Trânsito](#)

[Baixar livros de Educação Física](#)

[Baixar livros de Engenharia Aeroespacial](#)

[Baixar livros de Farmácia](#)

[Baixar livros de Filosofia](#)

[Baixar livros de Física](#)

[Baixar livros de Geociências](#)

[Baixar livros de Geografia](#)

[Baixar livros de História](#)

[Baixar livros de Línguas](#)

[Baixar livros de Literatura](#)
[Baixar livros de Literatura de Cordel](#)
[Baixar livros de Literatura Infantil](#)
[Baixar livros de Matemática](#)
[Baixar livros de Medicina](#)
[Baixar livros de Medicina Veterinária](#)
[Baixar livros de Meio Ambiente](#)
[Baixar livros de Meteorologia](#)
[Baixar Monografias e TCC](#)
[Baixar livros Multidisciplinar](#)
[Baixar livros de Música](#)
[Baixar livros de Psicologia](#)
[Baixar livros de Química](#)
[Baixar livros de Saúde Coletiva](#)
[Baixar livros de Serviço Social](#)
[Baixar livros de Sociologia](#)
[Baixar livros de Teologia](#)
[Baixar livros de Trabalho](#)
[Baixar livros de Turismo](#)