



MINISTÉRIO DA CIÊNCIA E TECNOLOGIA
INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS

SER-301: ANÁLISE ESPACIAL DE DADOS GEOGRÁFICOS

Bárbara Maria Giacom Ribeiro

RELATÓRIO DE ATIVIDADES LABORATÓRIO Nº 2: GEOESTATÍSTICA NÃO LINEAR

INPE
São José dos Campos
2008

1 INTRODUÇÃO

O *Laboratório 2* teve como objetivo a prática e o entendimento de um dos procedimentos da geoestatística não-linear implementado no SPRING: *Krigeagem por Indicação*.

A *Krigeagem por Indicação* busca obter uma grade regular de valores a partir dos dados amostrados pontualmente.

O módulo de *Krigeagem por Indicação* implementado no SPRING baseia-se na sub-rotina "ik3d" da GSLIB (DEUTSCH e JOURNAL, 1992). Este módulo possibilita a espacialização, segundo uma grade regular, de atributos espaciais de natureza contínua e categórica. Juntamente com o mapa de atributos, é gerado um mapa de incerteza, também com uma representação de grade regular, de estimação.

O exemplo prático deste *Laboratório* refere-se à estimativa de cotas de altimetria, de uma região fictícia, cujas amostras foram obtidas ao longo de um rio. A proposta foi adensar o conjunto de dados e obter os valores de altimetria para o restante da área.

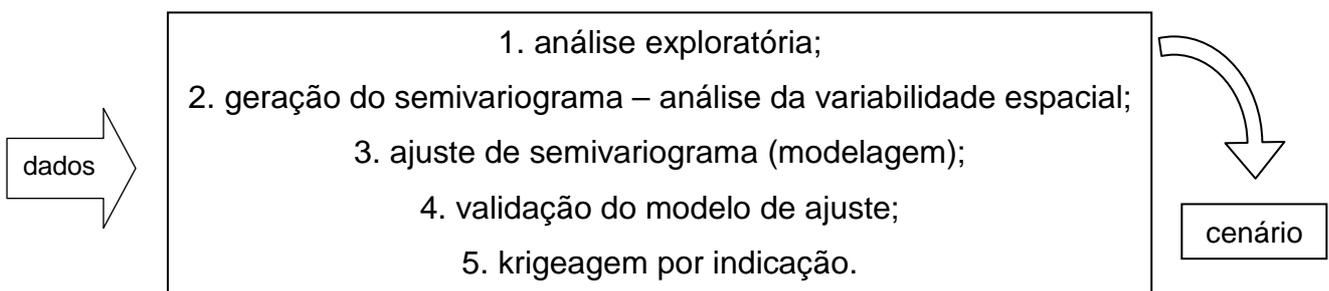
2 DADOS

Os dados disponíveis estavam contidos no banco de dados "Mancha_Teste", composto por um projeto chamado "Inundação", cujas amostras referem-se a valores de inundação gerados por um modelo hidrológico.

3 DESENVOLVIMENTO

Os passos desenvolvidos neste *Laboratório 2* resumem-se em: (1) análise exploratória dos dados, (2) análise estrutural (cálculo e modelagem do semivariograma) e (3) realização de inferências pelo procedimento de *Krigeagem por Indicação*.

A análise geoestatística realizada no *software* Spring segue-se as seguintes etapas:



O banco de dados utilizado foi o Mancha_Teste, e o projeto Inundação, com projeção *UTM/SAD69*, determinado pelas seguintes coordenadas: 49° 47' 38,98" O e 25° 51' 44,13" S; 48° 42' 50,67" O e 25° 16' 38,08" S.

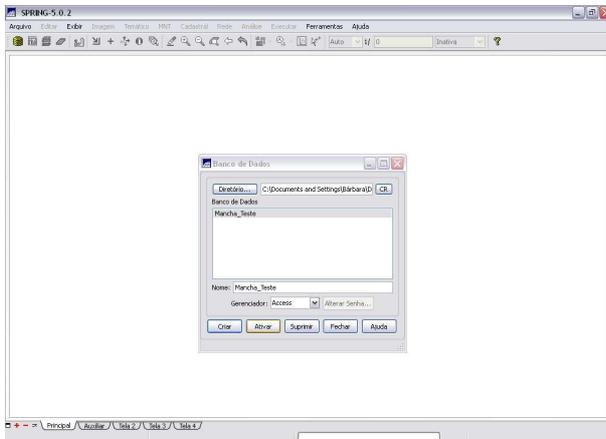


Figura 1 – ativação do banco de dados Mancha_Teste.

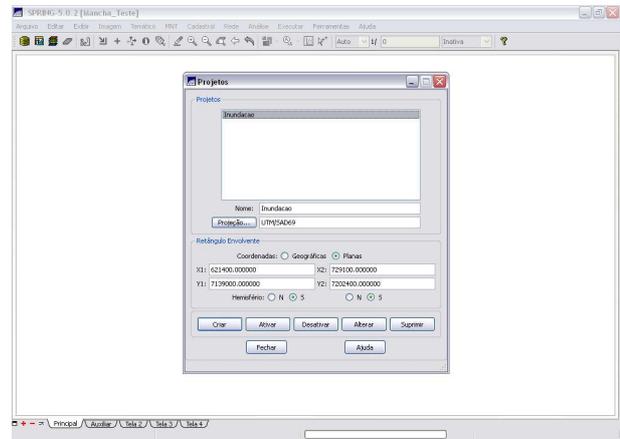


Figura 2 – ativação do projeto Inundação dentro do banco de dados Mancha_Teste.

Inicialmente ativou-se o Banco de Dados e o Projeto com os respectivos dados. Os dados pontuais – amostras, e a delimitação do limite da área puderam ser visualizados, e então se procedeu a fase de realização da análise exploratória.

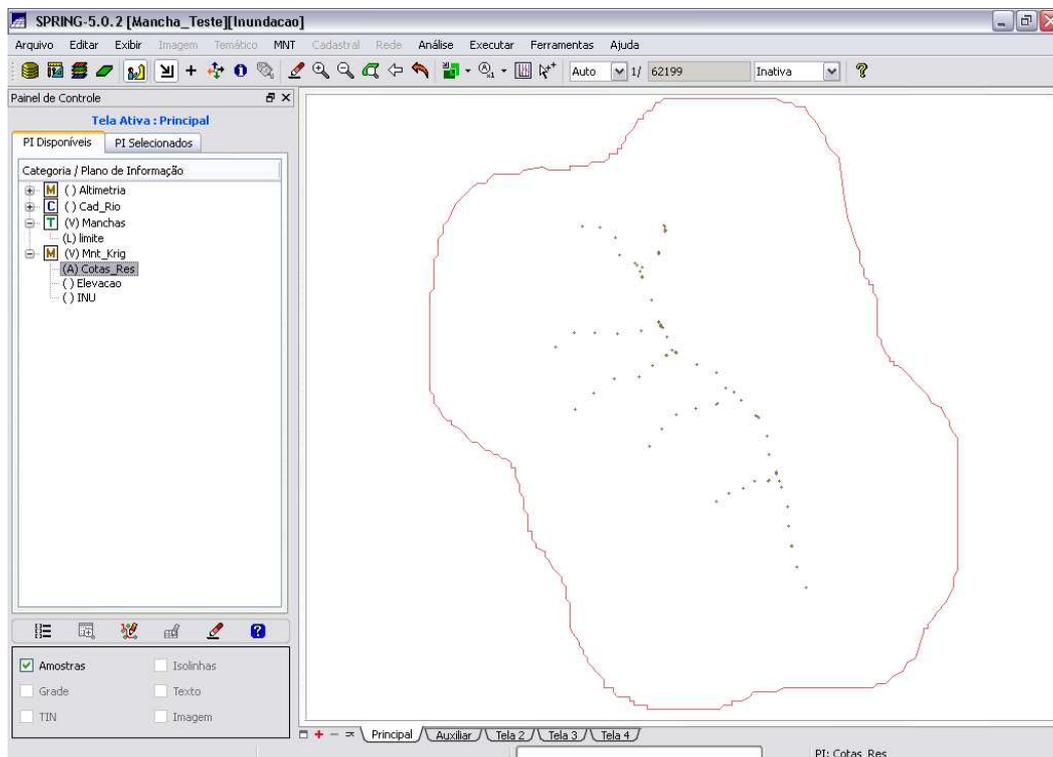


Figura 3 – visualização dos PIs: Limite (categoria: Mancha) e Cotas_Res (categoria: MNT_Krig).

3.1 Análise Exploratória em Geoestatística

No Spring a análise exploratória dos dados é realizada por meio de estatísticas univariadas e bivariadas.

As estatísticas univariadas fornecem um meio de organizar e sintetizar um conjunto de valores, e é realizada, principalmente, por meio do histograma. As características importantes do histograma são organizadas em três grupos:

1. *medidas de localização*: média, valor mínimo, quartil inferior, mediana, quartil superior e valor máximo;
2. *medidas de dispersão*: variância e desvio padrão;
3. *medidas de forma*: coeficiente de assimetria, coeficiente de curtose e coeficiente de variação.

As estatísticas bivariadas fornecem meios de descrever o relacionamento entre duas variáveis, isto é, entre dois conjuntos de dados ou de duas distribuições. Esta relação pode ser visualizada através do diagrama de dispersão e o grau da relação linear entre as variáveis pode ser medido através do coeficiente de correlação.

A primeira fase da análise geoestatística realizada neste *Laboratório* foi a *análise exploratória*, que fornece os padrões dos dados. Inicialmente, optou-se pela *estatística descritiva*, que fornece diversas informações sobre os dados utilizados, como número de amostras, média, desvio padrão, valor máximo e mínimo das amostras, entre outras (Figura 4).

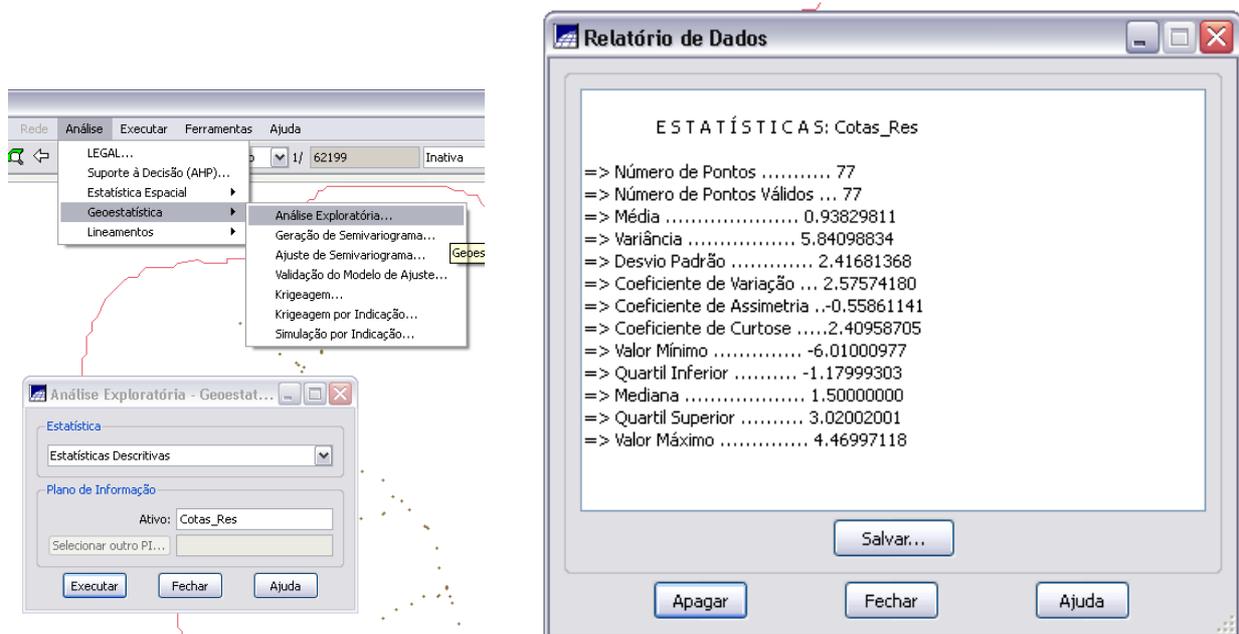


Figura 4 – Análise >> Análise exploratória >> Estatísticas Descritivas.

A segunda opção dentre as estatísticas descritivas se refere ao *Histograma*, que apresenta a distribuição dos dados em classes e uma curva Gaussiana para fins de

comparação. Para efeito de comparação, foram realizados três histogramas com 5, 10 e 20 classes de dados, respectivamente (Figura 5).

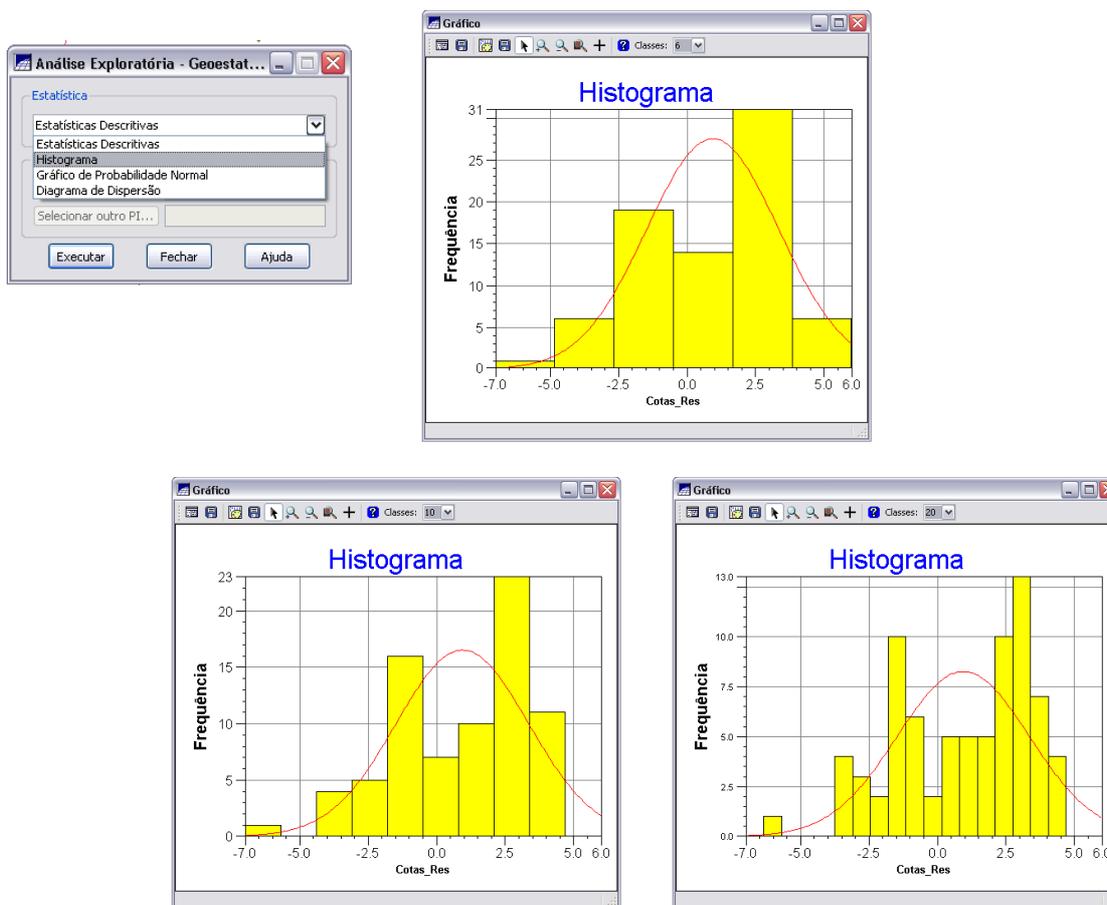


Figura 5 – Histogramas apresentados em 5, 10 e 20 classes de dados (em amarelo), com as respectivas distribuições Gaussianas, com médias e desvios padrões calculados a partir dos dados amostrados (em vermelho). A distribuição dos dados é negativamente assimétrica com coeficiente de assimetria igual a -0,5261.

O recurso do Gráfico da Probabilidade Normal também costuma ser utilizado para uma melhor caracterização dos dados (representa a probabilidade em relação à normalização da variável) (Figura 6).

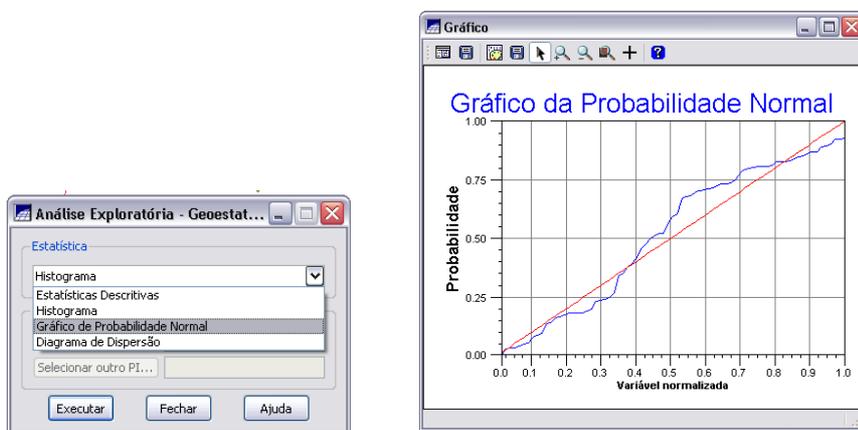


Figura 6 – Gráfico da Probabilidade Normal.

3.2 Análise Variabilidade Espacial por Semivariograma

Na geoestatística, a *análise da variabilidade espacial por semivariograma* é a etapa mais importante de todo processo, pois o modelo de semivariograma escolhido corresponde à interpretação da estrutura de correlação espacial a ser utilizada nos procedimentos inferenciais da krigagem.

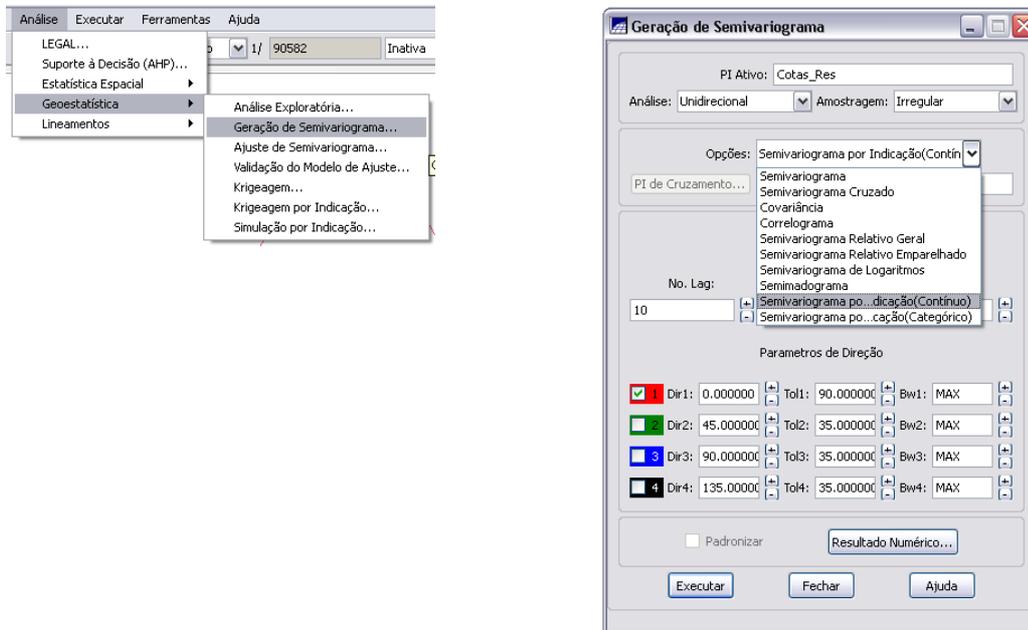


Figura 7 – Análise >> Geoestatística >> Geração de semivariograma >> semivariograma por indicação.

Geração de semivariograma por indicação para dados contínuos

No processo de krigagem por indicação os valores do atributo são transformados, segundo uma função não linear. A codificação por indicação, sobre um conjunto de dados amostrais numéricos, da VA $Z(u)$, para um valor de corte z_k , gera um conjunto amostral por indicação $I(u; z_k)$ do tipo: (FELGUEIRAS, 1999)

$$I(\mathbf{u}; z_k) = \begin{cases} 1, & \text{for } Z(\mathbf{u}) \leq z_k \\ 0, & \text{for } Z(\mathbf{u}) > z_k \end{cases}$$

Neste exercício, os semivariogramas foram gerados para valores de corte considerando os quartis. Cada quartil define um valor de corte (z_k), a saber: $z_{k1} = -1,24$; $z_{k2} = 1,42$ e $z_{k3} = 3,02$.

Por se tratar de valores numéricos, utiliza-se a opção de semivariograma por indicação para dados contínuos. No caso de informações temáticas, a opção será semivariograma para dados categóricos.

Assumindo isotropia para cada z_k definido acima, um semivariograma por indicação (omnidirecional) será gerado e ajustado.

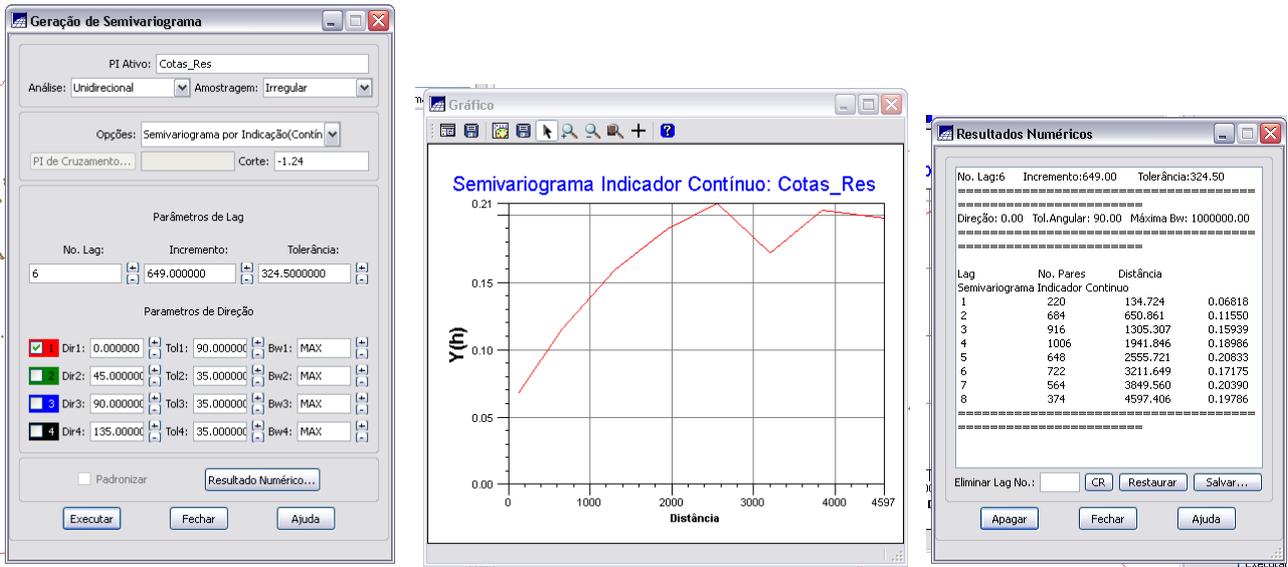
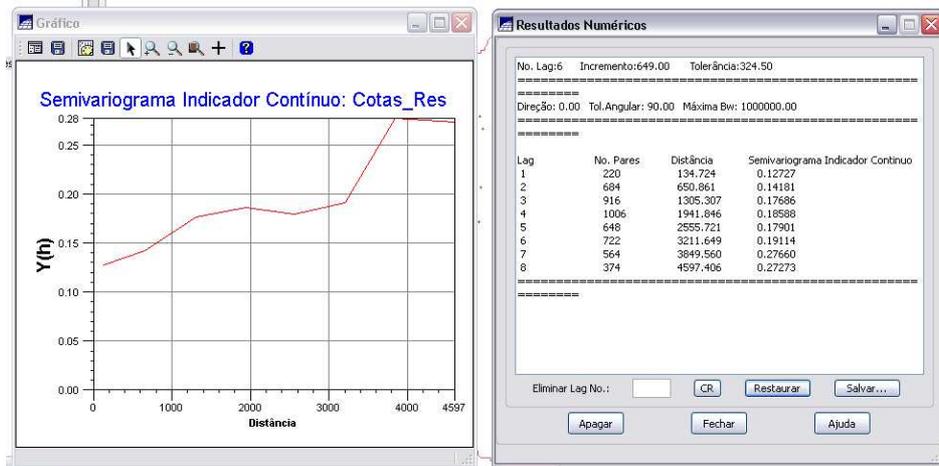
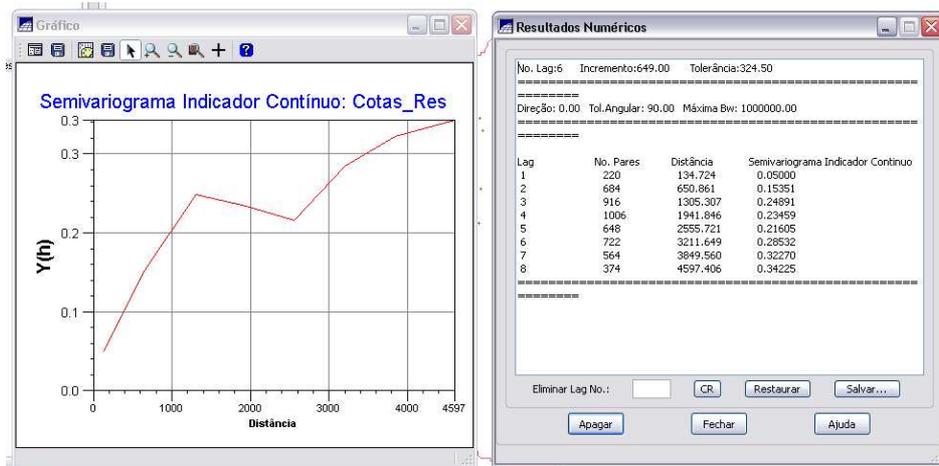


Figura 8 – Semivariograma por indicação: corte: -1.24, nº Lag: 6, incremento: 649,0 , tolerância: 324,5. É possível visualizar os resultados numéricos para melhor entendimento dos resultados.



Figuras 9 e 10 – Semivariogramas por indicação (e respectivos resultados numéricos): cortes: 1.42 e 3.02, nº Lag: 6, incremento: 649,0 , tolerância: 324,5.

□ Os parâmetros de Lag foram alterados buscando melhorar o semivariograma até que se obtivesse um resultado considerado adequado.

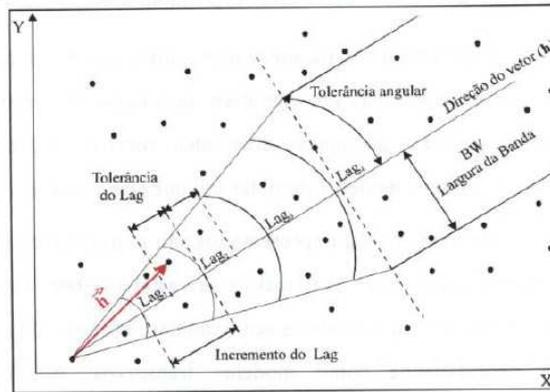


Figura 11 – Parâmetros para o cálculo do semivariograma a partir de amostras irregularmente espaçadas em duas dimensões (Fonte: CAMARGO, 1997).

Segundo Camargo (1997), *Lag* refere-se a uma distância pré-definida que é utilizada no cálculo do semivariograma. Tomando como exemplo a Figura 11, e como referência o *Lag₂*, supõe-se um incremento de *Lag* igual a 100 metros, com tolerância de 50 metros. Considere a direção de medida a 45° com tolerância angular de 22,5°. Desta forma, qualquer par de observações cuja distância esteja compreendida entre 150 e 250 metros e entre 22,5° e 67,5° será incluído no cálculo do semivariograma de *Lag₂*. Este processo se repete para todos os *Lags* (CAMARGO, 1997).

Ainda com referência na Figura 11, a largura de banda – *BW* – refere-se a um valor de ajuste a partir do qual se restringe o número de pares de observações para o cálculo do semivariograma (CAMARGO, 1997).

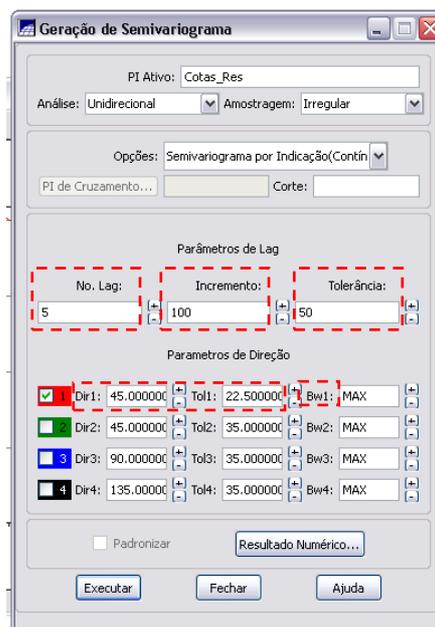
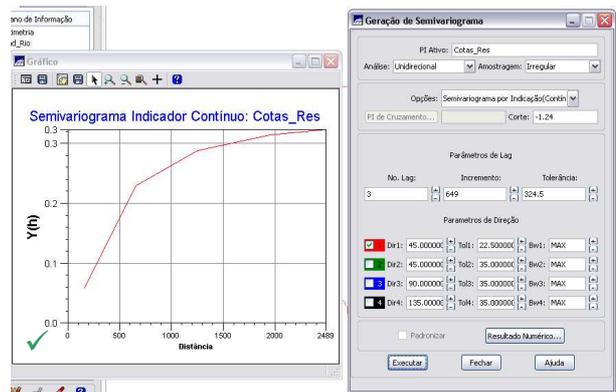
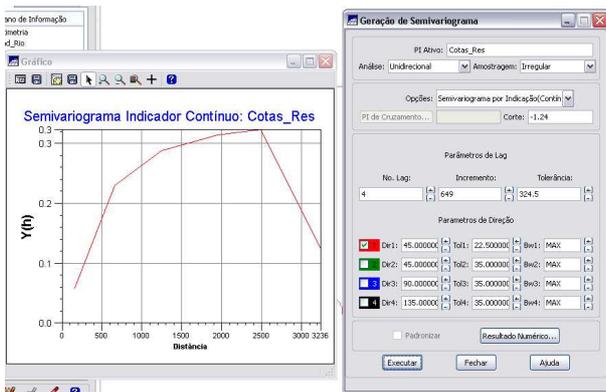
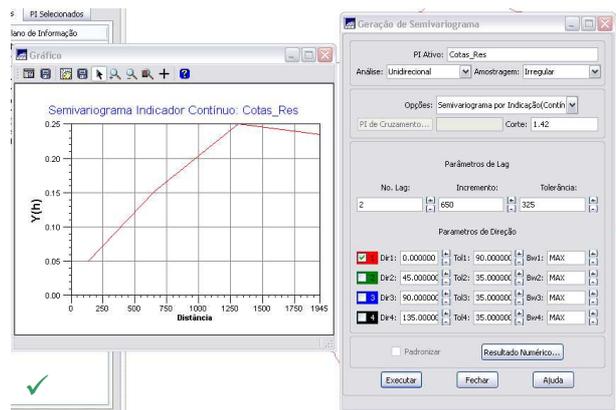
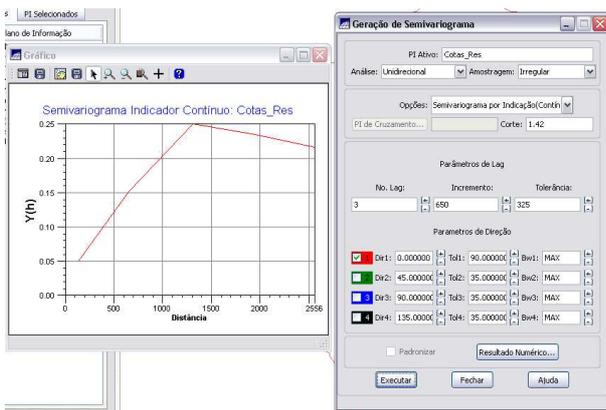


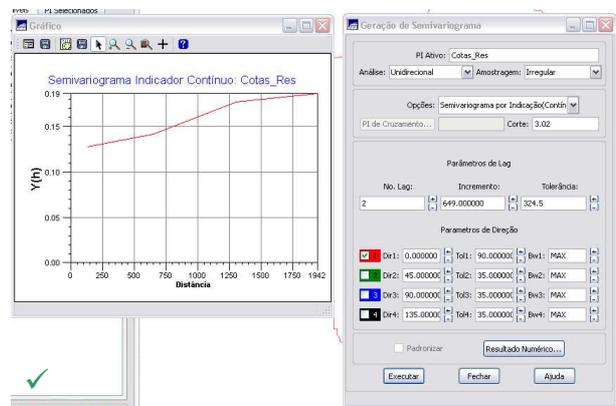
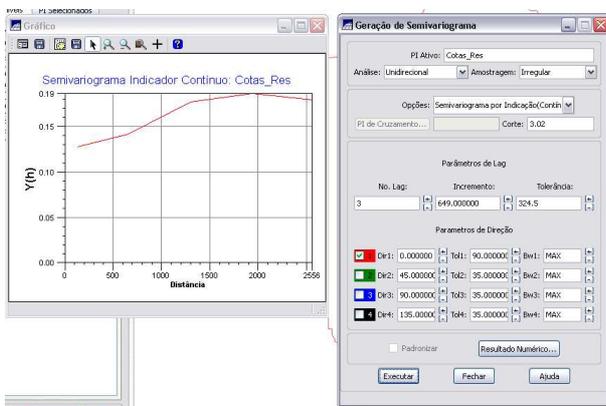
Figura 12 – Janela de definição dos parâmetros para geração do semivariograma no Spring.



Figuras 13 e 14 – Semivariogramas por indicação (e respectivos resultados numéricos): corte: -1.24, nº Lag: 4 e 3, incremento: 649,0 , tolerância: 324,5. O melhor resultado foi obtido com Lag = 3.



Figuras 15 e 16 – Semivariogramas por indicação (e respectivos resultados numéricos): corte: 1.42, nº Lag: 3 e 2, incremento: 649,0 , tolerância: 324,5. O melhor resultado foi obtido com Lag = 2.



Figuras 17 e 18 – Semivariogramas por indicação (e respectivos resultados numéricos): corte: 3.02, nº Lag: 3 e 2, incremento: 649,0 , tolerância: 324,5. O melhor resultado foi obtido com Lag = 2.

3.3 Ajuste ou Modelagem do Semivariograma

O gráfico do semivariograma experimental, $\gamma(h)$, é formado por uma série de valores, conforme ilustram as Figuras 8, 9, 10, 13, 14, 15, 16, 17 e 18, sobre os quais objetiva-se ajustar uma função (modelo). É importante que o semivariograma experimental possua variações semelhantes ao de um modelo teórico (esférico, exponencial, gaussiano, potência) a ser ajustado. Isto garante que o ajuste seja mais

representativo, ou seja, que o modelo ajustado represente a tendência de $\gamma(h)$ em relação a h . Deste modo, as estimativas obtidas a partir da krigagem serão mais exatas e, portanto mais confiáveis.

Uma vez gerado o semivariograma omnidirecional o passo seguinte é o ajuste ou modelagem do mesmo.

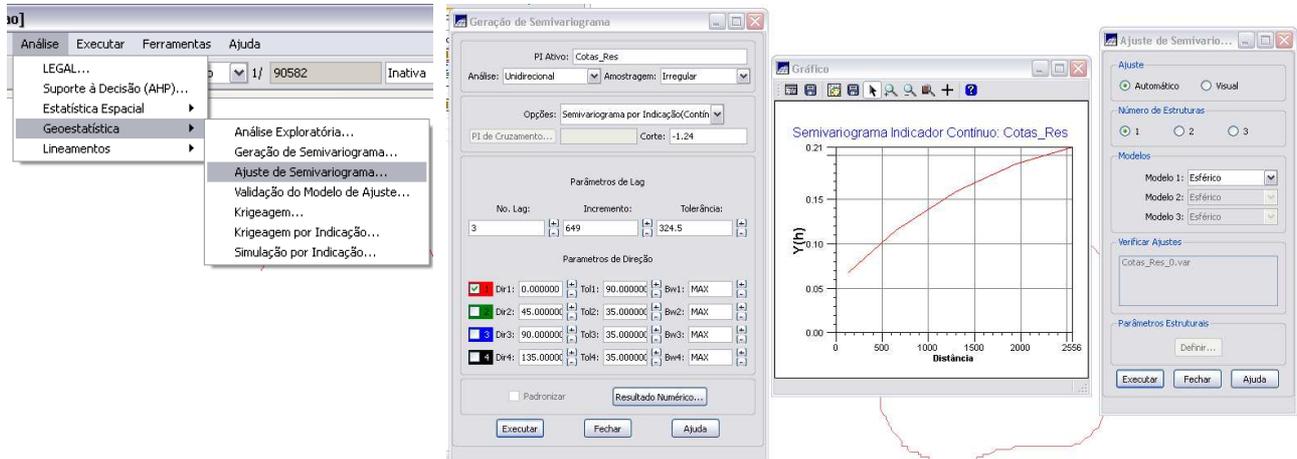
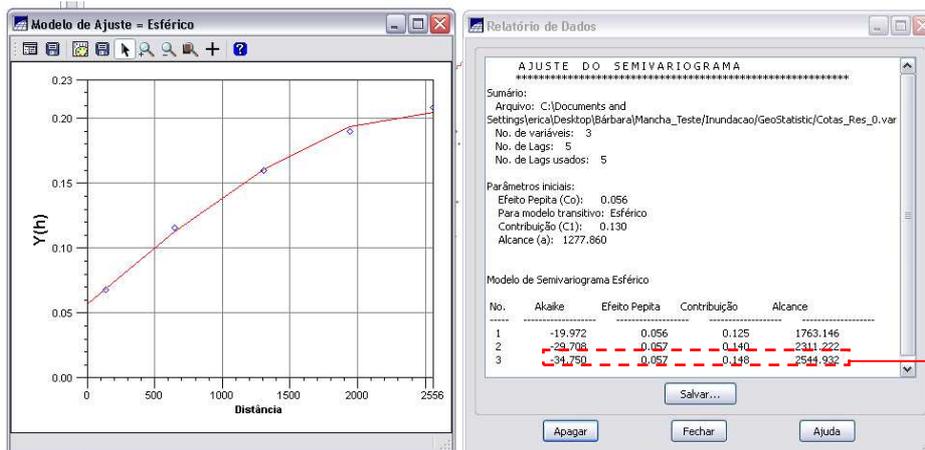
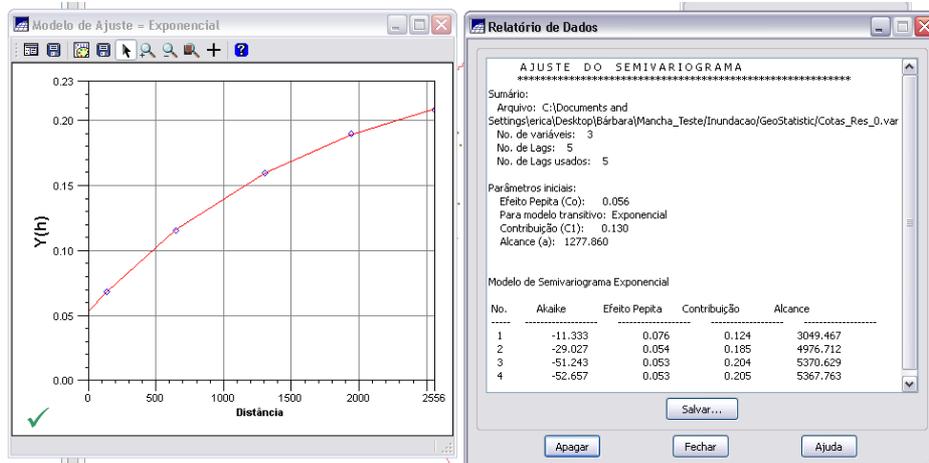


Figura 19 – Análise >> Geoestatística >> Ajuste de semivariograma >> semivariograma por indicação.

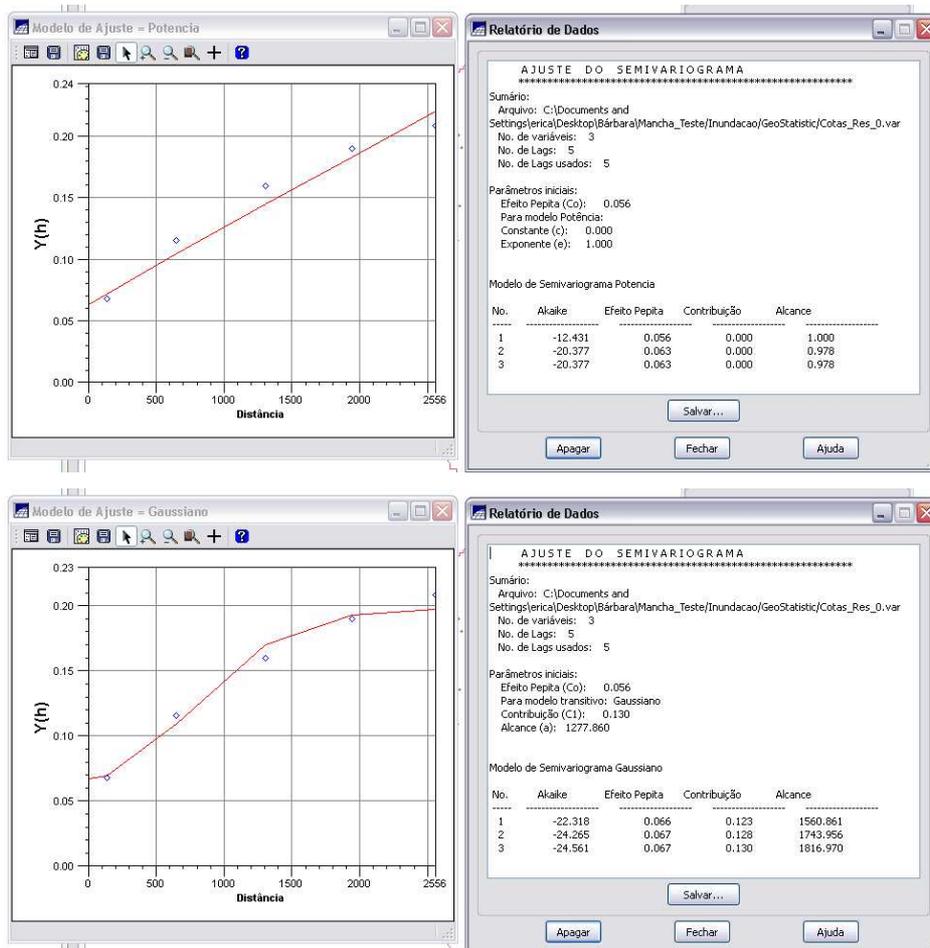


Parâmetros do modelo (efeito pepita, contribuição e alcance), são tomados sempre referentes ao menor valor de Akaike



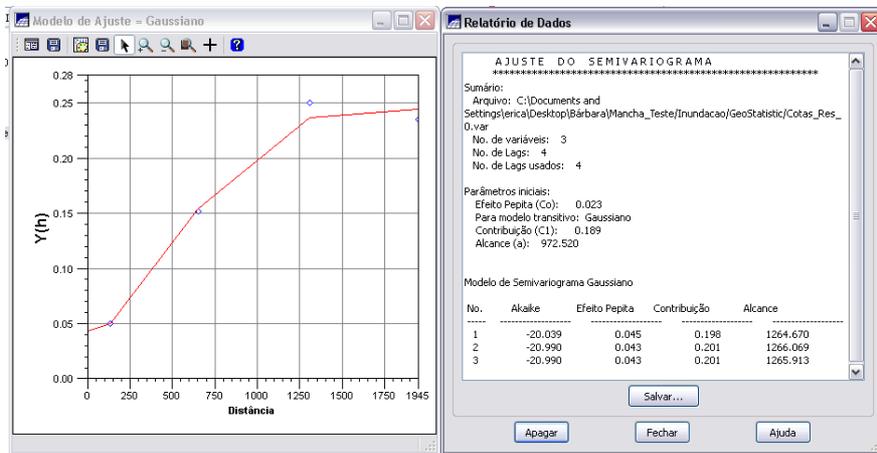
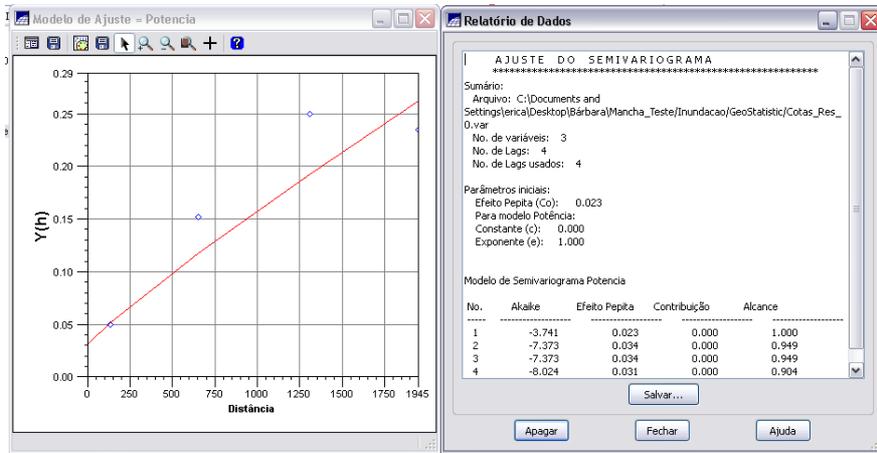
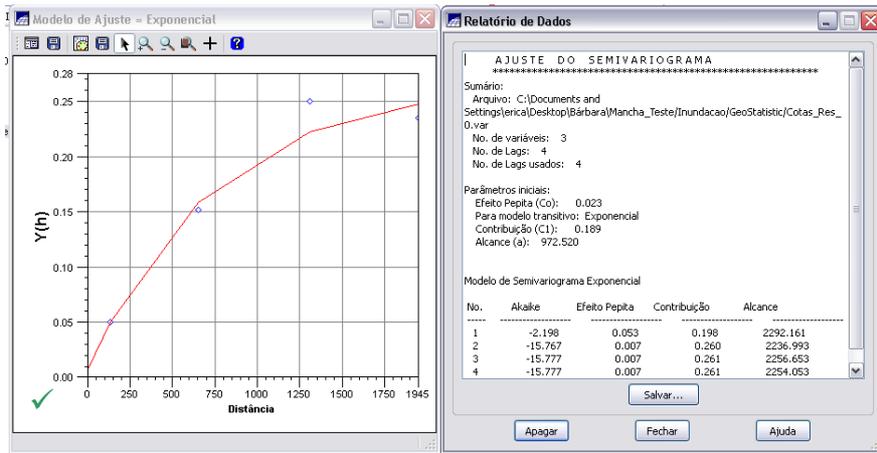
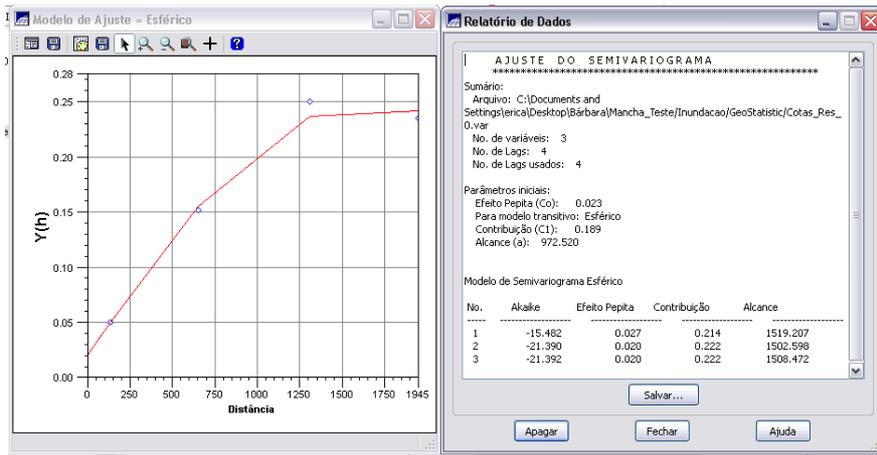
Figuras 20 e 21 – Modelos de ajuste esférico e exponencial para o semivariograma de corte = -1.24 e lag = 3.

Foram gerados quatro modelos de ajuste (*esférico*, *exponencial*, *potência* e *gaussiano*) do último semivariograma produzido (de corte = -1.24 e lag = 3; Figura 14). Ao se comparar os modelos, o que melhor representou os dados foi o *exponencial*, cujos resultados (gráfico e parâmetros do modelo) são mostrados na Figura 21. O modelo *esférico* produziu resultados muito semelhantes. Já os modelos *potência* e *gaussiano* não foram capazes de representar o semivariograma (Figuras 22 e 23).

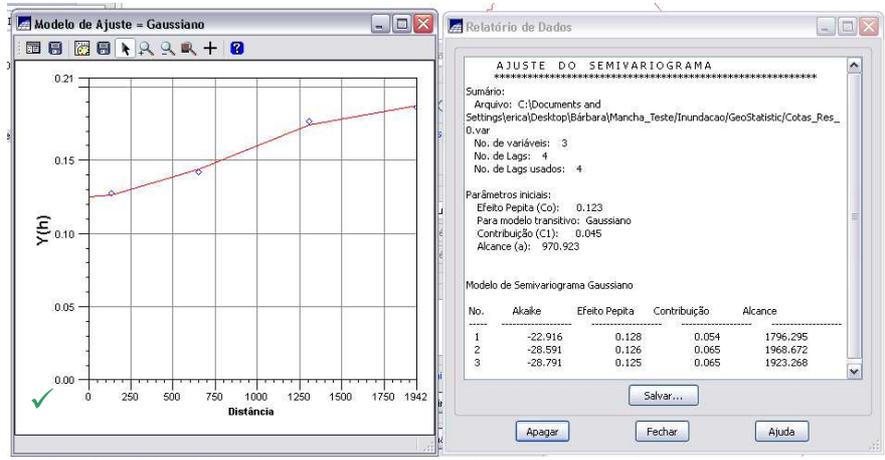
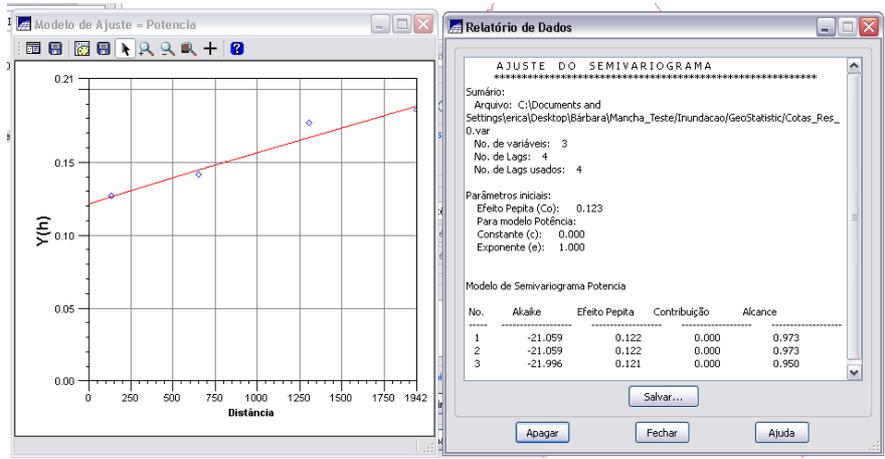
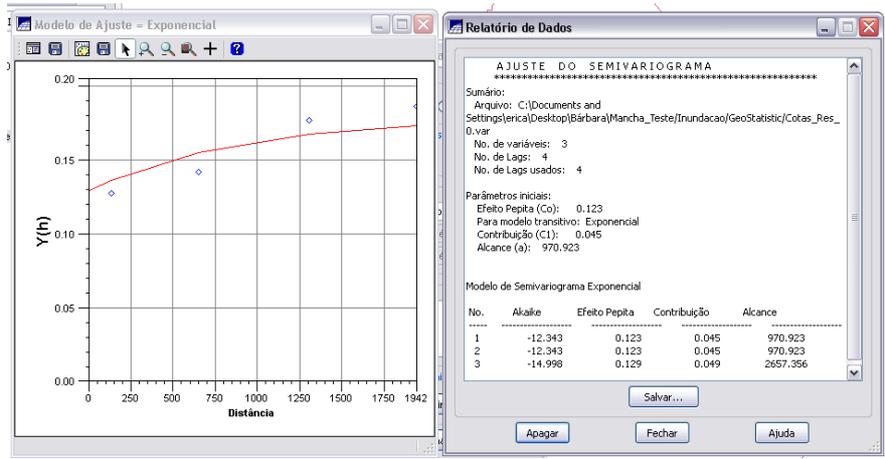
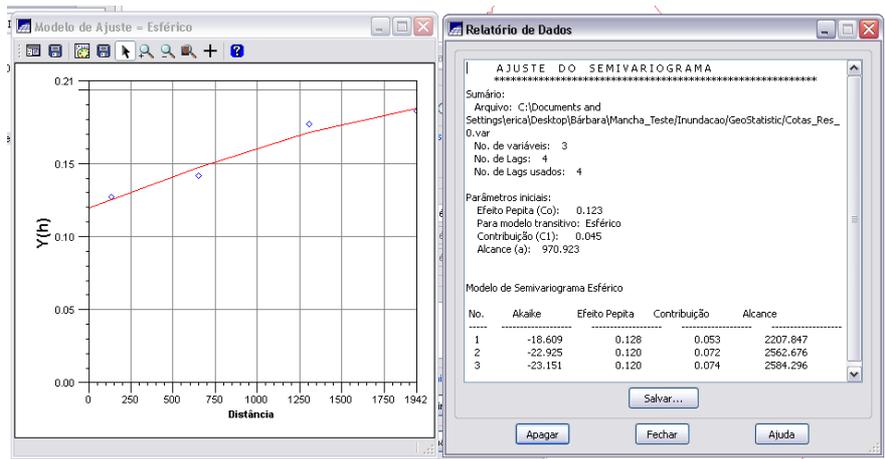


Figuras 22 e 23 – Modelos de ajuste *potência* e *gaussiano* para o semivariograma de corte = -1.24 e lag = 3.

O mesmo procedimento foi repetido para os valores de corte 1.42 e 3.02 (lag = 2 para ambos). Os quatro modelos de ajuste gerados para cada um dos casos são mostrados nas Figuras 24, 25, 26, 27, 28, 29, 30 e 31. A comparação dos resultados de cada grupo de modelos indicou que o modelo de ajuste que melhor representou os dados para o corte = 1,42 foi o *exponencial*, cujos resultados (gráfico e parâmetros do modelo) são mostrados na Figura 25; para o corte = 3.02, o modelo *gaussiano* foi o mais representativo de seu semivariograma (Figura 31).



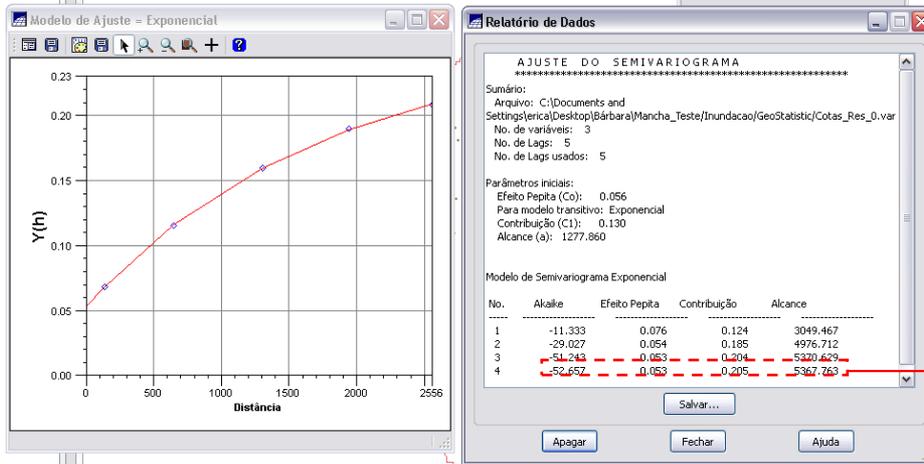
Figuras 24, 25, 26 e 27 – Modelos de ajuste esférico, exponencial, potência e gaussiano para o semivariograma de corte = 1.42 e lag = 2.



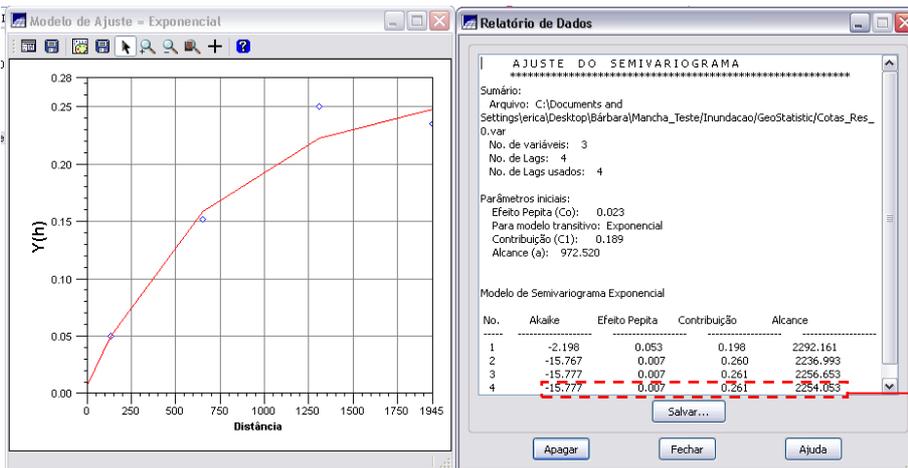
Figuras 28, 29, 30 e 31 – Modelos de ajuste esférico, exponencial, potência e gaussiano para o semivariograma de corte = 3.02 e lag = 2.

3.4 Validação do Modelo de Ajuste do Semivariograma

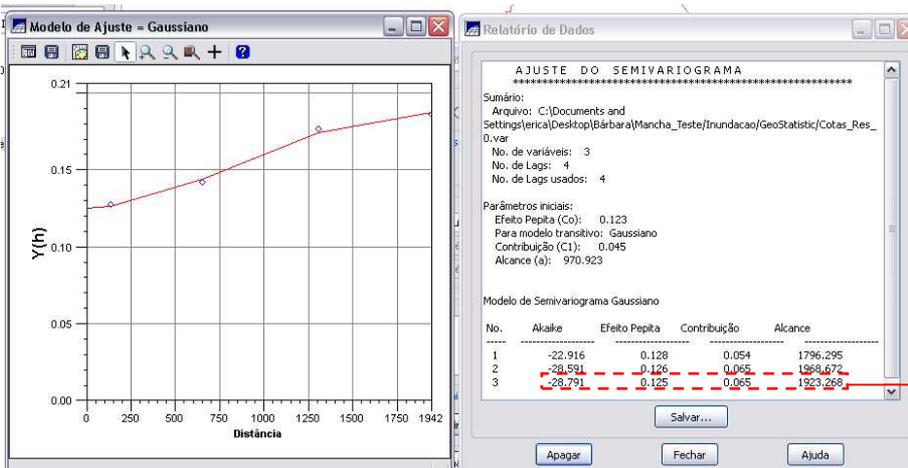
Uma vez realizado o procedimento de ajuste do semivariograma, os parâmetros do modelo serão utilizados em sua validação. Estes dados são coletados no *Relatório de Dados*, e então inseridos nos respectivos campos da interface de *Parâmetros Estruturais*.



- corte = -1.24 lag = 3
- Akaike (menor valor): -52,657
 - efeito pepita: 0,053
 - contribuição: 0,205
 - alcance: 5367,763



- corte = 1.42 lag = 2
- Akaike (menor valor): -15,777
 - efeito pepita: 0,007
 - contribuição: 0,261
 - alcance: 2254,053



- corte = 3.02 lag = 2
- Akaike (menor valor): -28,791
 - efeito pepita: 0,125
 - contribuição: 0,065
 - alcance: 1923,268

Figuras 32, 33 e 34 – Parâmetros dos modelos de ajuste para cada semivariograma produzido.

A análise do semivariograma compreende o levantamento do semivariograma experimental e posteriormente o ajuste a uma família de modelos teóricos. Em toda esta seqüência, existe sempre um certo grau de incerteza sobre os parâmetros ajustados aos

modelos. Esta incerteza é o *erro da estimativa*, o qual pode ser obtido através do procedimento chamado *validação do modelo*, processo de validação envolve a re-estimação dos valores conhecidos através dos parâmetros ajustados ao modelo do semivariograma.

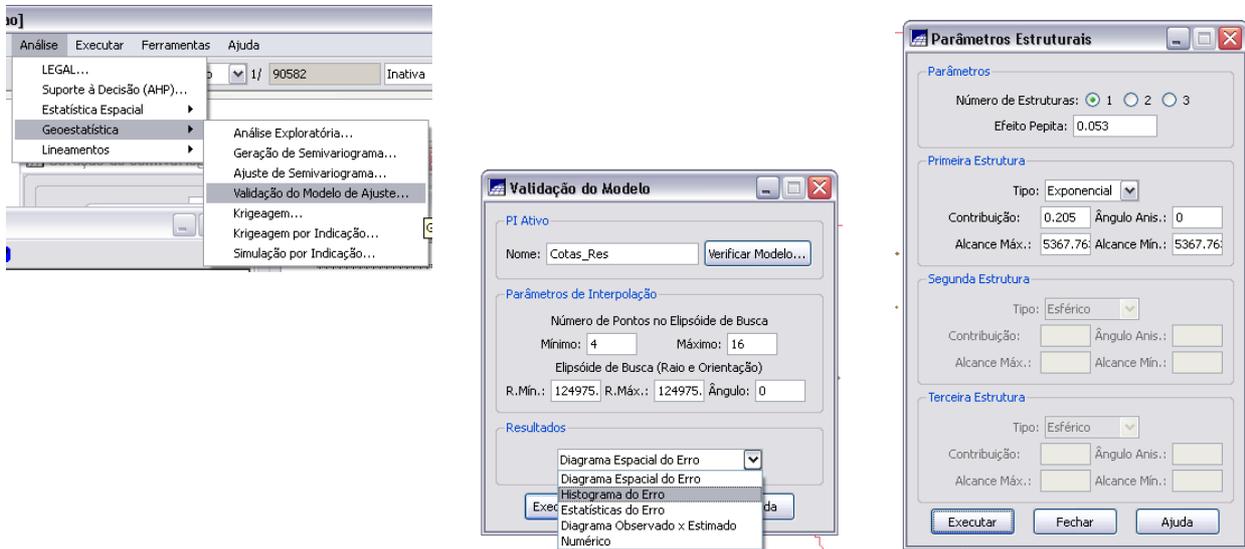
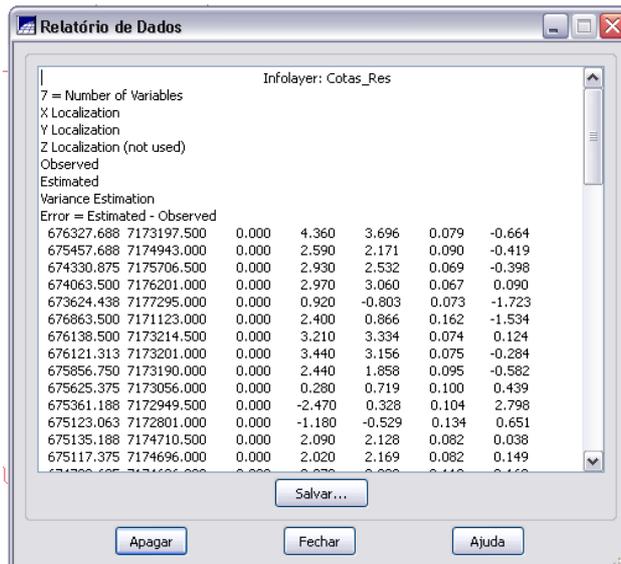
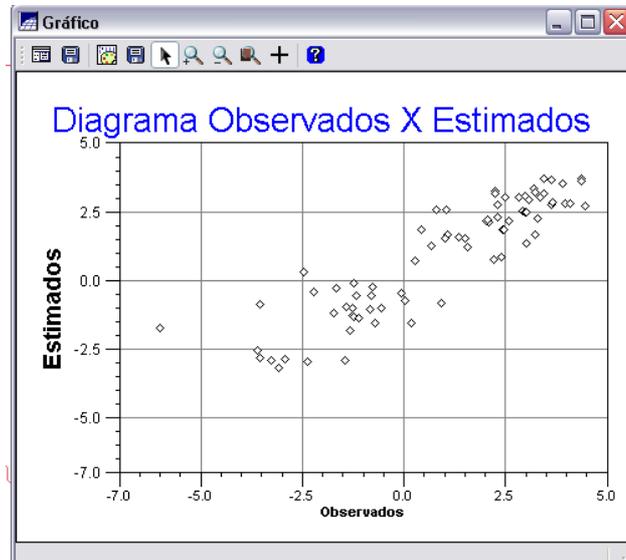
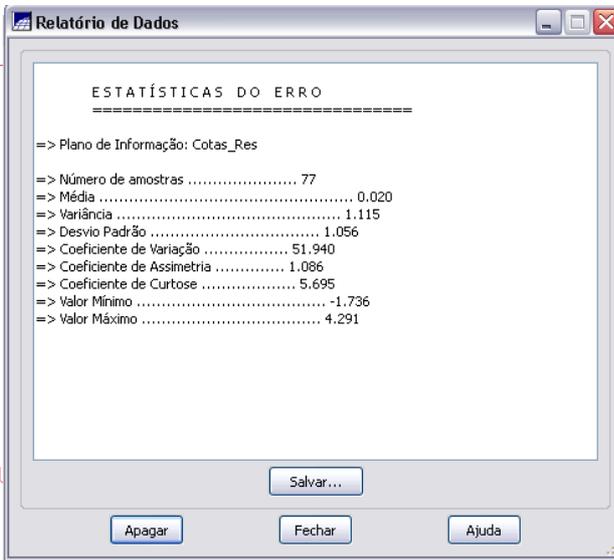
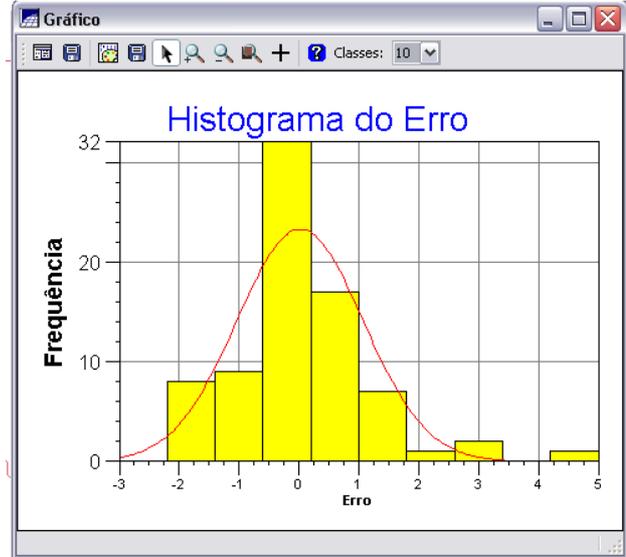
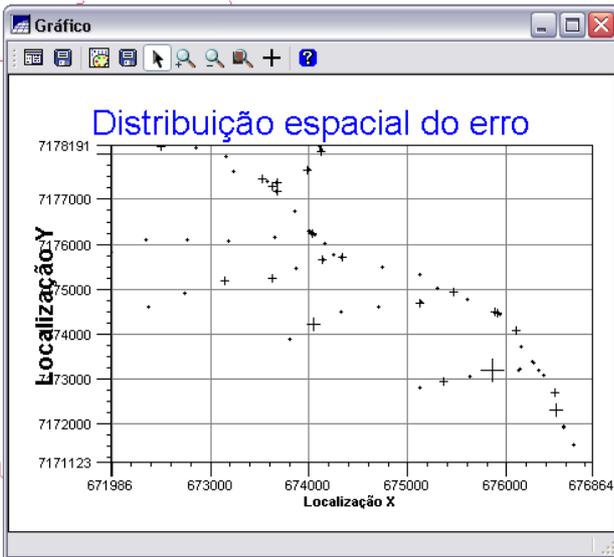


Figura 35 – *Análise >> Geoestatística >> Validação de Modelo de Ajuste >> Parâmetros Estruturais.*

Antes de executar a krigeagem é recomendável verificar os resultados da validação. Problemas óbvios podem ser identificados com os parâmetros de entrada (por exemplo, a especificação do semivariograma) ou com os dados (*outliers*, por exemplo).

O módulo de validação desenvolvido no Spring utiliza a subrotina "kt3d" da GSLIB (DEUTSCH e JOURNAL, 1992) e fornece as saídas: diagrama espacial do erro, histograma do erro, estatísticas do erro, diagrama dos valores observados x estimados, e os resultados numéricos.

Os *Parâmetros de Interpolação Mínimo e Máximo* referem-se ao *Número de Pontos no Elipsóide de Busca*. São preenchidos com valores *default* (4 e 16, respectivamente). São definidos os raios e a orientação do *Elipsóide de Busca*. Os campos *R.min*, *R.max* e *Ângulo* são inicializados, para um caso isotrópico, com seguintes valores *default*: *R.min* e *R.max* equivalem, em metros, à diagonal do retângulo envolvente do Projeto e o *Ângulo* possui um valor qualquer, por exemplo igual a zero. Evidentemente que se a anisotropia faz-se presente, esses parâmetros devem ser ajustados e escolhidos de acordo (DEUTSCH e JOURNAL, 1992).



Figuras 36, 37, 38, 39 e 40 – resultados obtidos na validação do modelo de ajuste do semivariograma (corte=-1.24, lag=3): Diagrama Espacial do Erro: os símbolos tipo cruz na Figura 36 indicam a localização geográfica das amostras e a magnitude do erro (para os símbolos pequenos o erro é menor e vice-versa); Estatísticas do Erro, Histograma do Erro, Diagrama Observado X Estimado, Numérico (Relatório de Erros).

3.5 Krigeagem por Indicação

A etapa final do processo geoestatístico é a inferência dos valores nos pontos da grade não amostrados, utilizando o estimador de krigeagem por indicação.

O módulo de *Krigeagem por Indicação* implementado no SPRING baseia-se na subrotina da GSLIB (DEUTSCH e JOURNEL, 1992). Este módulo possibilita a espacialização, segundo uma grade regular, de atributos espaciais de natureza contínua e categórica. Neste caso os dados utilizados referem-se a fenômenos contínuos. Além disso, é gerada uma representação de grade regular, com valores de desvio padrão, representativa das incertezas associadas às estimativas do atributo.

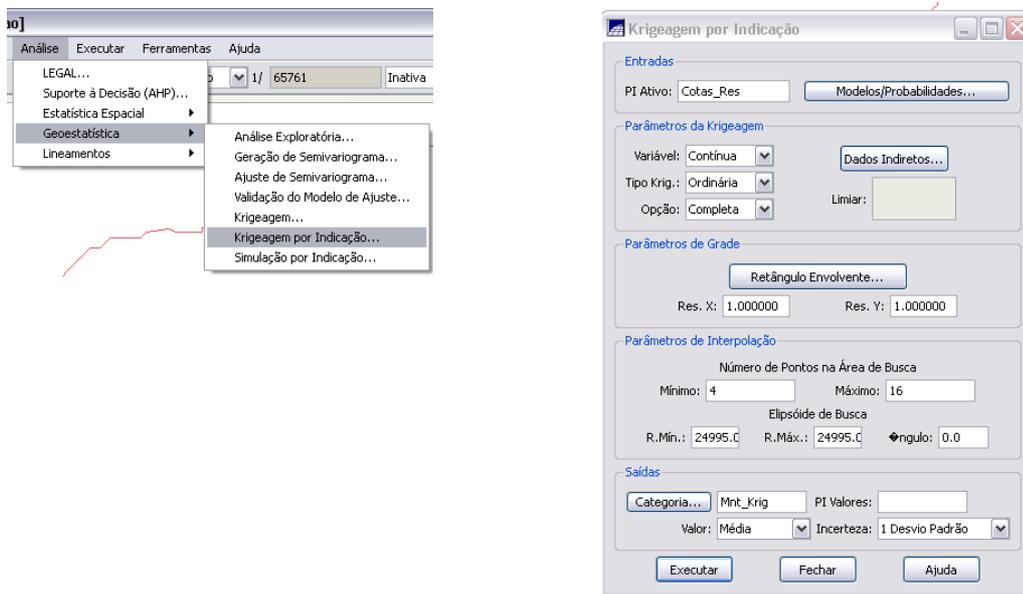


Figura 41 – Análise >> Geoestatística >> Krigeagem por Indicação >> Modelos Probabilidades (Parâmetros Estruturais)

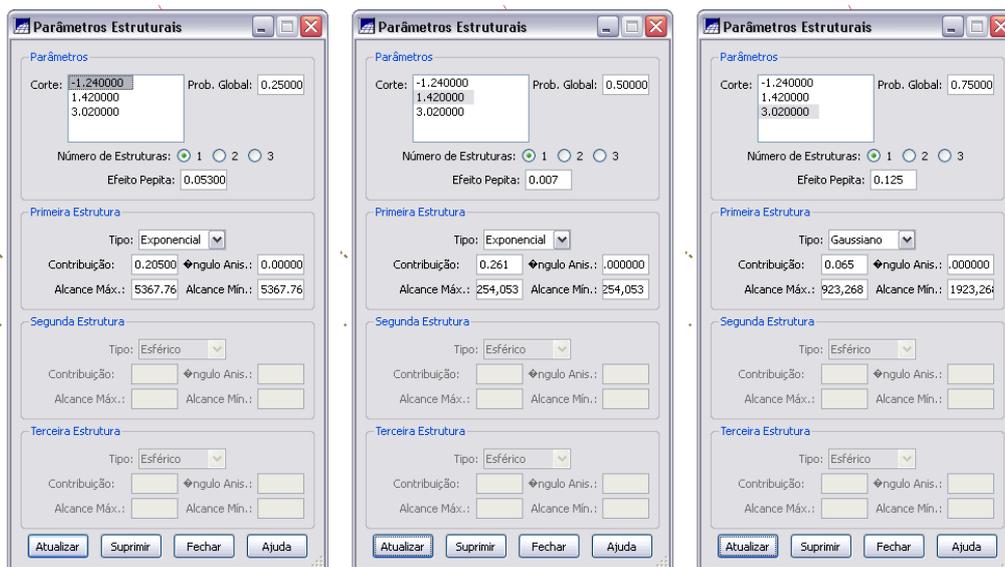


Figura 42 – Parâmetros estruturais dos modelos de ajuste dos semivariogramas relativos aos três cortes.

Definição dos Parâmetros Estruturais: foram selecionados cada valor de corte apresentado na lista “Corte” e então os demais campos da janela *Parâmetros Estruturais* foram atualizados automaticamente. Definiu-se também um valor de probabilidade global para o valor de corte selecionado. Cada valor de corte precisa ter uma probabilidade global que varie entre 0 e 1 e a soma das probabilidades globais associadas aos valores de corte deve ser igual a 1.

Após a definição dos parâmetros estruturais para os valores de corte considerados, iniciou-se a execução da *krigeagem por indicação*.

Inicialmente, o tipo de krigeagem escolhido foi o “Ordinária”. Os campos *Res.X* e *Res.Y* permaneceram preenchidos com valores *default*, segundo as definições para o *Plano de Informação* ativo (optou-se por não alterar estas resoluções).

Os *Parâmetros de Número de Pontos na Área de Busca Mínimo e Máximo* referem-se aos números mínimo e máximo de pontos no *Elipsóide de Busca*. Foram utilizados, inicialmente, os valores *default* 4 e 16, respectivamente.

Foram definidos também os raios e a orientação do *Elipsóide de Busca*. Os campos *R.min*, *R.max* e *Ângulo* são inicializados, para um caso isotrópico, com seguintes valores *default*: *R.min* e *R.max* equivalem, em metros, ao alcance do variograma isotrópico e o *Ângulo* igual a zero.

As saídas da *Krigeagem por Indicação* são dois *Planos de Informação* com representações em grade regular, uma com valores do atributo (*Krig_A*) e outra com incertezas da estimação (*Krig_A_Inc*). Por se tratar de atributos contínuos, foram escolhidas a média (poderia ser a mediana) como *Valor* e intervalos de confiança baseados em desvios padrões (poderia ser quantis) como *Incerteza*.

A grade de krigeagem gerada é apresentada na Figura 43 e representa os valores do atributo estimado através da média. Uma representação numérica da incerteza associada à estas estimativas também foi gerada: *PI krig_A_Inc*.

A variabilidade espacial é melhor visualizada por meio da Imagem da grade numérica gerada. Desta forma, tem-se uma visão imediata do comportamento espacial da variável (Figura 44). Os níveis de cinza mais escuros correspondem aos valores baixos do atributo estimado, e os mais claros, por sua vez, correspondem a valores altos.

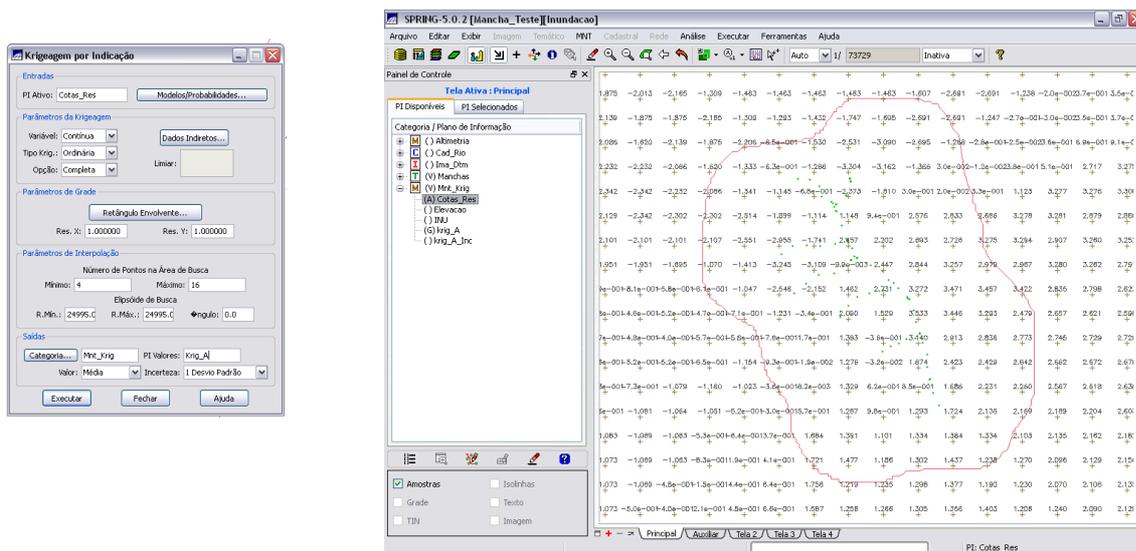


Figura 43 – Após a definição dos parâmetros estruturais de cada modelo de ajuste, para visualizando da grade de krigagem gerada, basta executar a *krigagem por indicação* e selecionar a opção “Grade” no plano de informação gerado (krig_A).

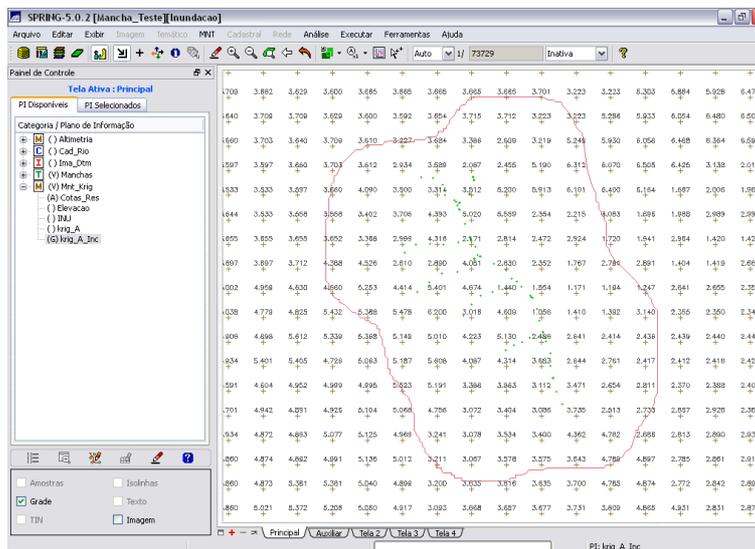


Figura 44 – Grade numérica de representação da incerteza associada às estimativas geradas com a *krigagem por indicação*.

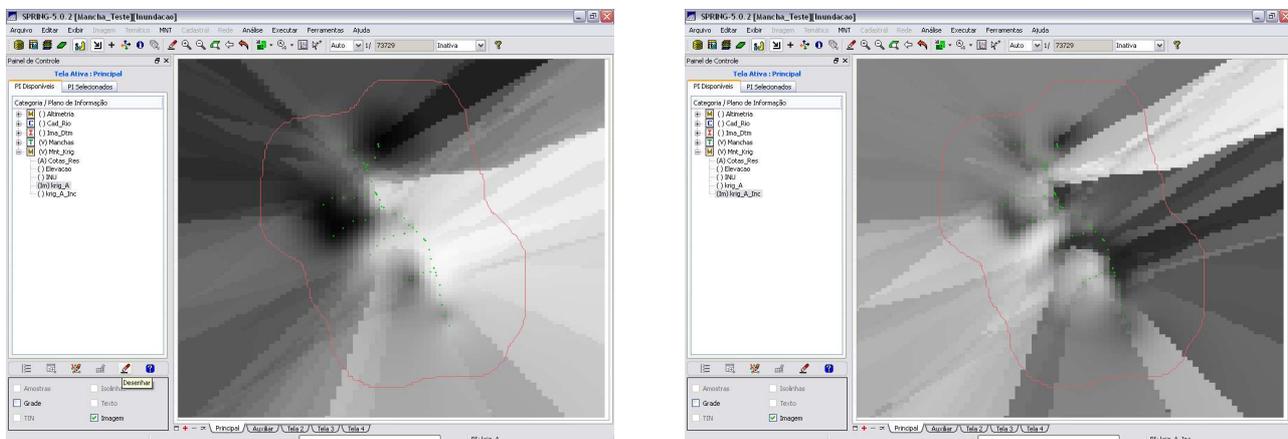


Figura 45 – Imagens das grades numéricas da *krigagem por indicação* e das incertezas associadas.

A porção extra da imagem que ultrapassa o contorno externo da área de estudo pode ser eliminada através de um recorte de imagens (Figura 47). Isto é realizado através de um programa escrito em LEGAL (*Linguagem Espacial para Geoprocessamento ALgébrico*). Entretanto, é necessário criar um PI de imagem em uma nova categoria de imagem (Figuras 46).

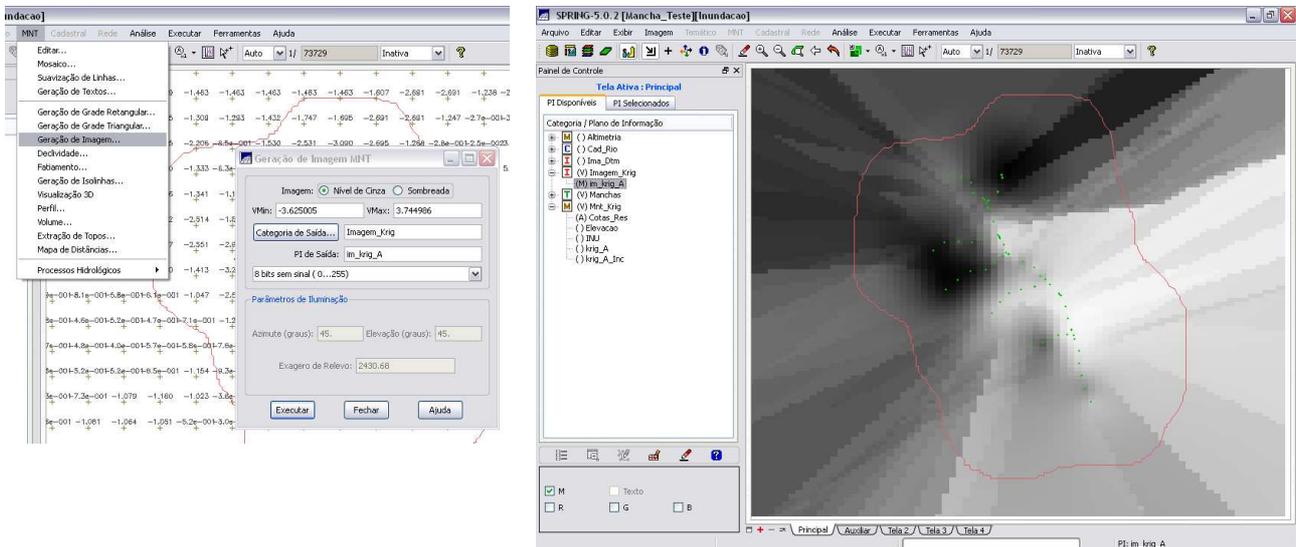


Figura 46 – Geração de PI de imagem da grade de *krigeagem por indicação* em categoria 'Imagem'.

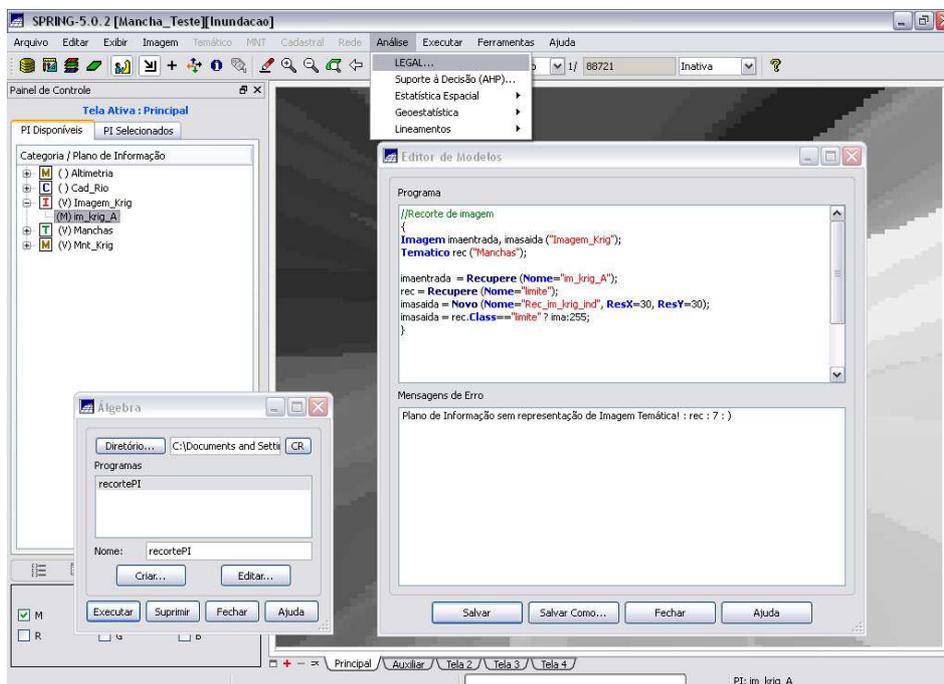


Figura 46 – Geração de PI de imagem da grade de *krigeagem por indicação* em categoria 'Imagem'.

O mesmo procedimento foi realizado para gerar a imagem da grade de incerteza, que também foi recortada conforme o limite da área de estudo (Figura 48).

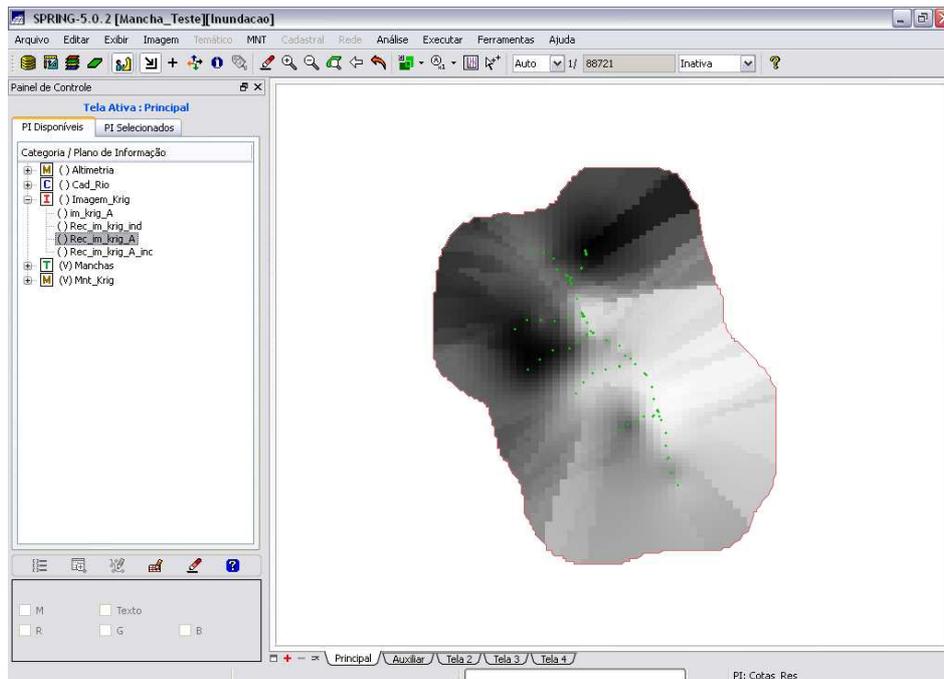


Figura 47 – PI da imagem da *krigeagem por indicação* recortado conforme o limite da área de estudo.

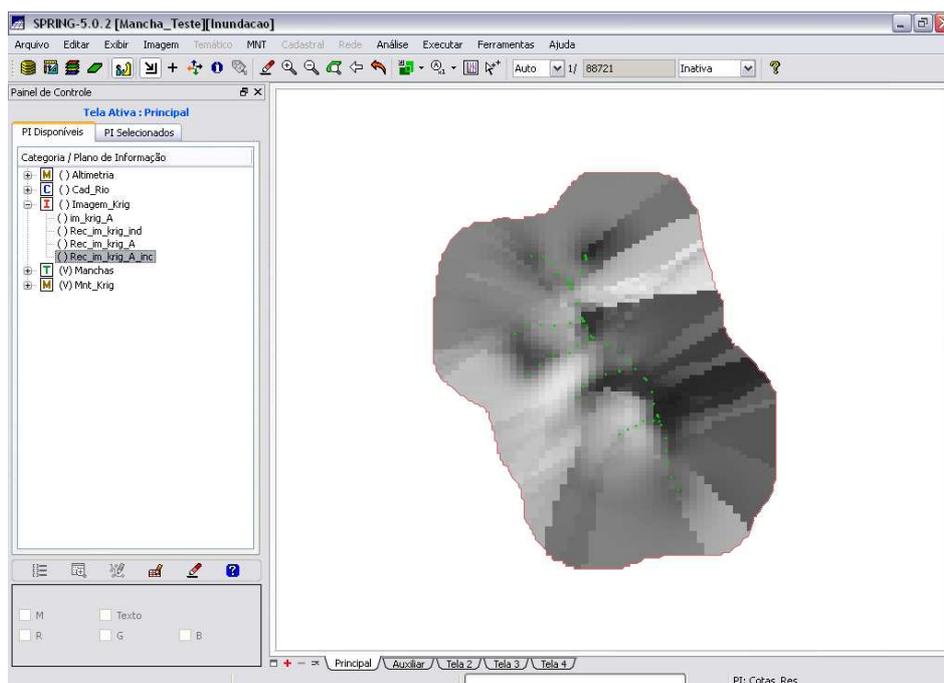


Figura 48 – PI da imagem da grade de incerteza gerada pela *krigeagem por indicação* recortado conforme o limite da área de estudo.

4 CONCLUSÃO

A *krigeagem por indicação* consiste numa operação não-paramétrica e não considera nenhum tipo de distribuição de probabilidade a priori para a variável aleatória. Ao invés

disso, ela possibilita a construção de uma aproximação discretizada da fdc de $Z(u)$. Os valores de probabilidades discretizados podem ser usados diretamente para se estimar valores característicos da distribuição, tais como: valor médio, variância, moda, quantis e outros.

Por meio deste *Laboratório 2* foi possível obter conhecimentos de análise geoestatística, aplicando-os no *software* SPRING e analisando os seus resultados a partir das imagens, semivariogramas e relatórios gerados.

Referências Bibliográficas

CAMARGO, E. C. G. *Desenvolvimento, implementação e teste de procedimentos geoestatísticos (krigeagem) no Sistema de Processamento de Informações Georeferenciadas (SPRING)*. **Dissertação** (Mestrado em Sensoriamento Remoto) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, 1997.

DEUTSCH, C. V.; JOURNAL, A. G. **GSLIB Geostatistical Software Library and User's Guide**. Oxford University Press, 1998.

FELGUEIRAS, C. A. *Modelagem Ambiental com Tratamento de Incertezas em SIG: O paradigma Geoestatístico por Indicação*. **Tese** (Doutorado em Computação Aplicada), São José dos Campos, INPE, 1999.