



# **KRIGAGEM INDICATIVA APLICADA À ELABORAÇÃO DE MAPAS PROBABILÍSTICOS DE RISCOS**

**PAULO M. BARBOSA LANDIM**

Professor Emérito da Universidade Estadual Paulista  
Professor Voluntário do Depto. Geologia Aplicada  
UNESP/Rio Claro

**JOSÉ RICARDO STURARO**

Professor Assistente Doutor do Depto. Geologia Aplicada  
UNESP/Rio Claro

UNESP/campus de Rio Claro  
Departamento de Geologia Aplicada - IGCE  
Laboratório de Geomatématica  
Texto Didático 06  
2002

*Reprodução autorizada desde que citada a fonte*

*Norma 6023-2000/ABNT ( <http://www.abnt.org.br>):*

*LANDIM, P.M.B. & STURARO, J.R. Krigagem indicativa aplicada à elaboração de mapas probabilísticos de riscos. DGA,IGCE,UNESP/Rio Claro, Lab. Geomatématica, Texto Didático 06, 19 pp. 2002. Disponível em <<http://www.rc.unesp.br/igce/aplicada/textodi.html>>.*

*Acesso em:....*

## INTRODUÇÃO

A geoestatística calcula estimativas dentro de um contexto regido por um fenômeno natural com distribuição no espaço e, desse modo, supõe que os valores das variáveis, consideradas como regionalizadas, sejam espacialmente correlacionados. Devido a essa característica tem tido grande aplicação, principalmente para efetuar estimativas e/ou simulações de variáveis em locais não amostrados.

De uma forma geral, a metodologia geoestatística procura extrair, de uma aparente aleatoriedade dos dados coletados, as características estruturais probabilísticas do fenômeno regionalizado, ou seja, uma função de correlação entre os valores situados numa determinada vizinhança e direção no espaço amostrado. O método de estimativa básico utilizado é o da *krigagem*. Trata-se de um processo de estimativa por médias móveis, de valores de variáveis distribuídas no espaço a partir de valores adjacentes, enquanto considerados como interdependentes por uma função denominada *variograma*. Como no cálculo dessa função a somatória de diferenças ao quadrado é dividida por 2\* número de pares de valores o termo correto seria *semi-variograma*, porém é usual o emprego do termo variograma, mais sintético.

Se uma variável regionalizada  $v(i)$  for coletada em diversos pontos  $i$ , o valor de cada ponto estará relacionado com valores obtidos a partir de pontos situados a uma certa distância  $h$  e a influência será tanto maior quanto menor for a distância entre os pontos. O grau de relação entre pontos numa certa direção pode ser expresso pela covariância, sendo os pontos regularmente espaçados por múltiplos inteiros de  $h$ .

O vetor  $h$  apresentando-se infinitamente pequeno faz com que a variância seja mínima e a covariância máxima. Haverá um valor  $h$  para o qual ambas podem apresentar valores aproximadamente iguais, porém a medida que  $h$  aumenta a covariância diminui enquanto a variância aumenta, porque ocorre progressivamente maior independência entre os pontos a distâncias cada vez maiores. A semivariância distribui-se assim de 0, quando  $h=0$ , até um valor maior aproximadamente igual a variância das observações, se os dados forem estacionários, isto é, não ocorrer a presença de deriva.

Sendo  $v(1), v(2), \dots, v(i), \dots, v(n)$ , realizações de uma variável regionalizada, a estimativa não tendenciosa da semivariância é dada por

$$\gamma(h) = \frac{1}{2n} \sum (v_{i+h} - v_i)^2$$

Tais relações são mostradas quando a função  $g(h)$  é colocada em gráfico contra  $\Delta h$  para originar o *semivariograma*. Este expressa o comportamento espacial da variável regionalizada e mostra, segundo a Figura 1:

- a) a *amplitude* ( $a$ ), que indica a distância a partir da qual as amostras passam a não possuir correlação espacial e a relação entre elas torna-se aleatória; toda amostra cuja distância ao ponto a ser estimado for menor ou igual à amplitude fornece informações sobre o ponto;
- b) o *patamar* ( $C + C_0$ ), que indica o valor segundo o qual a função estabiliza-se no campo aleatório, correspondente à distância “a”; mostra a variabilidade máxima entre pares de valores, isto é, a variância dos dados e, conseqüentemente, covariância nula;
- c) a *continuidade*, pela forma do variograma, em que para  $h \cong 0$ ,  $g(h)$  já apresenta algum valor. Esta situação é conhecida como *efeito pepita* e é representada por  $C_0$ ; o efeito pepita pode ser atribuído a erros de medição ou ao fato de que os dados não foram coletados a intervalos suficientemente pequenos, para mostrar o comportamento espacial subjacente do fenômeno em estudo, isto é, não é capturado um fenômeno numa escala maior;
- d) a *anisotropia*, quando os semivariogramas mostram parâmetros diferentes para diferentes direções de amostragem.

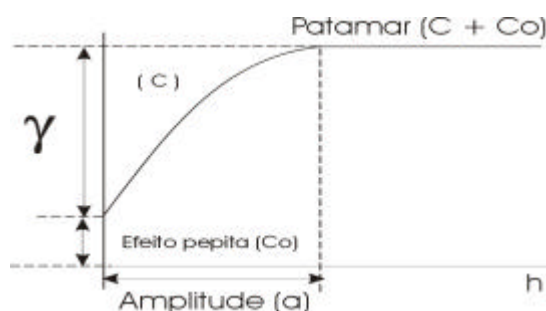


Figura 1. Modelo variográfico

É necessário verificar se o variograma estabiliza-se ou não em um patamar com o aumento de  $h$ , quando isto não ocorre há evidências de uma tendência dos dados e não estacionaridade do fenômeno em estudo.

A krigagem usa informações a partir do variograma para encontrar os pesos ótimos a serem associados às amostras com valores conhecidos que irão estimar pontos desconhecidos. É entendida como uma série de técnicas de análise de regressão que procura minimizar a variância estimada a partir de um modelo prévio, que leva em conta a dependência estocástica entre os dados distribuídos no espaço. É considerado o melhor

estimador linear não enviesado (*best linear unbiased estimator, BLUE*), em que a variância da krigagem é utilizada para definir intervalos de confiança do tipo gaussiano.

As formas mais usuais são a *krigagem simples* e a *krigagem ordinária*. A *krigagem simples* é utilizada quando a média é assumida como estatisticamente constante para toda a área e a *krigagem ordinária*, por sua vez, considera a média flutuante ou móvel por toda a área.

As variâncias de krigagem, sendo condicionadas apenas pelo arranjo geométrico dos pontos e, portanto, independentes dos valores das amostras, não são normalmente medidas de acurácia da estimativa local. Para satisfazer esta necessidade uma das soluções apontadas é a *krigagem indicativa*. O enfoque passa a ser, neste caso, não estimar um determinado valor, como na krigagem ordinária, mas sim definir áreas com maior ou menor probabilidade que um determinado evento ocorra.

A *krigagem indicativa*, utilizada neste texto, consiste basicamente na aplicação da krigagem ordinária para a variável transformada, ou seja, a variável resultante da aplicação da função não linear  $f(z) = 0$  ou  $1$ . O conceito inicial foi apresentado por **Journel (1983)** como uma proposta para construir uma função de distribuição de probabilidades acumuladas (*cumulative distribution function, "cdf"*) para a estimativa de distribuições espaciais. O conceito da transformação indicativa é dos mais simples e amigável, visto que os variogramas indicativos são os mais fáceis de modelar.

## **ESTIMATIVA DA DISTRIBUIÇÃO DE PROBABILIDADES PELA KRIGAGEM INDICATIVA**

No processo básico da krigagem, a estimativa é feita para determinar um valor médio em um local não amostrado. Pode-se, porém, também fazer estimativas baseadas em valores que se situam abaixo ou acima de um determinado nível de corte (*cutoff*).

Este procedimento, estabelecido para vários níveis de corte (percentis, decis e/ou quartis, por exemplo) de uma distribuição acumulada, conduzirá a uma estimativa de vários valores dessa distribuição em um determinado local, cuja função pode ser ajustada.

Segundo a metodologia geoestatística os valores de um determinado atributo num determinado ponto do espaço  $x$  podem ser considerados como uma realização de uma variável aleatória (VA), descrita como  $Z(x)$ . Na ponto  $x$ , portanto,  $Z(x)$  pode assumir diferentes valores para o atributo considerado, com cada valor associado a uma determinada probabilidade. Desse modo, uma variável aleatória, contínua ou discreta, após ordenada pode ser caracterizada pela sua função de distribuição acumulada condicionada, isto é, uma função de distribuição acumulada condicionada aos  $n$  dados amostrados (*conditional cumulative distribution function, "ccdf"*).

Para se atingir estes objetivos o primeiro passo, na krigagem indicativa, é transformar os dados originais em indicadores, isto é, transformar os valores que estão acima de um determinado nível de corte em zero (0) e os que estão abaixo em um (1):

$$i_j(v_c) = \begin{cases} 1 & \text{se } v_j \leq v_c \\ 0 & \text{se } v_j > v_c \end{cases},$$

onde  $v_c$  = nível de corte e  $v_j$  é o valor observado.

A frequência acumulada de valores observados, por exemplo, abaixo do nível de corte pode ser expressa por:

$$F(v_c) = \frac{1}{n} \sum_{i=1}^n i_j(v_c)$$

De modo idêntico, a proporção de valores abaixo do nível de corte pode, também ser considerada como a média ponderada dos indicadores, no caso 1, situados na vizinhança do local avaliado segundo:

$$\hat{F}(v_c) = \sum_{j=1}^n w_j \cdot i_j(v_c),$$

onde  $w_j$  são os pesos, cuja soma deve ser 1 pela condição de não viés;  $i_j$  os indicadores e  $v_c$  o nível de corte

Desta forma, são calculados os semivariogramas experimentais indicativos para determinados níveis de corte e estabelecidos os modelos variográficos para os mesmos. Os semivariogramas indicativos podem ser estimados pela função:

$$\gamma_i(h, v_c) = \frac{1}{2N_h} \sum_{i=1}^{N_h} [i(x+h, v_c) - i(x, v_c)]^2,$$

onde:

$h$  = passo (*lag*) básico

$v_c$  = nível de corte (*cutoff*)

$N$  = número de pares

Efetuada-se a krigagem ordinária pontual nos valores transformados, obtém-se a probabilidade de  $v_i < v_c$ . À medida que se incrementa  $v_c$ , obter-se-á valores estimados da função de distribuição de probabilidades acumuladas, assim expresso:

$$F\left(v; \frac{v_c}{n}\right) = E\left[\frac{i(v; v_c)}{n}\right],$$

com  $i(v; v_c) = 1$ , se  $v_i \leq v_c$ .

Definidas as funções da distribuição acumulada, pode-se, portanto, obter qualquer intervalo probabilístico da variável, ou seja:

$$F(v_j) - F(v_i)$$

onde:  $v_j > v_i$ .

Por fim, de posse dessas proporções para os vários níveis, estabelece-se a função de distribuição acumulada condicionada para os diversos locais de ocorrência da variável sob análise.

Se não há níveis de corte com especial significado com relação à variável sob estudo, o usual é escolher 9 níveis correspondentes aos decis da distribuição. Independentemente do número de níveis distribuição acumulada da curva será sempre em função de um número finito de pontos. Para uma estimativa completa haverá necessidade de interpolações, entre os níveis considerados, e extrapolações para as além do primeiro e do último nível.

Antes de efetuar a krigagem indicativa, é necessário que para cada nível de corte seja encontrado um semi-variograma e uma boa aproximação, se possível, é procurar encontrar o mesmo modelo para todos eles, principalmente aquele correspondente à mediana (**Isaaks e Srivastava, 1989**).

Exemplos podem ser encontrados em **Isaaks e Srivastava (1989)**, **Sturaro e Landim (1996)** e **Sturaro et al. (2000)**. Existem à disposição, em diversos “softwares”, programas para o cálculo da krigagem indicativa, inclusive na biblioteca do GSLIB (**Deutsch & Journel, 1998**).

Neste texto é apresentado apenas para três níveis de corte, sem preocupação com a função de distribuição acumulada condicional, uma maneira muito simples para a aplicação desta metodologia com o auxílio de um popular *software* de livre acesso, ou seja, o GEOEAS<sup>®</sup>, associado à um outro de fácil aquisição, o SURFER<sup>®</sup>, na versão 6. É esperado que o leitor, ou leitora, esteja familiarizado com esses programas.

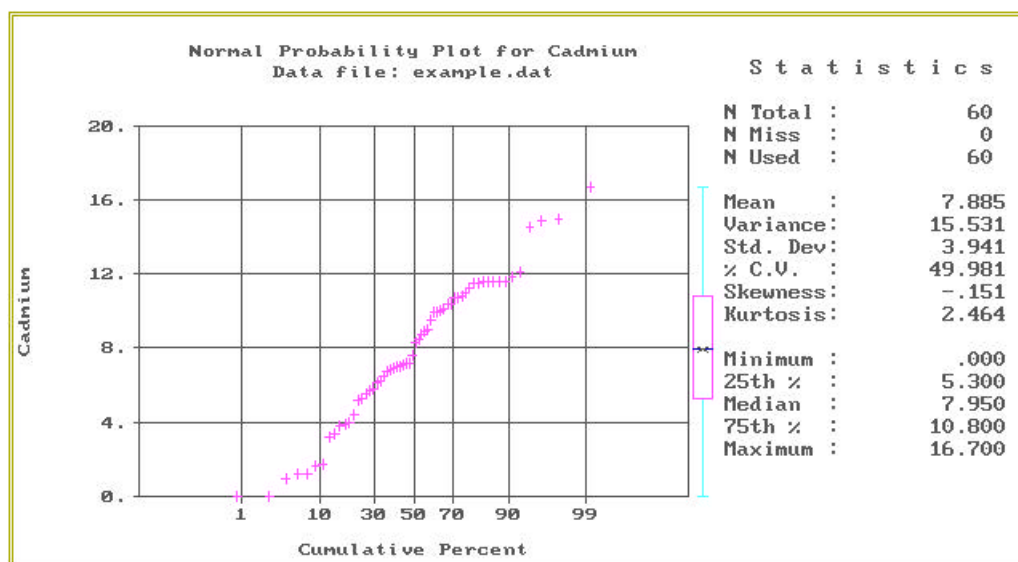
## EXEMPLO

A variável escolhida, *Cádmio* com teores em ppm, encontra-se na matriz de dados do arquivo “example.dat” que acompanha o GEOEAS.

## Escolha dos níveis de corte

Para a aplicação da krigagem indicativa o passo inicial é a escolha dos níveis de corte, segundo os quais serão obtidos os mapas de probabilidades de ocorrência. O objetivo tanto pode ser a procura de valores acima do nível de corte, como na determinação de teores anômalos de um determinado bem mineral, como valores abaixo do nível de corte, como em análise ambiental para a determinação de níveis de poluição abaixo de um certo teor. Esta decisão, portanto, é de fundamental importância, podendo tal escolha dar-se por um conhecimento “a priori”, quando já se tem informações pertinentes sobre certos valores considerados críticos em relação à variável sob estudo, ou por manipulação matemática, como no cálculo de distribuições de probabilidades acumuladas que revelarão valores de percentis. Neste texto este foi o caminho seguido.

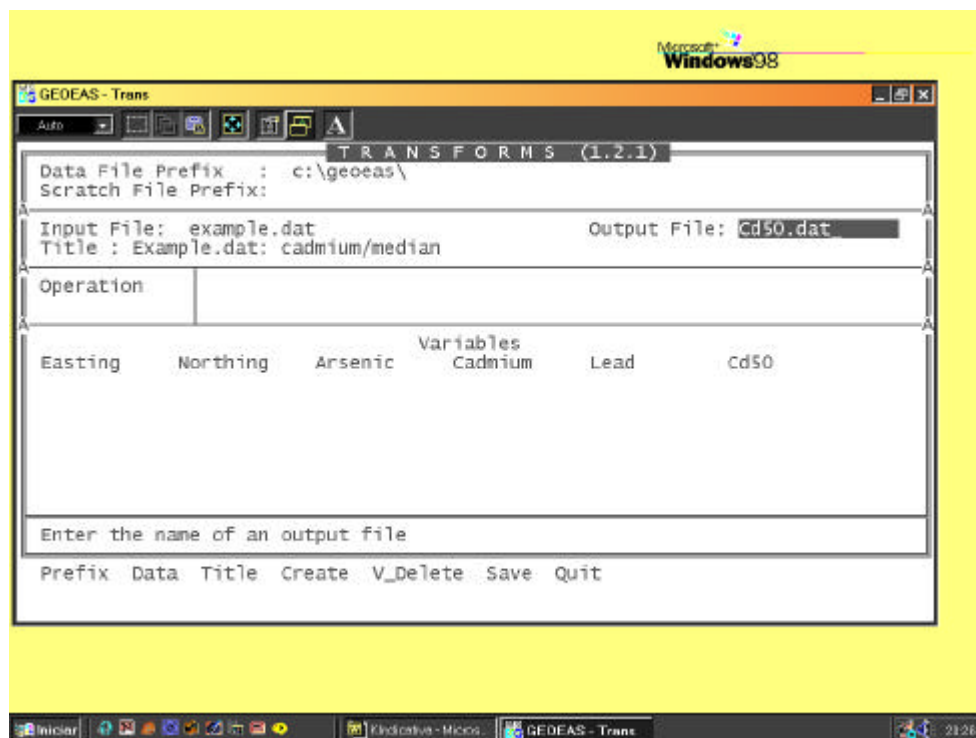
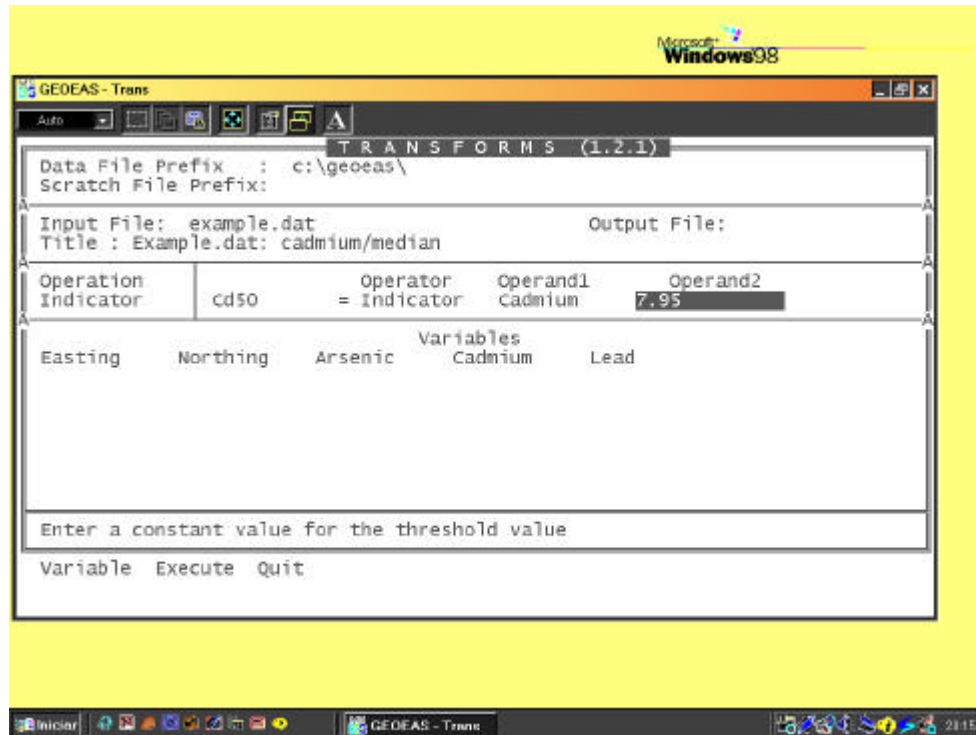
No pacote GEOEAS deve-se escolher o programa “Stat1” e em “Results” a opção “Probability Plot”. Na curva de distribuição acumulada escolhe-se os percentis julgados mais apropriados. Normalmente são escolhidos diversos, geralmente cinco com destaque para a mediana e o percentil 95. O primeiro indica a distribuição dos valores em duas porções e o segundo os valores correspondentes ao 5% superior da distribuição. Neste texto foram escolhidos, simplesmente por conveniência, os percentis 25, a mediana e percentil 75 que já aparecem calculados. Tais valores são 5.38 ppm, 7.95ppm e 10.8 ppm, que passam a ser, então, os níveis de corte.



## Transformação dos valores originais em escala binária (0 ou 1)

Escolher o programa “Trans” para transformar os valores de cádmio em indicativos do tipo 0 ou 1. Para tanto utilizar os comandos “Create”/”New variable” e “Transform

indicator". Este comando requer dois operadores: a variável existente a ser transformada e o valor limite, no caso, cada um dos percentis. Dar um novo nome a essa variável indicativa, onde valores 0 significam valores acima do nível de corte e valores 1 valores abaixo. Gravar o novo arquivo de dados, por exemplo Cd50.dat, e sair.





Abaixo, no formato GEOEAS, a porção inicial da matriz de dados transformados para o nível de corte referente à mediana de cádmio.

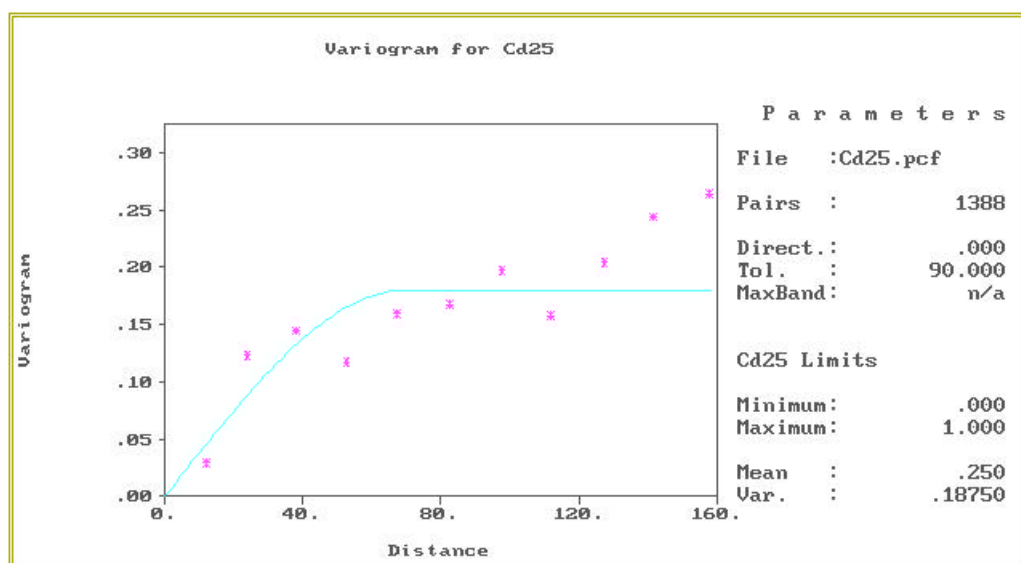
Example.dat: 50th Cd = 7.95

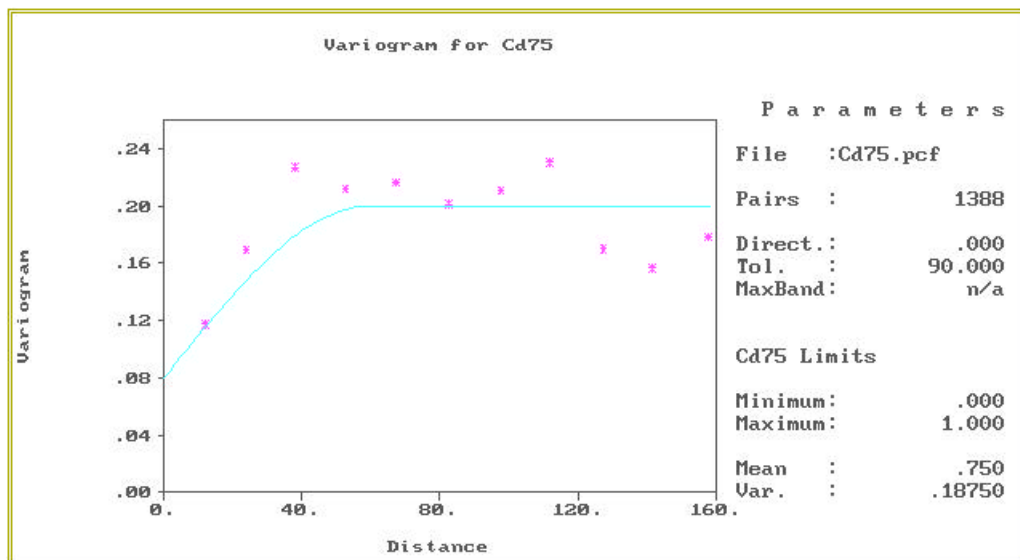
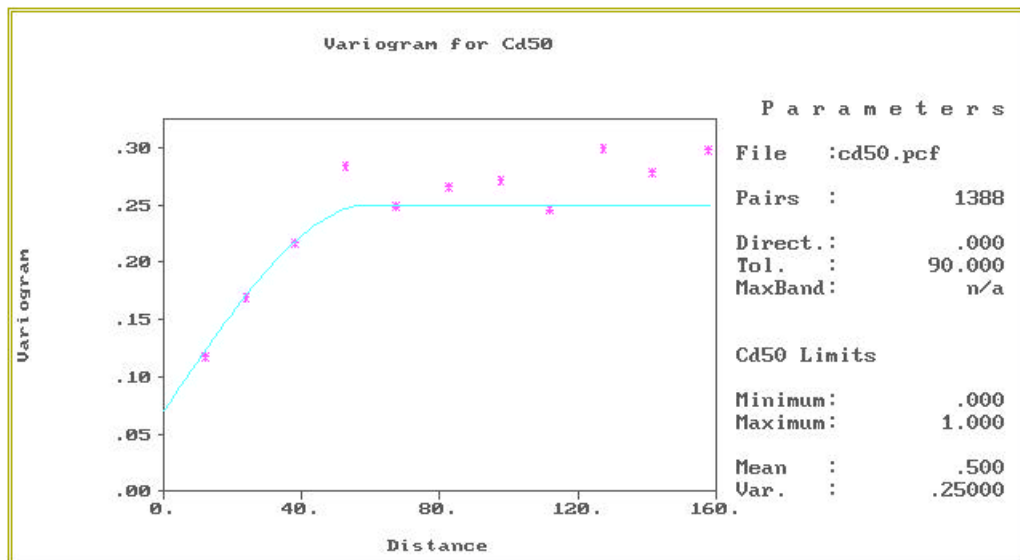
```

4
X
Y
Cadmium
Cd50
288.0 311.0 11.50 0
285.6 288.0 8.50 0
273.6 269.0 7.00 1
280.8 249.0 10.70 0
273.6 231.0 11.20 0
276.0 206.0 11.60 0
285.6 182.0 7.20 1
288.0 164.0 5.70 1
292.8 137.0 5.20 1

```

A partir deste ponto seguem-se as operações normais existentes no GEOEAS para a obtenção do semi-variograma experimental, sua modelagem e gravação do arquivo \*.grd, que contem o reticulado de valores para a impressão do mapa de probabilidades.





	Modelo	Co	C	a
Cd25	esférico	0.0	0.18	70.0
Cd50	esférico	0.07	0.18	60.0
Cd75	esférico	0.08	0.12	60.0

No programa “Krige” é obtido o arquivo \*.grd que contem o reticulado com os valores calculados pelo método da krigagem ordinária. Nesta etapa é de fundamental importância a definição das dimensões desse reticulado, ou seja, quantos pontos ao longo do eixo X e quantos ao longo do eixo Y. Neste exemplo a escolha foi 25 x 20 pontos.

A saída gráfica no GEOEAS, seja pelo programa “Krige”, seja pelo “Conrec” é muito ruim e deve-se, então, procurar uma outra maneira para a impressão final ficar esteticamente mais agradável e de fácil manuseio.

### **Transformação do arquivo \*.grd em \*.dat, a ser utilizado no SURFER**

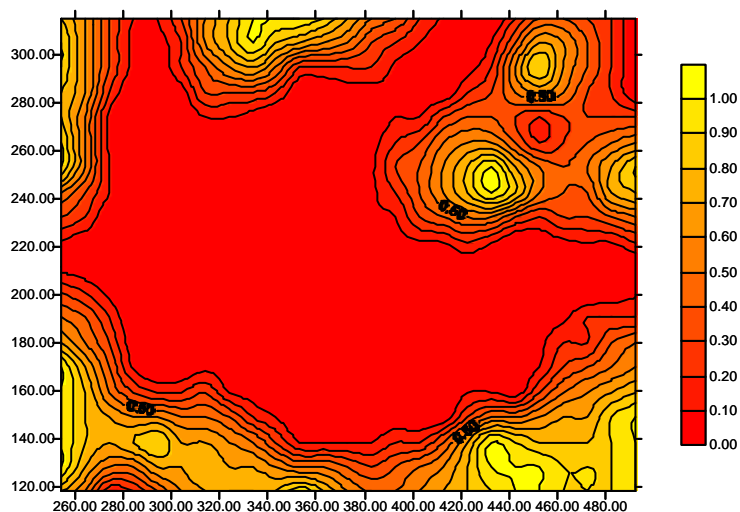
A escolha para a saída dos mapas resultantes foi o pacote SURFER e para tanto deve-se renomear o arquivo \*.grd para \*.dat e ao editar este novo arquivo retirar as primeiras linhas e substituí-las apenas por X, Y e “variável transformada”.

### **Obtenção dos mapas de probabilidades de ocorrência**

Dentro do pacote SURFER, obter os mapas de probabilidades. Neste caso, como o reticulado de dados calculados pelo GEOEAS é regular, deve-se utilizar o método “*Nearest Neighbor* (vizinho mais próximo). A razão está no fato que este método não interpola valores, mas sim honra os valores recebidos já que os nós do reticulado coincidem com a matriz XY de dados. Os resultados para os três níveis de corte estão nas figuras 2, 3 e 4.

As escalas desses mapas são definidas por 0, 0.1, 0.2,...0.8, 0.9 e 1.0, onde zero (0) significa que a probabilidade de ocorrência estar acima do limite definido pelo nível de corte é de 100%, já que inicialmente foi estabelecido que valores acima do nível de corte seriam substituídos pelo valor 0. Assim sendo as cores vermelhas, nos mapas, é que indicam tal situação.

Caso a distribuição dos pontos apresente agrupamentos, o que não acontece no presente caso, sugere-se uma regularização da malha, com dimensões a serem determinadas em função das distâncias existentes, com o auxílio do algoritmo “Inverso do Quadrado da Distância”. Maiores detalhes podem ser encontrados em **Sturaro (1994)**.



**Figura 2. Mapa de probabilidades de ocorrência referente ao nível de corte Cd25 (5.3 ppm)**

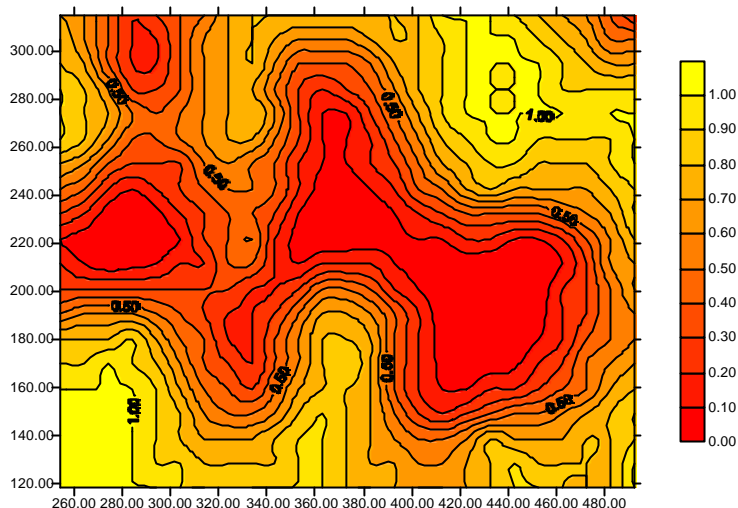


Figura 3. Mapa de probabilidades de ocorrência referente ao nível de corte Cd50 (7.95 ppm)

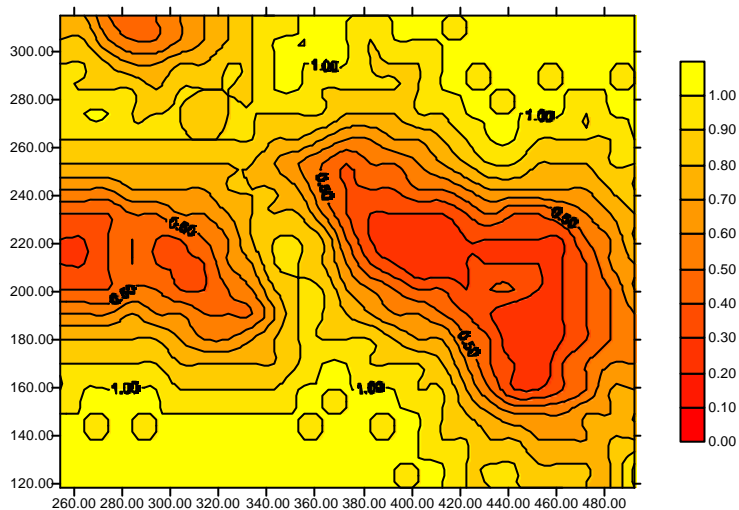


Figura 4 Mapa de probabilidades de ocorrência referente ao nível de corte Cd75 (10.8 ppm)

## KRIGAGEM INDICATIVA PARA DUAS VARIÁVEIS

Sturaro e Landim (1995) propuseram a aplicação da krigagem indicativa para duas variáveis em conjunto e por pressuporem que os dois eventos fossem independentes aplicaram a regra multiplicativa de probabilidades para eventos independentes, ou seja:

$$P(X1 \geq v_c) \times P(X2 \geq v_c) = \text{valor combinado de probabilidades,}$$

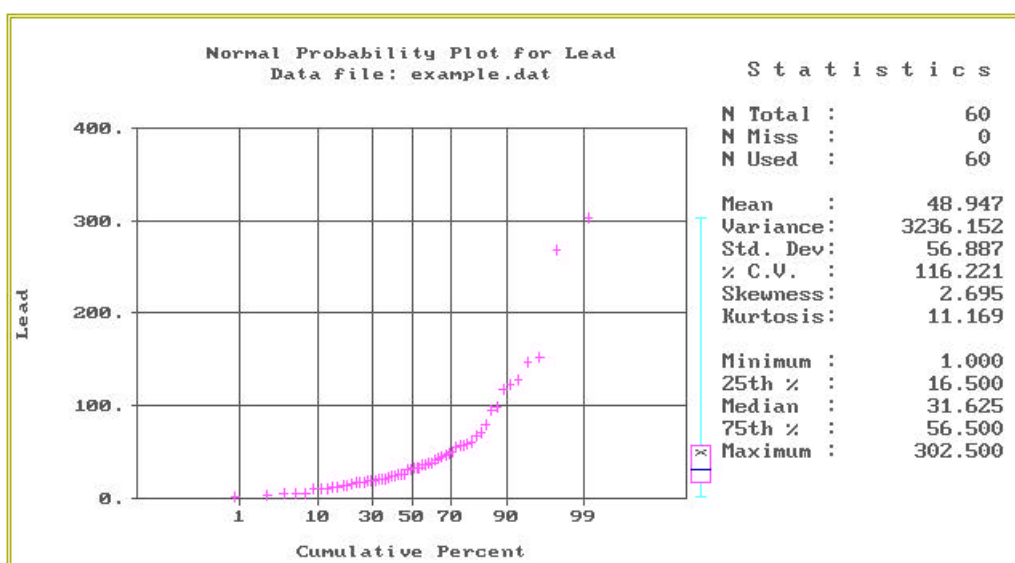
onde  $v_c$ , representa os valores de corte de interesse para cada variável. O resultado fornecido é um mapa combinado mostrando as probabilidades de ocorrência dos dois

eventos simultaneamente. Este tema pode ser encontrado mais facilmente em **Sturaro et al. (2000)**

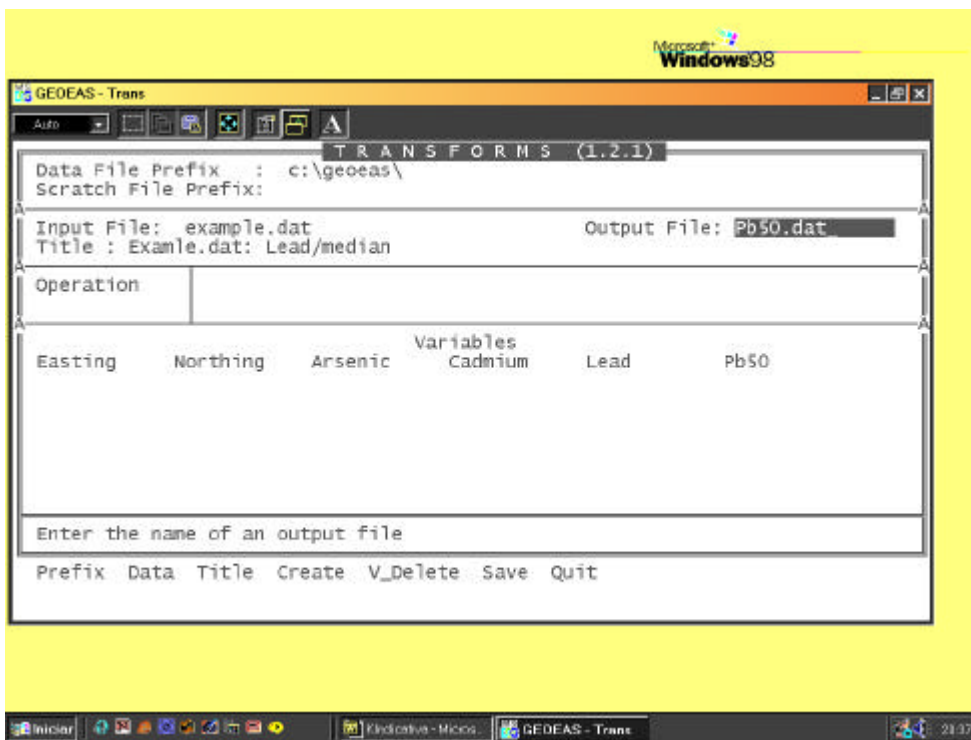
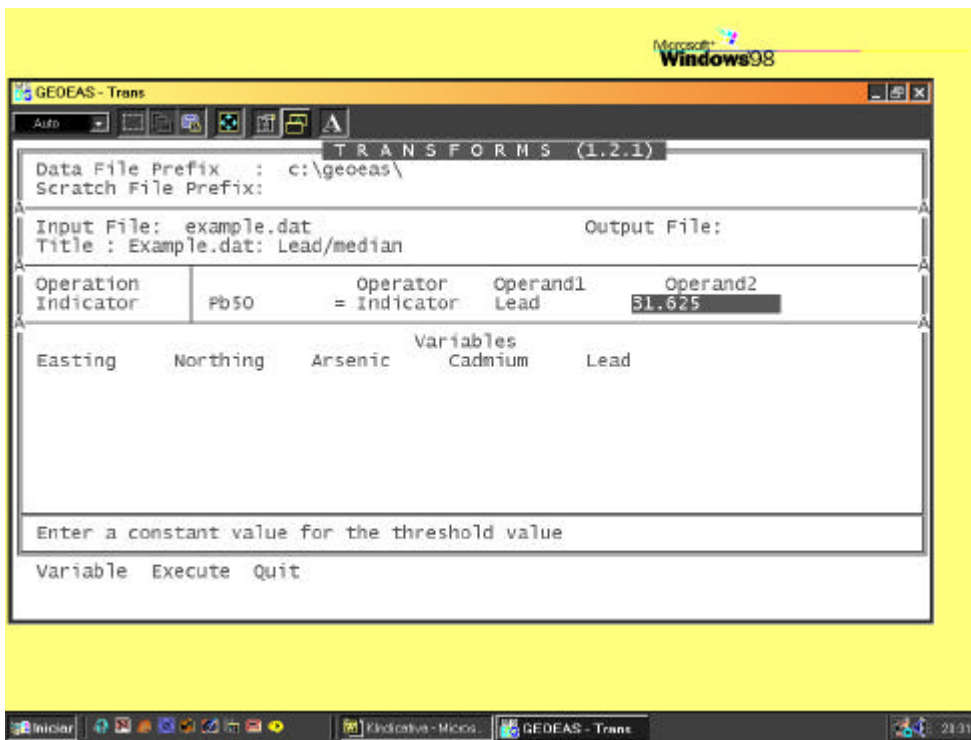
### Exemplo

Foi escolhida como segunda variável o *Chumbo*, com teores também em ppm, que apresenta um coeficiente de correlação com *Cádmio* da ordem de 0.46.

A curva de probabilidades acumuladas mostrou os valores 16.50 ppm, 31.625ppm e 56.50 ppm para o 1º. quartil, mediana e 3º. quartil, respectivamente.



Escolheu-se o valor da mediana para efetuar a transformação binária para o chumbo, que resultou na criação do arquivo PB50.dat.



Abaixo, no formato GEOEAS, a porção inicial da matriz de dados transformados para o nível de corte referente à mediana do chumbo.

Example.dat: Lead median = 31.625

4

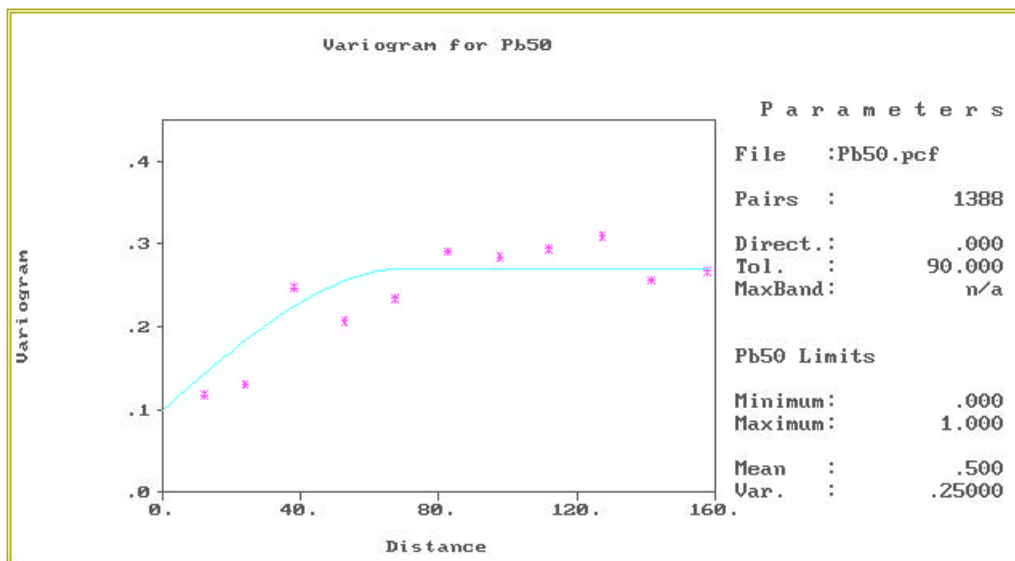
X

Y

Lead

**Pb50**

288.0	311.0	18.25	1
285.6	288.0	30.25	1
273.6	269.0	20.00	1
280.8	249.0	19.25	1
273.6	231.0	151.50	0
276.0	206.0	37.50	0
285.6	182.0	80.00	0
288.0	164.0	46.00	0
292.8	137.0	10.00	1



A aplicação da krigagem indicativa revelou os seguintes valores após a modelagem variográfica:

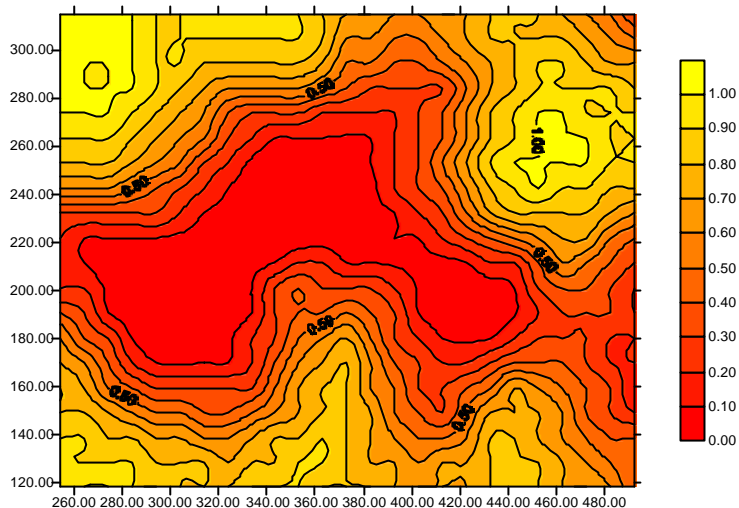
Modelo esférico

Efeito pepita: 0.1

Patamar: 0.17

Alcance: 70

O mapa de probabilidades de ocorrência de valores da mediana de chumbo é apresentado na Figura 5.

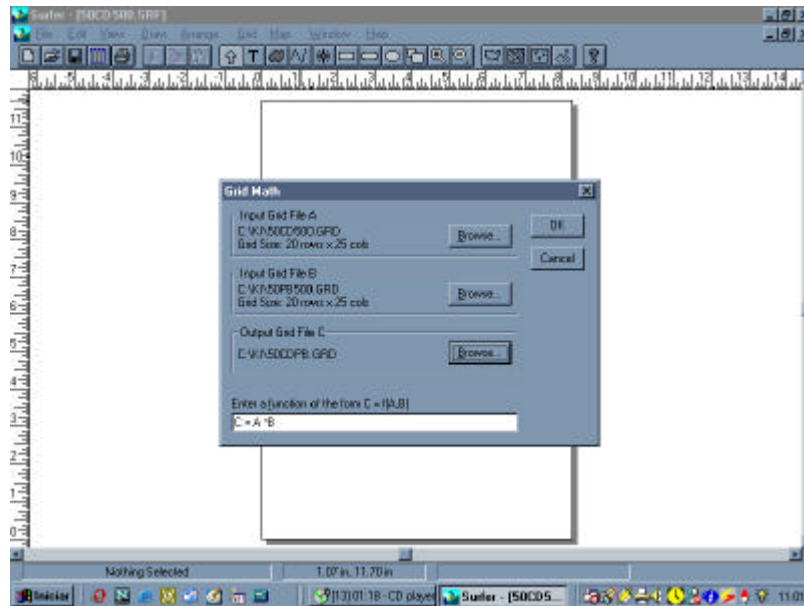


**Figura 5. Mapa de probabilidades de ocorrência referente ao nível de corte Pb50 (31.62 ppm)**

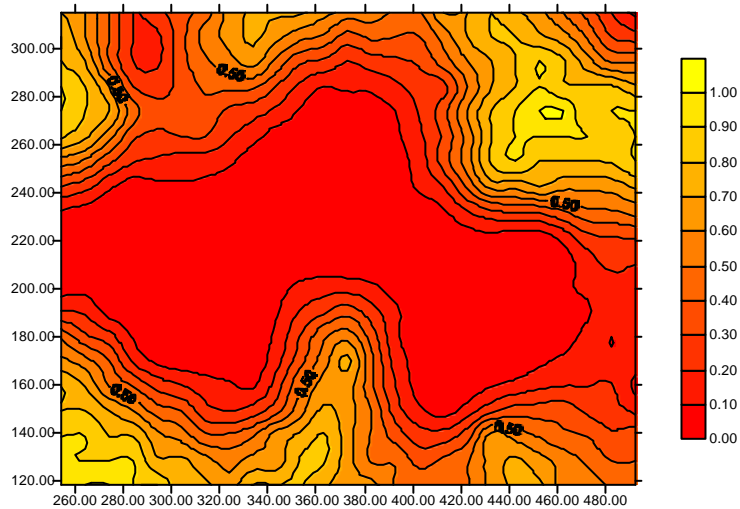
Para a aplicação da krigagem indicativa para duas variáveis, optou-se pelos valores das medianas, ou seja 7.95 ppm de Cd e 31.625 ppm de Pb, para a aplicação da dupla krigagem indicativa. Isso significa que o interesse é obter mapas probabilísticos que indiquem, conjuntamente, áreas com níveis superiores a 7.97 ppm de cádmio e 31.625 ppm de chumbo.

Para a obtenção do mapa combinado de probabilidades utilizar o SURFER. Entrar em "Grid" e, em seguida, em "Math" e atribuir ao "input grid file A" o arquivo que contem os valores de Cd50 e ao "input grid file B" o arquivo que contem os valores de Pb50. Para a função  $C = f(A,B)$  estabelecer  $C = A*B$ , que originará o arquivo CdPb.grd. Observar que as dimensões dos arquivos A e B são as mesmas, isto é, 25 colunas por 20 linhas, como definido para a obtenção do reticulado \*.grd proveniente do GEOEAS.



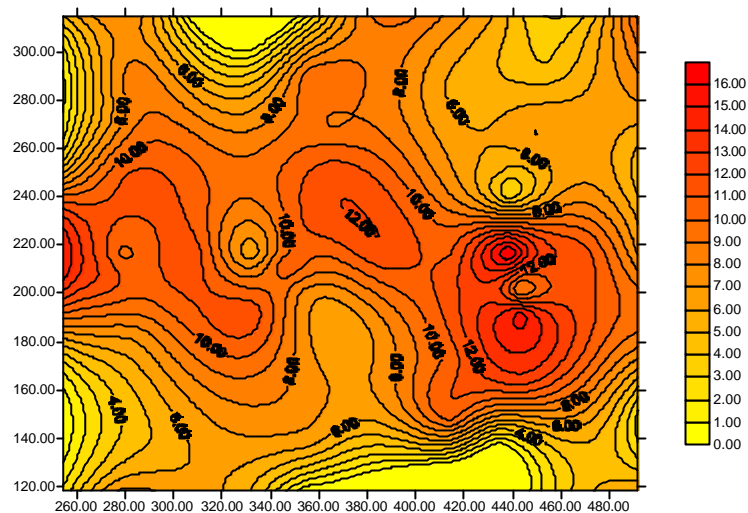


O mapa resultante para a probabilidade conjunta de ocorrências de valores acima de 7.95 ppm de Cd e também de 31.625 de Pb é mostrado na Figura 6.

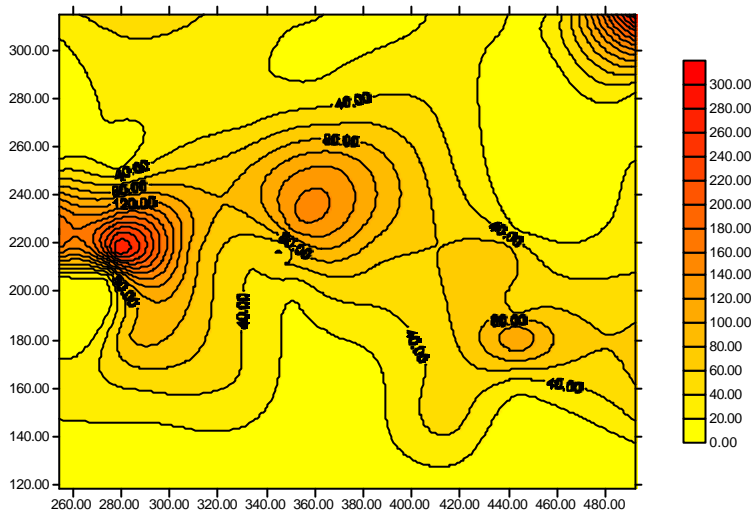


**Figura 6. Mapa de probabilidades de ocorrências combinadas para o nível de corte medianas de Cd (7.95 ppm) e Pb (31.62 ppm)**

Este resultado, que aponta as maiores probabilidades de concentrações conjuntas para Cd e Pb na região central é coerente com a distribuição de isotores desses dois elementos . Ver nas Figuras 7 e 8 os mapas de isotores.



**Figura 7. Mapa de isotores para o elemento cádmio.**



**Figura 8. Mapa de isotores para o elemento chumbo.**

Neste exemplo apenas duas variáveis foram utilizadas, porém a metodologia pode ser aplicada a diversas variáveis combinadas. Tal emprego da krigagem indicativa, com conotação multivariada, consiste em uma alternativa para modelagem simultânea de diversas variáveis com propósitos, por exemplo, de análise ambiental, fornecendo um método viável para estimar incertezas referentes à distribuições de diversas variáveis regionalizadas.

## **SUGESTÕES**

Dúvidas, questões, sugestões, etc. sobre o texto deverão ser encaminhadas para o endereço [plandim@rc.unesp.br](mailto:plandim@rc.unesp.br), as quais serão sempre bem recebidas.

## Referências bibliográficas

DEUTSCH, C.V. & JOURNEL, A.G. (1998) - *GSLIB-Geoestatistical Software Library and User's Guide*. Oxford University Press, 2th. Edition.

ENGLUND, E. & SPARKS, A. (1991) - *GEO-EAS 1.2.1., Geostatistical Environmental Assessment Software. User's Guide*. United States Environmental Protection Agency. Conjunto de programas, disponível em <http://www.sph.umich.edu/~aelon/geoeas/>

JOURNEL, A.G. (1983) - *Non-parametric estimation of spatial distribution*. *Mathematical Geology*, **15**(2):445-468.

STURARO, J.R. (1994) - *Mapeamento geoestatístico de propriedades geológico-geotécnicas obtidas em sondagens de simples reconhecimento*. Tese de Doutorado em Geotecnia, Escola de Engenharia de São Carlos - USP, São Carlos, 183 pp.

STURARO, J. R. & LANDIM, P. M. B.; (1995) - *Geoestatística indicativa aplicada à análise espacial*. VII Simpósio Latinoamericano de Percepção Remota. La Sociedad Latinoamericana de Percepção Remota y Sistemas de Información Espacial, Bol. Resumos Expandidos, 68-73.

STURARO, J. R.; LANDIM, P. M. B. & RIEDEL, P. S. (2000) - *O emprego da técnica geoestatística da krigagem indicativa em Geotecnia Ambiental*. *Revista Solos e Rochas*, **23**(3):157-164

SURFER 6.0 (1995) – *User's guide. Contouring and 3D Surface Mapping for Scientists and Engineers*: Golden Software, Inc.