

MEDIDAS DE INCERTEZAS EM MODELAGEM DE ATRIBUTOS ESPACIAIS

1	Introdução.....	2
2	Aspectos Conceituais	4
2.1	<i>O paradigma da geoestatística para modelagem de atributos espaciais</i>	<i>4</i>
2.2	<i>Determinação da da $F(u; z/(n))$ de uma VA.....</i>	<i>5</i>
2.3	<i>Medidas de incertezas para atributos numéricos.....</i>	<i>7</i>
2.4	<i>Medidas de incertezas para atributos temáticos.....</i>	<i>9</i>
3	Estudo de casos com análises.....	11
3.1	<i>Caracterização da região espacial considerada.....</i>	<i>11</i>
3.2	<i>Modelagem de dados numéricos de altimetria.....</i>	<i>12</i>
3.3	<i>Modelagem de dados temáticos de textura de solo</i>	<i>15</i>
4	Conclusões.....	19
	Referências Bibliográficas	20

MÉTRICAS DE INCERTEZAS EM MODELAGEM DE ATRIBUTOS ESPACIAIS

Carlos Alberto Felgueiras¹
Suzana Druck Fuks²
Antônio Miguel Vieira Monteiro¹

¹ Divisão de Processamento de Imagens (DPI/INPE)

² Centro de Pesquisas Agropecuárias do Cerrado (CPAC/EMBRAPA)

1 Introdução

Atualmente os Sistemas de Informação Geográfica (SIG) vem sendo, cada vez mais, utilizados para modelagem de fenômenos sócio-econômicos-ambientais, simples ou complexos, que envolvem uma ou diversas variáveis relacionadas à geografia terrestre. *Variáveis numéricas*, representando atributos contínuos, e *variáveis temáticas*, representando atributos categóricos, podem ser utilizadas nesses modelos. Modelos de erosão, de fluxo hídrico e de previsão de safras são exemplos típicos de análises complexas realizadas em um SIG. As variáveis espaciais desses modelos matemáticos precisam ser representadas computacionalmente de forma a facilitar a utilização do SIG como ferramenta eficiente para análises.

A modelagem de atributos espaciais consiste na obtenção de modelos computacionais, matemáticos ou numéricos, que representam a variabilidade do atributo dentro de uma região de interesse. Assim, para um determinado atributo, constroem-se representações computacionais individuais que serão posteriormente utilizadas nas análises realizadas no ambiente do SIG. As estruturas matriciais, também conhecidas como grades retangulares regulares, são bastante utilizadas para representar atributos espaciais. Nesta, cada atributo é representado em um nível separado, plano de informação, e cada célula da grade pode estar associado a um valor escalar diferente. O valor de cada ponto da matriz é obtido, em, geral por um processo de interpolação. A interpolação consiste na predição do valor de um atributo, em uma localização espacial não amostrada, a partir de medidas feitas em outras localizações, que se situam dentro de uma vizinhança dada (Burrough, 1998). As medidas, ou amostras, de atributos

espaciais podem ser obtidas a partir de trabalhos em campo, de mapeamentos sistemáticos, de dados de sensoriamento remoto ou outros.

Várias soluções para o problema de modelagem de atributos espaciais, amostrados pontualmente, tem sido apresentadas na literatura, porém com a deficiência principal de não apresentar medidas de incerteza das estimativas. (Obs. Aqui o termo *incerteza* não deve ser confundido com *erro* que pressupõe o conhecimento do valor real do atributo na posição inferida. A incerteza é uma estimativa da magnitude do erro condicionada aos conhecimentos a priori, por exemplo variabilidade espacial e amostras vizinhas, em relação ao atributo).

A informação de incerteza, relacionada com os modelos de inferência adotados, é útil para se avaliar a qualidade do modelo assumido para um determinado atributo. Além disso, a incerteza se propaga para produtos de análises, que envolvem diversas variáveis, qualificando os produtos gerados no ambiente do SIG. Essa incerteza quantifica o risco assumido em tomadas de decisões apoiadas nesses produtos. Heuvelink (1998), apresenta várias metodologias que podem ser utilizadas em propagação de incertezas para atributos numéricos.

Procedimentos geoestatísticos consideram o valor do atributo, em uma posição qualquer do espaço, como uma variável aleatória. O conjunto de valores do atributo na região de interesse é considerado um campo aleatório. Esses procedimentos possibilitam a predição, ou inferência, do valor do atributo, em uma posição qualquer, utilizando-se um conjunto amostral pontual local e informação da variabilidade espacial do atributo na região. O estudo da variabilidade espacial do atributo é feita a priori, utilizando-se todo o conjunto amostral. O variograma, resultado desse estudo, representa a variabilidade do atributo em função da distância e da direção consideradas. Procedimentos de modelagem determinísticos são mais simplistas por não comportarem esse estudo de variabilidade. Nestes, em geral, a variabilidade do atributo com a distancia é definida pelo sentimento do usuário e a direção de variabilidade não é considerada. Isaaks e Srivastava (1989) apresentam um a introdução aos conceitos principais da geoestatística, incluindo aspectos importantes sobre o estudo da variografia dos atributos espaciais.

Procedimentos geoestatísticos por indicação, krigeagem e simulação estocástica, possibilitam a modelagem de atributos espaciais, numéricos e temáticos, e também a estimação de incertezas associadas aos modelos gerados. Atualmente há uma forte tendência de se integrar o SIG com a geoestatística, incorporando procedimentos geoestatísticos no SIG (Camargo, 1997; Felgueiras, 1999a e Felgueiras et al., 1999b). O SIG SPRING (SPRING V 3.5, 2001), por exemplo, contém procedimentos geoestatísticos lineares e por indicação nos seu módulo de análise espacial.

Este artigo descreve, exemplifica e analisa *métricas de incertezas* aplicáveis para qualificar representações de atributos espaciais. Essas métricas pressupõem o uso dos procedimentos geoestatísticos por indicação no processo de predição. Para atributos numéricos são exploradas medidas de incerteza definidas por intervalos de confiança, baseadas em desvios padrão e quantis, e por probabilidades de pertinência a intervalos. Para atributos temáticos são apresentadas medidas de incerteza baseadas na moda, valor de máxima probabilidade, e na entropia de Shannon.

Quanto a organização deste texto, a seção 2 apresenta aspectos conceituais relacionados com a geoestatística por indicação e com as métricas de incerteza aqui adotadas. A seção 3 mostra exemplos de modelagens de atributos espaciais utilizando-se de amostras de atributo numérico, dados de elevação, e amostras de atributo temático, dados de textura de solo,. Nessa seção são apresentadas, ainda, análises dos mapas gerados dando-se ênfase às informações de incerteza. Na seção 4 algumas conclusões consideradas relevantes são reportadas.

2 Aspectos Conceituais

2.1 O paradigma da geoestatística para modelagem de atributos espaciais

A geoestatística modela os valores de um atributo espacial, dentro de uma região $A \subset \mathfrak{R}^2$ da superfície terrestre, como uma função aleatória. Para cada posição $\mathbf{u} \in A$ o valor do atributo, de um dado espacial, é modelado como uma *variável aleatória* (VA) $Z(\mathbf{u})$. Isto significa que, na posição \mathbf{u} , a VA $Z(\mathbf{u})$ pode assumir diferentes valores desse atributo, cada valor com uma probabilidade de ocorrência associada. Nas n posições amostradas, \mathbf{u}_α , $\alpha=1,2,\dots,n$, os valores $z(\mathbf{u}_\alpha)$ são considerados determinísticos, ou seja, podem ser tomados como VA's cujo valor observado tem uma probabilidade de 100%

de ocorrência. A função de distribuição de $Z(\mathbf{u})$ condicionada aos dados amostrado, $F(\mathbf{u}; z|(\mathbf{n}))$, é definida por:

$$F(\mathbf{u}; z|(\mathbf{n})) = Prob\{Z(\mathbf{u}) \leq z|(\mathbf{n})\} \text{ quando o atributo é numérico} \quad (2.1)$$

e

$$F(\mathbf{u}; z|(\mathbf{n})) = Prob\{Z(\mathbf{u}) = z|(\mathbf{n})\} \text{ quando o atributo é temático} \quad (2.2)$$

A $F(\mathbf{u}; z|(\mathbf{n}))$ modela a incerteza sobre valores $z(\mathbf{u})$, em posições \mathbf{u} não amostradas, considerando-se as \mathbf{n} amostras (Deutsch e Journel, 1998). Esta propriedade é a base dos procedimentos de estimativa de incerteza apresentados neste artigo. Assim, obtendo-se a $F(\mathbf{u}; z|(\mathbf{n}))$, ou uma aproximação da mesma, pode-se derivar diferentes estimadores e incertezas para valores não amostrados $z(\mathbf{u})$. Esse é o tema da seção que segue.

2.2 Determinação da $F(\mathbf{u}; z|(\mathbf{n}))$ de uma VA

Procedimentos geoestatísticos conhecidos por *krigeagem por indicação* e *simulação estocástica por indicação* possibilitam a estimativa de uma aproximação para a $F(\mathbf{u}; z|(\mathbf{n}))$, que representa um atributo espacial de interesse. Um aspecto importante a se destacar consiste no fato de que estes procedimentos são *não paramétricos*. Isto significa que não se considera, a priori, nenhum modelo paramétrico de distribuição relacionado com a $F(\mathbf{u}; z|(\mathbf{n}))$ a ser estimada. A idéia básica consiste na obtenção de valores estimados $F^*(\mathbf{u}; z_k|(\mathbf{n}))$, da $F(\mathbf{u}; z|(\mathbf{n}))$ real da VA.

Para se obter um conjunto de pontos de controle da $F(\mathbf{u}; z|(\mathbf{n}))$, os procedimentos geoestatísticos por indicação requerem uma transformação, a priori, sobre os dados amostrais do atributo. Esta transformação é conhecida por *codificação por indicação*. A codificação por indicação da VA $Z(\mathbf{u}=\mathbf{u}_\alpha)$, em uma localização \mathbf{u}_α e para um *valor de atributo, chamado valor de corte, $z = z_k$* , gera a VA $I(\mathbf{u}=\mathbf{u}_\alpha; z_k)$ segundo a seguinte função de mapeamento não linear:

$$I(\mathbf{u}_a; z_k) = \begin{cases} 1, & \text{se } Z(\mathbf{u}_a) \leq z_k \\ 0, & \text{se } Z(\mathbf{u}_a) > z_k \end{cases} \quad \text{para atributos numéricos e,} \quad (2.3)$$

$$I(\mathbf{u}_a; z_k) = \begin{cases} 1, & \text{se } Z(\mathbf{u}_a) = z_k \\ 0, & \text{se } Z(\mathbf{u}_a) \neq z_k \end{cases} \quad \text{para atributos temáticos} \quad (2.4)$$

A esperança condicional da VA por indicação $I(\mathbf{u}; z_k)$ é calculada por:

$$\begin{aligned} E\{I(\mathbf{u}; z_k) / (\mathbf{n})\} &= 1 \cdot \text{Prob}\{I(\mathbf{u}; z_k) = 1 / (\mathbf{n})\} + 0 \cdot \text{Prob}\{I(\mathbf{u}; z_k) = 0 / (\mathbf{n})\} \\ &= 1 \cdot \text{Prob}\{I(\mathbf{u}; z_k) = 1 / (\mathbf{n})\} = F^*(\mathbf{u}; z_k / (\mathbf{n})) \end{aligned} \quad (2.5)$$

A equação acima apresenta um resultado muito importante para a inferência da distribuição de probabilidade de uma variável aleatória: “A esperança condicional de $I(\mathbf{u}; z_k)$ fornece, para o valor de corte $z = z_k$, uma estimativa do valor da função de distribuição condicionada (fdc) de $Z(\mathbf{u})$ no caso de atributos temáticos e uma estimativa da função de distribuição acumulada condicionada (fdac) para atributos numéricos“. Sob a hipótese de estacionariedade e da existência do variograma para o conjunto amostral por indicação, a esperança condicional da equação 2.5 pode ser obtida pela aplicação de um interpolador de krigeagem linear, krigeagem ordinária por exemplo, sobre esse conjunto por indicação. Este interpolador fornecerá o melhor estimador linear, não tendencioso, para a $F(\mathbf{u}; z | (\mathbf{n}))$ em $z = z_k$ (Oliver e Webster, 1990; Isaaks e Srivastava, 1989; Deutsch e Journel, 1998).

Utilizando-se K valores de corte, $z = z_k$. $k=1,2,\dots,K$, pode-se obter uma aproximação discretizada $F^*(\mathbf{u}; z | (\mathbf{n}))$, da $F(\mathbf{u}; z | (\mathbf{n}))$ real de $Z(\mathbf{u})$, a partir de um conjunto de estimações, $F^*(\mathbf{u}; z_k | (\mathbf{n}))$. A $F^*(\mathbf{u}; z | (\mathbf{n}))$ estimada é, então, utilizada para se representar o comportamento local de cada variável aleatória em estudo. Dessa representação obtém-se parâmetros estatísticos que possibilitam a inferência de valores típicos e incertezas relacionados à VA.

No caso de atributos temáticos, os valores de corte são os valores definidos para as classes consideradas para o atributo. Assim, uma aproximação da fdc é totalmente definida após a aplicação da equação (2.5) para todas as classes. A partir dessa aproximação obtém-se parâmetros estatísticos, para a VA categórica, do tipo: valores de mínima e máxima probabilidade e valor da moda (ou valor mais presente).

Para atributos numéricos, os valores de corte são definidos em função do número de amostras. É necessário que a quantidade de amostras codificadas com valor 1 seja suficiente para se definir, com sucesso, um modelo de variografia para cada valor de corte (Journel, 1983). Os valores estimados $F^*(\mathbf{u}; z_k | (\mathbf{n}))$, para um conjunto de valores de

corte z_k , são utilizados como pontos de controle para o ajuste de uma curva bidimensional. Essa curva representa uma função $F^*(\mathbf{u}; z|(n))$ que aproxima a fdac da VA. Desta aproximação é possível obter-se parâmetros estatísticos da VA numérica, tais como: média, mediana, desvio padrão, quantis, etc....

Além das estimativas para cada valor de corte, a $F^*(\mathbf{u}; z|(n))$, para atributos numéricos e temáticos, requer a utilização de procedimentos de *correção de relações de ordem* que garantem que essa função pode ser usada como uma representação da função de distribuição de probabilidade da VA $Z(\mathbf{u})$. Deutsch e Journel (1998) abordam com mais detalhes esses procedimentos.

Para finalizar, a aproximação $F^*(\mathbf{u}; z|(n))$, estimada e corrigida, é utilizada para se obter medidas de incertezas associadas à VA $Z(\mathbf{u})$. As duas próximas seções apresentam algumas dessas medidas.

2.3 Medidas de incertezas para atributos numéricos

Como já observado anteriormente, estimativas das fdacs de variáveis aleatórias numéricas são obtidas por um conjunto de probabilidades acumuladas univariadas $p_k(\mathbf{u})$, $k=1, \dots, K$, em K valores de corte pré-definidos. Essas probabilidades são utilizadas para inferência do valor e da incerteza da VA na localização \mathbf{u} não amostrada.

Quando a natureza do atributo espacial é numérica, é comum expressar-se incertezas em função de intervalos de confiança. Um *intervalo de confiança* descreve um intervalo no qual um valor estimado, por exemplo a média, estará incluído com um nível específico de certeza, por exemplo 90%, (Myers 1997). O intervalo de confiança consiste de um valor mínimo e máximo em conjunto com a probabilidade de um valor, não conhecido, estar dentro desse intervalo.

Quando a fdac apresenta um alto grau de similaridade e pode-se supor a hipótese de normalidade para a distribuição, a incerteza pode ser especificada por intervalos de confiança, centrados no valor médio estimado, $\mathbf{m}_z(\mathbf{u})$, do tipo:

$$\begin{aligned}
& Prob\{Z(\mathbf{u}) \in [\hat{\mathbf{i}}_Z(\mathbf{u}) \pm \mathbf{s}(\mathbf{u})]\} \cong 0.68 \\
& \text{ou} \\
& Prob\{Z(\mathbf{u}) \in [\hat{\mathbf{i}}_Z(\mathbf{u}) \pm 2\mathbf{s}(\mathbf{u})]\} \cong 0.95
\end{aligned} \tag{2.6}$$

onde $\mathbf{s}^2 = E\{(Z(\mathbf{u}) - E\{Z(\mathbf{u})\})^2\}$ é a variância da distribuição e o valor médio \mathbf{m} obtido a partir de $F^*(\mathbf{u}; z|(n))$, é obtido pela seguinte formulação:

$$\hat{\mathbf{i}}_Z(\mathbf{u}) = \int_{-\infty}^{\infty} z \cdot dF(\mathbf{u}; z|(n)) \approx \sum_{k=1}^{K+1} z'_k [F^*(\mathbf{u}; z_k|(n)) - F^*(\mathbf{u}; z_{k-1}|(n))] \tag{2.7}$$

onde: os valores das $F^*(\mathbf{u}; z_k|(n))$, $k=1,2,\dots,K$, são os valores estimados das fdc's acumuladas para cada valor z_k do atributo, $z_0 = z_{min}$, $z_{K+1} = z_{max}$, $z'_k = (z_k + z_{k-1})/2$, $F(\mathbf{u}; z_0|(n)) = 0$ e $F(\mathbf{u}; z_{K+1}|(n)) = 1$.

Para distribuições altamente assimétricas, uma medida mais robusta é o intervalo interquantil, que é definido como a diferença entre dois quantis, simétricos em relação a mediana. A partir da função de distribuição acumulada condicionada inferida, $F^*(\mathbf{u}; z|(n))$, pode-se derivar vários intervalos de probabilidade tais como o intervalo 95%, $[q_{0.025}; q_{0.975}]$, tal que:

$$Prob\{Z(\mathbf{u}) \in [q_{0.025}; q_{0.975}]|(n)\} = 0.95 \tag{2.8}$$

com $q_{0.025}$ e $q_{0.975}$ sendo os quantis .025 e .975, respectivamente, da fdac, ou seja, $q_{0.025}$ é tal que $F^*(\mathbf{u}; q_{0.025}|(n)) = 0.025$ e $q_{0.975}$ é tal que $F^*(\mathbf{u}; q_{0.975}|(n)) = 0.975$. Os valores do atributo, referentes aos quantis, são estimados a partir da função de ajuste e dos valores de corte usados na krigeagem por indicação.

A incerteza pode ser estimada, também, para intervalos de valores do atributo. A probabilidade de um valor $z(\mathbf{u})$ estar dentro de um intervalo $(a, b]$ qualquer, chamado de *probabilidade do intervalo*, é computado como a diferença entre os valores da fdac para os limiares b e a , ou seja:

$$Prob\{Z(\mathbf{u}) \in (a, b]|(n)\} = F(\mathbf{u}; b|(n)) - F(\mathbf{u}; a|(n)) \tag{2.9}$$

Uma probabilidade do intervalo de 70%, ou .7, significa que $z(\mathbf{u})$ tem 70% de probabilidade de estar dentro e, portanto, 30% de estar fora do intervalo (a, b]. Estas incertezas são úteis para qualificar mapas temáticos obtidos por fatiamento, classificação por faixas, de atributos numéricos.

Quando $b = +\infty$ obtêm-se a probabilidade de se exceder um limiar a , ou seja:

$$Prob\{Z(\mathbf{u}) \in (a, +\infty] | (n)\} = Prob\{Z(\mathbf{u}) > a | (\mathbf{n})\} = 1 - F(\mathbf{u}; a | (n)) \quad (2.10)$$

Esta probabilidade é particularmente importante em aplicações ambientais focadas em riscos de se exceder limites regulatórios (Goovaerts, 1997).

2.4 Medidas de incertezas para atributos temáticos

Como já explanado acima, uma aproximação $F^*(\mathbf{u}; z | (n))$ da fdc de $Z(\mathbf{u})$, para variáveis aleatórias temáticas, é obtida pela estimação das probabilidades univariadas $p_j(\mathbf{u})$, $j=1, \dots, L$, das L classes definidas no domínio da VA. Estas probabilidades, após uma correção das relações de ordem, são utilizadas para inferência do valor, ou da classe, e da incerteza da VA numa localização qualquer \mathbf{u} não amostrada. O valor estimado para a VA temática é, em geral, a classe de maior probabilidade. A incerteza em \mathbf{u} , $Inc(\mathbf{u})$, pode ser determinada pelo valor de máxima probabilidade, *incerteza por moda*, ou por uma combinação dos valores de probabilidades das L classes, *incerteza por entropia*.

Dado o conjunto de probabilidades, $p_j(\mathbf{u})$ com $j=1, \dots, L$, a *incerteza por moda* é definida por:

$$Inc(\mathbf{u}) = 1 - p_{jmax}(\mathbf{u}) \quad (2.11)$$

onde $p_{jmax}(\mathbf{u})$ é a probabilidade da classe mais provável da função de distribuição de probabilidade em \mathbf{u} , ou seja, $p_{jmax}(\mathbf{u}) = Max(p_j(\mathbf{u})) \quad j=1, \dots, L$,

A *incerteza por entropia* é calculada a partir da medida de entropia, H , proposta por Shannon, 1948. Informalmente, a entropia pode ser entendida com uma medida relacionada a organização espacial de um atributo. Ela mede a confusão, ou a desordem, relacionada aos valores, ou possíveis estados, associados a um atributo. Quando a VA é

temática, ou resultado da discretização de uma VA numérica, (Journel, 1993), a entropia de Shannon, de sua distribuição univariada, tem a seguinte formulação:

$$Inc(\mathbf{u}) = H = -\sum_{j=1}^L p_j(\mathbf{u}) \ln(p_j(\mathbf{u})) \quad (2.12)$$

Embora outras medidas possam ser derivadas, as incertezas por moda e por entropia são as duas medidas de incerteza, que utilizam a fdc aproximada de VAs temáticas, mais usadas na prática.

Para completar, é importante salientar que as medidas de incerteza podem refinar os mapas de inferência de atributos temáticos, ou seja, a informação de incerteza pode ser usada como restrição para se classificar um dado ponto do mapa de inferências. Por exemplo, pode-se gerar um mapa de classes de vegetação restrito a um valor de incerteza menor que 30%. Isto significa que apenas as localizações, do mapa de inferências, com risco menor que 30% serão associados a uma classe de vegetação enquanto as demais serão marcadas como não classificadas. Exemplos deste tipo de classificação, com dados reais de textura do solo, serão apresentados na seção 3 que segue.

3 Estudo de casos com análises

Nesta seção serão apresentados exemplos da aplicação dos procedimentos geostatísticos por indicação, krigagem e simulação estocástica, para modelagem de atributos espaciais a partir de amostras pontuais. Análises acompanharão os resultados mostrados dando-se ênfase para as medidas de incertezas relacionadas com as estimativas obtidas. Dados amostrais de elevação serão usados como exemplo para modelagem de atributos numéricos. A modelagem de atributos temáticos será exemplificada com dados amostrais de classes de textura do solo.

3.1 Caracterização da região espacial considerada

A área considerada pertence à fazenda Canchim, base física do Centro de Pesquisa Pecuária do Sudeste (SPPSE) da Empresa Brasileira de Pesquisa Agropecuária, localizada no município de São Carlos, estado de São Paulo, Brasil. Esta região cobre uma área de 2660 hectares, entre as coordenadas $21^{\circ}55'00''$ à $21^{\circ}59'00''$, latitudes sul, e $47^{\circ}48'00''$ à $47^{\circ}52'00''$, longitudes oeste. A Figura 3.1 ilustra a localização da área.

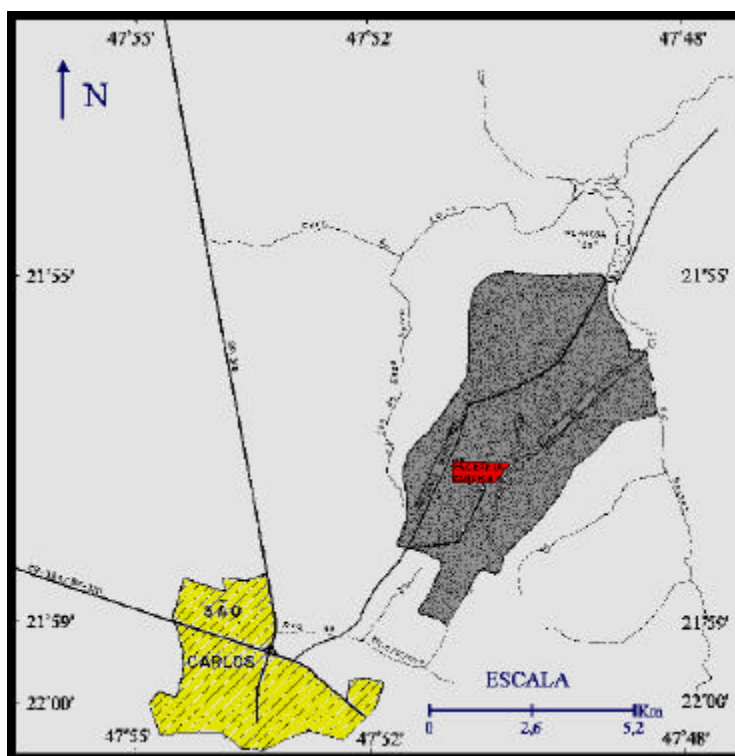


Figura 3.1 - Localização da Fazenda Canchim.
FONTE: Calderano Filho et al. (1996, p. 4).

3.2 Modelagem de dados numéricos de elevação do terreno

Para exemplificar a modelagem de atributos numéricos com medidas de incerteza utilizou-se um conjunto amostral de altimetria, observada na região de Canchim, com distribuição espacial apresentada na Figura 3.2. Nessa figura os dados amostrais estão superpostos a uma imagem em níveis de cinza que representa, de uma forma simplista, a variabilidade do atributo na região considerada. Essa imagem foi obtida por um interpolador de vizinhança mais próxima que associa a cada ponto da imagem um valor proporcional ao valor da amostra mais próxima a esse ponto.

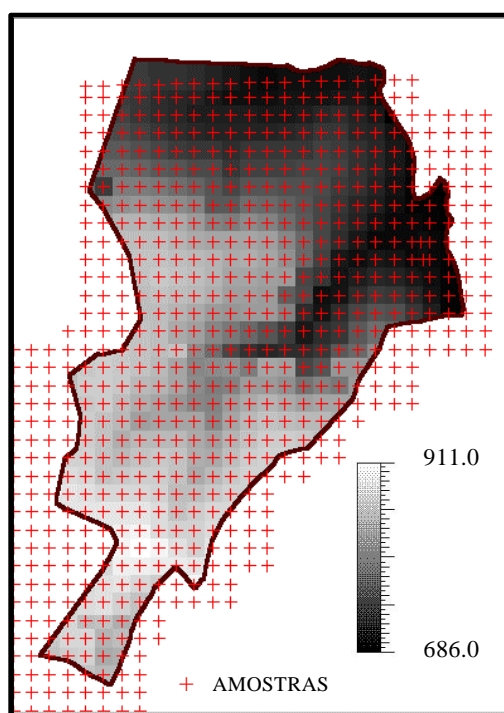


Fig. 3.2: Distribuição de amostras de elevação observadas na região de Canchim.

Para obtenção da aproximação da fdac, inicialmente, o conjunto amostral de altimetria foi dividido em 10 subconjuntos, correspondentes aos decis, a partir da definição de 9 valores de corte. Os subconjuntos foram transformados por indicação e, a partir de análises exploratórias, seus modelos de variabilidade foram determinados. Para esse fim utilizou-se os procedimentos de geração e ajuste de variogramas do módulo de geoestatística do SPRING. Na sequência, os modelos de variografia obtidos foram utilizados como parâmetros de entrada dos procedimentos de krigeagem e de simulação por indicação do SPRING.

Os mapas (imagens) da Figura 3.3 ilustram os resultados da aplicação do procedimento de simulação por indicação, condicionado ao conjunto amostral da Figura 3.2, para espacializar a informação de elevação e gerar medidas locais de incertezas. O mapa da esquerda representa a variabilidade do atributo elevação. Comparando-se esse mapa com a imagem da Figura 3.2 observa-se que ambos apresentam uma conformação global parecida, tendência global semelhante, com regiões de elevação altas e baixas bem definidas. A representação da Figura 3.3, porém, apresenta um comportamento mais suave nas transições entre regiões de altimetria distintas. Isto é uma característica típica dos interpoladores de krigagem por basearem-se na minimização da variância do erro de estimação.

O mapa da direita da Figura 3.3 apresenta uma imagem de incertezas locais. As incertezas foram estimadas pela metodologia baseada nos parâmetros estatísticos de média e de desvio padrão apresentada na seção 2.3. Esse mapa mostra que as regiões de maior incerteza, aquelas que aparecem mais brancas no mapa de incerteza, estão localizadas em áreas de maior variabilidade do atributo.

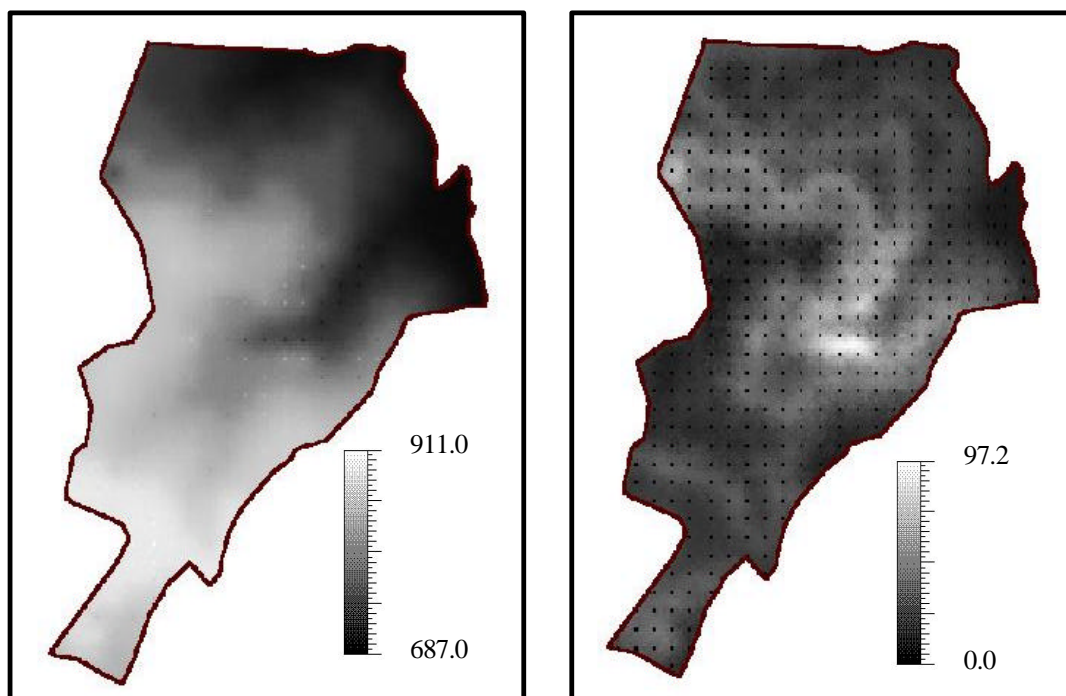


Figura 3.3: Mapas de valores estimados de elevação (a esquerda) e de incertezas (a direita) baseados nos parâmetros estatísticos de média e de 1 desvio padrão.

Os mapas da Figura 3.4 são equivalentes aos mapas da Figura 3.3 considerando-se os parâmetros estatísticos de mediana, para inferência dos valores de elevação, e de intervalos interquartil, como medidas locais de incertezas. Comparando-se qualitativamente as imagens de valores de elevação inferidos, mostrados nessas figuras, observa-se que os mesmos tem tendências globais semelhantes mas apresentam diferenças locais bem determinadas. Portanto pode-se concluir, neste caso, que o atributo elevação na região de Canchim tem distribuições locais não simétricas. Assim a hipótese de normalidade deve ser descartada. Os mapas da Figura 3.4 são mais representativos para o atributo altimetria na região de Canchim.

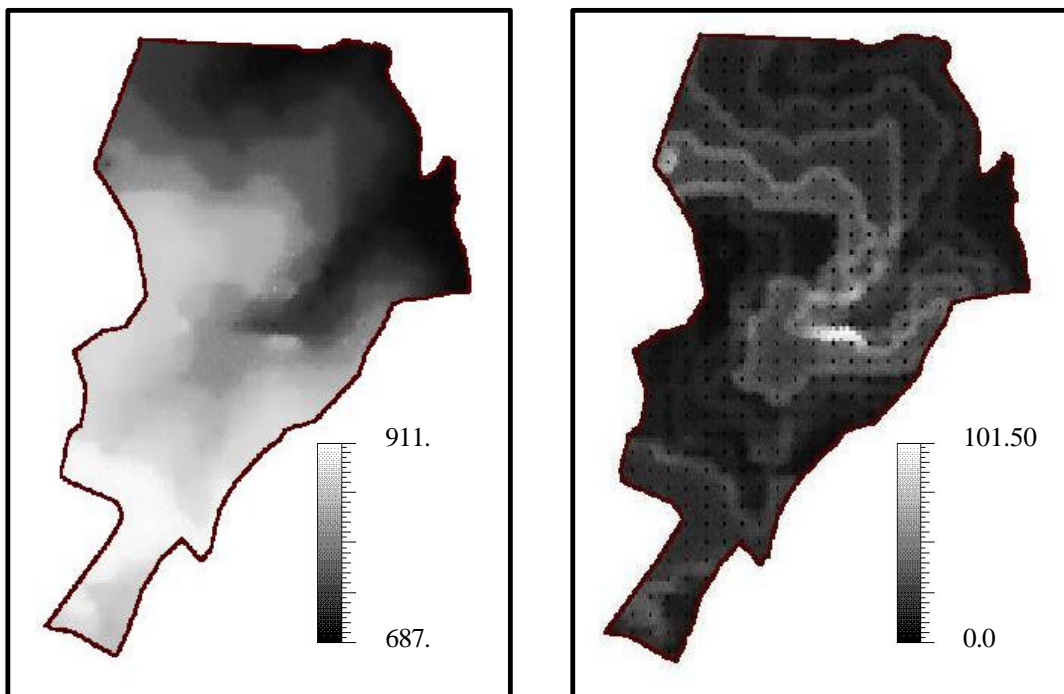


Figura 3.4: Mapas de valores estimados de elevação (a esquerda) e de incertezas (a direita) baseados nos parâmetros estatísticos de mediana e de intervalo interquartil.

A Figura 3.5 apresenta uma alternativa à representação dos dados de altimetria para quando se quer trabalhar com classes, ou intervalos, de valores de elevação. Neste caso utilizou-se o resultado das probabilidades de pertinência a intervalos de elevação para se atribuir uma classe e uma incerteza a cada posição espacial da grade. O resultado apresentado no mapa classificado, mapa esquerdo, mostra coerência com os valores inferidos por média e mediana. Elevações baixas, médias e altas foram agrupadas gerando as regiões classificadas segundo as classes <744.1 , entre 744.1 e

841.1, incluindo-se os limites, e > 841.1 respectivamente. As incertezas por moda, mostradas no mapa da direita, aparecem com valores mais altos exatamente nas regiões de transição entre as classes. Nestas regiões a classificação é mais incerta pois as probabilidades máxima de pertinência a duas ou mais classes tendem a se igualar.

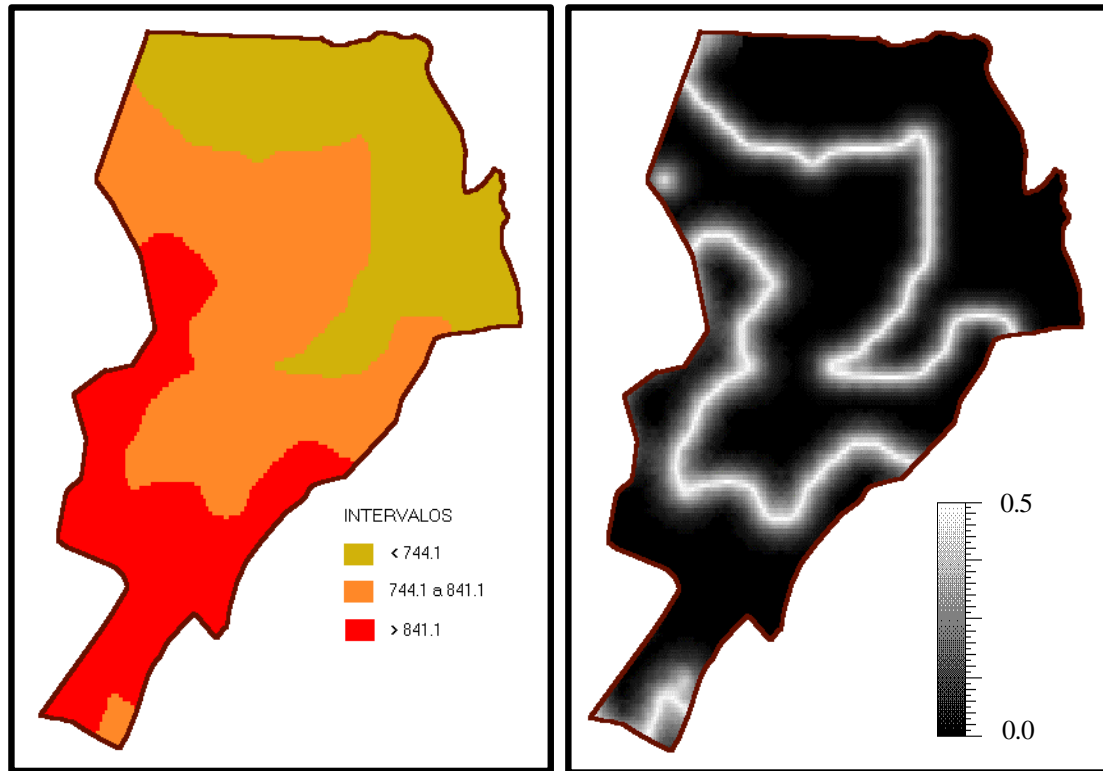


Figura 3.5: Mapa de classes de altimetria (a esquerda), inferido por probabilidades do intervalo, com respectivo mapa de incertezas por moda (a direita).

3.3 Modelagem de dados temáticos de textura de solo

Como exemplo de modelagem de atributos temáticos com medidas de incertezas utilizou-se um conjunto amostral pontual de textura do solo. A Figura 3.6 mostra a distribuição espacial, dentro da região de Canchim, dessas amostras. Nessa figura os dados amostrais estão superpostos a uma imagem colorida que representa uma partição do espaço segundo regiões de influência de cada amostra. Pode-se considerar que essa imagem foi gerada por um classificador de vizinhança mais próxima, que atribui a cada ponto da grade a classe de textura de solo da amostra mais próxima. Assim, a partir da Figura 3.6 pode-se, facilmente, identificar a classe de textura do solo de cada amostra.

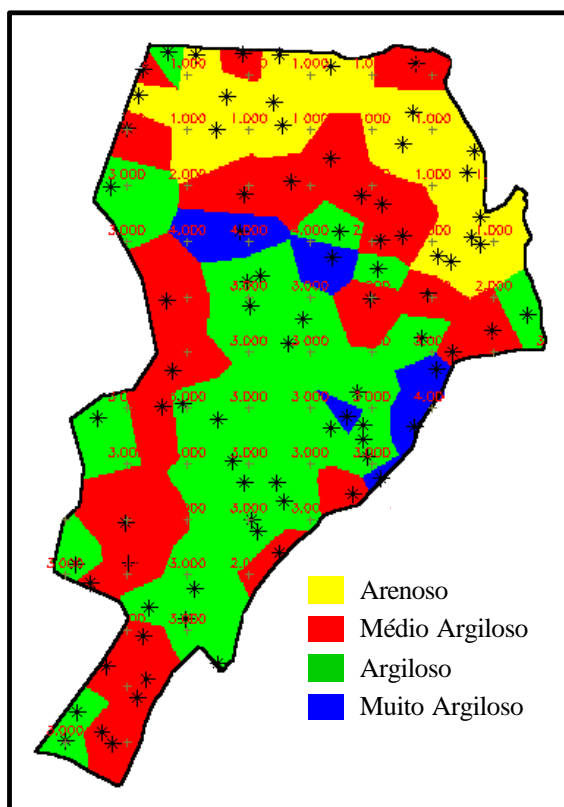


Figura 3.6 - Distribuição das amostras de textura de solo na região de Cachim.

Para obtenção da aproximação da fdc, inicialmente, o conjunto amostral de textura de solo foi dividido em 4 subconjuntos a partir da definição de 4 valores de corte, 1 para cada classe. Os subconjuntos foram transformados por indicação e, a partir dos procedimentos de geração e ajuste de variogramas do módulo de geoestatística do SPRING, seus modelos de variabilidade foram determinados. Na sequência, os modelos de variografia obtidos foram utilizados como parâmetros de entrada dos procedimentos de krigeagem e de simulação por indicação do SPRING.

Os mapas (imagens) da Figura 3.7 ilustram os resultados da aplicação do procedimento de krigeagem por indicação, condicionado ao conjunto amostral da Figura 3.6, para espacializar a informação de textura do solo e gerar medidas locais de incertezas. Neste caso, limitou-se a 16 o número máximo de pontos condicionantes da inferência local. O valor da classe de cada ponto e a incerteza da inferência foram estimados considerando-se a máxima probabilidade de ocorrência, valor da moda, entre as classes para esse ponto. Numa análise qualitativa observa-se que o mapa de valores

estimados apresenta coerência com a distribuição das amostras de solo e, como esperado, as incertezas locais por moda aparecem nas regiões de transição entre classes.

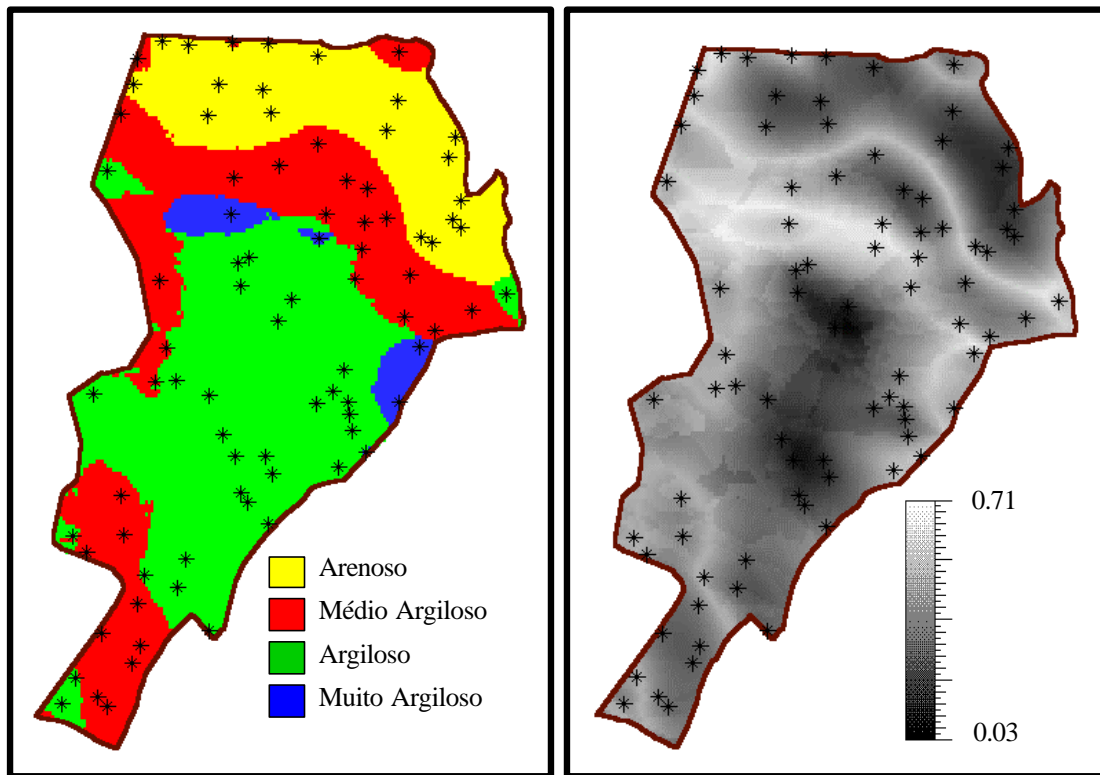


Figura 3.7 - Mapas de valores estimados de textura do solo (a esquerda) e incertezas associadas (a direita) considerando-se o critério de máxima probabilidade

A Figura 3.8 mostra mapas de textura classificados considerando-se limiares de incertezas ao se atribuir uma classe válida a um ponto. Somente localizações com incerteza máxima menor que o limiar adotado foi classificada com uma das 4 classes de textura do mapa da Figura 3.7. Com esta metodologia pode-se conseguir mapas de áreas de interesse, valores estimados válidos, assumindo-se um valor máximo de risco.

A Figura 3.9 apresenta um mapa de incertezas, associadas ao estimador apresentado na Figura 3.7, gerado pelo critério da entropia de Shannon. Diferentemente do mapa de incertezas por máxima probabilidade, que mostra as maiores incertezas nas regiões de transição entre classes, no mapa de incerteza por entropia de Shannon os maiores valores de incerteza estão situados em regiões de maior confusão entre classes de textura. A escolha do mapa de incerteza mais adequado depende, basicamente, da aplicação que utilizará este mapa.

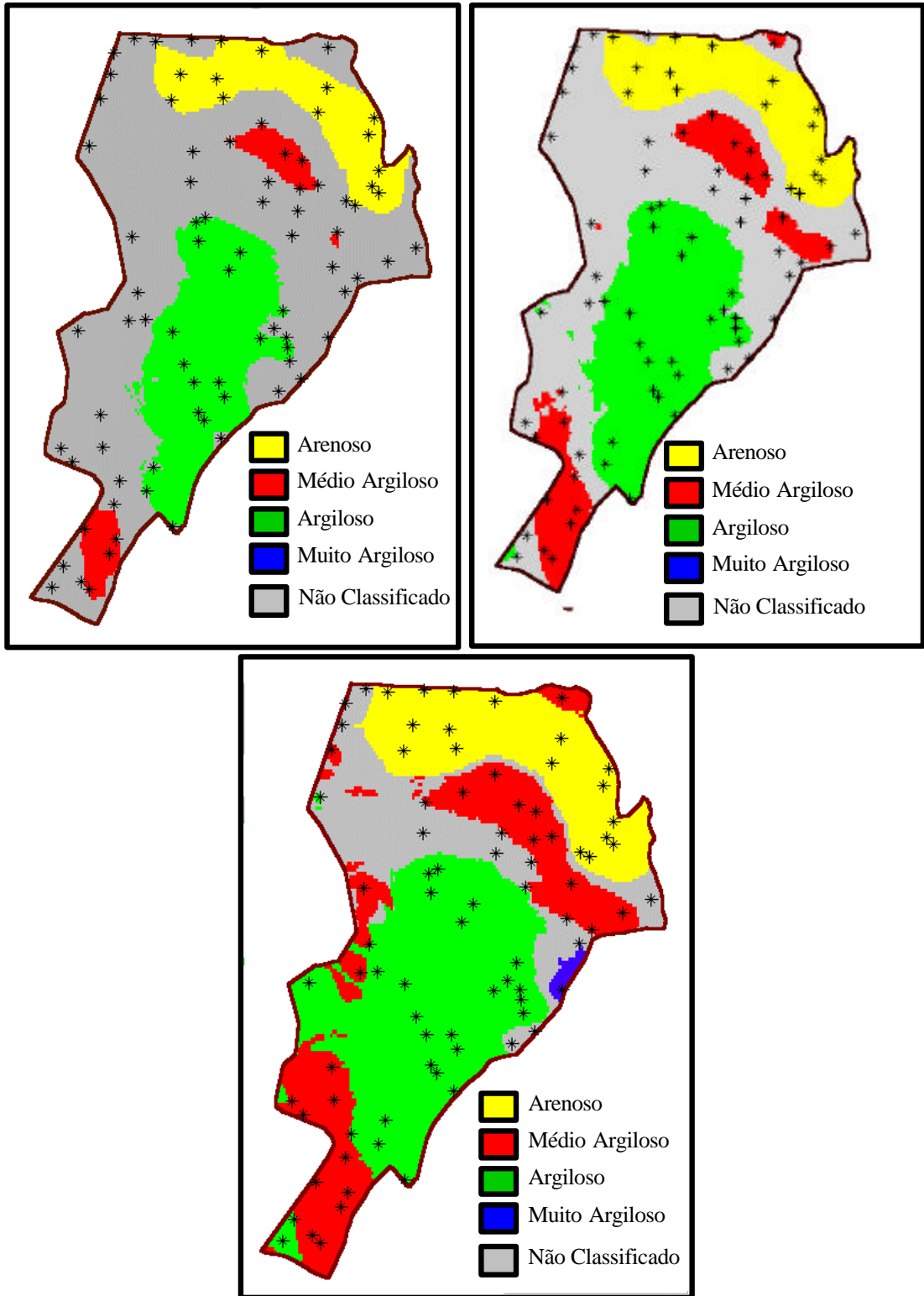


Figura 3.8 - Espacialização do atributo textura do solo condicionada a diferentes limiares (0.38, 0.4 e 0.5) máximos de incerteza

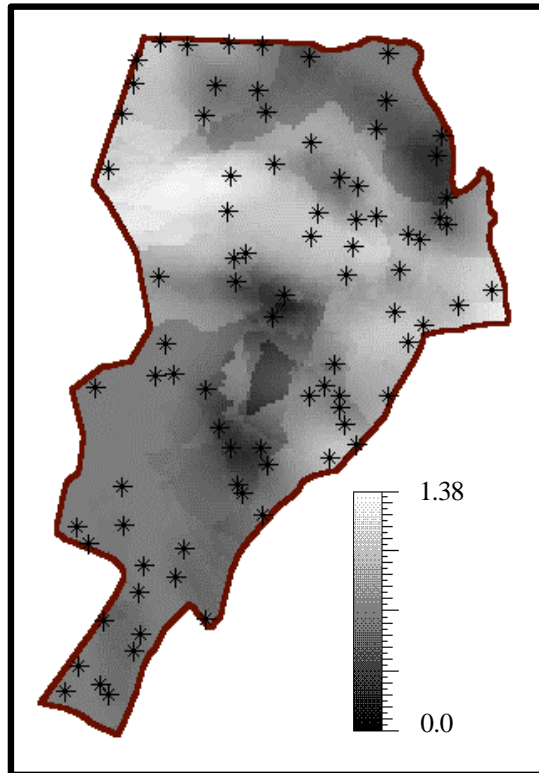


Figura 3.9 - Mapa de incerteza por entropia de Shannon estimado a partir do procedimento de krigeagem por indicação usado para inferir o mapa da Figura 3.7

4 Conclusões

Os conceitos, os resultados e as análises apresentados neste trabalho mostram que as metodologias de inferência, baseadas na geoestatística por indicação, são opções interessantes para modelagem de atributos espaciais. O trabalho mostrou que essas metodologias se aplicam a atributos numéricos e temáticos e, também, que elas possibilitam a associação de informação de incertezas relacionadas com as inferências. Algumas métricas de incerteza foram exploradas e o uso de cada uma delas vai depender da natureza, numérica ou temática, do atributo modelado e do tipo de aplicação que fará uso do modelo criado.

Nas análises dos mapas de incerteza apresentados observou-se que regiões espaciais com maior variabilidade do atributo são as candidatas mais fortes a

apresentarem informação de incerteza mais alta (regiões mais brancas nos mapas de incerteza). Além disso, é importante ressaltar que a incerteza medida pode ser diminuída, local e globalmente, com amostragens mais refinadas nessas regiões. Este fato indica que as informações de incerteza desses modelos são também úteis para planejamento de reamostragem ou de amostragem mais fina dentro de uma área de interesse.

Finalmente, uma outra vantagem dos procedimentos de geoestatística por indicação, não explorada neste texto, está no fato de se poder integrar dados indiretos, informação proveniente de outro atributo correlacionado, na modelagem do atributo em estudo. Esta propriedade pode melhorar o processo de modelagem, diminuir a incerteza das inferências e auxiliar no planejamento de reamostragens futuras.

Referências Bibliográficas

Burrough, P. A.; McDonnell, R. A.,1998. **Principles of Geographical Information Systems**. New York, Oxford University Press, 1998. 333p.

Calderano Filho, B.; Fonseca, O. O. M.; Santos, H. G.; Lemos A. L. **Levantamento Semidetalhado dos Solos da Fazenda Canchim São Carlos - SP**. Rio de Janeiro, EMBRAPA- CNPS, 1996. 261p.

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.; Journel, A. G. **GSLIB Geostatistical Software Library and User's Guide**. New York, Oxford University Press, 1998. 369p.

Felgueiras C. A. **Modelagem Ambiental com Tratamento de Incertezas em Sistemas de Informação Geográfica: O Paradigma Geoestatístico por Indicação**. Tese (Doutorado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, Publicado em <http://www.dpi.inpe.br/teses/carlos/>, 1999a.

Felgueiras, C. A.; Monteiro, A. M. V.; Camargo, E. C. G.; Câmara, G. e Fuks, S. D.. **"Integrating Geostatistical Tools in Geographical Information Systems"**. In: Proceedings of the I Brazilian Workshop on GeoInformatics, 20 e 21/10, Campinas, 1999b. Proceedings. São Paulo, Geo-Info 1999. Palestra Técnico-Científica. p.40-44

Goovaerts, P. **Geostatistics for natural resources evaluation**. New York: Oxford University Press, 1997. 481p.

Heuvelink G. B. M. **Error Propagation in Environmental Modeling with GIS**, Bristol, Taylor and Francis Inc, 1998.

Isaaks, E. H.; Srivastava, R. M. **An Introduction to Applied Geostatistics**. New York, Oxford University Press, 1989. 561p.

Journel, A. G. Nonparametric Estimation of Spatial Distributions. **Mathematical Geology**, v. 15, n. 3, p. 445-468, 1983

Journel, A. G.; Deutsch, C. V. Entropy and Spatial Disorder. **Mathematical Geology**, v. 25, n. 3, p. 329-355, 1993.

Myers, J. C. **Geostatistical error management**: quantifying uncertainty for environmental sampling and mapping. New York: Van Nostrand Reinold, 1997. 571p.

Oliver, M. A.; Webster, R. Kriging: a method of interpolation for geographical information systems. **International Journal of Geographical Information Systems**, v. 4, n. 3, p. 313-332, 1990.

Shannon, C. E.; Weaver, W. **The Mathematical Theory of Communication**. Urbana, The University of Illinois Press, 1949. 117p.

SPRING V 3.5, (DPI/INPE) **Sistema de Processamento de Informações Georeferenciadas** – Divisão de Processamento de Imagens (DPI) do Instituto Nacional de Pesquisas Espaciais (INPE), <http://www.dpi.inpe.br/spring/> , 2001.