

INPE-8180-TDI/760

# MODELAGEM AMBIENTAL COM TRATAMENTO DE INCERTEZAS EM SISTEMAS DE INFORMAÇÃO GEOGRÁFICA: O PARADIGMA GEOESTATÍSTICO POR INDICAÇÃO

Carlos Alberto Felgueiras

Tese de Doutorado em Computação Aplicada, orientada pela Dra. Suzana Druck Fuks e pelo Dr. Antônio Miguel Vieira Monteiro, aprovada em 21 de dezembro de 1999.

INPE São José dos Campos 2001 550.34.012 : 519.766.4

## FELGUEIRAS, C. A.

Modelagem ambiental com tratamento de incertezas em sistemas de informação geográfica: o paradigma geostatístico por indicação / C. A. Felgueiras - São José dos Campos: INPE, 1999.

212p. – (INPE-8180-TDI/760).

1.Sistemas de Informação Geográfica (SIG). 2.Geostatística. 3.Krigeagem. 4.Simulação estocástica. 5.Modelagem de incertezas. 6.Qualidade de dados. I. Título.

Aprovado pela Banca Examinadora em cumprimento a requisito exigido para a obtenção do Título de Doutor em Computação Aplicada.

Dr. Gilberto Câmara Neto	Presidente
Dr. Antônio Miguel Vieira Monteiro	Muickello Grientador
Dra Suzana Druck Fucks	Suzana Druck Lubs Orientadora
Dr <sup>a</sup> Corina da Costa Freitas	Membro da Banca
Dr. Eduardo Delgado Assad	Membro da Banca Convidado
Dr. Paulo Cézar Pinto Carvalho	Membro da Banca Convidado

Candidato (a): Carlos Alberto Felgueiras

São José dos Campos, 21 de dezembro de 1999.

e o que bate abrir-se-lhe-á. ( Lucas, 11:9,10)

A meus pais de criação Anibal (em memória) e Eugênia, à minha esposa Débora e à minha filha Carla

### **AGRADECIMENTOS**

A Jesus Cristo, por ter me guiado espiritualmente nesta caminhada de crescimento pessoal e profissional.

Aos orientadores deste trabalho, Dra. Suzana Druck Fuks e Dr. Antônio Miguel Vieira Monteiro, pela confiança, apoio, paciência e amizade.

Aos estimados amigos de trabalho da Divisão de Processamento de Imagens do Instituto Nacional de Pesquisas Espaciais, pela camaradagem, cooperação e amizade.

À chefia da Divisão de Processamento de Imagens do Instituto Nacional de Pesquisas Espaciais, pela oportunidade, confiança e apoio logístico.

Aos membros da banca examinadora, pela predisposição em analisar este trabalho e pelas sugestões recebidas.

Finalmente, agradeço a todos que direta ou indiretamente colaboraram na elaboração desta tese.

### **RESUMO**

Os produtos de modelagens envolvendo propriedades e processos do meio físico natural, gerados no ambiente computacional de Sistemas de Informação Geográfica (SIG) como mapas de aptidão agrícola, mapas de vulnerabilidade à erosão, zoneamento ecológico-econômico, para citar alguns, deveriam fornecer informação quantitativa sobre a qualidade dos produtos disponibilizados, uma vez que são destinados a apoiar processos de tomada de decisão. A tecnologia atual dos SIG não incorpora, em seus procedimentos, a noção de qualidade aos diferentes cenários que é capaz de produzir através da integração e operação sobre dados espaciais de diversas naturezas. Assim, em relação ao uso sistemático de ambientes de SIG, no atual estágio da tecnologia, para as mais diversas modelagens em apoio a processos decisórios, o que efetivamente tem acontecido chega a ser temerário. Na imensa maioria dos casos, nenhuma informação quantitativa sobre a qualidade dos produtos gerados é agregada ao produto final, e o que era para ser um cenário, uma possível resposta a uma determinada modelagem para os dados envolvidos no processamento, passa a ser a expressão da verdade. É neste contexto que se insere este trabalho. Uma proposta de sistematização, baseada no paradigma da modelagem geoestatística por indicação, para uso em modelagem computacional desenvolvida em ambiente de SIG. As técnicas e métodos. apontados pela sistematização aqui proposta, são explorados na teoria e na prática, através de sua implementação, e são demonstradas em estudos de caso. Este trabalho dedica especial atenção à formulação de técnicas de estimativa e métricas de incerteza que possam abordar, de forma unificada, dados geográficos de natureza temática (categórica) e numérica (contínua). apresentando resultados promissores. Discute-se e testa-se a hipótese do uso de dados indiretos, quando há pouca disponibilidade de amostras, como forma de se procurar melhorar a acurácia das inferências utilizadas. A integração destes procedimentos aos ambientes de SIG, em seu atual estágio de desenvolvimento tecnológico, é discutida e os requisitos básicos para esta integração são apresentados. Com este trabalho espera-se estar contribuindo na direção de orientar arquitetos e desenvolvedores de SIG para incorporar as ferramentas de modelagem espacial necessárias à manipulação de dados geográficos, com suas medidas de incertezas associadas, e à propagação das incertezas para os produtos de modelagem ambiental, que irá qualificar o produto resultante do uso da tecnologia.

## ENVIRONMENTAL MODELING WITH UNCERTAINTY TREATMENT IN GEOGRAPHICAL INFORMATION SYSTEMS: THE INDICATOR GEOSTATISTICAL PARADIGM

#### **ABSTRACT**

The use of Geographic Information Systems (GIS) to model environmental phenomena limits decision-making processes because the products of GIS modeling, such as agriculture suitability and vulnerability to erosion maps or economic-ecological zoning, do not encompass quantitative information on data quality. Current GIS technology does not associate the notion of data quality to the various scenarios resulting from integrating and combining different spatial data. This situation requires attention and jeopardizes decision-making processes based on GIS. In most cases there is no information on data quality aggregated to the final products. As a consequence, what should be a scenario of possible answers to a certain modeling turns out to be the expression of truth. The present research is concerned with issues that come from this context. This thesis is a proposal towards a systematic computational modeling in GIS based on the paradigm of indicator geostatistic modeling. The proposed techniques and methods are implemented and demonstrated through case studies. This work is especially concerned with the formulation of estimation techniques and uncertainty metrics that can treat both thematic (categorical) and numerical (continuous) geographic data. It also discusses and tests the hypothesis of using soft data to improve the estimate accuracy when only a few samples are available. The integration of these procedures in a GIS environment, as well as the basic requirements for such integration, is discussed and presented. This research contributes to guide GIS designers and developers to incorporate spatial modeling tools required for an appropriate handling of geographic data, taking their uncertainties into consideration when the data set is directly involved in environmental modeling.

## SUMÁRIO

	<u>Pág</u>
LISTA DE FIGURAS LISTA DE TABELAS	
LISTA DE SÍMBOLOS	
CAPÍTULO 1 - INTRODUÇÃO	27
1.1 - Motivação	27
1.2 - Definições preliminares	29
1.3 - Objetivos do trabalho	31 32
1.5 - Organização do trabalho	34
CAPÍTULO 2 - MODELAGEM AMBIENTAL EM SISTEMAS DE	
INFORMAÇÃO GEOGRÁFICA COM TRATAMENTO DE INCERTEZAS	37
2.1 - Introdução	37 38
2.2.1 - A natureza da modelagem ambiental em meio digital	39
2.2.1.1 - A dimensão espaço-temporal e a natureza contínua dos	39
processos da natureza	40
2.2.1.3 - Representações de continuidade a partir de dados amostrais	41
<ul><li>2.2.2 - Modelagem de dados espaciais em SIG</li><li>2.2.3 - Modelagem computacional para modelos ambientais em SIG</li></ul>	42 46
2.3 - Propagação de incertezas em modelagem ambiental com SIG	49
2.3.1 - Método de Taylor de primeira ordem	51 52
2.3.3 - Método de Rosenblueth	53
2.3.4 - Método de Monte Carlo	53 54
2.5 - Conclusões	56
CAPÍTULO 3 - INFERÊNCIAS COM ESTIMATIVA DE INCERTEZAS	
POR KRIGEAGEM	59
3.1 - Introdução	59
3.2 - Conceitos gerais da geoestatística	60
<ul><li>3.3 - O estimador de krigeagem linear</li><li>3.4 - A variância de krigeagem linear como medida de incerteza: uma</li></ul>	61
análise crítica	68
3.5 - O Estimador de krigeagem por indicação	69 70

<ul> <li>3.5.2 - A krigeagem por indicação para atributos temáticos</li></ul>	75 77 79 82 85
CAPÍTULO 4 - INFERÊNCIAS COM ESTIMATIVA DE INCERTEZAS POR SIMULAÇÃO ESTOCÁSTICA CONDICIONADA	87
4.1 - Introdução	87
4.2 - O procedimento de simulação sequencial condicionada	88
4.3 - Inferência de estatísticas a partir dos valores simulados	93
4.3.1 - Estatísticas sobre realizações de atributos numéricos	93
4.3.2 - Estatísticas sobre realizações de atributos temáticos	94
4.4 - Estimativa de incertezas utilizando simulação sequencial	
condicionada	94
4.4.1 - Estimativa de incertezas locais: atributos numéricos simulados	95
4.4.2 - Estimativa de incertezas locais para atributos temáticos simulados	95
4.4.3 - Métricas globais de incerteza para atributos simulados	95
4.5 - Comparação entre simulação condicionada e krigeagem	97
4.6 - Conclusões	98
CAPÍTULO 5 - INFERÊNCIAS COM ESTIMATIVA DE INCERTEZAS	
COM O USO DE DADOS INDIRETOS	101
COM O USO DE DADOS INDIRETOS	<b>101</b> 101
5.1 - Introdução	101
5.1 - Introdução	101 102
5.1 - Introdução	101 102 106
<ul> <li>5.1 - Introdução</li> <li>5.2 - Os problemas de subamostragem e de amostragem inadequada para inferências baseadas na geoestatística</li> <li>5.3 - Codificação por indicação de dados amostrais diretos e indiretos</li> <li>5.4 - A krigeagem por indicação com dados diretos e indiretos</li> </ul>	101 102 106 108
5.1 - Introdução	101 102 106 108 109
5.1 - Introdução	101 102 106 108 109 111
5.1 - Introdução 5.2 - Os problemas de subamostragem e de amostragem inadequada para inferências baseadas na geoestatística	101 102 106 108 109
5.1 - Introdução	101 102 106 108 109 111 112
5.1 - Introdução	101 102 106 108 109 111 112 113
5.1 - Introdução	101 102 106 108 109 111 112 113 113
5.1 - Introdução	101 102 106 108 109 111 112 113 115
5.1 - Introdução 5.2 - Os problemas de subamostragem e de amostragem inadequada para inferências baseadas na geoestatística 5.3 - Codificação por indicação de dados amostrais diretos e indiretos 5.4 - A krigeagem por indicação com dados diretos e indiretos 5.5 - A cokrigeagem por indicação 5.6 - Otimização pelo modelo Markov-Bayes 5.7 - Otimização para dados colocalizados 5.8 - Inferência e estimativa de incertezas locais e globais com o uso de dados diretos e indiretos 5.9 - Conclusões  CAPÍTULO 6 - ESTUDOS DE CASOS E ANÁLISES  6.1 - Introdução 6.2 - Caracterização da área utilizada nos estudos de caso 6.3 - Estudo da propriedade altimetria sob representação numérica: caso Canchim	101 102 106 108 109 111 112 113 115
5.1 - Introdução 5.2 - Os problemas de subamostragem e de amostragem inadequada para inferências baseadas na geoestatística 5.3 - Codificação por indicação de dados amostrais diretos e indiretos 5.4 - A krigeagem por indicação com dados diretos e indiretos 5.5 - A cokrigeagem por indicação 5.6 - Otimização pelo modelo Markov-Bayes 5.7 - Otimização para dados colocalizados 5.8 - Inferência e estimativa de incertezas locais e globais com o uso de dados diretos e indiretos 5.9 - Conclusões  CAPÍTULO 6 - ESTUDOS DE CASOS E ANÁLISES  6.1 - Introdução 6.2 - Caracterização da área utilizada nos estudos de caso 6.3 - Estudo da propriedade altimetria sob representação numérica: caso Canchim 6.3.1 - Inferências com estimativa de incertezas por krigeagem ordinária	101 102 106 108 109 111 112 113 115 115
5.1 - Introdução 5.2 - Os problemas de subamostragem e de amostragem inadequada para inferências baseadas na geoestatística 5.3 - Codificação por indicação de dados amostrais diretos e indiretos 5.4 - A krigeagem por indicação com dados diretos e indiretos 5.5 - A cokrigeagem por indicação 5.6 - Otimização pelo modelo Markov-Bayes 5.7 - Otimização para dados colocalizados 5.8 - Inferência e estimativa de incertezas locais e globais com o uso de dados diretos e indiretos 5.9 - Conclusões  CAPÍTULO 6 - ESTUDOS DE CASOS E ANÁLISES  6.1 - Introdução 6.2 - Caracterização da área utilizada nos estudos de caso 6.3 - Estudo da propriedade altimetria sob representação numérica: caso Canchim	101 102 106 108 109 111 112 113 115 115 117

6.3.3 - Inferências com estimativa de incertezas por simulação	120
sequencial condicionada	130 132
6.4 - Estudo da textura do solo sob representação temática: caso	102
Canchim	135
6.4.1 - Inferências com estimativa de incertezas por krigeagem por	
indicação	136
6.4.2 - Inferências com estimativa de incertezas por simulação	
sequencial condicionada	147
6.4.3 - Análise e comparação dos resultados obtidos	151
6.5 - Uso de dados indiretos	154
6.5.1 - Uso do mapa de solos para estimar o atributo textura do solo	155
6.5.2 - Análise dos resultados e comparação com estimativas sem dados	
indiretos	159
6.6 - Medidas globais de incerteza	161
6.7 - Conclusões	162
CAPÍTULO 7 - INTEGRAÇÃO DE PROCEDIMENTOS GEOESTATÍSTI-	400
COS AOS SISTEMAS DE INFORMAÇÃO GEOGRÁFICA	163
7.4 Jahradus **	400
7.1 - Introdução	163
7.2 - SIG e Geoestatística	164
7.2.1 - Importância do SIG para implementação das facilidades da	1 <i>CE</i>
geoestatística	165
7.2.2 - Importância dos procedimentos de geoestatística para um SIG	166
7.2.3 Integração da geoestatística a um SIG	167 168
7.3 - Módulos de geoestatística integrados a um SIG	168
7.3.1 - Requisitos básicos de um módulo de geoestatística para um SIG.	170
7.3.2 - O módulo de geoestatística do SPRING	170
7.3.3 - O módulo de geoestatística do IDRISI	
7.3.4 - O módulo de geoestatística do PCRaster	180 182
7.3.5 - Outros módulos de geoestatística	183
7.4 - Linguagens para modelagem ambiental com tratamento de incerteza	103
7.4.1 - ADAM: uma ferramenta para modelagens com propagação de	105
incertezas	185
	187
especificação preliminar de requisitos	
7.5 - Conclusões	190
CAPÍTULO 8 - CONCLUSÕES	193
CAPITULU 0 - CUNCLUSUES	133
8.1 - Considerações finais	193
8.1 - Considerações finais	
8.2 - Sugestões para futuras investigações	195
REFERÊNCIAS BIBLIOGRÁFICAS	197
APÊNDICE A - DADOS UTILIZADOS NO TRABALHO	203

## **LISTA DE FIGURAS**

		Dác
2.1	Operações locais: (a) pontuais e (b) de vizinhança em modelagem computacional. Fonte: Baseada em Heuvelink (1998)	<u>Pág.</u> 50
3.1	Variograma teórico ajustado sobre variograma experimental obtido a partir das amostras.	65
3.2	Codificação por indicação dos dados amostrais numéricos para um valor de corte $z = z_k$	71
3.3	Codificação por indicação dos dados amostrais temáticos para uma classe de corte $z = z_k$	75
3.4	Correção dos desvios de relação de ordem para dados numéricos Fonte: Baseada em Deutsch e Journel (1998)	84
4.1	Obtenção de um valor da VA a partir da função de distribuição de probabilidade acumulada inferida para um atributo numérico	91
4.2	Obtenção de um valor da VA a partir da função de distribuição de probabilidade acumulada inferida para um atributo temático	92
5.1	Exemplo de ajuste para dados de amostragem reduzida e pouco representativa (sem alguns valores máximos, mínimos e pontos de inflexão).	103
5.2	Exemplo de ajuste para uma amostragem com valores de máximos, mínimos e pontos de inflexão.	104
6.1	Localização da Fazenda Canchim. FONTE: Calderano Filho et al. (1996, p. 4)	116
6.2	Distribuição espacial das amostras, sobreposta à representação em imagem nível de cinza, da altimetria de Canchim	118
6.3	Mapa de valores inferidos por krigeagem ordinária utilizando o conjunto amostral de altimetria	119
6.4	Histograma do conjunto amostral de altimetria de Canchim	121
6.5	Mapa de incertezas locais obtido a partir da variância do erro de krigeagem ordinária.	122
6.6	Mapa de valores médios inferidos a partir das fdac locais construídas pelo procedimento de krigeagem por indicação	123

6.7	Mapa de incertezas locais obtido a partir das variâncias inferidas dos modelos de distribuição probabilística construídos pela krigeagem por indicação	125
6.8	Mapa de medianas inferidas a partir das fdac locais construídas pelo procedimento de krigeagem por indicação	126
6.9	Mapa de incertezas locais obtido a partir dos quartis, primeiro e terceiro, inferidos dos modelos de distribuição probabilística construídos pela krigeagem por indicação	127
6.10	Mapa de classes de altimetria inferido por intervalos de probabilidade a partir das fdac inferidas pela krigeagem por indicação.	128
6.11	Mapa de incertezas por intervalos de probabilidades referente à classificação apresentada na Figura 6.10	129
6.12	Mapa de valores médios inferidos a partir de um conjunto de 400 realizações geradas por simulação sequencial condicionada as amostras de altimetria apresentadas na Figura 6.2	130
6.13	Mapa de incertezas obtido a partir das variâncias inferidas dos modelos de distribuição probabilística construídos pela simulação sequencial condicionada aos dados de altimetria	131
6.14	Distribuição das amostras de textura de solo na região de Canchim.	135
6.15	Mapas de valores de textura do solo inferidos pelo procedimento de krigeagem por indicação com 12 pontos condicionantes	137
6.16	Mapa de incerteza por moda estimado a partir do procedimento de krigeagem por indicação usado para inferir o mapa da Figura 6.15	139
6.17	Classificação do mapa de incerteza apresentado na Figura 6.16	140
6.18	Mapas de valores estimados de textura do solo e incerteza associadas com, no máximo, 4 e 16 pontos condicionantes	141
6.19	Mapa de incerteza por entropia de Shannon estimado a partir do procedimento de krigeagem por indicação usado para inferir o mapa da Figura 6.15	142
6.20	Espacialização do atributo textura do solo condicionada a diferentes limiares (0.38, 0.4 e 0.5) máximos de incerteza	143
6.21	Gráficos de Médias de Incerteza por Número Máximo de PC's	145

6.22	Gráficos de Variâncias de Incerteza por Número Máximo de PC's	146
6.23	Exemplos de realizações do atributo textura do solo geradas pela simulação sequencial	148
6.24	Mapa de valores de textura do solo inferidos pelo critério de frequência máxima a partir do conjunto de realizações gerados pela simulação sequencial	149
6.25	Incertezas associadas aos valores de textura de solo mostrados na Figura 6.24	150
6.26	Mapa de Solos da região da fazenda Canchim	154
6.27	Mapa de valores de textura inferido por simulação sequencial condicionada a dados diretos e indiretos	158
6.28	Mapa incertezas, por moda, associadas aos valores de textura inferidos por simulação sequencial condicionada a dados diretos e indiretos	159
6.29	Mapas das Figuras 6.27 e 6.28 com os contornos do mapa de solo sobrepostos.	160
7.1	Estrutura do Módulo de Geoestatística no SPRING. Fonte: Camargo et al. (1999)	171
7.2	Tela principal do SPRING 3.3 com a interface de interação para análise de continuidade espacial por semivariograma	173
7.3	Interface do SPRING 3.3 para ajuste de um modelo teórico para o semivariograma experimental	174
7.4	Interface do SPRING 3.3 para espacialização do atributo por krigeagem simples ou ordinária	175
7.5	Ilustração do Modelador de Dependência Espacial do IDRISI32. Fonte: http://www.idrisi.com/03prod/ldrisi.htm	178
7.6	Interface de ajuste de curvas do sistema IDRISI32. Fonte: http://www.idrisi.com/03prod/ldrisi.htm	179
7.7	Interface de krigeagem e simulação do sistema IDRISI32. Fonte: http://www.idrisi.com/03prod/ldrisi.htm	180
7.8	Exemplo de ajuste de variograma pelo programa GSTAT	181

## **LISTA DE TABELAS**

		<u>Pág.</u>
3.1	Exemplo de cálculos de incertezas para várias distribuições	81
6.1	Parâmetros estatísticos sobre o conjunto amostral de altimetria de Canchim	120
6.2	Parâmetros dos variogramas teóricos definidos para os dados de altimetria transformados por indicação nos valores de corte apresentados	124
6.3	Classes de textura com probabilidades a priori	136
6.4	Parâmetros dos variogramas teóricos definidos para os dados de textura do solo transformados por indicação	138
6.5	Estatísticas dos valores de incertezas, por moda e por entropia, obtidos por krigeagem por indicação para números máximos de PC distintos	144
6.6	Extensão dos tipos de solos da região de Canchim	155
6.7	Probabilidade de ocorrência das classes de textura de solo para as classes de solo de Canchim	156
6.8	Estatísticas das medidas de imprecisão com e sem dados indiretos	158

## LISTA DE SÍMBOLOS

A	- região ou área de estudo
$B(z_k)$	<ul> <li>coeficientes de calibração da variável secundária a partir da primária</li> </ul>
<i>C</i> ( <b>h</b> )	- covariância entre variáveis aleatórias, VA, em função de <b>h</b>
$C(\mathbf{u}_{\beta} - \mathbf{u}_{\alpha})$	- covariância entre as amostras de um atributo posicionadas em $\mathbf{u}_{\beta}$ e $\mathbf{u}_{\alpha}$
$C_l(\mathbf{h}; z_k)$	- autocovariância da VA por indicação $I$ em função de $\mathbf{h}$ e de $\mathbf{z}_k$
$C_I(\mathbf{h}_{\alpha\beta}; \mathbf{z}_k)$	- autocovariância da VA por indicação $\emph{I}$ definida para $\mathbf{h}_{lphaeta}$ e $\mathbf{z}_{\emph{k}}$
$C_l(\mathbf{h}_{\alpha}; \mathbf{z}_k)$	- autocovariância da VA por indicação $I$ definida para $\mathbf{h}_{\alpha}$ e z= $\mathbf{z}_{k}$
$C_Y(\mathbf{h}; z_k)$	- autocovariância da VA Y em função de ${f h}$ e de ${f z}_k$
$C_{IY}(\mathbf{h}; z_k)$	- covariância cruzada entre as VA $\ I$ e Y em função de $\mathbf{h}$ e de $\mathbf{z}_k$
$C_R(\mathbf{h})$	- autocovariância da função residual
$d_{lpha}$	- distância euclidiana da amostra $lpha$ e o ponto a ser interpolado
<i>E</i> { <i>Z</i> }	- valor esperado da variável aleatória Z
$E\{Z(\mathbf{u})\}$	- esperança da VA Z definida para a posição espacial <b>u</b>
$E\{I(\mathbf{u};z_k) (n)\}$	- valor esperado da VA por indicação em $z_k$ condicionado a $n$ amostras
f(z)	- função de densidade de probabilidade de uma VA Z
$f(\mathbf{u},\mathbf{z} (n)$	<ul> <li>função densidade de probabilidade de Z(u), condicionada a n amostras.</li> </ul>
<i>F</i> ( <b>u</b> ; <i>z</i> )	- função de distribuição acumulada da variável aleatória $Z(\mathbf{u})$
$F(\mathbf{u}; z (n))$	<ul> <li>função de distribuição acumulada condicionada, fdac, a n amostras da VA Z(u)</li> </ul>
$F^*(z_k)$	- valor estimado da média estacionária em $A$ para $z = z_k$
$F^*(\mathbf{u}; \mathbf{z}_k)$	- valor estimado para a fdac de $Z(\mathbf{u})$ em $z = z_k$
$F^*(\mathbf{u}; z_k (n))$	- valor estimado para a fdac de $Z(\mathbf{u})$ em $z_k$ , condicionada a $n$ amostras
$F^*(\mathbf{u}; \mathbf{z}_k   (n+K))$	- valor estimado para a fdac de $Z(\mathbf{u})$ em $z_k$ , condicionada a $n$ - amostras diretas e $K$ observações indiretas
$F(z_1,,z_m)$	- função de distribuição conjunta de m variáveis aleatórias
$F(z_1,,z_m (n))$	<ul> <li>função de distribuição conjunta de m VA condicionada a n amostras</li> </ul>
g	<ul> <li>função que relaciona atributos de entrada numa modelagem espacial</li> </ul>
h	- vetor de separação entre duas posições espaciais
$\mathbf{h}_{lpha}$	- vetor de separação definido por $\mathbf{u}_{\alpha}$ e uma posição $\mathbf{u} \in A$
$h_{lphaeta}$	- vetor de separação definido pelas posições amostrais $\mathbf{u}_{lpha}$ e $\mathbf{u}_{eta}$

Н	volor de entrania de Channan
	- valor da entropia de Shannon
H(u)	- valor da entropia de Shannon definida para a posição <b>u</b>
H(h)	- entropia referente ao vetor de separação <b>h</b>
$H_R(\mathbf{h})$	- entropia relativa normalizada referente ao vetor de separação h
$I(\mathbf{u}; \mathbf{z}_k)$	- conjunto amostral codificado por indicação no valor de corte $z_k$
$i(\mathbf{u}_{\alpha}, \mathbf{z}_{k})$	- valor da amostra direta em $\mathbf{u}_{\alpha}$ e codificada por indicação em $z_k$
$j(\mathbf{u}_{\alpha}; \mathbf{z}_{k})$	- valor da amostra indireta em $\mathbf{u}_{lpha}$ codificada por indicação em $\mathbf{z}_{k}$
<i>Inc</i> (u)	- valor de incerteza determinado para a posição espacial <b>u</b>
k	<ul> <li>matriz de covariâncias entre amostras e posição u do valor a ser estimado</li> </ul>
K	<ul> <li>número total de valores de corte usados na codificação por indicação</li> </ul>
K	- matriz de covariâncias entre amostras de um atributo espacial
L	<ul> <li>número total de realizações de uma variável ou de uma função aleatória</li> </ul>
m	<ul> <li>número total de variáveis aleatórias que formam um campo aleatório</li> </ul>
$m^{(0)}(z_k)$	- valor esperado de VA secundária $Y(\mathbf{u}, z_k)$ dado que a VA primária, $I(\mathbf{u}; z_k)$ , é igual a 0
$m^{(1)}(z_k)$	- valor esperado de VA secundária $Y(\mathbf{u}, z_k)$ dado que a VA primária, $I(\mathbf{u}; z_k)$ , é igual a 1
n	- número total de amostras (diretas) de um atributo espacial
<i>n</i> ( <b>u</b> )	- número total de amostras vizinhas a uma posição espacial <b>u</b>
N	- número de atributos de entrada de uma modelagem espacial
$N(\mathbf{h})$	- número de pares de amostras separadas pelo vetor <b>h</b>
$p_j(\mathbf{u})$	- probabilidade de ocorrência da classe j na posição <b>u</b>
$p_i^{(l)}$	<ul> <li>probabilidade de ocorrência da i-ésima VA na I-ésima realização</li> </ul>
$p_{k,k'}(\mathbf{h})$	- probabilidade bivariada, <i>Prob</i> { <i>Z</i> ( <b>u</b> ) ∈ a classe <i>k</i> e <i>Z</i> ( <b>u</b> ') ∈ a classe <i>k</i> '}, referente ao vetor de separação <b>h</b>
$q_j$	- <i>i</i> -ésimo quantil de uma distribuição de probabilidade
<b>9</b> 0.5	- mediana de uma distribuição de probabilidade
$S_k$	- k-ésima classe de uma variável aleatória categórica
u	- vetor de coordenadas
$u_lpha$	- coordenadas referentes as amostras
Υ	<ul> <li>variável aleatória de um atributo de saída de uma modelagem espacial</li> </ul>
$Y(\mathbf{u}; \mathbf{z}_k)$	- conjunto amostral indireto codificado por indicação no corte $z_k$
$y(\mathbf{u}_{\alpha}; \mathbf{z}_{k})$	- valor de probabilidade a priori, em u = $\mathbf{u}_{\alpha}$ e $z = z_k$ , obtida a partir de dados indiretos

Z	- realização de uma variável aleatória Z
$\boldsymbol{z}_i$	- <i>i</i> -ésima realização da variável aleatória Z <sub>i</sub>
z <sub>im</sub>	- valor médio da <i>i</i> -ésima variável aleatória Z <sub>i</sub>
$z(u_lpha)$	- valor do atributo na amostra vizinha localizada em $\mathbf{u}_{lpha}$
$z^{*}(\mathbf{u})$	- valor do atributo inferido numa posição <b>u</b>
$z_{\mathcal{S}}^{*}(u)$	- valor do atributo inferido, por krigeagem simples, na posição <b>u</b>
$z_0^*(u)$	<ul> <li>valor do atributo inferido, por krigeagem ordinária, na posição u</li> </ul>
$z^{(\prime)}(u)$	- <i>l</i> -ésima realização da variável aleatória Z em <b>u</b>
$\mathbf{z_i}^{(l)}$	- valor obtido da <i>i</i> -ésima variável aleatória na <i>l</i> -ésima realização
Z	<ul> <li>variável aleatória representando um atributo espacial</li> </ul>
<i>Z</i> (u)	- variável aleatória definida para a posição <b>u</b>
$Z_i$	- <i>i</i> -ésima variável aleatória
$Z_i(\mathbf{u})$	- <i>i</i> -ésima variável aleatória definida para a posição <b>u</b>
$\phi(u)$	- multiplicador de Lagrange referente à posição <b>u</b>
$\phi(\mathbf{u}; \mathbf{z}_k)$	- multiplicador de Lagrange para posição ${f u}$ e $$ valor de corte ${f z}_k$
γ <b>(h</b> )	- semivariograma estacionário para o vetor de separação <b>h</b>
$\lambda_{lpha}\left(\mathbf{u} ight)$	- valor de ponderação referente à amostra $lpha$ e à posição ${f u}$ .
$\lambda_{Slpha}\left(u ight)$	- valor de ponderação de krigeagem simples referente à amostra $\alpha$ e à posição ${\bf u}$ .
$\lambda_{0lpha}\left(\mathbf{u} ight)$	- valor de ponderação de krigeagem ordinária referente à amostra $\alpha$ e à posição ${\bf u}$ .
$\lambda_{S\alpha}(\mathbf{u}; \mathbf{z}_k)$	- ponderador de krigeagem por indicação simples para $z = z_k$ , amostra $\alpha$ e posição <b>u</b>
$\lambda_{O\alpha}(\mathbf{u}; \mathbf{z}_k)$	- ponderador de krigeagem por indicação ordinária para $z=z_k$ , amostra $\alpha$ e posição espacial $\mathbf{u}$
$\mu$	- média estacionária de uma função aleatória
$\mu\!(u)$	- valor esperado ou média de uma variável aleatória em <b>u</b>
$\mu_{ m Y}$	<ul> <li>valor esperado ou média da variável aleatória Y</li> </ul>
$\upsilon_{\beta}(\mathbf{u}; \mathbf{z}_k)$	- peso de cokrigeagem da variável secundária para $z=z_k$ e observada em ${f u}$
$oldsymbol{ ho}_{ij}$	- coeficiente de correlação entre duas variáveis aleatórias i e j.
$\sigma_{Y}^{2}$	- variância da variável aleatória Y
$\sigma_{\!\scriptscriptstyle  m O}{}^2$	- variância do erro do estimador de krigeagem ordinária
$\sigma_{\!\scriptscriptstyle Y}$	- desvio padrão da variável aleatória Y

## CAPÍTULO 1 INTRODUÇÃO

## 1.1 MOTIVAÇÃO

A adoção, cada vez mais frequente, da tecnologia de *Sistemas de Informação Geográfica* (SIG) por organismos públicos e privados, e a nova inserção desta tecnologia, como base efetiva de informação para sistemas de apoio à decisão gerencial, em seus diversos níveis, têm mobilizado pesquisadores, usuários e desenvolvedores na busca de uma caracterização quantitativa para a avaliação da qualidade dos produtos gerados através da tecnologia SIG disponível.

Ao longo de muitos anos tem-se construído um consenso de que a capacidade de suporte do meio físico ao desenvolvimento de atividades econômicas é finita, e que o desenvolvimento econômico de sustentação prolongada precisa e deve observar as "necessidades" ambientais do lugar geográfico onde se insere. O que se vê, hoje, no Brasil e no mundo, através de seus organismos internacionais de representação, é uma aproximação, ainda que questionável, do discurso ambiental e das práticas de financiamento. Assim, os SIG assumem um papel fundamental e necessário como ferramenta básica para o planejamento e gestão de recursos naturais em um contexto de sistema.

O produto em um SIG é sempre, em essência, o resultado da manipulação de dados espaciais armazenados em meio digital. A informação, necessária aos processos decisórios, é o resultado da modelagem destes dados espaciais, buscando representar uma realidade geográfica, e das operações possíveis sobre estas representações. Este procedimento de modelagem computacional usando SIG é aplicado a modelos computacionais que representam fenômenos ambientais. A diversidade da realidade geográfica conduziu a uma noção de "camadas" para representar a informação espacial em meio digital. Cada "camada" é entendida como uma representação de um conjunto de elementos da realidade geográfica, um nível de informação espacial sob modelagem. A

partir desta definição, um produto em SIG é quase sempre o resultado de operações sobre estas "camadas", ou seja, operações sobre diferentes níveis de representação. Propriedades ambientais, tais como, temperatura, teor de ferro em uma rocha, textura do solo, relevo, e outras, são "camadas" de dados importantes em análises utilizando SIG. Deste modo, a capacidade de integração de dados espaciais dos SIG significa operar sobre níveis de informação espacial diferenciados. Como a cada nível cabe a representação de um conjunto de elementos da realidade modelada, o resultado de uma operação multinível precisa se preocupar com a distorção da realidade geográfica em modelagem, que pode ocorrer como resultado da "integração" de representações diferenciadas e com características particulares. Assim, muitos esforços têm sido empregados na formulação de uma estratégia para que o SIG, com sua capacidade de análise multinível, forneça, também, informação quantitativa com respeito a qualidade dos produtos gerados pelas operações envolvidas na integração. A qualidade desses produtos depende da qualidade dos dados espaciais, em primeira instância, da qualidade de cada nível de informação utilizado na análise e, também, do tipo e dos parâmetros da equação utilizada para integrar as diferentes "camadas". Informação especializada é a base para a tomada de decisão e informação com qualidade é a base para uma decisão mais acertada.

A motivação essencial deste trabalho é estudar, avaliar e propor métodos e técnicas que possam ser incorporados ao processo de modelagem computacional, no estágio atual da tecnologia de SIG, permitindo assim que estes possam ser efetivamente empregados no desenho de sistemas de apoio à decisão que incorporem informação de qualidade e onde as variáveis modeladas são, principalmente, variáveis ambientais relacionadas ao meio físico.

## 1.2 DEFINIÇÕES PRELIMINARES

Um *dado espacial*  $^1$  é caracterizado por sua localização espaço-temporal, (x, y, z, t), e por atributos, geralmente não espaciais. Os atributos ambientais podem ser de natureza categórica, referenciados, neste texto, como *atributos temáticos*, ou de natureza contínua, aqui referenciados como *atributos numéricos*. Elevação, temperatura e teor de argila no solo são exemplos de propriedades numéricas. Classes de textura do solo, tipos de rocha e classes de vegetação são exemplos de atributos categóricos. Os termos *dados temáticos* e *dados numéricos* serão utilizados, neste texto, como referência a dados espaciais cujos atributos são temáticos e numéricos, respectivamente. Também, este trabalho se restringirá a representação de atributos espaciais considerando-se sua localização bidimensional  $(x, y) \in \Re^2$ .

A geoestatística trata os valores dos atributos ambientais segundo a teoria das variáveis regionalizadas (Matheron 1971; Huijbregts 1973 e Oliver e Webster 1990). Neste contexto os valores dos atributos são representados como variáveis ou campos aleatórios, cujas funções de distribuição de probabilidade modelam a *incerteza*<sup>2</sup> dos valores assumidos. Estas incertezas devem ser consideradas como parte da representação do atributo e se propagam em análises realizadas no ambiente de modelagem computacional de um SIG.

A *krigeagem* é a âncora dos procedimentos geoestatísticos. Ela possibilita inferências de valores, a partir de amostras pontuais de um atributo espacial. Estas inferências são usadas, geralmente, para finalidades de mapeamento digital de atributos espaciais. Supondo-se a existência de um conjunto de n amostras pontuais,  $z(\mathbf{u}_{\alpha})$ , de valores de um atributo, dentro de uma região A da superfície terrestre,

$$z(\mathbf{u}_{\alpha}), \mathbf{u}_{\alpha} \in A \subset \mathcal{R}^2 e \alpha=1,2,...,n$$

<sup>&</sup>lt;sup>1</sup> O termo dado espacial é um dado com referência espaço-temporal relacionado à superfície terrestre.

o procedimento da krigeagem permite a estimação de valores do atributo em localizações  $\mathbf{u} \neq \mathbf{u}_{\alpha}$ . Além disso, este procedimento possibilita a inferência da variância de krigeagem, que, em certas condições, que serão discutidas neste trabalho, pode ser usada como uma medida de incerteza do atributo em  $\mathbf{u}$ .

O procedimento de inferência por krigeagem linear é um interpolador de média ponderada dos valores de um conjunto de amostras vizinhas. Os pesos das amostras são determinados a partir do estudo da variabilidade do atributo, na região de interesse, e respeitando *o critério de mínima variância do erro de estimação*. A representação da variabilidade baseia-se na hipótese de *estacionariedade de segunda ordem* do atributo, ou seja, requer que o valor médio do atributo seja constante e que a covariância entre dois valores do atributo, definidos em duas localizações  $\mathbf{u}_i$  e  $\mathbf{u}_j$ , dependa somente do vetor de separação,  $\mathbf{h} = \mathbf{u}_i - \mathbf{u}_j$ , dentro da região A. A partir desta hipótese, pode-se definir um modelo de variabilidade do atributo, um *variograma teórico* (ajustado sobre um variograma experimental que é obtido a partir do conjunto amostral). O variograma teórico é usado na determinação dos valores de covariância, entre amostras e entre cada amostra e o ponto a ser estimado, necessários para o cálculo dos ponderadores das amostras utilizadas numa inferência.

A krigeagem pode, também, ser utilizada para a construção de modelos probabilísticos de incerteza relacionados aos valores não amostrados. É esta propriedade o foco central deste trabalho. Os modelos probabilísticos são inferidos por um procedimento de krigeagem não linear chamado krigeagem por indicação (Goovaerts 1997; Deutsch e Journel 1998 e Felgueiras et al. 1999a). A krigeagem por indicação é, também, a base dos procedimentos de simulação estocástica condicionada que possibilitam a criação de campos aleatórios equiprováveis de atributos numéricos e temáticos. A formulação e inserção, no SIG, destes procedimentos geoestatísticos por indicação, para

\_

<sup>&</sup>lt;sup>2</sup> No contexto do trabalho, o termo *incerteza* é usado como uma medida indicativa da magnitude do erro, uma vez que não se pode determinar com exatidão o erro de estimação do atributo. Para isto seria necessário o conhecimento do valor real do atributo em cada posição estimada.

ampliar sua capacidade de modelagem de dados espaciais, é a essência da contribuição deste trabalho.

### 1.3 OBJETIVOS DO TRABALHO

## **Objetivo Geral:**

O objetivo principal deste trabalho é sistematizar uma metodologia para modelagem ambiental para SIG, baseada no paradigma geoestatístico por indicação, que considere informação de incerteza. Procedimentos de inferência e métricas de incerteza para atributos espaciais amostrados pontualmente são estudados e testados. Especificam-se os requisitos para sua integração ao ambiente de SIG, através de um módulo de análise espacial geoestatístico e através de seu uso em ferramentas de modelagem computacional. A finalidade maior é contribuir para que informação de qualidade possa ser incorporada nas representações e nos produtos gerados por modelagens computacionais desenvolvidas em ambiente de SIG.

## **Objetivos Específicos:**

- Explorar os procedimentos geoestatísticos por indicação, krigeagem e simulação estocástica, para inferência de valores, a partir de um conjunto de amostras pontuais, de atributos ambientais temáticos e numéricos.
- 2) Apresentar e discutir metodologias de cálculo de incertezas, locais e globais, baseadas nos modelos de distribuição probabilística das variáveis e campos aleatórios que modelam os atributos.
- Explorar o uso de informação indireta, relacionada ao atributo a ser modelado, para melhorar a acurácia das estimativas baseadas em krigeagem por indicação.

4) Discutir e especificar os requisitos básicos para integração dos procedimentos geoestatísticos em SIG através de um módulo de análise espacial geoestatístico e através de seu uso em ferramentas, módulos ou linguagens, de modelagem computacional.

## 1.4 CONTRIBUIÇÕES DA TESE

Vários autores, tais como McCullagh (1988); Oliver e Webster (1990); Camargo (1997) e Fuks (1998), têm se dedicado ao problema de estimativa de atributos espaciais numéricos a partir de um conjunto amostral pontual. As soluções apresentadas podem ser classificadas em dois grupos, as que tratam o problema com procedimentos determinísticos e aquelas que utilizam procedimentos geoestatísticos para as inferências.

As inferências para atributos espaciais por procedimentos geoestatísticos apresentam, pelo menos, duas vantagens principais em relação aos determinísticos: se baseiam no estudo da variabilidade do atributo dentro da região de interesse e possibilitam a determinação da variância da estimação. Heuvelink (1998), por exemplo, utiliza a variância do erro da krigeagem linear para modelar incerteza das estimativas de atributos numéricos. Entretanto, essa variância só pode ser considerada medida de incerteza em situações muito particulares, ou seja, quando a distribuição das variáveis aleatórias, VA, na região de interesse, tem um comportamento multigaussiano (Deutsch e Journel 1998) e quando o patamar do variograma teórico, obtido das amostras, representa exatamente o valor da variância do atributo, ou seja, não está escalonado (Isaaks e Srivastava 1989). Essas restrições são abordadas com mais detalhes nos Capítulos 2 e 3 deste trabalho.

Por outro lado, como já citado na Seção 1.2, os procedimentos de estimação por indicação se baseiam na construção dos modelos de distribuição de probabilidade das VA e FA, que representam os atributos em localizações não conhecidas. Essa metodologia é não paramétrica, no sentido de que não é determinada por parâmetros de uma distribuição padrão definida a priori, e

possibilita a obtenção de parâmetros estatísticos que definem os valores estimados e as incertezas relacionadas ao comportamento estocástico do atributo. Além disso, estes procedimentos podem ser aplicados a atributos temáticos, além dos numéricos, o que não é possível com krigeagem linear ou outro tipo de interpolador determinístico que considere mais de um vizinho num processo de inferência por médias. Uma outra vantagem importante, é a possibilidade de se utilizar informação amostral complementar, referida como dados indiretos neste texto, ao conjunto original de amostras do atributo. Os dados indiretos são incorporados aos procedimentos de inferência para melhorar a acurácia das estimativas.

Neste contexto, como contribuição principal, o presente trabalho propõe o uso de um *novo paradigma de modelagem computacional baseado na modelagem de atributos espaciais por procedimentos geoestatísticos por indicação*. Para isto, este trabalho investiga e explora o potencial dos procedimentos geoestatísticos por indicação no sentido de incorporar, ao estágio tecnológico atual dos SIG, estimadores mais confiáveis, informação de incerteza das estimativas e estimadores, baseados numa metodologia única, que possibilitem inferências para atributos numéricos e temáticos. O trabalho apresenta, nos Capítulos 3 4 e 5, as formalizações matemáticas destes procedimentos. Estudos de casos, envolvendo atributos numéricos e temáticos, com análises comparativas com outras metodologias são apresentados no Capítulo 6. Outras contribuições importantes são:

- Estudo e formalização matemática, apresentados no Capítulo 5, dos procedimentos geoestatísticos por indicação para uso de dados indiretos, como dados complementares às amostras diretas do atributo.
   Um estudo de caso com o atributo textura do solo e dados indiretos obtidos do mapa de solos é apresentado e analisado no Capítulo 6.
- Discussão e especificação dos requisitos básicos para integração de procedimentos geoestatísticos no SIG. Apresentam-se, no Capítulo 7, os

requisitos de procedimentos de um módulo de geoestatística integrado ao SIG e também os requisitos básicos de uma linguagem de análise e consulta com tratamento de incertezas.

## 1.5 ORGANIZAÇÃO DO TRABALHO

Este trabalho tem a seguinte organização:

O Capítulo 1 é uma introdução objetivando, principalmente, apresentar a motivações, objetivos e contribuições desta tese.

O Capítulo 2 apresenta e discute aspectos importantes relacionados à modelagem ambiental em SIG com tratamento de incertezas. Dá-se ênfase à modelagem de dados e à modelagem computacional considerando-se o tratamento das incertezas envolvidas neste processo. Reportam-se, ainda, alguns métodos clássicos de propagação de incerteza para modelagens com atributos numéricos. Finalmente, apresenta-se a proposta de uso do paradigma de modelagem geoestatística por indicação analisando-se as suas vantagens em relação às modelagens computacionais tradicionalmente utilizadas até o presente momento.

O Capítulo 3 conceitua e formaliza os procedimentos de inferência de krigeagem linear e de krigeagem por indicação. Este Capítulo apresenta, ainda, alternativas para o cálculo de incertezas locais associadas aos atributos, temáticos e numéricos, inferidos pelo método de krigeagem por indicação. Ao final, listam-se as vantagens e desvantagens do procedimento de krigeagem por indicação comparando-o com o de krigeagem linear.

O Capítulo 4 descreve o procedimento de simulação sequencial condicionada e as metodologias para se inferir estatísticas e incertezas a partir dos campos simulados. Compara-se, ainda, o procedimento de simulação sequencial por indicação com o procedimento de krigeagem por indicação.

O Capítulo 5 explora o potencial dos procedimentos geoestatísticos, krigeagem e simulação estocástica, para utilização de informação indireta em conjunto com os dados amostrais diretos de um atributo. O objetivo é melhorar a acurácia da estimação. Apresenta-se a codificação e formaliza-se o estimador de krigeagem por indicação para uso de dados diretos e indiretos. Em seguida, sugerem-se algumas otimizações que agilizam este procedimento, que é custoso computacionalmente.

O Capítulo 6 apresenta estudos de casos nos quais se aplicam os conceitos dos capítulos anteriores em atributos amostrais de problemas reais. São mostrados e analisados casos de espacialização<sup>3</sup>, por krigeagem e por simulação, do atributo numérico altimetria e do atributo temático textura do solo. Ilustra-se, também, o uso de dados indiretos, obtidos de um mapa de solos, para melhorar a acurácia de estimação da textura do solo. Todos os resultados apresentados são analisados e comparados com outros procedimentos e resultados.

O Capítulo 7 é dedicado ao tema de integração dos procedimentos geoestatísticos com os SIG. Apresentam-se os requisitos básicos para inclusão de um módulo de geoestatística em um SIG. Descrevem-se as funcionalidades incorporadas nos primeiros módulos disponíveis em alguns SIG do mercado. Finalmente, abordam-se as alterações que se fazem necessárias para se considerar as análises e incertezas, provenientes dos procedimentos da geoestatística, em ferramentas de modelagem computacional disponibilizadas no SIG.

No Capítulo 8 apresentam-se conclusões a respeito deste trabalho, com sugestões para futuros desenvolvimentos em relação aos assuntos aqui abordados.

<sup>&</sup>lt;sup>3</sup> A *espacialização* compreende a inferência de valores de um atributo para se obter uma representação computacional do atributo no espaço. Em ambientes de SIG é comum representar-se atributos com distribuições espaciais na forma matricial. Neste caso diz-se que o atributo foi espacializado segundo uma representação de grade regular.

### **CAPÍTULO 2**

# MODELAGEM AMBIENTAL EM SISTEMAS DE INFORMAÇÃO GEOGRÁFICA COM TRATAMENTO DE INCERTEZAS

#### 2.1 Introdução

Cada vez mais, os computadores vêm sendo utilizados como ferramentas de apoio a procedimentos de estudos, de análises e de simulações em vários campos do conhecimento humano. Sistemas complexos para análises e para modelagens foram desenvolvidos para se trabalhar dados relacionados com áreas específicas como finanças, transportes, geologia, solos, etc.. Nessa mesma tendência, sistemas de armazenamento, manipulação e apresentação de dados espaciais, conhecidos como SIG, foram criados e estão sendo utilizados no campo das ciências ambientais. Modelos matemáticos, aritméticos e lógicos, buscando representar propriedades e processos do meio físico natural, têm sido implementados, nos SIG, com o objetivo de facilitar o seu estudo e compreensão para que se possa atuar sobre o meio ambiente de forma responsável e cooperativa. Para que isto ocorra, este trabalho defende a idéia de que é imperativo que se considere o tratamento das incertezas das representações dos dados espaciais envolvidos em um modelo matemático ambiental apoiado em SIG. As incertezas destas representações, uma vez propagadas para os resultados finais das modelagens, qualificam os produtos gerados em SIG para o apoio efetivo aos processos de tomadas de decisão baseados nesses produtos.

Considerando-se esse contexto os objetivos deste capítulo são:

 apresentar conceitos importantes relacionados com a representação dos dados espaciais e com seu uso em modelagem ambiental produzida utilizando-se da tecnologia de SIG;

- tratar os temas de estimativa de incertezas para dados espaciais e de propagação destas incertezas em modelagens computacionais desenvolvidas em SIG e;
- apresentar, analisar e propor o uso do paradigma da modelagem geoestatística por indicação para modelagem computacional em SIG envolvendo dados espaciais com atributos temáticos e numéricos.

O presente capítulo tem a seguinte organização:

- na Seção 2.2 são apresentados os aspectos gerais da natureza da modelagem ambiental, tratados neste trabalho, os conceitos gerais sobre SIG, os detalhes da modelagem de dados espaciais, com e sem inclusão de incertezas, e as características importantes da modelagem computacional no ambiente de um SIG;
- na Seção 2.3 discute-se o problema da propagação de incertezas em modelagens computacionais, com a descrição de alguns modelos de propagação envolvendo atributos de natureza numérica.
- a Seção 2.4 apresenta um histórico evolutivo da modelagem computacional enfatizando-se a necessidade de se adotar o paradigma da modelagem geoestatística por indicação para o efetivo uso de modelagem computacional e;
- a Seção 2.5 encerra este capítulo tecendo algumas conclusões.

#### 2.2 MODELAGEM AMBIENTAL E SISTEMAS DE INFORMAÇÃO GEOGRÁFICA

A potencialidade principal de um SIG está na sua capacidade de realizar análises complexas a partir da integração, em uma base de dados única, de representações de dados espaciais. Como já definido, na Seção 1.2, um dado espacial é caracterizado por sua posição espaço-temporal, (x, y, z, t), e por atributos a ele associados. Um dado de temperatura, por exemplo, pode ser

observado em várias posições do espaço tridimensional (x, y, z) e em diferentes períodos de tempo t. As medidas de temperatura são os valores do atributo observados em posições desse sistema espaço-temporal.

Os procedimentos de análise espacial, desenvolvidos no ambiente de um SIG, possibilitam, no estágio tecnológico atual, a análise de processos, alguns simples e outros mais complexos, do mundo real. Para isto é necessário a criação de modelos ambientais, que representem adequadamente o fenômeno<sup>4</sup> natural em estudo. Assim, a *modelagem ambiental*, no contexto deste trabalho, consiste na criação de modelos matemáticos, determinísticos ou estocásticos, que relacionam atributos ambientais na tentativa de representar o comportamento de um processo ocorrendo na natureza. Os modelos ambientais são, então, transformados em *modelos computacionais* para serem executados no ambiente de um SIG.

## 2.2.1 A natureza da modelagem ambiental em meio digital

## 2.2.1.1 A dimensão espaço-temporal e a natureza contínua dos processos da natureza

As pessoas observam o mundo, diretamente ou com auxílio de instrumentos, e percebem fenômenos da natureza que se modificam com maior ou menor intensidade em função do tempo e do espaço (Burrough e McDonnell 1998), ou seja, apresentam uma *variação contínua espaço-temporal*. Segundo Steyaert (1993), os processos ambientais no mundo real são, tipicamente, tridimensionais, dependentes do tempo e complexos. Essa complexidade pode incluir comportamento não linear, componentes estocásticos e realimentações em múltiplas escalas de tempo e de espaço. Assim sendo, a natureza espaço temporal dos fenômenos ambientais, devido a sua importância para compreensão do fenômeno em estudo, coloca algumas questões à modelagem ambiental. Estas questões se referem a criação de modelos matemáticos e

<sup>-</sup>

<sup>&</sup>lt;sup>4</sup> Os termos fenômeno e processo são usados como sinônimos neste texto com o significado de sequência de estados de um sistema que se transforma.

computacionais com representações, uma vez que o fenômeno deve ser, de alguma forma, representado.

A modelagem deve considerar que os processos da natureza resultam de interações espaço-temporais complexas entre os diversos elementos que os compõem, ou seja, as propriedades ambientais. No modelo matemático de um processo, as propriedades ambientais são tratadas como variáveis do modelo enquanto que suas inter-relações são representadas por operações aritméticas ou lógicas.

A representação digital envolve procedimentos de discretizações e quantizações, nos domínios do espaço-tempo e do atributo, para se obter estruturas computacionalmente manipuláveis. Essas estruturas representam digitalmente o comportamento do atributo no espaço-tempo e são usadas diretamente pelos modelos computacionais.

#### 2.2.1.2 O mundo contínuo e o mundo discreto dos computadores digitais

Como já citado, os dados espaciais, assim como os processos ambientais, apresentam a característica de continuidade no espaço e no tempo. Por outro lado, atualmente, a tecnologia de computadores digitais vem sendo intensivamente utilizada para representar dados espaciais e simular processos ambientais com o objetivo de melhor compreendê-los para, então, atuar sobre o meio ambiente de forma cooperativa. Por exemplo, modelos de simulação ambiental fornecem diagnósticos e saídas preditivas que podem ser combinadas com dados sócio-econômicos para avaliação de riscos ambientais regionais e locais ou decisões relacionadas com o gerenciamento de recursos naturais (Steyaert 1993).

Entretanto, os computadores digitais trabalham com um número fixo de campos, ou bytes, para armazenar qualquer valor numérico. Isto vale para os valores das posições espaciais e também para os valores dos atributos relacionados aos dados espaciais. Assim, uma base de dados computacional

contém aproximações, ou representações simplificadas, dos dados ou fenômenos espaciais. Goodchild e Guoging (1992) argumentam que: "por que uma base de dados computacional é um armazenador discreto e finito, é necessário amostrar, abstrair, generalizar ou, até mesmo, comprimir informações do mundo real".

#### 2.2.1.3 Representações de continuidade a partir de dados amostrais

As observações de propriedades e processos da natureza contam, atualmente, com um conjunto cada vez mais numeroso de tecnologias. As imagens de sensoriamento remoto orbital, por exemplo, possibilitam acompanhar diversos processos ambientais, tais como, as modificações na paisagem por ações humanas de engenharia e agricultura, os aumentos de áreas desertificadas, erodidas ou reflorestadas, o tamanho e a direção de crescimento de áreas urbanas, a extensão de áreas inundadas ou queimadas, etc.. Equipamentos de medidas com grande grau de acurácia, tais como, os sistemas de posicionamento globais, GPS, as plataformas de coletas de dados meteorológicos e outros, são hoje, cada vez com mais frequência, utilizados para se obter amostras da distribuição de um atributo espacial dentro de uma área de interesse.

O computador digital não pode armazenar um atributo de forma contínua e, também, por razões de custo de aquisição, os atributos da natureza são representados digitalmente por *conjuntos de amostras* observadas, direta ou indiretamente, do atributo. A quantidade e a distribuição das amostras, que são coletadas, devem representar a variação do atributo dentro de uma região espacial de interesse. Valores de máximo, de mínimo e de inflexão relacionados aos atributos de um dado espacial são informações obrigatórias de um conjunto amostral.

Os dados amostrais podem ser pontuais, quando o valor de cada dado amostrado está relacionado exclusivamente a uma posição observada. Por exemplo, o valor da elevação, em relação ao nível do mar, em uma posição do

terreno, para levantamento topográfico de uma região, só é válido para essa localização. Por outro lado, um dado amostral por área tem um valor do atributo que é válido para uma área predeterminada. Um pixel de uma imagem de sensoriamento remoto representa uma amostra de uma região do espaço definida pela resolução espacial do sensor. Este trabalho supõe a existência de amostragens pontuais para os atributos a serem modelados por procedimentos geoestatísticos de krigeagem e simulação.

#### 2.2.2 Modelagem de dados espaciais em SIG

Existe uma grande quantidade de definições para SIG, cada uma delas baseada no tipo de usuário e no domínio da aplicação (Maguirre et al. 1991). A metodologia de banco de dados define o SIG como um banco de dados geográficos não convencional que possibilita gerenciamento de dados espaciais. A visão orientada a processos considera o SIG como uma coleção de subsistemas integrados, onde dados passam por uma sequência de procedimentos de transformação. A definição de aplicação ou utilização conceitua o SIG de acordo com o tipo de problema a ser solucionado e o tipo de dado manipulado (De Oliveira et al. 1997). Burrough e McDonnell (1998) definem o SIG como "um conjunto poderoso de ferramentas para coleta, armazenamento, recuperação, transformação e apresentação de dados espaciais do mundo real para um conjunto particular de propósitos". Esta definição considera o SIG como uma caixa de ferramentas computacionais para se trabalhar com dados espaciais. No contexto deste trabalho, é esta a definição que será enfocada. Segundo Câmara e Medeiros (1998), os SIG, como instrumentos computacionais para geoprocessamento, possibilitam a realização de análises complexas ao integrar dados de diversas fontes e ao criar banco de dados georreferenciados.

Um dado espacial, para ser trabalhado dentro de um SIG, deve ser representado por um modelo computacional. Um modelo é uma abstração de fatos ou de entidades do mundo real. A modelagem de dados geográficos é o

processo de discretização que converte a realidade geográfica complexa em um número finito de registros ou objetos (Goodchild 1993a).

Num nível de abstração mais alto, ou *nível conceitual*, o mundo real é representado segundo duas visões complementares: *o modelo de campos* e o *modelo de objetos* (Goodchild 1993a; Câmara e Medeiros 1998). É este o modelo que será adotado neste trabalho.

O modelo de campos representa os dados espaciais cujos atributos têm distribuição contínua no espaço. Um campo é formalizado como uma função matemática cujo domínio é uma região geográfica e cujo contradomínio é o conjunto de valores que o campo pode tomar (Câmara et al. 1996b). Os campos podem ser especializados em *campos numéricos*, quando o contradomínio pode conter um número infinito de valores numéricos e *campos temáticos*, quando o conjunto de valores do contradomínio é um conjunto limitado de classes, ou temas. Dados de elevação e temperatura são exemplos de campos numéricos, enquanto que classes de solos e classes de vegetação são exemplos típicos de campos temáticos.

O modelo de objetos é utilizado, tipicamente, para representar facilidades construídas pelo homem. Este modelo considera que o mundo está povoado por um conjunto de objetos, ou entidades. Os objetos têm identificações únicas e são caracterizados por suas propriedades geométricas e topológicas e por valores de atributos não espaciais. Neste modelo, uma posição espacial pode, até mesmo, ser ocupada por um ou mais objetos. Tubos de água ou de esgoto, prédios, ginásios esportivos e postes de eletricidade são entidades do mundo real modeladas como objetos.

Este trabalho vai lidar com propriedades ambientais cujas variáveis podem ser representadas no modelo de campos. A informação inicial se constitui de dados de propriedades ambientais coletados em forma de amostras pontuais georeferenciadas no espaço.

Amostras de dados espaciais, modelados como campos, podem ser espacializadas utilizando-se algoritmos de inferência ou interpolação. O procedimento de espacialização consiste em inferir valores do atributo, em posições não amostradas, a partir dos valores observados. Dessa forma, podese obter o valor do atributo em qualquer posição do espaço, ainda que a representação por campo seja discreta. É muito comum, no ambiente de um SIG, a criação de estruturas de representação por grades regulares retangulares onde os valores do atributo nos vértices da grade são obtidos por procedimentos de interpolação local a partir do conjunto de amostras.

Os algoritmos para inferência de atributos de dados espaciais, representados por amostras pontuais, podem ser classificados em dois (2) tipos básicos: interpoladores determinísticos e interpoladores estocásticos.

Os algoritmos de interpolação determinísticos mais utilizados na prática são os de média móvel ponderada (Burrough 1987; Felgueiras 1987; McCullagh 1988), que têm a seguinte formulação geral:

$$z^{*}(\mathbf{u}) = \frac{\sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}(\mathbf{u}) z(\mathbf{u}_{\alpha})}{\sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}(\mathbf{u})} e \lambda_{\alpha}(\mathbf{u}) = \frac{1}{d_{\alpha}(\mathbf{u})}$$
(2.1)

onde  $z^*(\mathbf{u})$  é o valor do atributo inferido numa posição  $\mathbf{u}$  do espaço,  $z(\mathbf{u}_{\alpha})$  o valor do atributo na amostra vizinha  $\alpha$ , onde  $\alpha$ =1,..., $n(\mathbf{u})$  com  $n(\mathbf{u})$  igual ao número total de amostras vizinhas de  $\mathbf{u}$ , e  $\lambda_{\alpha}(\mathbf{u})$  é o valor de ponderação relativo à amostra  $\alpha$  e à posição  $\mathbf{u}$ . Em geral, o valor de ponderação é definido igual ao inverso, simples ou ao quadrado, da distância euclidiana,  $d_{\alpha}(\mathbf{u})$ , da amostra  $\alpha$  à posição  $\mathbf{u}$ . Estes interpoladores são denominados locais pois utilizam um subconjunto das amostras na vizinhança de cada ponto a ser interpolado. Variações deste algoritmo também são usadas, tais como: interpolação por vizinho mais próximo, quando o número de vizinhos é igual a

1; interpolação por média simples, quando os pesos de interpolação são todos iguais a 1; interpolação por quadrantes, quando as amostras vizinhas são escolhidas por quadrantes no sentido de resolver problemas causados por amostras aglomeradas e; interpolação por cotas, quando as amostras estão representadas por curvas de nível e se deseja evitar tendenciosidade em relação à uma das curvas.

Os valores inferidos pelos interpoladores determinísticos são tratados como dados sem erros, ou seja, os valores obtidos não estão contaminados por erros nos dados de entrada ou pelo algoritmo de interpolação. Esta é a grande desvantagem dos interpoladores determinísticos em relação aos estocásticos.

Os interpoladores estocásticos utilizam as ferramentas da geoestatística para inferências de valores de atributos com estimativas de incertezas. A geoestatística considera que a distribuição espacial do atributo define uma função aleatória, FA, dentro de uma região de interesse A (Delfiner e Delhomme 1975; Isaaks e Srivastava 1989; Journel 1988; Cressie 1991; Goovaerts 1997; Deutsch e Journel, 1998). O valor z do atributo em estudo, em uma localização qualquer de A, é considerado uma realização de uma VA Z, cuja distribuição de probabilidade modela a incerteza local a respeito de z.

Os procedimentos inferenciais da geoestatística, conhecidos como métodos de *krigeagem*, baseiam-se na análise e na modelagem da variabilidade espacial do atributo a partir de um conjunto amostral pontual desse atributo. Supõem, ainda, a hipótese de estacionariedade de segunda ordem para a propriedade que está sendo modelada, ou seja, a média é constante, em todas as posições do campo, e a covariância só depende da distância entre as amostras. *A krigeagem ordinária* é o estimador estocástico mais utilizado na prática e tem a seguinte formulação:

$$z_{o}^{*}(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{o\alpha}(\mathbf{u}) \cdot z(\mathbf{u}_{\alpha})$$
(2.2)

onde  $z_0^*(\mathbf{u})$  é o valor do atributo inferido, por krigeagem ordinária, numa posição  $\mathbf{u}$  do espaço,  $z(\mathbf{u}_\alpha)$  o valor do atributo na amostra vizinha  $\alpha$ , onde  $\alpha=1,...,n(\mathbf{u})$  e  $n(\mathbf{u})$  é o número total de amostras vizinhas de  $\mathbf{u}$ , e  $\lambda_{0\alpha}(\mathbf{u})$  é o valor de ponderação da krigeagem ordinária relativo à amostra  $\alpha$  e à posição  $\mathbf{u}$ . A soma dos valores de ponderação deve ser igual a 1, para evitar tendenciosidade do estimador. Os pesos da krigeagem, diferentemente dos pesos do interpolador por média móvel, são obtidos a partir da hipótese de minimização da variância do erro de estimação e a partir do estudo de variabilidade do atributo, ou, mais formalmente, a partir de modelos de covariância inferidos sobre o conjunto amostral do atributo. Além disso, os procedimentos de krigeagem possibilitam a estimativa de incertezas relacionadas aos valores inferidos para o atributo.

Um valor observado, ou sorteado, é conhecido como uma *realização* de uma variável aleatória. A *simulação estocástica*, outro procedimento geoestatístico, se utiliza da krigeagem para obtenção de realizações de variáveis e campos aleatórios. Os valores obtidos, em várias realizações de uma mesma VA, possibilitam inferências de valores e estimativas de incertezas para o atributo.

O Capítulo 3, deste trabalho, é dedicado a formalização dos procedimentos de krigeagem linear e de krigeagem por indicação, enquanto que o Capítulo 4 formaliza os procedimentos de simulação estocástica condicionada e descreve as principais diferenças destes com os procedimentos de krigeagem. Esses capítulos mostram, ainda, que os procedimentos geoestatísticos, baseados em krigeagem por indicação, possibilitam inferências para atributos temáticos, além de atributos numéricos, o que não pode ser realizado com estimadores determinísticos ou com a krigeagem linear.

## 2.2.3 Modelagem computacional para modelos ambientais em SIG

Modelos ambientais, aqui também chamados modelos matemáticos, são representações matemáticas criadas para representar fenômenos ou processos do mundo real. Estes modelos são simplificações da realidade, de

onde se abstraem os elementos mais importantes para uma aplicação, e são construídos a partir da observação dos dados espaciais e seus relacionamentos. Os modelos ambientais são usados para aumentar o conhecimento sobre um processo, predizer valores ou comportamentos em áreas não observadas e comprovar, ou não, hipóteses feitas sobre processos. Estes modelos variam de equações empíricas simples, tais como, equações de regressão linear, até conjuntos de equações diferenciais complexas derivadas dos fundamentos da física (Moore et al. 1993)

Modelagem computacional em SIG é a implementação de um modelo matemático, que representa um fenômeno natural, no contexto de um Sistema de Informação Geográfica. Segundo Heuvelink (1998), os modelos espaciais podem ser classificados em 3 tipos, lógicos, empíricos e conceituais.

Os *modelos lógicos* computam um atributo de saída, resultado do modelo, pela aplicação de regras lógicas simples sobre os atributos de entrada. Por exemplo, um mapa de risco de erosão pode ser obtido pelo cruzamento dos dados de declividade, cobertura vegetal e tipos de solo.

Os *modelos empíricos* baseiam-se em experiências ou em conhecimentos obtidos por percepção dos fenômenos ambientais. Em geral são formulados por regressões cujos coeficientes são definidos experimentalmente e, via de regra, se aplicam somente às áreas usadas para derivá-los. Burrough e McDonnell (1998) apresenta o relacionamento empírico de dois modelos de perdas de solo por erosão: o modelo "Universal Soil Loss Equation" (USLE) e o modelo "Soil Loss Estimation Model for Southern Africa" (SLEMSA). O modelo USLE, por exemplo, depende de alguns atributos, tais como, o comprimento de rampa L e a declividade S que, por sua vez, são, também, derivados de regressões empíricas.

Os *modelos conceituais*, também conhecidos como *modelos físicos*, são concebidos a partir do entendimento dos processos físicos do fenômeno que está sendo modelado. Têm aplicações mais gerais e seus coeficientes referem-

se às propriedades físicas, já comprovadas ou aceitas, do mundo real. A maioria dos modelos atmosféricos e hidrológicos são construídos como modelos conceituais. Lee et al. (1993) descreve modelos atmosféricos baseados nas leis da conservação da física. Maidment (1993) apresenta aspectos importantes relacionados com a modelagem hidrológica em ambiente de SIG.

Na prática, muitos dos modelos usados nas ciências ambientais contém componentes empíricas e conceituais (Heuvelink 1998), como exemplos típicos podem ser citados os modelos de erosão de solo e os modelos de aptidão agrícola. Além disso, Burrough et al. (1996) argumenta que muitos modelos físicos e empíricos são lineares porque modelos lineares são fáceis para manipular computacionalmente e eles têm um comportamento previsível quando possuem realimentações.

Numa formulação geral, a modelagem computacional estima o valor de um atributo de saída Y, para uma determinada posição, ou região, a partir dos valores de N atributos, ou variáveis, de entradas,  $Z_1$ ,  $Z_2$ ,..., $Z_N$ , ou seja:

$$Y = g(Z_1, Z_2, \dots, Z_N)$$
 (2.3)

Algumas considerações podem ser feitas em relação à esta abordagem:

- Os valores dos atributos fontes (z<sub>1</sub>,...,z<sub>N</sub>) não são exatos, ou seja,
   z<sub>i</sub>=z<sub>im</sub>+ε onde z<sub>im</sub> é o valor médio do atributo de entrada e ε é uma componente de erro aleatório com média zero;
- Os valores dos atributos z<sub>i</sub>'s podem ser independentes ou apresentar dependência entre eles;
- A função g é uma relação aritmética quando a natureza dos atributos dos dados dos mapas fontes for quantitativa, dados numéricos;
- A função g é uma relação lógica quando a natureza dos atributos dos mapas fontes for qualitativa, dados temáticos;

 A função g pode ser uma composição de relações aritméticas e lógicas para aplicações complexas envolvendo atributos numéricos e temáticos.

Atualmente, modelos computacionais simples são executados diretamente nos SIG através de operações básicas, lógicas e aritméticas, contidas nos seus módulos de análise ou de álgebra de dados espaciais. Modelos complexos são, muitas vezes, executados fora do ambiente do SIG, por sistemas de modelagem específicos. Nestes casos, os SIG são usados como base de armazenamento de dados espaciais e também como ferramentas de visualização para os dados de entrada e de saída dos modelos. Aqui, coloca-se um dilema entre se dotar o SIG de todas as potencialidades de análises ou criar interfaces inteligentes entre o SIG e os sistemas específicos de modelagem.

Outra questão importante se refere à inclusão de informação de incertezas nos modelos computacionais. Como já apresentado neste capítulo, espacialização de um atributo pode ser feita por algoritmos de inferência determinísticos e estocásticos. Os modelos computacionais estocásticos, diferentemente dos determinísticos, consideram as incertezas associadas aos dados e aos métodos envolvidos no modelo. Assim, quando um modelo computacional integra dados espaciais, modelados segundo o paradigma estocástico, as incertezas, atribuídas aos dados de entrada, se propagam para o dado de saída. A estimativa e a propagação de incertezas são temas ainda pouco explorados e usados, porém são de extrema relevância, uma vez que objetivam adicionar informação de qualidade aos dados e produtos disponibilizados pela tecnologia dos SIG.

#### 2.3 Propagação de incertezas em modelagem ambiental com SIG

A propagação de incertezas nos atributos espaciais, em modelagem ambiental com SIG, tem sido abordada com frequência na literatura. Trabalhos como os de Arbia (1993); Burrough (1992); Goodchild e Guoging (1992); Goodchild (1993b); Heuvelink e Stein (1989); Heuvelink et al. (1993) e Heuvelink (1998)

são exemplos reais dessa preocupação. Entretanto, segundo Heuvelink (1996), desenvolvimentos teóricos determinam que é relativamente direto avaliar como incertezas de atributos são propagadas através de modelagem espacial quantitativa em SIG. Mais recentemente, Heuvelink (1998) apresentou os fundamentos teóricos necessários para a solução dos problemas de propagação de incerteza em modelagens espaciais considerando que:

- os dados espaciais estão representados por campos com atributos de natureza contínua, ou seja, modelados como campos numéricos, e cada campo, de entrada ou saída, é modelado como uma função aleatória e;
- a modelagem espacial envolve classes de operações locais, pontuais ou de vizinhança, conforme ilustrado na Figura 2.1.

Segundo esta abordagem, os valores de saída  $y(\mathbf{u})$ , para cada posição  $\mathbf{u}$ , são derivados dos N atributos de entrada, segundo uma relação do tipo:

$$Y(\mathbf{u}) = g(Z_1(\mathbf{u}), Z_2(\mathbf{u}) ..., Z_N(\mathbf{u}))$$
(2.4)

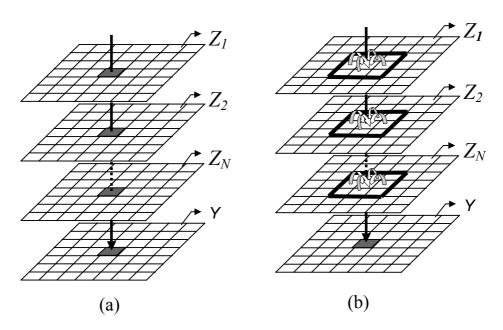


Fig. 2.1 - Operações locais: (a) pontuais e (b) de vizinhança em modelagem computacional.

Fonte: Baseada em Heuvelink (1998).

Supondo-se conhecidos os campos de incerteza de cada uma das representações das entrada, pode-se, também, obter um campo de incertezas do campo resultante Y. Isto é feito por procedimentos de propagação de incerteza, que consideram os tipos de dados, temáticos ou numéricos, e os tipos de operações, lógicas ou aritméticas, e seus parâmetros, envolvidos nas relações definidas sobre os dados de entrada. Heuvelink (1998) apresenta o formalismo de quatro métodos de propagação de incertezas para campos numéricos. O autor assume que cada campo de incerteza, dos atributos de entrada, é modelado por intervalos de confiança, (Isaaks e Srivastava 1989), definidos pelas variâncias do erro de estimação obtidas por procedimentos de krigeagem linear. As principais características dos métodos de propagação de incerteza, formalizados por Heuvelink (1998), são apresentadas e analisadas, nas quatro seções que seguem, segundo restrições de aplicação e vantagens.

## 2.3.1 Método de Taylor de primeira ordem

A expansão da série de Taylor de primeira ordem, em torno do vetor de médias,  $\vec{\mu}_z = (\mu_{z_1}, ..., \mu_{z_n})$ , de N variáveis aleatórias  $Z_i$ , i = 1,...,N, é dada por:

$$Y = g(\vec{\mu}_z) + \sum_{i=1}^{N} \left\{ \left( Z_i - \mu_{Z_i} \right) \cdot \left( \frac{\partial g}{\partial Z_i} (\vec{\mu}_z) \right) \right\} + \text{resíduo}$$
 (2.5)

A partir desta expansão, e desconsiderando-se o resíduo, obtêm-se as seguintes aproximações para média  $\mu_Y$  e variância  $\sigma_y^2$  da VA Y:

$$\mu_{Y} = E(Y) \approx g(\vec{\mu}_{Z})$$
 (2.6)

$$\boldsymbol{\sigma}_{Y}^{2} \approx \sum_{i=1}^{N} \sum_{j=1}^{N} \left[ \frac{\partial \boldsymbol{g}}{\partial z_{i}} (\boldsymbol{\mu}_{Z}) \cdot \frac{\partial \boldsymbol{g}}{\partial z_{j}} (\boldsymbol{\mu}_{Z}) \cdot \boldsymbol{\sigma}_{Z_{i}} \cdot \boldsymbol{\sigma}_{Z_{j}} \cdot \rho_{ij} \right]$$
(2.7)

onde  $\sigma_{Z_k}$  é o desvio padrão da k-ésima VA de entrada, k =1,..,N, e  $\rho_{ij}$  é o coeficiente de correlação entre as VA  $Z_i$  e  $Z_i$ .

As seguintes considerações, sobre essa metodologia, podem ser ressaltadas:

- o método aplica-se somente para operações aritméticas, Z<sub>i</sub>'s numéricos;
- o valor médio de Y depende apenas dos valores médios das entradas;
- a variância da saída depende dos desvios padrão das variáveis de entrada, dos coeficientes de correlação entre elas e, também, das derivadas parciais de g em relação às entradas. Isto implica que este método só é aplicável para funções g continuamente diferenciáveis e;
- quando as variáveis de entrada são independentes, o coeficiente de correlação, ρ<sub>ij</sub>, é igual a 0 para i≠j, e é igual a 1 para i=j. Neste caso, e somente neste caso, a formulação da variância é simplificada para:

$$\boldsymbol{\sigma}_{Y}^{2} = \sum_{i=1}^{N} \left[ \frac{\partial \boldsymbol{g}}{\partial z_{i}} (\overrightarrow{\boldsymbol{\mu}}_{Z}) \right]^{2} \cdot \boldsymbol{\sigma}_{Z_{i}}^{2}$$
(2.8)

## 2.3.2 Método de Taylor de segunda ordem

O método de Taylor de segunda ordem é uma extensão do método de Taylor de primeira ordem, incluindo-se o termo de segunda ordem na expansão, ou seja:

$$Y = \mathbf{g}(\vec{\boldsymbol{\mu}}_{z}) + \sum_{i=1}^{N} \left\{ \left( Z_{i} - \boldsymbol{\mu}_{z_{i}} \right) \cdot \left( \frac{\partial \mathbf{g}}{\partial z_{i}} (\vec{\boldsymbol{\mu}}_{z}) \right) \right\}$$

$$+ \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left\{ \left( Z_{i} - \boldsymbol{\mu}_{z_{i}} \right) \cdot \left[ Z_{j} - \boldsymbol{\mu}_{z_{j}} \right] \cdot \left( \frac{\partial^{2} \mathbf{g}}{\partial z_{i} \delta z_{j}} (\vec{\boldsymbol{\mu}}_{z}) \right) \right\} + \text{resíduo}$$

$$(2.9)$$

Neste caso, a aproximação da média  $\mu_Y$  é dada por:

$$\left| \boldsymbol{\mu}_{\boldsymbol{Y}} = E(\boldsymbol{Y}) \approx \boldsymbol{g}(\vec{\boldsymbol{\mu}}_{\boldsymbol{Z}}) + \frac{1}{2} \sum_{i=1}^{N} \sum_{j=1}^{N} \left\{ \boldsymbol{\rho}_{i,j} \boldsymbol{\sigma}_{z_i} \boldsymbol{\sigma}_{z_j} \frac{\partial^2 \boldsymbol{g}}{\partial_{z_i} \partial_{z_j}} (\vec{\boldsymbol{\mu}}_{\boldsymbol{Z}}) \right\} \right|$$
(2.10)

Contrário ao que se obteve com o método de Taylor de  $1^{\underline{a}}$  ordem, a média  $\mu_{\gamma}$ , da equação 2.10, pode diferir do valor de g aplicado às médias das entradas. Além disso, a estimativa da variância, para o método de Taylor de segunda ordem, tem uma formulação mais complexa, pois requer o cálculo do  $1^{\underline{o}}$ ,  $2^{\underline{o}}$ ,  $3^{\underline{o}}$  e  $4^{\underline{o}}$  momentos das entradas e, também, derivadas parciais de primeira e segunda ordens de g em relação às entradas (Heuvelink 1998). Assim, o método só é aplicável para g continuamente diferenciáveis até segunda ordem.

#### 2.3.3 Método de Rosenblueth

O método de Rosenblueth é equivalente ao de Taylor de  $1^{\underline{a}}$  ordem, podendo ser usado também nos casos em que a função g não é continuamente diferenciável. Como no método de Taylor de  $1^{\underline{a}}$  ordem, os valores de média e a variância da VA Y de saída são estimados a partir da aplicação da função g e do  $1^{\underline{a}}$  e  $2^{\underline{a}}$  momentos das VA  $Z_i$  de entrada, (Heuvelink 1998).

#### 2.3.4 Método de Monte Carlo

A idéia principal do método de simulação de Monte Carlo é obter realizações da VA de saída Y a partir de realizações, das VA de entrada Z/s, obtidas a partir da distribuição conjunta destas variáveis. Os valores realizados de Y são calculados pela aplicação da função g sobre os valores realizados das entradas. As estatísticas da saída, como média e variância, são obtidas a partir dos valores realizados de Y. Considerando-se operações pontuais com N entradas, o método de Monte Carlo consiste dos seguintes passos:

```
Início

Para cada localização \mathbf{u} \in A

Início

Repita \mathbf{k} vezes:

Início

Obtenha um conjunto de realizações z_i(\mathbf{u}), i=1,...,N.

Calcule e armazene realizações de Y, y(\mathbf{u}) = \mathbf{g}(z_I(\mathbf{u}),...z_N(\mathbf{u}))

Fim

Calcule estatísticas da distribuição das \mathbf{k} saídas y_i de Y em \mathbf{u}

Fim

Fim
```

As estatísticas de média e variância de Y são calculadas, em u, por:

A principal vantagem do método de Monte Carlo é que ele é de aplicação geral. Pode ser aplicado a modelos lógicos, conceituais ou empíricos. O método considera a função **g** como uma caixa preta cujas perturbações nos dados de entrada são estudadas a partir do resultado da saída.

Por outro lado, como os resultados não são obtidos numa forma analítica, análises de sensibilidade do modelo são mais difíceis, uma vez que a simulação deve ser executada várias vezes. Isto requer recursos computacionais com processamento rápido e possibilidade de armazenamento de grandes volumes de dados. O número de vezes que o algoritmo é repetido é, usualmente, maior que uma centena mas pode chegar a centenas de milhões.

#### 2.4 O PARADIGMA DA MODELAGEM ESTOCÁSTICA POR INDICAÇÃO

Em função de um melhor entendimento do contexto deste trabalho, a modelagem computacional em ambiente de SIG será classificada segundo 3 (três) etapas de evolução distintas, a modelagem determinística, a modelagem estocástica<sup>5</sup> linear e a modelagem estocástica, não linear, por indicação.

A modelagem ambiental determinística se caracteriza pela modelagem dos dados espaciais segundo o paradigma determinístico. Essa modelagem é praticada desde os primeiros modelos computacionais implementados em um SIG. Neste tipo de modelagem, os modelos de dados espaciais não contém componente aleatória e são, portanto, considerados exatos. Consequentemente, os produtos da modelagem computacional não contém informação de qualidade. A qualidade dos resultados é neglicenciada ou é

54

<sup>&</sup>lt;sup>5</sup> O termo *modelagem estocástica* é utilizado, neste texto, como referência exclusiva à modelagem que se utiliza de procedimentos geoestatísticos para representação de atributos espaciais.

inferida subjetivamente pelos usuários do sistema. Este tipo de modelagem ainda é praticado nos tempos atuais, mas obviamente é pobre e tende a ser substituído por modelos mais representativos.

A modelagem estocástica linear é mais recente e se caracteriza pela modelagem dos atributos espaciais segundo o paradigma geoestatístico linear, na qual os atributos espaciais são considerados variáveis aleatórias e são tratados pela teoria das variáveis regionalizadas. A modelagem dos atributos se baseia nos estimadores de krigeagem linear, principalmente a krigeagem ordinária. Neste modelo, o valor estimado e a variância do erro de estimação são considerados como parâmetros de média e de variância, respectivamente, de um modelo de distribuição gaussiana associado ao atributo. Assim, a variância do erro de estimação, ou o desvio padrão correspondente, é usada como unidade de medida de incerteza da modelagem do atributo. Também, uma combinação da média e da variância do erro de estimação é utilizada para definição de intervalos de confiança. Além disso, nesta etapa da modelagem são introduzidas, formalizadas e implementadas teorias de propagação de incertezas que possibilitam qualificar os produtos da modelagem. O trabalho atualmente mais expressivo e completo desta etapa é sistematizado e apresentado por Heuvelink (1998). Porém, existem problemas com este tipo de modelagem. Primeiramente, a variância de krigeagem linear está condicionada ao tipo de estimador utilizado e não é dependente dos valores dos atributos (Isaaks e Srivastava, 1989). Além disso, a hipótese de distribuição paramétrica multigaussiana para o atributo deve ser verificada, o que é praticamente impossível (Deutsch e Journel 1998). No Capítulo 3 será apresentado uma análise crítica do uso dessa variância como medida de incerteza. Apesar destas restrições, essa medida de incerteza vem sendo usada, com frequência, por vários usuários de SIG, por falta de conhecimento ou de outra alternativa. O segundo problema é o fato de a krigeagem linear permitir a modelagem espacial apenas para atributos numéricos. Assim, o problema de atribuição de incerteza para dados temáticos não é resolvido e, portanto, também não se resolve o problema de propagação de incertezas para modelos computacionais lógicos que envolvam variáveis temáticas.

Do conhecimento do autor, a modelagem estocástica por indicação é mais recente e não havia sido, até então, sistematizada para uso em modelagem computacional desenvolvida em SIG. Esta tese é o primeiro esforço neste sentido. Este tipo de modelagem computacional se caracteriza por utilizar os procedimentos não lineares da geoestatística, a krigeagem por indicação e a simulação estocástica por indicação, para modelar a variabilidade dos atributos espaciais. Estes procedimentos possibilitam a inferência de uma aproximação discretizada do modelo de distribuição de probabilidade do atributo que é, então, utilizada para modelagem da incerteza sobre seus valores. Assim, temse uma modelagem espacial não paramétrica que pode, portanto, ser usada sem restrições ao tipo de distribuição do atributo. Os modelos de incerteza são obtidos diretamente da distribuição, independem de um estimador escolhido e estão relacionados ao comportamento de variabilidade do atributo. Metodologias de estimativa de incertezas para procedimentos geoestatísticos por indicação são apresentados e formalizados nos Capítulos 3, 4 e 5 deste trabalho. Outra vantagem importante destes procedimentos é a possibilidade de se modelar dados temáticos, além dos dados de natureza numérica. Assim, pode-se trabalhar com propagação de incertezas para modelos computacionais que envolvam atributos numéricos e temáticos.

#### 2.5 CONCLUSÕES

O presente capítulo apresentou uma introdução aos conceitos de modelagem computacional, no ambiente de um SIG, para modelagem de processos ambientais. Ressaltou-se que, a modelagem dos dados espaciais, envolvidos em um modelo computacional, pode ser feita segundo paradigmas determinístico ou estocástico.

O paradigma estocástico, que utiliza procedimentos da geoestatística para espacialização de atributos de natureza ambiental, possibilita a estimação das

incertezas associadas a esses atributos. Este capítulo descreve e propõe o uso de uma formulação estocástica mais rigorosa para modelar incerteza de atributos numéricos e temáticos, a modelagem estocástica por indicação. As incertezas inferidas por esta modelagem são utilizadas por metodologias de propagação de incerteza que devem, necessariamente, ser incluídas no processo da modelagem ambiental.

A incerteza propagada para o resultado de uma modelagem computacional fornece informação quantitativa sobre o produto obtido. A qualidade do produto gerado é utilizada para avaliação da qualidade do modelo, para análises de sensibilidade do modelo aos dados de entrada e para a quantificação dos riscos assumidos ao se tomar decisões apoiadas nesse produto.

Entretanto, apesar da sua importância, as informações de incerteza ainda são pouco, ou quase nunca, consideradas nas aplicações desenvolvidas no ambiente de SIG. Isto ocorre, principalmente, devido à falta de ferramentas disponibilizadas nos SIG atualmente em uso. Este trabalho é uma contribuição para a mudança desse cenário, uma vez que, apresenta, analisa e propõe o uso de novas ferramentas e de novos procedimentos em modelagens computacionais implementadas em SIG.

.

#### **CAPÍTULO 3**

## INFERÊNCIAS COM ESTIMATIVA DE INCERTEZAS POR KRIGEAGEM

#### 3.1 Introdução

Nos capítulos anteriores, destacou-se a necessidade da associação de informação de incertezas aos modelos de dados utilizados no SIG. Enfatizou-se, ainda, que a geoestatística fornece ferramentas de inferências baseadas na hipótese de representação dos atributos espaciais como variáveis ou funções aleatórias, cujos funções de distribuição de probabilidade possibilitam a modelagem da incerteza em relação a um valor atribuído a uma variável ou a um campo aleatório. Mais especificamente, os algoritmos geoestatísticos de inferência, conhecidos como *krigeagem*, permitem a construção de modelos de distribuição de variáveis aleatórias que são usados para inferências de valores de atributos, numéricos e temáticos, e para estimativas de incertezas associadas a esses valores. Neste contexto, o presente capítulo tem por objetivos principais:

- apresentar, formalmente, os procedimentos geoestatísticos de krigeagem linear e não linear, por indicação, comparando-os e explorando suas vantagens e desvantagens em relação ao uso dos mesmos;
- apresentar e discutir metodologias para estimativa de incertezas locais baseadas nos modelos de distribuição probabilística das variáveis aleatórias que modelam os atributos temáticos e numéricos;

O capítulo está organizado da seguinte forma: a Seção 3.2 apresenta os aspectos conceituais importantes para a formalização dos estimadores de krigeagem linear; a Seção 3.3 faz uma análise crítica do uso da variância de krigeagem linear como medida de incerteza do atributo espacial em análise; a Seção 3.4 apresenta a krigeagem não linear por indicação, ou *krigeagem por indicação*, para a inferência direta de valores da função de distribuição

probabilística de atributos numéricos e temáticos; na Seção 3.4 explora-se, ainda, alternativas de estimativas de incertezas utilizando os modelos probabilísticos inferidos; a Seção 3.5 descreve e analisa as principais vantagens e desvantagens do uso da krigeagem por indicação e; finalmente, na Seção 3.6, apresentam-se as principais conclusões deste capítulo.

#### 3.2 CONCEITOS GERAIS DA GEOESTATÍSTICA

A geoestatística modela os valores de um atributo, dentro de uma região da superfície terrestre  $A \subset \Re^2$ , como uma função aleatória. Para cada posição  $\mathbf{u} \in A$  o valor do atributo de um dado espacial é modelado como uma VA  $Z(\mathbf{u})$ . Isto significa que, na posição  $\mathbf{u}$ , a VA  $Z(\mathbf{u})$  pode assumir diferentes valores para o atributo, cada valor com uma probabilidade de ocorrência associada. Nas n posições amostradas,  $\mathbf{u}_{\alpha}$  com  $\alpha$  =1,2,...,n, os valores  $z(\mathbf{u}_{\alpha})$  são considerados determinísticos, ou ainda, podem ser considerados VA cujo valor medido tem uma probabilidade de 100% de ocorrência. Uma VA é caracterizada pela sua função de distribuição acumulada, fda, univariada,  $F(\mathbf{u}; z)$ , definida como:

$$F(\mathbf{u};z) = Prob\{Z(\mathbf{u}) \le z\}$$
(3.1)

A função de distribuição acumulada condicionada, fdac, univariada de  $Z(\mathbf{u})$ , condicionada aos n dados amostrados,  $F(\mathbf{u}; z|(n))$ , é dada por:

$$F(\mathbf{u}; z \mid (n)) = Prob\{\mathbf{Z}(\mathbf{u}) \le z \mid (n)\}$$
(3.2)

Uma função aleatória, ou campo aleatório, é um conjunto de VA definidas sobre a região A. A função de distribuição acumulada multivariada, que caracteriza a FA composta por K VA, é definida por:

$$F(\mathbf{u}_1,...,\mathbf{u}_K;z_1,...,z_K) = Prob\{Z(\mathbf{u}_1) \le z_1,...,Z(\mathbf{u}_K) \le z_K\}$$
(3.3)

A função de distribuição acumulada univariada modela a incerteza local de uma variável aleatória. A função de distribuição acumulada multivariada modela a incerteza global de um campo aleatório.

O modelo de distribuição de uma VA, ou de uma FA, pode ser determinado segundo dois enfoques:

- o paramétrico, que estabelece um modelo de distribuição a priori que é
  determinado completamente por um conjunto limitado de parâmetros.
   Um exemplo típico é o modelo de distribuição normal que é totalmente
  determinado pelos valores da média e da variância da distribuição.
- o não paramétrico, que não estabelece nenhum um modelo de distribuição probabilístico a priori e que, portanto, não pode ser determinado por um conjunto limitado de parâmetros. Neste caso a função de distribuição de probabilidade, fdp, é obtida por um conjunto de valores estimados que representam uma aproximação discretizada do modelo de distribuição.

Uma vez determinados, os modelos de distribuição probabilísticos são usados:

- para estimativa de valores do atributo em posições não conhecidas;
- para modelagem da incerteza dos valores do atributo e;
- para inferência de valores, realizações, em procedimentos de simulação estocástica.

#### 3.3 O ESTIMADOR DE KRIGEAGEM LINEAR

A krigeagem linear, ou simplesmente krigeagem, foi inicialmente utilizada como um estimador de valores de atributos numéricos, em posições não observadas, para mapeamentos por médias ponderadas dos valores das amostras locais. Atualmente a krigeagem vem sendo utilizada, com maior frequência, na construção de modelos probabilísticos de incerteza sobre os valores dos

atributos. A krigeagem linear pode ser usada para inferência dos parâmetros, média e variância, de um modelo probabilísticos gaussiano, enquanto que, a krigeagem por indicação, uma krigeagem não linear, é usada para se construir uma aproximação discretizada de um modelo probabilístico qualquer. Mais formalmente, o estimador de krigeagem é definido como uma "coleção de técnicas de regressão linear generalizadas para *minimizar a variância do erro de estimação* a partir de um modelo de covariância definido a priori" (Deutsch e Journel 1998). A krigeagem possibilita a inferência de um valor médio do atributo, em uma posição espacial  $\mathbf{u}$ , a partir de um quantidade  $n(\mathbf{u})$  de amostras vizinhas  $z(\mathbf{u}_{\alpha})$ ,  $\alpha=1,...,n(\mathbf{u})$ . A formulação geral para um estimador por regressão linear generalizada é:

$$z^{*}(\mathbf{u}) - \mu(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}(\mathbf{u}) \cdot [z(\mathbf{u}_{\alpha}) - \mu(\mathbf{u}_{\alpha})]$$
(3.4)

onde  $\mu(\mathbf{u})$  é a tendência, ou o valor médio, do atributo na posição  $\mathbf{u}$ ,  $\mu(\mathbf{u}_{\alpha})$  é a média em cada posição amostrada  $\mathbf{u}_{\alpha}$  e  $\lambda_{\alpha}(\mathbf{u})$  são os ponderadores das amostras vizinhas de  $\mathbf{u}$ .

Sob a *hipótese de estacionariedade de segunda ordem*, a média de um atributo é constante e não depende da localização dentro de uma região de interesse, ou seja,  $\mu(\mathbf{u}) = \mu \ \forall \ \mathbf{u} \in A$ . Neste caso a equação 3.4 é modificada para se definir o estimador de *krigeagem simples* como:

$$z_{S}^{*}(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{S\alpha}(\mathbf{u}) \cdot z(\mathbf{u}_{\alpha}) + \left[1 - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{S\alpha}(\mathbf{u})\right] \boldsymbol{\mu}$$
(3.5)

onde  $\lambda_{Sa}(\mathbf{u})$  são os ponderadores de krigeagem simples das amostras vizinhas,  $\alpha$ , de  $\mathbf{u}$ , e  $\mu$  é a *média estacionária* da função aleatória caracterizada pelas variáveis aleatórias definidas numa região estacionária A.

A hipótese de estacionariedade de segunda ordem implica, também, em se considerar que a covariância entre duas variáveis aleatórias só depende do vetor de separação **h**, determinado pelas localizações espaciais dessas variáveis, ou seja,  $C(\mathbf{u}, \mathbf{u}+\mathbf{h}) = C(\mathbf{h})$ . Os pesos do estimador de krigeagem simples,  $\lambda_{S\alpha}(\mathbf{u})$  com  $\alpha=1,...,n(\mathbf{u})$ , são determinados considerando-se as seguintes premissas:

- o estimador deve ser não tendencioso, ou seja, o valor esperado do erro, ou do resíduo, de estimação deve ser zero,  $E\{Z^*(\mathbf{u}) Z(\mathbf{u})\} = 0$ , e;
- a variância do erro de estimação,  $\sigma^2 = Var\{Z^*(\mathbf{u}) Z(\mathbf{u})\}$ , é a mínima entre todos os estimadores lineares não tendenciosos.

A aplicação destes dois critérios possibilita o cálculo dos pesos  $\lambda_{Sa}(\mathbf{u})$ , pela solução do seguinte sistema de equações de krigeagem simples:

$$\sum_{\beta=1}^{n(\mathbf{u})} \lambda_{\beta}(\mathbf{u}) C(\mathbf{u}_{\beta} - \mathbf{u}_{\alpha}) = C(\mathbf{u} - \mathbf{u}_{\alpha}) \text{ para } \alpha = 1,...,n(\mathbf{u})$$
(3.6)

onde:  $C(\mathbf{u}_{\beta} - \mathbf{u}_{\alpha})$  é a covariância entre as amostras em  $\mathbf{u}_{\beta}$  e  $\mathbf{u}_{\alpha}$ ;  $C(\mathbf{u} - \mathbf{u}_{\alpha})$  é a covariância entre a amostra observada em  $\mathbf{u}_{\alpha}$  e a posição  $\mathbf{u}$ .

Uma desvantagem do estimador de krigeagem simples é a necessidade de se conhecer, a priori, a média estacionária da região A. Uma variante dessa formulação é o estimador de *krigeagem ordinária* que possibilita a inferência do atributo, numa posição  $\mathbf{u}$ , sem a necessidade de se conhecer a média estacionária  $\mu$ . Sob a condição de que a somatória dos ponderadores da krigeagem ordinária  $\lambda_{O\alpha}(\mathbf{u})$  é igual a 1, ou seja:

$$\sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{o\alpha}(\mathbf{u}) = 1$$
 (3.7)

obtêm-se a seguinte formulação para o estimador de krigeagem ordinária:

$$z_{o}^{*}(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{o\alpha}(\mathbf{u}) \cdot z(\mathbf{u}_{\alpha})$$
(3.8)

Pode-se mostrar que a formulação do estimador de krigeagem ordinária é equivalente a uma krigeagem simples, com médias estimadas localmente a partir das amostras vizinhas. A substituição de uma única média estacionária por médias locais, ou tendências locais, explica a extrema robustez do algoritmo de krigeagem ordinária (Deutsch e Journel 1998).

A aplicação dos critérios de mínima variância do erro de estimação e de não tendenciosidade do estimador possibilita o cálculo dos pesos,  $\lambda_{O\alpha}(\mathbf{u})$ , pela solução do seguinte sistema de equações de krigeagem ordinária:

$$\begin{cases}
\sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{o_{\alpha}}(\mathbf{u}) C(\mathbf{u}_{\alpha}, \mathbf{u}_{\beta}) + \phi(\mathbf{u}) = C(\mathbf{u}, \mathbf{u}_{\beta}) & \text{para } \beta = 1, ..., n(\mathbf{u}) \\
\sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{o_{\alpha}}(\mathbf{u}) = 1
\end{cases}$$
(3.9)

onde:  $C(\mathbf{u}_{\alpha}, \mathbf{u}_{\beta})$  é a covariância entre as amostras observadas em  $\mathbf{u}_{\alpha}$  e  $\mathbf{u}_{\beta}$ ;  $C(\mathbf{u}, \mathbf{u}_{\beta})$  é a covariância entre a amostra observada em  $\mathbf{u}_{\beta}$  e a posição  $\mathbf{u}$  e;  $\phi(\mathbf{u})$  é o multiplicador de Lagrange, necessário para a minimização da variância do erro e associado com a restrição  $\sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{o\alpha}(\mathbf{u}) = 1$ .

Em notação matricial, o sistema de krigeagem ordinária da equação 3.9 pode ser expresso como:  $\mathbf{K} \cdot \lambda = \mathbf{k} \implies \lambda = \mathbf{K}^{-1} \cdot \mathbf{k}$ , com:

$$\mathbf{K} = \begin{bmatrix} C_{11} C_{12} & \cdots & C_{1n} & 1 \\ C_{21} C_{22} & \cdots & C_{2n} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ C_{n1} C_{n2} & \cdots & C_{nn} & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix}; \quad \lambda = \begin{bmatrix} \lambda_{\mathbf{O}_{1}} \\ \lambda_{\mathbf{O}_{2}} \\ \vdots \\ \lambda_{\mathbf{O}_{n}} \\ \phi \end{bmatrix} \text{ e } \mathbf{k} = \begin{bmatrix} C_{1u} \\ C_{2u} \\ \vdots \\ C_{nu} \\ 1 \end{bmatrix}$$
(3.10)

onde: **K** e **k** são matrizes das covariâncias;  $C_{\alpha\beta}$ , com  $\alpha$  e  $\beta$  = 1,.., $n(\mathbf{u})$ , é a covariância entre as posições das amostras  $\alpha$  e  $\beta$ ;  $C_{\alpha u}$ , com  $\alpha$  = 1,.., $n(\mathbf{u})$ , é a

covariância medida entre a posição da amostra  $\alpha$  e a posição **u**, na qual se quer inferir um valor para o atributo e;  $\lambda$  é o vetor dos pesos.

Pode-se, ainda, expressar, na forma matricial, a variância do erro de estimação por krigeagem ordinária por (Journel 1988):

$$\sigma_0^2 = \mathbf{C}(0) - (\lambda)^{\mathsf{T}} \mathbf{k}$$
 (3.11)

As covariâncias  $C(\mathbf{u}_{\beta}, \mathbf{u}_{\alpha})$  e  $C(\mathbf{u}, \mathbf{u}_{\alpha})$ , apresentadas em 3.10, são calculadas a partir do *modelo de variograma teórico*,  $2\gamma(\mathbf{h})$ , ajustado sobre o *variograma experimental* que é determinado, a partir do conjunto amostral, por:

$$2\gamma(\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum_{(\alpha,\beta)/\mathbf{h}_{\alpha\beta} = \mathbf{h}} (z(\mathbf{u}_{\alpha}) - z(\mathbf{u}_{\beta}))^{2}$$
(3.12)

onde  $N(\mathbf{h})$  é o número de pares de amostras separadas, aproximadamente, pelo vetor  $\mathbf{h}$ . A Figura 3.1 mostra um variograma experimental com ajuste por um modelo de variografia teórico.

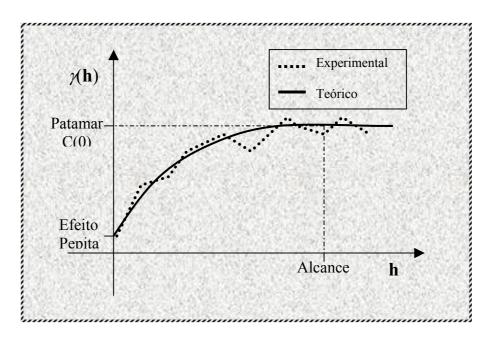


Fig. 3.1 - Variograma teórico ajustado sobre variograma experimental obtido a partir das amostras.

Os modelos teóricos mais utilizados são os modelos esférico, exponencial, gaussiano e potência. Os valores de efeito pepita, de patamar e de alcance, obtidos na variografia do dado, são utilizados na definição dos parâmetros do modelo teórico escolhido para o ajuste. Caso o comportamento do atributo seja anisotrópico, o ajuste deve ser realizado com modelos teóricos compostos baseados nas direções de máxima e mínima continuidade. Para mais detalhes sobre o ajuste de variogramas teóricos podem ser encontrados em Camargo (1997); Deutsch e Journel (1998) e Isaaks e Srivastava (1989). Sob as hipóteses de estacionariedade da covariância em A e  $\sum_{\alpha=1}^{n(\mathbf{u})} \lambda = 1$  a seguinte relação entre o semivariograma,  $\chi(\mathbf{h})$ , e a covariância,  $C(\mathbf{h})$ , é válida (Deutsch e Journel 1998):

$$\gamma(\mathbf{h}) = C(0) - C(\mathbf{h}) \tag{3.13}$$

A formulação para o estimador de krigeagem ordinária, equação 3.8, é semelhante à formulação dos estimadores determinísticos por média móvel ponderada pelo inverso da distância, doravante citado apenas como estimador por média móvel. A diferença aparece na determinação dos ponderadores das amostras vizinhas. Na krigeagem, os pesos são obtidos a partir de uma análise geoestatística da variabilidade do atributo dentro da região A. No estimativa por média móvel, estes pesos são baseados em um modelo geral e possivelmente inapropriado (Burrough e McDonnell 1998). Resumidamente, o estimador de krigeagem ordinária apresenta as seguintes vantagens, em relação ao estimador por média móvel:

 automaticamente considera aglomerados quando da definição dos pesos das amostras. Dependendo da distribuição espacial, algumas amostras podem, inclusive, ser associadas a pesos negativos. Isto permite o aparecimento de valores inferidos fora do intervalo de valores das amostras, ou seja, valores estimados maiores que o maior valor, ou menores que o menor valor, observados na amostragem;

- permite o tratamento de anisotropia de uma forma unificada pela definição de um único modelo de variograma teórico que é capaz de representar a variabilidade do atributo para várias direções e distâncias.
- possibilita o cálculo da variância do erro de estimação, que mostra a variância dos resíduos do estimador em função da distribuição espacial das amostras. Uma análise crítica da variância do erro de estimação será apresentada na Seção 3.4 deste capítulo.

É importante destacar que, a anisotropia do atributo e o problema de agrupamento de amostras também podem ser considerados ao se utilizar um estimador determinístico por média ponderada móvel. A anisotropia pode ser considerada a partir da definição de raios de busca, dos vizinhos, que variam com as direções. O problema de agrupamento pode ser minimizado utilizandose estratégia de busca por quadrantes ou octantes. Porém, nestes casos, a definição destas condicionantes, de busca dos vizinhos locais, é considerada subjetiva e depende do especialista. O especialista define os parâmetros de busca e de escolha dos vizinhos de acordo com seus conhecimentos, ou com sua percepção, a respeito do atributo em análise. A estimação por krigeagem adota um formalismo mais rigoroso, para determinação dos pesos das amostras vizinhas, ao se basear no modelo de correlação espacial, a variografía. Com um formalismo mais rigoroso, na estimação por krigeagem, os pesos das amostras vizinhas são calculados a partir de um modelo de correlação espacial obtido do conjunto amostral.

A principal desvantagem do estimador por krigeagem, em relação ao estimador de média móvel, é o custo de se realizar o estudo de correlação espacial entre as amostras. Esta é uma fase crítica do processo e requer, do especialista, o conhecimento e domínio das técnicas e das ferramentas, geralmente interativas, da geoestatística para a obtenção de um ajuste teórico de variografia que seja uma boa representação da variabilidade da propriedade em estudo.

## 3.4 A VARIÂNCIA DE KRIGEAGEM LINEAR COMO MEDIDA DE INCERTEZA: UMA ANÁLISE CRÍTICA

A partir dos critérios de não tendenciosidade do estimador e de minimização da variância do erro de estimação, apresentados na Seção 3.2, chega-se a seguinte formulação para a variância do erro do estimador de krigeagem ordinária,  $\sigma_{\varrho}^2$ , (Isaaks e Srivastava 1989):

$$\sigma_{o}^{2} = \sigma^{2} + \sum_{\alpha=1}^{n(\mathbf{u})} \sum_{\beta=1}^{n(\mathbf{u})} \lambda_{o\alpha}(\mathbf{u}) \lambda_{o\beta}(\mathbf{u}) C(\mathbf{u}_{\alpha}, \mathbf{u}_{\beta}) - 2 \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{o\alpha}(\mathbf{u}) C(\mathbf{u}, \mathbf{u}_{\alpha})$$
(3.14)

onde  $\sigma^2 = C(\mathbf{0})$  é a variância global do atributo e representa o quão errático é o atributo. Atributos com comportamentos mais erráticos têm variância global alta, contribuindo, assim, para um aumento no valor da variância de estimação. Sabendo-se que a covariância entre duas localizações diminui a medida que a distância entre elas aumenta, observa-se que: as covariâncias entre amostras,  $C(\mathbf{u}_{\alpha}, \mathbf{u}_{\beta})$ , aumentam a variância de estimação sempre que as amostras estão próximas entre si e, portanto, penalizam aglomerados de amostras e; a variância de estimação diminui com o aumento da covariância  $C(\mathbf{u}, \mathbf{u}_{\alpha})$ , ou seja, com a proximidade da amostra em relação à posição  $\mathbf{u}$ .

Portanto, a variância de krigeagem fornece uma medida que incorpora fatores relevantes para determinação da incerteza do estimador. Por isso, ela vem sendo utilizada como medida de incerteza por intervalo de confiança centrado na média de uma distribuição gaussiana. Entretanto, isto requer que duas condições básicas sejam satisfeitas (Isaaks e Srivastava 1989):

- a distribuição de probabilidade do atributo deve ser multigaussiana e;
- o patamar definido na variografia deve corresponder ao valor de variância do atributo.

A primeira premissa é difícil de ser verificada e raramente ocorre para atributos ambientais. Na prática, verificando-se a normalidade da distribuição bivariada,

supõe-se também a normalidade da distribuição multivariada (Goovaerts 1997 e Deutsch e Journel 1998). A segunda condição é necessária porque a variância é afetada pela escalabilidade do variograma (Isaaks e Srivastava 1989). A variância do modelo é influenciada por artefatos, criados ao se ajustar o variograma experimental por um modelo de variograma teórico. Caso isto ocorra, o variograma deve ser reescalonado para que o seu patamar coincida com o valor de variância global do atributo em estudo.

Finalmente, analisando-se a equação 3.14, observa-se que o valor de variância de krigeagem está relacionado à distribuição espacial das amostras, e não depende diretamente dos valores do atributo. Deutsch e Journel (1998) argumentam que "as variâncias dos erros da krigeagem linear, sendo independentes dos dados, fornecem apenas uma comparação entre diferentes configurações geométricas de dados". Entretanto, modelos de incerteza baseados na variância do erro de krigeagem linear vêm sendo utilizados sem que as premissas e análises acima sejam consideradas.

Na Seção seguinte descreve-se a krigeagem por indicação que possibilita a estimativa de incertezas sem a necessidade de se adotar um modelo de distribuição de probabilidade a priori para o atributo em estudo.

#### 3.5 O ESTIMADOR DE KRIGEAGEM POR INDICAÇÃO

Como já observado, os estimadores de krigeagem são considerados estimadores lineares por estimarem um valor, em uma posição espacial não observada, segundo uma combinação linear dos valores de um subconjunto amostral local. Além dos problemas com a estimativa de incerteza, apresentados na Seção 3.4, estes estimadores são usados apenas para inferir valores de variáveis de natureza numérica. A krigeagem linear não pode ser usada para inferir valores entre classes ainda que exista ordenação entre elas.

Um estimador de krigeagem não linear é um estimador de krigeagem linear aplicado sobre um conjunto amostral cujos valores do atributo foram

modificados segundo uma transformação não linear, por exemplo, uma transformação gaussiana, uma transformação lognormal ou outra (Deutsch e Journel 1998). O procedimento de krigeagem por indicação requer uma transformação não linear, chamada de codificação por indicação, que transforma cada valor do conjunto amostral  $Z(\mathbf{u}_{\alpha})$  em valores por indicação. Nesta seção serão introduzidos os conceitos que formalizam o estimador de krigeagem por indicação, o qual possibilita inferências para dados numéricos e para dados temáticos e, também, estimativas de incertezas associadas a estes atributos. Esta técnica tem como principal vantagem ser não paramétrica, ou seja, nenhum tipo de distribuição para a VA é considerado a priori. Ela possibilita a estimativa da função de distribuição da VA que, por sua vez, permite a determinação de incertezas e a inferência de valores do atributo, em localizações espaciais não amostradas. Além disso, diferentemente da krigeagem linear, este procedimento consegue modelar atributos com alta variabilidade espacial, sem a necessidade de se ignorar os dados amostrados cujos valores estão muito distantes de uma tendência (Journel 1983). A Seção 3.5.5 apresenta um detalhamento das principais vantagens e desvantagens do procedimento de krigeagem por indicação.

### 3.5.1 A krigeagem por indicação para atributos numéricos

Como já observado, a fdac,  $F(\mathbf{u}; \mathbf{z}|(n))$ , modela a incerteza sobre os valores de  $Z(\mathbf{u})$ , em posições  $\mathbf{u}$  não amostradas. Uma aproximação da fdac pode ser obtida pelo procedimento de krigeagem por indicação. Esta krigeagem é uma técnica de estimação com os mesmos fundamentos da krigeagem linear, porém, aplicada sobre os valores do atributo transformados segundo um mapeamento não linear, a *codificação por indicação*. A codificação por indicação, sobre um conjunto de dados amostrais numéricos,  $Z(\mathbf{u}=\mathbf{u}_{\alpha})$ , para um valor de corte  $z_k$ , gera um *conjunto amostral por indicação l*( $\mathbf{u}=\mathbf{u}_{\alpha}$ ;  $z_k$ ) do tipo:

$$I(\mathbf{u}; z_k) = \begin{cases} 1, & \text{se } Z(\mathbf{u}) \le z_k \\ 0, & \text{se } Z(\mathbf{u}) > z_k \end{cases}$$
(3.15)

A codificação por indicação, ilustrada na Figura 3.2, é aplicada sobre todo conjunto amostral criando, para cada valor de corte, um conjunto amostral por indicação cujos valores das amostras são transformados em 0 ou 1. Os K valores de corte,  $z_k$ , k=1,2...,K, são definidos em função do número de amostras.

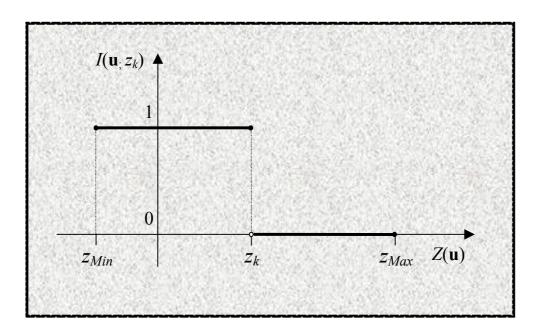


Fig. 3.2 - Codificação por indicação dos dados amostrais numéricos para um valor de corte  $z = z_k$ 

A definição de um modelo de variografia depende da existência de uma distribuição mínima de 0's (zeros) e 1's (uns) no conjunto de amostras codificados por indicação. Pode-se mostrar que a melhor variografia é obtida para o valor de corte igual ao valor da mediana obtido a partir do conjunto amostral (Isaaks e Srivastava 1989). Assim, é possível usar um único valor de corte, igual ou muito próximo do valor da mediana das amostras, para se gerar uma codificação por indicação única, conhecida como codificação por indicação pela mediana. O uso de um único valor de corte é vantajoso no sentido de eficiência computacional mas, apresenta restrições quanto ao seu uso quando o modelo de variografia, próximo a mediana, é muito diferente dos obtidos para outros valores de corte (Isaaks e Srivastava 1989).

Da mesma forma que se utiliza o conjunto amostral do atributo para inferir valores para a VA numérica  $Z(\mathbf{u})$ , o conjunto de dados amostrais por indicação é utilizado para inferir valores para variáveis aleatórias por indicação  $I(\mathbf{u}; z_k)$ , com  $\mathbf{u} \neq \mathbf{u}_{\alpha}$ .

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

$$E\{I(\mathbf{u}; z_k) | (n)\} = 1 \cdot Prob\{I(\mathbf{u}; z_k) = 1 | (n)\} + 0 \cdot Prob\{I(\mathbf{u}; z_k) = 0 | (n)\}$$

$$= 1 \cdot Prob\{I(\mathbf{u}; z_k) = 1 | (n)\} = F^*(\mathbf{u}; z_k | (n))$$
(3.16)

A equação 3.16 apresenta um resultado extremamente importante no que diz respeito 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 da função de distribuição acumulada condicionada,  $F^*(\mathbf{u}; z_k | (n))$ , para atributos numéricos". A esperança condicional  $E\{I(\mathbf{u}; z_k) | (n)\}$  pode ser estimada por algoritmos de krigeagem que utilizam as VA numéricas codificadas por indicação.

A *krigeagem por indicação simples* é um procedimento de krigeagem linear simples aplicado a um conjunto amostral codificado por indicação, nos valores de corte  $z = z_k$ , e tem a seguinte formulação:

$$F_{\mathbf{S}}^{*}(\mathbf{u}; z_{k} \mid (n)) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{S\alpha}(\mathbf{u}; z_{k}) i(\mathbf{u}_{\alpha}; z_{k}) + \left[1 - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{S\alpha}(\mathbf{u}; z_{k})\right] F^{*}(z_{k})$$
(3.17)

onde  $F^*(z_k)$  é a média da FA da região estacionária e os pesos  $\lambda_{Sa}(\mathbf{u}; z_k)$  são determinados com o objetivo de minimizar a variância do erro de estimação.

Considerando-se a somatória dos pesos de ponderação igual a 1, obtêm-se uma variante mais simplificada da krigeagem por indicação simples, a krigeagem por indicação ordinária, cuja expressão de estimação se resume a:

$$F_{o}^{*}(\mathbf{u}; z_{k} \mid (n)) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{O\alpha}(\mathbf{u}; z_{k}) i(\mathbf{u}; z_{k})$$
(3.18)

sendo que os pesos  $\lambda_{O\alpha}(\mathbf{u}; \mathbf{z}_k)$  são obtidos pela solução do seguinte sistema de equações de krigeagem por indicação ordinária:

$$\begin{cases}
\sum_{\beta=1}^{n(\mathbf{u})} \lambda_{O\beta}(\mathbf{u}; z_k) C_I(\mathbf{h}_{\alpha\beta}; z_k) + \phi(\mathbf{u}; z_k) = C_I(\mathbf{h}_{\alpha}; z_k) & \forall \alpha = 1, 2, ..., n(\mathbf{u}) \\
\sum_{\beta=1}^{n(\mathbf{u})} \lambda_{O\beta}(\mathbf{u}; z_k) = 1
\end{cases}$$
(3.19)

onde  $\phi(\mathbf{u}; z_k)$  é o multiplicador de Lagrange,  $\mathbf{h}_{\alpha\beta}$  é o vetor de separação definido pelas posições  $\mathbf{u}_{\alpha}$  e  $\mathbf{u}_{\beta}$ ,  $\mathbf{h}_{\alpha}$  é o vetor definido entre posições  $\mathbf{u}_{\alpha}$  e  $\mathbf{u}$ ,  $C_I(\mathbf{h}_{\alpha\beta}; z_k)$  é a autocovariância definida por  $\mathbf{h}_{\alpha\beta}$  e  $C_I(\mathbf{h}_{\alpha}; z_k)$  é a autocovariância definida por  $\mathbf{h}_{\alpha}$  em  $z = z_k$ . As autocovariâncias são determinadas pelo modelo de variografia teórico definido pelo conjunto I para  $z = z_k$ .

A krigeagem por indicação, simples ou ordinária, fornece, para cada valor de corte  $z_k$ , uma estimativa que é também a melhor estimativa mínima quadrática da esperança condicional da VA  $I(\mathbf{u}; z_k)$ . Utilizando esta propriedade, pode-se calcular estimativas dos valores da fdac de  $Z(\mathbf{u})$  para vários valores de  $z = z_k$ , pertencentes ao domínio de  $Z(\mathbf{u})$ . O conjunto dos valores das fdac's, estimados nos valores de corte, é considerado uma *aproximação discretizada* da fdac real de  $Z(\mathbf{u})$ . Quanto maior a quantidade de valores de corte, melhor é a aproximação. A aproximação é complementada pela definição de *uma função de ajuste para a distribuição*, que deve ser utilizada para se inferir a fdac para valores diferentes dos valores de corte. Um ajuste linear é o mais simples de se definir, porém funções de maior grau podem ser usadas.

Assim, a krigeagem por indicação é dita não paramétrica, pois 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 fdac de  $Z(\mathbf{u})$ . Os valores de probabilidades discretizados podem ser usados diretamente para se estimar valores estatísticos característicos da distribuição, tais como: valor médio, variância, mediana, quantis e outros.

Por exemplo, pode-se obter uma estimativa,  $\mu_Z(\mathbf{u})$ , do valor esperado  $E[Z(\mathbf{u})]$ 

$$E[Z(\mathbf{u})] = \int_{-\infty}^{\infty} z \cdot f(\mathbf{u}; z \mid (n)) dz = \int_{-\infty}^{\infty} z \cdot dF(\mathbf{u}; z \mid (n))$$
(3.20)

a partir da função de densidade de probabilidade condicionada as n amostras,  $f(\mathbf{u}, \mathbf{z}|(n))$ , e usando-se os K valores de corte,  $\mathbf{z}_k$ , pela aproximação:

$$\boldsymbol{\mu}_{Z}(\mathbf{u}) = \int_{-\infty}^{\infty} z \cdot dF(\mathbf{u}; z \mid (n)) \approx \sum_{k=1}^{K+1} z_{k}' \left[ F^{*}(\mathbf{u}; z_{k} \mid (n)) - F^{*}(\mathbf{u}; z_{k-l} \mid (n)) \right]$$
(3.21)

onde: os valores das  $F^*(\mathbf{u}; z_k|(n))$ , k=1,2,...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  $F(\mathbf{u}; z_{K+1}|(n)) = 1$ .

De forma equivalente, é possível obter-se uma estimativa da variância da distribuição condicionada,  $(\sigma^2)^*(\mathbf{u})$ , pela seguinte formulação:

$$(\boldsymbol{\sigma}^{2})^{*}(\mathbf{u}) = \int_{-\infty}^{\infty} [z - \mu_{Z}(\mathbf{u})]^{2} dF(\mathbf{u}; z \mid (n))$$

$$\approx \sum_{k=1}^{K+1} [z'_{k} - \mu_{Z}(\mathbf{u})]^{2} [F^{*}(\mathbf{u}; z_{k} \mid (n)) - F^{*}(\mathbf{u}; z_{k-1} \mid (n))]$$

$$(3.22)$$

O valor da mediana  $q_{0.5}(\mathbf{u})$ , da distribuição,

$$q_{0.5}(\mathbf{u}) = F^{-1}(\mathbf{u}; 0.5 \mid (n))$$
 (3.23)

é inferido aplicando-se a função de ajuste da distribuição sobre os valores de corte com probabilidades acumuladas vizinhas ao valor .5. Para distribuições com alto grau de assimetria, a mediana é um estimador mais robusto do que a média, enquanto que, uma medida mais robusta de espalhamento é um intervalo interquantil. Por exemplo, o intervalo interquantil,  $q_R(\mathbf{u})$ , é definido por:

$$q_{R}(\mathbf{u}) = q_{0.75}(\mathbf{u}) - q_{0.25}(\mathbf{u}) = \mathbf{F}^{-1}(\mathbf{u}; 0.75 \mid (n)) - \mathbf{F}^{-1}(\mathbf{u}; 0.25 \mid (n))$$
 (3.24)

# 3.5.2 A krigeagem por indicação para atributos temáticos

A diferença básica entre a krigeagem por indicação para dados temáticos e a krigeagem por indicação para dados numéricos, é a codificação por indicação das VA temáticas. Para dados temáticos, esta codificação tem a seguinte formulação:

$$I(\mathbf{u}; z_k) = \begin{cases} 1, & \text{se } Z(\mathbf{u}) = z_k \\ 0, & \text{se } Z(\mathbf{u}) \neq z_k \end{cases}$$
(3.25)

onde os valores de corte  $z_k$ , k=1,...,K, são valores das K classes que pertencem ao domínio da função aleatória  $Z(\mathbf{u})$  definida na região A. A Figura 3.3 ilustra essa transformação.

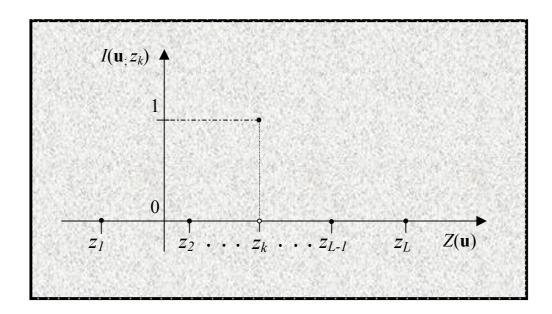


Fig. 3.3 - Codificação por indicação dos dados amostrais temáticos para uma classe de corte  $z = z_k$ 

Cada conjunto amostral codificado por indicação deve possibilitar a definição de um modelo de variografia relacionado à classe usada na codificação.

Quando o conjunto amostral por indicação possui poucos valores iguais a 0, ou iguais a 1, torna-se mais difícil a definição de um modelo de variografia para este conjunto. Isto significa que, para conjuntos amostrais com uma das classes com cardinalidade muito pequena, ou muito grande, a definição do modelo de variografia por indicação, relativo a essa classe, pode ser problemática. A esperança condicional da VA temática por indicação  $I(\mathbf{u}; \mathbf{z}_k)$  é definida como:

$$E\{I(\mathbf{u}; z_k) | (n)\} = 1 \cdot Prob\{I(\mathbf{u}; z_k) = 1 | (n)\} + 0 \cdot Prob\{I(\mathbf{u}; z_k) = 0 | (n)\}$$

$$= Prob\{I(\mathbf{u}; z_k) = 1 | (n)\} = F^*(\mathbf{u}; z_k | (n))$$
(3.26)

Este resultado mostra que é possível inferir valores da função de distribuição de probabilidade condicionada (fdpc), de uma classe  $z_k$ , a partir da esperança condicional do conjunto amostral por indicação  $I(\mathbf{u}; z_k)$ . A esperança condicional pode ser inferida pelos procedimentos de krigeagem por indicação descritos na seção anterior.

Para dados temáticos, a krigeagem por indicação, simples ou ordinária, fornece, para cada classe  $z_k$ , uma estimativa que é, também, a melhor estimativa mínima quadrática da esperança condicional da VA  $I(\mathbf{u}; z_k)$ . Utilizando esta propriedade, pode-se calcular estimativas dos valores da fdpc de  $Z(\mathbf{u})$  para todas as classes  $z_k$  do domínio de  $Z(\mathbf{u})$ . O conjunto destes valores representa uma estimativa da fdpc da VA temática em  $\mathbf{u}$ .

Em resumo, o estimador de krigeagem por indicação, aplicado a uma VA temática, possibilita a inferência dos valores de probabilidade  $p_k(\mathbf{u})$ , para cada classe  $z_k$ , k=1,...,K, em uma determinada localização  $\mathbf{u}$ .

Pelo estimador de moda, o valor para a VA temática  $Z(\mathbf{u})$ , na posição  $\mathbf{u}$ , é a classe de maior probabilidade, ou seja:

$$Z(\mathbf{u}) = z_k \operatorname{sse} p_k(\mathbf{u}) > p_i(\mathbf{u}) \ \forall i = 1,...,K \ e \ k \neq i$$
(3.27)

Uma variação do estimador de moda foi apresentada por Soares (1992). O autor argumenta que o algoritmo de krigeagem por indicação tende a superestimar as classes que apresentam grande probabilidade de ocorrer na amostragem considerada. Por outro lado tende a subestimar as classes com menor probabilidade de ocorrência no conjunto amostral. A idéia básica é manter a distribuição das classes nos valores inferidos iguais à distribuição das classes no conjunto amostral, ou seja, o histograma do campo inferido deve ser igual, ou muito próximo, ao histograma das amostras.

Esse estimador por moda, restrito ao histograma das amostras, apresenta 2 desvantagens:

- requer que a amostragem não seja tendenciosa, ou seja, a amostragem represente, com um certo grau de acurácia, a distribuição global do atributo e;
- é um procedimento que tem computação mais custosa pois requer o gerenciamento da probabilidade total dada a cada classe durante o processo de espacialização do atributo.

#### 3.5.3 Estimativa de incertezas locais para atributos numéricos

O conhecimento da fdac, em uma localização  $\mathbf{u}$ , possibilita uma estimativa direta da incerteza, sobre o valor não conhecido  $z(\mathbf{u})$ , anterior a, e independente da, escolha de um estimador para  $z(\mathbf{u})$  (Goovaerts 1997).

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,  $\mu_Z(\mathbf{u})$ , do tipo:

$$Prob\{Z(\mathbf{u}) \in [\boldsymbol{\mu}_{Z}(\mathbf{u}) \pm \sigma(\boldsymbol{u})]\} \cong 0.68$$
ou
$$Prob\{Z(\mathbf{u}) \in [\boldsymbol{\mu}_{Z}(\mathbf{u}) \pm 2\sigma(\boldsymbol{u})]\} \cong 0.95$$
(3.28)

onde  $\sigma^2 = E\{(Z(\mathbf{u}) - E\{Z(\mathbf{u})\})^2\}$  é a variância da distribuição.

Como já observado, 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,  $F^*(\mathbf{u}; \mathbf{z}|(n))$ , pode-se derivar vários intervalos de probabilidade tais como o intervalo 95%, [ $q_{0.025}$ ;  $q_{0.975}$ ], tal que:

$$Prob\{Z(\mathbf{u}) \in [q_{0.025}; q_{0.975}] | (n)\} = 0.95$$
 (3.29)

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 *intervalo de probabilidade*, é computado como a diferença entre os valores da fdac para os limiares b e a, ou seja:

$$Prob\{Z(\mathbf{u}) \in (a,b] \mid (n)\} = F(\mathbf{u};b\mid (n)) - F(\mathbf{u};a\mid (n))$$
(3.30)

Um intervalo de probabilidade 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:

$$|Prob\{Z(\mathbf{u}) \in (a, +\infty] | (n)\} = Prob\{Z(\mathbf{u}) > a | (\mathbf{n})\} = 1 - F(\mathbf{u}; a | (n))|$$
(3.31)

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

## 3.5.4 Estimativa de incertezas locais para atributos temáticos

Nesta seção, apresentam-se duas formulações para cálculo da incerteza local para dados temáticos: a incerteza definida pela moda da distribuição e a incerteza estimada por uma medida de entropia, a entropia de Shannon.

A incerteza em  $\mathbf{u}$ ,  $Inc(\mathbf{u})$ , pode ser determinada por um único valor de probabilidade, a probabilidade da classe  $s_k$ , que foi associada à posição  $\mathbf{u}$ . Esta medida, que será referenciada, neste texto, como *incerteza pela moda*, tem a seguinte formulação:

$$Inc(\mathbf{u}) = 1 - p_k(\mathbf{u}) \tag{3.32}$$

onde  $p_k(\mathbf{u}) = Prob\{Z(\mathbf{u}) = s_k\}$ 

Outra possibilidade é medir a incerteza a partir da combinação dos valores de probabilidades das K classes em  $\mathbf{u}$ . A *incerteza por entropia* pertence a esse grupo e é calculada a partir da medida de entropia proposta por Shannon e Weaver (1949). Informalmente, a entropia pode ser entendida com uma medida da organização espacial de um atributo. Ela mede a confusão, a desordem, relacionada com os valores, ou possíveis estados, de um atributo. Para uma VA contínua, cuja função de densidade de probabilidade é f(z), a entropia de Shannon é definida por (Journel e Deutsch 1993):

$$H_f = \int_{-\infty}^{\infty} [ln(f(z))] f(z) dz$$
 (3.33)

Quando a VA é temática, ou resultado da discretização de uma VA contínua, a entropia de Shannon, de sua distribuição univariada discreta, tem a seguinte formulação:

$$H(\mathbf{u}) = -\sum_{k=1}^{K} p_k(\mathbf{u}) ln(p_k(\mathbf{u}))$$
(3.34)

onde  $p_k(\mathbf{u})$ , k=1,...,K, são as probabilidades univariadas das K classes, ou categorias, do domínio da variável aleatória. A entropia de Shannon, para VA discretas, mede o grau de confusão entre as classes pois considera as probabilidades de todas as classes de forma unificada. Assim, utilizando-se a medida de Shannon, a incerteza por entropia, para VA discretizadas, é igual a entropia da distribuição discreta univariada, ou seja,  $Inc(\mathbf{u}) = H(\mathbf{u})$ .

A medida de entropia de Shannon, *H*, possui as seguintes propriedades:

- H é não negativa, ou seja H ≥ 0, porque as probabilidades associadas às variáveis discretas não podem ser maior do que 1;
- H = 0 se a probabilidade de uma das categorias é igual a 1 e a probabilidade das demais categorias é igual 0, ou seja, p<sub>i</sub> =1 e p<sub>j</sub> = 0 para j=1,...,K e j ≠ i;
- Para um determinado número K de classes, H = ln(K) é o valor máximo para H e ocorre quando tem-se uma distribuição uniforme, todos p<sub>k</sub>'s são iguais a 1/K, ou seja, H = ln(K) para p<sub>k</sub>=1/K ∀ k=1,...,K. Além disso, para distribuições uniformes, H cresce com o aumento do número de classes.

As duas metodologias de cálculo de incertezas, consideradas nesta seção, não são as duas únicas abordagens possíveis, porém apresentam, entre si, diferenças significativas. A Tabela 3.1 mostra exemplos de valores de incertezas, pela moda e por entropia, calculados para 6 distribuições de probabilidades com 4 classes distintas. Analisando-se essas distribuições observa-se que:

- a distribuição 1 da tabela mostra que, quando a probabilidade de uma das classes é igual a 1.0 e a das demais é igual a 0.0, as incertezas pela moda e por entropia são iguais e têm valor 0.0;
- Para as distribuições 2, 3 e 4, a incerteza pela moda se mantém constante, enquanto que a incerteza por entropia aumenta a medida que mais classes apresentam probabilidades diferente de 0.0;
- As distribuições 5 e 6 mostram que, para distribuições uniformes, o valor de incerteza por entropia, que é máximo e igual a ln(K), cresce quando K cresce, ou seja, quando o número de classes com probabilidade diferente de 0.0 aumenta.

TABELA 3.1- EXEMPLOS DE CÁLCULOS DE INCERTEZAS PARA VÁRIAS DISTRIBUIÇÕES

Distribuições	Probabilidades				Incertezas	
	P(Classe=1)	P(Classe=2)	P(Classe=3)	P(Classe=4)	Por moda	Por entropia
1	1.000	0.000	0.000	0.000	0.000	0.000
2	0.520	0.480	0.000	0.000	0.480	0.692
3	0.520	0.240	0.240	0.000	0.480	1.025
4	0.520	0.160	0.160	0.160	0.480	1.220
5	0.333	0.333	0.333	0.000	0.666	1.099
6	0.250	0.250	0.250	0.250	0.750	1.386

O objetivo da criação da Tabela 3.1 e das observações tecidas sobre a mesma, é mostrar, quantitativamente, as principais diferenças entre os dois procedimentos de estimativa de incertezas para atributos temáticos.

# 3.5.5 Vantagens e desvantagens da krigeagem por indicação

A partir das considerações apresentadas em Isaaks e Srivastava (1989); Deutsch e Journel (1998) e Gouvaerts (1997), pode-se ressaltar as seguintes vantagens, específicas do procedimento de krigeagem por indicação:

- a krigeagem por indicação é não paramétrica. 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: quantis, valor esperado e variância. Portanto não se restringe a modelagem de atributos com distribuições simétricas como, por exemplo, a gaussiana;
- a krigeagem por indicação fornece uma metodologia única para espacialização, com estimativa de incertezas, para atributos espaciais de natureza temática e numérica;
- diferentemente da krigeagem linear, que estima a variância do erro de estimação em função do estimador e da distribuição geométrica das amostras, a krigeagem por indicação possibilita a estimativa de incertezas, utilizando a função de distribuição acumulada condicionada da VA que representa o atributo, independentemente do estimador;
- a krigeagem por indicação pode ser usada para modelar atributos com alta variabilidade espacial sem a necessidade de se filtrar amostras cujos valores estão muito distantes de uma tendência ("outliers");
- a krigeagem por indicação permite melhorar a qualidade de estimação com o uso de amostras indiretas, retiradas de fontes auxiliares, em conjunto com o conjunto amostral do atributo, amostras diretas. Este procedimento será apresentado, em detalhes, no Capítulo 5.

Além das vantagens, expostas acima, os procedimentos de krigeagem por indicação apresentam algumas desvantagens. Este procedimento requer, do especialista, um alto grau de interatividade para se definir a quantidade e os valores de corte a serem utilizados. Também, exige que seja definido um variograma para cada valor de corte considerado. Além disso, a aproximação da fdc apresenta alguns problemas, conhecidos como desvios de relação de ordem, que devem ser corrigidos automaticamente pelo procedimento. Os desvios de ordem, para atributos numéricos e temáticos, são apresentados e analisados a seguir.

Quando o atributo é numérico, os valores de probabilidades acumuladas condicionadas, para cada valor de corte, são inferidos independentemente. Para que esses valores de probabilidade constituam uma distribuição legítima, eles devem ser corrigidos para verificar as seguintes relações de ordem:

1) Os valores inferidos de probabilidade acumulada,  $F^*(\mathbf{u}; z_k|(n)) = Prob\{Z(\mathbf{u}) \le z_k|(n)\}$ , não podem ser menores que 0 ou maiores que 1, ou, seja:

$$0 \le F^*(\mathbf{u}; z_k \mid (n)) \le 1 \quad \forall z_k, \ k = 1,...,K$$
 (3.35)

2) O valor estimado da fdac para um valor de corte  $z_j$  não deve ser maior do que a fdac, para um valor de corte  $z_k$ , quando  $z_j \le z_k$ , ou seja:

$$F^*(\mathbf{u}; z_j \mid (n)) \le F^*(\mathbf{u}; z_k \mid (n)) \text{ se } z_j \le z_k$$
(3.36)

A condição 1 pode ser garantida quando todos os pesos do estimador são positivos e somam 1. A krigeagem não garante que os pesos sejam todos positivos. Por isso é possível a inferência de valores da fdac fora do intervalo [0,1]. A solução, para este problema, é ajustar os valores estimados para as bordas, ou seja, valores negativos são mapeados para 0 e valores maiores que 1 são mapeados para 1.

A segunda condição é garantida com o uso de ponderadores positivos que somam 1, e com a utilização dos mesmos pesos de estimação para todos os valores de corte. A krigeagem por indicação mediana usa os mesmos pesos de estimação mas não se pode garantir que os pesos sejam todos positivos. Portanto, estas inconsistências podem ocorrer no caso da krigeagem por indicação e devem ser corrigidas. Um procedimento simples de correção é verificar pares de fdac's estimadas, em valores sucessivos de cortes, e ajustálos para o valor médio das duas, sempre que a relação de ordem, apresentada em 3.36, não for satisfeita (Isaaks e Srivastava 1989). A Figura 3.4 ilustra os problemas e as soluções referentes às duas condições acima descritas.

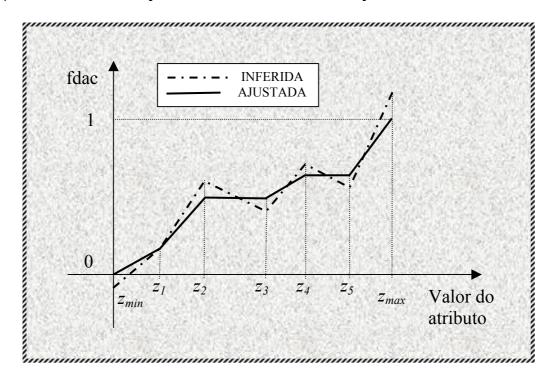


Fig. 3.4 - Correção dos desvios de relação de ordem para dados numéricos Fonte: Baseada em Deutsch e Journel (1998).

Também, para atributos temáticos, os valores de probabilidade são estimados independentemente, para cada valor de corte. Para que esses valores de probabilidade constituam uma distribuição legítima eles devem ser corrigidos para verificar as seguintes relações de ordem:

1) Os valores de probabilidade inferidos,  $F^*(\mathbf{u}; z_k|(n)) = Prob\{ Z(\mathbf{u}) = z_k|(n) \}$ , não podem ser menores do que 0 ou maiores do que 1, ou, seja:

$$0 \le F^*(\mathbf{u}; z_k \mid (n)) \le 1 \quad \forall z_k, \ k = 1, ..., K$$
(3.37)

2) A somatória dos valores de probabilidade, inferidos para as *K* classes, deve ser igual a 1, ou seja:

$$\sum_{k=1}^{K} F^{*}(\mathbf{u}; z_{k} | (n)) = 1$$
 (3.38)

Para a condição 1 valem as mesmas considerações apresentadas para o caso de inferência dos atributos numéricos.

A segunda condição é um pouco mais complicada de se obter. Uma solução, não muito complicada, é remapear linearmente todos os valores de probabilidade inferidos para valores de probabilidade cuja somatória total seja igual a 1.

### 3.6 CONCLUSÕES

O presente capítulo apresentou a ferramenta geoestatística de krigeagem que é utilizada para inferir valores de atributos, em posições não observadas, e também incertezas associadas aos valores inferidos. Mostrou-se que a krigeagem por indicação tem aplicação mais geral, principalmente porque não supõe nenhum tipo de distribuição de probabilidade a priori e pode ser usado com atributos numéricos e temáticos. Por exemplo, a krigeagem por indicação permite a inferência de valores temáticos e, portanto, pode ser considerada um classificador estocástico, que fornece estimativas de incertezas associadas aos valores das classes atribuídos a cada ponto do espaço.

Apresentou-se, ainda, alternativas para estimativas de incertezas que devem ser escolhidas de acordo com a natureza do atributo, que está sendo modelado, e também de acordo com os objetivos de uma aplicação.

Porém a krigeagem não é a única ferramenta geoestatística atualmente em uso para modelagem de dados espaciais. A simulação estocástica também pode ser usada e apresenta algumas vantagens, em relação à krigeagem, que serão descritas e analisadas no Capítulo 4.

# **CAPÍTULO 4**

# INFERÊNCIAS COM ESTIMATIVA DE INCERTEZAS POR SIMULAÇÃO ESTOCÁSTICA CONDICIONADA

#### 4.1 INTRODUÇÃO

Simular significa representar por semelhança. Neste capítulo apresentam-se os formalismos matemáticos do procedimento geoestatístico de simulação estocástica condicionada, que possibilita a criação de campos aleatórios semelhantes, segundo critérios probabilísticos. Estes campos, realizações de funções aleatórias, são usados para caracterizar distribuições de probabilidade que, como já enfatizado nos procedimentos de krigeagem do Capítulo 3, modelam as incertezas associadas aos valores de variáveis ou de campos aleatórios. Entretanto, as funções de distribuição probabilística são construídas de maneiras distintas pela krigeagem e pela simulação estocástica condicionada. As diferenças e semelhanças entre estes procedimentos são abordadas, com detalhes, no desenvolvimento das seções deste capítulo.

Os objetivos principais deste capítulo são:

- apresentar, formalmente, o procedimento geoestatístico de simulação estocástica condicionada, ao conjunto amostral e aos valores previamente simulados, sequencial;
- apresentar e discutir metodologias para estimativa de valores e de incertezas, locais e globais, baseadas nos valores simulados para variáveis e funções aleatórias representando atributos temáticos e numéricos e;
- apresentar uma análise comparativa entre os procedimentos geoestatísticos de krigeagem e de simulação estocástica condicionada.

O presente capítulo apresenta a seguinte organização: na Seção 4.2 apresentam-se os formalismos matemáticos do procedimento de simulação sequencial condicionada proposto por Deutsch e Journel (1998); na Seção 4.3, mostra-se como inferir parâmetros estatísticos, para atributos numéricos e temáticos, a partir das realizações geradas pela simulação sequencial; a Seção 4.4 apresenta alternativas para estimativas de incertezas, para dados numéricos e temáticos, também a partir das realizações da simulação sequencial; a Seção 4.5 dedica-se a uma análise comparativa entre os procedimentos geoestatísticos de krigeagem e simulação condicionada e; por fim, na Seção 4.6 apresentam-se conclusões sobre os temas deste capítulo.

#### 4.2 O PROCEDIMENTO DE SIMULAÇÃO SEQUENCIAL CONDICIONADA

Uma simulação estocástica é um processo de construção de realizações alternativas conjuntas, igualmente prováveis, das VA que compõem um modelo de uma função aleatória. As realizações  $\{z^{(l)}(\mathbf{u}), \mathbf{u} \in \mathbf{a} \text{ uma região de estudo } A$  e  $l = 1,2...,L\}$ , representam L imagens possíveis da distribuição espacial dos valores de um atributo Z sobre a área A.

As variáveis aleatórias de um campo aleatório, que representa a distribuição de um atributo espacial, não são independentes. Como já observado no Capítulo 3, existe uma correlação entre essas variáveis que, sob a hipótese de estacionariedade de segunda ordem, depende do vetor de separação,  $\mathbf{h}$ , determinado pela posição espacial das amostras. Assim, a simulação conjunta de m variáveis aleatórias, que compõem um campo aleatório, requer a inferência e a modelagem da matriz de covariância cruzada entre essas variáveis. Entretanto, este procedimento é muito custoso computacionalmente e, por isso, não é adotado na prática. Deutsch e Journel (1998) apresentam uma metodologia eficiente que aproxima a simulação conjunta de várias variáveis aleatórias, denominado simulação sequencial condicionada. Essa simulação usa a função de distribuição acumulada condicionada, às VA mais correlacionadas, para obter valores  $z(\mathbf{u})$ , de uma VA Z em cada posição  $\mathbf{u} \in A$ .

O condicionamento considera os dados amostrais originais e também os valores pré-simulados dentro da vizinhança de  ${\bf u}$ .

Considere a distribuição acumulada conjunta de m VA,  $F(z_1,...,z_m \mid (n))$ , condicionadas a um conjunto de n amostras do atributo:

$$F(z_1,...,z_m | (n)) = Prob\{Z_i \le z_i, i = 1,...,m | (n)\}$$
 (4.1)

A implementação da simulação sequencial condicionada, para obtenção de valores em *m* posições não amostradas, segue as etapas apresentadas no algoritmo abaixo (baseado em Deutsch e Journel 1998):

# Para cada nova realização I faça:

#### Início

- 1) Obtenha um valor  $z_1^{(l)}$  a partir da fdac univariada de  $Z_1$  condicionada aos n dados originais: Prob{ $Z_1 \le z_1 | (n)$ };
- 2) Atualize o conjunto de dados originais (n) para:  $(n+1) = (n) \cup \{Z_1 = z_1^{(l)}\};$
- 3) Obtenha um valor  $z_2^{(l)}$  a partir da fdac univariada de  $Z_2$  condicionada aos (n+1) dados: Prob $\{Z_2 \le z_2 | (n+1)\}$ .
- 4) Atualize o conjunto de dados originais (n+1):  $(n+2) = (n+1) \cup \{Z_2 = z_2^{(l)}\};$
- 5) Sequencialmente considere todas a m VA  $Z_i$ 's;

# Fim

Cada conjunto  $\{z_i^{(l)}, i=1,...,m\}$ , criado a partir deste algoritmo, representa uma realização, simulada em conjunto, das m VA dependentes,  $Z_i$ 's. É comum, no ambiente de um SIG, que as m posições estejam distribuídas regularmente no espaço, compondo uma estrutura de representação de grade regular retangular.

Do algoritmo 4.1, observa-se que o procedimento de simulação sequencial requer a determinação de m fdac's univariadas:  $Prob\{Z_1 \le z_1 | (n)\}$ ,  $Prob\{Z_2 \le z_2 | (n+1)\}$ , ...,  $Prob\{Z_m \le z_m | (n+m-1)\}$ . Estas fdac's podem ser estimadas pelos algoritmos de krigeagem apresentados no Capítulo 3.

Como abordado, no Capítulo 3, a krigeagem por indicação deve ser usada quando não se pode estabelecer uma função de distribuição a priori, para a VA ou para a FA, e também quando o atributo é de natureza temática. A *simulação* sequencial por indicação utiliza os procedimentos de krigeagem por indicação para estimar as fdac's requeridas pelo algoritmo 4.1.

A simulação sequencial por indicação é utilizada para se obter realizações para atributos contínuos e temáticos. A determinação das fdac's para variáveis temáticas exige a definição de uma ordenação prévia para as *K* categorias, do domínio da VA, ainda que não exista uma razão que justifique a ordenação.

A etapa seguinte, à ordenação da VA, é a obtenção de valores simulados a partir das fdac's inferidas. A obtenção de um valor simulado  $z_i^{(l)}$ , i=1,...,m, da realização l, se inicia com a geração de um valor aleatório  $\in [0,1]$  de uma distribuição uniforme. Esse valor aleatório corresponde a uma probabilidade,  $p_i^{(l)}$ , de ocorrência da variável aleatória  $Z(\mathbf{u})$ . Em seguida essa probabilidade é mapeada para um valor de  $Z(\mathbf{u})$ ,  $z_i^{(l)}$ , utilizando-se a fdac inferida para a posição  $\mathbf{u}$ . A Figura 4.1 ilustra esse procedimento para o caso de obtenção de realizações para atributos numéricos. Observa-se, neste caso, que uma interpolação se faz necessária sempre que a probabilidade  $p_i^{(l)}$  difere dos valores das probabilidades acumuladas inferidas nos valores de corte  $z = z_k$ .

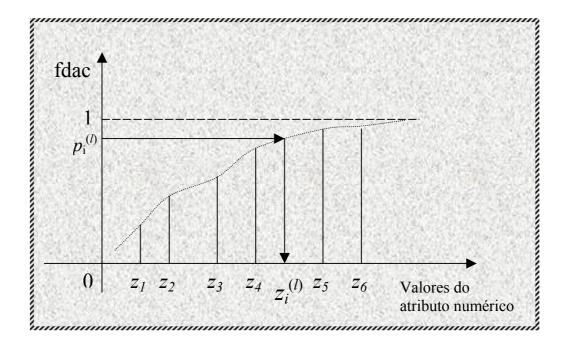


Fig. 4.1 - Obtenção de um valor da VA a partir da função de distribuição de probabilidade acumulada inferida para um atributo numérico.

A Figura 4.2 ilustra o procedimento de obtenção de um valor simulado, para um atributo de natureza temática, a partir de um valor de probabilidade  $p_i^{(l)}$  obtido da distribuição uniforme  $\in$  [0,1]. Neste caso o valor de probabilidade é sempre mapeado para uma das classes do domínio do atributo.

Neste ponto, pelo menos duas observações são importantes:

1) No procedimento de krigeagem, o valor do atributo, inferido para cada posição  $\mathbf{u}$  de uma região de interesse, é único. Geralmente, infere-se a média ou a mediana para atributos numéricos e a moda para atributos temáticos. Por outro lado, utilizando-se o procedimento de simulação, o valor inferido,  $\mathbf{z}_i^{(l)}$ , pode ser qualquer valor do domínio do atributo e, portanto, não é único.

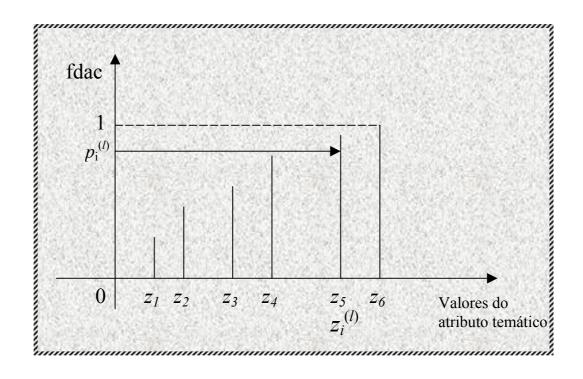


Fig. 4.2 - Obtenção de um valor da VA a partir da função de distribuição de probabilidade acumulada inferida para um atributo temático.

2) O procedimento de simulação é mais completo do que o de krigeagem no que diz respeito ao modelo de covariância considerado para o atributo. Isto ocorre porque as fdac's estimadas pela krigeagem estão condicionadas apenas às amostras e, portanto, as incertezas, estimadas a partir destas, são locais. As fdac's estimadas para a simulação consideram também os valores pré-simulados e, portanto, fornecem informação sobre a incerteza conjunta em várias localizações. Dos valores simulados obtém-se, por exemplo, a probabilidade que todos os valores, simultaneamente, excedam um limiar para uma sub-área da, ou para toda a, área de interesse.

Resumindo-se, o procedimento de simulação estocástica permite a geração de um conjunto de campos aleatórios equiprováveis, superfícies que têm

características probabilísticas similares, relacionado a um atributo. Neste procedimento, a função de distribuição de probabilidade, inferida pela krigeagem para uma determinada localização, é usada para se obter realizações para o valor do atributo. Uma simulação estocástica condicionada combina os dados nas posições observadas, os dados já simulados e a informação do variograma para computar novas realizações para localizações ainda não consideradas. O conjunto de valores realizados, em uma determinada posição possibilita a inferência de estatísticas e incertezas locais a respeito do atributo. O conjunto de campos obtidos pela simulação permite a inferência de estatísticas e incertezas globais a respeito do atributo.

#### 4.3 INFERÊNCIA DE ESTATÍSTICAS A PARTIR DOS VALORES SIMULADOS

As duas subseções seguintes mostram alternativas para criação de campos aleatórios que fornecem estatísticas sobre as *L* realizações obtidas em uma simulação sequencial. Apresentam-se soluções para atributos numéricos e para atributos temáticos.

# 4.3.1 Estatísticas sobre realizações de atributos numéricos

Pelo procedimento de simulação sequencial condicionada, quando o atributo é de natureza numérica, obtém-se, para cada posição  $\mathbf{u} \in A$  de cada realização I, um valor numérico para a VA em análise. O conjunto de valores obtidos para uma localização  $\mathbf{u}$ , nas L realizações, possibilita a estimativa do valor da probabilidade acumulada,  $F^*(\mathbf{u}; z_k)$ , para qualquer valor de corte  $z = z_k$  do atributo, pela formulação:

$$F^*(\mathbf{u}; z_k) = Prob\{\mathbf{Z}(\mathbf{u}) \le z_k\} = \frac{1}{L} \sum_{l=1}^{L} I_l(\mathbf{u}; z_k)$$
(4.2)

onde  $I_l(\mathbf{u}; z_k)$  é a variável por indicação, determinada pelo valor de corte  $z = z_k$  sobre Z, e definida como:  $I_l(\mathbf{u}; z_k) = 1$  se  $z^{(l)}(\mathbf{u}) \le z_k$  e  $I_l(\mathbf{u}; z_k) = 0$  caso contrário.

Utilizando-se um conjunto de valores de corte pode-se, assim, obter uma aproximação discretizada, com qualquer grau de acurácia desejado, da fdac da variável aleatória definida em **u**. A inferência de parâmetros estatísticos, tais como médias e medianas, podem ser obtidos a partir da aproximação da fdac seguindo as formulações apresentadas na Seção 3.5.1.

#### 4.3.2 Estatísticas sobre realizações de atributos temáticos

Pelo procedimento de simulação sequencial, quando o atributo é de natureza temática, obtém-se, para cada posição  $\mathbf{u} \in A$  de cada realização I, um valor temático para a VA em análise. O conjunto de valores obtidos para uma localização  $\mathbf{u}$ , nas I realizações, permite o cálculo da probabilidade, pelo paradigma frequencista, para qualquer valor de classe do atributo. Assim podese inferir os valores da função de distribuição condicionada da variável aleatória,  $F^*(\mathbf{u}; \mathbf{z}_k)$ , definida em  $\mathbf{u}$ , para cada classe  $\mathbf{z} = \mathbf{z}_k$ , por:

$$F^*(\mathbf{u}; z_k) = Prob\{\mathbf{Z}(\mathbf{u}) = z_k\} = \frac{1}{L} \sum_{l=1}^{L} I_l(\mathbf{u}; z_k)$$
(4.3)

o onde  $I_l(\mathbf{u}; z_k)$  é a variável por indicação, determinada pelo valor de corte  $z = z_k$  sobre Z, e definida como:  $I_l(\mathbf{u}; z_k) = 1$  se  $z^{(l)}(\mathbf{u}) = z_k$  e  $I_l(\mathbf{u}; z_k) = 0$  caso contrário.

Aplicando-se a equação 4.3 para todas as *K* classes do domínio da VA, que representa o atributo temático, obtém-se uma estimativa da função de distribuição de probabilidade do atributo. Esta aproximação é usada na inferência do estimador pelo valor de moda segundo a formulação apresentada na Seção 3.5.2.

# 4.4 ESTIMATIVA DE INCERTEZAS UTILIZANDO SIMULAÇÃO SEQUENCIAL CONDICIONADA

Da mesma forma que foi possível obter estimativas de incertezas para campos aleatórios obtidos por procedimentos de krigeagem, pode-se, também, estimar incertezas a partir das *L* realizações geradas pela simulação sequencial

condicionada. Estas medidas são locais, quando baseadas na distribuição univariada do atributo, que é definida pelos L valores do atributo simulados para uma determinada localização  $\mathbf{u}$ . Medidas de incerteza global, ou conjunta, se baseiam na distribuição multivariada, que é definida pelos L campos alternativos simulados. A estimativa de incertezas, locais e globais, para atributos numéricos e temáticos, são abordados nas seções seguintes.

#### 4.4.1 Estimativa de incertezas locais: atributos numéricos simulados

Os valores da distribuição de probabilidade univariada em **u**, inferidos pelo procedimento descrito em 4.3.1, podem ser utilizados nas metodologias de cálculo de incerteza apresentadas na Seção 3.5.3.

#### 4.4.2 Estimativa de incertezas locais para atributos temáticos simulados

Os valores da distribuição de probabilidade univariada em **u**, obtidos pelo procedimento descrito em 4.3.2, podem ser utilizados nas metodologias de cálculo de incerteza apresentadas na Seção 3.5.4.

# 4.4.3 Métricas globais de incerteza para atributos simulados

Esta seção é dedicada à estimativa de incertezas globais utilizando a medida de entropia de Shannon para um conjunto de variáveis aleatórias pertencentes à uma função aleatória. Considerando-se a função aleatória discreta  $Z(\mathbf{u})$ , seu conjunto de probabilidades bivariadas  $p_{k,k'}(\mathbf{h}) = Prob\{Z(\mathbf{u}) \in \text{à classe } k \in Z(\mathbf{u}') \in \text{à classe } k'\}$ , k, k'= 1,...,K, e seu conjunto de probabilidades marginais  $p_k = \sum_{l=1}^{L} p_{k,l}(\mathbf{h}) = Prob\{Z(\mathbf{u}) \in \text{à classe } k\}$ , k = 1,...,K. Sejam, ainda,  $Z(\mathbf{u}) \in Z(\mathbf{u}')$ 

duas variáveis aleatórias separadas pelo vetor de separação **h**, tal que, **u**'= **u**+**h**. A entropia associada a esse conjunto de probabilidades bivariadas é definida como:

$$H(\mathbf{h}) = -\sum_{k=1}^{K} \sum_{k'=1}^{K} p_{k,k'}(\mathbf{h}) ln(p_{k,k'}(\mathbf{h}))$$
(4.4)

Da formulação apresentada na equação 4.4, pode-se fazer a seguinte análise:

- Quando  $\mathbf{h} = 0$ ,  $p_{kk'}(\mathbf{0}) = 0$ ,  $\forall k \neq k'$  e  $p_{kk}(\mathbf{0}) = p_k$ . Assim  $H = H(\mathbf{0})$ ;
- Quando  $\mathbf{h} \to +\infty$ , as VA  $Z(\mathbf{u})$  e  $Z(\mathbf{u}')$  são consideradas independentes, ou seja,  $p_{kk'=} p_k p_{k',} \forall k, k'$  e, neste caso, chega-se à  $H(\infty) = 2H(\mathbf{0})$ .

Portanto, a partir dessas análises, tem-se que  $H(\mathbf{h}) \in [H(\mathbf{0}), 2H(\mathbf{0})]$ . Este resultado possibilita a definição de uma medida normalizada, valores entre 0 e 1, para a entropia bivariada, chamada *entropia relativa normalizada*, que é definida por:

$$H_R(\mathbf{h}) = \frac{H(\mathbf{h}) - H(\mathbf{0})}{H(\mathbf{0})} \in [0,1]$$
 (4.5)

Journel e Deutsch (1993) definem uma *entropia relativa média* sobre um campo A, discretizado em m localizações cujas coordenadas são  $\mathbf{u}_i$ , i=1,...,m., de medida |A| como:

$$\overline{H_R}(A, A) \cong \frac{1}{m^2} \sum_{i=1}^m \sum_{j=1}^m H_R(\mathbf{u}_i - \mathbf{u}_j)$$
(4.6)

Como já visto anteriormente, quando  $A \to \infty$ ,  $H_R(\infty) \to 1$  e, portanto,  $\overline{H_R}(A,A) \to 1$ . Ainda, segundo Journel e Deutsch (1993), a medida de entropia relativa média pode ser usada como uma medida global da entropia espacial bivariada. Uma medida global multivariada deveria usar uma função de entropia multivariada. Porém, o autor ressalta que, da mesma forma que as funções de distribuição de probabilidades multivariadas são raramente acessíveis na prática, isto também é verdade para funções de entropia multivariada.

#### 4.5 COMPARAÇÃO ENTRE SIMULAÇÃO CONDICIONADA E KRIGEAGEM

O estimador de krigeagem, apresentado no Capítulo 3, possibilita a construção de funções de distribuição de probabilidade, paramétricas ou não, para um atributo em estudo. A distribuição de probabilidade, definida em cada posição de uma região estacionária de interesse, é usada para se inferir, em cada posição u, um valor único do atributo e uma incerteza associada ao atributo. Dessa forma, a krigeagem, usada como um estimador, cria um único campo aleatório cujos valores compõem uma superfície suavizada. A variabilidade do estimador no espaço é uma versão suavizada da verdadeira variabilidade e não reflete exatamente as flutuações reais (Huijbregts 1973). Esta superfície tem variância menor do que o conjunto amostral, pois o valor estimado para cada variável aleatória é obtido a partir da hipótese de mínima variância do erro de estimação. Duas restrições de uso surgem em relação a esse procedimento:

- 1) O critério de mínima variância do erro de estimação, para a superfície inferida, pode não ser o que se necessita para uma determinada aplicação. Quando se aplica este critério, por ser um filtro passa-baixa, perde-se detalhes que podem ser importantes para modelos ambientais que são sensíveis a padrões de variabilidade local.
- 2) Muitos procedimentos de modelagem espacial, como por exemplo o método de Monte Carlo, trabalham sobre vários valores dos atributos, realizações estocásticas para uma dada localização, para estudar o comportamento do modelo em relação a essa gama de valores.

Resumidamente, as diferenças atribuídas aos procedimentos de krigeagem e simulação estocástica (Deutsch e Journel 1998 e Englund 1993) são:

 A krigeagem é um interpolador que gera um único campo aleatório segundo os critérios de mínima variância e não tendenciosidade do estimador. Portanto, o campo interpolado tem menor grau de variabilidade do que as amostras e reproduz apenas a média das amostras. A simulação, por sua vez, cria vários campos aleatórios que reproduzem características globais e estatísticas, de ordem maior que 1, das amostras. Por exemplo, o histograma das amostras é reproduzido pelos valores simulados.

- 2) A krigeagem fornece um conjunto de representações locais onde a acurácia local prevalece. A simulação fornece representações globais alternativas, onde prevalece a representação de padrões de continuidade espacial, que permite estimativas de acurácia global quando várias localizações são consideradas conjuntamente.
- 3) O procedimento de krigeagem é menos custoso computacionalmente do que a simulação. Para se reproduzir, com um alto grau de acurácia, momentos estatísticos de ordem maior que 1, são necessárias de 500 a dezenas de milhares de campos simulados.

Apesar das diferenças, podem-se destacar pelo menos 3 semelhanças entre os procedimentos de krigeagem e de simulação estocástica:

- São procedimentos geoestatísticos que utilizam um modelo de variografia, definido sobre o conjunto amostral, para estabelecer o comportamento de variabilidade do atributo numa região de interesse.
- 2) São procedimentos que honram o conjunto amostral original, ou seja, os valores atribuídos às amostras não são modificados.
- Possibilitam estimativas de estatísticas e incertezas sobre o atributo em estudo

#### 4.6 CONCLUSÕES

Este capítulo apresentou o procedimento de simulação estocástica condicionada sequencial, que pode ser considerado uma ferramenta mais completa do que a krigeagem pois possibilita inferências baseadas em

distribuições conjuntas. A simulação estocástica deve ser usada, em lugar da krigeagem, nas seguintes situações:

- o objetivo é a criação de modelos que representem, com alto grau de acurácia, momentos estatísticos de ordem maior que 1;
- deseja-se estimar incertezas globais que consideram distribuições conjuntas de variáveis aleatórias dependentes;
- necessita-se de vários campos equiprováveis de um ou mais atributos para ser utilizados em modelagens computacionais e procedimentos de propagação de incertezas, como por exemplo, para a simulação de Monte Carlo.

# **CAPÍTULO 5**

# INFERÊNCIAS COM ESTIMATIVA DE INCERTEZAS COM O USO DE DADOS INDIRETOS

#### 5.1 INTRODUÇÃO

Nos Capítulos 4 e 5, os procedimentos geoestatísticos, de krigeagem e de simulação estocástica, foram formalizados sob a hipótese de existência de um conjunto de amostras pontuais do atributo em estudo. Este conjunto amostral pode ser chamado *dado amostral direto*, ou *dado direto*, por ter sido obtido por medida direta dos valores do atributo de interesse.

Tipicamente, o dado direto é obtido em trabalho de levantamento de campo. Também pode resultar de medidas por instrumentos, localizados em pontos estratégicos, que transmitem continuamente informações sobre o atributo, como por exemplo, as plataformas de coletas de dados meteorológicos.

Por outro lado, podem existir outras fontes de dados, em geral mais abundantes e mais baratas de serem obtidas, que estão correlacionados com o atributo em estudo. Sempre que for possível determinar essas correlações, as fontes de dados auxiliares podem ser incorporadas ao conjunto amostral de um atributo. O conjunto de dados auxiliares é denominado *dado amostral indireto*, ou *dado indireto*, uma vez que resultam de informações obtidas indiretamente de fontes de dados auxiliares.

Uma das maiores vantagens dos procedimentos geoestatísticos por indicação é a possibilidade de uso de dados indiretos, complementares aos dados diretos, para inferências e estimativas de incertezas para valores de atributos temáticos e numéricos. O objetivo deste capítulo é apresentar os conceitos e formalismos matemáticos envolvidos nesta metodologia.

O presente capítulo tem a seguinte organização: a Seção 5.2 discute o problema de subamostragem e de amostragem inadequada para inferências

baseadas nos procedimentos geoestatísticos; na Seção 5.3 aborda-se o problema de codificação por indicação dos dados diretos e indiretos; a Seção 5.4 dedica-se a formalização do procedimento de krigeagem por indicação utilizando dados diretos e indiretos codificados por indicação; a Seção 5.5 descreve e formaliza o procedimento de cokrigeagem por indicação. As Seções 5.6 e 5.7 apresentam otimizações, pelo modelo de Markov-Bayes e por dados colocalizados, respectivamente, que, aplicados ao procedimento de cokrigeagem por indicação, aumentam a eficiência computacional deste procedimento. A Seção 5.8 aborda os temas de inferências e estimativas de incertezas, locais e globais, com o uso de dados indiretos. Finalmente, a Seção 5.9 apresenta conclusões sobre os temas abordados neste capítulo.

# 5.2 OS PROBLEMAS DE SUBAMOSTRAGEM E DE AMOSTRAGEM INADEQUADA PARA INFERÊNCIAS BASEADAS NA GEOESTATÍSTICA

A amostragem deve ser representativa do comportamento espacial do atributo. O problema de subamostragem se apresenta quando o conjunto amostral não contém um número suficiente de amostras para representar o fenômeno com um certo grau de acurácia. A *amostragem inadequada* se caracteriza quando temos um conjunto grande de amostras, porém mal distribuídas ou mal localizadas. Nos dois casos pode-se dizer que a amostragem não representou bem a variação do atributo de interesse.

Uma amostragem mal feita, ou seja, pouco representativa, não pode ser compensada por algoritmos de interpolação sofisticados. *Pontos característicos* do atributo, como valores máximos e mínimos são fundamentais em relação a representatividade das amostras. Também os pontos de inflexão são importantes na amostragem.

A Figura 5.1 ilustra, para um caso unidimensional, problemas com a interpolação por quantidade reduzida de amostras e por desconsideração de alguns valores característicos, pontos de máximo, de mínimo e de inflexão do atributo, na amostragem. Nesta figura, o eixo Z representa valores do atributo

enquanto que o eixo X representa sua posição unidimensional. A curva vermelha da figura representa os valores reais do atributo em função de X. O ajuste, curva azul, foi realizado por um interpolador que é exato, ou seja, passa pelas amostras. Pode-se observar que, neste caso, o ajuste apresentou erros de estimação significativos.

A Figura 5.2 apresenta o resultado do ajuste, pelo mesmo interpolador, tendose, como dado de entrada, um conjunto amostral mais representativo. Neste caso, o número de amostras é maior e as amostras estão melhor distribuídas, ou seja, localizadas nos pontos característicos do atributo. Observa-se que os erros de estimação diminuem em relação aos da Figura 5.1.

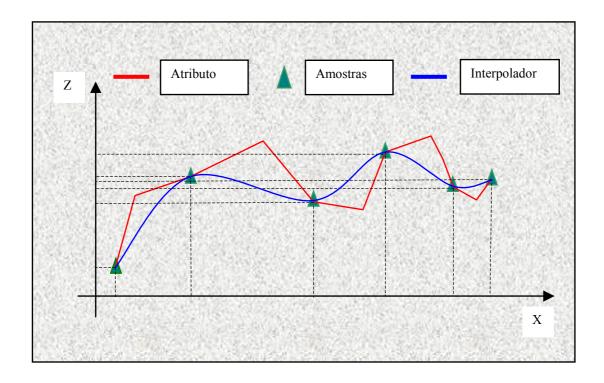


Fig. 5.1 - Exemplo de ajuste para dados de amostragem reduzida e pouco representativa (sem alguns valores máximos, mínimos e pontos de inflexão).

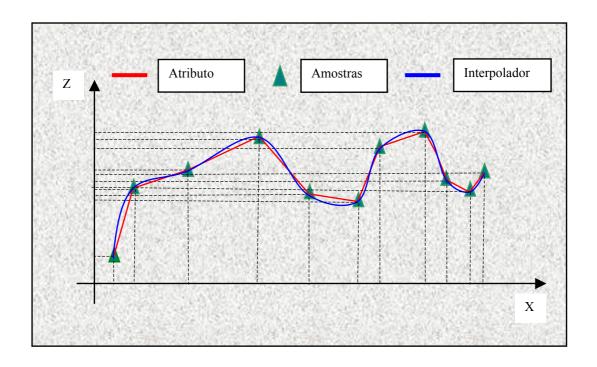


Fig. 5.2 - Exemplo de ajuste para uma amostragem com valores de máximos, mínimos e pontos de inflexão.

No caso bidimensional, linhas características, representando inflexões e descontinuidades, são importantes para informar os interpoladores sobre mudanças bruscas no comportamento do atributo. Amostragens com agrupamentos de dados, com valores redundantes, podem "enganar" (no sentido de criar tendências) algoritmos de interpolação que não consideram a configuração espacial das amostras.

Burrough e MacDonnell (1998) argumentam que quando a quantidade de dados é grande, as amostras são abundantes, a maioria das técnicas de interpolação apresentam resultados similares. Entretanto, quando os dados são esparsos, as suposições feitas sobre o comportamento de variabilidade do atributo no espaço e a escolha do método de interpolação e dos seus parâmetros pode ser crítica quando se quer evitar resultados inesperados.

Como já apresentado no Capítulo 3, as inferências baseadas na geoestatística se utilizam de modelos teóricos ajustados a modelos experimentais de variabilidade, autocovariogramas ou semivariogramas experimentais, obtidos a partir do conjunto amostral. O modelo teórico de variabilidade, em conjunto com as premissas básicas do estimador, não tendenciosidade e mínima variância para o erro de estimação, determinam os valores dos ponderadores de cada amostra usada na estimação. Estes procedimentos consideram, automaticamente, o agrupamentos de amostras. Também o problema de anisotropia do atributo deve ser considerado durante o processo de modelagem do semivariograma (Deutsch e Journel 1998)

Quando o conjunto amostral é esparso e não permite a definição da variografia, dificilmente um interpolador dará resultados com acurácia desejada, principalmente se o atributo tem um comportamento predominantemente errático. Por exemplo, quando na modelagem do variograma predomina o efeito pepita, não se pode determinar uma correlação entre as amostras e, portanto, obtêm-se resultados imprevisíveis ao se inferir valores do atributo a partir dos valores observados.

As razões para uma amostragem ser pouco representativa, esparsa e pouco numerosa, podem ser: dificuldade de acesso às localizações, custo da amostragem, etc.. Como já apresentado nos Capítulos 3 e 4, quando uma amostragem é esparsa, mas possibilita a definição de um ou mais modelos teóricos de variografia, os estimadores geoestatísticos, krigeagem e simulação condicionada por indicação, são capazes de inferir valores e também estimar incertezas associadas aos valores inferidos.

Entretanto, nos Capítulos 3 e 4 foram utilizadas *amostras pontuais diretas*, ou dados diretos, observadas diretamente sobre o atributo em análise. As estimativas baseadas na geoestatística por indicação permitem, também, o uso de *amostras indiretas* que, quando usadas em conjunto com os dados diretos, melhoram a qualidade da estimação. Goovaerts (1997) argumenta que: "A

maior vantagem da metodologia por indicação é sua habilidade para incorporar informações indiretas de vários tipos, além das medidas diretas, no atributo de interesse".

As amostras indiretas são, em geral, mais abundantes e mais fáceis de serem obtidas. Por exemplo, se o comportamento de um atributo tem correlação com a declividade do terreno, pode-se utilizar dados de topografia, geralmente abundantes, para inferência da declividade da região.

Amostras indiretas podem ser observadas em mapas preexistentes, tais como mapas de solos, mapas de geologia e outros. Outras fontes importantes de dados indiretos são: dados pontuais de uma variável determinística adicional; dados pontuais de uma variável estocástica adicional e; informação a priori sobre os valores e tipo da distribuição da variável aleatória.

Stein (1994) argumenta que faz parte do domínio da estatística espacial o uso da informação a priori durante a análise estatística do dado espacial. A informação a priori, sobre as localizações não amostradas, pode ser obtida da distribuição dos dados indiretos, como por exemplo as distribuições de um atributo numa determinada classe de solos. Também, probabilidades globais, provenientes da distribuição global dos dados diretos, podem ser usadas como fontes de informação a priori.

Esta abordagem estende o conceito de *conjunto amostral* para dados obtidos direta e indiretamente. As seções seguintes mostram como esse conjunto amostral é codificado por indicação e, também, como ele é usado pelos procedimentos de krigeagem e simulação estocástica.

#### 5.3 CODIFICAÇÃO POR INDICAÇÃO DE DADOS AMOSTRAIS DIRETOS E INDIRETOS

Segundo Deutsch e Journel (1998), o conjunto amostral, composto de amostras diretas e indiretas, é utilizado para definir probabilidades a priori, em posições não amostradas, e também para atualizá-las para probabilidades a

posteriori, segundo o paradigma Bayesiano. O autor apresenta formalmente 4 fontes distintas de dados a priori, com respectivas codificações por indicação segundo um valor de corte  $z = z_k$ . Estas fontes são:

(a) dados locais diretos  $z(\mathbf{u}_{\alpha})$ , com a seguinte codificação por indicação :

$$i(\mathbf{u}_{\alpha}; z_{k}) = \begin{cases} 1, & \text{se } z(\mathbf{u}_{\alpha}) \leq z_{k} \\ 0, & \text{se } z(\mathbf{u}_{\alpha}) > z_{k} \end{cases} \text{ para atributos numéricos}$$
ou
$$i(\mathbf{u}_{\alpha}; z_{k}) = \begin{cases} 1, & \text{se } z(\mathbf{u}_{\alpha}) = z_{k} \\ 0, & \text{se } z(\mathbf{u}_{\alpha}) \neq z_{k} \end{cases} \text{ para atributos temáticos}$$

$$(5.1)$$

(b) dados locais indiretos com restrições de intervalo,  $z(\mathbf{u}_{\alpha}) \in (\mathbf{a}_{\alpha}, \mathbf{b}_{\alpha}]$ , codificados por indicação pela seguinte formulação:

$$j(\mathbf{u}_{\alpha}; z_{k}) = \begin{cases} 0, & \text{se } z_{k} \leq \mathbf{a}_{\alpha} \\ \text{indefinido, se } z_{k} \in (\mathbf{a}_{\alpha}, \mathbf{b}_{\alpha}] \\ 1, & \text{se } z_{k} > \mathbf{b}_{\alpha} \end{cases}$$
 (5.2)

(c) *informação global a priori*, comum a todas as posições **u**, dentro da região estacionária *A*, e codificada como:

$$F(z_k) = Prob\{Z(\mathbf{u}) \le z_k\}, \forall \mathbf{u} \in A$$
(5.3)

(d) dados locais indiretos,  $y(\mathbf{u}_{\alpha}; \mathbf{z}_{k})$ , que fornecem probabilidades a priori sobre os valores  $\mathbf{z}(\mathbf{u}_{\alpha})$ , ou seja:

$$y(\mathbf{u}_{\alpha}; z_{k}) = Prob\{Z(\mathbf{u}_{\alpha}) \le z_{k} \mid \text{informação local}\}$$
 (5.4)

onde  $y(\mathbf{u}_{\alpha}; z_k) \in [0,1]$  e  $y(\mathbf{u}_{\alpha}; z_k)$  é diferente da informação global  $F(z_k)$ .

A seção seguinte mostra como os dados a priori, após serem codificados por indicação, são usados nos procedimentos de krigeagem por indicação.

#### 5.4 A KRIGEAGEM POR INDICAÇÃO COM DADOS DIRETOS E INDIRETOS

Considere o estimador de krigeagem por indicação definido em 3.17. Nessa formulação, o valor de  $F^*(z_k)$ , a média da função aleatória da região estacionária, que não depende da posição  $\mathbf{u}$ , representa um dado global a priori que é comum para todas as posições não amostradas. Esta média pode ser substituída pela informação de probabilidade local a priori obtida a partir de dados indiretos. Considerando-se, como exemplo, um mapa de solos com suas K classes  $\{s_1, s_2, ..., s_k\}$  de solo, a média  $F^*(z_k)$  pode ser redefinida como:

$$F^*(z_k) = y(\mathbf{u}, z_k) = F^*(z_k \mid s_l) = \text{Prob}\{Z(\mathbf{u}) < z_k \mid \text{classe de solo} = s_l\}$$
 (5.5)

Incluindo-se a informação a priori do dado amostral indireto, *o estimador de krigeagem por indicação simples*, condicionado às *n1* amostras diretas e às *K* classes do dado indireto, é definido pela seguinte formulação:

$$F_{S}^{*}(\mathbf{u}; z_{k} \mid (n+K)) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{S} (\mathbf{u}; z_{k}) i(\mathbf{u}_{\alpha}; z_{k}) + \left[1 - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{S} (\mathbf{u}; z_{k})\right] y(\mathbf{u}; z_{k})$$

$$= y(\mathbf{u}; z_{k}) + \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{S} (\mathbf{u}; z_{k}) [i(\mathbf{u}_{\alpha}; z_{k}) - y(\mathbf{u}_{\alpha}; z_{k})]$$
(5.6)

Neste caso, o dado indireto,  $Y(\mathbf{u}; z_k)$ , informa sobre as tendências locais do dado direto codificado por indicação  $I(\mathbf{u}; z_k)$  (Goovaerts e Journel 1995). A expressão 5.6 representa uma atualização da probabilidade a priori  $y(\mathbf{u}; z_k)$ , na posição  $\mathbf{u}$ , utilizando-se os valores dos dados diretos e indiretos, codificadas por indicação, da vizinhança de  $\mathbf{u}$ . Obtêm-se os pesos de krigeagem,  $\lambda_{S\alpha}(\mathbf{u}; z_k)$ , pela solução do sistema de equações de krigeagem por indicação simples:

$$\sum_{\beta=1}^{n(\mathbf{u})} \lambda_{S\beta}(\mathbf{u}; z_k) C_R(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}; z_k) + \phi(\mathbf{u}; z_k) = C_R(\mathbf{u}_{\alpha} - \mathbf{u}; z_k) \qquad \forall \alpha = 1, 2, ..., n(\mathbf{u})$$
(5.7)

sendo  $\phi(\mathbf{u}; z_k)$  o multiplicador de Lagrange e  $C_R(\mathbf{h})$ , com  $\mathbf{h} = \mathbf{u}_{\alpha} - \mathbf{u}_{\beta}$  ou  $\mathbf{h} = \mathbf{u}_{\alpha} - \mathbf{u}_{\beta}$  a covariância da função aleatória residual,  $R(\mathbf{u}; z_k) = I(\mathbf{u}; z_k) - Y(\mathbf{u}; z_k)$ , em  $z = z_k$ .

#### 5.5 A COKRIGEAGEM POR INDICAÇÃO

A krigeagem por indicação simples, apresentada na seção anterior, requer o conhecimento da informação indireta nas localizações das amostras diretas,  $\mathbf{u}_{\alpha}$ , e dos pontos inferidos,  $\mathbf{u}$ . Quando os dados indiretos são observados em posições distintas das posições das amostras diretas, os valores observados para os dados locais diretos e indiretos,  $i(\mathbf{u}; \mathbf{z}_k)$  e  $y(\mathbf{u}; \mathbf{z}_k)$ , podem ser interpretados como realizações de duas variáveis aleatórias correlacionadas,  $I(\mathbf{u}; \mathbf{z}_k)$  e  $Y(\mathbf{u}; \mathbf{z}_k)$ . Esses dados, codificados por indicação, podem ser combinados através do procedimento de *cokrigeagem por indicação* considerando-se  $I(\mathbf{u}; \mathbf{z}_k)$  e  $Y(\mathbf{u}; \mathbf{z}_k)$  como variáveis primárias e secundárias, respectivamente.

O estimador de cokrigeagem por indicação ordinária, que atualiza a probabilidade a priori  $y(\mathbf{u}; z_k)$ , numa localização  $\mathbf{u}$  não amostrada, usando os dados por indicação diretos e indiretos, tem a seguinte formulação:

$$F^{*}(\mathbf{u}; z_{k} \mid (n1+n2)) = \upsilon(\mathbf{u}; z_{k}) y(\mathbf{u}; z_{k}) + \sum_{\alpha=1}^{n1(\mathbf{u})} \lambda_{\alpha}(\mathbf{u}; z_{k}) i(\mathbf{u}_{\alpha}; z_{k}) + \sum_{\beta=1}^{n2(\mathbf{u})} \upsilon_{\beta}(\mathbf{u}; z_{k}) y(\mathbf{u}'_{\beta}; z_{k})$$
(5.8)

onde: n1 é o número de dados diretos; n2 é o número de dados indiretos;  $v(\mathbf{u}; \mathbf{z}_k)y(\mathbf{u}; \mathbf{z}_k)$  define um valor de probabilidade global em  $\mathbf{u}$ , a partir dos dados indiretos;  $\lambda_{\alpha}(\mathbf{u}; \mathbf{z}_k)$  é o peso de cokrigeagem da variável primária observada em  $\mathbf{u}_{\alpha}$  e  $v_{\beta}(\mathbf{u}; \mathbf{z}_k)$  é o peso de cokrigeagem da variável secundária observada em  $\mathbf{u}_{\beta}$ .

Os pesos de cokrigeagem são determinados a partir da solução do seguinte sistema de cokrigeagem ordinária de  $(n1(\mathbf{u})+n2(\mathbf{u})+1)$  equações:

$$\begin{bmatrix}
\sum_{\beta=1}^{nI(\mathbf{u})} \lambda_{\beta}(\mathbf{u}; z_{k}) C_{I}(\mathbf{u}_{\beta} - \mathbf{u}_{\alpha}; z_{k}) + \\
\sum_{\beta=0}^{n2(\mathbf{u})} \upsilon_{\beta}(\mathbf{u}; z_{k}) C_{IY}(\mathbf{u}'_{\beta} - \mathbf{u}_{\alpha}; z_{k}) + \phi(\mathbf{u}; z_{k}) = C_{I}(\mathbf{u} - \mathbf{u}_{\alpha}; z_{k}) \quad \forall \alpha = 1, ..., n I(\mathbf{u})
\end{bmatrix}$$

$$\begin{cases}
\sum_{\beta=1}^{nI(\mathbf{u})} \lambda_{\beta}(\mathbf{u}; z_{k}) C_{IY}(\mathbf{u}_{\beta} - \mathbf{u}'_{\alpha}; z_{k}) + \\
\sum_{\beta=0}^{n2(\mathbf{u})} \upsilon_{\beta}(\mathbf{u}; z_{k}) C_{Y}(\mathbf{u}'_{\beta} - \mathbf{u}'_{\alpha}; z_{k}) + \phi(\mathbf{u}; z_{k}) = C_{IY}(\mathbf{u} - \mathbf{u}'_{\alpha}; z_{k}) \quad \forall \alpha = 0, ..., n 2(\mathbf{u})
\end{cases}$$

$$\begin{bmatrix}
\sum_{\beta=1}^{nI(\mathbf{u})} \lambda_{\beta}(\mathbf{u}; z_{k}) + \sum_{\beta=0}^{n2(\mathbf{u})} \upsilon_{\beta}(\mathbf{u}; z_{k}) = 1
\end{bmatrix}$$
(5.9)

onde:  $\phi(\mathbf{u}; z_k)$  é o parâmetro de Lagrange;  $C_l(\mathbf{h}; z_k)$  e  $C_Y(\mathbf{h}; z_k)$  são as autocovariâncias das funções aleatórias por indicação primária e secundária, respectivamente e;  $C_{lY}(\mathbf{h}; z_k)$  é a covariância cruzada entre essas funções. Outras condições de não tendenciosidade devem ser consideradas caso  $E\{l(\mathbf{u}; z_k)\}$  e  $E\{Y(\mathbf{u}; z_k)\}$  sejam diferentes (Deutsch e Journel 1998).

A solução do sistema de equações definido em 5.9 requer, para cada valor de corte  $z = z_k$ , o cálculo das duas autocovariâncias,  $C_l(\mathbf{h}; z_k)$  e  $C_Y(\mathbf{h}; z_k)$ , e da covariância conjunta cruzada,  $C_{lY}(\mathbf{h}; z_k)$ , das duas variáveis. Essas covariâncias podem ser inferidas diretamente quando se tem uma quantidade suficientemente grande de dados diretos e indiretos. Na prática isto raramente ocorre. Além disso, este procedimento é complexo pois requer computação intensa, para cálculo das covariâncias cruzadas entre as classes, e apresenta severas restrições quando o número de classes cresce levando a um aprofundamento dos problemas de desvios de relação de ordem (Goovaerts e Journel 1995). Por essas razões, algumas otimizações foram sugeridas no sentido de se minimizar esse esforço procedimental. Duas destas otimizações são apresentadas nas seções seguintes.

#### 5.6 OTIMIZAÇÃO PELO MODELO MARKOV-BAYES

Uma primeira otimização, denominada modelo de Markov-Bayes, propõe a inferência das covariâncias  $C_Y(\mathbf{h}; z_k)$  e  $C_{IY}(\mathbf{h}; z_k)$ , a partir de  $C_I(\mathbf{h}; z_k)$ , utilizando as seguintes relações:

$$C_{IY}(\mathbf{h}, z_k) = B(z_k) C_I(\mathbf{h}, z_k), \forall \mathbf{h}$$

$$C_Y(\mathbf{h}, z_k) = B^2(z_k) C_I(\mathbf{h}, z_k), \forall \mathbf{h} > 0$$

$$= |B(z_k)| C_I(\mathbf{h}, z_k), \mathbf{h} = 0$$
(5.10)

Os coeficientes,  $B(z_k)$ 's, são obtidos pela calibração das variáveis secundárias com as variáveis primárias segundo a expressão:

$$B(z_k) = m^{(1)}(z_k) - m^{(0)}(z_k) \in [-1, +1]$$
(5.11)

Essa calibração requer o conhecimento de um conjunto de dados codificados por indicação,  $\{y(\mathbf{u}_{\alpha}, \mathbf{z}_k), i(\mathbf{u}_{\alpha}, \mathbf{z}_k), \alpha=1,...,\mathbf{n}\}$ , onde os valores das variáveis secundárias são comparadas com os valores das variáveis primárias. A partir desse conjunto de dados, pode-se calcular as médias:

$$m^{(1)}(z_k) = \mathbb{E}[Y(\mathbf{u}, z_k) | I(\mathbf{u}, z_k) = 1]$$
 (5.12)

que é a média dos valores de y correspondentes a i=1, e

$$m^{(0)}(z_k) = E[Y(\mathbf{u}, z_k) | I(\mathbf{u}, z_k) = 0]$$
 (5.13)

que é a média dos valores de y correspondentes a i=0, para  $z=z_k$ .

O parâmetro  $B(z_k)$  é um *índice da acurácia para a informação indireta*. Ele mede a capacidade da variável secundária separar os casos  $i(\mathbf{u}; z_k)=1$  e  $i(\mathbf{u}; z_k)=0$ . A melhor situação ocorre quando  $B(z_k)=1$ , ou seja, quando  $m^{(1)}(z_k)=1$ , caso em que todos os valores de y predizem exatamente a saída  $i(\mathbf{u}; z_k)=1$ , e quando  $m^{(0)}(z_k)=0$ . Neste caso, a variável secundária  $y(\mathbf{u}; z_k)$  é tratada como

proveniente de um dado direto e não é atualizada. Quando  $B(z_k) = 0$  a variável secundária é ignorada.

Testes experimentais com o modelo de Markov-Bayes mostram que a covariância cruzada  $C_{IY}(\mathbf{h}; z)$  é mais acurada do que a autocovariância da variável secundária  $C_{Y}(\mathbf{h}; z)$ . Portanto, a validade da aproximação de Markov-Bayes deve ser verificada, particularmente o valor de autocovariância da variável secundária (Deutsch e Journel 1998).

#### 5.7 OTIMIZAÇÃO PARA DADOS COLOCALIZADOS

Uma segunda metodologia de otimização, sugerida por Almeida e Journel (1993), supõe ser conhecido o valor da variável secundária por indicação,  $y(\mathbf{u}; \mathbf{z}_k)$ , em cada ponto  $\mathbf{u}$  a ser estimado. Neste caso, a expressão de cokrigeagem por indicação, apresentada em (5.8), fica simplificada para:

$$F^*(\mathbf{u}; z_k \mid (n \mid I)) = \upsilon(\mathbf{u}; z_k) y(\mathbf{u}; z_k) + \sum_{\alpha=1}^{n \mid l(\mathbf{u})} \lambda_{\alpha}(\mathbf{u}; z_k) i(\mathbf{u}_{\alpha}; z_k)$$
(5.14)

A desvantagem deste procedimento é a obrigatoriedade de se dispor dos dados indiretos por indicação nas posições a serem estimadas. Caso isto não ocorra, porém existe uma grande quantidade de dados indiretos, pode-se estimar os valores da variável secundária, nas posições  $\mathbf{u}$ 's, a partir do procedimento de simulação estocástica, condicionada aos dados secundários existentes. Por outro lado, a simplificação proposta tem as seguintes vantagens (Goovaerts e Journel 1995): é rápida e fácil de ser implementada; evita problemas de instabilidade, nas matrizes de cokrigeagem, associados com múltiplas redundâncias das variáveis secundárias; não requer a inferência do modelo de autocovariância,  $C_Y(\mathbf{h}; z_k)$ , da variável secundária e; o valor da covariância cruzada  $C_{IY}(\mathbf{h}; z_k)$  pode ser obtido, a partir da covariância  $C_I(\mathbf{h}; z_k)$ , pelo modelo de Markov-Bayes apresentado na Seção 5.6.

### 5.8 INFERÊNCIA E ESTIMATIVA DE INCERTEZAS LOCAIS E GLOBAIS COM O USO DE DADOS DIRETOS E INDIRETOS

Os valores de distribuição,  $F(z_k, \mathbf{u} \mid \text{conjunto amostral})$ , estimados segundo as formulações apresentadas nas Seções 5.4, 5.5 e 5.7, são utilizados para se construir uma aproximação discretizada da fdac da VA na posição  $\mathbf{u}$ . A diferença, em relação as metodologias apresentadas nos Capítulos 3 e 4, consiste na consideração de um conjunto amostral formado por dados indiretos em conjunto com os dados diretos do atributo. A fdac aproximada é usada na inferência de estatísticas da distribuição da VA e, também, na estimativa das incertezas locais e globais da FA. Os procedimentos utilizados nas inferências das estatísticas e das incertezas, para atributos numéricos e contínuos, são os mesmos já apresentados nas Seções 3.5.1 a 3.5.4 e subseções das Seções 4.3 e 4.4.

#### 5.9 CONCLUSÕES

Este capítulo apresentou os formalismos matemáticos para se integrar amostras indiretas, em conjunto com amostras diretas, nos procedimentos de krigeagem e cokrigeagem por indicação.

Como já citado no Capítulo 4, o procedimento de simulação sequencial requer a determinação de m fdac's univariadas:  $Prob\{Z_1 \le z_1 | (n)\}$ ,  $Prob\{Z_2 \le z_2 | (n+1)\}$ , ...,  $Prob\{Z_m \le z_m | (n+m-1)\}$ . Estas fdac's podem ser estimadas pelos algoritmos de krigeagem e cokrigeagem por indicação apresentados neste capítulo. Portanto pode-se afirmar que os procedimentos de simulação estocástica sequencial condicionada também podem fazer uso dos conceitos e das vantagens do uso de dados indiretos, apresentados neste capítulo.

A principal razão para uso de dados indiretos, em conjunto com dados diretos, é a de se obter resultados de estimativas, que utilizam os procedimentos de inferência da geoestatística, com um grau maior de acurácia. Esta premissa é

testada na prática, no Capítulo 6, em estudos de casos que comparam as metodologias de inferência com e sem uso de dados indiretos.

# CAPÍTULO 6 <u>ESTUDOS DE CASOS E ANÁLISES</u>

#### 6.1 INTRODUÇÃO

Este capítulo desenvolve estudos de casos para aplicar, testar e analisar os conceitos, principalmente os de inferências e de estimativas de incertezas, apresentados e formalizados nos capítulos anteriores. Para isto são utilizados conjuntos amostrais pontuais de natureza numérica, amostras de altimetria, e de natureza temática, amostras de classes de textura de solo. Apresenta-se, também, um exemplo de uso de informação indireta, inferida de um mapa de solos, para melhorar a acurácia da inferência do atributo temático textura de solo. Todos os dados amostrais, diretos e indiretos, foram obtidos em trabalho de levantamento de campo executado na fazenda Canchim, uma fazenda experimental da Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA). O apêndice A contém uma listagem dos dados utilizados neste capítulo. Os resultados dos estudos de caso são apresentados, na forma de mapas ou tabelas, e são analisados e comparados com outras metodologias tradicionalmente utilizadas. O objetivo principal é mostrar as vantagens de se utilizar o paradigma de modelagem estocástica por indicação apresentado no Capítulo 2 deste trabalho.

O presente capítulo está organizado da seguinte forma: a Seção 6.2 apresenta e analisa estudos de caso com o atributo numérico altimetria enfatizando-se as diferenças entre as estimativas obtidas pelos procedimentos de krigeagem linear e de krigeagem e simulação por indicação; a Seção 6.3 está direcionada para estudos de caso com a propriedade temática textura de solo onde se comparam resultados obtidos por krigeagem e por simulação por indicação; a Seção 6.4 mostra e analisa um exemplo de uso de dados indiretos para espacializar a textura de solo; na Seção 6.5 apresenta-se um exemplo de medidas globais de incerteza baseado na distribuição bivariada de um atributo

espacial e; finalmente, a Seção 6.6 reporta conclusões referentes aos resultados e análises apresentados neste capítulo.

#### 6.2 CARACTERIZAÇÃO DA ÁREA UTILIZADA NOS ESTUDOS DE CASO

A área de estudo 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°55'00" à 21°59'00", latitudes sul, e 47°48'00" à 41°52'00", longitudes oeste. A Figura 6.1 mostra a localização da área de estudo. Os dados de localização da fazenda, em coordenadas UTM, estão listados na Seção 1 do apêndice A.

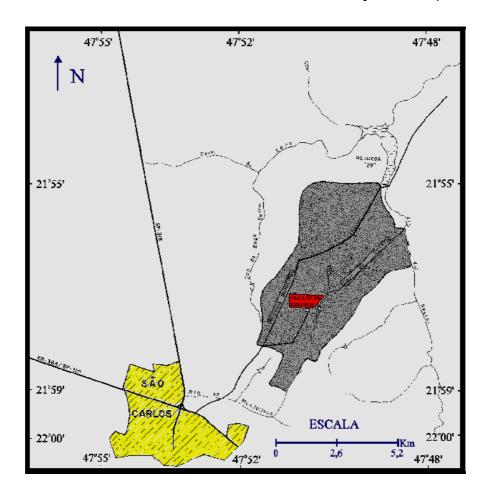


Fig. 6.1 - Localização da Fazenda Canchim. FONTE: Calderano Filho et al. (1996, p. 4).

As principais características da região de Canchim são (Camargo 1997):

- a região ocupa o reverso de cuestas, pertencente à província geomorfológica "Cuestas Basálticas" (Almeida 1964, citado por Calderano Filho et al. 1996);
- o clima, segundo a classificação de Köppen, é temperado úmido de invernos secos e verões brandos;
- o relevo predominante é suave ondulado a ondulado, com pendentes longas, dissecado por forças erosivas, com desníveis variáveis;
- o material de origem é composto além do produto do intemperismo de rochas eruptivas básicas e sedimentares retrabalhadas, por depósitos mais recentes e:
- a vegetação primária é constituída por cerradões, cerrados e florestas tropicais subcaducifólias, estando sua ocorrência muito correlacionada com o material de origem, textura e umidade do solo.

O levantamento de solos da área na escala 1:10000 foi finalizado em 1997. Este levantamento objetivou identificar as distribuições das classes de solos de forma a orientar a pesquisa no planejamento do uso dos solos, principalmente no que se refere a pastagem. Além das classes de solo foram levantados outros atributos importantes tais como: altimetria, teores de cálcio e magnésio, quantidade de argila do solo, etc..

### 6.3 ESTUDO DA PROPRIEDADE ALTIMETRIA SOB REPRESENTAÇÃO NUMÉRICA: CASO CANCHIM

Nesta seção estuda-se a espacialização de um atributo de natureza contínua, a altimetria da região. No levantamento de solos, adquiriu-se um conjunto de amostras de dados numéricos de elevação, distribuídos na região de Canchim, como mostra a Figura 6.2. Os dados das localizações, em coordenadas UTM, e

dos valores de altimetria, em metros, deste conjunto amostral estão listados na Seção 2 do Apêndice A.

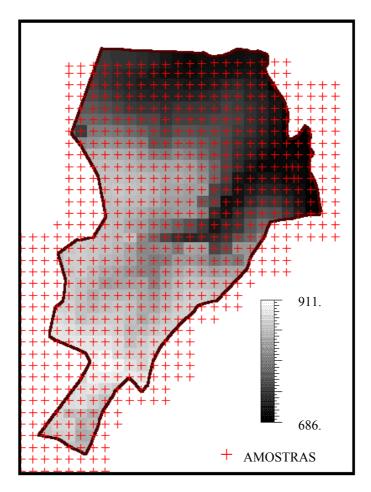


Fig. 6.2 - Distribuição espacial das amostras, sobreposta à representação em imagem nível de cinza, da altimetria de Canchim

A imagem em níveis de cinza, que aparece como "pano de fundo" da Figura 6.2, foi obtida por um procedimento de interpolação por vizinho mais próximo. Essa é a forma mais simples de se espacializar um atributo a partir de um conjunto amostral pontual. A cada pixel dessa imagem atribuiu-se um valor de nível de cinza proporcional valor da altimetria da amostra mais próxima do pixel. Assim, a Figura 6.2, mostra, além da distribuição espacial das amostras, a variação do atributo altimetria, segundo o estimador por vizinho mais próximo, dentro da região de interesse.

#### 6.3.1 Inferências com estimativa de incertezas por krigeagem ordinária

A Figura 6.3 apresenta o resultado do uso do procedimento de krigeagem ordinária para espacializar, segundo uma grade regular, a propriedade altimetria utilizando o conjunto amostral ilustrado na Figura 6.2.

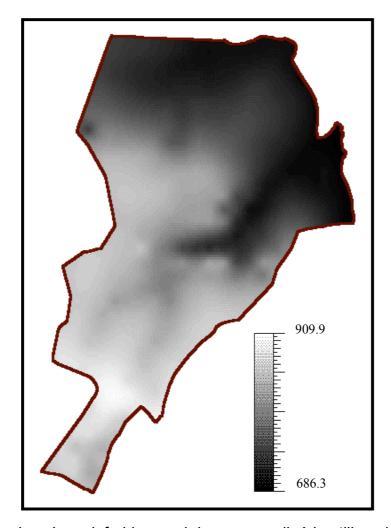


Fig. 6.3 - Mapa de valores inferidos por krigeagem ordinária utilizando o conjunto amostral de altimetria

A geração da grade regular foi precedida por uma análise exploratória dos dados amostrais para se definir um modelo de variabilidade para esse atributo.

Utilizando-se procedimentos de análise exploratória do módulo de geoestatística (Camargo 1997 e Felgueiras et al. 1999b) do Sistema de Processamento de Informações Georeferenciadas (SPRING) (Câmara et al. 1996a), obteve-se um conjunto de estatísticas, apresentado na Tabela 6.1, sobre os dados amostrais de altimetria.

TABELA 6.1 - PARÂMETROS ESTATÍSTICOS REFERENTES AO CONJUNTO AMOSTRAL DE ALTIMETRIA DE CANCHIM

Estatística	Valor
Número de amostras	662
Média	800.596
Variância	4481.662
Desvio Padrão	66.945
Coeficiente de Variação	0.084
Coeficiente de Assimetria	-0.296
Coeficiente de Curtose	1.562
Valor Mínimo	687.000
Quartil Inferior	732.500
Mediana	827.000
Quartil Superior	859.500
Valor Máximo	911.000

As ferramentas de análise exploratória do módulo de geoestatística do SPRING possibilitam, também, cálculo e apresentação de histogramas das amostras, de gráficos de probabilidade normal e de gráficos de dispersão para duas variáveis aleatórias escolhidas. A Figura 6.4 mostra o histograma do conjunto amostral de altimetria de Canchim.

Este módulo de geoestatística permite, ainda, construir variogramas experimentais, ajustar esses variogramas a modelos teóricos, validar os modelos de ajuste e estimar valores, e variâncias da estimação, a partir de um algoritmo de krigeagem ordinária. A quase totalidade destas ferramentas foram utilizadas durante o desenvolvimento deste capítulo. Assim, utilizando-se este módulo, chegou-se, para o conjunto amostral de altimetria em estudo, a um modelo teórico de variografia do tipo esférico e isotrópico com os seguintes

parâmetros: efeito pepita igual a 0.02, contribuição igual a 0.782 e alcance igual a 1.025 m.

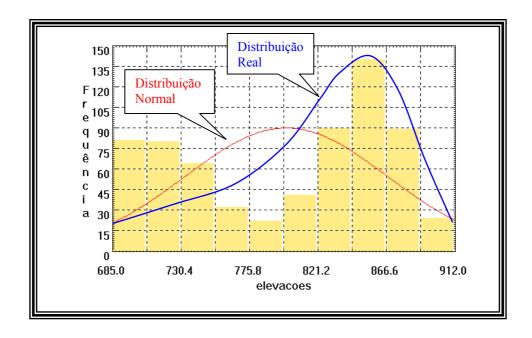


Fig. 6.4 - Histograma do conjunto amostral de altimetria de Canchim

Para a espacialização da altimetria por krigeagem ordinária, utilizou-se o programa kt3d.exe da Geostatistical Software Library (GSLIB) (http://ekofisk.stanford.edu/SCRF.html), uma biblioteca pública de funções geoestatísticas. Definiu-se a informação planimétrica da grade regular: por um ponto de referência, inferior esquerdo, com abscissa igual a 204017.5 e ordenada igual a 7565025.0; pelas suas resoluções horizontais e verticais iguais a 35 m e 50 m, respectivamente e; pelo tamanho da grade igual a 200 linhas por 200 colunas. Estabeleceu-se, ainda, um raio de busca máximo igual a 2000 m e o número de pontos condicionantes mínimo e máximo iguais a 1 e 12, respectivamente.

Os valores do atributo, mostrados na Figura 6.3, representam valores de altimetria inferidos pelo interpolador de krigeagem ordinária. Em conjunto com

os valores estimados, inferiu-se os valores de variância do erro de estimação utilizando-se a formulação apresentada na equação 3.14. Estas variâncias foram utilizadas para gerar o campo da Figura 6.5, que mostra o mapa de incertezas gerado por valores de incerteza local iguais a 2 (dois) desvios padrão no total, que equivale, para distribuições gaussianas, a um intervalo de confiança de 68% de probabilidade, como apresentado na equação 3.28. Este é o campo de incertezas que comumente se infere quando a modelagem de um atributo numérico é realizada por procedimentos de krigeagem linear. Este mapa deve ser comparado com os mapa de incertezas gerados, e apresentados nas seções seguintes, por procedimentos baseados na krigeagem por indicação.

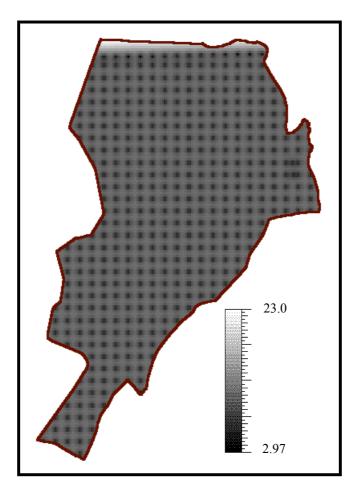


Fig. 6.5 - Mapa de incertezas locais obtido a partir da variância do erro de krigeagem ordinária.

## 6.3.2 Inferências com estimativa de incertezas por krigeagem por indicação

A Figura 6.6 mostra o resultado do uso do procedimento de krigeagem por indicação para espacializar, segundo uma grade regular, a propriedade altimetria. Para isto utilizou-se o conjunto amostral apresentado na Figura 6.2. Os valores do atributo, para cada ponto da grade, foram estimados pelo valor médio, como formulado pela aproximação apresentada na equação 3.21.

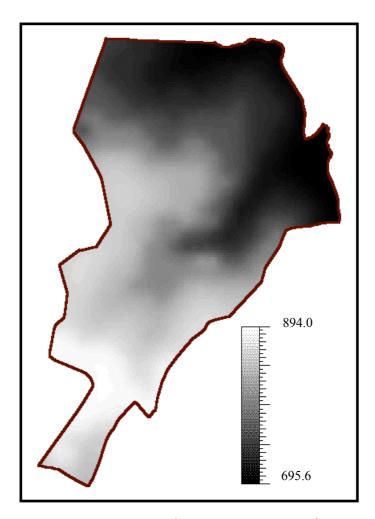


Fig. 6.6 - Mapa de valores médios inferidos a partir das fdac locais construídas pelo procedimento de krigeagem por indicação

Para realizar a krigeagem por indicação para os dados de altimetria, dividiu-se o conjunto amostral em 10 subconjuntos selecionando-se 9 valores de cortes. Cada um dos 10 subconjuntos foi definido com, aproximadamente, a mesma cardinalidade. Para cada valor de corte realizou-se a transformação por indicação, como definido em 3.15, sobre as amostras originais. Criaram-se, assim, 9 conjuntos amostrais por indicação. Para cada um desses conjuntos, utilizando-se o módulo de geoestatística do SPRING, definiu-se um modelo teórico de variografia para representar a variabilidade espacial da variável por indicação. A Tabela 6.2 mostra os resultados obtidos para cada valor de corte. Ajustou-se um modelo teórico de variografia do tipo esférico e isotrópico para todos os subconjuntos por indicação.

TABELA 6.2 - PARÂMETROS DOS VARIOGRAMAS TEÓRICOS DEFINIDOS PARA OS DADOS DE ALTIMETRIA TRANSFORMADOS POR INDICAÇÃO EM VÁRIOS VALORES DE CORTE

Valor de corte	Efeito pepita	Contribuição	Alcance
703.1	0.020	0.060	3172.0
719.1	0.014	0.150	4874.0
744.1	0.015	0.228	5955.0
779.1	0.011	0.202	4855.0
826.1	0.010	0.218	4950.0
841.1	0.026	0.200	5049.0
854.1	0.030	0.162	4016.0
863.1	0.024	0.123	3606.0
876.1	0.014	0.065	2061.0

A fdac, para cada ponto da grade, foi construída pelo programa de krigeagem por indicação, *ik3d.exe*, fornecido pela biblioteca GSLIB. Para executar este procedimento, definiu-se a distribuição planimétrica da grade regular por um ponto de referência, inferior esquerdo, com abscissa igual a 204017.5 e

ordenada igual a 7565025.0, pelas suas resoluções horizontais e verticais iguais a 35 m e 50 m, respectivamente, e pelo tamanho da grade igual a 200 linhas por 200 colunas. Estabeleceu-se, ainda, um raio de busca máximo igual a 2000 m e o número de pontos condicionantes entre 1 e 12.

A Figura 6.7 apresenta um campo de incertezas locais obtido a partir das fdacs construídas pelo procedimento de krigeagem por indicação. Utilizou-se a aproximação da equação 3.22 para se obter, para cada ponto da grade, um valor de variância, e portanto de desvio padrão, em torno do valor médio.

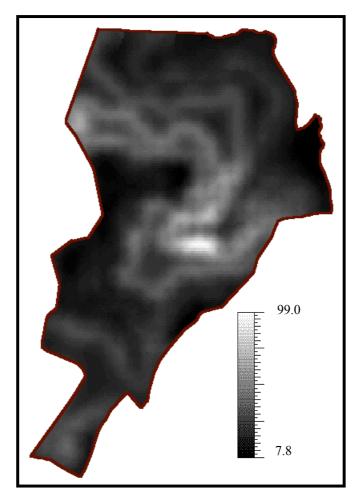


Fig. 6.7 - Mapa de incertezas locais obtido a partir das variâncias inferidas dos modelos de distribuição probabilística construídos pela krigeagem por indicação.

O mapa da Figura 6.7 foi gerado por valores de incerteza local iguais a 2 (dois) desvios padrão no total. Isto equivale, para distribuições gaussianas, a um intervalo de confiança de 68% de probabilidade, como apresentado na equação 3.28

A Figura 6.8 apresenta o mapa de valores de altimetria estimados por medianas. O procedimento de espacialização foi idêntico ao utilizado para gerar o mapa da Figura 6.6, substituindo-se os valores médios pelos valores das medianas obtidos segundo a equação 3.23.

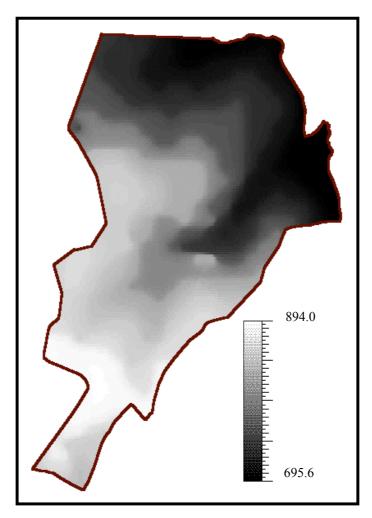


Fig. 6.8 - Mapa de medianas inferidas a partir das fdac locais construídas pelo procedimento de krigeagem por indicação

Um mapa de incertezas obtido pelos valores de uma grade de intervalos interquartis, diferença entre o primeiro e o terceiro quartil, de altimetria, e estimados segundo a equação 3.24, está apresentado na Figura 6.9. Os valores de altimetria do primeiro e terceiro quartis foram estimados das fdac construídas pelo procedimento de krigeagem por indicação.

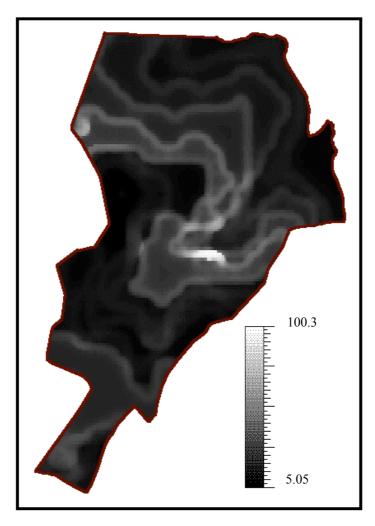


Fig. 6.9 - Mapa de incertezas locais obtido a partir dos quartis, primeiro e terceiro, inferidos dos modelos de distribuição probabilística construídos pela krigeagem por indicação.

A Figura 6.10 mostra o resultado de um classificador por intervalos de probabilidade. Utilizou-se as fdac, construídas pelo procedimento de krigeagem por indicação a partir das amostras da Figura 6.2, para se gerar um mapa temático classificado em 3 faixas distintas de valores de altimetria.

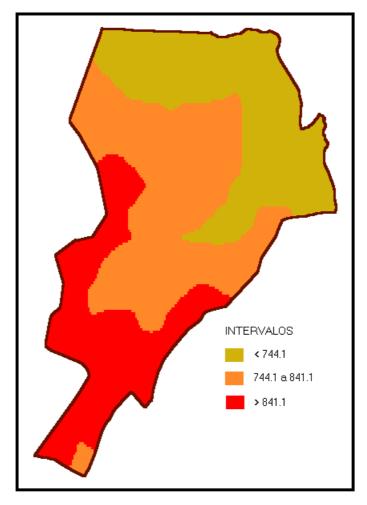


Fig. 6.10 - Mapa de classes de altimetria inferido por intervalos de probabilidade a partir das fdac inferidas pela krigeagem por indicação.

Em cada ponto da grade, as probabilidades de pertinência a cada um dos intervalos de valores, definidos para as classes, foram calculadas pela formulação apresentada na equação 3.30. Para classificação de cada ponto da grade, utilizou-se o critério de máxima probabilidade, ou seja, atribuiu-se, ao

ponto da grade, a classe de maior probabilidade. Para isto, utilizou-se os modelos de distribuição probabilística inferidos pelo procedimento de krigeagem por indicação condicionado às amostras de altimetria. Estes modelos foram usados, também, para estimar-se o mapa de incertezas da classificação mostrado na Figura 6.11. Os valores de incerteza foram calculados a partir da probabilidade do valor de cada ponto da grade pertencer à classe estimada para o ponto. Assim, calculou-se a incerteza como:

$$Inc(\mathbf{u}) = 1 - Prob\{z(\mathbf{u}) \in s_k(\mathbf{u}), k=1,2 \text{ ou } 3\},\$$

onde  $s_k(\mathbf{u})$  é a classe atribuída a localização  $\mathbf{u}$ .

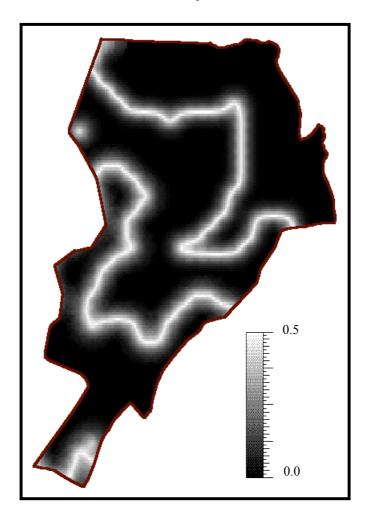


Fig. 6.11 - Mapa de incertezas por intervalos de probabilidades referente à classificação apresentada na Figura 6.10

### 6.3.3 Inferências com estimativa de incertezas por simulação sequencial condicionada

A Figura 6.12 mostra o resultado do uso do procedimento de simulação sequencial condicionada para espacializar, segundo uma grade regular, a propriedade altimetria utilizando-se o conjunto amostral de altimetria.

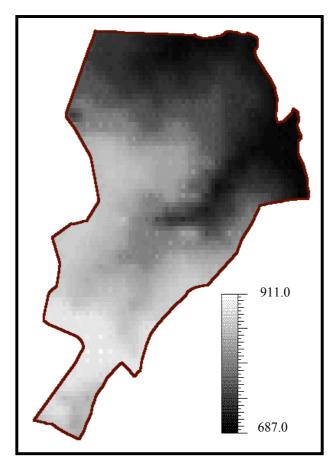


Fig. 6.12 - Mapa de valores médios inferidos a partir de um conjunto de 400 realizações geradas por simulação sequencial condicionada as amostras de altimetria apresentadas na Figura 6.2.

Para realizar a simulação sequencial condicionada, com o conjunto de amostras de altimetria, utilizou-se o conjunto de valores de corte e os modelos teóricos de variografia da Tabela 6.2. O procedimento de simulação foi executado com o programa *sisim.exe* da GSLIB. Geraram-se 400 grades simuladas, cada uma com as seguintes características espaciais: um ponto de

referência, inferior esquerdo, com abscissa igual a 204035.0 e ordenada igual a 7565050.0; resoluções horizontais e verticais da grade iguais a 70 m e 100 m, respectivamente e; tamanho da grade igual a 100 linhas por 100 colunas. Definiu-se, também, um raio de busca máximo igual a 2000 m e o número de pontos condicionantes entre 1 e 12. Os valores do atributo, mostrados na Figura 6.12, são valores médios obtidos das 400 realizações em cada ponto da grade, como descrito na Seção 4.3.1. O resultado da simulação sequencial possibilitou, também, a estimação de um mapa de incertezas, mostrado na Figura 6.13. A elaboração deste mapa é semelhante àquele da Figura 6.7, com os valores de desvio padrão sendo obtidos a partir do conjunto de realizações geradas no procedimento de simulação condicionada às amostras de altimetria.

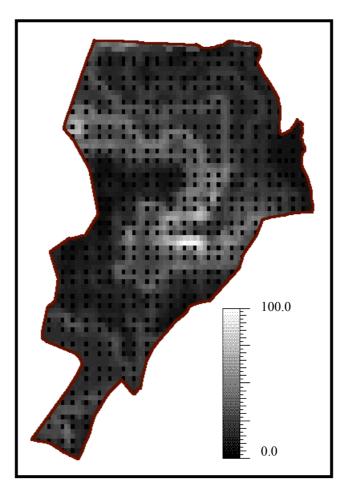


Fig. 6.13 - Mapa de incertezas obtido a partir das variâncias inferidas dos modelos de distribuição probabilística construídos pela simulação sequencial condicionada aos dados de altimetria.

#### 6.3.4 Análise e comparação dos resultados obtidos

Comparando-se o mapa da Figura 6.2 e com o mapa da Figura 6.3 observa-se que o estimador de krigeagem ordinária gera uma superfície mais suavizada mostrando a característica de minimização da variância do erro de estimação, própria desta classe de estimadores. Os resultados apresentados nestas duas figuras são parecidos porque a quantidade de amostras de altimetria é bastante grande.

Qualitativamente, os mapas das Figuras 6.3, 6.6 e 6.12 apresentam resultados semelhantes. Todos são mapas de médias, dos valores do atributo em estudo, inferidas por procedimentos baseados em krigeagem. As diferenças são pequenas e ocorrem devido, basicamente, à:

- aproximações da função de distribuição inferida pelos procedimentos baseados em krigeagem por indicação;
- tamanhos de grade diferentes, a grade gerada pela simulação tem 100 linhas por 100 colunas, um limite imposto pelo programa sisim.exe da GSLIB, enquanto que as grades geradas pelas krigeagens possuem 200 linhas por 200 colunas e;
- estratégias de busca das amostras vizinhas, o programa sisim.exe foi executado com a opção de busca em espiral, realocando-se os dados amostrais para os nós mais próximos da grade regular, para otimizar o tempo de geração das realizações.

Da Figura 6.5 observa-se que o campo de incerteza obtido apresenta um padrão de textura definido pela distribuição do conjunto amostral. A variância da krigeagem ordinária aumenta a medida que o ponto estimado está mais distante das amostras. Por outro lado, a variância diminui para pontos estimados que estão mais próximos das amostras. No limite, quando o ponto inferido está sobre uma amostra, a variância se iguala a zero, pois o estimador

de krigeagem é exato, ou seja, assume o valor da amostra nas posições observadas. Isto ocorre porque, como já observado e analisado no Capítulo 3, a variância de estimação da krigeagem linear depende da distribuição geométrica do conjunto amostral, sem levar em consideração os valores do atributo.

Em contrapartida, os campos de incerteza, mostrados nas Figuras 6.7, 6.9, 6.11 e 6.13, apresentam variações proporcionais ao comportamento do atributo na região. Nas áreas em que existe uma variação maior do valor do atributo, os valores de variância são maiores. Nas áreas em que o atributo tem variação mais suave, ou não varia, observam-se valores de incertezas menores. Assim, os resultados dos procedimentos baseados em krigeagem por indicação mostram que os valores estimados de incerteza estão relacionados diretamente com os valores do atributo.

As diferenças observadas nos mapas das Figuras 6.7 e 6.13 são explicadas pelas seguintes características próprias do algoritmo de simulação usado, otimização de busca de vizinhos em espiral, agregação dos valores estimados ao conjunto amostral e limitação no tamanho da grade estimada. Também, o aumento na quantidade de realizações obtidas gera distribuições mais acuradas, ao custo de aumento no tempo de processamento.

Os campos de valores de altimetria apresentados nas Figuras 6.6 e 6.8 apresentam diferenças porque, como mostrado na Figura 6.4, a distribuição do atributo, na região de estudo, não é simétrica. A Tabela 6.1 reporta um valor negativo para o coeficiente de assimetria do conjunto amostral. Também, os campos apresentados nas Figuras 6.7 e 6.9 refletem a incerteza associada com a variabilidade dos valores estimados apresentados nas Figuras 6.6 e 6.8. Portanto, o estimador de mediana em conjunto com um estimador de incerteza por intervalos interquantis, são os mais apropriados para se representar o atributo altimetria a partir do conjunto amostral disponibilizado.

Os resultados apresentados nas Figuras 6.10 e 6.11 mostram que os procedimentos baseados em krigeagem por indicação podem ser usados para obtenção de campos temáticos, com informação de incertezas, a partir de um conjunto amostral numérico. Este resultado é relevante pois ainda é comum utilizar-se modelagens computacionais, em SIG, que envolvem operações lógicas com dados representados por campos temáticos. A desvantagem de representação dos atributos por regiões homogêneas pode ser compensada, neste caso, pela inclusão de análises dos mapas de incertezas nos processos de tomadas de decisão sobre os produtos dos SIG.

### 6.4 ESTUDO DA TEXTURA DO SOLO SOB REPRESENTAÇÃO TEMÁTICA: CASO CANCHIM

A área de estudo é a mesma descrita na Seção 6.2. Nesta seção o atributo a ser modelado para a região é de natureza temática. No levantamento de solos da área, finalizado em 1997, na escala 1:10000, obteve-se um conjunto de 86 amostras de textura de solo (Calderano et al. 1996) classificadas como: solo arenoso, solo médio argiloso, solo argiloso e solo muito argiloso. A Figura 6.14 mostra a localização e a classificação das amostras de textura de solo adquiridas dentro dos limites da fazenda Canchim. Uma listagem numérica dessas localizações, em coordenadas UTM e com respectivos valores de classes, se encontra na Seção 3 do Apêndice A.

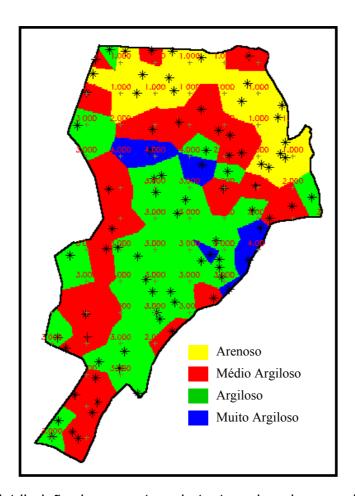


Fig. 6.14 - Distribuição das amostras de textura de solo na região de Canchim.

A classificação, apresentada na Figura 6.14, utilizou o critério de vizinhança mais próxima, ou seja, a cada ponto da grade atribuiu-se a classe da amostra mais próxima do ponto. Assim, essa figura apresenta, além da distribuição espacial das amostras, também, os valores de cada amostra.

A partir desse conjunto amostral, calculou-se um valor de probabilidade a priori para cada classe. Esses valores, que estão reportados na Tabela 6.3, foram calculados pelo paradigma frequencista, ou seja, dividindo-se o número de amostras de cada classe pelo número total de amostras.

TABELA 6.3 - CLASSES DE TEXTURA COM PROBABILIDADES A PRIORI.

Classe de Textura	Frequência Relativa
1 - Arenosa	0.20
2 - Média Argilosa	0.34
3 - Argilosa	0.38
4 - Muito Argilosa	0.07

## 6.4.1 Inferências com estimativa de incertezas por krigeagem por indicação

A Figura 6.15 mostra o resultado do uso do procedimento de krigeagem por indicação para espacializar, segundo uma grade regular, a propriedade textura do solo, a partir do conjunto amostral apresentado na Figura 6.14.

Para realizar a krigeagem por indicação, dividiu-se o conjunto amostral em 4 subconjuntos, considerando-se que os valores de cortes são determinados pelas classes de textura. Assim, para cada valor de corte, ou para cada classe de textura, realizou-se uma transformação por indicação, como definida na

equação 3.25, sobre o conjunto amostral original. Criaram-se, assim, 4 conjuntos amostrais por indicação. Para cada um desses conjuntos e utilizando-se o módulo de geoestatística do SPRING, definiu-se um modelo experimental de variografia e ajustou-se, a este, o modelo teórico de variografia que melhor representou a variabilidade espacial da variável por indicação.

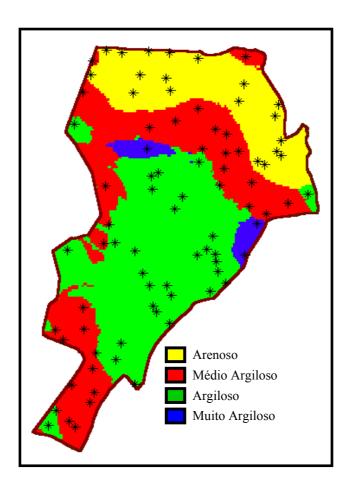


Fig. 6.15 - Mapas de valores de textura do solo inferidos pelo procedimento de krigeagem por indicação com 12 pontos condicionantes.

A Tabela 6.4 mostra os parâmetros dos variogramas teóricos obtidos para cada valor de corte. Considerou-se o modelo teórico de variografia do tipo esférico e anisotrópico para todos os subconjuntos por indicação.

TABELA 6.4 - PARÂMETROS DOS VARIOGRAMAS TEÓRICOS DEFINIDOS PARA OS DADOS DE TEXTURA DO SOLO TRANSFORMADOS POR INDICAÇÃO

CLASSE	ÂNGULO	EFEITO PEPITA	CONTRIBUIÇÃO	ALCANCE MÁXIMO	ALCANCE MÍNIMO
Arenoso	135°	0.07	0.126	1795	1380
Médio Argiloso	135°	0.08	0.06	1753	3
	135°		0.09	1753	919
Argiloso	0°	0.092	0.098	3899	3
	0°		0.07	3899	1835
Muito Argiloso	0°	0.015	0.02	1072	3
	90°		0.05	2517	1072

Para a espacialização da textura do solo, por krigeagem por indicação, utilizouse o programa *ik3d.exe* da GSLIB. Definiu-se a informação planimétrica da grade regular: por um ponto de referência, inferior esquerdo, com abscissa igual a 204017.5 e ordenada igual a 7565025.0; pelas suas resoluções horizontais e verticais iguais a 35 m e 50 m, respectivamente e; pelo tamanho da grade igual a 200 linhas por 200 colunas. Estabeleceu-se, também, um raio de busca máximo igual a 2000 m e o número de pontos condicionantes entre 1 e 12. Os valores do atributo, mostrados na Figura 6.15, representam valores de maior probabilidade, inferidos pela moda, como definido na equação 3.27.

O procedimento de krigeagem por indicação possibilitou, também, a determinação dos valores de incerteza por moda, como definido na equação 3.32, associados aos valores estimados. A Figura 6.16 mostra o campo de incerteza por moda relativo ao mapa de valores estimados da Figura 6.15. A Figura 6.17 apresenta esse mapa de incerteza classificado segundo 3 faixas de incerteza.

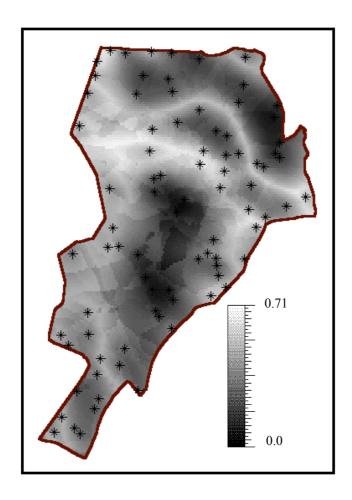


Fig. 6.16 - Mapa de incerteza por moda estimado a partir do procedimento de krigeagem por indicação usado para inferir o mapa da Figura 6.15

Com a finalidade de facilitar comparações, a Figura 6.18 apresenta 2 mapas de valores de textura de solo, com os mapas de incertezas associados. Os mapas superiores, desta figura, foram inferidos utilizando-se, no máximo, 4 (mapa da esquerda) e 16 (mapa da direita) pontos condicionantes. Os mapas inferiores representam as incertezas, por moda, associadas aos respectivos mapas superiores.

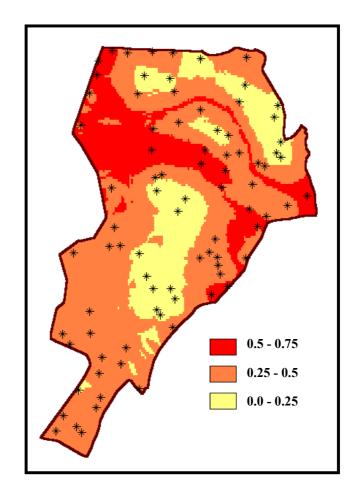


Fig. 6.17 - Classificação do mapa de incerteza apresentado na Figura 6.16

Utilizando-se as probabilidades de cada classe, para cada ponto da imagem, determinou-se, também, os valores de incerteza local por entropia de Shannon, como definida em 3.34, associados aos valores estimados. A Figura 6.19 mostra o campo de incerteza por entropia de Shannon, relativo ao mapa de valores estimados apresentado na Figura 6.15.

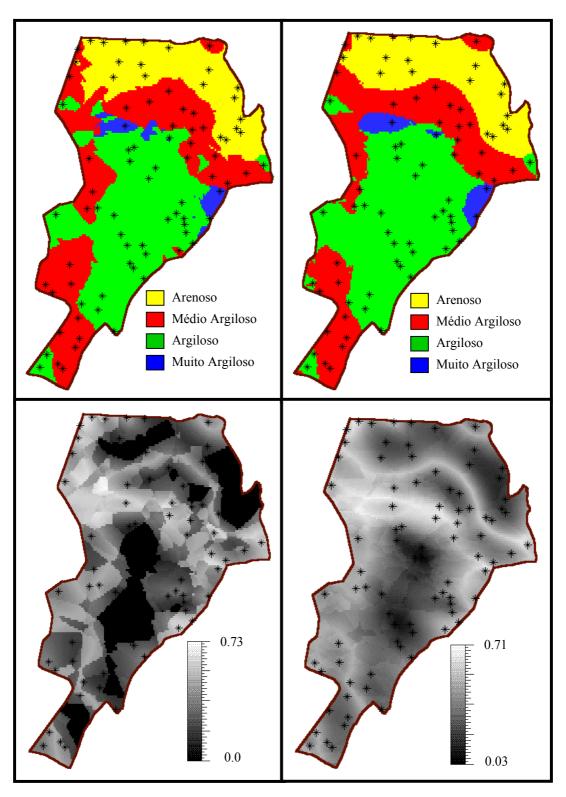


Fig. 6.18 - Mapas de valores estimados de textura do solo e incertezas associadas com, no máximo, 4 e 16 pontos condicionantes

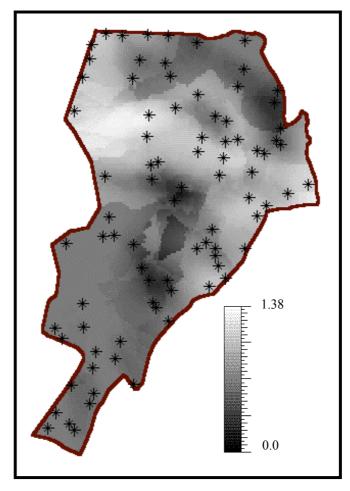


Fig. 6.19 - Mapa de incerteza por entropia de Shannon estimado a partir do procedimento de krigeagem por indicação usado para inferir o mapa da Figura 6.15

O resultado de um procedimento de espacialização, condicionada a diferentes limiares de probabilidade máxima de incerteza, é apresentado na Figura 6.20. Utilizaram-se limiares de incerteza iguais a .32, .4 e .5 para aceitação dos valores inferidos. Valores inferidos com incerteza maior que estes limiares foram rejeitados. O procedimento de krigeagem por indicação, utilizado para inferir os valores dos mapas da Figura 6.20, é o mesmo usado para inferir os valores da Figura 6.15 que não foram condicionados a uma incerteza máxima.

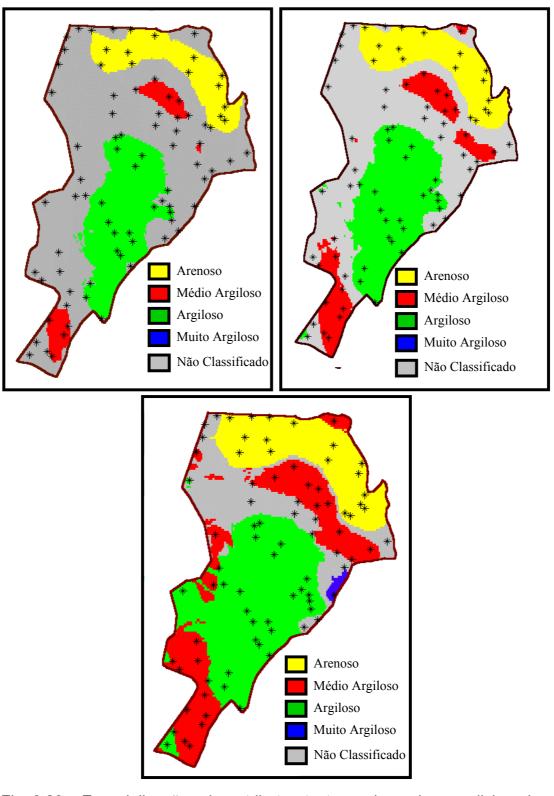
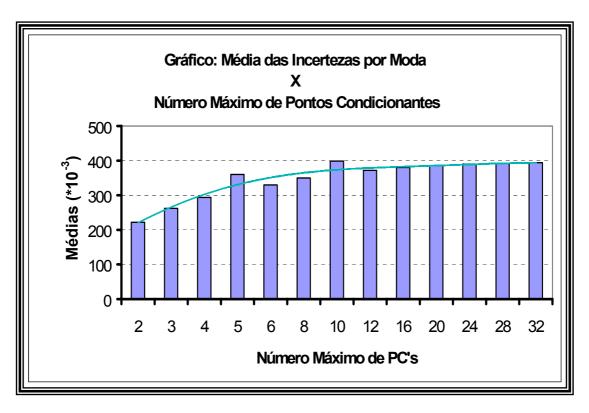


Fig. 6.20 - Espacialização do atributo textura do solo condicionada a diferentes limiares (0.38, 0.4 e 0.5) máximos de incerteza

A Tabela 6.5 apresenta os valores mínimo, máximo, médio e a variância dos mapas de incerteza, por moda e por entropia, para diferentes valores máximos de pontos condicionantes. As Figuras 6.21 e 6.22 mostram os gráficos de médias e variâncias reportados na Tabela 6.5.

TABELA 6.5 - ESTATÍSTICAS DOS VALORES DE INCERTEZAS, POR MODA E POR ENTROPIA, OBTIDOS POR KRIGEAGEM POR INDICAÇÃO PARA NÚMEROS MÁXIMOS DE PC DISTINTOS

Máximo de PC's	Incerteza por Moda				Incerteza por Entropia				
	Mínima	Máxima	Média	Variância	Mínima	Máxima	Média	Variância	
2	0.0	0.500	0.222	0.052	0.0	0.693	0.336	0.118	
3	0.0	0.667	0.262	0.041	0.0	1.099	0.487	0.132	
4	0.0	0.729	0.294	0.038	0.0	1.380	0.556	0.120	
5	0.0	0.689	0.360	0.030	0.0	1.356	0.683	0.094	
6	0.0	0.708	0.330	0.028	0.0	1.366	0.662	0.085	
8	0.0	0.703	0.350	0.024	0.0	1.366	0.716	0.067	
10	0.019	0.686	0.399	0.022	0.092	1.353	0.777	0.063	
12	0.030	0.697	0.372	0.020	0.133	1.340	0.770	0.056	
16	0.019	0.690	0.380	0.018	0.097	1.349	0.798	0.050	
20	0.062	0.697	0.386	0.018	0.233	1.335	0.815	0.046	
24	0.068	0.690	0.390	0.017	0.256	1.320	0.829	0.045	
28	0.049	0.680	0.392	0.017	0.198	1.318	0.838	0.044	
32	0.057	0.683	0.394	0.017	0.219	1.309	0.846	0.043	



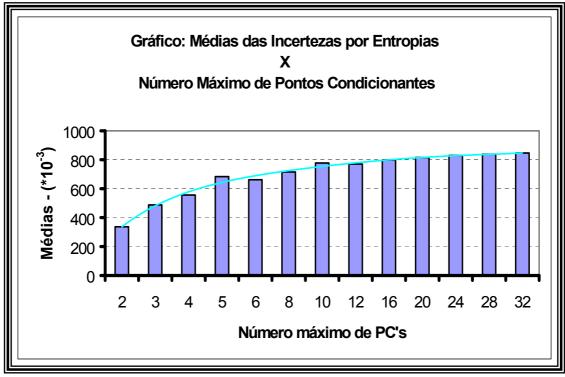
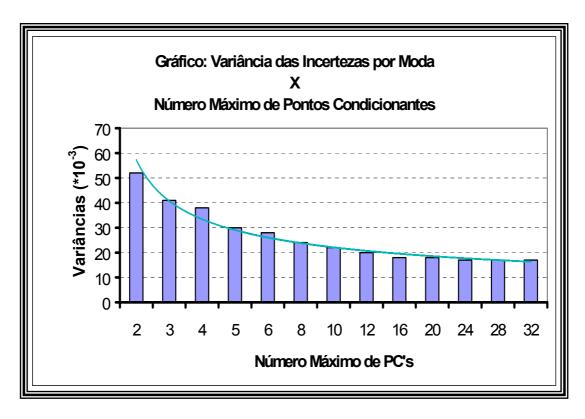


Fig. 6.21 - Gráficos de Médias de Incerteza por Número Máximo de PC's



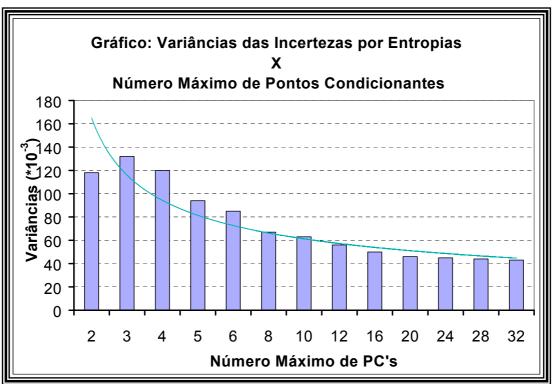


Fig. 6.22 - Gráficos de Variâncias de Incerteza por Número Máximo de PC's

# 6.4.2 Inferências com estimativa de incertezas por simulação sequencial condicionada

Para realizar a simulação sequencial condicionada ao conjunto de amostras de textura do solo, utilizou-se o mesmo conjunto de classes de corte e os modelos teóricos de variografia definidos na Tabela 6.4. O procedimento de simulação foi executado com o programa *sisim.exe* da GSLIB. Geraram-se 400 grades simuladas, cada uma com as seguintes características espaciais: um ponto de referência, inferior esquerdo, com x = 204035.0 e y = 7565050.0; resoluções horizontais e verticais da grade iguais a 70 m e 100 m, respectivamente e; tamanho da grade igual a 100 linhas por 100 colunas. Definiu-se raio de busca máximo igual a 2000 m e o número de pontos condicionantes entre 1 e 16. A Figura 6.23 mostra 4 exemplos de realizações, extraídas do total de 400, geradas pelo procedimento de simulação para o atributo temático textura do solo

A partir do conjunto total de realizações, inferiu-se o mapa da Figura 6.24, que representa a distribuição das classes de textura de solo inferidas pelos valores de moda, em cada ponto da grade, como descrito na Seção 4.3.2.

A Figura 6.25 representa a medida de incerteza associada aos valores inferidos por moda, e apresentados na Figura 6.24, a partir dos campos simulados.

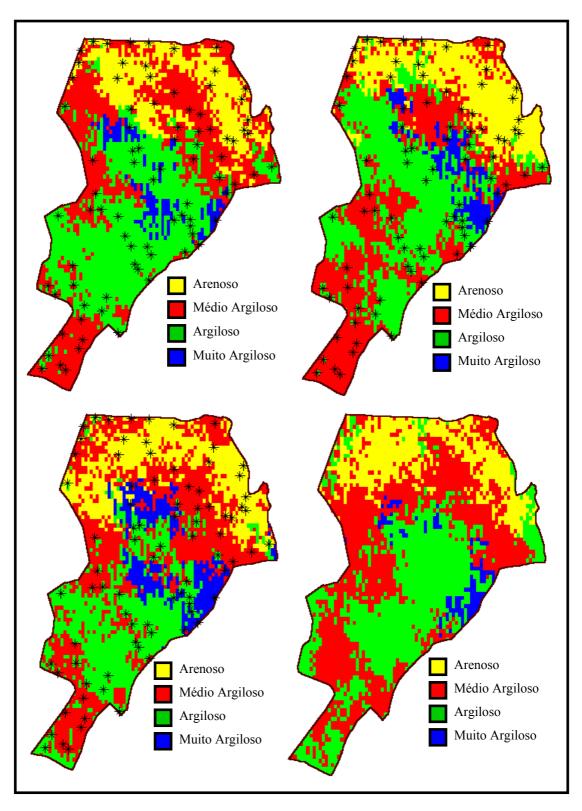


Fig. 6.23 - Exemplos de realizações do atributo textura do solo geradas pela simulação sequencial

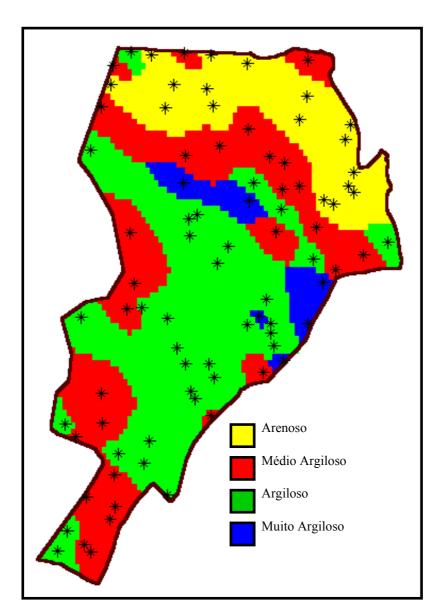


Fig. 6.24 - Mapa de valores de textura do solo inferidos pelo critério de frequência máxima a partir do conjunto de realizações gerados pela simulação sequencial.

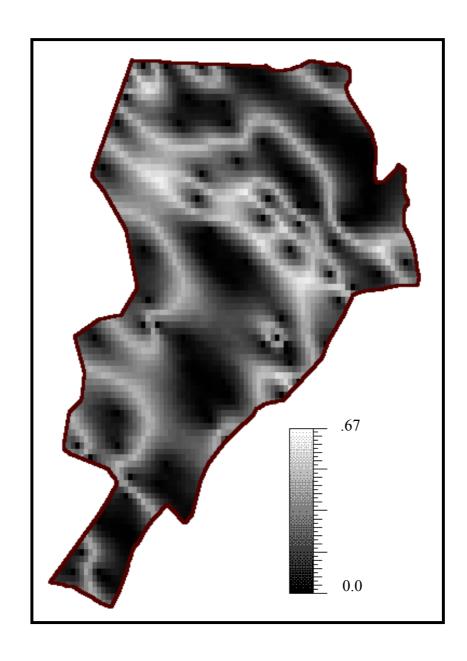


Fig. 6.25 - Incertezas associadas aos valores de textura de solo mostrados na Figura 6.24

## 6.4.3 Análise e comparação dos resultados obtidos

As Figuras 6.15, 6.18 e 6.20 apresentam resultados dos procedimentos de krigeagem por indicação para espacialização de atributos temáticos. Isto não é um resultado trivial, pois os interpoladores lineares, incluindo-se a krigeagem linear, só se aplicam a atributos numéricos.

A incerteza por moda, associada ao mapa da Figura 6.15, e apresentada na Figura 6.16, mostra um campo com variação proporcional ao comportamento do atributo na região. Nas regiões de transição entre as classes, os valores de incerteza por moda aumentam. Afastando-se dessas regiões, as incertezas diminuem. Este resultado confirma o que já havia sido concluído para atributos numéricos, ou seja, que os procedimentos de estimação de incertezas, baseados em krigeagem por indicação, fornecem medidas relacionadas diretamente com os valores do atributo. Isto é um resultado que coincide com o comportamento real dos atributos na natureza. Por exemplo, a transição de uma classe de vegetação para outra se faz de forma não abrupta indicando existir uma maior incerteza, quanto ao valor da classe correta, nas transições entre essas classes.

O mapa de incertezas, por moda ou por entropia, pode ser classificado, como ilustrado na Figura 6.17. Esta classificação é útil quando se quer mapear áreas por intervalos de incerteza, independente do estimador utilizado, ou seja, da classe atribuída a cada localização da área de interesse. Neste caso utilizou-se um conjunto de 3 classes que agregam os valores de incerteza entre 0 e 0.25, > 0.25 e <= 0.5 e >0.5.

Dos resultados mostrados nas Figuras 6.15, 6.16 e 6.18, observa-se que o resultado da inferência dos valores e das incertezas de um atributo depende do número de pontos condicionantes escolhidos para o procedimento de krigeagem. A medida que se aumenta o número de pontos condicionantes o mapa resultante da espacialização tende a ter menor variabilidade, ou seja, a superfície gerada é mais suave. Assim, um número grande de pontos

condicionantes deve ser usado para propriedades com comportamento suave. Por outro lado, quanto mais errático é o comportamento do atributo, um número menor de pontos condicionantes deve ser utilizado.

A Figura 6.19 apresenta um mapa de incerteza por entropia de Shannon, uma alternativa ao mapa de incerteza por moda. Comparando-se este mapa com o da Figura 6.16, obtidos com um mesmo número de pontos condicionantes, pode-se analisar as diferenças existentes entre o mapa de incertezas por moda e o mapa de incertezas por entropia. As diferenças são mais aparentes nas regiões onde várias classes se confundem. Isto é um resultado já esperado, uma vez que, nestas regiões a distribuição de probabilidade das variáveis aleatórias está mais próxima de uma distribuição uniforme. Nestas áreas, a incerteza por entropia tem seus valores maximizados. A incerteza por moda mostra um aumento de incerteza a partir do parte central de uma classe em direção as bordas de transição. Os valores máximos de incerteza por moda aparecem nas bordas entre as classes e não têm influência do número de classes próximos as bordas. Dependendo da aplicação, o especialista é responsável por decidir sobre qual mapa de incerteza será usado. Quando a confusão entre classes é importante deve-se optar pela incerteza por entropia. Caso o interesse seja somente nas transições entre as classes a entropia por moda deve ser priorizada.

O resultado apresentado na Figura 6.20 mostra que os dados de incerteza, inferidos conjuntamente com os valores do atributo, podem ser utilizados em procedimentos de inferência condicionada à um limiar máximo de incerteza. Numa aplicação de uso de solos, por exemplo, pode ser requerido um nível de qualidade mínimo para que uma área seja considerada pertencente a uma determinada classe de ocupação.

Os dados reportados na Tabela 6.5 e apresentados graficamente nas Figuras 6.21 e 6.22, permitem uma análise quantitativa da dependência dos dados de incertezas, por moda e por entropia, em relação a quantidade de pontos

condicionantes, utilizado no procedimento de krigeagem por indicação. Os resultados para 2 e 3 pontos condicionantes podem ser tendenciosos pois o número de classes, consideradas nesta aplicação, é igual a 4. Analisando-se os resultados a partir de 4 pontos condicionantes, conclui-se que:

- os valores mínimos de incerteza, por moda e por entropia, têm variações pequenas, com uma tendência de crescimento de 0.0 a um patamar, com o aumento do número de PC's;
- os valores máximos de incerteza, por moda e por entropia, têm variações pequenas, com uma tendência de diminuição para um determinado patamar, a medida que o número de PC's aumenta;
- os valores médios de incerteza, por moda e por entropia, aumentam, até um certo patamar, a medida que o número de PC's aumenta e;
- os valores de variância de incerteza, por moda ou por entropia, diminuem, tendendo a um certo patamar, com o aumento do número de pontos de controle.

As Figuras 6.23, 6.24 e 6.25 ilustram o uso da simulação sequencial para inferência dos valores e das incertezas associadas de atributos de natureza temática. As razões das diferenças, observadas entre estes mapas e os mapas inferidos por krigeagem por indicação, são as mesmas reportadas na Seção 6.2.4, ou seja, essas diferenças são explicadas, basicamente, pelo procedimento de otimização de busca em espiral, pela agregação dos valores estimados ao conjunto amostral e pelo número limitado de realizações geradas pela simulação. O aumento do número de realizações gera distribuições mais acuradas ao custo de aumento no tempo de processamento.

#### 6.5 USO DE DADOS INDIRETOS

Esta seção apresenta resultados do uso de dados indiretos, neste caso fornecidos por um mapa de solos, para melhorar a acurácia dos mapas produzidos pelos procedimentos baseados em krigeagem por indicação.

A região de Canchim possui diferentes classes de solos, tais como, latossolos, terra roxa estruturada, podzólicos, areias quartzosas, solos orgânicos, etc.. Estas classes estão representadas em 29 unidades de mapeamento composta de unidades simples e unidades taxonômicas e apresentadas no mapa de solo da Figura 6.26. A Tabela 6.6 reporta a extensão, em hectares e porcentagem em relação ao total, de cada classe de solo da região.

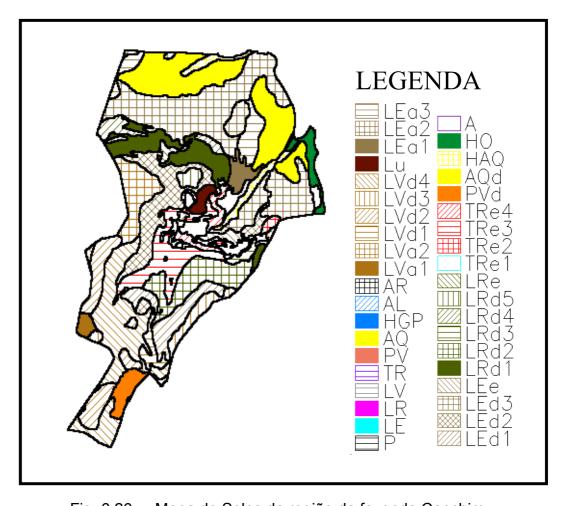


Fig. 6.26 - Mapa de Solos da região da fazenda Canchim

TABELA 6.6 - EXTENSÃO DOS TIPOS DE SOLOS DA REGIÃO DE CANCHIM

Solos	Extensão (ha)	Porcentagem (%)
Latossolo Vermelho - Amarelo	572.99	20.1
Latossolo Vermelho - Escuro	943.94	32.9
Latossolo V Amarelo, Var. Una	198.48	6.0
Latossolo Roxo	380.52	13.4
Terra Roxa Estruturada	280.40	9.9
Podzólico Vermelho - Amarelo	42.47	1.5
Areias Quartzosas	328.46	11.6
Areias Quartzosas Hidromórficas	56.81	2.0
Solos Orgânicos	36.68	1.3
Solos Aluviais	8.06	0.3
Afloramentos de Rocha	1.58	0.1
Total	2841.38	100.00

## 6.5.1 Uso do mapa de solos para estimar o atributo textura do solo

O mapa de solos, apresentado na Figura 6.26, serviu de fonte de dados indiretos para se estimar o atributo textura do solo segundo a metodologia apresentada no Capítulo 5. Assim, para cada classe de solo dessa região, determinou-se os valores de probabilidade de ocorrência das classes de textura. A Tabela 6.7 apresenta essas probabilidades, que foram definidas por especialistas, pedólogos, envolvidos na construção desse mapa de solos.

TABELA 6.7 - PROBABILIDADE DE OCORRÊNCIA DAS CLASSES DE TEXTURA DE SOLO PARA AS CLASSES DE SOLO DE CANCHIM

SOLO	ARENOSO	MÉDIO ARGILOSO	ARGILOSO	MUITO ARGILOSO
LVA1	0	0	1	0
LVA2	0	1	0	0
LVD1	0	0	1	0
LVD2	0	0	1	0
LVD3	0	1	0	0
LVD4	0	1	0	0
LU	0	0	1	0
LEA1	0	0.4	0.6	0
LEA2	0	1	0	0
LEA3	0	1	0	0
LED1	0	0	1	0
LED2	0	0	1	0
LED3	0	1	0	0
LEe	0	0	1	0
LRD1	0	0	0	1
LRD2	0	0	0.8	0.2
LRD3	0	0	0.7	0.3
LRD4	0	0	1	0
LRD5	0	0	1	0
LRe	0	0	0.4	0.6
TRe1	0	0	0.4	0.6
TRe2	0	0	0	1
TRe3	0	0	1	0
TRe4	0	0	0.7	0.3
PVd	0	1	0	0
AQd	1	0	0	0
Haq	0.8	0	0.2	0
Но	0	0	1	0
A	0	0	1	0

A informação indireta foi, inicialmente, espacializada, dentro da região de interesse, segundo uma grade regular com a mesma distribuição espacial da grade de textura a ser inferida. Para cada ponto da grade obteve-se um conjunto de valores de probabilidade a priori referentes à classe de solo definida para o ponto.

Para realizar a simulação sequencial condicionada ao conjunto de amostras de textura do solo e aos dados indiretos, utilizou-se o mesmo conjunto de classes de corte e os modelos teóricos de variografia definidos na Tabela 6.4. O procedimento de simulação foi executado com o programa *sisim\_lm.exe* da GSLIB. Este programa permitiu realizar-se uma cokrigeagem por indicação, utilizando-se dados diretos de textura do solo e dados indiretos do mapa de solos co-localizados, como descrito na Seção 5.7. Geraram-se 400 grades simuladas, cada uma com as seguintes características espaciais: um ponto de referência, inferior esquerdo, com x = 204035.0 e y = 7565050.0; resoluções horizontais e verticais da grade iguais a 70 m e 100 m, respectivamente e; tamanho da grade igual a 100 linhas por 100 colunas. Estabeleceu-se raio de busca máximo igual a 2000 m e o número de pontos condicionantes entre 1 e 16.

A Figura 6.27 mostra o mapa de valores do atributo textura do solo resultante do uso dos dados indiretos segundo o procedimento acima descrito.

A Figura 6.28 apresenta o mapa de incertezas associadas aos valores inferidos mostrados na Figura 6.27.

Os mapas mostrados na Figura 6.29 são aqueles apresentados nas Figuras 6.27 e 6.28 com os contornos do mapa de solo sobreposto aos mesmos. Este procedimento foi realizado no sentido de facilitar a análise dos resultados obtidos.

A Tabela 6.8 apresenta resultados estatísticos dos mapas de incertezas obtidos com, e sem, o uso de dados indiretos. Para estes mapas, observaram-

se os valores mínimos e máximos e calcularam-se as médias e variâncias das incertezas inferidas.

TABELA 6.8 - ESTATÍSTICAS DAS MEDIDAS DE IMPRECISÃO COM, E SEM, DADOS INDIRETOS

In cer te za	Simulação apenas com dados diretos				Simulação com dados diretos e indiretos			
	min	max	μ	$\sigma^2$	min	max	μ	$\sigma^2$
$\overline{\phi}$	0.0	0.665	0.240	0.027	0.0	0.655	0.157	0.018

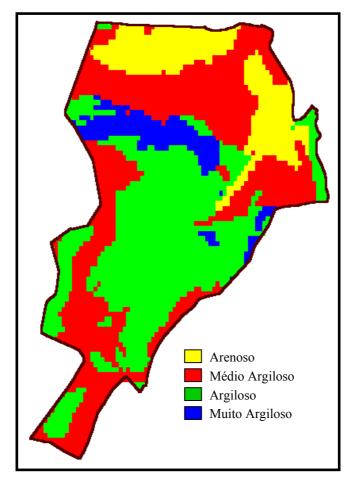


Fig. 6.27 - Mapa de valores de textura inferido por simulação sequencial condicionada a dados diretos e indiretos

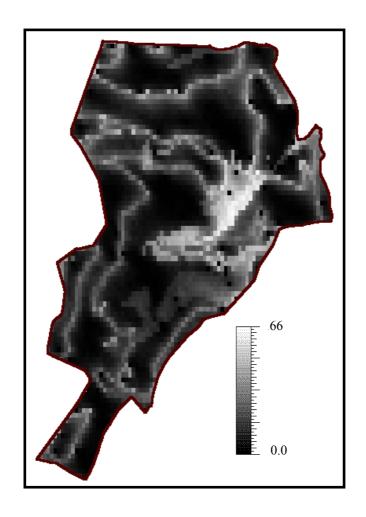


Fig. 6.28 - Mapa incertezas, por moda, associadas aos valores de textura inferidos por simulação sequencial condicionada a dados diretos e indiretos

## 6.5.2 Análise dos resultados e comparação com estimativas sem dados indireto

Comparando-se o mapa de textura de solo obtido com uso de dados indiretos, Figura 6.27, com o mapa da Figura 6.24, observa-se uma diferença importante entre estes. O mapa da Figura 6.27 tem sua conformação final de classes moldada pelo mapa de solos. A Figura 6.29 ilustra melhor esse resultado ao apresentar os limites das classes de solo sobrepostos ao o mapa de textura obtido. Observa-se, portanto, que este mapa consegue mesclar os resultados

puramente matemáticos, fornecidos pelos procedimentos geoestatísticos e os dados diretos, com o conhecimento, muitas vezes abstrato ou empírico, dos especialistas, sobre uma propriedade a ser estudada.

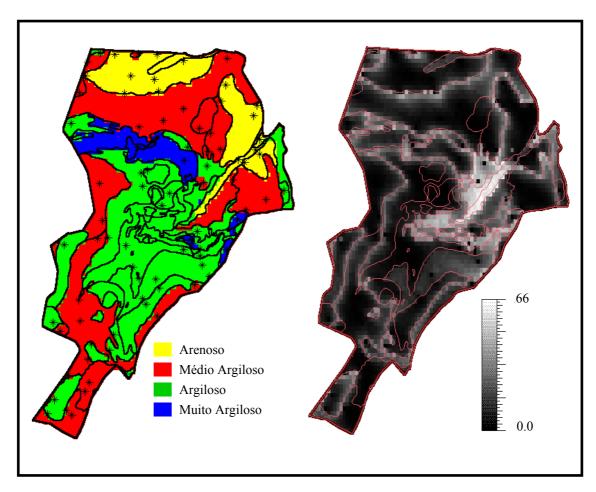


Fig. 6.29 - Mapas das Figuras 6.27 e 6.28 com os contornos do mapa de solo sobrepostos.

Como já observado para os outros mapas de incertezas por moda, apresentados anteriormente, o mapa da Figura 6.28 reflete, no geral, o resultado obtido com o mapa de textura obtido com os dados diretos e indiretos, ou seja, as áreas de incerteza são maiores nas transições entre as classes do mapa de textura. Entretanto, apareceram regiões de alta incerteza, principalmente nas classes de solo Haq e TRe1, para as quais as informações indiretas se contrapõem às inferências obtidas somente com dados diretos.

Estas regiões de alta incerteza são as candidatas mais importantes para serem amostradas com mais detalhes, caso isto seja possível.

Visualmente, o mapa de incerteza da textura, inferido com uso de informação indireta, mostra mais áreas escuras do que o mapa de incerteza da Figura 6.22, associado ao mapa de textura sem essa informação. Os resultados apresentados na Tabela 6.8 possibilitam uma comparação quantitativa global entre esses dois mapas de incerteza. Observa-se, dessa tabela, que a média das incertezas diminuiu com o uso de dados indiretos. Também, as variâncias ficaram menores, indicando uma dispersão menor em torno da média. A partir destes resultados, observou-se que, neste caso, o uso de dados indiretos, em conjunto com as amostras diretas, melhorou a acurácia da estimação. Esta conclusão não pode ser generalizada. Pode acontecer da incerteza média total aumentar, quando existir muita informação indireta se contrapondo às inferências obtidas com os dados diretos.

## 6.6 MEDIDAS GLOBAIS DE INCERTEZA

Mediu-se, ainda, a entropia relativa média bivariada,  $\overline{H}_{R}(A,A)$ , para os resultados de probabilidades das classes, utilizando-se apenas os dados diretos e utilizando-se dados diretos e indiretos nas estimativas, segundo a metodologia apresentada na Seção 4.4.3. A idéia é avaliar quantitativamente o comportamento global da entropia, antes e depois do uso de dados indiretos na simulação. Os valores obtidos foram  $\overline{H_R}(A,A) = .818$  para a incerteza obtida sem uso de dado indireto e  $\overline{H_R}(A,A)$  = .795 para a incerteza com o uso do mesmo. Este resultado sugere uma melhora na acurácia do estimador após o uso de dados indiretos. No entanto, caso houvessem mais regiões com informações indiretas se contrapondo com às inferências diretas.  $\overline{H_{R}}(A,A)$  poderia até aumentar com o uso de dados indiretos.

## 6.7 CONCLUSÕES

É comum encontrar-se, na literatura, trabalhos que exploram algoritmos de inferência a partir de amostras pontuais de atributos ambientais numéricos. Mais raros são os artigos que exploram a espacialização de atributos temáticos devido, principalmente, à complexidade dos procedimentos de inferência para esses tipos de dados. Entretanto, propriedades temáticas são importantes quando se deseja modelar fenômenos ambientais relacionados, por exemplo, com dados de solos. Também, é grande a quantidade de atributos qualitativos que não podem ser descartados só por que seu tratamento é complexo.

Este capítulo apresentou estudos de caso para o uso de procedimentos geoestatísticos de inferência não linear, krigeagem por indicação e simulação sequencial condicionada, envolvendo dados numéricos e temáticos. Os resultados obtidos mostram que esses procedimentos apresentam vantagens, em relação aos procedimentos lineares, principalmente na estimação de atributos temáticos e na inferência das incertezas associadas aos valores estimados. Além disso, possibilitam o uso de dados indiretos, em conjunto com os dados diretos, que podem melhorar a acurácia das estimativas. A desvantagem principal é a necessidade de se estimar modelos de variabilidade para diversos valores de corte, o que demanda uma carga de trabalho maior.

Foram apresentadas várias métricas de incerteza. O uso destas métricas não foi explorado intensivamente num domínio suficientemente amplo de aplicações. Assim, não é possível se fazer assertivas mais rigorosas sobre as vantagens e desvantagens de cada uma em relação as aplicações. Porém, sob um enfoque mais geral, pode-se dizer que, para atributos numéricos, é apropriado utilizar-se de métricas de incerteza baseadas em intervalos de probabilidade e que consideram a simetria do modelo de distribuição do atributo e, para atributos temáticos, sugere-se utilizar incertezas por entropia de Shannon sempre que uma estimativa da confusão entre classes for importante, caso contrário, a incerteza por moda deve ser a escolhida.

## **CAPÍTULO 7**

# INTEGRAÇÃO DE PROCEDIMENTOS GEOESTATÍSTICOS AOS SISTEMAS DE INFORMAÇÃO GEOGRÁFICA

## 7.1 INTRODUÇÃO

Nos capítulos anteriores apresentaram-se teorias e estudos de casos relacionados a procedimentos geoestatísticos para estimativa de atributos espaciais com inferência de incertezas. Mostrou-se, 9também, que esses procedimentos podem ser utilizados com frequência, em SIG, para espacialização de atributos numéricos e temáticos. Neste capítulo aborda-se o tema de integração dos procedimentos de geoestatística ao ambiente de um SIG. Busca-se, aqui, especificar os requisitos de um módulo de geoestatística, implementado como parte do ambiente operacional de um SIG, e discutir a extensão das linguagens de modelagem espacial para considerar incertezas nas modelagens computacionais executadas neste ambiente.

O presente capítulo tem a seguinte organização:

A Seção 7.2 discute a importância da integração das ferramentas geoestatísticas para manipulação e análise de dados espaciais. Aborda, também, os requisitos necessários para se disponibilizar, num SIG, os procedimentos e os produtos da geoestatística.

Na Seção 7.3 descrevem-se os componentes e os aspectos funcionais de um módulo de geoestatística para ser integrado a um SIG. Apresentam-se as características dos primeiros módulos sendo disponibilizados no mercado.

A Seção 7.4 apresenta aspectos relevantes relacionados às linguagens de modelagem espacial com tratamento de incertezas. Nesta seção descrevem-se as principais características da linguagem ADAM (Heuvelink 1998) para modelagens de atributos numéricos com propagação de incertezas. Apresentam-se, ainda, os requisitos básicos, na forma de uma especificação

preliminar, para inclusão de tratamento de incertezas na linguagem de manipulação algébrica LEGAL ( Câmara 1995 ).

A Seção 7.5 encerra o capítulo apresentando algumas considerações sobre a temática abordada.

#### 7.2 SIG E GEOESTATÍSTICA

Sistemas de informação geográfica e pacotes geoestatísticos têm, no mínimo, duas características básicas que os atraem:

- 1) Manipulam representações digitais de dados espaciais. As ferramentas da geoestatística são usadas para entendimento e modelagem da variabilidade espacial de atributos (Deutsch e Journel 1998). Os SIG, por sua vez, possibilitam que dados espaciais do mundo real sejam inseridos, consultados, analisados, transformados e apresentados para um conjunto particular de propósitos (Burrough 1987).
- 2) São multidisciplinares com interesses comuns. As disciplinas de interesse dos dois sistemas são muitas (por exemplo: ciências de solo, agricultura, epidemologia, geologia de exploração para engenharia de petróleo e mineração, hidrologia, ciência ambiental, etc.) e praticamente as mesmas, pois estão relacionadas ao estudo e compreensão dos fenômenos e processos da natureza.

Entretanto, até recentemente, as funcionalidades dos SIG e da geoestatística foram disponibilizadas à comunidade por sistemas separados e independentes. Neste contexto, as modelagens computacionais, desenvolvidas no SIG e que requerem procedimentos geoestatísticos, são executadas em duas fases, primeiro exporta-se o dado espacial para um sistema de análise geoestatística e, então, importa-se os resultados da análise para ser integrado com outros dados da base do SIG. Este procedimento é, em geral, trabalhoso, demorado e requer do usuário um trabalho extra de mapeamento, quando possível, dos

elementos conceituais e de representação utilizados por cada sistema. Esse tempo e esforço despendidos seriam melhor utilizados pelos usuários em atividades específicas requeridas pelo processo de modelagem dos dados e do fenômeno em estudo. Essas são as principais razões para a recente tendência de incorporação de facilidades geoestatísticas em ambientes de SIG. Nesse contexto, esta seção apresenta e discute a importância das funcionalidades da geoestatística para o SIG e vice-versa. Apresentam-se, também, os aspectos vantajosos de se ter uma integração forte entre essas duas ferramentas.

# 7.2.1 Importância do SIG para implementação das facilidades da geoestatística

Dados e procedimentos típicos de um SIG podem ser aproveitados para suportar dados e procedimentos da geoestatística.

Inicialmente, pode-se pensar na utilização direta da *capacidade de* armazenamento e de visualização do SIG para os dados amostrais e para os produtos das análises realizadas pelos procedimentos geoestatísticos.

Procedimentos e dados do SIG podem servir para *garantir consistência* dos dados geoestatísticos. Pode-se verificar, por exemplo, se todos os dados amostrais estão dentro de uma região de interesse, pode-se editar, inclusive eliminar, amostras erroneamente observadas (erro de atributo ou posicionamento), etc..

Dados previamente armazenados no SIG podem servir para *limitação de atuação ou de validade* dos procedimentos e produtos geoestatísticos. Extrapolações geradas por procedimentos de krigeagem ou simulação podem ser eliminadas por mapas de fronteiras. Procedimentos de krigeagem por blocos e de desagrupamento de amostras requerem a definição de polígonos ou áreas que devem estar previamente armazenados no SIG.

Dados previamente armazenados no SIG podem servir para explicar e, ou, melhorar resultados de análises geoestatísticas. Por exemplo, algumas anisotropias só podem ser explicadas, ou compreendidas, quando os dados amostrais da geoestatística são sobrepostos à outros dados espaciais. Outro exemplo é o uso de mapas de solo para obtenção de dados adicionais indiretos para uso em inferências geoestatísticas segundo o paradigma bayesiano. No Capítulo 6, obteve-se informação indireta de um mapa de solo, que poderia estar previamente armazenado no SIG, combinada com os dados diretos para melhorar a acurácia de inferência da textura de solo.

Finalmente, pode-se utilizar módulos ou linguagens de análise espaciais, embutidas no SIG, no sentido de combinar dados e produtos da geoestatística em modelagens ambientais complexas.

## 7.2.2 Importância dos procedimentos de geoestatística para um SIG

Os formalismos matemáticos e os resultados, apresentados nos capítulos anteriores, mostram que os procedimentos geoestatísticos são importantes para o ambiente operacional de um SIG pelas seguintes razões:

- Fornecem novas metodologias de espacialização para atributos de dados espaciais, representados por amostras pontuais, baseadas no estudo e modelagem da variabilidade espacial do atributo;
- Possibilitam o uso de uma abordagem unificada para espacialização de atributos de natureza numérica e temática amostrados pontualmente;
- Permitem a estimação de incertezas associadas aos valores espacializados. As incertezas informam sobre a qualidade dos dados espaciais e dos produtos de modelagens computacionais.
- Possibilitam a geração de realizações equiprováveis de campos numéricos e temáticos para uso em modelagens computacionais e

técnicas de propagação de incerteza. O exemplo típico é o uso da simulação de Monte Carlo em modelagens ambientais.

## 7.2.3 Integração da geoestatística a um SIG

Os procedimentos de geoestatística, com estimativas de incerteza, incluem, obrigatoriamente, um novo componente aos dados espaciais, ou seja, a informação de incerteza. Essa informação, além de qualificar o dado espacial, se propaga em análises simples, como por exemplo a geração de um mapa de declividade a partir de um modelo digital de elevação, e em modelagens computacionais que envolvam várias representações de dados espaciais e também diversas formas de operações relacionando esses dados.

Assim, o uso de geoestatística, integrada a um SIG, exige algumas modificações na estrutura tradicional de um SIG.

Inicialmente, o SIG deve ser provido de funcionalidade para armazenar e recuperar, além do dado espacializado, também a informação de incerteza associada e, em alguns casos, informação sobre a correlação espacial entre os dados da base. As incertezas podem ser armazenadas como novos mapas ou como metadados e devem incluir explícita, ou implicitamente, informação sobre a quais dados espaciais elas se referem. Em relação as estruturas de armazenamento e recuperação, deve-se, também, buscar soluções eficientes para a imensa quantidade de dados, muitas vezes temporários, gerados pela simulação estocástica.

Além das estruturas de armazenamento, deve-se implementar um conjunto mínimo de procedimentos da geoestatística como novas funcionalidades do SIG. A Seção 7.3, que segue, especifica os requisitos para integração de um módulo geoestatístico ao SIG.

Finalmente, os procedimentos de análise de um SIG, incluindo-se as linguagens de modelagem espacial, devem considerar as incertezas das

representações dos dados e dos parâmetros envolvidos nas análises. Essas incertezas devem ser propagadas para os produtos da modelagem. Isto não tem uma implementação trivial, uma vez que, no SIG armazenam-se dados de diferentes categorias, existem várias formas para integração dos dados e os modelos de propagação de incerteza não são únicos, apresentando restrições em relação às operações e aos tipos de dados usados na análise. O problema de propagação de incertezas é uma consequência natural do uso da geoestatística com estimativa de incerteza.

#### 7.3 MÓDULOS DE GEOESTATÍSTICA INTEGRADOS A UM SIG

Inicialmente pode-se pensar em agregar funções geoestatísticas a um SIG via inclusão de um módulo de geoestatística. Camargo et al. (1999) argumenta que os SIG, devido a sua crescente utilização, vêm incorporando novos princípios e procedimentos de análise espacial. Este esforço deve-se ao consenso, hoje existente, de que o futuro sucesso da tecnologia dos SIG depende de sua capacidade de gerar cenários confiáveis, através de um conjunto de procedimentos e ferramentas para a modelagem espacial, que possam representar adequadamente, em meio digital, as propriedades, os processos e suas imprecisões. Os procedimentos geoestatísticos têm uma contribuição importante a oferecer nessa direção. Por isso, as versões mais recentes do SPRING, do IDRISI, do PC-RASTER e do ARC-INFO, para citar alguns, contém módulos de geoestatística integrados ao sistema. Nas seções seguintes serão descritas as principais características dos módulos de geoestatística que foram, ou estarão sendo, integrados a esses SIG.

## 7.3.1 Requisitos básicos de um módulo de geoestatística para um SIG

É importante salientar que o módulo de geoestatística do SIG deve ter um conjunto mínimo de ferramentas básicas para caracterizá-lo como tal. Por exemplo, alguns SIG que oferecem procedimento de krigeagem ainda requerem o uso de ferramentas externas para a definição da variabilidade espacial do atributo. Idealmente o procedimento de krigeagem do módulo

deveria utilizar parâmetros de variografia obtidos no próprio ambiente de trabalho do SIG.

Como já observado, a geoestatística faz uso de amostras pontuais como dados de entrada. Procedimentos de importação/exportação, digitalização e edição das amostras fazem parte do conjunto de funcionalidades típicas de um SIG e, portanto, não serão consideradas como parte do módulo. Os procedimentos específicos do módulo de geoestatística são:

- 1) Procedimentos de análise exploratória dos dados de entrada. Estes procedimentos devem informar sobre estatísticas gerais, univariadas e bivariadas, referentes aos dados. São procedimentos típicos da análise exploratória, o cálculo e apresentação de estatísticas descritivas simples, tais como, média, mediana, moda, variância, desvio padrão, mínimo, máximo, quantis, coeficientes de curtose, coeficientes de correlação e outros, e o cálculo e apresentação de histogramas, de funções de distribuição de probabilidade, de gráficos de probabilidade normal e de diagramas de dispersão.
- 2) Procedimentos de transformação dos dados. São procedimentos para corrigir, melhorar ou otimizar as amostras objetivando uma determinada aplicação. Fazem parte deste grupo os procedimentos de filtragem de valores muito distantes de uma tendência ("outliers"), de desagrupamento ("decluster") das amostras, de transformação para normalização e retorno à distribuição original ("back transformation"), de transformação logarítmica e de transformação por indicação.
- 3) Procedimentos para modelagem da variabilidade espacial do atributo: Análises para se definir o comportamento da variabilidade espacial de um atributo em função da distância, linear e angular, para o mesmo atributo, autocovariância, ou para outro atributo correlacionado, covariância cruzada, são realizadas por estes procedimentos. Exemplos típicos são a análise de estacionariedade por janelas móveis, a análise espacial por

semivariograma, a análise por variografia de superfície para detecção de anisotropia, a análise de dimensão fractal e a variografia exploratória com geração de variograma experimental e a sua modelagem por um variograma teórico;

- 4) Procedimentos para inferência de valores de atributos e incertezas. Estes procedimentos possibilitam a inferência dos valores do atributo, em posições não amostradas, considerando o conjunto amostral e um modelo de correlação espacial que é definido pelo semivariograma teórico. Procedimentos de krigeagem linear simples, ordinária e universal, de krigeagem por blocos, de krigeagem não linear por indicação, de cokrigeagem linear e por indicação e de simulação estocástica condicional, são os procedimentos mais comuns neste grupo. Estes procedimentos devem considerar restrições de bordas, definidas por linhas ou polígonos, durante as inferências. Amostras além de limites pré-estabelecidos não podem fazer parte das inferências. Estes limites servem, também, para restringir o alcance dos interpoladores, ou seja, para limitar a região de validade das inferências.
- 5) Procedimentos de validação dos resultados: São procedimentos para avaliação da qualidade dos modelos de variografia e de inferência utilizados. Tipicamente são os procedimentos de validação cruzada e de análise de erros e resíduos.

## 7.3.2 O módulo de geoestatística do SPRING

O SPRING é um SIG no estado da arte com funções de processamento de imagens, análise espacial, modelagem numérica de terreno e consulta a bancos de dados espaciais. O SIG SPRING é um software de domínio público que está disponibilizado na Internet no endereço: http://www.dpi.inpe.br/spring/.

O SPRING na sua versão 3.3, para plataforma Windows 9x/NT, contém um módulo de geoestatística cuja estrutura básica é apresentada na Figura 7.1.

Esse módulo tem como objetivo a análise, em duas dimensões, de dados espacialmente distribuídos, no que diz respeito a inferência de superfícies a partir de amostras pontuais georeferenciadas (Camargo et al. 1999). A estrutura implementada utiliza a biblioteca pública de funções geoestatísticas GSLIB, e possibilita a realização dos seguintes procedimentos:

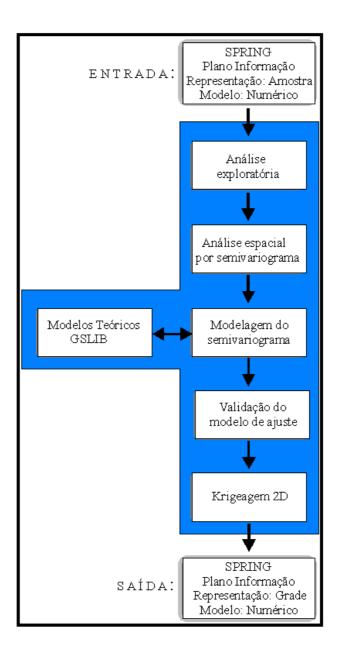


Fig. 7.1 - Estrutura do Módulo de Geoestatística no SPRING. Fonte: Camargo et al. (1999)

- Entrada de dados amostrais pontuais relativos a um atributo em estudo.
   No modelo de dados do SPRING, os dados amostrais dos atributos são armazenados em um plano de informação, PI, pertencente à uma categoria do modelo numérico.
- Análise exploratória sobre os valores de um, ou dois, atributos, de um PI com amostras pontuais, através de estatísticas descritivas univariadas, incluindo, média, variância, desvio padrão, coeficiente de variação, coeficiente de assimetria, curtose, mediana, quartil inferior e superior, de determinação e apresentação do gráfico de histograma e de probabilidade normal, e de apresentação de estatística bivariada, entre os valores de duas variáveis aleatórias, pelo diagrama de dispersão.
- Análise da continuidade espacial do atributo baseada em semivariogramas experimentais. A Figura 7.2 mostra a tela principal do SPRING 3.3 com a interface de interação para análise da continuidade espacial por semivariograma. Este é um procedimento interativo que permite, ao usuário, a experimentação de diferentes valores para parâmetros necessários à definição do semivariograma experimental. Pode-se definir os parâmetros lineares, ou seja, o número de passos ("lags"), o incremento de distância e a tolerância do incremento. Também é possível a definição dos parâmetros angulares, de direção, tolerância e largura de banda. A análise de continuidade espacial compreende, além do levantamento do semivariograma experimental, o ajuste a uma família de modelos de semivariogramas teóricos incluindo os modelos linear, esférico, exponencial e gaussiano e a validação cruzada do modelo ajustado. A Figura 7.3 ilustra o procedimento de ajuste de um modelo teórico, para o semivariograma experimental, no SPRING 3.3.
- Para finalizar, a inferência dos dados é realizada pelas técnicas de krigeagem simples e ordinária e os resultados, campo de valores

inferidos e campo de variâncias de estimação, armazenados em planos de informação do sistema SPRING. A Figura 7.4 mostra interface de interação, do SPRING 3.3, para espacialização de um atributo por krigeagem simples ou ordinária.

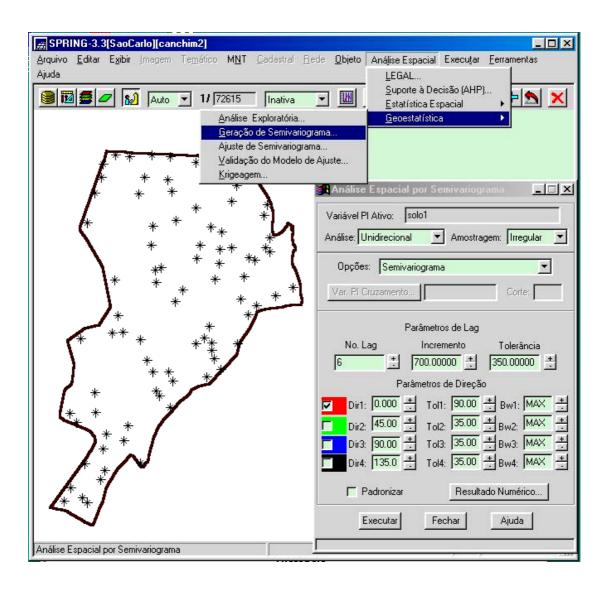


Fig. 7.2 - Tela principal do SPRING 3.3 com a interface de interação para análise de continuidade espacial por semivariograma

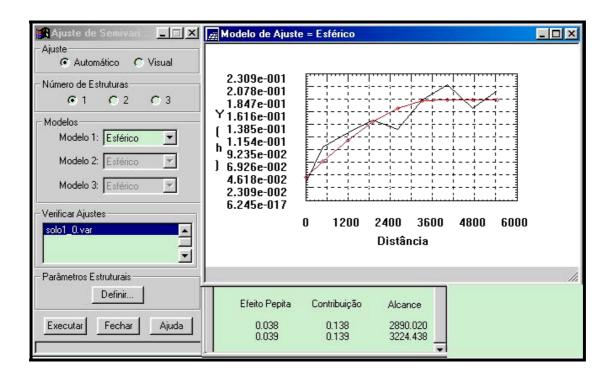


Fig. 7.3 - Interface do SPRING 3.3 para ajuste de um modelo teórico para o semivariograma experimental

Em resumo, o módulo de geoestatística do SPRING 3.3 contém: vários procedimentos de análise exploratória das amostras; procedimentos para análise da continuidade espacial de um atributo e procedimentos para validação do modelo e espacialização do atributo por krigeagem linear. Comparando essa funcionalidade com os requisitos descritos em 7.3.1, esse módulo necessita ser complementado com procedimentos de transformação das amostras e com outros procedimentos de espacialização, tais como: krigeagem universal, krigeagem por indicação, cokrigeagem e simulação estocástica linear e por indicação.

Os estudos e resultados deste trabalho deverão ser utilizados para implementação de uma extensão do módulo atual do SPRING. A idéia é adicionar, ao módulo, as funcionalidades dos procedimentos geoestatísticos baseados na krigeagem por indicação. Assim, o sistema permitirá interpolação

e estimativa de incertezas utilizando a técnica de krigeagem por indicação e estimação de campos equiprováveis, interpolação e estimativa de incertezas utilizando a técnica de simulação estocástica condicional por indicação.

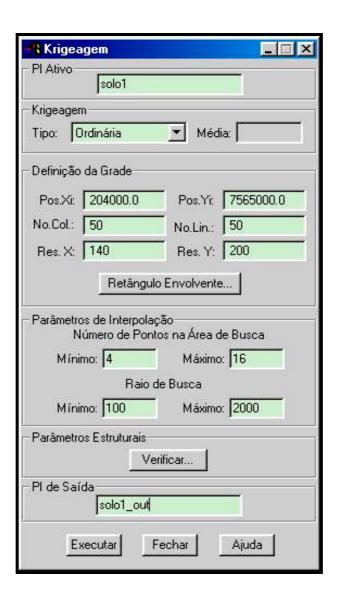


Fig. 7.4 - Interface do SPRING 3.3 para espacialização do atributo por krigeagem simples ou ordinária

Essas novas funcionalidades, uma vez integradas ao SPRING, possibilitarão:

- a espacialização de atributos sem a necessidade de se considerar um modelo de distribuição a priori para o atributo em questão;
- a espacialização de atributos numéricos e, também, de atributos temáticos:
- a espacialização de atributos com uso de informação indireta, complementar as amostras diretas, para diminuir a incerteza do resultado;
- a criação de mapas de médias e incertezas por variância, ou desvio padrão, para atributos numéricos;
- a criação de mapas de mediana e incerteza por intervalos interquantis para atributos numéricos;
- a geração de mapas classificados por moda, com incerteza complementar à máxima probabilidade, para atributos temáticos;
- a geração de mapas classificados por moda, com incerteza por entropia de Shannon, para atributos temáticos e;
- a geração de campos numéricos e temáticos equiprováveis, por simulação sequencial condicionada, para serem utilizados em espacialização com estimativa de incertezas e, também, em análises e propagação de incertezas para modelagens computacionais.

Como já visto anteriormente, os procedimentos baseados em krigeagem por indicação requerem, para cada valor de corte, o ajuste de um modelo de variografia teórico. Operacionalmente isto pode ser realizado com o mesmo procedimento de análise de continuidade espacial, atual do SPRING 3.3, sobre os valores previamente transformados por indicação.

## 7.3.3 O módulo de geoestatística do IDRISI

Segundo informação distribuída em http://www.idrisi.com/03prod/ldrisi.htm, o IDRISI é um software com funcionalidade analítica matricial, cobrindo completamente o espectro de necessidades de SIG e sensoriamento remoto desde consulta a banco de dados e modelagem espacial até classificação e melhorias ("enhancement") em imagem. O sistema inclui facilidades especiais para monitoramento ambiental e gerenciamento de recursos naturais, incluindo análises por séries espaço-temporais, suporte a decisão multi-objetiva e multi-critério, análise de incerteza, incluindo, análise Bayesiana e por conjuntos nebulosos e modelagem por simulação.

O IDRISI32, uma versão do sistema IDRISI para 32 bits, fornece um conjunto de ferramentas geoestatísticas para a inferência e modelagem de características de superfícies a partir de amostras pontuais. As ferramentas integradas ao IDRISI se baseiam no **gstat** (http://www.geog.uu.nl/gstat/), um programa para modelagem e predição geoestatística multivariável e simulação de variáveis em 1, 2 ou 3 dimensões.

O sistema IDRISI32 possui um conjunto de módulos para análise exploratória, dos quais podem-se destacar: apresentações interativas, análise de histogramas, estatísticas de janelas móveis, cálculo de autocorrelação espacial, análises de tendência e análises por diagrama de dispersão. Além destes módulos, o IDRISI32 fornece 3 interfaces interativas para funções da gstat (http://www.idrisi.com/17News/17news.htm):

O Modelador de Dependência espacial, ilustrado na Figura 7.5, é uma interface interativa que utiliza funções da gstat para criar diagramas de dispersão, modelos de variograma de superfície, modelos de variograma direcional e um sumário de estatísticas adicionais. A continuidade espacial pode ser modelada usando-se os seguintes métodos: semivariograma, covariograma, correlograma, variograma cruzado, covariograma cruzado e correlograma cruzado. Transformações de dados possibilitam aplicações de

transformações normais e logarítmicas e transformações por indicação. Os variogramas podem ser omnidirecionais ou direcionais. A representação por superfície pode ser usada para se determinar as direções de continuidade espacial mínima e máxima.

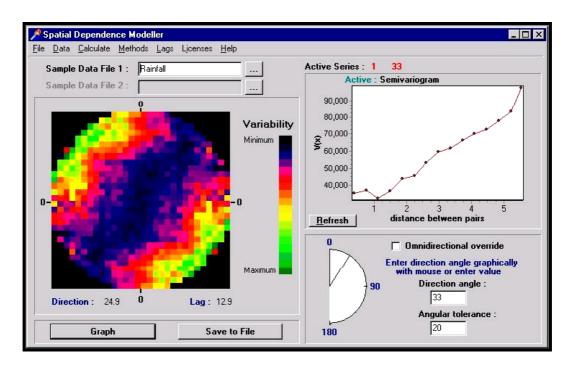


Fig. 7.5 - Ilustração do Modelador de Dependência Espacial do IDRISI32. Fonte: http://www.idrisi.com/03prod/Idrisi.htm

- A interface de ajuste de curvas, ilustrada na Figura 7.6, possibilita o ajuste interativo de diferentes estruturas matemáticas para o variograma. Estão disponíveis 10 tipos de estruturas: esférica, exponencial, gaussiana, linear com patamar, potência, logarítmica, circular, pentaesférica e periódica. Os métodos de ajuste incluem mínimos quadrados ponderados, mínimos quadrados ponderados com pesos normalizados e probabilidade máxima restrita ("Restricted Maximum Likelihood" REML).
- A interface de krigeagem e simulação, ilustrada na Figura 7.7, fornece acesso para os estimadores de krigeagem e cokrigeagem simples e

ordinária, de krigeagem estratificada e de simulação gaussiana condicional e incondicional, incluindo múltiplas simulações. Ela inclui transformação de normalização e logarítmica para os dados, múltiplos métodos para definição dos parâmetros de vizinhança e validação cruzada dos dados amostrais. Uma imagem máscara pode ser usada para limitar as regiões a serem interpoladas. As máscaras podem ser usadas também para definir regiões estratificadas que serão interpoladas separadamente.

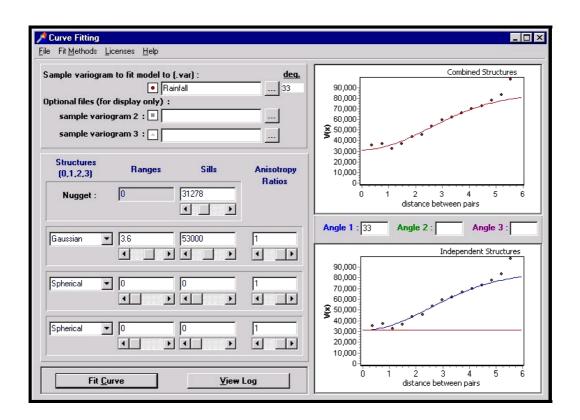


Fig. 7.6 - Interface de ajuste de curvas do sistema IDRISI32. Fonte: http://www.idrisi.com/03prod/Idrisi.htm

Comparando-se as funcionalidades dos módulos de geoestatística do SPRING 3.3 e do IDRISI32, este último oferece, a mais, opções de krigeagem estratificada, cokrigeagem e simulação Gaussiana. Por outro lado o IDRISI32

também não oferece krigeagem e simulação condicionada por indicação o que torna o sistema limitado para inferência de incertezas de distribuições não gaussianas. Para finalizar, ressalta-se que o módulo de geoestatística do IDRISI32 foi prometido para ser lançado em Outubro de 1999. Porém, o que existe, até então, é uma versão beta-teste e uma versão oficial ainda não foi disponibilizada.

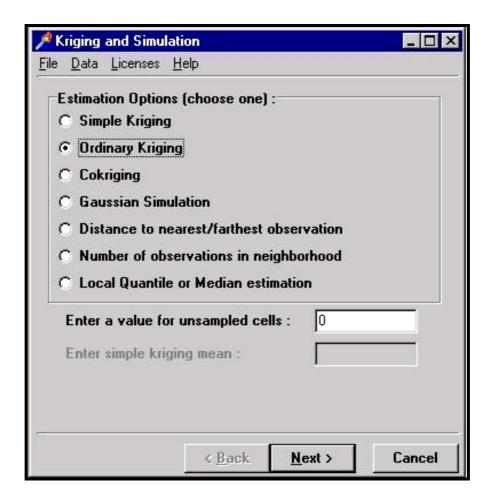


Fig. 7.7 - Interface de krigeagem e simulação do sistema IDRISI32. Fonte: http://www.idrisi.com/03prod/ldrisi.htm

#### 7.3.4 O módulo de geoestatística do PCRaster

O PCRaster é um SIG matricial que integra funcionalidades de um SIG, de modelagem ambiental e de modelagem geoestatística

(http://www.geog.uu.nl/pcraster/what.html). O sistema contém uma Linguagem de Modelagem Dinâmica para construção interativa de modelos ambientais espaço-temporais integrada ao SIG. A modelagem geoestatística é feita pelo pacote Gstat 2.0.g (http://curie.ei.jrc.it/software/Gstat.htm) que está fortemente acoplado ao sistema. A Figura 7.8 mostra o ajuste de um semivariograma pelo Gstat. O sistema contém, também, um pacote para modelagem de propagação de erro, chamado ADAM, que será abordado na Seção 7.4.

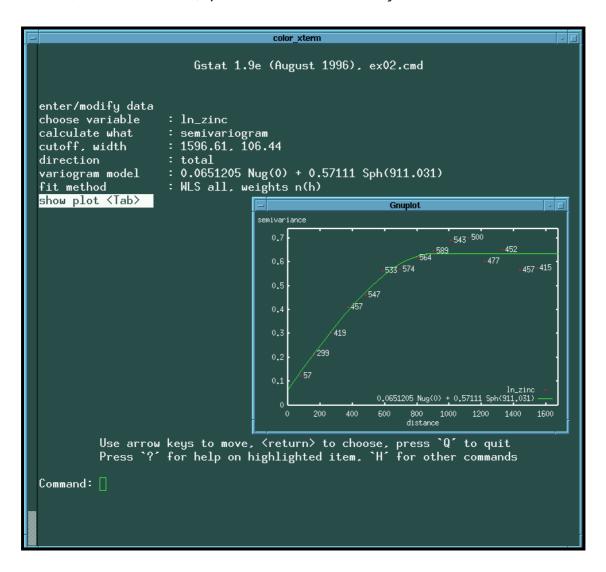


Fig. 7.8 - Exemplo de ajuste de variograma pelo programa GSTAT

- O módulo de geoestatística do PCRaster possibilita (http://curie.ei.jrc.it/software/pcraster.htm):
- Análise exploratória: O PCRaster possibilita operações de vizinhança e contem um menu de modelagem de variograma para modelagem de variogramas simples, múltiplos, isotrópicos e anisotrópicos.
- Interpolação de dados pontuais: O sistema oferece opção de interpolação para mapas do PCRaster ou para tabelas de dados pontuais formatada. Os procedimentos de interpolação disponíveis são a krigeagem com um atributo, a cokrigeagem, a krigeagem por blocos, a krigeagem estratificada, a krigeagem universal e a interpolação pelo inverso da distancia.
- Simulação: simulação de campos aleatórios correlacionados espacialmente com base no variograma modelado.

O PCRaster funciona por linguagem de comandos no sistema operacional DOS ou UNIX. Para apresentação de gráficos, como os variogramas por exemplo, o PCRaster necessita do programa Gnuplot, que é distribuído sem custo.

Comparando a funcionalidade do PCRaster com o IDRISI32 parece não haver diferenças significativas de procedimentos. Também, como o SPRING 3.3 e o IDRISI32, o PCRaster não oferece krigeagem e simulação condicionada por indicação o que torna o sistema limitado para inferência de incertezas de distribuições não gaussianas.

#### 7.3.5 Outros módulos de geoestatística

Está fora do escopo deste trabalho realizar uma busca exaustiva de todos os SIG que contem, ou estão em vias de implementar, módulos de geoestatística. Os três exemplos, apresentados nas seções anteriores, são amostras representativas do estado atual no que se refere à integração de um módulo de geoestatística em um SIG.

Além dos SIG apresentados acima, vale a pena ainda citar que, para a versão 8 do ArcInfo, está sendo prevista uma extensão chamada "Geostatistical Analyst Designed for Advanced Spatial Modeling" (http://www.esri.com/news/arcnews/spring99articles/07-geostatistical.html).

Esse módulo conterá ferramentas desenvolvidas especificamente para geração de superfícies, com análise de erro da estimação resultante, usando ferramentas geoestatísticas. A extensão conterá vários tipos de krigeagem e ferramentas preditivas. Segundo os responsáveis pela nova versão do ArcInfo, a extensão será diferente de todas as outras em dois aspectos: alta integração com a funcionalidade do SIG e a interface com o usuário que será poderosa e fácil de usar.

Outra referência interessante é uma extensão chamada "Splus for Arcview & S+Spatial Stats 1.0" (http://www.mathsoft.com/splus/splsprod/arcview.htm) que objetiva acoplar a funcionalidade dos pacotes de softwares SPLUS e S+Spatial Stats (http://www.mathsoft.com/splus/) com o SIG da ESRI denominado Arcview (http://www.esri.com/software/arcview/index.html).

#### 7.4 LINGUAGENS PARA MODELAGEM AMBIENTAL COM TRATAMENTO DE INCERTEZA

Nos primeiros SIG do mercado, as operações, envolvendo diversos dados espaciais, eram realizadas por comandos ou módulos de álgebra de mapas que impunham várias restrições aos modelos, principalmente aos mais complexos.

Para facilitar a implementação de modelos espaciais, no ambiente de um SIG, foram desenvolvidas linguagens para modelagem, que realizam consultas e manipulação de dados espaciais armazenados na base geográfica. Essas linguagens são desenvolvidas para se implementar modelos matemáticos que representam fenômenos espaciais. Wesseling e Heuvelink (1993) argumentam que para se ter uma linguagem de propósito geral no ambiente de um SIG, o projeto da linguagem e o modelo de dados do SIG devem ser consistentes entre eles.

Outra opção é a implementação um módulo interativo, denominado *módulo* para modelagem espacial, que possibilita a construção dos modelos com posteriori criação, automática, dos programas na linguagem hospedeira do SIG. Caso o módulo de modelagem esteja bem projetado, o esforço do usuário é diminuído, em comparação com a necessidade de se escrever um programa, uma vez que ele não precisa aprender a programar nessa linguagem. Também, os programas gerados pelo módulo são livres de erros, pelo menos de sintaxe.

Após ser implementado, testado e aprovado, o programa, que implementa um modelo ambiental, pode ser reutilizado para novos conjuntos de dados, equivalentes aos utilizados na implementação do programa. Esses dados podem ser de uma mesma região, obtidos em épocas diferentes, ou de outras regiões. Outra vantagem do uso de uma linguagem é a possibilidade de se realizar análises de sensibilidade do modelo aos seus parâmetros de entrada. Esses parâmetros podem ser modificados no programa, de forma simples e rápida, e este pode ser executado diversas vezes até se conseguir um patamar de sensibilidade aceitável. Finalmente, o programa de modelagem pode servir de documentação do modelo.

Como já ressaltado, este trabalho defende a tese de que deve-se agregar informação de qualidade aos dados espaciais armazenados na base de dados de um SIG. Esta nova concepção influencia diretamente no projeto de linguagens de modelagem espacial, pois, a informação de qualidade dos dados deve ser considerada e propagada para o resultado da modelagem. Operadores específicos para análise e propagação de incertezas devem ser incluídos na linguagem, além dos que já existem. As incertezas dos dados espaciais, as incertezas dos coeficientes do modelo, os tipos de dados e os tipos de operações envolvidas na modelagem são os parâmetros típicos a serem considerados no projeto desses novos operadores.

Segundo esse novo paradigma, um ambiente computacional amigável, linguagem ou módulo, para modelagem ambiental deve:

- facilitar a definição dos parâmetros de entrada, coeficientes e variáveis, do modelo:
- facilitar a definição das estatísticas dos parâmetros de entrada incluindo,
   média, variância, probabilidades, tipo de distribuição, etc.;
- facilitar a definição das estatísticas multivariadas, covariância ou correlação cruzada por exemplo, entre as variáveis de entrada.
- facilitar a definição do modelo matemático de integração, ou seja, das operações envolvidas na modelagem;
- sugerir métodos apropriados de análise e propagação das incertezas e;
- informar sobre a sensibilidade do modelo em relação aos parâmetros de entrada.

As duas subseções seguintes são dedicadas ao tema de linguagens para modelagem espacial com propagação de incerteza. Na Seção 7.4.1 descreve-se a ferramenta **ADAM**, uma solução desenvolvida para dados espaciais de natureza numérica. Na Seção 7.4.2 aborda-se a linguagem **LEGAL** do SPRING sugerindo-se modificações na linguagem para inclusão de facilidades para o tratamento de incertezas.

## 7.4.1 ADAM: uma ferramenta para modelagens com propagação de incertezas

**ADAM** é uma ferramenta de software para acompanhar a propagação de erro em modelagem espacial quantitativa com SIG (Wesseling e Heuvelink 1993).

A principais características da ferramenta ADAM são:

 O usuário da ADAM define o modelo, suas entradas e seus parâmetros estatísticos, na linguagem de descrição de modelos MDL;

- A ferramenta ADAM funciona como um compilador, transformando o problema em uma sequência de operações padrão que serão processadas por um SIG. Ela emprega um método de diferenciação simbólica, ao invés de numérica, quando derivadas de primeira e segunda ordem são requeridas pelo método de propagação;
- Durante sua execução, a ADAM auxilia o usuário na seleção da técnica de propagação de erro mais apropriada. Para isto ela usa um conjunto de regras predefinidas.

Segundo Heuvelink (1998), uma ferramenta como a ADAM tem as seguintes vantagens:

- ao criar uma sequência de operações, a serem executadas pelo SIG, utilizam-se as estruturas e o gerenciador da base de dados e, também, as capacidades analíticas do próprio sistema hospedeiro, adicionando, ao invés de substituir, funcionalidades ao SIG;
- ela é flexível no sentido de que pode manipular uma variedade de operações do SIG, pode ser integrada a vários tipos de SIG e não é dependente das estruturas, vetorial ou matricial, de representação e de integração dos dados;
- a ferramenta contém uma interface com o usuário que é eficiente para obtenção dos dados requeridos por determinados tipos de modelagem.
   Por exemplo, um parâmetro estatístico pode ser recuperado da base ou calculado automaticamente.
- sua interface com o usuário possibilita o exame, numa forma exploratória, da variação do erro de saída em relação a mudanças nos parâmetros de erro de uma ou mais entradas;

 ela trabalha com algoritmos numéricos eficientes e técnicas de otimização de código comuns com a finalidade de minimizar o tempo de computação

Uma desvantagem da ferramenta ADAM é sua limitação para tratamento de modelos ambientais, pois aceita somente modelos com dados numéricos. Seria interessante que essa ferramenta fosse estendida para modelos lógicos envolvendo atributos temáticos. Neste caso a ferramenta ADAM deveria incorporar operações e técnicas de propagação de incerteza para dados temáticos.

## 7.4.2 A linguagem LEGAL com tratamento de incertezas: Uma especificação preliminar de requisitos

O SPRING tem uma linguagem de consulta e manipulação espacial denominada Linguagem Espacial para Geoprocessamento Algébrico (LEGAL). O objetivo da LEGAL é prover um ambiente geral para análise geográfica, incluindo operações sobre geo-campos e geo-objetos (Câmara et al. 1996b).

A implementação atual da linguagem LEGAL manipula dados espaciais determinísticos. Os dados de entrada, os geo-campos, os geo-objetos e os coeficientes, de um modelo ambiental são considerados informações determinísticas, ou seja, não possuem componentes de incerteza. O objetivo desta seção é apresentar e discutir alguns requisitos básicos para que a funcionalidade da Linguagem LEGAL seja estendida para manipular incertezas dos dados espaciais e propaga-las. Os seguintes requisitos são considerados:

 Possibilitar a definição de variáveis estocásticas, além das determinísticas já consideradas. Como a LEGAL é fortemente tipada, ou seja, as variáveis utilizadas nas manipulações são definidas, a priori, como pertencentes a um determinado tipo (IMAGEM, TEMÁTICO, NUMÉRICO, etc.), é necessário estender o conjunto destes tipos para:

- que complementem o conjunto de tipos atuais da linguagem, adicionando tipos equivalentes, porém com uma componente estocástica. Um exemplo seria a criação do tipo IMAGEM\_EST, que é o equivalente ao tipo IMAGEM complementado com o componente estocástico, mapa de incerteza. Outro exemplo é a criação do tipo REAL\_EST, que é equivalente ao tipo REAL, porém contém um valor de incerteza associado. Outra solução, mais elegante, seria a redefinição dos tipos atuais para permitir instanciações com, e sem, a componente aleatória.
- que representem informação de correlação espacial entre duas variáveis. Por exemplo, um tipo IMAGEM\_CORREL pode ser criado para se definir variáveis representando índices de correlação entre pixels de duas imagens.
- que representem modelos de variografia. Por exemplo, o tipo VARIOG pode ser criado para se definir variáveis com os parâmetros de um variograma teórico, por exemplo, o modelo de ajuste, as direções de anisotropia e os valores do efeito pepita, do patamar e do alcance.
- que representem modelos de distribuição de uma variável aleatória.
   Por exemplo, o tipo MODEL\_PROB pode ser criado para se definir variáveis com os parâmetros de um modelo probabilístico (curva, valor médio e desvio padrão)
- Implementar formas de instanciação para os novos tipos criados.
   Exemplos:

```
    a = (10.754, 1.33),
    sendo a uma variável do tipo REAL com componente aleatória;
    b = (30.75, 10%),
    sendo b uma variável do tipo REAL com componente aleatória;
```

- c = (Spheric, 45, 135, 0.22, 5.33, 1200.83, 3825.55),
  sendo c uma variável do tipo VARIOG;
  d = (Normal, 2525.74, 333.45),
  sendo d uma variável tipo MODEL PROB.
- Modificar os comandos de instanciação, RECUPERE e NOVO, para recuperar e criar, respectivamente, dados da base com suas informações de incerteza.
- Implementar novos operadores para cálculo de curtose, de autocorrelação, de correlação cruzada, de estimação por krigeagem e por simulação estocástica, de validação do modelo, etc.
- Implementar operadores de propagação de incerteza. A linguagem deve fornecer operadores para vários métodos de propagação de incerteza. Além disso, a linguagem deve verificar se o método escolhido é aplicável para as operações e os dados envolvidos numa determinada modelagem.
- Inserir capacidade de diferenciação simbólica ou numérica. Alguns métodos de propagação necessitam de derivadas parciais de 1<sup>a</sup> e 2<sup>a</sup> ordem.
- Inserir mecanismos de consistência na linguagem para que ela seja capaz de sugerir os melhores métodos de propagação em função dos dados e operadores do modelo.

O conjunto de requisitos descritos acima não está completo. Está fora do escopo deste trabalho uma especificação completa da extensão da linguagem Legal para tratamento de incertezas. Espera-se que a contribuição acima motive novos trabalhos nessa direção.

#### 7.5 CONCLUSÕES

No presente capítulo explorou-se o tema de integração de procedimentos geoestatísticos com os SIG. A solução inicial, que já está sendo adotada, é a incorporação de módulos de geoestatística ao SIG. Na prática, os módulos implementados estão baseados em softwares de geoestatística disponíveis no mercado há algum tempo. Neste capítulo apresentaram-se algumas soluções utilizando os softwares GSLIB, Gstat e Splus. Assim, para implementação do módulo, o trabalho principal, dos desenvolvedores de SIG, está relacionado com a completeza da funcionalidade e com a facilidade de uso dos módulos. Estes devem conter um conjunto mínimo de procedimentos geoestatísticos e devem interagir com o usuário de forma amigável, por sistemas de janelas e menus por exemplo.

Entretanto, somente a integração de um módulo de geoestatística não explora todo o potencial que a geoestatística pode oferecer para as aplicações desenvolvidas em um SIG. Além do módulo de geoestatística, os SIG que possuem linguagem de modelagem espacial devem incorporar funcionalidades geoestatística linguagem. inclui, nessa Isto principalmente, procedimentos de espacialização por krigeagem e por simulação estocástica. Como mostrado neste trabalho, uma das principais vantagens dos procedimentos da geoestatística é a possibilidade de inferência de incertezas associadas às espacializações. O dado de incerteza constitui um novo elemento a ser tratado pelo SIG. Ele deve ser armazenado e recuperado, sempre com uma referência ao dado espacial associado. Além disso, a incerteza é propagada nas análises espaciais realizadas no ambiente do SIG. Assim, as linguagens de modelagem espacial devem conter, também, novos operadores que possibilitem propagar as incertezas dos dados de entrada para os produtos das análises e modelagens.

Em resumo, pode-se dizer que a integração de funcionalidades da geoestatística no ambiente de um SIG possibilita:

- analisar e inferir valores de atributos de dados espaciais de natureza numérica e temática;
- calcular incertezas associadas aos valores inferidos e;
- trabalhar com modelagem espacial que faz uso de realizações inferidas por procedimentos de simulação, que considera incertezas dos parâmetros e dos dados espaciais do modelo e que propaga as incertezas para o(s) produto(s) da modelagem.

Para finalizar, os benefícios reportados neste capítulo explicam o crescente interesse da comunidade, de desenvolvedores de software e de usuários, por procedimentos da geoestatística integrados ao contexto de um SIG. Por isso, pode-se prever que, num futuro próximo, todos os principais SIG do mercado possuirão funcionalidades da geoestatística e estarão oferecendo ferramentas de tratamento de incertezas que possibilitarão qualificar os produtos gerados por esses sistemas.

# CAPÍTULO 8 CONCLUSÕES

#### 8.1 CONSIDERAÇÕES FINAIS

O presente trabalho apresentou uma proposta de sistematização, baseada nos procedimentos da geoestatística por indicação, para ser utilizada em modelagens computacionais desenvolvidas em ambiente de SIG. Esta proposta apresenta vantagens, que foram apresentadas, discutidas e analisadas nos capítulos anteriores deste trabalho, em relação aos procedimentos anteriormente utilizados, a modelagem determinística e modelagem estocástica linear.

De forma sucinta, os elementos que fizeram parte dessa sistematização foram:

- a exploração teórica e prática dos procedimentos estocásticos não linear, a krigeagem e a simulação condicionada por indicação, para inferência de atributos numéricos e temáticos. Os Capítulos 3 e 4 apresentaram os formalismos teóricos desses procedimentos e análises das vantagens desses estimadores em relação aos procedimentos determinísticos e estocástico linear. Os resultados apresentados nos estudos de caso do Capítulo 6 mostram que o uso dessas técnicas é viável e promissor;
- 2) a proposição, testes e comparações de métodos de estimativa de incertezas baseados nos modelos de distribuição inferidos pelos procedimentos estocásticos por indicação. Foi proposto o uso de diferentes métricas de incerteza para atributos numéricos e temáticos. Os resultados apresentados no Capítulo 6 mostram que as incertezas estimadas refletem o comportamento da variabilidade do atributo modelado.

- 3) a exploração teórica e prática do uso de dados indiretos, em conjunto com um quantidade insuficiente de amostras diretas, para melhorar a acurácia das inferências dos atributos espaciais. O Capítulo 5 apresentou a formalização teórica dos procedimentos geoestatísticos por indicação para suportar a inclusão de dados indiretos codificados como valores de probabilidades. O estudo de caso apresentado no Capítulo 6 mostra que a inclusão de dados indiretos melhorou a acurácia da inferência e gerou um mapa de inferências que reflete a distribuição das amostras e a informação subjetiva presente no mapa de solos (que serviu de fonte para os dados indiretos). Este é um resultado relevante pois, em muitos casos, a amostragem é escassa e mal distribuída.
- 4) a especificação dos requisitos básicos para integração dos procedimentos geoestatísticos por indicação no estágio atual de desenvolvimento tecnológico dos SIG. O Capítulo 7 discute e apresenta os requisitos para integração de um módulo de geoestatística a um SIG e a especificação de requisitos para incluir informação de incerteza numa linguagem de modelagem espacial que considera e propaga incertezas do modelo.

Com a integração das funcionalidades descritas acima, mais a inclusão de incerteza nos demais dados espaciais da base, espera-se que os usuários possam ter acesso a informação quantitativa sobre os riscos assumidos nos planejamentos, nas simulações e tomadas de decisões apoiados sobre os produtos dos SIG.

#### 8.2 SUGESTÕES PARA FUTURAS INVESTIGAÇÕES

Obviamente, este trabalho não considera todas as possibilidades em relação ao potencial da geoestatística por indicação. Propõe-se mais investigações envolvendo os seguintes temas:

- Pesquisar novas metodologias para cálculo de incertezas locais e, principalmente, globais a partir dos resultados da geoestatística por indicação;
- Investigar e desenvolver metodologias de propagação de incertezas em modelagens ambientais envolvendo atributos temáticos com e sem correlação espacial entre os atributos;
- Investigar novas metodologias de simulação estocástica;
- Fazer um estudo comparativo em relação às inferências resultantes da krigeagem linear, da krigeagem por indicação por mediana e da krigeagem por indicação com vários valores de corte.

Como consequência natural deste trabalho propõe-se as seguintes tarefas envolvidas com a evolução do SIG SPRING:

- Complementar o módulo de geoestatística do SPRING com a implementação dos procedimentos de krigeagem e simulação por indicação e de estimativa de incertezas testados e analisados neste trabalho.
- Estender a linguagem LEGAL para considerar a informação de incerteza nas análises implementadas na linguagem.

### REFERÊNCIAS BIBLIOGRÁFICAS

- Almeida, A.; Journel, A. Joint simulation of multiple variables with a Markov-type coregionalization model. **Mathematical Geology**, v. 26, n. 5, p. 565-588, 1993.
- Arbia, G. The use of SIG in spatial statistical surveys. **International Statistical Review**, v.61, n. 2, p. 339-359, 1993.
- Burrough, P. A. **Principles of geographical information systems for land resources assessment.** Oxford: Clarendon Press, 1987. 193p.
- Burrough, P. A. Development of intelligent geographical information systems. **International Journal of Geographical Information Systems**, v. 6, n. 1, p.1-11, 1992.
- Burrough, P. A.; van Rijn, R.; Rikken, M. Spatial data quality and error analysis issues: GIS functions and environmental modeling. In: Goodchild, M. F.; Steyaert, L. T.; Parks, B. O.; Johnston, C.; Maidment, D.; Crane, M.; Gendinning, S. ed. **GIS and Environmental Modeling:** Progress and Research Issues. Fort Collins: GIS World, Inc, 1996. Cap. 5, p. 29-34.
- Burrough, P. A.; McDonnell, R. A. **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.
- Câmara, G. Modelos, linguagens e arquiteturas para bancos de dados geográficos. [on line]. São José dos Campos. 264 p. Tese (Doutorado em Computação Aplicada http://www.dpi.inpe.br/teses/gilberto/) Instituto Nacional de Pesquisas Espaciais, 1995.

- Câmara, G.; Souza, R. C. M.; Freitas U. M.; Garrido, J. SPRING: Integrating remote sensing and GIS by object-oriented data modelling. **Computer & Graphics**, v. 20, n. 3, p. 395-403, 1996a.
- Câmara, G.; Casanova M.; Hemerly, A.; Medeiros, C. B.; Magalhães, G.

  Anatomia de sistemas de informação geográfica. Campinas: UNICAMP Instituto de Computação, 1996b. 193p.
- Câmara, G.; Medeiros, J. S. Princípios básicos em geoprocessamento. In:
  Assad, E. D.; Sano, E. E. ed. **Sistemas de Informações Geográficas. Aplicações na Agricultura.** Brasília: Embrapa-SPI / Embrapa-CPAC, 1998.
  Cap 1, p. 3-11.
- Camargo, E. C. G. **Desenvolvimento, implementação e teste de procedimentos geoestatísticos (krigeagem) no Sistema de Processamento de Informações Georeferenciadas (SPRING).** São José dos Campos. 124 p. (INPE-6410-TDI/620). Dissertação (Mestrado em Sensoriamento Remoto) Instituto Nacional de Pesquisas Espaciais, 1997.
- Camargo, E. C. G; Monteiro, A. M. V.; Felgueiras, C. A.; Fuks, S. D. Integração de geoestatística e sistemas de informação geográfica: uma necessidade. [CD-ROM]. In: Congresso e Feira para Usuários de Geoprocessamento da América Latina, 7, Salvador, 1999. **Anais**. Salvador: gisbrasil'99. Seção de Posters Técnico-Científicos.
- Cressie, N. A. C. **Statistics for spatial data**. New York: John Wiley, 1991. 900p.
- Delfiner, P.; Delhomme, J. P. Optimum interpolation by kriging. In: Davis, J. C.; McCullagh, M. J. ed. **Display and Analysis of Spatial Data**. New York: John Wiley, 1975. p. 96-114.
- De Oliveira, J. L.; Pires, F.; Medeiros, C. B. An environment for modeling and design of geographic applications. **Geoinformatica**, v. 1, p. 29-58, 1997.

- Deutsch, C. V.; Journel, A. G. **GSLIB Geostatistical software library and user's guide**. New York: Oxford University Press, 1998. 369p.
- Englund, E. J. Spatial simulation: environmental applications. In: Goodchild, M.F.; Bradley, O. P.; Steyaert L. T. ed. Environmental modeling with GIS.New York: Oxford University Press,1993. p. 432-437.
- Felgueiras, C. A. **Desenvolvimento de um sistema de modelagem digital de terreno para microcomputadores.** São José dos Campos. 202 p. (INPE-4406-TDI/303). Dissertação (Mestrado em Computação Aplicada) Instituto Nacional de Pesquisas Espaciais, 1987
- Felgueiras, C. A.; Monteiro, A. M. V.; Fuks, S. D.; Camargo, E. C. G.
  Inferências e estimativas de incertezas utilizando técnicas de krigeagem não linear. [CD-ROM]. In: Congresso e Feira para Usuários de
  Geoprocessamento da América Latina, 7, Salvador, 1999a. Anais.
  Salvador: gisbrasil'99. Seção de Palestras Técnico-Científicas.
- Felgueiras, C. A.; Monteiro, A. M. V.; Camargo, E. C. G.; Câmara, G.; Fuks, S. D. Integrating geostatistical tools in geographical information systems. In: Brazilian Workshop on GeoInformatics, 20-21 out, Campinas, 1999b.
  Proceedings. São Paulo: Geo-Info, 1999. Palestra Técnico-Científica. p. 40-44
- Fuks, S. D. Novos modelos para mapas derivados de informações de solos.
   In: Assad, E. D.; Sano, E. E. ed. Sistemas de informações geográficas.
   aplicações na agricultura. Brasília: Embrapa-SPI / Embrapa-CPAC, 1998.
   p. 373-410
- Goodchild, M. F.; Guoging, S. Development and test of an error model for categorical data. **International Journal of Geographical Information Systems**, v. 6, n.2, p.87-104, 1992.
- Goodchild, M. F. The state of GIS for environmental problem solving. In: Goodchild, M. F.; Bradley, O. P.; Steyaert L. T. ed. **Environmental modeling with GIS**. New York: Oxford University Press, 1993a. p. 8-15.

- Goodchild, M. F. Data models and data quality: problems and prospects. In: Goodchild, M. F.; Bradley, O. P.; Steyaert L. T. ed. **Environmental modeling with GIS**. New York: Oxford University Press, 1993b. p. 94-103.
- Goovaerts, P.; Journel, A. G. Integrating soil map information in modelling the spatial variation of continuous soil properties. **European Journal of Soil Science**, v. 46, n. 3, p. 397-414, 1995.
- \*Goovaerts, P. **Geostatistics for natural resources evaluation**. New York: Oxford University Press, 1997. 481p.
- Heuvelink, G.B.M.; Stein P. A. Propagation error in spatial modelling with GIS. **International Journal of Geographical Information System**, v. 3, p.303-322, 1989.
- Heuvelink, G. B. M.; Burrough, P. A.; Stein, A. Error propagation in cartographic modelling using boolean logic and continuous classification.

  International Journal of Geographical Information Systems, v. 7, n. 3, p.231-246, 1993.
- Heuvelink, G. B. M. Identification of field attribute error under different models of spatial variation. **International Journal of Geographical Information System,** v. 10, n. 8, p. 921-936, 1996.
- Heuvelink, G. B. M. Error propagation in environmental modelling with GIS. Bristol: Taylor and Francis Inc, 1998. 127p.
- Huijbregts, C. J. Regionalized variables and quantitative analysis of spatial data. In: Davis, J. C.; McCullagh, M. J. ed. **Display and Analysis of Spatial Data**. New York: John Wiley, 1975. p. 39-53.
- Isaaks, E. H.; Srivastava, R. M. **An introduction to applied geostatistics**. New York: Oxford University Press, 1989. 561p.
- Instituto Nacional de Pesquisas Espaciais. Departamento de Processamento de Imagens. (INPE/DPI). **Sistema de processamento de informações**

- **georreferenciadas (SPRING)**. http://www.dpi.inpe.br/spring/, Mar, 1999. versão 3.3.
- Journel, A. G. Nonparametric estimation of spatial distributions. **Mathematical Geology**, v. 15, n. 3, p. 445-468, 1983.
- Journel, A.G. **Fundamentals of geostatistics in five lessons**. California: Stanford Center for Reservoir Forecasting Applied Earth Sciences Department, 1988. 85p.
- Journel, A. G.; Deutsch, C. V. Entropy and spatial disorder. **Mathematical Geology**, v. 25, n. 3, p. 329-355, 1993.
- Lee, T. J.; Pielke, R.A.; Kittel, T. G. F.; Weaver, J. F. Atmospheric modeling and its spatial representation of land surface characteristics. In: Goodchild, M. F.; Bradley, O. P.; Steyaert L. T. ed. **Environmental modeling with GIS**. New York: Oxford University Press, 1993. p. 108-122.
- Maguirre, D. J.; Goodchild, M. F.; Rhind, D. W. **Geographical information** systems Volume I. New York: John Wiley and Sons, 1991. 649p.
- Maidment, D. R. GIS and hydrologic modeling. In: Goodchild, M. F.; Bradley,O. P.; Steyaert L. T. ed. Environmental modeling with GIS. New York:Oxford University Press, 1993. p. 147-167.
- Matheron, G. The theory of regionalized variables and its applications.

  Paris: Le Cahiers du Centre de Morphologie Mathematique de Fontainebleu,
  1971. 211p.
- McCullagh, M. J. Terrain and surface modeling systems: theory and practice. **Photogrammetric Record,** v. 12, n. 2, p. 747-779, 1988.
- Moore, I. D.; Turner, A. K.; Wilson, J. P.; Jenson, S. K.; Band, L. E. GIS and land surface subsurface process. In: Goodchild, M. F.; Bradley, O. P.; Steyaert L. T. ed. **Environmental Modeling with GIS**. New York: Oxford University Press, 1993. p. 196-230.

- 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.
- Soares, A. Geostatistical estimation of multi-phase structure. **Mathematical Geology**, v. 24, n. 2, p. 149-160, 1992.
- Stein, A. The use of prior information in spatial statistics. **Geoderma**, v. 62, p.199-216, 1994.
- Steyaert, L. A perspective on the state of environmental simulation modeling. In: Goodchild, M. F.; Bradley, O. P.; Steyaert L. T. ed. **Environmental Modeling with GIS**. New York: Oxford University Press, 1993. p. 16-30.
- #Wesseling, C. G.; Heuvelink, G. B. M. **ADAM user's manual**. Departament of Physical Geography, University of Utrecht, The Netherlands. 1993

## <u>APÊNDICE A</u> <u>DADOS UTILIZADOS NO TRABALHO</u>

#### A.1 DADOS DE CONTORNO DA FAZENDA CANCHIM

X(m)	Y(m)
204382.1	
204412.3	7565759.5
204457.2	7565751.0
204526.5	7565738.0
204642.8	7565713.5
204720.5	7565696.5
204766.6	7565685.0
204854.6	7565631.0
204988.7	7565551.0
205075.1	7565496.5
205098.5	7565483.0
205183.2	7565441.0
205272.1	7565398.5
205346.5	
205403.3	7565479.0
205449.9	
205507.2	7565715.0
205559.8	7565846.5
205593.5	7565930.5
205620.3	7566043.0
205631.6	7566092.5
205662.0	7566178.5
205683.0	7566242.0
205712.9	
205759.3	7566382.5
205795.1	7566431.5
205818.0	7566461.5
205871.7	7566557.0
205913.3	7566630.5
205944.5	7566687.0
206014.4	7566762.0
206091.1	7566843.5
206197.3	7566953.0
206240.0	
206281.5	
206367.3	7566868.0
206466.3	7566748.5
206513.3	7566694.5
206597.3	7566808.5
206640.8	7566833.5
206675.4	7567055.0
206762.0	7567384.5
206868.5	7567626.0
207007.1	7567812.0
207149.0	7567978.5

7568211.0
7568348.0
7568409.5
7568504.0
7568534.5
7568571.5
7568643.0
7569321.0
7569595.5
7569963.0
7570064.0
7570224.0
7570290.0
7570309.5
7570352.0
7570405.0
7570438.5
7570445.5
7570452.0
7570466.5
7570464.5
7570593.5
7570741.5
7570888.0
7571004.5
7571177.0
7571294.0
7571373.0
7571440.0
7571457.5
7571547.0
7571590.0
7571653.0
7571689.0
7571706.0
7571736.5
7571765.0
7571809.5
7571889.5
7571951.5
7571985.5
7572011.5
7572043.5
7572089.0
7572150.0
7572182.5

209995.3	7572195.5
209977.2	7572219.5
209959.5	7572240.5
209977.6	7572263.5
209986.2	7572288.0
209956.6	7572315.0
209937.4	7572338.5
209910.5	7572341.0
209881.6	7572378.0
209863.3	7572353.0
209846.3	7572319.0
209827.6	7572285.5
209740.2	7572177.0
209664.7	7572121.0
209607.7	7572121.0
209576.7	7572104.5
209544.4	7572076.0
209544.4	7572076.0
209515.4	7572192.0
209512.8	
209511.7	7572273.0
	7572307.5
209518.6	7572373.5
209525.3	7572438.0
209530.0	7572515.5
209532.2	7572566.0
209529.8	7572620.5
209514.6	7572698.0
209499.7	7572750.0
209486.3	7572805.5
209478.9	7572831.0
209428.8	7572929.0
209406.0	7572968.5
209310.8	7573127.0
209197.8	7573302.5
209106.1	7573454.0
209082.6	7573502.0
209069.5	7573541.0
209056.9	7573615.5
209112.2	7573712.0
209122.4	7573834.5
209082.0	7573902.5
208989.5	7573951.5
208914.4	7573981.5
208870.6	7573948.0
208831.1	7573956.5
208795.5	7573957.5
•	

#### Conclusão

208750.6	7573969.0
208682.4	7573992.0
208607.5	7573999.5
208563.5	7574012.5
208473.8	7574009.5
208360.4	7573940.5
208206.0	7573899.0
208100.9	7573878.5
208008.7	7573864.5
207911.0	7573879.0
207863.5	7573900.0
207815.1	7573923.0
207784.8	7573948.5
207719.4	7573932.0
206696.1	7573979.5
206274.2	7574002.0
205685.3	7574034.5

205631.9	7573869.5
205536.4	7573629.5
205067.0	7572282.0
205054.7	7572208.5
205243.1	7571964.0
205574.6	7571371.5
205600.2	7571289.5
205767.3	7570425.5
205500.9	7570001.0
205202.1	7569922.0
205173.1	7569937.0
204777.0	7569649.0
204935.0	7569096.5
204890.5	7568746.5
204872.3	7568655.0
204860.2	7568623.5
204723.2	7568502.5

204696.6	7568455.5
204598.8	7567972.0
204599.3	7567873.0
204700.3	7567801.0
205301.4	7567547.5
205341.5	7567505.0
205379.5	7567458.0
205401.7	7567415.0
205415.5	7567383.0
205431.5	7567316.5
205388.6	7567230.0
205346.0	7567149.5
205298.9	7567069.0
205217.9	7566939.5
205128.3	7566802.0
204382.1	7565765.5

## A.2 DADOS AMOSTRAIS DE ALTIMETRIA

V ()	\// \	7/>
X (m)	Y(m)	Z(m)
205000	7565500	859.00
205250	7565500	853.00
205500	7565500	850.00
205750	7565500	862.00
204000	7565750	831.00
204250	7565750	851.50
204500	7565750	861.00
204750	7565750	870.00
205000	7565750	875.00
205250	7565750	838.00
205500	7565750	842.50
205750	7565750	862.50
204000	7566000	830.00
204250	7566000	853.50
204500	7566000	866.00
204750	7566000	877.50
205000	7566000	880.90
205250	7566000	848.00
205500	7566000	855.00
205750	7566000	862.00
206000	7566000	879.00
204000	7566250	850.00
204250	7566250	842.50
204500	7566250	862.50
204750	7566250	878.00
205000	7566250	858.70
205250	7566250	859.00
205500	7566250	844.00
205750	7566250	855.00
206000	7566250	877.50
204000	7566500	858.00
204250	7566500	848.00
204500	7566500	855.00
204750	7566500	872.50
205000	7566500	883.50
205250	7566500	888.00
205500	7566500	872.50
205750	7566500	860.00
206000	7566500	869.00
206250	7566500	894.00
206500	7566500	900.50
206750	7566500	898.70
207000	7566500	899.00
204000	7566750	866.00
204000	7566750	859.00
204230		843.50
204500	7566750 7566750	
204750		867.80 880.80
203000	7566750	000.00

<b>.</b>		-
205250	7566750	887.50
205500	7566750	892.50
205750	7566750	877.50
206000	7566750	868.00
206250	7566750	900.00
206500	7566750	903.50
206750	7566750	902.00
207000	7566750	895.00
204000	7567000	869.00
204250	7567000	872.00
204500	7567000	861.30
204750	7567000	852.00
205000	7567000	865.00
205250	7567000	877.00
	7567000	
205500		901.00
205750	7567000	902.50
206000	7567000	895.00
206250	7567000	896.50
206500	7567000	892.00
206750	7567000	896.50
207000	7567000	886.50
207250	7567000	864.50
207500	7567000	885.80
204000	7567250	865.50
204250	7567250	877.50
204500	7567250	880.00
204750	7567250	872.00
205000	7567250	865.30
205250	7567250	889.50
205500	7567250	905.40
205750	7567250	911.40
206000	7567250	905.70
206250	7567250	886.50
206500	7567250	875.30
206750	7567250	889.70
207000	7567250	882.50
207250	7567250	865.80
207500	7567250	867.00
204000	7567500	860.00
204250	7567500	877.20
204500	7567500	885.10
204750	7567500	884.70
205000	7567500	872.50
205250	7567500	886.30
205500	7567500	902.00
205750	7567500	903.70
206000	7567500	882.30
206250	7567500	875.00
206500	7567500	872.50
200000	1301300	012.00

i <del></del>		
206750	7567500	884.70
207000	7567500	878.30
207250	7567500	864.00
207500	7567500	876.50
204000	7567750	865.00
204250	7567750	876.70
204500	7567750	881.50
204750	7567750	892.00
205000	7567750	891.70
205250	7567750	889.00
205500	7567750	892.50
205750	7567750	892.00
206000	7567750	881.50
206250	7567750	862.30
206500	7567750	867.30
206750	7567750	879.70
207000	7567750	877.30
207250	7567750	865.00
207500	7567750	862.00
204000	7568000	878.00
204250	7568000	879.50
204500	7568000	885.00
204750	7568000	889.00
205000	7568000	885.00
205250	7568000	878.00
205500	7568000	879.00
205750 206000	7568000	879.50
+	7568000	874.00
206250	7568000	855.00 857.40
206500 206750	7568000	
	7568000	865.00
207000	7568000	872.00
207250	7568000	873.00
207500	7568000	867.50
204000	7568250	882.00
204250	7568250	883.00
204500	7568250	883.00
204750	7568250	882.00
205000	7568250	878.00
205250	7568250	860.00
205500	7568250	864.00
205750	7568250	870.50
206000	7568250	870.00
206250	7568250	861.50
206500	7568250	839.70
206750	7568250	852.70
207000	7568250	865.00
207250	7568250	872.70
207500	7568250	875.90
207750	7568250	868.00
208000	7568250	850.00
204000	7568500	878.00

l		
204250	7568500	877.00
204500	7568500	874.00
204750	7568500	872.00
205000	7568500	872.50
205250	7568500	860.00
205500	7568500	850.00
205750	7568500	861.00
206000	7568500	863.00
206250	7568500	860.50
206500	7568500	844.50
206750	7568500	837.00
207000	7568500	856.50
207250	7568500	869.00
207500	7568500	873.30
207750	7568500	872.00
208000	7568500	865.00
208250	7568500	855.50
208500	7568500	832.00
204000	7568750	871.50
204250	7568750	869.50
204500	7568750	864.00
204750	7568750	858.00
205000	7568750	869.50
205250	7568750	860.50
205500	7568750	842.00
205750	7568750	851.50
206000	7568750	853.50
206250	7568750	853.00
206500	7568750	838.80
206750	7568750	830.00
207000	7568750	845.50
207250	7568750	860.30
207500	7568750	869.00
207750	7568750	868.00
208000	7568750	863.00
208250	7568750	860.00
208500	7568750	845.00
204000	7569000	860.30
204250	7569000	859.00
204500	7569000	855.00
204750	7569000	853.50
205000	7569000	867.00
205250	7569000	862.00
205500	7569000	852.00
205750	7569000	828.50
206000	7569000	834.50
206250	7569000	834.00
206500	7569000	832.50
206750	7569000	808.00
207000	7569000	835.00
207250 207500	7569000 7569000	851.50 859.00
207300	7 308000	003.00

207750	7560000	055 50
207750	7569000	855.50
208000	7569000	851.70
208250	7569000	851.00
208500	7569000	843.50
208750	7569000	828.50
204000	7569250	842.00
204250	7569250	845.00
204500	7569250	834.00
204750	7569250	856.50
205000	7569250	866.50
205250	7569250	867.00
205500	7569250	858.00
205750	7569250	844.20
206000	7569250	836.50
206250	7569250	820.00
206500	7569250	807.50
206750	7569250	803.50
207000	7569250	826.50
207250	7569250	840.00
207500	7569250	845.00
207750	7569250	837.00
208000	7569250	829.00
208250	7569250	832.00
208500	7569250	832.50
208750	7569250	828.50
209000	7569250	810.00
209250	7569250	757.50
209500	7569250	725.00
204000	7569500	848.80
204250	7569500	832.00
204500	7569500	842.50
204750	7569500	857.50
205000	7569500	867.00
205250	7569500	870.80
205500	7569500	863.50
205750	7569500	851.50
206000	7569500	846.00
206250	7569500	834.50
206500	7569500	810.00
206750	7569500	800.00
207000	7569500	805.00
207250	7569500	825.00
207500	7569500	824.00
207750	7569500	817.00
208000	7569500	790.00
208250	7569500	805.00
208500	7569500	792.00
208750	7569500	827.00
209000	7569500	803.80
209250	7569500	787.00
209500	7569500	729.50
204000	7569750	847.00

204250	7560750	922.00
	7569750	832.00 840.50
204500	7569750	
204750	7569750	856.30
205000	7569750	865.00
205250	7569750	870.30
205500	7569750	866.50
205750	7569750	859.00
206000	7569750	851.20
206250	7569750	834.50
206500	7569750	809.00
206750	7569750	811.50
207000	7569750	818.00
207250	7569750	799.00
207500	7569750	835.00
207750	7569750	830.00
208000	7569750	749.00
208250	7569750	762.50
208500	7569750	820.00
208750	7569750	821.50
209000	7569750	805.00
209250	7569750	771.00
209500	7569750	743.00
204000	7570000	843.00
204250	7570000	825.00
204500	7570000	836.00
204750	7570000	852.20
205000	7570000	860.00
205250	7570000	864.00
205500	7570000	864.00
205750	7570000	862.30
206000	7570000	855.30
206250	7570000	862.50
206500	7570000	813.50
206750	7570000	787.50
207000	7570000	748.00
207250	7570000	732.00
207500	7570000	723.00
207750	7570000	728.50
208000	7570000	728.50
208250	7570000	735.00
208500	7570000	767.50
208750	7570000	781.00
209000	7570000	810.00
209250	7570000	762.30
209500	7570000	746.00
209750	7570000	716.00
210000	7570000	696.00
210250	7570000	693.00
210500	7570000	710.00
204750	7570250	852.00
205000	7570250	851.00
205250	7570250	854.00
200200	1010200	007.00

205500	7570050	056.00
205500	7570250	856.80
205750	7570250	861.00
206000	7570250	856.00
206250	7570250	838.20
206500	7570250	828.00
206750	7570250	808.00
207000	7570250	795.00
207250	7570250	779.00
207500	7570250	762.50
207750	7570250	742.00
208000	7570250	710.00
208250	7570250	728.30
208500	7570250	737.50
208750	7570250	767.50
209000	7570250	787.50
209250	7570250	776.50
209500	7570250	742.50
209750	7570250	711.00
210000	7570250	693.50
210250	7570250	693.00
210500	7570250	717.50
205000	7570500	832.70
205250	7570500	832.70
205500	7570500	850.00
205750	7570500	856.50
206000	7570500	855.30
206250	7570500	848.00
206500	7570500	840.00
206750	7570500	824.80
207000	7570500	821.00
207250	7570500	827.50
207500	7570500	826.00
207750	7570500	804.00
208000	7570500	727.50
208250	7570500	708.00
208500	7570500	726.00
208750	7570500	749.30
209000	7570500	762.00
209250	7570500	758.00
209500	7570500	736.50
209750	7570500	713.00
210000	7570500	694.50
210250	7570500	693.50
210500	7570500	715.00
205000	7570750	816.50
205250	7570750	825.00
205500	7570750	848.50
205750	7570750	856.70
206000	7570750	861.00
206250	7570750	856.80
206500	7570750	850.00
206750	7570750	842.50

00=000		004.50
207000	7570750	831.50
207250	7570750	830.00
207500	7570750	826.30
207750	7570750	770.00
208000	7570750	753.00
208250	7570750	700.00
208500	7570750	705.00
208750	7570750	728.70
209000	7570750	742.00
209250	7570750	738.00
209500	7570750	724.80
209750	7570750	708.50
210000	7570750	694.30
210250	7570750	694.00
210500	7570750	715.00
205000	7571000	810.00
205250	7571000	837.00
205500	7571000	850.60
205750	7571000	862.00
206000	7571000	864.00
206250	7571000	861.50
206500	7571000	857.00
206750	7571000	847.00
207000	7571000	835.00
207250	7571000	834.50
207500	7571000	829.50
207750	7571000	825.00
208000	7571000	751.00
208250	7571000	737.50
208500	7571000	704.80
208750	7571000	715.00
209000	7571000	724.50
209250	7571000	718.30
209500	7571000	709.50
209750	7571000	702.50
210000	7571000	694.00
210250	7571000	693.50
210500	7571000	713.50
205000	7571250	827.00
205250	7571250	842.50
205500	7571250	855.80
205750	7571250	861.50
206000	7571250	861.40
206250	7571250	857.50
206500	7571250	852.00
206750	7571250	848.00
207000	7571250	838.00
207250	7571250	841.00
207500	7571250	836.80
207750	7571250	817.50
208000	7571250	780.00
208250	7571250	747.50

1-		
208500	7571250	712.50
208750	7571250	702.00
209000	7571250	711.30
209250	7571250	704.50
209500	7571250	698.50
209500	7571250	698.50
209650	7571250	702.00
209750	7571250	698.50
210000	7571250	692.50
210250	7571250	699.00
210500	7571250	718.00
205000	7571500	812.50
205250	7571500	840.00
205500	7571500	854.00
205750	7571500	857.30
206000	7571500	855.70
206250	7571500	848.00
206500	7571500	838.70
206750	7571500	834.00
207000	7571500	832.00
207250	7571500	835.50
207500	7571500	830.00
207500	7571500	808.60
208000	7571500	779.50
208250	7571500	752.50
208500	7571500	727.50
208750	7571500	709.30
209000	7571500	695.50
209250	7571500	698.00
209500	7571500	698.00
209630	7571500	701.00
209750	7571500	698.00
210000	7571500	691.50
210250	7571500	702.50
210500	7571500	717.50
205000	7571750	802.00
205250	7571750	825.00
205500	7571750	845.00
205750	7571750	846.50
206000	7571750	843.40
206250	7571750	840.00
206500	7571750	826.00
206750	7571750	804.00
207000	7571750	799.00
207250	7571750	810.00
207500	7571750	805.00
207750	7571750	787.50
208000	7571750	767.70
208250	7571750	749.00
208500	7571750	732.40
208750	7571750	717.00
209000	7571750	705.00

l <del>a.</del>		•
209250	7571750	695.00
209500	7571750	697.50
209750	7571750	694.00
210000	7571750	689.00
210250	7571750	703.00
210500	7571750	717.50
205000	7572000	791.30
205250	7572000	809.00
205500	7572000	818.00
205750	7572000	819.00
206000	7572000	818.50
206250	7572000	812.50
206500	7572000	802.50
206750	7572000	767.80
207000	7572000	766.50
207250	7572000	786.30
207500	7572000	780.00
207750	7572000	772.50
208000	7572000	762.50
208250	7572000	749.00
208500	7572000	736.30
208750	7572000	723.30
209000	7572000	714.00
209250	7572000	701.80
209500	7572000	692.50
209750	7572000	692.50
210000	7572000	688.00
210250	7572000	703.00
210500	7572000	713.50
205000	7572250	778.50
205250	7572250	743.50
205500	7572250	799.00
205750	7572250	799.50
206000	7572250	792.00
206250	7572250	784.00
206500	7572250	778.50
206750	7572250	765.00
207000	7572250	752.50
207250	7572250	765.00
207500	7572250	763.00
207500	7572250	758.50
208000	7572250	758.50
208250	7572250	753.50
208230	7572250	742.50
208300	7572250	732.00
209000	7572250	718.50
209000	7572250	707.80
209230	7572250	694.00
209300	7572250	688.00
210000	7572250	688.50
210250	7572250	700.00
210230	7572250	705.00
210000	1012200	1 700.00

l-		
205000	7572500	768.00
205250	7572500	780.00
205500	7572500	782.50
205750	7572500	777.00
206000	7572500	768.00
206250	7572500	762.20
206500	7572500	755.70
206750	7572500	753.00
207000	7572500	738.80
207250	7572500	747.50
207500	7572500	745.30
207750	7572500	745.00
208000	7572500	750.00
208250	7572500	750.80
208500	7572500	744.30
208750	7572500	734.30
209000	7572500	723.00
209250	7572500	710.00
209500	7572500	696.00
209750	7572500	689.00
210000	7572500	689.00
210250	7572500	697.00
210230	7572500	707.30
205000	7572750	758.30
205250	7572750	768.50
205250	7572750	767.00
205750	7572750	759.50
206000	7572750	750.00
206250	7572750	743.50
206500	7572750	737.20
206750		737.20
	7572750	
207000	7572750	730.00
207250 207500	7572750	738.00
	7572750	736.50
207750	7572750	737.70
208000	7572750	739.50
208250	7572750	745.50
208500	7572750	740.00
208750	7572750	732.00
209000	7572750	722.00
209250	7572750	710.00
209500	7572750	695.00
209750	7572750	687.50
210000	7572750	689.00
210250	7572750	698.50
210500	7572750	702.00
205000	7573000	753.00
205250	7573000	757.30
205500	7573000	756.00
205750	7573000	746.00
206000	7573000	737.00
206250	7573000	726.30

206500	7573000	722.50
206750	7573000	719.00
207000	7573000	721.00
207250	7573000	721.30
207500	7573000	719.70
207750	7573000	722.00
208000	7573000	727.60
208250	7573000	736.50
208500	7573000	736.30
208750	7573000	724.00
209000	7573000	717.50
209250	7573000	706.50
209500	7573000	688.00
209750	7573000	695.00
210000	7573000	693.50
210250	7573000	700.00
210500	7573000	704.30
205000	7573250	749.00
205250	7573250	749.30
205500	7573250	745.00
205750	7573250	735.00
206000	7573250	727.30
206250	7573250	722.30
206500	7573250	716.40
206750	7573250	713.20
207000	7573250	713.50
207250	7573250	713.00
207500	7573250	709.00
207750	7573250	712.50
208000	7573250	715.00
208250	7573250	724.00
208500	7573250	726.50
208750	7573250	719.50
209000	7573250	709.50
209250	7573250	690.00
209500	7573250	694.00
209750	7573250	707.00
210000	7573250	715.00
210250	7573250	709.50
210500	7573250	708.20
205000	7573500	755.00
205250	7573500	750.00
205500	7573500	742.50
205750	7573500	737.50
206000	7573500	730.00
206250	7573500	723.80
206500	7573500	719.10
206750	7573500	711.50
207000	7573500	703.70
207250	7573500	703.50
207500	7573500	703.00
207750	7573500	706.90

## Conclusão

208000	7573500	707.50
208250	7573500	708.50
208500	7573500	712.30
208750	7573500	711.20
209000	7573500	700.00
209250	7573500	700.00
209500	7573500	713.00
205000	7573650	758.00
205250	7573655	752.40
205500	7573660	746.80
205750	7573665	741.50
206000	7573670	733.00
206250	7573675	731.00
206500	7573680	726.00

206750	7573685	718.00
207000	7573690	712.00
207250	7573695	702.30
207500	7573700	699.00
207750	7573705	703.00
208000	7573710	703.50
208250	7573715	703.00
208500	7573720	703.50
208750	7573725	706.50
209000	7573730	702.50
209250	7573735	692.00
209500	7573740	715.00

## A.3 DADOS AMOSTRAIS DE CLASSES DE TEXTURA DO SOLO

X(m)	Y(m)	Classe
204834.8	7567954.	3
205937.9	7570221.	2
205869.2	7571042.	2
206180.3	7567671.	3
206079.3	7567323.	3
205077.2	7569672.	3
205675.2	7567465.	3
206455.1	7566798.	3
204705.5	7565905.	3
204855.0	7566228.	3
204992.3	7567744.	2
207162.2	7568091.	2
207990.6	7568790.	2
205174.2	7566770.	2
205820.7	7569813.	2
205404.5	7568439.	2
205428.8	7567970.	2
205606.5	7567125.	2
205129.8	7565994.	2
205242.9	7565872.	2
207433.0	7570823.	3
207267.3	7570549.	3
208277.5	7571422.	3
208196.7	7571922.	2
208190.7	7571002.	2
209590.7	7570686.	2
209390.7	7570000.	2
208342.1	7570444. 7572165.	2
205411.2	7572105. 7573066.	2
208572.5	7573000. 7571797.	2
205598.2	7573741.	2
		2
208309.8	7571749.	3
206911.7	7568354.	
207206.7	7568706.	3
206798.6	7571268.	
206822.8	7570985.	3
206059.1	7569833.	3
208790.7	7570605.	3
207293.8	7572436.	2
206740.2	7573938.	2
206754.1	7572295.	2
207749.6	7572707.	2
208717.9	7573822.	2
208851.3	7571122.	2
208055.2	7569967.	3
207766.6	7571553.	4
208709.8	7569566.	4
207121.8	7568916.	3

206834.9	7568475.	3
208321.9	7568968.	4
208164.3	7569219.	3
205230.8	7572371.	3
207841.1	7571846.	3
205888.7	7573955.	3
206936.0	7571333.	3
206713.7	7571850.	4
207930.0	7569696.	4
208111.8	7569579.	3
208956.3	7570237.	4
206455.1	7569659.	3
206616.7	7569174.	3
206758.2	7568920.	3
207740.1	7569562.	3
205634.8	7566616.	2
205541.9	7566406.	2
208693.5	7573236.	1
209449.3	7571696.	1
205558.1	7573437.	1
206433.0	7573049.	1
207097.5	7573363.	1
207745.5	7573778.	1
206205.9	7573911.	1
207160.9	7573915.	1
206563.2	7573430.	1
207191.0	7573092.	1
209314.7	7572541.	1
209387.3	7572788.	1
208576.5	7572869.	1
209113.9	7571506.	1
209360.4	7571785.	1
208972.5	7571571.	1
209449.3	7572020.	1
209994.8	7570880.	3
208119.9	7569417.	3
·		

Palavras Chaves

Sistemas de Informação Geográfica

Geoestatística

Krigeagem

Simulação estocástica

Modelagem de incertezas