

XIX SBSR

Simpósio Brasileiro de Sensoriamento Remoto



14 a 17 de abril de 2019

Santos, SP Mendes Convention Center

Mini Curso 06: Geoestatística e Aplicações em Geoprocessamento

Carlos Alberto Felgueiras
Jussara de Oliveira Ortiz
Eduardo Celso Gerbi Camargo



MINISTÉRIO DA
CIÊNCIA, TECNOLOGIA,
INOVAÇÕES E COMUNICAÇÕES



PÁTRIA AMADA
BRASIL
GOVERNO FEDERAL

Geoestatística não Linear: por Indicação ("Indicator Kriging and Indicator Simulation")

Conteúdo

Introdução

Definições Iniciais da Geoestatística

Estimador de Krigagem

Variâncias de Krigagem

A Abordagem Multigaussiana x Indicadora (Paramétrica)

Modelagem por Indicação para Variáveis Contínuas e Categóricas

Transformações por indicação e propriedades

Avaliação dos modelos de incerteza

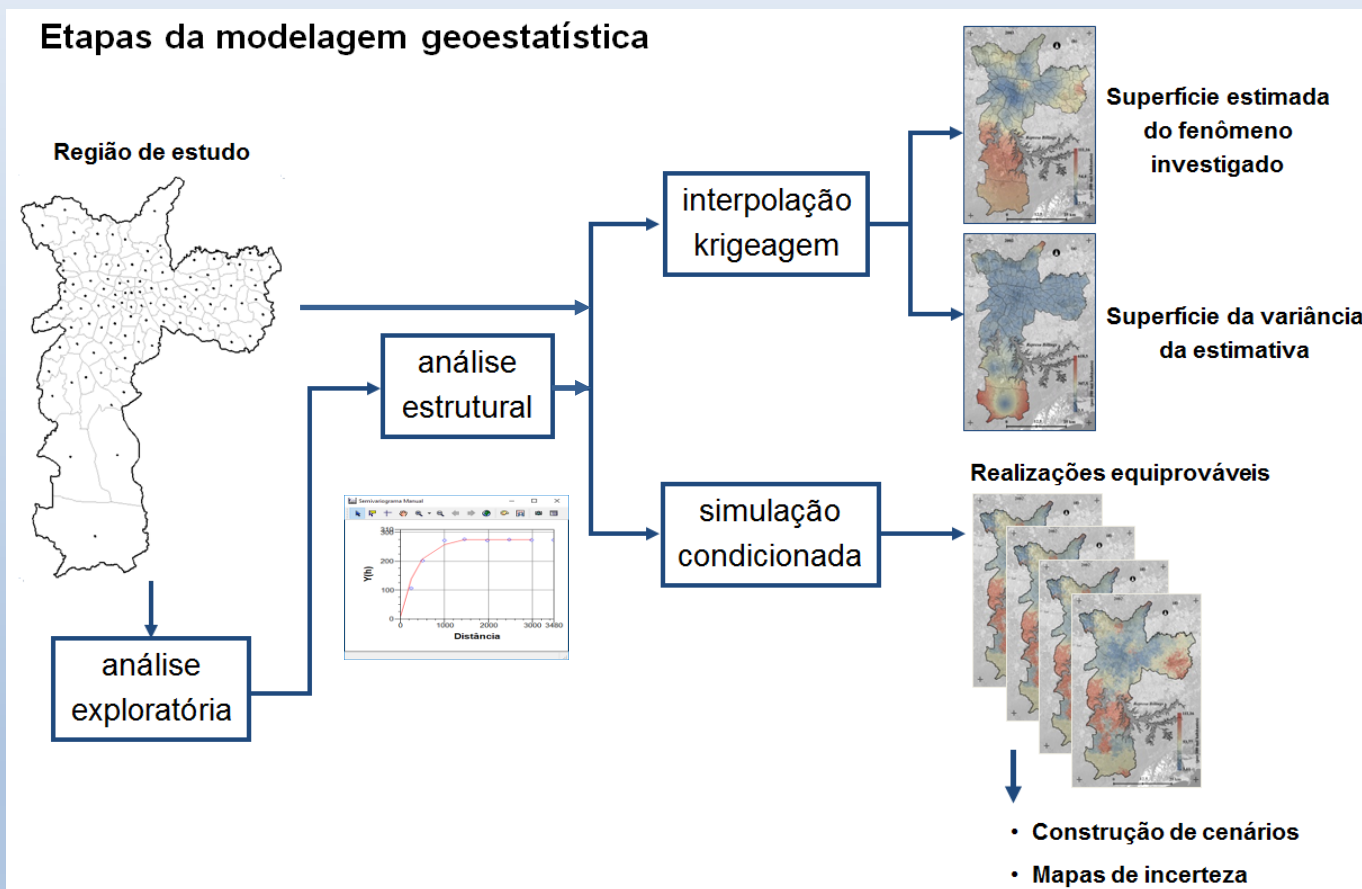
Correção dos desvios de relação de ordem

Estimativa dos parâmetros da VA (médias, variâncias, quantis,...)

Estimativa de Incertezas por Intervalos de confiança

Introdução as copulas espaciais

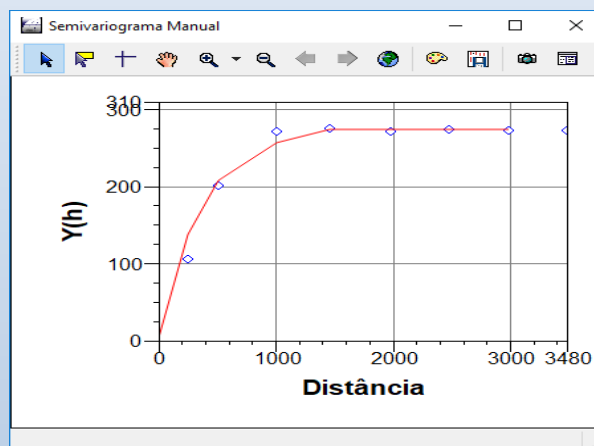
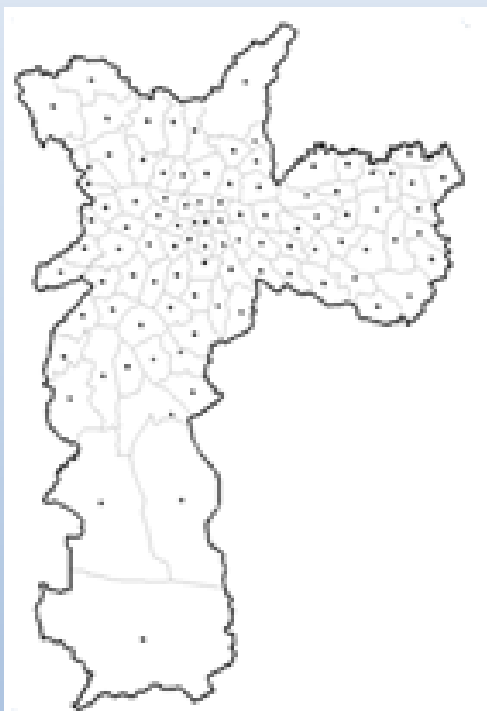
INTRODUÇÃO: Definições Iniciais da Geoestatística



INTRODUÇÃO: Variografia – Análise estrutural de dependência espacial

$$\gamma(\mathbf{h}) = \frac{1}{2} E\{[z(\mathbf{u}) - z(\mathbf{u} + \mathbf{h})]^2\}$$

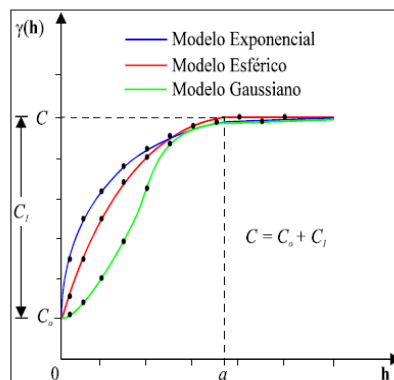
$$\hat{\gamma}(\mathbf{h}) = \frac{1}{2N(\mathbf{h})} \sum_{i=1}^{N(\mathbf{h})} [z(\mathbf{u}_i) - z(\mathbf{u}_i + \mathbf{h})]^2$$



- Semivariogramas
- ◇ experimental
 - conceitual (teórico ou matemático)

Modelos Teóricos para ajuste do Semivariograma Experimental (Eduardo C.G. Camargo)

Modelos Transitivos



EXPONENCIAL

$$\gamma(\mathbf{h}) = \begin{cases} 0 & , |\mathbf{h}| = 0 \\ C_0 + C_I \left[1 - \exp\left(-3 \frac{|\mathbf{h}|}{a}\right) \right] & , |\mathbf{h}| \neq 0 \end{cases}$$

ESFÉRICO

$$\gamma(\mathbf{h}) = \begin{cases} 0 & , |\mathbf{h}| = 0 \\ C_0 + C_I \left[\frac{3}{2} \left(\frac{|\mathbf{h}|}{a}\right) - \frac{1}{2} \left(\frac{|\mathbf{h}|}{a}\right)^3 \right] & , 0 < |\mathbf{h}| \leq a \\ C_0 + C_I & , |\mathbf{h}| > a \end{cases}$$

GAUSSIANO

$$\gamma(\mathbf{h}) = \begin{cases} 0 & , |\mathbf{h}| = 0 \\ C_0 + C_I \left[1 - \exp\left(-\frac{|\mathbf{h}|^2}{a}\right) \right] & , |\mathbf{h}| \neq 0 \end{cases}$$

Principais conceitos teóricos

Predições: Determinística x Estocástica $z^*(\mathbf{u}) = K$

○ Amostras

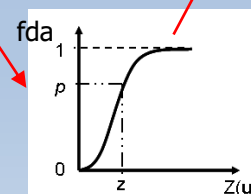
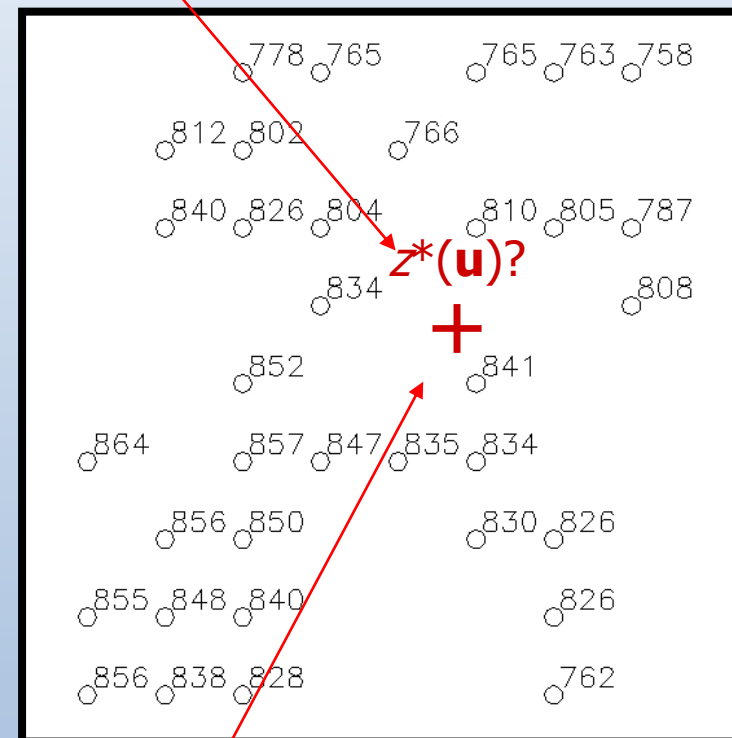
+ Estimativas

- **Determinística**

- O valor $z^*(\mathbf{u})$ é estimado como uma **Variável Determinística**. Um único valor é associado a sua localização espacial.
- As incertezas não são associadas às estimativas.

- **Estocástica**

- O valor $z^*(\mathbf{u})$ é considerado como **Variável Aleatória**, que tem a sua *função de distribuição de probabilidade* associada aos seus possíveis valores.
- As incertezas podem ser associadas às estimativas.
- Importante: as amostras são realizações das Variáveis Aleatórias.



Avaliação da Incerteza Local

- **Introdução (de Deutsch and Journel, 1998)**

- **A Krigagem (por exemplo, krigagem simples)** possibilita estimativas de valores de um atributo espacial considerando o seguinte estimador linear:

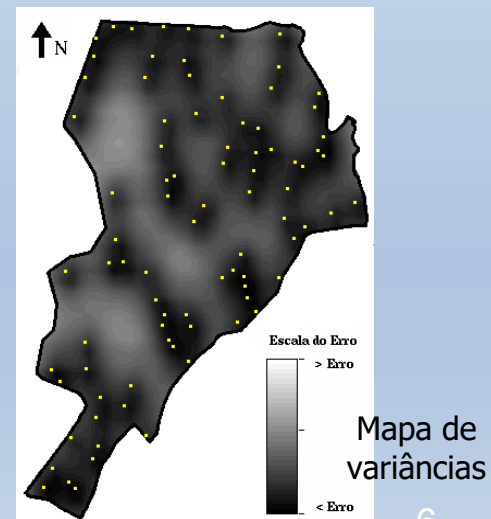
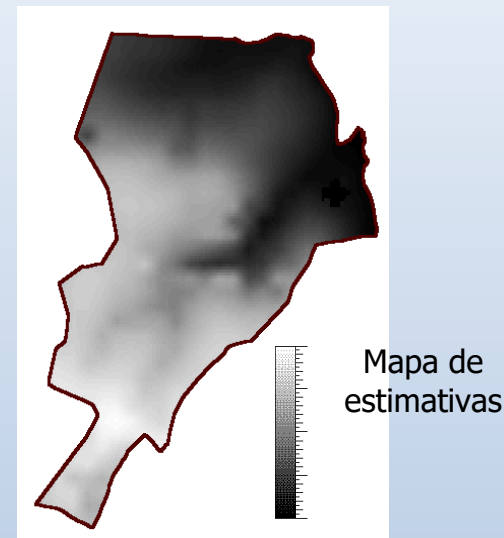
$$Z^*(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}(\mathbf{u}) \cdot Z(\mathbf{u}_{\alpha}) + \left[1 - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}(\mathbf{u}) \right] \cdot m$$

- Os **pesos** λ_{α} são avaliados para minimizar o erro da variância, também chamado "variância de estimativa". Essa minimização resulta em um conjunto normal de equações do tipo:

$$\sum_{\beta=1}^{n(\mathbf{u})} \lambda_{\beta} C(\mathbf{u}_{\alpha} - \mathbf{u}_{\beta}) + \mu = C(\mathbf{u}_{\alpha} - \mathbf{u}_u), \quad \alpha = 1, \dots, n(\mathbf{u})$$

- A **variância de estimativa minimizada**, ou variância de krigagem, é dada por:

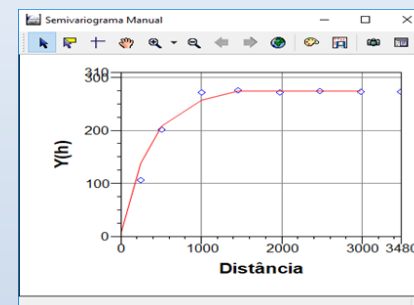
$$\sigma_{SK}^2(\mathbf{u}) = C(0) - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}^{SK}(\mathbf{u}) \cdot C(\mathbf{u}_{\alpha} - \mathbf{u})$$



Krigeagem Ordinária: Cálculo dos pesos das amostras

Segundo *Journal* (1988): $\mathbf{K}\lambda = \mathbf{k} \Rightarrow \lambda = \mathbf{K}^{-1}\mathbf{k}$

$$\begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \alpha \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} & C_{13} & C_{14} & 1 \\ C_{21} & C_{22} & C_{23} & C_{24} & 1 \\ C_{31} & C_{32} & C_{33} & C_{34} & 1 \\ C_{41} & C_{42} & C_{43} & C_{44} & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}^{-1} \begin{bmatrix} C_{01} \\ C_{02} \\ C_{03} \\ C_{04} \\ 1 \end{bmatrix}$$



Os elementos das matrizes de covariâncias são calculados da seguinte forma (*Journal*, 1988):

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

Substituindo os valores de C_{ij} nas matrizes encontram-se os pesos $\lambda_1, \lambda_2, \dots, \text{e } \lambda_n$.

Estimador de Krigeagem (*Journal*, 1988): $Z_{x_0}^* = \sum_{i=1}^n \lambda_i Z(x_i)$

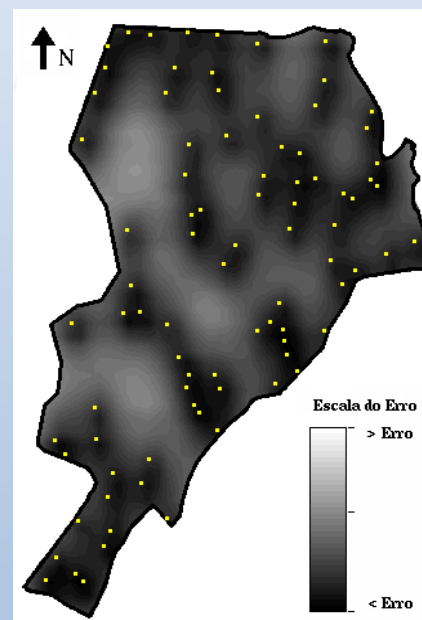
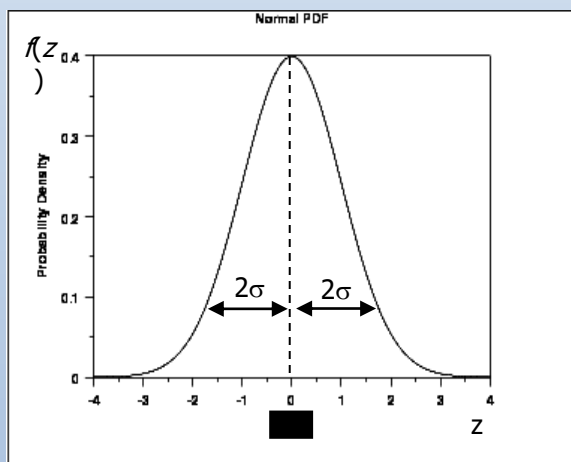
Variância de Krigeagem (*Journal*, 1988): $\sigma_{k_0}^2 = (C_0 + C_1) - \lambda^T \mathbf{k}$

Avaliação da Incerteza Local

• Introdução (de Deutsch and Journel, 1998)

- O mapa de variância de krigagem mostra que seus valores são proporcionais as distâncias das localizações \mathbf{u} das amostras:

$$\sigma_{SK}^2(\mathbf{u}) = C(0) - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}^{SK}(\mathbf{u}) \cdot C(\mathbf{u}_{\alpha} - \mathbf{u})$$



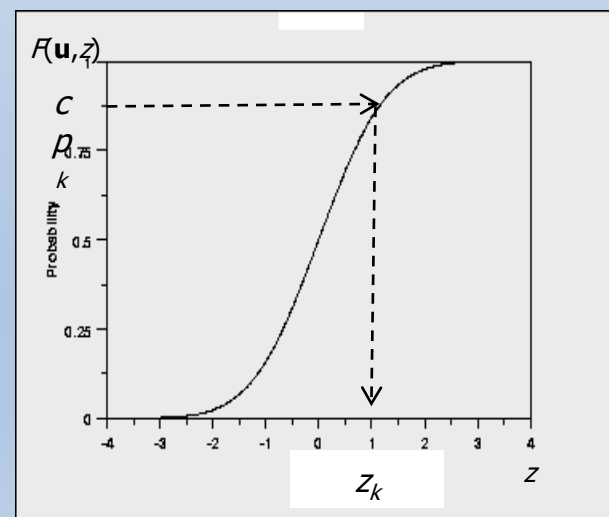
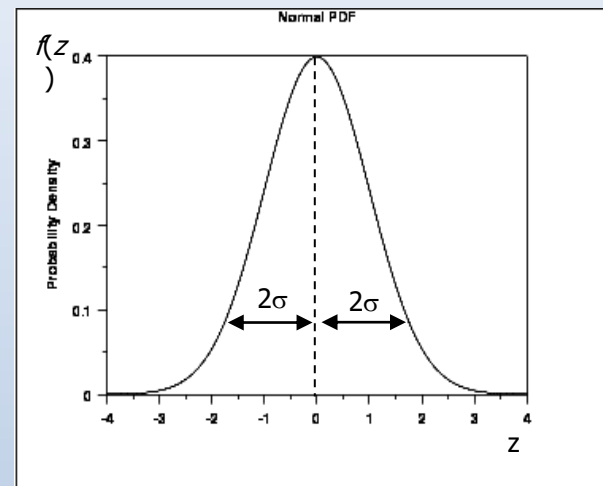
- **Importante:** A variância de krigagem, sendo independente dos valores dos dados, somente fornece uma comparação geométrica alternativa das configurações das amostras. As variâncias de krigagem não são, usualmente, medidas de acurácias das estimativas locais.

Modelagem da Incerteza Local

• Introdução (de Deutsch and Journel, 1998)

- **A Abordagem MultiGaussiana (MG):** Se o modelo da FA de $Z(\mathbf{u})$ é Gaussiano Multivariado, então as estimativas e variâncias da krigeagem linear identificam a média e variância de suas fda posteriores. Além disso, desde que a fda é Gaussiana, ela é total/e determinada por esses 2 parâmetros (**Abordagem Paramétrica**). Como verificar que FA é multigaussiana?

- **A abordagem da Krigeagem por Indicação (IK):** Se o interesse é só por avaliar valores de média e variância, os procedimentos da geoestatística linear são suficientes. Caso contrário, a **Krigeagem por Indicação "Indicator kriging"** fornece métodos não lineares para a estimativa aproximada de uma fdac , $F(\mathbf{u}, z)$, que modela a incerteza sobre $z(\mathbf{u})$. **Abordagem, ou procedimento, não paramétrica.**



Modelagem por Geoestatística por Indicação

Parte I

Krigeagem por Indicação para variáveis contínuas

Procedimento por Indicação

- Variáveis contínuas
- Transformação por Indicação e propriedades

Ao invés da VA $Z(\mathbf{u})$, considere sua transformação binária por indicação $I(\mathbf{u}; z_k)$ definida como:

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

A Krigeagem da VA por Indicação RV $I(\mathbf{u}; z)$ fornece uma estimativa que é também a melhor estimativa não tendenciosa e de mínimos quadrados (LS) da expectativa condicional de $I(\mathbf{u}; z)$. Essa estimativa condicional de $I(\mathbf{u}; z)$ é igual à fdac local de $Z(\mathbf{u})$, ou seja:

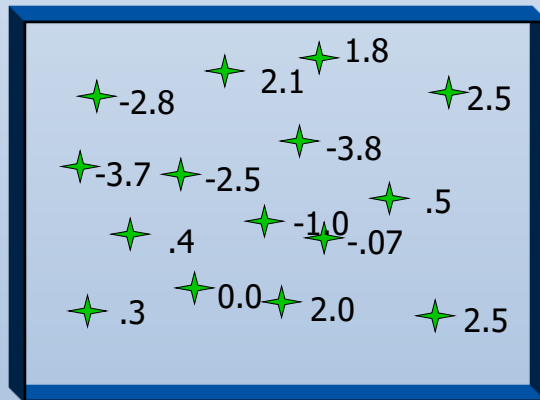
$$\begin{aligned} E\{I(\mathbf{u}; z_k) | (n)\} &= 1 \cdot \text{Prob}\{I(\mathbf{u}; z_k) = 1 | (n)\} + 0 \cdot \text{Prob}\{I(\mathbf{u}; z_k) = 0 | (n)\} \\ &= 1 \cdot \text{Prob}\{I(\mathbf{u}; z_k) = 1 | (n)\} = \text{Prob}\{Z(\mathbf{u}) \leq z_k | (n)\} = F(\mathbf{u}; z_k | (n)) \end{aligned}$$

Importante: Assim, a Krigeagem por Indicação não objetiva estimar valores de média e variância de $z(\mathbf{u})$, mas fornecer um modelo de incerteza, a fdac, sobre $z(\mathbf{u})$.

Procedimento por Indicação

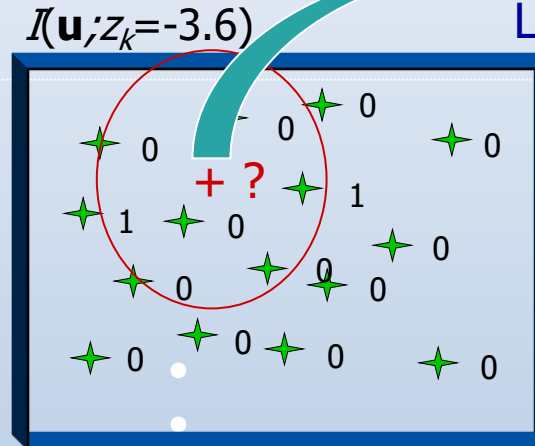
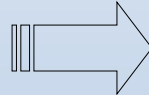
- Variáveis Contínuas
- Estimativa do Modelo de Incerteza

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



$$z_{\min} = -4 \text{ e } z_{\max} = 3$$

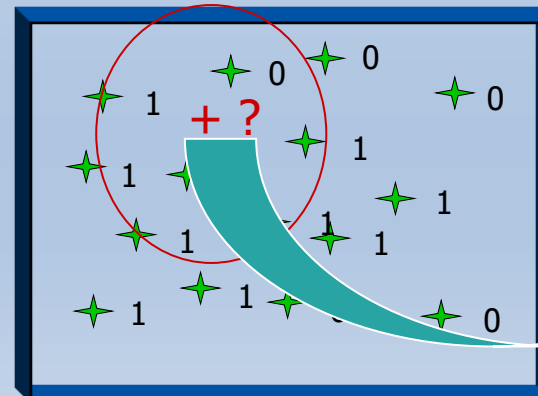
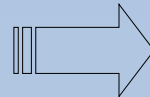
$$z_k = -3.6$$



-
-
-

$$I(\mathbf{u}; z_k = 1.1)$$

$$z_k = 1.1$$



Krigagem
Linear

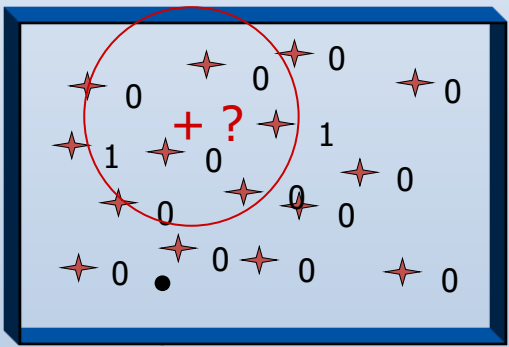
z	$fdac$
-4.0	0.0
-3.6	0.13
-2	0.375
-.55	.6
1.1	.9
3	1

Krigagem
Linear

Procedimento por Indicação

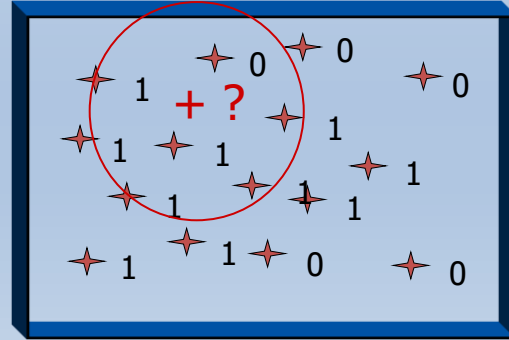
- Variáveis Contínuas
 - Avaliação do modelo de incerteza

$I(\mathbf{u}; z_k = -3.6)$

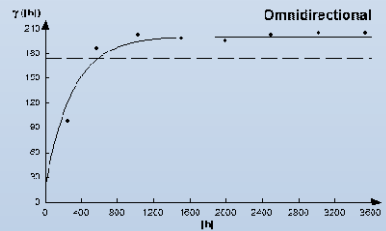


- Probabilístico (Indicador)
- Campos Aleatórios

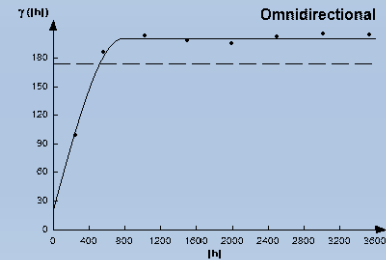
$I(\mathbf{u}; z_k = 1.1)$



Semivariogramas por Indicação: Um para cada corte

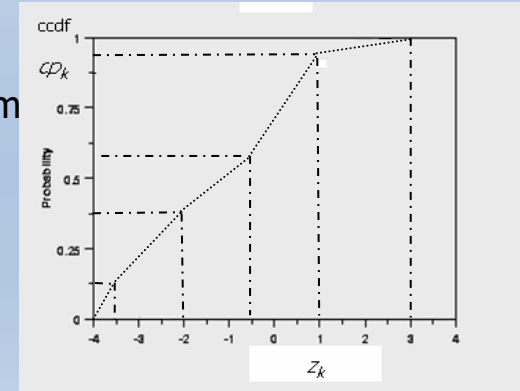


Krigeagem Linear



Krigeagem Linear

z	$fdac$
-4.0	0.0
-3.6	0.13
-2	0.375
-.55	.6
1.1	.9
3	1

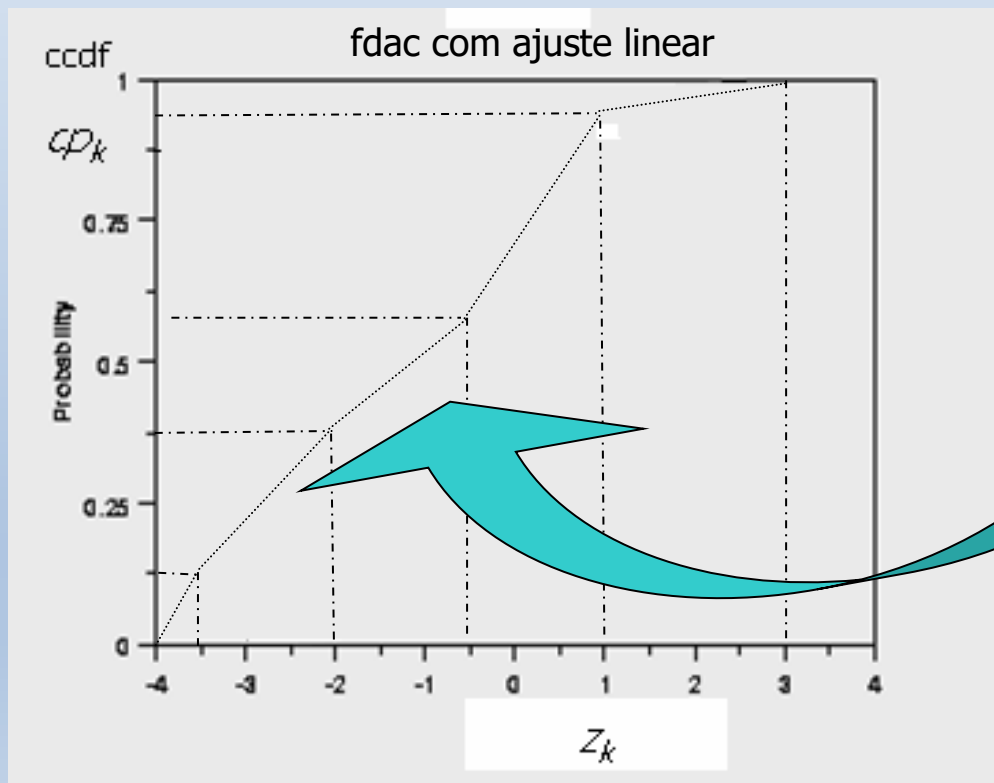


Procedimento por Indicação

- **Variáveis Contínuas**

- **Avaliação do modelo de incerteza (Ilustração)** : Usando-se K valores de cortes, ou limiares, em conjunto com os valores mínimos e máximos (neste caso $K=4$, $z_{min}=0$ and $z_{max}=3$)

z	$fdac$
0.0	0.0
-3.6	0.13
-2	0.375
-.55	.6
1	.9
3	1

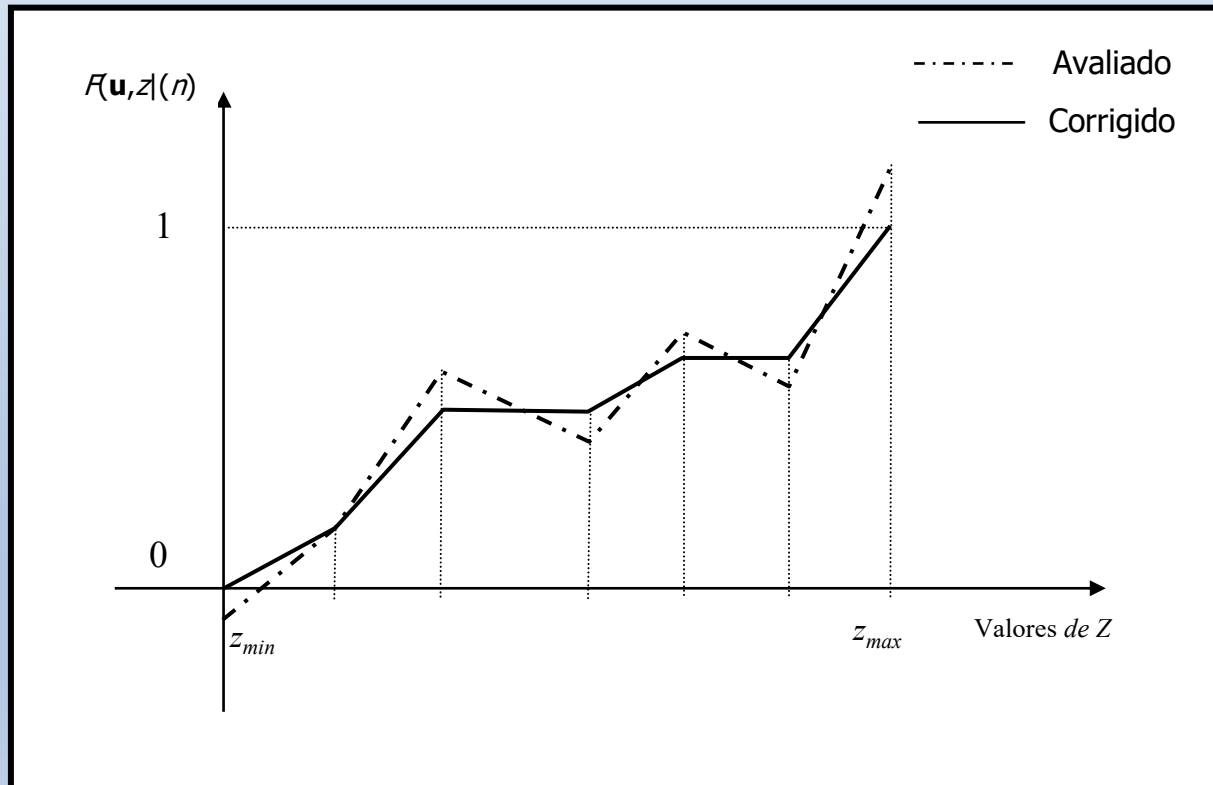


Modelo de
Incerteza de Z
(fdac ou $F(\mathbf{u}, z | (n))$)

Procedimento por Indicação

- **Variáveis Contínuas**

- **Correção de desvios de relação de ordem (Goovaerts, 1997)** – a representação da fdac final, $F(\mathbf{u}, z|(n))$, deve ser uma função não decrescente e deve ficar no intervalo $[0,1]$, tornando, assim, necessário essas correções.



Procedimento por Indicação

• Variáveis Contínuas

- Estimando-se parâmetros da VA (**média e variância**)
- O **valor médio** pode ser estimado, a partir de uma fdac discreta, por:

$$E[Z] = \int_{-\infty}^{\infty} z \cdot f(z) dz$$

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

$$\text{onde } z'_k = (z_k + z_{k-1})/2$$

- Usando-se o valor esperado acima, a **variância** pode ser estimada, similarmemente ao valor médio, por:

$$Var[Z] = E[(Z - E(Z))^2] = \int_{-\infty}^{\infty} (z - E(Z))^2 \cdot f(z) dz$$

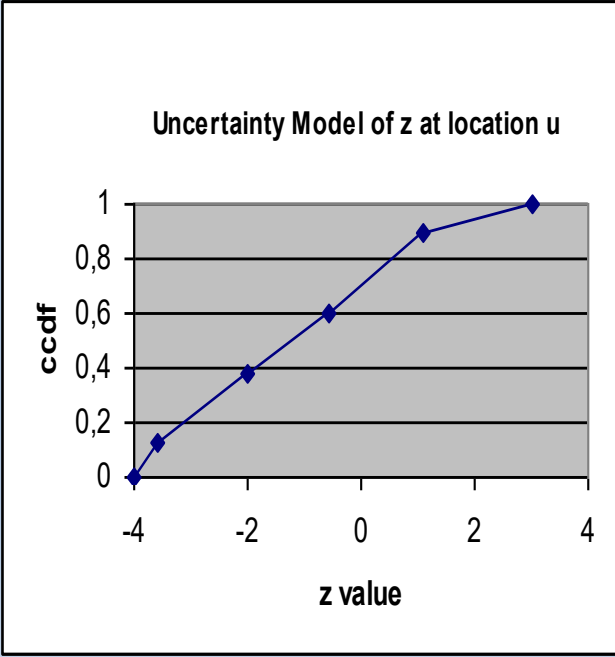
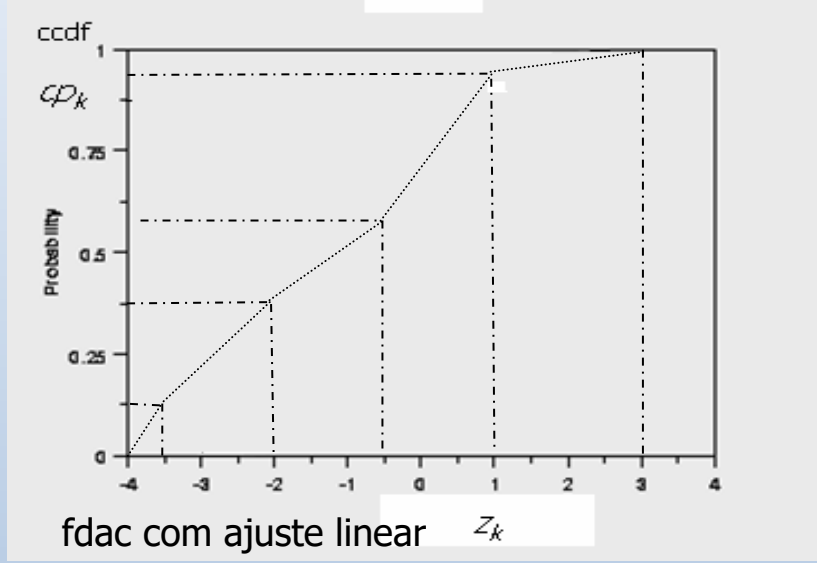
$$Var[z(\mathbf{u})]_E^* \approx \sum_{k=1}^{K+1} (z'_k - [z(\mathbf{u})]_E^*)^2 \cdot [F(\mathbf{u}; z_k | (n)) - F(\mathbf{u}; z_{k-1} | (n))]$$

Procedimento por Indicação

• Variáveis contínuas

- Exemplo de **Estimativa de parâmetros** (média e variância) **de um VA discretizada** usando Excel.

<i>z</i>	<i>fdac</i>
0.0	0.0
-3.6	0.13
-2	0.375
-.55	.6
1	.9
3	1



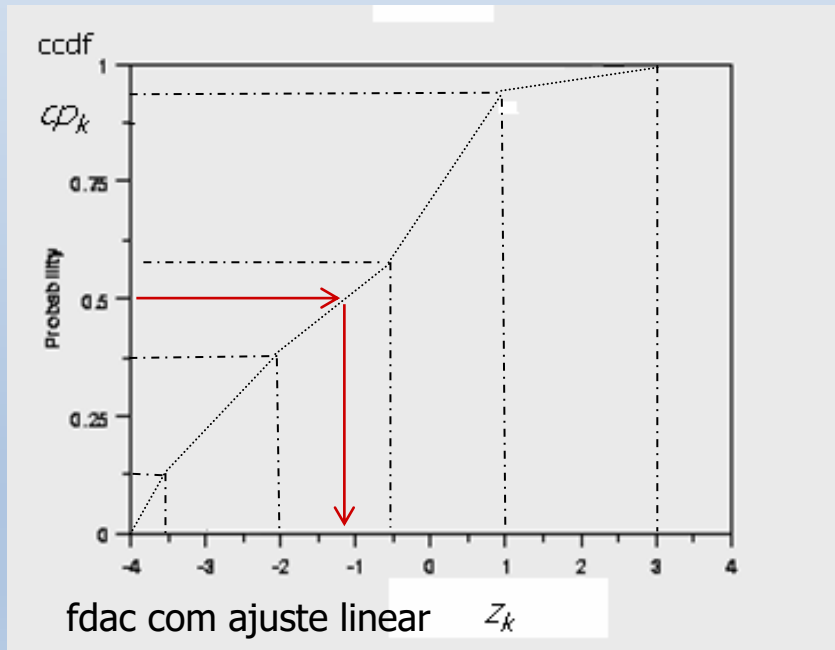
Z Value	Probability	$(z_k+z_{k-1})/2$	$F(z_k)-F(z_{k-1})$	C*D	$(C-\text{mean})^2$	F*D
-4,000	0,000	-3,800	0,130	-0,494	6,868	0,893
-3,600	0,130	-2,800	0,245	-0,686	2,626	0,643
-2,000	0,375	-1,275	0,225	-0,287	0,009	0,002
-0,550	0,600	0,275	0,300	0,083	2,115	0,635
1,100	0,900	2,050	0,100	0,205	10,429	1,043
3,000	1,000					
			Mean Value	-1,179	Variance	3,216
					Stand. Dev.	1,793

Procedimento por Indicação

• Variáveis Contínuas

- Estimando-se parâmetros da VA (**mediana e quantis**)
- O valor da **mediana**, $q_{.5}$, pode ser estimado, a partir de uma fdac discretizada, como o valor de z com probabilidade igual a 0.5.

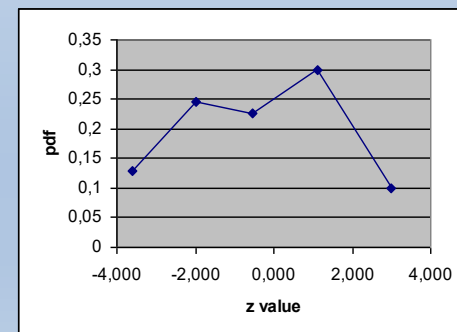
z	$fdac$
-4.	0.0
-3.6	0.13
-2	0.375
-0.55	.6
1.1	.9
3	1



$$\frac{z - (-2)}{-0.55 - (-2)} = \frac{0.5 - 0.375}{0.6 - 0.375}$$

$$z = \frac{(2 - 0.55) * 0.125}{0.225} - 2 = -1.194$$

O valor da mediana é mais robusto do que o valor da média para distribuições assimétricas

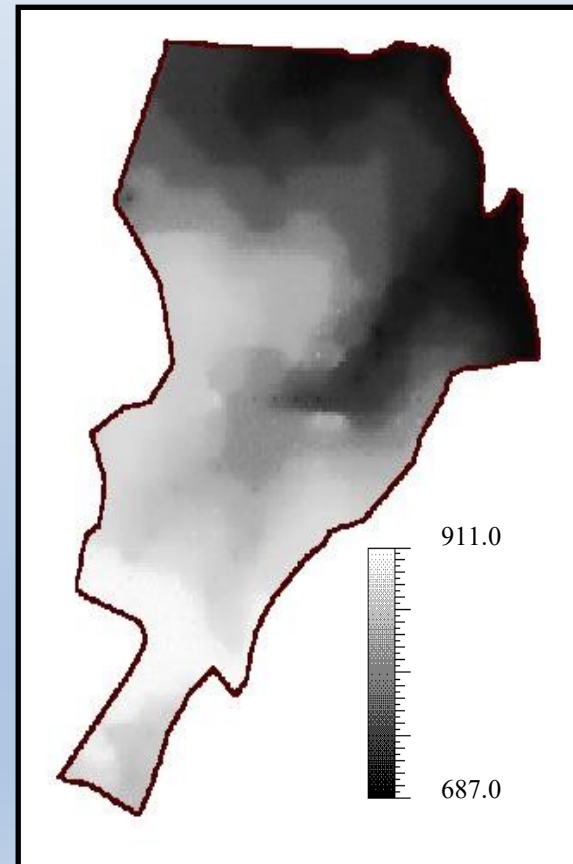
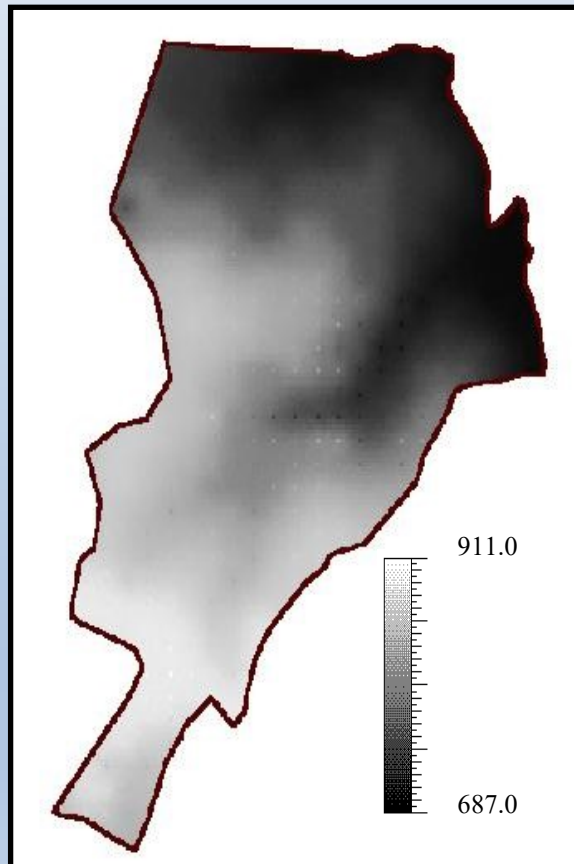
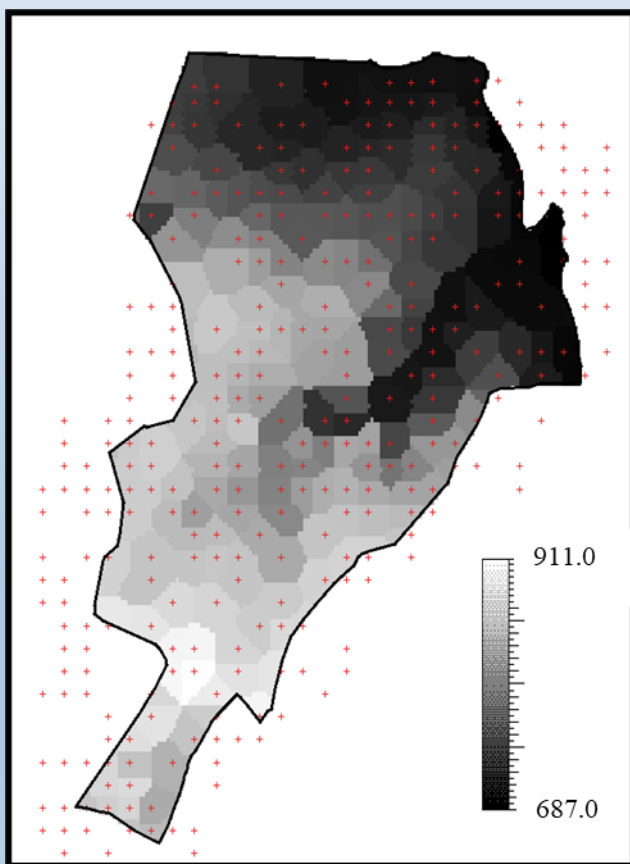


- Similarmente pode-se estimar qualquer valor de quantil q .

Procedimento por Indicação

- **Variáveis Contínuas**

- Exemplos de Mapas de Predições (Mapas de Dirichlet, de média e de mediana)



Procedimento por Indicação

- **Variáveis Contínuas**

- **Avaliação de Incertezas por Intervalos de Confiança**

- **Desvios Padrão:** desvio do valor médio. Considerando-se somente um desvio padrão define-se o valor incerteza como:

$$\text{Inc}(\mathbf{u}) = 2\sigma(\mathbf{u}) \text{ onde } \text{Prob}\{Z(\mathbf{u}) \in [\mu(\mathbf{u})_Z \pm \sigma(\mathbf{u})]\} \approx 0.68$$

- **Quantis:** separa as realizações por probabilidades. Por exemplo, o intervalo interquartil é definido como:

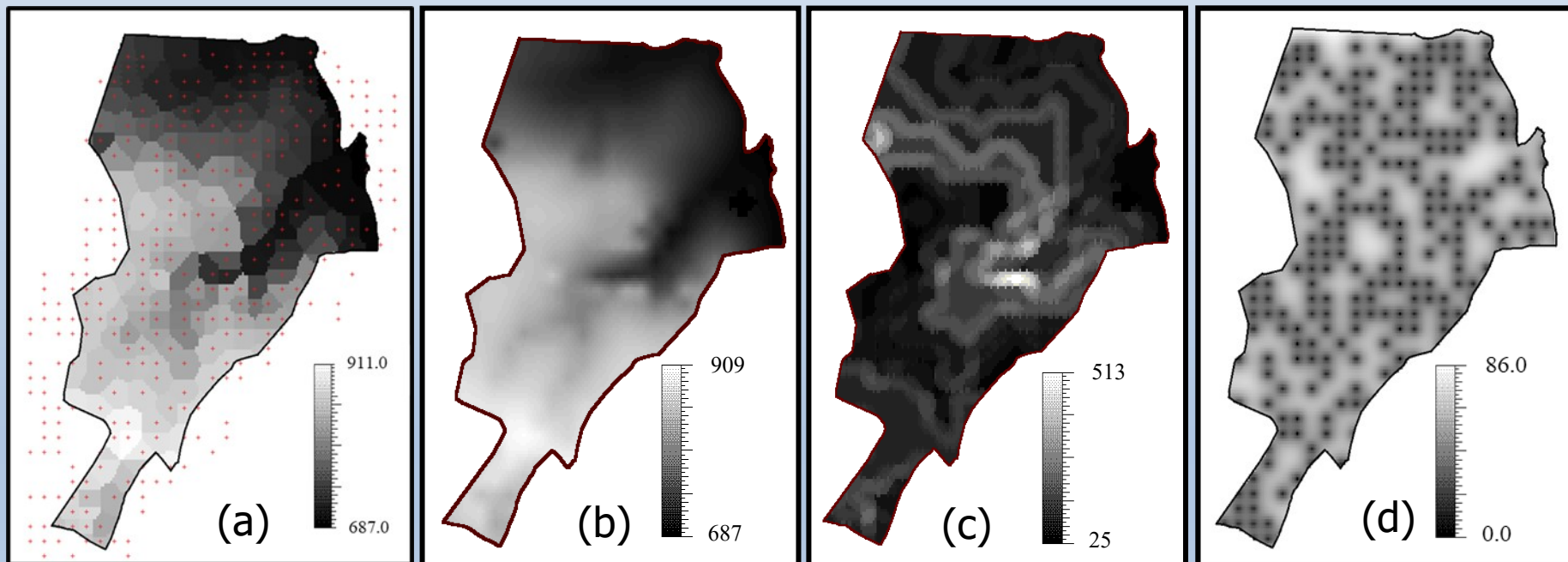
$$\text{Inc}(\mathbf{u}) = [q_{0.25}; q_{0.75}] \text{ onde } \text{Prob}\{Z(\mathbf{u}) \in [q_{0.25}; q_{0.75}] | (n)\} = 0.50$$

Procedimento por Indicação

• Variáveis Contínuas

- Exemplos de Mapas de Estimativas e de Incertezas

Mapas: (a) médias com amostras, (b) médias, (c) incertezas por 1 desvio padrão e (d) incertezas por variância de krigagem linear



Como interpretar o mapa (c)? Porque os mapas (c) e (d) são tão diferentes?

Modelagem por Geoestatística por Indicação

Parte II

Krigeagem por Indicação para variáveis categóricas (temáticas)

Modelagem por Indicação – Variáveis Categóricas

- **Introdução**

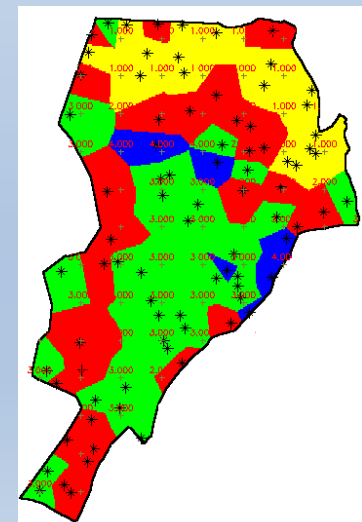
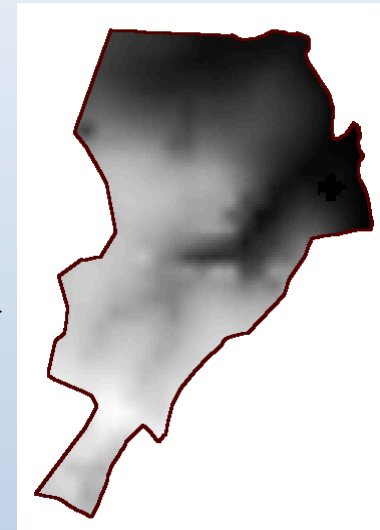
- Modelagem espacial: **Variáveis Contínuas x Variáveis Categóricas (Temáticas)**

- **A Krigagem linear (simples, ordinária,...)** é usada para estimativas de atributos contínuos pela equação:

$$Z^*(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}(\mathbf{u}) \cdot Z(\mathbf{u}_{\alpha}) + \left[1 - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}(\mathbf{u}) \right] \cdot m$$

- Como modelar variáveis categóricas? É possível utilizar a geoestatística para obter-se modelos de incerteza a partir de amostras pontuais de VA categóricas? Como?

- Solução Determinística: Estimador por vizinho mais próximo. Quais as desvantagens desse modelo? Como avaliar a qualidade dessa modelagem determinística?



Modelagem por Indicação – Variáveis Categóricas

• Introdução

• Variáveis Aleatórias Categóricas

• *Definição:* Uma VA categórica $S(\mathbf{u})$ é composta por um conjunto de valores categóricos (discretos) que estão associados a um modelo de incerteza, uma função de distribuição de probabilidade (fdp), representando a distribuição de probabilidade de seus valores.

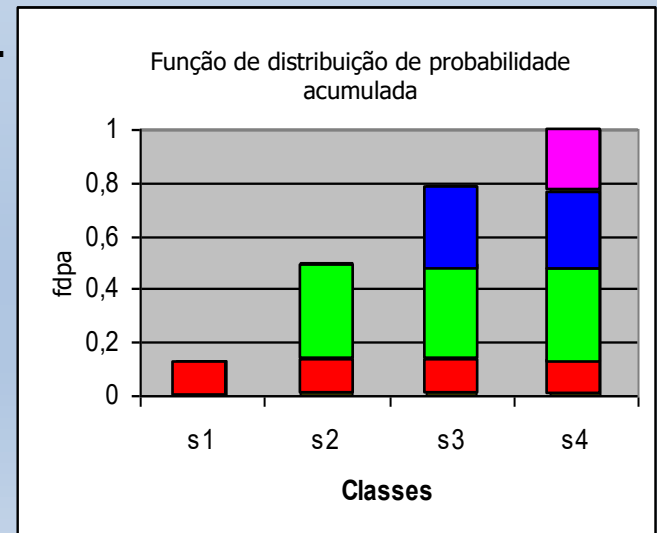
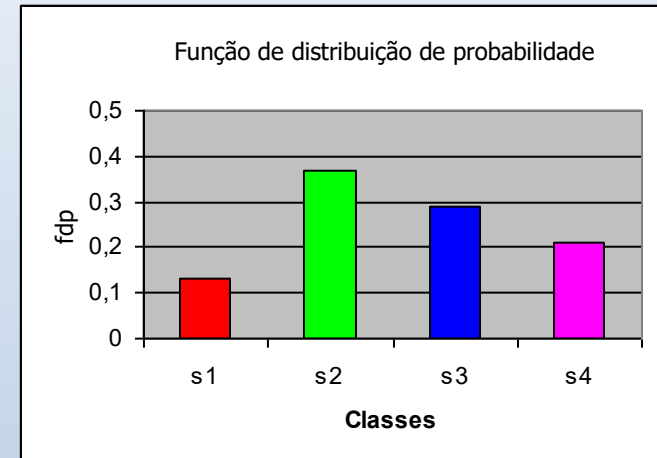
• Propriedades

• Cada um dos s_k valores, ou classes, (A, B, C, ...) está associado a uma probabilidade de ocorrência $p_k \in [0,1]$.

• A soma das probabilidades de todas as classes em uma localização espacial \mathbf{u} é igual a 1

$$\sum_{k=1}^K p_k(\mathbf{u}) = \sum_{k=1}^K p(\mathbf{u}; s_k) = 1$$

• Importante: Qualquer fdp categórica pode ser transformada em uma fdpa, acumulada, considerando-se uma ordenação a priori das classes.



Procedimento por Indicação

- **Variáveis Categóricas**
- **Transformação por Indicação e propriedades**

Ao invés da FA $S(\mathbf{u})$, considere sua transformação binária por indicação $I(\mathbf{u}; z_k)$ como:

$$I(\mathbf{u}; s_k) = \begin{cases} 1, & \text{se } S(\mathbf{u}) = s_k \\ 0, & \text{caso contrário} \end{cases}$$

A Krigeagem da VA por Indicação RV $S(\mathbf{u}; z)$ fornece uma estimativa que é também a melhor estimativa não tendenciosa e de mínimos quadrados (LS) da expectativa condicional de $I(\mathbf{u}; z)$. Essa esperança condicional de $I(\mathbf{u}; z)$ é igual à fdp local de $S(\mathbf{u})$, ou seja:

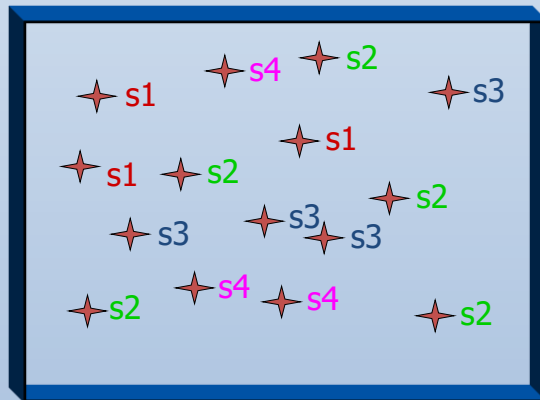
$$\begin{aligned} E\{I(\mathbf{u}; s_k) | (n)\} &= 1 \cdot \text{Prob}\{I(\mathbf{u}; s_k) = 1 | (n)\} + 0 \cdot \text{Prob}\{I(\mathbf{u}; s_k) = 0 | (n)\} \\ &= 1 \cdot \text{Prob}\{I(\mathbf{u}; s_k) = 1 | (n)\} = \text{Prob}\{S(\mathbf{u}) = s_k | (n)\} = p(\mathbf{u}; s_k | (n)) \end{aligned}$$

Importante: Assim, a Krigeagem por Indicação para variáveis categóricas não objetiva estimar classes $s(\mathbf{u})$, mas fornecer um modelo de incerteza, a fdp de $S(\mathbf{u})$ condicionada às amostras.

Procedimento por Indicação

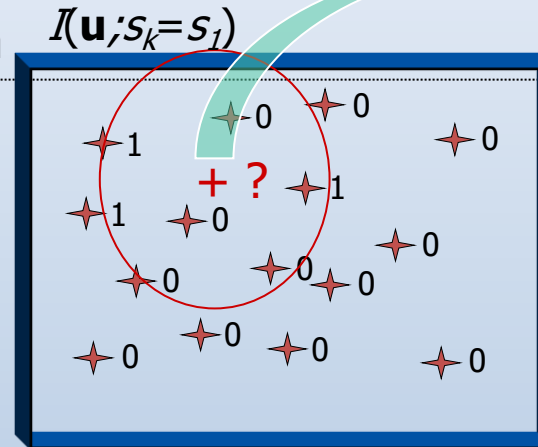
- Variáveis Categóricas
- Estimativa do Modelo de Incerteza

$$I(\mathbf{u}; s_k) = \begin{cases} 1, & \text{se } S(\mathbf{u}) = s_k \\ 0, & \text{caso contrário} \end{cases}$$



Supondo $K = 4$ (Classes s_1, s_2, s_3 e s_4)

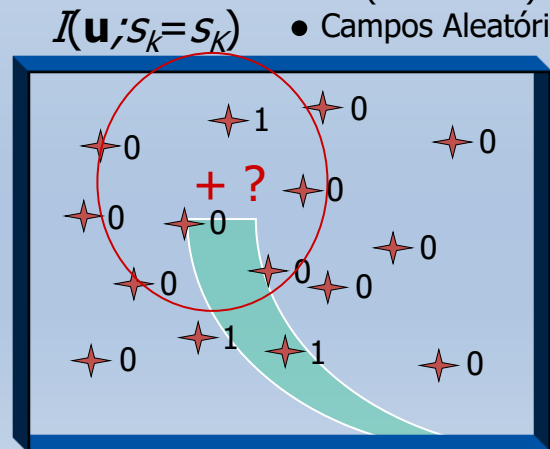
$S_k = s_1$



-
-
-

- Probabilística (Indicadora)
- Campos Aleatórios

$S_k = s_K$

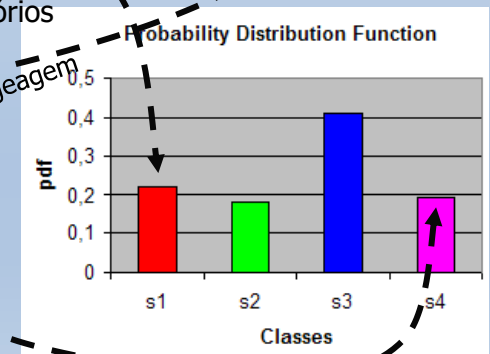
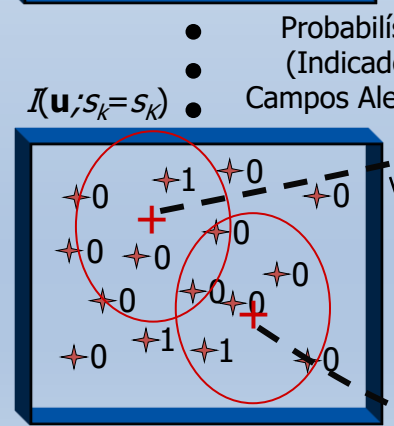
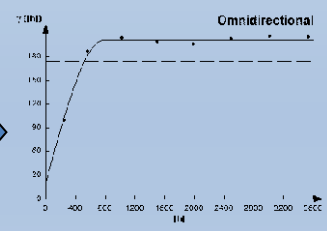
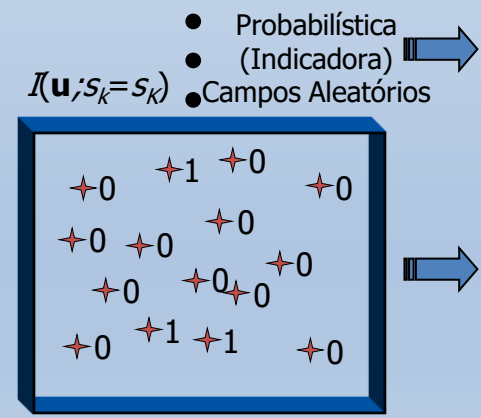
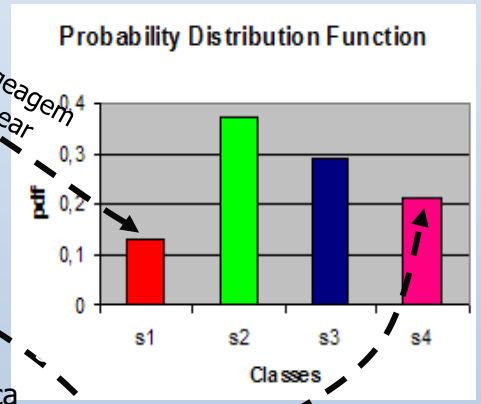
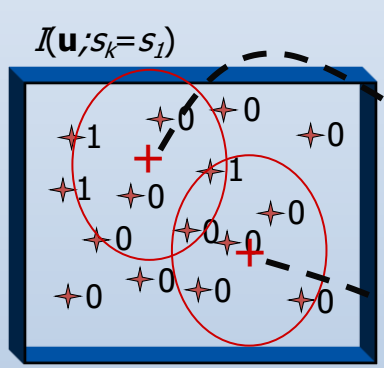
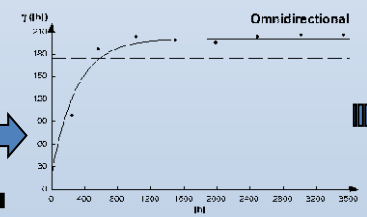
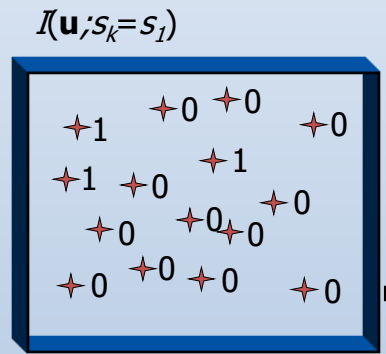


s	$fdpc$
s1	0.13
s2	0.37
s3	0.29
s4	0.21

Procedimento por Indicação

- Variáveis Categóricas
- Estimativa do Modelo de Incerteza (mais detalhes)

Variogramas por Indicação
Um para cada campo (classe)



Krigeagem linear

Krigeagem

Krigeagem Linear

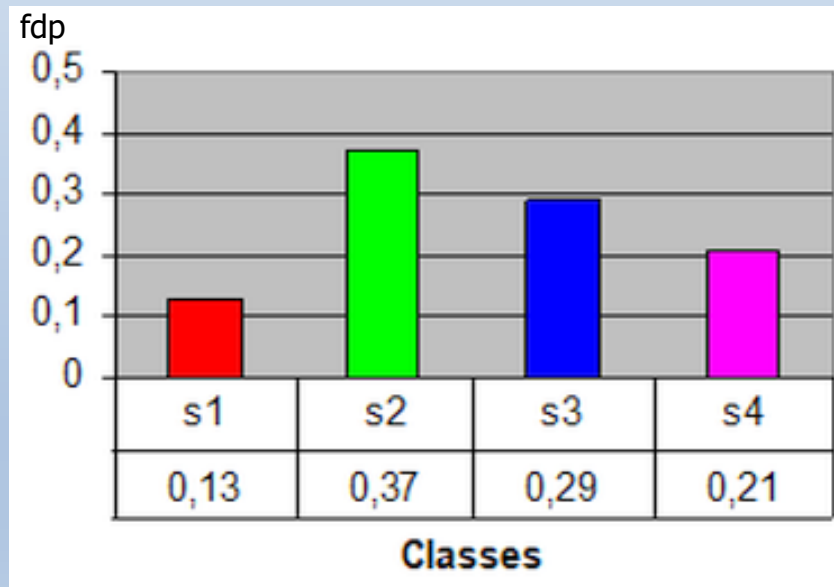
Procedimento por Indicação

• Variáveis Categóricas

- **Estimativa do Modelo de Incerteza (Ilustração)** : Usando-se K valores de cortes, ou limiares (Neste exemplo, para 4 classes, $K=4$)

Modelo de Incerteza sobre a VA S em \mathbf{u} .
 $p(\mathbf{u}, s | n)$ - Função de Distribuição de Probabilidade Condicionada a n amostras

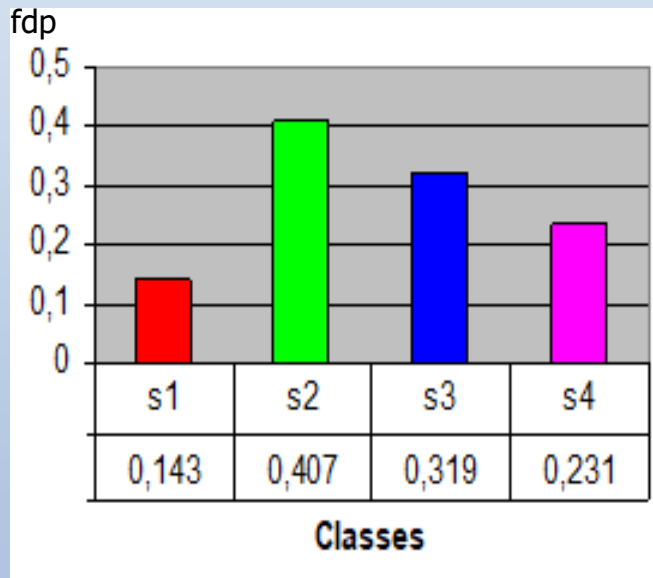
s	$fdpc$
s1	0.13
s2	0.37
s3	0.29
s4	0.21



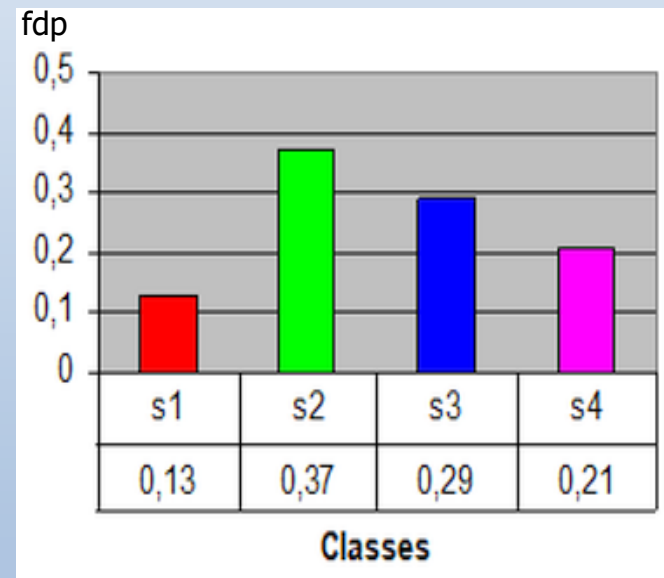
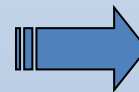
Procedimento por Indicação

• Variáveis Categóricas

- **Correção dos desvios de relação de ordem (Goovaerts, 1997)** – A somatória dos valores de probabilidade das classes da fdp deve ser igual a 1. Assim, será necessária a correção para garantir essa propriedade (escalonamento das probabilidades)



$$\sum_{k=1}^K p_k(\mathbf{u}) = \sum_{k=1}^K p(\mathbf{u}; s_k) > 1$$



$$\sum_{k=1}^K p_k(\mathbf{u}) = \sum_{k=1}^K p(\mathbf{u}; s_k) = 1$$

Procedimento por Indicação

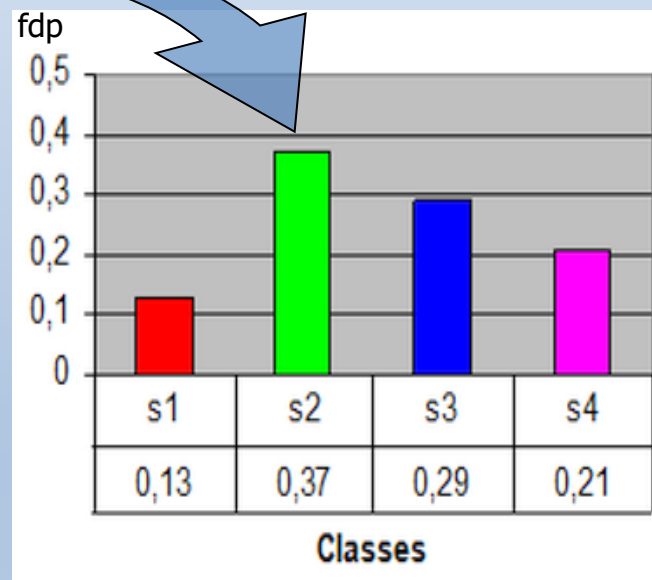
• Variáveis Categóricas

- **Estimando Parâmetros da VA: Critério da Moda (Probabilidade Máxima)**
- Estima-se o **valor da classe**, numa localização espacial \mathbf{u} , a partir da fdpc da VA S como:

$$s^*(\mathbf{u}) = \text{Max}_{k=1}^K (p(\mathbf{u}; s_k)) = \text{Max}_{k=1}^K (p_k(\mathbf{u}))$$

Neste exemplo: $s^*(\mathbf{u}) = s2$

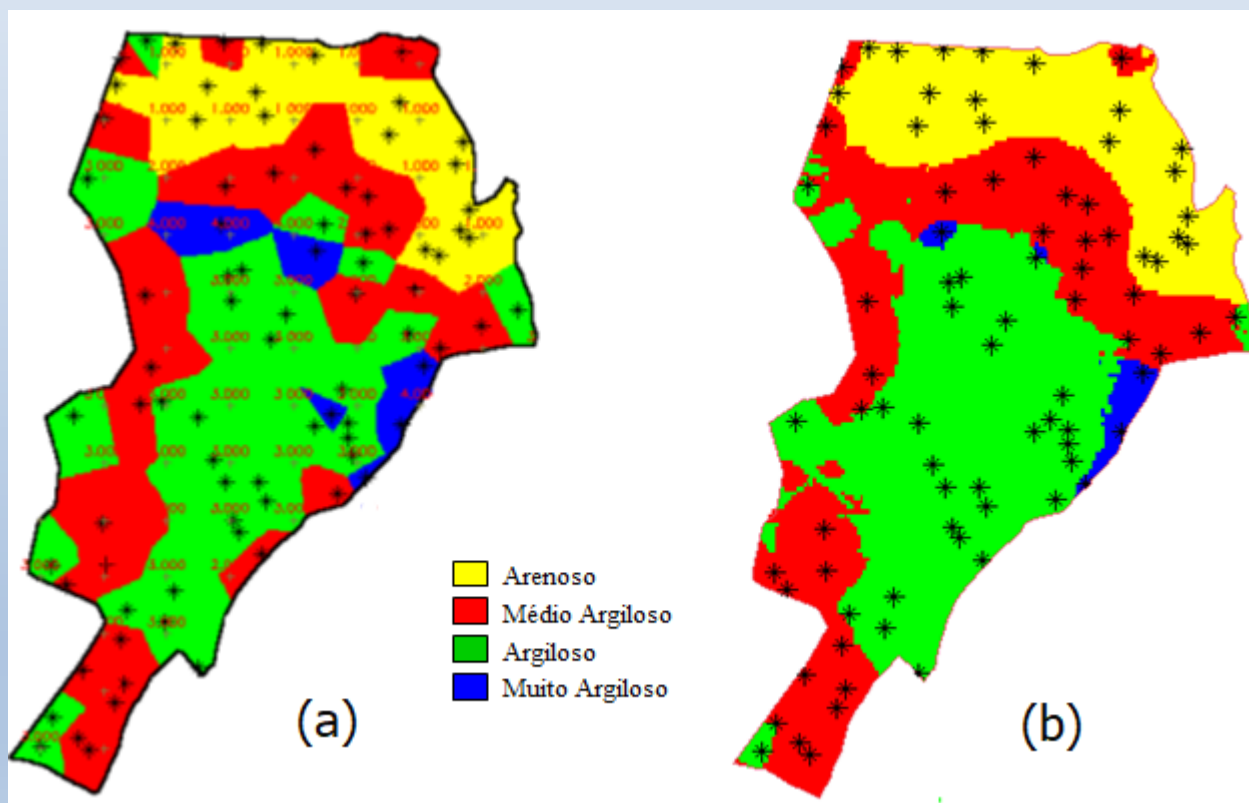
Outro estimador qualquer ?



Procedimento por Indicação

- **Variáveis Categóricas**

- Exemplos de Mapas de Estimativas: (a) Dirichlet e (b) Máxima Probabilidade



Mapas de Textura de Solo

Procedimento por Indicação

- **Variáveis Categóricas**

- **Estimativa de incertezas a partir da função de distr. de probabilidade**

- **Moda**: estimador de máxima probabilidade.

$$\text{Inc}(\mathbf{u}) = 1 - p(\mathbf{u}; s^*(\mathbf{u})) = 1 - \text{Max}_{k=1}^K (p_k(\mathbf{u}))$$

- **Entropia de Shannon**: mede a desordem (confusão) ou a aleatoriedade entre elementos de um sistema fechado.

$$\text{Inc}(\mathbf{u}) = - \sum_{k=1}^K p_k(\mathbf{u}) \cdot \ln(p_k(\mathbf{u}))$$

Principais propriedades da Entropia de Shannon:

- É um valor não negativo
- É igual a 0 quando uma das probabilidades é igual a 1 (as outras igual a 0)
- Seu máximo ocorre quando a distribuição for uniforme (máxima confusão)

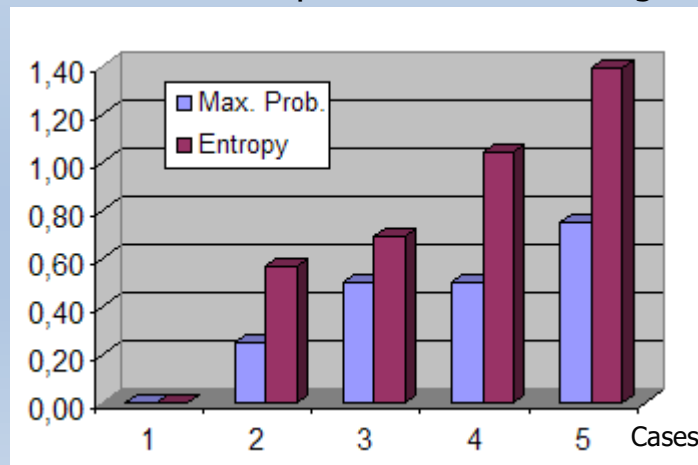
Procedimento por Indicação

- **Variáveis Categóricas**

- **Exemplo numérico: Estimativas de incerteza a partir de fdps**
- **Incerteza por Máxima Probabilidade x Entropia de Shannon**

Cases	$p(u,s1)$	$p(u,s2)$	$p(u,s3)$	$p(u,s4)$	Max Prob	Shannon
1	0,00	0,00	0,00	1,00	0,00	0,00
2	0,00	0,00	0,25	0,75	0,25	0,57
3	0,00	0,00	0,50	0,50	0,50	0,69
4	0,00	0,25	0,25	0,50	0,50	1,04
5	0,25	0,25	0,25	0,25	0,75	1,39

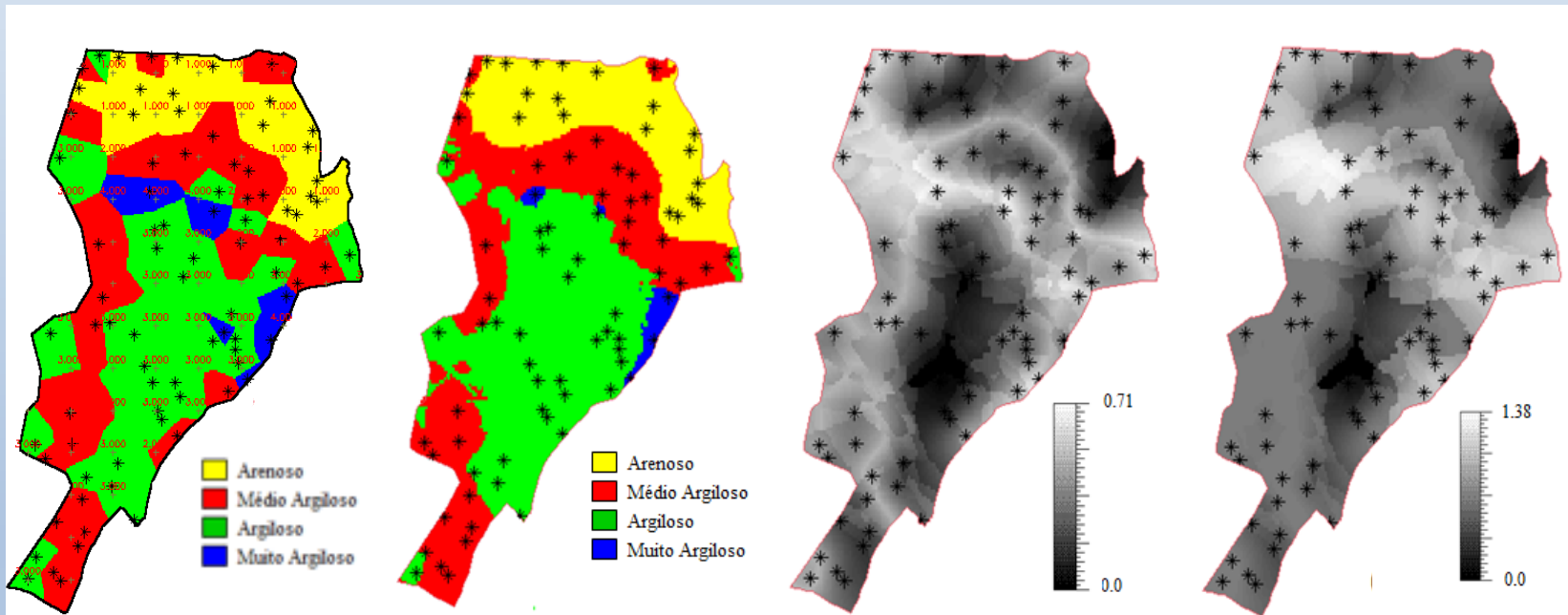
Tabela: Medidas de incerteza para diferentes configurações de probabilidades



Procedimento por Indicação

• Variáveis Categóricas

- Mapas de Estimativas: (a) Estimativas por Moda, (b) Incerteza por Máxima e (c) Incerteza por Entropia de Shannon



Mapas de Estimativa e Incertezas de Textura de Solo

Procedimento por Indicação

Parte III

Simulação por indicação para variáveis contínuas e categóricas

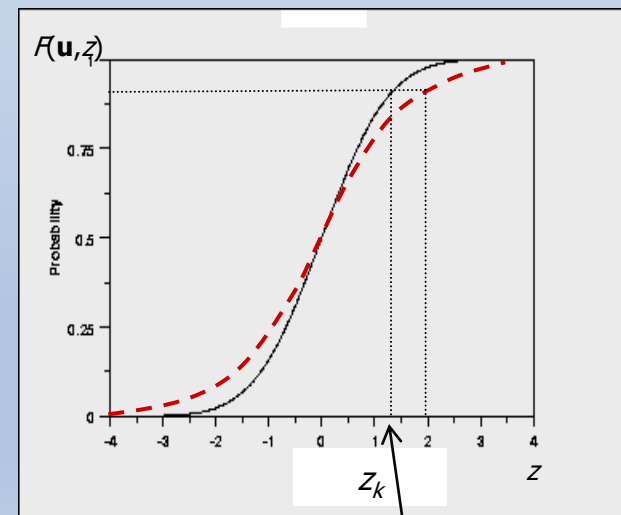
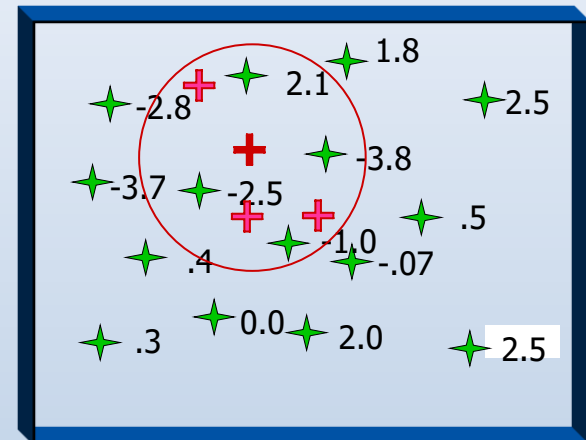
Procedimento por Indicação

- **Introdução**
- **Estimativa x Simulação (Modelagem Local x Espacial)**
- **Problemas com Estimador de Krigagem:**

- Considera apenas as amostras locais (não valores pré-estimados). Modelo de Covariância definido só pelas amostras.

$$Z^*(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}(\mathbf{u}) \cdot Z(\mathbf{u}_{\alpha}) + \left[1 - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_{\alpha}(\mathbf{u}) \right] \cdot m$$

- Tende a suavizar detalhes da variabilidade do atributo. (minimiza variância da estimativa)
- Valores baixos são superestimados e valores altos são subestimados (problema de valores extremos)
- Problemas com reprodução de covariância (próximo slide)



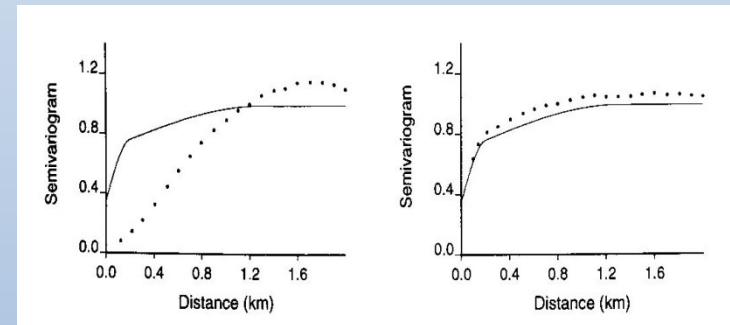
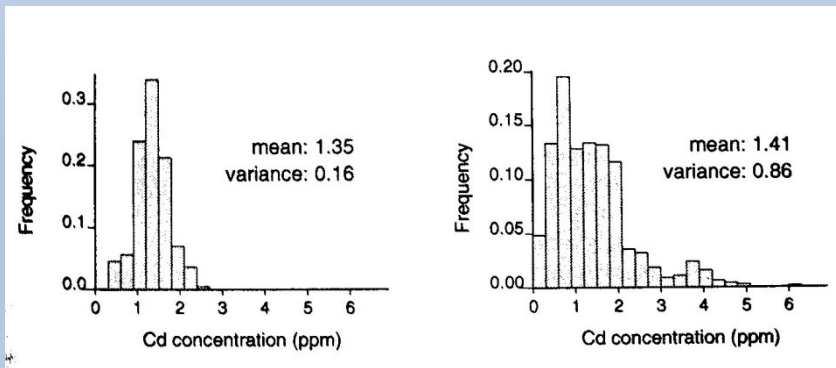
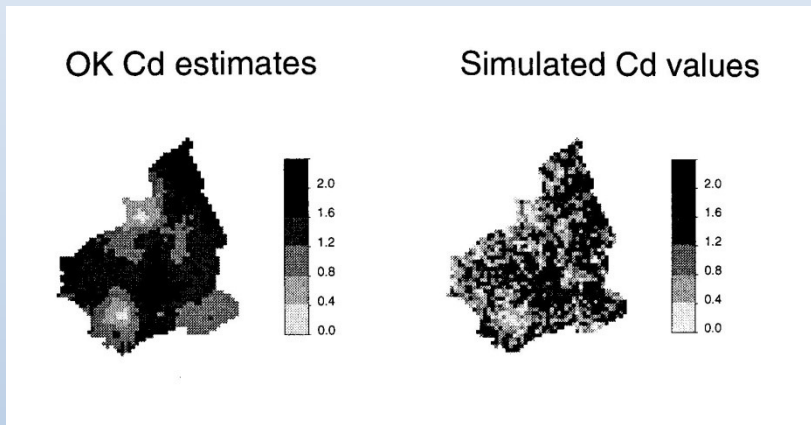
Valor de Z é subestimado pela krigagem

Procedimento por Indicação

• Simulação Estocástica (Goovaerts, 1997)

Estimativa x Simulação

Reproduzindo estatísticas do modelo (ilustrações)



Note o efeito de suavização da krigagem que leva a uma subestimação da variabilidade de "short-range" dos valores de Cádmió (Goovaerts)

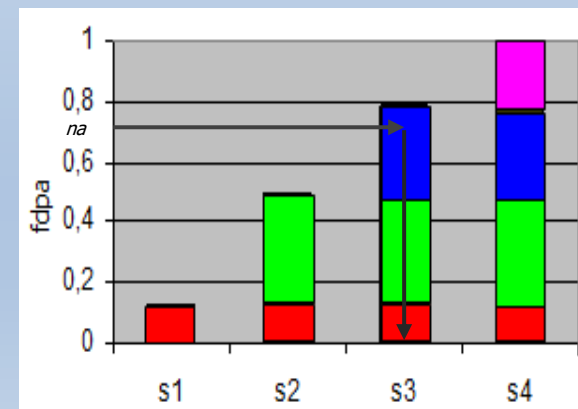
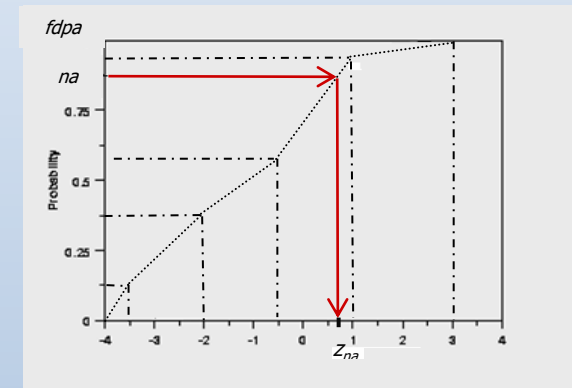
Procedimento por Indicação

• Simulação Estocástica – Obtendo realizações das fdacs

• Simulação: processo de obter, sortear, realizações de uma função de distribuição de probabilidades acumulada (condicionada ou não). Usa a fdac e um gerador de números aleatórios x

• N realizações de cada VA Z podem ser sorteadas de uma fdac repetindo-se n vezes os seguintes passos:

- Sorteia-se um número aleatório na entre 0 e 1 de uma distribuição uniforme.
- Mapeia-se o valor de na para um valor do atributo z_{na} ou s_{na} usando-se a fd estimada.
- A fd pode ser paramétrica, ex. gaussiana: média e variância, ou não paramétrica construída pela simulação sequencial por indicação apresentada a seguir.



Procedimento por Indicação

• Simulação estocástica: O método da Simulação Sequencial

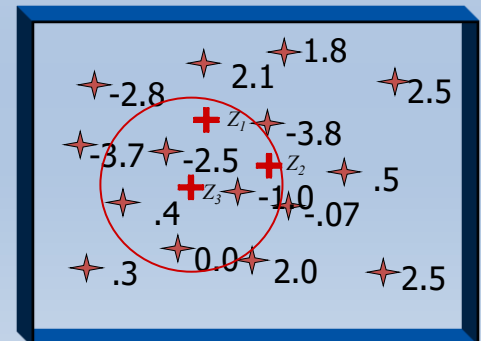
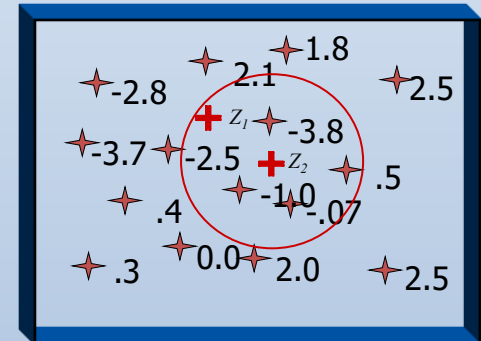
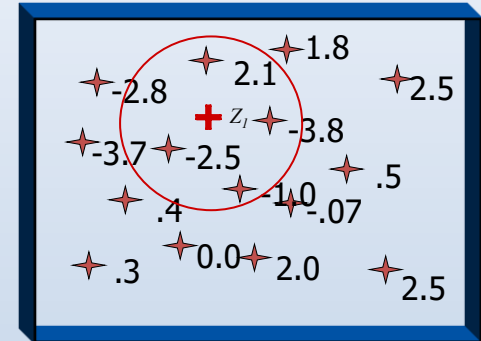
- Sorteia-se um valor $z_1^{(l)}$ da fdac univariada de Z_1 , $\text{Prob}\{Z_1 \leq z_1(n)\}$, condicionada às (n) amostras originais.
- Atualiza-se o conjunto de dados amostral original (n) para um novo conjunto de informação $(n+1)$:

$$(n+1) = (n) \cup \{Z_1 = z_1^{(l)}\};$$

- Sorteia-se um novo valor $z_2^{(l)}$ da fdac univariada de Z_2 , $\text{Prob}\{Z_2 \leq z_2(n+1)\}$, condicionada o conjunto de informação $(n+1)$:
- Atualiza-se o conjunto de informação $(n+1)$ para um novo conjunto de informação $(n+2)$:

$$(n+2) = (n+1) \cup \{Z_2 = z_2^{(l)}\};$$

- Sequencialmente consideram-se todas as J VAs Z_j 's.
- Repete-se a sequencia acima para cada nova realização l , até se obter L Campos Aleatórios.



Procedimento por Indicação

- **Simulação estocástica: *Predições a partir dos L campos simulados***
 - Para *atributos Contínuos*, considera-se os valores simulados em cada posição espacial (\mathbf{u}) e nas L realizações
 - **Média:** $[z(\mathbf{u})]_{\text{E}}^* = m^* = \left(\frac{1}{L}\right) \cdot \sum_{l=1}^{l=L} z(\mathbf{u})^l$
Exemplo: 2, 3, 9, 8, 2, 7, 9- $m^* = 40/7 = 5.714$
 - **Variância:** $\sigma^2 = \left(\frac{1}{L}\right) \cdot \sum_{l=1}^{l=L} (z(\mathbf{u})^l - m^*)^2$
 - **Mediana:** Ordena-se os valores $z(\mathbf{u})^l$ e obtém-se o valor que divide o conjunto ao meio (50%). Exemplo:
 - Outros: Moda, Quantis,...
 - Para *atributos Categóricos*, considera-se as classes simuladas em cada posição espacial (\mathbf{u}) e nas L realizações
 - **Moda** - Classe de maior probabilidade ($\text{Prob}(C_i) = \text{Nro de elementos da classe } C_i / L$)
 - **Incerteza** – $1.0 - \text{Prob}(\text{Moda})$
 - **Entropia de Shannon ?**

2, 2, 3, **7**, 8, 9, 9

Mediana = **7**

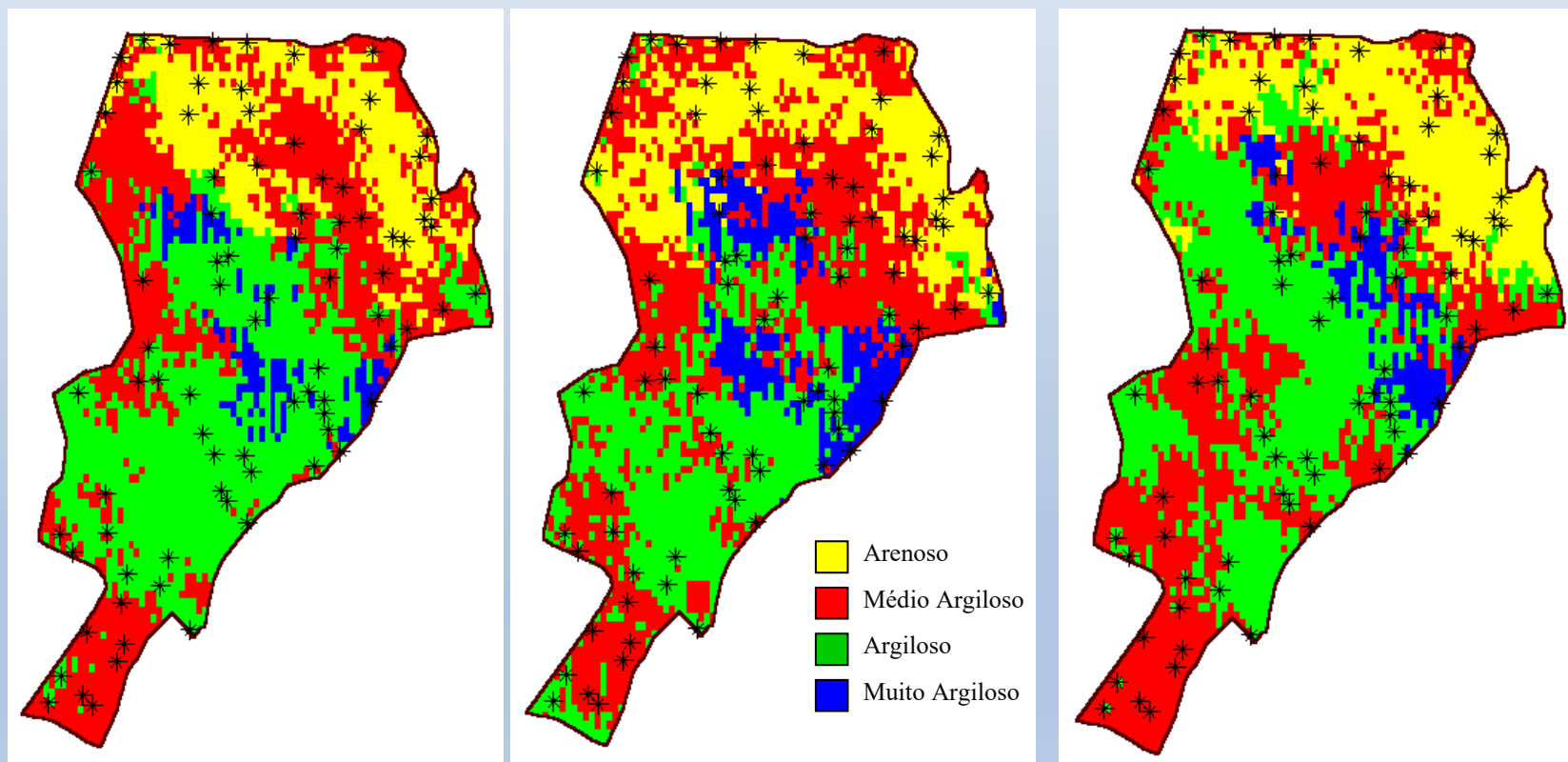
1, 4, 4, **5, 6**, 7, 7, 7

Mediana = $(5+6) \div 2$

= **5.5**

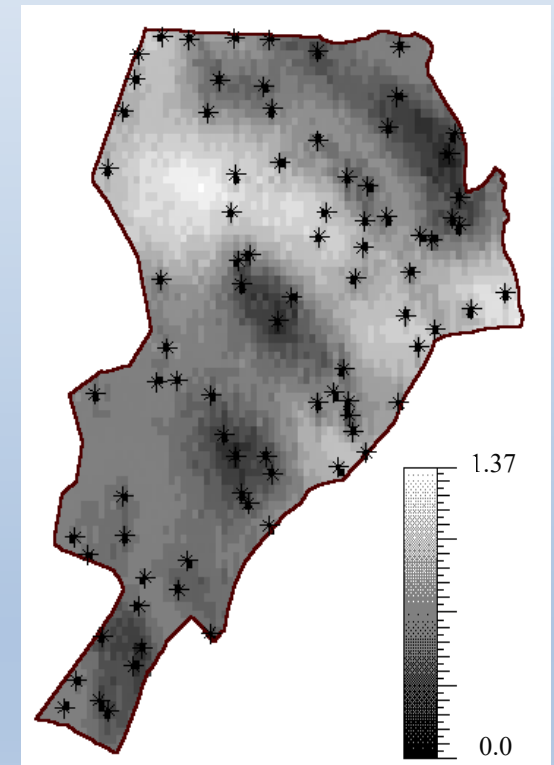
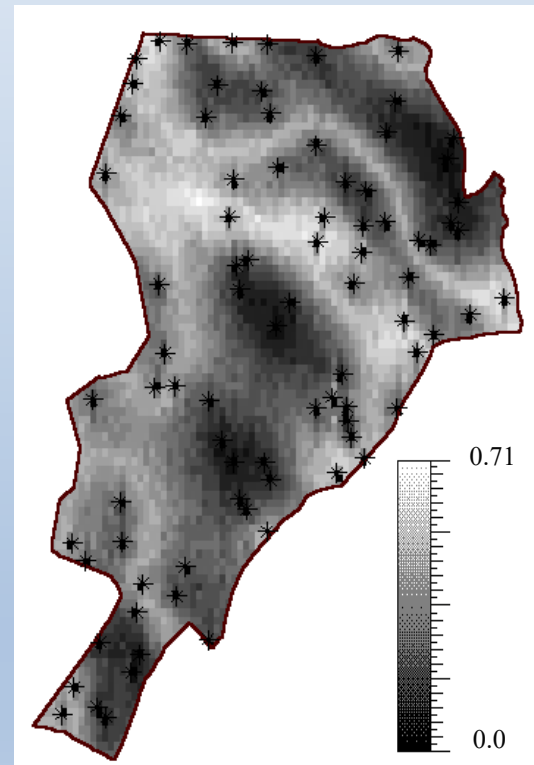
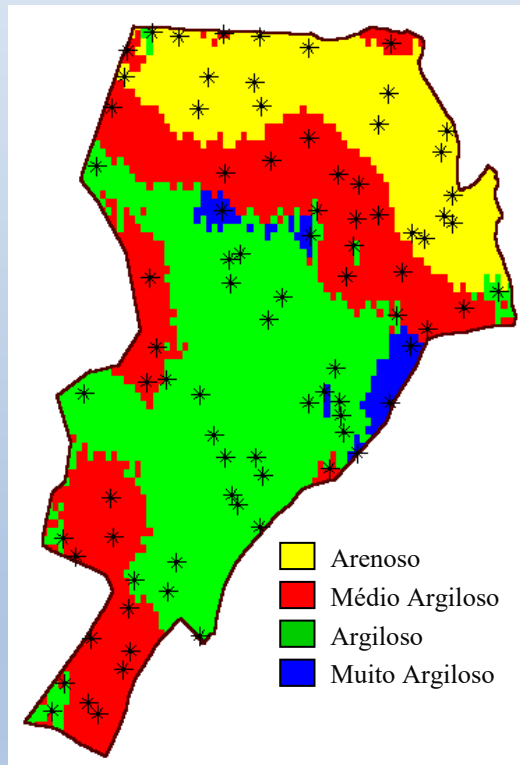
Procedimento por Indicação

- **Introdução – Simulação Estocástica (Goovaerts, 1997)**
 - **Exemplos de campos simulados para variáveis categóricas**



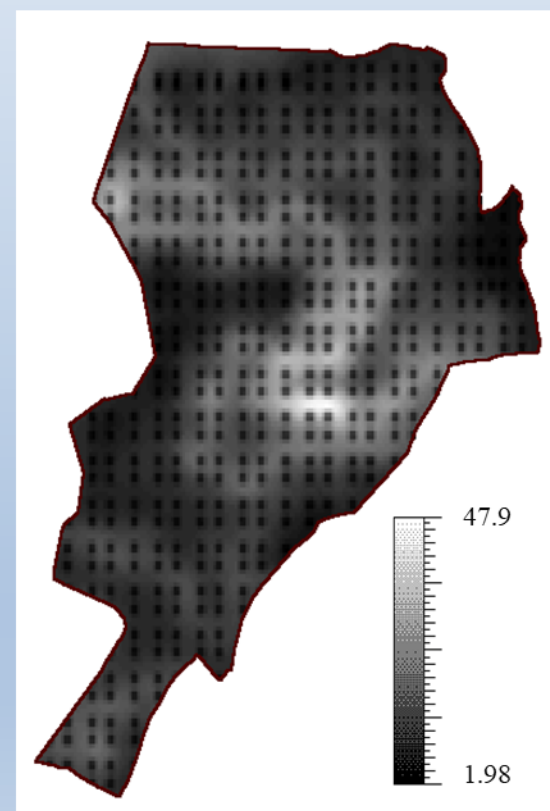
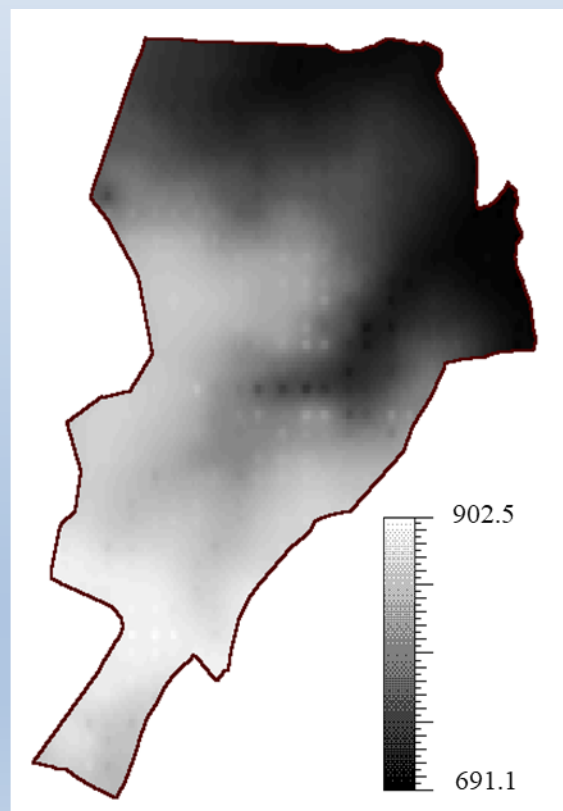
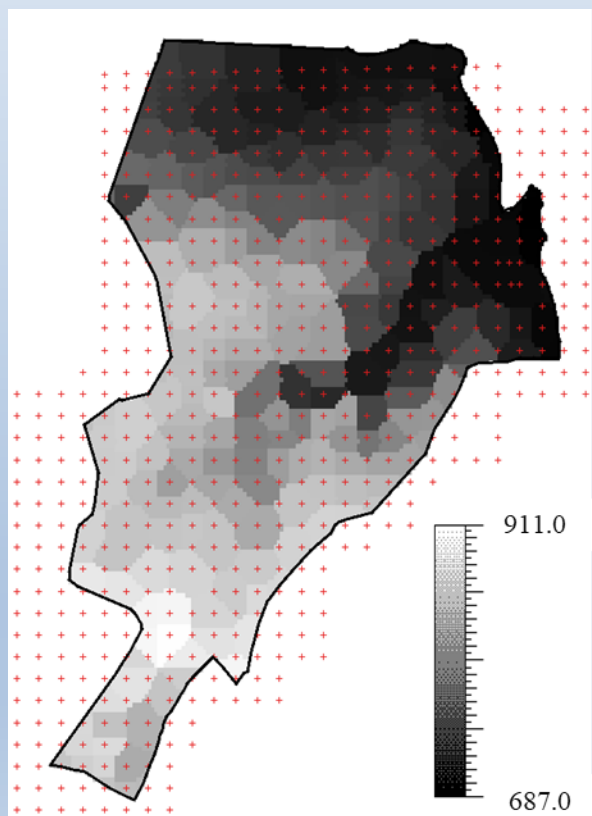
Procedimento por Indicação

Exemplo de Mapas de Estimativas e de Incertezas, obtidos das realizações, para atributos categóricos.



Procedimento por Indicação

- **Introdução – Simulação Estocástica (Goovaerts, 1997)**
 - **Exemplo de Mapas de Estimativas e de Incertezas, obtidos das realizações, para atributos contínuos.**



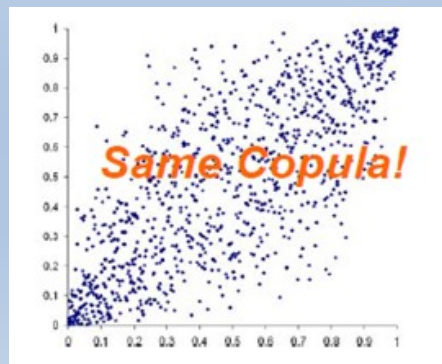
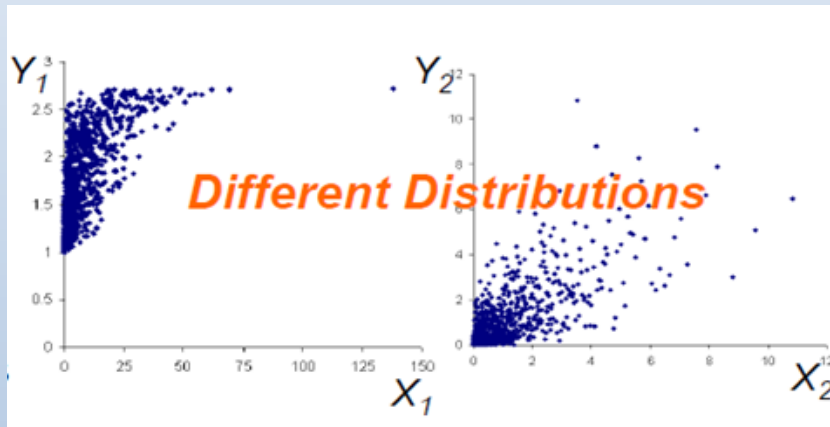
Modelagem por Geoestatística por Indicação

Parte II

Introdução às Copulas Espaciais

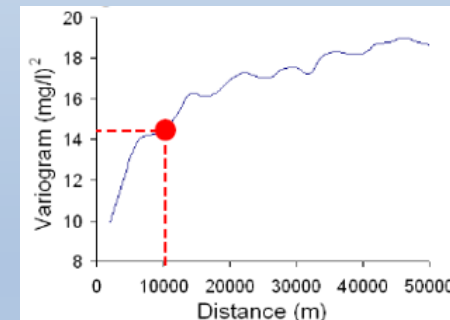
Geoestatística por Copulas

- Definição – Copulas** representam a estrutura de dependência (“correlação”) entre VAs (2 ou mais) num sistema de referência de probabilidade $[0, 1]$. As probabilidades são obtidas das distribuições marginais F (univariadas) das VAs consideradas, $p_i = F(z_i)$



Cópuas capturam a dependência (correlação) sem influência das marginais.

Cópuas bivariadas descrevem a dependência espacial sobre todo o intervalo de quantis e não somente a dependência média quadrática como nos variogramas.



Cópuas mostram estrutura de correlação (covariância) no espaço de probabilidades
 $u = F(z)$

$$\rho_S = 12 \int_0^1 \int_0^1 u_1 u_2 dC(u_1, u_2) - 3$$

Geoestatística por Copulas

- Definição** – Exemplos de estruturas bivariadas, e respectivas densidades de copulas. A esquerda temos diferentes copulas com covariâncias (0.7) e distribuições marginais normal padrão idênticas (Gräler, 2010). À direita densidade de copula e copula acumulada gaussianas bivariadas.

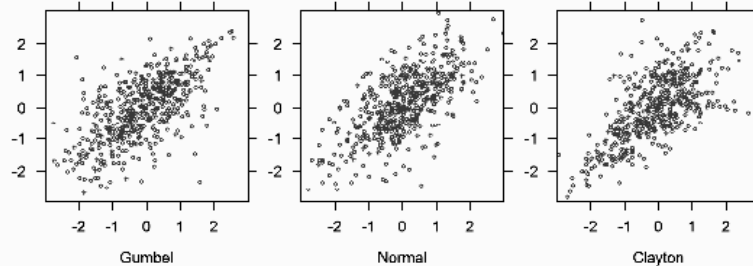


Figure 1: A comparison of different dependence structures with identical covariance (0.7) and standard normal distributed margins. In the *Gumbel* case the scatter plot shows higher dependence in the upper-right corner in contrast to the *Clayton* plot which gets tighter in the lower-left corner. The plot of the bivariate *normal* distribution exhibits the characteristic elliptical shape.

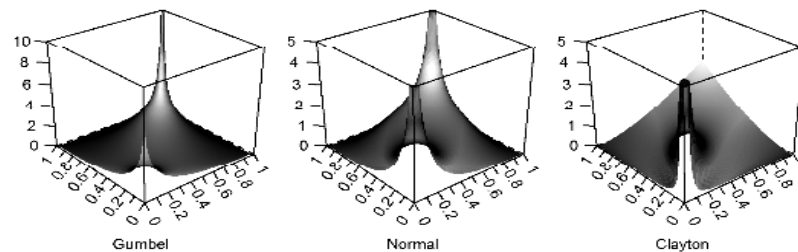
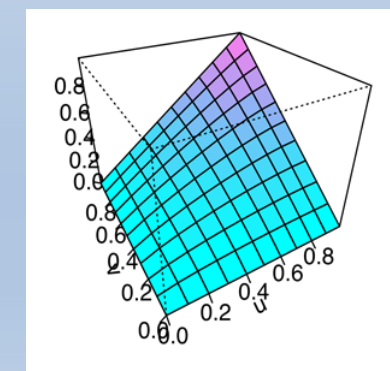
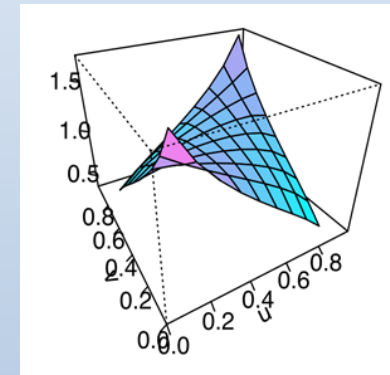


Figure 2: The densities of the underlying copulas of Figure 1.



Geoestatística por Copulas

- **Conceitos – (Bárdossy e Li, 2008)**

Definição: A copula é definida como uma **função de distribuição** no cubo unitário n dimensional. Todas as distribuições marginais estão uniformizadas em $[0, 1]$. Formally:

$$C: [0, 1]^n \rightarrow [0, 1]$$

Propriedades:

$$C(\mathbf{u}^{(n)}) = u_i \text{ se } \mathbf{u}^{(n)} = (1, \dots, 1, u_i, 1, \dots, 1)$$

Se pelo menos um dos u_i 's for igual a 0, então $C(\mathbf{u}^{(n)}) = 0$

Copulas e distribuições multivariadas são ligadas entre si pelo **teorema de Sklar's** [Sklar, 1959]. *Sklar provou que cada distribuição multivariada $H(z_1, \dots, z_n)$ (ou $F(t_1, \dots, t_n)$) pode ser representada pelo auxílio de uma copula como:*

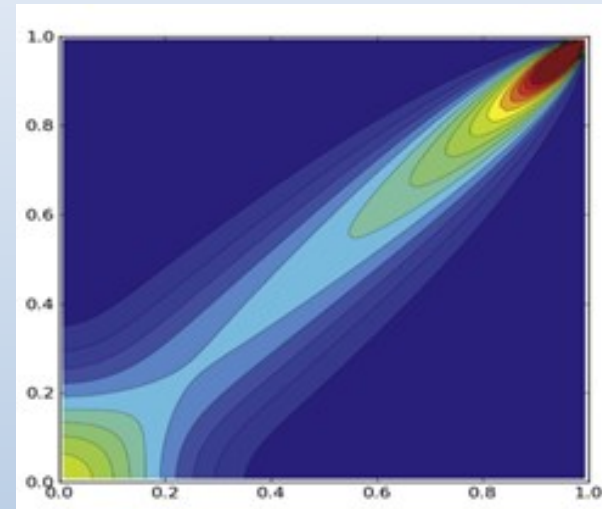
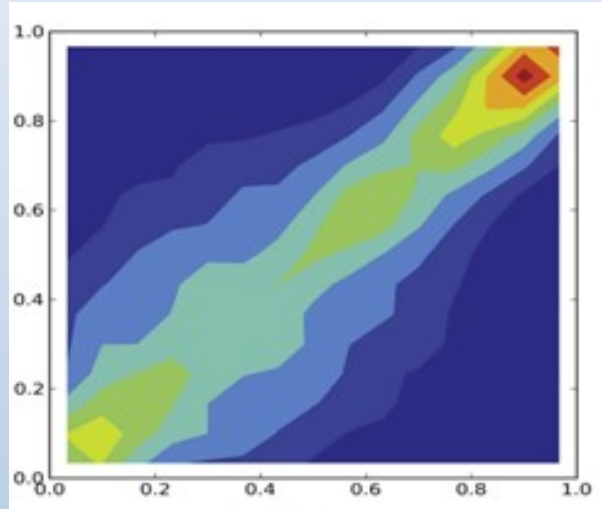
$$F(t_1, \dots, t_n) = C(F_1(t_1), \dots, F_n(t_n))$$

Copulas podem ser construídas a partir de funções de distribuição, como descrito por Nelsen [1999]:

$$C(\mathbf{u}) = C(u_1, \dots, u_n) = F(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n))$$

Geoestatística por Copulas

Densidades de Copulas Experimentais e Teóricas (Modelos matemáticos)



A partir de um conjunto de amostras bidimensionais (tipo pares de z para uma distancia h) podemos obter representações experimentais de densidade de copulas.

A partir da representação experimental pode se *ajustar um modelo matemático teórico* para a densidade da copula.

Os modelos teóricos de densidade de copulas são utilizados para inferências espaciais (estimativas e incertezas)

Geoestatística por Copulas

Conceitos - Copulas e interpolação espacial (Bárdossy, 2006)

Spatial Interpolation Using Copula-Based Geostatistical Models

309

H is an n -dimensional distribution function with margins F_1, \dots, F_n . Moreover, if $F_1^{-1}, \dots, F_n^{-1}$ are the inverse distribution functions of F_1, \dots, F_n , we get

$$C(u_1, \dots, u_n) = H(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n)). \quad (2)$$

If C is an absolutely continuous copula, its density can be written as

$$c(u_1, \dots, u_n) = \frac{\partial^n C(u_1, \dots, u_n)}{\partial u_1 \dots \partial u_n} = \frac{h(F_1^{-1}(u_1), \dots, F_n^{-1}(u_n))}{\prod_{i=1}^n f_i(F_i^{-1}(u_i))}. \quad (3)$$

where h denotes the density of H and the f_i denote the densities of F_i . One of the

Exemplo: Para distribuições gaussianas f e h só depende de vetores de médias, variâncias e covariâncias obtidas das informações das amostras.

$$\mathbf{f} = p(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{1}{2\sigma^2}(x - \mu)^2\right)$$

$$\mathbf{h} = p(x; \mu, \Sigma) = \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(x - \mu)^T \Sigma^{-1} (x - \mu)\right)$$

Geoestatística por Copulas

Conceitos - Copulas e interpolação espacial (Bárdossy, 2008)

Para outras distribuições f e h são obtidos de acordo com o modelo estatístico adotado. Os parâmetros do modelo são ajustados por máxima verossimilhança usando-se as amostras: Exemplo:

The v -Transformed Multivariate Normal Copula (Bardossy, 2008 – Bayesian Approach)

Let Y be $N(0, G)$, an n dimensional normal random variable with $0T = (0, \dots, 0)$ mean and G correlation matrix. All marginal's are supposed to have unit variance. Let X be defined for each coordinate $j = 1, \dots, n$ as:

$$X_j = \begin{cases} k(Y_j - m) & \text{if } Y_j \geq m \\ m - Y_j & \text{if } Y_j < m \end{cases}$$

All one dimensional **marginal** of X are identical and have the distribution function:

$$H_1(x) = P(X < x) = \Phi\left(\frac{x}{k} + m\right) - \Phi(-x + m)$$

and density:

$$h_1(x) = \frac{1}{k} \phi\left(\frac{x}{k} + m\right) + \phi(-x + m)$$

The corresponding **multivariate** and density distribution functions are:

$$H_n(x_1, \dots, x_n) = P(X_1 < x_1, \dots, X_n < x_n) \\ = \sum_{i=0}^{2^n-1} (-1)^i \Phi(\zeta_i + \mathbf{m})$$

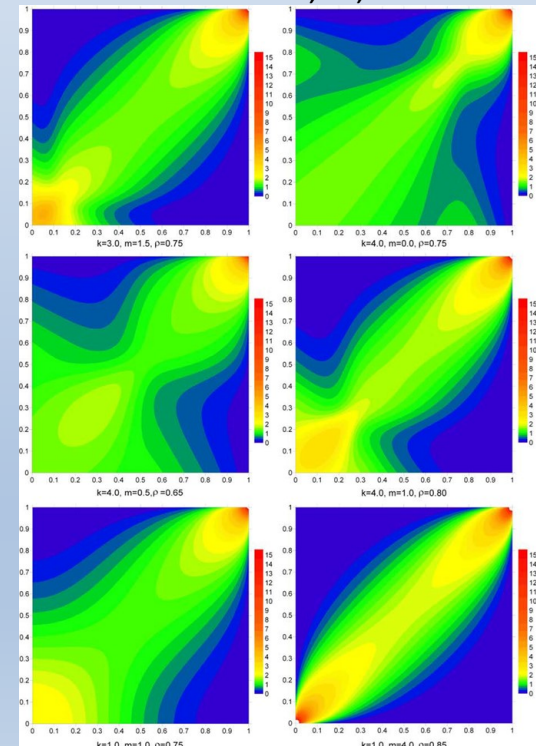
$$h_n(x_1, \dots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Gamma^{-1}|^{\frac{1}{2}}} \sum_{i=0}^{2^n-1} \frac{1}{k^{n-\sum_{j=0}^{n-1} i_j}} \\ \cdot \exp\left(-\frac{1}{2}(\zeta_i + \mathbf{m})^T \Gamma^{-1}(\zeta_i + \mathbf{m})\right)$$

$$\zeta_i^T = (b((-1)^{i_1})x_1, \dots, b((-1)^{i_n})x_n)$$

$$i = \sum_{j=0}^{n-1} i_j 2^j$$

$$b(x) = \begin{cases} -1 & \text{if } x < 0 \\ \frac{1}{k} & \text{if } x > 0 \end{cases}$$

Bivariate v -transformed normal copula density with different combinations of k , m , and r values.



Geoestatística por Copulas

Conceitos - Copulas e interpolação espacial

No caso da krigagem por indicação, usamos a definição de média estatística aplicada a valores discretizados da fda para obter média, ou seja:

$$E[Z] = \int_{-\infty}^{\infty} z \cdot f(z) dz$$

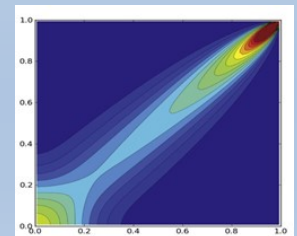
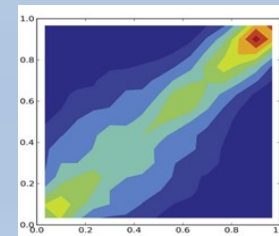
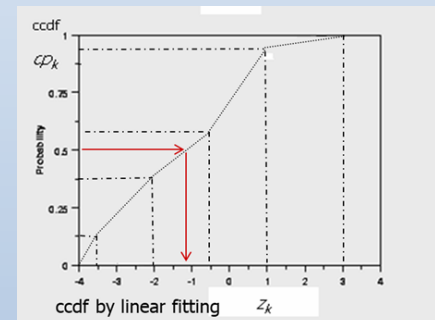
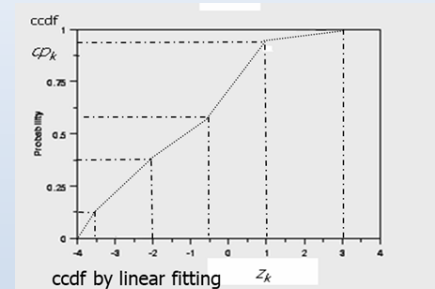
$$[z(\mathbf{u})]_E^* = \int_{-\infty}^{\infty} z \cdot dF(\mathbf{u}; z | (n)) \approx \sum_{k=1}^{K+1} z'_k \cdot [F(\mathbf{u}; z_k | (n)) - F(\mathbf{u}; z_{k-1} | (n))]$$

A mediana, quantil .5, também pode ser obtida a partir da fda discretizada (figura ao lado)

Ao definirmos um modelo matemático teórico para a *densidade da copula*, valores de médias e medianas podem ser obtidos por (Modelagem Bayesiana):

$$z_a^* = \int_0^1 F_Z^{-1}(u) c_{\mathbf{x},n}(u | u_1 = H_Z(z'_1), \dots, u_n = H_Z(z'_n)) du$$

$$z_m^* = F_Z^{-1}(u = C_{\mathbf{x},n}^{-1}(0.5))$$



Geoestatística por Copulas

Copulas e interpolação espacial (Bárdossy e Li, 2008). Mapas Estimativas e Incertezas

Mapas de Médias e Medianas

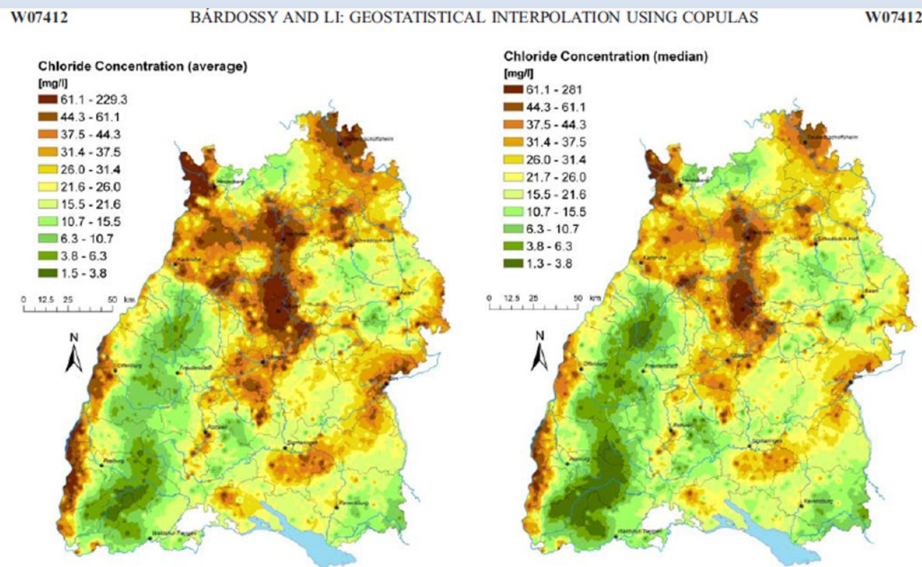


Figure 10. Interpolation maps for chloride using v-transformed normal copula mean (left) and median (right).

Mapas de Incertezas por intervalos de confiança

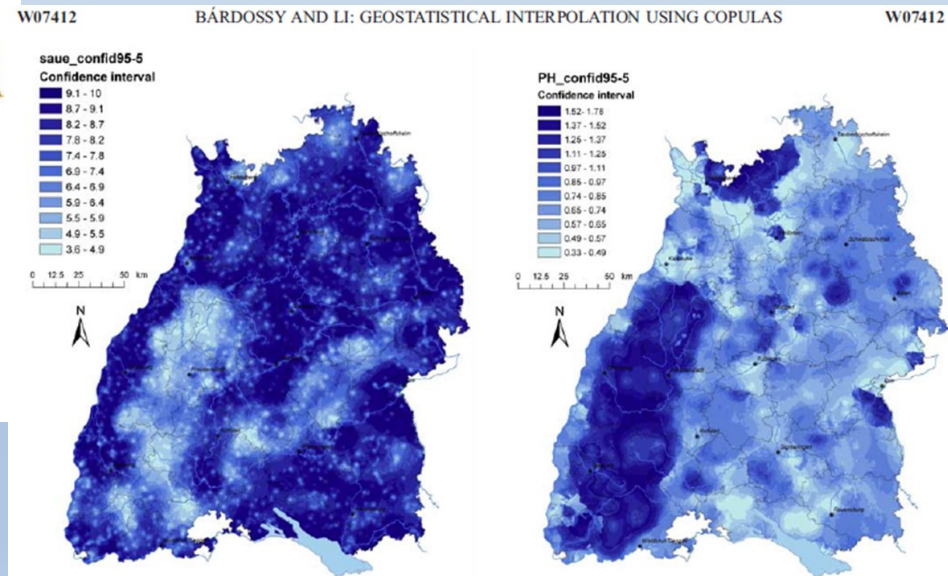
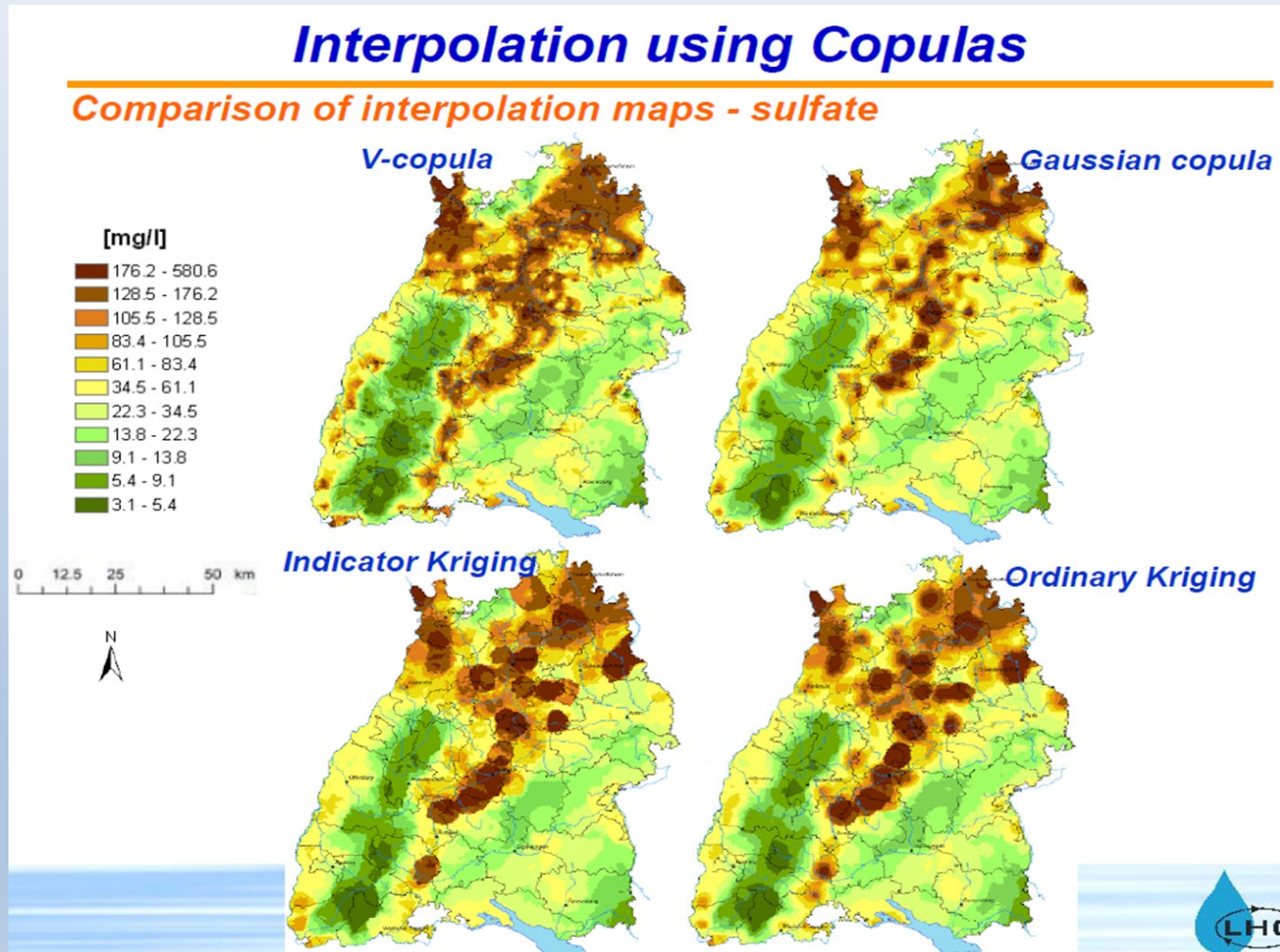


Figure 12. Length of the 90% confidence intervals obtained from the v transformed normal copula for dissolved oxygen (left) and pH (right).

Geoestatística por Copulas



Geoestatística por Copulas

Interpolation using Copulas

Crossvalidation results

Mean absolute error

	Chloride [mg/l]	Nitrate [mg/l]	pH [-]	Dissolved oxygen [mg/l]	Sulfate [mg/l]
V-copula	14.861	13.689	0.192	1.876	34.992
G-copula	15.380	13.938	0.194	2.049	38.128
O.Kriging	16.817	13.853	0.198	1.911	42.365
I.Kriging	16.561	15.501	0.200	1.989	43.979

Referências Bibliográficas

- BURROUGH , P. A.; McDONNELL , R.A. **Principles of geographical information systems**. New York: Oxford university pres, 1998. 344p.
- CAMARGO, E. C. G. **Desenvolvimento, implementação e teste de procedimentos geoestatísticos (Krigeagem) no sistema de processamento de informações georreferenciadas (SPRING)**. 1997. 148 p. IBI: <6qtX3pFwXQZ3r59YD6/GPfam>. (INPE-6410-TDI/620). Dissertação (Mestrado em Sensoriamento Remoto) - Instituto Nacional de Pesquisas Espaciais (INPE), Sao Jose dos Campos, 1997.
- CAMARGO, E.C.G, FELGUEIRAS, C. A.; MONTEIRO, A .M.V.; **A Importância da Modelagem da Anisotropia na Distribuição Espacial de Variáveis Ambientais Utilizando Procedimentos Geoestatísticos**. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 10., 2001, Foz do Iguaçu. Anais. Paraná, 2001 [CDROM].
- DEUTSCH, C.V.; JOURNEL, A.G. **GSLIB: geostatistical software library and user's guide**. New York: Oxford University Press, 1998. 369p, 1CD.
- DRUCK, S.; Carvalho, M.S.; Câmara, G.; Monteiro, A.V.M. (eds) "Análise Espacial de Dados Geográficos". Brasília, EMBRAPA, 2004 (ISBN: 85-7383-260-6).
- FELGUEIRAS, C.A. **Modelagem ambiental com tratamento de incertezas em sistemas de informação geográfica: o paradigma geoestatístico por indicação**. 1999. 165p. Tese (Doutorado em Computação Aplicada) – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, Publicado em <<http://www.dpi.inpe.br/teses/carlos/>>, 1999.
- FELGUEIRAS, C. A.; MONTEIRO, A .M.V.; FUCKS, S.D.; CAMARGO, E.C.G. **Uso de simulação estocástica não linear para inferências de atributos espaciais numéricos**. In:X SBSR, Foz do Iguaçu. Anais. Paraná, 2001. [CD-ROM]. p. 437-446.

Referências Bibliográficas

FELGUEIRAS, C. A.; FUCKS, S.D.; MONTEIRO, A .M.V. Classificação de atributos espaciais baseada em informações de incertezas: uma metodologia de apoio a decisões. In: **SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO**, 11., 2003, Belo Horizonte. Anais. Minas Gerais, 2003. INPE, São José dos Campos, SP, [CD-ROM]. p. 967-974.

GOOVAERTS, p. **Geostatistics for natural resources evaluation**. New York : Oxford University Press, 1997. 483p.

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

ISAAKS, E. H; SRIVASTAVA, R.M. **An introduction to applied geostatistics**: New York: Oxford university press, 1989. 561p.

ORTIZ, J.O.; FELGUEIRAS, C.A.; DRUCK, S.; MONTEIRO, A.M.V. Modelagem de fertilidade do solo por simulação estocástica com tratamento de incertezas. **Pesquisa Agropecuária Brasileira-PAB**, Brasília, v.39. n.4, p.379-389, 2004.

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

SOARES, A. **Geoestatística para ciências da terra e do ambiente**. Lisboa: IST Press. 2000. 206p.

Wackernagel, H. **Multivariate geostatistics**. New York, Springer, 1998. 291p.

Referências Bibliográficas

Copulas

Bárdossy A. **Copula-based geostatistical models for groundwater quality parameters**. Water Resources Research 2006;42.

Bárdossy A, Li J. **Geostatistical Interpolation using copulas**. Water Resources Research 2008;44.

FELGUEIRAS, C. A.; CAMARGO, E. C. G.; ORTIZ, J. O. Abordagem geoestatística por indicação com uso de copulas bivariadas empíricas para modelagem de incertezas associadas a imagens de sensoriamento remoto. In: SIMPÓSIO BRASILEIRO DE SENSORIAMENTO REMOTO, 17. (SBSR), 2015, João Pessoa. **Anais...** São José dos Campos: INPE, 2015. p. 6373-6380.

Kazianka H, Pilz J. **Bayesian spatial modeling and interpolation using copulas**. Computers & Geosciences 2010.

Nelsen R. **An Introduction to Copulas**. 2nd ed. New York: Springer 2006.

Gräler Benedikt, Hannes Kazianka, Giovana Mira de Espindola. **Copulas, a novel approach to model spatial and spatio-temporal dependence** . GIScience for Environmental Change, Campos do Jordão (São Paulo), Brazil, November 27, 2010.

Gräler Benedikt . **Spatial Copulas course**. 2011. Course available at http://ifgi.uni-muenster.de/~b_grae02/