Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

Master of Science in Geospatial Technologies

Geostatistics
Assessment of Spatial Uncertainty with Indicator Geostatistics

Carlos Alberto Felgueiras
cfelgueiras@isegi.unl.pt

Assessment of Spatial Uncertainty

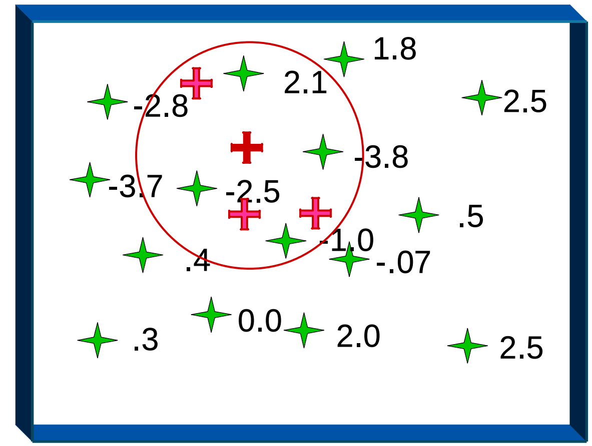
Contents

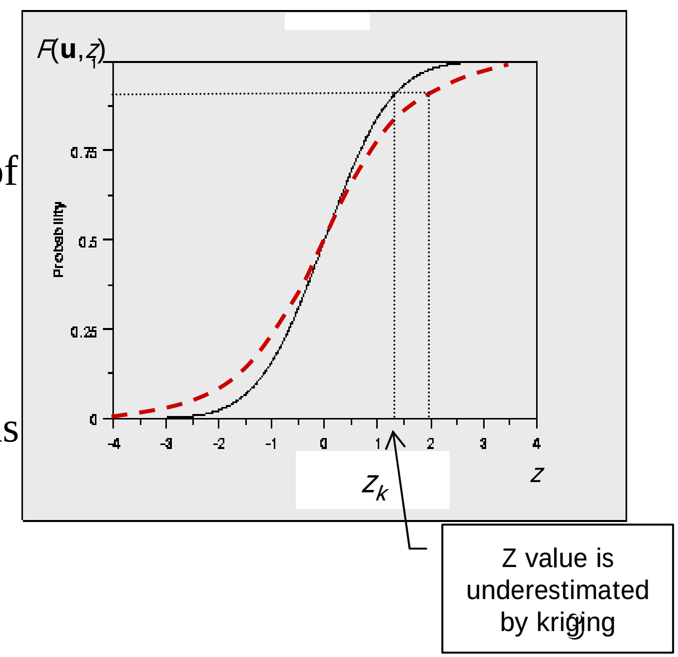
Assessment of Spatial Uncertainty

- **Introduction**
- **Estimation x Simulation**
- **Problems with Kriging Estimation:**
  - Account only for local samples (not the values already estimated). Consider covariances only with samples.

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

  - Tend to smooth out local details of the spatial variation of the attribute. (local error variance is minimum.)
  - Small values are *overestimated*. Large values are *underestimated* (extreme values problem)
  - Estimates appears more variable in densely sampled areas than in sparsely sampled areas





Assessment of Spatial Uncertainty

- **Introduction**
- **Estimation x Simulation – Reproducing model statistics**
- **Stochastic Simulation (Goovaerts, 1997)**
  - Generates a map or a realization of *z*-values, say, $\{z^{(l)}(\mathbf{u}), \mathbf{u} \in A\}$ with l denoting the lth realization.
  - Requisites for simulated maps
    1. Data values are honored at their n locations

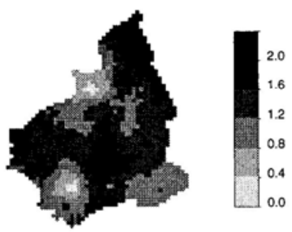
       $$z^{(l)}(\mathbf{u}) = z(\mathbf{u}_\alpha) \qquad \forall \mathbf{u} = \mathbf{u}_\alpha, \alpha = 1,\ldots, n$$

       The realization is said to be conditional (to the samples)
    2. The histogram of simulated values reproduces closely the declustered sample histogram.
    3. The covariance model $C(\mathbf{h})$, or better, the set of indicator covariance models $C_I(\mathbf{h}; z_k)$, for various thresholds z_k are reproduced.

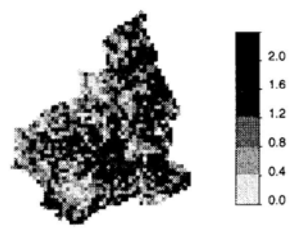
Assessment of Spatial Uncertainty

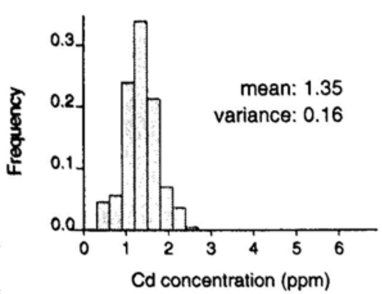
- **Stochastic Simulation (Goovaerts, 1997)**

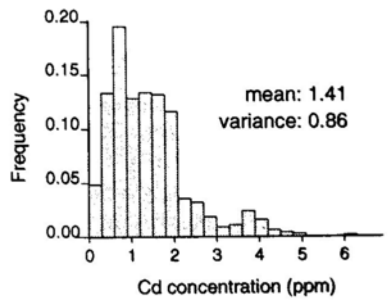
OK Cd estimates



Simulated Cd values



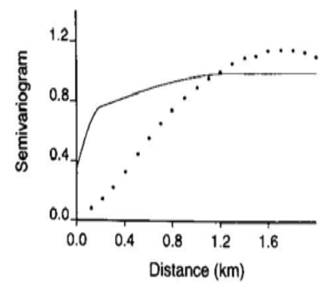


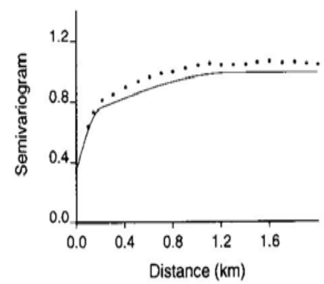


Estimation x Simulation

Reproducing model statistics

(ilustrations)





Note the smoothing effect of kriging that leads to underestimation of the short-range variability of Cd values (Goovaerts)

Assessment of Spatial Uncertainty

- **Stochastic Simulation (Goovaerts, 1997)**
  - **Modeling Spatial Uncertainty**
  - The multiGaussian and Indicator kriging based algorithms provide a measure only of *local uncertainty* because each conditional cdf relates to a single location $\mathbf{u}_j$ (single-point ccdfs)

$$F\left(\mathbf{u}_j; z_c|(n)\right), j = 1,\ldots,J$$

  - The *joint probability*, probability that the *z*-values at *J* locations $\mathbf{u}_j$ are **jointly** no greater than critical threshold z_{cj} is given by:

$$\text{Prob}\left\{Z(\mathbf{u}_j) \le z_{cj},\ j = 1,\ldots, J \mid (n)\right\} = F\left(\mathbf{u}_1,\ldots, \mathbf{u}_j; z_1,\ldots z_j \mid (n)\right)$$

  - **The joint probability may be assessed numerically from a set of *L* realizations of the spatial distribution of z-values over the *J* locations $\mathbf{u}_j$, that is:**

$$\text{Prob}\left\{Z(\mathbf{u}_j) \le z_{cj},\ j = 1,\ldots, J \mid (n)\right\} \approx \frac{1}{L}\sum_{l=1}^{L}\prod_{j=1}^{J} i^{(l)}\left(\mathbf{u}_j; z_{cj}\right)$$

  - where $i^{(l)}(\mathbf{u}_j; z_c)$ =1 if the simulated *z*-value at $\mathbf{u}_j$ does not exceed z_c and 0 otherwise.
  - Problem: How to draw the L realizations?

Assessment of Spatial Uncertainty

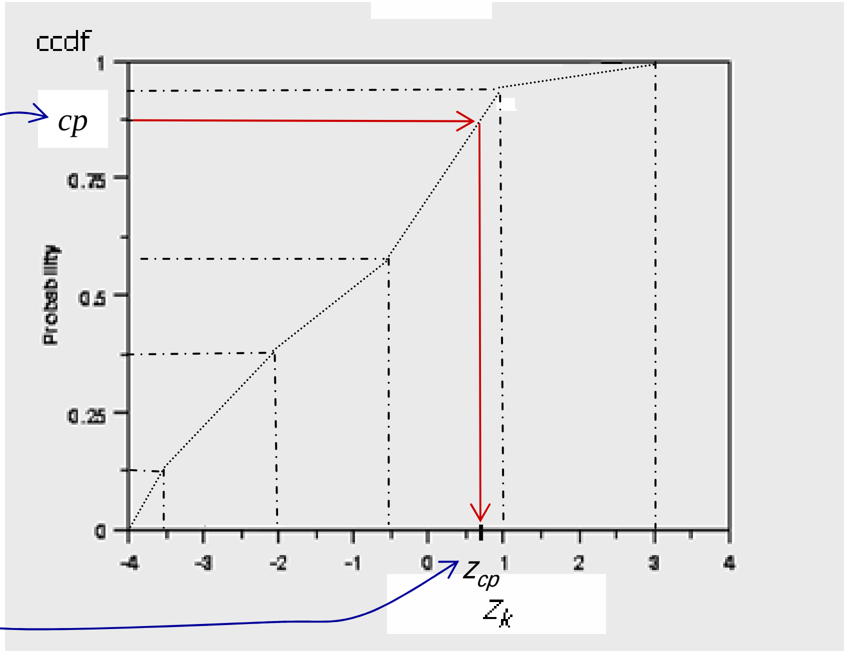
- **Stochastic Simulation (Goovaerts, 1997)**
  - **Modeling Spatial Uncertainty – Numerical Example (not extensive) with $z_{c1} = z_{c2} = 2$**

Realizations		*Indicator Values*		
z_1	z_2	$i^{(l)}(\mathbf{u}_1; z_c)$	$i^{(l)}(\mathbf{u}_2; z_c)$	$\prod_{j=1}^{2} i^{(l)}(\mathbf{u}_j; z_c)$
0.5	1.1	1	1	1
0.8	1.8	1	1	1
1.7	2.1	1	0	0
1.9	2.3	1	0	0
2.7	1.7	0	1	0
3.5	2.9	0	0	0
1.1	1.7	1	1	1
2.8	2.7	0	0	0
2.9	1.8	0	1	0
3.2	2.3	0	0	0
Correlation 0.717		Prob($z_1<z_c$) = 5/10=.5	Prob($z_2<z_c$) = 5/10=.5	Prob($z_1 < z_{c1}$; $z_2 < z_{c2}$) = 3/10=.3

Assessment of Spatial Uncertainty

- **Stochastic Simulation - Drawing realizations from ccdfs**
- Simulation: process of drawing realizations from a cumulative (conditioned or not) distribution function. Uses the ccdf and a random number generator.

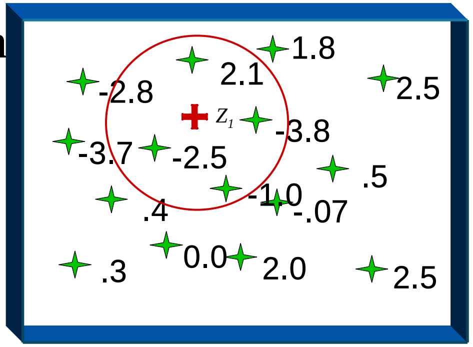
- **N** realizations of each RV *Z* can be drawn from a ccdf repeating *n* times the following steps:
    - Generating a random number between 0 and 1 (*cp* - cumulative probability value).
    - Mapping the *cp* value to the z_{cp} value using the given ccdf.



- Problem: How to obtain the ccdf for spatial uncertainty assessment?

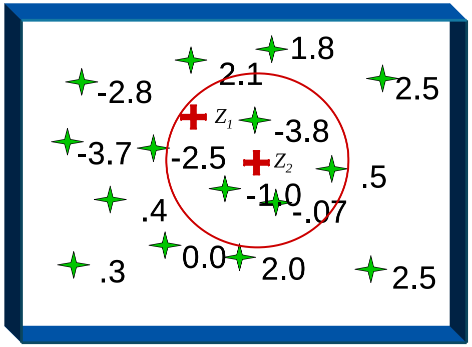
Assessment of Spatial Uncertainty

- **Stochastic Simulation- The Sequential Simulation Approach**



  - Draw a value $z_1^{(l)}$ from the univariated ccdf of Z_1, Prob$\{Z_1 \leq z_1|(n)\}$, conditioned to the (n) original samples.
  - Update the original sample data set (n) to a new information set (n+1) :

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

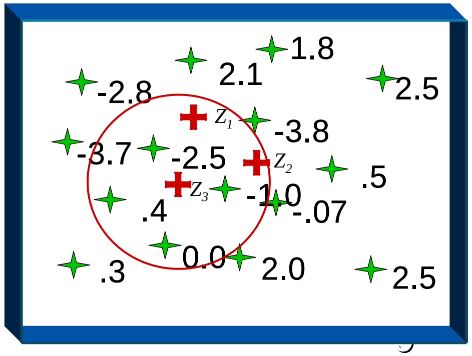


  - Draw a new value $z_2^{(l)}$ from the univariated ccdf of Z_2, Prob$\{Z_2 \leq z_2|(n+1)\}$, conditioned to the information set (n+1):
  - Update the information set (n+1) to a new information set (n+2) :

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

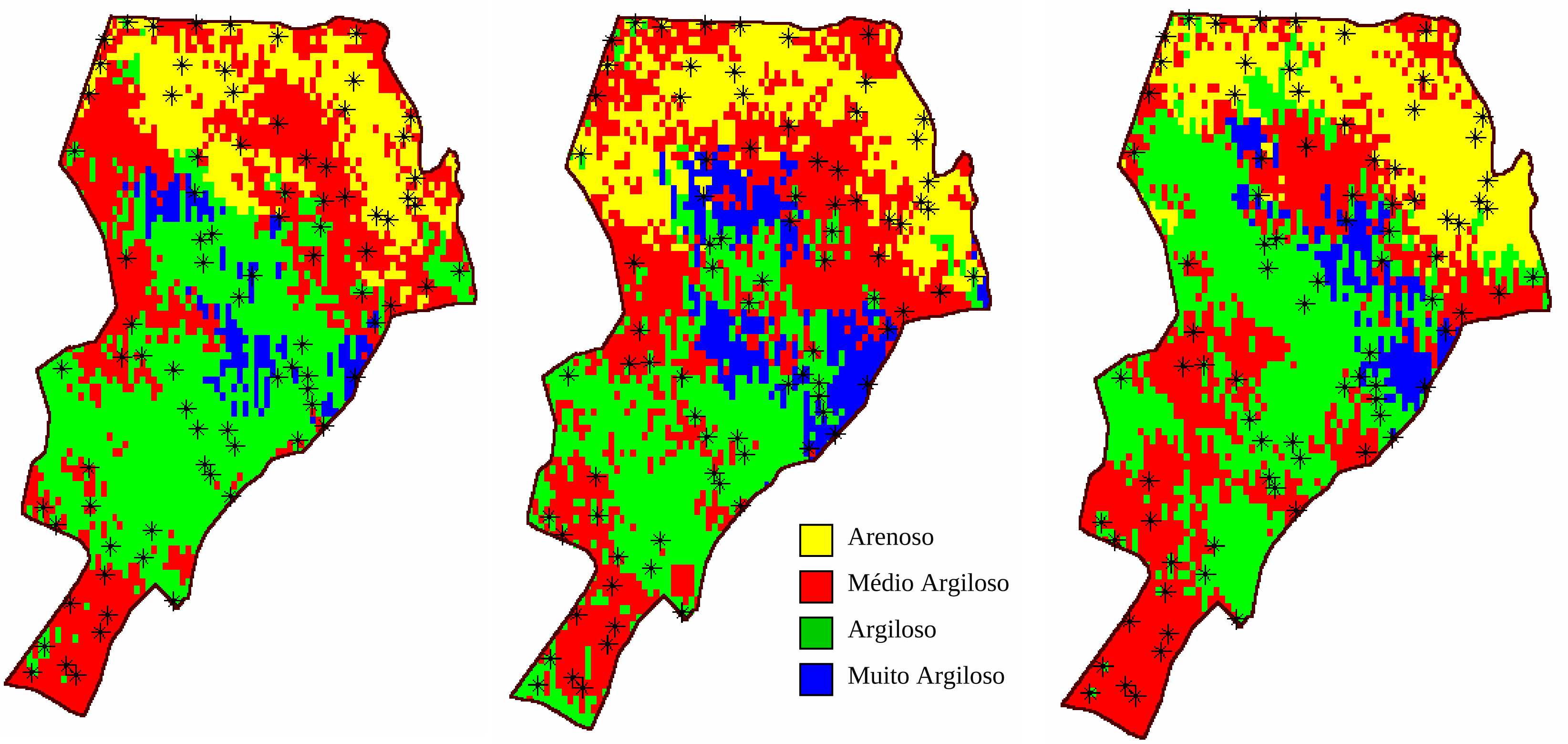
  - Sequentially consider all the J Random Variables Z_j*'s*.
  - Repeat the above sequence for a new l realization (up till L Random Fields)

.
.



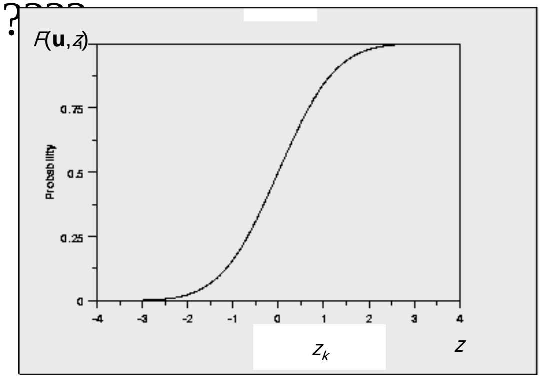
Assessment of Spatial Uncertainty

- **Introduction - Stochastic Simulation (Goovaerts, 1997)**
  - **Example of simulation fields for categorical variables**

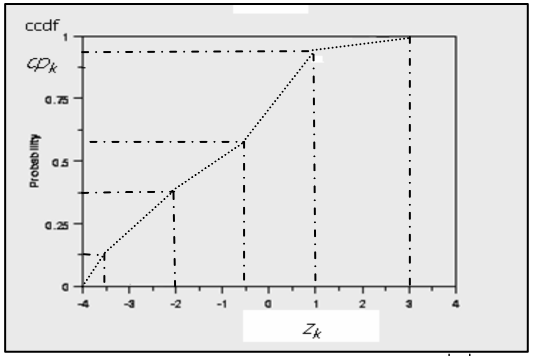


Assessment of Spatial Uncertainty

- **Introduction - Stochastic Simulation (Goovaerts, 1997)**
  - **Sequential Gaussian Simulation x Sequencial Indicator Simulation**
  - The sequential simulation principle is independent of the algorithm or model used to establish the sequence of univariate ccdf's
  - see continuation in the pag 124 of the Deutsch book ???



  - *Sequential Gaussian Simulation. Parametric.*
    - Relies on that the ccdf's at each location u is Gaussian (must be checked). The ccdf is determined by the mean and the standard deviation of the distribution.



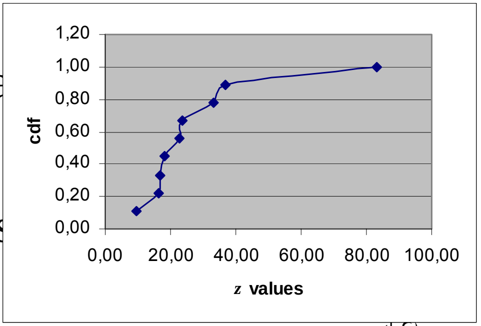
  - *Sequential Indicator Simulation. NonParametric*.
    - No assumption on the shape of the ccdf's.
    - Each ccdf is approximated by probabilities is evaluated for K thresholds z_k.

Assessment of Spatial Uncertainty

- **Estimates and Uncertainties from realizations**
  - Given a set of L realization values at a location **u** it is possible to assess:
  - **For continuous variables**
    - The *mean value* μ: add all values and divide result by L.
    - The *variance value* σ^2 : mean value of the squared difference between each value and the mean value μ.
    - The *standard deviation* σ: square root of the variance value
    - The *median value* $q_{.5}$: sorting the values and getting a value that splits the data set in the middle (50% or p=.5).
    - Any q_p *quantile*: sorting the values and getting a value tha split the data set considering the probability p.
    - *Confidence intervals* with standard deviations or quantiles
    - *Cumulative Distribution Function* (cdf)

mean = 29.01

Realizations		
values	sorted values	cdf
22.5	9.6	1/9
17	16.5	2/9
33.3	17	3/9
36.7	18.4	4/9
18.4	22.5	5/9
23.7	23.7	6/9
16.5	33.3	7/9
83.4	36.7	8/9
9.6	83.4	1



Assessment of Spatial Uncertainty

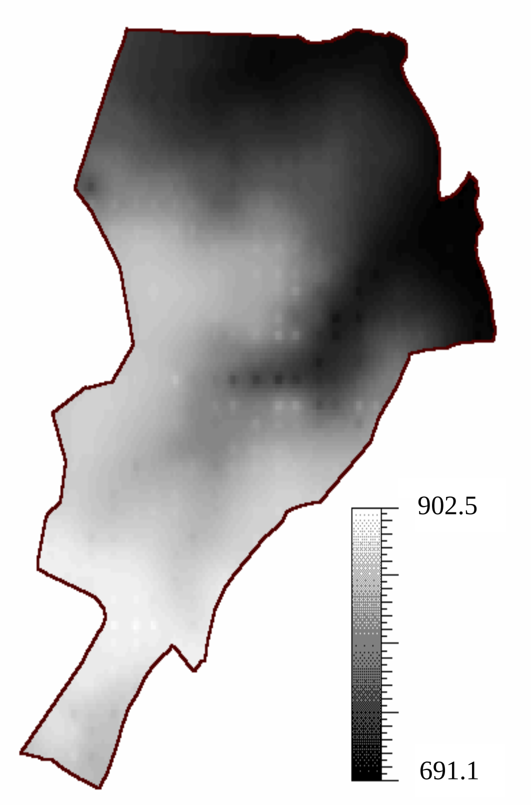
- **Estimates and Uncertainties from realizations**
  - Given a set of *L* realization values at a location **u** it is possible to assess:
  - **For categorical variables**
    - pdfs**:** counting the frequency of each class
    - Estimates based on mode values, for example.
    - Uncertainties based on the probability of the mode value.
    - Uncertainties based on the overall probabilities (Shannon Entropy, for example).

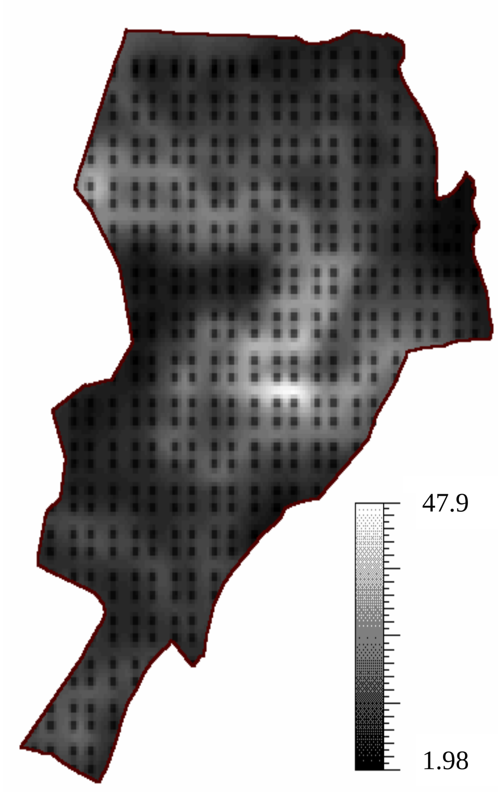
Realizations
classes
AB
AB
S
S
AB
AB
S
A
AB

class	freq.
A	1
AB	5
S	3

Assessment of Spatial Uncertainty

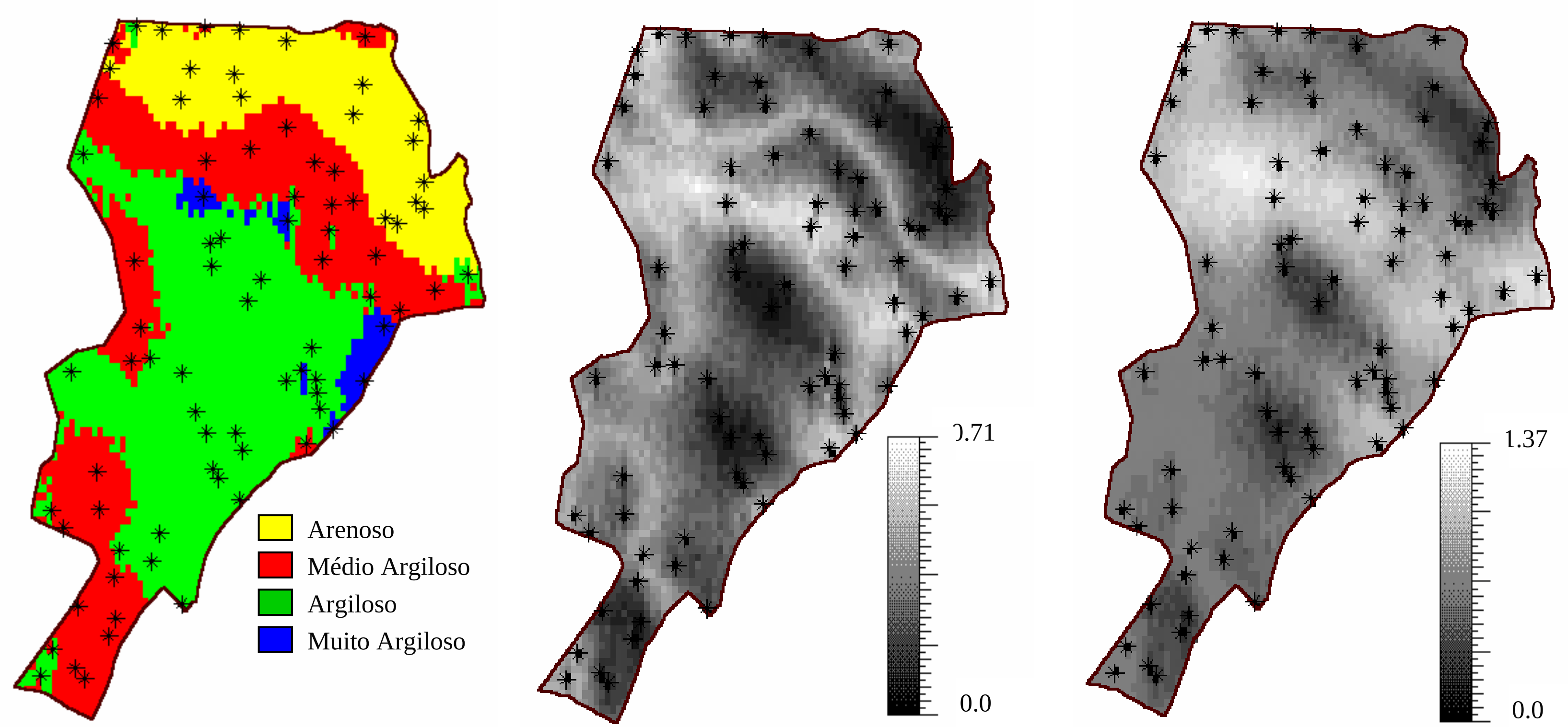
Example of Maps of Estimates and Uncertainties of continuous attributes from realizations.





Assessment of Spatial Uncertainty

Example of Maps of Estimates and Uncertainties of categorical attributes from realizations.



Assessment of Spatial Uncertainty

- **Accounting for secondary information on modeling spatial uncertainties**

Why accounting for secondary information ?

• generally primary data are sparse and poorly correlated in space.

• the estimation can be improved when secondary denser information is taken into consideration.

• Important: the secondary information must be correlated with the primary data

Assessment of Spatial Uncertainty

- **Acounting for secondary information in predictions (Goovaerts book)**

Accounting for Secondary Information in Kriging

- *Exhaustive* secondary information (*locations* **u** *being estimated and* $\mathbf{u}_{\alpha}$)
  - Kriging within strata
  - Simple kriging with varying local means
  - Kriging with an external drift
  - Co-located kriging

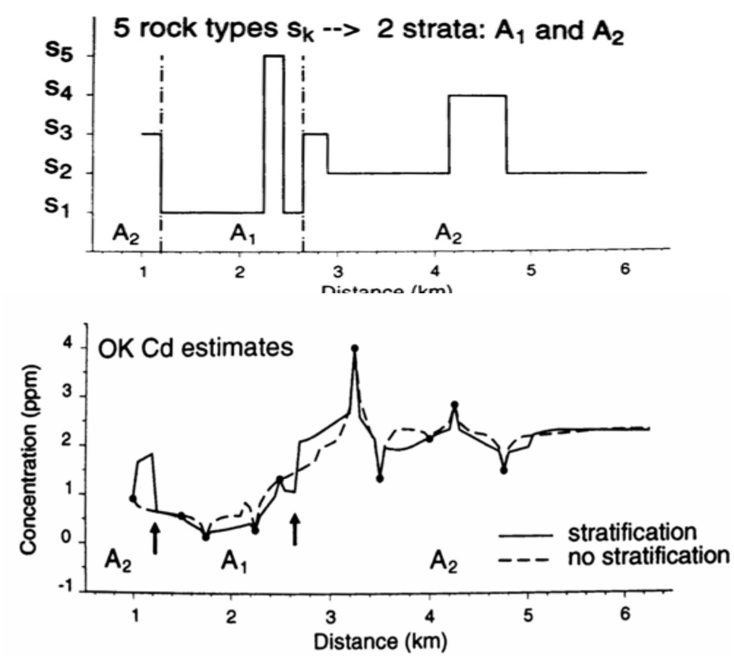
Accounting for Secondary Information in Kriging

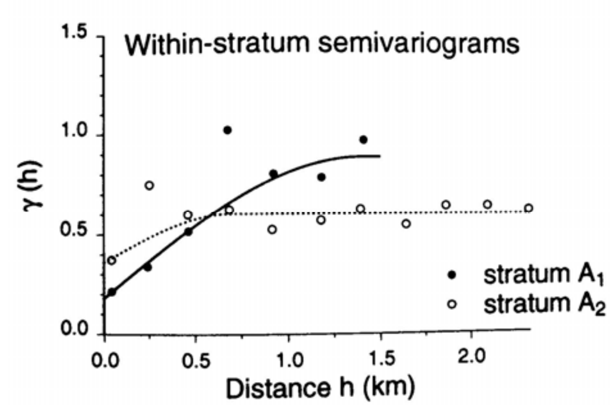
- Better sampled secondary information
  - Cokriging
  - Cross semivariogram estimation
  - Linear model of coregionalization
  - Markov models

Assessment of Spatial Uncertainty

- **Acounting for secondary information in predictions**

- *Kriging within strata (example Goovaerts)*
  - Stratify the study area considering the second information (soil map, for example)
    - Interpolation within each stratum separetely using stratum-specific semivariogram.





Assessment of Spatial Uncertainty

- **Accounting for secondary information in predictions**

- *Simple kriging with varying local means*

- The know stationary mean m , required by the simple kriging, can be replaced by known estimated varying means $m^*_{SK}(\boldsymbol{u})$.

$$Z^*_{SK}(\boldsymbol{u}) - \mathrm{m} = \sum_{\alpha=1}^{\mathrm{n}(\boldsymbol{u})} \lambda_\alpha^{SK}(\boldsymbol{u})\left[Z(\boldsymbol{u}_\alpha) - \mathrm{m}\right]$$

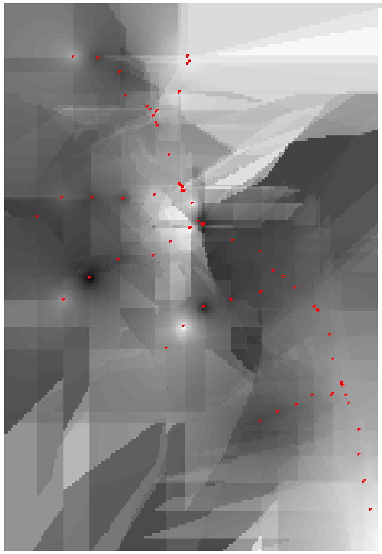
$$Z^*_{SK}(\boldsymbol{u}) - m^*_{SK}(\boldsymbol{u}) = \sum_{\alpha=1}^{\mathrm{n}(\boldsymbol{u})} \lambda_\alpha^{SK}(\boldsymbol{u})\left[Z(\boldsymbol{u}_\alpha) - m^*_{SK}(\boldsymbol{u})\right]$$

- Different estimates of the primary local mean can be used, depending on the secondary information available.
  - specialist knowledge
  - scattergrams (regression between primary and secondary co-located data)

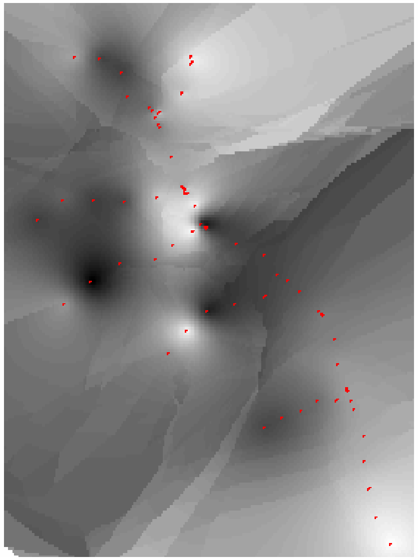
Assessment of Spatial Uncertainty

- **Accounting for secondary information in predictions**

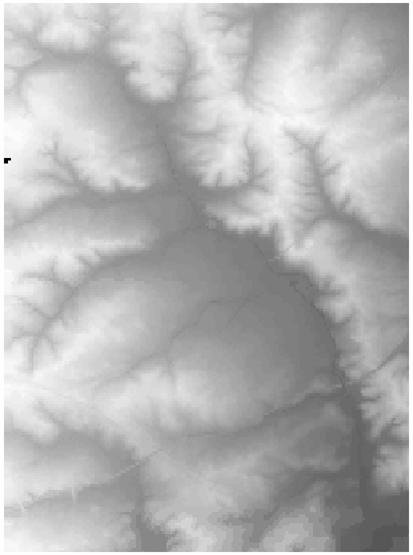
- *Simple kriging with varying local means example (modelling flooding information)*



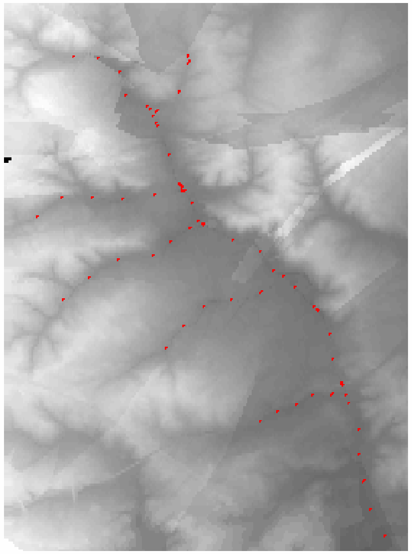
(a) Weighted mean



(a) Ordinary kriging



(a) Grid of flood mean values

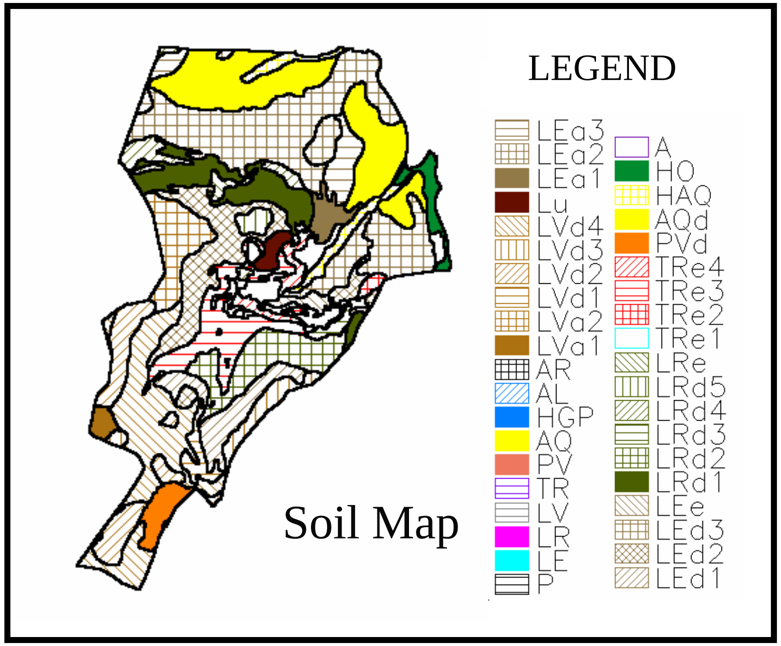


(a) SK with varying local means

The flood samples were found having a high correlation with topography (secondary data) (ρ ~.97)

Assessment of Spatial Uncertainty

- **Acounting for soft (imprecise) information in the estimation (indicator simulation with varying local means)**
  - The probability a priori for each grid cell is considered. (Soft information)
  - The probability a priori is updated considering the hard data (Bayesian approach)
  - Different estimates of the prob. a priori can be used, depending on the secondary information available.

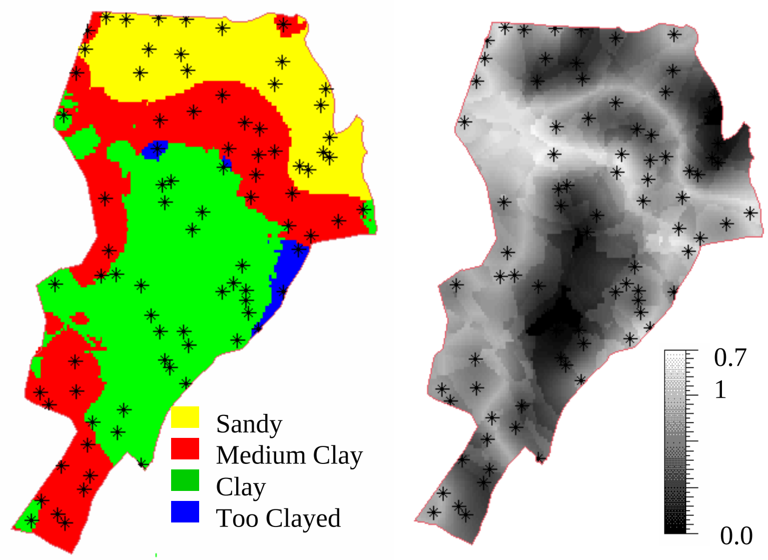


SOLO	ARENOSO	MÉDIO ARGILOSO	ARGILOSO	MUITO ARGILOSO
LVA1	0	0	1	0
LVA2	0	1	0	0
LVD1	0	0	1	0
LVD2	0	0	1	0
LVD3	0	1	0	0
LVD4	0	1	0	0
LU	0	0	1	0
LEA1	0	0.4	0.6	0
LEA2	0	1	0	0
LEA3	0	1	0	0
LED1	0	0	1	0
LED2	0	0	1	0
LED3	0	1	0	0
LEe	0	0	1	0
LRD1	0	0	0	1
LRD2	0	0	0.8	0.2
LRD3	0	0	0.7	0.3
LRD4	0	0	1	0
LRD5	0	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
Ho	0	0	1	0
A	0	0	1	0

Assessment of Spatial Uncertainty

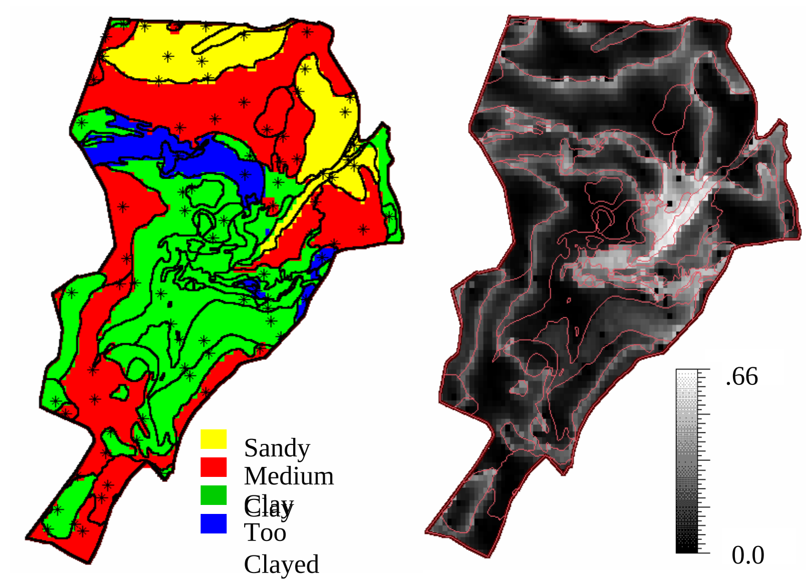
- **Acounting for secondary information in predictions**

Methodology 3: Indicator Simulation with varying local means



Modeling of soil texture
with hard data only

Modeling of soil texture
with hard + soft data



- More detailed information
- Candidate areas for resampling

Assessment of Spatial Uncertainty

- **Acounting for secondary information in predictions**

  - *Cokriging* – Considering Z_1 primary variable and Z_i other variables

$$Z_1^*(\mathbf{u}) - \mathrm{m}_1(\mathbf{u}) = \sum_{\alpha_1=1}^{\mathrm{n}_1(\mathbf{u})} \lambda_{\alpha_1}(\mathbf{u})\left[Z_1(\mathbf{u}_{\alpha_1}) - \mathrm{m}_1(\mathbf{u}_{\alpha_1})\right]$$
$$+ \sum_{i=2}^{N_v} \sum_{\alpha_i=1}^{\mathrm{n}_i(\mathbf{u})} \lambda_{\alpha_i}(\mathbf{u})\left[Z_i(\mathbf{u}_{\alpha_i}) - \mathrm{m}_1(\mathbf{u}_{\alpha_i})\right]$$

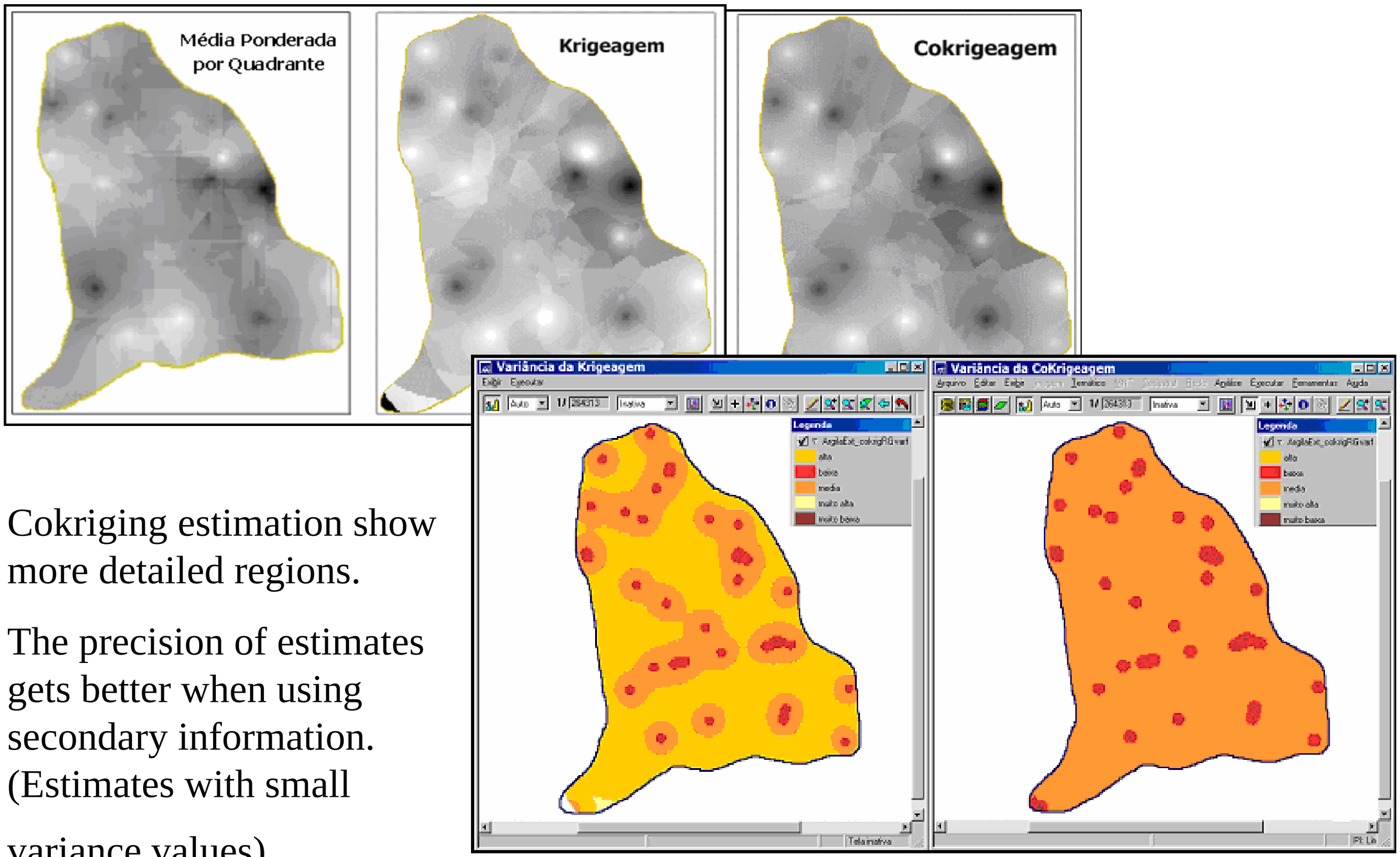
- Variants: simple, ordinary, with trends

- Requires inference of direct and cross semivariograms

- Useful if secondary variables are better sampled and there is reasonable correlation among the variables.

Assessment of Spatial Uncertainty

- **Acounting for secondary information in predictions** *Cokriging (example)*



Cokriging estimation show more detailed regions.

The precision of estimates gets better when using secondary information. (Estimates with small

variance values)

Assessment of Spatial Uncertainty

Summary and Conclusions

- Local Uncertainties based on estimatings with kriging (and variants) tend to smooth the attribute variability.
- Spatial Uncertainty modeling can be acomplished from sequential simulation approaches
- Spatial Models should be used when the histogram and the Covariance of the samples have to be reproduced
- Secondary variables can be used to improve the uncertainty models generated with geostatistics

Assessment of Spatial Uncertainty

END

of Presentation