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

Master of Science in Geospatial Technologies

Geostatistics

Carlos Alberto Felgueiras

carlos@dpi.inpe.br

cfelgueiras@isegi.unl.pt

September 20, 2007

Presentation

Carlos Alberto Felgueiras
carlos@dpi.inpe.br

Brazilian National Institute for Space Researches

- Bachelor in Electronic Engeneering (UNICAMP)
- MsC in Applied Computation (Digital Terrain Modeling - INPE)
- PhD in Applied Computation (Uncertainties using Indicator Geoestatistical Approaches - INPE)
- Manager of the development of the SPRING project (Image Processing Division – INPE)

Geostatistics – Program of the course

1. Introduction, Initial Concepts and Motivation (20/09)
2. Exploratory Data Analysis (EDA/ESDA) (27/09 and 04/10)
   i. Univariate Description
   ii. Bivariate Description
   iii. Spatial Description
   iv. Spatial Continuity Analysis (Variograms)
   v. Spatial Isotropy/Anisotropy
3. Deterministic Estimation Procedures (DEP) (11/10)
   i. Triangulation (TIN)
   ii. Local Sample Mean
   iii. Inverse Distance Methods (IDW)
   iv. Examples
   v. Problems with DEPs

Geostatistics – Program of the course

4. Linear Geoestatistical Estimation (18/10)
   i. Simple, Ordinary and Universal Kriging
   ii. Cokriging
   iii. Validation and CrossValidation
   iv. Advantages of using Geoestatistical Estimation
   v. Examples
5. Stochastic Simulation and Cosimulation (25/10)
6. Indicator Geostatistical Approaches (continuous and categorical information) (08, 15 and 22/11)
   i. Indicator Estimation for continuous variables (C.V.)
   ii. Assessment of Local Uncertainties for C.V.s
   iii. Indicator Estimation for Discrete Variables (D.V.)
   iv. Assessment of Local Uncertainties for D.V.s

Geostatistics – Program of the course

v. Indicator Simulation for CVs and DVs
vi. Assessment of Global Uncertainty
vii. Account for Secondary information
viii. Examples

7. Advantages of using Indicator Approaches (29/11)
8. Advanced Topics (06/12)
    1. Decision Making in the face of Uncertainty
    2. Propagation of uncertainties in spatial modeling
    3. Spatial/Temporal Geostatistics
9. Final Remarks, Final Test and Presentations (13/12)

Geostatistics – References

Basic Bibliography

Burrough, P. A., 1986. *Principles of Geographical Information Systems for Land Resources Assessment*. Clarendon Press – Oxford – London.

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

Cressie, N., 1991. Statistics for Spatial Data. John Wiley and Sons, New York, USA.

Deutsch, C. V.; Journel, A. G., 1998. *Geostatistical Software Library and User's Guide*. Oxford University Press, New York, USA.

Felgueiras, C. A., 1999 *Modelagem ambiental com tratamento de incertezas em sistemas de informação geográfica: o paradigma geoestatístico por indicação*. 165p. PhD Dissertation in Applied Computer – Instituto Nacional de Pesquisas Espaciais, São José dos Campos, Available at http://www.dpi.inpe.br/teses/felgueiras , Dec,1999.

Goovaerts, P., 1997. *Geostatistics for Natural Resources Evaluation*. Oxford University Press, Inc, New York, USA.

Geostatistics – References

Basic Bibliography

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

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

Internet: http://www.ncgia.ucsb.edu/giscc/units/u128/u128_f.html,

http://en.wikipedia.org/wiki/Geostatistics .....

Soares, A. 2000. *Geoestatística para as Ciências da Terra e do Ambiente*. Instituto Superior de Tecnico, IST Press. Lisboa, Portugal.

SPRING – Sistema de Processamento de Informações Geográficas – www.dpi.inpe.br/spring . Divisão de Processamento de Imagens –INPE-Brasil.

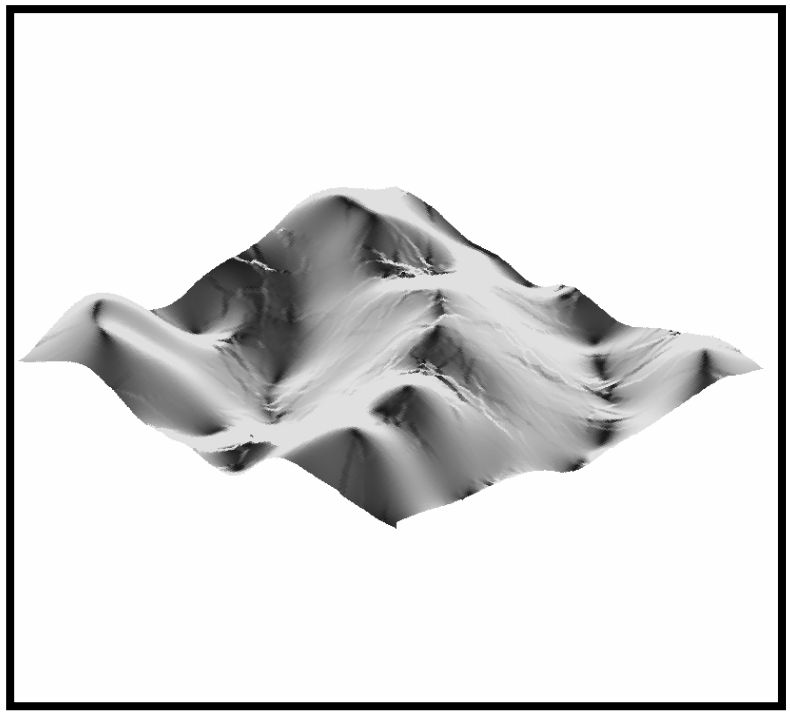
Geostatistics – Dinamic of the course

- Theoretical Classes with examples, mainly in SPRING
- Labs
- 2 Tests
- Student Presentations
- What software will be used? Free Choice
  - ArcView, ArcGIS module, GSLIB, R, SAS, SPRING, VarioWin, ...

Geostatistics – Initial Concepts

What is the Geostatistics? Some Definitions

Geostatistics is concerned with "the study of phenomena that fluctuate in space" and/or time. Geostatistics offers a collection of deterministic and statistical tools aimed at understanding and modeling spatial variability. (Deutsch and Journel)



Geostatistics – Initial Concepts

What is Geostatistics? Some Definitions

Geostatistical offers a way of describing the spatial continuity that is an essential feature of many natural phenomena and provides adaptations of classical regression techniques to take the advantage of this continuity. (Isaaks and Srivastava)

Geostatistics provides a set of statistical tools for incorporating the spatial and temporal coordinates of observations in data processing. (Goovaerts)

Geostatistics – Initial Concepts

Geostatistics = Theory of regionalized variables

Statistical tools for analyzing space/time information

Geostatistics – Application Fields

Mining, petroleum, geophysics, geochemistry soil science, forestry, agriculture (esp. in precision farm), environmental control, landscape ecology, remote sensing, hydrology, oceanography, meteorology, ...

Geostatistics – Initial Concepts

What is Geostatistics? http://en.wikipedia.org/wiki/Geostatistics

Geostatistics explains not only its applications within Geographic Information Systems but also the numerous applications of mathematical analysis on varied spatial datasets, the most prominent being a digital elevation model, from which any number of analyses may be derived.

Geostatistics is also applied in varied branches of human geography, particularly those involving the spread of disease (epidemiology), the practice of commerce and military planning (logistics), and the development of efficient spatial networks.

Geostatistics – Initial Concepts

Sampling

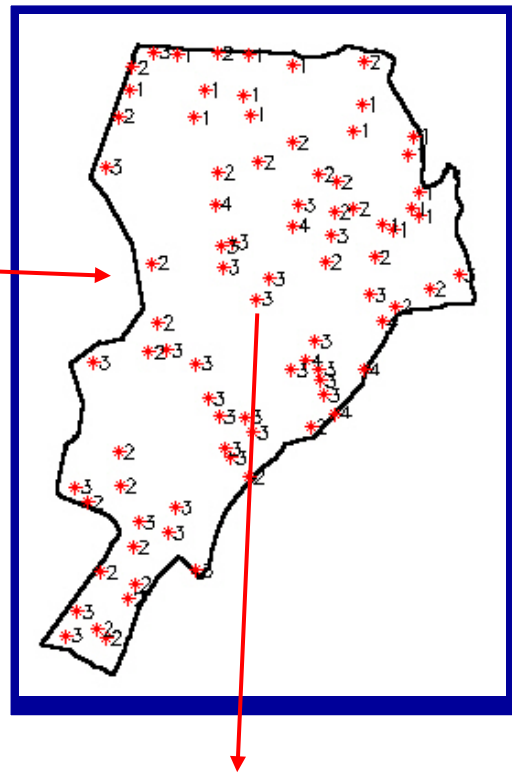
Spatial sampling involves determining a limited number of locations in a geo-space for faithfully measuring phenome-na that are subject to dependency and heterogeneity.

Dependency suggests that since one location can predict the value of another location, we do not need observations in both places.

Heterogeneity suggests that this relation can change across space, and therefore we cannot trust an observed degree of dependency beyond a region that may be small.

Basic spatial sampling schemes include random, clustered and systematic.

(http://en.wikipedia.org/wiki/Spatial_analysis



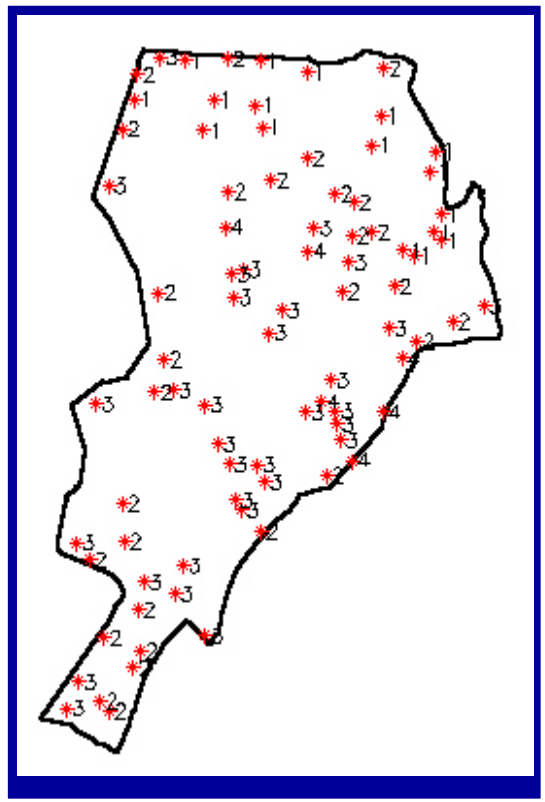
Each sample point α is represented by its (x,y,z) coord.

(x,y) is the 2-d space location

z is the attribute value

Geostatistics – Initial Concepts

• **Earth sciences data are typically distributed in space and/or time. Knowledge of an attribute value is of interest only if location and/or time of measurement are known and accounted for in the data analysis. (Goovaerts)**



• **Continuous x Categorical Variables (Attribute representations)**

Continuous variables: may take on any value within a finite or infinite interval. Examples: elevation, temperature, mineral grade, pollutant concentration, weight, ...)

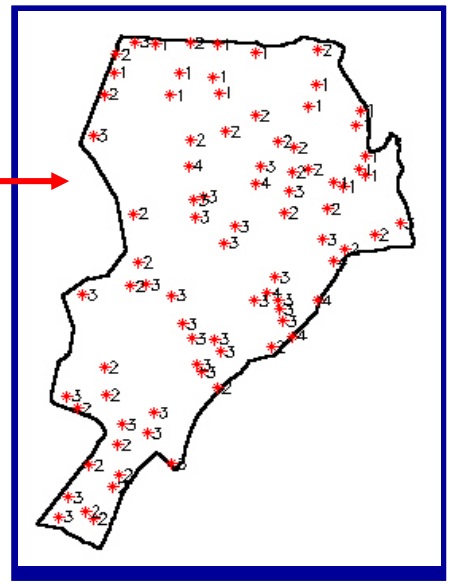
Categorical variables examples: may take discrete values or categories. Examples: type of vegetation, class of soil, number of population, ….)

Geostatistics – Initial Concepts

Input:

A Set of Points sampled sparsely distributed in a spatial region

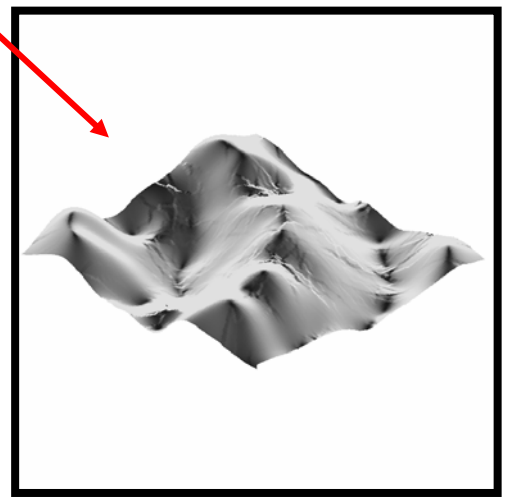
Each point represents a measurement of a variable (spatial attribute) that occurs in that spatial location.



Output:

A SPATIAL DATA MODEL

(computer/mathematical representation) that allows one to perform estimations and/or simulations for attribute values at spatial/temporal locations not sampled.

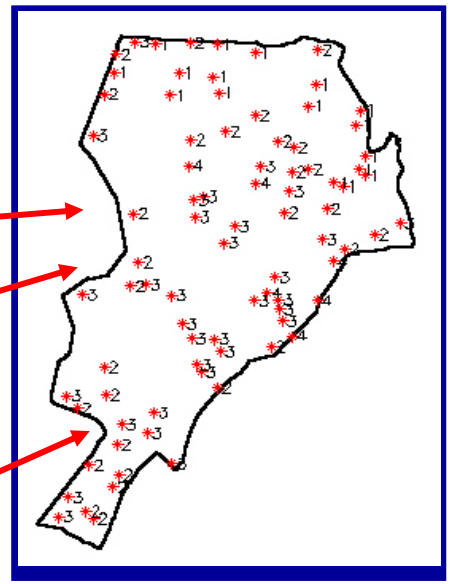


DETERMINISTIC X STOCHASTIC MODELS

Geostatistics – Initial Concepts

QUESTIONS ???????

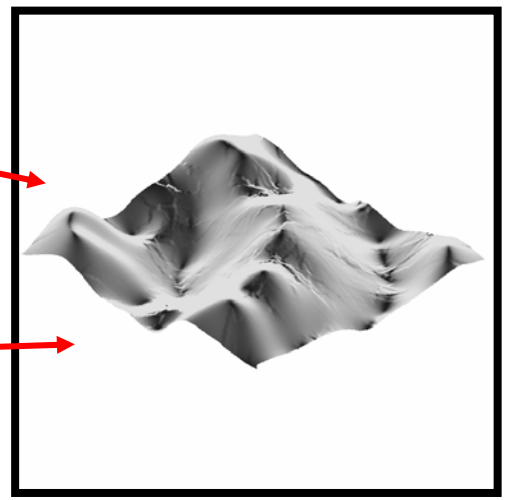
How many samples we need as input?
How they must be distributed in space?



Are the Samples a good representation of the attribute variation in my region of interest?

What are the problems with undersampling, oversampling and clusters?

Can my Sample Set lead to a good spatial model, with the geostatistical tools?



What is a good Data Model? How to measure the quality of the output Data Model?

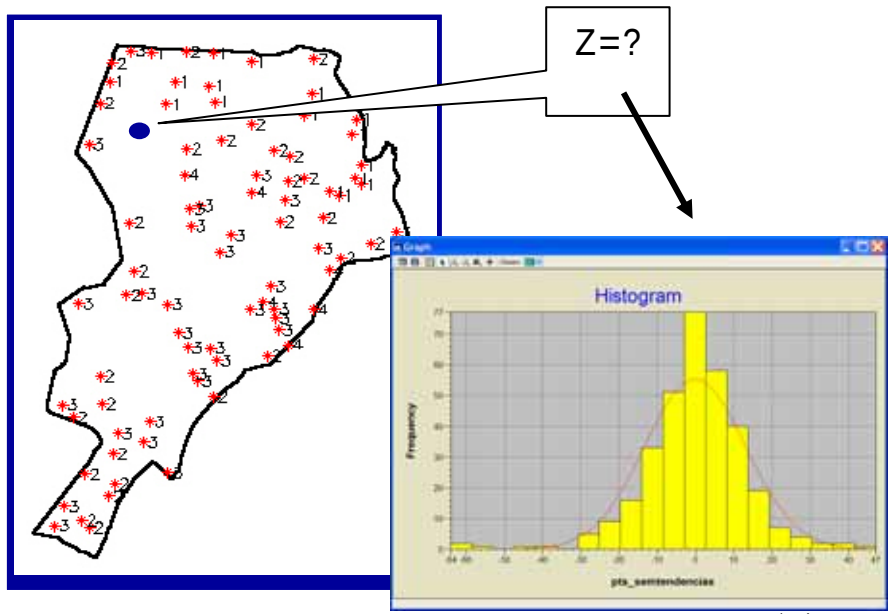
Geostatistics – Initial Concepts

- **Basic hypothesis for predictions in geographic space:**

A fundamental concept in geography is that nearby entities often share more similarities than entities which are far apart. This idea is often labelled 'Tobler's first law of geography' and may be summarized as "everything is related to everything else, but nearby objects are more related than distant objects". **Correlation an AutoCorrelation concepts.**

Geostatistical Paradigm

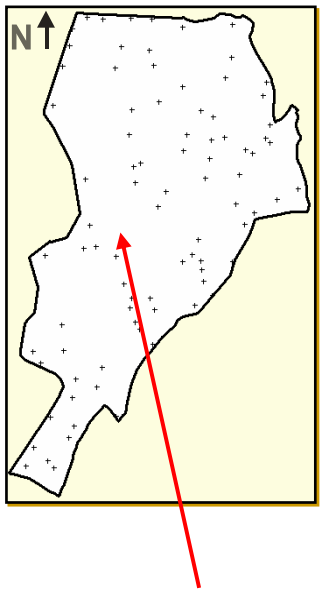
The basic paradigm of predictive statistics is to characterize any unsampled (unknown) value *z* as a Random Variable (RV) Z, the probability distribution of which models the uncertainty about *z*. (Deutsch and Journel)



Geostatistics – Initial Concepts and Motivation

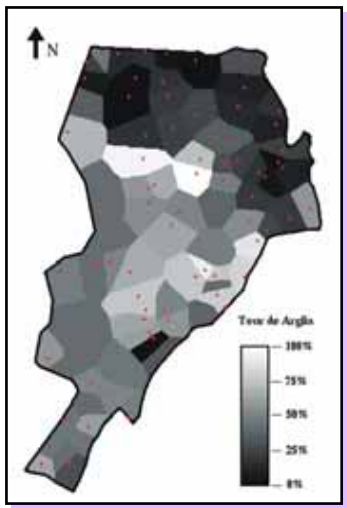
Why using geostatistical procedures ? Data Modeling

Study Area

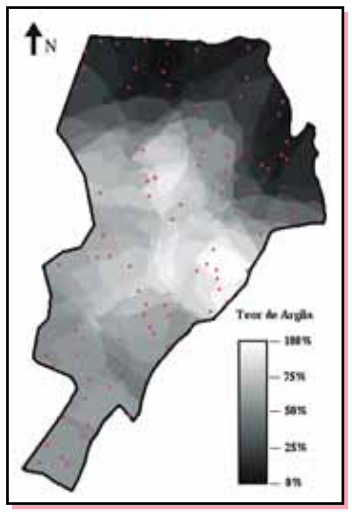


Sample Set

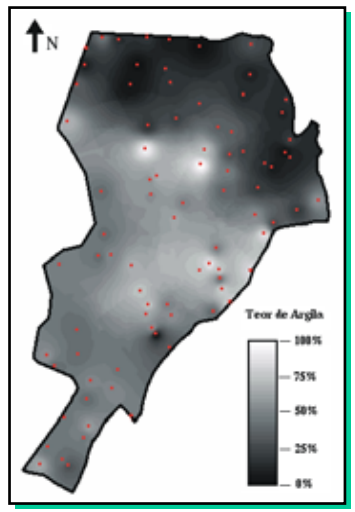
Deterministic Procedures



Nearest Neighbours

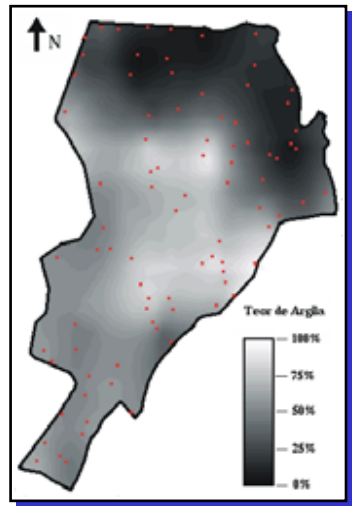


Simple Mean



Inverse Distance Weighted

Geostatistics

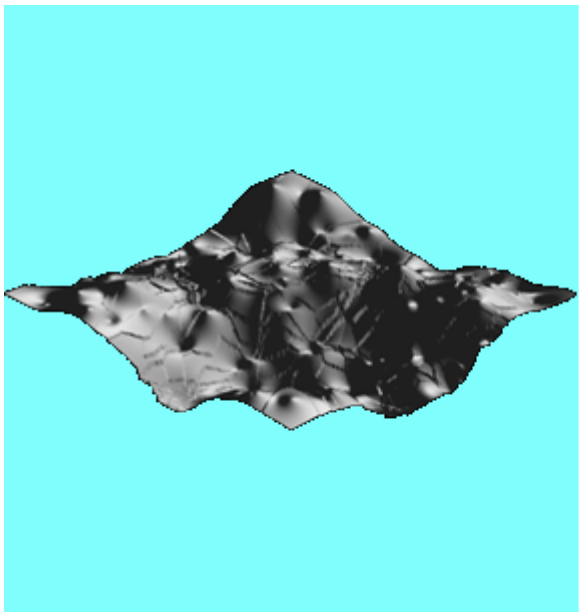


Kriging

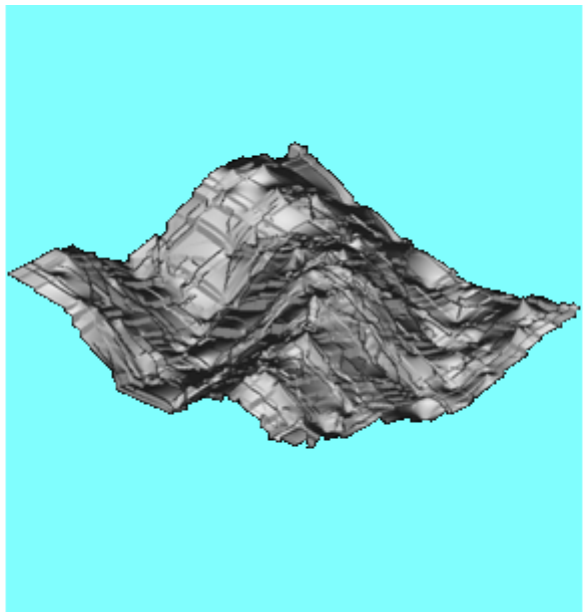
Geostatistics – Initial Concepts and Motivation

Why using geostatistical procedures ? Data Modeling

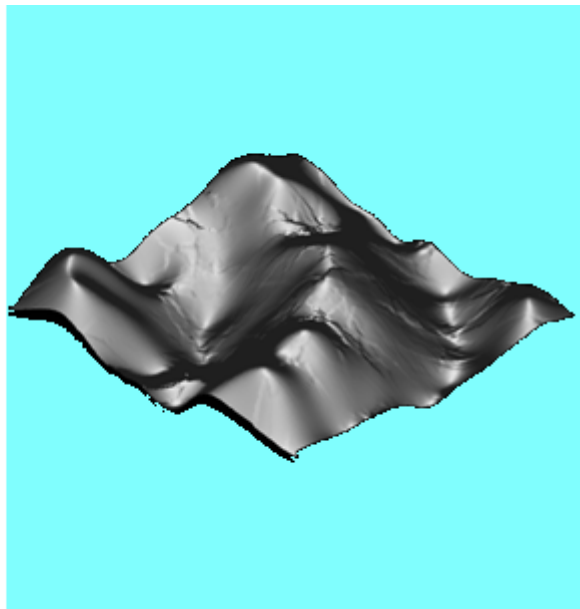
• When the primary data are sparse the geostatistical approaches perform better.
Example: Using only 50 random samples of a sine(d) function



(a) Simple mean



(b) Inverse Distance Weighted

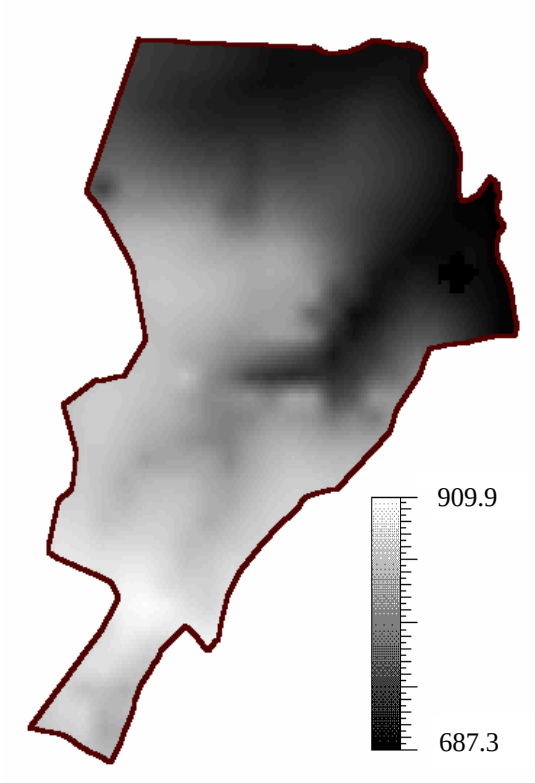


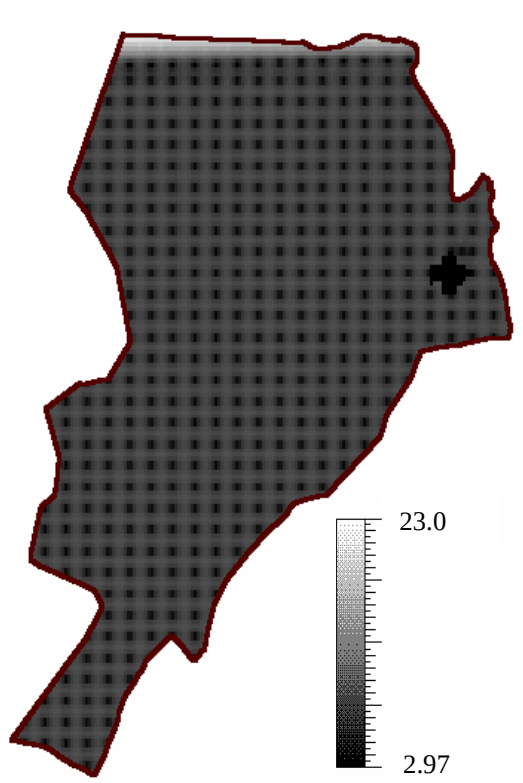
(c) Ordinary kriging

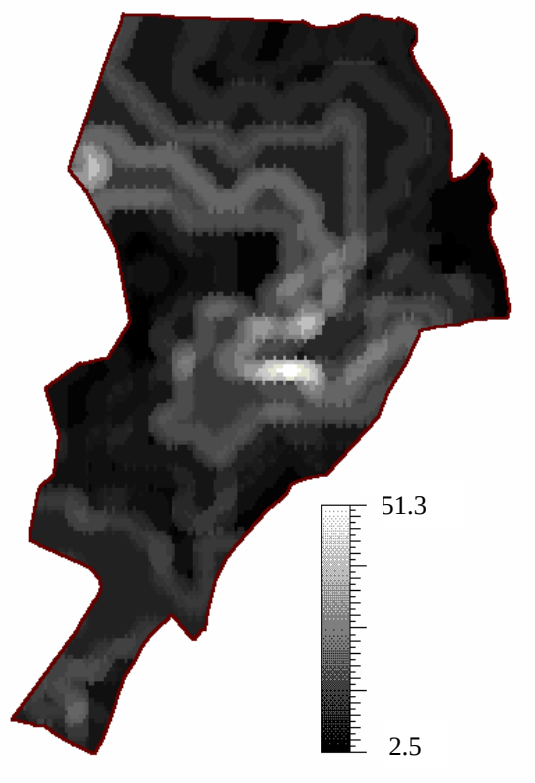
Geostatistics – Initial Concepts and Motivation

Why using geostatistical procedures ? Data Modeling + Uncertainties

- Geostatistical Indicator approaches allow estimation of continuous variables along with related uncertainties. Examples:





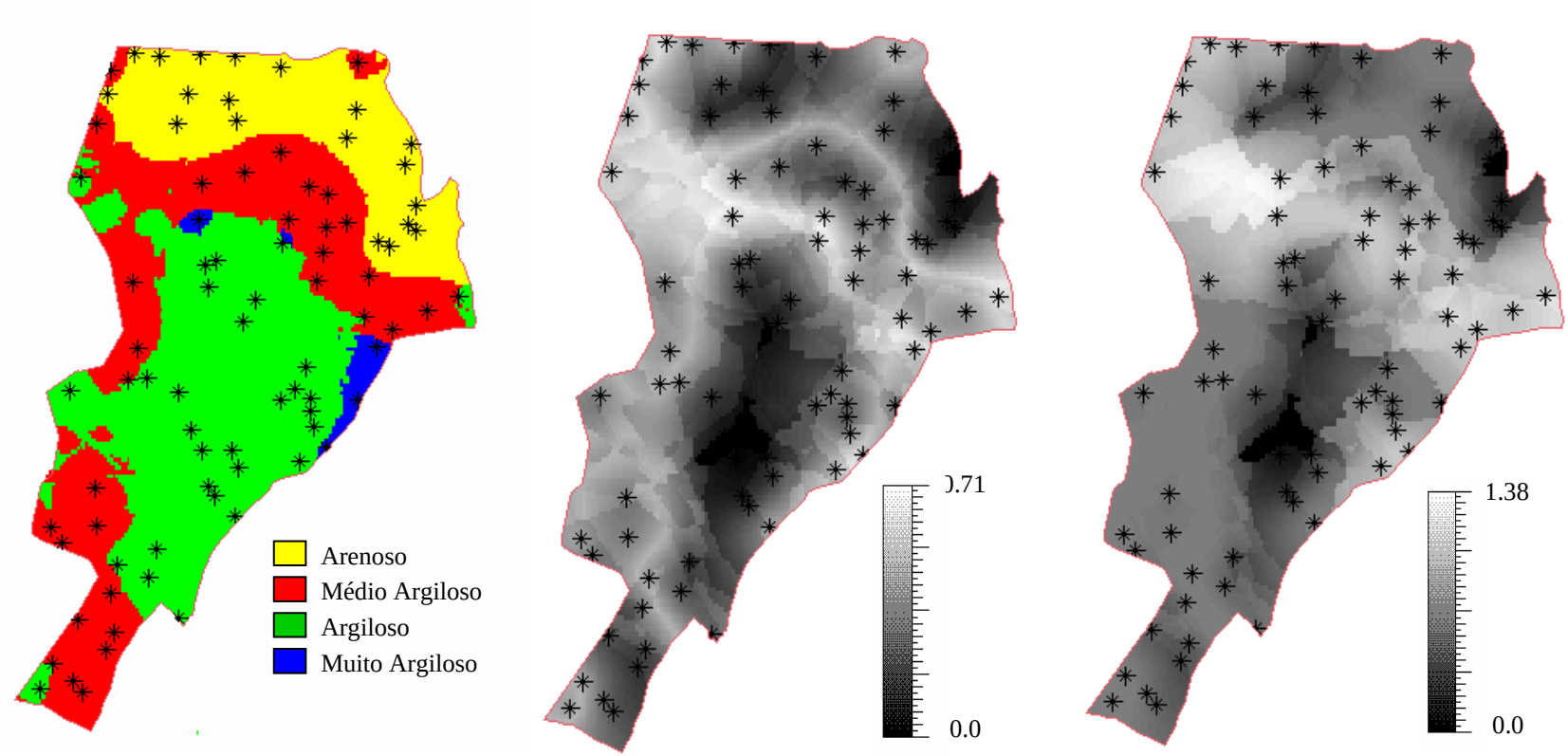


Important : The uncertainty maps can be used to qualify the estimations

Geostatistics – Initial Concepts and Motivation

Why using geostatistical procedures ? Categorical Data Modeling + Uncertainty

- Geostatistical Indicator approaches allow estimation of categorical variables along with related uncertainties. Example:

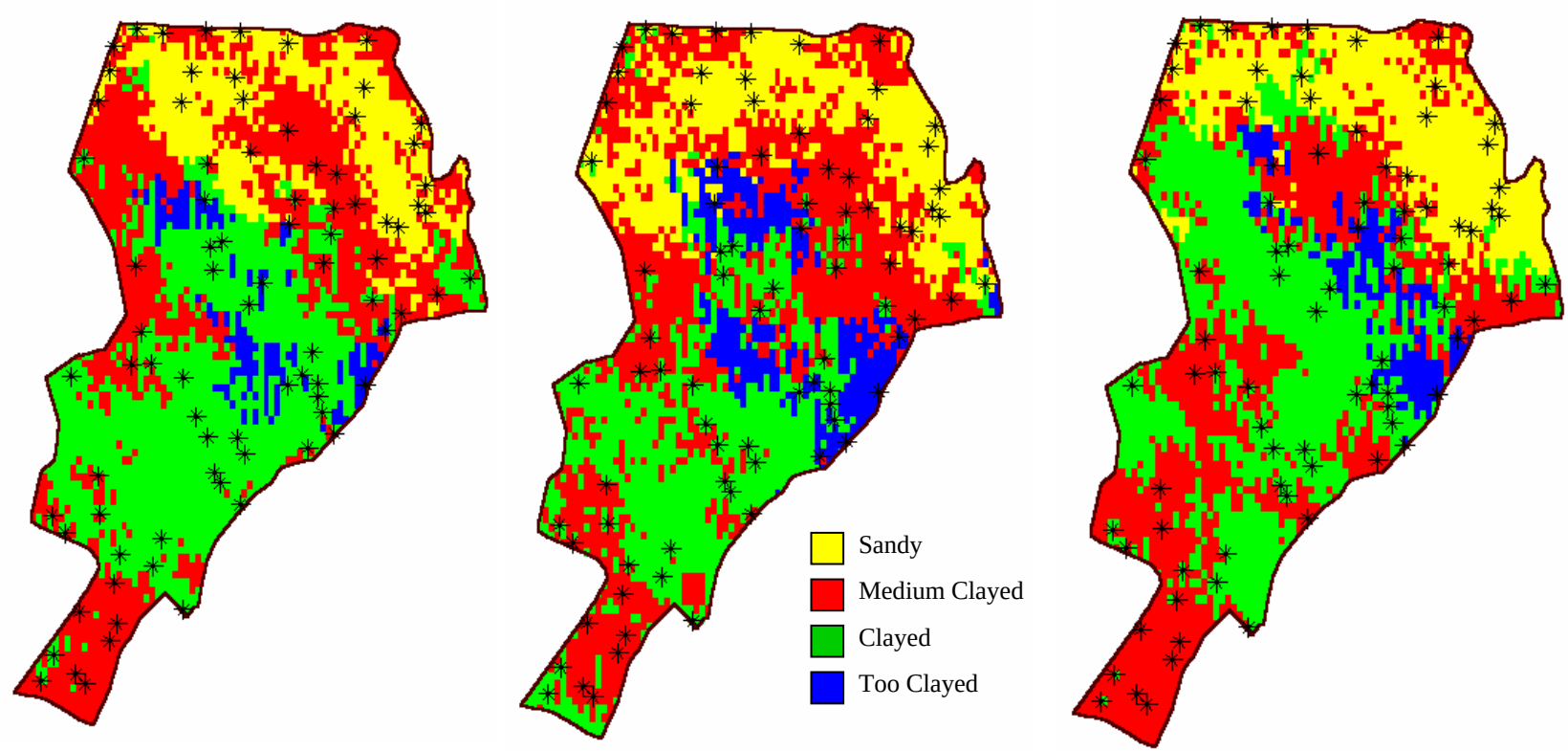


Important : The uncertainty maps can be used to qualify the estimations

Geostatistics – Initial Concepts and Motivation

Why using geostatistical procedures ? Categorical Data Realizations

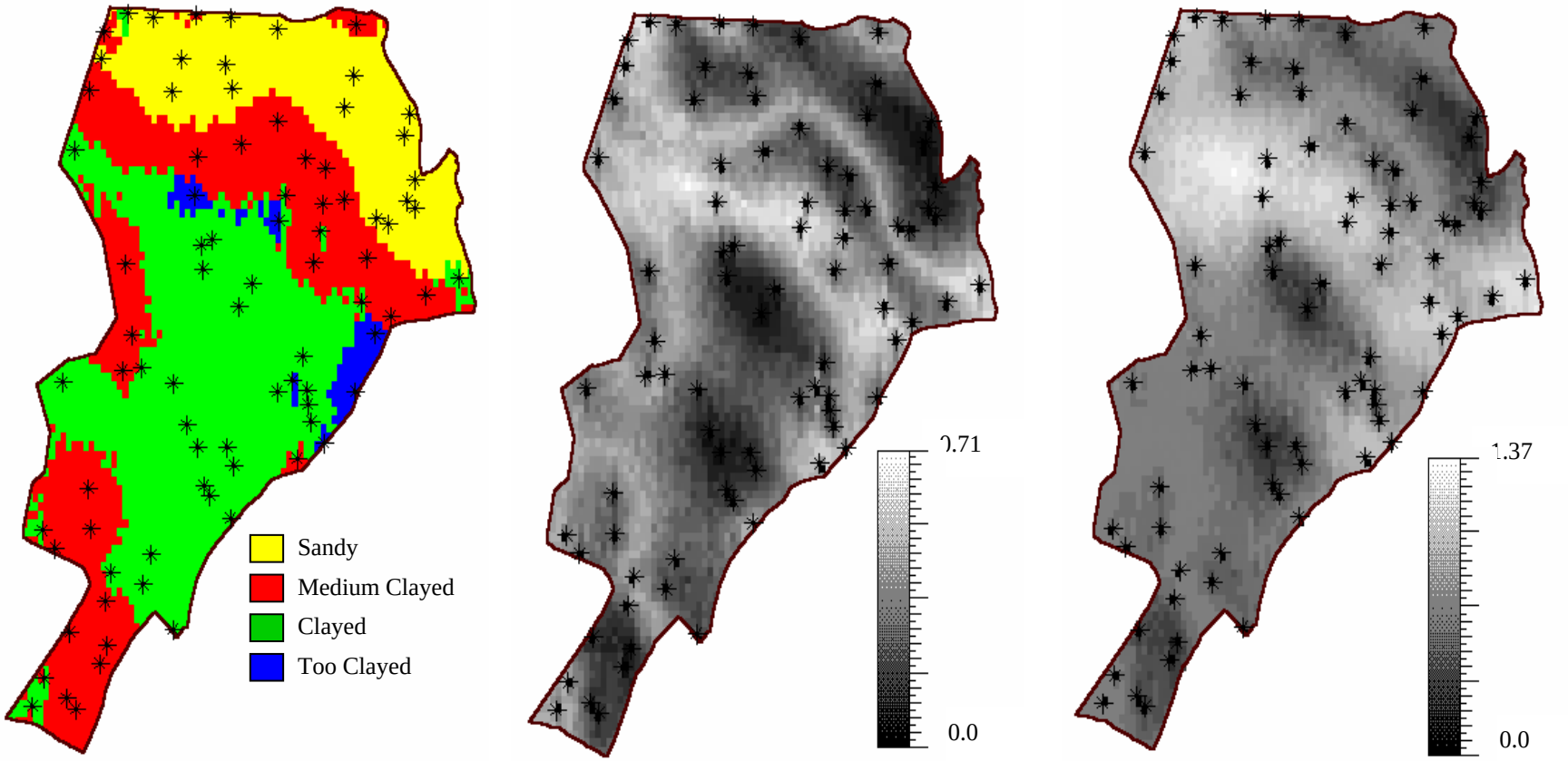
- Geostatistical Indicator approaches allow simulation of categorical variables along with realization maps. Example:



Geostatistics – Initial Concepts and Motivation

Why using geostatistical procedures ? Categorical Data Realizations

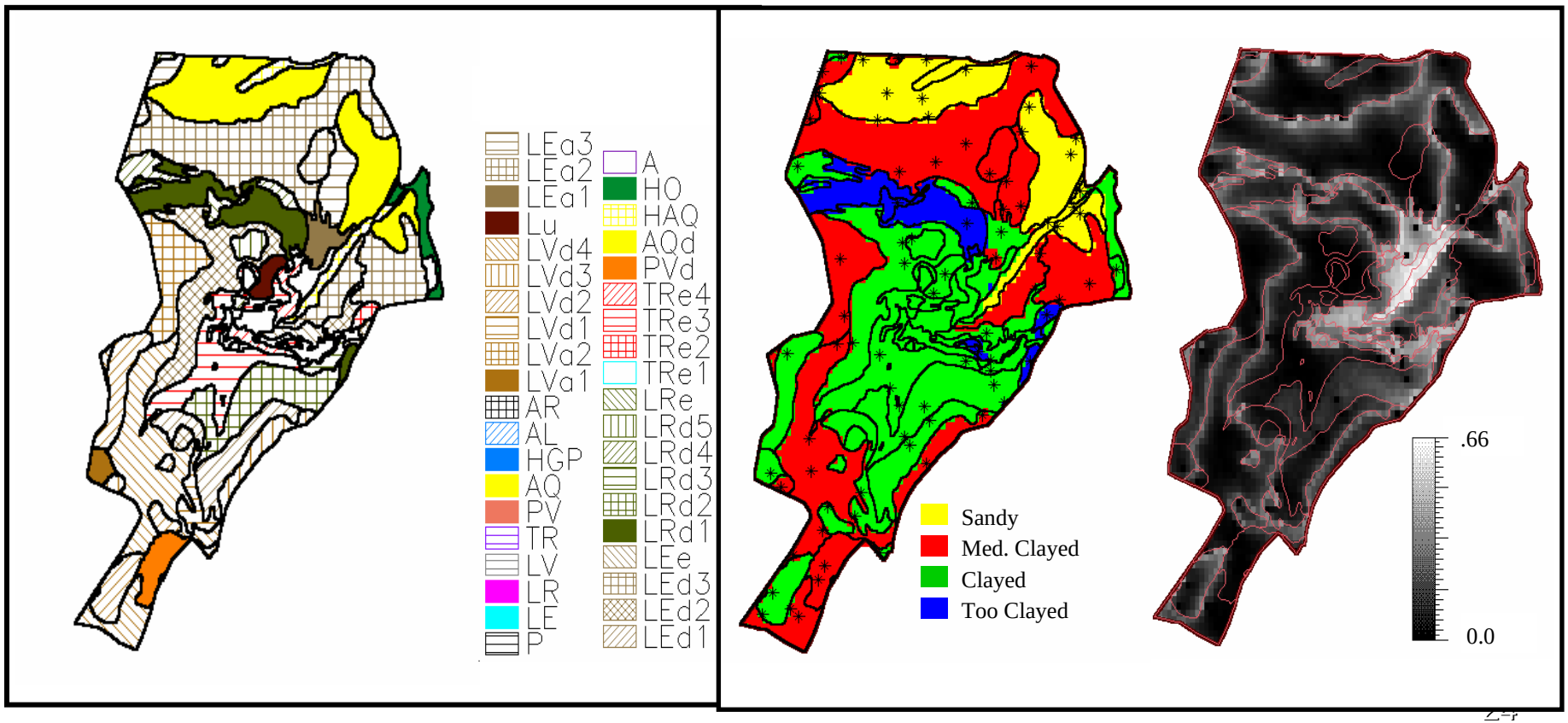
- Geostatistical Indicator approaches allow simulation of continous and categorical variables along with related uncertainties. Example:



Geostatistics – Initial Concepts and Motivation

Why using geostatistical procedures ? Account for secondary information

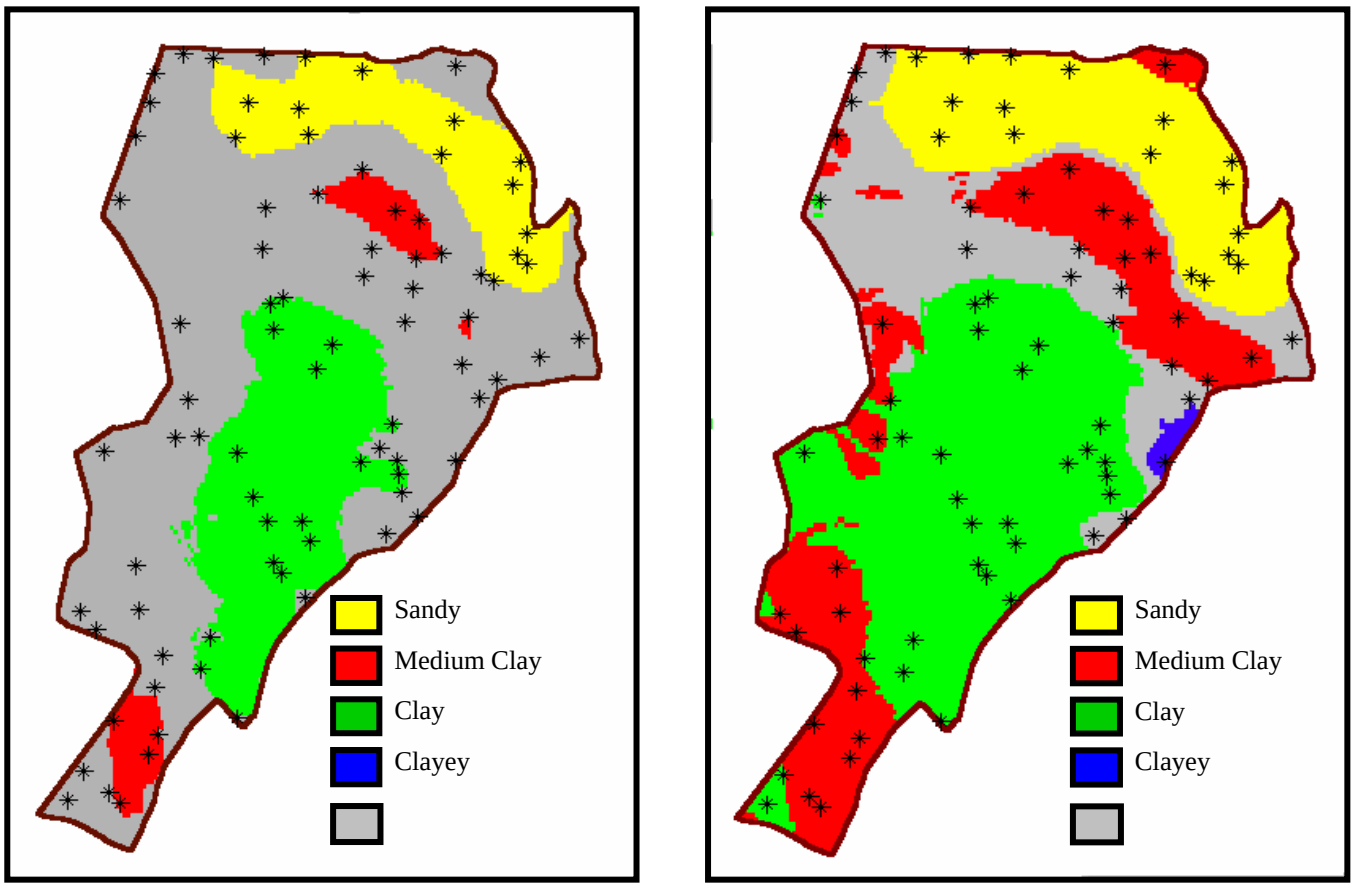
- Geostatistical Indicator approaches allow the incorporation of secondary information as a soft variable to enhance the accuracy of inferences of the hard data. Example:



Geostatistics – Initial Concepts and Motivation

Why using geostatistical procedures ?. Decision Making in the face of Uncertainty, using the local uncertainty to constrain the classification.

Example:



Geostatistics – Initial Concepts and Motivation

Summary

The Geostatistical Process

1. *Exploratory Data Analysis* (Knowledge and Edition of the INPUT data)
2. *Structural Analysis -Spatial Continuity Modeling* (Autocorrelation Models)
3. *Predictions* (Local and Global Estimations/Realizations of Continuous and Categorical attributes at unvisited locations)
4. *Assessment of Uncertainties* (Local and Spatial Uncertainty about unsampled values)

Results are used in Spatial Modeling in a GIS Environment

Geostatistics – Initial Concepts and Motivation

The Geostatistical Process

Region of Study

Exploratory Analysis

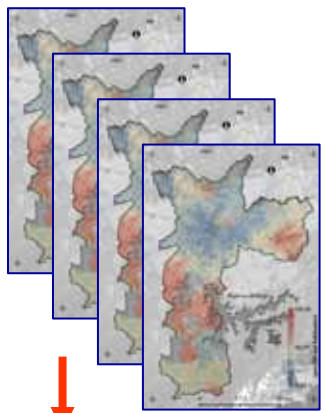
Structural Analysis

Interpolation kriging

Conditional Simulation

Estimated Surface of the phenomenon of interest

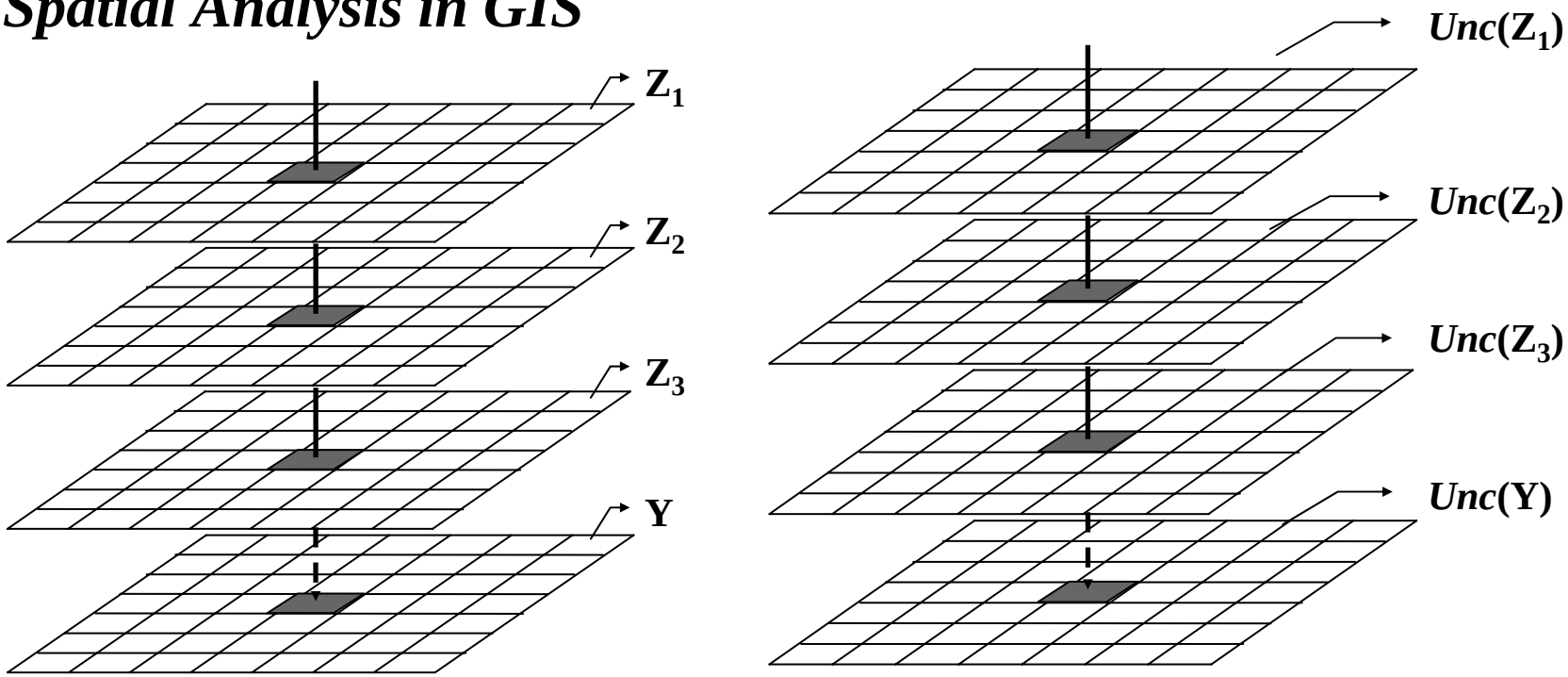
Realizations Fields



- **Derived Scenarios**
- **Uncertainty Maps**

Geostatistics – Motivation

- ***Spatial Analysis in GIS***



Spatial Model: $Y(\mathbf{u}) = g(Z_1(\mathbf{u}),...,Z_n(\mathbf{u}))$ for n inputs

The *Uncertainties* of the *Input representations* propagate to the the *Uncertainty* of the *Output* representation.

Geostatistics – Initial Concepts and Motivation

END of Presentation

Geostatistics – An Introduction to SPRING

Software Installation

Presentation

Data Modeling in SPRING

Data Importation

Visualization of Data

Handling Visualization