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

Master of Science in Geospatial Technologies

Geostatistics
Decision Making in the face of Uncertainty

Carlos Alberto Felgueiras
cfelgueiras@isegi.unl.pt

Decision making in the face of Uncertainty

Contents

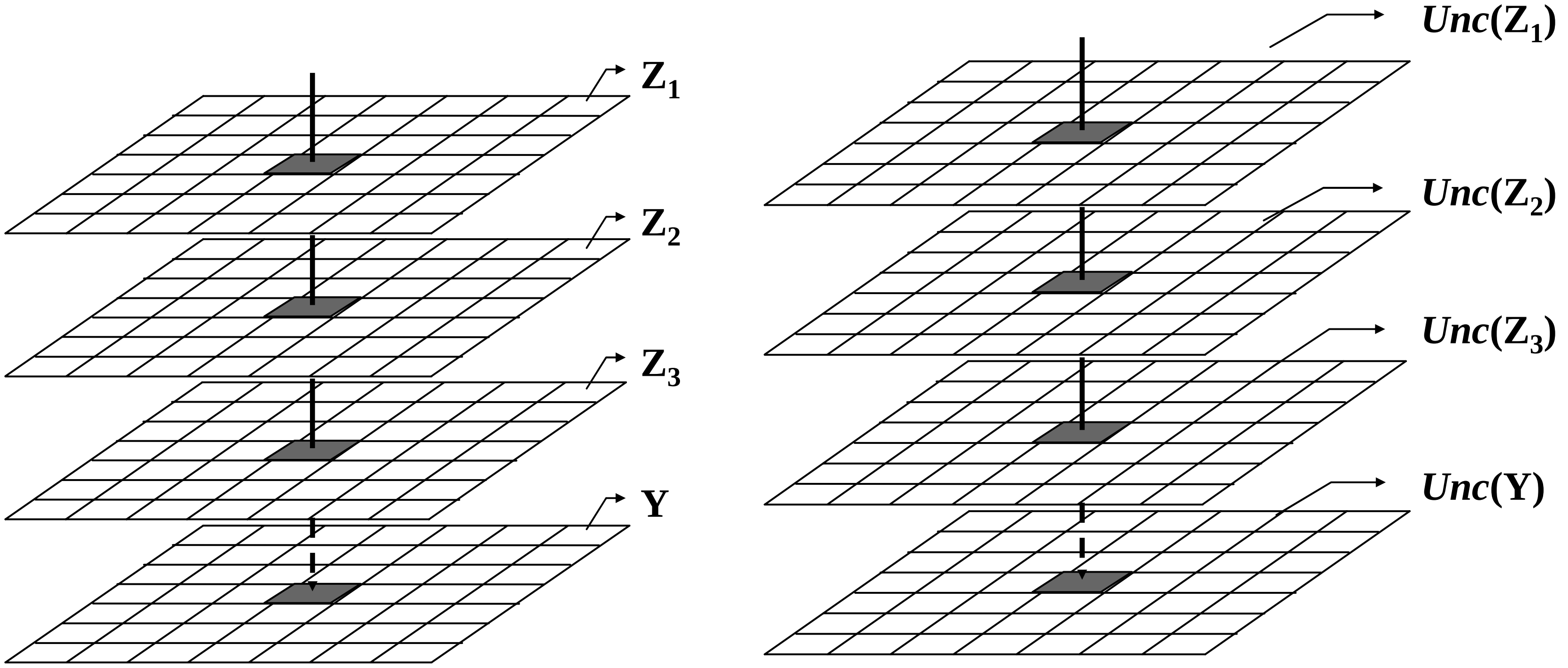
Decision making in the face of Uncertainty

Introduction

- Uncertainties of data modeling can be accomplished by geostatistics approaches
- Standard Kriging – kriging variance for multiGaussian models
- Indicator Kriging – local uncertainties (local cdf or pdf)
- Indicator Simulation – global uncertainties (global cdf or pdf)
- Uncertainties (cdfs or pdfs) qualify estimates
- Uncertainties should be accounted in decision making processes because the estimates are uncertain.
- How to use the uncertainties in GIS Analysis and Decision Makings ?

Decision making in the face of Uncertainty

Uncertainty Propagation on Spatial Modeling



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

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

Decision making in the face of Uncertainty

Uncertainty Propagation on Spatial Modeling

Heuvelink, 1998, presents 4 methods to assess the problem of "*Error Propagation in Environmental Modelling with GIS*".

1. **First Order Taylor Method**
2. **Second Order Taylor Method**
3. **Rosenblueth's Method**
4. **Monte Carlo's Method**

Decision making in the face of Uncertainty

Uncertainty Propagation on Spatial Modeling

1. *First Order Taylor Method* (Heuvelink, 1998)

The first order expansion of the Taylor series, around the mean vector of the *n* input variables, $\vec{\mu}_Z = \left(\mu_{Z_1}, \dots, \mu_{Zn}\right)$, is given by:

$$Y = g(Z) = g\left(\vec{\mu}_Z\right) + \sum_{i=1}^{n} \left\{ \left(Z_i - \mu_{Z_i}\right) \cdot \left(\frac{\partial g}{\partial z_i} \left(\vec{\mu}_Z\right)\right)\right\} + \text{residue}$$

- From this expansion, and not considering the residue, one can obtain the following approximations for the mean and variance values of the output R.V. *Y*:

$$\mu_Y = E(Y) \approx g\left(\vec{\mu}_Z\right)$$

$$\sigma_Y^2 \approx \sum_{i=1}^{n}\sum_{j=1}^{n} \left[\frac{\partial g}{\partial z_i}\left(\vec{\mu}_Z\right) \cdot \frac{\partial g}{\partial z_j}\left(\vec{\mu}_Z\right) \cdot \sigma_{Z_i} \cdot \sigma_{Z_j} \cdot \rho_{ij} \right]$$

$$\sigma_Y^2 \approx \sum_{i=1}^{n} \left[\frac{\partial g}{\partial z_i}\left(\vec{\mu}_Z\right) \right]^2 \cdot \sigma_{Z_i}^2 \quad \text{when } \rho_{ij} = 0 \text{ for } i \neq j$$

Decision making in the face of Uncertainty

Uncertainty Propagation on Spatial Modeling

Considerations about the *First Order Taylor Method*

- Used only for operations with continuous attributes
- Aplicável somente à operações com atributos qualitativos.
- The mean value of the output Y is depends only of the means of the input values. The standard deviations do not affect the output mean value.
- The output variances depend on the standard deviations and correlations of the inputs. Also there is a dependency related to the partial derivatives over the input mean vectors. This method is applicable only to g functions continually differentiable.
- For independent input variables the correlation coefficient ρ_{ij} is equal 0 for i≠j and is equal 1 for i=j. In this case the formulation of the variance is simplified to:

$$\sigma_Y^2 = \sum_{i=1}^{n}\left[\frac{\partial g}{\partial z_i}\left(\vec{\mu}_Z\right)\right]^2 \cdot \sigma_{Z_i}^2$$

Decision making in the face of Uncertainty

Uncertainty Propagation on Spatial Modeling (Heuvelink, 1998)

2. Second Order Taylor Method

- Extension of the First Order Taylor method, including the term of the second order of the Taylor series.

$$Y = g(Z) = g(\vec{\mu}_Z) + \sum_{i=1}^{n} \left\{ \left(Z_i - \mu_{Z_i} \right) \cdot \left(\frac{\partial g}{\partial z_i} (\vec{\mu}_Z) \right) \right\}$$

$$+ \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left\{ \left(Z_i - \mu_{Z_i} \right) \cdot \left[Z_j - \mu_{Z_j} \right] \cdot \left(\frac{\partial^2 g}{\partial z_i \delta z_j} (\vec{\mu}_Z) \right) \right\} + \text{resídue}$$

- In this case the approximation for the output mean value is assessed by:

$$\mu_Y = E(Y) \approx g(\vec{\mu}_Z) + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \left\{ \rho_{z_i z_j} \sigma_{z_i} \sigma_{z_j} \frac{\partial^2 g}{\partial_{z_i} \partial_{z_j}} \right\}$$

- Important: The output mean value can differ of the g value applied to the input mean values.

Decision making in the face of Uncertainty

Uncertainty Propagation on Spatial Modeling (Heuvelink, 1998)

2. Second Order Taylor Method

- The approximation for the variance of *Y* is given by:

$$\sigma_Y^2 \approx \sum_{k=1}^{n}\sum_{l=1}^{n}\left\{\frac{\partial g}{\partial z_k}(\vec{\mu}_Z)\frac{\partial g}{\partial z_l}(\vec{\mu}_Z)\sigma_{Z_k}\sigma_{Z_l}\rho_{kl}\right\}$$

$$+\sum_{k=1}^{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{\left(E\left[(z_k-\mu_k)(z_i-\mu_i)(z_j-\mu_j)\right]-\sigma_{Z_k}\sigma_{Z_l}\rho_{ij}\right)\frac{\partial g}{\partial z_k}(\vec{\mu}_Z)\frac{\partial^2 g}{\partial z_i\,\delta z_j}(\vec{\mu}_Z)\right\}$$

$$+\frac{1}{4}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}\left\{\left(E\left[(z_i-\mu_i)(z_j-\mu_j)(z_k-\mu_k)(z_l-\mu_l)\right]-\rho_{ij}\sigma_{Z_i}\sigma_{Z_j}\rho_{kl}\sigma_{Z_k}\sigma_{Z_l}\right)\frac{\partial^2 g}{\partial z_i\,\delta z_j}(\vec{\mu}_Z)\frac{\partial^2 g}{\partial z_k\,\delta z_l}(\vec{\mu}_Z)\right\}$$

- The evaluation of the variance value requires the calculation of the 1º, 2º, 3º e 4º moments and the partial derivatives of first and second orders.

- Comparing with the first order method:
  - Approximations close to μ_z are better than those far from μ_z. Therefore the variance can be worst.
  - Second order method is better when the g is quadratic.

Decision making in the face of Uncertainty

Uncertainty Propagation on Spatial Modeling (Heuvelink, 1998)

3. Rosenblueth's Method

- It is equivalent to the first order Taylor method.
- Must be used when g is not continually differentiable over μ_z
- The output mean is evaluated by the relation:

$$\mu_{\mathrm{Y}} = \mathrm{E}(Y) \approx \sum_{\mathrm{k}=1}^{2^m} r_{\mathrm{k}}\, \mathrm{g}(d_{\mathrm{k}}),$$

$$\text{onde } d_{\mathrm{k}} = (d_{\mathrm{k}}, \ldots, d_{\mathrm{m}}) \; e \; d_{\mathrm{i}} = \mu_{\mathrm{i}} + \sigma_{\mathrm{i}} \; ou \; d_{\mathrm{i}} = \mu_{\mathrm{i}} - \sigma_{\mathrm{i}}$$

$$r_{\mathrm{k}} = \frac{1}{2^{\mathrm{m}}}\left(\sum_{\mathrm{i}=1}^{m-1} \sum_{\mathrm{j}=\mathrm{i}+1}^{m} \delta_{\mathrm{ij}}(\mathrm{k})\rho_{\mathrm{ij}} + 1 \right)$$

$$\delta_{\mathrm{ij}}(\mathrm{k}) = +1 \quad \text{quando} \quad d_{\mathrm{i}} = \mu_{\mathrm{i}} + \sigma_{\mathrm{i}} \; e \; d_{\mathrm{j}} = \mu_{\mathrm{j}} + \sigma_{\mathrm{j}}$$

$$\delta_{\mathrm{ij}}(\mathrm{k}) = +1 \quad \text{quando} \quad d_{\mathrm{i}} = \mu_{\mathrm{i}} - \sigma_{\mathrm{i}} \; e \; d_{\mathrm{j}} = \mu_{\mathrm{j}} - \sigma_{\mathrm{j}}$$

$$\delta_{\mathrm{ij}}(\mathrm{k}) = -1 \quad \text{quando} \quad d_{\mathrm{i}} = \mu_{\mathrm{i}} + \sigma_{\mathrm{i}} \; e \; d_{\mathrm{j}} = \mu_{\mathrm{j}} - \sigma_{\mathrm{j}}$$

$$\delta_{\mathrm{ij}}(\mathrm{k}) = -1 \quad \text{quando} \quad d_{\mathrm{i}} = \mu_{\mathrm{i}} - \sigma_{\mathrm{i}} \; e \; d_{\mathrm{j}} = \mu_{\mathrm{j}} + \sigma_{\mathrm{j}}$$

Decision making in the face of Uncertainty

Uncertainty Propagation on Spatial Modeling (Heuvelink, 1998)

3. *Rosenblueth's Method*

- The variance of *Y* is given by:

$$\sigma_Y^2 \approx \sum_{k=1}^{2^{\mathbf{m}}} \left\{ \mathbf{r}_k \left[g(\mathbf{d}_k) - \sum_{l=1}^{2^{\mathbf{m}}} \mathbf{r}_l \, g(\mathbf{d}_k)_l \right]^2 \right\}$$

- When m=1:

$$\mu_Y \approx \frac{1}{2}\left(g(\mu_z + \sigma) + g(\mu_z - \sigma)\right)$$

$$\sigma_Y^2 \approx \frac{1}{4}\left(g(\mu_z + \sigma) + g(\mu_z - \sigma)\right)^2$$

- Comparing with the first order Taylor method, this method uses a smoothed approximation of the first partial derivative over μ_z.

Decision making in the face of Uncertainty

Uncertainty Propagation on Spatial Modeling (Heuvelink, 1998)

4. Monte Carlo's method

- Compute Y repeatedly from z_i input values randomly sampled from its **joint distribution (requires as joint simulation if inputs are not independent)**.

- The method follows the below steps:
    - At each spatial location **u**
        - Repeat n time:
            - Draw a set of R realizations z_{i}, i=1,...,R.
            - Evaluate and store the output value $y=g(z_1,...z_m)$.
        - Define the cdf, or pdf, model from the n output values. Evaluate statistics, μ e σ^2 for ex., from the output values as:

$$\mu_Y(\mathbf{u}) = \frac{1}{n}\sum_{\text{i}=1}^{n} y_{\text{i}}(\mathbf{u}) \quad \text{and} \quad \sigma_Y^2(\mathbf{u}) = \frac{1}{n-1}\sum_{\text{i}=1}^{n}\left(y_{\text{i}}(\mathbf{u}) - \mu_Y(\mathbf{u})\right)$$

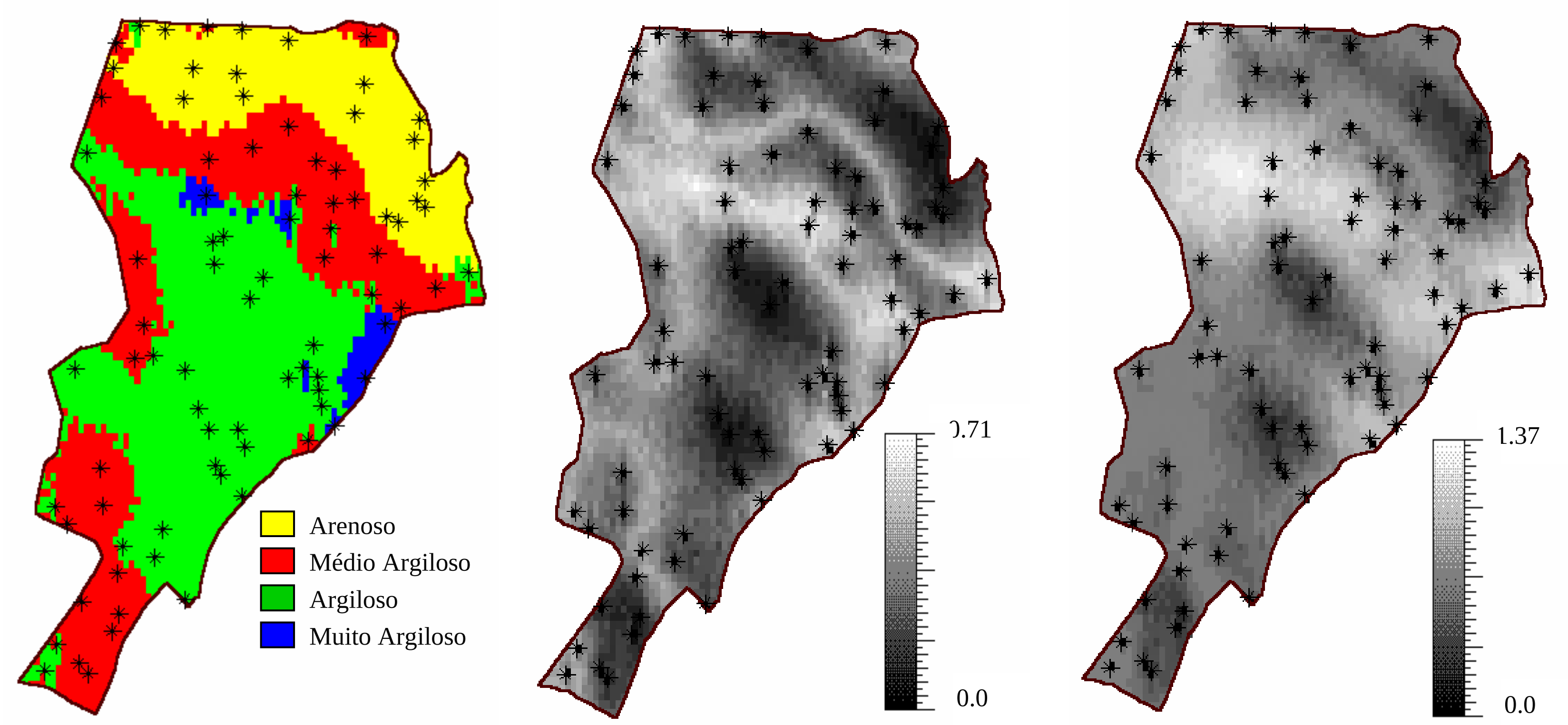
Decision making in the face of Uncertainty

Constraints applied to the estimates

- Deriving estimate maps taken into account uncertainties: Uncertainty constraints can be applied to estimates for categorical and continuous attributes.
- Continuous attribute
  - Maps with constrained areas: derived map keeps only the locations where the probability of the estimates are greater (or smaller) than a given probability value. A dummy value is set to the other locations.
  - Classified maps: derived map can be a classified map taken into account user defined probability intervals.
- Categorical attribute
  - Maps with constrained areas: derived map keeps only the locations where the probability of the estimates are greater (or smaller) than a given z value. A dummy z value is set to the other locations.
  - Classified maps: derived map can be a reclassified map taken into account user defined probability intervals.

Decision making in the face of Uncertainty

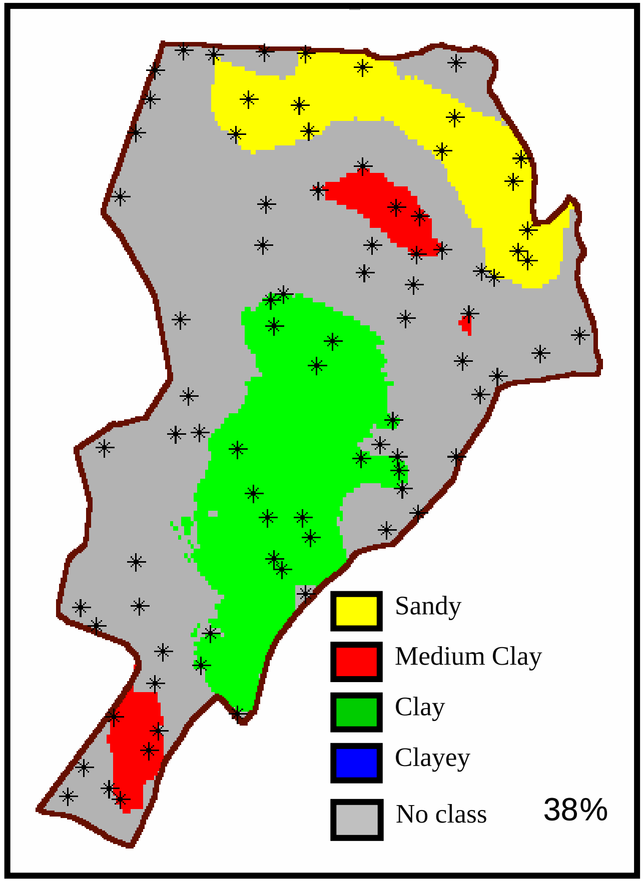
Example of Maps of Estimates and Uncertainties of categorical attributes from realizations.

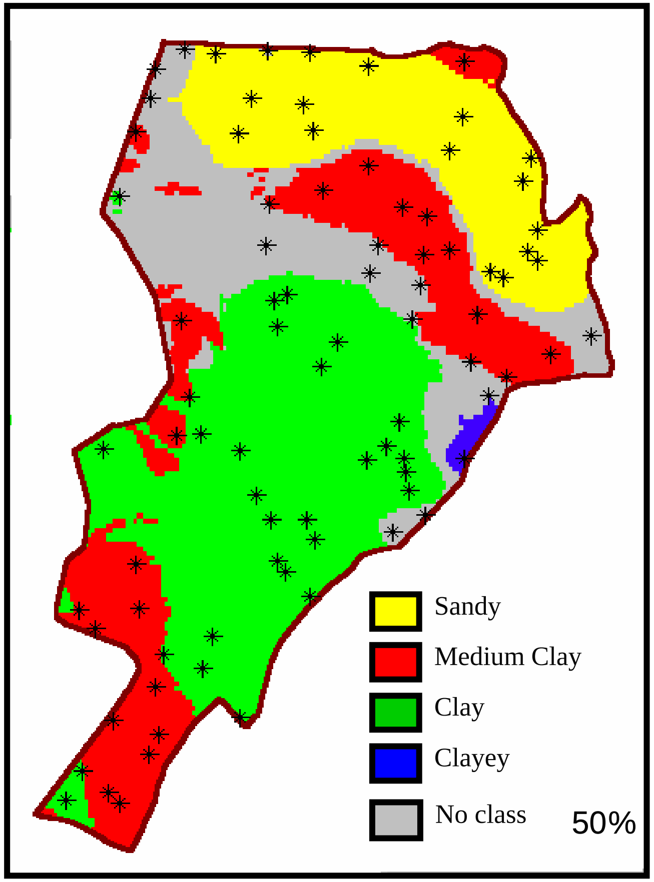


Decision making in the face of Uncertainty

Constraints applied to the estimates

Uncertainty constraints applied to estimates for categorical variables (example)





Decision making in the face of Uncertainty

Exceeding a probability threshold

• Defining a critical *z* value, z_c, of a continuous or a categorical attribute, it is possible to create *probability maps* from the uncertainty models of the RVs.

• For *continuous variables*: probabilities of your attribute value be lower (or greater) than the threshold *z* value.

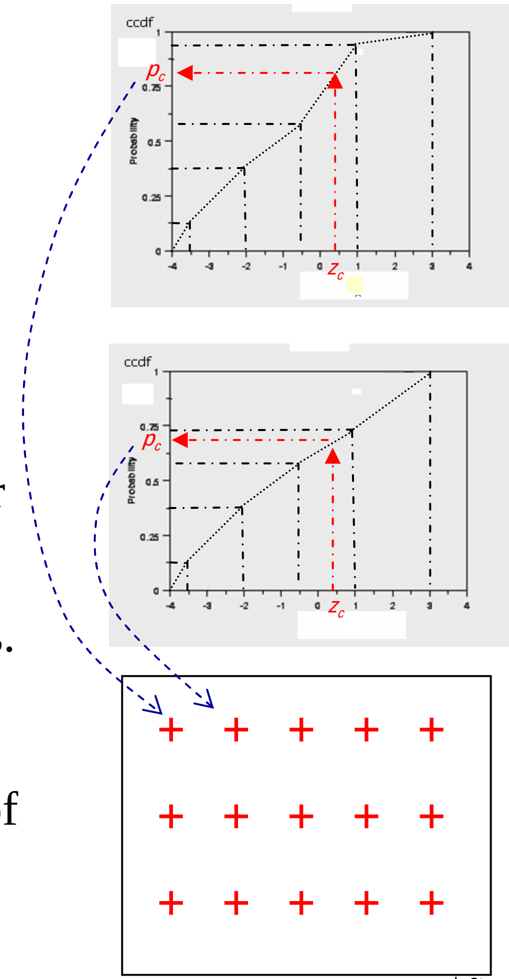
• For *categorical variables*: probabilities of your attribute value be equal to the *z* value (user defined class).

• The probability values of the probability maps can be ranked or classified (safe areas, hazardous areas, suitable regions, …)

• The results are taken into account in decision making processes.

• Goovaerts example: Areas contaminated by Cadmium

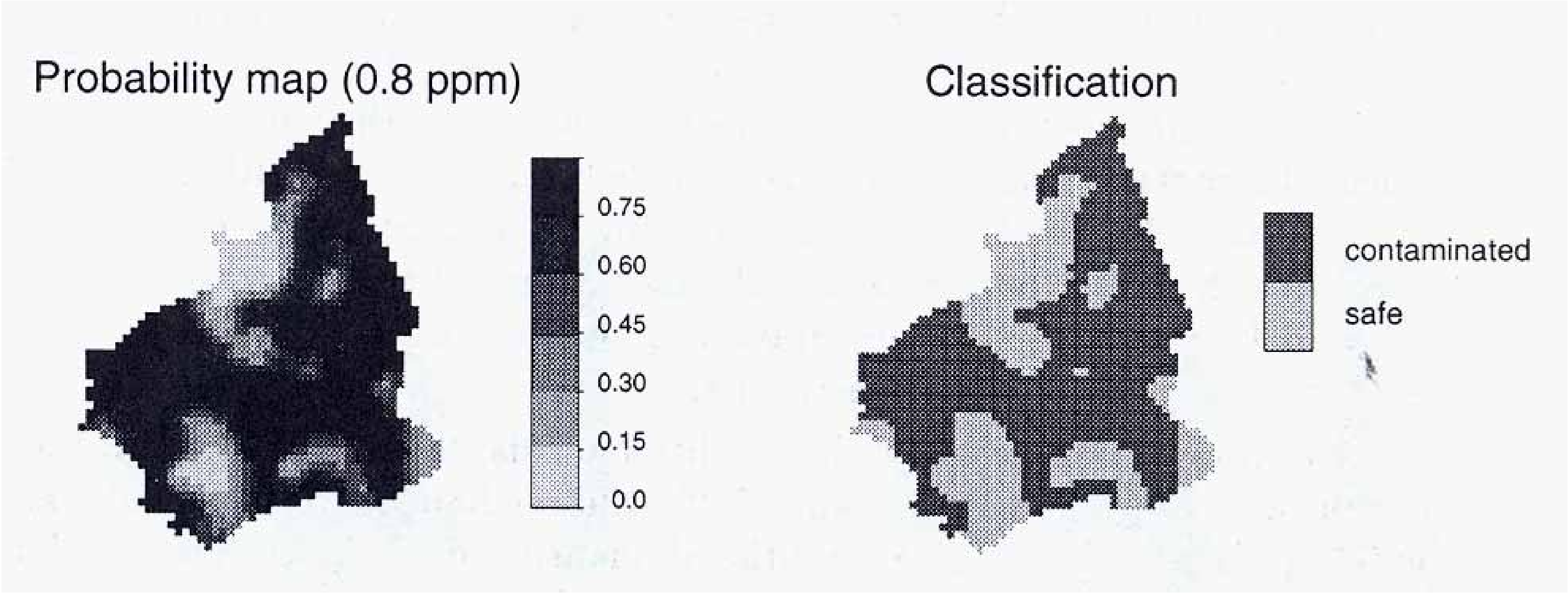
  • Consider cleaning all the locations where the probability of exceeding the tolerable maximum 0.8 ppm is larger than a marginal probability threshold.



Decision making in the face of Uncertainty

Exceeding a probability threshold - Example

Figure 7.47 - Goovaerts, 1998



Classification of locations as contamined by cadmium on the basis that the probability of exceeding the critical threshold 0.8 ppm is larger than the marginal probability of contamination (0.625)

Decision making in the face of Uncertainty

Exceeding a physical threshold (Risk Analyses)

Given a map of optimal estimates Z_L^*, a critical value z_c and the cdf $F(\mathbf{u};z|(n))$ at **u**

Two misclassification risks can be assessed

1. The **risk α(u)** of wrongly classifying a location u as hazardous (false positive) is:

$$\alpha(\mathbf{u}) = \text{Prob}\{Z(\mathbf{u}) \leq z_c \mid z_L^* > z_c, (n)\} = F(\mathbf{u}; z_c \mid (n))$$

for all locations **u** such that the estimate $z_L^*(\mathbf{u}) > z_c$

2. The **risk β(u)** of wrongly classifying a location **u** as safe (false negative) is:

$$\beta(\mathbf{u}) = \text{Prob}\{Z(\mathbf{u}) > z_c \mid z_L^* \leq z_c, (n)\} = 1 - F(\mathbf{u}; z_c \mid (n))$$

for all locations **u** such that the estimate $z_L^*(\mathbf{u}) \leq z_c$

Decision making in the face of Uncertainty

Exceeding a physical threshold

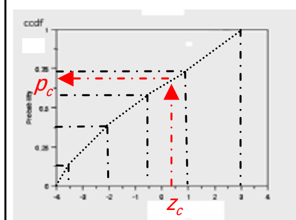
Example Risk analyses - **risk α(u)** and **risk β(u)** for $z_c = .8$

Map of estimates

+.81	+.87	+.91	+.79	+.75
+.85	+.86	+.79	+.66	+.72
+.92	+.89	+.81	+.80	+.77
+.95	+.85	+.82	+.75	+.77

Risk α ← → Risk β

Not Safe / Safe



$z_c = .8$

Map of Risks - Probabilities

+.11	+.27	+.31	+.19	+.25
+.25	+.36	+.79	+.46	+.22
+.42	+.29	+.81	+.40	+.47
+.55	+.35	+.22	+.35	+.27

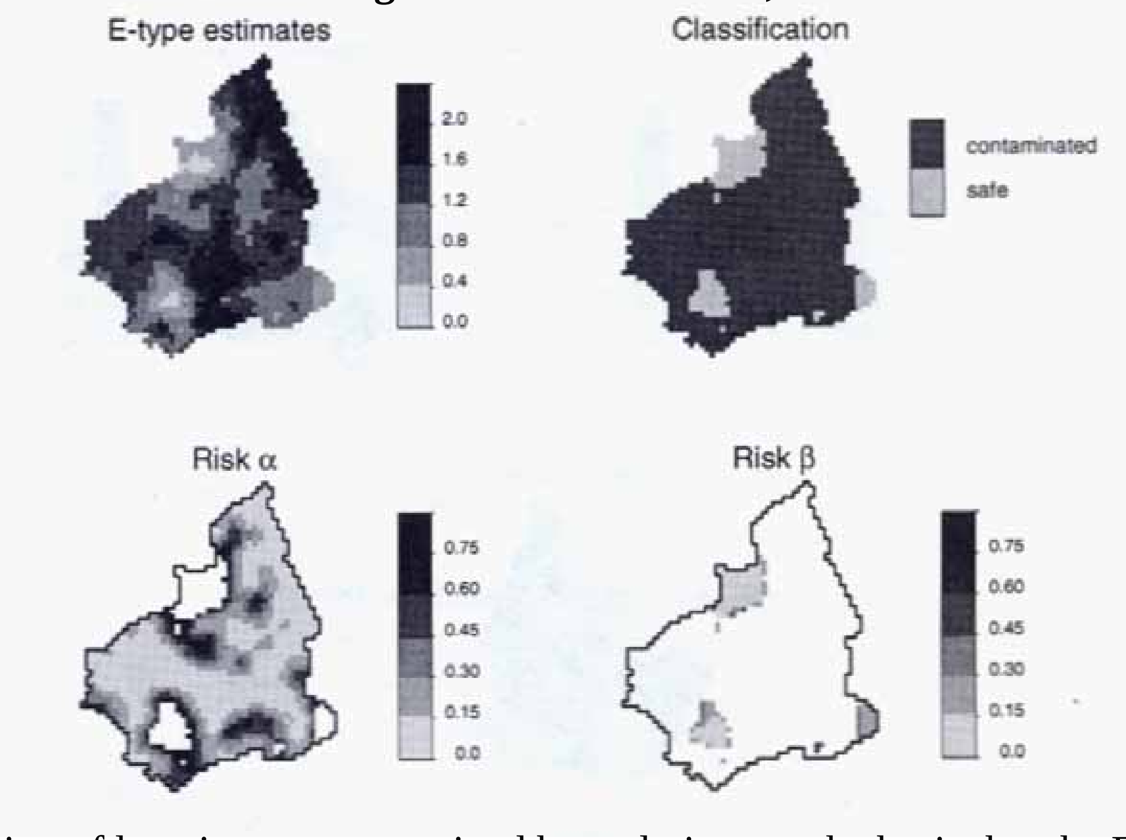
Risk α ← → Risk β

Misclassification risks can be used to rank locations candidates to additional sampling

Decision making in the face of Uncertainty

Exceeding a physical threshold - Example

Figure 7.48 - Goovaerts, 1998



Classification of locations as contamined by cadmium on the basis that the E-type estimate exceeds the critical threshold 0.8 ppm. Bottom graphs show the corresponding risks of wrongly declaring that a location is hazardous (risk α) or safe (risk β)

Decision making in the face of Uncertainty

Minimization of the expected losses

Based on the specification of two *economic functions* that measure the impact of the two types of misclassification: safe and contaminated, for example.

The loss associated with classifying a location **u** as safe could be modeled as:

$$L_1(z(\mathbf{u})) = \begin{cases} 0 & \text{if } z(\mathbf{u}) \leq z_c \\ w_1[z(\mathbf{u}) - z_c] & \text{otherwise} \end{cases}$$

The loss associated with classifying a location **u** as contaminated (remediation cost) could be modeled as:

$$L_2(z(\mathbf{u})) = \begin{cases} 0 & \text{if } z(\mathbf{u}) > z_c \\ w_2 & \text{otherwise} \end{cases}$$

where w1 and w2 are constants

Decision making in the face of Uncertainty

Minimization of the expected losses

- The conditional cdf model $F(\mathbf{u}; z|(n))$ allows one to determine the expected loss attached to the two type of classifications

$$\varphi_i(\mathbf{u}) = E\left[L_i(Z[\mathbf{u}]) \mid (n)\right] = \int_{-\infty}^{\infty} L_i(z(\mathbf{u}))dF(\mathbf{u}; z \mid (n)) \quad i = 1,2$$

- which are in practice approximated as:

$$\varphi_i(\mathbf{u}) \cong \sum_{k=1}^{K+1} L_i(\bar{z}_k)\left[F(\mathbf{u}; z_k \mid (n)) - F(\mathbf{u}; z_{k-1} \mid (n))\right] \quad i = 1,2$$

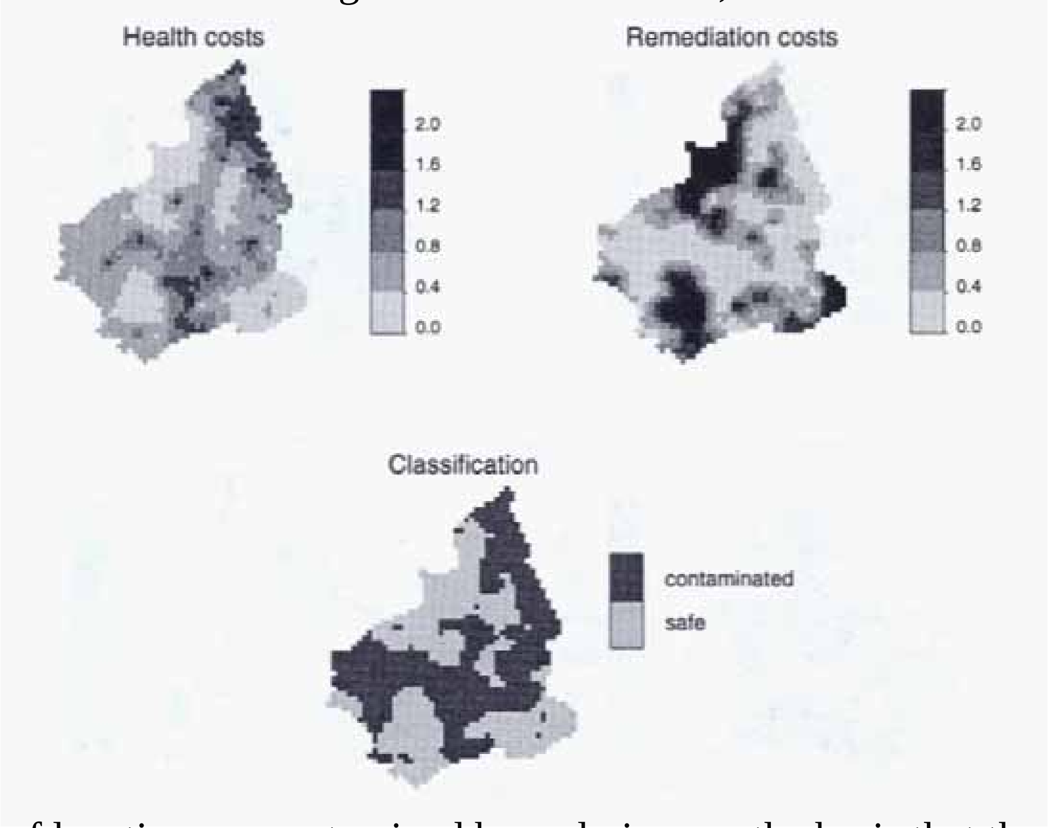
- the location **u** is then declared safe or contaminated so as to minimize the resulting expected loss:

$$\varphi_1(\mathbf{u}) > \varphi_2(\mathbf{u}) \Rightarrow \mathbf{u} \text{ is classified as contamined}$$
$$\varphi_1(\mathbf{u}) \le \varphi_2(\mathbf{u}) \Rightarrow \mathbf{u} \text{ is classified as safe}$$

Decision making in the face of Uncertainty

Minimization of the expected losses - Example

Figure 7.49 – Goovaerts, 1998



Classification of locations as contamined by cadmium on the basis that the resulting expected cost (unecessary cleaning) is smaller than the cost associated with wrongly classifying a location as safe.

Assessment of Global Uncertainty

Advanced Topics

- Researches in decision making processes using GIS information
- Non Parametrical Geostatistics (Baysean Approaches)
- Exploration of Binomial models (Suzana-Eduardo-Miguel) to risk analysis
- Spatio-Temporal Analysis using geostatistic (Suzana-Rodrigo)

See also: Short Course on Geostatistical Analysis of Environmental Data (Goovaerts) http://www.ai-geostats.org/index.php?id=206 and

http://conference.ifas.ufl.edu/soils/geostats/index.html

From Goovaerts short course

Space-time Geostatistics - Approaches available

1. Space-rich Time-poor information
2. Space-poor Time-rich information
3. Space-rich Time-rich information

- A space-time model (Sulfate in Europe)
- Comparison of space-time interpolation methods

Decision making in the face of Uncertainty

Summary and Conclusions

- The spatial data must be modeled considering the uncertainties related to these representations.
- The uncertainties of individual attributes propagates to the results obtained with spatial modeling. The correlation between the attributes must be considered in spatial modeling and in uncertainty propagation procedures.
- The uncertainty models of spatial information should be used in decision make processes in order to get more reliable answers for spatial problems.
- GISs and their tools should be used to perform complex spatial analysis, considering uncertainties in data, instead of being used only to create nice colored maps.

Decision making in the face of Uncertainty

Basic References

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.

Bussab, W. O.; Morettin, P. A.. 1987. *Estatística Básica*. Atual Editora Ltda, Sao Paulo, Brasil.

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

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.

Decision making in the face of Uncertainty

END

of the Course