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

Master of Science in Geospatial Technologies

Geostatistics
Assessment of Local Uncertainty with Indicator Geostatistics

Carlos Alberto Felgueiras
cfelgueiras@isegi.unl.pt

Assessment of Local Uncertainty

Contents

Assessment of Local Uncertainty

- **Introduction (from Deutsch and Journel, 1998)**

  - **The Kriging (simple kriging for example)** allows the estimation of an attribute value considering the following linear estimator.

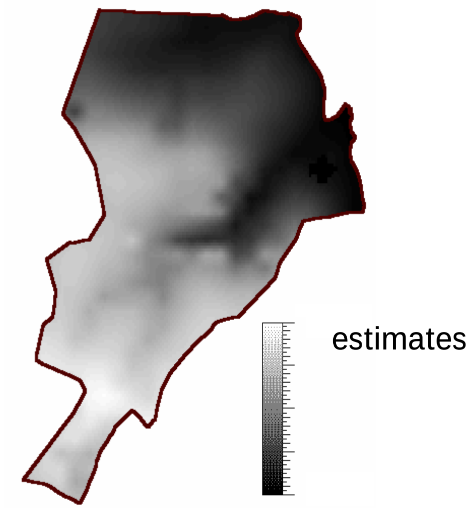
$$Z^*(\mathbf{u}) = \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u}) \cdot Z(\mathbf{u}_\alpha) + \left[1 - \sum_{\alpha=1}^{n(\mathbf{u})} \lambda_\alpha(\mathbf{u})\right] \cdot m$$

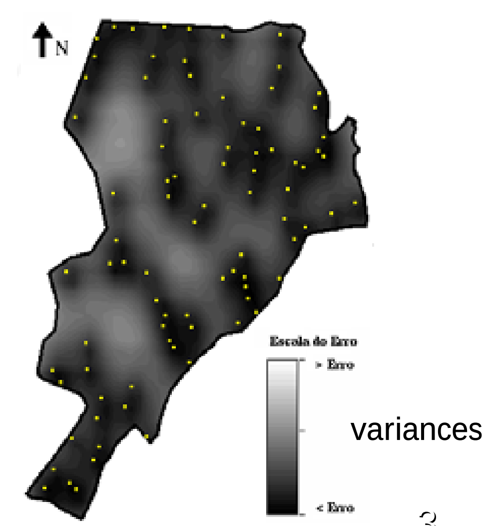
  - The **weights $\boldsymbol{\lambda}_\alpha$** are determined to minimize the error variance, also called the “estimation variance”. That minimization results in a set of normal equations

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

  - The corresponding **minimized estimation variance**, or kriging variance, is:

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





Assessment of Local Uncertainty

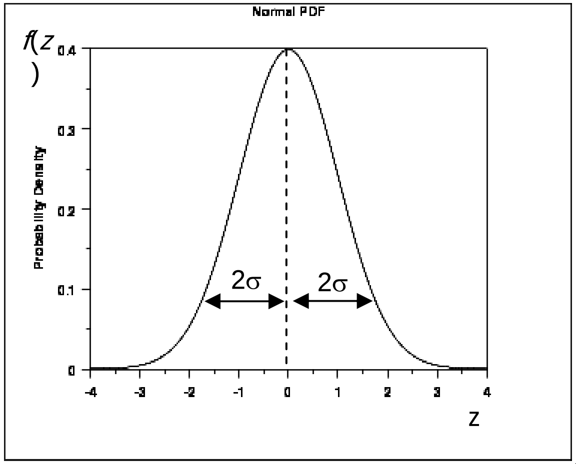
- **Introduction (from Deutsch and Journel, 1998)**

  - The corresponding **minimized estimation variance**, or kriging variance, is:

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

  - **Important:** The **kriging variances**, being independent of the data values, **only provides a comparison of alternative geometric data configurations. Kriging variances are usually not measures of local estimation accuracy.** Ex. gaussian-type confidence interval:

$$\text{Prob}\left\{Z(\mathbf{u}) \in \left[z^*_{SK}(\mathbf{u}) \pm 2\sigma_{SK}(\mathbf{u})\right]\right\} \cong 0.95$$

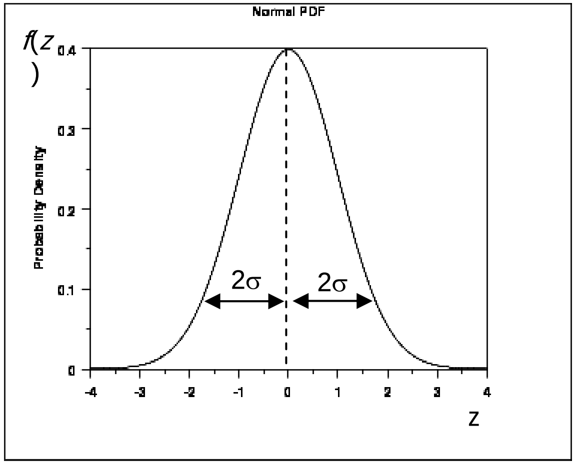


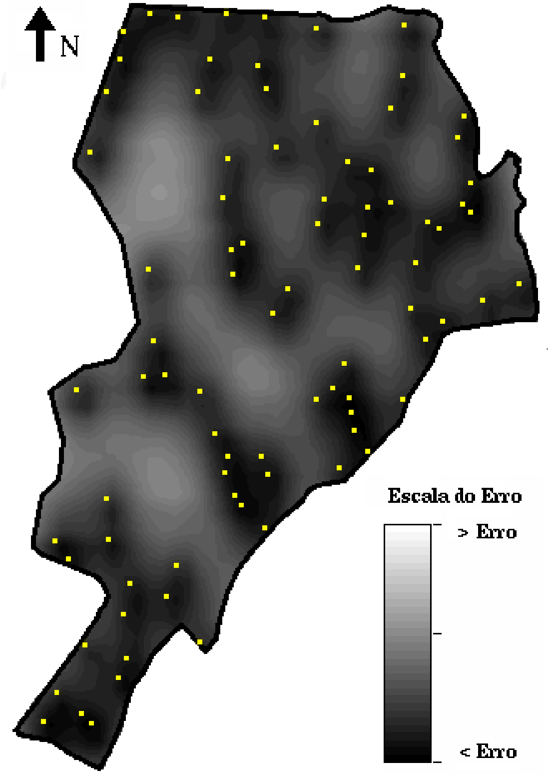
Assessment of Local Uncertainty

- **Introduction (from Deutsch and Journel, 1998)**

  - A kriging variance map showing that the confidence interval is proportional to the distances of the **u** locations and the sample locations, is:

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

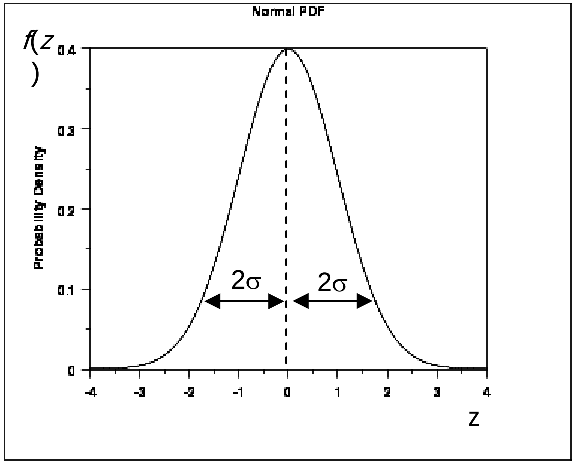




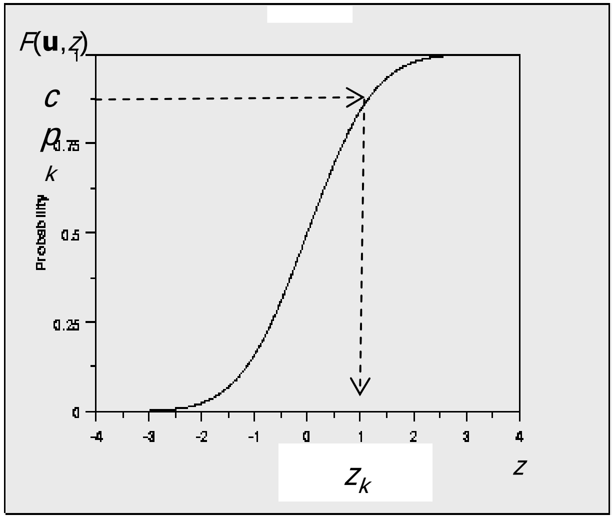
Assessment of Local Uncertainty

- **Introduction (from Deutsch and Journel, 1998)**

  - **The Multi-Gaussian Approach (MG):** If the RF model $Z(\mathbf{u})$ is multivariate Gaussian, then the simple kriging estimate and variance identify the mean and the variance of the posterior ccdf. In addition , since the ccdf is Gaussian, it is fully determined by these two parameters (Parametric Approach). How to check that the Random Function is multigaussian?



- **The Indicator Kriging approach (IK)**: If the value to be estimated is the expected value (mean) the standard krigings (simple, ordinary, cokriging, ...) are a priori the preferred algorithm. Otherwise, the ***indicator kriging*** provides tools for constructing an approximation of the ccdf, $F(\mathbf{u},z)$, model of uncertainty about $z(\mathbf{u})$. How to do this?



Assessment of Local Uncertainty

- **Indicator Approach for continuous variables**
  - **Indicator transformation and properties**

Instead of the Variable $Z(\mathbf{u})$, consider its binary indicator transform $I(\mathbf{u};z_k)$ as defined by the relation:

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

Kriging of the indicator RV $I(\mathbf{u};z)$ provides an estimate that is also the best LS estimate of the conditional expectation of $I(\mathbf{u};z)$. Now the conditional expectation of $I(\mathbf{u};z)$ is equal to the local ccdf of $Z(\mathbf{u})$; indeed:

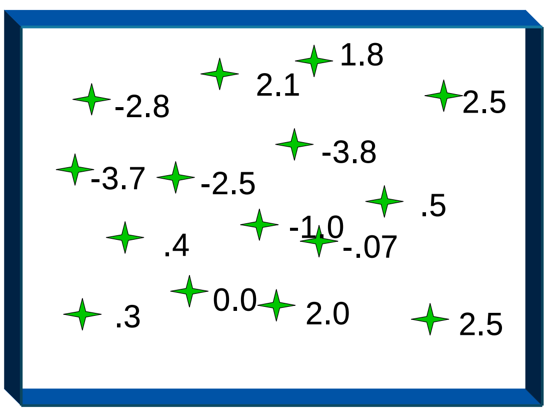
$$\begin{aligned}\mathrm{E}\{I(\mathbf{u};z_k)\,|\,(n)\} &= 1\cdot\mathrm{Prob}\{I(\mathbf{u};z_k)=1\,|\,(n)\}+0\cdot\mathrm{Prob}\{I(\mathbf{u};z_k)=0\,|\,(n)\}\\ &= 1\cdot\mathrm{Prob}\{I(\mathbf{u};z_k)=1\,|\,(n)\}=\mathrm{F}(\mathbf{u};z_k\,|\,(n))\end{aligned}$$

Important: So the indicator kriging is not aimed at estimating the unsampled value $z(\mathbf{u})$; or its indicator transform $i(\mathbf{u};z)$ but at providing a ccdf model of uncertainty about $z(\mathbf{u})$.

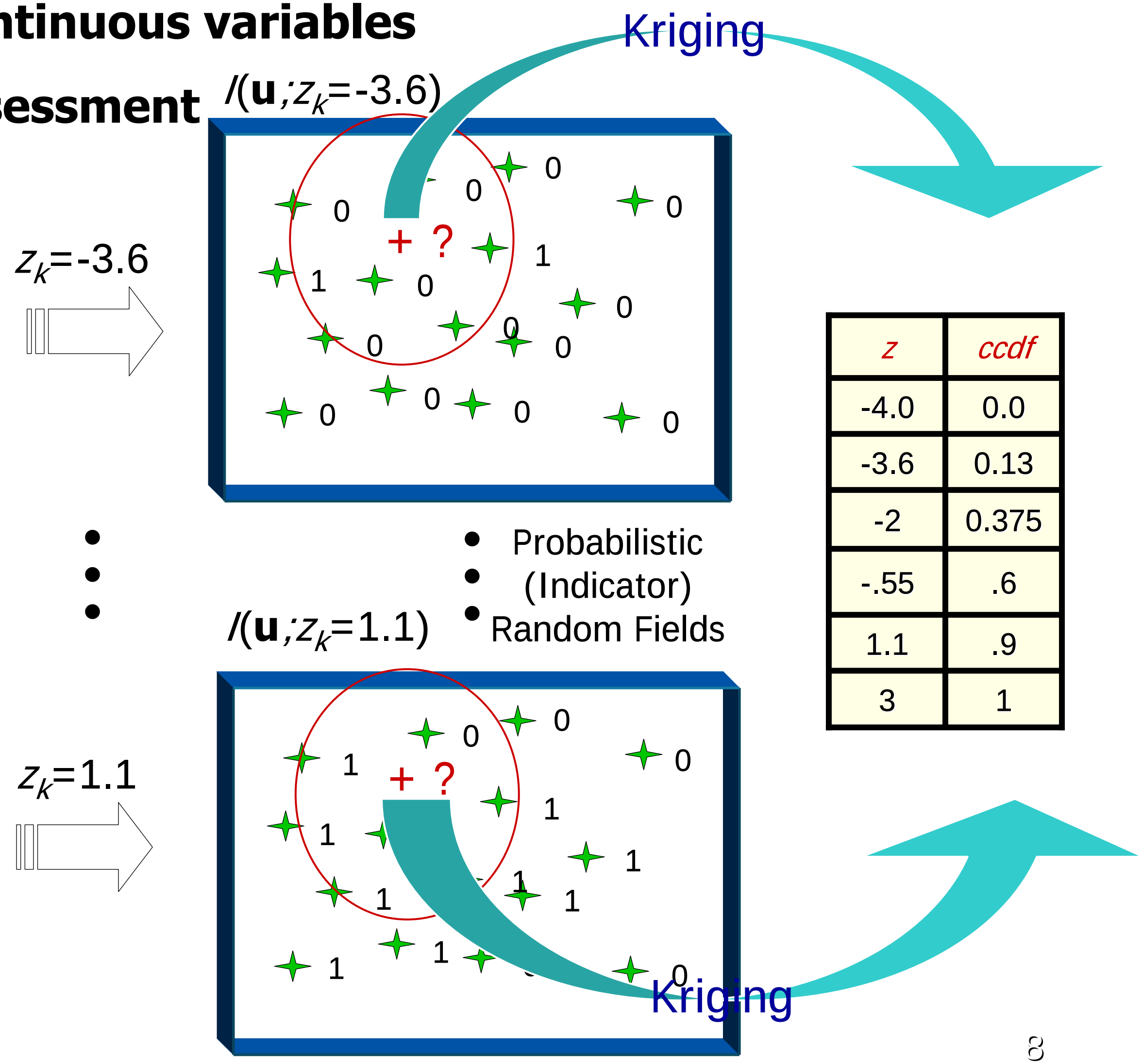
Assessment of Local Uncertainty

- **Indicator Approach for continuous variables**
- **The uncertainty model assessment**

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



z_{min} = -4 and z_{max} =3



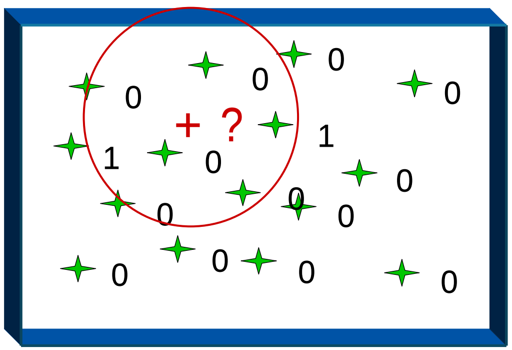
- Probabilistic
- (Indicator)
- Random Fields

z	*ccdf*
-4.0	0.0
-3.6	0.13
-2	0.375
-.55	.6
1.1	.9
3	1

Assessment of Local Uncertainty

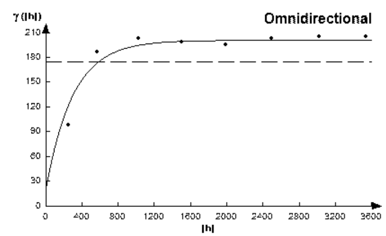
- **Indicator Approach for continuous variables**
  - **The uncertainty model assessment**

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



Indicator Variograms

One for each cutoff

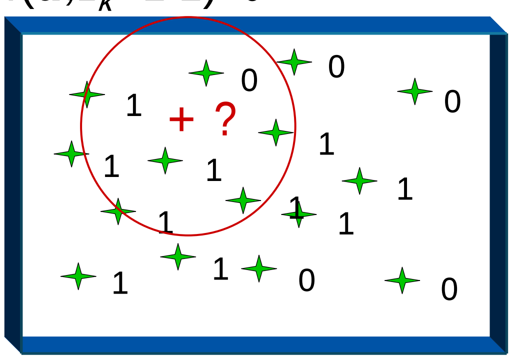


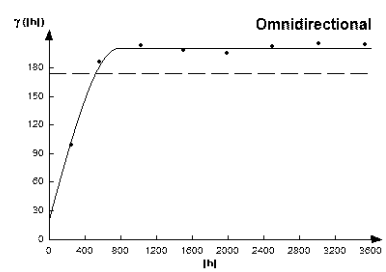
Kriging

z	*ccdf*
-4.0	0.0
-3.6	0.13
-2	0.375
-.55	.6
1.1	.9
3	1

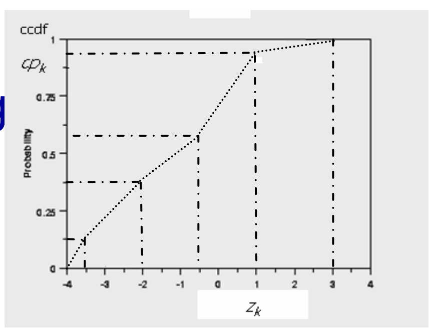
- Probabilistic
- (Indicator)
- Random Fields

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





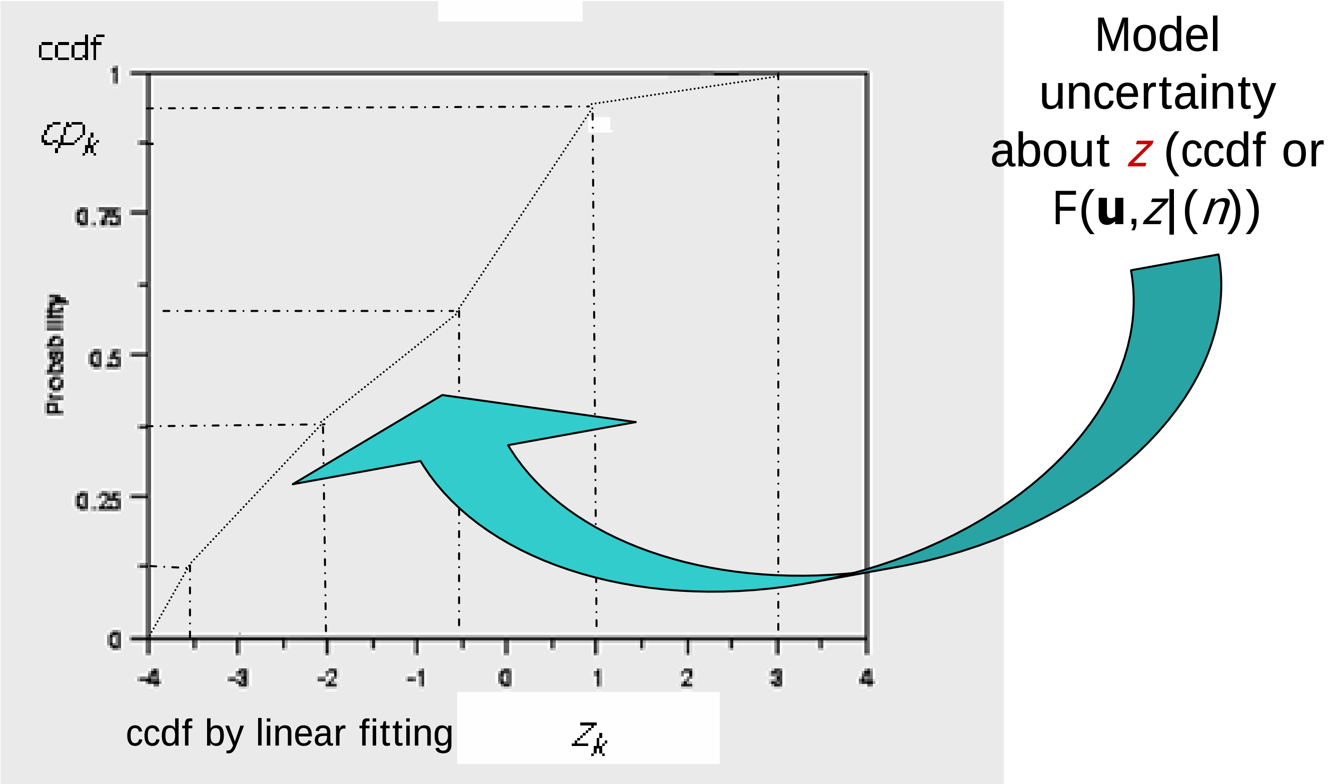
Kriging



Assessment of Local Uncertainty

- **Indicator Approach for continuous variables**
  - **The uncertainty model assessment (Ilustration) :** Using *K* cutoffs, or thresholds, values plus the minimum and maximum values (in this case *K*=4, z_{min}=0 and z_{max}= 3)

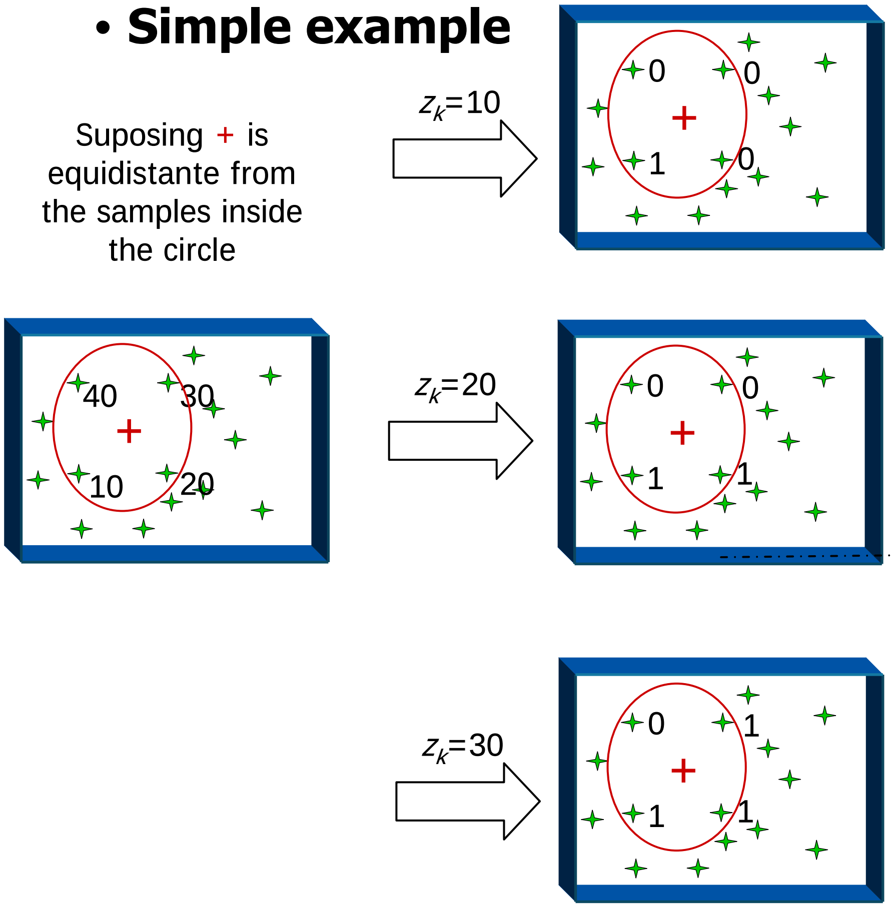
z	*ccdf*
0.0	0.0
-3.6	0.13
-2	0.375
-.55	.6
1	.9
3	1

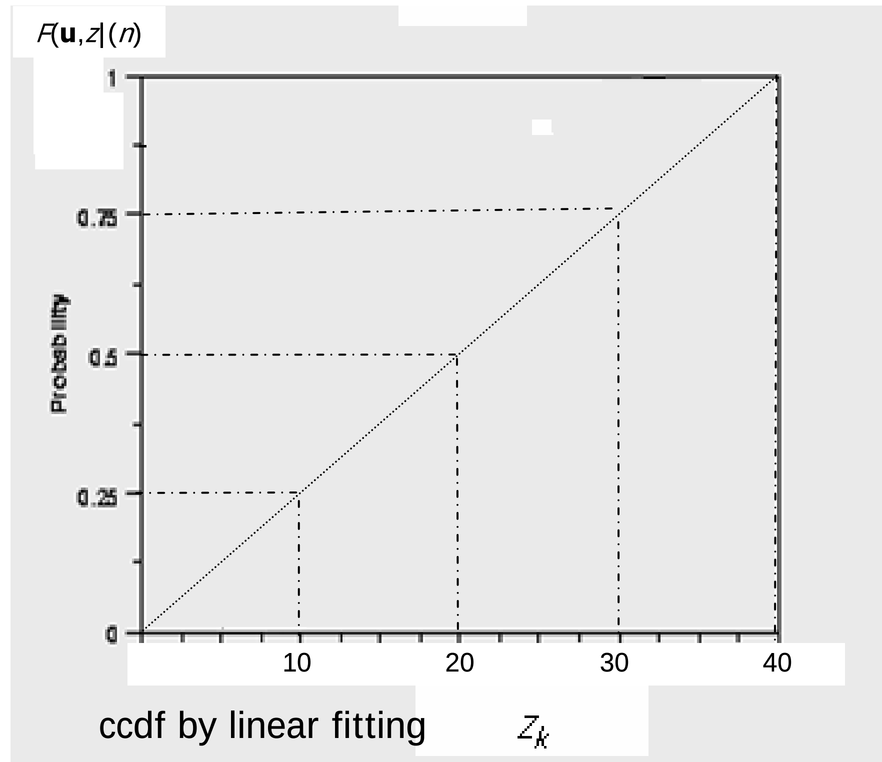


Assessment of Local Uncertainty

- **Indicator Approach for continuous variables**
  - **Simple example**

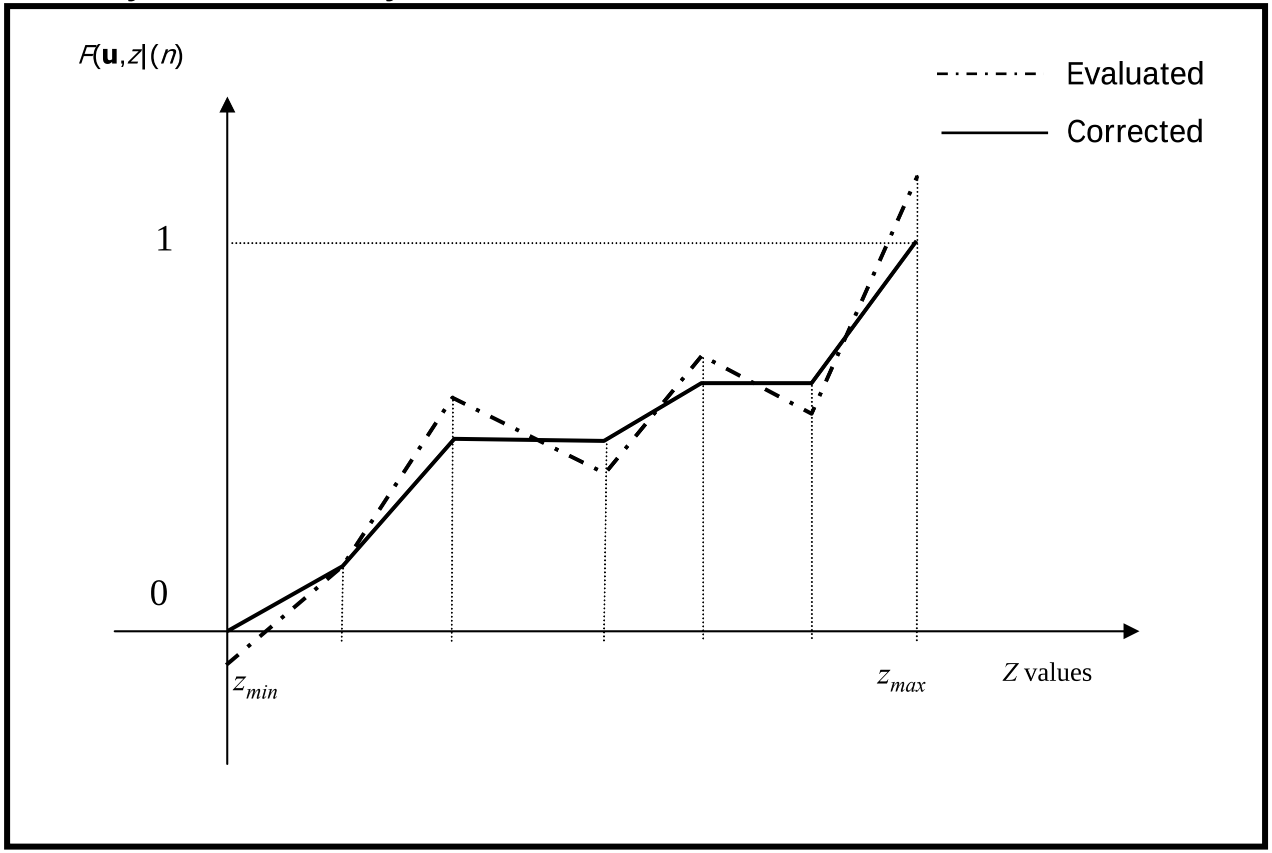
$$I(\mathbf{u};z_k)=\begin{cases}1, & \text{se} \quad Z(\mathbf{u})\le z_k\\ 0, & \text{se} \quad Z(\mathbf{u})> z_k\end{cases}$$





Assessment of Local Uncertainty

- **Indicator Approach for continuous variables**
  - **Correcting for order relation deviations (Goovaerts, 1997) –** the final ccdf representation, $F(\mathbf{u},z|(n))$, must be a non-decreasing function and must lie in the interval [0,1], so it may be necessary corrections for order relation deviations.



Assessment of Local Uncertainty

- **Indicator Approach for continuous variables**
  - **Estimating RV parameters** (mean and variance)
  - The ***mean value*** can be estimated from a discrete ccdf representation as:

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

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

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

  - Using the expected value above, the ***variance*** can be estimated, similarly to the mean value, as:

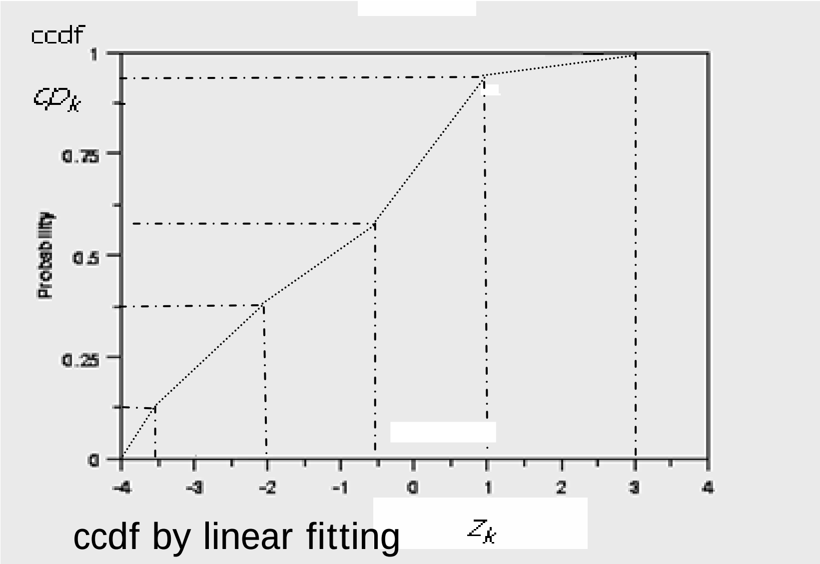
$$Var[Z] = E\left[(Z - E(Z))^2\right] = \int_{-\infty}^{\infty} (z - E(Z))^2 \cdot f(z)\,dz$$

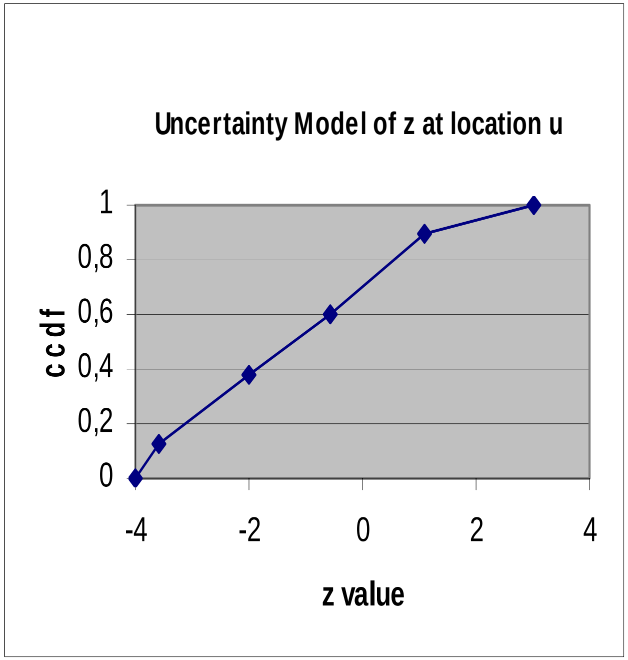
$$Var[z(\mathbf{u})]^* \approx \sum_{k=1}^{K+1} (z'_k - [z(\mathbf{u})]_E^*)^2 \cdot [F(\mathbf{u}; z_k \mid (n)) - F(\mathbf{u}; z_{k-1} \mid (n))]$$

Assessment of Local Uncertainty

- **Indicator Approach for continuous variables**
  - **Estimating RV parameters** (mean and variance – using Excel)

z	*ccdf*
0.0	0.0
-3.6	0.13
-2	0.375
-.55	.6
1	.9
3	1



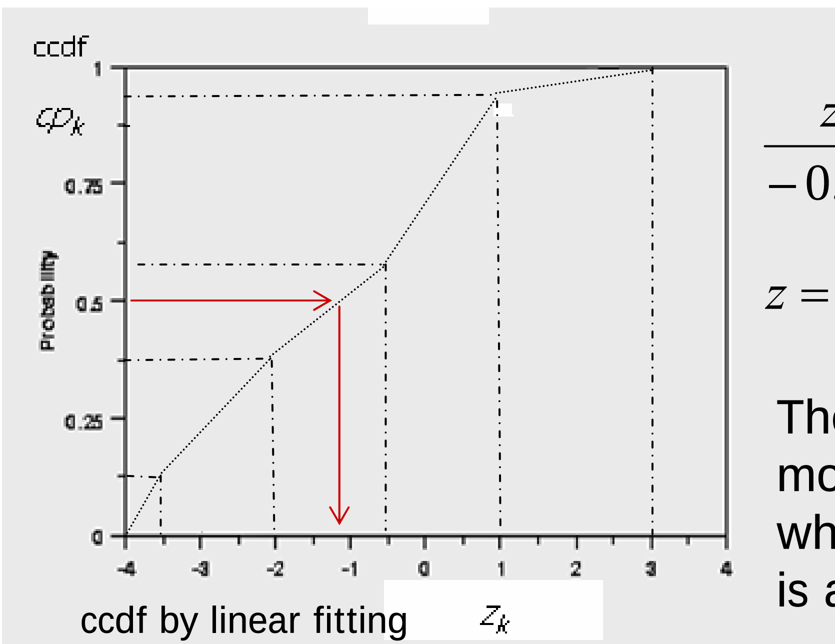


Z Value	Probability	(zk+zk-1)/2	F(zk)-F(zk-1)	C*D	(C-mean)^2	F*D
-4,000	0,000	-3,800	0,130	-0,494	6,868	0,893
-3,600	0,130	-2,800	0,245	-0,686	2,626	0,643
-2,000	0,375	-1,275	0,225	-0,287	0,009	0,002
-0,550	0,600	0,275	0,300	0,083	2,115	0,635
1,100	0,900	2,050	0,100	0,205	10,429	1,043
3,000	1,000		Mean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

Assessment of Local Uncertainty

- **Indicator Approach for continuous variables**
  - **Estimating RV parameters** (median, quantils, ….)
  - The **median value**, $q_{.5}$, can be estimated, from a discrete ccdf representation, as the z value whose probability equals .5:

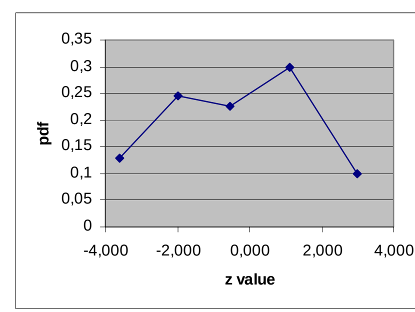
z	*ccdf*
-4.	0.0
-3.6	0.13
-2	0.375
-.55	.6
1.1	.9
3	1



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

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

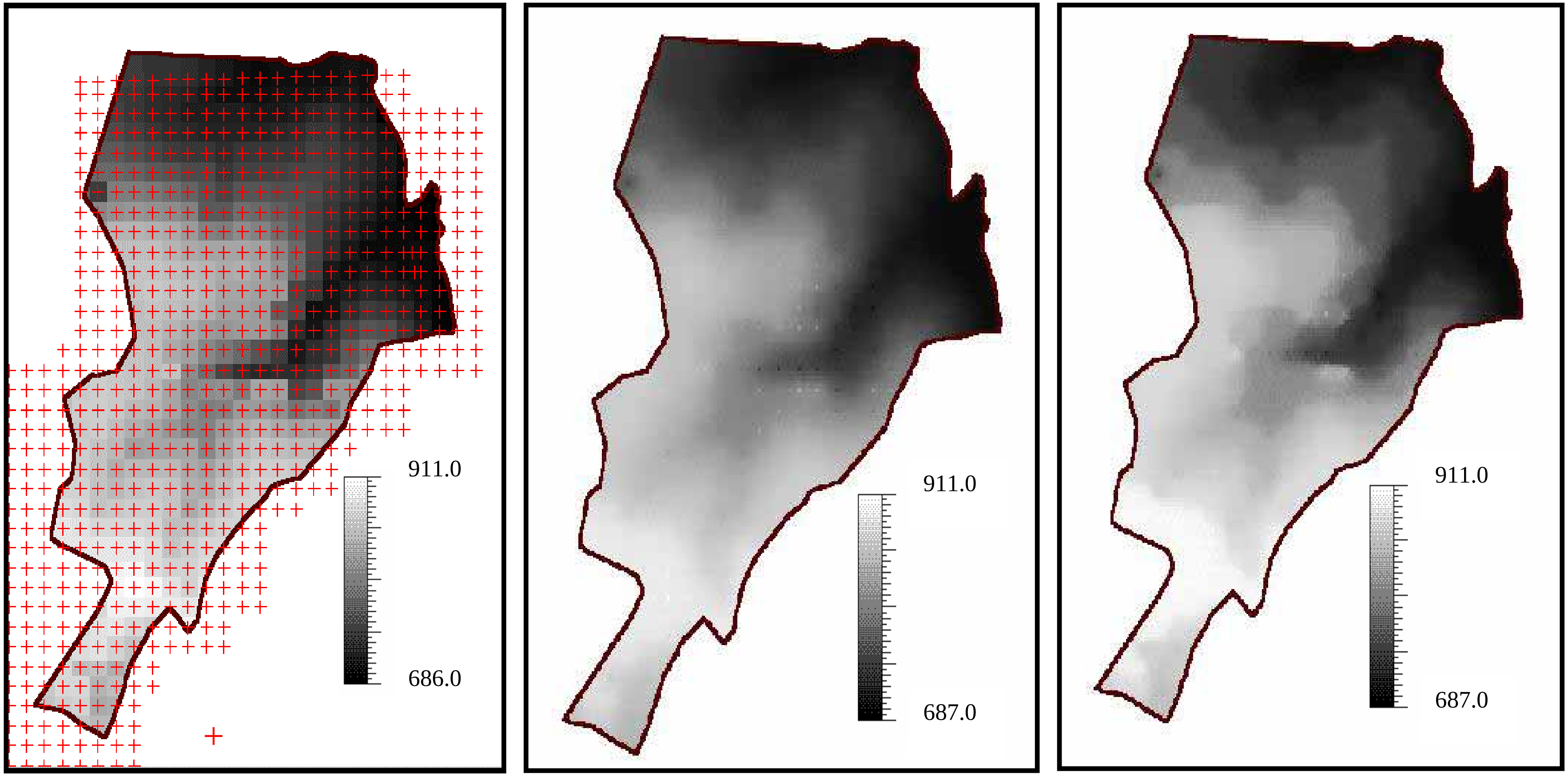
The median value is a more robust estimator when the distribution is asymmetric



- Similarly, it is possible to estimate any *q* quantil :

Assessment of Local Uncertainty

- **Indicator Approach for continuous variables**
  - Examples of Prediction Maps (Dirichlet, mean and medium maps)



Assessment of Local Uncertainty

- **Indicator Approach for continuous variables**
  - **Uncertainties evaluations from confidence intervals**
    - ***Standard Deviations***: deviation from the mean value. Example using only one standard deviation

$$\mathrm{Unc}(\mathbf{u}) = 2\sigma(\mathbf{u}) \text{ where } \mathrm{Prob}\{Z(\mathbf{u}) \in [\mu_Z(\mathbf{u}) \pm \sigma(\mathbf{u})]\} \cong 0.68$$

    - ***Quantiles***: split the realization in n subsets. Example of interquartil confidence interval

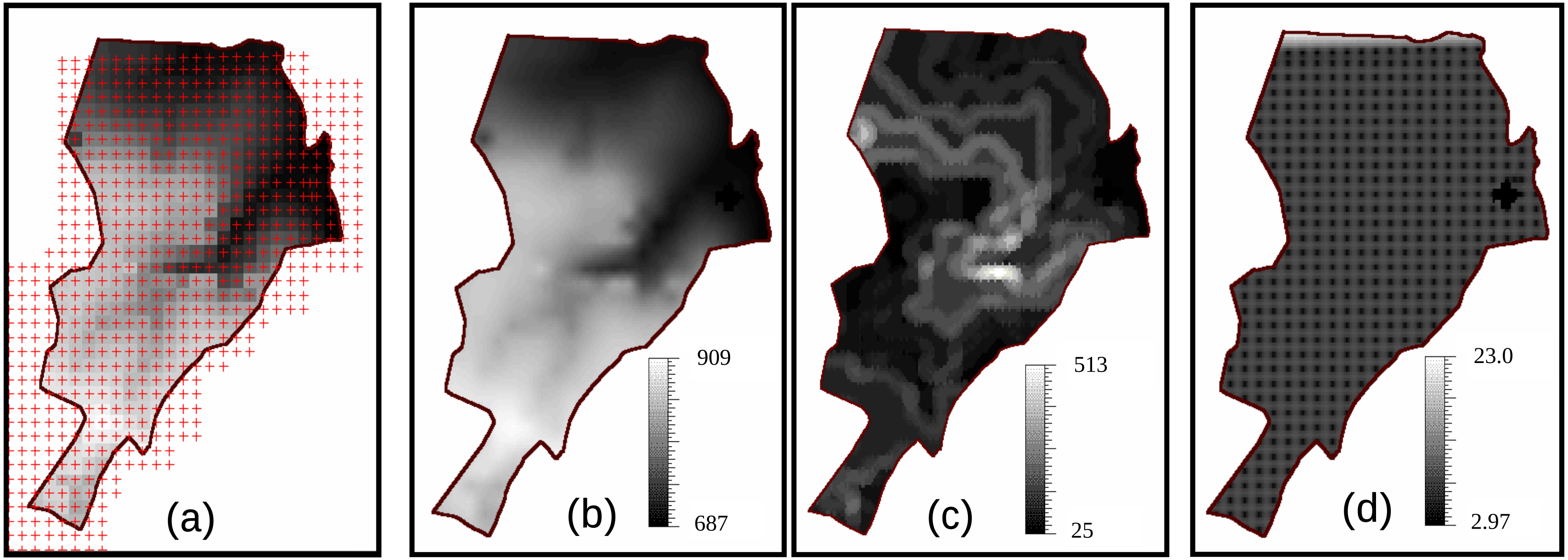
$$\mathrm{Unc}(\mathbf{u}) = [q_{0.25}; q_{0.75}] \text{ where } \mathrm{Prob}\{Z(\mathbf{u}) \in [q_{0.25}; q_{0.75}] \mid (n)\} = 0.50$$

    - ***Shannon Entropy:*** to be investigated

Assessment of Local Uncertainty

- **Indicator Approach for continuous variables**
  - Examples of Predictions and Uncertainties Maps

Maps: (a) of mean estimates with samples, (b) mean estimates, (c) uncertainties from standard deviations and (d) kriging variance



How to interpret map (c)? Why the maps (c) and (d) are so differents?

Assessment of Local Uncertainty

- **Problems with indicator geostochastic procedures**

The main drawback of using geostatistic approaches is the need of work on variogram generations and fittings for each cutoff . This work is interactive and requires from the user knowledge of the main concepts related to basics of the geostatistics, and indicator approaches, in order to obtain reliable results.

The definition of the number of cutoffs is important too. Ideally one must define many numbers of cutoffs. This will allow to get a more reliable approximation of the real uncertainty model (the ccdf). On the other hand, too many cutoffs lead to too many variogram generation and fitting. This means more work for the user. The variogram generation and fitting for extreme values, larger and smaller cuttofs, sometimes are hard to obtain because of the greater number of 0s (zeros) or 1s (ones) in the Random Fields.

Assessment of Local Uncertainty

- **Advantages on using indicator geostochastic procedures**

  - All the advantages of the geostatistic approaches because of the use of:
    - variograms to represent the variation of the attribute.
    - kriging to estimate the values considering covariance between samples and between the samples and the point to be estimated
  - Allows the assessment of the local uncertainty model at any **u** spatial location that can be used for getting:
    - estimates maps using different distribution parameters as mean, median or any quantil.
    - uncertainty maps based on confidence intervals of standard deviation or quantils

Assessment of Local Uncertainty

Summary and Conclusions

- If the value to be estimated is the expected value (mean), standard krigings (simple, ordinary, cokriging, ...) are a priori the preferred algorithm.
- Otherwise, the ***indicator kriging*** provides tools for constructing an approximation of the uncertainty model (ccdf) of the attribute for any spatial location **u**. The ccdfs model allow the creation of maps of estimates, other than the mean value, and maps of uncertainties based on confidence intervals.
- The uncertainties can be used to qualify the estimation at each spatial location **u** considered.
- Indicator approaches can be applied to continuous variables and to categorical variables (to be seen in next class).

Assessment of Local Uncertainty

Exercises

1. Run the Lab6 available in the geostatistics course area of ISEGI online. (The Portuguese version will be replaced by the English-Portuguese version before next Tuesday)

2. Given the pdf (right) of a continuous variable:

   2.1 Construct and plot the ccdf of the variable

   2.2 Evaluate the mean and the median parameters from the ccdf.

   2.3 Evaluate the standard deviation and the first and last decils and quartils of the distribution.

   2.4 Evaluate the confidence intervals based on 1 and 2 standard deviations.

   2.5 Evaluate the confidence intervals based on the quartil and the decil quantils.

3. Send a report to the professor about the above exercises, before 29/11/2007

z	*pdf*
10	.10
23	.14
34	.32
38	.27
43	.09
49	.08

Predictions with Deterministic Procedures

END

of Presentation