Risk Assessment of Sediment Pollution Using Geostatistical Simulations

Jean Golay and Mikhaïl Kanevski

University of Lausanne
Institute of Geomatics and Risk Analysis

Journée des laboratoires de SIG romands, 2012
This presentation was partly presented at the EGU-2012 conference in Vienna

Introduction

Objectives and Motivations

When dealing with risk assessment, extremes values must be retrieved and probabilites of exceeding given thresholds must be computed. Geostatistics provides several algorithms allowing to take into account these requisites. The present study focuses on:

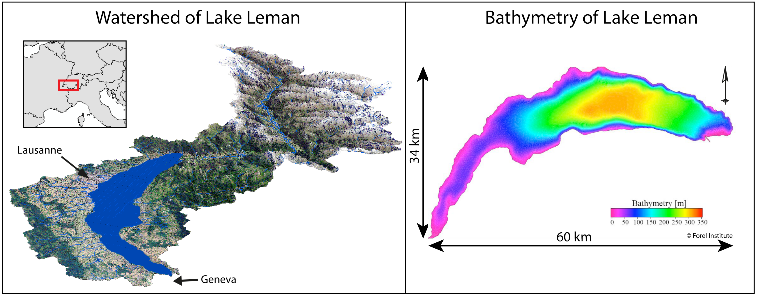
- a comparison of the effectiveness of three parametric algorithms in reproducing global statistics when applied to a complex dataset of sediment pollution.
- a comparison of the way they characterize the uncertainty of the contamination pattern.
- the visual analytics of the results in order to effectively communicate with experts.

Outline

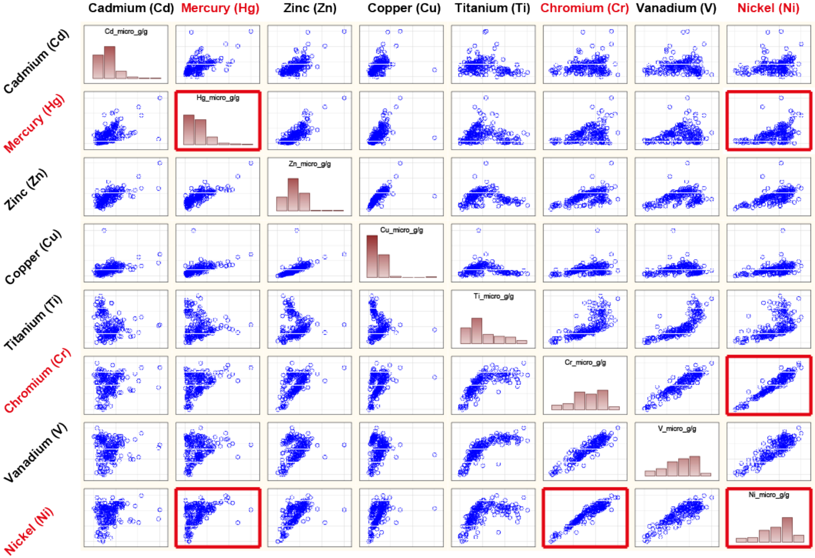
Outline

Lake Leman (or Lake Geneva)

- Lake Leman is one of the largest lakes in Western Europe with a water volume of $89km^3$.
- Approximately, one million people live within its watershed.
- Water pollution reached a peak in the 1980s and has decreased ever since.

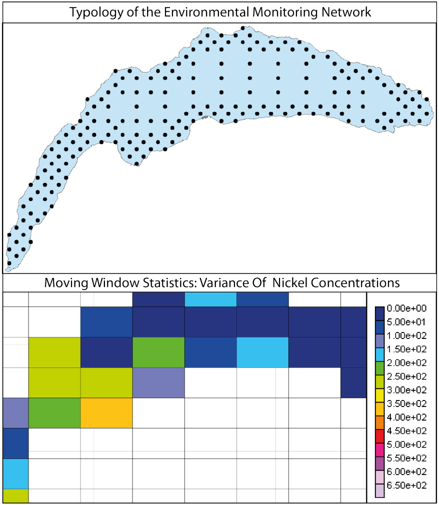


Correlation Matrix

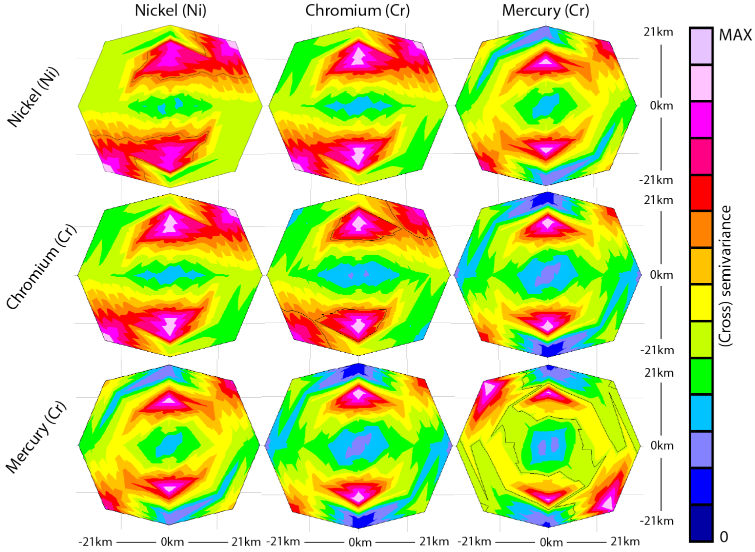


Measurement Network and Moving Window Statistics

- No significant clustering within the limit of the validity domain
- Moving Window Statistics show, among other things, that the variance of the Nickel concentrations is higher in the western part of the lake.

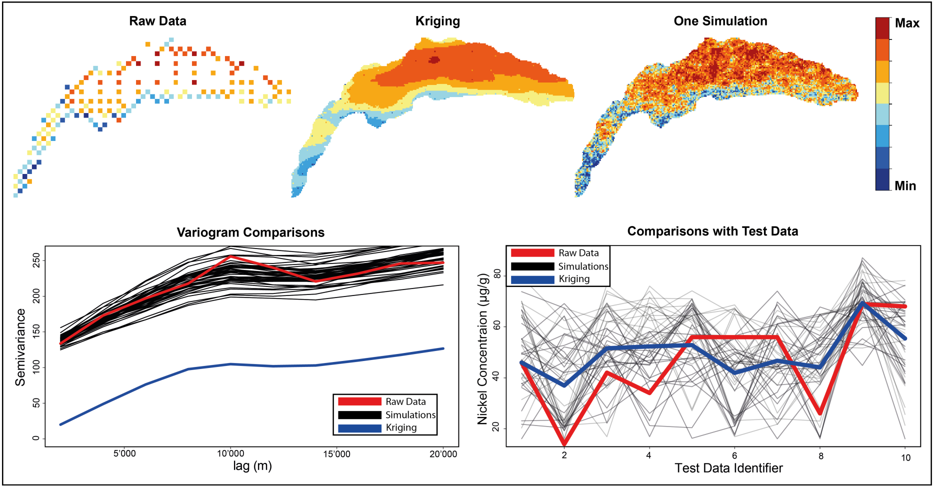


Variogram Rose Matrix



Outline

Why Do We Need Simulations?



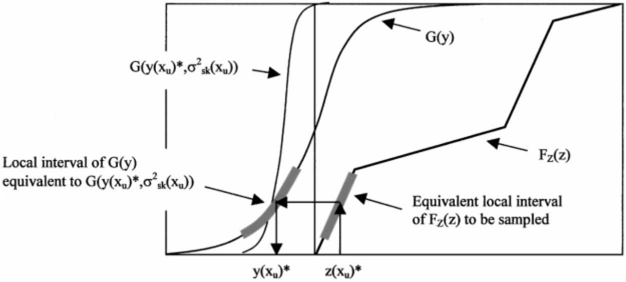
Sequential Gaussian (Co)Simulations

On the assumption that data are multinormal, sequential Gaussian simulation aims at preparing alternative, equally probable and high resolutions models of the spatial distribution of a phenomenon which has been sampled over a study area. In other words, it allows to easily solve the following equation in considering that the conditional distribution of Z_i is Gaussian with mean Z_i^* and variance $\sigma^2_{K_i}$ [Chilès and Delfiner, 1999]:

$$
\begin{aligned}
&\Pr\{z_{M+1} \le Z_{M+1} < z_{M+1} + dz_{M+1}, \ldots z_{N+1} \le Z_{N+1} < z_{N+1} + dz_{N+1} \mid z_1, \ldots z_M\} \\
&= \Pr\{z_{M+1} \le Z_{M+1} < z_{M+1} + dz_{M+1} \mid z_1, \ldots z_M\} \\
&* \Pr\{z_{M+2} \le Z_{M+2} < z_{M+2} + dz_{M+2} \mid z_1, \ldots z_M, z_{M+1}\} \\
&\bullet \\
&* \Pr\{z_N \le Z_N < z_N + dz_N \mid z_1, \ldots z_M, z_{M+1}, \ldots z_{N-1}\}
\end{aligned}
$$

Direct Sequential Simulations

Direct sequential simulation is similar to SGS, but it doesn't require to make any assumption regarding the possible multinormality of the original variable. It uses the estimated local mean and variance not to define the local cdf, but to sample from the global cdf [Soares, 2001].

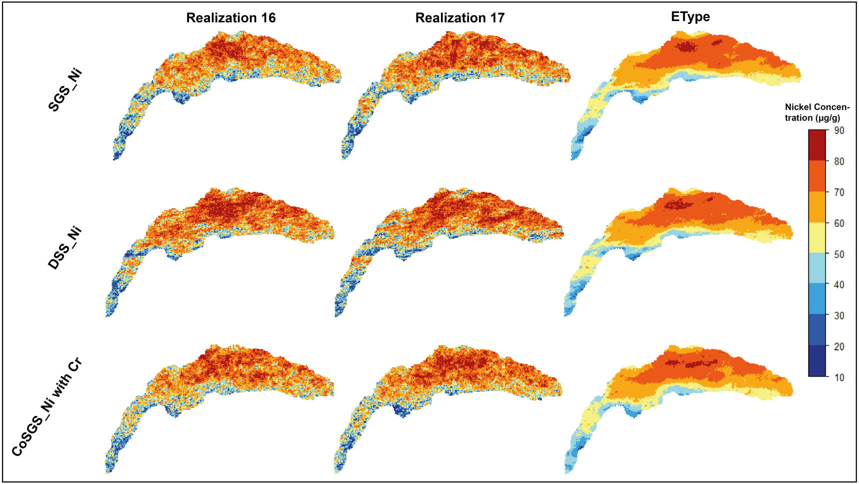


Outline

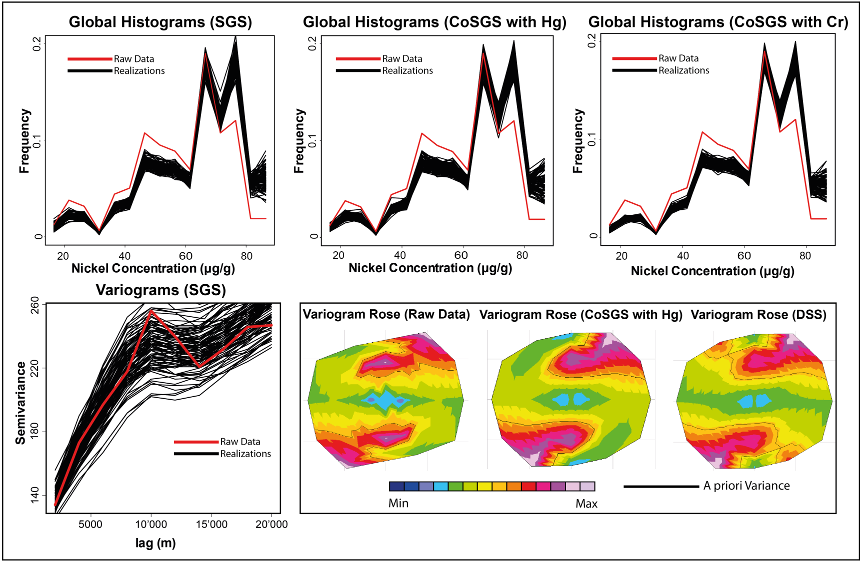
Procedure

- Four models were analyzed: SGS applied to the Nickel data, DSS applied to the Nickel data and two CoSGS applied to the Nickel data with Chromium and Mercury respectively
- Data were split into two subsets:
  - a training subset composed of 160 data points
  - a test subset composed of 40 data points
- The different algorithms were run on the training set to produce 100 simulations (for each model).
- The training set was used to assess the reproduction of the global statistics. The test set was used to check the generalization of the models (to avoid overfitting).
- Post-simulation analyses were carried out with a special focus on visual analytics and risk mapping.

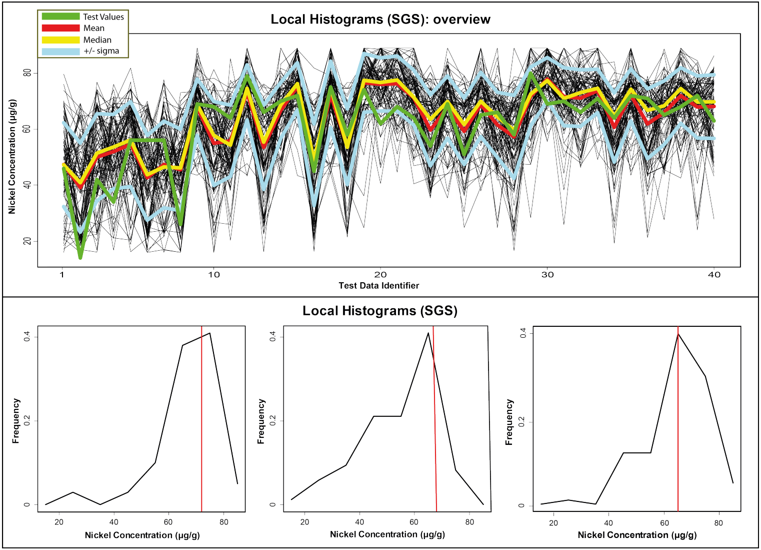
Direct Outputs



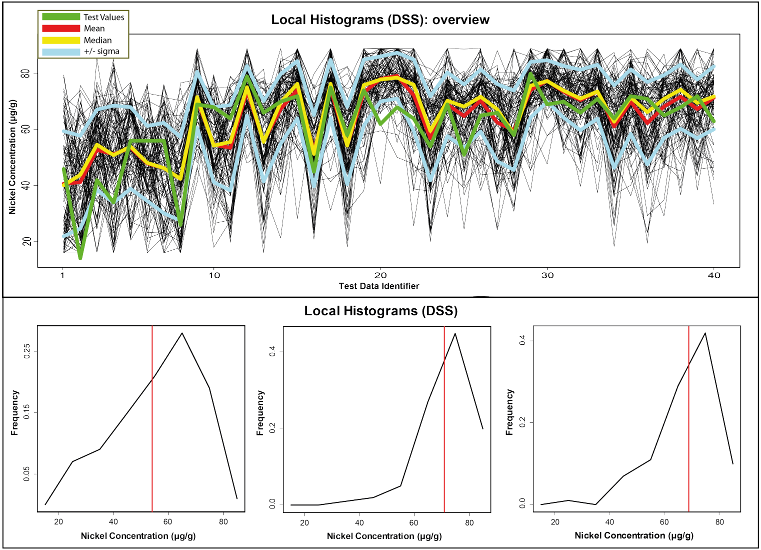
Reproduction of Global Statistics



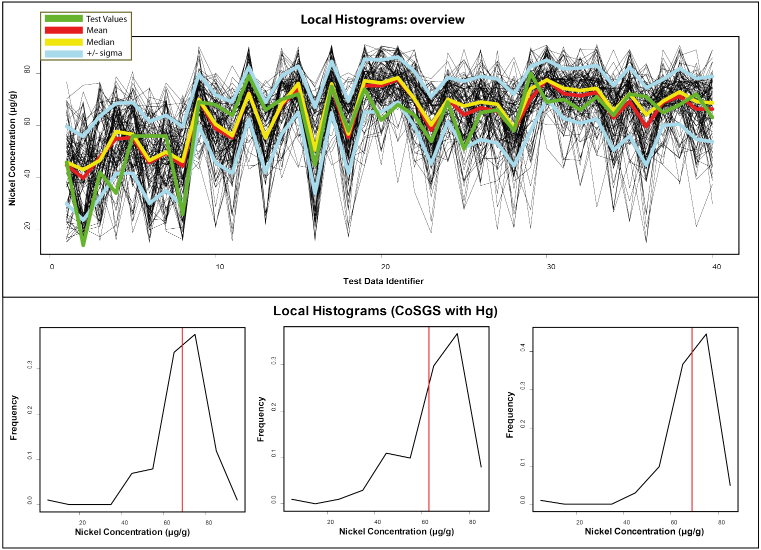
SGS applied to the Nickel Data



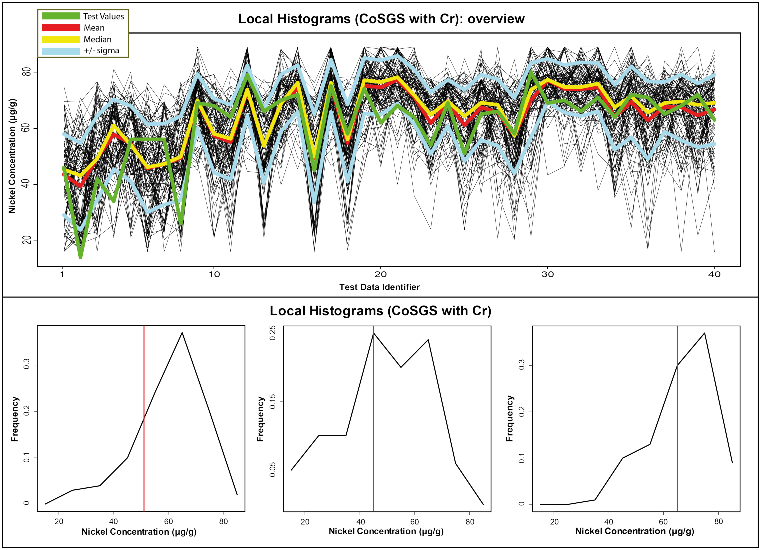
DSS applied to the Nickel Data



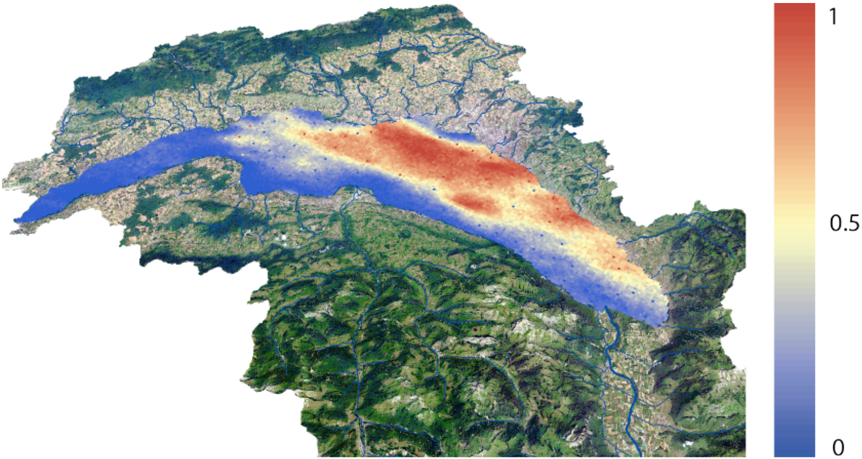
CoSGS applied to the Nickel Data with Hg



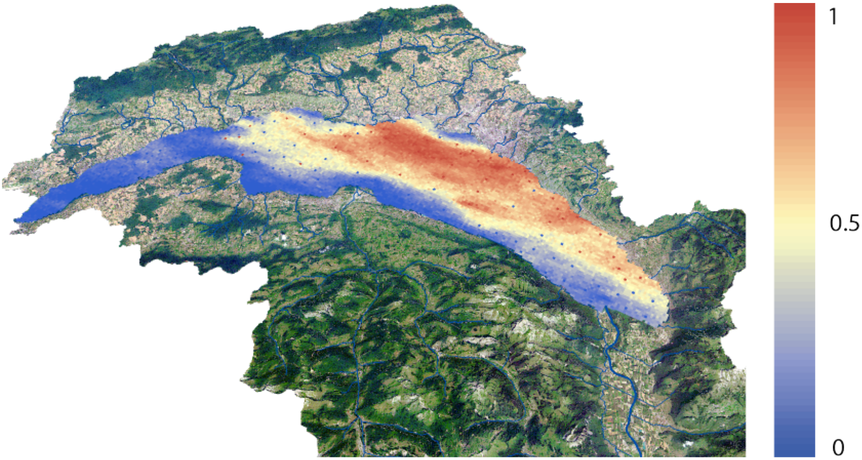
CoSGS applied to the Nickel Data with Cr



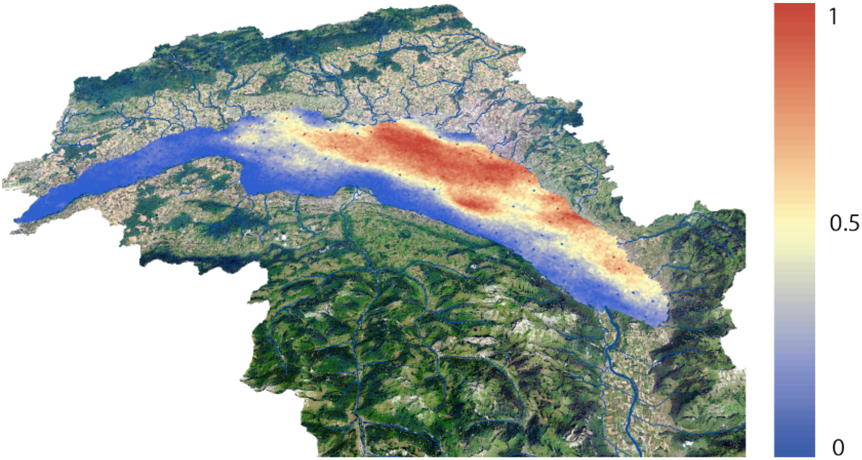
Probability of exceeding 70 μg/g (SGS_Ni)



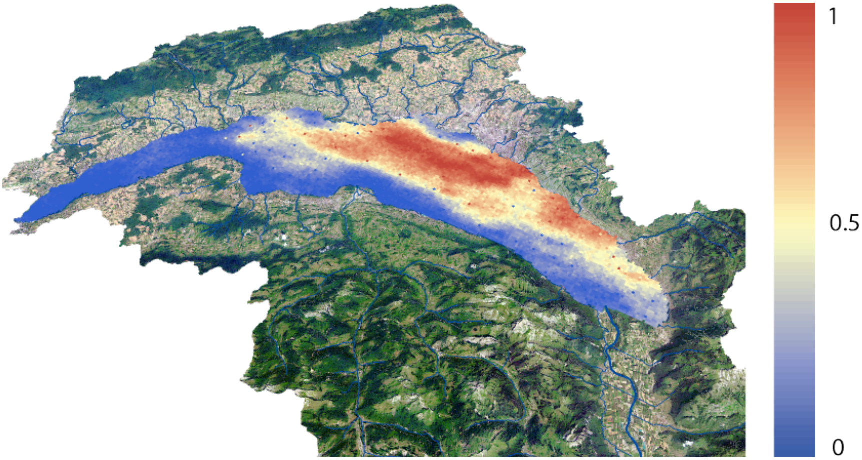
Probability of exceeding 70 μg/g (DSS_Ni)



Probability of exceeding 70 μg/g (CoSGS_Ni with Hg)



Probability of exceeding 70 μg/g (CoSGS_Ni with Cr)



Surface Above Different Thresholds

- In the previous slides, it was shown that the spatial distribution of the probability of exceeding 70 μg/g depended on the models.
- Here, it is highlighted that the choice of the model hardly affects the surface of the lake (in % of the total surface) exceeding given thresholds (based on Etype estimates).

Models	SGS_Ni	CoSGS_Hg	DSS_Ni	CoSGS_Cr
$\geq$50μg/g	84%	84%	84%	84%
$\geq$70μg/g	35%	36%	32%	33%
$\geq$80μg/g	0.2%	0.2%	0.1%	0.2%

Outline

Conclusion

- The models shown in this study are able to reproduce the global statistics of the nickel concentration data.
- The results of risk mapping provided by the models are different regarding the spatial distributions of probabilities, although the surface ratios exceeding given thresholds are similar.
- For future research, new variables of the same dataset will be analyzed and modeled, non-parametric algorithms will be tested and the visual analytics of the results will be developed.