Enrico Guastaldi

guastaldi@unisi.it

Geostatistical modelling of uncertainty for the risk analysis of contaminated sites

CGT, Centre for GeoTechnologies, University of Siena - Via Vetri Vecchi, 34
– 52027, San Giovanni Valdarno (AR) – Italy
Tel. +39 0559119483
Fax. +39 0559119439
e-mail: guastaldi@unisi.it
URL: www.geotecnologie.unisi.it

Geostatistical study

- **Study of the multivariate simulation of pollutants, in order to assess uncertainty for the risk analysis of a contaminated site.**
  - Terrain samples taken from boreholes analysed by a chemical laboratory for a remediation project of a steelworks in Settimo Torinese area
- **Data set comprises concentrations of several pollutants:**
  - Subset of ten organic and inorganic compounds was selected.
- **First part of study: univariate and bivariate statistical analysis**
  - Data set transformed to the Gaussian space so as to reduce the effects of extreme high values due to
    - contaminant hot spots and
    - because of the requirements of sequential Gaussian simulation.
- **Spatial correlation and cross-correlations from variography analysis**
  - This led to an hypothesized linear model of coregionalization for all variables.
- **Geostatistical simulations were applied in order to assess the uncertainty.**
  - Two types of simulation were performed:
    - Univariate Sequential Gaussian Simulation (SGS),
      - Correlation Correction Simulation and
    - Sequential Gaussian Co-Simulation (SGCOS).
- **Grade-tonnage curves were produced to assess basic environmental risk.**

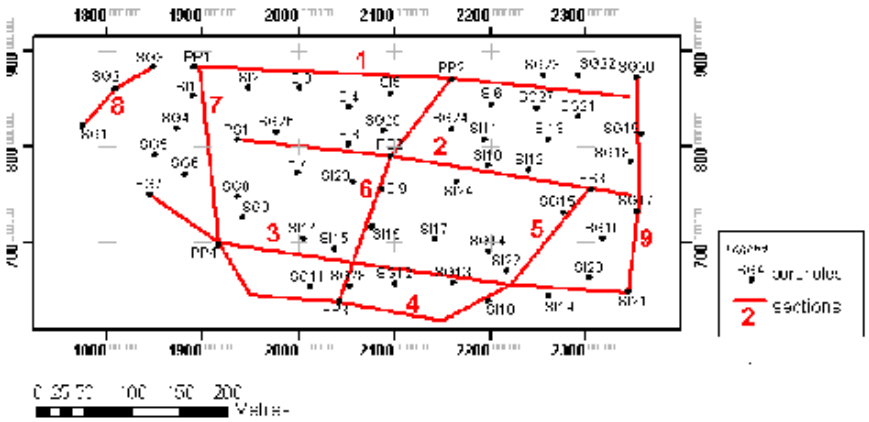
Objectives

- Assess the uncertainty of the spatial variability of contamination by heavy metals and heavy hydrocarbons in order to perform a risk analysis suitable for the reclamation a contaminated land site.

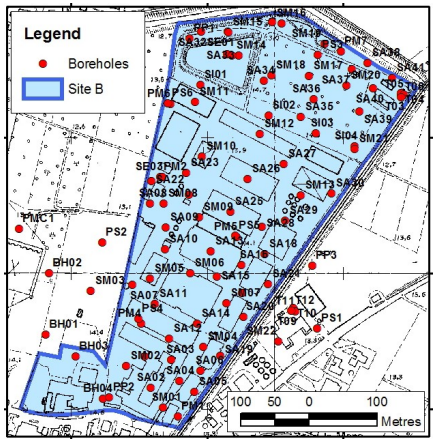
- In particular, the specific objectives of this study are to:
  - Analyse and interpret the data set;
  - Generate estimations of the average pollutant concentrations over blocks of terrain of a size suitable for reclamation and remediation;
  - Generate multivariate conditional simulations of the principal pollutants
  - Assess risk based on the conditional simulation approach
  - Provide recommendations for further sampling on the basis on the simulation results
  - Provide a basis for planning reclamation and remediation and identify future applications of this method for this particular site and for other similarly contaminated sites.

Project Overview

- Two case studies, North Italy, sampled in 2002
  - Site A: 10 environmental variables
  - Site B: 26 environmental variables



Site **A**



Site **B**

Main limitations of this study

•Both datasets used in this project tend to limit the potentiality of geostatistical investigation methods (usual condition when one has to deal with environmental data) (Guastaldi, 2005).

- both samples arrays were collected without a regular sampling grid, but just following some particular precept risen from several circumstantial fact, as survey experience of professionals in the field, geological aspects, a priori information on more contaminated areas, etc.

► Additionally, politico-economical reasons leaded to drill a relatively small number of boreholes outside both industrial areas of investigated sites.

- The two areas are similar as problems (two contaminated areas), even if they are quite different in terms of datasets (number of samples, core dimension, sampling grid, density of sampling, etc). So, different steps were completed at the beginning stages of work, in order to achieve dataset suitable for the further calculations to quantifying the uncertainty in contaminant concentrations for a preliminary reclamation study.

► Limitations to studying **Site A**

- Nature of the data set.
  - Boreholes are on an irregular sampling grid
  - Very few samples in each borehole and the Lengths of the boreholes vary from 5m to 20m, the lengths of cores vary from 0.1m to 1.0m) and sampling within a borehole is often discontinuous.
  - The usual approach is to recomposite the samples to approximately the same length but the discontinuous nature of much of the sampling made this difficult. A composite size of 0.5m was chosen and the data were recomposited to this length.
- Occasional presence of extremely high values of concentration of heavy metals and hydrocarbons
- Although this is exactly why a reclamation project is in progress, these outliers mask the underlying spatial variability quantified in variograms and cross-variograms. A significant amount of time was spent in investigating and overcoming the effects of outliers.
- Finally, this study must be considered as a parallel study to the reclamation project of this contaminated site and as an alternative methodology to be applied to similar case studies

Main limitations of this study

•Both datasets used in this project tend to limit the potentiality of geostatistical investigation methods (usual condition when one has to deal with environmental data) (Guastaldi, 2005).

- both samples arrays were collected without a regular sampling grid, but just following some particular precept risen from several circumstantial fact, as survey experience of professionals in the field, geological aspects, a priori information on more contaminated areas, etc.

► Additionally, politico-economical reasons leaded to drill a relatively small number of boreholes outside both industrial areas of investigated sites.

- The two areas are similar as problems (two contaminated areas), even if they are quite different in terms of datasets (number of samples, core dimension, sampling grid, density of sampling, etc). So, different steps were completed at the beginning stages of work, in order to achieve dataset suitable for the further calculations to quantifying the uncertainty in contaminant concentrations for a preliminary reclamation study.

► Limitations to studying **Site B**

► Nature of dataset. Data had no needs to be recomposited, however, their number in the vertical direction was no enough to carry on a complete three-dimensional study.

➢ Samples in each borehole weren't too much continuous for the whole length of boreholes, they are more concentrated in the upper part, near to the surface. Even if the superficial part is the most contaminated, the lower part should have been sampled more intensely, in order to found a truly representative "blank" distribution of each variable.

► Outliers had caused some problem to deal with the "natural" distribution of variable. However, there are no plant protection products in natural terrain, so it is impossible to eliminate the hot spots to get better distributions or better variograms, because the outliers are what we are looking for.

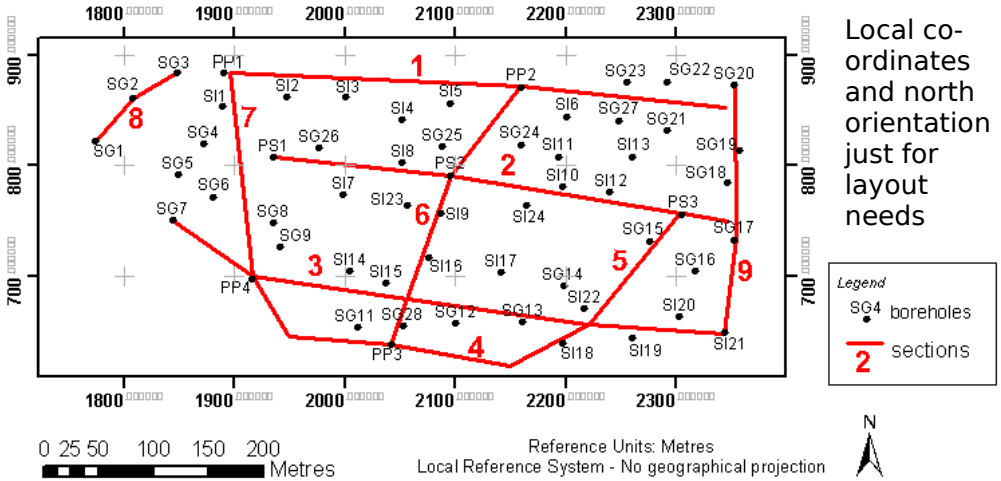
► Some variables have poor structure. Nevertheless these variables can be modelled in a coregionalization

➢ The most part of X-ESVs do not present good structures, even modellable.

➢ Directional ESVs are more erratic than the omnidirectional ones, as usual, because the large amount of data under the instrumental tolerance, the outliers and the general lack of data in vertical direction.

Dataset A: Boreholes and variables

Site **A**



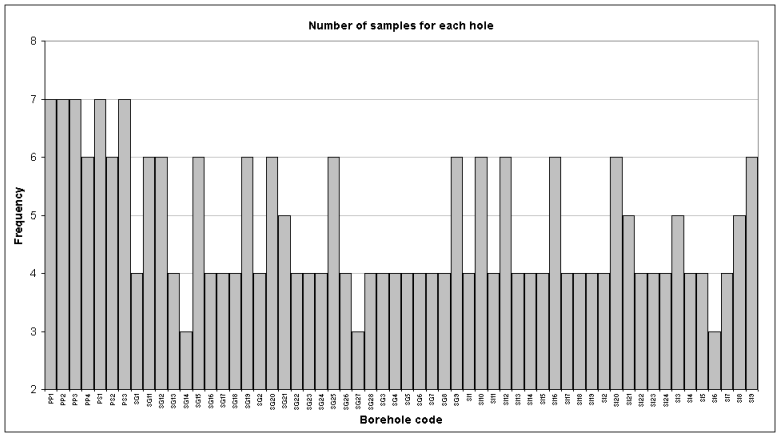
Local co-ordinates and north orientation just for layout needs

Code	*Frequency*	*Sampling collection methods*
SG	28	Percussion type boreholes (Geoprobe® type)
SI	24	Continuous log type boreholes
PP	4	Deep piezometers (~33m)
PS	3	Superficial piezometers (~22m)
BB	3	Surface samples for the determination of values in the bottom
WS	2	Bottom well sediments
	64	

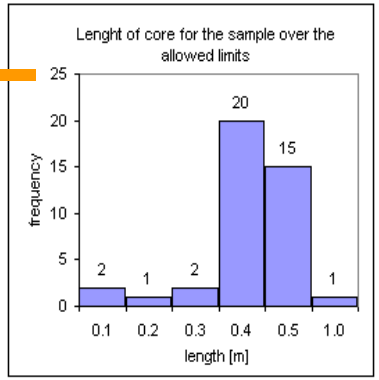
Studied variables:

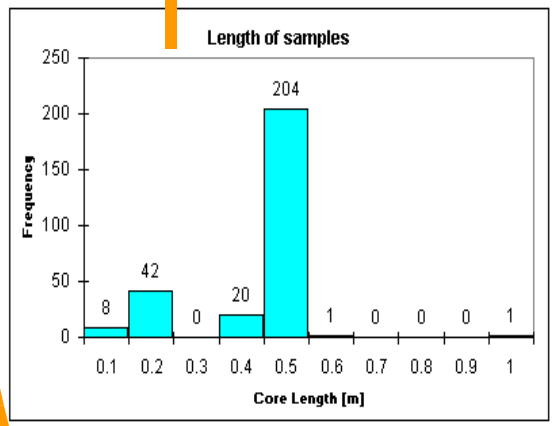
- Borehole code
- x co-ordinate
- y co-ordinate
- core lower limit
- core upper limit
- Cr [mg/kg]
- Co [mg/kg]
- Ni [mg/kg]
- Cu [mg/kg]
- Zn [mg/kg]
- As [mg/kg]
- Pb [mg/kg]
- Cd [mg/kg]
- Sn [mg/kg]
- Heavy Hydrocarbons [mg/kg] (called HY)

Dataset A: samples

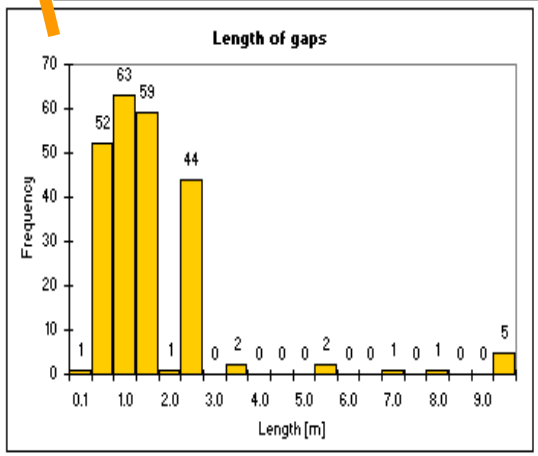


Samples Recomposition (=0.5m length)





Length of samples	
Mean	0.43
Standard Error	0.01
Median	0.50
Mode	0.50
Standard Dev.	0.13
Range	0.9
Minimum	0.1
Maximum	1
Sum	119.9
Count	276



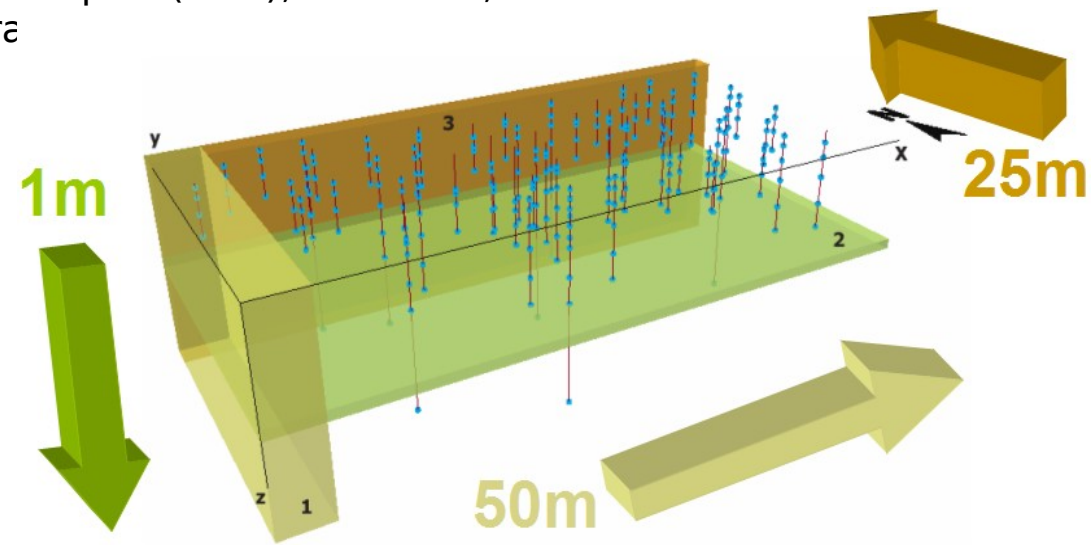
Length of gaps	
Mean	1.55
Standard Error	0.10
Median	1
Mode	1.5
Standard Deviation	1.55
Range	9.7
Minimum	0.1
Maximum	9.8
Sum	356.9
Count	231

- Number of samples per borehole varies between a minimum of 2 and 7 (on average 5)
  - The reason for the differing numbers of sample in boreholes is not known
  - No obvious pattern in the locations of the boreholes that have the most samples
  - They are fairly uniformly spread over the study area
- Length of samples in each drillhole varies from 0.1m to 1m
- Lengths of most gaps (i.e. missing data) are either between 0.5m and 1.5m, or between 2.5m and 3.0m.

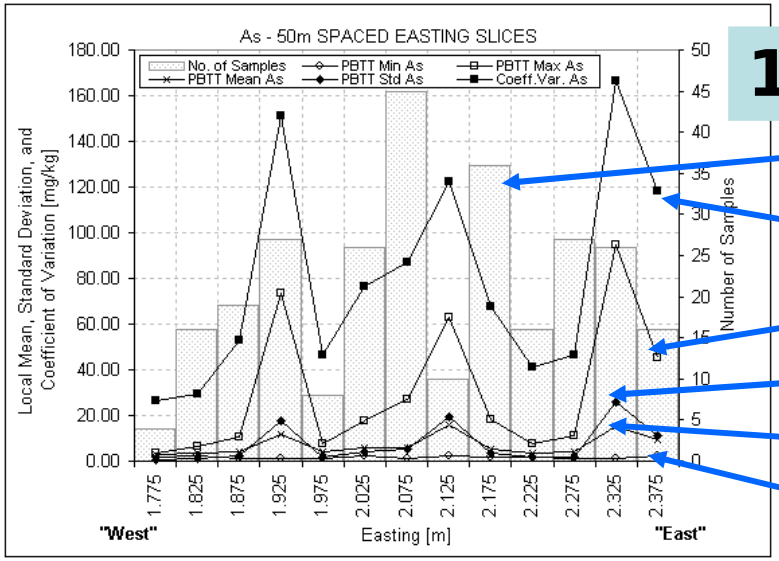
Site **A**

Spatial distribution of data

- The spatial distribution of the data is presented as:
  - Scatter plots of concentrations against depth
    - Scatter plot maps of the data values were plotted for variable to provide a rough indication of the variation in pollutant concentration with sampling depth
    - For each variable, two scatter plots were drawn showing concentration at two depth intervals (0.0-1.0m and 1.0m-2.0m), which were chosen because the most representative
  - Histograms of concentration and relative statistics of each variable under study.
    - Histograms based on slicing the three-dimensional volume.
    - For each "slice": number of samples (bars), minimum, maximum and mean value of concentra deviation and coefficient of variation
    - The results are plotted slice by slice in conventional histograms. The figure illustrates the manner in which the 3D volume is sliced and shows the notation used

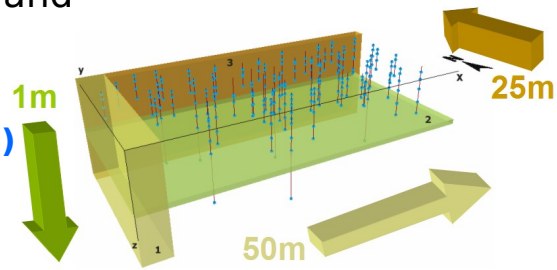


Spatial distribution of data



1

Variation No. of samples and principal descriptive statistics of Cr concentration



Number of samples (bars)

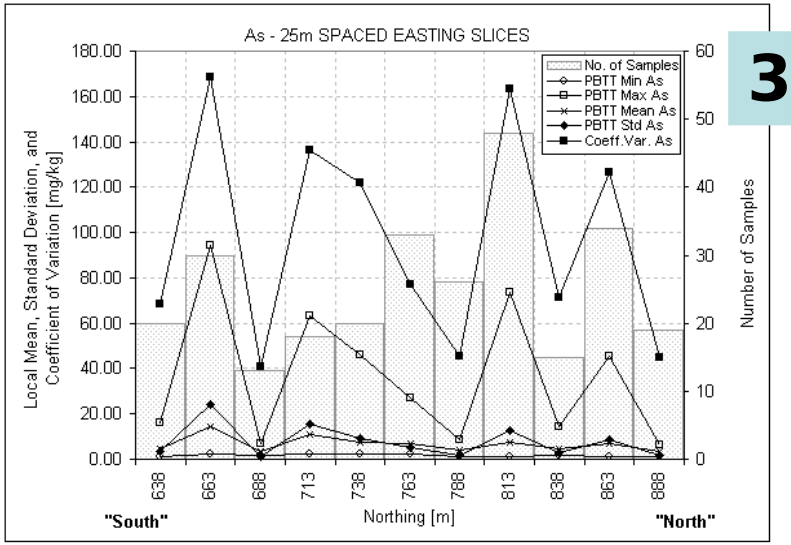
Coefficient of variation

Maximum value of concentration

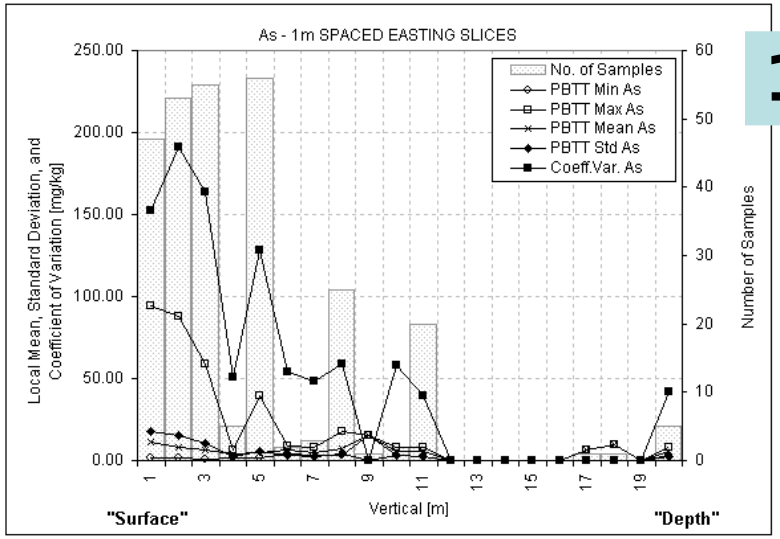
Standard deviation value of concentration

Mean value of concentration

Minimum value of concentration



3

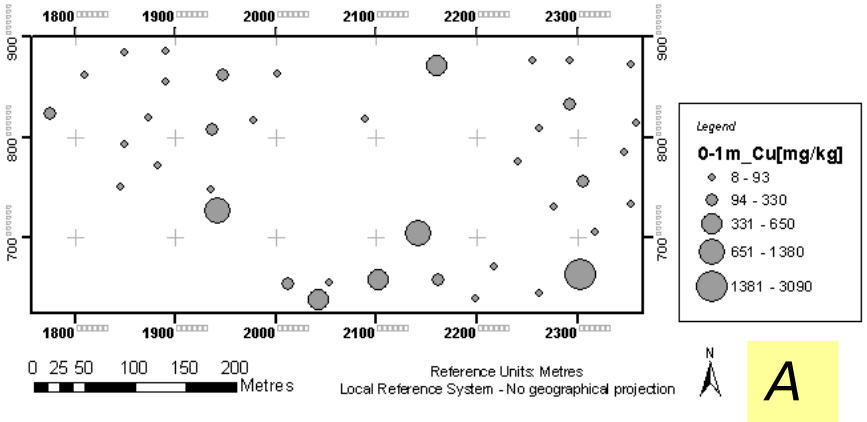


1

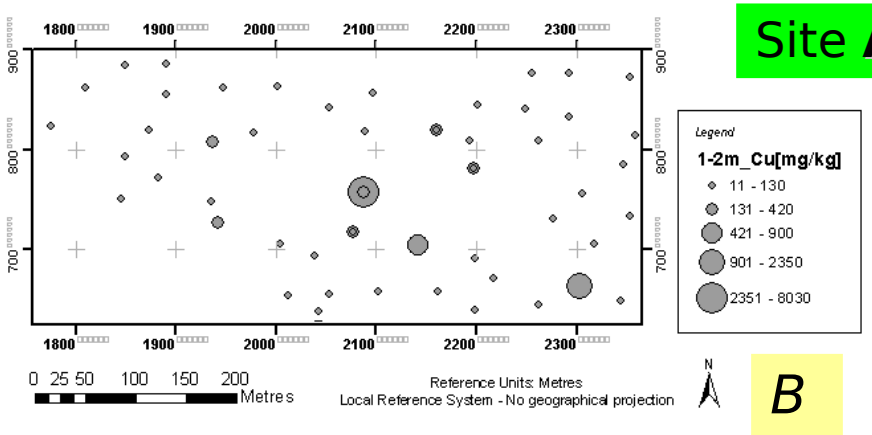
Spatial distribution of data

Scatter plots of **Cu** concentration at ***A***: 0-1m and ***B***: 1-2m

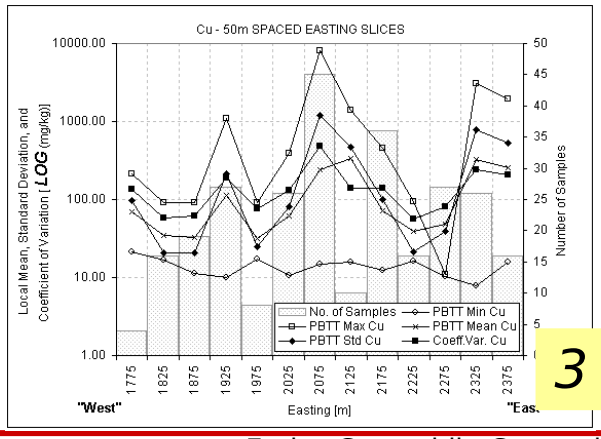
Variation No. of samples and principal descriptive statistics of **Cu** concentration



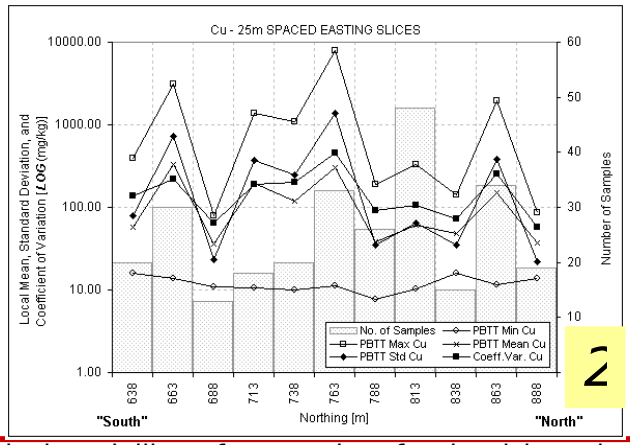
A



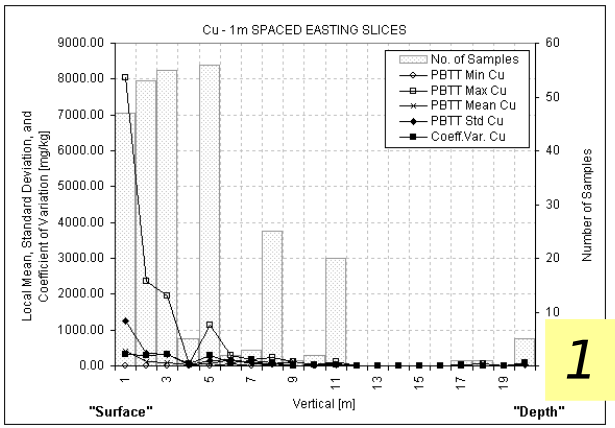
B



3



2



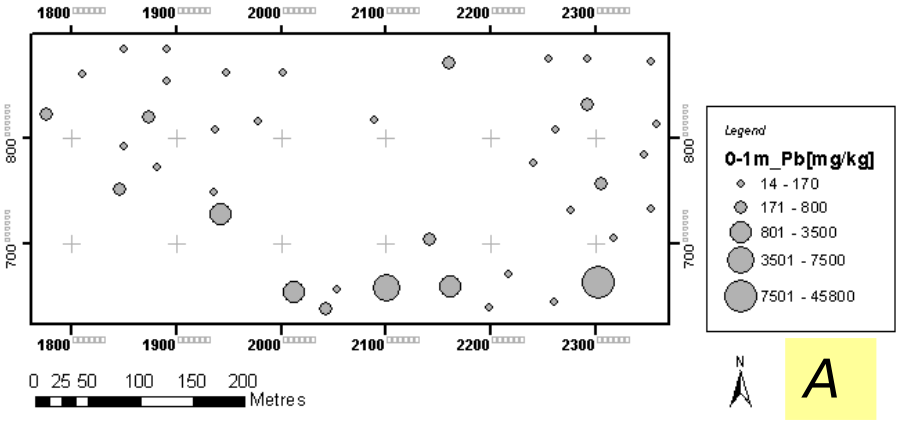
1

Spatial distribution of data

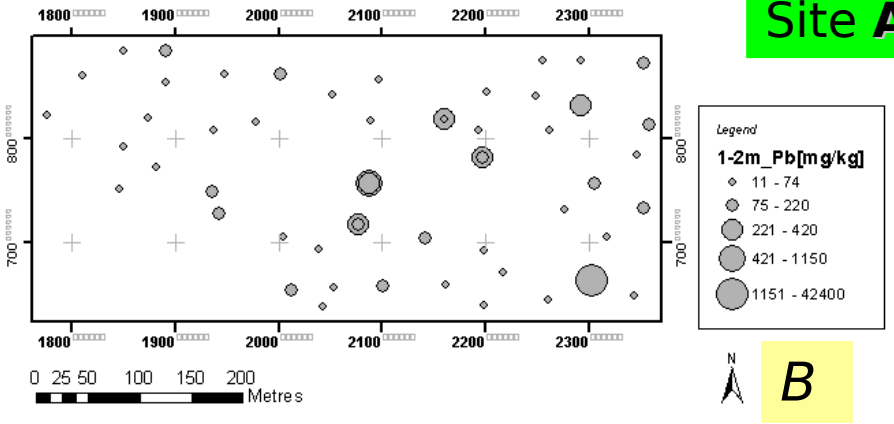
Scatter plots of **Pb** concentration at ***A***: 0-1m and ***B***: 1-2m

Variation No. of samples and principal descriptive statistics of **Pb** concentration

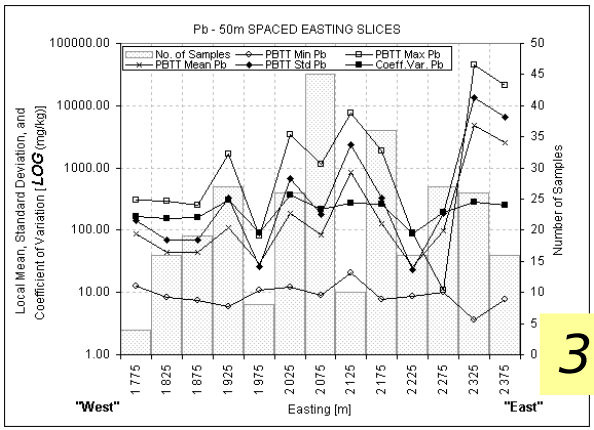
Site **A**



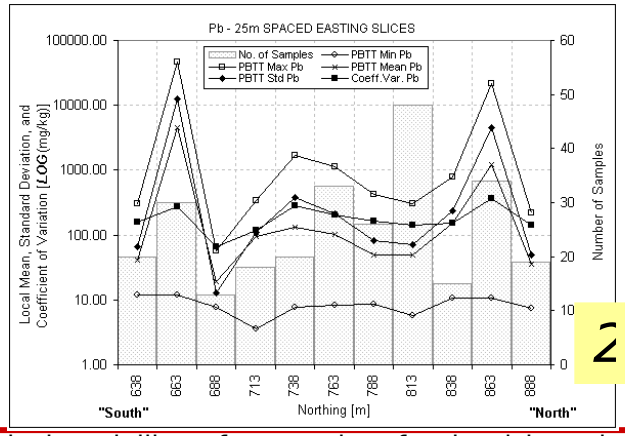
A



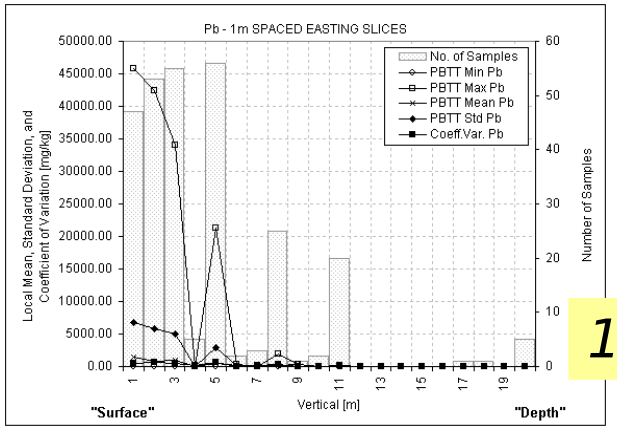
B



3



2



1

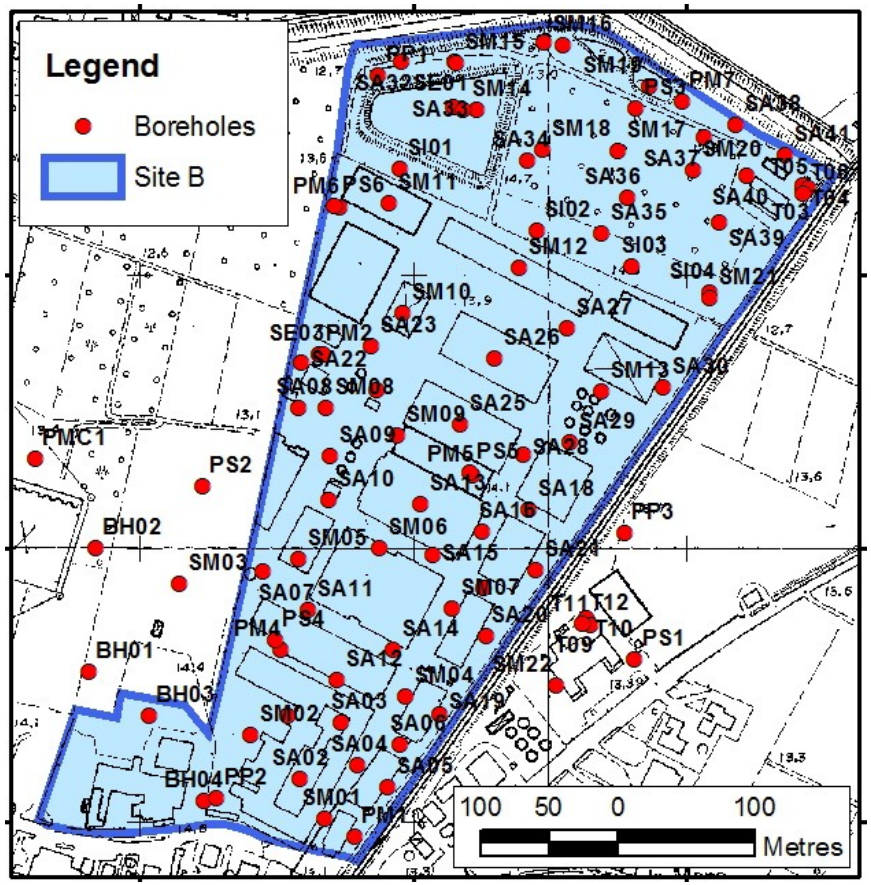
Univariate Statistical Analysis

Site A	Cr	Co	Ni	Cu	Zn	As	Pb	Cd	Sn	HY
Mean	294.91	21.59	269.26	134.54	4633.19	7.09	841.73	9.05	20.35	670.74
Standard Error	16.21	0.49	6.31	30.60	1839.33	0.82	337.34	3.74	4.35	181.17
Median	240.00	20.66	262.75	29.69	101.00	3.62	24.70	0.50	3.60	35.00
Mode	230	24	273	26	110	3.8	14	0.5	0.9	10
Stan. Deviation	242.66	7.29	94.37	458.01	27528.54	12.23	5048.79	56.04	65.15	2711.53
Variance	58882.5	53.2	8906.62	209769.7	757820314.4	149.6	25490275.6	3140.22	4245.01	7352421.1
Kurtosis	27.65	3.02	6.99	59.76	44.17	25.49	57.42	60.98	31.49	52.70
Skewness	4.92	1.10	1.69	7.13	6.68	4.74	7.45	7.63	5.34	6.67
Range	1930	48.773	723.48	4810.2	221486.4	93.6	45796.4	529.94	539.5	27594
Minimum	100	7.817	86.52	7.8	13.6	1	3.6	0.06	0.5	6
Maximum	2030	56.59	810	4818	221500	94.6	45800	530	540	27600
Sum	66060	4837.16	60315.18	30137.96	1037835.35	1588.4	188546.88	2027.89	4557.97	150246.2
Count	224	224	224	224	224	224	224	224	224	224
Conf.Lev (95%)	31.95	0.96	12.43	60.31	3624.68	1.61	664.77	7.38	8.58	357.03

Summary statistics of 0.5m recomposited raw data

- All frequency distributions show:
  - high positive values of skewness and high value of kurtosis
  - the mean, the median and the mode are never coincident
- Data are not from a Normal (Gaussian) distribution
- Gaussian transformation is required for sequential Gaussian simulation

Dataset B: samples

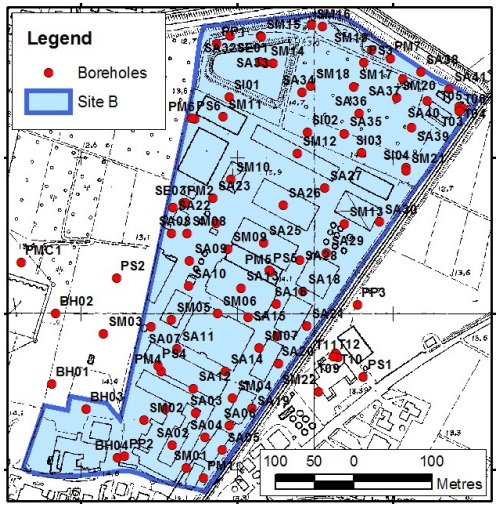


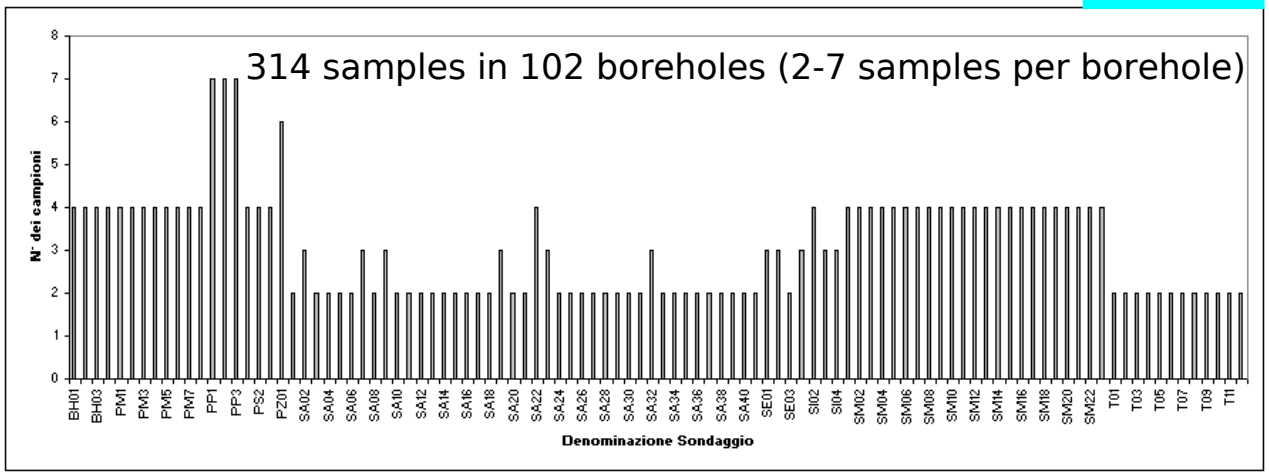
Site **B**

Variable	*Type of variable*
Alachlor	PPP
Aldrin	PPP
Atrazine	PPP
Cadmium	Heavy Metals
Chlordane	PPP
Total Chromium	Heavy Metals
DDD (diclorodifenildicloroetano) +**DDT** (diclorodifeniltricloroetano) +**DDE** (diclorodifenildicloroetilene)	PPP
Dichloromethane	PPP
Dieldrin	PPP
Endrin	PPP
Light Hydrocarbons	Hydrocarbons
Heavy Hydrocarbons	Hydrocarbons
PCB (Policlorobifenili)	PPP
Lead	Heavy Metals
Copper	Heavy Metals
α-HCH (α-HCH esaclorocicloesano)	PPP
β-HCH (β-HCH esaclorocicloesano)	PPP
Lindane	PPP
residue a 105°C	other
grain size (> 2mm)	other

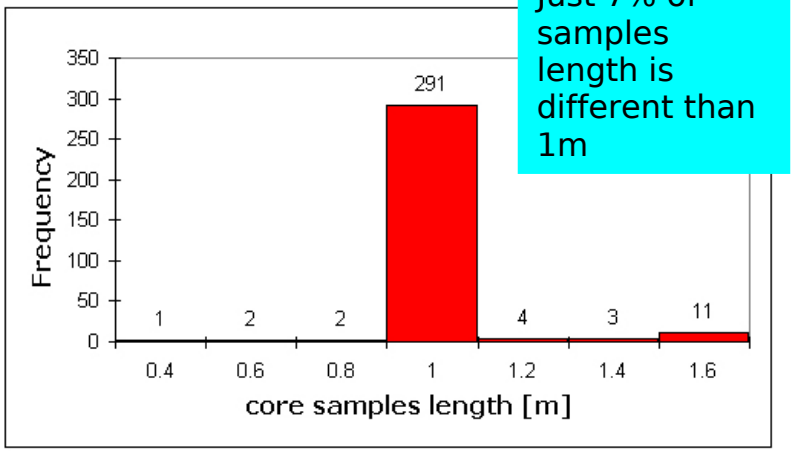
Dataset B: samples

Site **B**

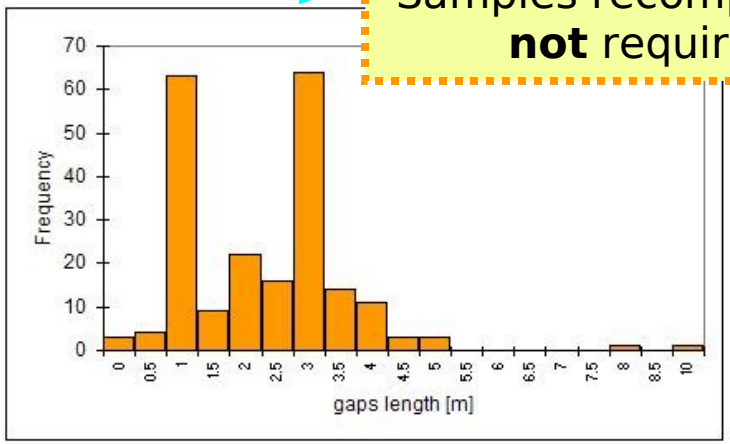




Just 7% of samples length is different than 1m



Samples recomposition **not** required



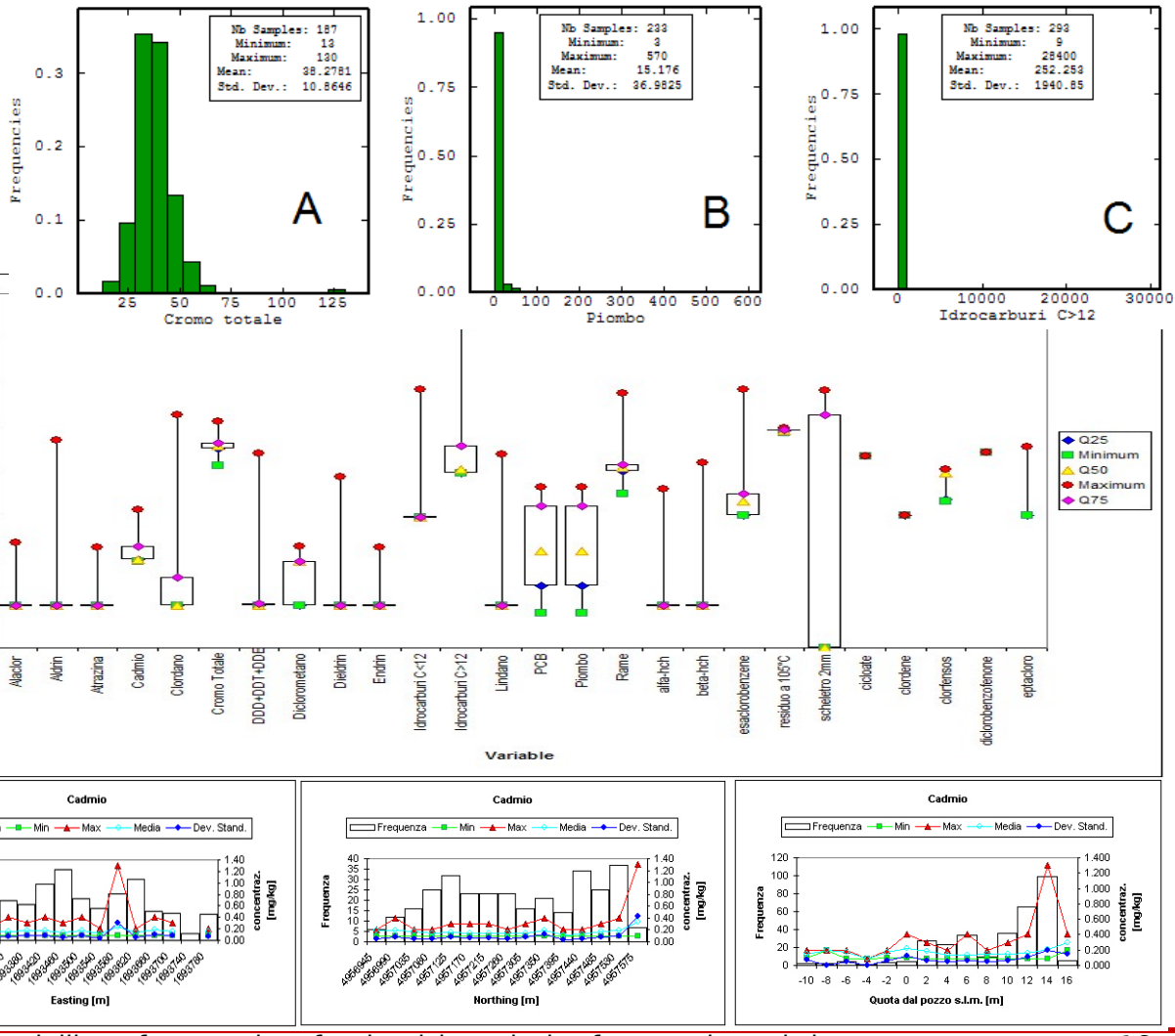
gaps	
Media	2.271963
Errore standard	0.085728
Mediana	2.5
Moda	3
Deviazione standard	1.254086
Varianza campionaria	1.572731
Curtosi	4.285883
Asimmetria	1.10637
Intervallo	9
Minimo	0
Massimo	9
Somma	486.2
Conteggio	214

Univariate Statistical Analysis

Site **B**

Three kinds of distribution

- Quasi-homogeneous (close to gaussian) (**A**)
- Positiv. Skewed (**B**)
- Very high positiv. skewed(most part of variables) (**C**)



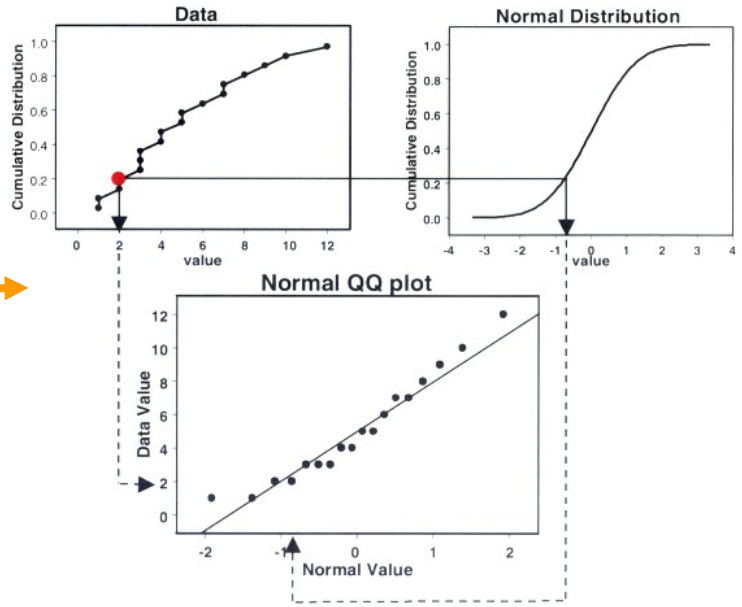
Univariate Statistical Analysis

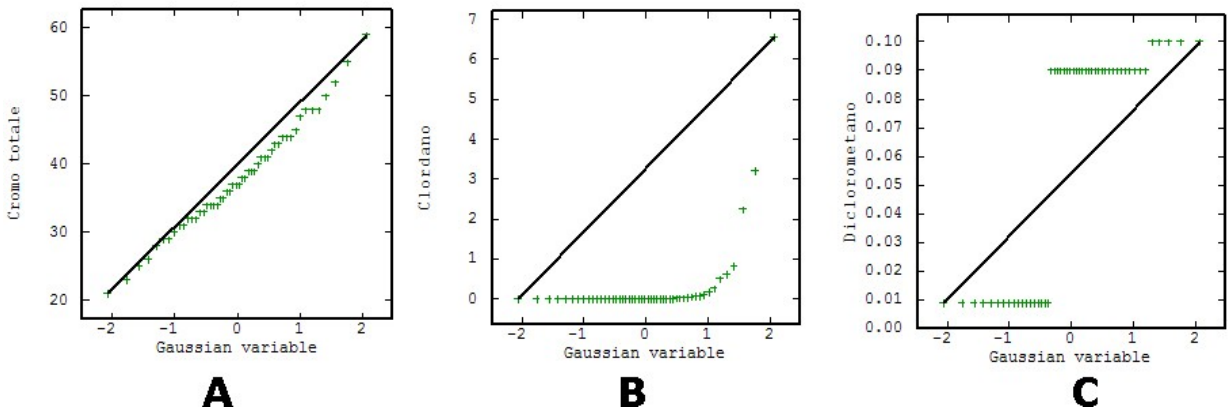
Site **B**

Three kinds of distribution

- Quasi-homogeneous (close to gaussian) (**A**)
- Positiv. Skewed (**B**)
- Very high positiv. skewed(most part of variables) (**C**)

Gaussian verifying





Correlation analysis

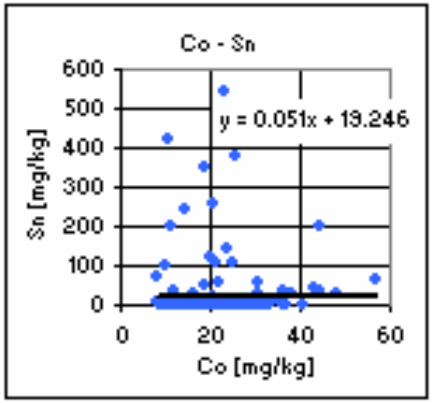
- This multivariate data set provides an opportunity to conduct a complete multivariate analysis and a multivariate co-simulation, which can form the basis of a much more realistic risk assessment than a sequence of independent univariate analyses.
- This observation is, of course, only valid if the two or more of the variables are correlated in situ and/or spatially correlated
- Correlations between pairs of variables can be assessed by scatter plots and by calculating the covariance and the **linear correlation coefficient** (Pearson Coefficient)

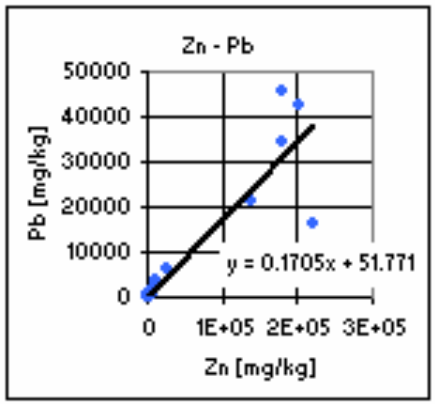
Site **A**

$$\rho = \frac{s_{xy}}{s_x \cdot s_y} \qquad s_{xy} = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{(n-1)} \qquad s_x = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{(n-1)}} \qquad s_y = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{(n-1)}}$$

	Cr	Co	Ni	Cu	Zn	As	Pb	Cd	Sn	HY
Cr	1									
Co	0.155	1								
Ni	0.204	0.403	1							
Cu	0.607	0.016	-0.088	1						
Zn	0.818	-0.065	-0.072	0.646	1					
As	0.739	0.241	-0.047	0.656	0.704	1				
Pb	0.833	-0.024	-0.076	0.659	0.930	0.753	1			
Cd	0.837	-0.037	-0.075	0.666	0.950	0.758	0.972	1		
Sn	0.716	0.006	-0.114	0.736	0.851	0.676	0.825	0.851	1	
HY	0.362	-0.038	-0.067	0.251	0.365	0.390	0.403	0.375	0.263	1

Correlation coefficients matrix of raw recomposited data of Site A





Correlation analysis

- This multivariate data set provides an opportunity to conduct a complete multivariate analysis and a multivariate co-simulation, which can form the basis of a much more realistic risk assessment than a sequence of independent univariate analyses.
- This observation is, of course, only valid if the two or more of the variables are correlated in situ and/or spatially correlated
- Correlations between pairs of variables can be assessed by scatter plots and by calculating the covariance and the **linear correlation coefficient** (Pearson Coefficient)

Site **B**

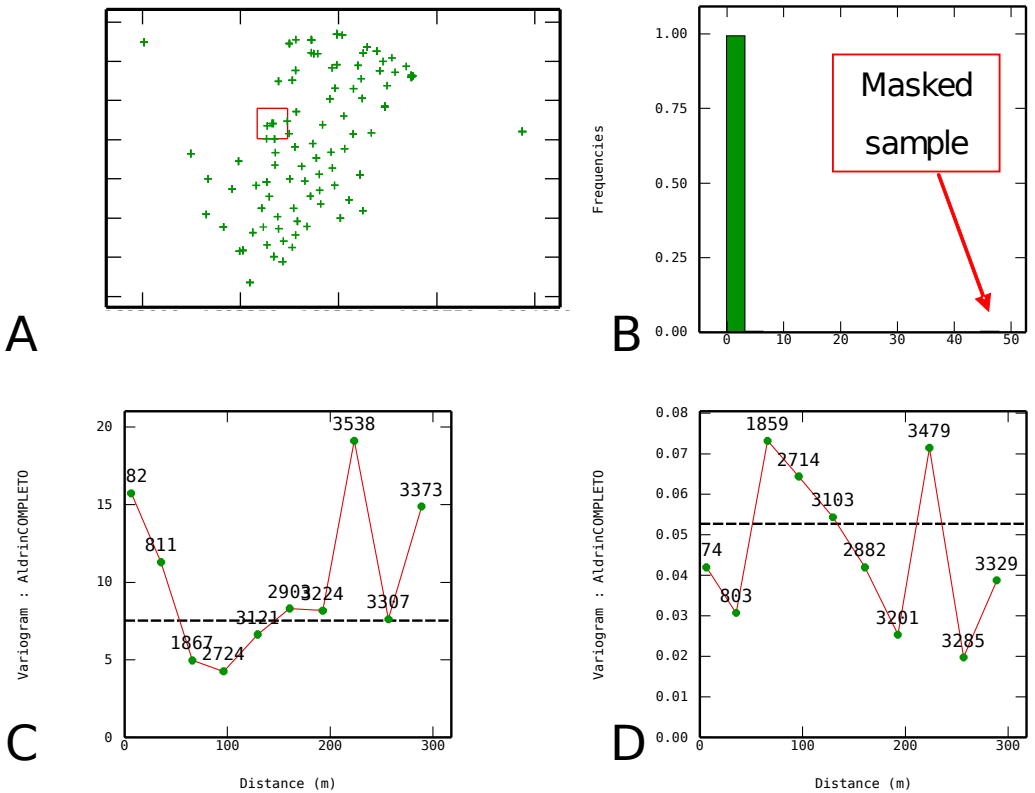
Correlation coefficients matrix of raw data of Site B

VARIABLE	Alaclor	Aldrin	Atrazina	Cadmio	Clordano	Cromo	DDD+DDT+DDE	Di cloro-metano	Dieldrin	Endrin	Idrocarburi C<12	Idrocarburi C>12	Lindano	Piombo	Rame	alfa-hch	beta-hch	residuo a 105°C	scheletro 2mm
Alaclor	1																		
Aldrin	0.053	1																	
Atrazina	0.668	0.05	1																
Cadmio	0.51	0.217	0.498	1															
Clordano	0.169	0.076	0.204	0.756	1														
Cromo	0.414	0.123	0.401	0.675	0.603	1													
DDD+DDT+DDE	0.435	0.076	0.615	0.758	0.61	0.61	1												
Diclorometano	-0.06	-0.02	0.085	0.039	-0.07	-0.04	0.01	1											
Dieldrin	0.539	0.077	0.673	0.755	0.579	0.612	0.91	0.004	1										
Endrin	0.72	0.054	0.968	0.502	0.211	0.414	0.628	0.061	0.696	1									
Idrocarburi C<12	0.234	0.025	0.339	0.087	0.229	-0	0.711	0.062	0.423	0.345	1								
Idrocarburi C>12	0.633	0.072	0.888	0.662	0.484	0.538	0.762	0.033	0.88	0.906	0.258	1							
Lindano	0.491	0.003	0.698	0.026	0.026	0.01	0.068	0.065	0.079	0.711	0.102	0.472	1						
Piombo	0.536	0.079	0.742	0.787	0.312	0.648	0.87	0.043	0.938	0.756	0.093	0.909	0.16	1					
Rame	0.085	0.009	0.122	0.243	0.538	0.115	0.408	0.134	0.338	0.126	0.926	0.284	0.152	0.197	1				
alfa-hch	0.589	0.073	0.836	0.701	0.244	0.577	0.6	0.06	0.806	0.851	0.046	0.94	0.379	0.963	0.153	1			
beta-hch	0.003	0.997	0.01	0.164	-0	0.079	0.004	0.038	0.005	0.012	0.007	0.013	0.009	0.011	-0	0.011	1		
residuo a 105°C	0.195	0	0.046	0.163	-0.13	-0.54	-0.03	-0.43	-0.016	0.096	0.024	-0.02	0.111	-0.301	-0.17	0.033	0.058	1	
scheletro 2mm	0.184	0	0.267	0.337	0	-0.22	0.235	-0.18	0.251	0.298	0.262	0.226	0.199	0.185	0.012	0.31	0.285	0.564	1

Transform of data in Gaussian Space

- The data set for all the variables contains outliers which can strongly affect the statistical analysis and the subsequent geostatistical analysis. There are several ways of dealing with these outliers (Goovaerts, 1997):

  - Consider them as erroneous data and remove them from the data set
  - Consider them as coming from a different population
  - Use robust statistics which are less sensitive to outliers
  - **Transform the data** to a different space to reduce the effects of outliers.
  - Discard the extreme values and reduce the data set to a more representative set of sample values.



Transform of data in Gaussian Space

- The Normal distribution can be completely defined by two parameters of the distribution: **mean** and **standard deviation**, which for a standard cumulative density function are zero and one respectively

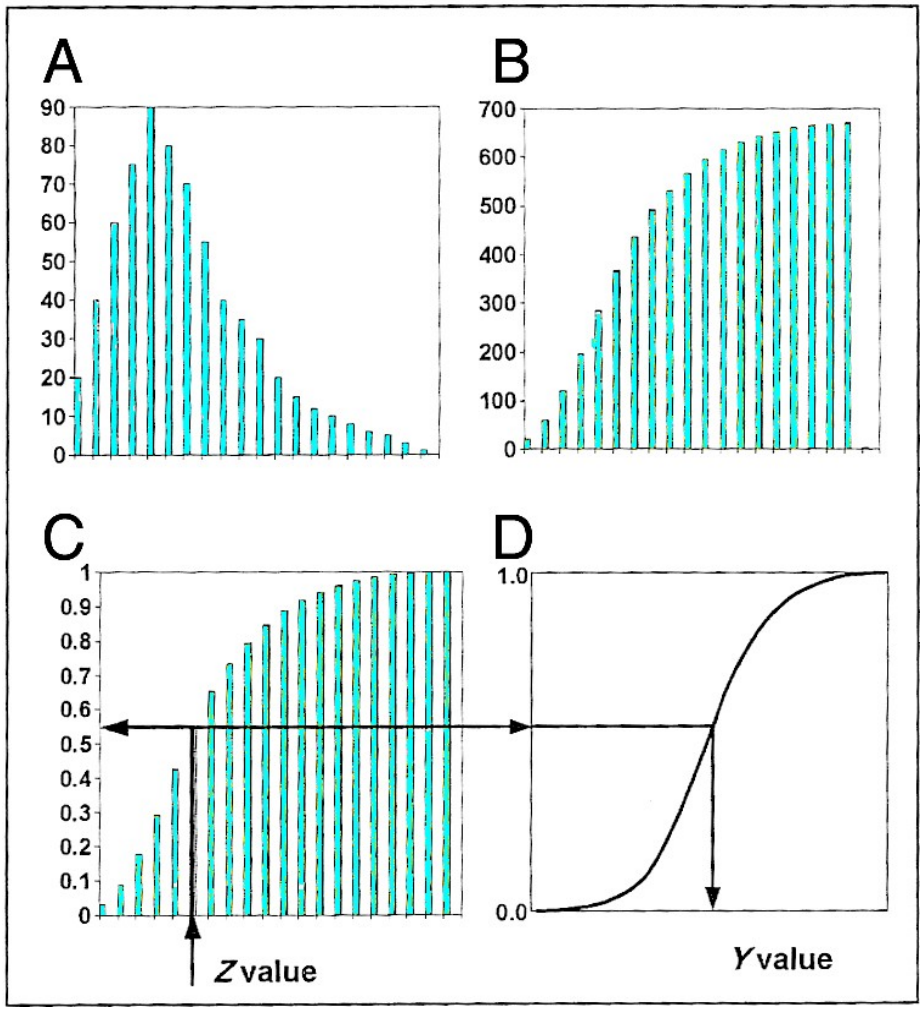
- Several methods to transform into Gaussian space:
  - Log-normal transform
  - ...
  - **Hermite polynomials transform** (used for the most part of Site B's variables) Site **B**
  - **Normal score transform** (used for the most part of Site A's variables) Site **A**
    - Non-parametric method
    - It is used to transform the data into the Normal space, and to back-transform the data after the estimation and/or simulation calculations
    - It does not require the strong assumptions (e.g. needed for the log-normal transform)
    - It provides satisfactory results, in terms of statistics and normal-probability plots

Normal Scores Transform

- Three phase forward Normal Scores transform (Goovaerts, 1997):
  - ranking original raw $z(x)$ data
  - Cumulative frequency of $z(x)$
  - Normal Scores transform of $z(x_k)$ of k^{th} position: coincidence of k^{th} quantile of ranked data with the corresponding quantile of standard Gaussian cumulative function.
    - This yield a set of transformed $y(x)$ data from -5 to +5, in Normal space.

$$Z(x) = \varphi[Y(x)]$$

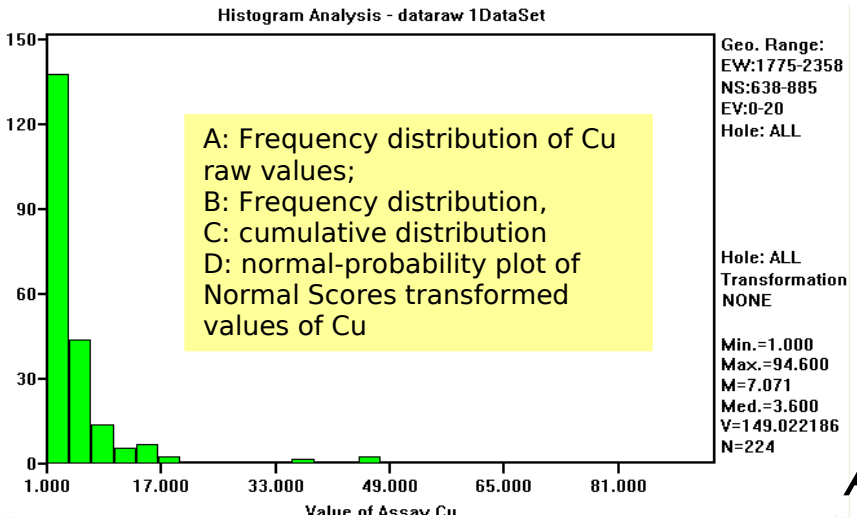
$$Y(x) = \varphi^{-1}[Z(x)]$$

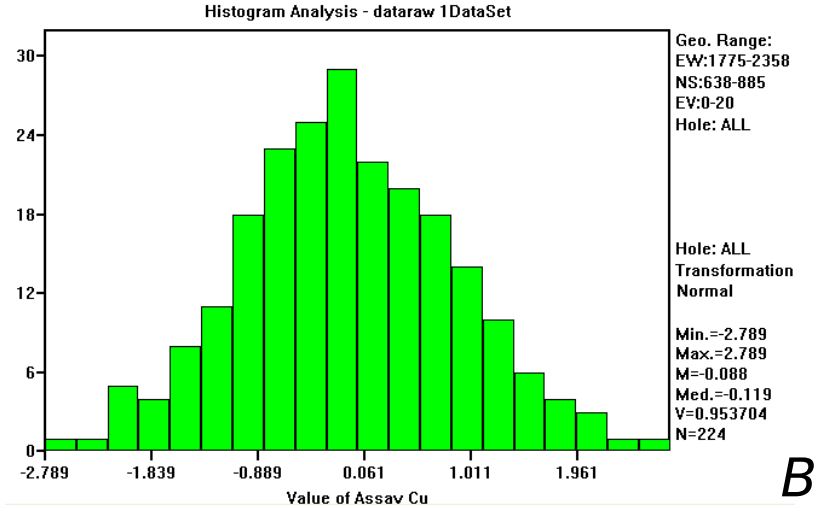


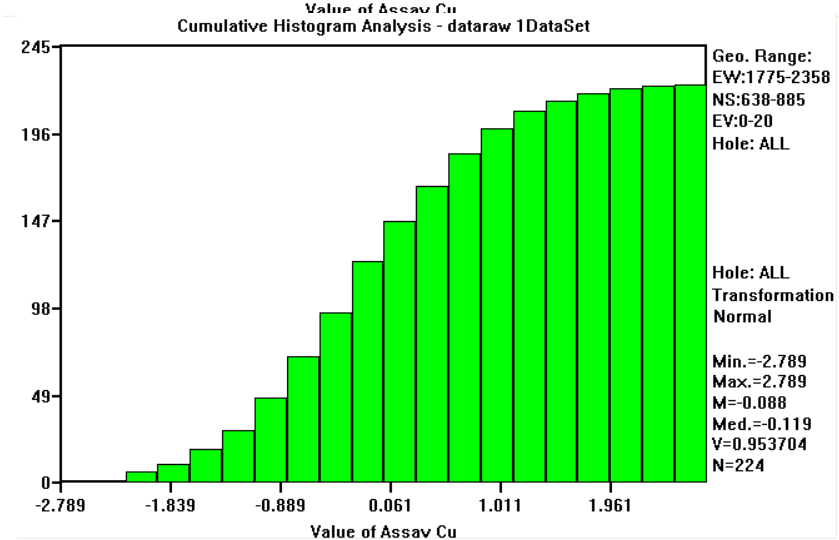
Normal Scores Transform

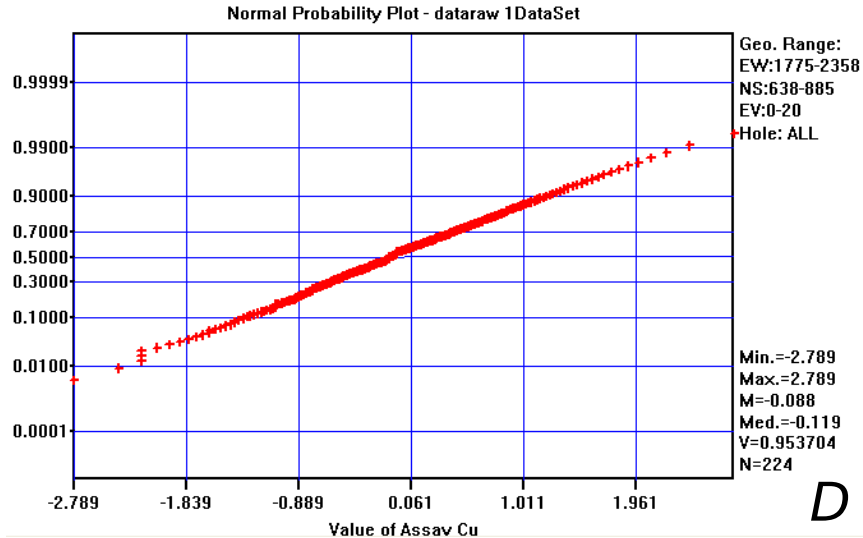
Site **A**

6 variables have means relatively close to zero and standard deviations close to one:
Co, Ni, Cu, Zn, As, Pb





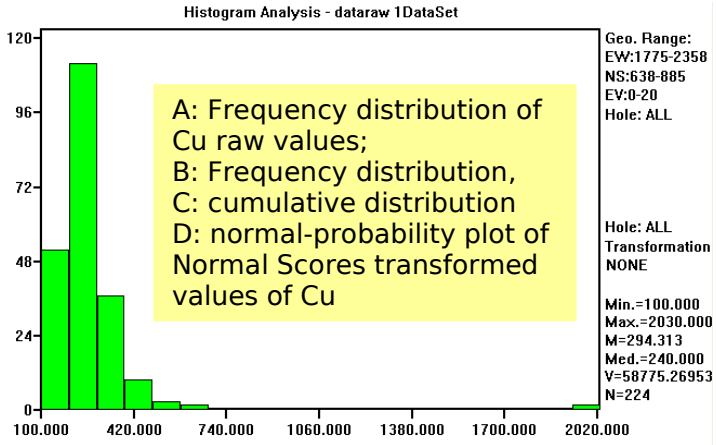




Normal Scores Transform

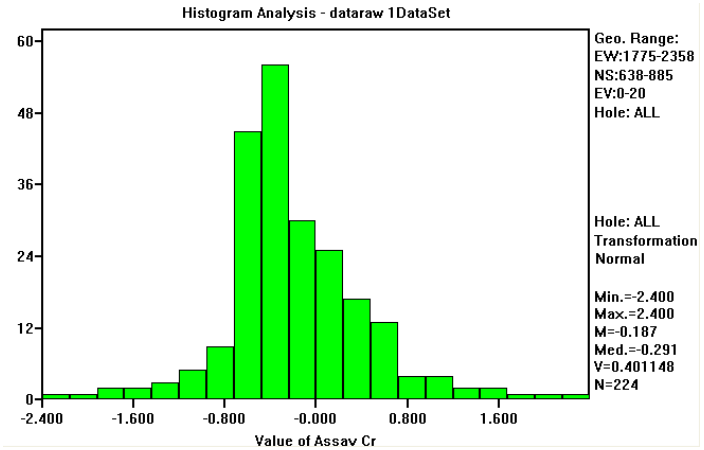
Site **A**

- However there are **4** variables - **Cr, Cd, Sn** and **HY** - for which this is not the case.
- For **Cd, Sn** and **HY** this can be explained by the use of a default minimum value equal to the lower detection limit of the analysis machine.
- **Cr**, however, is always well sampled and there are no apparent reasons for these differences, unless there are two or more populations of **Cr** concentration.

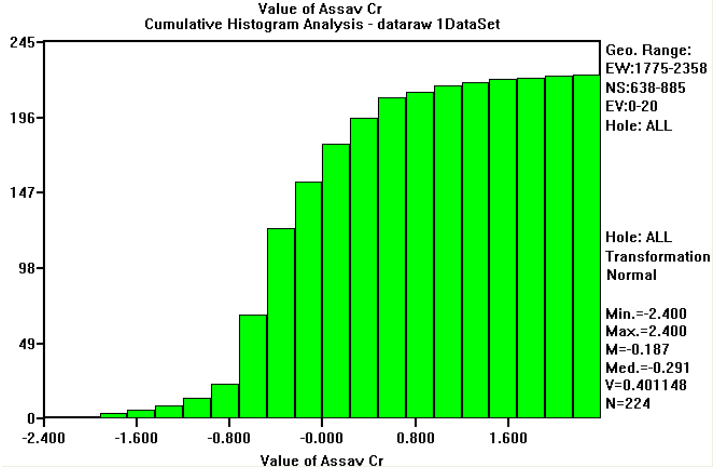


A: Frequency distribution of Cu raw values;
B: Frequency distribution,
C: cumulative distribution
D: normal-probability plot of Normal Scores transformed values of Cu

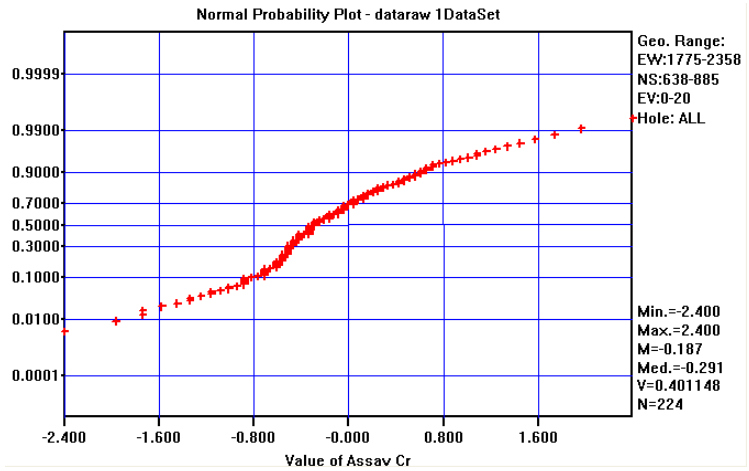
A



B



C



D

Hermite polynomials expansion

Transform function:

$$F\left[\varphi(y)\right]=G(y)$$

$$\Phi(y)=F^{-1}\left[G(y)\right]$$

$$\Phi^{-1}(z)=G^{-1}\left[F(z)\right]$$

Polynomials expansion of transform function

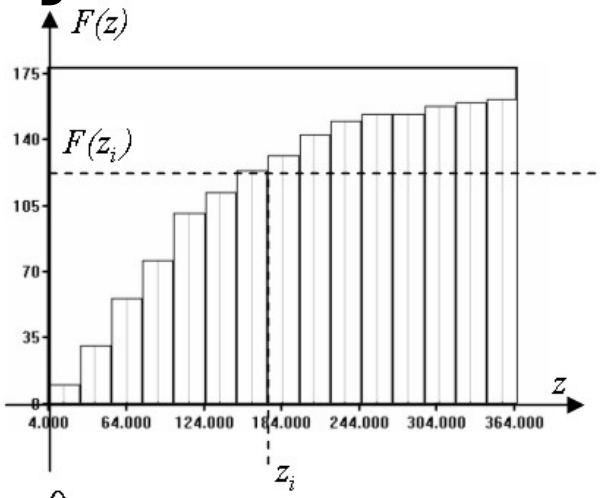
$$\Phi(y)=\sum_i b_i g^{(i)}(y)=\sum_i \frac{c_i}{i!}H_i(y)$$

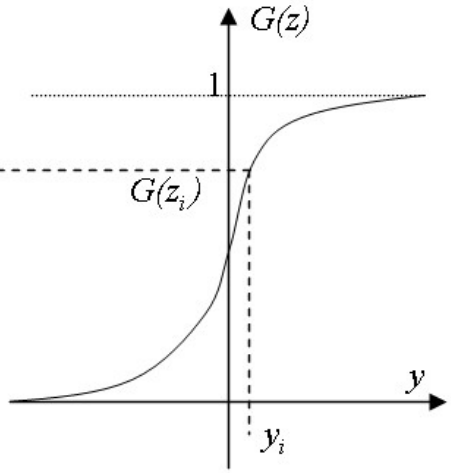
$$c_i=\int_{-\infty}^{+\infty}\Phi(y)H_i g(y)dy$$

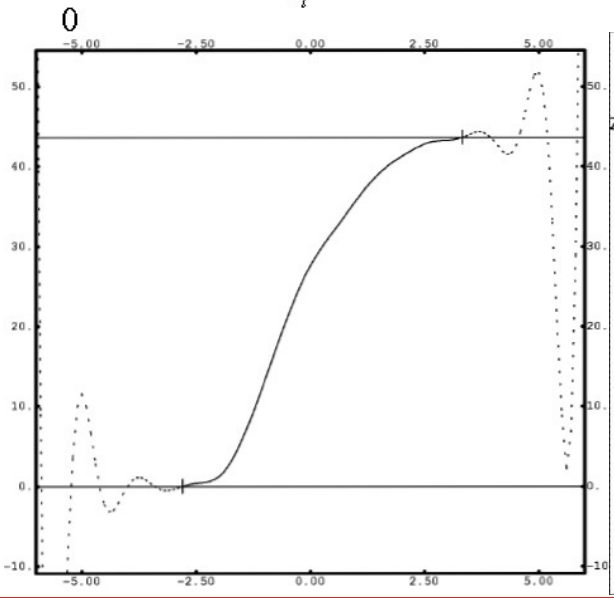
$$H_i=\frac{1}{\sqrt{i!}g(y)}\frac{d^i g(y)}{dy^i}$$

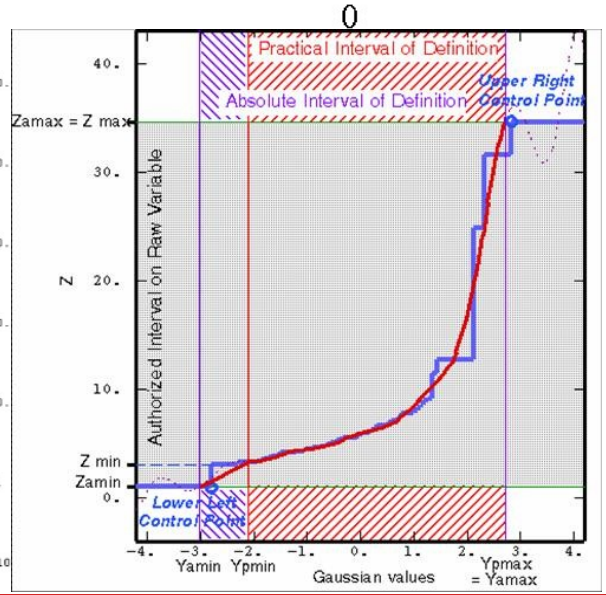
Mean: $E\left[\Phi(y)\right]=c_0$

Variance: $D^2\left[\Phi(y)\right]=\sum_{1=1}^{+\infty}\frac{c_i^2}{i!}$





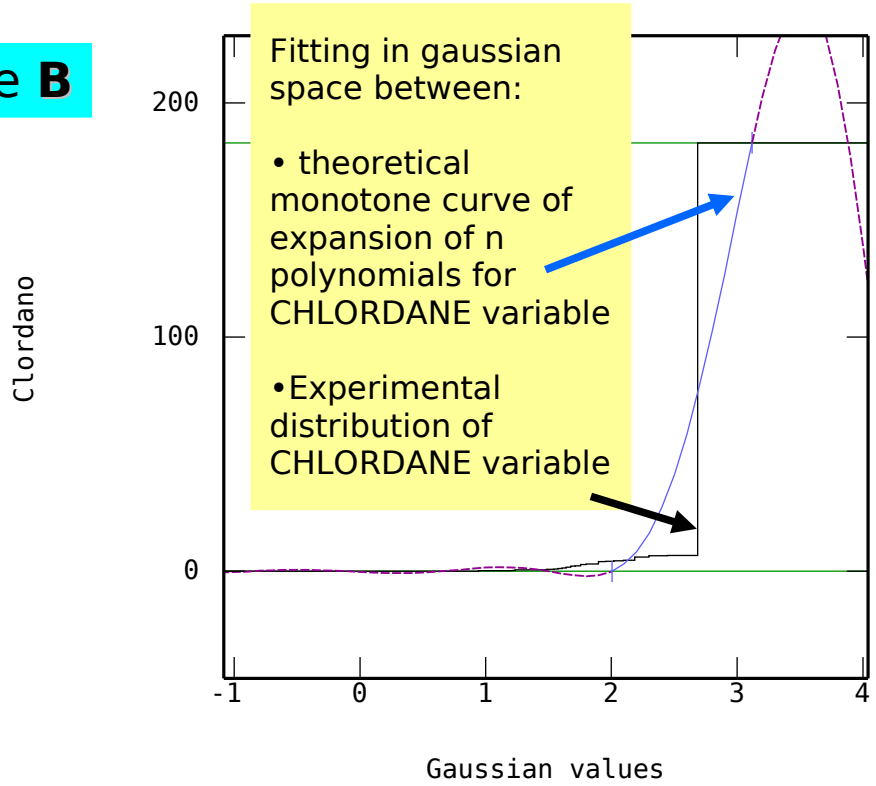


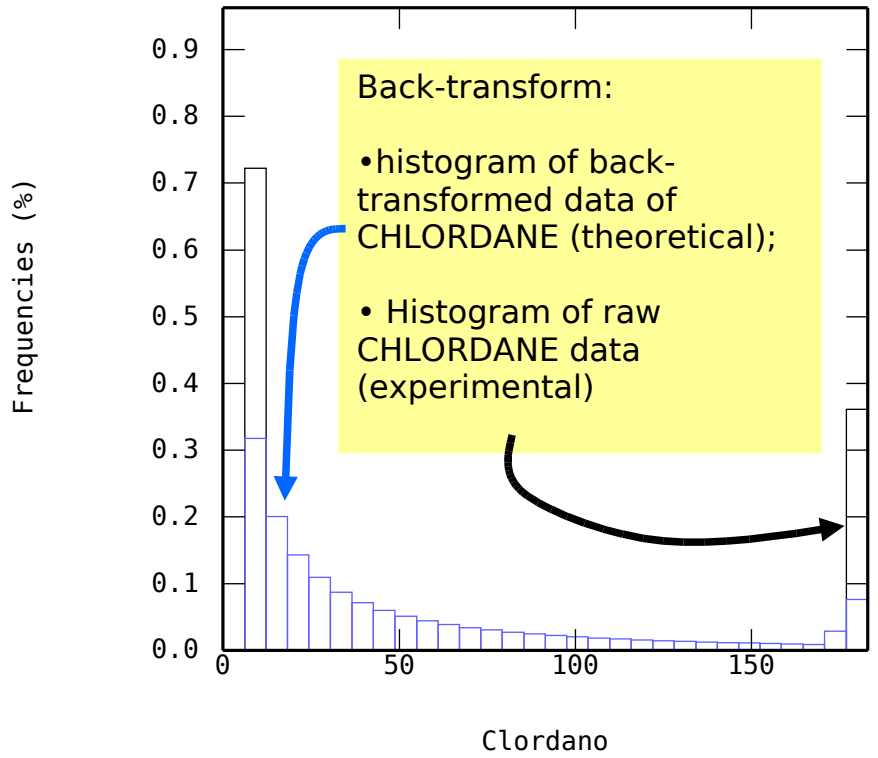


Hermite polynomials expansion

- Hermite polynomials expansion gave better results than Normal Scores on the most part of Site B's variables
  - **Back-transformed Values Distribution** (theoretical histogram) similar to the original one (raw data histogram) (Bleines et alii, 2004)

Site **B**





Hermite polynomials expansion

- Hermite polynomials expansion gave better results than Normal Scores on the most part of Site B's variables
  - ► **Back-transformed Values Distribution** (theoretical histogram) similar to the original one (raw data histogram)
  - ► **Definition intervals of variables** and Hermite polynomials utilised for variables transforming

Site **B**

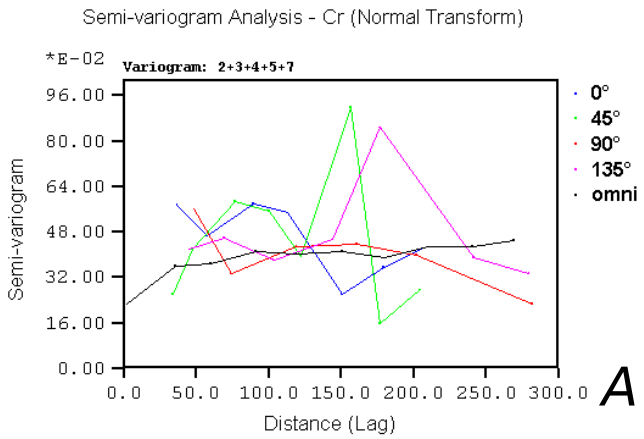
Variable	Definition of practical interval of				Number of Hermite polynomials
	Gaussian variable		raw variable		
	min	max	min	max	
ALACLOR	1.19	2.91	0.000000001	0.24	37
ALDRIN	1.45	2.71	0.000000001	0.18	30
ATRAZINE	1.36	2.9	0.000000001	0.06	40
CADMIUN	-1.66	2.19	0.000000001	0.4	20
CHLORDANE	2	3.12	0.000000001	183	17
TOTAL CHROMIUM	-2.77	2.63	13	65	30
DDD+DDT+DDE	0.62	3.1	0.000000001	14	17
DICHLOROMETHANE	0.83	3.21	0.000000001	0.2	17
DIELDRIN	0.98	2.92	0.000000001	2.84	33
HEAVY HYDROCARB.	1.37	3.25	0.000000001	8440	10
LINDANE	1.75	2.88	0.000000001	0.1	16
COPPER	1.65	3.12	7	3300	13

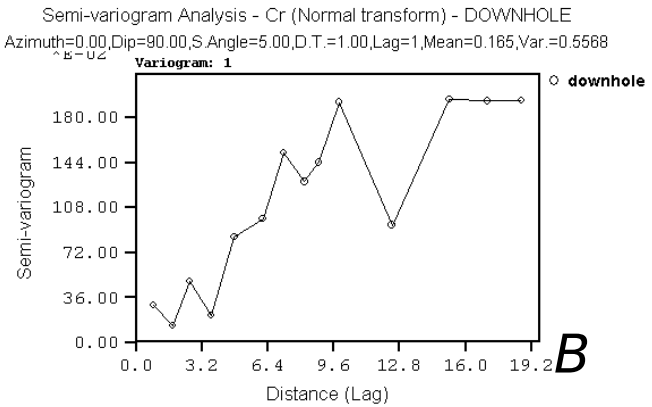
Structural analysis: Univariate approach

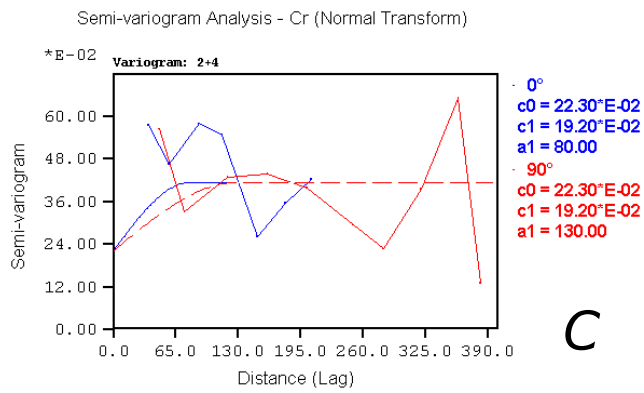
- **Structural analysis** is to
  - ► generate three dimensional Experimental Semi-Variograms (ESV) in the Normal space
  - ► fit models to them and
  - ► interpret the models in the context of the local geology and other possible factors conditioning the spatial distribution of the pollutants.
- However, the univariate approach does not allow to consider unsampled locations and spatial cross-correlation

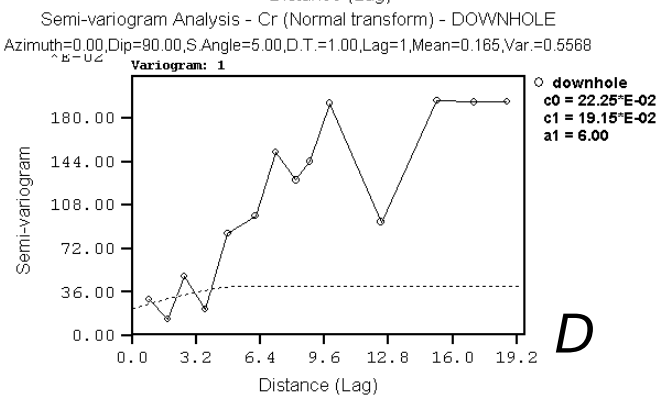
Site **A**

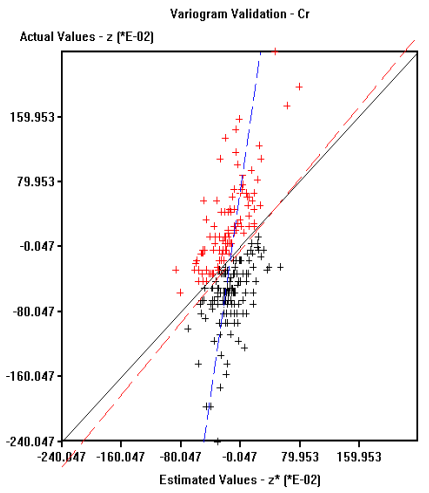
Structural analysis of **Cr**.











Linear Coregionalization Model

- **Experimental Semi-Variogram (ESV)** allow us to quantify the Spatial correlation, i.e. the measure of spatial variability of regionalized variable $Z(\mathbf{x})$ between samples separated by vector $\mathbf{h}$.

$$2\gamma^*(\mathbf{h}) = \mathrm{Var}\{Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h})\} = \mathrm{E}\{[Z(\mathbf{x}) - Z(\mathbf{x}+\mathrm{h})]^2\} - \mathrm{E}\{[Z(\mathbf{x}) - Z(\mathbf{x}+\mathbf{h})]\}^2$$

$$\gamma^*(\mathbf{h}) = \frac{1}{2m(\mathbf{h})} \sum_{i=1}^{m(\mathbf{h})} [\{Z(\mathbf{x}_i+\mathbf{h}) - Z(\mathbf{x}_i)\}^2]$$

- Possible spatial correlation between environmental variables statistically correlated
- **Experimental Cross Semi-Variogram (X-ESV)** quantifies spatial co-correlation

$$\gamma_{12}^*(\mathbf{h}) = \frac{\sum_{i=1}^{m(\mathbf{h})} [z_1(\mathbf{x}_i) - z_1(\mathbf{x}_i+\mathbf{h})][z_2(\mathbf{x}_i+\mathbf{r}) - z_2(\mathbf{x}_i+r+\mathbf{h})]}{2 \cdot m(\mathbf{h})}$$

- ESVs and X-ESVs do not describe properly the continuous spatial (co)variability of regionalized variable $Z(\mathbf{x})$ (Dowd, 2004a)
  - Continuous function is required to describe changes at any separation distance
  - **Variogram Model**
  - Similar spatial behavior of different variables (look at **variogram map**)
  - **Linear model of Coregionalization**
    - Generally two structures of variance (nugget effect + spherical model) (Wackernagel, 2003)

$$\gamma(\mathbf{h}) = \begin{cases} c_0 + c\left[\frac{3}{2}\frac{\mathbf{h}}{a} - \frac{1}{2}\left(\frac{\mathbf{h}}{a}\right)^3\right] & \text{per } |\mathbf{h}| \le a \\ c_0 + c & \text{per } |\mathbf{h}| > a \end{cases}$$

Linear Coregionalization Model

- The linear model of coregionalization is a sum of proportional covariance models. Proportional covariance models are models in which all covariances (or all variograms) are proportional to the same covariance (or variogram) function (Chilès and Delfiner, 1999).

- In practice:
  - The structural part of the variogram remains the same for every coefficient.
  - All **variograms** and **cross-variograms** have the same range in a particular direction

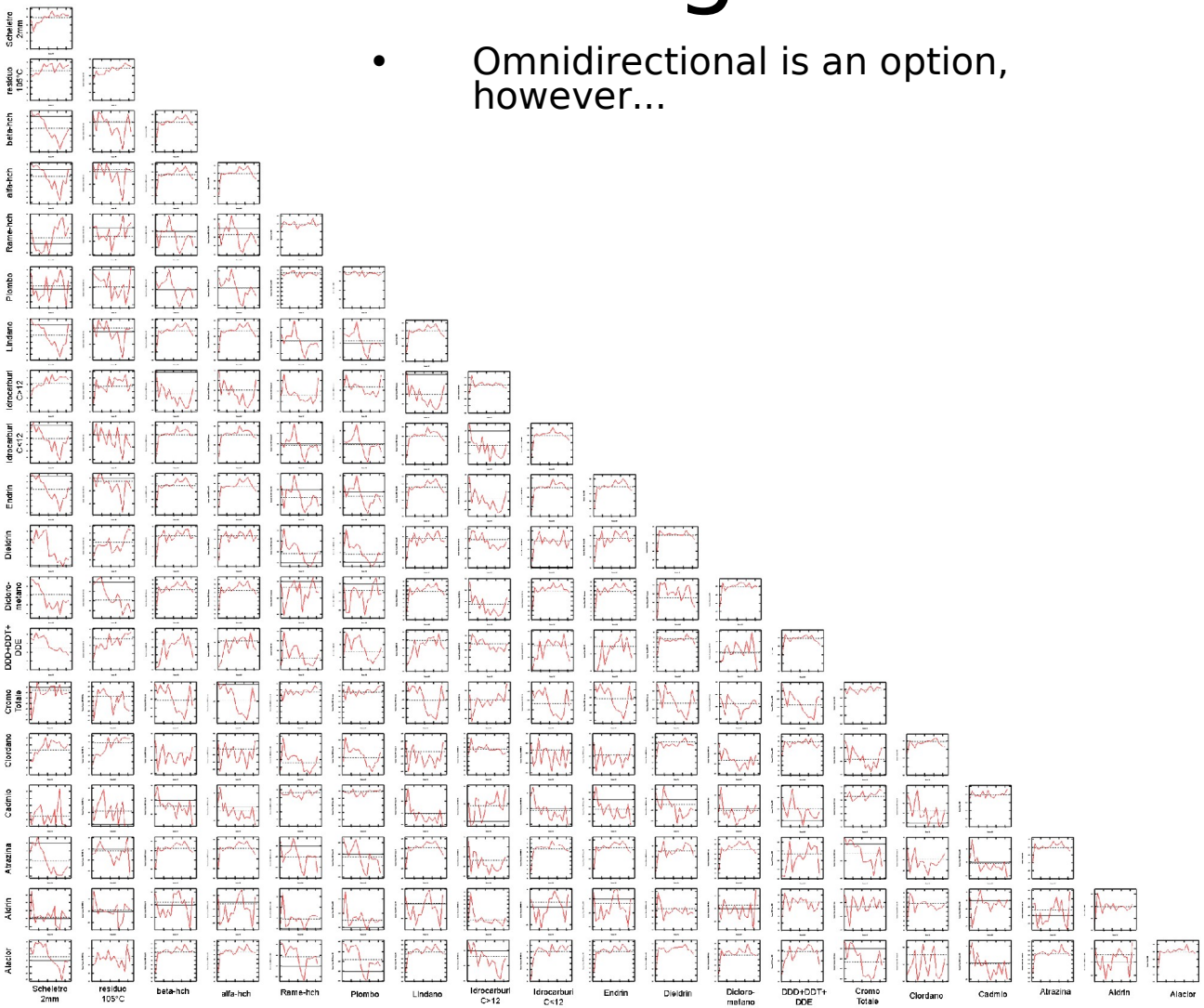
- There are automatic procedures that can fit linear models of coregionalization among whatsoever number of variograms and cross-variograms:

  - There is no guarantee that they will work properly in all applications.
  - In such cases manual adjustments of the fitted models are needed, respecting the following constraints:
    - Fixed range of anisotropy in any direction
    - Assumption of positive definite variance-covariance matrix

Linear Coregionalization Model

- Omnidirectional is an option, however...

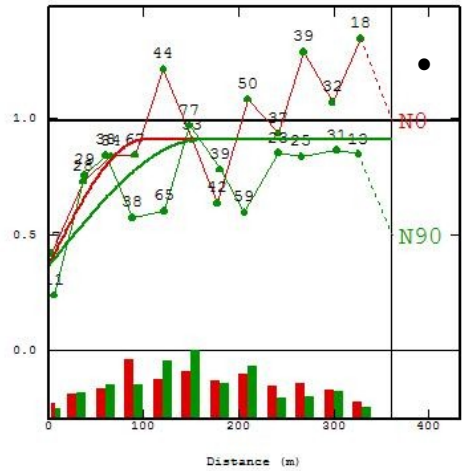
Site **B**

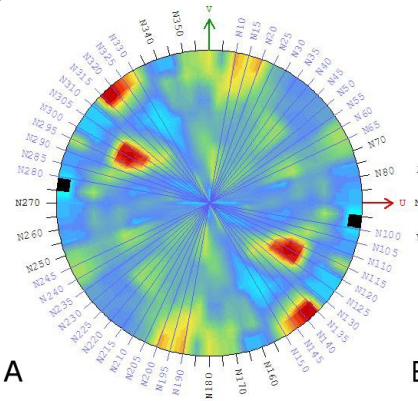


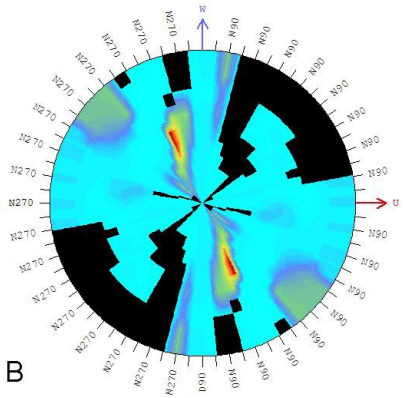
Linear Coregionalization Model

- Omnidirectional is an option, however...
- Variogram map revealed the different continuity in various three-dimensional directions for group of variables (i.e. some spatial correlated variables having similar behavior)

Site **B**

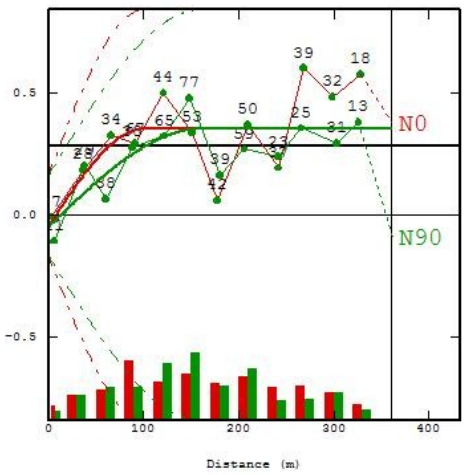


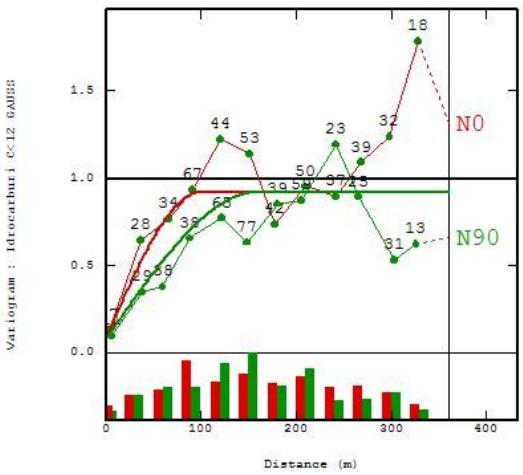


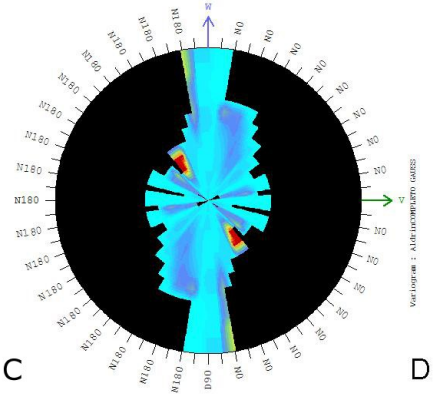


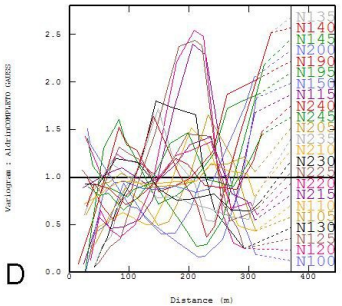
➢ **Anisotropy ratio**

❖ **Range in relation with underground structures**









Cross-Validation of Models

- Cross-Validation is a “back estimation” techniques for testing how different variogram models fit in ESV and X-ESV, by comparing true values of z at any sampled location x_i , and the estimated (by kriging) $z^*(x_i)$ values (Dowd, 2004)
- It's a powerfull validation methd which both assesses the performances of model and indicates how to improve it (Chiles & Delfiner, 1999) by checking the following statistics:

  - mean of kriging variance

$$\overline{\sigma_k^2}$$

  - mean squared error

$$\frac{1}{N}\sum_{i=1}^{N}\left[z(x_i)-z^*(x_i)\right]^2$$

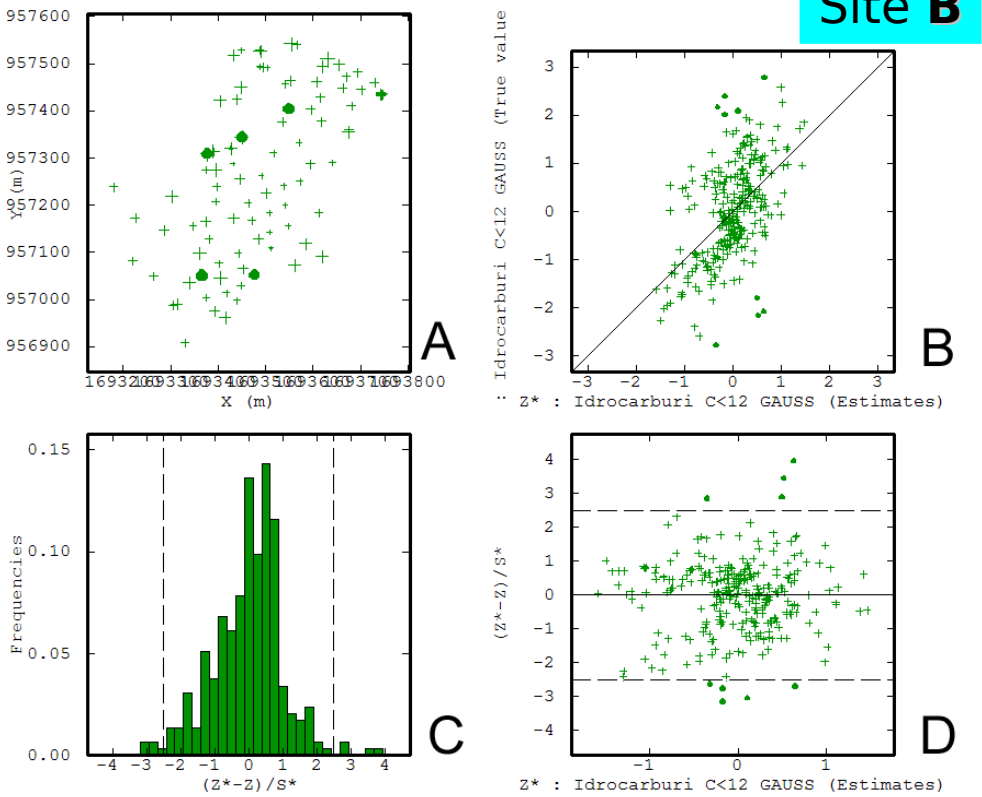
  - Mean Standard Errore

$$\frac{1}{N}\sum_{i=1}^{N}\left[z(x_i)-z^*(x_i)\right]$$

  - Standard deviation of estimated val.

$$\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left[z(x_i)-z^*(x_i)\right]^2}$$

Site **B**



Cross-Validation of Models

- Cross-Validation is a “back estimation” techniques for testing how different variogram models fit in ESV and X-ESV, by comparing true values of z at any sampled location x_i , and the estimated (by kriging) $z^*(x_i)$ values (Dowd, 2004)
- It's a powerfull validation methd which both assesses the performances of model and indicates how to improve it (Chiles & Delfiner, 1999) by checking the following statistics:

Site **B**

- mean of kriging variance

$$\overline{\sigma_k^2}$$

- mean squared error

$$\frac{1}{N}\sum_{i=1}^{N}\left[z(x_i)-z^*(x_i)\right]^2$$

- Mean Standard Errore

$$\frac{1}{N}\sum_{i=1}^{N}\left[z(x_i)-z^*(x_i)\right]$$

- Standard deviation of estimated val.

$$\sqrt{\frac{1}{N-1}\sum_{i=1}^{N}\left[z(x_i)-z^*(x_i)\right]^2}$$

	Mean		Variance	
Variable	Errore	Errore Stad.	Errore	Errore Stad.
Alaclor GAUSS	0.0026	0.0633	0.7755	1.4082
Aldrin GAUSS	0.0102	0.0188	0.4355	0.8435
Residuo a 105°C GAUSS	0.0222	0.0281	1.0053	1.3535
Scheletro 2 mm GAUSS	-0.0249	-0.0217	0.9006	0.9016
Cadmio GAUSS	-0.0307	-0.0350	1.0370	1.3442
Cromo Totale GAUSS	-0.0446	-0.0519	0.8657	1.3472
Piombo GAUSS	-0.0105	-0.0095	1.0202	0.9323
Rame GAUSS	0.0023	0.0046	0.8708	0.9237
Idrocarburi C<12 GAUSS	-0.0154	-0.0086	0.6965	1.0852
Idrocarburi C>12 GAUSS	-0.0113	-0.0076	0.9364	1.1105
Atrazina GAUSS	-0.0443	-0.0423	0.7694	1.0323
DDD+DDT+DDE GAUSS	0.0182	0.0186	0.9541	0.8465
Diclorometano GAUSS	-0.0522	-0.0502	0.6579	1.1305
Endrin GAUSS	-0.0486	-0.0440	0.6098	1.0456
Lindano GAUSS	-0.0444	-0.0342	0.6141	1.0461
Alfa-hch GAUSS	-0.0478	-0.0390	0.7138	1.1537
Beta-hch GAUSS	-0.0421	-0.0372	0.6281	1.1874
Clordano GAUSS	-0.0053	0.0015	0.7016	0.9284
Dieldrin GAUSS	-0.0017	0.0058	0.8348	1.5699

Statistics results of cross-validation of four coregionalization models of Site B.

Geostatistical Simulations

- The objective of geostatistical simulation (Journel & Huijbregts, 1978; Goovaerts, 1997; Chiles & Delfiner, 1999; Lantuejoul, 2002; Dowd, 2004c) is to provide alternate realizations of regionalized variables on any specified scale.

- It does not create data but provides at unsampled locations a possible reality statistically and geostatistically similar to the original sampled data (Dowd, 2004c).

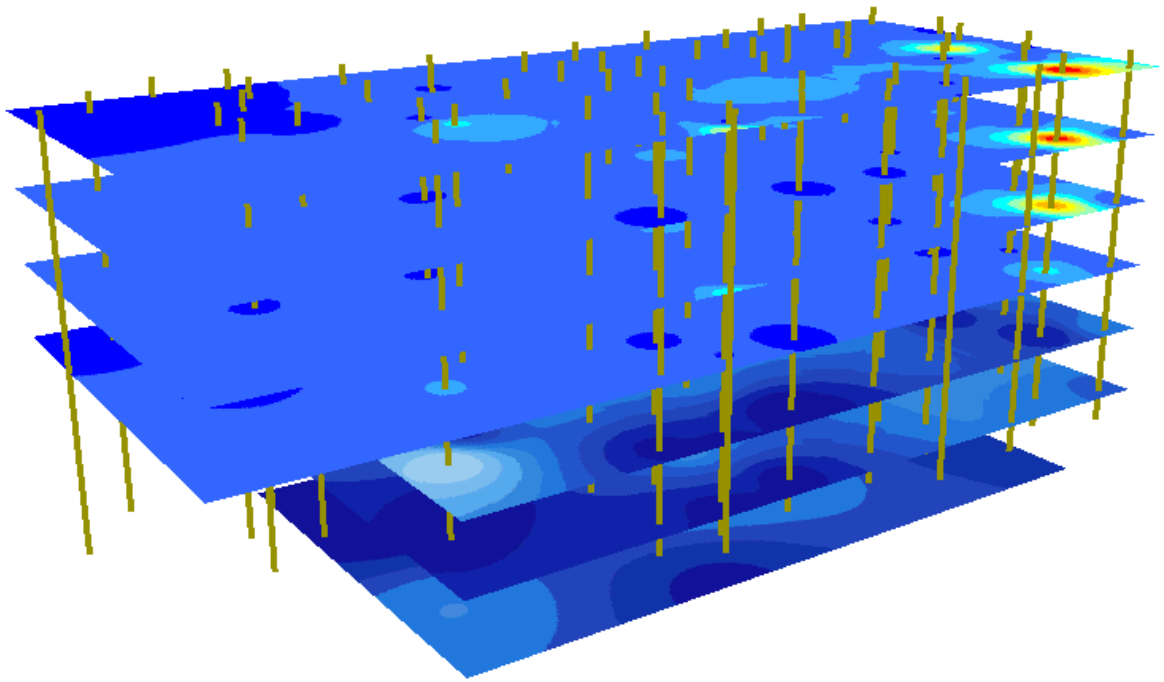
- Geostatistical simulation can be applied to:
  - the assessment of the variability of a regionalized variable and
  - the quantification of the uncertainty associated with the value of a regionalized variable at specified locations.
    - This second application is used in this project.

- Once the contaminants have been simulated, the volume can be subjected to any number of simulated operational activities and it can be used to assess the likely concentration of a metal above the imposed legal limit (Dowd, 1997).

Simulation vs Estimation

- Interpolation algorithms (estimations) tend to smooth the spatial variation of a variable;
  - they overestimate small values and underestimate large values.
  - This makes it difficult to detect patterns of extreme high values, for instance metal concentrations above legal limits.
- The estimation smoothing effect is not the same everywhere as it depends on the data configuration and it will be low for dense samples.

A smooth interpolator should not be used for applications in which the pattern of continuity of extremely high values is critical (Goovaerts, 1997).



Geostatistical Simulation: Methods

- Choice of method: based essentially on the application.
- For this project **sequential conditional methods** were chosen, because the need to respect the concentration values at each sample locations

- **Sequential conditional methods**:
  - ► easy to understand
  - ► provide a simple, robust and manageable implementation

- **Simulated values**:
  - ► Coincide with the actual values at all data locations
  - ► Spatial correlation (same variogram) as the data values
  - ► Same distribution (same histogram) as the data values
  - ► Coregionalized with other variables as the data values (spatial cross-correlation).

- **Simulation Methods**:
  1. Correlation correction of single univariate Sequential Gaussian Simulations
  2. Sequential Gaussian Co-Simulation

Geostatistical Simulations

- **Simulation Methods:**
  - Correlation correction of single univariate Sequential Gaussian Simulations (SGS)
  - Sequential Gaussian Co-Simulation (SGCOS)

    - SGS is performed independently for each variable and any spatial cross-correlation among the variables is ignored.
    - Correlation is common in environmental applications
    - Simulation results are improved by introducing a correlation correction between pairs of the independently simulated variables (Dowd & Xu, 2004)

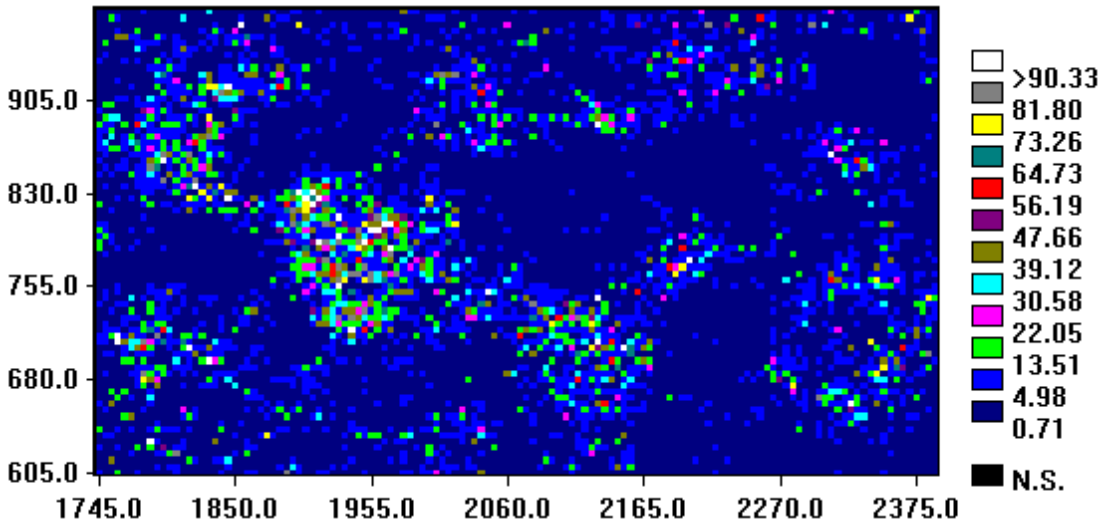
	Block centre	*Block sizes*	*Nos. Blocks*
E-W	1745 m	5m	130
N-S	605 m	5m	75
Vertical	0.5 m	1m	22

SGS grid definition of Site A

Geostatistical Simulations

- **Simulation Methods:**
  - Correlation correction of single univariate Sequential Gaussian Simulations (SGS)
  - Sequential Gaussian Co-Simulation (SGCOS)

Site **A**

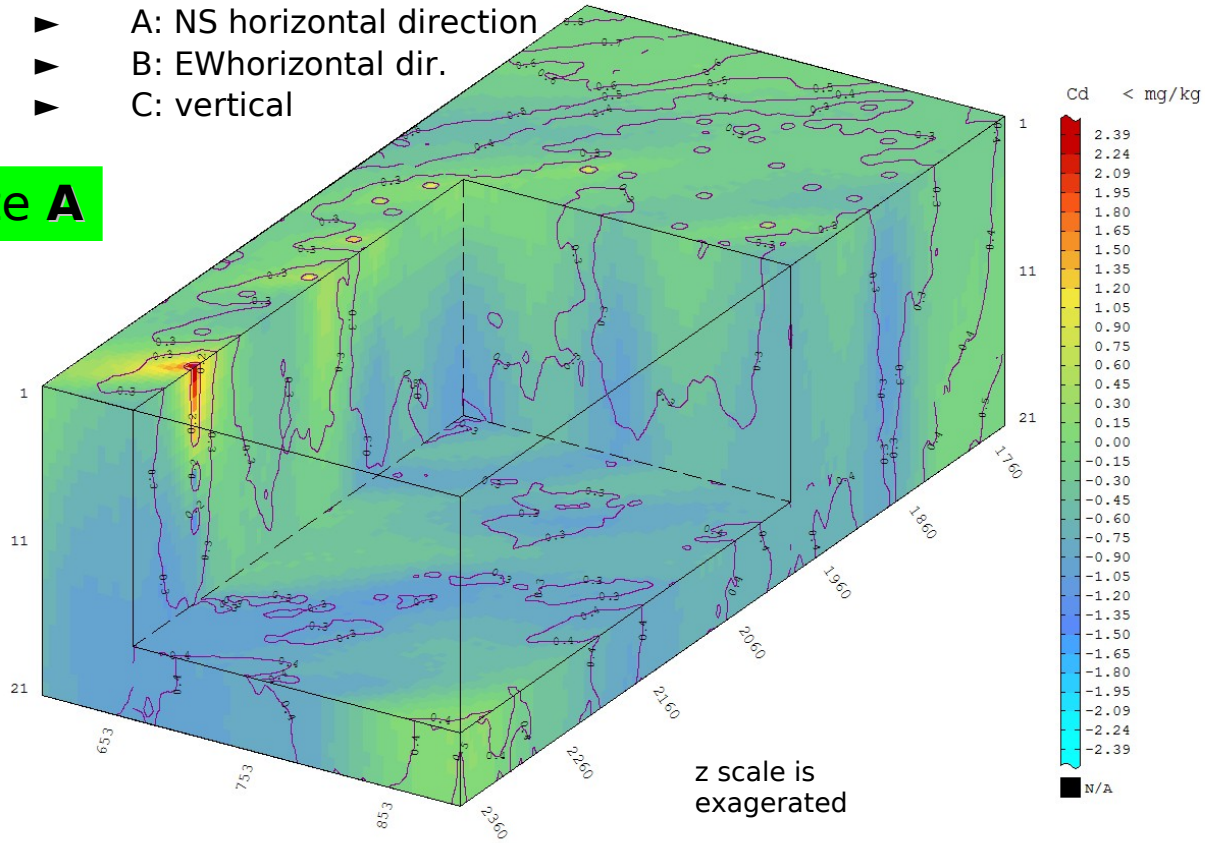


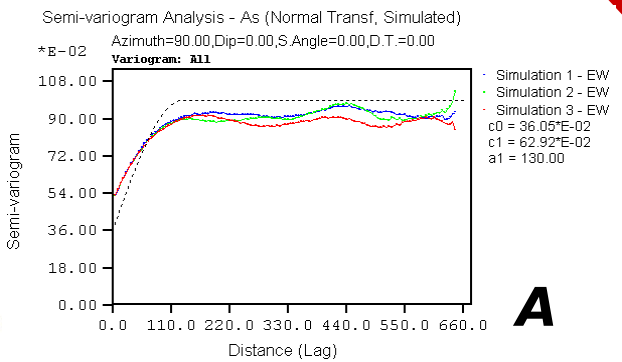
Example of a single simulation in SGS calculation for **AS** (legend unit: mg/kg)

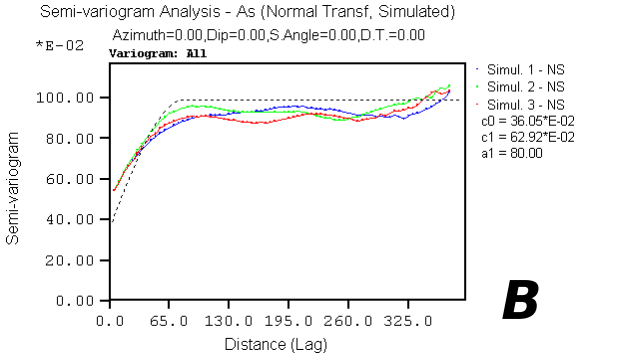
Geostatistical Simulations

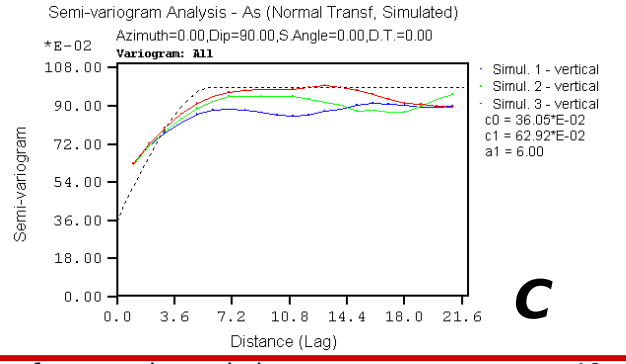
- Results of SGCOS Simulation of **Cd** (**Site A**) and other variables:
  - mean of 200 simulated realization (color) and standard deviation of estimation
- Geostatistical validation of simulation:
  - A: NS horizontal direction
  - B: EWhorizontal dir.
  - C: vertical

Site **A**









Geostatistical Simulations

- **Simulation Methods:**
  - Correlation correction of single univariate Sequential Gaussian Simulations (SGS)
  - Sequential Gaussian Co-Simulation (SGCOS)

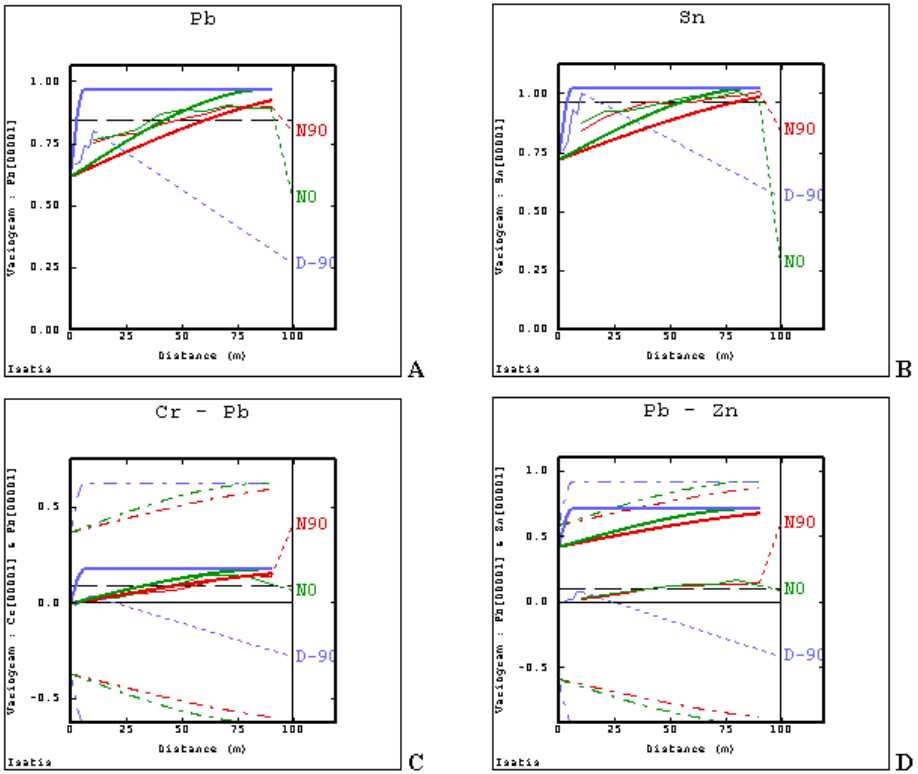
Site **A**

•SGCOS takes into account the spatial correlation among a set of regionalised variables by using the parameters of a linear coregionalization model.

•It is very similar to SGS, except that kriging is replaced by cokriging.

•The only drawback of this method is the significant increase in computing time over SGS and its is most effective for two or three variables and a relatively small dataset (Dowd & Xu, 2004).

Example of direct variograms and cross-variograms for co-simulated variables

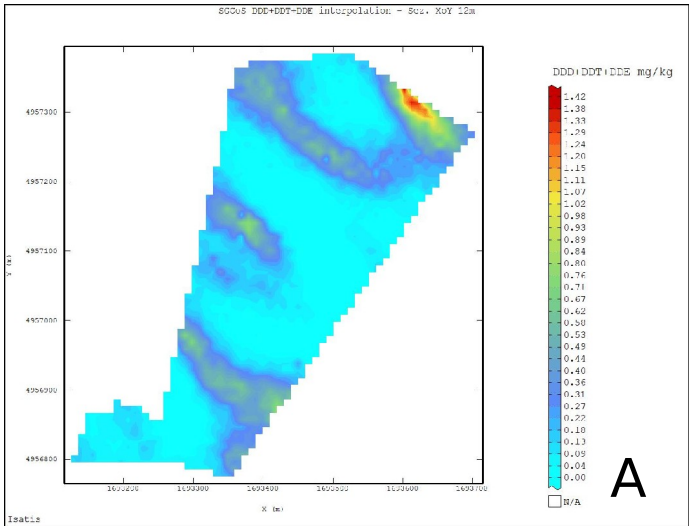


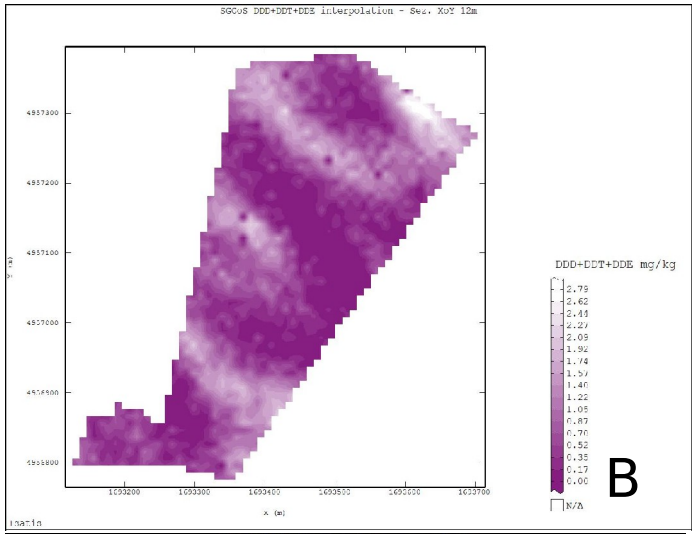
Geostatistical Simulations

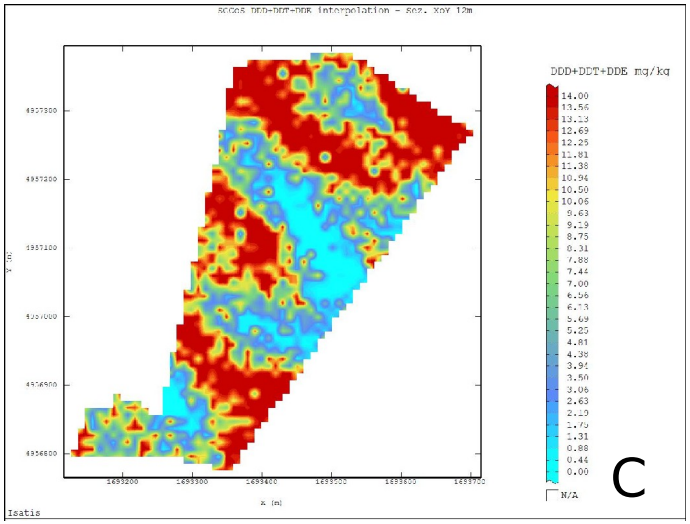
Site **B**

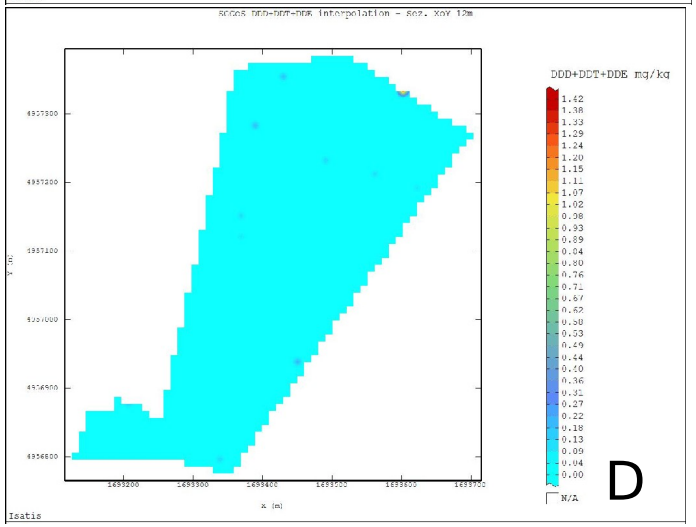
- Results of SGCOS Simulation of "**DDD+DDT+DDE**" variable and other variables (**Sito B**):

(A) Mean of 500 simulated realizations
(B) Standard deviation
(C) Largest simulation
(D) Smallest simulation









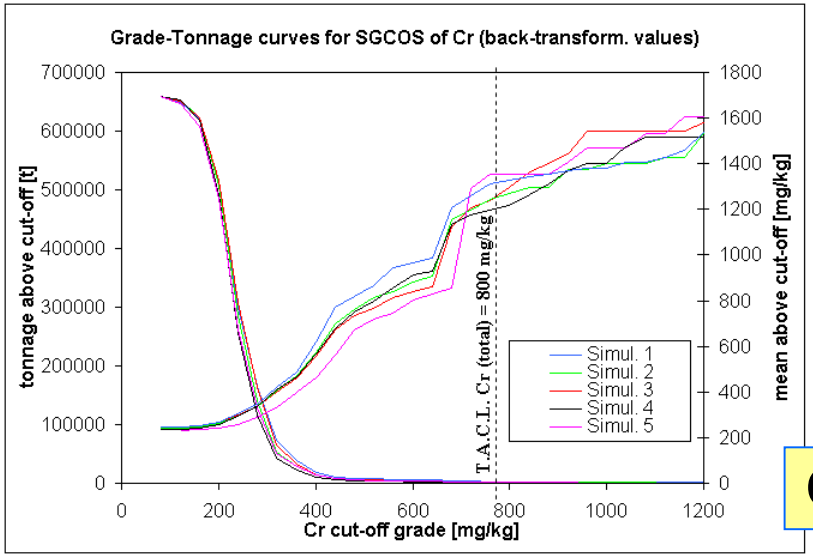
Risk Analysis: Grade-Tonnage Curves

- Risk quantification in the form of contaminant grade-tonnage curves is critical for capital investment in mining and environmental projects and can be obtained through geostatistical simulations of the studied volume (Dimitrakopoulos & Fonseca, 2003).

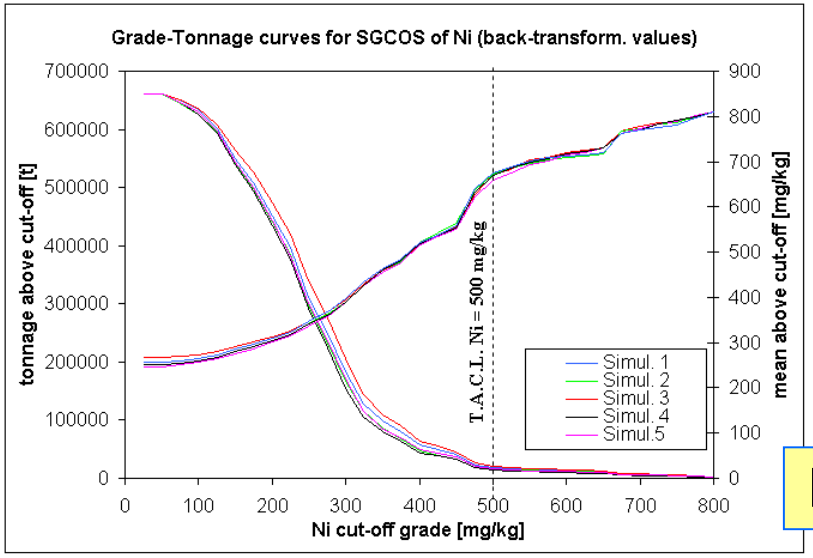
- These curves display simultaneously the tonnage of terrain above a particular threshold grade and the average concentration of the contaminant above that threshold (or cut-off).

- In practice:
  - grade-tonnage curves provide a means of determining how much of the population is likely to lie above or below a threshold value, i.e. the acceptable concentration limit
  - in addition it provides the average grade of the material above the threshold value (Clark & Harper, 2000). For instance, if terrain below the legal limit is ignored, the average value of the remaining terrain will be higher than the original average of the population.

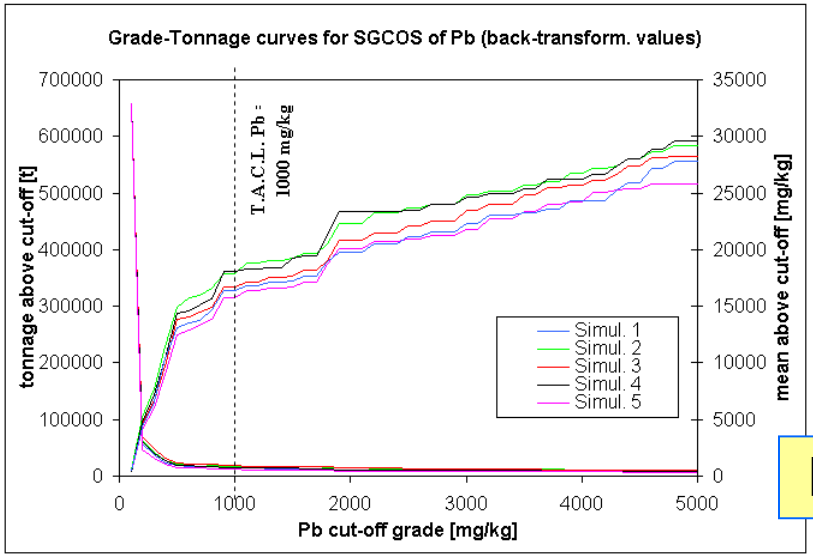
Risk Analysis: Grade-Tonnage Curves



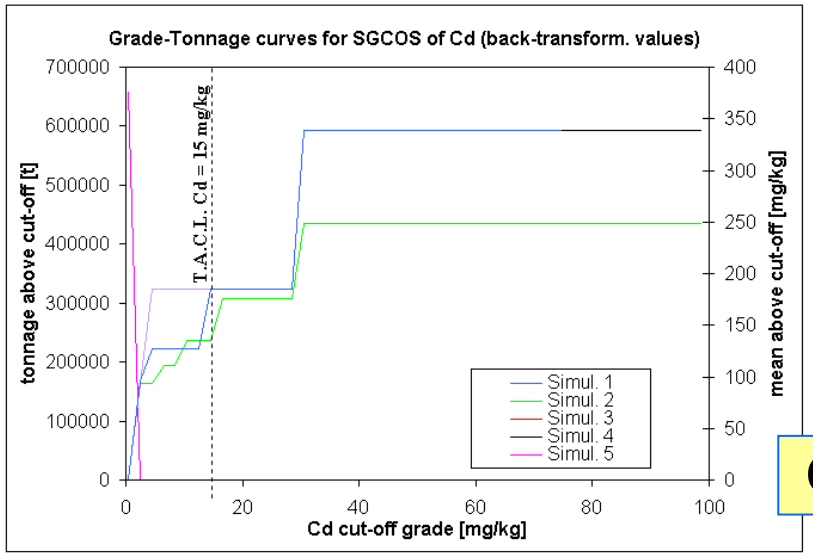
Cr



Ni



Pb



Cd

Risk Analysis: Contaminated Volumes

Site **B** Group	Variable	Terrain Volumes [m³] with concentration values greater than:					
		Location with habitative future destination			Location with industrial future destination		
		TACL (*) for houses terrain [mg/kg]	Volumes > TACL houses [m³]	Volumes > TACL houses [%]	TACL factories [mg/kg]	Volumes > TACL factories [m³]	Volumes > TACL factories[%]
Group 1	Alachlor	0.01	1442100	64.3	1	0	0.0
	Dieldrin	0.01	2242800	100.0	0.1	1303700	58.1
Group 2	Copper	120	114100	5.1	600	1100	0.0
	Cadmium	2	0	0.0	15	0	0.0
	Total Chromium	150	0	0.0	800	0	0.0
	Lead	100	200	0.0	1000	0	0.0
Group 3	Heavy Hydrocarb	10	163800	7.3	250	100	0.0
	Light Hydrocarb.	50	1381900	61.6	750	83000	3.7
Group 4	Aldrin	0.01	2226300	99.3	0.1	958400	42.7
	Atrazine	0.01	658300	29.4	1	0	0.0
	Chlordane	0.01	2224000	99.2	0.1	1573400	70.2
	DDD+DDT+DDE	0.01	2226100	99.3	0.1	951600	42.4
	Dichloromethane	0.1	101100	4.5	5	0	0.0
	Endrin	0.01	575700	25.7	2	0	0.0
	Lindane	0.01	2167100	96.6	0.5	41200	1.8
	alpha-HCH	0.01	1722700	76.8	0.1	49900	2.2
	beta-HCH	0.01	2129700	95.0	0.5	24800	1.1

(*) TACL – Total Accepted Value fixed by Law (Ministero dell'Ambiente,1999)

Discussion

Conclusions

1. Geostatistical simulation is particularly useful when data are sparse and variability is erratic
2. Co-simulation results were better than those those obtained by correcting the correlation between univariate simulations
3. This study can be taken as the basis for a complete risk assessment for further complete remediation projects, in parallel with a hydrogeological simulation of contaminants in the underground water

Recommendations

- Further sampling on the basis on the simulation results obtained.
- More continuous and denser sampling undertaken in a few new boreholes could significantly improve the simulation results.

References

[1] Bleines, C. J., Deraisme, F., Geffory, N., Jeannee, S., Perseval, F., Rambert, D., Renard, O., Torres, and Touffait, Y. (2004) Isatis Software Manual, Geovariances & Ecole Des Mines De Paris

[2] Clark, I. & Harper, W.V. (2000), Practical Geostatistics 2000. Vol. 1, Ecosse North America Llc., Columbus, Ohio, U.S.A. 342pp.

[3] Dimitrakopoulos, R. & Fonseca, M.B. (2003), *Assessing risk in grade-tonnage curves in a complex copper deposit, northern Brazil, based on an efficient joint simulation of multiple correlated variables*, in *Application of Computers and Operations Research in the Minerals Industries*, South African Institute of Mining and Metallurgy

[4] Dowd, P.A. (1997), *Risk in mineral projects: analysis, perception and management.* Trans. Insts Min. Metall. (Sect. A: Min. industry). **106**: p. A9-A18

[5] Dowd, P.A. (2004a), *MINE5270: Multivariate Geostatistics*. MSc in Mineral Resource and Environmental Geostatistics, University of Leeds, Leeds. 82pp.

[6] Dowd, P.A. (2004c), *MINE5290: Geostatistical Simulation*. MSc in Mineral Resource and Environmental Geostatistics, University of Leeds, Leeds. 156pp.

[7] Dowd, P.A. & Xu, C. (2004), *GeostatWinTM - User's manual*, University of Leeds, Leeds, England.

[8] Chiles, J.-P. & Delfiner, P. (1999), *Geostatistics: modeling spatial uncertainty*, Wiley, New York; Chichester. xi, 695 pp

[9] Goovaerts, P. (1997), *Geostatistics for natural resources evaluation*. Applied geostatistics series., Oxford University Press, New York; Oxford. xiv, 483 p.

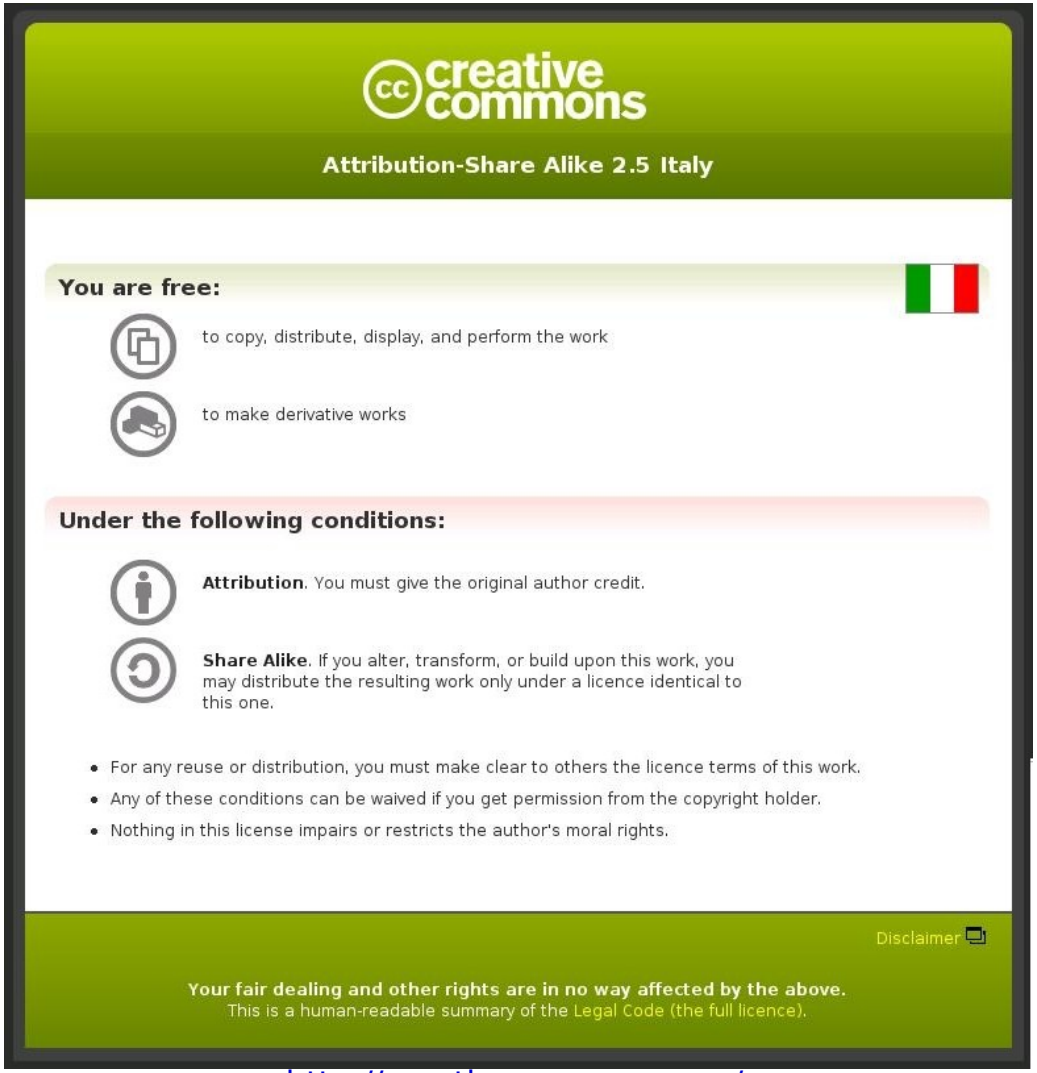
[10] Guastaldi, E. (2005), Risk analysis for remediation project of contaminated sites: the geostatistical approach. PhD Thesis, University of Siena, Italy

[11] Journel, A.G. & Huijbregts,C.J. (1978), *Mining Geostatistics*, Academic Press, London. 600.

[12] Lantuejoul, C. (2002). *Geostatistical Simulation: Models and Algorithms*: Springer Verlag.

[13] Ministero dell'Ambiente (1999). DECRETO LEGISLATIVO N. 471/99 - Valori di concentrazione limite accettabili nel suolo e nel sottosuolo riferiti alla specifica destinazione d'uso dei siti da bonificare (Vol. DM471/99).

[14] Wackernagel, H. (2003), *Multivariate Geostatistics: An Introduction with Applications*. 3rd ed, Springer-Verlag, Berlin, Germany. 387 pp



http://creativecommons.org/