IMPACT ANALYSIS AND SAMPLING DESIGN IN THE POLLUTION MONITORING PROCESS OF THE AZNALCOLLAR ACCIDENT USING GEOSTATISTICAL METHODS

C. Franco [a], J. Delgado [b,*], A. Soares [a]

[a] CMRP- Grupo do Ambiente. Instituto Superior Tecnico. Universidade Tecnica de Lisboa.
Av. Rovisco Pais – 1096-001 Lisboa, Portugal – cfranco@ist.utl.pt, ncrmp@alfa.ist.utl.pt
[b] Dpto. Ingeniería Cartográfica, Geodésica y Fotogrametría, Universidad de Jaén
c/ Virgen de la Cabeza, 2 – 23071 Jaén, Spain – jdelgado@ujaen.es

KEY WORDS: Spatial Information Systems, Cartography, Pollution, Impact analysis, Sampling, Statistics

ABSTRACT:

On the morning of the 25th of April 1998, an accidental spill of the waste damp basin of the Aznalcóllar mine (Seville, SW Spain) occurred. The basin, containing mud and acid waters, suffered a rupture spilling 4 hm^3 of acid waters and 2 hm^3 of acid mud directly into the Agrio and Guadiamar rivers. This accidental spill spread to an area of about 49 km^2 and they had very important ecological consequences due to those rivers are the basic water resource of the Doñana National Park aquifer (located 60km downstream the mine) that was declared as Biosphere Reserve by UNESCO in 1994. Remediation actions were taken immediately in order to protect the population health and agriculture and minimize the ecological impact in the Park. Once the primary remediation actions were applied, it was very important to characterize the residual contamination distribution patterns and to obtain environmental hazards maps in which areas with high heavy metals concentration (Cu, Pb, Zn, Cd and As) can be identified. This paper presents the most important results of the geostatistical treatment of this data. The basic objectives of this treatment were to establish the impact analysis of the contaminant (environmental hazards maps) using the direct cosimulation technique and remediation levels that defines if a certain zone can be considered as contaminated or clean and the design of a new sampling schema that would be applied in future sampling campaigns and planning the optimum sampling for the next campaigns.

1. INTRODUCTION

On the morning of the 25th of April 1998, the waste damp basin of the Aznalcóllar mine, West of Seville (South-West of Spain), containing mud and acid waters, suffered a rupture spilling 4hm^3 of acid waters and 2hm^3 of acid mud directly into the Agrio river and consecutively into the Guadiamar river (C.M.A, 2000) (figure 1). This accidental spill spread to an area of around 49 km^2. The affected area, situated 60km downstream the mine, is of great ecological importance because the Guadiamar river is the main hydrological resource of the National Park of Doñana (Biosphere Reserve, UNESCO 1994) (figure 2).

Figure 1. Geographical location map

Figure 2. Acid mud deposits in Guadiamar river (CMA, 2000)

The evaluation and management of environmental impacts due to residual soil contamination, was a main concern. Despite the direct remediation done to the whole area, primarily through mobilization and excavation of the acid mud, there still was a significant quantity of residual contamination, which can affect negatively all ecosystems. With this study, we intended to characterise the spatial dispersion of heavy metals – Cu, Pb, Zn, Cd and As – on the Guadiamar river margins, to be able to elaborate and create environmental hazards maps as basic tools for important decision-making, such as the delineation of target areas for remediation or for additional sampling and the optimum sampling grid design.

* Corresponding author.

Considering the geostatistical methodology presented, four main steps can be distinguished: a) obtaining spatial variability functions (variogram functions) for the hard data; b) estimation of average values of contaminant heavy metals by means of the ordinary kriging method, c) production of hazard maps based on simulated images of the spatial dispersion of heavy metals, using a stochastic simulation technique – direct sequential simulation and co-simulation, d) optimum sampling desing for the next campaigns.

2. DATA SET

The study area is a 2 km^2 region located approximately 10 km in the South of the Aznalcóllar mine. The information available was obtained through a soil sampling realized in August 1999, where 80 samples were collected. For the purpose of this study, only the samples located inside the study area, i.e. 40 samples, were used to characterize spatial dispersion of residual contamination with heavy metals (figure 3). Soil samples were collected from the topsoil and analytical results of Cu, Pb, Zn, Cd and As recorded in terms of total concentration (ppm). Figure 1 shows the position of the sampling locations and the basic statistics of the metal concentration are summarized in Table 1.

Figure 3. Sampling locations

	Cu	Pb	Zn	Cd	As
Mean	242.53	686.43	1042.50	3.30	336.55
Std.Dev.	184.03	848.17	820.40	2.98	448.00
Minimum	47.0	116.0	146.0	0.5	24.0
25th Perc.	104.5	203.5	362.5	1.0	77.0
Median	220.5	403.0	1031.5	3.0	193.5
75th Perc.	340.0	830.5	1421.5	4.0	415.0
Maximum	1074.0	5150.0	4460.0	17.0	2649.0
Range	1027.0	5054.0	4314.0	16.5	2625.0

Table 1. Basic statistics of the available data (in ppm)

3. SPATIAL VARIABILITY CHARACTERIZATION

The spatial variability characterization was one of the main objectives of this work. This characterization has been made using the variogram functions. Variograms of the metal contents value were calculated (Figure 4).

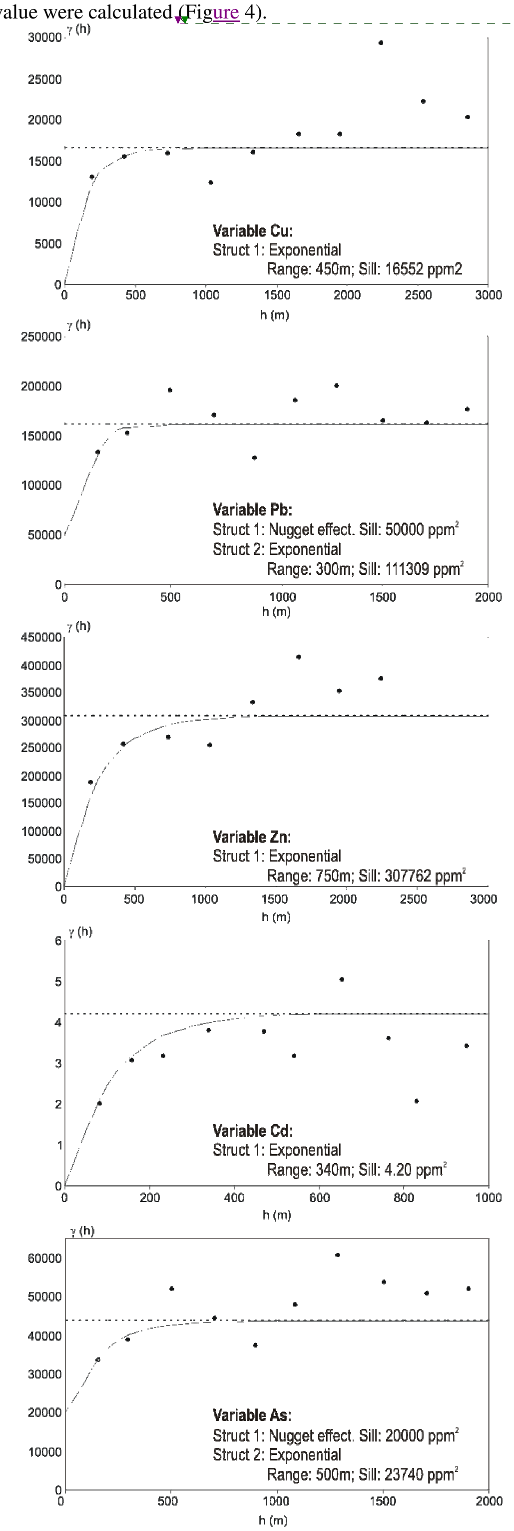

Figure 4. Variogram of the experimental data

Deleted: were fitted to them

Deleted: see

The experimental calculated has been fitted to exponential models (with ranges between 350m to 750m –influence distance–) and Pb and As variables present a nugget effect that reflects a discontinuity at local scale.

4. ESTIMATION MAPS

Using the information derived from the structural analysis a estimation process was made in order to obtain the most probable metal concentration maps (estimated maps).

The estimated maps were obtained using ordinary kriging geostatistical estimation methodology. The ordinary kriging is a stochastic estimation method that uses the information derived from the structural analysis (variogram analysis) in order to improve the estimation quality.

In the ordinary kriging methodology the weights that are applied to each data in order to obtain the final estimated value are calculated using the information provided with the structural analysis. This weights are calculated according to obtain an unbiased estimation and a minimum variance estimation. In figure 5, an example of a estimated map is show (As variable).

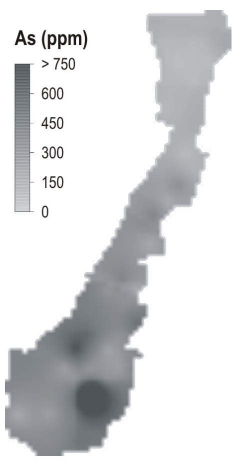

Figure 5. As metal concentration estimated map

5. HAZARD MAPS

5.1 Co-Simulation of the set of metals

The applied methodology relies on direct sequential simulation and cosimulation techniques (Soares A., 2001). The multivariate set of variables was co-simulated using the direct sequential co simulation technique: each variable is simulated based on the hard data – experimental samples – and an image (secondary information) given by the previously simulated map of another metal.

The simulation methods provide probable versions of the reality. All simulated versions have equal probability of occurrence. The simulated version must accomplish the following conditions:

– must have equal mean and variance than the experimental data.
– must have similar data distribution (histogram)
– must have a similar variability structure (similar variogram structure)
– the simulated version must be coincide with the experimental data in the known (sampled) points (conditional simulation).
– It is possible to obtain an unlimited number of simulations of the phenomenon in a specific grid.

One of the most important characteristics of the simulated version is that the variability remains constant. This fact allows calculating the probability that a defined limit value (risk value) has been exceeded in a determined location.

The high correlation values between these metals clearly reflect the common origin of contamination. Based on the correlations between the different variables and on the corresponding variograms (spatial continuity), it is possible to rank the variables to define the sequence of metals to be simulated. For this study the first variable simulated was the Cd because it showed better spatial continuity (variogram ranges) as well as a good correlation with the other heavy metals (greater than 0.92).

Given the high correlation coefficients between metals and the spatial continuity revealed by the variograms and after some tests, a set of 10 simulations was considered sufficient to represent the spatial uncertainty of those metals. Simulations and co-simulations were performed according to the following sequence:

- First, a set of 10 realizations of the first element Cd was simulated with direct sequential simulation.
- From the 10 previously simulated images, one is chosen to serve as soft data to simulate 10 images of the next element, Cu, using Direct Co simulation. The “soft” image of Cd is chosen according to the better match of the basic statistics (standard deviation and mean). This step was repeated for the next 3 metals: Zn, Pb and As.

The simulated maps show a similar concentration pattern for the different heavy metals (Figure 6), and reproduce quite well the variogram models.

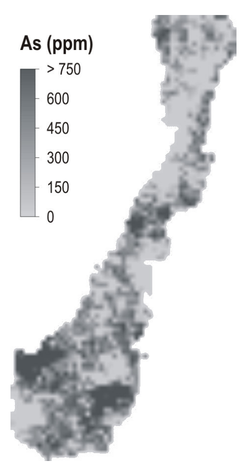

Figure 5 – Example of the As cosimulated map

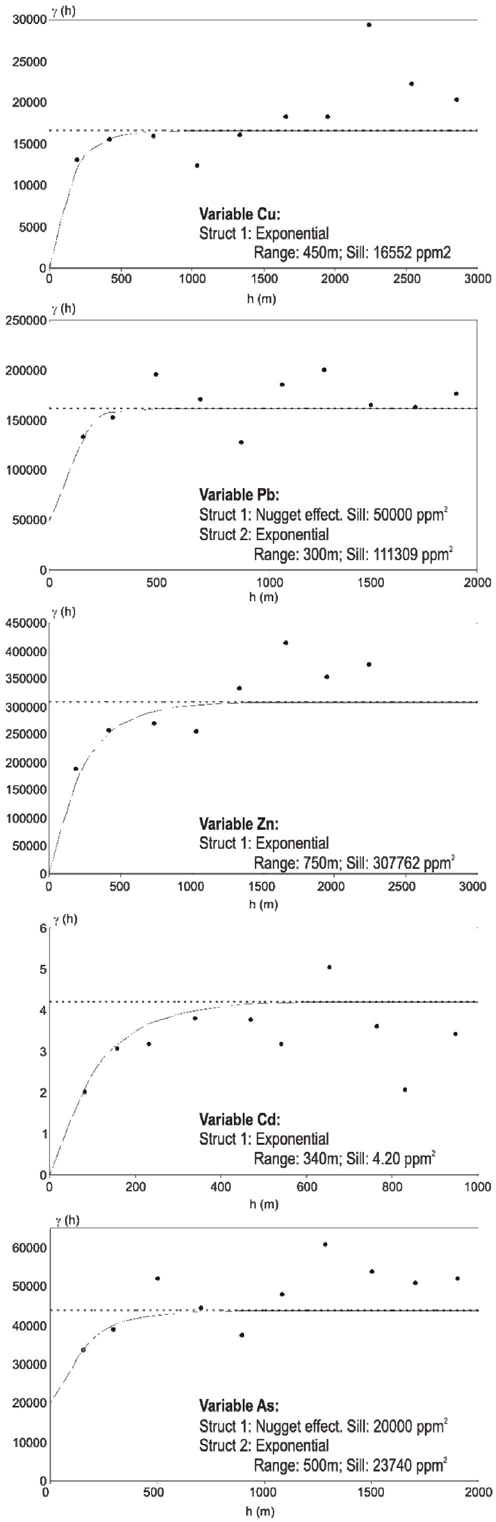

Figure 6. Variograms of the simulated data

5.2 Hazard Maps

One of the principal aims of this study was to delineate the areas that need future remediation based on intervention values. The Consejeria de Medio Ambiente de la Junta de Andalucía (Environmental Agency of the regional government of the South of Spain), C.M.A, (C.M.A., 2000) defined for each contaminant four different remediation levels: maximum level allowed, recommended investigation, compulsory investigation, compulsory treatment. Hence joint probabilities of different metals to be simulated simultaneously, above or under the remediation levels, can originate hazard maps of the region. The study area mainly consists of agriculture soils with pH values lower than 7.

Considering the thresholds: z_1 (maximum level allowed), z_2 (recommended investigation), z_3 (compulsory investigation) and z_4 (compulsory treatment), the following joint probabilities, at a given location x_0, can be identified with different hazard levels:

a) Prob $\{z_{Cd}(x_0)<z_{1Cd}$, $z_{Cu}(x_0)<z_{1Cu}$, $z_{Zn}(x_0)<z_{1Zn}$, $z_{Pb}(x_0)<z_{1Pb}$, $z_{As}(x_0)<z_{1As}\}$ corresponds to the most clean hazard scenario;
b) Prob $\{z_{Cd}(x_0)\leq z_{2Cd}$, $z_{Cu}(x_0)\leq z_{2Cu}$, $z_{Zn}(x_0)\leq z_{2Zn}$, $z_{Pb}(x_0)\leq z_{2Pb}$, $z_{As}(x_0)\leq z_{2As}\}$, corresponds to the intermediate clean hazard scenario, meaning that all metals at $x_0 \leq z_2$;
c) Prob $\{z_{Cd}(x_0)\geq z_{3Cd}$, $z_{Cu}(x_0)\geq z_{3Cu}$, $z_{Zn}(x_0)\geq z_{3Zn}$, $z_{Pb}(x_0)\geq z_{3Pb}$, $z_{As}(x_0)\geq z_{3As}\}$, corresponds to the intermediate contaminated hazard scenario, meaning that all metals at x_0 are greater or equal to z_3;
d) Prob $\{z_{Cd}(x_0)\geq z_{4Cd}$, $z_{Cu}(x_0)\geq z_{4Cu}$, $z_{Zn}(x_0)\geq z_{4Zn}$, $z_{Pb}(x_0)\geq z_{4Pb}$, $z_{As}(x_0)\geq z_{4As}\}$, corresponds to the most contaminated hazard scenario.

We can define a joint indicator using the corresponding indicator for the different metals: $Iz_1(x_0) = Iz_{1Cd}(x_0) \cdot Iz_{1Cu}(x_0) \cdot Iz_{1Pb}(x_0) \cdot Iz_{1Zn}(x_0) \cdot Iz_{1As}(x_0)$. The joint probability at x_0 – corresponding to scenario i) – can be estimated with the 10 simulated images:

$prob_{z1}(x_0) = \frac{1}{10}\sum_{i=1}^{10} I_{z1}(x_0,i)$ $I_{z1}(x_0,i)$ corresponds to $I_{z1}(x_0)$ of simulated image *i*.

Equivalent joint probabilities can be computed for the other scenarios: $prob_{z2}(x_0)$, $prob_{z3}(x_0)$ and $prob_{z4}(x_0)$.

Finally, a global hazard map was obtained by classifying each pixel in four defined scenarios (Figure 7). The results show that approximately 44% of the study area need compulsory treatment and 40% of the study area need compulsory investigation.

Since some European legislation impose treatment whenever one metal exceeds the highest threshold (compulsory treatment), an alternative for scenario iv was conducted: if at least one heavy metal exceeds the compulsory treatment threshold the pixel is considered to belong to scenario iv, i.e., treatment is imposed to that soil, with Prob $\{z_{Cd}(x_0)=z_{4Cd}$ or $z_{Cu}(x_0)=z_{4Cu}$ or $z_{Zn}(x_0)=z_{4Zn}$ or $z_{Pb}(x_0)=z_{4Pb}$ or $z_{As}(x_0)=z_{4As}\}$.

In this alternative, scenario iv has higher probabilities of occurring in comparison to the other remediation levels. The obtained probabilities for this scenario show that approximately 72% of the study area needed compulsory treatment and 22 % of the study area needed a compulsory investigation.

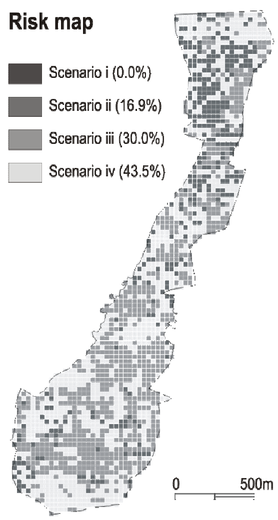

Figure 7. Global hazard map for scenario i

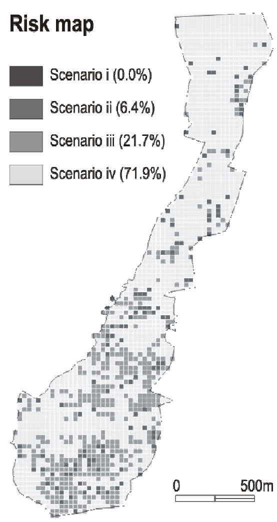

Figure 8. Global hazard map for scenario iv

5.3 Environmental hazard maps

Finally, with the hazards maps obtained for the two alternatives it is possible to intersect them with an environmental impact map. An impact map that classify to areas according its environmental importance was made by C.M.A. (C.M.A., 2000). The maps include 4 different impacts levels. The lowest impact level corresponds to an extensive culture occupation that occupies 56% of the total study area while the highest corresponds to the Guadiamar river margins (27% of the area).

Intersecting the impact map with the 2 hazards maps (resulting from the two alternatives for the scenario iv) 23% and 36% of the highest impact area needed treatment for the first and second alternative, respectively.

6. NETWORK SAMPLING DESIGN

According to the previous results, the residual contamination affected to a determined areas. For this reason, a new sampling campaign was considered as necessary. The new sampling must have a lower grid size in order to reach a better characterization of the variability characteristics of the residual contamination (local contamination).

The campaign was made in summer 2001. The new data tried to analyse the real situation of the affected area after two years from the previous data (1999) and examine the properties that have had in the zone the different cleaning tasks (by means of the organic and chemistries addition) thus own regenerative power of the environment.

The sampling network design has two principal conditions. First, only the north area -from the Aználcollar mine to the Doblas Brigde- must be considered due to the elevated contained which were detected in this zone. On the other hand, the total number of samples must be around 300.

For the sampling network design geostatistical techniques were applied. The method has a base in the fact than the associated errors to an estimation process can be calculated a priori inside to the geostatistical schema. These errors only depend of the variability function (obtained from the structural analysis -variogram analysis- of the available data that are considered as representative of the variable to sample) and of the position of the sampling points being totally independent of the proper variable values. Using this approach, it is possible to analyse the estimation error (that define the estimation uncertainty) obtained for the different sampling network considered in the design.

This analysis has been combined with the land-use and environmental hazard level made by the environmental technicians of the CMA. Both aspects are fundamental in the sampling design network planning.

The network optimization begins with the determination of the variability function that is considered as representative of the variable to sample. The used structure (in which the different variogram structures of the variables are take into account) are composed by a nugget effect (local variability around 20% variance) and an exponential model (with 400m range that represents 80% variance).

Three grid sampling design were considered. The sampling schemas are shown in figure 9. The used reference distance unit -the work unit over the error must be analysed- is 100m (divided into 10x10 elements of 10x10m grid size). The grid size is variable between 50m x 50m and 500m x 500m, as function of the desired information density for a determined location. In Table 2, an example of the application of this methodology is shown. The final result is obtained as a weighted mean of a priori errors considering the surface of the different impact units (that must be according with the future available information density).

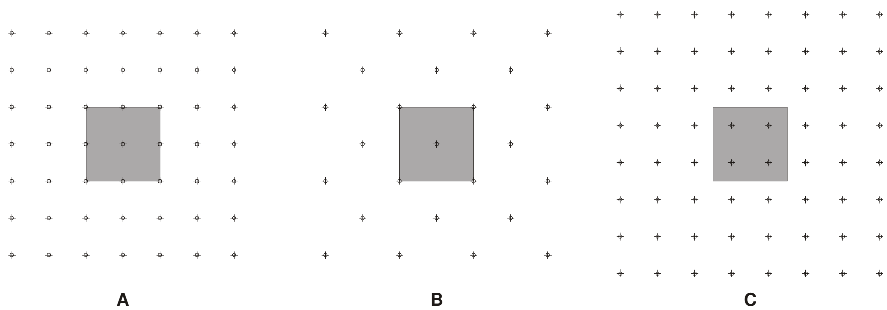

Figure 9. Sampling models used in the optimization process

Impact Zone	Surface (ha)	Spacing(m)	Density (data/km^2)	A priori Error	No.Samples
Gris type A					
2	198	500	9.00	0.8069	18
3	24	400	12.25	0.8019	3
4	342	300	18.80	0.7923	64
5	776	250	25.00	0.7826	194
Number of total samples:			311		
Weighted surface error level:			0.774		

Table 2. Optimization sampling design example of Sampling Schema A

7. DISCUSSION AND CONCLUSIONS

When, in soil quality evaluations, more than one pollutant occur simultaneously, it is necessary to determine the total area affected by any of the pollutants. This leads to the problem of defining the area contaminated by the different pollutants simultaneously. When the concentration values of the pollutants exceed the established intervention values a future remediation will be considered.

Usually theses actions are extremely expensive and, for this reason, a good interpretation of the spatial dispersion of all pollutants will be reflected in the remediation costs. For this type of contaminations the delineation of the remediation/treatment areas should not be defined considering each pollutant separately. With the methodology presented in this paper it is possible to account for the uncertainties in mapping the probability that different pollutants are simultaneously contaminating the soil.

Application of this methodology was made considering two alternatives regarding the treatment level, extreme scenario: when at least one heavy metal critically contaminates the soil; and, when all pollutants are simultaneously contaminating the soil. Depending on the aim of the soil quality investigation and on the costs associated to the remediation actions, it is possible to choose between these two alternatives. But, considering that European legislations imposes that the treatment actions should be carried out as long as one metal exceeds the highest threshold (compulsory treatment) in order to obtain a clean and safe area, the second alternative is certainly more appropriated.

Finally, it is very important emphasizing the main advantages derived from the kriging estimation variance utilization (as a parameter representative of the estimation uncertainty). This parameter can be use in order to establish a priori error for the sampling network optimization obtaining a optimum network. The optimization can be made considering the maximum expected sampling points (or the maximum admissible estimation error) and the impact environmental level information.

ACKNOWLEDGMENTS

The authors thank to the Oficina Técnica del Corredor Verde del Guadiamar for the data information and environmental expertise support.

REFERENCES

C.M.A., 2003, El Corredor Verde del Guadiamar, Consejería de Medio Ambiente. http://www.juntadeandalucia.es/medioambiente/guadiamar/indguadiamar.html (accessed 25-April-2004).

Soares A., 2001, Direct Sequential Simulation and Co-Simulation, *Math. Geol.*, vol. 33 (8), pp. 911-926.