

Application of Geostatistics to Soil and Groundwater Samples Contaminated by Petroleum Products – A Case Study

Carvalho, J.M., Fiúza, A.

University of Porto, Faculty of Engineering – Mining Department/ CIGAR

ABSTRACT: A geoenvironmental site investigation/characterization survey took place in a contaminated coastal area in the surroundings of an active oil refinery having an associated petrochemical complex, located north of Porto, in the north-western part of Portugal. In this context, several soil and groundwater samples were collected, respectively in trenches and boreholes, and chemically analyzed allowing the characterization of their chemical pollution, in terms of the different contamination products and respective concentrations, namely BTEX, TPH, PHA's and lead, as well as of some physical/chemical soil properties and characteristics, namely density, pH and organic matter content. At this stage of the investigation, a multivariate statistical analysis has been previously carried out. The present study is focused on comparing alternative geostatistical procedures, namely different kriging methods as well as strategies, used to characterize the qualitative and quantitative spatial contamination dispersion in the investigated area leading to different mapped models prior to the definition of remediation strategies. In addition, it is also summarized a risk analysis study, also based on geostatistical procedures, leading to a partition of the studied zone in terms of their assessed relative remediation needs.

Introduction

The overall study is part of a comprehensive survey aiming at assessing the state of contamination in the surrounding ground of an active industrial installation in which to base appropriate remediation strategies. The target site is an oil refinery with an associated petrochemical complex, located in the Atlantic seashore, north and nearby the city of Porto, Portugal. The studied area has 1600 m in the x (E-W), direction and 2150 m in the y (N-S) seashore direction.

The local topography is quite smooth with a generic downwards E-W gradient, also detected in the soil/bedrock interface (figures 33 and 34). Both ground and bedrock surfaces form two smooth valleys where the ground water flux tends to concentrate, again in E-W direction towards the sea, separated by a centrally located low crest. Geologically, the studied area comprises a granite/gneissic bedrock overlaid by residual soil from granite and in some parts by sandy deposits.

Both the topography and the geological conditions are dominant in the ground-water flux regime, conditioning the contamination spread.

An important intermediate objective of the referred survey was to obtain different and coherent patterns of contamination and credible spatial contamination models in which to base appropriate remediation strategies. A previous multivariate statistical analysis was done beginning from very simple proceedings towards more sophisticated ones (Fiúza, A.; Vila, C., 2003).

The present component of the study is focused on comparing alternative geostatistical procedures, namely different kriging methods as well as strategies, used to characterize the qualitative and quantitative spatial contamination dispersion in the investigated area leading to different mapped models. In addition, it is also summarized a risk analysis study, also based on geostatistical procedures used namely to estimate the spatial distribution of interval estimates associated to probability confidence intervals of exceeding selected contaminant thresholds, leading to a partition of the studied zone in terms of their assessed relative remediation needs. The following substances will be considered in the present phase: BTEX, TPH, lead (Pb) in water and soil, total carbon, organic matter (OC), pH and density. In addition it will also be presented some results concerning the mapping of the ground and bedrock topography. The sampling was done either in boreholes or trenches, at shallow depths.

Geostatistical tools are a useful methodology in modeling the spatial variability of such regionalized geoenvironmental variables as well as in their final mapping. However there are aspects that shall be taken in consideration namely those related to scale effects as the support/dispersion issue and variogram regularization, as well as those related to physical circumstances such as anisotropic behavior and possible presence of trend(s).

Statistical Data Characterization

A statistical data analysis was undertaken in order to characterize the various studied parameters based on their respective sample distributions.

In figures 1 and 2 are plotted the sampling location points, in water and soil, relative to the largest and smallest available data sets. In table 1 and in the next figures is presented a summary of some of the sample distributions most relevant features.

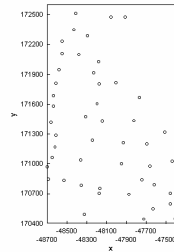


Fig. 1. Sampling points: largest data set

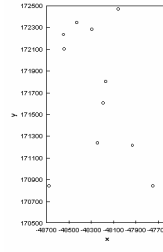


Fig. 2. Sampling points: smallest data set

Table 1. Summary statistics of the studied variables

	Mean	Median	Std. dev.	Variance	Kurtosis	Skewness	Iq. range
pH	6,5	6,4	1,0	1,0	2,2	0,5	1,6
Density	1003,6	986,3	96,4	9285,9	7,1	1,7	96
OC(%)	0,3	0,2	0,4	0,1	8,4	2,3	0,2
C(%)	0,2	0,1	0,2	0,0	8,3	2,3	0,1
TPH	2,2	1,6	1,7	2,8	3,0	1,0	2,2
BTEX	0,1	0,0	0,1	0,01	8,4	2,4	0,1
Pb($\mu\text{g/l}$)	0,4	0,0	1,3	1,8	23,5	4,6	0,2
Pb(ppm)	31,6	24,5	23,8	566,8	4,9	1,5	28,5

The summary statistics presented in Table 1 – mean, median, standard deviation (Std. Dev.), kurtosis, skewness and interquartile range (Iq. range) - corresponds to the sample distributions of the different studied variables in this phase of the study.

All the sample distributions, except for pH which is already naturally normalized, are very positively skewed. Excepting again the pH case, there is no evidence of a possible population's mixture. In figures 3 and 4 are the most positively skewed histograms and in figures 5 and 6 the least skewed among the observed.

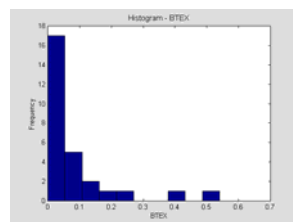


Fig. 3. Histogram of BTEX.

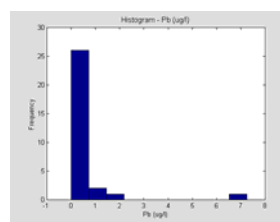


Fig. 4. Histogram of Pb in water.

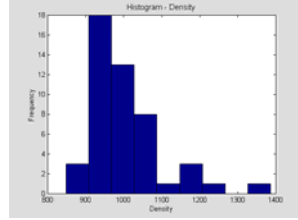


Fig. 5. Histogram of density

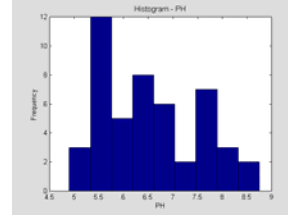


Fig. 6. Histogram of pH

In order to detect the existence of trends, which should be taken in consideration when choosing the most appropriate interpolation tools, data has been projected in the orthogonal main directions planes. In figures 7 and 8 can be seen the TPH data projections respectively in TPH,x and TPH,y planes, the most stationary observed global case, and in figures 9 and 10 the lead in soil data projections respectively in lead,x and lead,y planes, being the trend along the x direction the more significant among the observed cases. The dotted lines limit the 95% confidence interval.

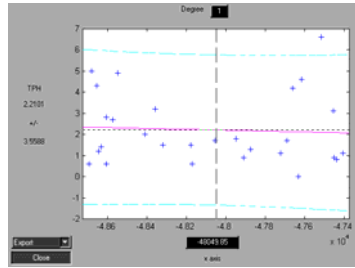


Fig. 7. TPH data projection in TPH,x plane

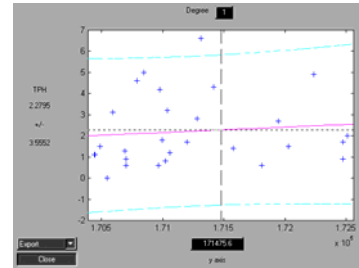


Fig. 8. TPH data projection in TPH,y plane

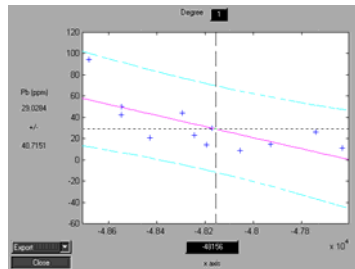


Fig. 9. Pb in soil data projection in Pb,x plane.

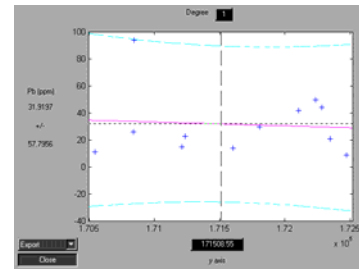


Fig. 10. Pb in soil data projection in Pb,y plane.

The issues, related to sample distributions lack of normality and some data apparent departure from stationarity, and whether or not should these facts be taken in consideration in the interpolation phase of the study, will be discussed ahead.

Multivariate Analysis Summary

In this initial approach data originated by the chemical analysis of groundwater samples, taken from thirty piezometers spread over the studied area, was used to formulate a string of different and complementary multivariate statistical analysis. Several classes of organic compounds, as well as some isolated hydrocarbons, were analyzed in each sample: BTEX, benzene, toluene, ethylbenzene, xylenes, PAH's, DRO, GRO hydrocarbons and lead. The first step of the analysis was the construction of the correlation coefficients matrix between variables (concentrations), showing high correlations between BTEX compounds, its individual components and the gasoline range hydrocarbon (GRO).

A Principal Component Analysis (PCA) allowed detecting three different patterns then a R-Mode Factor Analysis confirmed the previous statements and allowed to distinguish four composition patterns.

The resultant analytical model allows for explaining more than 95% of the variance for the following compounds: BTEX, toluene, ethylbenzene, xylene and GRO. Only 62,7% of the benzene variance is accounted. The model has very low efficiency for PAH (45,5%) and for lead (22,5%); and is totally inadequate for DRO accounting 0% of its total variance. This means that the patterns of the composition variations for DRO, lead and PAH's cannot be physically related with other concentrations. Its occurrence appears to be independent from the simultaneous occurrence of other contaminants and its physical origin distinctive.

If each sample is projected in the Cartesian system of coordinate axes defined by the two retained factors it produces a point. We can distinguish four variation groups: 1st group: sample 28 (high anomaly in BTEX); 2nd group: samples 18 and 5 (anomalies in benzene and ethylbenzene); 3rd group: sample 4 (anomaly in lead); 4th group: all the other samples.

A Q-Mode Factor Analysis allowed a net separation of the uncontaminated samples and all the others that were discriminated into three different patterns.

The analysis clearly pointed to the existence of four different groups: group 1: piezometers 5, 11, 18, 25 and 37: the common pattern is a high concentration in BTEX; Group 2: boreholes 4, 14, 22, 26 and 31: their water samples had an almost nought concentration in any contaminant; group 3: boreholes 7, 17, 33 and 36: physically they correspond to samples with higher values in DRO; group 4: piezometer 3; its composition is next to group one; nevertheless, it differs from the group by the fact that the contamination consists only on toluene.

A Correspondence Analysis (CA) where the objects (piezometers) and the variables (concentrations) were reduced to the same metric, allowed the net definition of different contamination patterns and simultaneously evidenced the nature of transition (progressive or abrupt) between these patterns.

The interpretation was the following: water from borehole 28: abnormal concentration in BTEX; water from piezometer 18: high concentration in xylenes; sample 5: high concentrations in benzene and ethylbenzene; location of samples 25, 11 and 37: determined by simultaneous concentrations of ethylbenzene, toluene and polyaromatic hydrocarbons; location of samples 7, 33 and 36 is originated

by higher concentration of hydrocarbons in the diesel range (DRO); the location of sample 4 results exclusively from the lead concentration, the other contaminants being absent; the location of samples 14, 17, 22, 26 and 31 represent water with no contamination; we could also deduce that concentration in GRO does not have any resemblance with other contamination patterns; the BTEX concentration is specially determined by the concentration in xylenes.

Geostatistical Analysis

A geostatistical analysis was then carried out, comprising the following different main steps: structural analysis; cross-validation, calculation of confidence intervals and mapping using kriging procedures.

The scarce and scattered data sets, the generically very positively skewed related distributions as well as some existing trends were some of the hindrances to deal with.

Data Structural Analysis

The structural analysis for the various parameters confirmed the expected impossibility of carrying out a reliable directional variography due to the scarce quantity of sampled points and the very irregular sampling pattern.

The highly positively skewed sample distributions previously shown raised the question if weather or not the data should be transformed in order to become normally distributed enough. Taking in consideration the fact that it was possible to obtain variograms considered not too much erratic as well as the often controversial back transformations, it has been decided not to transform the data. In addition the observed symmetry in the distribution of the estimated grid points minus the ones obtained through cross-validation supported that decision.

The omnidirectional experimental variograms for the different parameters as well as the respective fitted spherical models can be seen in figures 11 to 18. The structural analysis was performed using the software Variowin.

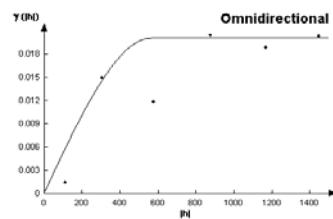


Fig. 11. BTEX: omnidirectional experimental variogram and spherical model

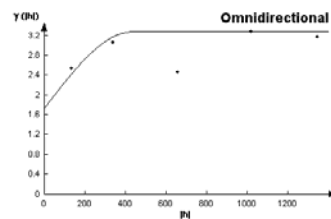


Fig. 12. TPH: omnidirectional experimental variogram and spherical model

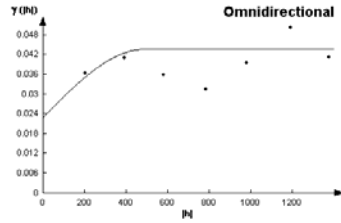


Fig. 13. Total carbon: omnidirectional experimental variogram and spherical model

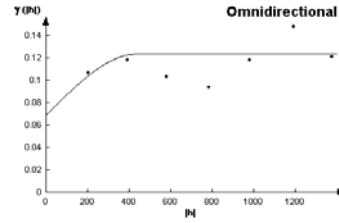


Fig. 14. Organic matter: omnidirectional experimental variogram and spherical model

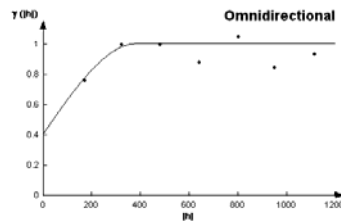


Fig. 15. pH: omnidirectional experimental variogram and spherical model

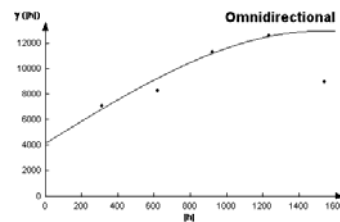


Fig. 16. Density: omnidirectional experimental variogram and spherical model

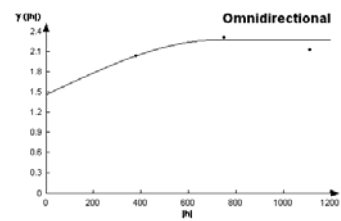


Fig. 17. Pb in water: omnidirectional experimental variogram and spherical model

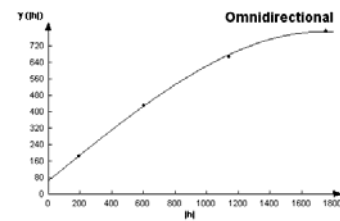


Fig. 18. Pb in soil: omnidirectional experimental variogram and spherical model

The next two figures show the omnidirectional variograms for the variables ground and bedrock topography. In the first case there is a gaussian model fit and in the second a linear fit. These results probably reveal a difference in both variables gradient influence in the x direction, being the trend related to the ground topography influent enough to lead to an unbounded model.

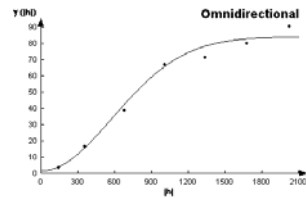


Fig. 19. Bedrock: omnidirectional experimental variogram and gaussian model

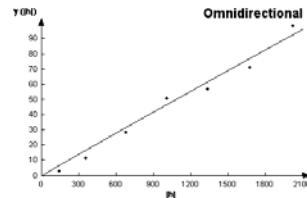


Fig. 20. Ground: omnidirectional experimental variogram and linear model

Cross-Validation

Cross-validation has been used in order to sense the selected variogram models behavior as well as the resulting distribution of estimation errors. Figures 21 to 24 show the cross-validation maps for BTEX and TPH using ordinary kriging (OK) as well as the respective estimation variance maps. As previously referred the obtained errors associated with the cross-validation procedures point towards the normality of the distribution kriging errors, which was a confidence factor for the legitimacy of using the procedures described further on for obtaining spatial distribution maps of interval estimates at chosen confidence levels.

Looking at figures 21 and 25 showing ordinary kriging estimation maps for BTEX, based on the sampled data and in cross-validated data, it is clear the effect of both the scarce amount of samples and the skewness of the related distribution. In fact, namely, the two isolated highest concentration values visible in the estimation map on figure 21, as expected are not represented in the cross-validation map on figure 25. The estimation variance maps reflect the irregular sample pattern, suggesting the location of extra sampling points if a more reliable mapping is thought necessary.

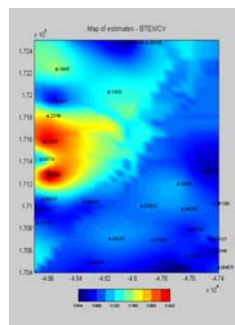


Fig. 21. BTEX: OK cross-validation map

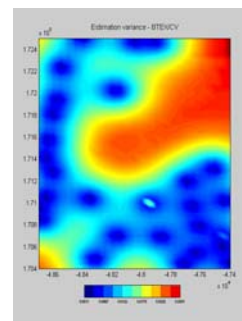


Fig. 22. BTEX: OK cross-validation estimation variance map

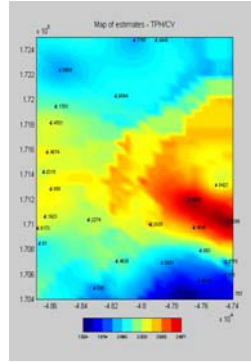


Fig. 23. TPH: OK cross-validation map

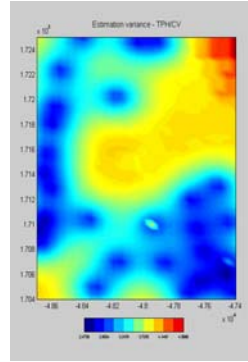


Fig. 24. TPH: OK cross-validation estimation variance map

Estimation

The following point estimation maps were obtained using ordinary kriging (OK) and kriging with a trend (KT). Figures 25 and 26 show BTEX and TPH OK estimation maps.

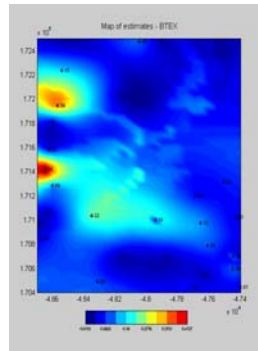


Fig. 25. BTEX: OK estimation map

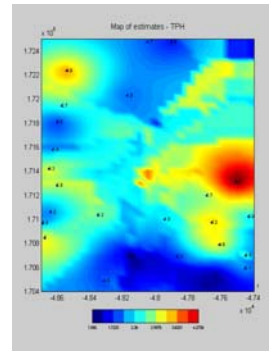


Fig. 26. TPH: OK estimation map

An example of the inverse square method, using the software Surfer, is shown for the variable pH (figure 28) and the OK counterpart in figure 27.

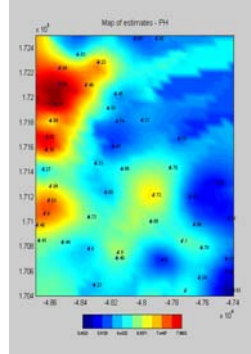


Fig. 27. pH: OK estimation map

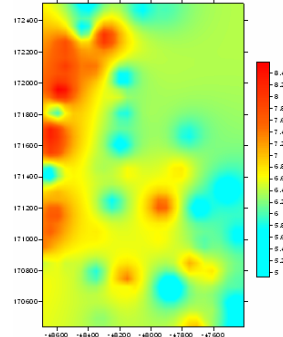


Fig. 28. pH: inverse square distance estimation map

The used algorithms were written in MatLab language. The printing capabilities of the referred software environment were also used in the mapping presentation, except for the experience with the inverse square method, as previously mentioned.

In figures 29 and 30 can be seen the lead in water OK estimation and estimation variance maps. In figures 31 and 32 can be seen the lead in soil KT and OK estimation maps. Taking in consideration that, as shown in figure 9, lead in soil has the most significant trendily behavior along the x direction, kriging with a trend was also used. One of the positive observed features, in comparison with OK results, is an increase in the spread of the estimated point values, reducing the usual smoothing effect of kriging.

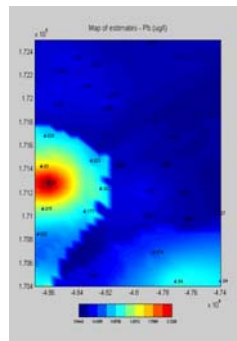


Fig. 29. Pb in water: OK estimation map

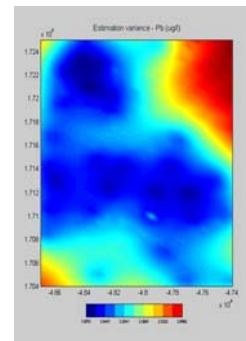


Fig. 30. Pb in water: OK estimation variance map

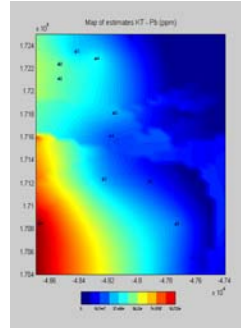


Fig. 31. Pb in soil: KT estimation map.

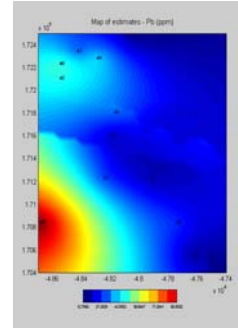


Fig. 32. Pb in soil: OK estimation map.

In the next two figures can be seen the bedrock topography mapping using ordinary kriging and the ground surface mapping using again kriging with a trend. It is obvious the quite similar generic pattern of spatial variation of both variables.

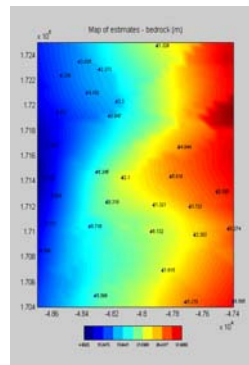


Fig. 33. Bedrock: OK estimation map

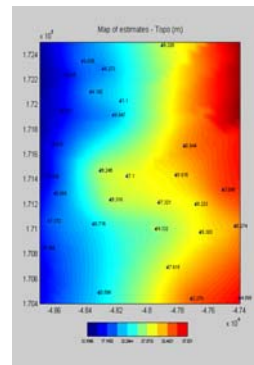


Fig. 34. Ground: KT estimation map

Risk Assessment

Aiming at obtaining probability distribution maps allowing an assessment of the risk related to the different contaminants, a controversial procedure was used which will be briefly outlined. Taking in consideration the fact that the kriging error distributions are often quite normally shaped and in addition the obtained cross-validation errors were rather symmetrically distributed it was decided to assume the normality of the estimation errors distribution and the unbiased condition of the estimator. It was then possible to use the usual standardization formula with the unknown regionalized variable “error”, the square root of the obtained estimation error variance and a zero mean, to compute for each grid point the variable values lower and upper limit of an estimated interval for a given confidence level.

As an example in figures 35 and 36 are shown respectively the lower and the upper limits for the central 95% confidence level interval for variable BTEX.

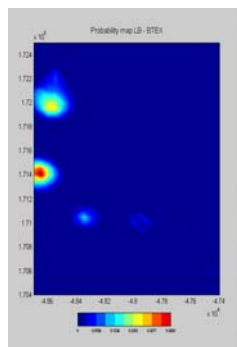


Fig. 35. BTEX: LB_951

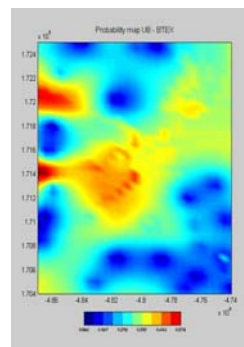


Fig. 36. BTEX: UB_951

It is then possible to obtain a partition of the studied zone in terms of their assessed relative remediation needs allowing a more detailed choice of the locally more appropriate remediation strategies.

Final Notes

The present short summary of a comprehensive geoenvironmental site investigation/characterization survey undertaken in a contaminated costal area in the surroundings of an active oil refinery having an associated petrochemical complex aims, namely, at giving an insight in some of the applied data treatment methodologies as well as some obtained related results and conclusions.

Although having scarce and scattered data sets with much skewed distributions it was decided not to try to normalize the data. This option caused an extra difficulty in obtaining enough well behaved variograms but permitted avoiding the consequences of having to deal with “artificial” distributions and often dubious back transformation effects.

It is worth mentioning the valuable generic advantage of using a geostatistical approach in such a study, due not only to the intrinsic characteristics and flexibility of the geostatistical tools but also in some advantageous although controversial possible “side effects” as it is an example the described risk assessment approach.

References

- Davis John, 1986. Statistics and Data Analysis in Geology, Wiley.
- Carlton, C., Critto, A., Marcomini, A., Nathanail, P., 2001. Risk based characterization of contaminated industrial site using multivariate and geostatistical tools. Environmental Pollution 111, 417–427.
- Fiúza, A., Vila, C., 2003. Application of multivariate statistics to underground water samples contaminated with petroleum products. Consoil 2003 Conference Proceedings, 746-753.

- Goovaerts, P., 1999. Geostatistics in soil science: state-of-the-art and perspectives. *Geoderma* 89, 1–45.
- Jobson, J.D., 1991. *Applied Multivariate Data Analysis*. Vol. I:
- Johnson, Wichern, 1992. *Applied Multivariate Statistical Analysis*, Prentice-Hall.
- Mcgrath, D., Zhang, C, Carton, O. T, 2004. Geostatistical analysis and hazard assessment on soil lead in Silvermines area, Ireland. *Environmental Pollution* 127, 239-248.
- Morrison, 1988. *Multivariate Statistical Methods*, McGraw-Hill, 8th edition.
- Webster, R., Oliver, M.A., 2001. *Geostatistics for Environmental Scientists*. John Wiley & Sons Ltd, Chichester.