Estimating Soil Contamination with Kriging Interpolation Method

Fatima-Zohra Benmostefa Largueche

Department of Mathematics, University Badji-Mokhtar of Annaba, B.P.12,Annaba ,23000, Algeria

Abstract: The main objective of this study was to investigate whether Kriging is a useful tool to estimate the spatial distribution of ground pollutants in contaminated land. The second objective of this work was a more practical one. It consists on the identification of areas that should be subjected to remedial actions and also on deciding which contaminant needs to be considered when remediation processes are taken. To achieve the described objectives, a contaminated site has been studied and the following steps have been followed: The contamination concentration limits beyond which action needs to be taken to remediate the ground contamination, in which case it is important to determine the areas that should be subjected to the appropriate remediation measures. A presentation of a case study will follow. A brief site description is given. Next, a spatial analysis of the site has been carried out. It consists essentially of: Firstly a primary process of the data which means that histograms and an unprocessed representation of the pollutant's distribution has been plotted for each contaminant. Secondly a graphic presentation of the pollution using Kriging interpolation technique is shown. Finally conclusions concerning Kriging applications are given. An assessment concerning Kriging is presented and a balance between the advantages and disadvantages of its use is discussed.

Key words: Ground contaminants, remediation, assessment, regionalised variables, semi-variogramms, geostatistics interpolations techniques

INTRODUCTION

A serious problem faced nowadays is the problem concerning contaminated land. This problem is not only a recent problem but also one inherited from the past. To satisfy industrial and economic needs, land was used cater for these demands without any thought for the future. Contaminated land will be defined as: Land that contains substances that when present in a sufficient quantities or concentrations are likely to cause harm, directly to Man, to the environment, or on occasions to other targets[1].

The size of contaminated land is considerable and according to recent surveys increases continually. Despite great efforts of land restoration it is obvious that new contaminated areas are emerging. According to the Department of the Environment (DOE) surveys, the areas like spoil heaps, excavations, pits, railway land, military land, justifying restoration or reclamation have fallen since 1974 and only the large increase in areas like industrial sites and redundant gas works, docks and power stations, has boosted the overall total. The physical, chemical or biological damage caused is so important that it has recently led governments to adopt, develop and maintain appropriate policies concerning waste management programmes which will certainly help to improve the present situation. An effective policy consists on firstly reducing the existing contaminated land and secondly preventing as far as possible the formation of any further contaminated land.

MATERIALS AND METHODS

The difficulties are now the restoration of these lands. It is obvious that before any piece of land can be used for a new purpose its old use has to identified. It is also important to identify the new use of the lands. Then the identification and the quantification of areas within contaminated land which require cleanup is required to enable the engineers to evaluate and reduce as much as possible the cost of any restoration or remediation.

For such purpose, it is essential to know the specific spatial distribution of the pollutants. So the objective here is to predict the spatial variation of one or more contaminants. In fact, the delineation of highly contaminated areas from areas of low contamination is necessary; since remedial action will only be applied to those areas where the average concentration of contaminants exceeds certain levels. A technique that would give a close approximation of the true spatial distribution of contaminants would therefore be very useful. Sampling is usually conducted to determine the pollutant concentration at several points. The resulting data are used to estimate the average concentration and extreme values. Few techniques are available to estimate the concentration of the contaminant at other points than the sampling points. For instance, one of the easiest methods used to describe the distribution of a Contaminant based upon a few data points is the linear interpolation. It consists on interpolating between data

Corresponding Author: Fatima-Zohra Benmostefa Largueche, Department of Mathematics, University Badji-Mokhtar of Annaba, B.P.12,Annaba, 23000, Algeria

points using predetermined weighting function. Such weighting schemes may not reflect the actual distribution of the contaminant. Another technique commonly used is trend surface analysis. This method assumes that the data can be described by a polynomial surface where random error is assumed to be responsible for a deviation between the observed and estimated values. Such an assumption is not reasonable for interpreting contaminated sites, since the factors which are causing a high concentration of a contaminant at a point are likely to cause the same effect at nearby points. A technique which takes into account the spatial correlation of the data is certainly more appropriate to describe the spatial distribution of the contaminants. A more detailed study of the problem shows that in certain ways the problems associated with describing the distribution of contaminants in contaminated lands are similar to the problems encountered in describing the distribution of mineral deposits.

RESULTS

Mining engineers found solutions for their problems in the field of Geostatistics. In fact this field was developed as a tool for interpreting geologic data and considerable research has been conducted to analyse their spatial distribution. Furthermore Geostatistics offer a choice of a few procedures which do not suffer from the same weaknesses as the previous techniques do. Thus the use of one of the techniques offered may be useful for estimating and interpreting the spatial distribution of a contaminant. The geostatistical estimation technique which may fulfil all the required criteria is called Kriging technique. This technique has been named after a mining engineer D.G.Krige who first applied it in 1951 in the mining field to estimate the average grade and total tonnage of ore reserve of the South African mines. Matheron[2] from Centre de Morphologie Mathematique de Fontainebleau in France, formalised this method and made a second generalisation in 1973. Following his definition this procedure provides the best linear unbiased estimator (BLUE) of the variable under study.

$$Z(X) = \sum \lambda i Z(Xi) \qquad (1)$$

Thus Kriging certainly overcomes some of the latest described problems. Kriging technique was primary developed to solve mining and geological problems but has since found through the latest few years wide applications in different fields such as groundwater, radiological, rainfall and medical applications. In a general way this procedure can be implemented in all cases where some spatial correlation between sampling points is observed. As Kriging has already produced excellent results when used in the mining field, if it can be used to estimate the spatial

distribution of contaminants in contaminated sites, the environmental engineers will possess a powerful tool for evaluating the exact sectors which require clean up. This will allow the reduction of the costs of remediation, which will help to make it more accessible than what it was in the past and thus more manageable and more affordable. The main reference of this work is based on the thesis submitted in partial fulfilment of the requirements for the degree of doctor of philosophy in environmental geotechnology.

$$\gamma \int h) = 1/2 VAR[Z(X) - Z(X+h)]$$

Case study: Introducing the site (ashted waste site): The site investigated is in Ashted parish, Aston park, in north Warwickshire (to the north-east of Birmingham), United Kingdom. This site also used to be a gas works site and it is also likely to have become contaminated with a wide range of chemical substances. This site is to be redeveloped into a non food retail outlet. Johnson Poole and Bloomer had provided a site location and tabulated results of the chemical analysis.

Spatial analysis: From the original tables the contaminants can be divided in two categories. The first category will represent the contaminants whose concentration at sampling points is under the value of the trigger action limits. The second category will include essentially the other contaminants that have at least one sampling point with concentration higher than the actual trigger action limits. In practice the contaminants contained in the first category should not be considered for the analysis of the sites, as we are interested in identifying which sectors of the site require cleanup. But from a theoretical point of view, as we are also interested in assessing the suitability of Kriging for analysing contaminated lands all the contaminants have been undertaken.

Results will be shown for pH, total chromium, total lead and elemental sulphur the latter three being significant contaminants. PH is not itself a contaminant as it is a measured parameter, but when presents in an certain level with other contaminants it can also be considered as a contaminant because of its effects.

The spatial analysis starts by plotting a histogram for each variable to allow evaluation of its distribution, followed by a presentation of 2D Dots maps showing the real location of every sampling point for each contaminant and the real values of every contaminant at their specific points.

To carry out the spatial analysis, the 2D Dots maps are followed by a graphic presentation of the pollutions using the Kriging interpolation technique and the bilinear interpolation technique.

To allow this graphic presentation to be done, the three guidelines were used in deciding which contaminant needs to be considered when determining the areas that require cleanup. These levels will

represent the legend for each contaminant, for both representations.

To terminate the analysis it is essential to decide which contaminant needs to be considered when determining the areas that require clean-up.

Primary process of the data: Before any statistical analysis of the data is done, it is important to first perform some elementary statistics. These elementary statistics are the primary process of the data. It consists of drawing the histogram for every contaminant concentration which provides the first information about the distribution of the probability law of every variable and secondly producing a 2D Dots map, for every contaminant which gives an overview about its real dispersion.

Histograms: Before developing and modelling any experimental semi-variogram, a histogram or scatter plots (correlation diagrams) could be used, to check for outliers and non homogeneity. In fact the histogram is a valuable tool in determining whether the sample distribution is reasonably symmetrical and to detect visually possible outliers, or sample values which are abnormally high or low. However the shape of the histogram is affected by the limits of the classes used to group the samples. A histogram has been drawn for every contaminant. These histograms plots did not show a normal sampling distribution for any of the contaminants under study. From these histograms it was apparent that the data for every contaminant are not normally distributed. A lognormal transformation was then performed to try to normalise the data. None of the histograms of the transformed data has shown a normal distribution. So the non transformed data has been used to perform the sites spatial analysis.

The histograms showing the sample distribution of elemental Sulphur, Carbon Disulphide and Sulphate reflect a constant sample distribution which means that normally these three contaminants would not be considered for any further analysis. Despite the shape of their histogram it was interesting to analyse these three contaminants.

Unprocessed representation of the pollutant's distribution: The unprocessed representation of the pollutant's distribution consists on the 2D Dots maps, which have also been plotted for each contaminant. These maps give the actual value for every contaminant at the exact location of the sampling points and can only be plotted for an irregular dataset. A legend is automatically created and displayed. It can be modified at any time. Titles and axes have been added to make the maps more accurate and easier to follow. These maps were very useful. They have been used to appreciate the level of every contaminant at every sampling point.

Graphic presentation of the distribution of pollutants using kriging interpolation technique: Before starting using the Kriging interpolation process, it is assumed that the intrinsic hypothesis for every contaminant is valid. So it is assumed that the expected mean value of the contaminants in the regions of interest is constant. It is also assumed that there is no significant drift present in the data so that the point Kriging program contained in UNIMAP could be used. Unimap is part of Uniras Interactive menu driven package for contouring and visualising spatial data in two or three dimensions. Isotropy was also assumed for both sites and has been checked by developing semi-variograms in the four main directions using the SPLIT facility provided by UNIMAP.

In order to give a graphic presentation of the distribution of the pollutants using Kriging interpolation process, a semi-variogram has been plotted for every contaminant as it is the first stage of Kriging. After selecting the relevant parameters, a theoretical model has been chosen between the four models provided by UNIMAP to fit these experimental semi-variograms.

In selecting the angle parameters a wide angle has been used to capture as many points as possible in an attempt to improve the fit of the semi-variograms and to avoid the lack of correlation where a default angle was used.

Despite the intrinsic hypothesis, the linear model has been implemented several times as a reasonable approximation since it has been seen that no other transition models could give a better fitting for the available data. Since h 0.7, near the origin both the spherical and the exponential models present a linear behaviour.

DISCUSSION

Identification of the contaminants and their locations for remediation purposes: Our final objective is then to specify both the contaminants and their exact locations (volumes) so that appropriate remediation plans can be undertaken.

pH: The value of pH is estimated to be higher than 10 by the Kriging method. So since the level of alkalinity is estimated to be higher than the action trigger value, remedial actions need to be taken.

Total chromium: In a particular region of the site the value of Chromium total is estimated to be higher than 800 mg/kg. As this level has been suggested as the action trigger level for this contaminant, total Chromium needs to be considered for remediation purposes.

Total lead: The 2D contour maps obtained by using Kriging technique shows that most of the site is typified by total Lead with a concentration higher than the 25

mg/kg the action trigger value. Since total Lead is considered as a source of danger, it needs also to be considered for a cleanup.

CONCLUSION

The use of Geostatistics in general and Kriging in particular was a useful tool to estimate the pollutants distribution in a contaminated site and also to give both the advantages and disadvantages associated with the use of Kriging. Conclusions and recommendations based on personal experiences for future Kriging use will also be given.

Assessment of Kriging: The literature review and the results of the analysis of the eighteen contaminants identified in the Ashted waste site several comments regarding the applicability of Geostatistics and Kriging for analysing contaminated lands can be made. It appears that Kriging could be a very powerful tool for analysing data originating from contaminated sites if some criteria like homogeneity can be warranted. There are, however several problems with the use of Geostatistics as well as some advantages. They will be discussed below.

Advantages of Kriging: The biggest advantage of the Kriging technique over many classical statistical procedures is that Kriging incorporates the spatial correlation of the data, while all the other classical statistical procedures do not. For instance, the use of multiple regressions, a statistical method in trend surface analysis, might seem to be an appropriate tool for analysing spatial data. However, if all the observations are spatially independent, have the same variance and if the mean is given by the fitted expression this method is then optimal which means that the estimates are unbiased and the variance between the estimated values and the true ones is minimum. If a spatial correlation exists between the data, trend surface analysis will not incorporate the additional information provided by the correlation structure and the least square estimates will not be optimal.

Another main advantage of Kriging over other contouring techniques is its ability of quantifying the estimation variance, which will lead to define the precision of the resulting estimates. The standard error map can be used effectively in identifying the areas for which further sampling is necessary in contaminated sites. In fact, the error map shows the confidence envelope that surrounds the estimated surface. It expresses the relative reliability of the map of the estimates values. In areas of poor sampling, the error map will show large values, indicating that the estimates are subject to high variability. In areas of dense sampling the error map will show low values and

at the sampling points themselves the estimation error will be zero.

Another advantage that Geostatistics and Kriging have got over any classical statistics is that they allow the incorporation of the data support. Since the spatial variability of the contaminants being studied is affected by the support size, it should be considered in whatever statistical procedure is used to analyse the data.

Kriging also estimates the average concentration in blocks of different sizes, which will determine which of the areas in the contaminated site require remedial actions.

Kriging will work regardless of the existence of spatial correlation between the data. When observations are independent, Kriging estimates will concur with estimates determined by using least-squares regression analysis.

Compared to the least square estimation, the Kriging method presents another advantage. While the polynomial determined using least square is simply a mathematical expression for the surface, without any physical meaning, the drift is defined as the expected value of the variable and thus has some physical meaning.

Compared to the bilinear interpolation process, the Kriging method has further advantages. While the bilinear interpolation technique estimates values of a contaminant in a sector where practically there is no sampling point so no data at all, the maps obtained by using Kriging estimation process show a blank which encourages the engineer for more sampling.

Disadvantages of Kriging: Kriging techniques are based on a wide range of methods. These methods are derived from the regionalised variable theory. While compared with pre-existing techniques for analysing the data they generally have got great advantages. In fact, they are based on strong theoretical basis. They also allow some estimation of the quality of estimates produced and they have some claim to statistical properties, as for example: unbiased ness, linearity and minimum variance. On the other hand, these methods require certain strong assumptions to be made; assumptions which are rarely met in nature.

In fact, fundamental regionalised variable theory requires that at least the intrinsic hypothesis form of stationarity is true: local variations in the mean are accepted, but the semi-variogram must be necessary stationary over the entire area of interest. In practice, real data sets rarely even approach stationary. Universal Kriging and the generalised covariance method deal with non-stationary data, but even in these methods, the types and amounts of non-stationary are restricted to a few idealised situations. To solve this problem, most authors on Geostatistics suggest that assuming stationary is not so significant since local stationary is assumed; however there is no general proof of this and

no statistical test to determine whether such an assumption is warranted: The fact remains that the theory of regionalised variable cannot be used under conditions where its defined form of stationary does not exist[3].

The second problem faced with this approach is the normality of the distribution. The normality of distribution is one of the assumptions on which the theory of regionalised variable is based. In practise, as most sets of data are not normally distributed, a family of techniques has been created, in an attempt to avoid this constraint. Lognormal Kriging and disjunctive Kriging are part of these methods. The aim of these methods is to transform the data to a normal distribution before using Kriging technique with standard equations. This procedure has got a main disadvantage which is that the variable to be estimated is then non-linear function of the original data. Thus this procedure will make the estimation process certainly quite complicated. Using standard Kriging equations on transformed data will minimise some function other than a simple variance. It will also possibly be biased as the unbiased estimate of a transformed value. So these methods will not produce the BLUE as required[3].

The pivot of parametric geostatistical methods is the semi-variogram which may be computed from the data set under investigation. The experimental semi-variogram obtained will often be different from all the theoretical models. Some skilful interpretations are required to fit one or more models to the empirical curve. It is also important to recognise breakdown of assumptions such as stationary, which will have direct distorting effects on the semi-variogram.

It is important to ensure that the semi-variogram computed from one data set will depend on this particular set. For other data sets the semi-variograms computed will be different.

The estimation of the semi-variogram may be difficult when there is a shortage of experimental points. This problem may be encountered in the case of a contaminated site, where ground conditions may prohibit access to certain areas of the site.

The choice of a technique has to be made. From the semi-variogram some departure from stationary has been diagnosed, it might be considered best to use universal Kriging or generalised covariance. On the other hand, if it is known that the data follow some complex non-normal distribution, it would probably be more appropriate to use disjunctive Kriging. When both situations occur, which should be then the recommended method? A method which will meet such a situation does not even exist and even if it combines the properties of both generalised covariance and disjunctive Kriging will be inevitably be too complex in terms of computing effort[3].

Generally, Kriging requires fewer samples than other spatial estimation techniques for obtaining an acceptable precision. Sometimes however, the number of samples required to estimate the semi-variogram may exceed the number of samples required to achieve a desired level of precision.

Another disadvantage of Kriging is the difficulty to understand and implement it. Much of the terminology and concepts on which Kriging is based are unique to Geostatistics and the majority of the literature is oriented towards mining applications. Therefore, the use of the Kriging technique may require learning both mining and geostatistical terminology. Also, judging from the available literature the use of Geostatistics in the field of environmental geotechnology is not widespread. Literature relative to the subject is very limited. Should problems arise with the analysis of the data, very few sources for assistance are available. On the contrary, numerous sources are available for assistance when implementing classical statistical methods, which may mislead the engineer to use classical statistics in problems where Geostatistics would be more appropriate. For instance, one can be tempted to use classical statistical methods, when, from the data under study it is apparent that the variance is not constant. Even when assuming that there is no correlation between samples, the multiple regression estimates would not be optimal.

The results presented herein show that further development on indicator Kriging is needed. The concentration of a contaminant is likely to be highly variable; this technique may eliminate many of the problems associated with highly variant data sets, for instance the estimation of the semi-variograms.

Since integrating the estimated values with soil guide-lines values to map the risk from those contaminants is a vital step for determining which areas require cleanup, a more complete guide-line may be required.

The experience of dealing with contaminated sites will contribute in reducing the element of uncertainty at all stages of the reclamation of polluted land and a database compiling these experiences will allow practitioners in contaminated land to improve the quality of the remediation.

REFERENCES

1. Fleming, G., 1991. Recycling Derelict Land. Institution of Civil Engineering, Thomas Telford Publishing Company.
2. Matheron, G., 1963. Principles of geostatistics. Econ. Geol., 58: 1246-1266.
3. Henley, S., 1981. Non-Parametric Geostatistics. Applied Science Publishers, pp: 145.