# GEOSTATISTICS COMBINED WITH THE FUNCTION OF INTERPOLATION IN GIS

## Truong Xuan Luan[1] and Truong Xuan Quang[2]

Hanoi University of Mining and Geology, Dong Ngac, Tu Liem, Hanoi

[1] Email:  txluan@netnam.vn,  cntt_mdc@humg.edu.vn

[2] Email: truongxuanquang@yahoo.com

## ABSTRACT

*Geostatistics appeared early in early sixties of the 20th century. First it was applied to research on mines, then in the fields such as: Hydrogeology, Engineering Geology, Geology, Geochemistry, Petroleum Geology, Environmental Geology, Geophysics, applied Hydrology, Environmental Science and Soil Science and it is one of the methods applying interpolation in GIS and remote sensing. Based on factual data of many projects (research on parameters of Nui Hong mine of Thai Nguyen province, Research on physico-mechanical characteristics of soil and groundwater quality in Hanoi area, etc.,), the Researchers proved the efficiency of the geostatistical method. Besides giving highly reliable results and accuracy, the geostatistical method also gives us the character and structure of the spatial variation of research objects as well as the isotropy and anisotropy of the properties under study which so far still cannot be solved completely by other methods.*

## 1.    INTRODUCTION

As we know, it is uneconomical and very difficult to determine the continuous spatial changes of an object of research. This variable is usually determined at a definite number of observation points. The estimation of properties at unobservable (unmeasurable) location is called interpolation. Interpolation is one of very familiar functions of GIS.

There are many interpolation methods and they are classified into 3 main groups.

- Local interpolation consists of: Interpolation by the nearest points (Thiessen area), Linear, Spline function, Average weight (usually arithmetic average weights are used)

- Global interpolation, of which the most common approach is the trend surface analysis, which is one of regression analysis method.

- Interpolation by kriging, which is inn deed geostatistic method. Although this method is difficult and time consuming, it is preferred owing to its good results. The authors have applied geostatistics to quantify the space and evaluate the quantity and quantity of research parameters of various fields. This method has given good results which are rather close to the reality and are highly valued. Within the framework of an article the authors can introduce briefly one of the results of such research. First of all, be aquatinted with some most basic concepts of geostatistics.

## 2.    METHODOLOGY

Objective of Geostatistics is to provide quantitative description of natural variables distributed in space. More concretely, in GIS kriging (of Geostatistics) is an interpolation method based on both "surface analysis" and "average weight" methods. The surface analysis method finds a mathematical equation to describe the general trend of a surface not taking into accounts the local irregularities. Local interpolation is used for calculating the deviation from the global trend due to the local irregularities. The average weight method is used for calculating this variation. The weights of the estimates are determined by the trend of deviations between the curve of the global surface and the observation points (Bourough, 1986).

With kriging, the unknown value to be estimated $Z(x)$ represents both components: the global trend $m(x)$ of the data and the local variation $e'(x)$. This local variation depends on the values of the surrounding data. Besides, there is a dependent non-spatial error to be considered, which is called $e''$ (Jones, 1997).

$$Z(x) = m(x) + e(x) + e''    \hspace{2cm} (1)$$

Due to the diversity of the input data, there are many types of geostatistics: linear, non-linear, stationary, non-stationary, multivariate. In each respective types there are many kinds of kriging: simple (SK), ordinary (OK), disjunctive (DK), Indicator (IK), universal (UK), multivariate (MK), etc. In GIS linear geostatistics is usually used with 2 mathematical hypotheses : second order stationary and intrinsic stationary. In most of commercial software packages not specialized in geostatistic it is usually set as default for the interpolation problem by using ordinary kriging.

In the most common form, the problem related with OK can be expressed as follows: There are n observation points with values $Z(x_1)$, $Z(x_2)$,..., $Z(x_n)$ at observation points $x_1$, $x_2$,...,$x_n$ distributed in the neighborhood of Xo. The best estimator at Xo is in the form of:

$$Z^*(x_o) = \sum_{i=1}^{N} \lambda_i Z(x_i) \hspace{2cm} i = 1,..., n \hspace{2cm} (2)$$

$\lambda$ - kriging weight are determined such as to minimize the error variance

$\sigma^2_k$ = var $[Z^*(x_o) - Z(x_o)]$ checking for the unbiasedness constraint : $E[Z^*(x_o) - Z(x_o)] = 0$ (mean error = 0). To satisfy this condition, it has been proved that $\sum_{\alpha=1}^{N} \lambda i = 1$. This is the first optimal condition. The second optimal condition is: minimum mean squared error, i.e. $E[Z^*(x_o) - Z(x_o)] \rightarrow$ minimum (The mean squared error is called the variance error or variance of estimation).

To be in accordance with the above optimal conditions, i.e. there is no systematic bias and the estimated mean square error is minimum, by the Lagrange multipliers ethod, we obtain the following kriging system for determining $\lambda\alpha$ as follows:

$$\begin{cases} \sum_{i=1}^{n} \lambda_j \gamma(X_1 - Y_1) + \mu = \gamma(x_i - x_0) \\ \sum_{i=1}^{n} \lambda_i = 1 \qquad\qquad \alpha = 1,...,n \end{cases} \qquad (3)$$

Where:    $\gamma$ (xi, xj) - value of variogram for distance $|$xi, xj$|$

$\mu$ - Lagrange multipliers

Kriging variance $\sigma^2_k$ is determined as follows:

$$\sigma_k^2 = \sum_{i=1}^{n} \lambda_i \gamma(x_i - x_j) - \gamma(x_o - x_o) + \mu \qquad (4)$$

$\sigma^2_k$ is very important in evaluating the properties under study, based on it we know the reliability of the results of the problem solution.

For interpolation by kriging, one must determine correctly variograms (also called structural functions). Variogram is defined as a half of mathematical expectation of random variables. Thus,

$$\gamma (h) = \frac{1}{2} [Z(x) - Z(x+h)]^2$$

A practical variogram is determined as follows:

$$\gamma(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(xi) - Z(xi + h)]^2 \qquad (5)$$

Where Z(x), Z(x+h): two known values with separation distance h

N (h): number of pairs of data points with separation distance h

It is important to transfer $\gamma$(h) values from the practical model to theoretical model. There are many kinds of variogram models: linear, spherical, power, exponential, gaussian, etc. In reality (especially when the parameters (property) under study varies strongly and in complicated manner), in variograms usually occur nugget effect and also pure nugget effect. Thus, the attribution of variograms to a certain appropriate kind of model and their exploitation is very important. This is because through variograms we get to know:

- The extent, characteristics and structure of the variation of the parameters under study.

- Decision of fitting the isotropy or anisotropy of a parameter under study

- Basis for determining the kind of suitable kriging to give good estimation results.

## 3.    STEPS OF RESERCH AND RESULTS

### 3.1    Steps of research

The researcher, besides mastering algorithms, must understand well the object of research. The steps of interpolation by kriging can be summarized as follows:

- Data preparation

- Data control analysis and determination of statistical parameters: mean, variance, coefficient of variation (v), correlation coefficient (in case of multivariate) suitable with one of probability distributions: normal, log normal, gamma normal, chi-square, etc., The distribution functions are usually tested by histogram, skewness, kurtosis, probability methods.

- Spatial quantification by variograms: Determination of experimental variograms. Modelization of variograms. Exploitation, structural analysis by variograms (occasionally also covariances)

- Kriging :Selecting suitable type of kriging. Selecting suitable neighborhood (radius of study), depending on the radius of influence, isotropy, anisotropy, etc.Calculation (interpolation)

- Evaluating the reliability of the interpolation by kriging performed, drawing 2D and 3D maps and conclusion.

### 3.2    Results

#### 3.2.1    Data

The authors have applied geostatistics in investigating many kinds of data of various fields. Within the scope of this paper we set forth only the results of research and evaluation of the radioactive substances in the Nui Hong coal mine.

The mining area is located in Thai Nguyen province (about 110 km north of Hanoi). The coal was formed in the swamp and lake condition, so its distribution area is small, the seams are not stable, in the form of lenses, formed in Norian - Rhaetian stages ($T_3 n$-$r$). It is now being exploited and used. It is specially notable that the coal has high contents of radioactive substances, especially germanium (Ge). The concentration of these elements are extremely variable. It has been only recognized that Ge content is high where the coal thickness is large. To investigate the radioactive elements, some geophysical methods have been applied, including the gamma spectrometric survey. The survey network was arranged in lines with spacing 250 m, with the survey points on each line at intervals of 50 m. The total number of survey points is 464.

#### 3.2.2    Results

- Statistical characteristics of the parameters under research  show:

    - The radioactive elements are on high, variable intensity. The coefficient of variation [V] according to the data which has been transferred to *ln* form is equal only 1/6 that of the original data.

- The distribution is nearly standard log normal.

• In reality due to many reasons the radiometric survey network is uneven. For calculation, we have grouped the survey points in certain $\alpha$ direction, where each value of certain $Z(x)$ is combined with all information within the range of $(\alpha \pm d\alpha)$ which fluctuate around $\alpha$. The grouping by angle $\alpha$ is carried out at the same time with the grouping by distance $\alpha [r \pm \varepsilon (r)]$.

For studying the isotropy and anisotropy we established $\gamma(h)$ values in many directions. To make full use of the measured points and to meet the actual conditions at the mine, we have established the $\gamma(h)$ values in 4 directions, i.e. $\alpha = 0^o$, $\alpha = 45^o$, $\alpha = 90^o$ and $\alpha = 135^o$; $d\alpha = \pm 22.5^o$

The results of research allow to confirm that:

+ The radioactive elements vary irregularly, with large nugget effect, expressed in the fact that survey points located near each other (about 50 m) and spatially related with each other are quite different in value.

+ All four directions have two structures: nugget effect and spherical. However they have different sills and ranges.

+ The parameter under study indicate a clear local anisotropy, with the dominant direction being NE-SW and with anisotropy index of 2.7.

The orientation of the radioactive anomaly fully coincides with the distribution of the coal seams underground.

The results of research by this method for coal quality parameters show that the radioactivity is very dispersed

+ To be economical, it is possible to reduce the density of the survey network to 300 x 100 m, with the dominant direction of NE-SW and the survey may be non-linear.

• When interpolating by kriging, besides implementing strictly the above mentioned steps, ensuring the precision of the interpolation, one must pay attention to time consumption in running the computer, in particular:

- Reducing the number of directions of the kriging system (selecting appropriate neighborhood).

- Reducing the number of kriging equations (grouping kriging weights together)

- Establishing quickly $\gamma(h)$ values.

- Selecting suitable kriging methods.

The results of interpolation has been used for compiling two maps.

# 4. CONCLUSION AND DISCUSSION

• The results of research on structural functions are of many benefits, contributing to the quantification of the object of study; justifying the density and shape of the survey (sampling) network.

• The Nui Hong coal basin has high content of radioactive elements, which varies complicatedly, therefore they must be studied and used as useful elements. At present, as these elements still cannot be separated from the coal, besides causing the loss of the natural resources, the coal extraction and utilization also causes environmental pollution.

• The interpolation by kriging, although is difficult and time consuming, is of high accuracy and reliability. It can overcome the deficiencies of other interpolation methods. In our opinion, this method should be considered as a leading method in GIS.
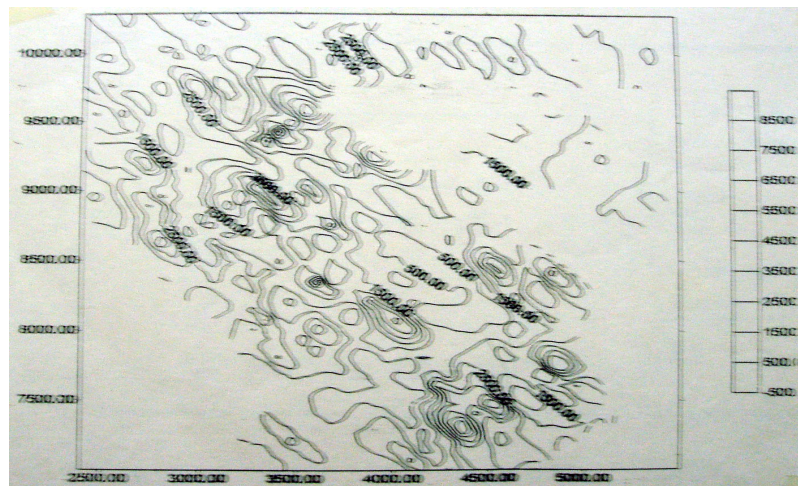


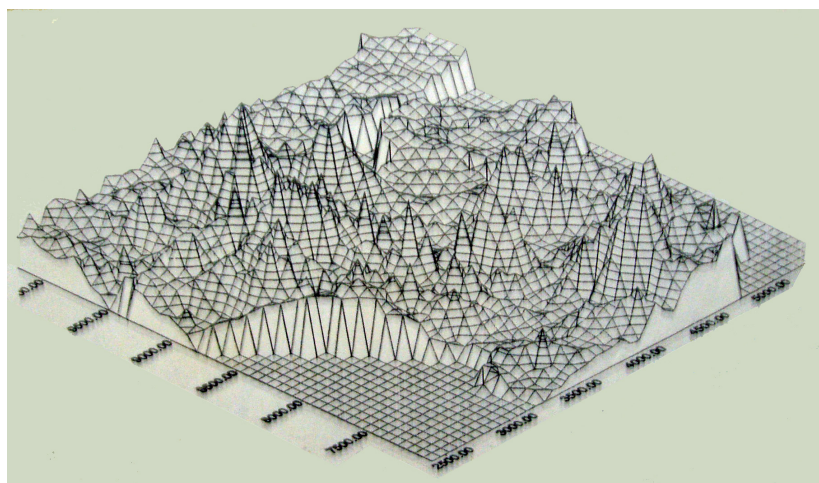**Figure 1. Iso-content map of radioactive elements (2D)**



**Figure 2. Surface map showing the content of radioactive elements (3D).**

## 5. REFERENCES

[1] Goovarert p., 1997. *Geostatistics for natural resource evaluation*. Stanford University, USA

[2] Jones C., 1997. *GIS and computer cartography*. Addison Wesley Longman Limited, Essex.

[3] Luan Truong Xuan. 1995. Geostatistics and Infomatics in research of some coal mines. Ministerial level research project.

[4] Matheron G., 1965. *Les variables RÐgionalisÐes et leur Estimation*, Maisson, Paris

[5] Matheron G., 1982. *Pour une analysis krigeante des donnÐes rÐgionalisÐes*. Internal report N 732. Centre de Geostatistique, Fontainebleau.