Available online www.jocpr.com

Journal of Chemical and Pharmaceutical Research, 2014, 6(6):1796-1799

Research Article

ISSN : 0975-7384
CODEN(USA) : JCPRC5

Research on geostatistical analysis approaches

Linan Wang[1], Jinxin He[2*], Tianyu Zhang[2] and Pan Zhu[1]

[1]*Jilin University of Finance and Economics, Changchun, China*
[2]*Jilin University, Changchun, China*

ABSTRACT

Geostatistical analysis methods are widely used in many fields, has become an important branch of statistics. For a long time, Geo-statistics can not be combined with GIS analysis model very well, which became a big regret of GIS software. ESRI ArcGIS Geo-statistical Analyst made a bridge between geo-statistics and GIS, making the complex geostatistical approaches be easily implemented in this software, reflects the people-oriented, visual development trend. This combination have an important groundbreaking, predicted by measuring the surface of statistical error, GIS people could be able to quantified the quality of the surface of the model for the first time.

Keywords: Geostatistical; Spatial analysis; ArcGIS; Analysis model; Geostatistical analysis

---

INTRODUCTION

Geostatistics is also called the geological statistics, is famous in French statisticians G. Matheron gradually formed on the basis of a large number of theoretical research of a new branch of statistics [1]. It is based on regionalized variable, with the aid of the variation functions [2]. The research is both random and structured, or spatial correlation and dependence of natural phenomena of a science [3]. All structural and spatial data and randomness, or the spatial correlation and dependence, or on the research of spatial pattern and variation, and the data is used to estimate the optimal unbiased interpolation, and simulate the characteristic of these data, volatility, all can use geostatistics theory and approaches.

Geostatistics and classical statistics are in common, since they are both on the basis of a large number of samples, based on the sample frequency distribution or the mean and variance of attribute values and their corresponding rules of analysis, determine the spatial distribution pattern and correlation [4]. But the biggest characteristic of Geostatistics different from classical statistics is that Geo-statistics considers the size of the sample values and attaches great importance to the sample space location and the distance between samples, so which makes up the defect of the classical statistics that ignore the spatial orientation [5]. Thus, Geo-statistical analysis includes assumptions, regionalized variable theory base, variation analysis and space valuation.

1. The premise of geostatistics

1.1 Random Process

Same as the classical statistics, geostatistics is also on the basis of a large number of samples, through the analysis of the rule of sample room, explore its distribution, and to make predictions [6]. Geostatistics regards all sample values in the study area as the result of a random process, namely all samples are not independent of each other and they follow certain inherent law. Therefore, the destination of geostatistics is to reveal the inherent law and to make some predictions.

1.2 Normal Distribution

In the statistical analysis, it was assumed that a large number of samples are to obey the normal distribution, hence

geostatistics is no exception. After achieving the data to deal with the analysis in the first place, geostatistics deal with data transformation to accord with normal distribution and try to select reversible transform, if the data do not conform to the normal distribution assumption.

1.3 Stability

Repeating for statistical point of view is the theoretical basis of geostatistics. Thinking of statistics from a large number of repeating observations can forecast, estimate and understand the estimates of variability and uncertainty. For most of the geospatial data, the stationarity assumption is reasonable. It includes two kinds of stability: one is the mean is smooth, that is, assuming that the mean is constant and has nothing to do with the position; the other is related to the covariance function of the second order smoothly and the half variation function of intrinsic stability. The second smooth order is to assume that the same distance and direction of any two points of covariance is the same, the covariance is only associated with the value of these two points and has nothing to do with where they are. Intrinsic stationary hypothesis refers to any two points with the same distance and direction of variance (i.e., the variation function) is the same. The second smooth order and the intrinsic smooth are to have the basic law and the basic false. Setting by covariance function and variable function can predict and estimate the uncertainty of prediction results.

2. The regional variation

When a variable presents certain spatial distribution as regionalized variables, it often reflects some characteristics or phenomenon of the area. The difference between Regionalized variable and general random variable is that the general random variable values accord with certain probability distribution, and regionalized variable values accord with the different location within the region and take different values. As regionalized variable values determine the location in the territory, it shows the general random variables, that is to say, it is related to the location of a random variable. In the actual analysis, it often uses the method of sampling for the value of the regionalized variables in a certain area, which is characterized as regionalized variable space point function [6]:

$$Z(x) = Z(xu, xv, xw)$$

According to its definition, regionalized variable has two significant features: randomness and structure. First of all, regionalized variable is a random variable, it has the local, random and abnormal characteristics; Secondly, the structure characteristics of the regionalized variable has a certain, namely variable at point x and the deviation distance with the value of x + h in h point of Z (x) and Z (x + h) have a certain degree of similarity, namely the self correlation degree depends on the distance between two points h and variable characteristics. Besides, regionalized variables also has the space limitations (i.e., the structural performance for a certain range), degree of continuity and different degree of anisotropy (i.e., all directions show the correlations).

3. Variation analysis

3.1 covariance function

Covariance is also called the semi-variance, which is the difference between the two random variables. In probability theory, random variables are named X and Y, and Variance is defined as [7]:

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y)]$$

The spatial covariance function can be represented as [8]:

$$C(h) = 1/\sum[Z(xi) - Z(xi)][Z(xi + h) - Z(xi + h)]$$

Among them, the Z (x) is a regional random variable and satisfy the second-order stationary hypothesis, namely spatial distribution regularity of random variable Z (x) is not due to the displacement change; H for two sample points of space distance; Z (xi) is the Z (x) in the space point xi sample values; Z + h (xi) is the distance Z (x) in the xi deviates from the value of the samples of h (I = 1, 2,..., N (h)]; N separation distance for h (h) is the total number of sample points; Z (xi) and Z + h (xi) of Z are respectively (xi) and Z + h (xi) sample mean, namely [8]:

$$Z(xi) = \sum Z(xi)$$
$$Z(xi + h) = \sum Z(xi + h)$$

On the type of n is regarded as sample unit number. In general, Z (xi) indicates the Z + h (xi) (special circumstances can be considered approximately equal).

3.2 Half variation function

Variation function is also called half variation function, is the unique function of spatial statistical analysis. Regionalized variable Z (x) at point x and x + h value of Z (x) and Z (x + h) are half of the variance of poor known as regionalized variable Z (x). The half variation function is remember to r (h), 2 r (h) as variable function According to the definitions are [8]:

$$r(x, h) = Var[Z(x) - Z(x + h)]$$

3.3 The variation analysis

The half variation function and covariance function have the size of the statistical correlation coefficient as a function of the distance, which is the quantitative geography close similarity theorem. Fig. 1 depicts one and a half of typical variation function and its corresponding covariance function diagram.

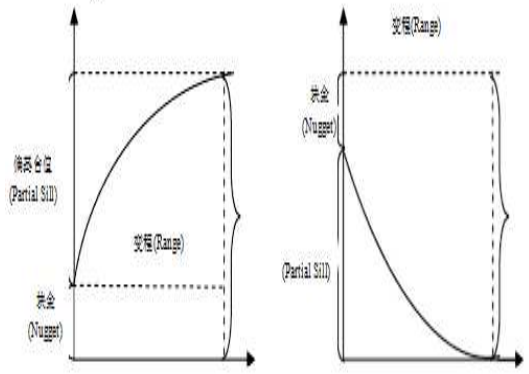

Fig. 1 One and a half of the variation function diagram

The half variation function curve and covariance function curve reflect a sampling point with adjacent sampling point of spatial relations. Moreover, they effect on abnormal sample point that has good detection. ESRI ArcGIS geostatistical analysis module can be used in any one of the two, generally uses the half variation function. In the half variation curve has two very important points: interval for zero point and half variant function when reaching steady turning point, is produced by the two point and four corresponding parameters: the piece of gold value (Nugget), cheng (Range), base station value (Sill) and Partial base stations (Partial Sill). Their meaning is as follows:

Nugget value (Nugget): in theory, when the distance between the sampling points is 0, the half variation function value should be set zero, but due to the presence of measurement error and variation space, when the two sampling points are very close to their half variant function value is not zero, the piece of gold value is produced. Internal error caused by the instrument, measurement errors space mutation is a natural phenomenon in a certain space within the scope of the change. The either side or both together produce a piece of gold value.

Base stations values (Sill) : when the distance between the sample point h increases, the variation function r (h) and a half from the original piece of gold value reached a relatively stable constant, the constant value is called base stations. When the half variation function value is more than base stations, namely function value does not change according to the sampling point distance, spatial correlation does not exist.

Partial base stations values (Partial Sill): the value of base stations and the piece of gold value difference.

Variation (Range): when the half variation function value with the initial piece of gold value reached base stations, the sample point distance called variation. Variation expressed in some scales and the scope of spatial correlation, its size is limited by observation scale. Within the scope of the variation, the smaller the distance between sample points, the similarity, namely spatial correlation is larger. When h > R, regionalized variable spatial correlation of Z (x) does not exist, as some point with the distance is greater than the variation of known points, point data cannot be used for interpolation or extrapolation. When a limited amount of sample points is after hours, which may appear on the graph with all r (h) material Nugget, namely curve is an approximate straight line parallel to the abscissa, and now the half variation function of pure Nugget will effect on it. This is due to the limited sample interval, point to point is a big change, namely each sample is random, do not have spatial correlation and average of sample points in the area.

This is the best estimate. At this time it only increases the sample interval, which can reflect the spatial correlation

between the samples. The strength of the spatial correlation can be reflect by Partial Sill/Sill, and the value is greater than the spatial correlation. Accordingly, the Nugget/Sill called basal effect and the variation characteristics between the values are the greater of the variation between the said samples is more caused by random factors.

4. Spatial valuation

A complete process of geostatistical analysis or spatial valuation process is generally as follows: the first one is to obtain the original data, inspect and analysis of data, such as looking for the data implied characteristics and laws e.g., whether it is normal distribution, is there a trend effect, anisotropy, etc; Then it select the suitable prediction model of surface, including the half variation model of choice and selection of prediction model; Final the inspection model or several models that is reasonable were compared. Although in the case of ArcGIS using spatial statistical analysis module to complete the above process is very simple, but follow a structured process is still very important [8].

(1) Data show

It is added in the ArcMap data view window and display the data layer to be analysed.

(2) Data check

It analysis on the statistical properties of data sets to deeply understand the data. Data check content includes inspected data distribution, finding outliers, global trend analysis, detection of spatial autocorrelation, direction variation, and covariance analysis of multiple data sets.

(3) Model fitting

Based on the understanding of the data, it chooses one preliminary thought suitable model to create the surface. Comprehensive data check helps to choose the suitable model.

(4) Model diagnosis

The output of the assessment model can understand the selected model's prediction effect of unknown values. Diagnosis includes forecasting accuracy and validity of the model.

(5) Model comparison

By setting different parameters or selecting multiple optional models to create the surface, which can be determined by comparing the analysis models that have better prediction effect on unknown values.

CONCLUSION

This paper introduces the principle of geostatistical analysis, provides an approach for geostatistical analysis, and statistical analysis for the further applications.

Acknowledgements

This research work was supported by Jilin Province Foundation of Applicable Sciences (20130102033JC). We also thank the anonymous reviewers for their helpful suggestions.

REFERENCES

[1] J Nian; L Tang; X Liu. *Oil Geophysical Exploration*, **2004**, 43(3), 278 -282.
[2] W Wang. *Journal of Nanjing Industry Professional Technology Institute*, **2003**, 3 (2), 9-13.
[3] F Chen; D Du. *Journal of Wuhan University (Information Science Edition)*, **2002**, 27(4), 391-396.
[4] J Zhang; Y Zhu; S Li. *Northwest Population*, **2010**, (3), 91-94.
[5] P Wang; X Li; Y Zhang; et al. *Journal of Geological Engineering*, **2013**, (1), 115-122.
[6] X Li; Y Zhou; T Wang; et al. *Journal of Rock Mechanics and Engineering*, **2011**, 30(10), 2049-2056.
[7] DSYSF Hoerger; K Yan. *Journal of Geological Science and Technology*, **1988**, 7(1), 110-115.
[8] H Ver; M Jay; K Konstantin; et al. Using ArcGIS Geostatistical Analyst, 1st Edition, ESRI, Redlands, **2001**; 89-90.