Heterogeneity Classification and Structural Analysis with Geometrical Characteristics

LIU Feng[1], GUO Jianwen[2]

[1] Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences,
320 Donggang West Road, Lanzhou, Gansu, China.
Telephone: (0086-13679445079)
Email:liufeng@lzb.ac.cn

[2] Cold and Arid Regions Environmental and Engineering Research Institute, Chinese Academy of Sciences,
320 Donggang West Road, Lanzhou, Gansu, China.
Telephone: (0086-18993161162)
Email: guojw@lzb.ac.cn

1. Introduction

The expressions of directional heterogeneity activation function (DHAF) and directional heterogeneity function (DHF) were defined at first, and a classification of spatial heterogeneity based on these two functions was proposed. Then the existing spatial interpolation approaches for different heterogeneity classification were discussed. Last but not least, the principle and application of structural analysis of directionally discrete heterogeneity (DDH) based on geometrical characteristics was put forward in detail.

All the study was supposed to be limited to 2-dimension space.

2. Heterogeneity Classification of Spatial Data

DHAF and DHF can be defined in equation 1 and equation 2, respectively.

$$N(\gamma(\Delta h), D) = \begin{cases} 1, & \gamma(\Delta h) \text{ is valid in } D \\ 0, & \gamma(\Delta h) \text{ is invalid in } D \end{cases}, \quad \Delta h \to 0, \tag{1}$$

$$\int_{D_0-\pi}^{D_0+\pi} N(\gamma(\Delta h), D)\, dD = M, \tag{2}$$

where D is an arbitrary spatial direction of data sample point and $N(\gamma(\Delta h), D_0) = 1$.

According to the value of M, the heterogeneities of spatial data can be classified into 3 groups. For all the sample points,

i) If $M = 2\pi$, the heterogeneity is directionally continuous. In this case, DHAF must be uniformly continuous in $[D_0 - \pi, D_0 + \pi]$.

ii) If $M = 0$, DDH can be found.

iii) If the equation $0 < M < 2\pi$ can be obtained by one or more points, the heterogeneity will be neither continuous nor discrete in these points.

3. Structural Analysis of Spatial Data (Review)

Spatial Data with a directionally continuous heterogeneity has been studied sufficiently. Traditional Kriging approaches, on the assumption that semivariogram is valid in any arbitrary angle, have been efficiently applied in many fields, such as spatial interpolation of soil moisture. Nevertheless, what has not been adequately addressed is the structural analysis of DDH. The latest advances are as follows:

Using LVA (Locally Varying Anisotropy) Field Model. Subdividing the research region into sub-regions as required, and calculating the major direction of anisotropy in each sub-region, the LVA can generate the heterogeneity distribution pattern. One method is Local Anisotropy Kriging (LAK, Chris et al. 2005). LAK is an iterative approach which calculates the gradient algorithm and kriging with local anisotropy repeatedly, and condition for the termination is that change of estimate E is less than a preset value. Another way to employ LVA is searching an optimal point-to-point route with the maximum attainable covariance, following a solution of kriging equations (Boisvert et al. 2009). Although Kriging with LVA can simulate the essential characters of spatial data, points excluded from research region are necessary in estimated process, which, however, may not be obtained in DDH.

Using stream distance and regarding flow as weight (Ver Hoef et al. 2006). These "compound Euclidean distances" constitute semivariograms, and might be worthless except spatial statistics for streams and rivers.

Using Non-Euclidean distance measures in semivariograms with the concept of isometric embedding (Curriero 2006). This method cannot describe the characteristics of study area accurately unless taking advantage of exponential covariance function.

4. Structural Analysis of DDH

A suggestion that research region with DDH displays a series of geometrical characteristics conspicuously is sound. It is therefore principal to analyse the geometrical characteristics quantitatively when semivariogram in such area is about to be calculated.

4.1 Analysis of Geometrical Characteristics

The route between two sample points can be counted as a curve C with a twice differentiable subsection function. In each section, the curve's geometrical characteristics comprise arc length ς, curvature κ and torsion τ, which must be introduced entirely when considering the range. if $\chi = \chi(\varsigma, \kappa, \tau, \Delta h)$ stands for a quantification of geometrical characteristics, semivariogram would be defined as

$$\gamma(h)=\gamma[h(\varsigma,\kappa,\tau,\Delta h)]=\gamma[\varsigma(1+\chi)]. \tag{3}$$

Consequently the main issue is how to derive a analytic formula of χ.

This problem can be solved separately. Total curvature, the curvilinear integral with respect to curvature pluses sum of angle between consecutive tangents, illustrates an accumulation of degree of crook, but relates to the length of arc. Therefore, the new concept "relative total curvature" (RTC), a ratio of total curvature of subsection c_i to the one of cylinder helix, was proposed. The expression of cylinder helix is given in equation 4,

$$r(t)=(a\cos t, a\sin t, bt), a=\frac{|\Delta h_i|}{2}, b=\frac{|\Delta h_i|^2}{2\varsigma}, t\in\left[0,\frac{2\varsigma_i}{|\Delta h_i|}\right], \tag{4}$$

where Δh_i and ς_i are straightaway distance and arc length between the two ends of c_i respectively. In view of its arc length, together with projected length in the vector Δh_i, are calculated as same as c_i's, RTC is found to be

$$\mathrm{K}_i^{\circ}=\left|2\bar{\kappa}_i \middle/ \Delta h_i\right|, \tag{5}$$

where $\bar{\kappa}_i$ and $\frac{\Delta h_i}{2}$ represent the Lagrange mean curvature value of c_i and cylinder helix respectively, so RTC has nothing to do with the length of arc.

The same method can be used to define a relative angle between consecutive tangents as $\theta^{\circ}=\theta/\pi$, and relative total torsion equals to zero when it's confined to 2-dimensional space. So it is reasonable that replacing χ with the equation 6,

$$\mathrm{K}^{\circ}=\sum \mathrm{K}_i^{\circ}+\sum \theta_i^{\circ}=\sum\frac{2\bar{\kappa}_i}{|\Delta h_i|}+\sum\frac{\theta_i}{\pi}. \tag{6}$$

4.2 Experimental Semivariogram

Lemma 1. Let the sample points $\{u_i \mid u_0=u,\quad u_1,\cdots,u_{n-1},\quad u_n=u+\Delta h\}$ be a partition set of C, then $S(\varsigma)$, parametric equations based on arc length determined by interpolation of sample points, exists eternally. Here, the equations should satisfy some conditions as follows:

1) $S_i^{(k)}(\varsigma)=\lim\limits_{\Delta\varsigma\to 0}\dfrac{S_i^{(k-1)}(\varsigma+\Delta\varsigma)-S_i^{(k-1)}(\varsigma)}{\Delta\varsigma}<\infty$, $\varsigma\in(\varsigma_{i-1},\varsigma_i), k=1,2$.

2) if $\Delta\varsigma = \max(\varsigma_j) \to 0$, $\overline{\kappa}_{Si}$, the Lagrange mean curvature value of $S_i(\varsigma)$, uniformly converges to $\overline{\kappa}_i$. Additionally, θ_{Si}, angle between consecutive tangents determined by $S_{i-1}(\varsigma)$ and $S_i(\varsigma)$, uniformly converges to θ_i。

It can be proved with the "not-a-knot" cubic spline interpolation (de Boor 2001). Based on this theorem, some conclusions will be drawn:

i) Quantification of geometrical characteristics could be obtained by interpolation as long as sampling appropriately.

ii) if $\Delta\varsigma = \max(\varsigma_j) \to 0$, K_S°, RTC of $S(\varsigma)$, will uniformly converges to K°.

It is therefore reasonable that replacing K° with K_S°.

As a result, the experimental semivariogram of DDH will be expressed as

$$\gamma(h^*) = \frac{1}{2N(h^*)} \sum [Z(u) - Z(u+h^*)]^2, \tag{7}$$

where $h^* = \sum \varsigma_i [1 + 2\left|\frac{m_i - m_{i-1}}{\Delta h\varsigma_i}\right|]$, and $\{m_i\}$ are solutions of equations consisting of the "not-a-knot" boundary conditions (Bica 2012, Su et al. 1980).

4.3 Example

Without any intersections, there is a significant curvilinear characteristic in the Qinghai-Tibetan Railway, making it possible and pretty easy to analyze DDH.

Three kinds of data types, recorded at 33 monitor locations scattered in roadbed, were input parameters passed to equation 7: embankment settlement data, longitude and latitude, railway mileage. Fig. 1 displays semivariogram utilizing exponential model with a range of 35.47km, a sill of 0.42km and a nugget of 0.20km.

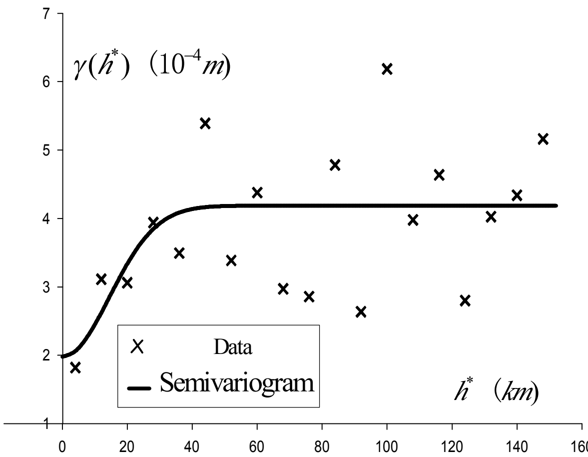

Figure 1. Semivariogram for the embankment settlement data

4.4 Conclusions and Discussions

Improving and supplementing the spatial heterogeneity, the above technique should still confront with and overcome some problems. For instance, assessing spatial uncertainty, 3D analysis, tolerances and nesting other semivariogram models. In addition, research on spatial interpolation based on geometry characteristics should be continued in depth.

5. Acknowledgements

Financial support for this work was provided by Basic Research on Permafrost Engineering in Qinghai-Tibet Plateau (973 Program, 2012CB026106). Thanks to Li Xin, Laboratory of Remote Sensing and Geospatial Science, CAREERI, CAS, for comments on a previous draft.

6. References

Chris B M and Judith J J C, 2005, Mapping Curvilinear Structures with Local Anisotropy Kriging. *Mathematical Geology*, 37(6): 635-649.

Boisvert J B, Manchuk J G and Deutsch C V, 2009, Kriging in the Presence of Locally Varying Anisotropy Using Non-Euclidean Distances. *Mathematical Geosciences,* 41: 585–601.

Ver Hoef J M, Peterson E and Theobald D, 2006, Spatial Statistical Models that Use Flow and Stream Distance. *Environmental and Ecological Statistics*, 13: 449–464.

Curriero F C, 2006, On the Use of Non-Euclidean Distance Measures in Geostatistics. *Mathematical Geology,* 38(8), 907-926.

C. de Boor, 2001, *A Practical Guide to Splines.* Springer-Verlag, New York, USA.

Bica A M, 2012, Fitting Data Using Optimal Hermite Type Cubic Interpolating Splines. *Applied Mathematics Letters*, doi: 10.1016/j.aml.2012.04.016.

Su Buqing, Liu Dingyuan, 1980, *Computation Geometry.* Shanghai Scientific and Technical Publishers, Shanghai, China.