International Association for Mathematical Geology
XI$^{th}$ International Congress
Université de Liège - Belgium

# GSLIB-type program to enhance possible anisotropies in variogram kernel based estimation

**C. Hervada-Sala[1], E. Jarauta-Bragulat[2], P. Buenestado-Caballero[2], M. El-Ghaidouni[2]**
*[1]Dept Physics and Nuclear Engineering, Universitat Politècnica de Catalunya (UPC). Terrassa, Spain.*
*[2]Dept Applied Mathematics III, Universitat Politècnica de Catalunya (UPC). Barcelona, Spain.*
*Corresponding author:* carme.hervada@upc.edu

ABSTRACT
Estimation of the variogram is an essential tool in standard geostatistics. Estimation techniques range from empirical non-parametric to model-based parametric methods. Kernel based techniques, being essentially non-parametric, represent an interesting compromise to obtain smooth, reliable and flexible variogram estimates. Typical challenges are the detection of anisotropies in vector fields and the admissibility of the estimated variogram. A GSLIB-type program has been developed in order to implement the kernel-based approach by Egozcue, El-Ghaidouni, Pawlowsky-Glahn (2002). The approach estimates the cross-variogram as a weighted linear combination of multidimensional, admissible, interpolating variograms fitted to each point of the variogram cloud. The corresponding weights are determined by radial and angular kernels that take into account the influence of each element in the cloud. Relevant characteristics of such estimator are: a) it is sensible to anisotropies; b) local-admissibility of the estimator is readily checked. Some examples of 2D-regionalized 2-component-vector field show the performance of the developed program.

*KEYWORDS: Variogram estimation, kernel, anisotropy.*

## 1. Introduction

One of the most essential and important stages in a geostatistical study is variogram modelling from the experimental variogram values. Several estimation methods have been proposed and are available in Deutsch and Journel (1998). A classification of estimation methods may be into parametric and non-parametric estimators.

The parametric models present the difficulty that they assume a limited family of variogram, whose adequacy in poorly testable. The non-parametric estimators often appear in applications although they are affected by several problems: dependence of spatial sampling and sample size. The question about admissibility of the final estimated variogram is an important one. Normally, the estimation procedures follow two steps: One, a non-parametric estimation, and two, an admissible parametric model of variogram are fitted to the previous non-parametric results. These points make the estimation of the variogram a sensible item in geostatistics.

The present aim is to introduce a non-parametric estimator of a variogram of a 2D regionalized variable. It is based on Egozcue et al. (2002). The main idea is that a point in the variogram cloud should influence the estimation of the variogram at a different lag. This influence should be modelled in two ways: (a) projection (or prediction) of cloud point value (which is actually a matrix) in the lag-radial direction, and (b) a weighting that takes into account how large is the lag-distance between the cloud point and the estimation point. Goal (a) is achieved by using the so-called projection variogram as a predictor of a value at different radial-lags. The weighting (b) is defined by using a radial kernel and angular one.

The estimator, at a given lag, is obtained by a linear combination of projection variograms whose weights are determined by radial and angular kernels. The combination changes from lag to lag and so the estimator is not a global linear combination of projection variograms. The standard non-parametric estimator of the variogram may be viewed as a special case of this general approach.

## 2. Kernel-based approach

Let us consider a 2D-space where a random row vector of $n_v$ components is observed al each sampling point $x_j$; and let Z( $x_j$ ) be the observed vectors at the sampling points. From them, we obtain the (cross) variogram cloud which values are the ( $n_v \times n_v$ ) matrices

$$V(h_i) = \left[Z(x_j) - Z(x_k)\right]\left[Z(x_j) - Z(x_k)\right] \tag{1}$$

where $h_i = x_j - x_k$ denotes the lag and the prime stands for transposition. Assume n points in the cloud. We define a 2D-kernel as a non-negative function which maximum value is attained at the origin. Two dimensional probability densities, centred at their mode, are acceptable models. Particularity, a factorization in polar coordinates seems to be a convenient approach and the kernel is denoted as $K_\theta(\theta) \times K_r(r)$ factored into the angular and the radial kernels. When estimating the variogram at a lag $h = h(\cos\theta, \sin\theta)$, a weight is assigned to each point of the variogram cloud by $K_\theta(\theta_i - \theta) \times K_r(h_i - h)$.

Define also a projection (cross)variogram $\Gamma_i(h)$ associated with the ith-point in the variogram cloud such that (a) it is an admissible variogram (i.e. non-positive definite); (b) it satisfies $\Gamma_i(h_i) = V(h_i)$ and (c) it is isotropic. Condition (a) is adequate to assure admissibility of the estimator; (b) is the condition for a proper projection and (c) is due to convenience. A natural definition of the projection variogram associated with the ith-point in the variogram cloud is

$$\Gamma_i(h) = \frac{\left[Z(x_j) - Z(x_k)\right]'}{\left\|Z(x_j) - Z(x_k)\right\|} \gamma_{0i}(h) \frac{\left[Z(x_j) - Z(x_k)\right]}{\left\|Z(x_j) - Z(x_k)\right\|} \tag{2}$$

where $\gamma_{0i}(h_i) = \left\|Z(x_j) - Z(x_k)\right\|^2$, thus fulfilling the condition (b) mentioned above. Moreover, we assume $\gamma_{0i}(h)$ does not depend on the angle $\theta$ of the lag and condition (c) is thus satisfied. Consequently, also $\Gamma_i(h)$ only depends on the absolute value of the lag. Projection variograms defined in (2) are admissible whenever $\gamma_{0i}(h)$ is an admissible, univariate variogram and therefore some well-known models are available. We define the variogram estimator

$$\hat{\Gamma}(h) = \frac{1}{W(h)} \sum_{i=1}^{n} K_\theta(\theta_i - \theta) K_r(h_i - h) \Gamma_i(h), \qquad W(h) = \sum_{i=1}^{n} K_\theta(\theta_i - \theta) K_r(h_i - h) \tag{3}$$

Admissibility of this kind of estimator has not been yet completely studied. But we guess that, for a regular distribution of sampling points $x_j$, and positive definite kernels, the estimator (3) is admissible. Among the possible choices of projection variograms, radial and angular kernel functions, we have selected an exponential projection variogram, a trigonometric rational angular kernel and a normal radial kernel. The projection variogram is fully defined when $\gamma_{0i}(h)$ is given. We use the following two parameter exponential expression

$$\gamma_{0i}(h) = c_i(1 - \exp(-\lambda_i h)), \quad c_i > 0, \quad \lambda_i > 0 \tag{4}$$

Parameter $c_i$, which is actually the projection sill, deserves special attention. It can be estimated directly from the data, e. g. by selecting an a priori range, $R$ and, then, identifying $c_i$ with the mean or median value of $\gamma_{0i}(h_k)$, conditional to $h_k > R$ and $\gamma_{0k}(h_k) > \gamma_{0i}(h_i)$.

Another appealing possibility, is to randomly choose $c_i = \gamma_{0k}(h_k)$ from the sample distribution of the cloud, conditional on $\gamma_{0k}(h_k) > \gamma_{0i}(h_i)$ and $h_k > R$. The expressions of the radial and angular kernel functions selected are the following:

$$K_r(h_i - h) = \frac{1}{\sqrt{2\pi}\alpha} \exp\left[-\frac{(h_i - h)^2}{2\alpha^2}\right], \quad K_\theta(\theta_i - \theta) = \frac{1 - a^2}{\pi(1 - 2a \quad \cos(2(\theta_i - \theta)) + a^2)} \qquad (5)$$

where $\alpha$ and $a$ are bandwidth parameters.

## 3. GSLIB-type program

A FORTRAN program has been elaborated to estimate variograms from experimental data. The method used for variogram estimation is kernel based approach, developed by Egozcue, El-Ghaidouni, Pawlowsky-Glahn (2002). The program is adapted to GSLIB package, so that everyone used to his format can use it easily, this has also the advantage that data have to be in Geo-EAS format, which is widely used. The structure of the file is well-known and identical to all GSLIB files, so user has only to specify the parameters in each case.
This is, the design of the program has the following structure:

| | |
|---|---|
| INPUT: | Data file in Geo-EAS format |
| | Parameter file |
| FUNCTIONS: | Variogram calculation |
| | Kernel estimation |
| OUTPUT: | Output file |

Parameter file made by the user needs the following information:
a) file with data
b) columns for X, Y coordinates
c) number of variables, column numbers
d) trimming limits
e) file for variogram output
f) number of lags
g) lag separation distance
h) lag tolerance
i) number of directions
j) azimuth, angular tolerance, bandwidth (for each direction)
k) estimated range
l) angular parameter
m) scale parameter
As usual, output is an ASCII file; for graphic output, other programs must be used then.

### 4. Discussion

### 4.1 Anisotropies

Since it has been said before, one of the greater utilities of this program is to obtain results that allow considering possible anisotropies in a bidimensional data set. Results, suitably processed, allow visualizing possible anisotropies, as it can be seen in Figure 1.

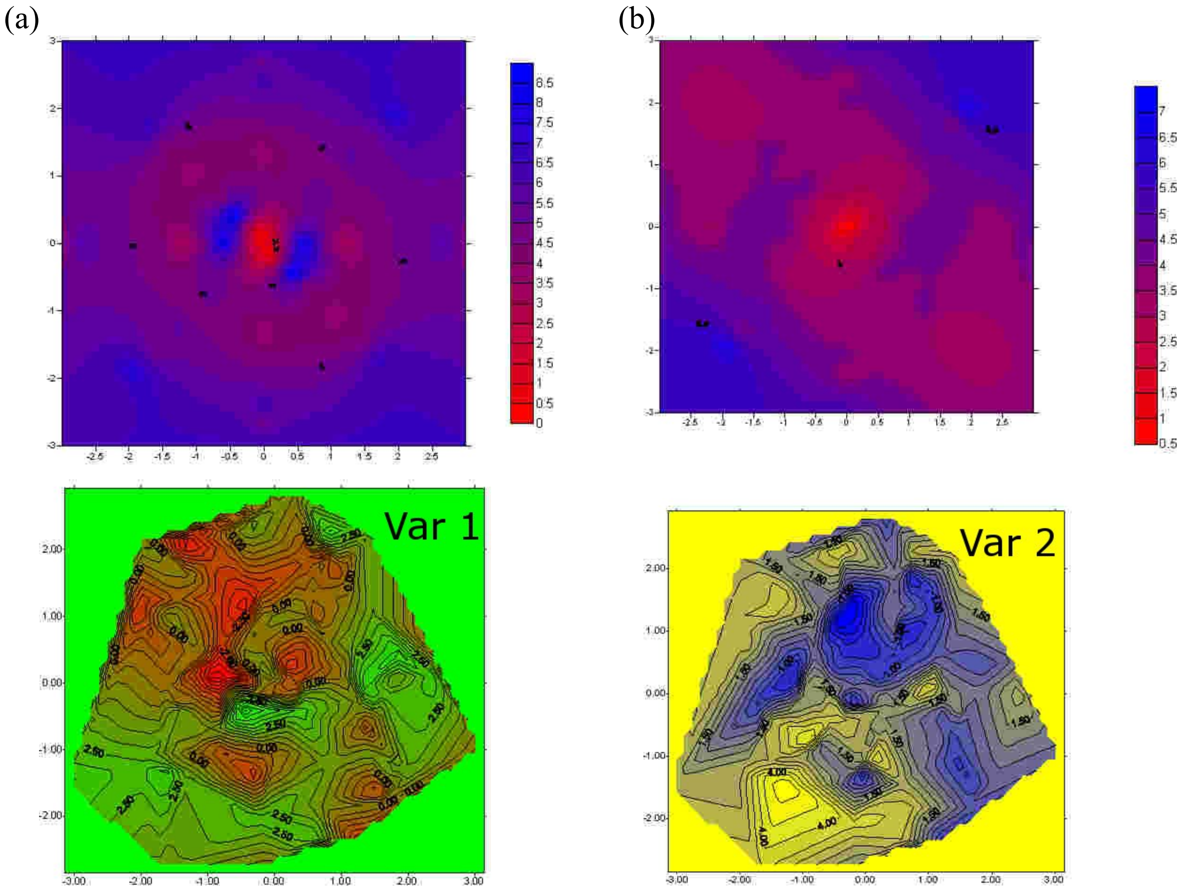


Figure 1. (a) Variogram estimation and data for variable z1. (b) The same for variable z2

**4.2 Local admissibility**

The program we present evaluates local admissibility. It contains a subroutine which estimates Fourier transform coefficients. Such coefficients provide information on whether the bandwidth parameter satisfies the admissibility condition or not.

**Conclusions**

1) This FORTRAN GSLIB-type program allows estimating variograms by means of a kernel based approach, for bidimensional variables.
2) Results provided by this program allow for an automatic variogram estimation
3) On top of that, this program detects anisotropies and assures local admissibility.

*Acknowledgements. Authors wish to thank ETSECCPB, EUETIB and DMA3 sponsorship.*

REFERENCES

Egozcue, J.J., El-Ghaidouni, M., Pawlowsky-Glahn, V., 2002. A kernel type estimator of the 2D-multivariate variogram. *Terra Nostra*, 57 – 62.

Deutsch, C.V., and Journel, A.G., 1998. GSLIB. Geostatistical Software Library and User's Guide. Oxford U. Press, New York (USA), 369 p.