EVALUATING THE ESTIMATION OF REGULARIZED COVARIOGRAMS

Luis Paulo Vieira BRAGA[1]
Cassio ALMEIDA[2]
Claudio BETTINI[3]

- *ABSTRACT: In this work, we present the effeciency of the approximation of a regularized variable´s covariogram over a large panel (2D) by the covariogram of a regularization via accumulation of order 2 (montée) of a three-dimensional regionalized variable. The regularized covariogram can be obtained through a numerical approximation of the inverse Fourier transform of the* G_1 *Fourier transform of the regularized covariogram by accumulation of order two of the fitted point covariogram. Such accumulation is obtained through the* G_3 *Fourier transform of the point covariogram by equating two of its coordinates to zero. Two data sets were approached, the first is simulated according to a Matérn law and the second one is based on observed temperature values recorded along extensive time periods, over a wide region located in the Brazilian Amazon region. In this area, we considered the mean values of the daily temperatures in January, along nine decades, namely, from 1901-1910 to 1981-1990. Cross validation interpolations as well as simulations were performed, which were compared with the true values observed in each decade, thus checking the accuracy of the approximation and the range of the temperature variation .*

- *KEYWORDS: Change of support; regularization; non-separable space-time models; geostatistics; global warming.*

1 Introduction

For the following developments (Matheron, 1965), we are going to consider covariograms instead of semivariograms. The covariogram is defined for a square integrable function by the expression: $g(h) = \int f(x) f(x+h) dx$ or through convolution, where $f^v(x) = f(-x)$:

$$g(h) = f * f^v(h) \tag{1}$$

[1]Instituto de Matemática, Universidade Federal do Rio de Janeiro - UFRJ, CEP 21945-970, Rio de Janeiro, RJ, Brasil, E-mail: *lpbraga@im.ufrj.br*
[2]Escola Nacional de Ciências Estatísticas, Instituto Brasileiro de Geografia e Estatística - IBGE, CEP 20231-050, Rio de Janeiro, RJ, Brasil, E-MAIL: *cassiofreitas@ibge.gov.br*
[3] Instituto de Geociências, Universidade Federal do Rio de Janeiro – UFRJ, CEP 21949-900, Rio de Janeiro, RJ, Brasil, E-mail: *bettini@geologia.ufrj.br*

From another definition, the regularization of a function f by a weighting function p is given by:

$$\int f(x+h)\,p(h)dh \tag{2}$$

If $p(h)$ is the indicator function, then the functions mean grade (t) and quantity of metal (q) can be considered as regularizations of function f:

$$t(x) = \frac{1}{p}\int f(x+h)\,p(h)dh \tag{3}$$

$$p = \int p(h)dh \tag{4}$$

$$q(x) = \int f(x+h)\,p(h)dh \tag{5}$$

Using the convolution notation, t and q can be represented as:

$$t(x) = \frac{1}{p} f * p^{v} \tag{6}$$

$$q(x) = f * p^{v} \tag{7}$$

The covariogram of q, the regularization of f by p, can also be represented by:

$$g_q = (f * p^{v}) * (f * p^{v})^{v} = (f * f^{v}) * (p * p^{v}) = g * P \tag{8}$$

Through analogous procedure, the covariogram of t is:

$$g_t = \frac{1}{p^2} g * P \tag{9}$$

If desired, this result can be expressed in terms of integrals, assuming that p is an indicator function defined on a volume V:

$$g_t(h) = \frac{1}{V^2}\int_V \int_V g(h+x-y)dxdy \tag{10}$$

The above expression suggests a form of regularization known as accumulation (montée), in which, through integration, dimension is reduced by one and V is considered unlimited. Denoting fn(x1, x2, ...xn) to be the value of the n-dimensional variable, the regularization over an unlimited domain (montée) is given by:

$$f_{n-1}(x_1,...,x_{n-1}) = \int f_n(x_1...x_{n-1}, x_n)dx_n \tag{11}$$

Graphically, Eq. (11) represents the quantity of contained metal by the straight lines parallel to the x_n axis, (regularization by accumulation of order 1). When another direction is chosen, it is sufficient to perform an adequate rotation. Proceeding this way, one obtains

the quantity of contained metal by the planes parallel to the axes x_{n-1} and x_n (regularization by accumulation of order 2), and so forth, see (12).

$$f_1(x_3) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_3(x_1 x_2 x_3)\, d x_1\, d x_2 \tag{12}$$

If the value of the variable f is a grade, then the regularization by accumulation has a quite concrete meaning.

The process of regularization by accumulation can be interpreted with the aid of the Fourier transform G_n of the covariogram g_n of the function f_n . The Fourier transform of the function f_3 is given by (13)

$$F_3(u_1,u_2,u_3) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_3(x_1,x_2,x_3) e^{-2i\pi(u_1x_1+u_2x_2+u_3x_3)} dx_1 dx_2 dx_3 \tag{13}$$

We can easily see that

$$F_3(0,0,u_3) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_3(x_1,x_2,x_3) e^{-2i\pi(0x_1+0x_2+u_3x_3)} dx_1 dx_2 dx_3 = \int_{-\infty}^{\infty}\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_3(x_1,x_2,x_3) dx_1 dx_2\} e^{-2i\pi u_3 x_3} dx_3 =$$

$$= \int_{-\infty}^{\infty} f_1(x_3) e^{-2i\pi u_3 x_3} dx_3 = F_1(u_3) \tag{14}$$

The three dimensional non centered covariogram, according to (1) is

$$g_3(h_1,h_2,h_3) = \int_{-\infty}^{\infty} f_3(x_1,x_2,x_3)\, f_3(h_1-x_1, h_2-x_2, h_3-x_3)\, dx_1 dx_2 dx_3 \tag{15}$$

Let $G_3(u_1, u_2, u_3)$ represent the Fourier transform of the covariogram, then (14) implies that:

$$G_1(u_3) = G_3(0,0,u_3) \tag{16}$$

The relationship between the transforms for each order of accumulation is given by:

$$G_{n-k}(u_1,u_2,...u_{n-k}) = G_n(u_1,u_2,...u_{n-k},0...0) \tag{17}$$

Therefore the result holds for isotropic covariograms. While using an appropriate transform, one could work with only one dimension, where h would represent the norm of the vector whose components are ui. But this approach will not be adopted, while the rectangular coordinate notation will be kept. Eq. (17) allows an interesting application to

the calculation of the covariogram of an accumulation of order k from the covariogram gn, which is possible by applying the inverse transform on G_{n-k} to obtain g_{n-k} . The inverse Fourier transform of a three-dimensional function h is given by (18).

$$F_3^{-1}(h(u_1,u_2,u_3)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(u_1,u_2,u_3)\, e^{i2x(u_1x_1+u_2x_2+u_3x_3)}\, du_1 du_2 du_3 \qquad (18)$$

If the covariogram g_3 is isotropic then the covariogram g_1 may be derived from sucessive applications of the Fourier transform. Since generation of closed forms for both the transform and its inverse is restricted to particular cases, the use of numerical methods becomes necessary. Therefore, let us apply the fast versions for both the multidimensional Fourier transform and its inverse versions to calculate g_{n-k} . The fast versions for both the transform and its inverse only take advantage of the calculation order, to avoid repetition of operations, thus reducing the calculation complexity from N^2 to $N \log N$.

2 Simulated data set

In particular, let us consider elements of the Matérn family of covariances (Matérn, 1986) represented by Eq. (19), where Γ is the gamma function, K_ν is the modified Bessel function of third kind and order *k, a* is the range and *x* is the lag:

$$g(x) = [2^{\nu-1}\Gamma(\nu)]^{-1}(x/a)^{\nu} K_\nu(x/a) \qquad (19)$$

For the following discussion we are going to consider semivariograms because of their practical appeal. The purpose of the implementation is to evaluate the proposed methodology by checking the calculation of the regularized semivariogram and using it to predict the mean value of a panel .

The calculation of the regularized semivariogram was checked by comparing the experimental semivariogram of an unlimited panel sample and the numerical regularized semivariogram, as it was explained in section 1.

The panel values were obtained by averaging simulated values in 3D space, but belonging to a single panel. Panels were placed at every 500 length units along an orthogonal axis. See Fig. (1). The values were simulated according to the turning band method with Matérn semivariogram parameters as in Table 1. See Fig. (2).

The experimental semivariogram was obtained from the sample values, while the numerical regularized semivariogram was generated from Eq. (16). Graphical representations of both curves are shown in Fig. (3) outlining a good fit between them.

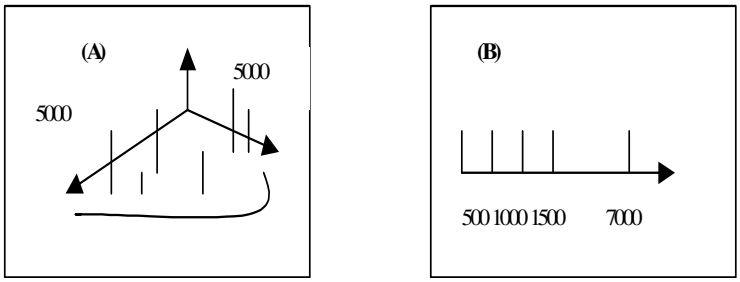

Figure 1 - A. Values per panel. B. Orthogonal axis.

Table 1 - Data generation

Panel configuration	Semivariogram Configuration
X: -5000 a 5000	ν: 0.5
Y: -5000 a 5000	Mean: 0
Z: from 0 to 7000 by 500	Variance: 1
Z: from 0 to 7000 by 500	Variance: 1
Points per panel: 7000	*a*: 10000

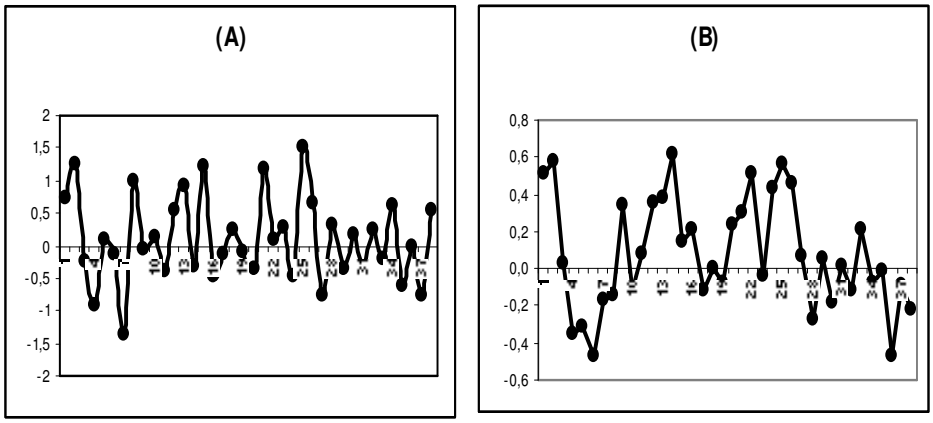

Figure 2 - A. Simulated panel values along the ortogonal axis. B. Smoothed panel values along the orthogonal axis, for kriging purposes only.

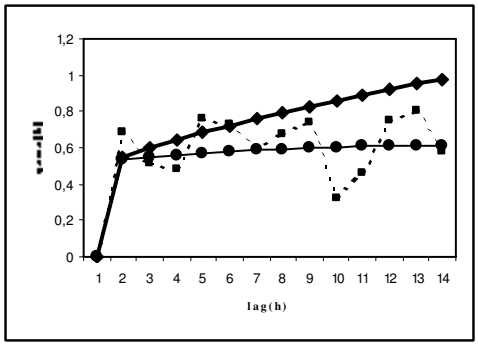

Figure 3 - The thinnest continuous line is the graph of the numerical approximation of the regularized semivariogram of the panel values. The dotted line is the graph of the experimental semivariogram of the simulated panel values. The thickest continuous line is the graph of the point semivariogram, for visualization purposes only.

A regularized semivariogram model was not uncovered, but numerical approximations were stored for future calculation. An ordinary kriging was performed in order to compare the observed simulated values and the predicted ones. A moving neighbourhood along the orthogonal axis, including eight values around a central point (four at each side), provided the data for the kriging system. Two different semivariograms were considered – the point semivariogram (posv) and the regularized semivariogram (rsv). An equal weight (0.125) moving average (ewma) prediction was also calculated. The results of the (rsv) are shown in Figure 4 and Table 2.

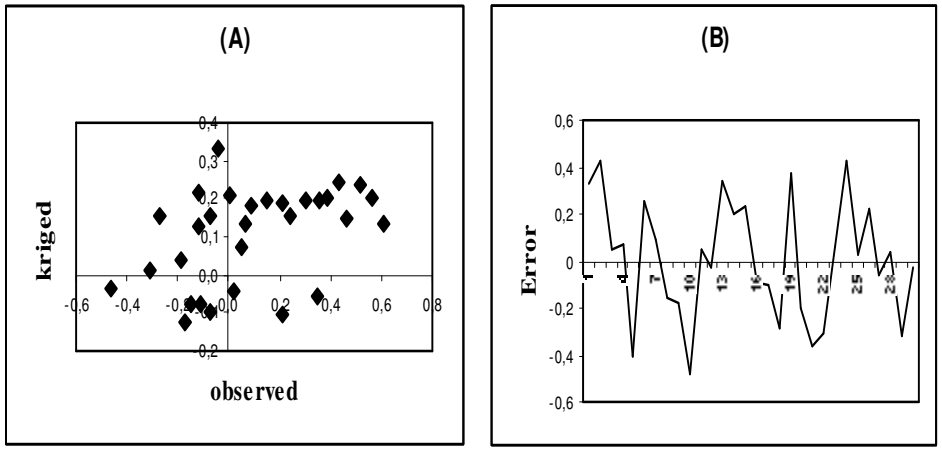

Figure 4 - A. Kriged values (rsv) versus Observed simulated values scattergram. B. Relative error dot plot (rsv).

Table 2 - Summary of several statistics related to the prediction errors

	ewma	rsv	posv
Standard deviation of errors:	0.258926	0.253283	
Mean of errors:	0.006540	0.005969	0.004425
Median of errors	0.028788	0.029560	0.010250
Correlation between observed and predicted values:	0.381423	0.423569	0.503200

The case study exemplifies the use of the proposed methodology, which enables a straightforward sequence of steps to predict or to simulate (by way of the rsv) regularized variables by accumulation. Clearly the smoothing effect of the kriging predictor is also observed and justifies the advantage of the point kriging (posv) if the purpose is interpolation. But, if the goal is to reproduce the fluctuation of the mean value of the panels, then (rsv) is recommended. All codes were developed in the R environment (R Development Core Team, 2005) as well as its implementations of the turning band method, the fast Fourier transform, the linear system solving, and the sample semivariogram generation. The results presented may be easily reproduced since the parameters are kept the same and the simulator seed is reinitiated.

3 Amazon rain forest data set

Consider a set of meteorological variables observed within a time interval, within a geographic region delimited by a set of meridians and parallels. Using the UTM projection, such region may be represented as a panel in three-dimensional space, in which the first two coordinates record the position, while the third one represents the time. By taking time series of values of a given variable of interest in monitoring stations, one can fit a point three-dimensional semivariogram, from which it is possible to simulate values in different locations of the region, as well as in different times. Moreover, it is possible to obtain a regularized semivariogram of the whole region, thus allowing the mean value interpolation or simulation for the entire region at any time.

Based on data released by the Data Distribution Center of the Intergovernmental Panel on Climate Change, we defined a region limited by meridians 54^oW and 48^oW, and by parallels 8^oS and 0^o, located in Brazilian Amazon. Within this region, we considered average temperature data in January, across nine decades, located in a network of $0{,}5^o$ lat by $0{,}5^o$ long cells. The articles by (New, Hulme & Jones 1999, 2000) explain how such data were obtained. The used decades span the periods from 1901-1910 to 1981-1990. The maps in Figure 5 represent the average temperature variations in the region by decade. The lighter the color, the higher the temperature. However representing smoothed values, the maps indicates a trend in the variation of temperature.

A lognormal transform, namely, $t' = log(274-t)$, was applied to the raw temperature data, multiplied by 100, before the calculation of the experimental semivariogram. Thirty six experimental semivariograms were obtained, corresponding to four directions represented by the azimuths 0^o,45^o,90^o and 135^o, for each of the nine decades.

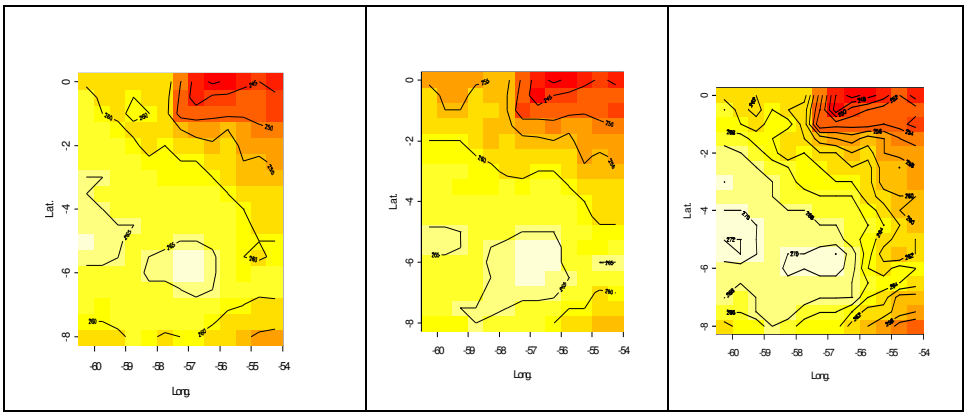

Figure 5 - Isotherms, 20, 50, and 80 decades.

A space-time model was fitted to the transformed data, considering an anisotropy correction in which the N-S and E-W axes plus a time axis were treated as the main axes of the ellipsoid defined by the relationship 0.015 to 0.020 to 0.01, respectively. The model (20) belongs to the family of non separable space-time covariances proposed by (Gneiting, 2001).

$$C(x,t) = \frac{\sigma^2 x}{[(t^{k_1}+1)^{k_2}]^4 [2^{\nu-1}\Gamma(\nu)]} K_\nu \left(\frac{x}{(t^{k_1}+1)^{k_2}} \right) \qquad (20)$$

In such model, K_ν is the modified Bessel function of third kind and order 3 and σ^2, ν, k_1, k_2 are the parameters to be estimated. The calculations were performed with the aid of the *Random Fields* package of the computacional system *R* (R Development Core Team, 2005). Figure 6 shows the results for three decades in East_West direction, where an increasing oscillation with time is observed, as well as in other directions. The fitted values of the parameters were the following: σ^2=0.1315876; ν=9.335935; k_1=1 e k_2=0.5.

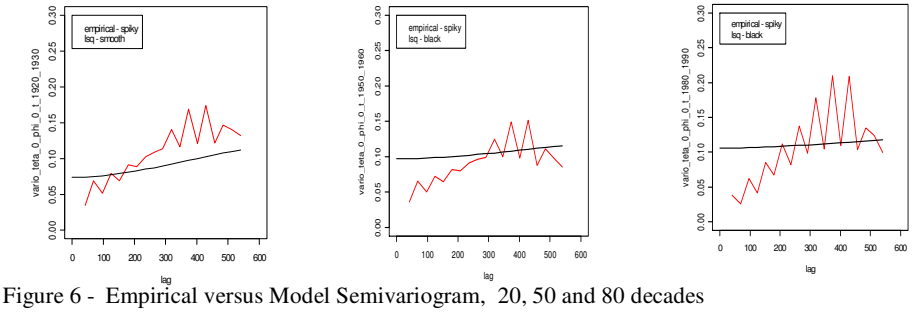

Figure 6 - Empirical versus Model Semivariogram, 20, 50 and 80 decades

According to the introduction, the regularized covariogram can be obtained through a numerical approximation of the inverse Fourier transform of the G1 Fourier transform of the regularized covariogram by accumulation of order two of the fitted point covariogram. Such accumulation is obtained through the G3 Fourier transform of the point covariogram by equating two of its coordinates to zero. For the fitted model (20), the numerical approximation of the regularized covariogram is shown in Table 3 and plotted in Figure 7. The values express the temporal dependence among the transformed average temperatures by decade.

Table 3 - Regularized Covariogram

Lag	0	100	200	300	400
Reg.Covar.	0.003570	0.003279	0.002543	0.001661	0.000897
Lag	500	600	700	800	900
Reg.Covar.	0.000373	0.000078	0.000000	0.000000	0.000000

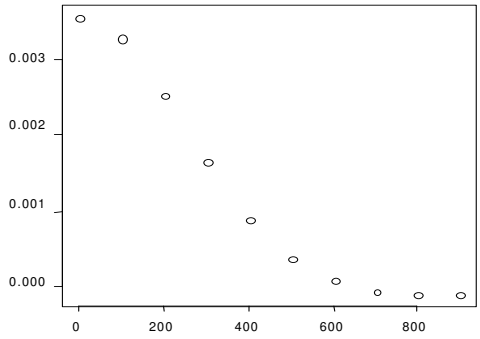

Figure 7 - Regularized Covariogram Plot

Applying a window over the sequence of transformed mean values of each panel by decade, we performed an ordinary kriging interpolation for the decades from 20 to 60, considering a moving neighborhood with two previous and two posterior observations. The kriging matrix was based on the values presented in Table 3. The errors between interpolated and observed values are summarized in Table 4, showing that the regularized covariogram produce a good interpolation.

Table 4 - Cross Validation Errors

Min	1s Qu	Median	Mean	3rd Qu	Max
-0.243300	0.006051	0.028590	0.066750	0.073590	0.468800

A simulation method based on the Cholesky decomposition (Olea,1991) was adopted, whose basic equation is given by equation (21), where Y2 is the vector with S2 simulations; L12 corresponds to the covariances between conditional and simulated decades; L11 corresponds to the covariances among the the conditioning decades; Z1 is

the vector with S1 conditioning decades; L22 corresponds to the covariances among the simulated decades and N_2 to a vector with S_2 random numbers with zero mean and unit

$$Y_2 = L_{12}^T L_{11}^{-1} Z_1 + L_{22} N_2 \quad (21)$$

standard deviation.

The decade 50 was chosen as an example for simulation. Table 5 shows the observed mean values by decade. Figure 8 illustrates the results of some realizations of such simulation.

Table 5 - Mean temperature by decade

Decade	0	10	20	30	40	50	60	70	80
Celsius	26.24	25.99	26.06	25.99	25.87	26.04	26.52	25.98	26.21

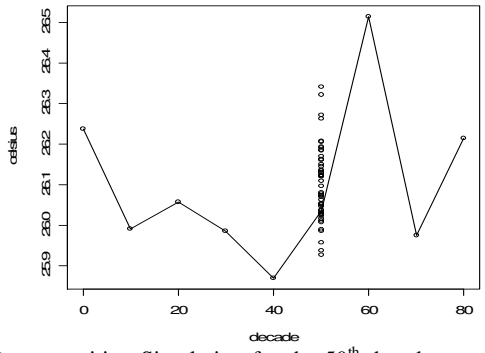

Figure 8 - Cholesky Decomposition Simulation for the 50th decade: continuous line – actual data; dots – simulated data.

A gradual increase in average temperatures can be observed mainly in the last three of the considered decades. The unstability of covariogram in these decades also was observed in all directions. Finally the time dependence was apparent at most between close decades. As in the previous data set all codes were developed in the R environment (R Development Core Team, 2005).

Conclusions

Predicting the realization of a regionalized variable with a given support, based on its realizations with different supports, is a usual problem in geosciences. In mining, petroleum field development and environmental studies, the measurements of an attribute of interest can be obtained through distinct techniques, thus generating data with distinct supports. It is common practice to treat such data sets as related with point support.

In the first study case , we showed the implications for the prediction error when the change of support is ignored. Then we obtained the predictions taking the different supports into account. For this purpose, the regularized covariograms were inferred according to an adaptation of the method proposed by (Matheron, 1965) who originally

used the closed forms of the direct and inverse Fourier transforms, which are not applicable to more complex situations like the Matérn model.

The second study case included exhaustive testing in order to understand better the performance of the method and its application in a real problem, assuming that its features are close to the considered hypotheses.

Both case studies exemplified the use of the proposed methodology, which enables a straightforward sequence of steps to predict or to simulate regularized variables by accumulation.

The results presented in this paper are preliminary, as many aspects are under research, namely: the studies of zonal anisotropy and temporal drift; the inference of space-time covariance parameters with alternative minimization criteria; improvement of the back-transformation of lognormal kriging and the study of other observed atributes such as the maximum and minimum temperatures.

BRAGA, L. P. V.; ALMEIDA, C.; BETTINI, C. Avaliando a estimação do covariograma regularizado. *Rev. Mat. Estat.,* São Paulo, v.24, n.3, p.149-160, 2006.

- RESUMO: Neste trabalho apresentamos a eficácia da aproximação do covariograma regularizado sobre um amplo painel (2D) pelo covariograma de uma regularização por acumulação de ordem 2 de uma variável regionalizada a três dimensões. O covariograma regularizado pode ser obtido por uma aproximação numérica da transformada inversa de Fourier da transformada de Fourier G1 do covariograma por acumulação de ordem 2 do modelo pontual ajustado. Esse covariograma é obtido atavés da transformada de Fourier G3 do covariograma pontual, igualando-se duas de suas coordenadas a zero. Dois conjuntos de dados foram testados, o primeiro é simulado a partir da lei de Matérn e o segundo é baseado em temperaturas observadas ao longo de extensos períodos de tempo sobre uma vasta área da floresta amazônica brasileira. Nessa área, consideramos valores médios de temperaturas diárias durante o mês de Janeiro, ao longo de nove décadas, de 1901-1910 a 1981-1990.Foram realizados testes de interpolação assim como simulações, que foram comparados aos valores observados, verificando-se assim a precisão das aproximações e seu intervalo de variação.

- PALAVRAS-CHAVE: Mudança de suporte; regularização; modelos espaço-temporais não separáveis; geoestatística; aquecimento global.

References

GNEITING, T. Nonseparable, stationary covariance functions for space-time data. *J. Am. Stat. Assoc.,* New York, v.97, p.590-600, 2001.

MATÉRN, B. *Spatial variation*. 2nd ed. New York: Springer Verlag. 1986. 149p.

MATHERON, G. *Les variables régionalisées et leur estimation*. Paris: Masson et C^{IE}, 1965. 305p.

NEW, M.; HULME, M.; JONES, P. Representing twentieth-century space-time climate variability. Part I: Development of a 1961-90 mean monthly terrestrial climatology. *J. Climate*, Boston, v.12, p.829-856, 1999.

NEW, M.; HULME, M.; JONES, P. Representing twentieth-century space-time climate variability. Part II: Development of a 1901-96 monthly grids of terrestrial surface climate. *J. Climate*, Boston, v.13, p.2217-2238, 2000.

OLEA, R. A. *Geostastistical glossary and multilingual dictionary*. New York: Oxford University Press. 1991. 177p.

R DEVELOPMENT CORE TEAM. R: A language and environment for statistical computing. R foundation for statistical computing. Disponível em <http://www.R-project.org>. Acesso em 01 jun. 2006. Viena, 2005.

Recebido em 29.07.2006.

Aprovado após revisão em 10.10.2006.