

Uso da Regressão Não Linear para Inferência de Variogramas em Suportes Amostrais Distintos Um estudo para a família exponencial

L.Braga (1), C.Almeida (2), C.Bettini (3)

(1) UFRJ-Instituto de Matemática, (2) IBGE - ENCE, (3) UFRJ-Instituto de Geociências

Cassio Freitas Pereira de Almeida - cassiofreitas@ibge.gov.br

Luis Paulo Vieira Braga, C.P. 68530, 21945-970 Rio de Janeiro, Brasil, braga@dme.ufrj.br

Abstract: The change of support has been thoroughly discussed within the geostatistical community, being considered one of its most relevant problems, owing to the implications in petroleum and ore reserve estimation. Among the precursor and favorite methods for dealing with this problem are both the Affine Correction, applicable to normally distributed populations, and the Generalized Lognormal Permanence, for other continuous populations. Afterwards non-linear estimation methods were developed such as the Disjunctive Kriging of Isofactorial Models, valid for discrete and continuous distributions, and the Estimation of Local Conditional Distribution through Indicators. This work turned back to the precursor approaches to the change of support problem, which had attempted to establish regression models between point and regularized variables. However, instead of searching for a regression between variables, a non-linear dependence was found between the support size and the parameters of the semivariogram. We successfully found a good adjustment for the spherical model (ALMEIDA, BRAGA & BETTINI 2001). In this paper we propose a regression model between the parameters of the exponential family derived from the Matérn model and the support dimensions:

$$\gamma(h) = \{1 - \exp[-3(h/a)^\nu]\} \quad (1)$$

The scaling factor “3” is chosen so that the semivariogram at $\nu = 0$ is $1 - e^{-3} \cong 0.05$. When $\nu=2$ the semivariogram is called the Gaussian semivariogram model. Note that $\nu=0$ corresponds to a white noise process. These semivariograms gradually change shape for different values of ν , producing sigmoidal or “S-shaped” curves for ν near 2. Besides that, we know that the change of support also induces an “S-shaped” semivariogram. Numerous mathematical functions have been proposed for modeling sigmoidal curves. We opted for a model derived from the Weibull family, $Y = \alpha - \beta \exp(-\gamma X^\delta)$, (RATKOWSKY 1983). In this paper we restricted ourselves to an unidimensional support l and to arguments $h \gg l$. The calculation of the regression parameters was performed through the function NLM of the freeware package R, (VENABLES & SMITH 2002). The mean and the regularized semivariograms were obtained through numerical integration and through regression. Finally some conclusions and future research steps are presented.

1. Introdução

A fim de compor o modelo de regressão, assim como a representação do variograma regularizado, várias etapas precisam ser cumpridas, a saber: (1) representação do variograma regularizado em termos do variograma pontual e do variograma médio; (2) cálculo do variograma médio por integração numérica; (3) ajuste de um modelo da família Weibull aos dados numéricos, obtendo-se um modelo predito para o variograma médio; (4) análise de erros; (5) representação final do variograma regularizado e (6) Comparação entre os variogramas regularizados obtidos por integração e predição.

A expressão do variograma regularizado em função do variograma médio é dado pela fórmula

$$\gamma_l(h) = \bar{\gamma}(l, l_h) - \bar{\gamma}(l, l) \quad (2)$$

É a regularização que provoca a mudança de forma do semivariograma. Quando h é muito maior que l o primeiro termo do segundo membro pode ser aproximado por $\gamma(h)$ ficando o variograma regularizado expresso por (3).

$$\gamma_l(h) = \gamma(h) - \bar{\gamma}(l, l) \quad (3)$$

O problema, portanto, reside no cálculo do variograma médio quando ele percorre o suporte l .

Devido a se considerar a distância como argumento para o variograma, é preciso ter controle sobre as diferenças

positivas e negativas quando os limites de integração varrem o suporte, assim a fórmula para o cálculo das integrais é dada por (4).

$$\bar{\gamma}(l, l) = \frac{1}{l^2} \left\{ \int_0^l du \left[\int_0^u \gamma(u-u') du' + \int_u^l \gamma(u'-u) du' \right] \right\} \quad (4)$$

A próxima etapa é a resolução da integral, neste caso (1) não tem primitiva analítica, portanto será necessário utilizar uma aproximação numérica. Aplicaremos o método da quadratura gaussiana nos dois níveis de integração. Denominando as duas integrais internas como F(u) e G(u) respectivamente, temos:

$$\bar{\gamma}(l, l) = \frac{1}{l^2} \left\{ \int_0^l du [F(u) + G(u)] \right\} \quad (5)$$

A fim de preparar a integral externa para a aplicação do método da quadratura gaussiana é necessária uma transformação de variáveis de modo que seu intervalo fique entre -1 e 1.

$$\begin{aligned} \bar{\gamma}(l, l) &= \frac{1}{l^2} \left\{ \int_{-1}^1 \left[F\left(l + \frac{tl-l}{2}\right)(l/2) + G\left(l + \frac{tl-l}{2}\right)(l/2) \right] dt \right\} \\ &= \frac{1}{l^2} \frac{l}{2} \left\{ \int_{-1}^1 \left[F\left(l + \frac{tl-l}{2}\right) + G\left(l + \frac{tl-l}{2}\right) \right] dt \right\} \end{aligned} \quad (6)$$

Denominando $v = l + (tl - l)/2$ vamos aplicar a transformação de variáveis de forma que os intervalos das integrais internas também fiquem entre -1 e 1.

$$F(v) = \int_{-1}^1 \gamma\left(v + \frac{tv-v}{2}\right)(v/2) dt \quad (7)$$

$$G(v) = \int_{-1}^1 \gamma\left(\frac{t(l-v)+v+l}{2} - v\right)\left(\frac{l-v}{2}\right) dt \quad (8)$$

O método da quadratura gaussiana consiste na representação da integral por um somatório da forma :

$$\int_{-1}^1 f(x) dx = \sum_{i=1}^N H_i f(x_i) \quad (9)$$

Por exemplo, escolhendo-se N=10 os valores de x_i e $H(x_i)$ seriam dados pela tabela abaixo, aonde os valores foram

arredondados para sete dígitos a partir de (ABRAMOWITZ E STEGUN, 1970).

X_i	H_i
$\pm 0,9739065$	0,0666713
$\pm 0,8650634$	0,1494513
$\pm 0,6794096$	0,2190864
$\pm 0,4333954$	0,2692602
$\pm 0,1488743$	0,2955242

Tabela I : Valores para quadratura gaussiana

Aplicando (9) nas integrais (6), (7) e (8) obtemos :

$$\begin{aligned} \bar{\gamma}(l, l) &\cong \frac{1}{l^2} \left\{ \sum_{i=1}^{10} H_i \left[\sum_{j=1}^{10} H_j \gamma\left(v_i - \frac{t_j v_i + v_i}{2}\right)(v_i/2) + \right. \right. \\ &\quad \left. \left. + \sum_{j=1}^{10} H_j \gamma\left(\frac{t_j(l-v_i)+v_i+l}{2}\right)\left(\frac{l-v_i}{2}\right) \right] \right\} \end{aligned} \quad (10)$$

Esta fórmula será a fórmula básica para o cálculo do variograma médio.

2. Cálculo do variograma médio

Tendo em vista a utilização da regressão, interessa-nos reduzir o número de variáveis independentes, no caso, os parâmetros a e v e o suporte l . Consideremos portanto a expressão (4) do variograma médio com o argumento dividido por a .

Procedendo as seguintes mudanças de variáveis nas integrais internas:

$$\bar{\gamma}(l, l) = \frac{1}{l^2} \left\{ \int_0^l du \left[\int_0^u \gamma\left(\frac{u-u'}{a}\right) du' + \int_u^l \gamma\left(\frac{u'-u}{a}\right) du' \right] \right\} \quad (11)$$

$$v' = \frac{u-u'}{a} \Rightarrow dv' = -\frac{du'}{a} \quad (12)$$

$$v'' = \frac{u'-u}{a} \Rightarrow dv'' = \frac{du'}{a} \quad (13)$$

Substituindo (12) e (13) em (11) obtemos

$$\bar{\gamma}(l, l) = \frac{1}{l^2} \left\{ \int_0^l du \left[\int_{\frac{u}{a}}^0 \gamma(v')(-a)dv' + \int_0^{\frac{l-u}{a}} \gamma(v'')(a)dv'' \right] \right\} \quad (14)$$

Fazendo uma nova mudança de variáveis,

$$u=aw \text{ e } du=adw \quad (15)$$

obtemos:

$$\bar{\gamma}(l,l) = \frac{a^2}{l^2} \left\{ \int_0^{\frac{l}{a}} dw \left[-\int_w^0 \gamma(v') dv' + \int_0^{\frac{l}{a}-w} \gamma(v'') dv'' \right] \right\} \quad (16)$$

Por outro lado consideremos agora a expressão do variograma médio para um suporte com tamanho l/a para um variograma no qual o argumento não está dividido por a .

$$\bar{\gamma}(l/a, l/a) = \frac{a^2}{l^2} \left\{ \int_0^{\frac{l}{a}} du \left[\int_0^u \gamma(u-u') du' + \int_u^{\frac{l}{a}} \gamma(u'-u) du' \right] \right\} \quad (17)$$

Fazendo uma mudança de variáveis:

$$v' = u - u' \Rightarrow dv' = -du' \quad (18)$$

$$v'' = u' - u \Rightarrow dv'' = du' \quad (19)$$

$$\bar{\gamma}(l/a, l/a) = \frac{a^2}{l^2} \left\{ \int_0^{\frac{l}{a}} du \left[\int_u^0 \gamma(v') (-dv') + \int_0^{\frac{l}{a}-u} \gamma(v'') dv'' \right] \right\} \quad (20)$$

Obviamente (20) \equiv (16), portanto provamos que o variograma médio para um suporte l com argumento h/a pode ser equivalentemente calculado como o variograma médio para um suporte l/a com argumento h . Para a avaliação do variograma médio no caso exponencial vai se considerar diversos valores de l/a para um argumento h . A partir da fórmula (10) foram gerados 380 valores de variogramas médios a partir da variação de l/a de 0,1 a 1 de 0,05 em 0,05, e do grau ν de 0,1 a 2 de 0,1 em 0,1. As

figuras abaixo

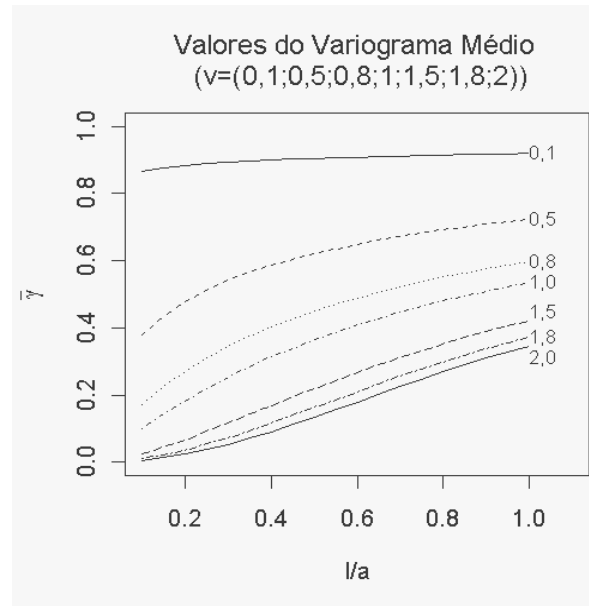


Figura I Valores do variograma médio

Uma visualização de todos os valores calculados é apresentada na Figura 2:

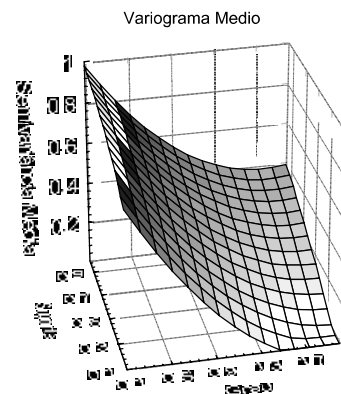


Figura II Visualização do variograma médio calculado

3. O modelo de regressão não-linear

O próximo passo é substituir a fórmula (10) por um modelo aproximado para o variograma médio. Isto foi obtido por regressão não linear para o seguinte modelo:

$$\bar{\gamma}^*(l,l) = [1 - \exp(-\alpha(l/a)^\beta)][\exp(-\alpha'(v)^\beta)] \quad (22)$$

Seguindo recomendação de (RATKOWSKY 1983) utilizamos $\exp(\alpha)$ no lugar de α e $\exp(\alpha')$ no lugar de α' . Usando a rotina NLM do pacote R obtivemos os seguintes valores para os quatro parâmetros:

α	β	α'	β'
1,4079847	0,7911549	-0,2424082	1,1831294

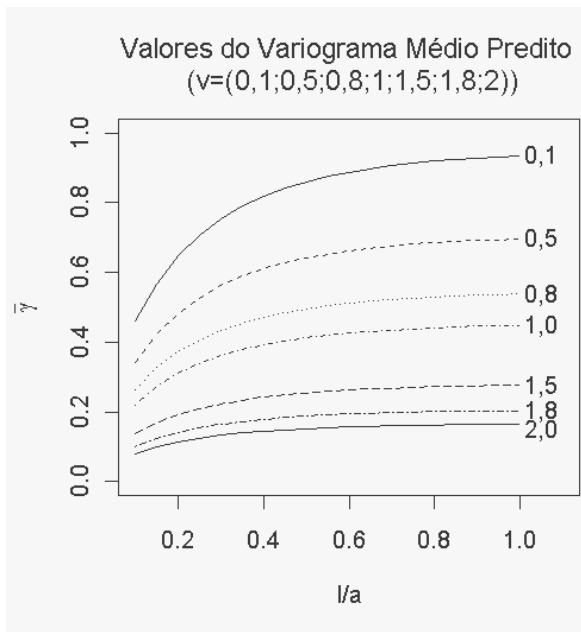


Figura III Variograma médio (predito)

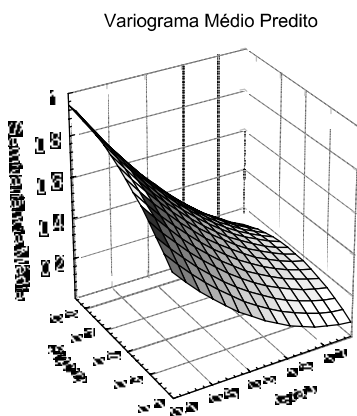


Figura IV Variograma médio (predito)

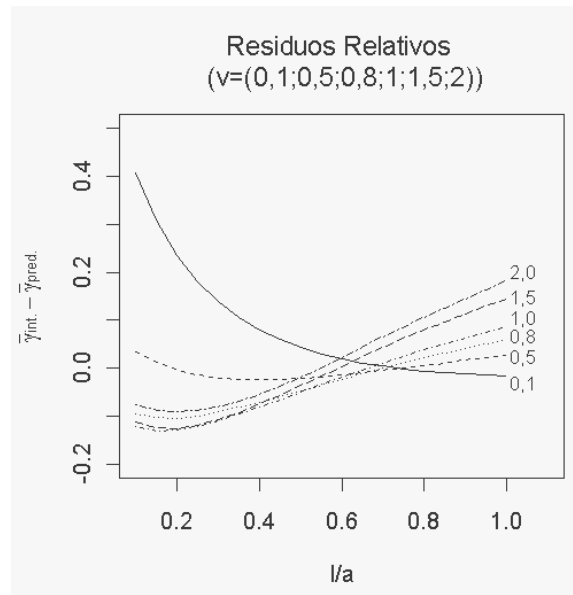


Figura V Resíduos relativos (Variograma médio calculado – predito)

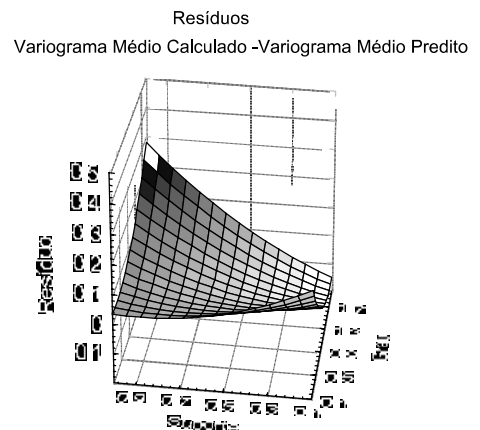


Figura VI Resíduos (Variograma médio calculado – predito)

Para estes valores o gradiente da função objetivo teve as suas componentes da ordem de 10^{-7} , a mediana dos erros foi igual a $-0,0008067$ e a média igual a $-0,0002460$. Em seguida recalculamos os valores do variograma médio usando o modelo ajustado para os mesmos valores de l/a e v . Na figuras III e IV, análogas às Figuras I e II, podemos comparar os valores obtidos. Na Figura V apresentamos o gráfico da diferença entre os dados da Figura I com a Figura III e na Figura VI a visualização da diferença para todos os valores.

Dos gráficos percebe-se que os piores resultados ocorrem

nas bordas da região considerada e também para valores baixos e elevados dos graus.

4. Cálculo do Variograma Regularizado

Obtivemos em seguida os valores do variograma regularizado segundo a fórmula (3), ou seja para valores bem maiores do que l . As figuras seguintes mostram os resultados.

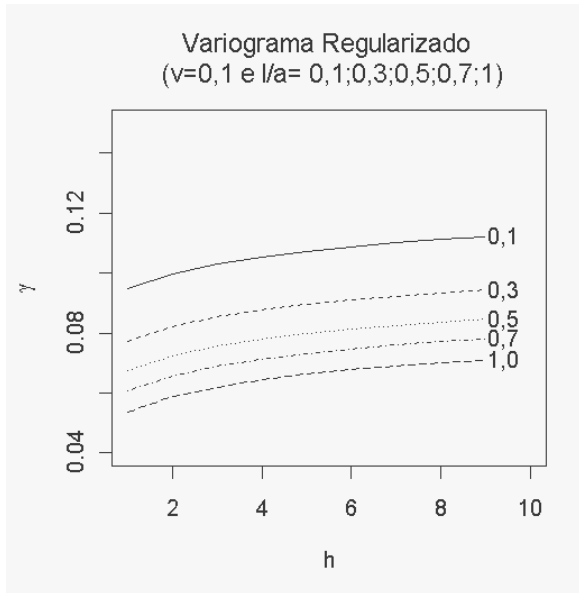


Figura VII Variograma regularizado calculado

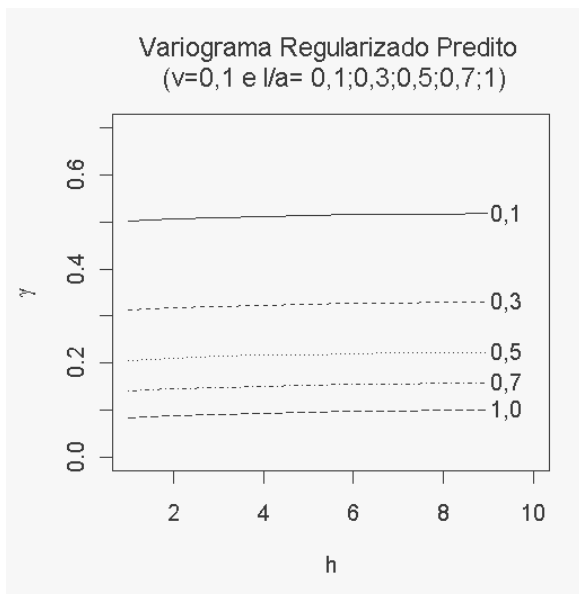


Figura VIII Variograma regularizado predito

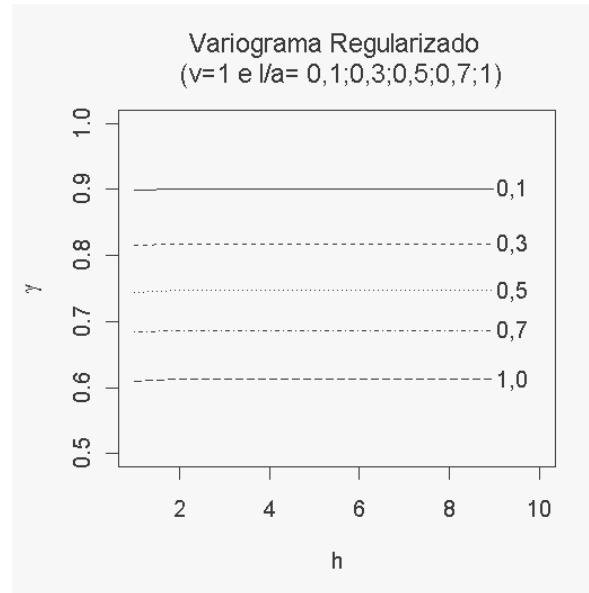


Figura IX Variograma regularizado calculado

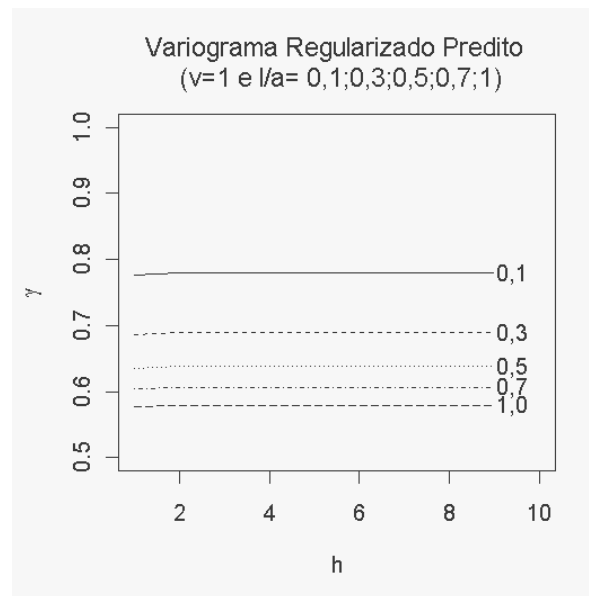


Figura X Variograma regularizado predito

5. Conclusões e trabalhos futuros

Em ambos os casos, calculado e predito, o patamar

decrece com o tamanho do suporte. Por outro lado o patamar cresce com o grau. A coincidência entre variograma calculado e predito não é boa para graus e suportes baixos, assim como para valores elevados de graus. Os melhores resultados são em regiões intermediárias.

Desenvolvimentos deste trabalho supõem : utilizar a fórmula (2) de modo a trabalhar com h qualquer, melhorar o modelo de regressão não linear, utilizar suportes bi-dimensionais, aplicar a metodologia para outros membros da família Matérn.

6. Referências

Abrahamsen, P., 1997: A Review of Gaussian Random Fields and Correlation Functions. Norwegian Computing Center, pp. 64.

Abramowitz, M., Stegun, I.A., 1970: Handbook of Mathematical Functions. Dover Publications Inc. pp. 1045.

Almeida, C.F.P., Braga, L.P.V., Bettini, C., 2001: A Variação do Suporte Amostral e as Mudanças na Estrutura do Semivariograma. III Workshop Brasileiro de Geoinformática, pp. 11-18.

Journel, A .G., Huijbregts, Ch., 1978: Mining Geostatistics. Academic Press, pp. 599.

Olea, Ricardo. Geostatistical Glossary and Multilingual Dictionary. Oxford University Press, 1991, 177 páginas.

Ratkowsky, D.A., 1983: Nonlinear Regression Modeling. Marcel Dekker, pp.276

Ratkowsky, D.A., 1989: Handbook of Nonlinear Regression Models. Marcel Dekker, pp. 241.

Rivoirard, Jacques. Change of support in global and local estimation. in Introduction to Disjunctive Kriging and Non-Linear Geostatistics, Part One , Sections 10 and 11, pp73-95, Clarendon Press, 1994, 181 páginas.

Venables, W., Smith, D., 2002: An Introduction to R Version 1.4.1, www.r-projects.org. 105.