

USO DE INDICADORES PARA ESCOLHA DO MELHOR MODELO DE SEMIVARIOGRAMA

LAURIMAR GONÇALVES VENDRUSCULO¹; JOSÉ RUY PORTO DE CARVALHO²

Escrito para apresentação no
XXXIII Congresso Brasileiro de Engenharia Agrícola
02 a 06 de Agosto de 2004 - São Pedro - SP

RESUMO: O semivariograma constitui-se numa ferramenta importante para representação quantitativa da variabilidade espacial e temporal de determinado atributo. O objetivo deste trabalho é apresentar os indicadores: coeficiente de determinação (R^2) e soma do quadrado dos desvios ponderados (SQDP) e o critério de informação de Akaike, que contribuem para escolha do melhor modelo matemático para a representação de estudos espaciais. Neste trabalho utilizou-se o software geostatístico GEOEST e dados de precipitação anual média do Estado de São Paulo. Para estes dados, dois foram os modelos dos semivariogramas teóricos pré-selecionados. O menor valor encontrado para o Critério de Akaike (AIC) correspondeu ao modelo Gaussiano e os melhores valores do coeficiente de determinação e soma do quadrado dos desvios ponderados coincidiram com o modelo exponencial. A disponibilidade destes indicadores no software GEOEST agilizou o processo de escolha do melhor modelo, diminuindo a subjetividade envolvida neste processo.

PALAVRAS-CHAVE: Geoestatística, Critério de informação de Akaike, software

Use of indicators for choice of best semivariogram models

ABSTRACT: The semivariogram is an important tool for quantitative representation of the spatial and temporal variability. The objective of this work is to present the indicators: determination coefficient and weighted sum of squared deviations and the criterion of information of Akaike, that contribute to choice of the best mathematical model for the representation of spatial studies. In this study the GEOEST, geostatistical software, and data of precipitation annual average of the State of São Paulo were used. For these data, two models of theoretical semivariograms were selected. The smallest value found for the Criterion of Akaike Information (AIC) it corresponded to the model Gaussiano and the best values of the determination coefficient and weighted sum of squared deviations coincided with the exponential model. To make available these indicators in the GEOEST software improve the process of choice of the best model, reducing the subjectivity involved in this process.

KEYWORDS: Geostatistics, Akaike Information Criterion, software

INTRODUÇÃO: O semivariograma constitui-se numa ferramenta importante para representação quantitativa da variabilidade espacial e temporal de determinado atributo. Segundo Vieira (2000), esta ferramenta é a mais adequada para medir dependência espacial.

Em geoestatística a escolha do melhor modelo de semivariograma é crucial para a correta interpretação de fenômenos que apresentem dependência espacial. A esta escolha, repercute diretamente a confiabilidade dos resultados oriundos do processo de interpolação (krigagem); ou seja, na confiabilidade dos valores estimados em pontos não amostrados.

De uma maneira geral, a modelagem de sistemas tais como os biológicos e químicos, constituem-se em uma atividade que contempla múltiplos e conflitantes objetivos. Estes objetivos envolvem a complexidade do(s) modelo(s), o(s) critério(s) de desempenho e a validação, que influenciam a seleção da estrutura do modelo matemático.

O objetivo deste trabalho é apresentar os indicadores: coeficiente de determinação, soma do quadrado dos desvios ponderados e o critério de informação de Akaike, que contribuem para escolha do melhor modelo matemático para a representação de estudos espaciais. Foi usado o software GEOEST (Vendrusculo, 2004), destinado à análise geostatística o qual implementa tais indicadores. Com o intuito de exemplificar o ajuste e a escolha do melhor modelo do semivariograma, utilizou-se dados

de precipitação anual média do Estado de São Paulo. Os dados provêm de mil e vinte e sete observações correspondentes as estações do DAEE, para o período de 1957 a 1997.

MATERIAL E MÉTODOS: Uma das maneiras de se encontrar a melhor estrutura de um modelo matemático para a representação de um sistema dinâmico pode se dar por meio da estimação de parâmetros para todas as possíveis estruturas e a consequente escolha baseado na comparação de alguns índices de desempenho.

No caso do semivariograma, os parâmetros estimados nos modelos matemáticos, são: C_0 – Efeito pepita, C – patamar e a – alcance. Os modelos matemáticos mais usados no contexto agropecuário, que contemplam estudos de variabilidade das variáveis do solo e agroclimatológicos são os modelos esférico, exponencial e gaussiano, representados na Figura 1.

O semivariograma pode ser definido pela Equação 1, onde $N(h)$ é o número de pares de valores medidos $Z(x_i)$ e $Z(x_i+h)$, separados pela distância h se a variável for escalar. (Carvalho et al., 2003)

- Efeito Pepita (C_0) - Quando o semivariograma tende a zero ($\sigma(h) = 0$) o valor C_0 é observado, revelando a descontinuidade do fenômeno para valores menores que a menor distância entre as amostras. Isaaks e Srivastava (1989) atribue às condições adversas na época de medição ou a variabilidade de menor escala não percebida pelo processo de amostragem.

- Alcance (a): distância na qual as amostras se encontram correlacionadas espacialmente.

- Patamar (C): corresponde ao alcance (a) no gráfico do semivariograma. Considera-se que a partir deste ponto não exista mais dependência espacial entre as amostras. É aproximadamente igual a variância dos dados.

- Variância Estrutural (C_1): É a diferença entre patamar e o efeito pepita.

Dentre os indicadores frequentemente utilizados para a tomada de decisão sobre o melhor modelo matemático citamos:

1) Coeficiente de determinação ou correlação múltipla (R^2) – O modelo que apresentar maior valor de R^2 ($0 < R^2 \leq 1$) é considerado o melhor. O valor de 1 para R^2 representa que o modelo teórico se adequou exatamente aos valores medidos no processo. Pode-se calcular o coeficiente de determinação pela Equação 2.

2) Soma dos Quadrados dos Desvios Ponderados (SQDP) – Neste caso escolhe-se o menor valor como o melhor. O cálculo deste índice é obtido pela Equação 3, onde P_i correspondem aos números de pares para cada classe de distância; $C(h)$ representa o número de classes de distâncias.

Estes critérios são largamente utilizados para a escolha do melhor modelo, porém não ponderam sobre o número de componentes usados para o modelo matemático estimado.

Para tanto um compromisso satisfatório entre o bom ajuste e o princípio de parcimônia pode ser alcançado aplicando-se o chamado Critério de informação de Akaike (AIC), apresentado por seu autor em Akaike (1974) como um procedimento para identificação de modelo estatístico.

$$AIC = -2 \log(\text{Máxima Verossimilhança}) + 2p \quad (4)$$

Onde p é número de parâmetros do modelo independentemente ajustados.

2.2 - Estudo de Caso

Um exemplo, no software GEOEST, da utilização dos indicadores e Critério de Akaike para dados de precipitação anual do estado de São Paulo é visualizado pela Figura 2.

O software destaca automaticamente o melhor valor de cada índice em cor vermelha, a medida que o usuário entra ou modifica os valores do alcance, efeito pepita e variância estrutural. A disponibilidade de tais informações auxilia a determinação dos parâmetros estruturantes do modelo a ser submetido à técnica de auto-validação Jack-knifing. Esta técnica re-estima todos os pontos medidos em função dos pontos vizinhos e do modelo matemático escolhido. Devem ocorrer várias iterações (ajustes

manuais) por parte do usuário optando pelo melhor modelo em função dos índices e da etapa de Jackknifing. O software realiza o cálculo dos índices em modelos matemáticos de estrutura simples, ou seja, não aninhados.

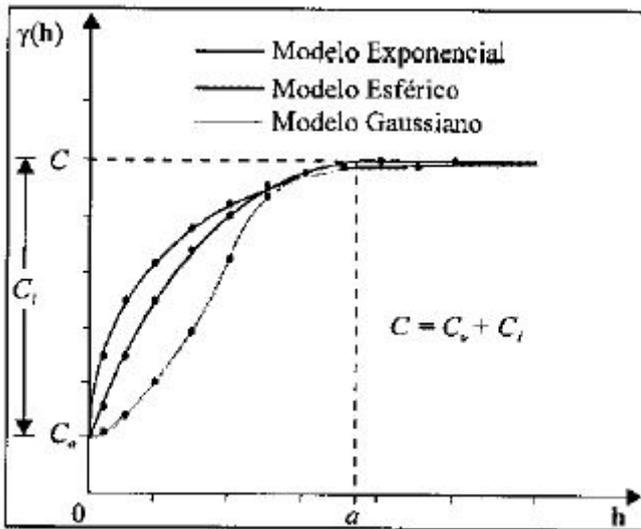


Figura 1 - Modelos teóricos de semivariograma

$$\gamma^*(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} [Z(X_i) - Z(x_i + h)]^2 \quad (1)$$

$$R^2 = \left(1 - \frac{\sum_{i=1}^{C(h)} (\gamma(h)_{\text{exp}} - \gamma(h)_{\text{mod}})^2}{\sum_{i=1}^{C(h)} (\gamma(h)_{\text{exp}} - \gamma(h)_{\text{orig médio}})^2} \right) \quad (2)$$

$\gamma(h)_{\text{exp}}$ corresponde a semivariância do semivariograma experimental;

$\gamma(h)_{\text{mod}}$ é a semivariância do modelo matemático;

$\gamma(h)_{\text{orig médio}}$ é a semivariância média do semivariograma experimental,

$$SQDP = \sum_{i=1}^{C(h)} \left(\frac{P_i}{\sum_{i=1}^{C(h)} P_i} * (\gamma(h)_{\text{mod}} - \gamma(h)_{\text{exp}})^2 \right) \quad (3)$$

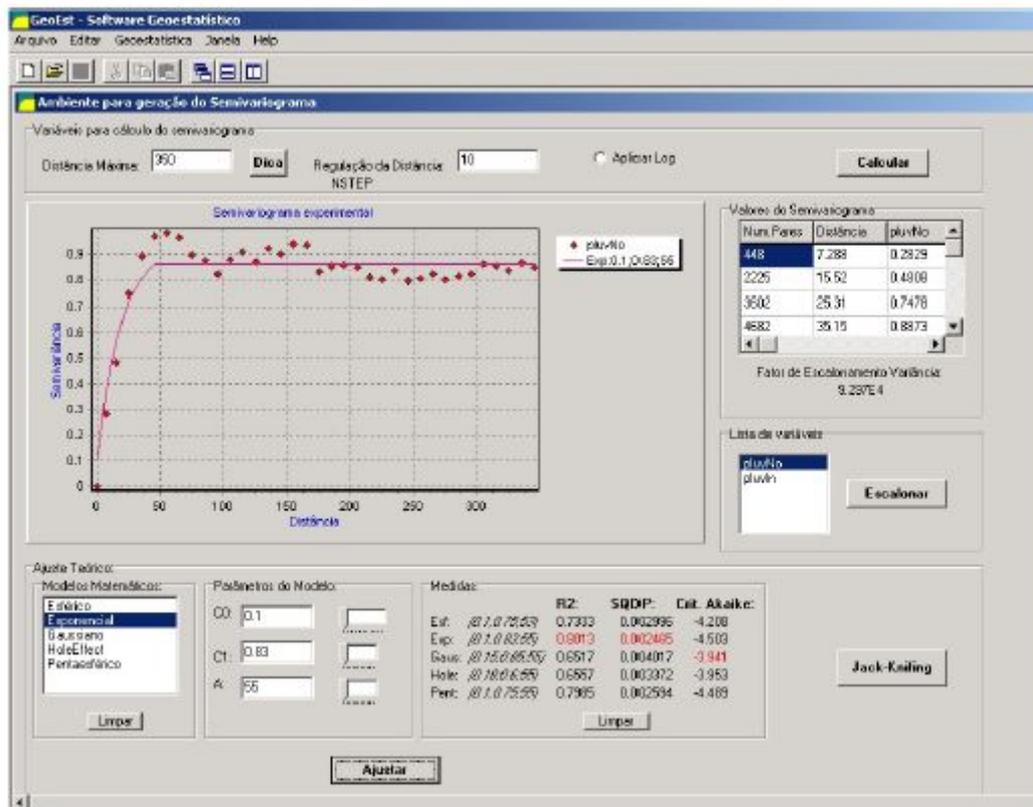


Figura 2 – Ambiente para modelagem do semivariograma destacando os índices de desempenho (R2 e SQDP) e Critério de Akaike (AIC) para dados de precipitação no estado de São Paulo

RESULTADOS E DISCUSSÃO: A situação ideal seria que os indicadores e AIC convergissem para um único modelo dos cinco (5) apresentados na Figura 2. Ou seja o melhor modelo seria aquele que apresentasse, concomitantemente, o maior valor de R2, menor valor de SQDP e menor valor de AIC. Na Figura 2, o menor valor de AIC corresponde ao modelo Gaussiano (-3.941) e os melhores valores de R2 e SQDP coincidem com o modelo exponencial.

Na situação estudada mais iterações dos parâmetros do modelo seriam necessárias ou pelo conhecimento da variabilidade do atributo, o usuário poderia optar pela escolha do modelo matemático que atendesse a maioria dos índices.

CONCLUSÕES: - uso de índices de desempenho e critério de informação de Akaike constituem-se em importante instrumento utilizado para a escolha do melhor modelo para ajuste do semivariograma, diminuindo a subjetividade que permeia este processo.

- Para aplicações que permitam a modelagem do semivariograma de forma manual e onde são necessárias grande número de interações a disponibilidade destes índices e do critério no Software GEOEST, e o destaque automático dos melhores valores agiliza o processo de modelagem do semivariograma, etapa fundamental na análise de variabilidade geoestatística.

REFERÊNCIAS BIBLIOGRÁFICAS:

- AKAIKE, H., A new look at statistical model identification, IEEE Trans. On Automatic Control, p. 716-723, 1974.
- ISAAKS, E. H., SRIVASTAVA, R.M. An Introduction to applied geostatistics, New York: Oxford University Press, 1989. 561p.
- VENDRUSCULO, L.G., MAGALHÃES, P.S.G., VIEIRA, S.R.; CARVALHO, J.R.P., Computational system for geostatistical analyses, Scientia Agricola, V. 61, n. 1, p. 100-107, 2004.
- VIEIRA, S. R. Geoestatística em estudos de variabilidade espacial do solo. In: NOVAIS, R. F. de; ALVAREZ V., V. H; SCHAEFER, C. E. G. R. (Ed.). Tópicos em ciência do solo. Viçosa: Sociedade Brasileira de Ciência do Solo, v. 1, p. 1-54, 2000.