

---

**Rotinas para Análises Geoestatísticas Utilizando a Linguagem R: um exemplo com dados agro-ambientais**

*Lucas Venezian Pova<sup>1</sup>, Rodrigo Lilla Manzione<sup>2</sup> e Edson Cesar Wendland<sup>3</sup>*

<sup>1</sup> Graduando em Análise de Sistemas, FATEC/Ourinhos, Av. Vitalina Marcusso, 1400 CEP: 19910-206 – Ourinhos/SP, lucas.pova@fatec.sp.gov.br, URL: <http://www.lucasvenezian.orgfree.com>

<sup>2</sup> Eng. Agrônomo, UNESP/Ourinhos, Av. Vitalina Marcusso, 1500 CEP: 19910-206 – Ourinhos/SP, manzione@ourinhos.unesp.br

<sup>3</sup> Eng. Civil, EESC/USP, Caixa Postal 359, CEP: 13560-970 São Carlos/SP, ew@sc.usp.br

**Resumo** - Este artigo apresenta análises geoestatísticas em dados agro-ambientais utilizando a linguagem R como ambiente para estatística computacional e visualização gráfica. O objetivo foi descrever de maneira simples e direta como utilizar um software livre e de domínio público e alguns pacotes disponibilizados por seus desenvolvedores e usuários para armazenar as informações espaciais, o modo de importá-las para o ambiente R e os recursos mínimos para a execução de análises geoestatísticas. Entre as operações desenvolvidas estão a variografia, predição espacial utilizando krigagem ordinária e validação cruzada dos resultados. Espera-se com isso encorajar a utilização do R para análises espaciais, uma vez que essa linguagem é um recurso completo, estável, confiável e desenvolvida e aprimorada por renomados pesquisadores na área. As rotinas descritas podem ser reproduzidas para fins de mapeamento, agricultura de precisão, levantamentos e inventários, entre diversas outras aplicações agro-ambientais.

**Palavras-chave:** variografia; krigagem; validação cruzada.

**Routines to Geostatistics Analysis Using R Language: an example with agri-environmental data**

**Abstract** - This paper presents geostatistical analysis on agro-ecological data using the R language as environment for statistical computing and visualization. The aim was to describe in a simple and direct way how to use and free and with public domain software and some packages available from developers and users to storage spatial information, import it to R environment and the minimal resources for implementation of geostatistical analysis. Among the operations carried out are variography, ordinary kriging and cross validation of the results. We expect with this work encourage the use of the R language for spatial analysis, since it is a complete, stable and reliable feature, developed and perfected by renowned researchers in this field. The routines described here could be reproduced for different targets, like mapping, precision farming, survey and inventory, among several other agro-ecological applications.

**Key words:** variography; kriging; cross-validation.

**Introdução**

R é uma linguagem aberta e gratuita com a finalidade de resolver problemas gráficos e de computação estatística, sendo uma implementação da linguagem S desenvolvida inicialmente por John Chambers, Rick Becker e Allan Wilks dos Laboratórios Bell (R Development Core Team, 2008). A linguagem R fornece padrões e métodos inovadores de análise estatística, e devido a sua característica de desenvolvimento aberto (*open source*) garante um contínuo aprimoramento de bibliotecas e rotinas por inúmeros profissionais de diversos pontos do mundo. Ser uma ferramenta estável e confiável fez com que milhares de usuários aderissem ao R como software estatístico para análises. A análise de dados espaciais também é muito difundida em R, e tem ganhado um grande número de recursos nos últimos anos. Especificamente em geoestatística, rotinas até então disponíveis em softwares comerciais e de alto valor no mercado são disponibilizadas em diversas bibliotecas. Além disso, o R dispõe de funções especiais para exibir, armazenar e organizar dados desse tipo. Contudo, a linguagem R por si só não possui artifícios para distinguir coordenadas de qualquer outro tipo de dado numérico. Uma equipe de pesquisadores escreveram um pacote denominado *sp* PEBESMA e BIVAND (PEBESMA; BIVAND, 2005) com classes e métodos específicos para dados espaciais, sendo esta biblioteca a matéria-prima de ferramentas em R para

processamento e análise de dados dessa grandeza. BIVAND et al. (2008) apresentam uma relação de pacotes disponíveis que dependem ou importam o pacote *sp* direta ou indiretamente e outros que utilizam ou sugerem utilizar o pacote *sp* sem declará-lo na descrição do pacote. Podemos citar os pacotes (ou bibliotecas) *mapproj*, *rgdal*, *splancs*, *geoR*, *gstat*, *spsurvey*, *trip*, entre muitos outros.

Desse modo, o objetivo deste trabalho foi descrever rotinas geoestatísticas em R que fossem aplicáveis a dados agro-ambientais, perfazendo variografia, krigagem ordinária e validação cruzada dos resultados a partir de dados pontuais.

## Material e Métodos

Para a execução das operações de geoestatística foi utilizado o programa R versão 2.10.0 (R Development Core Team, 2008) e os pacotes *sp* (PEBESMA; BIVAND, 2005), *gstat* (PEBESMA, 2004) e *lattice* (SARKAR, 2010). Os dados utilizados nos processos são pontos georreferenciados referentes a posição de poços de monitoramento de níveis freáticos em uma bacia hidrográfica em área de recarga do Sistema Aquífero Guarani (SAG) em Brotas (SP), a bacia do Ribeirão da Onça. São ao todo 20 poços georreferenciados, coordenadas planas UTM fuso 23 s. Também estão disponíveis dados vetoriais representando o limite da bacia e um modelo digital de terreno (MDT) em formato raster (WENGLAND et al., 2007). Os dados relativos aos poços foram dispostos da seguinte forma: a primeira linha representa o nome dos atributos, cada coluna é separada por uma tabulação (\t) e cada linha posterior a primeira representa o conjunto de dados de um determinado poço (Tabela 1).

**Tabela 1.** Dados pontuais de alturas do nível d'água e respectivas cotas altimétricas

X	Y	Z	E
194592,92	7537198,83	-17,45	753
194632,80	7537211,47	-5,42	754
197699,38	7543585,49	-26,38	748
198326,23	7544977,48	-18,97	736
197210,52	7544401,75	-25,30	731
196917,51	7545107,64	-9,80	715
197003,34	7545176,91	-6,62	714
197588,00	7545341,91	-7,63	706
195367,02	7544827,80	-5,49	707
195377,76	7544717,14	-12,43	712
195362,31	7544627,54	-15,95	717
195354,38	7544445,75	-16,20	720
196798,22	7545255,82	-2,89	694
196776,36	7545347,75	-0,45	687
197154,03	7545539,79	-1,93	691
197154,15	7545533,64	-0,99	690
194592,92	7537198,83	-17,45	753
194632,80	7537211,47	-5,42	754
197699,38	7543585,49	-26,38	748
198326,23	7544977,48	-18,97	736

Z=altura do nível d'água (metros); E = elevação (metros)

Essa tabela pode ser copiada e colada diretamente em algum aplicativo do tipo bloco de notas, gravada como formato texto (.txt) e lida e importada pela função *read.table* do R. É necessário nomear a tabela, no caso ALF (alturas de lençol freático). Em seguida a função *coordinates* transforma os dados carregados em dados do tipo *SpatialPointsDataFrame*, definindo as coordenadas x e y dos pontos. Os dados da bacia hidrográfica (limite e MDT) foram organizados seguindo o conjunto de regras do formato ASCII. Os processos executados para importação dessas informações foram através da função *read.asciigrid* que carrega os dados referentes a bacia e os transforma em um objeto do tipo *SpatialGridDataFrame*. Atribuir a

todos os dados a classe *Spatial*, que é um pré-requisito para a execução de análises espaciais no R.

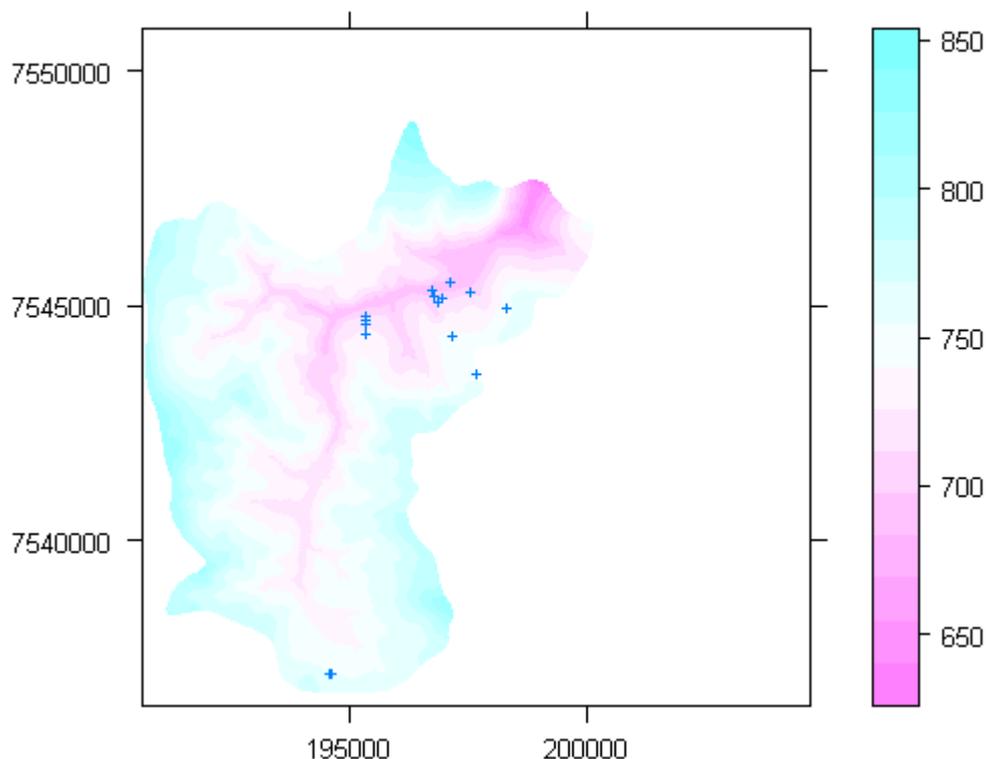
Após essa tarefa foram aplicadas funções do pacote *gstat* para análises geoestatísticas. As funções *variogram*, *vgm* e *fit.variogram*, foram utilizadas com a finalidade de gerar e ajustar um modelo teórico ao variograma. A função *krige* foi utilizada para gerar uma superfície estimada e a variância dessas previsões. A validação dos dados se deu através da função *krige.cv*. A visualização dos resultados é atribuída às funções *spplot* e *print*.

### Resultados e Discussão

Os resultados são expostos em vários fragmentos com a finalidade de melhor explorar cada ponto do *script* desenvolvido, e após cada ponto explorado é exibido o seu resultado. O elemento ">" no início de cada linha do código R deve ser descartado, uma vez que representa início de linha. Portanto, o código abaixo tem a finalidade de carregar os dados espaciais em objetos R e torná-los tipos derivados da classe *Spatial*:

```
> ALF = read.table( "pocos.txt", header=T );
> coordinates(WTD) = ~ X + Y;
> MDT = read.asciigrid("mdt.txt", plot.image = TRUE, colname = "E");
> spplot(MDT["E"], scales=list( draw=T ), sp.layout=list("sp.points", ALF,
pch="+") );
```

O código acima além de carregar os dados exibe um mapa com a representação do limite e altitudes da bacia hidrográfica junto aos pontos dos poços, como ilustra a Figura 1.

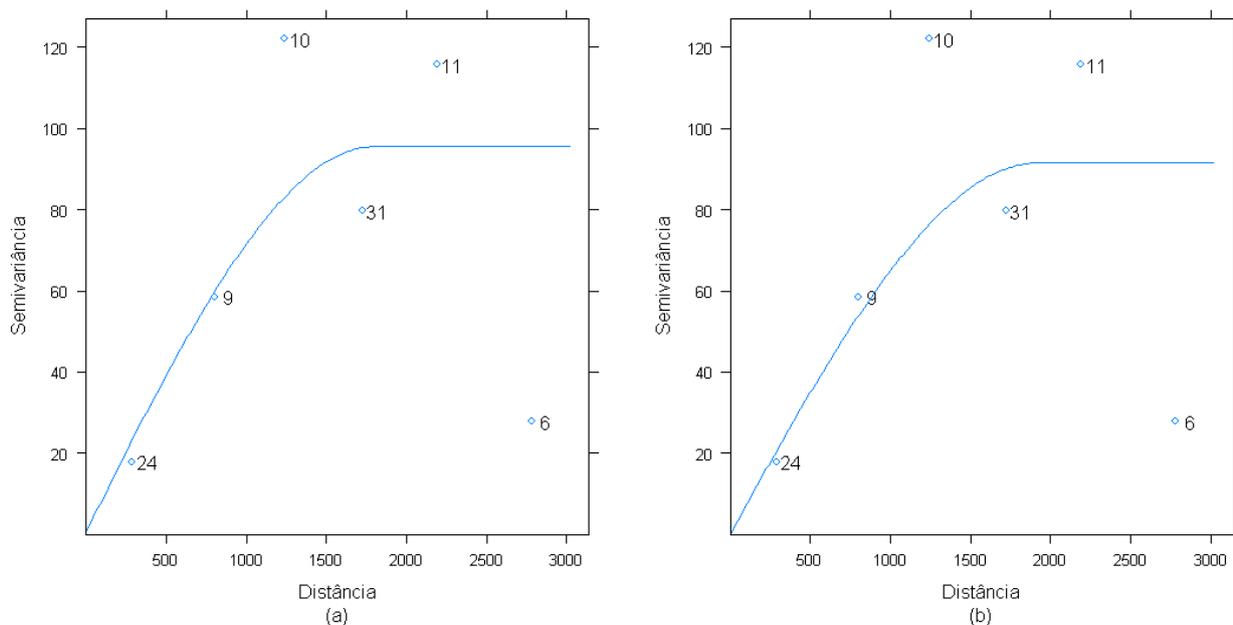


**Figura 1.** Visualização dos dados de poços (cruzes) sobre o modelo digital do terreno (metros) da Bacia do Ribeirão da Onça.

O trecho a seguir se refere à variografia e executa respectivamente o cálculo do variograma a partir de uma amostra de dados, ajusta um modelo teórico ao variograma amostral a partir de valores informados ou mesmo realiza um ajuste automático.

```
> variograma = variogram( Z ~ 1, ALF, cutoff=4000, width=500 );
> modelo = vgm( 95, "Sph", 1800, nug = 0.5);
> modelo_ajustado = fit.variogram(variograma, model = variograma);
> print( plot( variograma, model = modelo ), split=c(1,1,2,1), more=T);
> print( plot( variograma, model = modelo_ajustado ), split=c(2,1,2,1), more=F);
```

Primeiro deve-se definir as dimensões do variograma. Na função *variogram*, *cutoff* refere-se ao campo amostral do variograma e *width* a largura do passo definido. A seguir executa-se o ajuste a sentimento. Na função *vgm* são definidos primeiro o patamar do variograma (95), seguido pelo tipo de modelo teórico (esférico/Sph), alcance (1800) e efeito pepita (nug). Ainda é possível optar por um ajuste automático, por mínimos quadrados. Os resultados com os distintos modelos gerados são exibidos na Figura 2.



**Figura 2.** Variogramas utilizando o resultado da função *vgm* como modelo (a) e utilizando o resultado da função *fit.variogram* para ajuste automático do modelo (b).

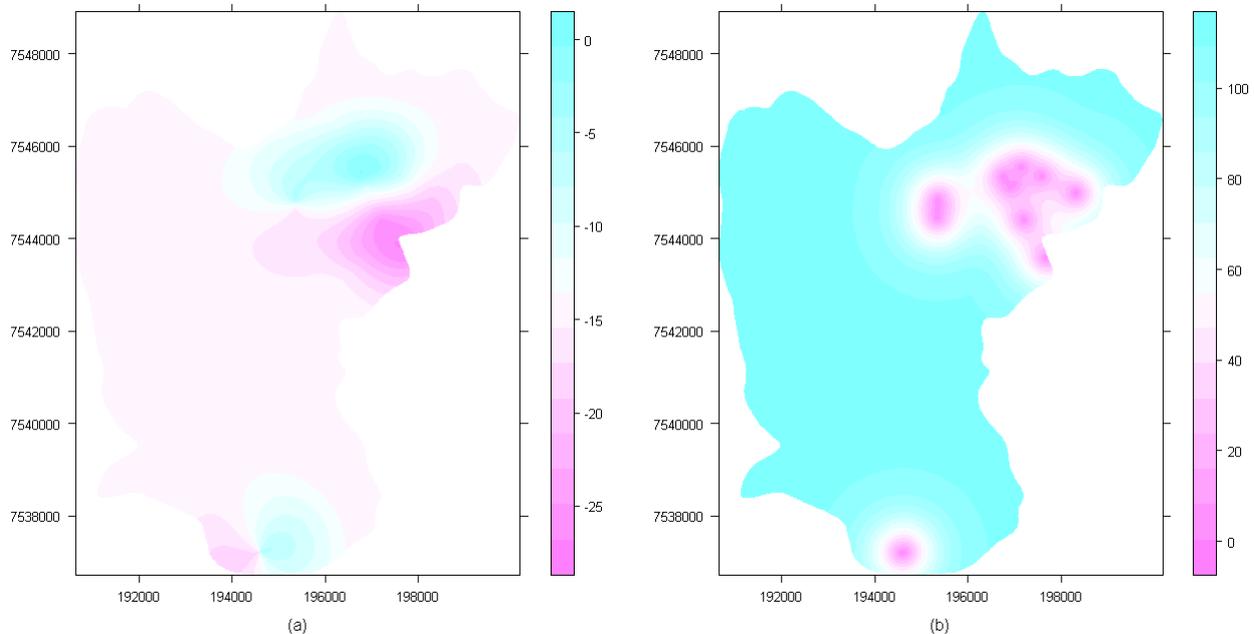
Por fim, essa última porção do código executa a interpolação dos dados por krigagem ordinária e realiza a validação dos resultados através da técnica de validação cruzada. Na função *krige* o primeiro parâmetro define a variável a ser interpolada no arquivo indicado (segundo parâmetro) e os parâmetros seguintes definem respectivamente o grid (células da malha) da bacia para executar a interpolação (no caso MDT) e o modelo do variograma usado nessa interpolação (*model*). Esse grid é necessário para que em cada célula o interpolador resolva o sistema de krigagem. Por fim, *nmax* define o número máximo de pares considerados na vizinhança para interpolação. A Tabela 2 exhibe os resultados da validação cruzada, reorganizados como distribuição de frequência e a Figura 3 ilustra os resultados da interpolação dos dados por krigagem ordinária.

```
> OK = krige(Z ~ 1, ALF, MDT, model = modelo_ajustado, nmax=40);
> CV.OK = krige.cv(Z ~ 1, ALF, modelo_ajustado, verbose=F)
> print( spplot(OK["var1.pred"], xlab="(a)"), split=c( 1, 1, 2, 1 ), more=T );
> print( spplot(OK["var1.var"], xlab="(b)"), split=c( 2, 1, 2, 1 ), more=F );
```

**Tabela 2.** Validação cruzada para a krigagem ordinária realizadas para os dados de alturas do lençol freático

	Pred	Var	Obs	Res	Zscore
Mínimo	-17,89	1,93	-26,38	-11,72	-4,28
1º Quartil	-14,50	8,44	-16,51	-3,32	-0,73
2º Quartil	-10,54	13,92	-8,72	-0,36	-0,19
3º Quartil	-5,16	32,94	-4,79	1,29	0,37
Máximo	-1,04	98,54	-0,45	11,94	4,36
Média	-9,94	27,77	-12,10	-0,93	-0,07
DesvPad	5,86	31,54	8,58	5,84	1,72

DesvPad=desvio padrão; Pred=valor predito; Var=variância da predição;  
Obs=valor observado; Res=resíduos; Zscore=(pred-obs)/var



**Figura 3.** Resultados da função *krige*: predição de alturas de lençol freático (metros) a partir da interpolação por krigagem ordinária (a) e a variância da krigagem ordinária (b).

Nossa proposta foi expor os recursos mínimos necessários para análises geoestatísticas em R. Para os resultados obtidos foram necessário conhecimentos básicos de geoestatística na interpretação das funções disponíveis para realização dos procedimentos listados. Todas as funções possuem documentação disponível na ajuda dos pacotes carregados (*help*), onde um conhecimento básico de inglês se faz necessário para compreender os textos. Aqui foram utilizadas apenas algumas partes de cada função, que podem ser mais complexas, com mais parâmetros, conforme o grau de conhecimento em R e em geoestatística que o analista tenha para utilizá-las. Espera-se assim encorajar novos usuários na utilização da linguagem R para análises espaciais, uma vez que é um recurso completo, estável, confiável e também gratuito, desenvolvido e aprimorado por renomados pesquisadores. As rotinas descritas aqui podem ser entendidas para análises mais complexas e avançadas como técnicas de krigagem indicativa, universal ou mesmo simulação seqüencial gaussiana, que não estão implementadas em todos os softwares geoestatísticos disponíveis no mercado. Além disso, é possível desenvolver suas próprias rotinas e bibliotecas em casos específicos definidos pelos usuários.



### **Conclusões**

O programa R se mostra uma alternativa viável, em relação a softwares comerciais, para análises espaciais e geoestatísticas, uma vez que importa, armazena, analisa e permite a visualização dos resultados. Para execução de tarefas através de seus pacotes e bibliotecas é necessário um conhecimento intermediário de geoestatística, uma vez que não se trata de um software intuitivo onde o usuário é guiado através de janelas a cada passo da análise. As rotinas descritas podem ser reproduzidas para fins de mapeamento, agricultura de precisão, levantamentos e inventários, entre diversas outras aplicações agro-ambientais.

### **Referências**

- BIVAND, R. S.; PEBESMA, E. J.; GÓMEZ-RUBIO, V. **Applied Spatial Data Analysis with R**. Nova York: Springer, 2008. 374 p.
- PEBESMA, E. J. Multivariable geostatistics in S: the gstat package. **Computers & Geosciences**, v. 30, p. 683-691, 2004.
- PEBESMA, E. J.; BIVAND, R. S. Classes and methods for spatial data in R. **R News**, v. 5, n. 2, p. 9–13, 2005.
- R DEVELOPMENT CORE TEAM. R: A Language and Environment for Statistical Computing. **R Foundation for Statistical Computing**, 2008. Disponível em: <<http://www.r-project.org>>. Acesso em: 29 Novembro 2010.
- SARKAR, D. lattice: Lattice Graphics. **R package version 0.18-3**, 2010. Disponível em: <<http://CRAN.R-project.org/package=lattice>>. Acesso em: 10 Dezembro 2010.
- WENDLAND, E, BARRETO, C. E. A. G., GOMES, L. H. Water balance in the Guarani Aquifer outcrop zone based on hydrogeologic monitoring. **Journal of Hydrology**, v. 342, p. 261-269, 2007.