

# SELEÇÃO DE MODELOS DE VARIABILIDADE ESPACIAL E ELABORAÇÃO DE MAPAS TEMÁTICOS DO TEOR DE CARBONO NO SOLO

**Rodrigo de Almeida Muniz**

**RESUMO:** O estudo do carbono orgânico no solo constitui-se um desafio através da estatística descritiva tradicional. Devido a variabilidade espacial do carbono a análise geoestatística torna-se um ferramenta poderosa para o estudo da variação espacial do carbono no solo, através da coleta em pontos conhecido e estimativa de um semivariograma experimental, por meio do semivariograma experimental é possível ajustar um modelo geoestatístico para a modelagem da variação de carbono no solo. Este trabalho teve por objetivo estimar, avaliar e elaborar mapas temáticos com os modelos escolhido de carbono no solo. Os estimadores dos parâmetros do modelo utilizados foram a máxima verossimilhança (ML) e mínimos quadrados ordinários (OLS), para a validação dos modelos foi utilizado, o critério de Akaike (AIC), o valor do mínimo quadrado ordinário (MMS), o máximo valor do logaritmo da função de máxima verossimilhança e análise de resíduos gerado na validação cruzada. Pelo critério utilizados, não foi possível detectar com segurança o melhor modelo a ser adotado, sendo imprescindível o conhecimento da variável em estudo, base bibliográfica e bom senso do pesquisador.

Palavras chaves: Carbono no solo, geoestatística, krigagem.

## INTRODUÇÃO

A presença de matéria orgânica nos solos é de grande importância uma vez que ela é o principal agente gerador de cargas elétricas negativas responsáveis pela retenção de nutrientes e água, pela agregação do solo, além de servir de substrato e contribuir para a manutenção da diversidade biológica do solo (Silva et al., 2004).

A conversão da vegetação nativa em sistema de produção com cultivo convencional, tem resultado em um declínio significativo da concentração de matéria orgânica do solo (Paustian et al., 2000; Lal, 2002). Além disto, há forte impacto ambiental com o aumento da emissão de gases do efeito estufa, especialmente CO<sub>2</sub> (Machado, et al., 2004). Práticas agrícolas que utilizam aração mecanizada no preparo do solo para

plantio, normalmente promovem, por diferentes mecanismos, a perda de matéria orgânica e conseqüentemente, a emissão de gases para a atmosfera.

A concentração de carbono é muito variável no solo, podendo ser influenciado por características intrínsecas e extrínsecas deste (Panosso, 2009). Diante de sua variabilidade o estudo de sua variação espacial no solo é um fator complicante, no entanto, através de análises geoestatísticas podem-se ajustar modelos geoestatísticos adequados que representem o seu comportamento espacial.

A geoestatística surgiu para o estudo de variáveis regionalizadas, ou seja, variáveis com condicionamento espacial, preocupando-se com o entendimento por meio de análise matemática, da gênese e leis naturais que governam fenômenos interpretados como regionais, com continuidade aparente e cujos valores são relacionados com a posição que ocupam, ou seja, não é inteiramente independente. Através de modelos matemáticos, utilizando métodos de interpolação de informações a partir de dados obtidos em locais convenientemente amostrados e modelados em um semivariograma experimental.

Para modelar um conjunto de dados com uma estrutura de correlação, considera-se um processo estocástico gaussiano  $\{Z(s), s \in S\}$ , em que  $S$  é um subconjunto de  $R^d$ , sendo  $d$  um espaço euclidiano  $d$ -dimensional. Supõe-se que os dados  $Z(s_1), \dots, Z(s_n)$  do processo são eventos conhecidas nos locais  $s_i, i = 1, \dots, n$  em que  $s_i$  é um vetor  $d$ -dimensional ( $d \geq 1$ ). Supõe-se que os dados, de modo geral, podem ser escritos como  $Z(s_i) = \mu(s_i) + \epsilon(s_i)$ , sendo  $\mu(s_i)$  uma função determinística que descreve a componente estrutural de  $Z$  em  $s_i$ , que pode ser expressa como  $\mu(s_i) = \sum_{k=1}^p f_k(s) \beta_k$ , sendo  $f_k$  uma função conhecida e  $\beta_k$  uma constante desconhecida a ser estimada, para  $k = 1, \dots, p$  e  $\epsilon(s_i)$  é um termo estocástico, que varia localmente e depende espacialmente de  $Z(s_i)$ . Assume-se que o termo estocástico  $\epsilon(s_i)$  tem média zero e a variação entre pontos no espaço é determinada pela função covariância  $C(s_i, s_j) = \text{Cov}\{\epsilon(s_i), \epsilon(s_j)\}$  (Mardia & Marshall, 1984).

A semivariância é uma função da distância  $h$ , que é estimada em um conjunto discreto de distâncias (*lags*). Em termos da covariância, a função semivariância define-se como  $\gamma(h) = C(0) - C(h)$ , sendo  $h = \|s_i - s_j\|$ ,  $C(h) = C(s_i, s_j)$  e  $C(0) = \sigma^2$  a variância das observações. A partir da estimativa de  $\gamma(h)$ , é possível ajustar um modelo com características espaciais e parâmetros desconhecidos que devem ser determinados.

Segundo Isaaks & Srivastava (1989), a função semivariância  $\gamma(h)$  é definida na equação 1 como:

$$\gamma(h) = \frac{1}{2}E[Z(s) - Z(s + h)]^2 \quad (1)$$

em que  $Z(s)$  é o valor da variável medida;  $s \in S \subset \mathbb{R}^d$ ,  $d = 1, 2$  ou  $3$ , a localização; e  $h$ , a distância entre duas amostras. O semivariograma experimental é um gráfico de dispersão de  $\gamma(h)$  em função de  $h$  que permite uma análise variográfica do comportamento de  $Z(s)$  (Cressie, 1993).

O estimador de momentos é o principal na construção do semivariograma experimental, este é conhecido como semivariograma de Matheron, para processos gaussianos, caso os dados não possuam uma distribuição normal, lança-se mão de métodos matemáticos para a normalização dos dados. A partir da semivariância empírica, ajusta-se um modelo teórico aos pontos obtidos.

Escolher um modelo adequado é obter estimadores dos parâmetros efeito pepita, patamar e alcance, com métodos estatísticos de otimização como: mínimos quadrados ordinários (Cressie, 1985) e máxima verossimilhança (Mardia & Marshall, 1984). Uma vez escolhido o modelo teórico de correlação espacial, resta saber se ele é eficiente para interpolar valores, permitindo estimativas confiáveis para construção de mapas temáticos (Cressie, 1985).

Os métodos de validação comparam valores teóricos do modelo geoestatístico escolhido e os valores empíricos obtidos na amostragem. Com base na análise desses erros de estimação, poderá ser selecionado o melhor modelo de semivariância. Entre os principais critérios para validação, encontram-se os de Informação de Akaike, de validação cruzada e o máximo valor do logaritmo da função verossimilhança.

O Critério de Informação de Akaike - AIC (Akaike's Information Criterion) procura uma solução satisfatória entre o bom ajuste e o princípio da parcimônia (Akaike, 1973). O método AIC considera que os modelos apresentam melhor desempenho quanto mais simples (menor valor de AIC) for o modelo, portanto, o método impõe uma penalidade à complexidade.

A validação cruzada, segundo Isaaks & Srivastava (1989), é uma técnica de avaliação de erros de estimativas que permite comparar os valores previstos com os

amostrados. O valor da amostra, em certa localização  $Z(s_i)$ , é temporariamente descartado do conjunto de dados e então é feita uma previsão por krigagem no local com ausência de amostragem, usando-se as amostras restantes. Uma vez que a estimação é feita, pode-se compará-la ao valor da amostra que foi inicialmente removida do conjunto de dados amostrais. Este procedimento, método de “deixar um fora”, é repetido para todas as amostras disponíveis.

Os métodos de estimação de mínimos quadrados ordinários (OLS) busca minimizar a soma de quadrados das diferenças entre valores observados e estimados. Já o método de estimação de máxima verossimilhança considera um estimador de parâmetros assintoticamente normal, não viciado e eficiente (Cox & Hinkley, 1974) – consiste em maximizar a função de densidade de probabilidade conjunta do processo  $Z(s)$  em relação aos efeitos fixos (processo estacionário) e os componentes dos efeitos aleatórios. O melhor modelo para um processo será aquele que apresentar o maior valor de maximização do logaritmo da função verossimilhança.

O objetivo deste trabalho foi descrever os comportamentos espaciais de dados de carbono no solo pela seleção de modelos de variabilidade espacial, usando os métodos de estimação de mínimos quadrados ordinários (OLS), (Cressie, 1985) e máxima verossimilhança (MV) (Mardia & Marshall, 1984), segundo os critérios de Akaike, validação cruzada e máximo valor do logaritmo da função verossimilhança (MLL). O trabalho também apresenta os mapas temáticos utilizando estrutura de dependência espacial, escolhida segundo os critérios utilizados.

## **MATERIAL E MÉTODOS**

As amostras de solo foram coletadas na área experimental com dimensões de 40 x 96 m do instituto agrônomo de Campinas, em Piracicaba –SP, na camada de 0 a 20 cm de solo, nos anos de agrícolas de 1996/1997. O relevo é suave ondulado, e o solo, um Latossolo Vermelho Amarelo. As amostras foram coletadas em uma grade regular com 6 x 6 m. Para a instalação da grade de amostragem foi utilizando um sistema de coordenadas arbitrário X e Y, iniciando em um dos cantos da área.

Para análise da estrutura de dependência espacial da variável em estudo, utilizou-se um semivariograma experimental, construídos empregando-se o estimador

de Matheron. Com a finalidade de se aplicarem os critérios de validação em estudo, ajustaram-se três modelos teóricos ao semivariograma experimental: exponencial, powered exponencial, matern e gaussiano, considerados adequados aos dados em análise, com e sem aplicação de um modelo linear de tendência. Na estimação dos parâmetros foram usados os métodos: dos mínimos quadrados ordinários (OLS), (Cressie, 1985) e máxima verossimilhança (ML) (Mardia & Marshall, 1984). Para estimação e ajuste de modelos e avaliação de critérios de ajuste foi utilizado o software R versão 2.13.1, no qual foi utilizado o pacote geoR (Ribeiro Jr. & Diggle, 2001), ambos livres e de acordo com a licença GPL (General Public Licence).

Com a comparação entre os resultados alcançados, foram analisados os comportamentos dos critérios de Akaiked, validação cruzada, valor dos mínimos quadrados ordinários e maior valor de maximização do logaritmo da função verossimilhança. Finalmente, construíram-se os mapas temáticos do carbono no solo, segundo os modelos que apresentaram melhor ajuste utilizando-se os critérios de validação de modelos.

## RESULTADOS E DISCUSSÃO

No Quadro 1 são apresentadas as estatísticas descritivas do carbono no solo. Observa-se que a variável não sofreu variações acentuadas com uma amplitude entre os dados que variou de 6 a 10,7 g kg<sup>-1</sup>, estando a média em 8,221 g kg<sup>-1</sup>, próxima a mediana que foi de 8,200 g kg<sup>-1</sup>, isto é um indicativo de normalidade dos dados. O coeficiente de variação indica homogeneidade (CV menor que 10 %), segundo Gomes & Garcia (2002).

Quadro 1. Análise estatística descritiva do carbono no solo na camada de 0 a 20 cm de profundidade.

N	Min.	1st Qu.	Mediana	Média	3rd Qu.	Max.	DP	Var	CV (%)
g kg <sup>-1</sup>									
119	6,000	7,600	8,200	8,221	8,800	10,700	0,9341	0,8725	11,36

N: Numero de amostras; DP: Desvio Padrão; CV: Coeficiente de Variação;

Na Figura 1 está apresentada a análise exploratória dos dados, através da análise e interpretação de gráficos, pode-se detectar a presença de normalidade, outliers, tendências entre outras características que auxiliaram na estratégia a ser adotada para o ajuste do modelo que melhor represente a variabilidade espacial da variável.

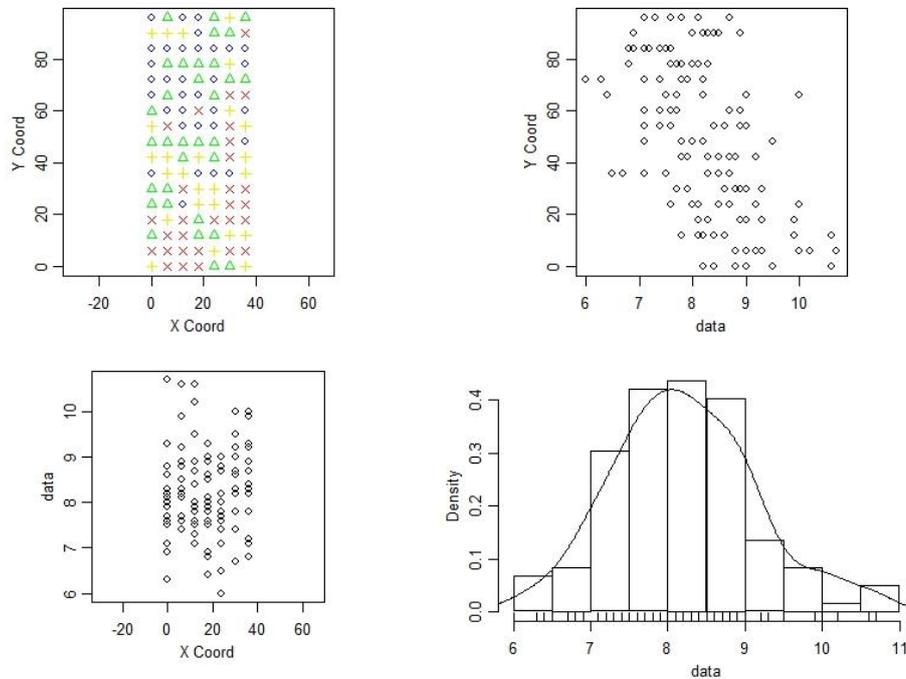


Figura 1. Análise exploratória para o teor de carbono no solo.

Através da Figura 1 pode se observar a distribuição espacial dos dados na área, coordenadas X e Y e a distribuição de densidade. A figura mostra que os dados possuem tendência no eixo x, embora esta não se apresente muito clara, optou-se aplicar um modelo linear para a retirada de tendência dos dados e a partir deste estimar o variograma experimental. O histograma dos dados mostra que a maior probabilidade de ocorrência dos valores está próxima a média e a mediana.

Nas Figuras 2 e 3 estão apresentados os variogramas experimentais com os modelos ajustados através do método dos mínimos quadrados ordinários e máxima verossimilhança, sendo a Figura 2 correspondente ao variograma experimental dos dados sem a aplicação do modelo de tendência linear e a Figura 3 apresenta o variograma experimental com a aplicação do modelo de tendência linear.

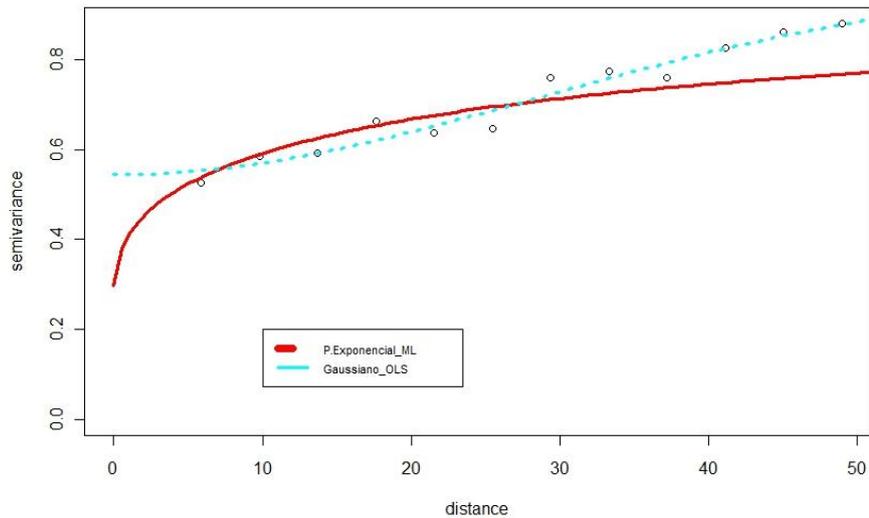


Figura 2. Semivariograma experimental dos dados sem aplicação do modelo linear de tendência e os modelos Powered exponencial e Gaussiano cujos parâmetros foram estimados respectivamente pelo método da máxima verossimilhança e mínimos quadrados ordinários.

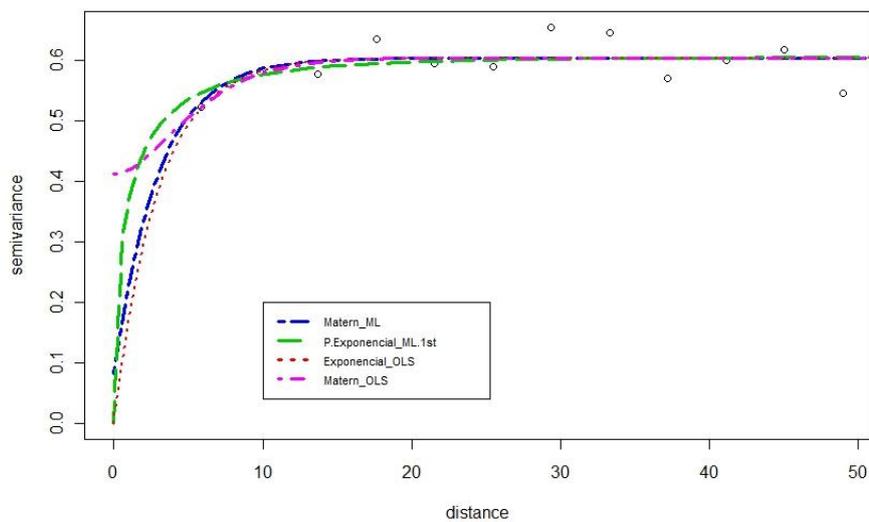


Figura 3. Semivariograma experimental dos dados com aplicação do modelo linear de tendência e os modelos Matern e Powered exponencial ajustado pela máxima verossimilhança e Exponencial e Matern ajustados mínimos quadrados ordinários.

O semivariograma experimental apresentado nas Figuras 2 segue uma tendência linear ascendente não atingindo o patamar onde a variância torna-se constante, nesta situação, a variância aumenta quanto maior for à distância, enquanto que o semivariograma da Figura 3 atinge o patamar próximo aos 10 m. A ausência de dados com distâncias menores que 6 m dificulta a estimativa do efeito pepita da variável, nesta situação opta-se por um modelo teórico que melhor se ajusta ao comportamento da

variável no solo, utilizando como referência a literatura específica, o conhecimento da área e do atributo.

No quadro 2 estão apresentados os parâmetros de ajuste dos modelos selecionados, calculados pelo método da máxima verossimilhança (ML), mínimos quadrados ordinários (OLS) e o alcance prático (AP).

Quadro 2. Parâmetros dos modelos semivariográficos, pelos métodos mínimos quadrados ordinários (OLS) e máxima verossimilhança (MV) para teor de carbono no solo

	<b>tausq</b>	<b>sigmasq</b>	<b>phi</b>	<b>AP</b>
<b>Matern_ML</b>	0,08	0,52	2,97	8,92
<b>P.Exponencial ML.1st</b>	0,00	0,61	1,08	9,75
<b>P.Exponencial ML</b>	0,30	0,62	23,97	215,17
<b>Exponencial OLS</b>	0,00	0,60	2,95	8,84
<b>Matern OLS</b>	0,41	0,19	2,20	13,07
<b>Gaussiano OLS</b>	0,54	0,43	40,06	69,34

Observa-se que os modelos ajustados através do variograma experimental com tendência obtiveram os maiores alcances práticos, enquanto que os modelos ajustados a partir de um variograma experimental sem tendência foi menor obtiveram um alcance prático menores, em torno de 10 m.

No quadro 3 e 4, encontram-se os resultados dos diferentes critérios de validação de ajustes dos modelos geoestatísticos. Segundo o critério de Akaike, o menor valor de AIC para o teor de carbono solo, aponta para o modelo matern estimado pelo método da máxima verossimilhança. Pelo máximo valor do logaritmo da função verossimilhança (ML), os valores são semelhantes em cada modelo, não se destacando nenhum modelo, embora dentre os modelos estimados pelo método da máxima verossimilhança o modelo matern tenha obtido melhor desempenho. Segundo o valor do mínimo quadrado ordinário o modelo gaussiano obteve o melhor desempenho.

Quando 3. Resultados dos critérios de AIC, valor máximo do logaritmo da função de máxima verossimilhança (ML) e valor do mínimo quadrado ordinários (MMS), para o teor de carbono no solo na camada de 0 a 20 cm.

	AIC	Log-like	MMS
<b>Matern_ML</b>	<b>286,79</b>	<b>-137,4</b>	-
<b>P.Exponencial ML.1st</b>	<b>292,98</b>	<b>-137,8</b>	-
<b>P.Exponencial ML</b>	<b>287,54</b>	<b>-142,5</b>	-
<b>Exponencial OLS</b>	-	-	<b>0,0107</b>
<b>Matern OLS</b>	-	-	<b>0,0107</b>
<b>Gaussiano OLS</b>	-	-	<b>0,0075</b>

A partir da técnica de validação cruzada, foi possível avaliar os erros na predição dos modelos geoestatísticos escolhidos. No Quadro 4 é apresentado os resultados da validação cruzada dos modelos selecionados. Em negrito é apresentado os modelos que obtiveram o melhor desempenho segundo esta análise. O modelo que apresentar menor resíduo tem um melhor desempenho.

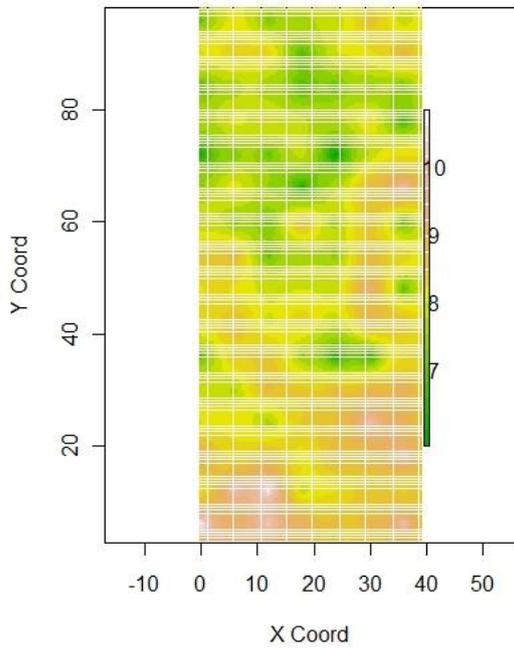
Pelo Quadro 4 é possível observar que o modelo matern estimado pelo método da máxima verossimilhança possui os menores resíduos, conseqüentemente menor desvio padrão, no entanto a diferença entre os modelos foi baixa.

Quadro 4. Resultado da validação cruzada dos modelos geoestatísticos escolhidos.

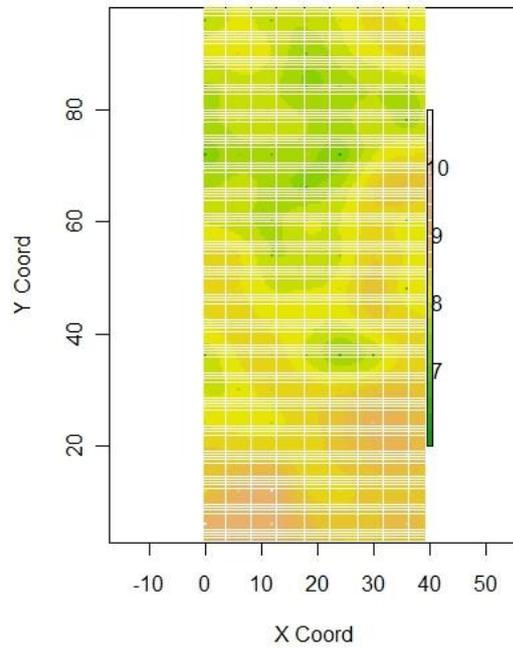
Validação Cruzada: Análise dos Resíduos								
		Min.	1st	Median	Mean	3rd	Max.	sd
Matern_ML	errors	-1,846	-0,5104	-0,02287	<b>0,000273</b>	0,5093	1,962	0,776
	std.errors	-2,428	-0,6683	-0,03009	<b>0,0001837</b>	0,6693	2,545	1,013
P.Exponencial_ML	errors	-1,803	-0,5293	-0,02007	-0,001273	0,5419	2,002	0,774
	std.errors	-2,366	-0,694	-0,02633	-0,0008072	0,7111	2,574	1,005
P.Exponencial_ML.1st	errors	-1,844	-0,4724	-0,03813	<b>0,0006268</b>	0,4926	1,992	0,777
	std.errors	-2,414	-0,617	-0,04985	<b>0,000411</b>	0,6439	2,574	1,009
Exponencial_OLS	errors	-1,843	-0,5255	-0,0234	0,04116	0,5328	1,955	0,78
	std.errors	-2,441	-0,6959	-0,03105	0,03466	0,7058	2,552	1,076
Matern_OLS	errors	-1,842	-0,5133	-0,02674	-0,02303	0,5118	1,954	0,776
	std.errors	-2,440	-0,6793	-0,03491	-0,007718	0,6778	2,548	1,019
Gaussiano_OLS	errors	-1,842	-0,5133	-0,02674	-0,02303	0,5118	1,954	0,776
	std.errors	-2,440	-0,6793	-0,03491	-0,007818	0,6778	2,548	1,019

Na figura 4 é apresentado os mapas temáticos obtidos através de krigagem ordinária, utilizando os modelos geoestatísticos selecionado.

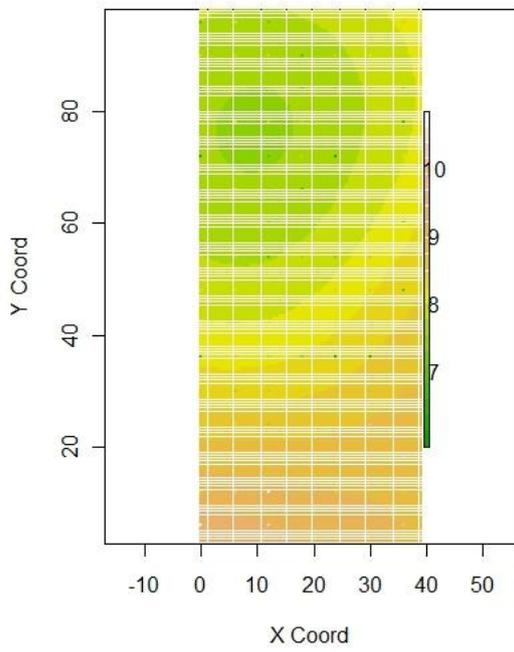
Modelo Exponencial OLS



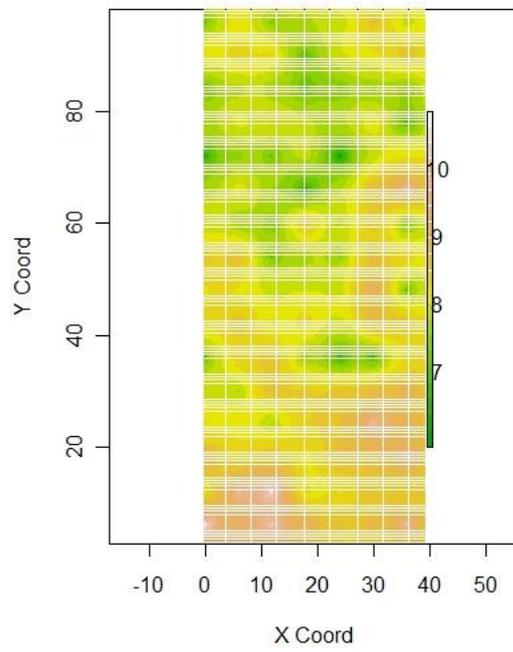
Modelo Matern OLS



Modelo Gaussiano OLS



Modelo Matern ML



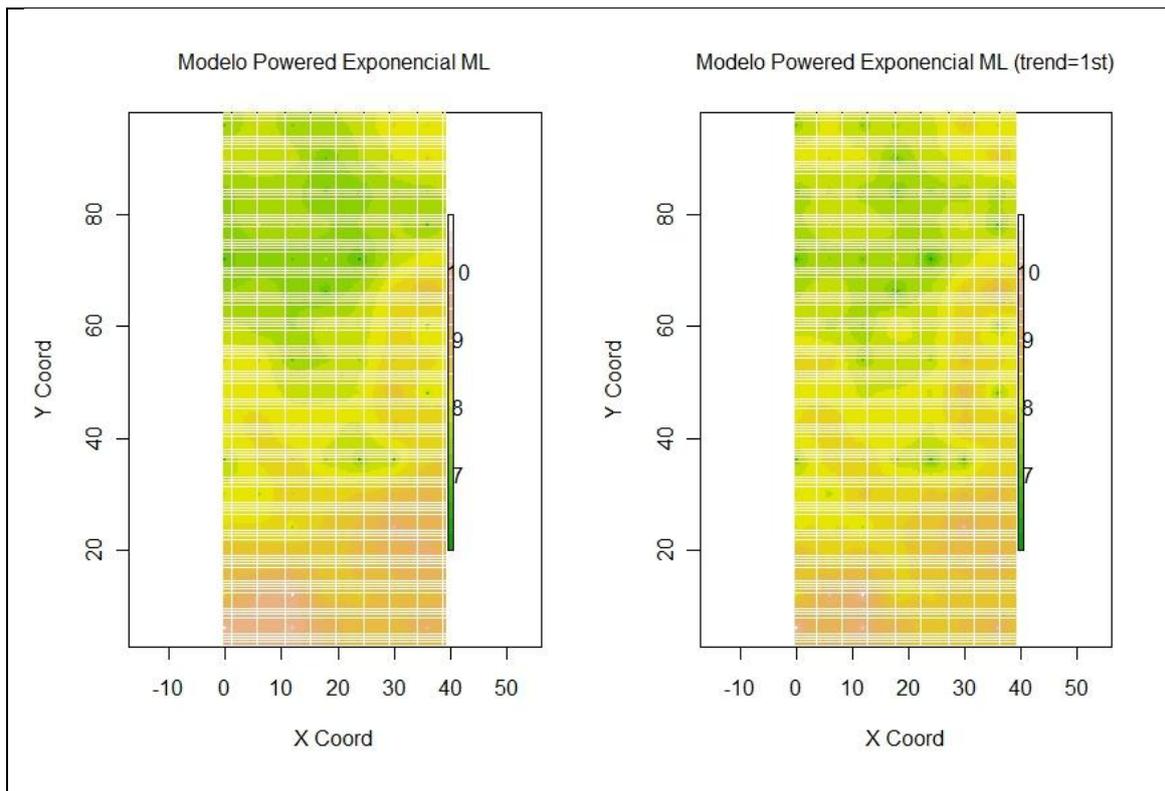


Figura 4. Mapas temáticos do teor de carbono no solo obtidos através de krigagem ordinária utilizando diversos modelos estatísticos.

Através da análise da Figura 4 observa-se que os modelos selecionados conseguiram captar de forma geral a tendência do teor de carbono no solo dentro da área. Observa-se que o modelo gaussiano obtido a partir dos dados com tendência espacial obteve uma maior suavização do atributo, pois o alcance estimado foi alto o que torna os valores próximos aos pontos coletados bastante semelhantes. Os outros modelos foram capazes de capturar a variação do teor de carbono no solo com menor generalidade, pois o alcance estimado foi menor, fazendo com que haja uma maior diferença a distâncias menores entre os pontos amostrados e estimados.

## CONCLUSÃO

Os critérios de seleção por validação cruzada, Akaike, máximo valor do logaritmo da função verossimilhança e valor mínimo dos mínimos quadrados ordinários não selecionam simultaneamente um mesmo modelo, devido em parte às distintas naturezas dos critérios. O critério de Akaike e do máximo valor do logaritmo da função

verossimilhança não foram conclusivos. O critério da validação cruzada através da análise dos resíduos foi considerado o mais adequado para a escolha do melhor ajuste.

Foram selecionados 6 modelos diferentes, sendo 3 obtidos através dos mínimos quadrados ordinários e 3 através da máxima verossimilhança. Pelo critério de Akaike o modelo matern (ML) obteve o melhor desempenho, pelo máximo valor do log da função de verossimilhança, o modelo exponencial (ML) obteve o melhor desempenho. Dos modelos ajustados pelo mínimo quadrado ordinário o modelo Gaussiano obteve o menor valor de MMS. Pela análise de resíduos, os modelos matern (ML) e powered exponencial com superfície de tendência obtiveram os menores resíduos.

À semelhança da construção de modelos empíricos de variabilidade espacial aos semivariogramas experimentais, em que o conhecimento do pesquisador relativo às variáveis em estudo conduz a resultados melhores, também no caso de verificação de qualidade dos ajustes o conhecimento e o bom senso do pesquisador são de grande valia.

## BIBLIOGRAFIA

- AKAIKE, H. Information theory as an extension of the maximum likelihood principle. In: INTERNATIONAL SYMPOSIUM ON INFORMATION THEORY, 2., Budapest, 1973. Proceedings. Budapest, Akadêmia Kiadó, 1973. p.267-281.
- COX, D.R. & HINKLEY, D.V. Theoretical statistics. London, Chapman and Hall, 1974. 532p.
- CRESSIE, N. Fitting variogram models by weighted least squares. *mathem. geol.*, 17:563-586, 1985.
- GOMES, F.P. & GARCIA, C.H. Estatística aplicada a experimentos agrônômicos e florestais. Piracicaba, FEALQ, 2002. V.11. 305P.
- ISAAKS, E.H. & SRIVASTAVA, R.M. an introduction to applied geostatistics. New York, Oxford University Press, 1989. 561P.
- MACHADO, P.L.O.A.; MADARI, R.B.B.; ALVES, B.J.R.; URQUIAGA, S. Os solos brasileiros e o seqüestro do carbono. *Sociedade Brasileira de Ciência do Solo: Boletim Informativo*, v. 29, n. 3, p. 21-25, 2004.
- MARDIA, K.; MARSHALL, R. Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, Alemanha, v.71, p.135-146, 1984.
- PANOSSO, A. R. Spatial and temporal variability of soil CO emission in a sugarcane area under green and slash-and-burn managements. *Soil & tillage research.*, n. 2, p. 275-282, 2009. ISSN 0167-1987.
- PAUSTIAN, K.; SIX, J.; ELLIOTT, E.T.; HUNT, H.W. Management options for redcing CO<sub>2</sub> missions from agricultural soils. *Biogeochemistry*, v.48, p.147-163, 2000.
- Lal, R. Soil carbon dynamics in cropland and rangeland. *Environmental Pollution*, v.116, p. 353-362, 2002.
- RIBEIRO JR., P.J.; DIGGLE P. J. geoR: A package for geostatistical analysis. *R-NEWS*, 01. Disponível em: <<http://cran.r-project.org/doc/Rnews>. 2001>. Acesso em: 3 jun. 2009.
- SILVA, I.R.; NOVAIS, R.F.; BARROS, N.F.; SILVA, E.F. Manejo de resíduos e matéria orgânica do solo em plantações de eucalipto: uma questão estratégia para a manutenção da sustentabilidade. *Sociedade Brasileira de Ciência do Solo: Boletim Informativo*, v. 29, n. 3, p. 10-20, 2004.