

# Escola Superior de Agricultura “Luiz de Queiroz” Universidade de São Paulo

## DISSERTAÇÃO DE MESTRADO

Bruno H F Fonseca - [bhffonse@esalq.usp.br](mailto:bhffonse@esalq.usp.br)

Estatística e Experimentação Agrônômica ESALQ/USP

Orientador: Paulo Justiniano Ribeiro Jr.

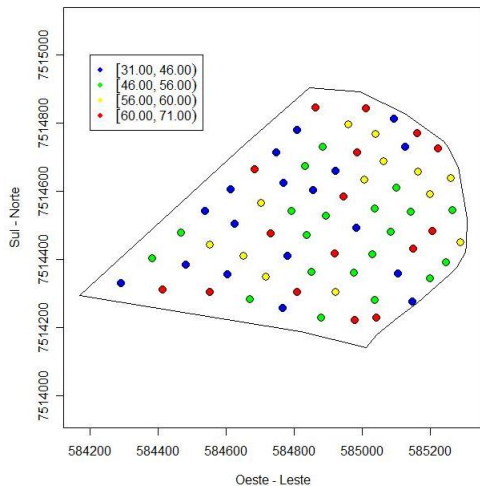
Laboratório de Estatística e Geoinformação LEG/UFPR

Piracicaba/2009

# UM ESTUDO SOBRE ESTIMAÇÃO E PREDIÇÃO EM MODELOS GEOESTATÍSTICOS BIVARIADOS

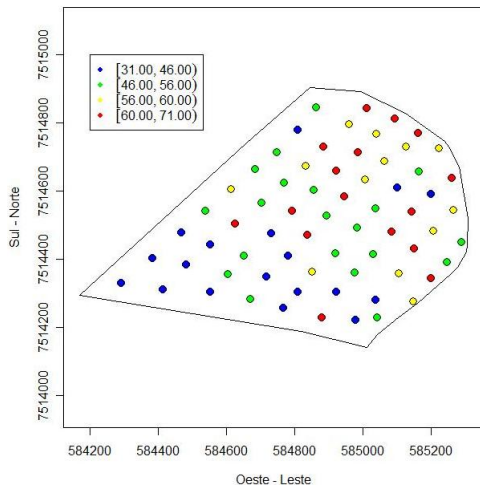
## Modelos lineares:

$$Y = X\beta + \epsilon,$$



Modelos geoestatísticos univariados:

$$Y = X\beta + Z + \epsilon,$$



Modelos geoestatísticos bivariados:

- Duas respostas de interesse
- Padrão espacial
- Forte correlação entre respostas
- Justificativa prática

- Revisão bibliográfica sobre modelos geoestatísticos bivariados
- Estudar a estimabilidade paramétrica dos modelos geoestatísticos bivariados
- Estudar a capacidade de predição espacial dos modelos geoestatísticos bivariados
- Avaliar a aplicabilidade de modelos geoestatísticos bivariados em dados observacionais
- Desenvolver computacionalmente a teoria dos modelos geoestatísticos bivariados

Uma notação para um campo aleatório  $Z$  na localização  $s_i$  pertencente ao espaço sob estudo  $G$  é:

$$(Z(s_i) : s_i \in G \subset R^d)$$

Segundo Schmidt e Sansó (2006) e Le e Zidek (2006), a descrição de um campo aleatório é obtida através da distribuição acumulada conjunta finita dimensional de  $Z(s)$ :

$$F_{s_1, s_2, \dots, s_n}(z_1, z_2, \dots, z_n) \equiv P(Z(s_1) \leq z_1, Z(s_2) \leq z_2, \dots, Z(s_n) \leq z_n)$$

sendo  $s = (s_1, s_2, \dots, s_n)$  um conjunto de localizações pertencentes a  $G$ .

# CAMPOS ALEATÓRIOS GAUSSIANOS

Assumindo que o campo aleatório é gaussiano, o processo aleatório é completamente especificado por:

- Vetor de média  $\mu = (\mu_1, \mu_2, \dots, \mu_n)^t$
- Matriz de covariâncias

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{1,2} & \cdots & \sigma_{1,n} \\ \sigma_{1,2} & \sigma_2^2 & \cdots & \sigma_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1,n} & \sigma_{2,n} & \cdots & \sigma_n^2 \end{bmatrix}$$

Sendo assim a distribuição conjunta de  $Z(s)$  é:

$$f_Z(z(s)) = (2\pi)^{-0.5n} |\Sigma|^{-0.5} \exp(-0.5(z(s) - \mu)^t \Sigma^{-1} (z(s) - \mu))$$



Em geral não é possível ter mais de uma realização do processo, sendo assim, outras suposições devem ser impostas sobre os campos aleatórios para análises inferenciais:

- Estacionariedade Forte
  - $f_s(z(s)) = f_{s+h}(z(s+h))$ , para qualquer incremento  $h$
- Isotropia

Dessa forma o campo aleatório é dito ser homogêneo.

Em alguns casos a suposição de homogeneidade não é válida:

- Problemas com a estacionariedade da média
  - Inclusão de efeito fixo no modelo
- Problemas com a estacionariedade da variância
  - Transformação de variáveis
- Problemas com a isotropia
  - Transformação do sistema de coordenadas

Se o campo aleatório é gaussiano e homogêneo,  $\Sigma$  deve:

- ser positiva definida
- assegurar o comportamento empírico utilizado em geoestatística

Estruturar  $\Sigma$  dessa forma não é trivial, no entanto:

- $Cov(Z(s_l), Z(s_k)) = \sigma^2 \rho(Z(s_l), Z(s_k))$ , para todo  $(s_l, s_k) \in s$
- existem funções de correlação que asseguram a validade de  $\Sigma$

Funções de correlação válidas:

- Família Matérn

$$\rho(h) = 2^{\kappa} - \Gamma(\kappa)^{-1} (h/\phi)^{\kappa} K_{\kappa}(h/\phi)$$

os parâmetros dessa função são  $\phi > 0$  e  $\kappa > 0$ , que são vinculados a escala com a dimensão de distância e suavidade do processo e  $K_{\kappa}(\cdot)$  é a função Bessel de ordem  $\kappa$ .

- Família Exponencial Potência

$$\rho(h) = \exp(-h/\phi)^{\kappa}$$

essa família também possui dois parâmetros, com as mesmas interpretações da família Matérn, no entanto agora  $\kappa$  limitado no intervalo  $[0, 2]$ .

# MODELO GEOESTATÍSTICO UNIVARIADO

Com as estruturas paramétricas definidas, pode-se pensar na modelagem:

$$Y(s) = X\beta + Z(s) + \epsilon, \quad s = (s_1, s_2, \dots, s_n), \quad (1)$$

- $Y(s)$  é o vetor dos valores observados da variável de interesse
- $X$  é a matriz de delineamento
- $\beta$  é o vetor de parâmetros associados a média da resposta
- $Z(s)$  é um campo gaussiano com vetor de médias nulo e  $\Sigma$  válida
- $\epsilon$  é o vetor dos ruídos brancos que são *i.i.d.* de uma  $N(0, \tau^2)$

Definida a modelagem, tem-se que:

$$Y(s) \sim N_n(\mu; \Sigma_Y),$$

- $\mu = X\beta$
- $\Sigma_Y = \Sigma + \tau^2 I$ ,  $I$  é uma matriz identidade.

Logo  $\theta = (\beta, \sigma^2, \phi^*, \tau^2)^t$  é o vetor paramétrico associado a (1). Sendo  $\phi^*$  o vetor de parâmetros associado a função de correlação adotada.

# ESTIMAÇÃO EM MODELO GEOESTATÍSTICO UNIVARIADO

O método da máxima verossimilhança é uma abordagem para estimar  $\theta$  e consiste em encontrar  $\hat{\theta}$  que maximize a função:

$$l(\theta; Y(s)) = -0.5[\ln(|\Sigma_Y|) + n \ln(2\pi) + (Y(s) - \mu)^t \Sigma_Y^{-1} (Y(s) - \mu)]$$

Diggle e Ribeiro Jr (2007) adotam a utilização da reparametrização  $\nu = \tau/\sigma$ , sendo assim:

- $\theta^* = (\beta, \sigma^2, \phi^*, \nu^2)^t$
- $\Sigma_Y = \sigma^2 V$

sendo  $V$  uma matriz que depende apenas de  $\theta_c = (\phi^*, \nu^2)^t$ .

# ESTIMAÇÃO EM MODELO GEOESTATÍSTICO UNIVARIADO

Estimadores de máxima verossimilhança com forma analítica:

- $\hat{\beta} = (X^t V^{-1} X)^{-1} (X^t V^{-1} Y(s))$
- $\hat{\sigma}^2 = n^{-1} (Y(s) - X\beta)^t V^{-1} (Y(s) - X\beta)$

As expressões de  $\hat{\beta}$  e  $\hat{\sigma}^2$  são funções de  $\theta_c$ , além das formas fechadas para os estimadores, é possível encontrar a matriz de informação de Fisher observada para os mesmos e portanto a matriz de covariâncias associada a  $(\hat{\beta}; \hat{\sigma}^2)^t$  (DIGGLE; RIBEIRO Jr., 2007).



# ESTIMAÇÃO EM MODELO GEOESTATÍSTICO UNIVARIADO

Utilizando  $\hat{\beta}$  e  $\hat{\sigma}^2$  na função de verossimilhança, encontra-se a função de verossimilhança concentrada, que depende apenas de  $\theta_c$ . Os métodos numéricos de estimação podem ser utilizados para encontrar:

- Estimativa de  $\theta_c$
- Estimativa da matriz Hessiana, denotada por  $H$

Logo, utilizando a propriedade de invariância dos estimadores de máxima verossimilhança pode-se encontrar  $\hat{\beta}$ ,  $\hat{\sigma}^2$  e  $\hat{\tau}^2$ .

# ESTIMAÇÃO EM MODELO GEOESTATÍSTICO UNIVARIADO

Utilizando as propriedades assintóticas dos estimadores de máxima verossimilhança e o método delta tem-se que:

- $\hat{\theta} \sim N(\theta; \Sigma_{\theta})$

sendo  $\Sigma_{\theta} = \Delta^t \Sigma_{\theta^*} \Delta$ , onde a  $i$ -ésima coluna de  $\Delta$  é o vetor  $\frac{\partial f(\theta_i)}{\partial \theta^*}$  e:

$$\Sigma_{\theta^*} = \begin{bmatrix} \Sigma_{\beta, \sigma^2} & O \\ O^t & \Sigma_{\theta_c} \end{bmatrix}$$

sendo  $\Sigma_{\beta, \sigma^2}$  a matriz de covariâncias de  $(\hat{\beta}, \hat{\sigma}^2)$ , que possui forma analítica,  $\Sigma_{\theta_c} = -H^{-1}$  é a matriz de covariâncias de  $\hat{\theta}_c$ , e  $O$  uma matriz de zeros, pois não é possível recuperar as covariâncias entre  $(\hat{\beta}, \hat{\sigma}^2)^t$  e  $\hat{\theta}_c$ .

# KRIGAGEM EM MODELO GEOESTATÍSTICO UNIVARIADO

Processo de predição do campo aleatório em localizações não amostradas, supondo que o interesse seja fazer predição de  $Z$  em  $s^*$ , tem-se que a distribuição de  $(Z(s^*)|Y(s))$  gera as predições com melhor precisão:

- $E(Z(s^*)|Y(s)) = \mu_Z + \Sigma_{Z,Y}\Sigma_Y^{-1}(Y(s) - X\beta)$
- $Var(Z(s^*)|Y(s)) = \Sigma_Z - \Sigma_{Z,Y}\Sigma_Y^{-1}\Sigma_{Z,Y}^t$

sendo  $\mu_Z$  a média de  $Z(s^*)$ ,  $\Sigma_{Z,Y}$  é a matriz de covariâncias cruzadas entre  $Z(s^*)$  e  $Y(s)$  e  $\Sigma_Z$  a matriz de covariâncias de  $Z(s^*)$ .

Quando existem dois campos aleatórios de interesse, pode-se pensar na modelagem:

$$Y_i = \mu_i + Z_i, \quad i = 1, 2 \quad (2)$$

- $Y_1$  e  $Y_2$  são os vetores dos valores observados das respostas
- $\mu_i$  é o vetor de médias associado a  $Y_i$  e pode incluir covariáveis
- $Z_i$  é um campo gaussiano com vetor de médias nulo e  $\Sigma_i$  válida

A distribuição de  $Y = (Y_1, Y_2)^t$  é de interesse,  $Y \sim N(\mu, \Sigma_Y)$ , sendo  $\mu = (\mu_1, \mu_2)^t$  e:

$$\Sigma_Y = \begin{bmatrix} \Sigma_1 & \Sigma_{1,2} \\ \Sigma_{1,2}^t & \Sigma_2 \end{bmatrix}$$

sendo  $\Sigma_i$  a matriz de covariância de  $Y_i$ ,  $i = 1, 2$  e  $\Sigma_{1,2}$  a matriz de covariâncias cruzadas entre as respostas.

A especificação do modelo deve assegurar que  $\Sigma_Y$  seja válida:

- Modelo gaussiano bivariado com componente de correlação parcialmente comum - *BGCCM*
- Modelo bivariado de co-regionalização - *BCRM*

Diggle e Ribeiro Jr. (2007) propuseram um modelo que decompõe os termos latentes de (2):

$$Z_i = \sigma_{0i}S_0 + \sigma_i S_i, \quad i = 1, 2, \quad (3)$$

- $\sigma^* = (\sigma_{01}, \sigma_1, \sigma_{02}, \sigma_2)^t$  um vetor de parâmetros de dispersão
- $S_0, S_1$  e  $S_2$  são campos aleatórios gaussianos mutuamente independentes, com vetores de médias nulos, variâncias unitárias e correlações determinadas por funções de correlação válidas

Com as decomposições (3) tem-se que (2) fica na forma:

$$\begin{cases} Y_1 = \mu_1 + \sigma_{01}S_0 + \sigma_1S_1 \\ Y_2 = \mu_2 + \sigma_{02}S_0 + \sigma_2S_2 \end{cases}$$

Utilizando propriedades básicas de covariância tem-se:

- $Var(Y_1) = \sigma_{01}^2 + \sigma_1^2$
- $Var(Y_2) = \sigma_{02}^2 + \sigma_2^2$
- $Cov(Y_{1,l}, Y_{1,k}) = \sigma_{01}^2\rho_0(h) + \sigma_1^2\rho_1(h)$
- $Cov(Y_{2,l}, Y_{2,k}) = \sigma_{02}^2\rho_0(h) + \sigma_2^2\rho_2(h)$
- $Cov(Y_{1,l}, Y_{2,k}) = \sigma_{01}\sigma_{02}\rho_0(h)$

sendo que os índices  $l$  e  $k$  referessem à duas localizações espaciais quaisquer sepradas pela distância  $h$ .

Sendo assim, a distribuição de probabilidade de  $Y = (Y_1, Y_2)^t$  está completamente estruturada e depende de:

$$\theta = (\beta^*, \sigma^*, \phi_0^*, \phi_1^*, \phi_2^*)^t$$

- $\beta^* = (\beta_1, \beta_2)^t$
- $\phi_0^*$ ,  $\phi_1^*$  e  $\phi_2^*$  parâmetros associados as escolhas das funções de correlação para  $S_0$ ,  $S_1$  e  $S_2$ , respectivamente.

Os processos de estimação e predição seguem a mesma técnica aplicada nos modelos univariados, porém existe a necessidade de um número maior de reparametrizações.



Gelfand et al. (2004) propôs outra abordagem para decompor os termos latentes de (2):

$$\begin{cases} Z_1 = \sigma_{11}S_1 \\ Z_2 = \sigma_{12}S_1 + \sigma_{22}S_2 \end{cases}$$

- $\sigma^* = (\sigma_{11}, \sigma_{12}, \sigma_{22})^t$  um vetor de parâmetros de dispersão
- $S_1$  e  $S_2$  são campos aleatórios gaussianos mutuamente independentes, com vetores de médias nulos, variâncias unitárias e correlações determinadas por funções de correlação válidas

Com as decomposições utilizadas no BCRM tem-se:

$$\begin{cases} Y_1 = \mu_1 + \sigma_{11}S_1 \\ Y_2 = \mu_2 + \sigma_{12}S_1 + \sigma_{22}S_2 \end{cases}$$

Utilizando propriedades básicas de covariância tem-se:

- $Var(Y_1) = \sigma_{11}^2$
- $Var(Y_2) = \sigma_{12}^2 + \sigma_{22}^2$
- $Cov(Y_{1,l}, Y_{1,k}) = \sigma_{11}^2\rho_1(h)$
- $Cov(Y_{2,l}, Y_{2,k}) = \sigma_{12}^2\rho_1(h) + \sigma_{22}^2\rho_2(h)$
- $Cov(Y_{1,l}, Y_{2,k}) = \sigma_{11}\sigma_{12}\rho_1(h)$

sendo que os índices  $l$  e  $k$  referessem à duas localizações espaciais quaisquer separadas pela distância  $h$ .

Sendo assim, a distribuição de probabilidade de  $Y = (Y_1, Y_2)^t$  está completamente estruturada e depende de:

$$\theta = (\beta^*, \sigma^*, \phi_1^*, \phi_2^*)^t$$

Os processos de estimação e predição seguem a mesma técnica aplicada nos modelos univariados, porém existe a necessidade de um número maior de reparametrizações.

Com intuito de comparar as capacidades de estimação sob o enfoque frequentista e de predição dos modelos bivariados, a pesquisa foi dividida em duas etapas:

- Estudo de Simulação
- Análise de Dados de Solos

Cabe ressaltar que todas análises foram conduzidas no ambiente R de programação (*R Development Core Team, 2006*), utilizando o pacote *geoR* (*Ribeiro Jr. e Diggle, 2001*). No entanto, para os modelos bivariados a maioria dos métodos utilizados não está implementada nos pacotes do R.

No estudo de simulação foi lançado mão das seguintes características nas duas abordagens de modelos bivariados:

- Três configurações amostrais de localizações espaciais
  - $n_1 = n_2 = 100$  e dados completamente co-locados
  - $n_1 = 100$  e  $n_2 = 50$  e dados completamente co-locados
  - $n_1 = n_2 = 100$  e apenas metade dos dados co-locados
- Simulação de 1000 conjuntos de dados em cada configuração
- Divisão dos dados simulados em um grupo de controle e um grupo de modelagem

Na simulação do BGCCM foram utilizadas as seguintes configurações paramétricas:

- $\mu_1 = 150$  e  $\mu_2 = 60$
- $\sigma_{01} = 8$ ,  $\sigma_1 = 4$ ,  $\sigma_{02} = 5$  e  $\sigma_2 = 2$
- Função de Matèrn foi adotada para as três funções de correlação
  - $\phi_0 = 0,25$ ,  $\phi_1 = 0,2$  e  $\phi_2 = 0,2$
  - $\kappa_0 = 0,5$ ,  $\kappa_1 = 0,5$  e  $\kappa_2 = 0,5$

Na simulação do BCRM foram levadas em consideração características de variabilidade e correlação parecidas com as do BGCCM:

- $\mu_1 = 150$  e  $\mu_2 = 60$
- $\sigma_{11} = 9$ ,  $\sigma_{12} = 5$  e  $\sigma_{22} = 2$
- Função de Matèrn foi adotada para as três funções de correlação
  - $\phi_1 = 0,25$  e  $\phi_2 = 0,2$
  - $\kappa_1 = 0,5$  e  $\kappa_2 = 0,5$

Para calcular as estimativas por máxima verossimilhança utilizou-se as seguintes reparametrizações:

- BGCCM

- $\sigma = \sigma_{01}$
- $\eta = \sigma_{02}/\sigma_{01}$
- $\nu_1 = \sigma_1/\sigma_{01}$
- $\nu_2 = \sigma_2/\sigma_{01}$

- BCRM

- $\sigma = \sigma_{11}$
- $\nu_1 = \sigma_{12}/\sigma_{11}$
- $\nu_2 = \sigma_{22}/\sigma_{11}$

Com essas reparametrizações é possível encontrar a função de máxima verossimilhança concentrada. Logo existe forma analítica para os parâmetros de média e para  $\sigma$ , os demais parâmetros foram estimados pelo método de Nelder e Mead (1965). E novamente as propriedades assintóticas e de invariância foram utilizadas para encontrar a distribuição de probabilidade do vetor de parâmetros.



Para avaliar a qualidade das estimativas foram analisados:

- erros absolutos médios
- erros quadráticos médios

Cada vetor de parâmetros estimados foram utilizados para fazer a predição nas localizações do grupo de controle:

- intervalos de confiança marginais de predição
- erros absolutos médios
- erros quadráticos médios

- Pesquisa realizada em 2006
- Propriedade agrícola com 51,8 ha
- Solo argissolo com textura média
- Dois históricos de manejo
- 67 localizações amostradas com grid regular
- Coordenadas no sistema UTM
- Variáveis químicas analisadas: pH e saturação por bases
- Forte correlação positiva entre respostas
- Justificativa prática para modelos bivariados

# RESULTADOS E DISCUSSÕES - SIMULAÇÃO

$\theta$	mínimo	1º quartil	mediana	média	3º quartil	máximo
$\beta_{01}$	137,03	147,39	150,01	149,99	152,48	161,82
$\beta_{02}$	52,18	58,56	60,02	59,98	61,43	66,32
$\sigma_{01}$	4,13	6,24	7,22	7,41	8,35	16,47
$\sigma_1$	> 0,001	2,14	3,58	3,18	4,29	7,81
$\sigma_{02}$	2,12	3,82	4,44	4,58	5,17	9,11
$\sigma_2$	> 0,001	0,38	1,65	1,54	2,40	5,20
$\phi_0$	0,05	0,14	0,19	0,22	0,27	1,00
$\phi_1$	> 0,001	0,11	0,15	0,19	0,22	1,00
$\phi_2$	> 0,001	0,11	0,17	0,21	0,28	1,00

**TABELA:** Estatísticas descritivas das estimativas por máxima verossimilhança dos parâmetros do BGCCM, usando as amostras simuladas do BGCCM com dados co-locados e balanceados

# RESULTADOS E DISCUSSÕES - SIMULAÇÃO

$\theta$	mínimo	1º quartil	mediana	média	3º quartil	máximo
$\beta_{01}$	138,39	147,67	150,07	149,99	152,45	162,05
$\beta_{02}$	53,86	58,63	60,03	59,98	61,37	67,17
$\sigma_{11}$	5,19	7,40	8,16	8,41	9,12	16,00
$\sigma_{12}$	2,69	4,10	4,54	4,68	5,14	9,55
$\sigma_{22}$	1,22	1,66	1,84	1,88	2,05	3,21
$\phi_1$	0,07	0,16	0,20	0,22	0,26	0,79
$\phi_2$	0,06	0,13	0,16	0,18	0,22	0,58

**TABELA:** Estatísticas descritivas das estimativas por máxima verossimilhança dos parâmetros do BCRM, usando as amostras simuladas do BCRM com dados co-locados e balanceados

# RESULTADOS E DISCUSSÕES - SIMULAÇÃO

$\theta$	mínimo	1º quartil	mediana	média	3º quartil	máximo
$\beta_{01}$	0,78	4,97	8,58	12,03	15,10	137,48
$\beta_{02}$	0,40	1,81	3,06	4,40	5,55	43,38
$\sigma_{01}$	0,04	0,10	0,13	0,14	0,17	0,68
$\sigma_1$	$> 0,001$	0,04	1,93	28,97	9,63	2467,42
$\sigma_{02}$	0,01	0,07	0,33	4,78	1,67	386,51
$\sigma_2$	$> 0,001$	0,12	0,42	6,93	1,92	797,07
$\phi_0$	0,00	0,00	0,002	0,01	0,009	2,48
$\phi_1$	0,00	0,00	0,003	0,04	0,01	2,56
$\phi_2$	0,00	0,00	0,002	0,07	0,02	7,70

**TABELA:** Estatísticas descritivas das variâncias das estimativas por máxima verossimilhança dos parâmetros do BGCCM, usando as amostras simuladas do BGCCM com dados co-locados e balanceados

# RESULTADOS E DISCUSSÕES - SIMULAÇÃO

$\theta$	mínimo	1º quartil	mediana	média	3º quartil	máximo
$\beta_{01}$	0,96	5,42	8,55	12,36	14,22	106,43
$\beta_{02}$	0,59	2,07	3,10	4,31	5,03	36,35
$\sigma_{11}$	0,07	0,14	0,17	0,18	0,21	0,64
$\sigma_{12}$	0,04	0,08	0,09	0,10	0,11	0,37
$\sigma_{22}$	0,04	0,13	0,19	0,23	0,29	1,41
$\phi_1$	> 0,001	0,003	0,005	0,02	0,01	0,48
$\phi_2$	> 0,001	0,002	0,003	0,007	0,07	0,15

**TABELA:** Estatísticas descritivas das variâncias das estimativas por máxima verossimilhança dos parâmetros do BCRM, usando as amostras simuladas do BCRM com dados co-locados e balanceados

# RESULTADOS E DISCUSSÕES - SIMULAÇÃO

Erro	$\beta_{01}$	$\beta_{02}$	$\sigma_{01}$	$\sigma_1$	$\sigma_{02}$	$\sigma_2$	$\phi_0$	$\phi_1$	$\phi_2$
AM	2,98	1,71	1,44	1,41	0,93	1,01	0,094	0,102	0,116
QM	14,04	4,79	3,17	3,65	1,27	1,52	0,015	0,021	0,024

**TABELA:** Erros absolutos médios e erros quadráticos médios das estimativas por máxima verossimilhança dos parâmetros do BGCCM, usando as amostras simuladas do BGCCM com dados co-locados e balanceados

Erro	$\beta_{01}$	$\beta_{02}$	$\sigma_{11}$	$\sigma_{12}$	$\sigma_{22}$	$\phi_0$	$\phi_1$
AM	2,89	1,67	1,33	0,76	0,26	0,08	0,06
QM	13,17	4,46	2,66	0,88	0,10	0,01	0,005

**TABELA:** Erros absolutos médios e erros quadráticos médios das estimativas por máxima verossimilhança dos parâmetros do BCRM, usando as amostras simuladas do BCRM com dados co-locados e balanceados

# RESULTADOS E DISCUSSÕES - SIMULAÇÃO

$\theta$	mínimo	1º quartil	mediana	média	3º quartil	máximo
$EAM_{Y_1}$	1,79	3,29	3,78	3,83	4,30	6,21
$EAM_{Y_2}$	1,14	2,00	2,30	2,31	2,58	4,06
$EQM_{Y_1}$	5,67	17,29	22,24	23,56	28,55	54,45
$EQM_{Y_2}$	2,05	6,32	8,13	8,50	10,20	21,95

**TABELA:** Estatísticas descritivas dos erros de krigagens com as estimativas paramétricas do BGCCM, usando as amostras simuladas do BGCCM com dados co-locados e balanceados

$\theta$	mínimo	1º quartil	mediana	média	3º quartil	máximo
$EAM_{Y_1}$	1,35	3,24	3,71	3,76	4,21	6,87
$EAM_{Y_2}$	1,00	2,00	2,27	2,29	2,58	3,75
$EQM_{Y_1}$	3,06	16,52	21,40	22,68	27,33	57,87
$EQM_{Y_2}$	1,64	6,22	7,97	8,38	10,04	21,09

**TABELA:** Estatísticas descritivas dos erros de krigagens com as estimativas paramétricas do BCRM, usando as amostras simuladas do BCRM com dados co-locados e balanceados



# RESULTADOS E DISCUSSÕES - SIMULAÇÃO

$\theta$	mínimo	1º quartil	mediana	média	3º quartil	máximo
$\overline{EP}_{Y_1}$	3,70	4,52	4,74	4,75	4,99	5,83
$\overline{EP}_{Y_2}$	2,21	2,71	2,84	2,85	2,98	3,68

**TABELA:** Estatísticas descritivas das médias dos erros padrões de krigagens com as estimativas paramétricas do BGCCM, usando as amostras simuladas do BGCCM com dados co-locados e balanceados

$\theta$	mínimo	1º quartil	mediana	média	3º quartil	máximo
$\overline{EP}_{Y_1}$	3,80	4,46	4,68	4,70	4,93	5,93
$\overline{EP}_{Y_2}$	2,26	2,71	2,86	2,86	3,00	3,54

**TABELA:** Estatísticas descritivas das médias dos erros padrões de krigagens com as estimativas paramétricas do BGCCM, usando as amostras simuladas do BGCCM com dados co-locados e balanceados

- As configurações das localizações não afeta muito
- O BCRM se mostrou um pouco mais eficiente
- o BGCCM se mostrou um pouco mais robusto

# RESULTADOS E DISCUSSÕES - DADOS DE SOLOS

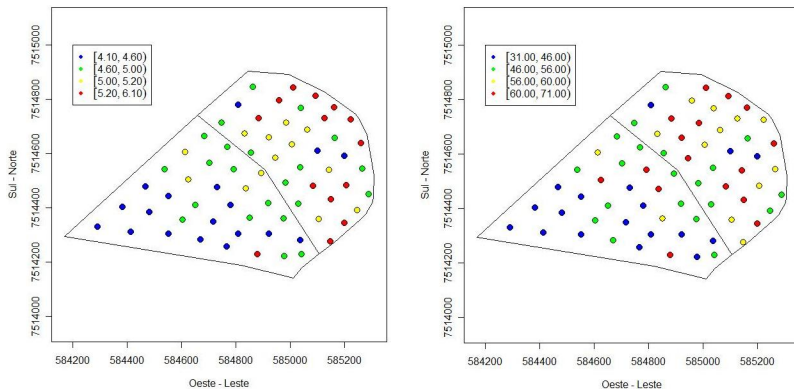


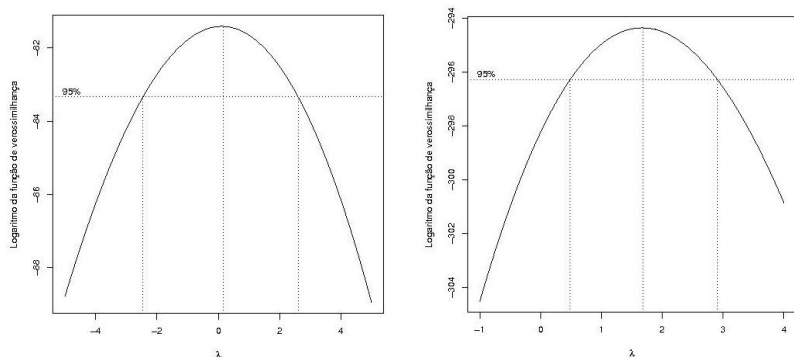
FIGURA: Gráficos de círculos - o gráfico a esquerda é para a variável pH e o da direita é da saturação por bases

# RESULTADOS E DISCUSSÕES - DADOS DE SOLOS

Variável	Mediana	Média	D.P.
Saturação	56,00	53,27	10,05
pH	5,00	4,94	0,42

TABELA: Médias, medianas e desvios padrões do pH e da saturação por bases

# RESULTADOS E DISCUSSÕES - DADOS DE SOLOS



**FIGURA:** Gráficos de possíveis transformações de variáveis de Box-Cox, o gráfico da esquerda é relativo ao pH e o gráfico da direita é relativo à saturação por bases

# RESULTADOS E DISCUSSÕES - DADOS DE SOLOS

$\beta$	$\tau^2$	$\sigma^2$	$\phi$	$\kappa$	$l(\theta)$	AIC
4,907	0,077	0,163	560	0,5	-25,49	58,98
4,905	0,101	0,149	510,6	1	-25,69	59,38
4,904	0,106	0,251	650,5	1,5	-25,57	59,15
4,903	0,106	0,288	569,7	2	-25,45	58,90
4,903	0,106	0,297	491,1	2,5	-25,36	58,73

**TABELA:** Estimativas de máxima verossimilhança dos parâmetros associados ao modelo para o pH com média constantes

# RESULTADOS E DISCUSSÕES - DADOS DE SOLOS

$\beta_0$	$\beta_1$	$\tau^2$	$\sigma^2$	$\phi$	$\kappa$	$l(\theta)$	AIC
4,725	0,397	0	0,125	49,19	0,5	-23,44	56,89
4,723	0,402	0	0,125	33,42	1	-23,35	56,70
4,723	0,403	0,014	0,111	28,43	1,5	-23,33	56,66
4,779	0,227	0,110	0,061	349,3	2	-24,86	59,73
4,786	0,204	0,109	0,086	353,1	2,5	-24,83	59,65

**TABELA:** Estimativas de máxima verossimilhança dos parâmetros associados ao modelo para o pH com média induzida pela área de manejo

# RESULTADOS E DISCUSSÕES - DADOS DE SOLOS

$\beta_0$	$\beta_1$	$\tau^2$	$\sigma^2$	$\phi$	$\kappa$	$l(\theta)$	AIC
-607,48	0,001	0	0,111	49,32	0,5	-19,66	49,31
-608,51	0,001	0	0,112	34,28	1	-19,44	48,89
-608,33	0,001	0	0,112	27,58	1,5	-19,39	48,77
-608,36	0,001	0,008	0,104	24,53	2	-19,36	48,73
-608,46	0,001	0,015	0,097	22,46	2,5	-19,35	48,70

**TABELA:** Estimativas de máxima verossimilhança dos parâmetros associados ao modelo para o pH com média induzida pelas coordenadas oeste-leste



# RESULTADOS E DISCUSSÕES - DADOS DE SOLOS

$\beta$	$\tau^2$	$\sigma^2$	$\phi$	$\kappa$	$l(\theta)$	AIC
49,53	47,66	105,33	700,00	0,5	-239,53	487,05
48,53	59,57	120,62	625,58	1	-239,76	487,51
47,98	62,40	124,31	516,78	1,5	-239,79	487,57
47,27	63,29	149,69	489,58	2	-239,75	487,50
47,18	63,51	139,18	400,00	2,5	-239,70	487,40

**TABELA:** Estimativas de máxima verossimilhança dos parâmetros associados ao modelo para a saturação por bases com média constantes

# RESULTADOS E DISCUSSÕES - DADOS DE SOLOS

$\beta_0$	$\beta_1$	$\tau^2$	$\sigma^2$	$\phi$	$\kappa$	$l(\theta)$	AIC
47,91	8,80	24,35	53,68	73,68	0,5	-238,05	486,10
47,92	8,83	38,78	39,28	59,18	1	-238,01	486,01
47,92	8,86	43,35	34,69	50,33	1,5	-237,98	485,96
47,92	8,88	45,53	32,47	44,34	2	-237,96	485,92
47,93	8,89	46,79	31,18	39,98	2,5	-237,95	485,90

**TABELA:** Estimativas de máxima verossimilhança dos parâmetros associados ao modelo para a saturação por bases com média induzida pela área de manejo

# RESULTADOS E DISCUSSÕES - DADOS DE SOLOS

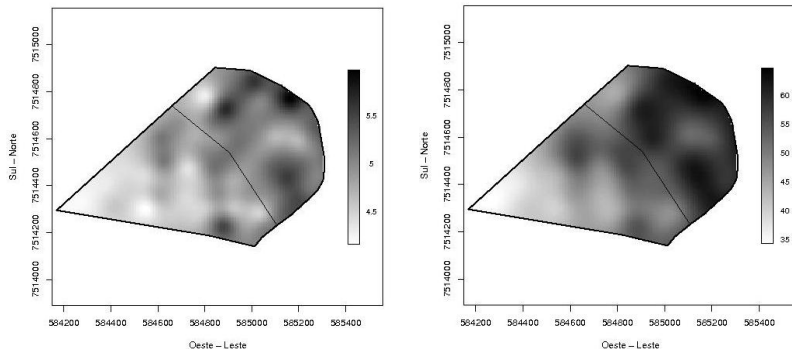
$\beta_0$	$\beta_1$	$\tau^2$	$\sigma^2$	$\phi$	$\kappa$	$l(\theta)$	AIC
-14297,01	0,025	8,98	59,98	54,51	0,5	-234,75	479,50
-14300,66	0,025	27,56	41,46	45,42	1	-234,72	479,44
-14300,92	0,025	33,53	35,53	39,51	1,5	-234,70	479,39
-14300,17	0,025	36,39	32,68	35,33	2	-234,68	479,36
-14298,98	0,025	38,06	31,03	32,19	2,5	-234,67	479,34

**TABELA:** Estimativas de máxima verossimilhança dos parâmetros associados ao modelo para a saturação por bases com média induzida pelas coordenadas oeste-leste

Utilizando o AIC e os máximos estimados das funções de vereossimilhança tem-se os modelos univariados finais:

- Média induzida pelas coordenadas oeste-leste
- parâmetros de suavidade da função de correlação igual a 2,5

# RESULTADOS E DISCUSSÕES - DADOS DE SOLOS



**FIGURA:** Gráficos de predições espaciais, o gráfico da esquerda é relativo ao pH e o gráfico da direita é relativo à saturação por bases

# RESULTADOS E DISCUSSÕES - DADOS DE SOLOS

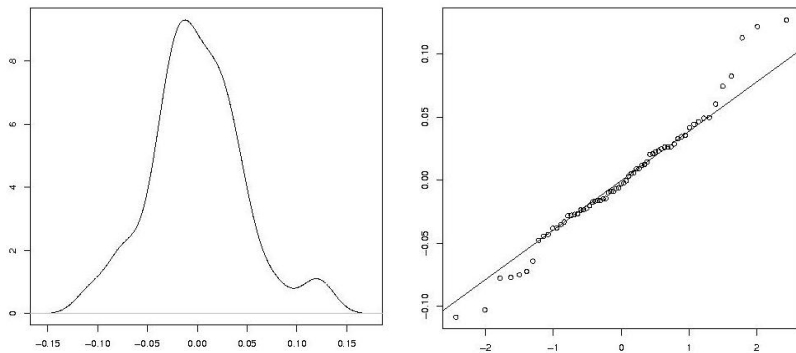
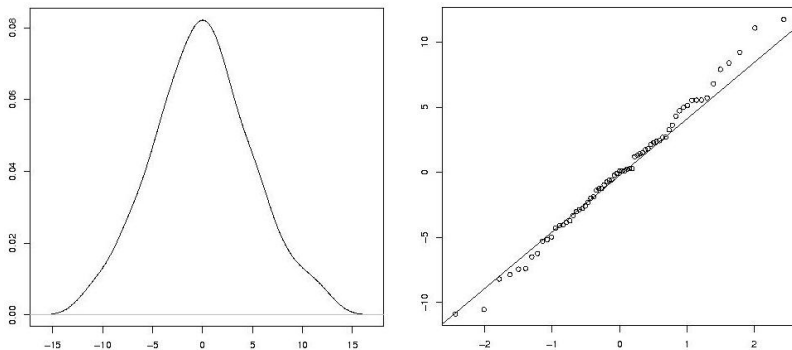


FIGURA: Densidade e gráfico de quartis dos resíduos para o modelo final do pH

# RESULTADOS E DISCUSSÕES - DADOS DE SOLOS



**FIGURA:** Densidade e gráfico de quartis dos resíduos para o modelo final da saturação por bases

# RESULTADOS E DISCUSSÕES - DADOS DE SOLOS

$\theta$	Constantes	Oeste-leste	Área de manejo
$\beta_{01}$	52,57	1,54e - 6	47,83
$\beta_{02}$	4,92	1,45e - 7	4,72
$\beta_1$	—	8,98e - 5	9,46
$\beta_2$	—	8,41e - 6	0,40
$\sigma_{01}$	8,51	8,35	6,85
$\sigma_1$	3,47	3,46	3,44
$\sigma_{02}$	0,38	0,38	0,31
$\sigma_2$	7e - 5	5e - 4	1e - 5
$\phi_0$	59,37	57,79	26,73
$\phi_1$	46,35	48,15	49,08
$\phi_2$	66,82	71,59	86,95
$\kappa_0$	1	1	1,5
$\kappa_1$	0,5	0,5	0,5
$\kappa_2$	0,5	0,5	0,5

TABELA: Estimativas de máxima verossimilhança dos parâmetros associados ao BGCCM para a saturação por bases e o pH

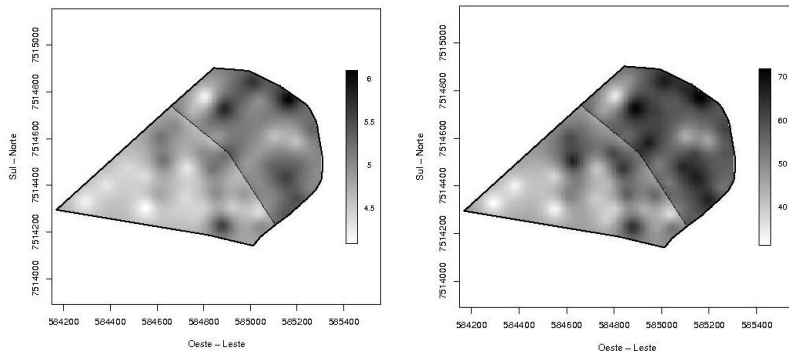


# RESULTADOS E DISCUSSÕES - DADOS DE SOLOS

$\theta$	Constantes	Oeste-leste	Área de Manejo
$\beta_{01}$	51,76	$1,51e - 6$	47,93
$\beta_{02}$	4,89	$1,43e - 7$	4,72
$\beta_1$	—	$8,85e - 5$	9,12
$\beta_2$	—	$8,36e - 6$	0,40
$\sigma_{11}$	8,26	99,53	7,82
$\sigma_{12}$	0,31	3,67	0,28
$\sigma_{22}$	0,13	1,54	0,14
$\phi_1$	100,63	100,21	53,32
$\phi_2$	21,43	21,38	19,34
$\kappa_1$	0,5	0,5	0,5
$\kappa_2$	2,5	2,5	2,5

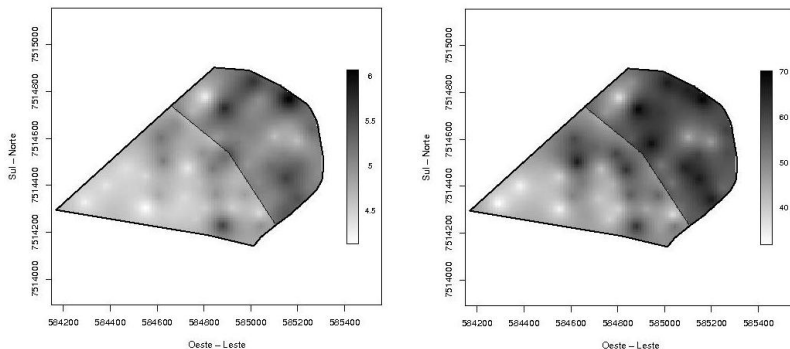
**TABELA:** Estimativas de máxima verossimilhança dos parâmetros associados ao BCRM para a saturação por bases e o pH

# RESULTADOS E DISCUSSÕES - DADOS DE SOLOS



**FIGURA:** Gráficos de predições espaciais com o BGCCM, o gráfico da esquerda é relativo ao pH e o gráfico da direita é relativo à saturação por bases

# RESULTADOS E DISCUSSÕES - DADOS DE SOLOS



**FIGURA:** Gráficos de predições espaciais com o BCRM, o gráfico da esquerda é relativo ao pH e o gráfico da direita é relativo à saturação por bases

# RESULTADOS E DISCUSSÕES - DADOS DE SOLOS

Obs.	Univ.	BGCCM	BCRM	Obs.	Univ.	BGCCM	BCRM
31	37,98	36,70	36,41	47	46,37	47,39	46,98
52	50,32	48,47	48,75	67	58,89	70,08	69,86
41	38,73	43,29	42,70	57	54,88	59,06	58,96
51	42,33	50,60	50,67	56	57,62	52,66	52,93
41	56,76	39,96	40,90	59	60,97	54,68	55,20
43	53,04	46,20	46,51	57	56,83	56,93	57,17
39	44,37	43,18	42,65	52	57,79	52,64	52,95
51	56,27	48,48	48,80	61	56,38	61,32	61,35
64	46,88	68,33	68,03	59	56,07	54,46	54,30
53	47,06	53,68	52,72	54	62,32	56,91	57,16

**TABELA:** Valores observados da saturação por bases e não utilizados na modelagem e suas respectivas predições utilizando os modelo univariado, o BGCCM e o BCRM

Estatística	Univariado	BGCCM	BCRM
Média	-0,34	-0,50	-0,50
Desvio Padrão	7,52	3,02	2,84

**TABELA:** Médias e desvios padrões dos erros de krigagem da saturação por bases nas 20 localizações omitidas no processo de estimação

# CONSIDERAÇÕES FINAIS

- Estudo de simulação
  - Estimativas precisas nas três configurações amostrais
  - Subestimação moderada dos parâmetros de variabilidade e correlação
  - Alta eficiência para predição espacial
  - Não existe vantagem evidente na escolha do modelo bivariado
  - Instabilidade computacional para estimação paramétrica
- Análise dos dados de solo
  - Os modelos bivariados foram eficientes na estimação e na predição
  - Possível barateamento de custos em análises futuras
- Sugestões de trabalhos
  - Estudo de simulação com outras configurações paramétricas
  - Estudo de outras formas para atribuir distribuições conjuntas
  - Estudo de técnicas bayesianas em conjunto com estimação MCMC

AZZALINI, A. **Statistical inference:** based on the likelihood. London: Chapman and Hall, 1996. 341 p.

BANERJEE, S.; CARLIN B.P.; GELFAND, A.E. **Hierarchical modelling and analysis for spatial data.** Boca Raton: Chapman and Hall, 2004. 453 p.

BANERJEE, S.; GELFAND, A.E. Predict, Interpolation and regression for spatial misaligned data points. **Sankhya**, Kolkata, v.64, p.227-245, 2002.

BICKEL, P.J.; DOKSUM K.A. **Mathematical statistics:** basic ideas and selected topics. San Francisco: Holden-Day, 1977. 492 p.

BOGNOLA, I. A.; RIBEIRO JR, P. J.; SILVA, E. A. A.; LINGNAU, C.; HIGA, A. R. Modelagem uni e bivariada da variabilidade espacial de rendimento de Pinus taeda L.. **Floresta**, Curitiba, v. 38, p. 373-385, 2008.

## REFERÊNCIAS

BROWN, P.J.; LE, N.D.; ZIDEK, J.V. Multivariate spatial interpolation and exposure to air pollutants. **The Canadian Journal of Statistics**, Vancouver, v.22, p.489-509, 1994.

CHRISTENSEN, W.F.; AMEMIYA, Y. Latent variable analysis of multivariate spatial data. **Journal of the American Statistical Association**, Alexandria, v.97, p.302-317, 2002.

CHRISTENSEN, W.F.; DIGGLE, P.J.; RIBEIRO Jr., P.J. Analysing positive-valued spatial data: the transformed gaussian model. **Geostatistics for environmental applications**, Boston, v.11, p.287-298, 2001.

DIGGLE, P.J.; RIBEIRO Jr., P.J. Bayesian inference in Gaussian model based geostatistics. **Geographical and Environmental Modelling**, New Jersey, v.6, p.129-146, 2002.

DIGGLE, P.J.; RIBEIRO Jr., P.J. **Model-Based geostatistics**. New York: Springer, 2007. 230p.



FARACO, M.A.; URIBE-OPAZO, M.A.; SILVA E.A.A.; JOHANN J.A.; BORSSOI J.A. Nonstationary multivariate process modeling through spatially varying coregionalization. **Revista Brasileira de Ciências do Solo**, Viçosa, v.32, p.463-476, 2008.

FERNANDES, M.V.M. **Modelos para Processos Espaço-Temporais Inflacionados de Zeros**, 2006. 128 p. Dissertação (Mestrado em Estatística) - Instituto de Matemática da Univeridade Federal do Rio de Janeiro, 2006.

FINLEY, A.O.; BANERJEE S.; CARLIN B.P. spBayes: an R package for univariate and multivariate hierarchical point-referenced spatial models. **Journal of Statistical Software**, Alexandria, v. 19, p.1-24, 2007.

FUENTES, M.; SMITH, R.L. **A new class of stationary spatial models**. North Caroline: Department of Statistics, North Caroline State University, 2001. Technical (Report, 2534).

## REFERÊNCIAS

GELFAND, A.E.; SCHMIDT, A.M.; BANERJEE S.; SIRMANS, C.F. Nonstationary multivariate process modeling through spatially varying coregionalization. **Test**, Valencia, v.13, p.263-312, 2004.

HANDCOCK, M.; STEIN, M. A Bayesian analysis of kriging. **Technometrics**, Alexandria, v.35, p.403-410, 1993.

HIGDON, D. **Quantitative methods for current environmental issues**. Chichester: Wiley, 2002. 185 p.

KITANIDIS, P.K. Parameter uncertainty in estimation of spatial functions: Bayesian analysis. **Water Resources Research**, New York, v.22 p.499-507, 1986.

LE, D.N.; ZIDEK, J.V. **Statistical analysis of environmental space-time processes**. New York: Springer, 2006. 327p.

MATÈRN, B. **Spatial variation**. Verlag, Berlin: Springer, 1986. 365 p.

NELDER, J.A.; MEAD R. A simplex method for function minimization. **The Computer Journal**, Oxford, v.7: p.308-313, 1965.

RIBEIRO Jr, P.J.; DIGGLE, P.J. geoR: A package for geostatistical analysis. **R-NEWS**, Vienna, v.1, n.2, p.14-18, 2001. ISSN 1609-3631.

ROYLE, A.J.; BERLINER, M.L. A hierarchical approach to multivariate spatial modeling and prediction. **Journal of Agricultural, Biological, and Environmental Statistics**, Alexandria, v.4, p.29-56, 1999.

SAMPSON, P.D.; GUTTORP, P. Nonparametric estimation of nonstationary spatial covariance structure. **Journal of American Statistical Association**, Alexandria, v.87, p.108-119, 1992.

SCHABENBERGER, O.; GOTWAY, C.A. **Statistical methods for spatial data analysis**. Boca Raton: Chapman and Hall / CRC, 2005. 488p.

SCHMIDT, A.M.; GELFAND, A.E. A Bayesian coregionalization approach for multivariate pollutant data. **Journal of Geophysical Research - Atmospheres**, Florida, v.108, p.8783, 2003.

SCHMIDT, A.M.; SANSÓ, B. Modelagem bayesiana da estrutura de covariância de processos espaciais e espaço-temporais. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 14, 2006. Caxambú, **Minicurso...** São Paulo: Associação Brasileira de Estatística, 2006. 151 p.

SCHMIDT, A.M.; O'HAGAN, A. Bayesian inference for nonstationary spatial covariance structure via spatial deformations, **Journal of Royal Statistical Society**, Oxford, v.65, p.743-758, 2003.

SCHULLER, J.K. A review and integrating analysis of spatially-variable crop control of crop production. **Fertilizer Research**, Amsterdam, v.33, p.1-34, 1992.

SILVA, A.S. **Modelos gaussianos geoestatísticos espaço-temporais e aplicações**, 2006. 70 p. Dissertação (Mestrado em Estatística e Experimentação Agrônômica) - Escola Superior de Agricultura "Luiz de Queiroz" da Univeridade de São Paulo, 2006.

SUN, W.; LE, N.D.; ZIDEK, J.V.; BURNETT, R. Assessment of a Bayesian multivariate interpolation approach for health impact studies. **Environmetrics**, Washington, v.9, p.565-586, 1998.

R: A Language and Environmental for Statistical Computing, R Development Core Team, R Foundation for Statistical Computing, Vienna, Austria, 2006, ISBN 3-900051-07-0. Disponível em: <http://www.R-project.org>. Acesso em 28 maio 2007.