

UNIVERSIDADE DE SÃO PAULO  
ESCOLA SUPERIOR DE AGRICULTURA “LUIZ DE QUEIROZ”  
PLANO DE DISSERTAÇÃO

PROGRAMA DE PÓS-GRADUAÇÃO EM “ESTATÍSTICA E EXPERIMENTAÇÃO AGRONÔMICA”  
CURSO DE MESTRADO

ANA JULIA RIGHETTO  
ORIENTADOR(A): CARLOS TADEU DOS SANTOS DIAS

AVALIAÇÃO DE MODELOS GEOESTATÍSTICOS MULTIVARIADOS

PIRACICABA  
Estado de São Paulo – Brasil  
Fevereiro / 2011

## SUMÁRIO

1 INTRODUÇÃO.....	2
2 OBJETIVO.....	3
3 REVISÃO DE LITERATURA .....	4
3.1 Geoestatística.....	4
3.1.1 Variáveis Regionalizadas .....	4
3.1.2 Semivariograma.....	5
3.1.3 Krigagem.....	7
3.1.4 Cokrigagem .....	7
4 METODOLOGIA.....	9
4.1 Material.....	9
4.2 Métodos .....	9
4.2.1 Modelos Lineares de Corregionalização .....	9
4.2.2 Semivariograma Cruzado .....	9
4.2.3 Modelos Geoestatísticos Bivariados Gaussianos.....	10
4.2.4 Modelo Bivariado de Co-regionalização .....	10
4.2.5 Modelagem por função de Correlação.....	11
4.2.5.1 Modelo Mátern Multivariado .....	12
5 CRONOGRAMA .....	13
REFERÊNCIAS .....	14

## 1 INTRODUÇÃO

Segundo Druck et al. (2004), compreender a distribuição espacial de dados provenientes de fenômenos ocorridos no espaço constitui um grande desafio para esclarecer questões centrais em diversas áreas do conhecimento, seja na área da saúde, em ambiente, em geologia, em agronomia, entre tantas outras áreas. Tais estudos tornam-se cada vez mais comuns, devido a disponibilidade de sistemas de informação geográfica (SIG). Desta forma, uma área da estatística que vem crescendo muito na atualidade é a estatística espacial.

A estatística espacial tem como intuito mensurar propriedades e relacionamentos, levando em conta a localização espacial do fenômeno estudado de forma explícita. Essa área da estatística é dividida em três áreas de estudo: geoestatística, dados de área e processos pontuais. Como o que distingue cada uma dessas categorias é o tipo de dado aleatório utilizado, é normal que em cada uma destas três áreas, existam métodos estatísticos diferentes para analisar e descrever os dados. Neste trabalho, será utilizado a geoestatística, mas especificamente na área de modelagem.

A geoestatística une o conceito de variáveis aleatórias com o conceito de variáveis regionalizadas, sendo estas últimas, variáveis que possuem um comportamento espacial. Este ramo da estatística espacial, incorpora tanto a interpretação da distribuição estatística como a correlação espacial das amostras, e segundo Matheron (1963), geoestatística é a aplicação do formalismo de funções aleatórias para o reconhecimento e estimação de fenômenos naturais.

Modelagem estatística é utilizada em diversos campos do conhecimento para tentar descrever o comportamento de um ou mais atributos que não podem ser descritos exclusivamente por modelos determinísticos. De uma forma geral, os modelos estatísticos tentam explicar, o máximo possível, a variabilidade dos processos estocásticos através de uma ou mais variáveis explanatórias que possuam alguma associação com a resposta de interesse (FONSECA,2008). Os modelos estatísticos lineares univariados, foram os primeiros a serem propostos e esses assumem erros aleatórios independentes e identicamente distribuídos com distribuição de probabilidade normal e todos os termos da parte explanatória do modelo fixos.

A modelagem geoestatística se diferencia destes modelos lineares nos pressupostos, no qual todas as observações não são assumidas independentes e existe efeito aleatório latente na parte explanatória do modelo. Nesta modelagem, existe um conjunto de técnicas para encontrar uma função aleatória para um ou mais atributos dos quais suas localizações espaciais são conhecidas. Desta forma, essas técnicas são importantes para capturar a correlação entre as observações das variáveis aleatórias de interesse, em que exista uma forte suspeita de que pontos espaciais próximos possuam valores parecidos da variável em estudo.

Em muitos casos, o interesse pode estar em não só estudar uma variável e sim, estudar duas ou mais variáveis e desta forma, tem-se dois ou mais atributos para modelar. Ocorrendo evidências de que esses processos não são independentes e existindo explicação prática para isto, modelos multivariados devem ser propostos para o estudo, devendo optar por um modelagem geoestatística multivariada.

Modelos geoestatísticos multivariados possuem, em geral, uma estrutura complexa combinando mais de um processo latente com elevado número de parâmetros e que podem apresentar problemas para estimação paramétrica. Além disto, em alguns casos, podem ocorrer problemas com a identificabilidade do modelo (FONSECA,2008).

Para este trabalho, um estudo detalhado de alguns métodos de modelagem multivariado será realizado, sendo que serão estudados o método de modelagem de modelos lineares de correção regionalização através de variogramas cruzados; modelos que implicam em estrutura de correlação e covariância, no qual será dada ênfase nos modelos BGCCM (*bivariate gaussian common component model*) e o BCRM (*bivariate coregionalisation model*). Mas o enfoque mesmo deste trabalho é na parte de modelagem por função de correlação bivariada (ou multivariada). Para esses estudos, serão utilizados dados simulados e dados reais.

## 2 OBJETIVO

Trabalhando com variáveis regionalizadas, inicialmente o intuito deste trabalho é de fazer uma revisão na modelagem de modelos lineares de corregionalização através de variogramas e variogramas cruzados. Também será feita uma revisão em modelos que implicam em estrutura de correlação/covariância, sendo estudados o modelo e o BCRM (*bivariate coregionalisation model*).

Além disso, um enfoque maior será dado em comparação de modelos de múltiplas estruturas com a modelagem através de função de correlação bivariada (ou multivariada), que foi proposta por Gneiting et al.(2009). Essas comparações e estudo dos resultados serão realizadas tanto em análise de dados reais como em dados obtidos através de simulações, trabalhando com o auxílio do software R, no qual será implementado tudo o que se fizer necessário.

### 3 REVISÃO DE LITERATURA

#### 3.1 Geostatística

Na atualidade, segundo Landim (2006), o termo Geostatística está relacionado como um tópico da estatística aplicada em que se trabalha com problemas relacionados às variáveis regionalizadas. Estas variáveis possuem um comportamento espacial, com características intermediárias entre as variáveis totalmente determinísticas e as verdadeiramente aleatórias.

##### 3.1.1 Variáveis Regionalizadas

As variáveis regionalizadas possuem um duplo aspectos contraditórios, pois é impossível uma previsão exata do valor da variável em um determinado ponto (aspecto aleatório), mas é esperado que o valor da variável seja parecido para pontos próximos (aspecto estrutural).

Na teoria das variáveis regionalizadas,  $Z(x)$  pode ser definida como uma variável aleatória que assume diferentes valores  $Z$  em função da posição  $x$  dentro de uma região  $A$ , e representa pares de coordenadas  $(x_i, y_i)$ , com mostrado na Figura 1. O ponto de referência para o sistema de coordenadas é arbitrário e fixado a critério do pesquisador.

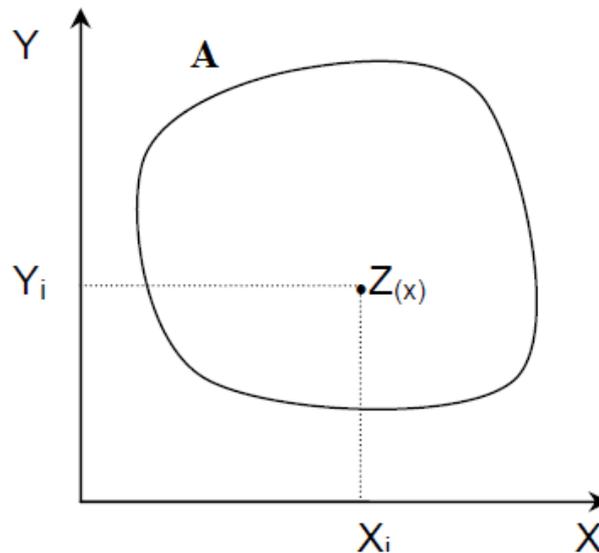


Figura 1 – Variável Aleatória Regionalizada  $Z(x)$ .

O conjunto dessas variáveis  $Z(x)$  medidas na totalidade da área  $A$ , pode ser considerada como uma função aleatória  $Z(x)$  pois, conforme Isaaks e Srivastava (1989), são variáveis aleatórias regionalizadas e assim, assume-se que a dependência entre elas seja especificada por algum mecanismo probabilístico.

Segundo Journel & Huijbregts (1978), a interpretação probabilística da variável regionalizada  $Z(x)$  ser uma particular realização de alguma função aleatória  $Z(x)$  é válida quando pode-se inferir toda ou parte da lei de distribuição de probabilidade que define esta função aleatória. Porém, na prática, em cada ponto  $x$  tem-se somente uma realização  $Z(x)$  e, como o número de pontos é finito, fica usualmente impossível inferir sobre a distribuição de  $Z(x)$ . Assim, hipóteses que envolvem diferentes graus de homogeneidade espacial, denominadas de hipóteses de estacionariedade, se fazem necessárias.

Estudando o comportamento das variáveis regionalizadas se faz necessário o uso de duas ferramentas dos métodos geoestatísticos: o semivariograma e a krigagem, explicados nas seções seguintes.

### 3.1.2 Semivariograma

Considera-se um variável regionalizada  $Z_{(x)}$  amostrada em diversos pontos que estão regularmente distribuídos em um região de estudo. Desta forma, o valor de cada ponto está relacionado de alguma maneira com valores obtidos a partir de pontos situados a certa distância, e pensando na Primeira Lei da Geografia de Tobler que diz :”Todas as coisas são relacionada, porém coisas mais próximas são mais parecidas que coisas distantes”, pode-se dizer que a influência é tanto maior quanto menor for a distância entre os pontos.

Para expressar essa relação é definido o vetor de distância  $\Delta\vec{h}$ , que possui uma orientação específica e o grau de relação entre pontos numa certa direção pode ser expresso pela covariância e, embora a covariância exista entre todas as distâncias possíveis ao longo de h, pode ser estipulado que somente sejam considerados valores entre pontos regularmente espaçados por múltiplos inteiros de  $\Delta h$ . (LANDIM,2006).

A covariância calculada entre os valores encontrados separados pela distância  $\Delta h$  ao longo de h é expressa por

$$C(h) = C(\Delta h) = \frac{1}{N(h)} \sum Z_x Z_{x+h} - m^2$$

em que  $m$  é a média da variável regionalizada e  $N(h)$  é o número de pares de valores medidos separados pela distância h.

A covariância depende do tamanho do vetor h, assim, se  $h = 0$ ,  $C(h)$  representará a variância, isto é

$$C(0) = E[Z^2] - E^2[Z] = Var[Z]$$

Assim, pode-se calcular a função denominada de semivariância, dada pela metade da variância das diferenças, representada por  $\gamma(h)$  em que

$$\gamma(h) = \frac{1}{2n} \sum_{i=1}^{N(h)} (Z_{x_i+h} - Z_{x_i})^2$$

Para analisar como os valores dos dados estão se comportando conforme varias distâncias estipuladas e assim, analisar o grau de dependência espacial da variável e definir os parâmetros necessários para estimar características em locais não amostrados, utiliza-se um gráfico de  $\gamma(h)$  por h denominado de semivariograma.

O semivariograma analisa o grau de dependência espacial entre amostras dentro de um campo experimental, além de definir parâmetros necessários para a estimativa de valores para locais não amostrados, através da técnica de krigagem (SALVIANO, 1996).

Segundo Landim (1996), o estudo do variograma é realizado em uma direção ao longo de uma linha ou ao longo de uma série de linhas paralelas, utilizando n possíveis distâncias h, para desta maneira realizar um estudo na variável para analisar se os dados são estacionários, isto é, se não há tendência nos valores.

O ajuste de um modelo matemático ao dados no gráfico, isto é, a uma função, define os parâmetros do semivariograma, os quais são: efeito pepita ( $C_0$ ), que representa o valor de  $\gamma$  quando  $h = 0$ ; quando o valor de h aumenta frequentemente, aumenta ate uma distância **a**, denominada de alcance (alcance da dependência espacial), também podendo ser expresso por  $\phi$ ; neste ponto **a**,  $\gamma(h)$  é chamado de patamar, dado por  $C+C_0$ , sendo **C** a variância estrutural. Na Figura 2, tem-se o grafico de  $\gamma(h)$  por h, no qual nota-se a disposição de cada parâmetro e pode-se observar um semivariograma experimental e o modelo matemático ajustado.

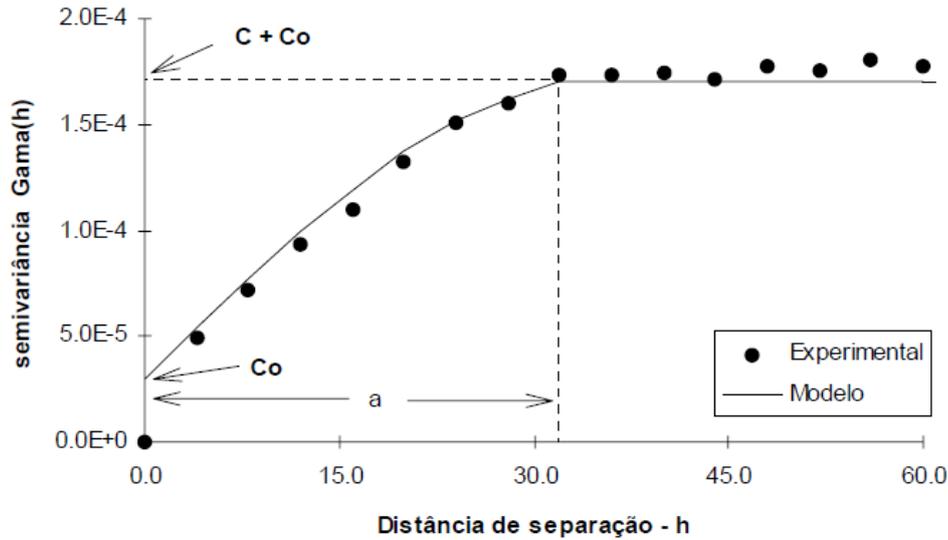


Figura 2 – semivariograma experimental e modelo matemático ajustado

O alcance é utilizado para definir o raio de ação (*range*) máximo de interpolação pela krigagem, em que os pesos utilizados podem afetar os valores estimados. Pontos amostrados que estão separados por distância menor do que o alcance, são ditos espacialmente dependentes, o que não é o caso de pontos separados com uma distância maior.

Segundo Vieira (200), em um semivariograma ajustado, o valor da semivariância aumenta a medida que aumenta a distância de separação entre os pontos, aumentando até atingir um patamar no qual se estabiliza. Este patamar (*sill*) é um parâmetro importante, pois permite a determinação da distância limite entre dependência e independência entre as amostras, e é atingido quando a variância dos dados fica constante com a distância entre os pontos amostrados.

O parâmetro  $C_0$ , efeito pepita, reflete o erro analítico, isto é, indica uma variabilidade não explicada de um ponto para outro, que pode ocorrer devido a erros de medidas ou microvariação não detectada por causa da distância de amostragem utilizada.

Uma outra consideração a ser feita é determinar o grau de aleatoriedade presente nos dados pela fórmula  $E = \frac{C_0}{C}$  (GUERRA, 1988 apud LANDIM, 1996), em que:

- $E < 0,15$ : componente aleatória pequena;
- $0,15 \leq E \leq 0,30$ : componente aleatória significativa;
- $E > 0,30$ : componente aleatória muito significativa.

O modelo matemático ajustado, estão divididos em duas categorias: modelos com patamar e sem patamar.

Modelo com patamar são aqueles em que a semivariância aumenta com o aumento da distância entre amostras, até atingir o patamar, no qual se estabiliza. Exemplos de modelos nessa categoria são:

- Modelo Esférico:  $\gamma(h) = C \left[ \frac{3}{2} \left( \frac{h}{a} \right) - \frac{1}{2} \left( \frac{h}{a} \right)^3 \right]$ , para  $h < a$ , e  $\gamma(h) = C$ , para  $h \geq a$ ;
- Modelo Exponencial:  $\gamma(h) = C \left[ 1 - e^{-3h/a} \right]$
- Modelo Gaussiano:  $\gamma(h) = C \left[ 1 - e^{(-3h/a)^2} \right]$

Uma outra função bastante utilizada é a função Matérn dada pela forma

$$\gamma(h) = \left\{ \left\{ 2^{\kappa-1} \Gamma(\kappa) \right\} \right\}^{-1} \left( \frac{h}{\phi} \right)^{\kappa} K_{\kappa} \left( \frac{h}{\phi} \right)$$

no qual  $\Gamma(\cdot)$  é a função gama,  $K_\kappa(\cdot)$  é a função Bessel de ordem  $\kappa$ , e  $h$  a distância entre duas localizações. Quando  $\kappa$  é igual a 0,5 a função Matérn é igual a função exponencial.

Os modelos sem patamar, são modelos que como próprio nome diz, não se estabilizam em um patamar. Exemplo de modelo sem patamar é o modelo potencial.

Segundo Landim(1996), após o ajuste de um variograma experimental, é importante que realize a ‘validação cruzada’, em que depois de obter o modelo variográfico, cada valor original da amostra é removido do domínio espacial e, utilizando os demais valores, um novo valor é estimado para o ponto retirado, podendo assim fazer uma comparação entre o real valor e o estimado e verificar se o erro de estimação esta muito grande. Está técnica da validação cruzada, não prova que o modelo escolhido é o mais correto, mas sim que o mesmo não é de todo incorreto.

Depois de estipulado o semivariograma da variável em estudo, e ocorrendo a dependência espacial entre os pontos amostrados, pode-se interpolar valores em qualquer posição na área em estudo, sem tendência e com variância mínima. O método de interpolação utilizado é a krigagem; e quando se trata de um modelos com mais de uma variável envolvida, utiliza-se a cokrigagem.

### 3.1.3 Krigagem

Krigagem é um processo de estimativa de valores de variáveis distribuídas no espaço, e/ou no tempo, a partir de valores adjacentes enquanto considerados como interdependentes pelo semivariograma (LANDIM,2006).

Utilizado como algoritmo estimador, esta técnica pode prever o valor pontual de uma variável regionalizada em um determinado local dentro da área de estudo, além de realizar o cálculo médio de uma variável regionalizada para um volume maior que o suporte geométrico. Em todos estes casos, este método fornece o “erro” associado a estimativa realizada, tornando-o um algoritmo melhor que os demais existentes.

Pela krigagem, o valor estimado da variável é expresso por:

$$Z_{(x_0)}^* = \lambda_0 + \sum_{i=1}^n \lambda_i Z_{(x_i)}$$

em que  $n$  é o número de vizinhos medidos,  $Z_{(x_i)}$  utilizados na estimativa da variável e  $\lambda_i$  são os ponderadores aplicados a cada  $Z_{(x_i)}$  que são selecionados de forma que a estimativa obtida seja não tendenciosa.

### 3.1.4 Cokrigagem

Segundo Bognola, I. A. et al.(2008),a partir do ajuste de um modelo é possível obter predições espaciais das variáveis de interesse através da cokrigagem, no qual o valor predito em um determinado ponto é dado por uma média ponderada dos dados das variáveis envolvidas. Isto é, segundo Landim,Sturaro&Monteiro(2002), cokrigagem é uma extensão multivariada do método da krigagem, quando para cada local amostrado obtém-se um vetor de valores no lugar de um único valor.

Esta técnica é bastante útil e evidente, quando duas ou mais variáveis são amostradas nos mesmos pontos amostrais dentro de um mesmo domínio espacial e apresentam significativo grau de correlação entre elas.

A cokrigagem, segundo Isaaks e Srivastava (1989), considera uma ou mais variáveis secundárias na estimativa de uma variável primária, podendo ser escrita em termos das covariâncias simples e cruzadas, sendo os pesos  $W$  usados na interpolação dados por  $W = C^{-1}D$ , em que:

$$C = \begin{pmatrix} Cov(Z_1Z_1) & Cov(Z_1Z_2) & 1 \\ Cov(Z_2Z_1) & Cov(Z_2Z_2) & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}; W = \begin{pmatrix} e \\ f \\ -\lambda_1 \\ -\lambda_2 \end{pmatrix} e D = \begin{pmatrix} Cov(Z_1Z_0) \\ Cov(Z_2Z_0) \\ 1 \\ 0 \end{pmatrix}$$

temos que  $Z_1$  e  $Z_2$ , variável primária e secundária, respectivamente.  $Z_0$  os valores que serão preditos;  $\mathbf{e}$  e  $\mathbf{f}$  os vetores de pesos atribuídos às variáveis primária e secundária, respectivamente, sendo que  $\sum e_i = 1$  e  $\sum f_i = 1$  são restrições para garantir a não-tendenciosidade, introduzidas pelo uso dos operadores de Lagrange  $\lambda_1$  e  $\lambda_2$ .

## 4 METODOLOGIA

### 4.1 Material

Para este trabalho, os dados utilizados serão tanto dados simulados, como dados reais. As simulações e análises serão realizadas no ambiente R de programação (R Development Core Team, 2006), utilizando em específico os pacotes geoR (Ribeiro Jr. e Diggle, 2001), gstat (Pebesma, 2011) e Random Fields (Schlather, 2011), e quando necessário, códigos próprios serão desenvolvidos no software.

### 4.2 Métodos

No caso de observações bi ou multivariadas, os modelos e procedimentos são estendidos de diferentes formas, visando modelar as covariâncias simples e cruzadas espaciais entre as variáveis envolvidas (WACKERNAGEL, 2003 apud BOGNOLA, I. A. et al., 2008). Existe várias técnicas de modelagem, dentre elas tem-se modelagem pelo semivariograma cruzado e modelagem através de função de correlação bi (ou multivariada).

#### 4.2.1 Modelos Lineares de Corregionalização

Segundo Landim, Sturaro & Monteiro (2002), utilizando o método da cokrigagem, formulado a partir da suposição de que as variáveis primária e secundária apresentam covariância, com a matriz sendo positiva definida, para esta ser considerada uma matriz de covariâncias cruzadas válida, e uma maneira simples para obter-se essa matriz é utilizando o modelo linear de corregionalização.

O modelo linear de corregionalização fornece um método para modelagem do auto-variograma e do variograma cruzado entre duas ou mais variáveis desde que a variância de qualquer combinação linear possível dessas variáveis seja sempre positiva. Cada variável é caracterizada por seu próprio auto-variograma amostral e cada par de variáveis por seu variograma cruzado amostral (ISAAKS e SRIVASTAVA, 1989).

O modelo linear de corregionalização, segundo Bellier et al. (2007), permite a modelagem conjunta do variograma simples e do variograma cruzado, sendo que o número total desses variogramas é dado por  $\frac{p(p+1)}{2}$  em que  $p$  é o número de variáveis presentes no conjunto de estudo.

Segundo Sartori (2006), para que este modelo seja compreendido, precisa-se que o modelo básico, utilizado no variograma cruzado, seja o mesmo utilizado nos auto-variogramas, mantendo-se assim o valor para o alcance.

A decomposição de um modelo linear de corregionalização, pode ser dada por:

$$\gamma_{ij}(h) = b_{ij}^{(l)} g^{(l)} + b_{ij}^{(k)} g^{(k)} + \dots + b_{ij}^{(q)} g^{(q)}$$

em que  $g^{(l)}$  são os modelos básicos (isto é, exponencial, esférico, entre outros), e  $b_{ij}^{(l)}$  são coeficientes que correspondem ao alcance parcial da função do variograma e variograma cruzados em cada escala considerada.

Assim, ao examinar o comportamento do variograma empírico, do simples e do variograma cruzado, o pesquisador escolhe o número e tipo de base modelos. Cada função do variograma utilizado para definir o modelo linear de corregionalização tem diferentes famílias e intervalos, em que cada um caracteriza uma escala específica da distribuição espacial da variável resposta em estudo.

#### 4.2.2 Semivariograma Cruzado

Sejam  $p$ -variáveis  $\{Z_{j(x)}; j = 1, \dots, p\}$  que foram medidas dentro de uma mesma região de estudo e nos mesmo pontos amostrais, com pares de coordenadas  $[x_i, y_i]$ . Dessas variáveis medidas, considerando apenas duas, temos  $\{[Z_{1(x)}]e[Z_{2(x)}]\}$  de tal forma que a covariância cruzada entre essas variáveis seja dada por:

e

$$Cov_{12}(h) = E\{[Z_{1(x+h)}Z_{2(x)}]\} - m_1m_2$$

$$Cov_{21}(h) = E\{[Z_{2(x+h)}Z_{1(x)}]\} - m_2m_1$$

Desta forma, o semvariograma cruzado será:

$$\gamma_{12}^*(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \{[Z_{1(x+h)} - Z_{1(x)}][Z_{2(x+h)} - Z_{2(x)}]\}$$

Segundo Silva et al.(2003 apud Vieira, 1996) a forma de obter os gráficos do semivariograma cruzado experimental tem significado diferente do obtido pelo semivariograma simples; pois no variograma cruzado o alcance (**a**) representa a distância máxima de dependência espacial entre as duas variáveis; o patamar, se existir, deve aproximar-se do valor da covariância entre as duas variáveis utilizadas. Quando estas duas variáveis apresentarem uma correlação inversa, a covariância será negativa e, desta forma, o semivariograma cruzado produzirá resultados negativos.

#### 4.2.3 Modelos Geoestatísticos Bivariados Gaussianos

Quando o interesse está em mais de um campo aleatória de interesse, suponha-se dois, muitos pensam que o passo inicial seria de ajustar um modelo geoestatísticos para cada um dos dois vetores das variáveis observadas. Mas nem sempre isso é o melhor caminho, pois há casos em que existe correlação entre as variáveis aleatórias estudadas, o que seria um bom caso para utilizar modelos geoestatísticos bivariados.

Existindo dois campos aleatórios gaussianos, utiliza-se o modelo dado por:

$$Y_i = \mu_i + Z_i, \text{ com } i = 1,2$$

em que  $Y_i$  é um vetor  $n_i \times 1$  dos valores observados no campo aleatório latente  $Z_i$ , o qual possui um vetor de médias nulos de dimensão  $n_i \times 1$  e matriz de covariância  $\Sigma_i$  com dimensão  $n_i \times n_i$ .  $\mu_i$  é um vetor, dos parâmetros de média associadas a  $Y_i$ .

Segundo Fonseca(2008), a distribuição de  $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2)$  é o interesse final, e esta distribuição é gaussiana n-variada, sendo  $n = n_1 + n_2$ , com vetor de médias  $\boldsymbol{\mu} = (\boldsymbol{\mu}_1, \boldsymbol{\mu}_2)$  e matriz de covariância  $\Sigma_Y$ , sendo esta matriz positiva definida e possui comportamento empírico de correlações utilizado na geoestatística, podendo ser particionada por:

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

sendo  $\Sigma_i$  uma matriz de covariância de dimensão  $n_i \times n_i$  da variável  $Y_i$  com  $i = 1,2$  e  $\Sigma_{1,2}$  a matriz de covariâncias cruzadas entre as variáveis  $Y_1$  e  $Y_2$ , de dimensão  $n_1 \times n_2$ . É necessário que a matriz  $\Sigma_Y$  seja válida.

#### 4.2.4 Modelo Bivariado de Co-regionalização

Gelfand et al. (2004) propuseram uma abordagem para conseguir a estrutura paramétrica válida para a matriz  $\Sigma_Y$ , que é denominada de *Bivariate Coregionalisation Model*, abreviado por BCRM.

Por esta abordagem, os termos latentes do modelo  $Y_i = \mu_i + Z_i$ ,  $i = 1,2$ , são decomposto da forma dada a seguir:

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

no qual  $S_1$  e  $S_2$  dois campos aleatórios gaussianos mutuamente independentes; possuem vetores de média nulo, variância unitária e correlações determinadas pela escolha de funções de correlação válidas. Nota-se

que o campo aleatório  $S_1$  é comum as duas decomposições, e desta forma, é induzida a correlação empírica entre as variáveis respostas.

Através do BCRM as observações do modelo em estudo são dadas por:

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

Considerando observações feitas da variável resposta  $Y_i$  dadas por  $Y_i(s_l)$  e  $Y_i(s_k)$ , em duas localizações,  $s_l$  e  $s_k$ , sendo estas observações separadas por uma distância  $h = h_{lk}$ , com  $l, k = 1, 2, \dots, n_i$  e  $i = 1, 2$ . Assim, tem-se que o elemento  $\Sigma_{1,(l,k)}$  é dado por  $Cov(h) = \sigma_{11}^2\rho_1(h)$  e o elemento  $\Sigma_{2,(l,k)}$  é representado por  $Cov(h) = \sigma_{12}^2\rho_1(h) + \sigma_{22}^2\rho_2(h)$ , no qual  $\rho_1$  e  $\rho_2$  são funções de correlação assumidas, respectivamente, para  $S_1$  e  $S_2$ .

Por propriedades básicas de covariâncias, encontra-se que  $\Sigma_{1,2}$  é equivalente a  $\sigma_{11}\sigma_{12}R_1$ , em que  $R_1$  é matriz das correlações cruzadas entre as respostas e depende somente da função de correlação adotada para  $S_1$ , e sua dimensão é  $n_1 \times n_2$ . Então,  $\Sigma_Y$  fica estruturada.

Supondo que existam apenas duas localizações que foram amostradas; assim, o vetor das variáveis  $(Y_i(s_l), Y_i(s_k), Y_j(s_l), Y_j(s_k))$  possui matriz de covariância expressa por:

$$\Sigma_Z = \begin{bmatrix} \sigma_{11}^2 & \sigma_{11}^2\rho_1(h) & \sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{12}\rho_1(h) \\ \sigma_{11}^2\rho_1(h) & \sigma_{11}^2 & \sigma_{11}\sigma_{12}\rho_1(h) & \sigma_{11}\sigma_{12} \\ \sigma_{11}\sigma_{12} & \sigma_{11}\sigma_{12}\rho_1(h) & \sigma_{12}^2 + \sigma_{22}^2 & \sigma_{12}^2\rho_1(h) + \sigma_{22}^2\rho_2(h) \\ \sigma_{11}\sigma_{12}\rho_1(h) & \sigma_{11}\sigma_{12} & \sigma_{12}^2\rho_1(h) + \sigma_{22}^2\rho_2(h) & \sigma_{12}^2 + \sigma_{22}^2 \end{bmatrix}$$

Desta maneira, segundo Fonseca (2008), a distribuição de probabilidade do vetor  $Y = (Y_1, Y_2)$  está estabelecida e depende do vetor de parâmetros  $\theta = (\beta^*, \sigma^*, \phi_1^*, \phi_2^*)$ , no qual  $\beta^*$  é o vetor de parâmetros associados a  $\mu$  e  $\phi_j^*$  é vetor de parâmetros relacionados a escolha da função  $\rho_j$ , com  $j = 1, 2$ .

Uma forma geral deste modelo, consiste na adição de um termo, sendo este de ruídos não correlacionados que podem ser associados a todas, ou algumas, variáveis  $Y_i$ . Desta forma, o modelo fica expresso:

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

em que  $U$  é um processo de ruído branco.

#### 4.2.5 Modelagem por função de Correlação

Em alguns casos multivariados, qualquer que seja o objetivo do pesquisador, segundo Gneiting, Kleiber & Schlater (2009), se faz necessário a modelagem de várias variáveis espaciais simultaneamente. O passo crítico é então identificar a estrutura de dependência espacial, não somente em cada variável como também entre as variáveis.

Gneiting, Kleiber & Schlater (2009) propuseram um ajuste por estrutura de covariância e covariância cruzada para um campo aleatório gaussiano. É uma proposta vantajosa, devido ao fato da facilidade em mudar os parâmetros do modelo e ajusta melhor aos dados. Para este tipo de modelagem, utiliza-se a função da família Matern, que tem sido muito utilizada na atualidade, que é uma função bem flexível, o que facilita na sua utilização.

Esta família específica uma função de covariância dada por  $\sigma^2 M(\mathbf{h}|\kappa, a)$  em que  $\sigma^2 > 0$  é a variância marginal e

$$M(\mathbf{h}|\kappa, a) = \frac{2^{1-\kappa}}{\Gamma(\kappa)} (a|\mathbf{h}|)^\kappa K_\kappa(a|\mathbf{h}|)$$

é a correlação espacial na distância  $|\mathbf{h}|$ , e como já dito  $K_\kappa$  é a função Bessel.

Em um processo multivariado, seja o vetor de variáveis  $\mathbf{Y}(s) = (Y_1(s), \dots, Y_p(s))'$ , em que cada localização  $s \in \mathbb{R}^d$  existem  $p$  componentes. Assumindo que o processo é gaussiano multivariado e estacionário de segunda ordem com vetor de média nulo e matriz de função de covariância dada por

$$C(h) = \begin{pmatrix} C_{11}(h) & \dots & C_{1p}(h) \\ \vdots & \ddots & \vdots \\ C_{p1}(h) & \dots & C_{pp}(h) \end{pmatrix}$$

em que cada  $C_{ii}(h) = E(Y_i(x+h)Y_i(x))$  é a função de covariância univariada, enquanto que  $C_{ij}(h) = E(Y_i(x+h)Y_j(x))$  é a função de covariância cruzada entre os processo  $1 \leq i \neq j \leq p$ , é equivalente a

$$C_{ij}(h) = C_{ji}(h) = \rho_{ij}\sigma_i\sigma_jM(\mathbf{h}|\kappa_{ij}, a_{ij})$$

sendo  $\rho_{ij}$  coeficiente de correlação.

#### 4.2.5.1 Modelo Mátern Multivariado

Como já mencionado, se o parâmetro  $\kappa$  da função Mátern foi igual a  $\frac{1}{2}$ , então esta função se reduz ao produto da função exponencial e polinômio, isto é

$$M\left(\mathbf{h} \mid n + \frac{1}{2}\right) = \exp(-a\|\mathbf{h}\|) \sum_{m=0}^n \frac{(n+m)!}{(2n)!} \binom{n}{m} (2a\|\mathbf{h}\|)^{n-m}$$

com  $n = 0, 1, \dots$

Segundo Gneiting, Kleiber & Schlater (2009), a flexibilidade e popularidade da classe Mátern se deve ao fato do parâmetro  $\kappa > 0$ , o qual controla a suavidade e é crítico em problemas de interpolação espacial.



## REFERÊNCIAS

- BELLIER, E.; MONESTIEZ, P.; DURBEC, J.P.; CANDAU, J.N. **Identifying spatial relationships at multiple scales: principal coordinates of neighbor matrices (PCNM) and geostatistical approaches.** *Ecography* 30: 385-399, 2007.
- BOGNOLA, I.A.; RIBEIRO JÚNIOR, 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, PR, v.38, n. 2, abr/jun.2008.
- DRUCK, S.; CARVALHO, M.S.; CÂMARA, G.; MONTEIRO, A.M.V. **Análise espacial de dados geográficos.** Brasília, EMBRAPA, 2004 (ISBN: 85-7383-260-6).
- FONSECA, B.H.F. **Um estudo sobre estimação e predição em modelos geoestatísticos bivariados.** Piracicaba, 2008. 74 p. Dissertação (Mestrado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.
- 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.
- GNEITING, T.; KLEIBER, W.; SCHLATHER, M. **Matérn Cross-Covariance Functions for Multivariate Random Fields.** Technical Report n.549. Department of Statistics, University of Washington.
- ISAAKS, E.H.; SRIVASTAVA, R.M. **An Introduction to Applied Geostatistics.** New York: Oxford University Press, 1989. 561 p.
- JOURNEL, A.G.; HUIJBREGTS, C.J. **Mining Geostatistics.** London: Academic Press, 1978. 600 p.
- LANDIM, P.M.B. STURARO, J.R. & MONTEIRO, R. C. **Exemplos de aplicação da cokrigagem.** DGA, IGCE, UNESP/Rio Claro, Lab. Geomatemática, Texto Didático 09, 17 pp. 2002. Disponível em <<http://www.rc.unesp.br/igce/aplicada/textodi.html>>. Acesso em: 01 de Fevereiro de 2011.
- LANDIM, P.M.B. **Sobre Geoestatística e mapas.** *Terrae Didatica*, 2(1):19-33. 2006. Disponível em: <<http://www.ige.unicamp.br/terradidatica/>>.
- MATHERON, G. **Principles of geostatistics.** *Economic Geology*, v. 58, p. 1246–1266, 1963.
- PEBESMA, E. **gstat package.** January 2011. Disponível em: <<http://cran.r-project.org/web/packages/gstat/>>
- R: A Language and Environment 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>>.
- RIBEIRO Jr, P.J.; DIGGLE, P.J. **geoR: A package for geostatistical analysis.** *R-NEWS*, Vienna, v.1, n.2, p.14-18, 2011. ISSN 1609-3631.
- SALVIANO, A.A.C. **Variabilidade de atributos de solo e de *Crotalaria juncea* em solo degradado do município de Piracicaba-SP.** Piracicaba, 1996. 91p. Tese (Doutorado) - Escola Superior de Agricultura "Luiz de Queiroz", Universidade de São Paulo.
- SARTORI, L.R. **Métodos para extração de informações a partir de imagens multiespectrais de escalas grandes.** Presidente Prudente : [s.n.], 2006. 127 f. Dissertação (Mestrado) - Universidade Estadual Paulista, Faculdade de Ciências e Tecnologia.

SCHLATHER, M. **Random Fields package**. February 2011. Disponível em:  
< <http://cran.r-project.org/web/packages/RandomFields/index.html> >

SILVA, E.A.A.; URIBE-OPAZO, M.A.; SOUZA, E.G.; ROCHA, J.V. **Um estimador robusto e o semivariograma cruzado na análise de variabilidade espacial de atributos de solo e planta**. Acta Scientiarum. Agronomy. Maringá, v. 25, no. 2, p. 365-371, 2003.

SOUZA, E.G *et al.* **Variabilidade espacial dos atributos químicos do solo em um Latossolo roxo Distrófico da região de Cascavel - PR**. *Engenharia Agrícola*, Jaboticabal, v.18, n.3, p.80-92, 1999.

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, 2000. v. 1, p. 1-54.