

## SUMÁRIO

LISTA DE FIGURAS . . . . .	8
LISTA DE TABELAS . . . . .	15
1 INTRODUÇÃO . . . . .	17
2 REVISÃO DE LITERATURA . . . . .	19
2.1 Geoestatística . . . . .	19
2.1.1 Modelos Geoestatísticos Univariados . . . . .	19
2.1.2 Estimaco dos Parâmetros do Modelo . . . . .	21
2.1.3 Váriaveis Regionalizadas . . . . .	23
2.1.3.1 Semivariograma . . . . .	24
2.1.3.2 Krigagem . . . . .	28
2.2 Geoestatística Multivariada . . . . .	29
2.2.1 Cokrigagem . . . . .	29
2.2.2 Modelos Lineares de Corregionalizao . . . . .	30
2.2.3 Semivariograma Cruzado . . . . .	31
2.2.4 Modelos Geoestatísticos Bivariados Gaussianos . . . . .	32
2.2.5 Modelo Bivariado de Co-regionalizao . . . . .	33
2.2.6 Modelo gaussiano bivariado com componente comum . . . . .	35
2.2.7 Modelagem por funo de correlao . . . . .	36
2.2.7.1 Modelo Mátern Multivariado . . . . .	37
2.2.7.1.1 Modelo Mátern Multivariado Parcimonioso . . . . .	37
REFERÊNCIAS . . . . .	39

**LISTA DE FIGURAS**

Figura 1 - Seminariograma experimental e Modelo matemático ajustado. . . . . 26



**LISTA DE TABELAS**



## 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 uma 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 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.

**FALAR SOBRE OS DADOS!!!!**

## 2 REVISÃO DE LITERATURA

### 2.1 Geoestatística

De acordo com Diggle, Ribeiro Jr e Christensen (2003), a geoestatística é um ramo da estatística espacial, em que os dados consistem de mensurações  $y_1, y_2, \dots, y_n$  obtidas em localizações  $\mathbf{s}_1, \mathbf{s}_2, \dots, \mathbf{s}_n$ , amostradas de uma região  $A$  espacialmente contínua, sendo que  $A \subset \mathbb{R}^d$ . Estas medidas estão associadas a um fenômeno espacial, tratado como a realização de um processo estocástico  $Z(\mathbf{s})$ ,  $\mathbf{s} \subset \mathbb{R}^d$  que não é diretamente observado. Um processo estocástico espacial pode ser definido como qualquer conjunto de variáveis aleatórias  $Z(\mathbf{s})$ ,  $\mathbf{s} \in \mathbb{R}^d$  definidas em um espaço de probabilidade comum, em que  $d$  é o número de linhas do vetor  $\mathbf{s}$ , no qual os elementos são as localizações amostrais. Além disso, segundo Soares (2006), um processo estocástico é um modo de abordagem do conjunto de dados provenientes das amostras experimentais então, o primeiro passo da modelização geoestatística é a concepção de um processo aleatório que melhor caracteriza o conjunto de dados experimentais das observações.

Foram Diggle, Tawn e Moyeed (1998) que introduziram o termo "geoestatística baseada em modelos", na qual a modelagem é realizada adotando-se um modelo paramétrico para a(s) variável(véis) em estudo, e utiliza-se métodos de inferência estatística baseada na verossimilhança aplicada a estudos geoestatísticos.

#### 2.1.1 Modelos Geoestatísticos Univariados

Segundo Diggle e Ribeiro Jr (2007), as suposições do modelo Gaussiano estacionário são:

- i.  $\{Z(s) : s \in \mathbb{R}^2\}$  é um processo Gaussiano com média  $\mu$  e variância  $\sigma^2 = Var\{Z(s)\}$  e função de correlação  $\rho(h) = Corr\{Z(s), Z(s')\}$ , em que  $h = \|s - s'\|$  e  $\|\cdot\|$  é a distância;
- ii. Condicional a  $\{Z(s) : s \in \mathbb{R}^2\}$ ,  $y_i$  são as realizações mutuamente independentes da variável aleatória  $Y_i$ , com distribuição Normal, média condicionada dada por  $E[Y_i|Z(\cdot)] = Z(s_i)$  e variância condicional  $\tau^2$ .

Assim, considerando que na região de estudo  $A$ , exista um campo aleatório gaussiano  $Z$  latente, é necessário realizar uma amostragem, obtendo-se valores da variável de inte-



resse, em  $n$  localizações. Desta forma, cria-se um vetor  $Y(s)$  de dimensão  $n \times 1$ , doo valores observados nas  $n$  localizações dadas pelo vetor  $\mathbf{s} = (s_1, \dots, s_n)$ . Desta forma, obtem-se o modelo:

$$Y(s) = \mu + Z(s) + \epsilon \quad (1)$$

em que  $\epsilon$  é o ruído branco, um vetor de dimensão  $n \times 1$ , com distribuição normal com média zero e desvio padrão  $\tau$ ;  $Z(s)$  é um campo aleatório gaussiano, definido nas suposições acima, que possui vetor de médias nulo, de dimensão  $n \times 1$  e matriz de covariância  $\Sigma$ ,  $n \times n$  e  $\mu = X\beta$ , sendo  $X$  uma matriz de dimensão  $n \times q$ , contendo  $q - 1$  covariáveis e  $\beta$  é um vetor  $q \times 1$  de parâmetros associados a essas covariáveis. Ainda, segundo Martins (2010), esse vetor de médias pode ser escrito por:

$$\mu = \beta_0 + \sum_{j=1}^q \beta_j X_j \quad (2)$$

Desta forma, utilizando o modelo (1), pode-se encontrar a distribuição de probabilidade de  $Y(s)$ , com vetor de médias  $X\beta$  e matriz de covariâncias dada por  $\Sigma_Y = \Sigma + \tau^2 I$ , em que  $I$  é uma matriz identidade  $n \times n$ . Então, tem-se um vetor a ser estimado,  $\theta = (\beta, \sigma^2, \phi^*, \tau^2)$ , no qual  $\phi^*$  é um vetor de parâmetros associados a função de correlação utilizada.

Segundo Martins (2010), quando a variável observada  $Y(\mathbf{s})$  não segue uma distribuição gaussiana, o valor do erro quadrático médio mínimo do preditor é afetado, podendo assim, obter aproximações ruins. Para contornar esse problema, realiza-se uma transformação nos dados, para que estes passem a seguir uma distribuição gaussiana.

Uma transformação muito conhecida, foi proposta por Box e Cox (1964):

$$Y^*(\mathbf{s}) = \begin{cases} \frac{Y^{\lambda(\mathbf{s})} - 1}{\lambda}, \lambda \neq 0 \\ \log Y(\mathbf{s}), \lambda = 0 \end{cases}$$

no qual o parâmetro  $\lambda$  introduz flexibilidade ao modelo. Alguns valores interpretáveis desse parâmetro são:

$$\left\{ \begin{array}{l} \lambda = 1 : \quad \text{sem transformação} \\ \lambda = 0.5 : \quad \text{transformação raiz quadrada} \\ \lambda = -1 : \quad \text{transformação recíproca} \end{array} \right.$$

### 2.1.2 Estimação dos Parâmetros do Modelo

Depois de obter um modelo adequado para os dados em estudo, deve-se estimar os parâmetros desse modelo. Para estimação dos parâmetros do modelo gaussiano normalmente utiliza-se o método dos mínimos quadrados e o método da máxima verossimilhança.

Pelo método da máxima verossimilhança, admitindo-se uma tendência polinomial para  $\mu(\mathbf{s})$ , tem-se:

$$\mathbf{Y}(\mathbf{s}) \sim N_n(\mathbf{X}\boldsymbol{\beta}; \sigma^2\mathbf{R} + \tau^2\mathbf{I}) \quad (3)$$

em que  $\mathbf{X}$  é a matriz  $n \times q$  de covariáveis,  $\boldsymbol{\beta}$  é o vetor de parâmetros e  $\mathbf{R}$  é a matriz de correlação.

A função de verossimilhança é:

$$L(\mathbf{Y}(\mathbf{s})) = (2\pi)^{-\frac{n}{2}} |\sigma^2\mathbf{R} + \tau^2\mathbf{I}|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2}(\mathbf{Y}(\mathbf{s}) - \mathbf{X}\boldsymbol{\beta})'(\sigma^2\mathbf{R} + \tau^2\mathbf{I})^{-1}(\mathbf{Y}(\mathbf{s}) - \mathbf{X}\boldsymbol{\beta}) \right\} \quad (4)$$

e o log da função de verossimilhança é dado por:

$$\begin{aligned} l(\boldsymbol{\beta}, \tau^2, \sigma^2, \phi) &= \ln[(2\pi)^{-\frac{n}{2}}] - \ln(|\sigma^2\mathbf{R} + \tau^2\mathbf{I}|^{\frac{1}{2}}) - \frac{1}{2}(\mathbf{Y}(\mathbf{s}) - \mathbf{X}\boldsymbol{\beta})'(\sigma^2\mathbf{R} + \tau^2\mathbf{I})^{-1}(\mathbf{Y}(\mathbf{s}) - \mathbf{X}\boldsymbol{\beta}) \\ &= -\frac{1}{2}\{n\ln(2\pi) + \ln(|\sigma^2\mathbf{R} + \tau^2\mathbf{I}|) + (\mathbf{Y}(\mathbf{s}) - \mathbf{X}\boldsymbol{\beta})'(\sigma^2\mathbf{R} + \tau^2\mathbf{I})^{-1}(\mathbf{Y}(\mathbf{s}) - \mathbf{X}\boldsymbol{\beta})\} \end{aligned}$$

Para proceder com a maximização do log da verossimilhança, considera-se  $\mathbf{Y}(\mathbf{s}) \sim N_n(\mathbf{X}\boldsymbol{\beta}; \sigma^2\mathbf{V})$  no qual tem-se a reparametrização na variância:

$$\sigma^2 \mathbf{R} + \tau^2 \mathbf{I} = \sigma^2 \left( \mathbf{R} + \frac{\tau^2}{\sigma^2} \mathbf{I} \right) = \sigma^2 (\mathbf{R} + v^2 \mathbf{I}) = \sigma^2 \mathbf{V}$$

Segundo Martins (2010), deve-se considerar os seguintes resultados da álgebra matricial:

$$\frac{\partial}{\partial \mathbf{T}} (\mathbf{C}\mathbf{T}) = \mathbf{C}' \quad (5)$$

$$\frac{\partial}{\partial \mathbf{T}} (\mathbf{T}'\mathbf{C}\mathbf{T}) = 2\mathbf{C}\mathbf{T}' \quad (6)$$

em que  $\mathbf{C}$  é uma matriz quadrada de orden  $n \times n$  e  $\mathbf{T}$  é um vetor de orden  $n \times 1$ .

Assim, pode-se reescrever o log da verossimilhança como:

$$l(\boldsymbol{\beta}, \sigma^2) = -\frac{1}{2} \{ n \ln(2\pi) + \ln(|\sigma^2 \mathbf{V}|) + (\mathbf{Y}(\mathbf{s}) - \mathbf{X}\boldsymbol{\beta})' (\sigma^2 \mathbf{V})^{-1} (\mathbf{Y}(\mathbf{s}) - \mathbf{X}\boldsymbol{\beta}) \}. \quad (7)$$

Em (7) tem-se que:

$$\begin{aligned} (\mathbf{Y}(\mathbf{s}) - \mathbf{X}\boldsymbol{\beta})' (\sigma^2 \mathbf{V})^{-1} (\mathbf{Y}(\mathbf{s}) - \mathbf{X}\boldsymbol{\beta}) &= \mathbf{Y}(\mathbf{s}) (\sigma^2 \mathbf{V})^{-1} \mathbf{Y}(\mathbf{s}) - \mathbf{Y}'(\mathbf{s}) (\sigma^2 \mathbf{V})^{-1} \mathbf{X}\boldsymbol{\beta} - \\ &- \boldsymbol{\beta}' \mathbf{X}' (\sigma^2 \mathbf{V})^{-1} \mathbf{Y}(\mathbf{s}) + \boldsymbol{\beta}' \mathbf{X}' (\sigma^2 \mathbf{V})^{-1} \mathbf{X}\boldsymbol{\beta}, \end{aligned}$$

Então, pode-se reescrever (7) por:

$$\begin{aligned} l(\boldsymbol{\beta}, \sigma^2) &= -\frac{1}{2} \left\{ n \ln(2\pi) + \ln(|\sigma^2 \mathbf{V}|) + \frac{1}{\sigma^2} \left[ \mathbf{Y}'(\mathbf{s}) \mathbf{V}^{-1} \mathbf{Y}(\mathbf{s}) - \right. \right. \\ &\quad \left. \left. - 2\mathbf{Y}'(\mathbf{s}) \mathbf{V}^{-1} \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\beta}' (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X}) \boldsymbol{\beta} \right] \right\} \quad (8) \end{aligned}$$

Então, aplicando-se (5) e (6) em (8), tem-se:

$$\frac{\partial}{\partial \boldsymbol{\beta}} l(\boldsymbol{\beta}, \sigma^2) = -\frac{1}{\sigma^2} \left[ -(\mathbf{Y}'(\mathbf{s}) \mathbf{V}^{-1} \mathbf{X})' + (\mathbf{X}' \mathbf{V}^{-1} \mathbf{X}) \boldsymbol{\beta} \right].$$

Considerando:

$$\frac{1}{\sigma^2} (\mathbf{X}' \mathbf{V}^{-1} \mathbf{Y}(\mathbf{s}) - \mathbf{X}' \mathbf{V}^{-1} \mathbf{X} \hat{\boldsymbol{\beta}}) = 0,$$

vem

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{V}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}^{-1}\mathbf{Y}(\mathbf{s}). \quad (9)$$

Na equação (9),  $\boldsymbol{\beta}$  só depende de  $\phi$  e  $v^2$ . E assim, reescrevendo (7), e considerando  $|\text{Var}(\mathbf{Y}(\mathbf{s}))| = |\sigma^2\mathbf{V}| = (\sigma^2)^n|\mathbf{V}|$ , tem-se:

$$l(\boldsymbol{\beta}, \sigma^2) = -\frac{1}{2} \left\{ n\ln(2\pi) + n\ln(\sigma^2) + \ln|\mathbf{V}| + \frac{1}{\sigma^2}(\mathbf{Y}(\mathbf{s}) - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{Y}(\mathbf{s}) - \mathbf{X}\boldsymbol{\beta}) \right\} \quad (10)$$

Derivando parcialmente (10) com relação a  $\sigma^2$

$$\frac{\partial}{\partial \sigma^2} l(\boldsymbol{\beta}, \sigma^2) = -\frac{1}{2} \left\{ \frac{n}{\sigma^2} + (\mathbf{Y}(\mathbf{s}) - \mathbf{X}\boldsymbol{\beta})'\mathbf{V}^{-1}(\mathbf{Y}(\mathbf{s}) - \mathbf{X}\boldsymbol{\beta}) \left( -\frac{1}{(\sigma^2)^2} \right) \right\}$$

desta forma, igualando esta derivada parcial a zero, obtém-se o estimador

$$\hat{\sigma}^2 = n^{-1} \left[ \mathbf{Y}(\mathbf{s}) - \mathbf{X}\hat{\boldsymbol{\beta}} \right]' \mathbf{V}^{-1} (\mathbf{Y}(\mathbf{s}) - \mathbf{X}\hat{\boldsymbol{\beta}}) \quad (11)$$

Substituindo (9) e (11) em (7), obtém-se o log da verossimilhança concentrada nos parâmetros  $v^2$  e  $\phi$ , dada por

$$l_c(\phi, v^2) = -\frac{1}{2} \{ n\ln(2\pi) + n\ln(\hat{\sigma}^2) + \ln|\mathbf{V}| + n \} \quad (12)$$

Assim, para obter as estimativas dos parâmetros, a equação (12) deve ser otimizada numericamente com relação aos parâmetros  $\phi$  e  $v^2$ . Obtendo as estimativas desses parâmetros, obtém-se a matriz  $\hat{\mathbf{V}}$ , que ao ser substituída em (9) obtém-se  $\hat{\boldsymbol{\beta}}$  e  $\sigma^2$ . Logo, substituindo  $\hat{v}^2$  e  $\hat{\sigma}^2$  em  $v^2 = \frac{\tau^2}{\sigma^2}$  obtém-se  $\hat{\tau}^2$

### 2.1.3 Variáveis Regionalizadas

Na atualidade, segundo Landim (2006), o termo Geoestatística 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.

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,  $Y(\mathbf{s})$  pode ser definida como uma variável aleatória que assume diferentes valores  $Y$  em função da posição  $\mathbf{s}$  dentro de uma região  $A$ .

O conjunto dessas variáveis  $Y_{(\mathbf{s})}$  medidas na totalidade da área  $A$ , pode ser considerada como uma função aleatória  $Y_{(\mathbf{s})}$  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  $Y_{(\mathbf{s})}$  ser uma particular realização de alguma função aleatória  $Y(\mathbf{s})$  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 localização  $\mathbf{s}$  tem-se somente uma realização  $Y_{(\mathbf{s})}$  e, como o número de pontos é finito, fica usualmente impossível inferir sobre a distribuição de  $Y_{(\mathbf{s})}$ . Assim, hipóteses que envolvem diferentes graus de homogeneidade espacial, denominadas de hipóteses de estacionariedade, se fazem necessárias.

As hipóteses de estacionariedade são (GRIPP, 1992):

- 1) Hipótese de estacionariedade: a função aleatória  $Y(\mathbf{s})$  é estacionária se a lei de probabilidades de  $Y(\mathbf{s})$  for a mesma em qualquer localidade  $\mathbf{s}$  considerada.
- 2) Hipótese de estacionariedade de ordem 2: a função aleatória  $Y(\mathbf{s})$  é estacionária de ordem 2 se  $E[Y(\mathbf{s})] = m$ , isto é, a esperança matemática de  $Y(\mathbf{s})$  é contante e não depende de  $\mathbf{s}$ .

Para o estudo do comportamento das variáveis regionalizadas, faz-se necessário utilizar duas ferramentas dos métodos geoestatísticos que são: Semivariograma e Krigagem.

### 2.1.3.1 Semivariograma

Considera-se um variável regionalizada  $Y(\mathbf{s})$  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  $\Delta\mathbf{h}$ , 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  $\Delta\mathbf{h}$  ao longo de  $\mathbf{h}$  é expressa por:

$$C(\mathbf{h}) = C(\Delta\mathbf{h}) = \frac{1}{N(\mathbf{h})} \sum Y_s Y_{s+\mathbf{h}} - m^2$$

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

A covariância depende do tamanho de  $\mathbf{h}$ , assim, se  $\mathbf{h} = 0$ ,  $C(\mathbf{h})$  representará a variância, isto é

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

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

$$\gamma(\mathbf{h}) = \frac{1}{2n} \sum_{i=1}^{N(\mathbf{h})} (Y_{s_i+\mathbf{h}} - Y_{s_i})$$

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  $\gamma(\mathbf{h})$  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  $\mathbf{h}$ , para desta maneira realizar um estudo na variável e analisar se os dados são estacionários, isto é, se não há tendência nos valores.

o ajuste de um modelo matemático aos dados no gráfico de  $\gamma(\mathbf{h})$  por  $h$ , isto é, a uma função, define os parâmetros do semivariograma, os quais são: efeito pepita ( $C_o$ ), que representa o valor de  $\gamma$  quando  $\mathbf{h}$ ; quando o valor de  $\mathbf{h}$  aumenta, frequentemente aumenta até um distância  $a$ , denominada de alcance (alcance da dependência espacial), também podendo ser expresso por  $\phi$ ; na distância  $a$ ,  $\gamma(\mathbf{h})$  é chamado de patamar, dado por  $C + C_o$ , sendo  $C$  a variância estrutural.

Na Figura 1, tem-se o gráfico de  $\gamma(\mathbf{h})$  por  $\mathbf{h}$ , no qual nota-se a disposição de cada parâmetro e observa-se um semivariograma experimental e o modelo matemático ajustado.

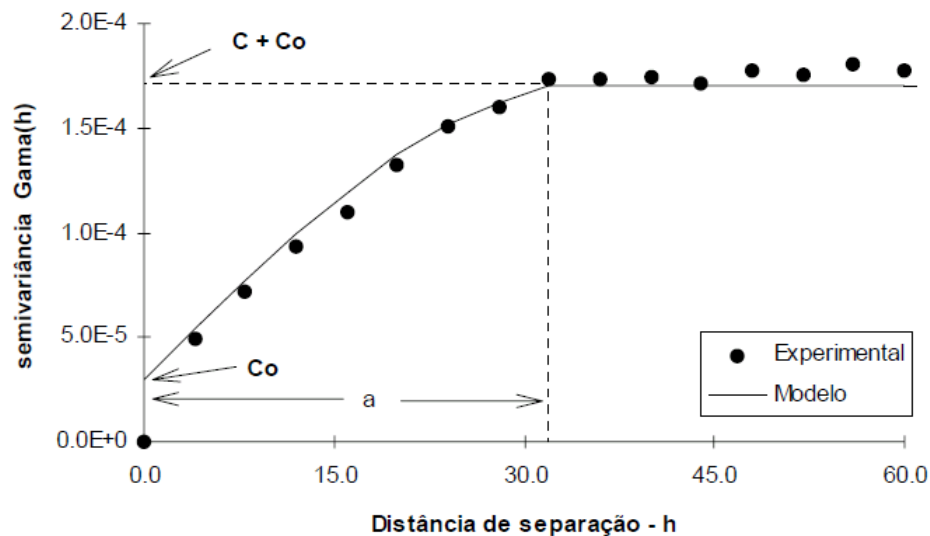


Figura 1 - Semivariograma experimental e Modelo matemático ajustado.

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 (2000), 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  $C_o$ , 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 formula  $E = \frac{C_o}{C}$  (GUERRA, 1988 apud LANDIM, 1996), em que:

- i.  $E < 0.15$ : componente aleatória pequena;
- ii.  $0.15 \geq E \geq 0.30$ : componente aleatória significativa;
- iii.  $E > 0.30$ : componente aleatória muito significativa.

O modelo matemático ajustado é dividido em duas categorias: modelo com patamar e sem patamar.

Modelo com patamar é aquele 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:

$$\text{i. Modelo Esférico: } \begin{cases} \gamma(\mathbf{h}) = C[\frac{3}{2}(\frac{h}{a}) - \frac{1}{2}(\frac{h}{a})^3], & \text{para } \mathbf{h} < a \\ \gamma(\mathbf{h}) = C, & \text{para } \mathbf{h} \geq a \end{cases}$$

$$\text{ii. Modelo Exponencial: } \gamma(\mathbf{h}) = C[1 - e^{-\frac{3h}{a}}]$$

$$\text{iii. Modelo Gaussiano: } \gamma(\mathbf{h}) = C[1 - e^{(-\frac{3h}{a})^2}]$$

Um outro modelo bastante utilizado é com a função Matérn que é dada por

$$\gamma(\mathbf{h}) = M(\mathbf{h}|v, a) = \frac{2^{1-v}}{\Gamma(v)} (a||\mathbf{h}||)^v K_v(a||\mathbf{h}||)$$



no qual  $\Gamma(\cdot)$  é a função Gama,  $K_v$  é a função Bessel de ordem  $v$ , e  $\|h\|$  é a distância euclidiana entre duas localizações. Um caso particular da função Matérn é quando o valor de  $v$  é 0.5, em que a função se iguala a função Exponencial.

Os modelos sem patamas, são modelos que como o próprio nome diz, não se estabilizam em um patamar. Exemplo dessa categoria de modelo tem-se o Modelo Linear e o Modelo Potencial

Segundo Landim(1996), após o ajuste de um variograma experimental, é importante realizar 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, verificando se o erro de estimação é 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.

### 2.1.3.2 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.

É um método para interpolação de dados. Quando está interpolação é realizada com apenas uma variável, denomina-se esse método de Krigagem e as mais utilizadas são Kri-

gagem Simples, na qual assume-se que a função média do processo estocástico é conhecida e o processo é estacionário, e Krigagem Ordinária, em que o processo estocástico é não estacionário. Quando tem-se a estimação de uma variável em um estudo conjunto com mais variáveis, tem-se a Cokrigagem, que será discutida mais adiante.

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

$$Y_{s_0}^* = \lambda_0 + \sum_{i=1}^n \lambda_i Y_{s_i}$$

em que  $n$  é o número de vizinhos medidos,  $Y_{s_i}$  valores utilizados na estimativa e  $\lambda_i$  são os ponderadores aplicados a cada  $Y_{s_i}$  que são selecionados de forma que a estimativa obtida seja não tendenciosa.

## 2.2 Geoestatística Multivariada

Segundo Bognola et al. (2008), os modelos multivariados são utilizados quando a variável que deseja-se estudar o comportamento é de difícil acesso, seja por causa de custo elevado ou algum outro motivo; porém, existe uma outra variável, variável secundária, correlacionada com a variável de interesse, sendo que esta secundária é de fácil acesso. Desta forma, combina-se algumas mensurações da variável de interesse com um número maior da secundária e a partir disso, pode-se realizar um estudo da variável de interesse sem muitas medições da mesma.

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), para a predição espacial utiliza-se o método da Cokrigagem.

### 2.2.1 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:

$$C = \begin{pmatrix} Cov(Y_1Y_1) & Cov(Y_1Y_2) & 1 \\ Cov(Y_2Y_1) & Cov(Y_2Y_2) & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}; W = \begin{pmatrix} \mathbf{e} \\ \mathbf{f} \\ -\lambda_1 \\ -\lambda_2 \end{pmatrix} eD = \begin{pmatrix} Cov(Y_1Y_0) \\ Cov(Y_2Y_0) \\ 1 \\ 0 \end{pmatrix}$$

temos que  $Y_1$  e  $Y_2$  são as variáveis primária e secundária, respectivamente,  $Y_0$  é referente aos valores que serão preditos;  $\mathbf{e}$  e  $\mathbf{f}$  são 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  $\lambda_2$ .

### 2.2.2 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 correionalização, 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 em 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 correionalização, 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 cruzado, em cada escala considerada.

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

### 2.2.3 Semivariograma Cruzado

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

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

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

Desta forma, o semivariograma cruzado será:

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

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.

#### 2.2.4 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:

$$Y_i = \mu_i + Z_i \quad i = 1, 2 \quad (13)$$

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 nulo 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 associado 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 que esta matriz é 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  $n_1 \times n_2$ .

### 2.2.5 Modelo Bivariado de Co-regionalização

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

Por esta abordagem, os termos latentes do modelo (3), são decomposto da forma 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  $S_2$  são 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 obsevações 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 a partição  $\Sigma_{1,(l,k)}$  da matriz de covariância é dado por  $Cov(h) = \sigma_{11}^2 \rho_1(h)$  e a partição  $\Sigma_{2,(l,k)}$  é dado 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 para  $S_1$  e  $S_2$ , respectivamente.

Por propriedades básicas de covariância, encontra-se que  $\Sigma_{1,2}$  é equivalente a  $\sigma_1 \sigma_2 R_1$  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  $n_1 \times n_2$ . Então,  $\Sigma_Y$  fica estruturada.

Supondo que existam apenas duas localizações que foram amostradas, então 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:

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

Desta maneira, segundo Fonseca (2008), a distribuição de probabilidade do vetor  $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2)$  está estabelecida e depende do vetor de parâmetros  $\boldsymbol{\theta} = (\boldsymbol{\beta}^*, \boldsymbol{\sigma}^*, \boldsymbol{\phi}_1^*, \boldsymbol{\phi}_2^*)$  no qual  $\boldsymbol{\beta}^*$  é o vetor de parâmetros associados a  $\mu$  e  $\boldsymbol{\phi}^*$  é o 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:

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

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

### 2.2.6 Modelo gaussiano bivariado com componente comum

Para problemas bivariados, Diggle e Ribeiro Jr (2007) propuseram uma abordagem denominada de *Bivariate Gaussian common component model* (BGCCM), em que no modelo (3), seu termo  $Z_i$ , campo aleatório latente, é decomposto em:

$$Z_i = \sigma_{0i}S_0 + \sigma_i S_i, \quad i = 1, 2 \quad (14)$$

sendo  $\boldsymbol{\sigma}^* = (\sigma_{01}, \sigma_1, \sigma_{02}, \sigma_2)$  um vetor de parâmetros de dispersão associados ao modelo (1);  $S_0$ ,  $S_1$  e  $S_2$  são campos aleatórios gaussianos mutuamente independentes, com vetores de médias nulos, variância unitárias e correlações obtidas por funções de correlação válidas. Segundo Fonseca (2008), as correlações destes campos aleatórios, quando combinadas desta forma, vão gerar correlações cruzadas válidas entre as variáveis estudadas  $Y_1$  e  $Y_2$ , já que o campo aleatório  $S_0$  é comum as duas variáveis. Desta forma, deve-se adotar três funções de correlação para  $S_0$ ,  $S_1$  e  $S_2$  e assim gerar uma matriz  $\Sigma_Y$  válida.

Então, com a decomposição de  $Z_1$  e  $Z_2$ , o modelo (1) pode ser reescrito por:

$$\begin{cases} Y_1 = \mu_1 + \sigma_{01}S_0 + \sigma_1 S_1 \\ Y_2 = \mu_2 + \sigma_{02}S_0 + \sigma_2 S_2 \end{cases}$$

Seja  $Y_i(s_l)$  e  $Y_i(s_k)$  observações da variável  $Y_i$  realizadas nas localizações  $s_l$  e  $s_k$ , separadas por uma distância euclidiana  $h = h_{lk}$ , com  $l, k = 1, \dots, n_i$  e  $i = 1, 2$ . Assim, o elemento  $\Sigma_{i,(l,k)}$  é representado por  $Cov(h) = \sigma_{0i}^2 \rho_0(h) + \sigma_i^2 \rho_i(h)$ , no qual  $\rho_0$  e  $\rho_i$  são as funções de correlação para  $S_i$ ,  $i = 0, 1, 2$ . Pelas propriedades básicas de covariância, encontra-se que  $\Sigma_{1,2}$  é igual a  $\sigma_{01}\sigma_{02}R_0$ , em que  $R_0$  é uma matriz, de dimensão  $n_1 \times n_2$ , das correlações cruzadas entre as variáveis e depende da função de correlação atribuída para  $S_0$ .

Considerando apenas duas localizações amostradas,  $s_k$  e  $s_l$ , e sendo  $h_{kl} = h$  a distância entre essas localizações, tem-se que a matriz de covariâncias para o vetor  $(Y_i(s_l), Y_i(s_k), Y_j(s_l), Y_j(s_k))$  é dada por:



$$\Sigma_Y = \begin{bmatrix} \sigma_{01}^2 + \sigma_1^2 & \sigma_{01}^2 \rho_0(h) + \sigma_1^2 \rho_1(h) & \sigma_{01} \sigma_{02} & \sigma_{01} \sigma_{02} \rho_0(h) \\ \sigma_{01}^2 \rho_0(h) + \sigma_1^2 \rho_1(h) & \sigma_{01}^2 + \sigma_1^2 & \sigma_{01} \sigma_{02} \rho_0(h) & \sigma_{01} \sigma_{02} \\ \sigma_{01} \sigma_{02} & \sigma_{01} \sigma_{02} \rho_0(h) & \sigma_{02}^2 + \sigma_2^2 & \sigma_{02}^2 \rho_0(h) + \sigma_2^2 \rho_2(h) \\ \sigma_{01} \sigma_{02} \rho_0(h) & \sigma_{01} \sigma_{02} & \sigma_{02}^2 \rho_0(h) + \sigma_2^2 \rho_2(h) & \sigma_{02}^2 + \sigma_2^2 \end{bmatrix}$$

Desta forma, a distribuição de probabilidade do vetor  $\mathbf{Y} = (\mathbf{Y}_1, \mathbf{Y}_2)$  está estabelecida e depende do vetor de parâmetros  $\boldsymbol{\theta} = (\boldsymbol{\beta}^*, \boldsymbol{\sigma}^*, \boldsymbol{\phi}_0^*, \boldsymbol{\phi}_1^*, \boldsymbol{\phi}_2^*)$  no qual  $\boldsymbol{\beta}^*$  é o vetor de parâmetros associados a  $\mu$  e  $\boldsymbol{\phi}_j^*$  é o vetor de parâmetros relacionados a escolha da função  $\rho_j$ , com  $j = 0, 1, 2$ .

Para a estimação dos parâmetros desse modelo e a predição espacial, utiliza-se as técnicas dos modelos geoestatísticos univariados porém, é necessário um reparametrização adicional em  $\boldsymbol{\theta}$  para obter a função de verossimilhança concentrada.

### 2.2.7 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 modelos e ajusta melhor aos dados. Para este tipo de modelagem, utiliza-se a função da família Mátern, que tem sido muito utilizada na atualidade, que é uma função bem flexível, o que facilita na sua utilização.

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

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

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

Em um processo multivariado, seja o vetor de variáveis  $\mathbf{Y}(\mathbf{s}) = (\mathbf{Y}_1(\mathbf{s}), \dots, \mathbf{Y}_p(\mathbf{s}))'$ , em que cada localização  $\mathbf{s} \in \mathbb{R}^d$  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

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

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

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

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

### 2.2.7.1 Modelo Mátern Multivariado

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

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

com  $n = 0, 1, \dots$

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

#### 2.2.7.1.1 Modelo Mátern Multivariado Parcimonioso

(AUMENTAR)



## 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.
- BOX, G.; COX, D. **An analysis of transformation.** *Journal of the Royal Statistical Society, Series B*, p. 211-252, 1964.
- DIGGLE, P. J.; RIBEIRO JR., P. J.; CHRISTENSEN, O. F. **An introduction to model-based geostatistics.** In: MØLLER, J. (Ed.). *Spatial Statistics and Computational Methods.* New York: Springer, 2003. p. 43-86.
- 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.
- GRIPP, A.H. **Utilização de Geoestatística para o tratamento de dados de prospecção geoquímica.** *Revista Brasileira de Geociências*, 22(2):248-251. Junho de 1992
- 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/>>.
- MARTINS, A.B.T. **Análise geoestatística de dados composicionais.** Curitiba, 2010. 200 p. Tese (Doutorado) - Universidade Federal do Paraná.

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. 91 p. 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.rproject.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.

SOARES, A. **Geoestatística para ciências da terra e do ambiente**. Segunda edição. IST Press, 2006

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.

