



UNIVERSIDADE FEDERAL DO PARÁ
INSTITUTO DE CIÊNCIAS EXATAS E NATURAIS
PROGRAMA DE PÓS-GRADUAÇÃO EM MATEMÁTICA E ESTATÍSTICA

Wanderson Cunha Pereira

Simulação Geoestatística via Métodos Convolutivos
Orientador: **Prof. João Marcelo Brazão Protázio, Dr.**

Belém

2011

Wanderson Cunha Pereira

Simulação Geoestatística via Métodos Convolutivos

Dissertação apresentada ao Programa de Pós-Graduação em Matemática e Estatística, da Universidade Federal do Pará, como requisito para a obtenção do grau de Mestre em Estatística.

Orientador: **Prof. João Marcelo Brazão Protázio, Dr.**
Área de Concentração: **Geoestatística**

Belém
2011

Wanderson Cunha Pareira

Simulação Geoestatística via Métodos Convolutivos

Esta Dissertação foi julgada e aprovada, para a obtenção do grau de Mestre em Estatística, no Programa de Pós-Graduação em Matemática e Estatística, da Universidade Federal do Pará.

Belém, 17 de novembro de 2011

Prof. Giovany de Jesus Malcher Figueiredo, *Dr.*

(Diretor do Programa de Pós-Graduação em Matemática e Estatística da UFPA)

Banca Examinadora

Prof. João Marcelo Brazão Protázio, *Dr.*

Universidade Federal do Pará

Orientador

Prof. João dos Santos Protázio, *Dr.*

Universidade Federal do Pará

Examinador

Prof. Mauro de Lima Santos, *Dr.*

Universidade Federal do Pará

Examinador

Prof. Paulo Cerqueira dos Santos, *Dr.*

Universidade Federal do Pará

Examinador

Prof. Sérgio Sacani Sancevero, *Dr.*

Unicamp

Examinador

À minha família que tanto amo.

Agradecimentos

Por saber que Deus é o único capaz de fornecer as condições necessárias ao homem e ajudá-lo a transpor tantos obstáculos que surgem no decorrer da vida, e por crer na sua existência, agradeço a Ele, primeiramente, por ter me dado força, saúde e inteligência para chegar até o fim de meu curso e deste trabalho.

Dedico meus sinceros agradecimentos:

- À minha família que deu todo o apoio necessário para continuar meu curso e que soube compreender-me nos momentos em que mais precisei;
- Ao meu grande companheiro Paulo André que tem me mostrado o verdadeiro sentido de felicidade, bem como sua maravilhosa família;
- Ao Professor e orientador João Marcelo Brazão Protázio que me ensinou o verdadeiro significado da palavra humildade;
- A minha *família postiça* que esteve sempre o meu lado e mostrou-me que, a união faz a força. Em especial: Admilson Alcântara e Shirley Reis;
- Aos amigos que nos momentos de necessidade me confortaram e me ajudaram, seja direta ou indiretamente;
- Aos colegas de estudo Gilson Prata, Jonelson Dias e Vanessa Pamplona, em especial a companheira de todas as horas Maura da Silva Costa Furtado;
- A todos os professores do PPGME que fizeram parte desse processo acadêmico.

*“O conhecimento torna a alma jovem e
diminui a amargura da velhice.
Colhe, pois, a sabedoria.
Armazena suavidade para o amanhã”.*

Leonardo da Vinci

Resumo

PEREIRA, Wanderson Cunha. Simulação Geoestatística via Métodos Convolutivos. 2011. Dissertação (Mestrado em Matemática e Estatística) - PPGME, UFPA, Belém, Pará, Brasil.

Este trabalho tem como principal objetivo propor um novo método de simulação mais eficiente - o Método de Convolução - e compará-lo com um mais utilizado na literatura - o Método de Krigagem, mostrando tal eficiência tanto em facilidade quanto em custo computacional, simulando diversos processos espaciais a partir do uso de diversos modelos de processos convolutivos. O trabalho envolve duas etapas principais: a utilização da Simulação via método de convolução e comparação dos resultados encontrados mostrando a eficiência do método proposto pelo trabalho. Na simulação via método de convolução mostra-se as vantagens de seu uso, onde as maiores delas são a especificação de um modelo espacial de covariograma e a simplificação de modelos que diminuem consideravelmente o esforço computacional. Para tal processo utilizou-se técnicas matemáticas, estatísticas e geoestatísticas, bem como, para alcançar os resultados desejados, a transformação Box-Cox e a transformada de Fourier. Nas realizações à 1D, utilizou-se um processo ruído-branco com discretização $n = 512$ em um suporte de comprimento 6, para a convolução entre o mesmo e o kernel associado ao modelo de covariância, e os modelos espaciais de covariância gaussiano e exponencial, comparou-se seus variogramas e chegou-se a conclusão desejada: o método de convolução aplicado à 1D é mais eficiente. Nas realizações à 2D, para o conjunto de dados utilizados – Data Set Jura, variável Cádmio – o método proposto mostrou-se mais robusto que a Krigagem.

Palavras-Chaves: Convolução, Krigagem, Simulação.

Abstract

PEREIRA, Wanderson Cunha. Simulation Geostatistic through Convolution Methods. 2009. Dissertation (Master in Mathematics and Statistics) - PPGME, UFPA, Belém, Pará, Brazil.

In this survey has been as a main target to propose a new method of simulation more efficient - the Convolution Method - and compares it with the most used in the literature - the Ordinary Kriging Method, showing its efficiency as easy as in computer cost, simulating various wide processes through the use of various models of convoluted processes. The paper involves two main steps: Simulation through the useful of a Convolution Method and comparison of the showing results and the efficiency of the method proposed for the survey. The simulations through the Convolution Method shows the advantages of its useful, wherever the largest of them are the specification of a spatial model covariogram and simplification of models that reduce the computational effort greatly. Such processing was used to the mathematical techniques, statistics and geostatistics to achieve the required results as well, the Box-Cox Transformation and the Fourier Transform. In the efforts to 1D, you used a white-noise process with discretization $n = 512$ on a support of length 6, for convolution between it and the kernel associated with the model covariance and the Covariance Models Spatial Gaussian and Exponential, compared to their variograms and came up the required conclusion: the method applied to 1D convolution is more efficient. In the efforts to 2D, for the Data Set Jura, variable cadmium – the proposed method was more robust than the Kriging, best reproducing the variability of the same Data Set Jura.

Key Words: Convolution, Kriging, Simulation

Índice

Resumo	vii
Abstract	viii
Lista de Figuras	xii
1 Introdução	
1.1 Aspectos Gerais.....	13
1.2 Justificativa e Importância.....	16
1.3 Objetivos.....	17
1.3.1 Objetivo Geral.....	17
1.3.2 Objetivos Específicos.....	17
2 Métodos	
2.1 História e Definição da Geoestatística.....	18
2.2 Krigagem Ordinária.....	20
2.2.1 Variância da Krigagem.....	21
2.2.2 Variância da Interpolação.....	21
2.3 Semivariograma.....	22
2.4 Principais Modelos de Variogramas.....	25
2.5 Simulação: o que é e para que serve?.....	28
2.6 Simulação Estocástica X Krigagem.....	29
2.7 Método de Kernel.....	32
2.8 Transformação Box-Cox.....	33
3 Convolução à 1D, à 2D e Resultados	
3.1 Modelo Espacial de Cavariância Gaussiano.....	35
3.2 Modelo Espacial de Cavariância Exponencial.....	35
3.3 Convolução à 1D.....	36
3.4 Convolução à 2D.....	37
3.4.1 Filtragem no Domínio Espacial.....	38

3.4.2	Filtragem no Domínio da Frequência.....	40
3.4.3	Metodologia para Simulação à 2D.....	41
3.5	Operações com Máscaras.....	42
3.6	Resultados 1D.....	46
3.6.1	Convolução à 1D.....	46
3.6.2	Testes realizados a partir do Modelo de Covariância Gaussiano.....	46
3.6.3	Testes realizados a partir do Modelo de Covariância Exponencial.....	48
3.6.4	Metodologia para Simulação à 1D.....	49
3.6.5	Simulação para 100 Realizações.....	51
3.7	Resultados 2D.....	52
4	Considerações Finais	58
	Referências Bibliográficas	

Lista de Figuras

2.1	Esquema de Semivariograma Experimental.....	22
2.2	Semivariogramas com Patamar.....	32
3.1	Esquema de uma imagem.....	37
3.2	(a) Imagem Monocromática, (b) Pixel de uma Imagem, (c) Imagem em Composição Colorida.....	37
3.3	Esquema do processo de convolução por filtragem espacial utilizando uma máscara $n \times m$	38
3.4	Vizinhança 3×3 ao redor de um ponto (x,y) em uma imagem.....	39
3.5	Esquema de processamento do Domínio da Frequência usando a Transformada de Imagem.....	40
3.6	Esquema ilustrando os passos da filtragem no Domínio de Fourier.....	42
3.7	Cálculo do primeiro valor da convolução de A por B.....	45
3.8	Ruído-Branco com discretização $n = 512$ e seu respectivo variograma.....	46
3.9	Modelo de Variograma Gaussiano com $a = 1$ e seu kernel associado.....	46
3.10	Modelo de Variograma Gaussiano com $a = 0,5$ e seu kernel associado.....	47
3.11	Modelo de Variograma Gaussiano com $a = 0,1$ e seu kernel associado.....	47
3.12	Modelo de Variograma Exponencial com $a = 1$ e seu kernel associado.....	48
3.13	Modelo de Variograma Exponencial com $a = 0,5$ e seu kernel associado.....	48
3.14	Modelo de Variograma Exponencial com $a = 0,1$ e seu kernel associado.....	49
3.15	Processo ruído-branco com respectivo Modelo Gaussiano associado.....	49
3.16	Comparação entre os variogramas dos Modelos: Gaussiano e Processo de Simulação.....	50
3.17	Processo ruído-branco com respectivo Modelo Exponencial associado.....	50
3.18	Comparação entre os variogramas dos Modelos: Exponencial e Processo de Simulação.....	51

3.19	Esquerda: Variograma das 100 simulações. Direita: Média das simulações comparadas com o variograma do Modelo Teórico Gaussiano.....	51
3.20	Esquerda: Variograma das 100 simulações. Direita: Média das simulações comparadas com o variograma do Modelo Teórico Exponencial.....	52
3.21	Mapa de Localização do Metal Cádmio.....	53
3.22	Esquerda: Histograma da Concentração de Metal Cádmio. Direita: Curva de Normalidade.....	54
3.23	Esquerda: Histograma dos dados transformados por Box-Cox da Concentração de Metal Cádmio. Direita: Curva de Normalidade.....	55
3.24	(a) Técnica de Krigagem. (b) Simulação via Kernel.....	55
3.25	(a) Histograma e Curva de Normalidade do Processo de Krigagem. (b) Histograma e Curva de Normalidade da Simulação via Kernel.....	56

Capítulo 1

Introdução

1.1 Aspectos Gerais

Nas aplicações da estatística nos estudos das ciências da terra usualmente se modela a medida de um determinado experimento ou observações por meio de uma variável aleatória. Entretanto as características peculiares destes fenômenos levou à modelagens mais complexas tais como vetores aleatórios e funções aleatórias.

A dependência espacial dos dados, ou seja, as posições relativas onde foram observados os diversos valores devem ser levados em conta no modelo (Braga, 1990).

Em outras palavras, na estatística clássica as amostras são aleatórias e independentes de uma distribuição de probabilidade simples, esta suposição é chamada estacionaridade, sua aplicação não envolve qualquer conhecimento da posição atual das amostras ou da relação entre elas. Já na geoestatística tem-se que a distribuição das diferenças de variáveis entre dois pontos amostrados é a mesma para toda área, e que isto depende somente da distância entre eles e da orientação dos pontos.

A busca de técnicas e estimadores mais eficientes tem provocado grandes preocupações entre os estudiosos que procuram determinar as características de uma área, ou uma porção da superfície terrestre, que são alvos não só na ciência do solo, mas também em outras áreas do conhecimento. Para facilitar esses estudos, a geoestatística surge levando em consideração a localização geográfica e a dependência espacial.

As propriedades naturais da superfície terrestre são espacialmente contínuas, sendo restritivo descrevê-la através de simples funções matemáticas que não respondem questões, tipo: qual o tamanho do domínio real ou da janela de estimativa? Que forma de orientação deve ter a janela para se obter uma estimativa ótima? Quais são os erros associados aos valores estimados?

A geoestatística é um ramo da estatística que introduz o conceito de variável aleatória regionalizada. Através desta é possível estimar o valor de uma dada propriedade (teor de minério, por exemplo), para cada posição da área de estudo, valor este condicionado aos dados existentes (amostra de solo, testemunho, etc ...) e a uma função de correlação espacial entre estes dados.

Este ramo da ciência nasceu na tentativa de solucionar parte dos problemas geológicos oriundos da mineração e foi concebida e idealizada por Matheron (1963). Constitui um critério científico e moderno de interpretação utilizado no estudo de grande parte dos fenômenos naturais.

As variáveis regionalizadas têm um comportamento espacial mostrando características intermediárias entre as variáveis verdadeiramente aleatórias e as totalmente determinísticas. Apresentam uma aparente continuidade no espaço, sendo representadas por funções numéricas ordinárias que assumem um valor definido a cada ponto no espaço e matematicamente descrevem um fenômeno natural (Garcês, 2009).

A continuidade geográfica atribuída se manifesta pela propriedade que a variável tem de apresentar valores muito próximos em dois pontos vizinhos e progressivamente mais diferentes à medida que os pontos vão ficando mais distantes (Landim, 1998). Existem outras propriedades que a variável regionalizada pode apresentar: localização, anisotropia e transição.

Por ser a variável regionalizada bastante complexa, impossibilita a formulação de um modelo teórico, deixando como única solução a determinação empírica ou relativa das probabilidades presentes. Como uma variável regionalizada é o resultado único de uma função aleatória, é possível fazer inferências estatísticas baseando-se em apenas uma amostra; para resolver esse impasse, a geoestatística utiliza uma restrição estacionária chamada de hipótese intrínseca e permite o uso de resultados de uma variável regionalizada por estimativas pelo método dos momentos. Nesta hipótese, a propriedade ocorre quando a média e a covariância, estimadas a partir de um conjunto restrito de valores, fornecem estimativas não-tendenciosas para o conjunto de total de valores.

Desta forma, os métodos geoestatísticos de estimativa constituem uma importante ferramenta na interpretação de dados geológicos. São, por definição, procedimentos matemáticos que ajustam uma função para determinados pontos de dados para assim se poder estimar o valor de uma variável em um ponto não amostrado (Yamamoto, 1998).

Furuie (2009) afirma que entre esses métodos é possível diferenciar os globais e os locais: os primeiros encontrando uma função que seja capaz de interpolar um valor qualquer local dentro do domínio dos dados, sendo que qualquer alteração em algum ponto se propaga por todo o restante do domínio; os métodos locais definindo funções válidas para pequenas porções do mapa que são aplicadas sucessivamente até cobrir toda área de interesse, com alterações pontuais afetando apenas os pontos mais próximos a estas alterações.

Entre os métodos globais usados na interpolação de dados geológicos, é possível citar a regressão polinomial (usada no ajuste de superfícies de Tendência) e as equações multiquadráticas (que visa representar superfícies por meio da soma de superfícies variadas matematicamente definidas). Por outro lado, os métodos locais seriam representados pela interpolação linear em triângulos, o inverso da distância e as funções de base radial. Estas últimas, se valendo de algumas funções base conhecidas para a determinação do valor da variável de acordo com a distância desta ao ponto amostrado, sendo que no caso de a função ser baseada em um modelo de variograma esta função passa a ser denominada de Krigagem.

O termo Krigagem é derivado do nome de Daniel G. Krige, que foi o pioneiro em introduzir o uso de médias móveis para evitar a superestimação sistemática de reservas em mineração (Delfiner e Delhomme, 1975).

À Krigagem é associado às técnicas de regressão linear que minimizam a variância das estimativas definidas por um modelo anterior para uma dada covariância. Porém, apesar de ser um ótimo procedimento de interpolação para se determinar precisamente dados espacialmente localizados, a Krigagem apresenta algumas limitações que restringem a sua aplicabilidade, tipo: ocorrência de pesos negativos e, conseqüentemente, teores negativos; caráter homoscedástico da variância de Krigagem; efeito de suavização das estimativas de Krigagem. Entre essas limitações, a mais vista na literatura, é a que se chama de efeito de suavização da Krigagem ordinária, que é uma das formas mais comuns de Krigagem. Além da Krigagem ordinária pode-se citar ainda, a Krigagem simples e universal.

Tal efeito justifica o fato da Krigagem ser considerada uma técnica de interpolação de precisão local (em contraste com as técnicas consideradas de precisão global), uma vez que este método tende a suavizar os valores de atributos dos dados, acarreta uma perda no modelo de variabilidade da distribuição como um todo, ou seja, seu modelo global (Furuie, 2009).

Como uma maneira de solucionar essas limitações do método da Krigagem ordinária, comuns a todas as técnicas de interpolação que visam a minimização dos erros, há outro

método alternativo para se tentar chegar a um modelo a partir de alguns poucos dados pontuais, este método sendo chamado de Simulação Estocástica. Essa técnica consiste basicamente na criação de algoritmos para se simular o comportamento de determinada variável, no caso, o seu comportamento espacial.

Ao se referir ao termo estocástico especificamente se refere ao fato destes algoritmos não serem determinísticos, ou seja, quando há mais de uma opção entre o valor a ser assumido em determinado ponto de execução do algoritmo, tal escolha será aleatória gerando inúmeras possíveis realizações a partir de um mesmo conjunto inicial de dados.

Assim, geoestatística é justamente um nome associado com uma classe de técnicas utilizadas para analisar e inferir valores de uma variável distribuída no espaço e/ou no tempo. Tais valores são implicitamente assumidos como correlacionados com outros, e o estudo de tal correlação é denominado de análise estrutural ou modelagem do semivariograma. Depois da análise estrutural, inferências em localizações não amostradas são realizadas usando Krigagem ou podem ser simuladas usando Simulações Condicionais.

Portanto, a incorporação de procedimentos geoestatísticos em um Sistema de Informações Geográficas (SIG), baseados em técnicas de Krigagem e/ou simulação condicional, são importantes, porque essa associação melhora os procedimentos tradicionais de tais sistemas devido à qualidade dos estimadores e, principalmente, pela acurácia nas análises dos dados

1.2 Justificativa e Importância

A Geoestatística, ao contrário da estatística clássica, possibilita separar a variabilidade explicada, pela relação existente entre as amostras e a variabilidade casual, sua aplicação, pode diminuir o número de amostras necessárias, para descrever as características de um determinado local, reduzindo custos de levantamento, mantendo a precisão das estimativas realizadas (Viera et al., 1983).

A simulação estocástica, método este utilizado neste trabalho é também um métodos que visa minimizar a variância dos erros e tentar chegar ao melhor modelo a partir de alguns poucos dados pontuais. Se criar algoritmos para simular o comportamento de determinada variável é a técnica básica deste método.

Este trabalho pretende, através da simulação, mais especificamente “Simulação Geoestatística via Métodos Convolutivos”, tema do mesmo, mostrar que tal método é mais vantajoso em relação a um outro método mais utilizado na literatura: o da Krigagem Ordinária. O modelo criado a partir da convolução entre um ruído branco e o kernel associado a uma função de covariância pode ser tanto vantajoso no seu custo computacional quanto na possibilidade da especificação do modelo espacial de covariância.

1.3 Objetivos

1.3.1 Objetivo Geral

O objetivo geral deste trabalho é propor um novo método de simulação - Método de Convolução - mais eficiente e com menor custo computacional do aquele utilizado com maior frequência na literatura - o da Krigagem Ordinária.

1.3.2 Objetivos Específicos

Os objetivos específicos desta dissertação podem ser resumidos como:

- (i) Apresentar os processos do Método de Convolução;
- (ii) Aplicar o Método de Convolução nos dados utilizados;
- (iii) Simular os dados utilizados através do Método Convolutivo;
- (iv) Comparar os métodos utilizados - o Convolutivo e a Krigagem Ordinária;
- (v) Mostrar a eficiência do método proposto.

Capítulo 2

Métodos

Neste capítulo serão apresentados os procedimentos metodológicos do estudo, onde inicialmente será abordada a área de estudo, as técnicas geoestatísticas que fundamentam esse trabalho e finalmente a técnica de simulação estocástica aqui utilizada - Método de Kernel (Convolução).

2.1 História e Definições da Geoestatística

A variabilidade espacial de algumas características do solo vem sendo uma das preocupações de pesquisadores praticamente desde o início do século. Os procedimentos usados na época baseavam-se na estatística clássica e utilizavam grandes quantidades de dados amostrais, visando caracterizar ou descrever a distribuição espacial da característica em estudo. Entende-se por estatística clássica aquela que se utiliza de parâmetros como média e desvio padrão para representar um fenômeno e se baseia na hipótese principal de que as variações de um lugar para o outro são aleatórias.

Krige (1951), ao trabalhar com dados de concentração de ouro, chegou a conclusão de que somente a informação dada pela variância seria insuficiente para explicar o fenômeno em estudo. Desta maneira, seria necessário levar em consideração a distância entre as observações. A partir daí surge o conceito de Geoestatística, que leva em consideração a localização geográfica e a dependência espacial.

Baseado nestas observações, Matheron (1963, 1971) desenvolveu uma teoria, a qual chamou de Teoria das Variáveis Regionalizadas, que contém os fundamentos da Geoestatística. Segundo ele, a geoestatística é uma função que varia de um lugar para o outro no espaço com certa aparência de continuidade. São variáveis cujos valores amostrados estão relacionados com a posição espacial que ocupam, onde assumem valores diferentes em diferentes lugares de observação, com certa independência de um lugar a outro.

Uma variável regionalizada é uma função numérica com distribuição espacial, que varia de um ponto a outro com continuidade aparente, mas cujas variações não podem ser representadas por uma função matemática simples. A continuidade atribuída às variáveis regionalizadas, se observa a partir da tendência de se tomar valores mais próximos em dois pontos amostrados, quanto menos afastados estão os referidos pontos. Segundo Guerra (1988), citado por Garcêz (2009), entende-se por continuidade uma flutuações mais ou menos importante entre valores de amostras vizinhas, refletindo um alto grau, de dependência ou independência, que existe entre um valor e outro.

Geograficamente, a continuidade atribuída se manifesta pela propriedade que a variável tem de apresentar valores muito próximos em dois pontos vizinhos e progressivamente mais diferentes à medida que os pontos vão ficando mais distantes (Landim,1998). Essa característica, que as variáveis regionalizadas possuem, representa a base de toda análise Geoestatística.

A Geoestatística envolve análise e inferência de fenômeno espacial e ou temporal. Atualmente, Geoestatística é justamente um nome associado com uma classe de técnicas usadas para analisar e inferir valores de uma variável distribuída no espaço e/ou no tempo. Tais valores são implicitamente assumidos ser correlacionados com outros, e se dá o nome de Análise Estrutural a esse estudo de correlações entre as variáveis.

Segundo Burrough (1987), a variação espacial de uma variável regionalizada pode ser expressa pela soma de três componentes: a) uma componente estrutural, associada a um valor médio constante ou uma tendência constante; b) uma componente aleatória, espacialmente correlacionada; c) um ruído aleatório ou erro residual.

Se x representa uma posição em uma, duas ou três dimensões, então o valor da variável Z , em x , é dada por (Burrough, 1987):

$$Z(h) = m(x) + \varepsilon'(x) + \varepsilon''$$

onde, $m(x)$ é uma função determinística que descreve a componente estrutural de Z em x ; $\varepsilon''(x)$ é um termo estocástico, que varia localmente e depende espacialmente de $m(x)$ e ε'' é um ruído aleatório não correlacionado, com distribuição normal com média zero e variância σ^2 .

2.2 Krigagem Ordinária

Dentre as várias técnicas abrangidas pela Krigagem, a Krigagem Simples (KS) é a sua forma mais básica já que seu modelo é o mais simples em termos de sua formulação matemática. É basicamente uma combinação linear de variáveis aleatórias nas localidades x_i , sendo seu estimador dado por:

$$Z^*_{KS}(x_0) = m + \sum_{i=1}^k \lambda_i (z(x_i) - m)$$

Onde $Z^*_{KS}(x_0)$ é o estimador na localidade x_0 , $Z(x)$ é a função de média m , e λ_i , são os ponderadores obtidos a partir da resolução de sistema linear de equação da Krigagem. Neste caso, a média é considerada constante em todo o domínio do fenômeno espacial.

A krigagem ordinária (KO), por sua vez, se diferencia da KS por não necessitar da média para se encontrar os pesos que minimizam a variância do erro da estimativa. No entanto, apresenta algumas exigências, tal como a estimativa da média local, que resulta na soma unitária dos ponderadores λ_i , e que a variável regionalizada seja estacionária, ou seja, não apresente deriva ou resíduos resultantes dela. Sua fórmula geral é:

$$Z^*_{KO}(x_0) = \sum_{i=1}^k \lambda_i z(x_i)$$

A KO é um método robusto que minimiza a variância do erro, não apresenta enviesamento, além de ser um interpolador exato, ou seja, honra os pontos dos dados amostrais. Outra característica da KO se refere à precisão local das estimativas resultantes em termos da correlação entre os valores estimados e os valores dos pontos de dados.

No entanto, em decorrência de ser um algoritmo de interpolação com base na minimização da variância do erro, a Krigagem como um todo tende a suavizar a variação espacial da informação em questão em suas estimativas. Uma decorrência desse fato é que valores abaixo da média serão superestimados e valores maiores que a média serão subestimados (Goovaerts, 1998).

Para se avaliar a qualidade do estimador procede-se por meio da incerteza associada, ou seja, por meio da dispersão em torno da estimativa. Da KO, se pode derivar dois tipos de variância, usados para se avaliar o grau de confiabilidade das estimativas: a variância da Krigagem e a variância da interpolação.

2.2.1 Variância da Krigagem

A variância da Krigagem é resultante da resolução do sistema de equações de Krigagem ordinária e, por definição, é igual a variância mínima, para um dado arranjo espacial:

$$\sigma^2_{KO} = C(0) - \sum_i \lambda_i \text{Cov}\{Z(x_i), Z(x_0)\} - \mu$$

A variância de Krigagem não depende dos valores dos pontos de dados utilizados na Krigagem e, portanto, apresenta um caráter homoscedástico.

2.2.2 Variância da Interpolação

Ao contrário da variância da Krigagem, a variância de interpolação depende dos valores dos dados, sendo definida como a média ponderada das diferenças ao quadrado entre os valores dos pontos dos dados e a estimativa $Z^*_{KO}(x_0)$. Desta forma, é calculada como:

$$S_0^2 = \sum_{i=1}^n \lambda_i [z(x_i) - z^*_{KO}(x_0)]^2$$

Como a variância de interpolação é dependente dos dados, tem-se que:

- i) Corresponde à propriedade da exatidão da Krigagem, assim se o ponto a ser estimado for um ponto de dado, o peso será 1; caso contrário, 0. O que resulta numa variância de interpolação igual a 0.
- ii) A variância de interpolação aumenta proporcionalmente à dispersão dos dados.
- iii) Leva em consideração a distância estrutural do variograma por conta dos ponderadores da Krigagem, ou seja, quanto mais influente o dado, maior seu peso.

Segundo (Olea, 1991 *apud* Yamamoto, 2000) uma outra vantagem da variância de interpolação em comparação a variância da Krigagem é o reconhecimento do efeito proporcional, condição que estabelece uma relação de proporção direta entre a variância do erro e alguma função média local dos dados amostrais.

2.3 Semivariograma

As técnicas geoestatísticas conseguem unir os aspectos topológicos (espaço) com o probabilístico (aleatório). Essas técnicas se baseiam na teoria das Variáveis Regionalizadas, onde é possível analisar a estrutura espacial e calcular o erro cometido na avaliação, levando em conta a posição real das amostras (longitude e latitude).

Daí surgiu a necessidade de se criarem novas ferramentas matemáticas que, de modo sintético, permitam estudar as duas características principais das Variáveis Regionalizadas: aspecto aleatório e aspecto espacial. Desta maneira, se tira da aparente desordem dos dados analisados, uma imagem que torna visível a variabilidade dos mesmos e também uma medida da correlação existente entre os valores tomados em dois pontos do espaço. Portanto esse é o objetivo da Análise Estrutural o do Semivariograma: estudar e mostrar a partir de ferramentas matemáticas a dispersão natural das Variáveis Regionalizadas.

Quando se utiliza as técnicas geoestatísticas, para a análise dos dados, algumas hipóteses básicas de trabalho são assumidas:

- i) A estacionaridade de 1ª ordem é a esperança do valor de um ponto no espaço ser igual à média admitindo-se que a média amostral é igual a média populacional;
- ii) A estacionaridade de 2ª ordem assume que há uma variação da média, mas a variância é constante dentro dos limites da continuidade espacial (Vieira, 1995);
- iii) A hipótese intrínseca diz que a probabilidades dos valores tomados entre pontos com a mesma distância de separação é igual, ou seja, a relação de dependência espacial mesma em qualquer posição de h , que é o vetor que separa dois pontos amostrais, assim sendo, é assumida a estacionaridade dentro do alcance da continuidade espacial

Para Diniz (1997), o semivariograma é dado por uma curva que representa o grau de continuidade de uma variável regionalizada. Experimentalmente, plota-se à distância h nas abscissas e nas ordenadas, o valor médio do quadrado da diferença entre os valores da variável regionalizada tomados a uma distância h de uma para outra variável. Em termos gerais, o semivariograma é uma função crescente com a distância h , desde que, em média, quanto maior a distância que ambas as amostras estiverem uma da outra, maiores serão as diferenças entre os seus valores. As características qualitativas da regionalização são muito bem expressas a partir do variograma.

Neste sentido, Huijbregts (1975), afirma que o variograma é uma ferramenta básica de suporte das técnicas geoestatísticas, que permite representar quantitativamente a variação de um fenômeno regionalizado no espaço. A natureza estrutural de um conjunto de dados (assumido pela variável regionalizada) é definida a partir da comparação de valores tomados simultaneamente em dois pontos, segundo uma determinada direção. A função de semivariância $\gamma(h)$ é definida como sendo a esperança matemática do quadrado da diferença entre os valores de pontos no espaço, separados por uma distância h , conforme

$$\gamma(h) = \frac{1}{2} \{ [Z_{(x)} - Z_{(x+h)}]^2 \}$$

E a partir de uma amostra $z(x_i), i = 1, 2, \dots, n$, pode ser estimado por

$$\hat{\gamma}(h) = \frac{1}{2N(h)} \sum_{i=1}^{N(h)} \{ [z_{(x_i)} - z_{(x_i+h)}]^2 \}$$

Onde

- i) $\hat{\gamma}(h)$ é a semivariância estimada para o intervalo h ;
- ii) $N(h)$ é o número de pares de valores medidos, $z_{(x_i)}$ e $z_{(x_i+h)}$, separados por um vetor distância h ;
- iii) $z_{(x_i)}$ e $z_{(x_i+h)}$ são valores da i -ésima observação da variável regionalizada, coletados nos pontos x_i e $x_i + h$ ($i = 1, 2, \dots, n$), separados pelo vetor h .

Segundo Souza (1999), ao representar o gráfico de $\gamma(h)$ versus h , o qual se chama de semivariograma, se obtém a estimativa do valor da variância para as diferentes combinações de pares de pontos e assim analisar o grau de dependência espacial da variável estudada e definir os parâmetros necessários para a estimativa de suas características em locais não amostrados. Conforme Figura abaixo

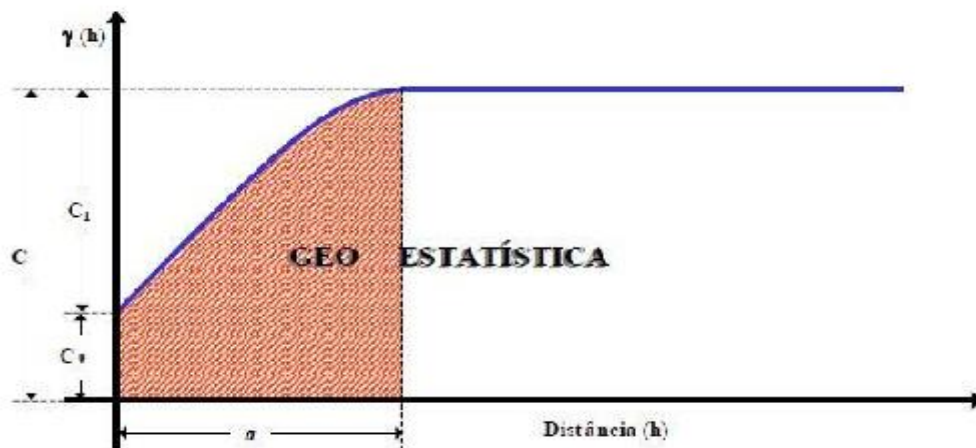


Figura 2.1: Esquema de Semivariograma Experimental

Conforme Silva (1988) e Camargo (1997), citado por Garcêz (2009), a medida que h aumenta $\gamma(h)$ também aumenta, pois é de se esperar que amostras tiradas a uma pequena distância entre si apresentem $[Z_{(x)} - Z_{(x+h)}]^2$ menores que aquelas tiradas a distâncias maiores.

O ajuste do modelo matemático aos dados no gráfico, ou seja, a uma função, define os parâmetros do semivariograma, que são: efeito pepita (C_0), que é o valor quando $h = 0$, quando h aumenta freqüentemente, aumenta até uma distância a , chamada de alcance a da dependência espacial; e a partir da qual $\gamma(h)$ neste ponto é chamado de patamar C , cujo valor é aproximadamente igual a variância dos dados, se ela existe, e é obtido pela soma do efeito pepita (C_0) e a variância estrutural (C_1).

De uma maneira mais explícita, o alcance representa a distância em que as amostras apresentam-se correlacionadas espacialmente; o patamar é o valor de semivariograma correspondente a seu alcance, sendo uma estimativa da variância da população, desse ponto em diante, considera-se que não existe mais dependência espacial entre as amostras, pois a variância da diferença entre pares de amostras torna-se invariante com a distância; o efeito pepita representa a descontinuidade do semivariograma para escalas menores que a distância entre as amostras. Parte desta descontinuidade pode ser também devida a erros de medição, mas impossível quantificar se a contribuição provém dos erros de medição ou da variabilidade de pequena escala não captada pela amostragem e a contribuição que é a diferença entre o patamar e o efeito pepita.

A relação entre os parâmetros (C_0) e (C_1) fornece o que se denomina efeito pepita relativo, dados pela equação C_0/C_1 , que expressa a aleatoriedade da regionalização. Para Royle (1977), os intervalos seguintes fornecem uma noção de influência do comportamento aleatório:

- i) $E < 0,15 \rightarrow$ Componente aleatória pequena;
- ii) $0,15 < E < 0,30 \rightarrow$ Componente aleatória importante;
- iii) $E > 0,30 \rightarrow$ Componente aleatória muito importante.

Um outro modo é o de analisar a razão $C_0/(C_0 + C_1)$, quando esta for maior que 0,8, a Estatística e a Geoestatística não possuem diferença (Journel e Huijbregts, 1978). Desta maneira, pode-se analisar quanto da variância total é causada por variações aleatórias.

O procedimento de ajuste não é direto e automático, como por exemplo, uma regressão, mas sim iterativo, pois nesse processo o interprete faz um primeiro ajuste e verifica a adequação do modelo teórico. De acordo com o ajuste obtido, pode ou não redefinir o modelo, até obter um que seja considerado satisfatório.

Os modelos considerados básicos, denominados de modelos isotrópicos por Issaks e Srivastava (1989), que serão detalhados no Capítulo 3, são divididos em dois tipos: modelos de tipo I, com patamar (esférico, exponencial e gaussiano) e modelos do tipo II, sem patamar. Modelos do primeiro tipo são denominados na Geoestatística como Modelos Transitivos.

2.4 Principais Modelos de Variogramas

Os modelos com patamar são normalmente ajustes que representam a estacionaridade de segunda ordem. A semivariância aumenta à medida que aumenta a distância entre as amostras, até atingir um patamar *sill*, onde se estabiliza. Este patamar deve ser teoricamente igual à variância amostral. A distância em que o semivariograma atinge o patamar é denominado de alcance *range*, que corresponde ao raio de dependência da variável. Os modelos mais utilizados dentro deste grupo estão representados na Figura abaixo.

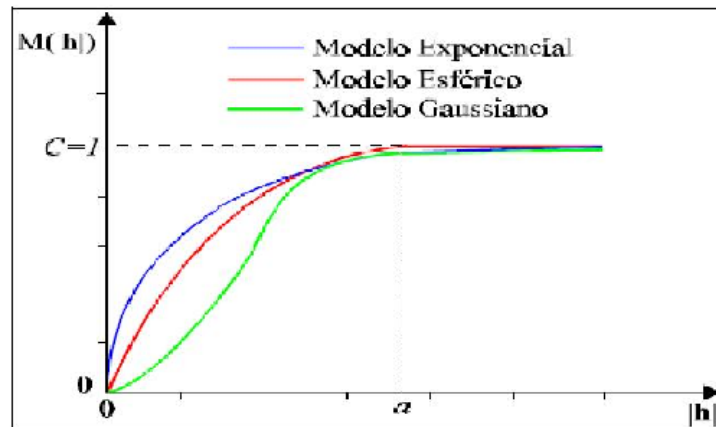


Figura 2.2 Semivariogramas com Patamar

a) Modelo de Variograma Espacial Esférico

O modelo esférico é um dos modelos mais utilizados e está representado em vermelho na Figura 3.1. A equação normalizada deste modelo é:

$$\gamma(h) = \begin{cases} C_0 + C_1 \left[1,5 \left(\frac{h}{a} \right) - 0,5 \left(\frac{h}{a} \right)^3 \right], & \text{se } 0 < h < a \\ C_0 + C_1, & \text{se } h \geq a \end{cases}$$

O modelo esférico é obtido selecionando-se os valores do efeito pepita, C_0 , e do patamar, C_1 , depois se passando uma reta que intercepte o eixo-y em C_0 e seja tangente aos primeiros pontos próximos de $h = 0$. Essa tangente cruzará o patamar à distância, $a' = 2/(3a)$. Assim, o alcance, $a = 3a'/2$. O modelo esférico é linear até aproximadamente $1/(3a)$.

b) Modelo de Variograma Espacial Exponencial

O modelo exponencial, também, é um dos mais utilizados, na Figura 2, está representado pela cor azul. Sua equação é dado por

$$\gamma(h) = C_0 + C_1 \left[1 - \exp\left(-3\frac{h}{a}\right) \right], 0 < h < d$$

onde d é a distância máxima na qual o semivariograma é definido. Uma diferença fundamental entre o modelo exponencial e o esférico é que o exponencial atinge o patamar apenas assintoticamente, enquanto que o modelo esférico o atinge no valor do alcance. O parâmetro a é determinado visualmente como a distância após a qual o semivariograma se estabiliza. Os parâmetros C_0 e C_1 para os modelos exponencial e gaussiano são determinados da mesma maneira que para o esférico.

c) Modelo de Variograma Espacial Gaussiano

O modelo gaussiano é na maioria das vezes usado para modelar fenômenos extremamente contínuos (Issak e Srivastava, 1989). Sua formulação é dada por:

$$\gamma(h) = C_0 + C_1 \left\{ 1 - \exp\left[-3\left(\frac{h}{a}\right)^2\right] \right\}, 0 < h < d$$

A função é parabólica próxima à origem. Este modelo apresenta um alcance extenso e o patamar semelhante ao modelo exponencial.

d) Modelo Aleatório (Efeito Pepita Puro)

$$\gamma(h) = C$$

para qualquer h . De acordo que aumenta a descontinuidade na origem do semivariograma, mais aleatório é o fenômeno que originou a variável em análise. Esta característica decorre de uma provável regionalização, inferior à escala de trabalho da malha de amostragem e/ou às variações espúrias associadas com a coleta e medição das amostras (Farias *et al.*, 2002a, 2002b).

e) Modelo sem Patamar ou Modelo Linear sem Patamar

Esses modelos correspondem a fenômenos que tem uma capacidade infinita de dispersão, e por isto, não tem variância finita e a covariância não pode ser definida. Eles podem ser escritos como

$$\gamma(h) = C_0 + Ah^B, 0 < B < 2$$

Os parâmetros A e B são constantes que definem o modelo, sendo que B tem que ser estritamente maior que zero e menor que 2 para garantir a condição de positividade definida condicional, ou seja, garantir que o semivariograma satisfaça a condição $\gamma(h) > 0$ e $\gamma(-h) = \gamma(h)$.

Depois da análise estrutural dos dados, inferências em localizações não amostradas são realizadas usando *Krigagem*, que é um método de inferência espacial, que estima dados em pontos não amostrados a partir de pontos amostrados, considerando a estrutura de dependência espacial do fenômeno em estudo ou *Simulados*, usando simulações condicionais.

Este trabalho limitar-se-á em realizar inferências utilizando a técnica de simulação, bem como comparar os dois métodos utilizados: a Simulação Gaussiana e Sequencial e o Método de Convolução.

2.5 Simulação: o que é e para que serve?

Como uma maneira de imitar um processo ou uma operação do mundo real, a simulação computacional, ou somente simulação, consiste na utilização de certas técnicas matemáticas, empregadas em computadores, as quais permitem imitar o funcionamento de, praticamente qualquer tipo de operação ou processo do mundo real, ou seja, é o estudo do comportamento de sistemas reais através do exercício de modelos.

Para Pegden *et. al.* (1990) a simulação é um processo de projetar um modelo computacional de um sistema real e conduzir experimentos com este modelo com o propósito de entender seu comportamento e/ou avaliar estratégias para sua operação. Desta maneira, pode-se entender a simulação como um processo amplo que engloba não apenas a construção do modelo, mas todo o método experimental que se segue, buscando: descrever o comportamento do sistema; construir teorias e hipóteses considerando as observações

efetuadas; usar o modelo para prever o comportamento futuro, isto é, os efeitos produzidos por alterações no sistema ou nos métodos empregados em sua operação.

Essa definição mostra que a simulação é uma relação entre o processo de construção e o de experimentação, inclui o ato de processo de projetar um modelo computacional em um sistema real e esta por sua vez inclui a formulação do problema, a especificação do funcionamento de tal sistema real, a análise de riscos, a construção do modelo e toda sua verificação e validação.

A simulação de sistemas é um processo amplo, vai desde sua criação até sua validação, já que se quer características mais próximas possível do real. A simulação consiste na utilização de determinadas técnicas matemáticas e estatísticas, empregadas em computadores, e que permitem imitar o comportamento de praticamente qualquer tipo de sistema, operação ou processo do mundo real, existente ou não.

Entre os diversos tipos de simulação existente a Simulação Geoestatística se utiliza de modelos geoestatísticos definidos que permitem analisar os melhores parâmetros para a construção do melhor modelo, ou seja, o que melhor se aproximar do real.

2.6 Simulação Estocástica X Krigagem

Apesar de ser um ótimo procedimento de interpolação para se determinar dados espacialmente localizados, a Krigagem apresenta algumas limitações que restringem sua aplicabilidade. As principais são: ocorrência de pesos negativos, e conseqüentemente, teores negativos; caráter homoscedástico da variância de Krigagem; efeito de suavização das estimativas de Krigagem; efeito de suavização das estimativas de Krigagem. Porém, uma das limitações mais discutidas na literatura, é o que se chama de efeito de suavização da Krigagem ordinária. Tal efeito justifica o fato da Krigagem ser considerada uma técnica de interpolação de precisão local, uma vez que o método tende a suavizar os valores de atributos dos dados, acarretando em uma perda no modelo de variabilidade da distribuição como um todo, ou seja, seu modelo global.

Contrapondo estas limitações, a Simulação, também, é um método alternativo para se tentar chegar a um modelo a partir de poucos dados pontuais.

Diversos autores propuseram como alternativa o uso da simulação estocástica em resposta à problemática do efeito de suavização no uso de interpolação derivadas de métodos de minimização da variância dos erros. Essa técnica consiste basicamente na criação de algoritmos para se simular o comportamento de determinada variável, no caso, o seu comportamento espacial. O termo “estocástico” se refere ao fato destes algoritmos serem não determinísticos, isto é, quando a mais de uma opção entre o valor a ser assumido em determinado ponto de execução do algoritmo, tal escolha será aleatória, gerando inúmeras possíveis realizações a partir de um mesmo conjunto inicial de dados.

É definido como o processo de se extrair realizações conjuntas, alternativas e equiprováveis de uma variável aleatória a partir de um modelo de uma função aleatória. Sendo possíveis L imagens possíveis da distribuição espacial dos valores de atributos $z(x)$ no campo A , cada imagem é denominada de “imagem estocástica”, refletindo as propriedades que foram impostas no modelo de função aleatória $Z(x)$. Assim sendo, quanto mais propriedades forem inferidas do conjunto amostral e incorporados no modelo $Z(x)$, melhor o modelo resultante (Deutsch e Journel, 1992).

Caers (2000), afirma que ao contrário da Krigagem, a simulação possui precisão global, isto é, reproduz a variabilidade espacial da amostragem em detrimento da manutenção da precisão local. Por reprodução da variabilidade espacial entende-se que as estatísticas multiponto globais das estimativas, tais como variograma e histograma, serão idênticas às da amostragem.

Apesar de modelos estocásticos serem raramente utilizados quantitativamente, a sua variabilidade auxilia na compreensão da complexidade que é encontrada nos fenômenos geológicos naturais, o que desafia a crença na simplificação exacerbada dos outros modelos. A modelada estocástica também pode ser usada para gerar apenas uma realização como base para se prever a realidade geológica. A vantagem de tal uso seria a sua habilidade de honrar um nível realístico de heterogeneidade, evitando modelos muito simplistas. Um dos maiores reveses da Simulação Estocástica talvez seja o fato de que serão geradas n imagens a partir de um mesmo conjunto de dados, todas equiprováveis, o que dificulta a escolha de uma realização ótima para qualquer situação. Modelos Estocásticos, oferecendo uma gama de possibilidades, permitem gerar múltiplos cenários de possibilidades, com modelos otimistas e pessimistas, permitindo se lidar com as incertezas (Srivastava, 1994).

No entanto, a ressalva de que podem haver várias realizações se aplica igualmente aos mapas interpolados: pode haver imagens estimadas alternativas a depender do método de interpolação/estimativa usado.

A partir da década de 1980 e meados de 1990 que se iniciou a consolidação dos métodos estocásticos aplicado na área da Geociência, notadamente no campo de caracterização de reservatório de hidrocarbonetos. Nessa época, diversas metodologias foram apresentadas, onde vários estudos de caso foram abrangidos, permitindo a abertura de um amplo espectro de abordagem possíveis, tanto do ponto de vista teórico quanto prático.

Mesmo a simulação estocástica tendo sido desenvolvida para fornecer medidas de incerteza espacial, algoritmos de simulação são cada vez mais utilizados para fornecer um único mapa estimado. Os algoritmos de simulação tem se mostrados muito mais versáteis na reprodução de todo o espectro de variabilidade dos dados espaciais e em considerar dados de diferentes tipos e fontes. Desta maneira, essas realizações são usadas em aplicações onde a reprodução de feições espaciais seja mais importante que a precisão local.

Uma outra característica que se pode citar aqui, é quanto a ergodicidade, que se refere às discrepâncias entre estatísticas da realização e os parâmetros do modelo correspondente. A função aleatória $Z(h)$ é dita ser ergódica se as estatísticas das realizações tenderem a média (μ) conforme se aumenta o tamanho do campo A , sendo este parâmetro geralmente considerado como uma média estacionária da função aleatória, ou seja, a média de sua Função de Distribuição Acumulada (CDF) $F(z)$. Assim, provado que $Z(x)$ seja estacionária e ergódica, e o campo de simulação seja grande suficiente, deve-se esperar que as estatísticas de qualquer realização reproduzam exatamente os parâmetros do modelo. É importante lembrar que, o tamanho do campo A para que se possa determinar a ergodicidade depende do tipo de função aleatória escolhida e na amplitude das suas heterogeneidades, conforme caracterizado pelo modelo de covariância, (quanto mais contínua a covariância em $h = 0$ e quanto maior a amplitude da covariância, o campo A deve ser maior para que a ergodicidade seja atingida.

Os questionamentos sobre ergodicidade surgem já que normalmente os usuários apenas retém uma única realização, ou seja, usando a simulação como um algoritmo melhorado de interpolação. Muitas vezes, esta realização retida não reproduz as estatísticas do modelo, sendo que isto é o pior cenário possível para o usuário. Para garantir a reprodução das estatísticas, a única maneira seria uma média levando-se em conta um número grande de realizações, sendo que quanto menos ergódico o modelo de função aleatória escolhido, maior

o número de realizações necessárias para se aproximar dos valores esperados (Deutsch e Journel, 1992).

São citados abaixo alguns métodos de simulação estocástica:

- i) Booleano, marked-point process e simulação baseada em objetos;
- ii) Estimativa mais erro simulado: simulação fractal e por bandas rotativas;
- iii) Simulated annealing;
- iv) Simulação de campo de probabilidade;
- v) Métodos de decomposição de matrizes;
- vi) Métodos iterativos;
- vii) Métodos hídricos.

2.7 Método de Kernel

Um modelo espacial contínuo ou discreto pode ser facilmente construído através da convolução entre um processo aleatório, o chamado ruído branco, e uma função Kernel escolhida convenientemente, ou seja, uma função Kernel específica encontrada a partir da função espacial de covariância.

Esta forma de construir processos espaciais oferece um grande número de vantagens e talvez a melhor delas seja a possibilidade da especificação do modelo espacial de covariograma, ou seja, fornecer uma informação à função espacial de covariância para se ter um processamento mais rápido dos dados, pois em geral, é necessário primeiro se ter o modelo para depois realizar a simulação. Em particular, tais processos convolutivos levam à simplificações que diminuem consideravelmente o esforço computacional.

O modelamento de processos espaciais através de processos gaussianos é muito comum na análise geoestatística (Matheron 1963, Journel and Huijbregts 1978, Ripley 1981, Cressie 1991, Wackernagel 1995 and Stein 1999).

Uma forma de se modelar a dependência espacial é através da função covariograma $c(\cdot)$, de modo que a covariância entre dois pontos dependa apenas da distância entre eles e maneira de se construir um processo gaussiano $z(s)$ em uma região S é envolvendo um ruído branco seja ele contínuo ou discreto $x(s); x \in S$ com uma função kernel $k(s)$, de modo que:

$$z(s) = \int k(u - s)x(u)du, s \in S$$

É a partir de então a função de covariância resultante dependerá apenas do vetor deslocamento $d = s - s'$ e é dada por:

$$c(d) = Cov[z(s), z(s')] = \int k(u - s)x(u)du, s \in S$$

Com isso é possível relacionar a função de covariância $c(s)$ com seu kernel associado $k(s)$. Esta relação se dá a partir da Transformada de Fourier:

$$c(s) \xrightarrow{TF} c(\omega) \xrightarrow{\sqrt{\cdot}} C^{0,5}(\omega) \xrightarrow{TF^{-1}} k(s)$$

onde TF e TF^{-1} são respectivamente a Transformada de Fourier e sua Inversa, que será abordada em mais detalhes no Capítulo 4 deste trabalho. Assim para simular um processo espacial que reproduza determinado modelo espacial de covariância $c(s)$, basta convolver um ruído branco com o kernel $k(s)$ associado a este modelo espacial de covariância.

2.8 Transformação Box-Cox

Uma das suposições mais freqüentes na área da estatística é que variáveis mensuráveis sigam uma distribuição normal. A distribuição normal, sendo uma das distribuições fundamentais da Estatística Moderna, possui a vantagem na facilidade de ser definida apenas com dois parâmetros: a média μ e o desvio padrão σ . Neste trabalho, a não-normalidade dos dados utilizados, levou ao uso da transformação dos dados originais através da transformação Box-Cox para chegar a valores transformados que demonstrem normalidade ou no mínimo normalidade aproximada. A transformação BOX-COX foi introduzida na literatura em 1964 para resolver o problema de estimação de regressões não lineares.

A transformação é relativamente simples:

$$y^\lambda = \frac{(y^\lambda - 1)}{\lambda}$$

Onde λ varia entre $(-1, 1)$. Essa transformação pode ser particularmente útil quando uma variável de resposta não cumpre com os pressupostos de normalidade e/ou homoscedasticidade. O valor transformado dos dados passa através de vários tipos de equações para cada valor de λ .

Se λ for igual a:	Então $y^\lambda = \frac{(y^\lambda - 1)}{\lambda}$ será igual a:
-1	$\frac{1}{y} + 1$
-0,5	$\frac{\frac{1}{\sqrt{y}} - 1}{-0,5}$
0	$\ln y$ (da Regra de L'Hospital)
0,5	$\frac{\sqrt{y} - 1}{0,5}$
1	$y - 1$

Tabela 1: Alguns valores transformados, para determinados valores de λ .

Capítulo 3

Convolução à 1D à 2D e Resultados

Para o experimento foram escolhidos dois modelos de covariância espacial: um para modelar fenômenos com maior continuidade e outro para fenômenos com menor continuidade que são respectivamente os modelos de covariância gaussiano e exponencial.

3.1 Modelo Espacial de Covariância Gaussiano

O modelo espacial de covariância gaussiano é dado por:

$$c(s) = C_0 e^{-\frac{s^2}{a^2}}$$

onde a é o alcance teórico do modelo, ou seja, a máxima distância onde há correlação espacial. O kernel associado à função de covariância gaussiana encontrado analiticamente é dado por:

$$k(s) = 2^{\frac{3}{4}} \frac{\sqrt{c}}{\sqrt{a}} e^{-2\frac{s^2}{a^2}}$$

3.2 Modelo Espacial de Covariância Exponencial

O modelo de covariância espacial exponencial é dado por:

$$c(s) = C_0 e^{-\frac{s}{a}}$$

onde a é o alcance teórico do modelo. O kernel associado ao modelo foi calculado numericamente utilizando-se a Transformada de Fourier Discreta ou do inglês Discrete Fourier Transform (DFT).

Segundo Neto (2010) a convolução é uma operação realizada por qualquer sistema linear invariante no tempo (1D) ou espaço (2D). A saída do sistema é dada pela convolução do sinal de entrada pela resposta ao impulso deste sistema, possibilitando operações de filtragem e processos de integração e diferenciação.

3.3 Convolução à 1D

A operação de convolução (*) unidimensional entre dois vetores f e h , denominada $f*h$, pode também ser entendida como o somatório produtos entre as amostras de f e h , sendo que inicialmente o vetor h é espelhado e, após cada soma de produtos é deslocado temporalmente de uma posição. Assim tem-se:

$$f(x) * h(x) = \sum_{i=1}^M f(x)h(i - x)$$

Como mostrado no Capítulo 2, deste trabalho, o kernel associado à função de covariância, é encontrado a partir da Transformada de Fourier e sua inversa. Para convolução à 1D, tem-se que se $f(x)$ é uma função contínua e integrável e se $F(u)$ for integrável a Transformada de Fourier (TF) é indicada por $\mathfrak{F}[f(x)]$, e definida pela equação

$$\mathfrak{F}[f(x)] = F(u) = \int_{-\infty}^{+\infty} f(x) \exp[-j2\pi ux] dx$$

onde $j = \sqrt{-1}$. Dado então $F(u)$, $f(x)$ pode ser obtida calculando-se a Transformada Inversa de Fourier $(TF)^{-1}$, indicada por $\mathfrak{F}^{-1}[F(x)]$ e definida por

$$\mathfrak{F}^{-1}[F(u)] = f(x) = \int_{-\infty}^{+\infty} F(u) \exp[j2\pi ux] du$$

3.4 Convolução à 2D

O raciocínio da convolução unidimensional pode ser expandido para o caso bidimensional, aplicado ao processamento de imagens. Imagem é uma função definida em uma superfície bidimensional cujo contradomínio é composto de valores originados de um espaço de cor

$$I = U \rightarrow C$$

onde I é a função da imagem, $U \subset \mathbb{R}^3$ é uma superfície que geralmente é um subconjunto plano chamado suporte da imagem e C é um espaço vetorial (espaço e cor), onde quando $C = 1$ a imagem é chamada monocromática.

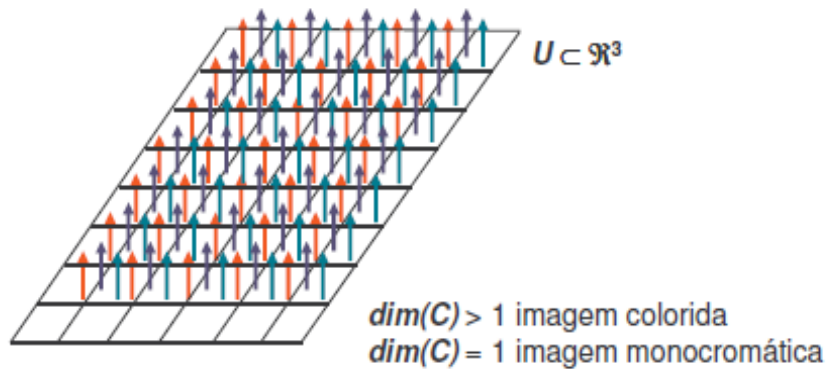


Figura 3.1: Esquema de uma Imagem

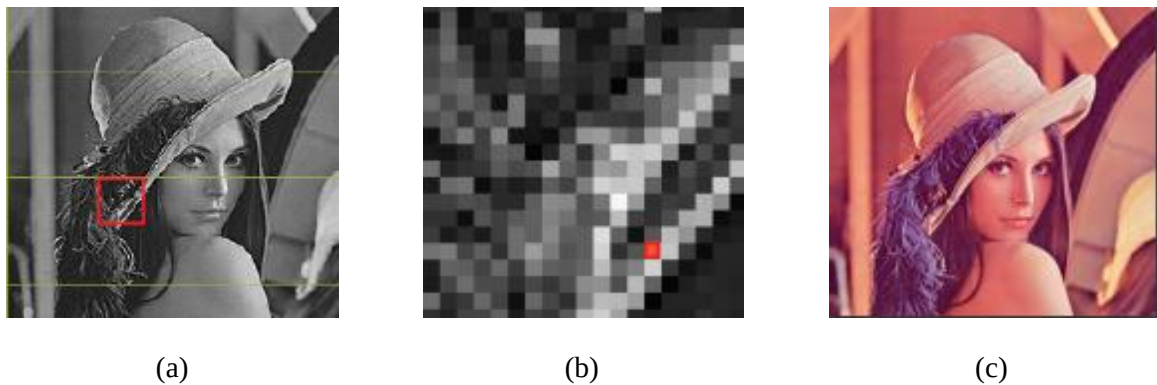


Figura 3.2: (a) Imagem Monocromática, (b) Pixel da Imagem, (c) Imagem em Composição Colorida

A Convolução bidimensional aplicada no processamentos de imagens pode ser realizada utilizando as técnicas de Filtragem, que podem ser: Filtragem no Domínio Espacial e no Domínio da Frequência. Filtragem é uma operação feita sobre um sinal ou imagem para obter outro sinal ou imagem.

3.4.1 Filtragem no Domínio Espacial

Os métodos de filtragem que trabalham no domínio espacial operam diretamente sobre os *pixels*, normalmente utilizando operações de convolução com máscaras. O uso de máscaras nas imagens no domínio espacial é usualmente chamado de filtragem espacial e as máscaras são chamadas de filtros espaciais.

$$g(x, y) = T[f(x, y)]$$

onde $f(x, y)$ é a imagem de entrada a ser filtrada; $g(x, y)$ é a imagem na saída, ou seja, a imagem já processada e T é um operador sobre f , definido sobre algum *pixel* da imagem.

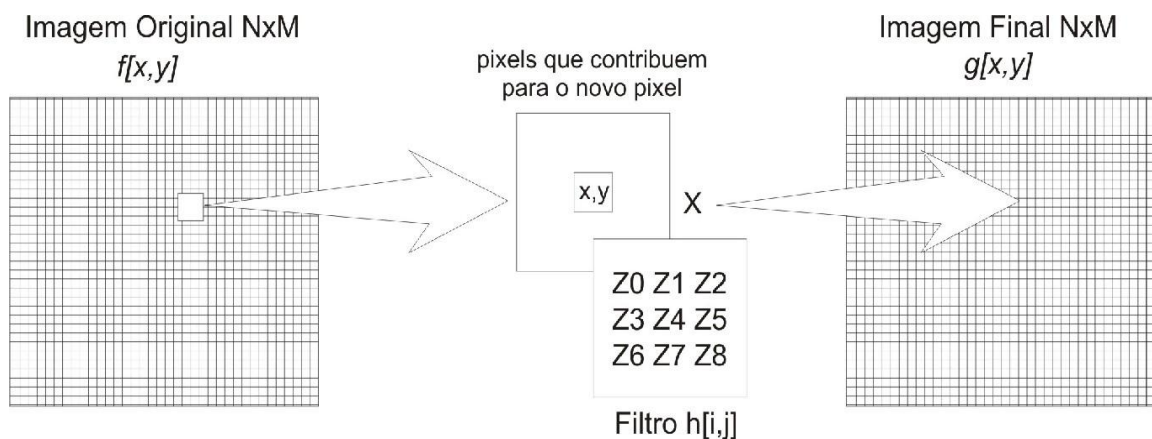


Figura 3.3: Esquema do processo de convolução por filtragem espacial utilizando uma máscara $n \times m$

A vizinhança normalmente definida ao redor de (x, y) é a 8-vizinhança do pixel de referência, o que equivale a uma região 3×3 na qual o pixel central é o de referência, como indica a Figura 3.3. O centro dessa região ou subimagem é movido pixel a pixel, iniciando no canto superior esquerdo da figura e aplicando a cada localidade o operador T para calcular o valor de g naquele ponto.

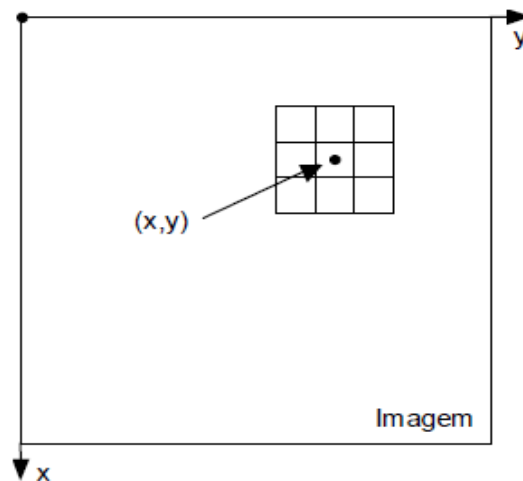


Figura 3.4: Vizinhança 3X3 ao redor de um ponto (x,y) em uma imagem

Nos casos em que a vizinhança é 1×1 , o operador T torna-se uma função de transformação (ou de mapeamento), do tipo:

$$s = T(r)$$

onde: r é o nível de cinza de $f(x,y)$ e s é o nível de cinza de $g(x,y)$ em um certo ponto. As técnicas de processamento de imagem pertencentes a este caso são freqüentemente denominadas técnicas ponto-a-ponto, que para (Marques Filho e Vieira Neto, 1999) são técnicas de modificação de histogramas, uma vez que o valor de tom de cinza de um certo pixel após o processamento depende apenas de seu valor original. Em contraste, nas técnicas de processamento orientadas a vizinhança, o valor resultante depende também, de alguma forma, dos pixels que circundam o elemento de imagem original.

Desta maneira no esquema da Figura 3.4, tem-se:

$$\begin{aligned} g(x,y) &= T[f(x,y)] \\ &= Z_1 f(x-1, y-1) + Z_2 f(x, y-1) + Z_3 f(x+1, y-1) + Z_4 f(x-1, y) \\ &\quad + Z_5 f(x, y) + Z_6 f(x+1, y) + Z_7 f(x-1, y+1) + Z_8 f(x, y+1) \\ &\quad + Z_9 f(x+1, y+1) \end{aligned}$$

3.4.2 Filtragem no Domínio da Frequência

Já se conhece que base matemática para o uso da técnica de Filtragem no Domínio da Frequência é a convolução. A imagem é transformada do Domínio do Espaço para o Domínio da Frequência a partir da Transformada de Fourier. Esta a ser processada é uma matriz bidimensional relativamente grande e corresponde a f , enquanto uma matriz de pequenas dimensões corresponde a h . A convolução (*), então é realizada pelo algoritmo:

$$g(x, y) = f(x, y) * h(x, y) = \sum_{i=1}^M \sum_{j=1}^N f(x, y) h(i - x, j - y)$$

O algoritmo consiste em processar cada valor da posição de índices (x, y) , percorrendo todos os vizinhos da janela dados pelos índices (i, j) , armazenando o resultado na posição (x, y) . Então, pela convolução, a seguinte relação é válida:

$$G(u, v) = F(u, v)H(u, v)$$

onde G , F e H são as Transformadas de Fourier de g , f e h respectivamente. E onde, $H(u, v)$, na terminologia de sistemas lineares, é denominado função de transferência do filtro. Na literatura, inúmeros problemas de processamento de imagens podem ser expressos na forma da equação de $G(u, v)$, assim como vários tipos de filtros que podem ser utilizados para o processamentos de imagens, porém esses detalhes fogem ao corpo deste trabalho.

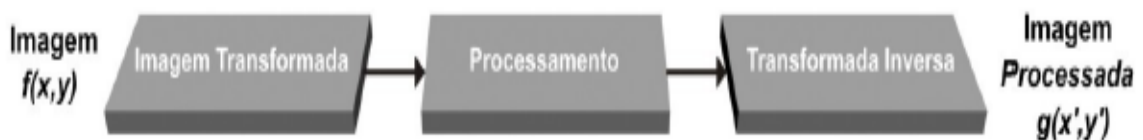


Figura 3.5: Esquema de processamento no Domínio da Frequência usando a transformada de imagem

Os conceitos unidimensionais da Transformada de Fourier estendem-se para o caso bidimensional, ou seja, uma função de duas variáveis: $f(x,y)$. Assim, se $f(x,y)$ é contínua e integrável e $F(u,v)$ é integrável então existem:

$$\mathfrak{F} = [f(x, y)] = F(u, v) = \int \int_{-\infty}^{+\infty} f(x, y) \exp[-j2\pi(ux + vy)] dx dy$$

e

$$\mathfrak{F}^{-1} = [F(u, v)] = f(x, y) = \int \int_{-\infty}^{+\infty} F(u, v) \exp[j2\pi(ux + vy)] dx dy$$

Inúmeras operações úteis em processamento digital de imagens são efetuadas a partir de um mesmo conceito básico, o de convolução com máscaras. A operação realizada na imagem depende da resposta ao impulso do sistema, ou seja, da máscara de convolução utilizada. A máscara, após ter sido espelhada tanto na horizontal quanto na vertical, deve percorrer todos os pontos da imagem, da esquerda para a direita e de cima para baixo, até o último pixel ter sido processado. O resultado normalmente é armazenado em uma matriz de dimensões similares às da imagem original.

3.4.3 Metodologia para Simulação à 2D

Os seguintes passos devem ser seguidos para o processo de convolução à 2D:

- i) A imagem $f(x,y)$ é transformada para o domínio de Fourier;
- ii) A imagem no domínio de Fourier é representada por $F(u,v)$ e é convoluída com o filtro $H(u,v)$;
- iii) Ao produto $F(u,v) H(u,v)$ é aplicada a inversa da transformada de Fourier para retornar ao domínio espacial, onde se tem a imagem processada.

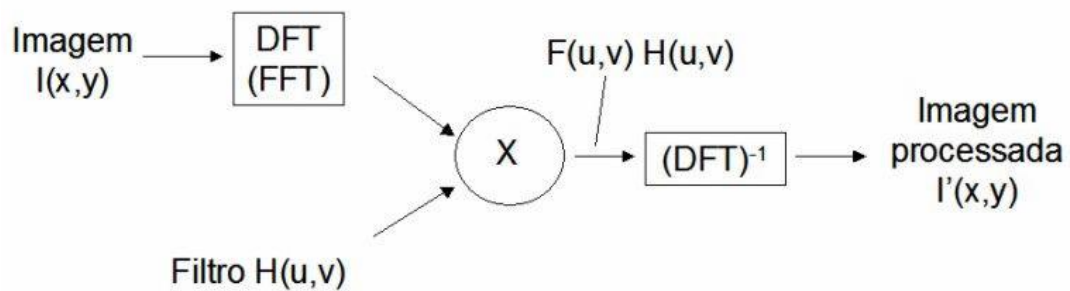


Figura 3.6: Esquema ilustrando os passos da filtragem no Domínio de Fourier

3.5 Operações com Máscaras

Nesta seção abordar-se-á em mais detalhes o funcionamento das operações de convolução e apresentar-se-á alguns exemplos típicos de máscaras e os resultados que elas produzem quando aplicadas a imagens monocromáticas, os exemplos foram extraídos de (Marques Filho e Vieira Neto, 1999).

Para Marques Filho e Vieira Neto (1999) as operações de convolução com máscaras são amplamente utilizadas no processamento de imagens. Uma seleção apropriada de coeficientes para o processo de convolução torna possível uma grande variedade de operações úteis, tais como redução de ruído, afinamento e detecção de características da imagem. Afirmam ainda que a operação de convolução com máscaras exige grande esforço computacional. Por exemplo, a aplicação de uma máscara 3×3 sobre uma imagem 512×512 requer nove multiplicações e oito adições para cada localização de pixel, num total de 2.359.296 multiplicações e 2.097.152 adições.

a) Operação à 1D

A operação de convolução unidimensional entre dois vetores A e B , denotada $A*B$, pode ser entendida como um conjunto de somas de produtos entre os valores de A e B , sendo que inicialmente o vetor B é espelhado e após cada soma de produtos é deslocado espacialmente de uma posição. Para ilustrar este conceito tem-se a seguir, a convolução do vetor $A = \{0, 1, 2, 3, 2, 1, 0\}$ com o vetor $B = \{1, 3, -1\}$.

1. Inicialmente, o vetor B é espelhado e alinhado com o primeiro valor de A. O resultado da convolução é $(0 \times (-1)) + (0 \times 3) + (1 \times 1) = 1$ (valores em branco assumidos como zero) e é colocado em A*B na posição correspondente ao centro do conjunto B.

A		0	1	2	3	2	1	0	
B	-1	3	1						
A*B		1							

2. O conjunto B é deslocado de uma posição. O resultado da convolução A*B é $(0 \times (-1)) + (1 \times 3) + (2 \times 1) = 5$.

A		0	1	2	3	2	1	0	
B		-1	3	1					
A*B		1	5						

3. O conjunto B é deslocado de uma posição. O resultado da convolução A*B é $(1 \times (-1)) + (2 \times 3) + (3 \times 1) = 8$.

A		0	1	2	3	2	1	0	
B			-1	3	1				
A*B		1	5	8					

4. O conjunto B é deslocado de uma posição. O resultado da convolução A*B é $(2 \times (-1)) + (3 \times 3) + (2 \times 1) = 9$.

A		0	1	2	3	2	1	0	
B				-1	3	1			
A*B		1	5	8	9				

5. O conjunto B é deslocado de uma posição. O resultado da convolução A*B é $(3 \times (-1)) + (2 \times 3) + (1 \times 1) = 4$.

A		0	1	2	3	2	1	0	
B					-1	3	1		
A*B		1	5	8	9	4			

6. O conjunto B é deslocado de uma posição. O resultado da convolução A*B é $(2 \times (-1)) + (1 \times 3) + (0 \times 1) = 1$.

A		0	1	2	3	2	1	0	
B						-1	3	1	
A*B		1	5	8	9	4	1		

7. O conjunto B é deslocado de uma posição. O resultado da convolução A*B é $(1 \times (-1)) + (0 \times 3) + (0 \times 1) = -1$. (valores em branco assumidos como zero)

A		0	1	2	3	2	1	0	
B							-1	3	1
A*B		1	5	8	9	4	1	-1	

O conjunto {1, 5, 8, 9, 4, 1, -1} é o resultado final da operação de convolução.

b) Operação à 2D

A imagem a ser processada é uma matriz bidimensional relativamente grande e corresponde ao conjunto A do exemplo anterior, enquanto uma matriz de pequenas dimensões (também chamada máscara ou janela) corresponde ao conjunto B. A máscara, após ter sido espelhada tanto na horizontal quanto na vertical, percorrerá todos os pontos da imagem deslocando-se ao longo de cada linha e entre as várias linhas, da direita para a esquerda, de cima para baixo, até ter processado o último elemento da matriz imagem. O resultado será armazenado em uma matriz de mesmas dimensões que a imagem original. Na prática sejam a matriz A (imagem) e a matriz B (máscara), a operação de convolução bidimensional produzirá uma matriz C como resultado do processo.

5	8	3	4	6	2	3	7
3	2	1	1	9	5	1	0
0	9	5	3	0	4	8	3
4	2	7	2	1	9	0	6
9	7	9	8	0	4	2	4
5	2	1	8	4	1	0	9
1	8	5	4	9	2	3	8
3	7	1	2	3	4	4	6

A

2	1	0
1	1	-1
0	-1	-2

B

20	10	2	26	23	6	9	4
18	1	-8	2	7	3	3	-11
14	22	5	-1	9	-2	8	-1
29	21	9	-9	10	12	-9	-9
21	1	16	-1	-3	-4	2	5
15	-9	-3	7	-6	1	17	9
21	9	1	6	-2	-1	23	2
9	-5	-25	-10	-12	-15	-1	-12

C

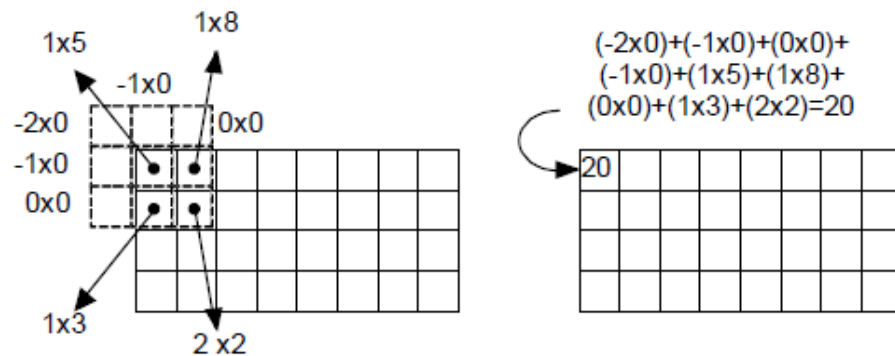


Figura 3.7: Cálculo do primeiro valor da convolução de A por B

A Figura 3.7 ilustra em detalhes o cálculo do resultado correspondente ao pixel no canto superior esquerdo da imagem.

Para calcular os valores resultantes dos pixels próximos às bordas da imagem, podem ser adotadas diversas estratégias, dentre elas: (Marques Filho e Vieira Neto, 1999)

- 1) Preencher com zeros o contorno da imagem, de maneira condizente com o tamanho de máscara utilizado;
- 2) Preencher o contorno da imagem com os mesmos valores da(s) primeira(s) e última(s) linha(s) e coluna(s).
- 3) Prevenir a eventual introdução de erros nas regiões de bordas da imagem causados por qualquer um dos métodos acima, considerando na imagem resultante apenas os valores para os quais a máscara de convolução ficou inteiramente contida na imagem original.

3.6 Resultados 1D

3.6.1 Convolução à 1D

Para a realização do teste em 1D inicialmente foi gerado um processo ruído branco com discretização $n=512$ em um suporte de comprimento 6, para processo de convolução, e em seguida seu respectivo variograma $\gamma(h)$, que teoricamente sua variância é igual a 1.

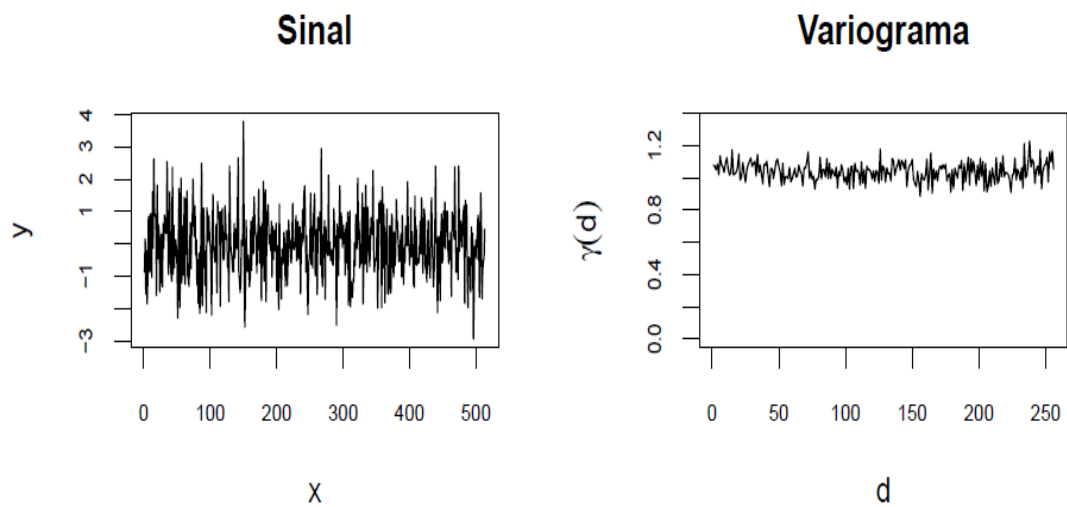


Figura 3.8 Ruído-Branco: $n=512$ e seu respectivo Variograma

3.6.2 Testes realizados a partir do Modelo de Covariância Gaussiano

a) Modelo de Covariância Gaussiano com $a = 1$

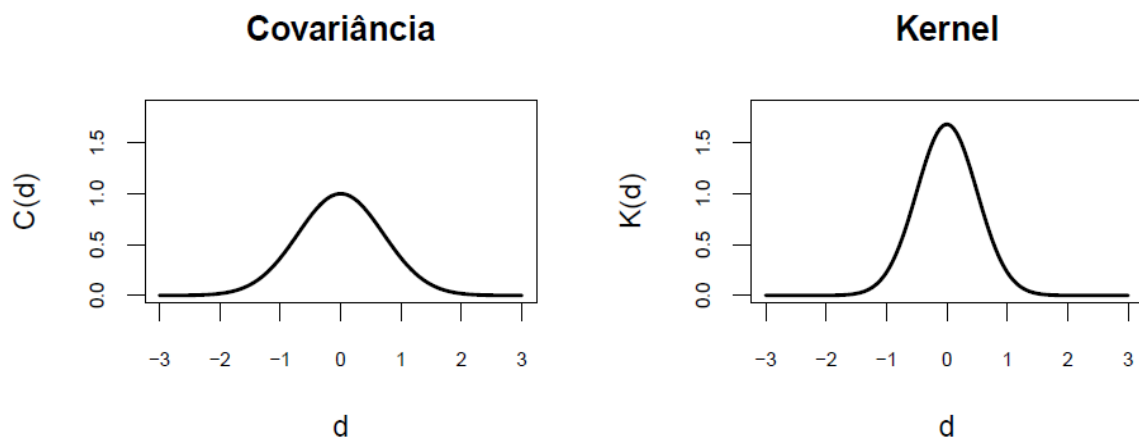


Figura 3.9 Modelo de Covariância Gaussiano com $a = 1$ e seu kernel associado

b) Modelo de Covariância Gaussiano com $a = 0,5$

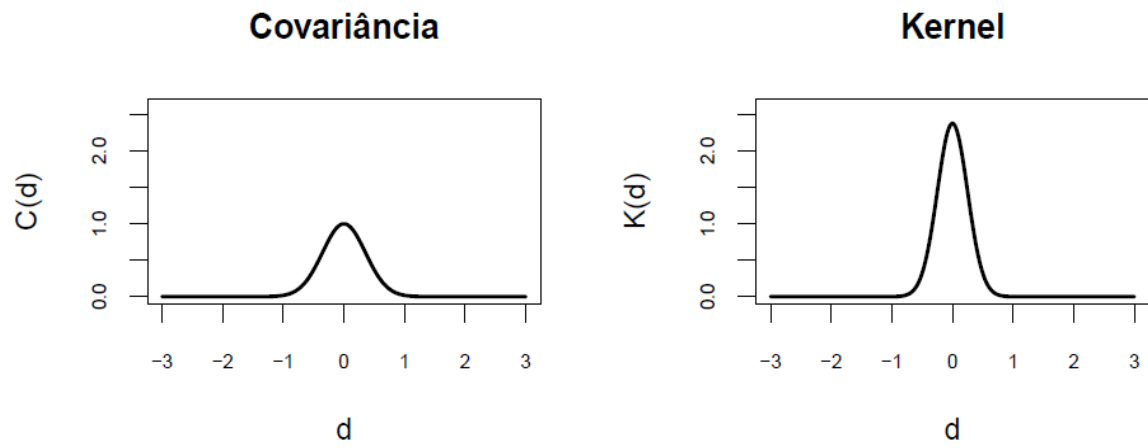


Figura 3.10 Modelo de Covariância Gaussiano com $a = 0,5$ e seu kernel associado

c) Modelo de Covariância Gaussiano com $a = 0,1$

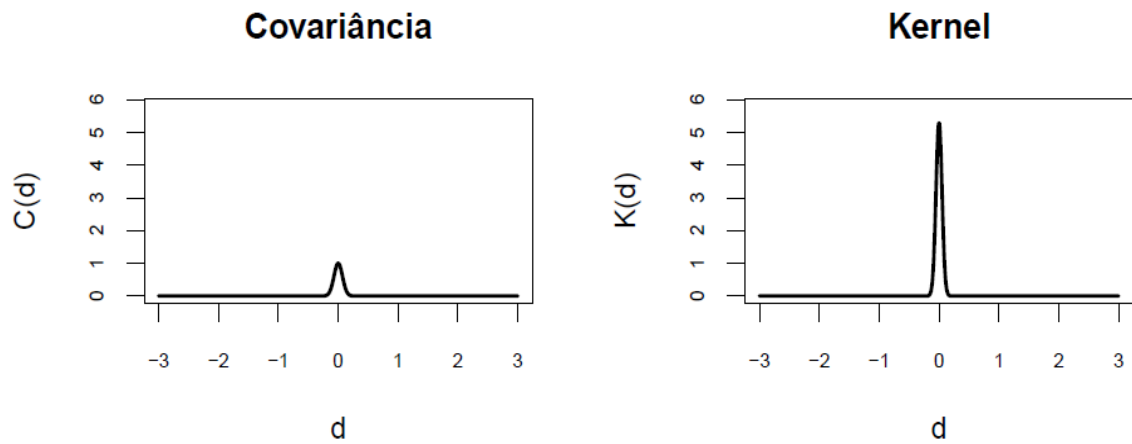


Figura 3.11 Modelo de Covariância Gaussiano com $a = 0,1$ e seu kernel associado

3.6.3 Testes realizados a partir do Modelo de Covariância Exponencial

a) Modelo de Covariância Exponencial com $\alpha = 1$

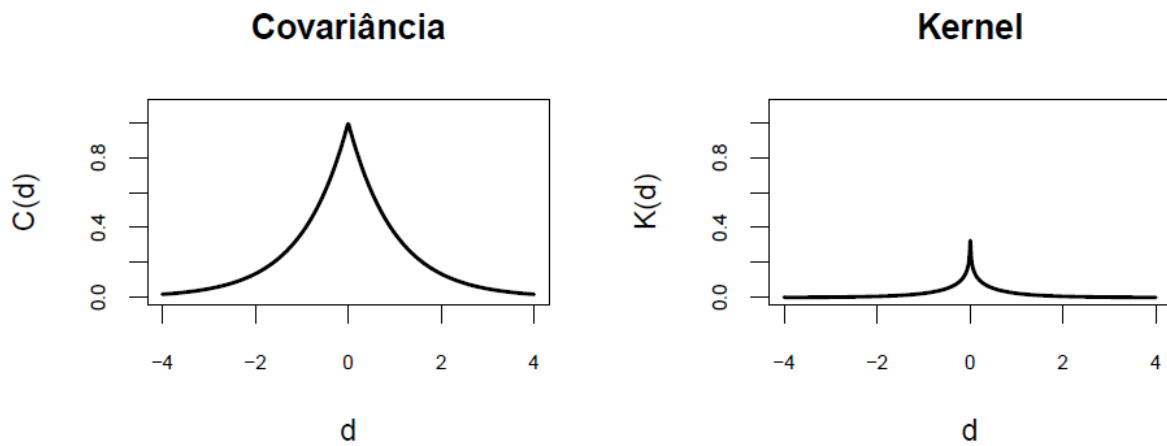


Figura 3.12 Modelo de Covariância Exponencial com $\alpha = 1$ e seu kernel associado

b) Modelo de Covariância Exponencial com $\alpha = 0,5$

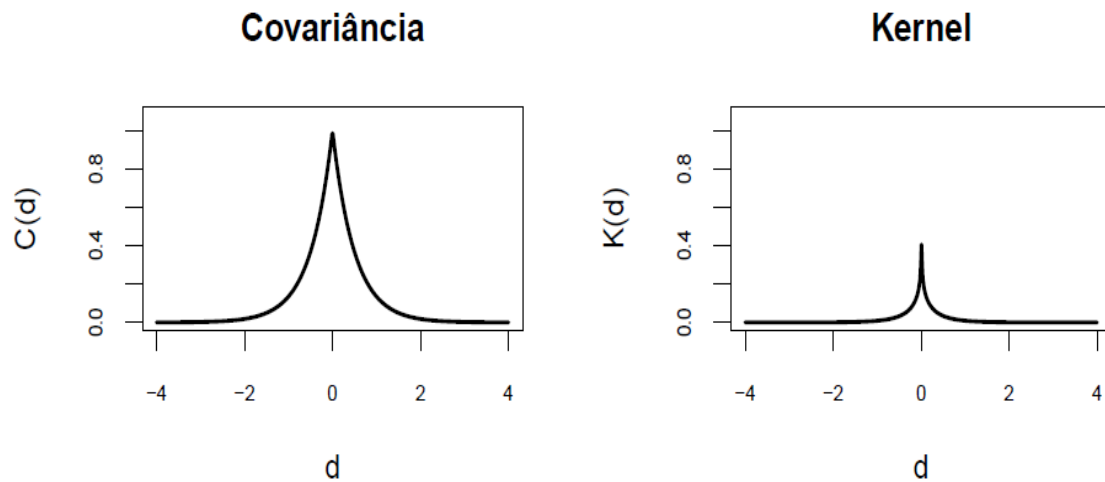


Figura 3.13 Modelo de Covariância Exponencial com $\alpha = 0,5$ e seu kernel associado

c) Modelo de Covariância Exponencial com $\alpha = 0,1$

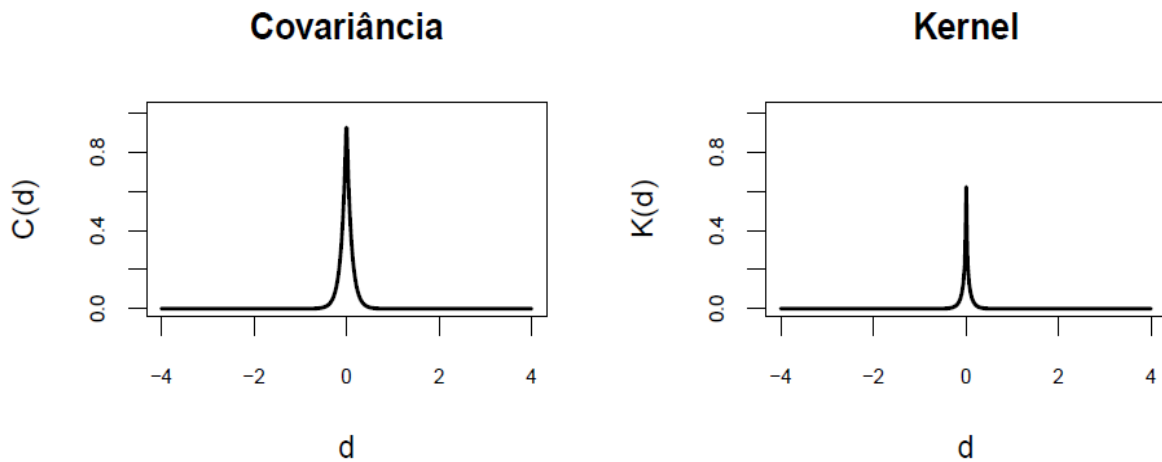


Figura 3.14 Modelo de Covariância Exponencial com $\alpha = 0,1$ e seu kernel associado

3.6.4 Metodologia para Simulação à 1D

Para testar o método simulou-se diversos fenômenos à 1D, seguindo os passos:

- i) Gerou-se o processo ruído-branco com discretização $n = 512$ e um suporte de comprimento 6;
- ii) Calculou-se o kernel associado ao modelo de covariância escolhido;
- iii) Realizou-se a convolução entre o kernel associado e o ruído;
- iv) Calculou-se o variograma do processo simulado e comparou-se o modelo teórico.

a) Realização para o Modelo Gaussiano

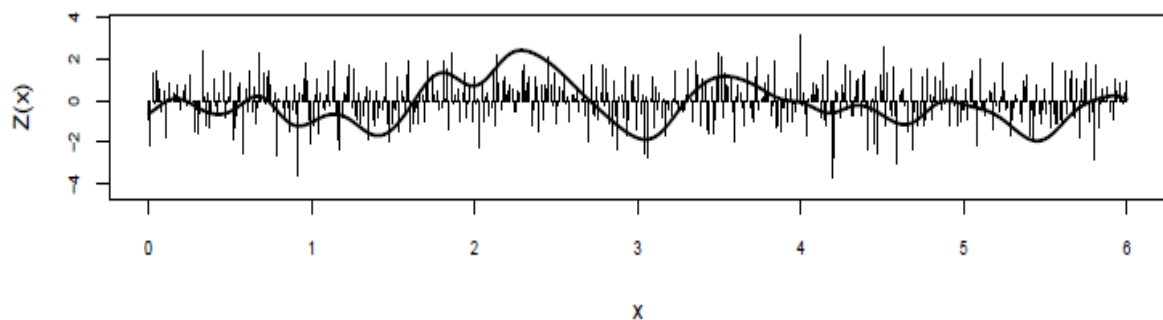


Figura 3.15 Processo ruído-branco com respectivo Modelo Gaussiano associado

Abaixo os respectivos variogramas dos modelos: em linha pontilhada o do modelo teórico (Gaussiano) e em linha contínua o do processo simulado.

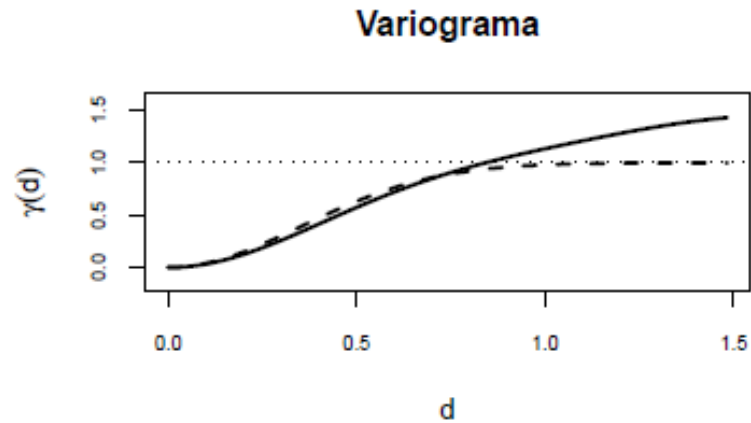


Figura 3.16 Comparação entre os variogramas dos modelos: Gaussiano e Processo de Simulação

b) Realização para o Modelo Exponencial

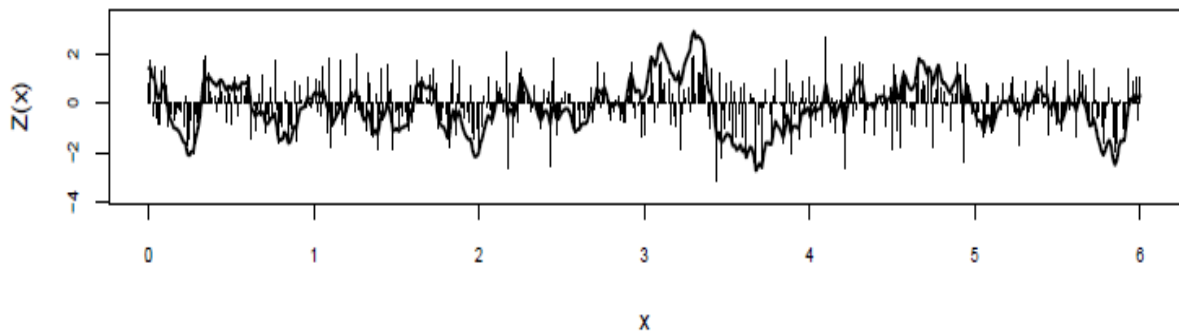


Figura 3.17 Processo ruído-branco com respectivo Modelo Exponencial associado

A seguir os respectivos variogramas dos modelos: em linha pontilhada o do modelo teórico (Exponencial) e em linha contínua o do processo simulado.

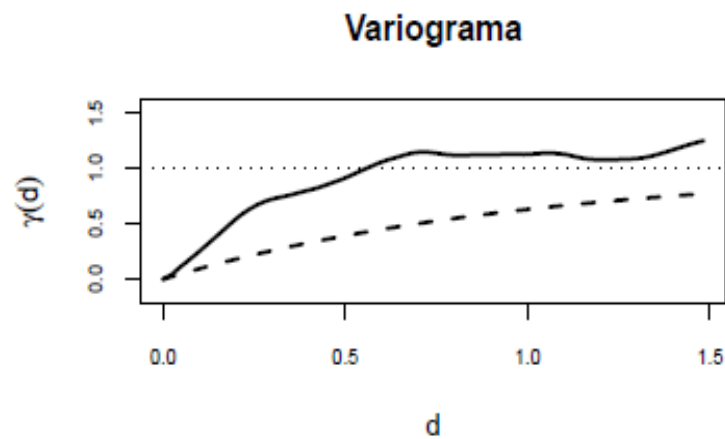


Figura 3.18: Comparação entre os variogramas dos modelos: Exponencial e Processo de Simulação

3.6.5 Simulação para 100 Realizações

No processo de simulação para 100 realizações o método mostra-se eficiente quando em média seus variogramas se aproximam ao variograma do modelo teórico, tanto para o modelo gaussiano, quanto para o modelo exponencial.

a) 100 Simulações para o Modelo Gaussiano

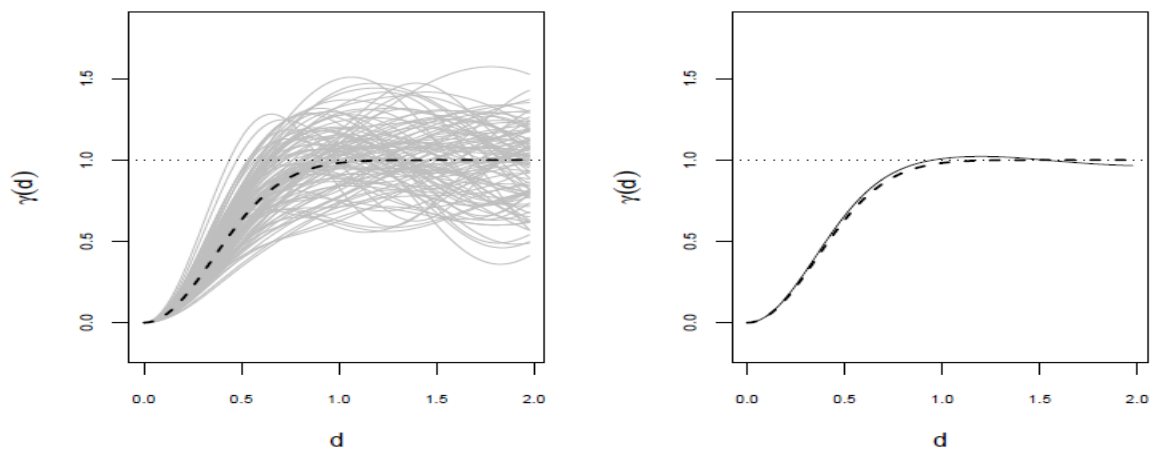


Figura 3.19: Esquerda: Variogramas das 100 simulações. Direita: Média dos Variogramas das simulações comparadas com o Variograma do Modelo Teórico Gaussiano

b) 100 Simulações para o Modelo Exponencial

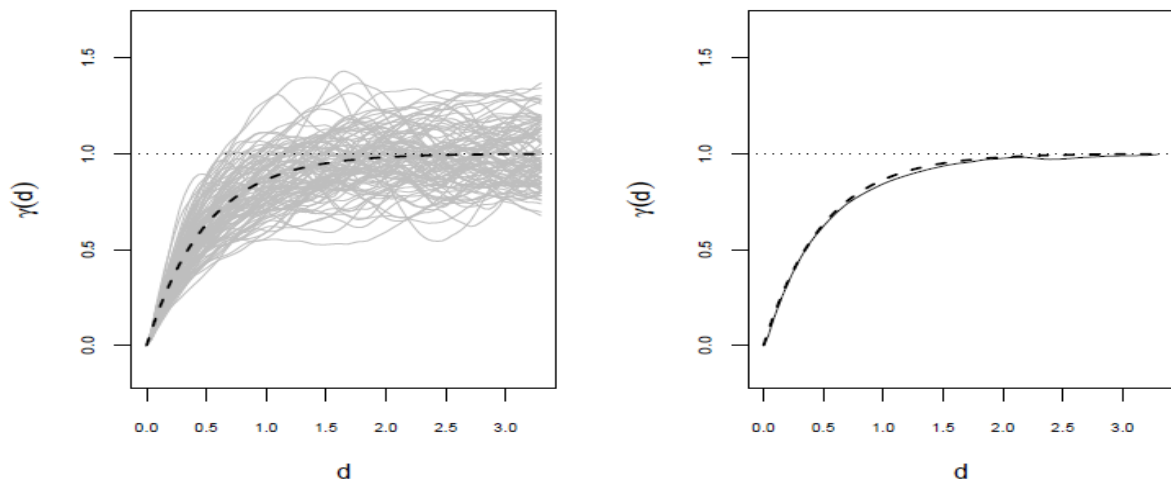


Figura 3.20: Esquerda: Variogramas das 100 simulações. Direita: Média dos Variogramas das simulações comparadas com o Variograma do Modelo Teórico Exponencial

3.7 Resultados 2D

Os dados utilizados neste trabalho, para a convolução à 2D, fazem parte do conjunto de dados Jura e foram coletados pelo Instituto Federal de Tecnologia da Suíça em Lausanne. Uma descrição detalhada desses dados, bem como os procedimentos laboratoriais e de amostragem é visto em (Goovaerts, 1997). Totalizando 359 pontos amostrais (259 de localização e 100 de validação) dispersos no espaço, dividem-se em sete tipos de metais: cádmio, cobalto, cromo, cobre, chumbo, níquel e zinco. Este conjunto de dados compartilha duas características comuns à maioria de dados obtidos em estudo geológicos: 1) os dados são correlacionados no espaço, 2) vários atributos estão envolvidos conjuntamente.

Aqui serão utilizado os 259 pontos amostrados de localização relacionados a concentração de metal cádmio. O cádmio é um elemento químico de símbolo Cd , de número atômico 48 (48 prótons e 48 elétrons) e de massa atômica igual a 112,4 u. À temperatura ambiente, o cádmio encontra-se no estado sólido. Está situado no grupo 12 (2B) da classificação periódica dos elementos. É um metal branco azulado, relativamente pouco abundante. É um dos metais mais tóxicos, apesar de ser um elemento químico essencial, necessário em quantidades muito pequenas, entretanto, sua função biológica não é muito clara. Normalmente é encontrado em minas de zinco, sendo empregado principalmente

na fabricação de pilhas. Foi descoberto em 1817 por Friedrich Strohmeyer.
<<http://pt.wikipedia.org/wiki/Cádmio>>

A Figura 3.21 mostra o mapa de localização dos pontos amostrados do metal, pontos mais largos indicam uma maior concentração do minério em partes por milhões (ppm) no Sistema Internacional (SI) a concentração é medida em $mg.kg^{-1}$.

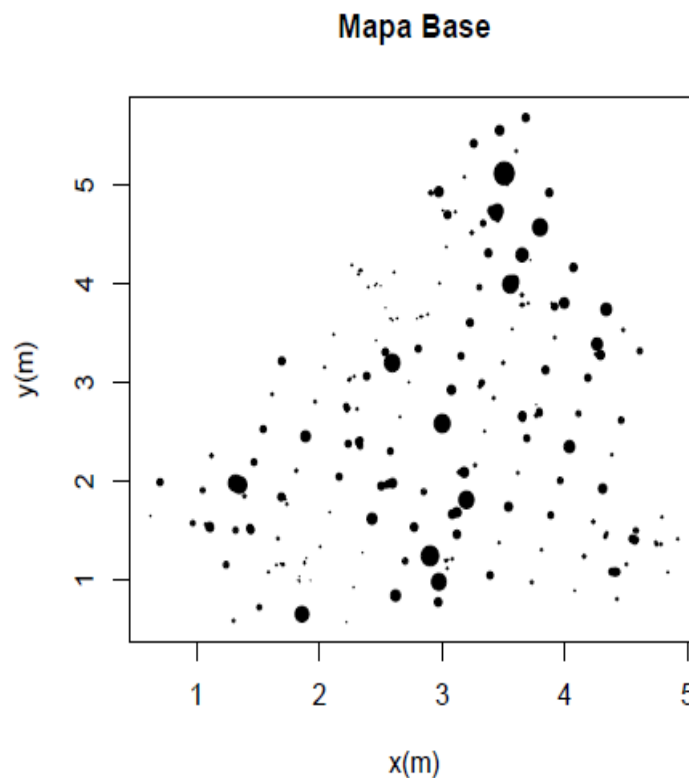


Figura 3.21: Mapa de Localização do metal Cádmio.

A partir do histograma dos dados mostrado na Figura 3.22 observa-se que a calda alongada a direita indica, além de uma assimetria positiva, a pouca concentração de metal. Tem-se que o conjunto é modal e senão o principal, mas um deles, o conjunto não segue uma distribuição normal. Podendo ser também verificada pela curva de normalidade, a grande maioria dos pontos não estão sobre a reta de 45° , isso indica não-normalidade dos dados. Da análise exploratória do conjunto encontra-se a média e o desvio padrão dos dados, sendo respectivamente 1, 30 e 0,46.

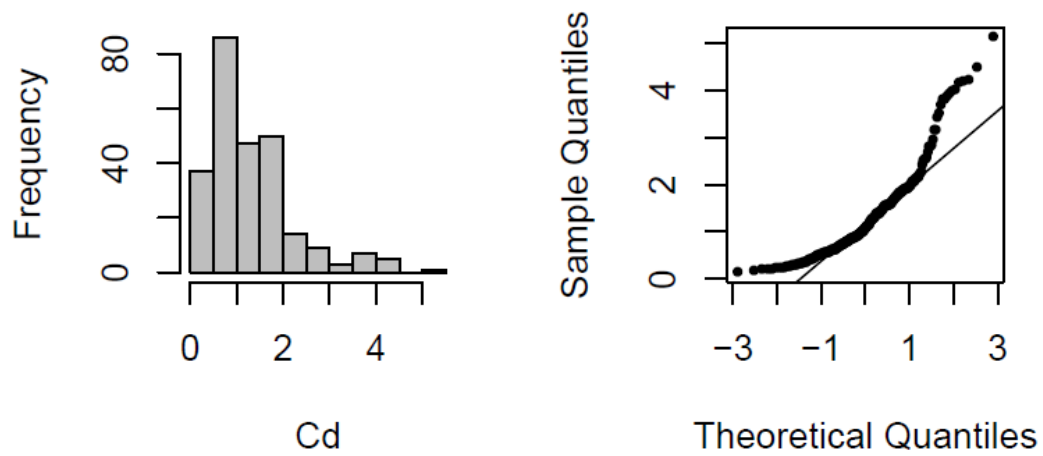


Figura 3.22: Esquerda: Histograma da concentração de metal cádmio. Direita: Curva de Normalidade

Verificada a não-normalidade do conjunto, foi realizada uma transformação, por Box-Cox, nos mesmos dados no intuito de aproximá-los a uma distribuição normal. O resultado é visto na Figura 3.23.

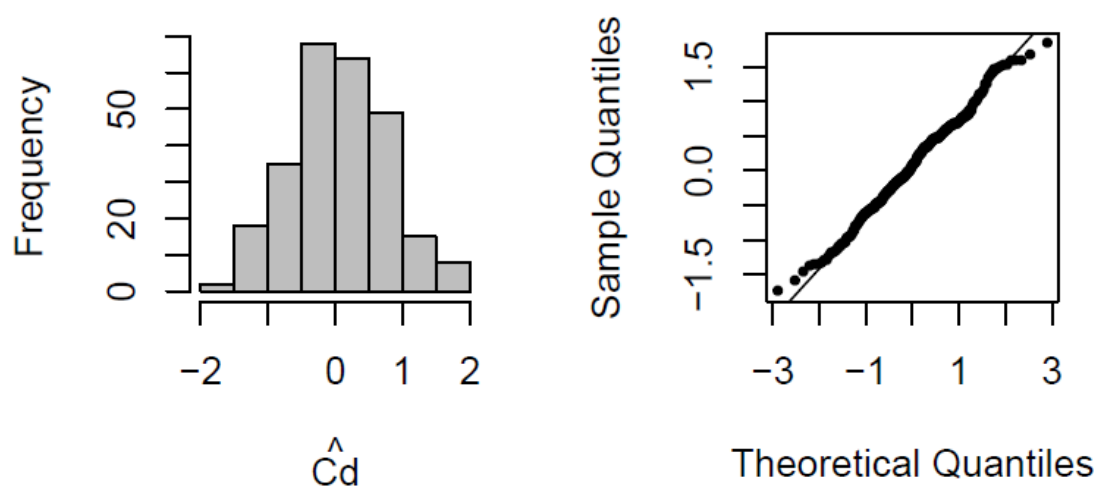


Figura 3.23: Esquerda: Histograma dos dados transformados por Box-Cox da concentração de metal cádmio. Direita: Curva de Normalidade

Observa-se no gráfico que os pontos caem próximos a reta, sugerindo que a amostra, a partir de então, segue aproximadamente uma distribuição normal.

Seguindo, apresentam-se os dois métodos proposto pelo trabalho bem como os resultados e comparações dos mesmos.

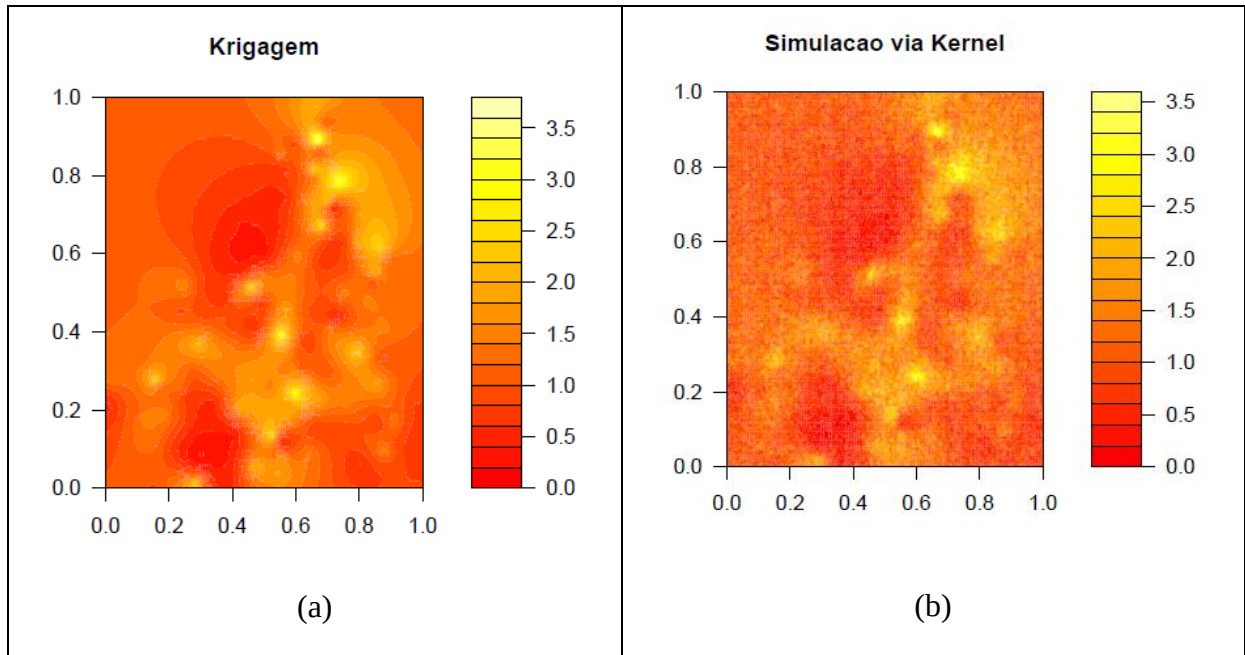


Figura 3.24: (a) Técnica de Krigagem. (b) Simulação via Kernel

Como resultado final dos Processos Geoestatísticos as Figuras 3.24(a) e 3.24(b) apresentam os resultados obtidos relativos à técnica de Krigagem Ordinária e de Simulação via Método de Convolução referentes aos dados de concentração de metal cádmio coletados pelo Instituto Federal de Tecnologia da Suíça em Lausanne.

Nos dois casos, cores mais claras indicam pouca variabilidade, ou seja, a concentração de metal é maior próximo dos pontos amostrados e à medida que se afasta (cores mais escuras) essa variabilidade aumenta, diminuindo a concentração de metal. Em outras palavras, próximos dos pontos amostrados o valor da variância de Krigagem e de Simulação via Kernel é baixo e aumenta à medida que se afasta dos mesmos.

O que torna um dos métodos, aplicados a esse conjunto de dados, mais adequado que o outro é justamente a análise exploratória das técnicas aplicadas, onde os valores da média e do desvio padrão dos processos devem ser iguais ou próximos do valor da média e do desvio padrão dos dados originais. As Figuras 3.25(a) e 3.25(b) mostram os resultados a partir da análise gráfica.

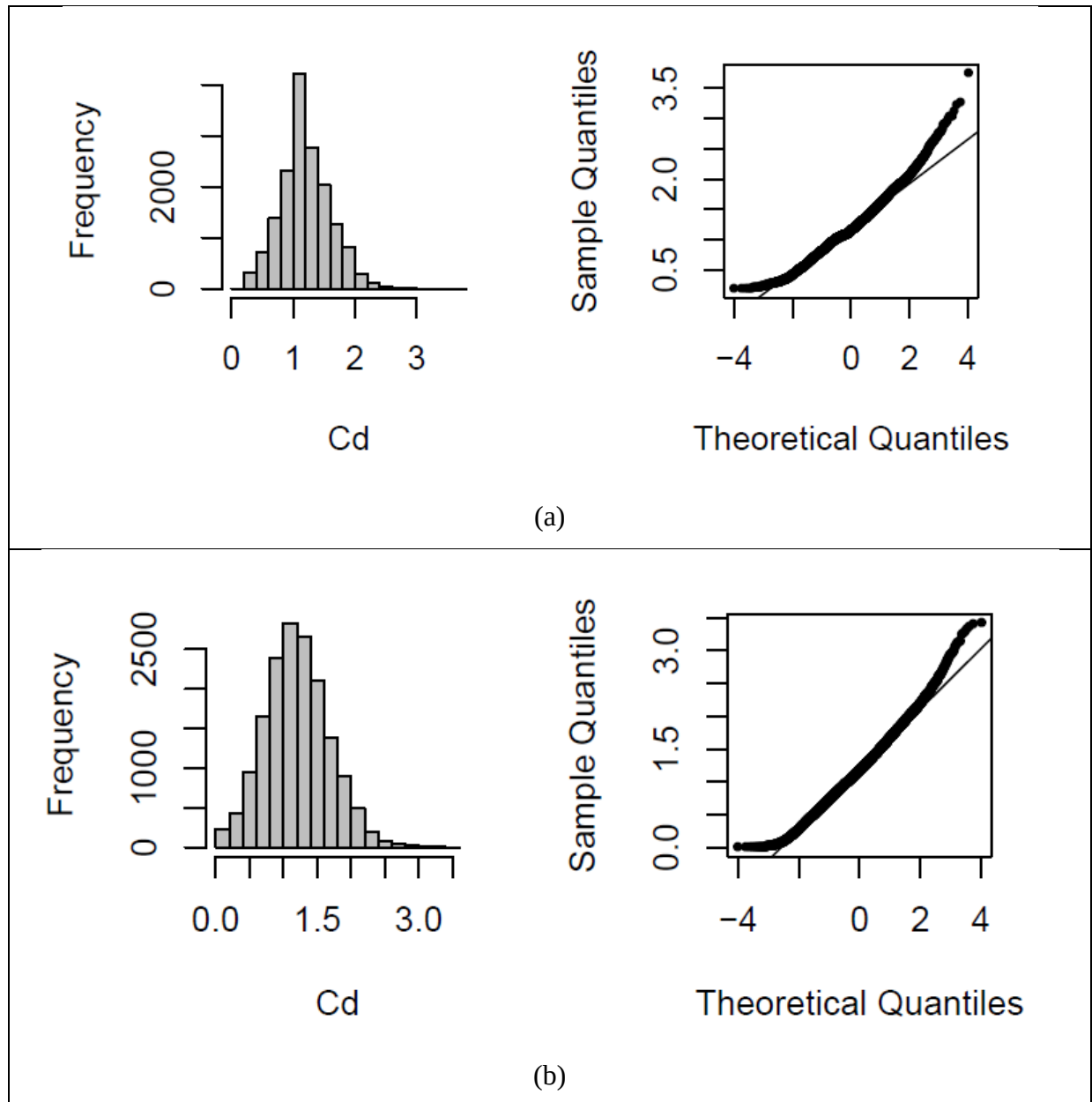


Figura 3.25: (a) Histograma e Curva de Normalidade do Processo de Krigagem. (b) Histograma e Curva de Normalidade do Processo de Simulação via Kernel

Comparando-se as duas imagens, nota-se que no gráfico de normalidade os pontos referentes ao processo de Krigagem se encontram um pouco mais afastados da reta de 45°, indicando que o conjunto de dados não segue uma distribuição normal. Ao contrário dos pontos referentes ao processo de convolução ou kernel. Outro fator, muito importante foi nos valores das médias e dos desvios padrões.

Conjunto de Dados	Média	Desvio Padrão
Originais	1,30	0,46
Krigagem	1,19	0,41
Via Kernel	1,29	0,47

Tabela 2: Comparação de Média e Desvio Padrão dos Métodos de Krigagem e de Simulação via kernel com os Dados Originais

Observando a Tabela 2, acima, pode-se afirmar que, para esse conjunto de dados, a Simulação Geoestatística via Métodos Conolutivos é mais adequada.

Considerações Finais

Para fenômenos naturais, a estatística clássica não envolve qualquer conhecimento da posição das amostras ou do relacionamento entre elas. À esta limitação, a geoestatística, introduz o conceito de variável aleatória regionalizada, ou seja, leva em consideração a posição geográfica da amostra ou do conjunto de amostras.

Os métodos geoestatísticos de estimativa constituem uma importante ferramenta na interpretação de dados geológicos. Entre os métodos geoestatísticos tem-se a Krigagem que se vale de algumas funções bases conhecidas para a determinação do valor da variável de acordo com a distância desta ao ponto amostrado, neste caso essa função se baseia num modelo de variograma e à esse método é associado as técnicas de regressão linear que minimizam a variância das estimativas.

Contrapondo-se algumas limitações da Krigagem, a simulação estocástica é a técnica que consiste basicamente na criação de algoritmos para se simular o comportamento de determinada variável, no caso, o seu comportamento espacial. Imitando um processo ou uma operação do mundo real, que permite praticamente qualquer tipo de operação do comportamento de sistemas reais através do exercício de modelos.

O objetivo deste estudo foi comparar a eficiência do método de Krigagem ordinária e de simulação estocástica por processo convolutivo. A Krigagem ordinária, por sua própria natureza minimiza a variância dos erros e mostra fidelidade na aplicação de alguns conjuntos de dados em termos de apresentarem maiores coeficientes de correlação e minimização dos erros. A simulação, também possui o mesmo objetivo da Krigagem, a diferença se encontra justamente nas vantagens que a simulação possui sobre a mesma, a principal delas é sua precisão global, isto é, reproduz a variabilidade espacial da amostragem em detrimento da manutenção da precisão local. Por reprodução da variabilidade espacial entende-se que as estatísticas multiponto globais das estimativas, tais como variograma e histograma, serão idênticas às da amostragem.

Nas realizações à 1D, as 100 simulações realizadas a partir dos modelos espaciais de covariância gaussiano e exponencial para três tipos de alcances diferentes ($a = 1$, $a = 0,5$ e $a = 0,1$) e convolvidos pelo kernel conveniente à função de covariância mostrou-se eficiente quando a média das realizações simuladas dos modelos espaciais aproximou-se, com pequena diferença, ao variograma do modelo escolhido.

Nas realizações à 2D, a partir das análises gráficas e exploratória do dados originais e os das técnicas propostas por este trabalho revelou que a Simulação Geoestatística via Métodos Convolutivos é o mais adequado para o conjunto de dados - os do Data Set Jura, 259 pontos amostrados de concentração de metal Cádmio.

Os resultados obtidos confirmam a eficácia do método de estimativa utilizado, isso também se dá pelo fato de o método de simulação ser um método de precisão global, e que trabalha com uma variância correspondente à realidade, permitindo muitas vezes uma modelagem melhor das heterogeneidades existentes na natureza e ótimas construções de modelos geológicos mais robustos.

Referências Bibliográficas

BRAGA, L. P. V. Geoestatística e Aplicações. In: Simpósio Brasileiro de Probabilidade e Estatística, 9o. São Paulo: IME/USP, 1990.

BURROUGHT, P. A. Principles of Geographical Information Systems for Land Resources Assessment. Xford, Clarendon Press, 1987.

CAMERGO, O. A. de; ALLEONI, L. P. F. Comparações do Solo e o Desenvolvimento das Plantas. Piracicaba: USP/ESALQ, 1997.

DELFINER, P.; DELHOMME, J. P. Optimum Interpolation by Kriging. In: DAVIS, J. C.; McCULLAGH, M. J. Display and Analysis of Spacial Data. New York, John Wiley, 1975.

DEUTSCH, P.; JOURNEL, A. G. GSLIB: Geostatistical Software Library and user's guid. New York, Oxford University Press, 1992.

DINIZ, R. A. V. Aplicação de Metodologia Geoestatística para a Classificação Geodinâmica de uma Camada da Hangwall na mina. São Bento, São Paulo, 1997. Dissertação de Mestrado em Geotecnia. Escola de Engenharia de São Carlos, Universidade de São Paulo.

FURUIE, Rafael de Aguiar. Estudo Comparativo de Métodos Geoestatísticos de Estimativas e Simulação Estocásticas Condicionais. Dissertação de Mestrado, Programa de Pós Graduação em Recursos Minerais e Hidrologia. São Paulo, 2009.

GARCÊS, Ana Cristina Alves. Avaliação da Distribuição Espacial de Dados Biológicos via Krigagem e Simulação Condicional. Dissertação de Mestrado, Programa de Pós-Graduação em Matemática e Estatística, Belém, 2009.

GUERRA, P. A. G. Geoestatística Operacional. Brasília: Ministério das Minas e Energia, 1988.

GOOVAERTS, P. Assessment of Local Uncertainty in Geostatistics for Natural Resouce Evaluation. Oxford University Press, 1997.

HUIJBREGTS, C. J. Regionalized Variables and Quantitative Analysis of Spatial Data. In: DAVIS, J. C.; McCULLAGH, M. J. Display and Analysis of Spatial Data. New York, John Wiley, 1975.

ISAAKS, E. H., SRIVASTAVA, R. M. Applied Geostatistics. New York: Oxford University Press, 1989.

JOURNEL, A. G., XU, W. Posterior Identification of Histograms Conditional to Local Data. Mathematical Geology, 1994.

KRIGE, D. G. A Statistical Approach to some Basic mine Evaluation Problems on the Witwatersrand. Johannesburg Chemistry Metallurgy Mining Society South African, 1951.

LANDIM, P. M. B. Análise Estatística de Dados Geológicos. São Paulo: Ed. UNESP, 1998.
JOURNEL, A. G., HUIJBREGTS, Ch. J. Mining Geostatistics, 1998.

MARQUES FILHO, Ogê; VIEIRA NETO, Hugo. *Processamento Digital de Imagens*, Rio de Janeiro: Brasport, 1999. ISBN 8574520098.

MATHERON, G. The Theory of Regionalized Variables and its Applications. Les Cahier du Centre de Morphologie Mathématique, 1971.

MATHERON, G. Les Variables Régionalisées et Leur Estimation. Paris: Masson et *C^{ie}*, 1963.

OLEA, R. A. Geostatistics for Engineers and Earth Scientists. Kluwer Academic Publishers, 1999.

ATTEIA, O.; J. P. Dubois; R. Webster. Corregionalization of Trace Metals in the Soil in the Swiss Jura. European Journal of Soil Science, 1994.

ATTEIA, O; J. P. Dubois; R. Webster. Geostatistical Analysis of Soil Contamination in the Swiss Jura. Environmental Pollution, 1994.

VIEIRA, S. R., HATFIELD, J. L., NIELSEN, D. R., BIGGAR, J. W. Geostatistical Theory and Application to Variability of some Agronomical Properties. Hilgardia, 1983.

SILVA, A. P da. Variabilidade Especial de Atributos Físicos do Solo. Piracicaba, 1988. Tese de Doutorado em Solos – Escola Superior de Agricultura “Luiz de Queiroz”, Universidade de São Paulo, Piracicaba, 1988.

SRIVASTAVA, R. M., 1994. An Overview of Stochastic Methods for Reservoir Characterization in Yarus, J. M. & Chamber, R. L. (Eds.), Stochastic Modeling and

Geostatistic: Principles, Methods and Case Studies. American Association of Petroleum Geologists – Computer Applications in Geology, 1994.

YAMAMOTO, J. K. A Review of Numerical Methods for the Interpation of Geological Data. Anais da Academia Brasileira de Ciências, vol. 70, 1998.

<<http://pt.wikipedia.org/wiki/Cádmio>> Acesso: 20 de outubro de 2011.