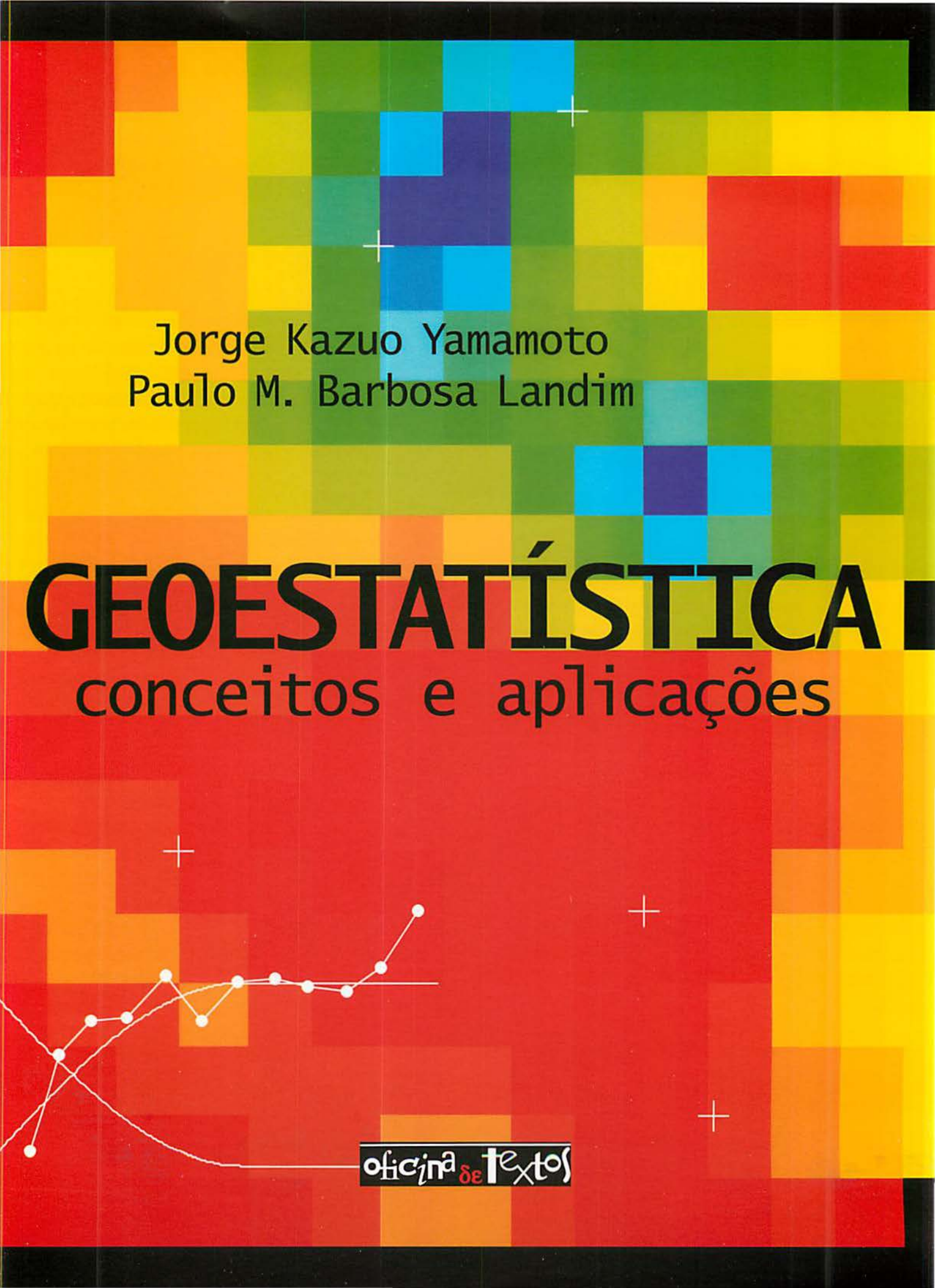
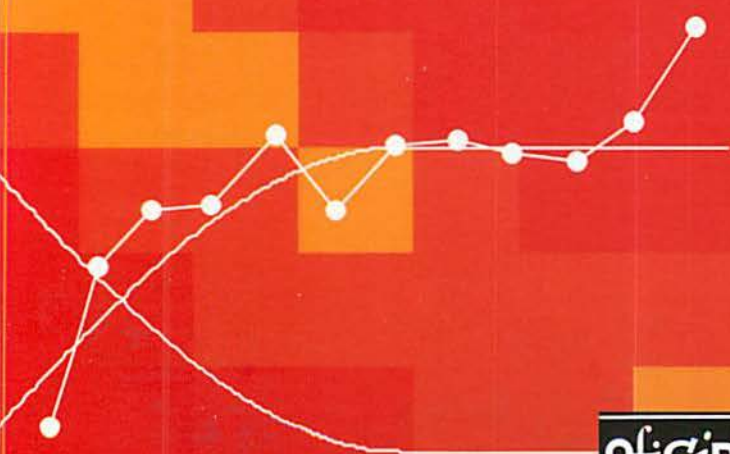




Jorge Kazuo Yamamoto
Paulo M. Barbosa Landim

GEOSTATÍSTICA

conceitos e aplicações



oficina de textos

Jorge Kazuo Yamamoto
Paulo M. Barbosa Landim

GEOESTATÍSTICA

conceitos e aplicações

oficina de textos

Copyright © 2013 Oficina de Textos

1ª reimpressão 2015

Grafia atualizada conforme o Acordo Ortográfico da Língua Portuguesa de 1990, em vigor no Brasil a partir de 2009.

Conselho editorial

Cylon Gonçalves da Silva; Doris C. C. K. Kowaltowski;
José Galizia Tundisi; Luis Enrique Sánchez; Paulo Helene;
Rozely Ferreira dos Santos; Teresa Gallotti Florenzano

Capa e projeto gráfico Malu Vallim

Diagramação Casa Editorial Maluhy Co.

Preparação de textos Cássio Pelin

Revisão de textos Hélio Hideki Iraha

Impressão e acabamento Gráfica VIDA & CONSCIÊNCIA

Dados Internacionais de Catalogação na Publicação (CIP)
(Câmara Brasileira do Livro, SP, Brasil)

Yamamoto, Jorge Kazuo
Geoestatística : conceitos + aplicações / Jorge
Kazuo Yamamoto, Paulo M. Barbosa Landim. --
São Paulo : Oficina de Textos, 2013.

ISBN 978-85-7975-077-9

1. Geoestatística 2. Geologia - Métodos
estatísticos I. Landim, Paulo M. Barbosa.
II. Título.

13-04311

CDD-551

Índices para catálogo sistemático:
1. Geoestatística 551

Todos os direitos reservados à Editora Oficina de Textos

Rua Cubatão, 959

CEP 04013-043 São Paulo SP

tel. (11) 3085 7933 fax (11) 3083 0849

www.ofitexto.com.br

atend@ofitexto.com.br



Apresentação

Geoestatística: conceitos e aplicações é um livro introdutório às bases e conceitos fundamentais da área. Trata-se de leitura essencial para todos aqueles que procuram na Geoestatística um conjunto de instrumentos para resolver problemas concretos na gestão de recursos naturais. Os autores, Jorge Yamamoto e Paulo Landim, cientistas ligados à prática das ciências da Terra, conceberam esta obra num formato que todos os livros fundamentais de ciências aplicadas deveriam ter: dos problemas para as soluções.

Começando por sublinhar o nascimento da Geoestatística num ambiente geológico e mineiro (com os “criadores” Georges Matheron, Daniel Krige, André Journel e Alain Marechal), os autores têm a preocupação de mostrar, ao longo de *Geoestatística: conceitos e aplicações*, a aplicabilidade dos métodos aos diversos domínios das ciências da Terra e do ambiente, isto é, à caracterização de fenômenos físicos de qualquer fenômeno natural estruturado no espaço. Como os autores citam, o livro “dedica-se à análise de dados geológicos controlados pela sua distribuição espacial, mas pode perfeitamente ser utilizado em outras áreas que também disponham de dados georreferenciados”.

Mas Jorge Yamamoto e Paulo Landim também são docentes, o que faz com que *Geoestatística: conceitos e aplicações* tenha um forte componente pedagógico, conferindo a todos os temas abordados uma clareza de exposição e uma grande preocupação com os detalhes dos formalismos matemáticos e seus algoritmos. Com efeito, numa altura em que a Geoestatística está difundida por inúmeros campos de aplicação, com algoritmos e metodologias implementados em *softwares* apelativos e amigáveis, a leitura desta obra é fundamental para a reeducação da maioria dos utilizadores da Geoestatística, cada vez mais transformada em *push-buttons*, que privilegiam o exercício experimental e repetitivo de menus imensos de métodos à sua compreensão e à avaliação do erro da sua má utilização.

Dividido em cinco capítulos, o livro começa pela análise de padrões espaciais dos fenômenos estruturados e modelos de instrumentos simples, como os variogramas e as covariâncias espaciais. Contudo, sua maior parte é dedicada aos métodos de inferência

espacial da extensa família de estimadores lineares, a krigagem. Nessa parte nobre do livro, fica evidente a intenção dos autores em referir e detalhar os métodos mais usuais da prática geoestatística. Eles finalizam a obra com um capítulo dedicado à quantificação da incerteza espacial pelos novos modelos de simulação estocástica.

Estou certo de que o ensino e a prática da Geoestatística no Brasil vão ficar substancialmente mais ricos com a publicação deste livro.

Prof. Dr. Amílcar Soares

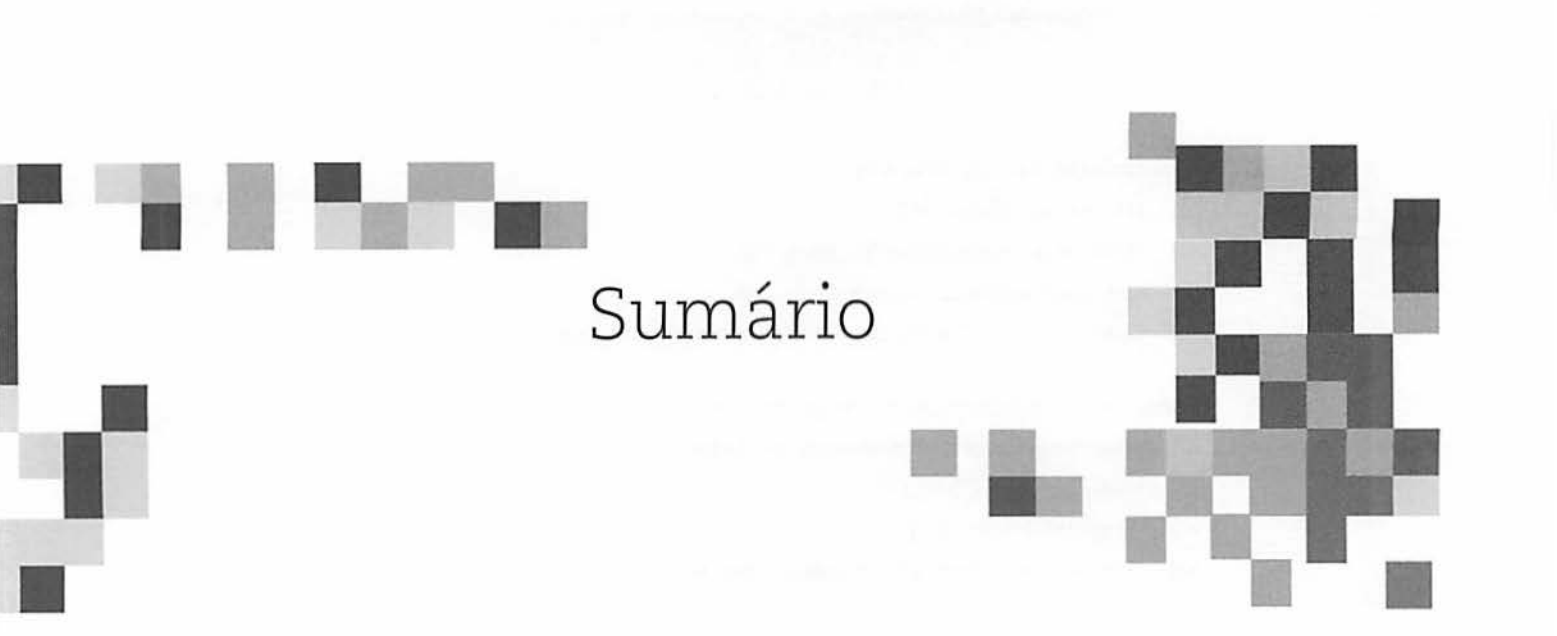
Diretor do Centre for Natural Resources and Environment (CERENA)
do Instituto Superior Técnico (IST) da Universidade Técnica de Lisboa, Portugal



Agradecimentos

Os autores expressam os seus agradecimentos:

- às respectivas universidades, Universidade de São Paulo (USP) e Universidade Estadual Paulista (Unesp), que proporcionaram as condições necessárias para suas atividades didáticas, bem como para o desenvolvimento de pesquisas cujos resultados estão consolidados nesta obra;
- ao Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela concessão de bolsas de produtividade em pesquisa que estimulam a produção científica no País;
- a Thelma Samara, da Seção de Ilustração Geológica do Instituto de Geociências da Universidade de São Paulo (USP), pela edição de parte das figuras desta obra;
- ao engenheiro Antonio Tadashi Kikuda, do Laboratório de Informática Geológica do Departamento de Geologia Sedimentar e Ambiental do Instituto de Geociências da USP, pelo auxílio no algoritmo para o teste de bigaussianidade utilizado nesta obra.



Sumário

Introdução, 9

Breve histórico da Geoestatística, 9

Objetivos, 12

Organização do livro, 12

1 Conceitos Básicos, 19

1.1 – Fenômeno espacial, 19

1.2 – Amostra e métodos de amostragem, 20

1.3 – Inferência espacial, 21

1.4 – Variáveis aleatória e regionalizada, 24

1.5 – Desagrupamento, 26

2 Cálculo e Modelagem de Variogramas Experimentais, 33

2.1 – Estatísticas espaciais, 33

2.2 – Cálculo de variogramas experimentais, 36

2.3 – Tipos de variogramas, 41

2.4 – Anisotropias, 43

2.5 – Comportamento do variograma próximo à origem, 47

2.6 – Considerações finais, 52

3 Estimativas Geoestatísticas, 55

3.1 – Transformação de dados, 56

3.2 – Estimativas geoestatísticas, 62

3.3 – Krigagem não linear, 83

3.4 – Interpolação de variáveis categóricas, 106

3.5 – Considerações finais, 117

4 Coestimativas Geoestatísticas, 121

4.1 – Cokrigagem, 123

4.2 – Krigagem com deriva externa, 135

4.3 – Considerações finais, 141

5 Simulação Estocástica, 145

5.1 – Erro de suavização, 147

5.2 – Métodos de simulação estocástica, 147

5.3 – Métodos sequenciais de simulação, 148

5.4 – Considerações sobre os métodos de simulação estocástica, 173

Anexo A – Fundamentos Matemáticos e Estatísticos, 175

A.1 – Métodos gráficos de apresentação de dados, 175

A.2 – Estatística descritiva, 177

A.3 – Estatística bivariada, 179

A.4 – Distribuições teóricas de probabilidades, 182

A.5 – Derivadas, 184

A.6 – Integral, 184

A.7 – Matrizes, 185

A.8 – Sistemas de equações lineares, 188

A.9 – Software, 192

Anexo B – Arquivos de Dados, 195

Sobre os autores, 216



Introdução

O professor Georges Matheron, inspirado inicialmente nos trabalhos pioneiros de H. J. de Wijs (De Wijs, 1951, 1953), professor da Universidade Técnica de Delft, na Holanda, e Daniel G. Krige (Krige, 1951), engenheiro de minas que trabalhou nas minas de ouro do Rand, na África do Sul, apresentou, no anos 1960, uma série de publicações que, por sua importante contribuição para o estudo e formalização da Teoria das Variáveis Regionalizadas, o distingue como criador da Geoestatística (Matheron, 1962, 1963, 1965, 1971).

Segundo Matheron (1971, p. 5), uma variável regionalizada é uma função $f(x)$ do ponto x , mas também é uma função irregular na qual se têm dois aspectos contraditórios ou complementares: um aspecto aleatório, cuja irregularidade não permite prever as variações de um ponto a outro; e um aspecto estruturado, que reflete as características estruturais do fenômeno regionalizado. Para Matheron, a Teoria das Variáveis Regionalizadas tem dois objetivos: teoricamente, descrever a correlação espacial; na prática, resolver problemas de estimativa de uma variável regionalizada com base em uma amostra.

BREVE HISTÓRICO DA GEOESTATÍSTICA

Matheron, formado pela École Normale Supérieure des Mines de Paris, criou, em 1968, em Fontainebleau, próximo a Paris, o Centre de Morphologie Mathématique, posteriormente subdividido em dois centros de pesquisa de importância fundamental para o estudo, difusão e formação de pesquisadores: Morfologia Matemática e Geoestatística.

André G. Journel e Michel David, ex-alunos de Matheron, foram os responsáveis por sua difusão na América do Norte e, entre outras obras, publicaram dois importantes livros: *Geostatistical ore reserve estimation* (David, 1977) e *Mining geostatistics* (Journel; Huijbregts, 1978). Michel David foi contratado pela Escola Politécnica de Montreal, no Canadá, e André Journel, pela Universidade de Stanford, nos Estados Unidos, onde criou o Stanford Center for Reservoir Forecasting (SCRF), do qual foi diretor, entre 1984 e 1997, e responsável pelo início da aplicação da Geoestatística na Geologia do Petróleo.

Esses professores, em suas escolas, também formaram alunos, dos quais se destaca Clayton V. Deutsch, que, após a pós-graduação na Universidade de Stanford, retornou

à Universidade de Edmonton, na qual se graduara em Engenharia de Minas e Petróleo. Clayton criou o Centre for Computational Geostatistics (CCG), que funciona da mesma forma que o SCRF. O CCG é mantido por empresas e universidades associadas, que recolhem uma taxa anual cuja receita é revertida em bolsas de estudo a alunos de pós-graduação. Clayton Deutsch colabora ativamente em periódicos internacionais e produziu obras como *Geostatistical reservoir modeling* (Deutsch, 2002), voltada à Geoestatística aplicada à modelagem de reservatórios de petróleo e gás.

Outro importante centro de aplicação não só da Geoestatística, mas também de desenvolvimento de técnicas de modelagem de reservatórios, é o Consórcio GoCad, na Universidade de Lorraine, na França. Ele foi criado em 1969 por Jean-Laurent Mallet, com o objetivo de apoiar as pesquisas desenvolvidas no âmbito acadêmico e solucionar problemas encontrados na indústria. O software GoCad, principal produto desse consórcio, é comercializado atualmente pela Paradigm, com o nome comercial de Skua. O Consórcio GoCad é suportado financeiramente por 18 empresas e 131 universidades, entre as quais a Universidade de São Paulo (USP), por meio do Instituto de Geociências. O professor Mallet foi responsável pelo consórcio da sua criação até 2006. Desde 2007, ele é dirigido pelo professor Guillaume Caumon.

As ideias de Matheron, porém, inicialmente suscitaram forte oposição por parte de geólogos e engenheiros de minas. Assim, por exemplo, com relação ao estimador da krigagem, Whitten (1966) preferia a interpolação por regressão polinomial, isto é, por análise de superfície de tendência. Matheron (1967) respondeu a essa crítica num artigo denominado *Kriging, or polynomial interpolation procedures?*.

A partir da década de 1980, a metodologia geoestatística passou a ter ampla aplicação, pois, além de Lavra e Prospeção Mineira, é utilizada em Agricultura de Precisão, Análise Espacial de Crimes, Cartografia, Climatologia, Ecologia da Paisagem, Engenharia Florestal, Epidemiologia, Geologia Ambiental, Geologia do Petróleo, Geotecnia, Hidrogeologia e Pedologia. Praticamente todas as últimas versões de softwares para confecção de mapas ou sistemas de informações georreferenciadas apresentam módulos com métodos geoestatísticos.

A Teoria das Variáveis Regionalizadas, já consagrada, tem por objetivo o estudo e a representação estrutural desse tipo de variável para a resolução de problemas de estimativa, com base em dados experimentais medidos sobre suportes que não abrangem totalmente tais domínios.

O melhor estimador para uma variável regionalizada deve levar em consideração as respectivas posições relativas e, portanto, a característica estrutural do fenômeno. Qualquer variável dependente do espaço que apresente, além do caráter aleatório, um caráter estrutural, pode ser tratada como variável regionalizada e sofrer uma análise segundo o formalismo desenvolvido pela Geoestatística. O termo geoestatística tem uma abrangência mais ampla do que a dada originalmente por Matheron (1971), e pode ser definido como uma subárea da Estatística que estuda variáveis regionalizadas.

Os métodos geoestatísticos fornecem um conjunto de técnicas necessárias para entender a aparente aleatoriedade dos dados, os quais apresentam, porém, uma possível estruturação espacial, estabelecendo, desse modo, uma função de correlação espacial.

Essa função representa a base da estimativa da variabilidade espacial em Geoestatística. Chilés e Delfiner (1999) e Soares (2006) apresentam uma revisão histórica sobre a Geoestatística com uma síntese sobre o desenvolvimento de suas técnicas, sendo o seu início ligado a problemas de lavra mineira.

A avaliação de reservas minerais é de extrema importância em todas as etapas de um projeto de mineração, da fase de pesquisa mineral até o estudo de viabilidade técnica e econômica do empreendimento. Além disso, no desenvolvimento da mina, a Geoestatística tem um papel fundamental no planejamento de lavra de curto, médio e longo prazos, pois, por meio de estimativas atualizadas das reservas minerais, pode auxiliar na tomada de decisões na operação da mina. As estimativas de reservas minerais são baseadas em amostras (sondagens, canaletas, galerias etc.) e, por isso, estão sujeitas a incertezas. Nesse sentido, o problema está em como avaliar as incertezas, as quais são baseadas em um modelo de distribuição de probabilidades.

É importante diferenciar erros de incertezas, pois os primeiros dependem do conhecimento dos valores verdadeiros da variável estimada. A avaliação de reservas minerais é sempre feita com base em blocos de cubagem, que devem ser estimados a partir de amostras coletadas em sua vizinhança. Seja, por exemplo, um bloco a ser estimado com base em cinco amostras (Fig. 1).

Supondo que ocorra uma correlação espacial entre os teores, os valores serão muito próximos em dois pontos vizinhos e progressivamente mais diferentes à medida que os pontos ficarem mais distantes. Nesse sentido, é de se esperar que o teor da amostra 3 seja similar ao teor médio do bloco. Isso significa que o teor da amostra 3 apresenta uma correlação com o teor do bloco. Pode-se esperar que as amostras 1, 4 e 5 também apresentem teores similares ao valor médio do bloco, mas não tanto como o teor em 3. Finalmente, com relação à amostra 2, mais distante em relação ao bloco, ela entraria com peso menor em relação às outras. Em outras palavras, amostras situadas perto do bloco deverão apresentar teores altamente relacionados com ele e poderão, portanto, ser utilizadas para estimar o seu valor médio, e, à medida que se situem a distâncias maiores, o seu relacionamento diminui até se tornarem independentes. A influência de cada amostra é inversamente proporcional à distância. Esse é um conceito compartilhado por diferentes métodos de estimativas, sejam elas geoestatísticas ou não. A diferença está na forma em que esses ponderadores são calculados. A Geoestatística proporciona um conjunto de métodos para a estimativa de reservas minerais, sempre fazendo o melhor uso da informação disponível. Isso significa que, para uma dada situação ou fase da pesquisa ou de desenvolvimento da mina, não se justifica amostragem adicional com a intenção de melhorar o variograma que será utilizado na krigagem. Entre os problemas operacionais que a Geoestatística pode resolver estão: definição da quantidade e localização de amostras vizinhas para estimativa de um bloco; reconhecimento e tratamento de amostras agrupadas por amostragens preferenciais ou detalhadas de zonas mais ricas em minério; tipo de mineralização em estudo (distribuição e variabilidade espaciais da variável de interesse); transformação de variáveis; geometria

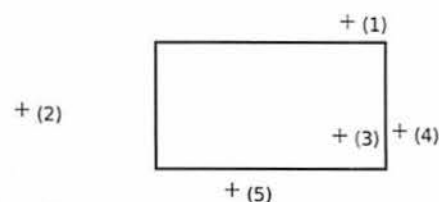


Fig. 1 Determinação do valor de uma área com base em cinco pontos com valores conhecidos

Fonte: desenho adaptado de Clark (1979, p. 3).

do corpo de minério; avaliação e mapeamento de incertezas; parametrização das reservas minerais em curvas teor/tonelagem, bem como variância global do depósito mineral.

Como fontes introdutórias são recomendados os livros de Clark (1979), Rendu (1981), Armstrong (1998), Brooker (1991), Clark e Harper (2000), Andriotti (2003), Landim (2003), Druck et al. (2004) e Olea (2009). Devem ser citados também diversos textos que tratam de aplicações da Geoestatística, como Journel e Huijbregts (1978), Valente (1982), Guerra (1988), Isaaks e Srivastava (1989), Deutsch e Journel (1992), Cressie (1993), Samper-Calvete e Carrera-Ramírez (1996), Goovaerts (1997), Hohn (1999), Olea (1999), Yamamoto (2001a), Soares (2006), Webster e Oliver (2007) e Oliver (2010).

Um extenso estudo bibliométrico sobre textos, tanto em livros como em artigos, relativos à Geoestatística é apresentado por Hengl, Minasny e Gould (2009). Nesse trabalho, como referência à origem geográfica dos autores, na América do Sul, são destaques as regiões de São Paulo/Brasil e Santiago/Chile (Hengl; Minasny; Gould, 2009, p. 508).

OBJETIVOS

O principal objetivo deste livro, baseado na experiência dos dois autores, é mostrar de maneira clara, simples e objetiva a metodologia geoestatística em suas diversas aplicações. Dedicar-se principalmente à análise de dados geológicos controlados pela sua distribuição espacial, mas pode perfeitamente ser utilizada em outras áreas que disponham também de dados georreferenciados. A teoria geoestatística foi baseada na literatura corrente, que foi referenciada com a maior precisão possível, indicando autor, ano e página.

ORGANIZAÇÃO DO LIVRO

Geoestatística: conceitos e aplicações está organizado em cinco capítulos. Evidentemente, o texto não tem a pretensão de cobrir todas as técnicas e campos de aplicação da Geoestatística, mas introduzir conceitos e técnicas fundamentais atualmente em uso.

O Cap. 1 aborda conceitos básicos envolvendo amostra e população (fenômeno espacial), métodos de amostragem, o problema da inferência espacial (Fig. 2) e a natureza das variáveis aleatórias contínuas e discretas.

É importante ressaltar que o estudo geoestatístico tem início com a coleta de uma amostra, que será usada para inferir as características da população ou do fenômeno espacial de interesse da pesquisa.

A amostragem deve ser feita em disposição regular ou o mais próximo disso, mas podem ocorrer amostragens preferenciais em zonas de maior interesse que acabam produzindo agrupamentos de pontos.

Esses agrupamentos devem ter seus efeitos atenuados para não distorcer as estatísticas globais, tais como o histograma e o variograma. Assim, são apresentadas duas técnicas de desagrupamento de amostras (polígonos e células).

Atualmente, os conceitos da Geoestatística podem ser aplicados tanto a variáveis contínuas como a discretas. Nesse sentido, abre-se uma gama de aplicações envolvendo

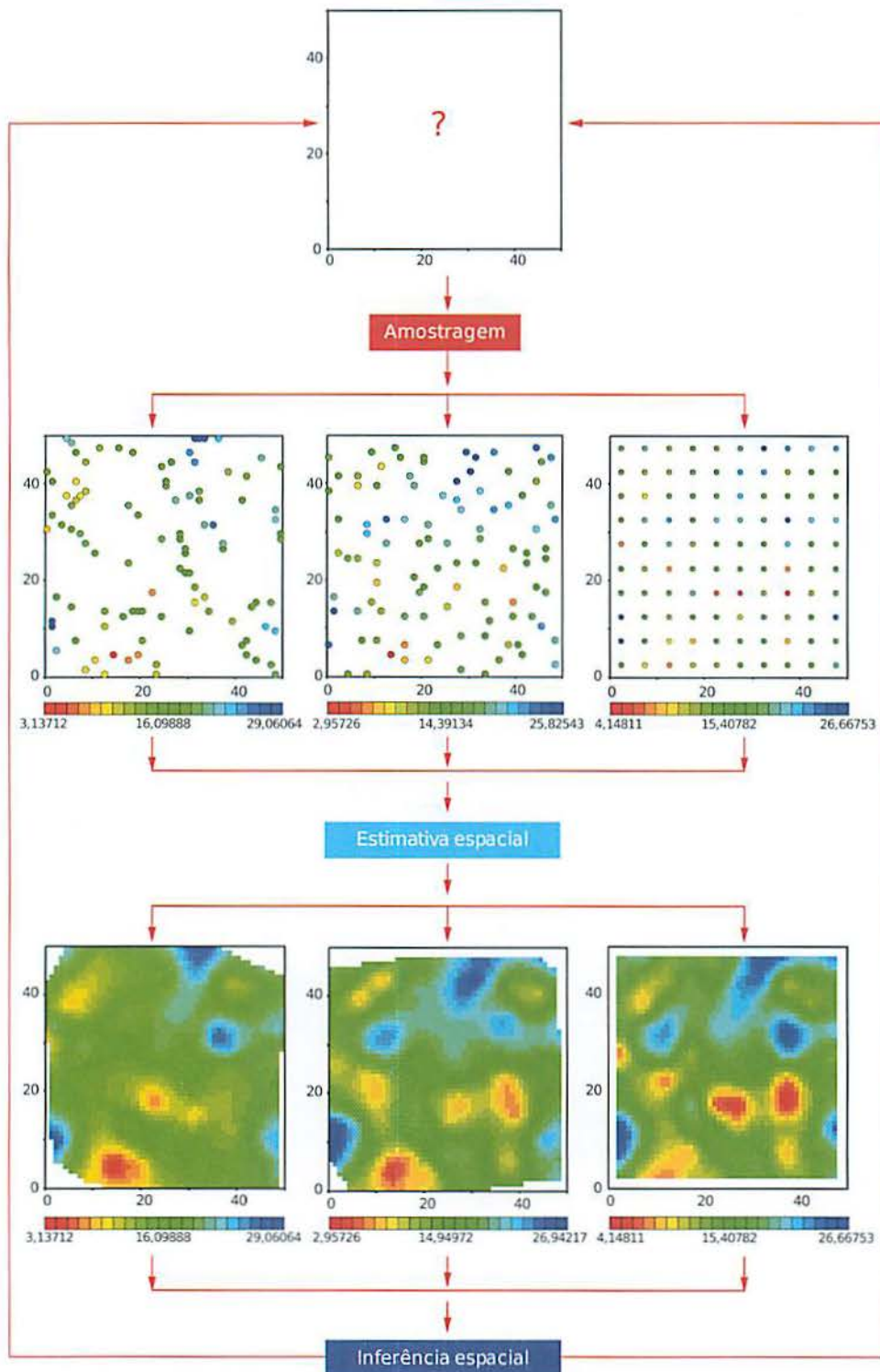


Fig. 2 Esquema mostrando o processo de inferência do fenômeno espacial com base na amostragem (seção 1.3)

variáveis discretas, pois elas são frequentemente observadas nos pontos de amostragem em que são feitas medidas de variáveis contínuas.

O Cap. 2 é voltado ao cálculo e modelagem de variogramas experimentais, e introduz os conceitos de estacionaridade, hipótese intrínseca, cálculo de variogramas expe-

rimentais, modelos teóricos de variogramas, anisotropias e graus de continuidade na origem.

Uma síntese do procedimento de cálculo e modelagem de variogramas experimentais pode ser vista na Fig. 3. O variograma depende fundamentalmente da direção e da distância, as quais permitem calcular o variograma experimental e verificar a hipótese intrínseca (Fig. 3C,D).

O Cap. 3 apresenta técnicas geoestatísticas de estimativa e interpolação para variáveis aleatórias contínuas e discretas (Fig. 4). Os métodos geoestatísticos de estimativa foram divididos em krigagem linear e não linear. As técnicas da krigagem simples, da média e ordinária foram incluídas como técnicas lineares, pois fazem uso da variável contínua na escala original de medida. Métodos que fazem uso da transformação não linear de dados foram classificados como krigagem não linear: krigagem multigaussiana, krigagem lognormal e krigagem indicadora. Além disso, esse capítulo apresenta uma seção especial

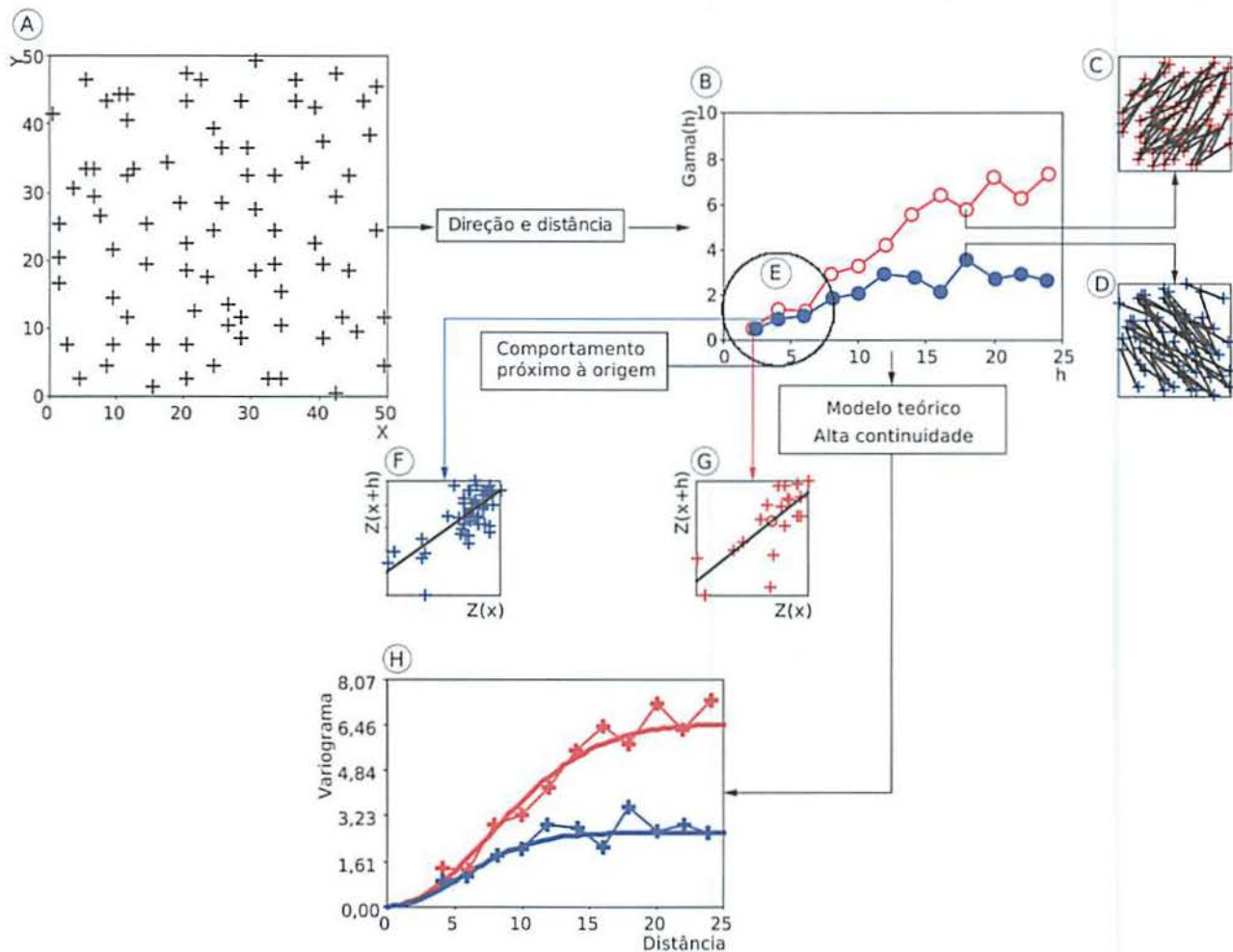


Fig. 3 Síntese do procedimento de cálculo e modelagem de variogramas experimentais: A) mapa de pontos; B) variogramas experimentais calculados para as direções de 45° (vermelho) e 135° (azul); C) vetores usados no cálculo do variograma experimental para a direção de 45°; D) vetores usados no cálculo do variograma experimental para a direção de 135°; E) destaque para o comportamento próximo à origem, com alta continuidade; F) interpretação geométrica de Journé (1989) para a direção de 135°; G) interpretação geométrica de Journé (1989) para a direção de 45°; H) modelos teóricos ajustados aos variogramas experimentais (seção 2.6)

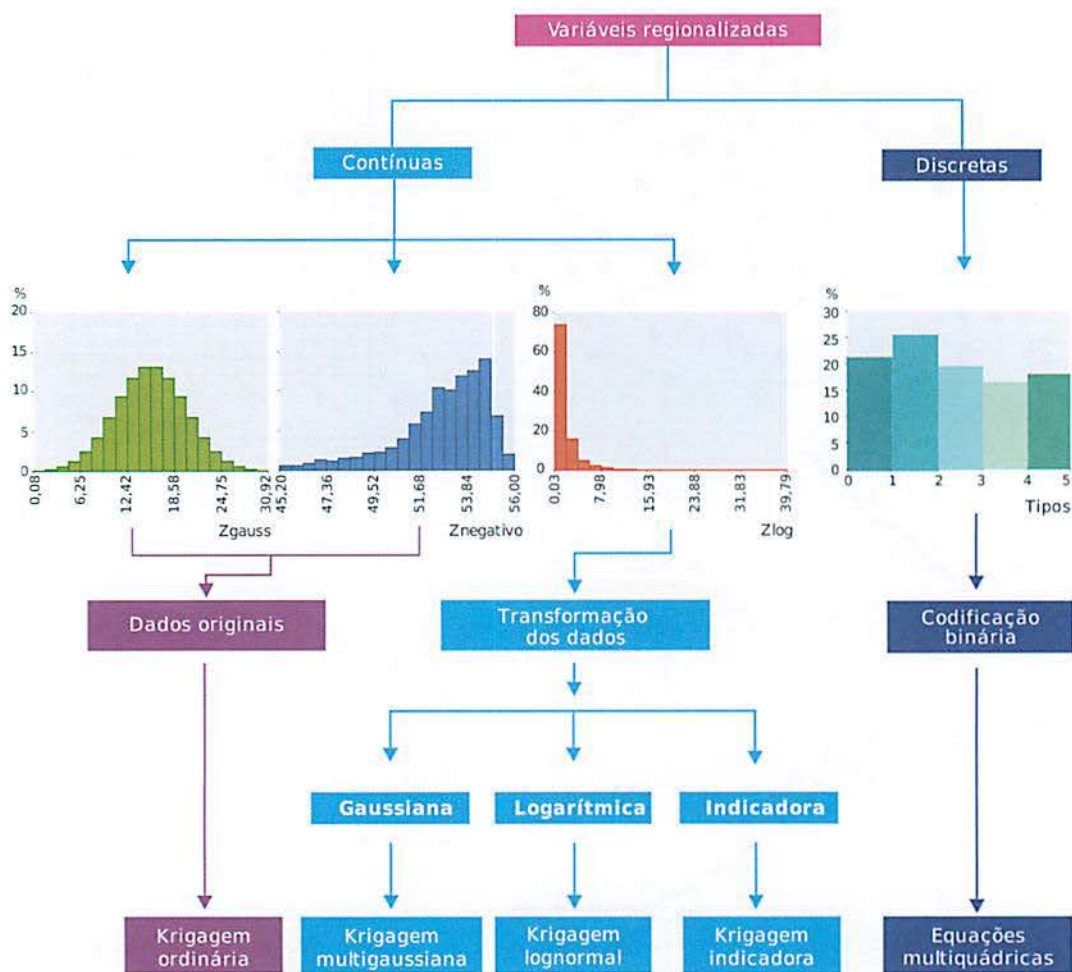


Fig. 4 Esquema ilustrando o processo de estimativa geoestatística ou interpolação de variáveis regionalizadas (seção 3.1)

sobre interpolação de variáveis categóricas baseada em equações multiquádricas, pois o cálculo de variogramas experimentais depende fortemente dos tipos e sua distribuição no espaço amostral.

O Cap. 4 trata das coestimativas geoestatísticas, como a cokrigagem ordinária, cokrigagem colocalizada e krigagem com deriva externa. Essas técnicas utilizam diferentes configurações de pontos de amostragem, que devem ser consideradas para fazer o melhor uso da informação disponível. A krigagem com deriva externa deveria ser abordada no Cap. 3, porém é tratada no Cap. 4 por compartilhar das mesmas amostras para o seu teste. Quando trataram da krigagem com deriva externa, no Cap. 4, os autores se depararam com dificuldades na obtenção do variograma residual. Desse modo, com base no cálculo do variograma da média com os dados de deriva externa, uma nova aproximação foi proposta para o cálculo do variograma residual. A síntese dos procedimentos de coestimativas geoestatísticas encontra-se na Fig. 5.

O Cap. 5 aborda a simulação estocástica, notadamente os métodos sequenciais, entre os quais são consideradas a simulação gaussiana sequencial, com opção tanto pela krigagem simples como pela ordinária, e a simulação indicadora sequencial, para variáveis contínuas

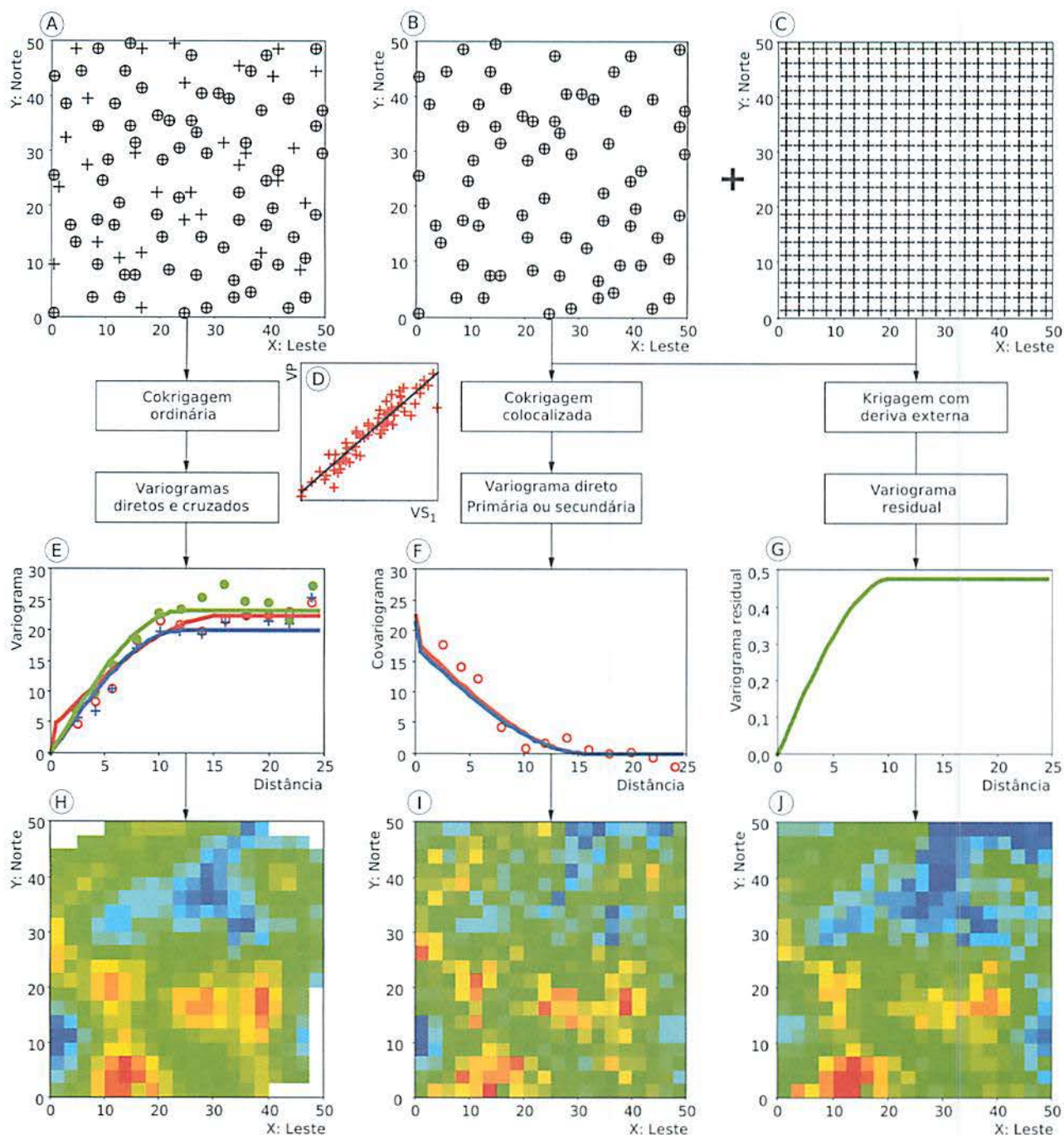


Fig. 5 Síntese dos métodos de coestimativas geoestatísticas: A) mapa de localização de pontos com heterotopia parcial; B) mapa de localização de pontos com isotopia; C) mapa de localização de pontos da variável secundária sobre os nós de uma malha regular; D) correlação entre a variável primária e a variável secundária; E) modelos de variogramas diretos (vermelho = variável primária; verde = variável secundária) e cruzado (vermelho); F) covariograma da variável primária (vermelho) e covariograma cruzado calculado por modelo de Markov 1 (azul); G) variograma residual; H) resultado da cokrigagem ordinária; I) resultado da cokrigagem colocalizada; J) resultado da krigagem com deriva externa (seção 4.3)

e discretas (Fig. 6). A opção pela krigagem ordinária para a simulação gaussiana sequencial foi incluída, pois a interpretação dos pesos da krigagem ordinária como probabilidades condicionais permite a determinação da função de distribuição acumulada condicional,

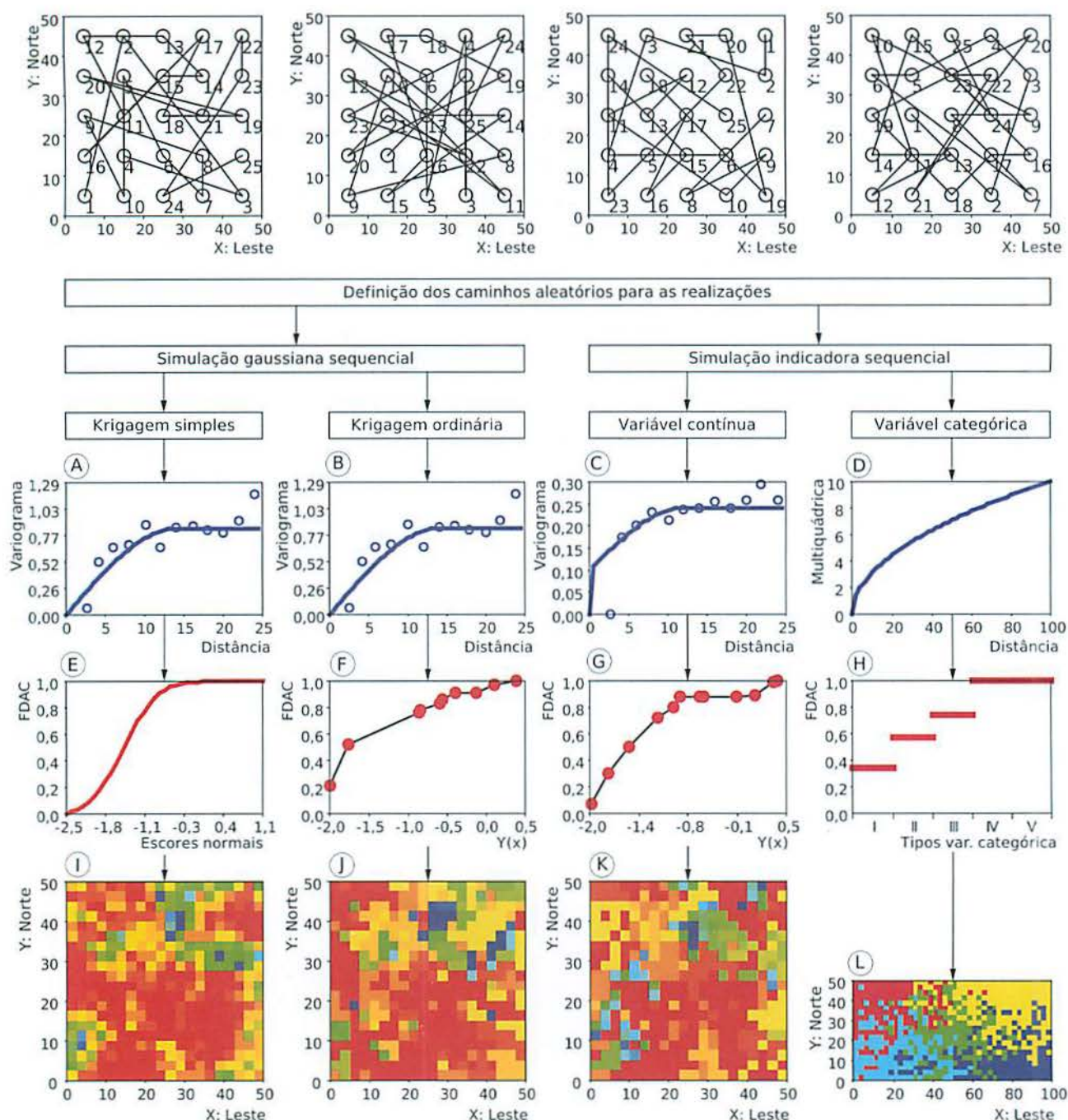


Fig. 6 Síntese dos métodos sequenciais de simulação estocástica. Definição dos caminhos aleatórios para as realizações (topo); variograma da variável transformada para escores normais (A e B); variograma indicadora da mediana (C); núcleo multiquádrico com constante nula (D); funções de distribuição acumulada condicional (E, F, G e H); resultado da simulação gaussiana sequencial – opção por krigagem simples (I); opção por krigagem ordinária (J); resultado da simulação indicadora sequencial – variável contínua (K) e variável categórica (L) (seção 5.4)

que pode ser amostrada por Monte Carlo. No caso de variáveis discretas, as realizações da simulação indicadora sequencial podem ser pós-processadas para determinação da imagem mais provável, assim como da zona de incerteza mapeada por meio da variância da proporção mais provável.

Também fazem parte da obra dois anexos: o primeiro, A, é uma introdução sobre os fundamentos de métodos matemáticos e estatísticos úteis para o entendimento das técnicas e conceitos empregados em Geoestatística; o segundo, B, apresenta as listagens dos dados utilizados nesta obra, que também podem ser obtidos no site do Laboratório de Informática Geológica do Departamento de Geologia Sedimentar e Ambiental da USP (<http://lig.igc.usp.br/geoestatistica/anexob>).

Todas as técnicas apresentadas são acompanhadas de cálculos mostrando os passos intermediários envolvidos para alcançar o resultado final. Assim, por exemplo, no caso da krigagem ordinária, para a estimativa de um ponto não amostrado, os pontos de dados vizinhos são listados e o sistema de equações de krigagem ordinária é montado e resolvido, dando origem aos ponderadores que são usados para a estimativa propriamente dita, bem como para o cálculo da incerteza associada. A apresentação de exemplos resolvidos passo a passo tem por objetivo mostrar ao leitor os algoritmos utilizados, permitir a aferição dos resultados apresentados e proporcionar um melhor entendimento das técnicas e conceitos apresentados.

Conceitos Básicos

1

O estudo geoestatístico tem como ponto de partida um conjunto de observações que constituem uma amostra. As observações, de natureza quantitativa ou qualitativa, são usadas para inferir as propriedades do fenômeno espacial em estudo. Na realidade, o fenômeno espacial desconhecido representa a população da qual uma amostra foi extraída. Nesse sentido, este capítulo tem a finalidade de introduzir os conceitos básicos empregados no estudo geoestatístico.

1.1 FENÔMENO ESPACIAL

A Geoestatística tem por objetivo a caracterização espacial de uma variável de interesse por meio do estudo de sua distribuição e variabilidade espaciais, com determinação das incertezas associadas.

O fenômeno espacial é o conjunto de todos os valores possíveis da variável de interesse, que define a distribuição e variabilidade espaciais dessa variável dentro de um dado domínio em 2D ou 3D. Representa, portanto, em termos estatísticos, a população que é o conjunto de todos os valores da qual uma amostra pode ser extraída. Para fins de ilustração de um fenômeno espacial, considerar uma variável de interesse que apresente a distribuição e variabilidade espaciais conforme apresentado na Fig. 1.1.

Dentro do domínio de 50 por 50 conhece-se o valor da variável em qualquer ponto. É preciso lembrar, porém, que, na prática, nada ou pouco se sabe sobre o fenômeno espacial a ser estudado. Assim, a Fig. 1.1 tem a finalidade didática de mostrar como se apresenta um fenômeno espacial em toda a sua extensão, conhecido como domínio de definição.

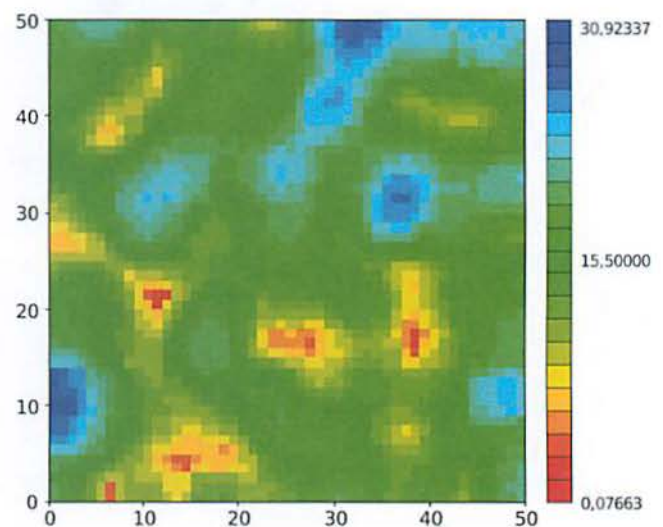


Fig. 1.1 Distribuição e variabilidade espaciais de uma variável de interesse caracterizando um fenômeno espacial em 2D (Arquivo completo 1, disponível em: <<http://lig.igc.usp.br/geoestatistica/anexob/download/Bell.txt>>)

Quando se decide estudar um fenômeno espacial do qual se tem pouco conhecimento sobre a variável de interesse, é necessária uma amostragem, pois é impossível analisar todo o conjunto de valores.

1.2 AMOSTRA E MÉTODOS DE AMOSTRAGEM

A amostra é um subconjunto de valores do fenômeno espacial que, se representativa, deve reproduzir a distribuição e variabilidade espaciais tanto em tamanho, isto é, número de pontos de dados, como em termos de distribuição dos pontos no domínio a ser estudado.

Qualquer estimativa baseada em pontos amostrais está, porém, sujeita a uma incerteza, e, nesse sentido, a metodologia geoestatística se destaca ao oferecer a incerteza associada à estimativa.

A amostragem é feita com base em um planejamento, que deve definir a coleta das unidades de amostragem de forma aleatória simples, aleatória estratificada ou sistemática.

1.2.1 Amostragem aleatória simples

Em Estatística, quando se fala em amostragem aleatória, a população constituída por N unidades é numerada sequencialmente e, assim, n unidades serão sorteadas sem reposição. A componente aleatória é, portanto, o número sequencial escolhido entre 1 e N . Nos estudos geoestatísticos, as observações são feitas em pontos de amostragem localizados dentro da região de estudo e, dessa maneira, a componente aleatória são as coordenadas geográficas a serem escolhidas casualmente.

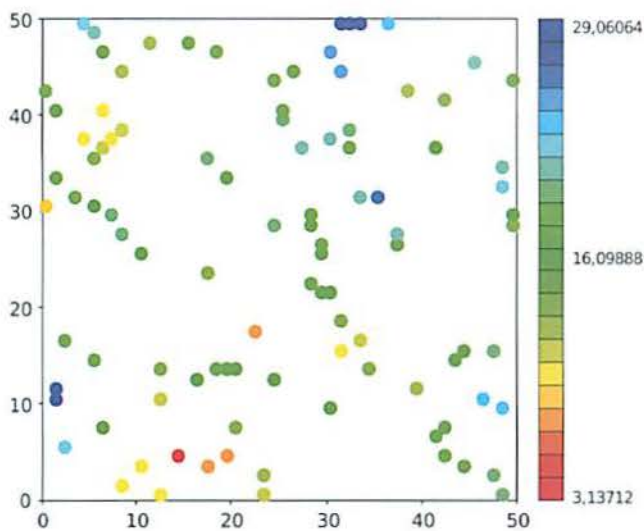


Fig. 1.2 Mapa de localização dos cem pontos de amostragem escolhidos aleatoriamente (Arquivo 1, Anexo B)

A Fig. 1.2 apresenta um mapa com cem pontos escolhidos aleatoriamente da população original (Fig. 1.1).

1.2.2 Amostragem aleatória estratificada

A amostragem aleatória estratificada é feita em estratos. Isso significa subdividir a região em estudo em células de dimensões fixas nas direções leste-oeste e norte-sul. Dentro de cada célula, as coordenadas geográficas de um ponto são escolhidas aleatoriamente e o ponto é selecionado. Assim, ao final desse processo, o número de unidades selecionadas será igual ao número de células. Para o exemplo da Fig. 1.1, a região de estudo foi subdividida em cem células de dimensões 5×5 e, dentro de cada célula, foi escolhido um ponto, resultando no mapa de localização da Fig. 1.3.

1.2.3 Amostragem sistemática

A amostragem sistemática é feita sobre os nós de uma malha regular definida com base em uma origem escolhida aleatoriamente. Teoricamente, a componente aleatória seria dada

pela escolha do ponto de origem, mas isso não é o que ocorre na prática, pois a malha regular é definida inicialmente pelo responsável pela amostragem para otimizar a coleta das unidades dentro da região de estudo. A amostragem sistemática em uma malha regular de 10×10 para o fenômeno espacial mostrado na Fig. 1.1 resulta no mapa de localização de pontos mostrado na Fig. 1.4.

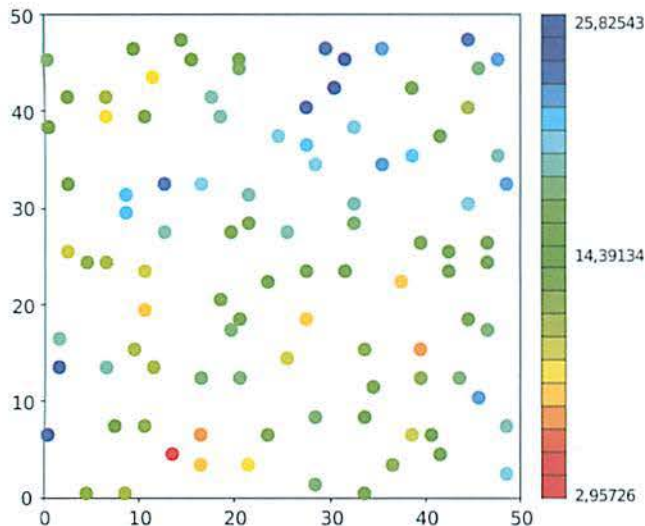


Fig. 1.3 Mapa de localização dos cem pontos da amostragem aleatória estratificada (Arquivo 2, Anexo B)

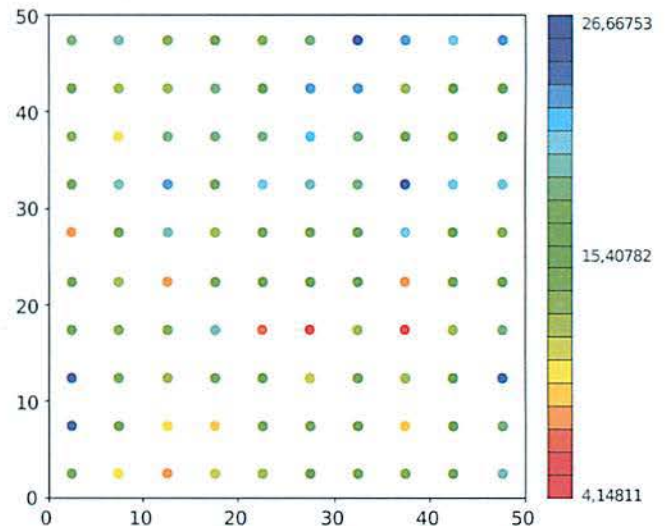


Fig. 1.4 Mapa de localização dos cem pontos da amostragem sistemática (Arquivo 3, Anexo B)

1.2.4 Considerações sobre os métodos de amostragem

Comparando-se os três métodos, verifica-se que a amostragem aleatória simples é a que oferece o pior resultado, haja vista áreas com pontos agrupados e áreas não amostradas; a amostragem aleatória estratificada é melhor que a anterior, mas ainda tem problemas na distribuição espacial dos pontos de amostragem; a amostragem sistemática é, sem dúvida, a que oferece o melhor resultado. Entretanto, nem sempre ela é possível, pois depende de uma série de fatores, tais como: acesso, acidentes geográficos (rios, lagos, topografia), vegetação etc.

Muitas vezes, a amostragem é feita ao longo de estradas, picadas e, portanto, resulta em uma distribuição semirregular. Independentemente, porém, do método de amostragem, a Geoestatística tem por objetivo extrair o máximo da informação disponível na amostra coletada.

1.3 INFERÊNCIA ESPACIAL

O processo de reprodução das características do fenômeno espacial baseado em pontos amostrais é denominado interpolação ou estimativa. A interpolação ou estimativa de um ponto não amostrado é feita por meio do ajuste de funções matemáticas locais (pontos mais próximos ao ponto não amostrado) ou globais (todos os pontos amostrais).

É preciso ressaltar que a interpolação ou estimativa em pontos não amostrados é sempre necessária, pois a amostragem nunca é feita em pontos muito próximos entre si, por causa, por exemplo, da limitação econômica. Geralmente, os pontos não amostrados são interpolados ou estimados em uma grade regular 2D ou 3D. Assim, a quantificação de recursos minerais ou a avaliação de contaminante em solo deve ser feita com base em medidas sistemáticas, ou seja, em pontos distribuídos regularmente no domínio do fenômeno espacial em estudo.

A grade regular resultante desse processo poderá ser usada para inferir a distribuição e variabilidade espaciais do fenômeno espacial em estudo. A qualidade dessa inferência espacial vai depender do tamanho da amostra e da distribuição espacial dos pontos amostrais.

Supondo que existe uma relação espacial entre os valores “ n ” conhecidos, regularmente distribuídos ou não, $Z_1, Z_2, \dots, Z_n$, o valor Z^* a ser interpolado para qualquer local será igual a: $Z^* = \sum p_i Z_i$.

A diferença fundamental entre os diversos métodos estimadores existentes baseia-se na maneira como os Z_i são escolhidos e os respectivos pesos p_i são calculados e aplicados durante o processo de estimativa. Uma divisão simples entre os métodos pode ser em modelos determinísticos e modelos estocásticos.

Os modelos determinísticos têm por base critérios puramente geométricos em que as distâncias são euclidianas e não fornecem medidas de incerteza como, por exemplo, o conhecido método do inverso do quadrado da distância (IQD).

Nos modelos estocásticos, os valores coletados são interpretados como provenientes de processos aleatórios e são capazes de quantificar a incerteza associada ao estimador. Os modelos geoestatísticos pertencem a essa categoria.

Para ilustrar o procedimento de inferência espacial, são consideradas três amostras, provenientes do fenômeno espacial exibido na Fig. 1.1 e obtidas pelos diferentes métodos de amostragem: aleatória simples, aleatória estratificada e sistemática.

Como método de estimativa é escolhido o ajuste pelas equações multiquádricas globais, por suas características de continuidade e suavidade da superfície resultante (Hardy, 1971, p. 1.907-1.908). A Fig. 1.5 ilustra, esquematicamente, todo o processo de inferência espacial, com base nas amostragens. Nesse caso, as amostras são de mesmo tamanho, mas com distribuições espaciais diferentes.

Os três métodos reproduzem, de modo geral, as características do fenômeno espacial mostrado na Fig. 1.1. O exame mais minucioso dos resultados mostra, porém, que a amostragem sistemática reproduz melhor a distribuição e variabilidade espaciais da variável de interesse.

Chegar a essa conclusão é possível à medida que se conheça o fenômeno espacial completo, mas isso não ocorre na prática e, então, deve-se usar o resultado da estimativa para fazer a inferência espacial, dentro da limitação da amostragem e do método de estimativa. Nesse caso, porém, não é possível analisar as incertezas associadas, pois o método das equações multiquádricas globais não permite o cálculo da incerteza.

Esse assunto será retomado no Cap. 3.

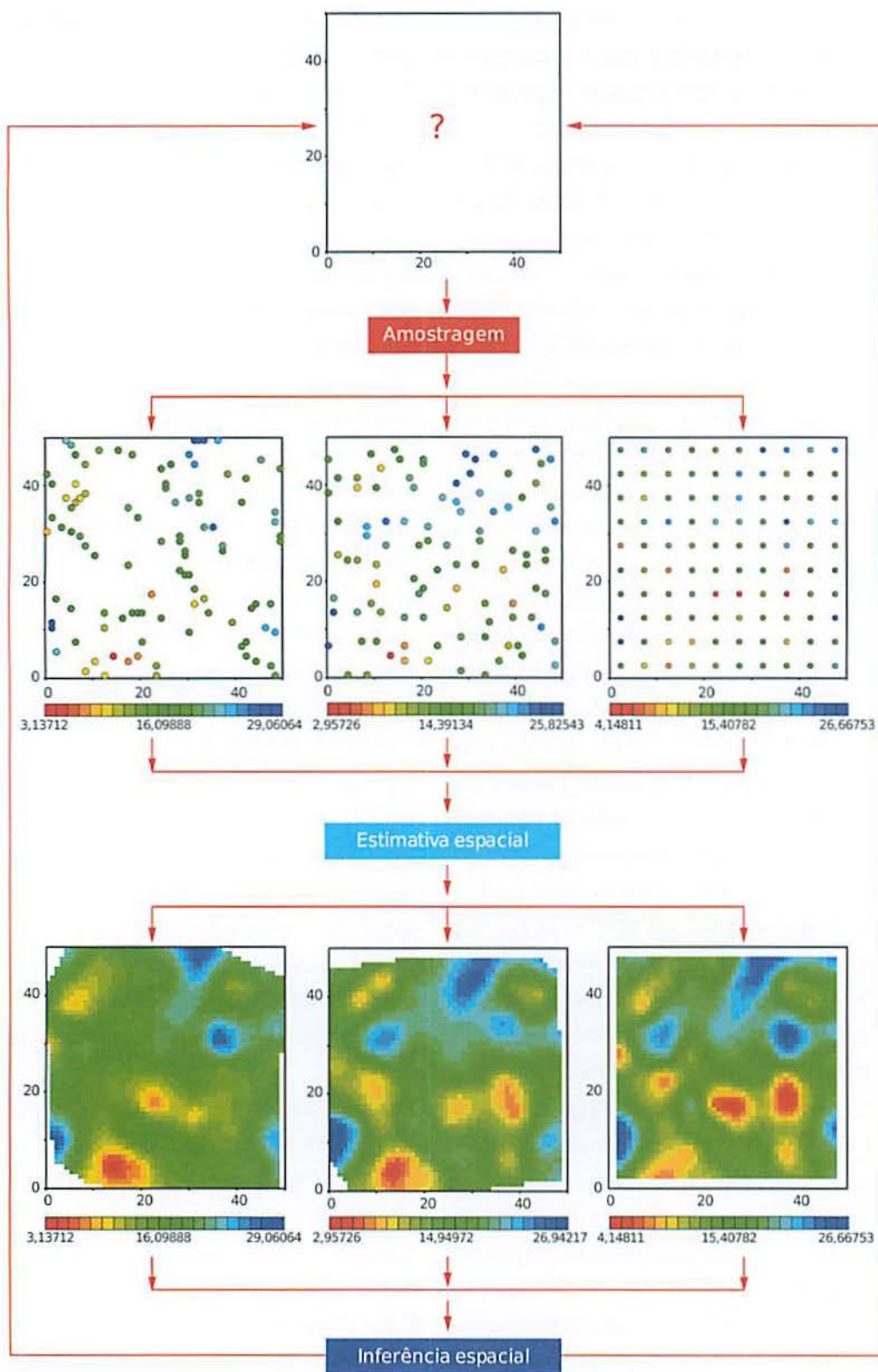


Fig. 1.5 Esquema mostrando o processo de inferência do fenômeno espacial com base na amostragem

1.4 VARIÁVEIS ALEATÓRIA E REGIONALIZADA

Na jogada de um dado, o resultado 1, 2, 3, 4, 5 ou 6 tem a mesma probabilidade de ocorrência, e o resultado atual não depende do anterior. Segundo esse exemplo, o processo de lançamento de dados pode ser repetido indefinidamente (condição A), e os resultados são independentes de lançamentos anteriores (condição B).

Nas Ciências da Terra, porém, quando se estudam teores de elementos metálicos em solos, porosidade e permeabilidade de rochas, características geotécnicas de maciços rochosos, concentração de poluentes em uma pluma de contaminação etc., ao se retirar uma amostra num determinado ponto, o teor da referida amostra é um valor único, fisicamente determinado, sendo impossível a repetição desse experimento. Se fosse retirada uma amostra em um ponto muito próximo, seria possível dizer que a condição A estaria satisfeita, porém, nesse caso, não se estaria respeitando a condição B.

O mesmo ocorre ao se subdividir uma unidade amostral. Essas frações, quando analisadas, resultarão em valores diferentes, mesmo muito próximos dentro da precisão do método analítico que for utilizado. Evidentemente, esses valores estarão correlacionados entre si, se o fenômeno apresentar alguma correlação espacial. Com base nisso, pode-se definir uma variável regionalizada como qualquer função numérica com uma distribuição e variação espacial, mostrando uma continuidade aparente, mas cujas variações não podem ser previstas por uma função determinística (Blais; Carlier, 1968 apud Olea, 1975).

Para melhor entender essa definição de variável regionalizada, apresentamos um exemplo proveniente da técnica da análise de superfícies de tendência, que foi largamente utilizada na década de 1970, baseada no trabalho clássico de Harbaugh e Merriam (1968).

Em geral, o ajuste de um polinômio aos pontos de dados não é exato, pois há uma diferença entre o valor estimado e o observado, qualquer que seja o grau do polinômio. Essa diferença, conhecida como resíduo, é, na realidade, a componente aleatória da variável de interesse, enquanto o valor estimado, tal como calculado pelo polinômio, é denominado componente regional, que apresenta grande continuidade. O polinômio ajustado é a função determinística que não pode prever as variações locais da variável de interesse.

O formalismo geoestatístico é baseado no conceito da dependência espacial e no entendimento de que cada ponto no espaço não apresenta um único valor, mas sim uma distribuição de probabilidade de ocorrência de valores.

No ponto x , a propriedade $Z(x)$ é uma variável aleatória com média m , variância S^2 e uma função de distribuição acumulada. No espaço existem infinitos pontos $\{x_i, i = 1, 2, \dots\}$ em que os valores $\{z(x_i), i = 1, 2, \dots\}$ são realizações das funções aleatórias com suas distribuições de probabilidade. O conjunto de variáveis aleatórias constitui uma função aleatória ou um processo aleatório ou processo estocástico, e o conjunto de valores reais de $Z(x)$, que inclui a realização da função aleatória, é conhecido como variável regionalizada.

Esse conceito é bem diferente do tradicional, que considera cada observação pontual como o resultado independente de uma variável casual. Uma variável regionalizada é entendida, porém, como uma única realização de uma função casual, possuindo dependência espacial. Desse modo, o seu entendimento pode descrever melhor o padrão espacial do fenômeno em estudo.

1.4.1 Notação

Variáveis aleatórias são representadas por letras maiúsculas: X, Y, Z etc. Os valores específicos dessas variáveis são representados por letras minúsculas, seguidas por índices que correspondem às observações. Por exemplo, seja Y a variável aleatória representando os teores de sílica; assim, $y_1 = 44,66\%$ representa o valor de sílica medido para a amostra 1.

A notação de uma função aleatória segue a mesma sistemática adotada para variáveis aleatórias, ou seja, letras maiúsculas para designar a função aleatória e letras minúsculas para designar valores dessa função em pontos específicos. A principal diferença é que a letra que representa a função aleatória vem acompanhada de um argumento que indica a sua localização no espaço. Assim, pode-se ter uma função aleatória $Z(x)$ representando teores de sílica e o valor em um ponto específico $z(x_1) = 44,66\%$. Nesse caso, x_1 indica a localização do ponto amostral que forneceu o valor de 44,66% de sílica. Na realidade, x é um vetor localização em uma, duas ou três dimensões (Fig. 1.6). Na Fig. 1.6A, o vetor aponta para a amostra $z(20)$. Da mesma forma, na Fig. 1.6B, o vetor aponta para a amostra $z(40,80)$, e na Fig. 1.6C, para a amostra $z(40,80,15)$.

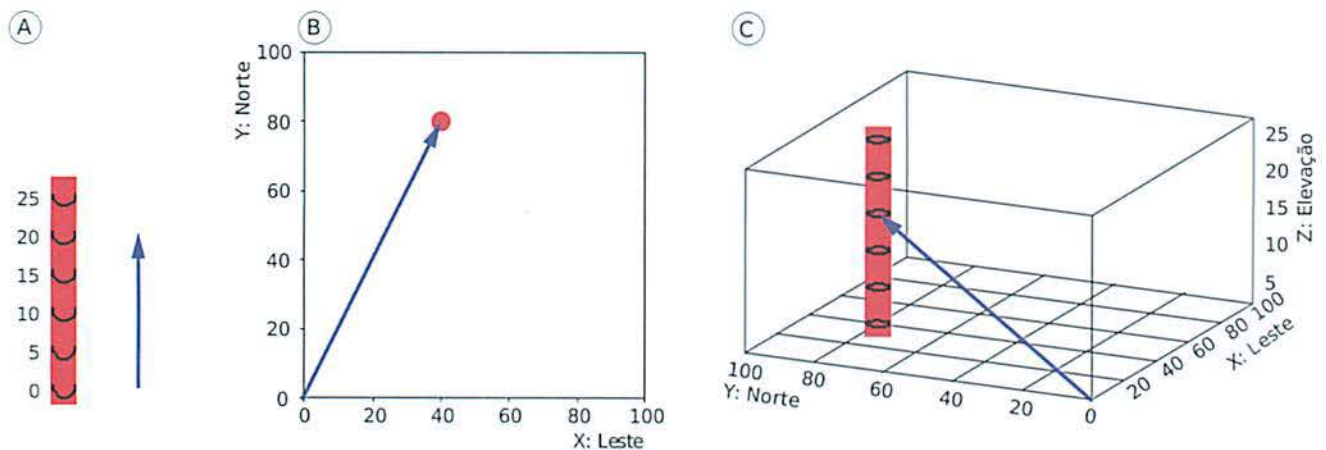


Fig. 1.6 O vetor localização para pontos em: A) uma; B) duas e C) três dimensões

1.4.2 Natureza das variáveis aleatórias e regionalizadas

As variáveis aleatórias podem ser subdivididas em contínuas e discretas, conforme proposta de Stevens (1946). A Fig. 1.7 mostra essa subdivisão, com exemplos das variáveis geológicas mais comuns.

As variáveis contínuas podem ser medidas pelas escalas *relacional* e *intervalar*. Podem ser medidas, pela escala relacional, as seguintes variáveis: *teores*, *espessuras*, *recuperação*, *densidade aparente*, *dados de perfilagem geofísica* e *rock quality designation (RQD)*.

Teores são medidas de razões, sejam percentuais ou em partes por milhão, sendo essas equivalentes a gramas por tonelada.

Espessuras são medidas diretamente nos testemunhos de sondagem.

Dados de recuperação são obtidos pela razão entre a metragem de testemunho recuperada sobre a espessura perfurada.

Variáveis aleatórias		Escala nominal		Escala ordinal
	Discretas	Litologia	Cor da rocha	Alteração
		Estrutura	Textura	Fraturamento
	Contínuas	Teores	Densidade	Temperatura
		Espessuras	Perf. Geof.	
		Recuperação	RQD	
		Escala relacional		Esc. intervalar

Fig. 1.7 Subdivisão das variáveis aleatórias (Stevens, 1946), com exemplos de variáveis geológicas

Densidade aparente é obtida pela razão entre a massa de minério (em base seca) e o volume ocupado por essa massa.

A perfilagem geofísica é realizada com o objetivo de obter indicação da litologia, mineralogia e da mineralização, por meio de medidas da intensidade de raios gama, resistividade e suscetibilidade magnética (Peters, 1978, p. 454-455).

A medida de RQD é obtida pela razão percentual entre a soma de segmentos do testemunho maiores que 10 cm dividida pela metragem perfurada (Deere et al., 1967).

Na escala intervalar, são encontradas medidas de temperatura feitas em prospecção geotérmica ou em determinação do grau geotérmico.

As variáveis discretas são medidas pelas escalas nominal e ordinal. Na escala nominal, as variáveis são *litologia*, *estrutura*, *cor da rocha* e *textura*. Cada uma dessas variáveis apresenta um número de tipos, dependendo da litologia. Esses tipos se encontram em tabelas proporcionadas por Blanchet e Godwin (1972, p. 799-806).

Graus de alteração e de fraturamento podem ser classificados na escala ordinal. Embora o grau de alteração possa ser usado para descrever o tipo de depósito, seja em termos de alteração hidrotermal e/ou intempérica, esse parâmetro é geralmente utilizado para estudo geomecânico do maciço.

Essa subdivisão de variáveis aleatórias persiste quando se trata também de variáveis regionalizadas. Embora a Geoestatística tivesse se desenvolvido com o foco inicial em variáveis quantitativas, as variáveis qualitativas são passíveis de tratamento e análise conforme a mesma metodologia, graças ao trabalho pioneiro de Journel (1983). Assim, torna-se possível a estimativa geoestatística de variáveis categóricas com determinação do tipo mais provável, bem como da incerteza associada, como será visto no Cap. 3.

1.5 DESAGRUPAMENTO

A pesquisa de recursos minerais requer que a amostragem seja planejada para fornecer as informações necessárias sobre uma malha perfeitamente regular. Entretanto, é muito

difícil que a amostragem reflita o plano inicial, por causa de vários motivos: dificuldade de acesso, áreas de proteção ambiental, rios, lagos, topografia etc. Além disso, muitas vezes, e especialmente na pesquisa mineral, uma região anômala, contendo valores extremos, pode ser detalhada (Olea, 2007, p. 453-454), resultando em uma amostragem semirregular com agrupamentos de pontos. A consequência disso é que uma amostragem planejada inicialmente para ser regular passa a apresentar agrupamentos de pontos em determinadas regiões. Segundo Pyrcz e Deutsch (2003, p. 1), a amostragem preferencial em áreas interessantes é intencional e facilitada por intuição geológica, por dados análogos ou por amostras prévias. De acordo com esses autores, a prática de coleta de amostras agrupadas ou espacialmente enviesadas é encorajada por limitações de ordem técnica e econômica, tais como objetivos de produção futura, acessibilidade e custos de laboratório. Muitas vezes, segundo eles, objetivos de produção futura podem encorajar amostragem agrupada ou espacialmente enviesada, e é comum iniciar a lavra em regiões de alto teor.

Agrupamentos de pontos amostrais acabam influenciando toda a área de interesse, na qual, por exemplo, teores mais elevados obtidos nas regiões anômalas acabam se propagando em torno da vizinhança dessas regiões. Em termos estatísticos, além do problema de agrupamento de pontos amostrais, há também o enviesamento da distribuição de frequências da variável de interesse. Por exemplo: regiões anômalas fornecem teores maiores e, assim, tanto a média como a mediana tendem para teores maiores quando, na verdade, deveriam ser menores para refletir a realidade.

Todos os problemas decorrentes de amostragem apresentando agrupamentos de pontos e vieses para teores altos devem ser corrigidos para que os tratamentos posteriores não sofram influência desses desvios. O objetivo é, portanto, obter uma distribuição representativa dos dados amostrais (Deutsch, 1989, p. 325).

Os procedimentos de desagrupamento atribuem pesos aos dados disponíveis conforme a sua configuração. Assim, pontos em regiões esparsamente amostradas têm pesos maiores, enquanto pontos em regiões com agrupamentos recebem pesos menores (Leuangthong; Khan; Deutsch, 2008, p. 21).

Existem quatro métodos de desagrupamento de dados bem estabelecidos (Leuangthong; Khan; Deutsch, 2008, p. 35): *poligonal, por células, krigagem e inverso da distância*. Desses quatro, apenas os métodos de desagrupamento poligonal e por células serão considerados aqui.

Para ilustrar os procedimentos de desagrupamento, considerar uma amostra com cem pontos de dados (Arquivo 4, Anexo B), conforme mapa de localização (Fig. 1.8). A amostra foi enviesada com o propósito de produzir agrupamentos em regiões de altos teores.

Esses agrupamentos de pontos em regiões de altos teores certamente irão influenciar as estatísticas

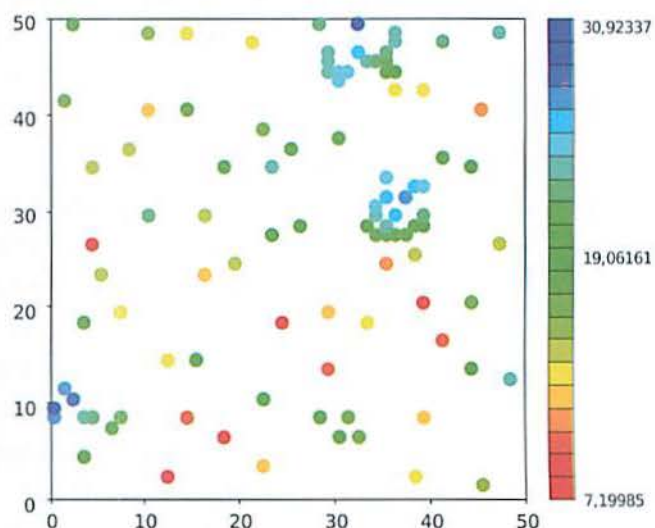


Fig. 1.8 Mapa de localização de pontos com amostragens preferenciais em regiões de altos teores (Arquivo 4, Anexo B)

globais. As distribuições de frequências simples e acumulada, bem como as estatísticas amostrais, podem ser vistas na Fig. 1.9.

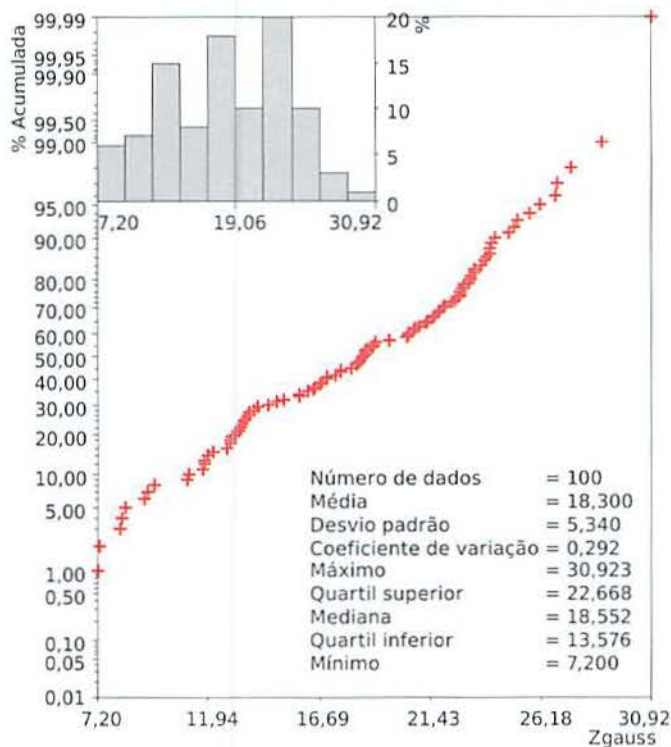


Fig. 1.9 Estatísticas amostrais para o Arquivo 4, Anexo B

Assim, na presença de agrupamentos preferenciais de pontos, as estatísticas globais devem ser calculadas aplicando-se os pesos de desagrupamento, conforme os algoritmos descritos a seguir.

1.5.1 Desagrupamento poligonal

Segundo Pyrcz e Deutsch (2003, p. 2), o método de desagrupamento poligonal é comumente aplicado em outras áreas das Ciências, como a Hidrologia. Esse método é baseado na construção de polígonos de influência em torno dos pontos de dados. Assim, tem-se um polígono para cada ponto. O peso de desagrupamento para o i -ésimo ponto de dado é igual à área do polígono dividida pela área total de interesse (Pyrcz; Deutsch, 2003, p. 3):

$$w_i = \frac{\text{área}_i}{\sum_{j=1}^n \text{área}_j}$$

Após a aplicação do desagrupamento poligonal, pontos de dados agrupados receberão pesos menores associados a pequenos polígonos de influência, enquanto pontos associados a grandes polígonos de influência terão pesos maiores como representativos de grandes áreas (Isaaks; Srivastava, 1989, p. 239).

Para a determinação dos pesos de desagrupamento usando esse método, faz-se a subdivisão da área de interesse em polígonos de influência, que pode ser obtida por meio do Diagrama de Voronoi (Hayes; Koch, 1984; Tipper, 1991; entre outros). Algoritmos para dados 2D são bem estabelecidos e funcionam muito bem. Contudo, para dados 3D, o equivalente ao Diagrama de Voronoi é computacionalmente muito complicado e, por isso, a solução mais simples é usar o método dos pontos mais próximos, no qual o valor de um ponto não amostrado é igual ao do ponto mais próximo, como sugerido por Pyrcz e Deutsch (2003, p. 3).

Outro problema associado ao método está relacionado ao limite na fronteira dos pontos de dados, no qual dados na periferia podem abrir os polígonos até um limite além da influência dos pontos amostrais, tradicionalmente calculados como a meia distância entre os pontos vizinhos próximos. Esses autores afirmam que a área associada a pontos periféricos é muito sensível à definição da borda. A Fig. 1.10 ilustra o problema da área dos pontos da periferia na área de interesse, na qual os polígonos estão abertos.

Uma possível solução proposta por Popoff (1966 apud Yamamoto, 2001b, p. 117) é a extrapolação da área de interesse pela aplicação da regra dos pontos mais próximos aos pontos da periferia da área de interesse (Fig. 1.11).

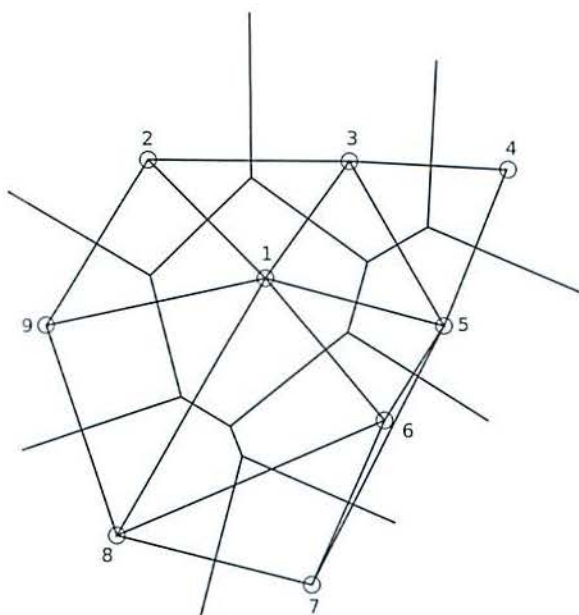


Fig. 1.10 Diagrama de Voronoi para um conjunto de pontos de dados de Popoff (1966 apud Yamamoto, 2001b, p. 117)

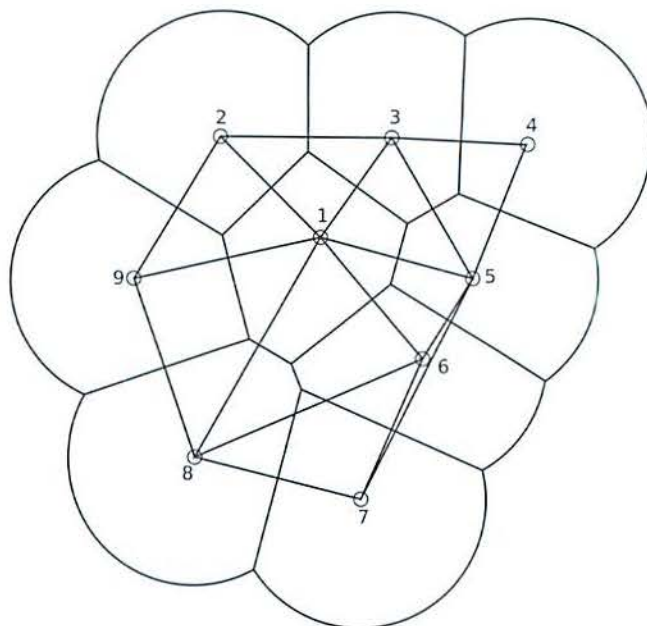


Fig. 1.11 Diagrama de Voronoi com aplicação da regra dos pontos mais próximos para determinação das áreas associadas aos pontos da periferia da área de interesse, segundo Popoff (1966 apud Yamamoto, 2001b, p. 117)

Considerando que os dados são confiáveis dentro do domínio de amostragem, bem como para evitar quaisquer extrapolações, pode-se simplesmente usar o limite definido pela fronteira convexa como a borda e, assim, determinar as áreas dos polígonos associados aos pontos da periferia.

Exemplo de aplicação do desagrupamento poligonal

Apesar de haver um algoritmo para o cálculo do Diagrama de Voronoi, optou-se por usar o método do ponto mais próximo (Yamamoto, 2001b, p. 91), que dá aproximadamente o mesmo resultado. Isso se justifica pela facilidade desse método em relação ao algoritmo de Voronoi, especialmente para casos em dados 3D, nos quais a geometria computacional envolvida é extremamente complicada. Para o desagrupamento poligonal usando essa aproximação, uma malha regular é interpolada, de modo que os seus nós recebem o valor do vizinho mais próximo. Ao final do processo, cada ponto de dado terá sua influência desenhada nos limites de um polígono convexo. A precisão dessa aproximação dependerá das dimensões das células nos eixos X e Y.

A amostra do Arquivo 4, Anexo B, foi submetida ao desagrupamento poligonal, conforme os polígonos desenhados na Fig. 1.12. Essa figura representa o resultado da interpolação

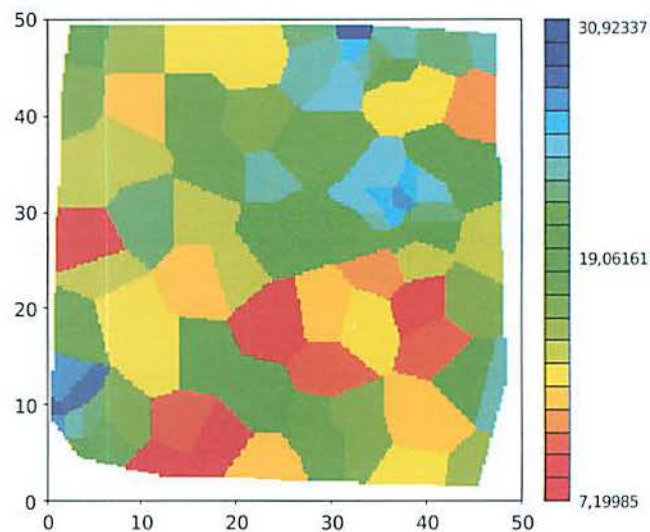


Fig. 1.12 Polígonos de influência para cálculo dos pesos de desagrupamento (Arquivo 4, Anexo B)

de uma malha regular com abertura igual a 0,25 nos dois eixos. Como se pode observar, os limites dos polígonos de Voronoi são quase retos, por causa do tamanho da célula usado. Nesse tipo de aproximação, quanto menor a abertura da malha regular, mais próximo o resultado será do valor teórico que seria fornecido pelo Diagrama de Voronoi (Tab. 1.1).

TAB. 1.1 Estatísticas amostrais após o desagrupamento poligonal aproximado por meio da interpolação de uma malha regular pelo método do ponto mais próximo

$DX = DY$	$\bar{X} = E[Z(x)]$	$S = \sqrt{Var[Z(x)]}$	$CV = S/\bar{X}$
0,10	15,791	4,694	0,297
0,25	15,796	4,698	0,297
0,50	15,802	4,699	0,297
0,63	15,807	4,677	0,296
1,00	15,856	4,683	0,295
1,25	15,784	4,743	0,300
2,00	15,939	4,740	0,297
2,50	15,711	4,816	0,307
3,33	16,044	4,794	0,299
5,00	16,025	4,567	0,285
10,00	15,662	4,241	0,271

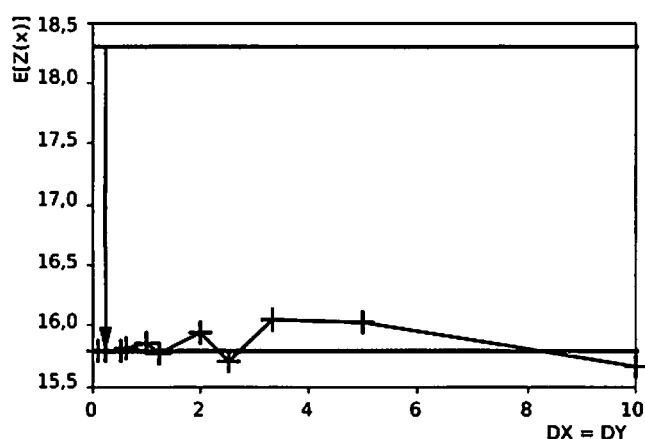


Fig. 1.13 Variação da média conforme as dimensões da malha regular e redução da média amostral pelo desagrupamento poligonal

Conforme a Tab. 1.1, a média obtida pelo desagrupamento poligonal tenderia a um valor muito próximo a 15,791. Essa aproximação dá resultados bons, basicamente, em uma abertura da malha regular $DX = DY = 1,00$. Como a ideia geral do desagrupamento é eliminar a forte influência dos agrupamentos de pontos em torno dos valores altos, a média global representativa deve ser a mais baixa possível após aplicação dos pesos de desagrupamento. Nesse caso, igual a 15,791, que é muito menor que a média amostral, igual a 18,300 (Fig. 1.9). A Fig. 1.13 mostra graficamente a variação da média conforme as dimensões da malha regular.

1.5.2 Desagrupamento por células

O método de desagrupamento por células é o mais comumente empregado em Geoestatística, pois não depende de extrapolações nos pontos da periferia e, por isso, é considerado mais robusto que o desagrupamento poligonal (Pyrz; Deutsch, 2003, p. 3). O método de desagrupamento por células está disponível na biblioteca de rotinas geoestatísticas do GSLib (Deutsch; Journel, 1992, p. 207-209). Conforme esse método, a área total é dividida

em regiões retangulares chamadas células (Isaaks; Srivastava, 1989, p. 241). Segundo esses autores, cada elemento da amostra recebe um peso inversamente proporcional ao número de elementos da amostra que existe dentro da mesma célula. O peso de desagrupamento pode ser calculado como (Leuangthong; Khan; Deutsch, 2008, p. 35):

$$w_i = \frac{1}{n_j \cdot J} \quad (1.1)$$

em que n_j é o número de elementos dentro da j -ésima célula e J é o número de células ocupadas por um ou mais elementos.

Assim, elementos dentro de agrupamentos receberão pesos menores, pois as células nas quais eles estão também irão conter outros elementos da amostra (Isaaks; Srivastava, 1989, p. 241), enquanto elementos distribuídos esparsamente receberão pesos maiores (Deutsch; Journal, 1992, p. 207).

A eficiência desse método depende da escolha correta do tamanho da célula, pois o peso de desagrupamento irá variar conforme o tamanho da célula. Assim, é comum o procedimento de calcular a média desagrupada para vários tamanhos de células e depois escolher a média ótima (Deutsch, 1989, p. 327).

Exemplo de aplicação do desagrupamento por células

O desagrupamento por células foi aplicado ao conjunto de dados do Arquivo 4, Anexo B, conforme os resultados da Tab. 1.2 e da Fig. 1.14.

TAB. 1.2 Estatísticas amostrais após desagrupamento por células

$DX = DY$	$\bar{X} = E[Z(x)]$	$S = \sqrt{Var[Z(x)]}$	$CV = S/\bar{X}$	J
0,10	18,300	5,340	0,292	100
0,50	18,300	5,340	0,292	100
1,00	18,300	5,340	0,292	100
2,00	17,709	5,344	0,302	88
2,50	17,339	5,382	0,310	80
5,00	16,719	5,110	0,306	62
5,55	16,066	5,009	0,312	58
6,25	15,894	4,786	0,301	59
8,33	15,987	4,839	0,303	36
10,00	16,775	4,894	0,292	25
25,00	17,055	5,318	0,312	4

De acordo com a Tab. 1.2, para células muito pequenas, nas quais se localizam apenas um ponto de dado, o desagrupamento por células não é efetivo. À medida que se aumenta o tamanho das células, um maior número de pontos é encontrado dentro delas, reduzindo, assim, a influência dos pontos agrupados, conforme a Eq. 1.1. Essa redução da média global ocorre até uma determinada dimensão da célula, então, a partir do valor ótimo e com o aumento do tamanho da célula, a média volta a subir, como mostra a Fig. 1.14.



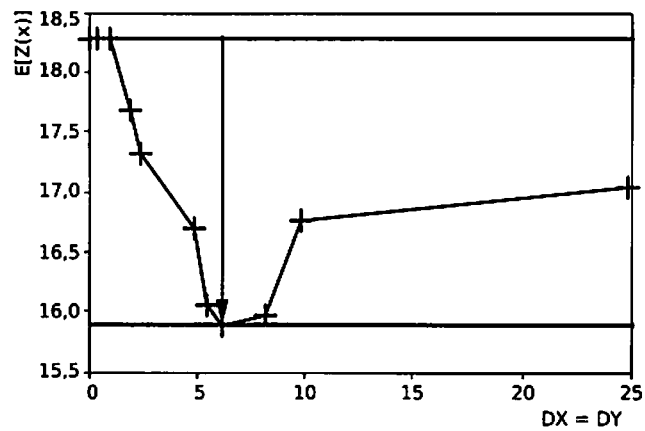


Fig. 1.14 Variação da média conforme as dimensões da célula e redução da média amostral pelo desagrupamento por células

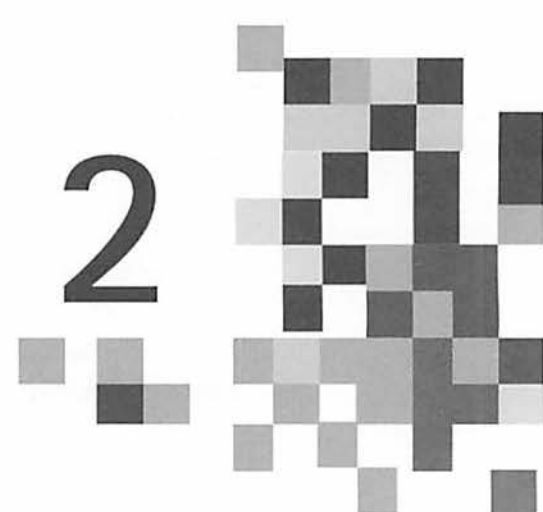
1.5.3 Considerações sobre os métodos de desagrupamento

Foram apresentados dois métodos de desagrupamento de dados: poligonal e por células. Os dois são efetivos tanto para dados 2D como para 3D. A aproximação do desagrupamento poligonal por meio da interpolação de uma malha regular pelo vizinho mais próximo é viável e simplifica bastante o procedimento de cálculo do Diagrama de Voronoi em 3D. Conforme essa aproximação, a malha regular deve ser interpolada com a menor dimensão possível para que o resultado obtido se aproxime do valor teórico encontrado com o Diagrama de Voronoi.



Cálculo e Modelagem de Variogramas Experimentais

2



Como definir e prever o comportamento espacial de uma variável regionalizada $\{Z(x_i), i = 1, n\}$ coletada em n pontos distribuídos em uma determinada região? Pretende-se responder a essa questão neste e no próximo capítulo por meio da metodologia geoestatística, com exemplos ilustrando aplicações.

Para entender a variação espacial do processo aleatório subjacente, deve-se levar em consideração a possibilidade de que o valor de cada ponto no espaço está relacionado, de algum modo, com valores obtidos de pontos situados a certa distância, sendo razoável supor que a influência é tanto maior quanto menor for a distância entre os pontos, conforme interpretação de Soares (2006, p. 18). Isso significa que a inferência da continuidade espacial de uma variável regionalizada pode ser feita com valores amostrais tendo como base a estatística de dois pontos. Aplicando-se as definições da função covariância e função variograma, verifica-se que elas dependem apenas de dois pontos x_1 e x_2 , situados a uma distância $h = x_1 - x_2$, então cada par de pontos é considerado uma realização diferente, o que torna possível a inferência estatística dessas funções (Journel; Huijbregts, 1978, p. 32). Para determinação do modelo de correlação espacial da variável regionalizada, calcula-se experimentalmente essa correlação usando os pontos amostrais e, em seguida, ajusta-se um modelo teórico. Esse modelo teórico permite determinar o valor da correlação espacial para qualquer distância dentro do espaço amostrado. Neste capítulo será apresentado como se calcula o modelo de correlação espacial, que é a ferramenta básica da Geoestatística para estimativas e simulações estocásticas.

2.1 ESTATÍSTICAS ESPACIAIS

Segundo Soares (2006, p. 18), o conjunto de variáveis aleatórias $\{Z(x_i), i = 1, n\}$ correlacionadas entre si constitui uma função aleatória cuja amostragem fornece uma realização $z(x_i)$. Por isso, de acordo com ele, com uma única realização torna-se impossível determinar as estatísticas no ponto x_i dessa função, tais como média e variância. Para ele, a solução consiste em assumir diversos graus de estacionaridade da função aleatória, como, por exemplo, admitindo que as variáveis aleatórias tenham a mesma média:

$$E[Z(x_1)] = E[Z(x_2)] = \dots = E[Z(x_n)] = E[Z(x)] = m$$

Desse modo, a média m passa a ser independente da localização e obtida como média aritmética das realizações das variáveis aleatórias (Soares, 2006, p. 18):

$$m = E[Z(x)] = \frac{1}{n} \sum_{i=1}^n Z(x_i)$$

Julgar, porém, que essa hipótese esteja correta significa supor que a média das amostras seja representativa da área estudada, isto é, que os valores são homogêneos (Soares, 2006, p. 18). A homogeneidade espacial raramente ocorre, sendo necessária a verificação da distribuição e variabilidade espaciais da função aleatória, como será visto neste capítulo.

A variância associada à média é calculada como:

$$\text{Var}[Z(x)] = E\{[Z(x) - m]^2\}$$

A hipótese de estacionaridade de 2ª ordem, além de definir que a esperança matemática, $E[Z(x)]$, existe e não depende do suporte x , define também que a correlação entre duas variáveis aleatórias depende somente da distância espacial, h , que as separa e é independente da sua localização (Journel; Huijbregts, 1978, p. 32).

Em Estatística, a covariância é uma medida da relação mútua entre duas variáveis aleatórias distintas, por exemplo, X e Y . Em Geoestatística, a covariância mede a relação entre valores da mesma variável, obtidos em pontos separados por uma distância h , conforme uma determinada direção. Isso significa que, ao alterar a direção, a covariância também pode se alterar e, nesse caso, há indicação de presença de fenômeno espacial anisotrópico (Fig. 2.1B).

Existem casos em que a covariância é a mesma em qualquer direção e, por isso, o fenômeno espacial é isotrópico (Fig. 2.1A). Assim, para detectar se o fenômeno espacial apresenta anisotropia ou não, a covariância é calculada para várias direções. Geralmente, quando o fenômeno em estudo está distribuído em 2D, calculam-se as covariâncias em quatro direções horizontais: 0°, 45°, 90° e 135°.

Para fenômenos espaciais 3D, além das direções horizontais, calculam-se as covariâncias para a direção vertical ou inclinada, conforme a estrutura geológica do corpo em profundidade.

A covariância de uma variável regionalizada para pontos separados por uma distância h pode ser calculada como:

$$C(h) = E\{[Z(x+h) - m][Z(x) - m]\}$$

em que h representa um vetor entre dois pontos x_1 e x_2 no espaço tridimensional.

É fácil verificar que a covariância para distância nula ($h = 0$) é igual à variância da variável regionalizada $Z(x)$.

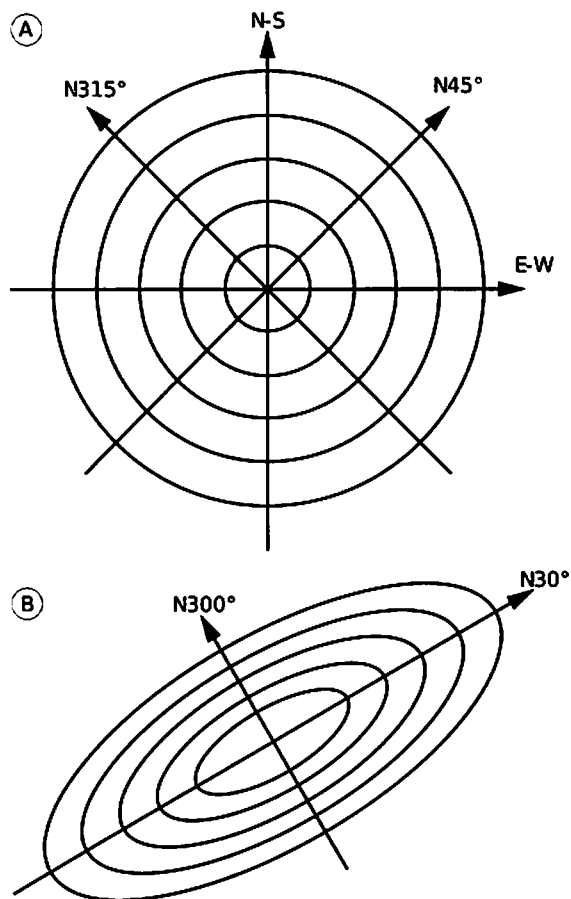


Fig. 2.1 Esquema ilustrando fenômenos espaciais: A) isotrópico e B) anisotrópico

A função variograma é definida como a variância do incremento $[Z(x+h) - Z(x)]$:

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

A hipótese de estacionaridade de 2ª ordem assume a existência da variância e, portanto, de uma variância *a priori* finita (Journel; Huijbregts, 1978, p. 33). Existem, porém, fenômenos físicos e, conseqüentemente, variáveis regionalizadas com uma capacidade infinita de dispersão, nos quais não se pode definir, *a priori*, nem a covariância nem a variância, mas se pode determinar um variograma (Journel; Huijbregts, 1978, p. 33).

Adota-se a hipótese intrínseca, que não requer a existência de uma média constante e variância finita para a função aleatória $Z(x)$, mas apenas que os incrementos da função aleatória $[Z(x+h) - Z(x)]$ sejam estacionários de 2ª ordem (Goovaerts, 1997, p. 71). Na realidade, segundo esse autor, a estacionaridade é uma propriedade do modelo de função aleatória necessária para a inferência estatística. Para todos os vetores h , o incremento $[Z(x+h) - Z(x)]$ tem uma variância finita, a qual não depende do suporte x (Journel; Huijbregts, 1978, p. 33):

$$\text{Var} [Z(x+h) - Z(x)] = E \{ [Z(x+h) - Z(x)]^2 \} = 2\gamma(h)$$

Com relação ao termo variograma, há uma confusão terminológica na literatura geoestatística. Alguns autores preferem essa terminologia, como Wackernagel (2003), por exemplo; outros, a denominação semivariograma, a exemplo de Journel e Huijbregts (1978). Segundo Bachmaier e Backes (2008), a confusão a respeito do prefixo *semi* surgiu porque Matheron (1965) tinha em mente a variância das diferenças $[Z(x+h) - Z(x)]$, mas o valor desejado, na prática, era a metade dessa diferença, que fornece a variância da diferença de pares de pontos separados por h . Na realidade, o prefixo *semi* se deve à divisão da média das diferenças ao quadrado por dois:

$$\begin{aligned} \gamma(h) &= \frac{1}{2} E \{ [Z(x+h) - Z(x)]^2 \} \\ &= \frac{1}{2n} \sum_{i=1}^n [Z(x+h) - Z(x)]^2 \end{aligned} \quad (2.1)$$

Portanto, $2\gamma(h)$ é chamado de variograma e $\frac{1}{2}\gamma(h)$, de semivariograma, por causa da divisão por dois. Muitos pesquisadores simplesmente chamam o semivariograma de variograma, mas, nos cálculos, sempre consideram a divisão por dois.

Pensava-se que a divisão por dois era empírica, mas Journel (1989, p. 6-7) demonstrou sua origem por meio de uma interpretação geométrica dos pares de pontos em um diagrama de dispersão (Fig. 2.2).

Nesse diagrama de dispersão, um par de pontos de coordenadas $(Z(x+h), Z(x))$ é representado. Esse ponto

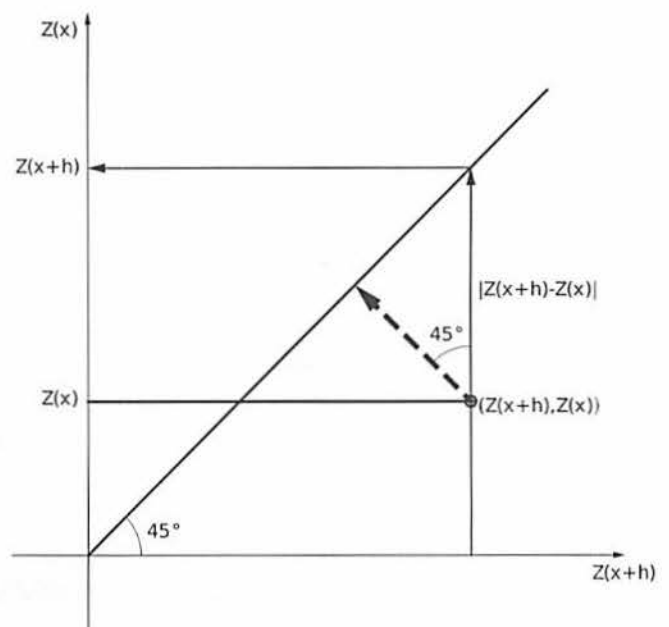


Fig. 2.2 Interpretação geométrica da função semivariograma em um diagrama de dispersão

Fonte: Journel (1989, p. 6).



é projetado na reta bissetriz, o que resulta na ordenada $Z(x+h)$; em seguida, determina-se a distância entre o ponto original e a reta bissetriz (vetor tracejado na Fig. 2.2). Esses três pontos formam um triângulo retângulo, cuja hipotenusa é a diferença em módulo entre $Z(x+h)$ e $Z(x)$. Sendo o i -ésimo par de coordenadas $(Z(x+h), Z(x))$, a distância para a reta bissetriz pode ser calculada como (Journel, 1989, p. 6):

$$d_i = |Z(x+h) - Z(x)| \cdot \cos 45^\circ$$

Elevando a i -ésima distância ao quadrado, tem-se:

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

Considerando n pares de pontos para uma determinada distância h , pode-se calcular a média das distâncias, a qual foi chamada por Journel (1989, p. 6) de momento de inércia:

$$\gamma_{x+h,x} = \frac{1}{n} \cdot \sum_{i=1}^n \frac{1}{2} [Z(x+h) - Z(x)]^2 = \frac{1}{2n} \cdot \sum_{i=1}^n [Z(x+h) - Z(x)]^2$$

Quanto maior a dispersão, maior o momento de inércia e menor a correlação. Se não houver dispersão, isto é, se todos os pares de pontos caem sobre a reta 45° , o momento de inércia é zero e o coeficiente de correlação é igual a 1 (máxima correlação). Journel (1989, p. 6-7) demonstrou que a fórmula do semivariograma não é empírica, mas resultante da interpretação geométrica dos pares de pontos em um diagrama de dispersão.

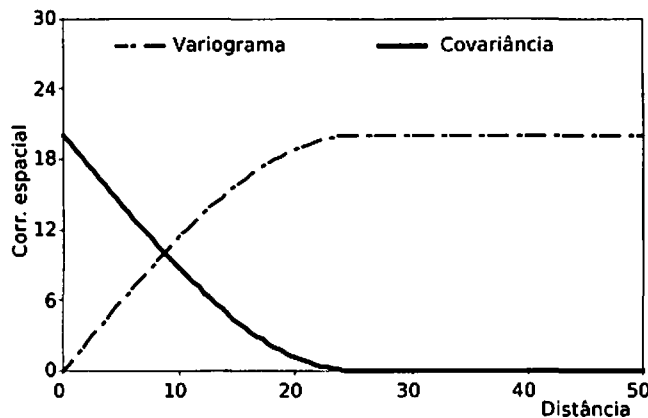


Fig. 2.3 Relação entre a função variograma e a função covariância

Como o variograma também usa a fórmula do semivariograma, é indiferente denominar variograma ou semivariograma, e, por simplicidade, o termo variograma será adotado neste livro.

Como $\gamma(h) = C(0) - C(h)$, isso faz com que, se o vetor h apresentar-se infinitamente pequeno, a variância seja mínima e a covariância, máxima.

Haverá um valor Δh para o qual as duas podem apresentar valores aproximadamente iguais, porém, à medida que Δh aumenta, a covariância diminui enquanto a variância aumenta, porque ocorre progressivamente maior independência entre os pontos a distâncias cada vez maiores (Fig. 2.3).

A função variograma distribui-se assim: de 0, quando $h = 0$, a um valor igual à variância das observações para um alto valor de h , se os dados forem estacionários, isto é, se não ocorrer a presença de tendência nos valores.

2.2 CÁLCULO DE VARIOGRAMAS EXPERIMENTAIS

O cálculo de variogramas experimentais não é algo simples e direto. Na verdade, o variograma é bastante sensível à distribuição dos pontos amostrais, bem como ao tipo de distribuição estatística associada. Com relação à distribuição espacial dos pontos amostrais, ela pode ser regular ou irregular.

2.2.1 Distribuição regular

É o caso em que o variograma pode ser calculado diretamente com base nos pontos amostrais. Os pares de pontos encontrados para uma determinada distância h , ao longo de uma direção, são usados para calcular as diferenças ao quadrado, as quais são acumuladas para o cálculo da média, conforme a Eq. 2.1. Como a malha é regular, as duas direções ortogonais são EW e NS; se a malha for quadrada, então se têm mais duas direções ortogonais, N45° e N315°; se a malha for retangular, as direções ao longo das duas diagonais do retângulo precisam ser calculadas com base nos lados do retângulo. A Fig. 2.4 ilustra uma malha quadrada e uma retangular. No caso da malha retangular do exemplo (Fig. 2.4B), as diagonais apresentam direções N33,6° e N326,4°.

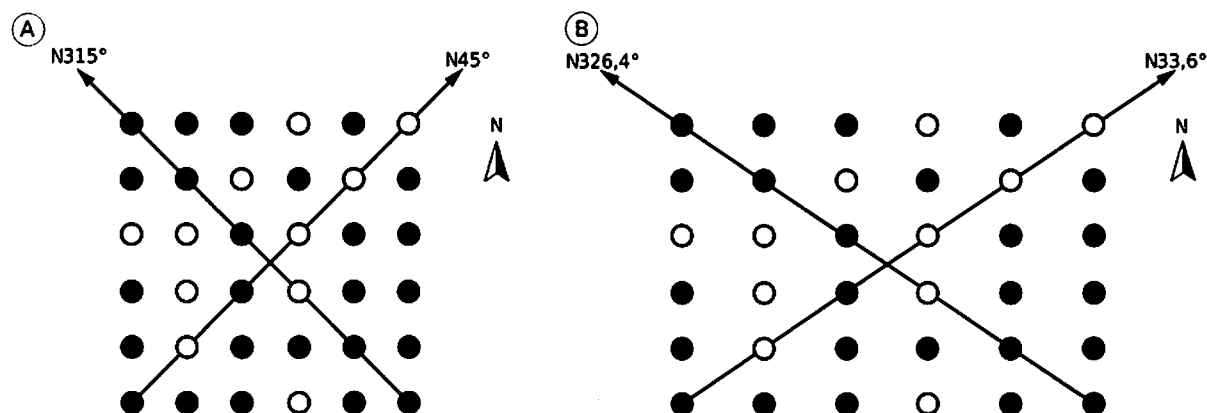


Fig. 2.4 A) Malha quadrada e B) malha retangular, com indicação das direções diagonais para cálculo dos variogramas experimentais. Círculo vazio = ponto não amostrado; círculo cheio = ponto amostrado

Para ilustrar o procedimento de cálculo de variogramas experimentais para dados com distribuição regular, sejam os dados de espessura de uma camada de carvão da região de Sapopema/PR (Tab. 2.1 e Fig. 2.5).

Embora a amostragem tenha sido planejada sobre uma malha regular, a figura mostra que muitos furos não foram feitos por diversos motivos: falta de acesso por causa de acidentes geográficos (lagos, rios, encostas íngremes etc.), bem como pela falta de interesse econômico, entre outros. Assim, a malha regular originalmente projetada pode se apresentar com dados irregulares.

Como descrito por Cava (1985) e Landim, Soares e Pumputis (1988), esse depósito situa-se a cerca de 20 km a noroeste de Figueira, no nordeste do Estado do Paraná, em sedimentos da parte superior do Membro Triunfo da Formação Rio Bonito.

Para calcular os variogramas em diversas direções, são encontrados os somatórios dos quadrados das

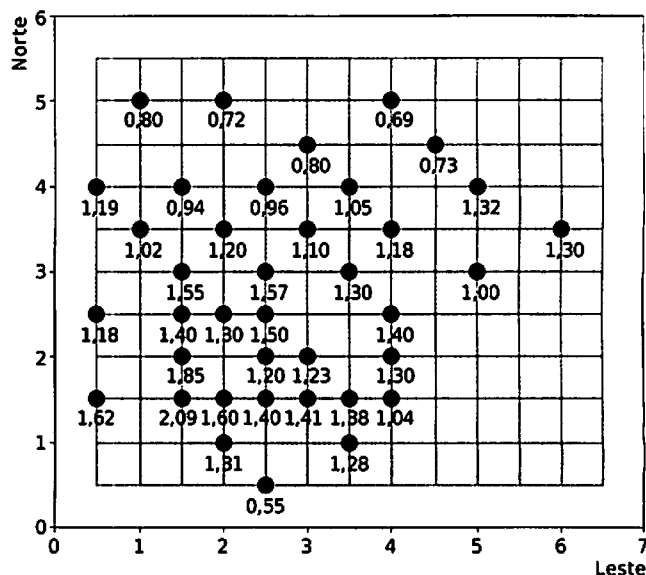


Fig. 2.5 Distribuição de valores da espessura de carvão, em rede regular

Fonte dos dados: Landim, Soares e Pumputis (1988).



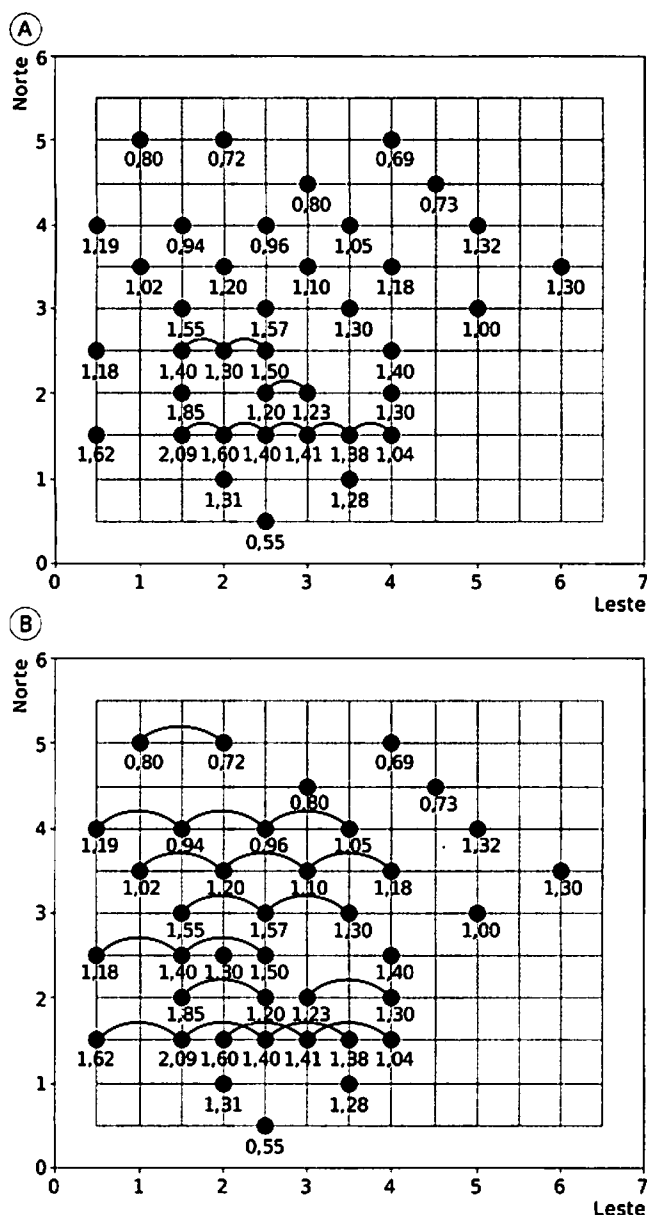


Fig. 2.6 Pares de pontos para o cálculo do variograma experimental na direção leste-oeste. Distância igual a: A) 0,5 m e B) 1,0 m

TAB. 2.1 Valores para a variável espessura da jazida de carvão em Sapopema/PR.

Ponto	X	Y	Esp.	Ponto	X	Y	Esp.
13	1,00	5,00	0,80	49	0,50	2,50	1,18
10	2,00	5,00	0,72	02	1,50	2,50	1,40
14	4,00	5,00	0,69	01	2,00	2,50	1,30
54	3,00	4,50	0,80	03	2,50	2,50	1,50
42	4,50	4,50	0,73	12	4,00	2,50	1,40
55	0,50	4,00	1,19	05	1,50	2,00	1,85
43	1,50	4,00	0,94	04	2,50	2,00	1,20
40	2,50	4,00	0,96	08	3,00	2,00	1,23
41	3,50	4,00	1,05	39	4,00	2,00	1,30
26	5,00	4,00	1,32	46	0,50	1,50	1,62
16	1,00	3,50	1,02	37	1,50	1,50	2,09
20	2,00	3,50	1,20	06	2,00	1,50	1,60
25	3,00	3,50	1,10	07	2,50	1,50	1,40
11	4,00	3,50	1,18	50	3,00	1,50	1,41
34	6,00	3,50	1,30	38	3,50	1,50	1,38
47	1,50	3,00	1,55	57	4,00	1,50	1,04
45	2,50	3,00	1,57	48	2,00	1,00	1,31
44	3,50	3,00	1,30	21	3,50	1,00	1,28
15	5,00	3,00	1,00	24	2,50	0,50	0,55

Fonte dos dados: Landim, Soares e Pumputis (1988).

diferenças e posteriormente se divide por duas vezes o número dessas diferenças. Assim, para a direção leste-oeste, inicia-se com o menor intervalo possível, ou seja, 0,5 m, da seguinte maneira, conforme os pares indicados na Fig. 2.6A:

$$\gamma^*(0,5) = \frac{1}{2 \times 8} [(1,4 - 1,3)^2 + (1,3 - 1,5)^2 + (1,2 - 1,23)^2 + (2,09 - 1,6)^2 + (1,6 - 1,4)^2 + (1,4 - 1,41)^2 + (1,41 - 1,38)^2 + (1,38 - 1,04)^2] = 0,028$$

Para o intervalo de 1,0 m, seguindo os pares de pontos da Fig. 2.6B:

$$\gamma^*(1,0) = \frac{1}{2 \times 18} [(0,8 - 0,72)^2 + (1,19 - 0,94)^2 + (0,94 - 0,96)^2 + (0,96 - 1,05)^2 + (1,02 - 1,2)^2 + (1,2 - 1,1)^2 + (1,1 - 1,18)^2 + (1,55 - 1,57)^2 + (1,57 - 1,3)^2 + (1,18 - 1,4)^2 + (1,4 - 1,5)^2 + (1,85 - 1,2)^2 + (1,23 - 1,3)^2 + (1,62 - 2,09)^2 + (2,09 - 1,4)^2 + (1,6 - 1,41)^2 + (1,4 - 1,38)^2 + (1,41 - 1,04)^2] = 0,043$$

E assim por diante, tanto para essa direção como para a norte-sul. Na Tab. 2.2, os variogramas experimentais foram calculados até uma distância máxima igual a 3,5 m. A distância máxima em que se pode calcular o variograma experimental é chamada de campo geométrico e é igual à metade do comprimento da linha na direção considerada (Journel; Huijbregts, 1978, p. 194).

No caso em estudo, o comprimento na direção leste é igual a 6 m, e na direção norte, igual a 4,5 m. Assim, o campo geométrico para a direção leste deveria ser igual a 3 m, e na direção norte, igual a 2,5 m. Mas nesse exemplo foi mantido um valor igual a 3,5 m para mostrar que, para distâncias grandes, há uma tendência à flutuação estatística da função variograma, pela diminuição do número de pares. Observar que as duas últimas distâncias na direção norte-sul, com apenas três pares cada, não têm significado estatístico.

Os variogramas experimentais obtidos para as duas direções consideradas encontram-se na Fig. 2.7. Como se pode verificar, a direção norte-sul apresenta maior variabilidade que a direção leste-oeste, significando que o comportamento é diferente conforme a direção pesquisada, o que indica, por sua vez, um fenômeno espacial anisotrópico.

2.2.2 Distribuição irregular

Para pontos com distribuição irregular, há necessidade de se definir parâmetros adicionais, além da distância e da direção. Isso é preciso para que a malha de pontos seja regularizada. Para cada ponto de dado, define-se uma janela, dentro da qual pode haver um ou mais pontos, ou nenhum. Essa janela é definida pela direção, tolerância angular e largura máxima, bem como pelo tamanho do passo (distância) e tolerância do passo (Fig. 2.8). O parâmetro largura máxima tem por objetivo limitar a abertura indefinida da janela de pesquisa dada pela tolerância angular.

O dispositivo de pesquisa é centrado em um ponto de dado. Por exemplo, na Fig. 2.8A, o dispositivo é centrado no ponto 1 e, nesse caso, o ponto 7 é encontrado dentro da janela. Então, a diferença ao quadrado entre os valores dos pontos 1 e 7 é considerada na Eq. 2.1. O dispositivo de pesquisa se movimenta para o ponto 2 e o ponto 6 é encontrado dentro da janela (Fig. 2.8B). Assim, a diferença ao quadrado entre os pontos 2 e 6 é somada na Eq. 2.1. E, assim, sucessivamente o processo é repetido até que todos os pontos do conjunto de dados sejam considerados. Mantendo-se a direção (azimute), todo o processo é repetido para os demais passos.

TAB. 2.2 Resultados do cálculo dos variogramas experimentais para dados de espessura de carvão referentes às direções leste-oeste e norte-sul

Distância	Leste-oeste		Norte-sul	
	$\gamma(h)$	Np	$\gamma(h)$	Np
0,5	0,028	8	0,028	11
1,0	0,043	18	0,097	15
1,5	0,051	12	0,069	13
2,0	0,047	12	0,147	7
2,5	0,158	6	0,216	9
3,0	0,015	5	0,133	3
3,5	0,104	4	0,178	3

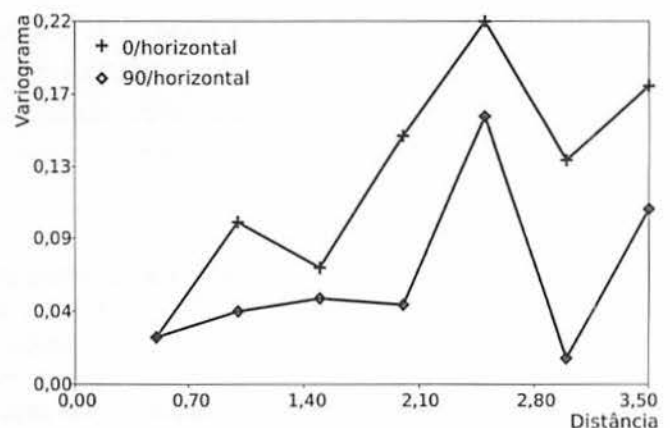


Fig. 2.7 Variogramas experimentais calculados para as direções norte-sul e leste-oeste

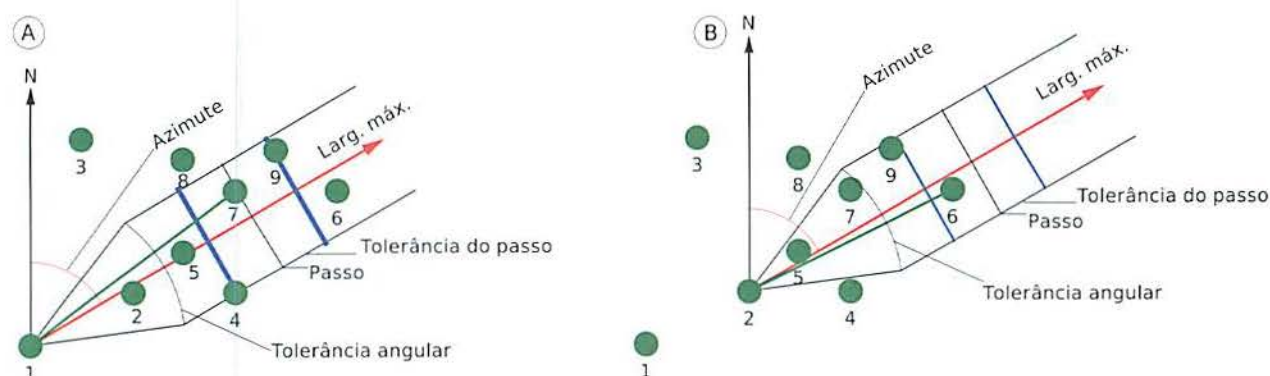


Fig. 2.8 Esquema mostrando a pesquisa de pares para cálculo de variogramas experimentais no caso de distribuição irregular: A) o dispositivo de pesquisa é centrado no ponto 1; B) o dispositivo de pesquisa se move e é centrado no ponto 2

Esse processo pode ser aplicado para dados com distribuição regular. Nesse caso, é preciso que sejam definidas as tolerâncias, tanto para o azimuth como para o passo.

Para ilustrar o procedimento de cálculo de variogramas experimentais para dados irregulares, seja o mesmo exemplo dos dados de carvão de Sapopema/PR, conforme Tab. 2.1 e Fig. 2.5. Os parâmetros do dispositivo de pesquisa foram estabelecidos de acordo com os valores da Tab. 2.3.

TAB. 2.3 Parâmetros para definição do dispositivo de pesquisa para cálculo de variogramas experimentais referentes a dados com distribuição irregular (Carvão de Sapopema/PR)

Azimute	Tol. angular	Larg. máxima	Passo	Tol. passo
0°	45°	2	0,5	0,25
90°	45°	2	0,5	0,25

Os resultados obtidos encontram-se na Tab. 2.4, e os gráficos, na Fig. 2.9.

TAB. 2.4 Resultados do cálculo dos variogramas experimentais nas direções leste-oeste e norte-sul para espessura da camada de carvão de Sapopema/PR

Direção leste-oeste			Direção norte-sul		
Distância	$\gamma(h)$	Np	Distância	$\gamma(h)$	Np
0,667	0,048	42	0,656	0,047	45
1,071	0,052	45	1,075	0,085	41
1,493	0,084	60	1,488	0,100	58
2,031	0,066	78	2,027	0,121	77
2,559	0,079	53	2,564	0,173	49
3,027	0,088	42	3,011	1,566	49
3,516	0,067	35	3,510	0,231	29

Comparando-se os variogramas calculados usando diferentes procedimentos, verifica-se que a aplicação do dispositivo de pesquisa (Tab. 2.3) resulta em variogramas com um número muito maior de pares, por causa da largura máxima igual a 2, considerada grande. Os variogramas resultantes (Figs. 2.7 e 2.9) são bastante semelhantes, mas o variograma calculado com maior número de pares é mais suavizado, facilitando o trabalho de ajuste do modelo teórico. O variograma experimental representado por um maior número de pares é estatisticamente mais significativo.

2.3 TIPOS DE VARIOGRAMAS

A função variograma mede a variância entre pontos separados por uma distância h . Assim, para pontos próximos, a diferença é pequena e, portanto, a variância é pequena. Ao aumentar a distância, os valores dos pontos tornam-se mais diferentes e, conseqüentemente, a variância aumenta. Muitas vezes, a variância se estabiliza em torno de uma variância máxima, a partir de certa distância. Isso significa que, mesmo com o aumento da distância, a função variograma irá oscilar em torno da variância máxima, denominada patamar. Esses casos definem os variogramas com patamar. Entretanto, há casos em que a variância continua aumentando indefinidamente com a distância, configurando os variogramas sem patamar.

2.3.1 Variogramas com patamar

Antes de introduzir os modelos de variogramas teóricos com patamar, é necessário conhecer as propriedades de um típico variograma com patamar (Fig. 2.10).

A distância segundo a qual $\gamma(h)$ atinge certo nível, denominado soleira ou patamar (sill), igual à variância *a priori* dos dados, é chamada de alcance ou amplitude (range). Geralmente, a soleira é representada por $C_0 + C$ e o alcance, por a . O efeito pepita C_0 é causado pela variância aleatória e C é denominada variância espacial. Ao se observar a origem do variograma mostrado na Fig. 2.10, verifica-se que, quando $\gamma(0) = 0$, ocorre o efeito pepita para distâncias muito pequenas, por exemplo, $\gamma(0,000000001) = C_0$.

O efeito pepita pode ser resultado tanto da variabilidade do fenômeno espacial em estudo como da escala de amostragem.

Teores são os melhores exemplos de variáveis regionalizadas que podem apresentar descontinuidade na origem.

O efeito pepita puro reflete um fenômeno que não é completamente conhecido, por falta de informação, mas não necessariamente um fenômeno espacial aleatório.

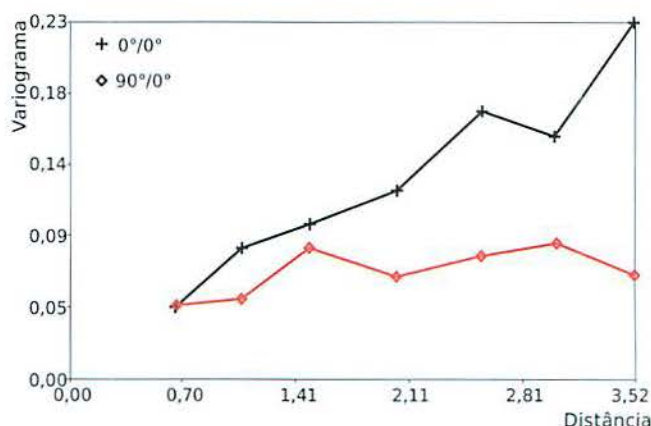


Fig. 2.9 Variogramas experimentais para as direções leste-oeste e norte-sul para os dados de espessura usando parâmetros do dispositivo de pesquisa da Tab. 2.3

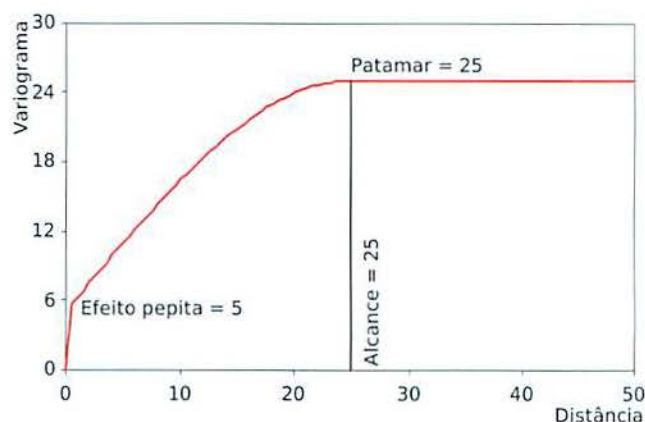


Fig. 2.10 Propriedades de um típico variograma com patamar

Outra consideração importante a ser feita é determinar o grau de aleatoriedade presente nos dados, conforme Guerra (1988):

$$E = \frac{C_o}{C}$$

O grau de aleatoriedade pode ser classificado em três intervalos, como pode ser visto na Tab. 2.5.

TAB. 2.5 Classificação dos graus de aleatoriedade

Grau de aleatoriedade	Componente aleatória
$E < 0,15$	Pequena
$0,15 \leq E \leq 0,30$	Significativa
$E > 0,30$	Muito significativa

Fonte: Guerra (1988).

O extremo dessa situação é o modelo efeito pepita puro, em que não ocorre correlação entre os valores e, portanto, a análise semivariográfica não se aplica, sugerindo o uso de outros métodos de interpolação.

Embora existam vários modelos de variogramas teóricos com patamar, apenas alguns são considerados como os mais comuns que podem explicar a variabilidade da grande maioria dos fenômenos espaciais (Fig. 2.11).

A Tab. 2.6. apresenta as equações dos modelos teóricos de variogramas ilustrados na Fig. 2.12.

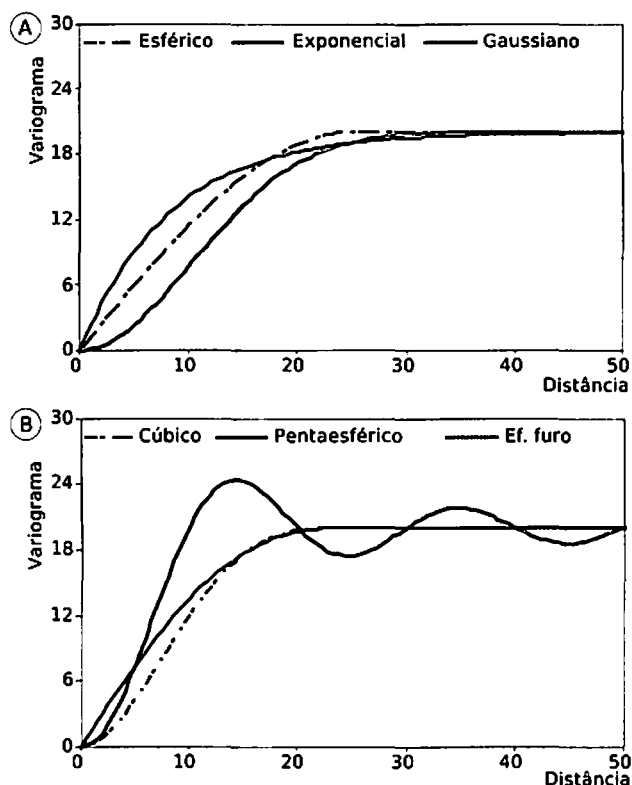


Fig. 2.11 Modelos de variogramas com patamar: A) esférico, exponencial e gaussiano; B) cúbico, pentaesférico e efeito furo, conforme equações disponíveis em Olea (1999, p. 76-79)

TAB. 2.6 Modelos teóricos de variogramas com patamar

Modelo	Equação
Esférico	$\begin{cases} \gamma(h) = C_o + C \left[1,5 \frac{h}{a} - 0,5 \left(\frac{h}{a} \right)^3 \right] & \text{para } h < a \\ \gamma(h) = C_o + C & \text{para } h \geq a \end{cases}$
Exponencial	$\gamma(h) = C_o + C \left[1 - \exp \left(- \frac{h}{a} \right) \right]$
Gaussiano	$\gamma(h) = C_o + C \left[1 - \exp \left(- \left(\frac{h}{a} \right)^2 \right) \right]$
Cúbico	$\begin{cases} \gamma(h) = C_o + C \left[7 \left(\frac{h}{a} \right)^2 - \frac{35}{4} \left(\frac{h}{a} \right)^3 + \frac{7}{2} \left(\frac{h}{a} \right)^5 - \frac{3}{4} \left(\frac{h}{a} \right)^7 \right] & \text{para } h < a \\ \gamma(h) = C_o + C & \text{para } h \geq a \end{cases}$
Pentaesférico	$\begin{cases} \gamma(h) = C_o + C \left[\frac{15}{8} \left(\frac{h}{a} \right) - \frac{5}{4} \left(\frac{h}{a} \right)^3 + \frac{3}{8} \left(\frac{h}{a} \right)^5 \right] & \text{para } h < a \\ \gamma(h) = C_o + C & \text{para } h \geq a \end{cases}$
Efeito furo	$\gamma(h) = C_o + C \left[1 - \frac{\sin \pi(h/a)}{\pi(h/a)} \right]$

Fonte: Olea (1999, p. 76-79).

Dos modelos teóricos apresentados (Fig. 2.12 e Tab. 2.6), os três primeiros explicam a maioria dos fenômenos espaciais. É importante lembrar que, para todos os modelos de variogramas, com ou sem efeito pepita, $\gamma(0) = 0$.

2.3.2 Variogramas sem patamar

Geralmente, quando a amostragem é insuficiente ou incompleta, ou até na presença de tendência nos dados, o variograma experimental não apresenta patamar. O modelo teórico para variogramas sem patamar pode ser representado pelo variograma de potência (Olea, 1999, p. 79):

$$\gamma(h) = \alpha h^\beta, \text{ com } 0 < \beta < 2$$

Nesse caso, α representa uma constante positiva que multiplica a distância elevada a uma potência β . Para $\beta = 1$, ocorre o modelo variograma linear. O caso extremo da potência β igual a 0 corresponde ao modelo de variograma efeito pepita puro. Os modelos de variogramas de potência que podem ser obtidos conforme os valores possíveis de β encontram-se na Fig. 2.12.

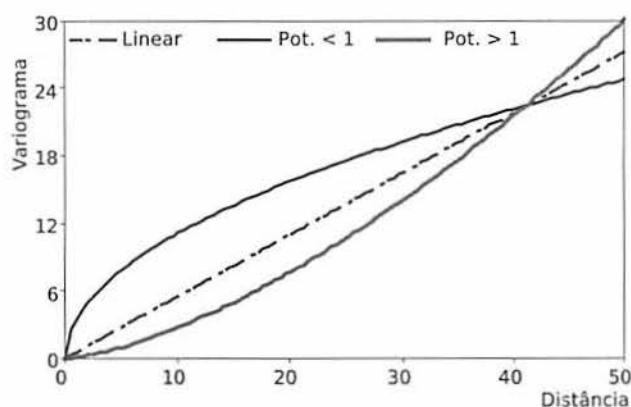


Fig. 2.12 Modelos de variogramas de potência (sem patamar)

2.4 ANISOTROPIAS

Os fenômenos espaciais podem apresentar anisotropias quando a função variograma muda conforme a direção (Fig. 2.1B). Quando a função variograma não se altera com a direção, diz-se que o fenômeno é isotrópico (Fig. 2.1A).

A Fig. 2.13 ilustra os tipos de anisotropias mais comuns encontrados na natureza.

A anisotropia geométrica (Fig. 2.13A) caracteriza-se pela existência de um único patamar e duas amplitudes diferentes. A Fig. 2.1B ilustra um caso de anisotropia geométrica, na qual a direção N30° apresenta maior continuidade que a N300°.

A anisotropia zonal (Fig. 2.13B) apresenta patamares diferentes conforme a direção analisada, mas todos sob um mesmo alcance.

Na anisotropia mista, tanto a amplitude como o patamar variam conforme a direção (Fig. 2.13C).

Ao detectar a presença de anisotropias, elas devem ser modeladas, ou seja, ajustadas a um modelo teórico de variograma. Na fase de modelagem, bem como na sua utilização para fins de estimativa e simulação, deve-se considerar a correção da anisotropia.

O objetivo da correção da anisotropia é a obtenção de um variograma isotrópico para o modelo de correlação espacial, ou seja, um modelo com parâmetros comuns (efeito pepita, variância espacial e amplitude) em todas as direções.

2.4.1 Correção da anisotropia para dados 2D

A correção da anisotropia geométrica (Fig. 2.14) é mais simples que a da zonal. Primeiro, mede-se o ângulo θ entre o norte e o eixo maior da elipse que representa a anisotropia geométrica.



Em seguida, faz-se a rotação dos eixos conforme o ângulo θ . Assim, dado o vetor distância $h = (h_x, h_y)$ entre dois pontos quaisquer, o novo vetor distância $h' = (h'_x, h'_y)$ após rotação de θ é obtido por:

$$\begin{bmatrix} h'_x \\ h'_y \end{bmatrix} = \begin{bmatrix} \cos(\theta) & \sin(\theta) \\ -\sin(\theta) & \cos(\theta) \end{bmatrix} \begin{bmatrix} h_x \\ h_y \end{bmatrix}$$

Após a rotação, faz-se o redimensionamento, de tal forma que a elipse ficará representada por um círculo de raio igual ao eixo menor:

$$\begin{bmatrix} h''_x \\ h''_y \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & \lambda \end{bmatrix} \begin{bmatrix} h'_x \\ h'_y \end{bmatrix}$$

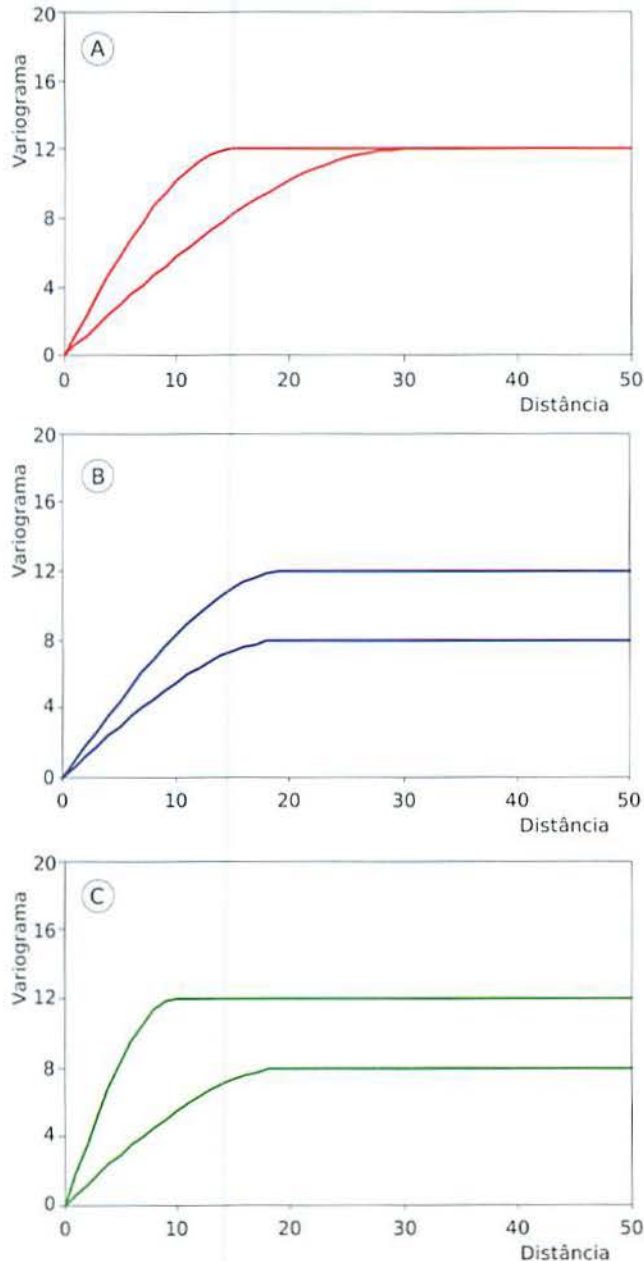


Fig. 2.13 Tipos de anisotropias em fenômenos espaciais: A) geométrica; B) zonal; e C) mista

Isso significa que, após a correção da anisotropia geométrica, será usado o variograma da direção de menor continuidade como variograma isotrópico.

A correção da anisotropia zonal se dá pela soma de um número de estruturas imbricadas igual ao número de patamares:

$$\gamma(h) = \gamma_1(h_1) + \gamma_2(h_2)$$

Supondo patamares diferentes nas duas direções, podem ocorrer no máximo duas estruturas imbricadas. Para cada estrutura imbricada que foi verificada de acordo com uma direção do variograma, fazem-se a rotação e o redimensionamento de coordenadas, de acordo com o que foi feito para a correção da anisotropia geométrica. Para a primeira estrutura imbricada usa-se o variograma de menor patamar, com o qual se calcula a componente $\gamma_1(h_1)$. Para a segunda estrutura imbricada, é usado o modelo de variograma dessa estrutura, mas com patamar correspondente à variância espacial entre o primeiro e o segundo variograma, e assim por diante.

Para ilustrar o procedimento de correção de anisotropia zonal, suponha o modelo de variograma com anisotropia zonal da Fig. 2.15A. Na realidade, trata-se de um variograma com anisotropia mista, pois há dois patamares e as amplitudes são diferentes nas duas direções (45° e 135°). Esse variograma com anisotropia mista pode ser decomposto em duas estruturas imbricadas em cada uma das direções (45° e 135° , respectivamente Figs. 2.15B e 2.15C).

Os parâmetros do variograma com anisotropia mista encontram-se na Tab. 2.7.

TAB. 2.7 Parâmetros para ajuste do variograma com anisotropia mista

Estrutura	Modelo	Var. esp.	θ	Amax	Amin
1	Esférico	80	45°	10	15
2	Esférico	60	45°	9.9e+30	15

Obs.: Amax = amplitude máxima; Amin = amplitude mínima.

Como se pode verificar na Fig. 2.15A e na Tab. 2.7, o variograma apresenta dois patamares e, portanto, são necessárias duas estruturas para a sua modelagem. A primeira estrutura tem como amplitude máxima e mínima: 15 e 10, respectivamente. A segunda estrutura tem como amplitude máxima um valor muito grande (Deutsch; Journel, 1992, p. 25) e a amplitude mínima igual à da primeira estrutura. Isso é necessário para liberar a segunda estrutura da primeira. Feito isso, a correção se processa fazendo a rotação de 45° dos eixos de coordenadas e, em seguida, o cálculo de cada uma das estruturas imbricadas, e o resultado é igual à soma dessas estruturas.

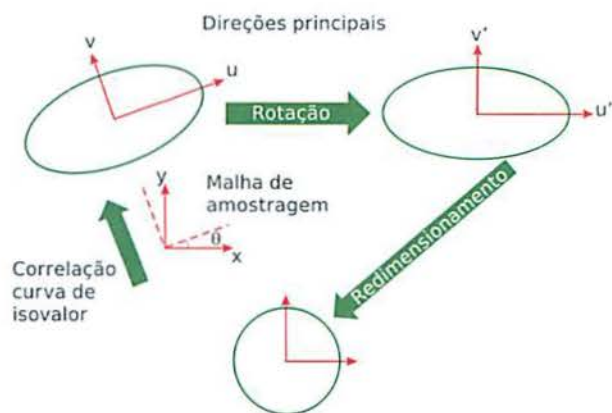


Fig. 2.14 Esquema mostrando a correção da anisotropia geométrica

Fonte: Chorti e Hristopulos (2008, p. 4.739).

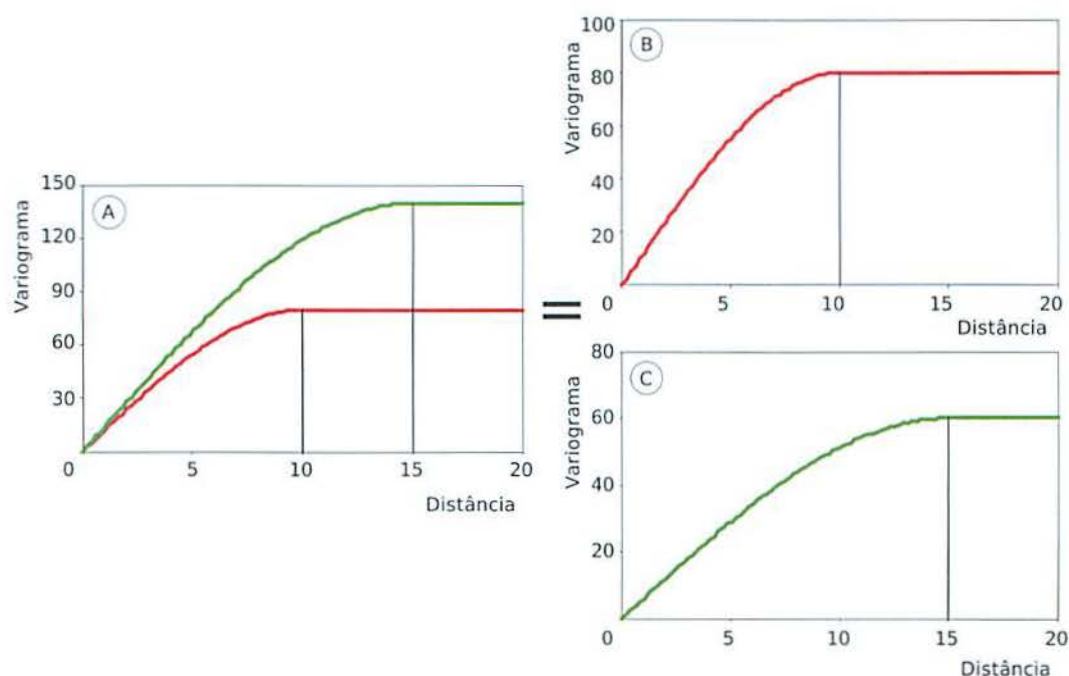


Fig. 2.15 A) Variograma com anisotropia mista (direção 45° - vermelho; direção 135° - verde) e sua decomposição em duas estruturas imbricadas: B) direção 45°; e C) direção 135°

2.4.2 Correção da anisotropia para dados 3D

A correção da anisotropia geométrica ou zonal em 3D envolve a rotação dos eixos de coordenadas no espaço, conforme os ângulos de direção, mergulho e *plunge* (Fig. 2.16). A direção é dada pela interseção entre o plano horizontal e a estrutura geológica, medida no

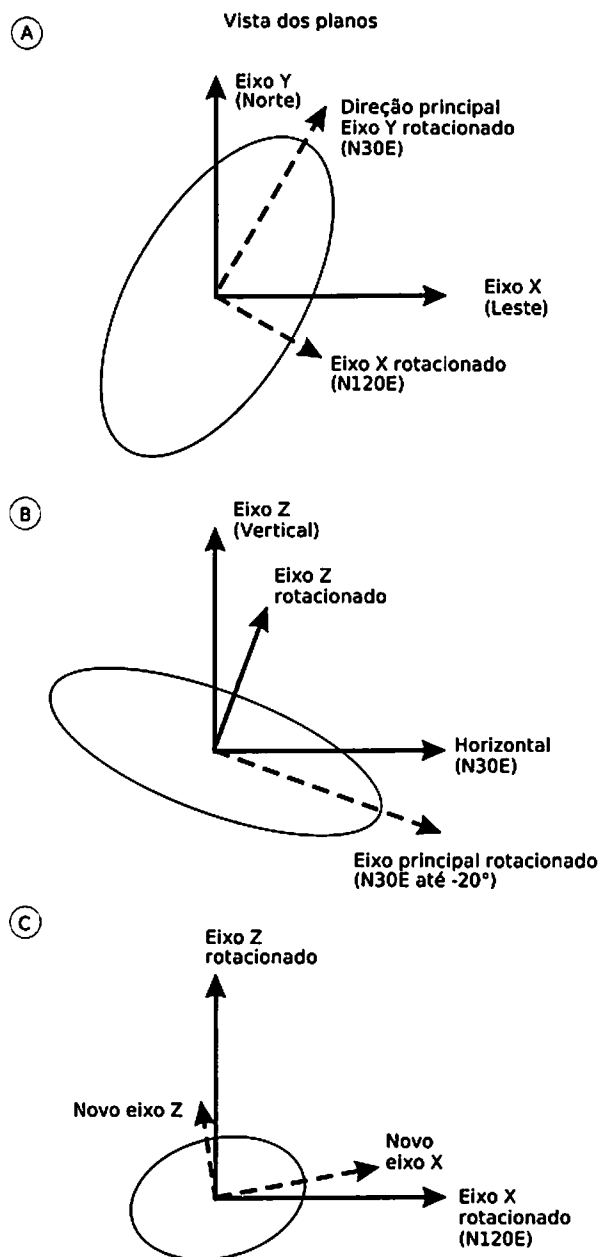


Fig. 2.16 Desenho ilustrando os três ângulos necessários para a rotação dos eixos de coordenadas em 3D
Fonte: Deutsch e Journel (1992, p. 26).

plano horizontal. Portanto, a direção é o próprio ângulo de azimuth, considerando que o sistema de coordenadas está orientado em relação ao norte. O mergulho é o ângulo entre o plano horizontal e a feição geológica, medido no plano vertical perpendicular à direção. *Plunge* é o ângulo vertical entre o plano horizontal e a linha de máxima elongação da feição geológica, por exemplo, o eixo de uma dobra.

A rotação dos eixos de coordenadas para o novo sistema, de acordo com a direção, mergulho e *plunge* da feição geológica, envolve a multiplicação da matriz rotação pelo vetor $h = (h_x, h_y, h_z)$. A matriz rotação passa a ser igual a 3×3 e envolve os três ângulos mencionados. Detalhes dessa matriz rotação poderão ser conferidos em Leuangthong, Khan e Deutsch (2008, p. 52-56).

Da mesma forma, como na correção da anisotropia geométrica e zonal em 2D, a rotação é feita para a direção de maior continuidade. Se a anisotropia for zonal ou mista, além da rotação há necessidade de se fazer a decomposição do variograma em um número de estruturas imbricadas igual ao número de patamares.

Para dados 3D, pode-se ter até três patamares, ou seja, um patamar para cada direção. O variograma com anisotropia zonal ou mista com três patamares pode ser decomposto em três estruturas imbricadas:

$$\gamma(h) = \gamma_1(h_1) + \gamma_2(h_2) + \gamma_3(h_3)$$

Um exemplo hipotético de anisotropia mista pode ser observado na Fig. 2.17A (p. 48), que pode ser decomposta em três estruturas imbricadas (Fig. 2.17B,C,D).

Os parâmetros necessários para modelagem e correção do variograma com anisotropia mista da Fig. 2.17 estão listados na Tab. 2.8.

TAB. 2.8 Parâmetros para correção da anisotropia mista da Fig. 2.17

Estrutura	Modelo	Var. esp.	θ	Amax	Amin	Aver
1	Esférico	30	30°	10	15	5
2	Esférico	40	120°	9.9e+30	15	5
3	Esférico	50	Vertical	9.9e+30	9.9e+30	5

Obs.: Amax = amplitude máxima; Amin = amplitude mínima; Aver = amplitude vertical.

Observar na Tab. 2.8, na segunda estrutura, que a amplitude máxima (A_{max}) foi colocada no infinito para liberar a segunda estrutura da primeira; na terceira estrutura, a amplitude máxima (A_{max}) e a amplitude mínima (A_{min}) foram colocadas no infinito para liberar a terceira estrutura da segunda e da primeira.

2.5 COMPORTAMENTO DO VARIOGRAMA PRÓXIMO À ORIGEM

As variáveis regionalizadas podem apresentar comportamentos distintos próximo à origem do variograma: parabólico, linear, efeito pepita e efeito pepita puro (Journel; Huijbregts, 1978, p. 38-39). Nesse sentido, os variogramas são classificados em contínuos (Fig. 2.18A,B) e descontínuos (Fig. 2.18C,D). Algumas variáveis regionalizadas, como a espessura, apresentam alta continuidade na origem (Fig. 2.18A), mostrando um comportamento parabólico característico de uma variabilidade espacial altamente regular, segundo Journel e Huijbregts (1978, p. 38).

Isso significa que, em pequenas distâncias, a variável não se altera tanto, mas as diferenças começam a surgir com distâncias maiores. Em geral, variáveis como teores podem apresentar média continuidade na origem, ou seja, um comportamento linear (Fig. 2.18B). Existem também variáveis descontínuas na origem (Fig. 2.18C,D) por causa do efeito pepita.

Teores são os melhores exemplos de variáveis regionalizadas que podem apresentar descontinuidade na origem, em geral pelo efeito pepita (Fig. 2.18C), que, segundo Journel e Huijbregts (1978, p. 39), seria causado por erros de medidas e por microvariabilidades da mineralização.

O efeito pepita, também denominado variância aleatória, reflete a incerteza em pequenas distâncias, principalmente pela falta de conhecimento da distribuição espacial da variável em estudo. Quanto maior o efeito pepita, maior a variabilidade e, conseqüentemente, a amostragem se torna insuficiente para esse nível de variabilidade espacial.

O efeito pepita puro (Fig. 2.18D) pode ser representado por um fenômeno de transição com um patamar igual ao efeito pepita e uma amplitude muito pequena em relação a distâncias de observações experimentais (Journel; Huijbregts, 1978, p. 39).

Com o objetivo de mostrar o efeito da amostragem no cálculo de variogramas experimentais, amostras de diversos tamanhos (25, 36, 49, 64, 81 e 100 pontos de dados) foram extraídas por amostragem aleatória estratificada do Arquivo completo 1 (disponível em: <<http://lig.igc.usp.br/geoestatistica/anexob/download/Bell.txt>>; Fig. 1.1). Com essas amostras (Arquivos 5 a 10 do Anexo B) foram calculados variogramas experimentais para duas direções, 45° e 135°, conforme os resultados apresentados na Fig. 2.19.

Todos os variogramas experimentais foram calculados com os mesmos parâmetros para definição do dispositivo de pesquisa, conforme a Tab. 2.9.

Os variogramas experimentais para a amostra com 25 pontos de dados (Fig. 2.19A) mostram praticamente o comportamento de um fenômeno espacial aleatório. Mas isso se deve à falta de amostragem a distâncias menores que aquelas consideradas nesses variogramas experimentais. Aumentando a amostra para 36 pontos de dados (Fig. 2.19B), verifica-se que os variogramas começam a apresentar alguma estruturação, apesar dos números de pares sempre menores que 30. Os variogramas experimentais para a amostra

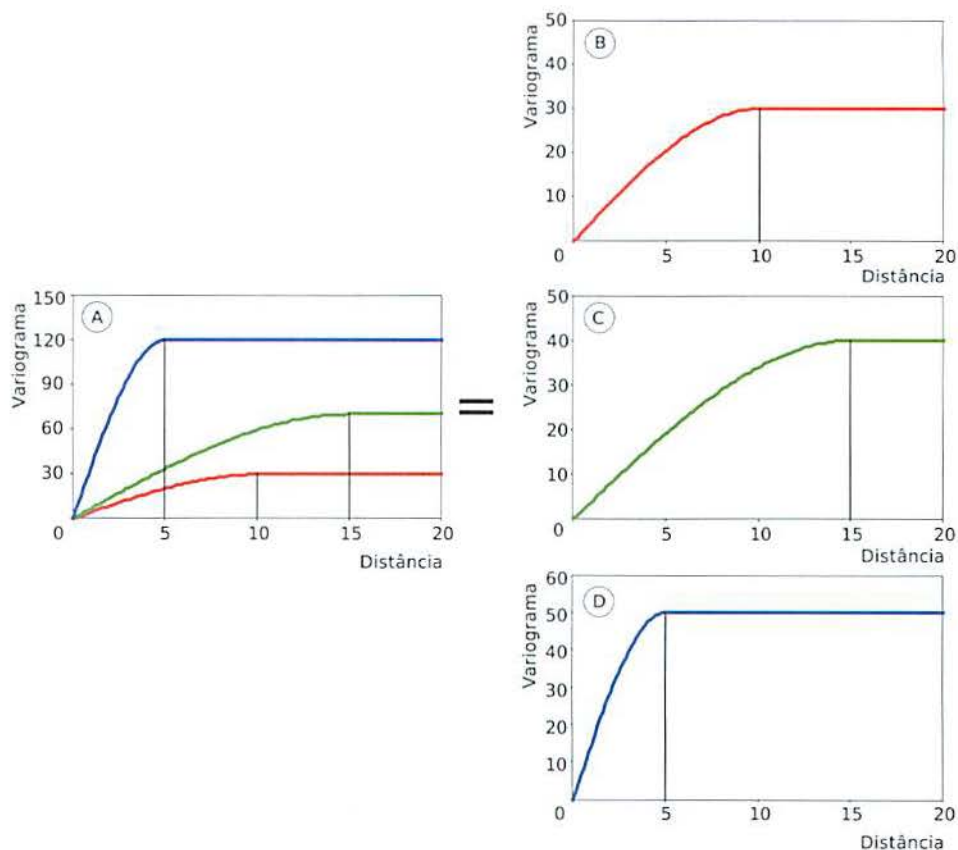


Fig. 2.17 Variograma com A) anisotropia mista em três direções (direção N30° - vermelho; direção N120° - verde; direção vertical - azul) e sua decomposição em três estruturas imbricadas: B) direção N30°; C) direção N120°; e D) direção vertical

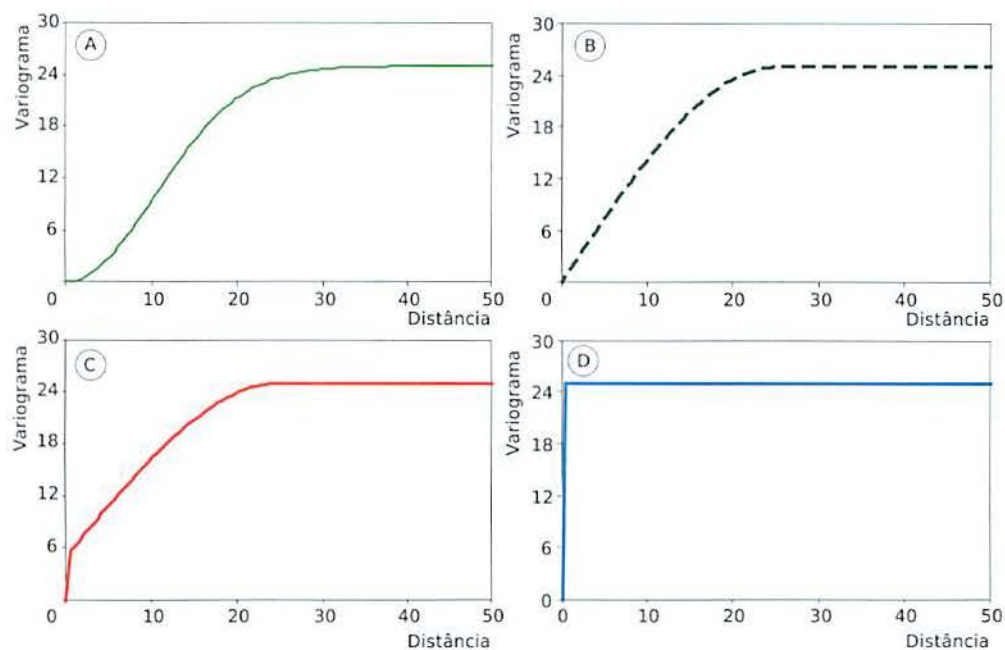


Fig. 2.18 Comportamento do variograma próximo à origem. Variogramas contínuos: A) alta continuidade na origem e B) média continuidade. Variogramas descontínuos: C) efeito pepita e D) efeito pepita puro

com 49 pontos (Fig. 2.19C) refletem melhor a estruturação do fenômeno espacial, mas calculados com número pequeno de pares de pontos, exceto alguns com mais de 30 pares. Todos os variogramas experimentais para amostras com tamanhos superiores a 49 pontos (Fig. 2.19D,E,F) já apresentam comportamentos contínuos, com aproximadamente as mesmas características em termos de alcance e patamar.

TAB. 2.9 Parâmetros para definição do dispositivo de pesquisa para cálculo de variogramas experimentais para amostras aleatórias estratificadas extraídas do Arquivo completo 1 (disponível em: <<http://liig.igc.usp.br/geoestatistica/anexob/download/Bell.txt>>; Fig. 1.1)

Azimute	Tol. angular	Larg. máxima	Passo	Tol. passo
45°	45°	8	2	1
135°	45°	8	2	1

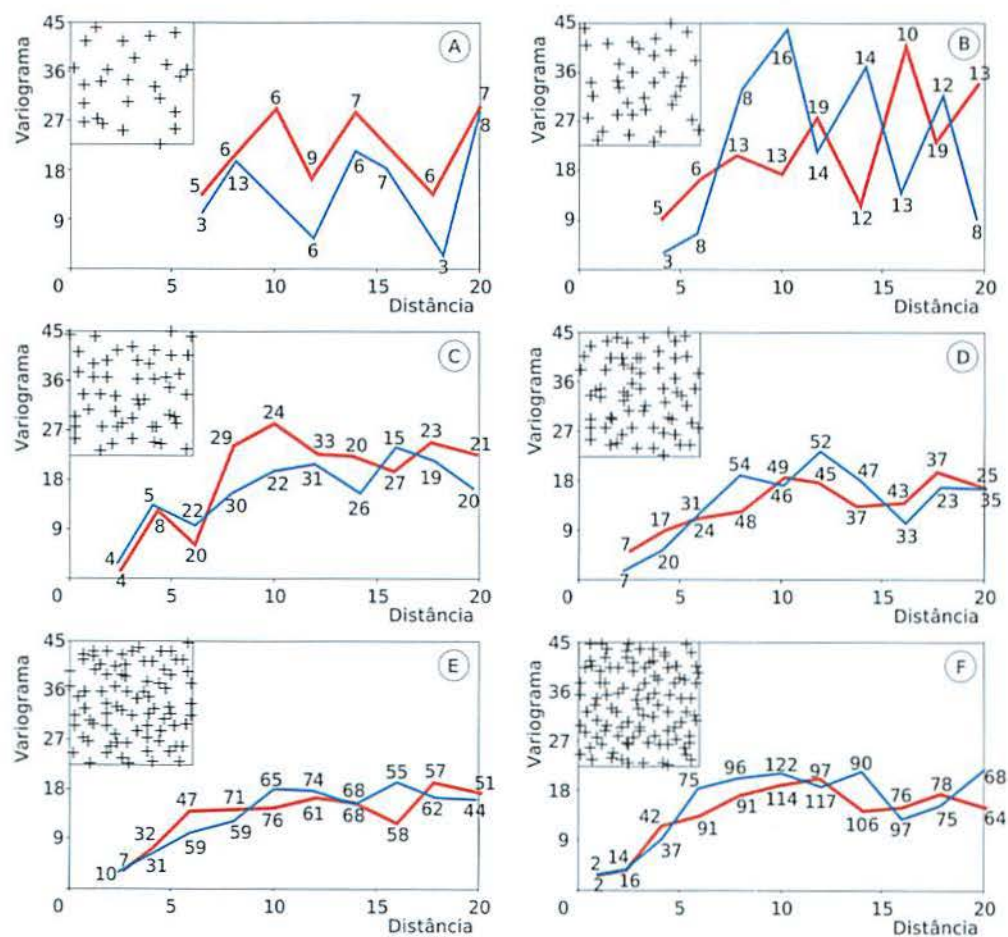


Fig. 2.19 Variogramas experimentais para amostras compostas por: A) 25 pontos; B) 36 pontos; C) 49 pontos; D) 64 pontos; E) 81 pontos; e F) 100 pontos. Linha vermelha: direção 45°; linha azul: direção 135°. No canto superior esquerdo de cada variograma, o mapa de localização de pontos. Os números indicam os pares encontrados para o cálculo dos variogramas experimentais

Segundo Journel e Huijbregts (1978, p. 194), o número de pares mínimo para os pontos do variograma experimental deve estar entre 30 e 50. Quando há informação suficiente, esse mínimo é facilmente alcançado, mas em situações de amostragem insuficiente dificilmente se consegue um número tão elevado de pares. Por exemplo, o variograma da Fig. 2.19C apresenta apenas três pontos com número de pares superior ou igual a 30, mas o variograma apresenta estrutura e poderia ser modelado. Portanto, a decisão em aceitar ou não um determinado variograma experimental dependerá do pesquisador. Mas em nenhuma hipótese se justifica o desenvolvimento de uma amostragem adicional para melhorar os pontos do variograma experimental, principalmente em mineração, na qual a pesquisa por sondagens é extremamente dispendiosa. A Geoestatística deve usar, portanto, a informação disponível da melhor maneira possível.

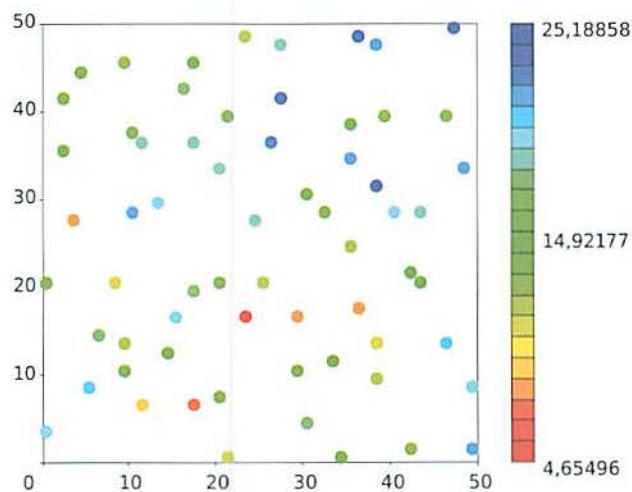


Fig. 2.20 Mapa de localização de pontos da amostra Arquivo 11, Anexo B

Exemplos de cálculo e modelagem de variogramas experimentais

Para ilustrar o procedimento de cálculo e modelagem de variogramas experimentais, foram consideradas três amostras aleatórias estratificadas. A primeira amostra, com 64 pontos, foi extraída do Arquivo completo 1 (disponível em: <<http://lig.igc.usp.br/geoestatistica/anexob/download/Bell.txt>>; Fig. 1.1) e denominada Arquivo 11, Anexo B. Outras duas amostras, com 64 e 100 pontos, denominadas, respectivamente, Arquivos 12 e 13, Anexo B, foram retiradas do conjunto completo de um fenômeno espacial apresentando distribuição lognormal.

Os mapas de localização de pontos para as amostras Arquivos 11, 12 e 13, Anexo B, encontram-se, respectivamente, nas Figs. 2.20, 2.21 e 2.22.

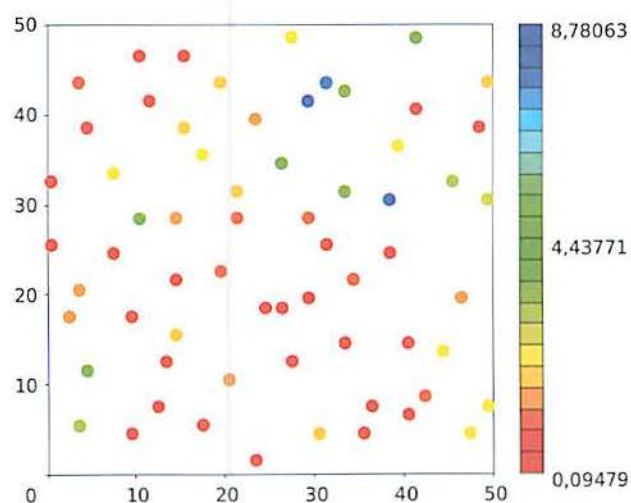


Fig. 2.21 Mapa de localização de pontos da amostra Arquivo 12, Anexo B

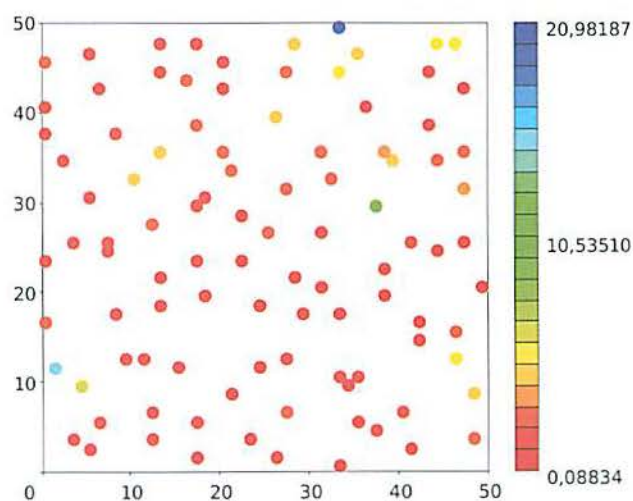


Fig. 2.22 Mapa de localização de pontos da amostra Arquivo 13, Anexo B

As estatísticas descritivas para essas amostras encontram-se resumidas na Tab. 2.10:

TAB. 2.10 Estatísticas descritivas para as amostras em estudo

Estatísticas	Amostras		
	Arquivo 11	Arquivo 12	Arquivo 13
<i>N</i>	64	64	100
Média	15,740	1,708	1,832
Desvio padrão	4,562	1,923	2,816
Coef. variação	0,290	1,126	1,538
Máximo	25,189	8,781	20,982
Quartil superior	18,870	2,113	2,107
Mediana	15,964	1,089	0,794
Quartil inferior	12,051	0,348	0,438
Mínimo	4,655	0,095	0,088

Nessa tabela é possível verificar que a amostra Arquivo 11, Anexo B, apresenta um baixo coeficiente de variação (0,290), enquanto as outras duas têm altos valores dessa estatística (1,126 e 1,538), comprovando o caráter lognormal dessas duas amostras. Essas características deverão influenciar os variogramas experimentais, pois dependem não apenas da distância e orientação, mas do tipo de distribuição de frequências. Todos os variogramas foram calculados com tolerância angular de 90°, ou seja, omnidirecionais. Os resultados encontram-se nas Figs 2.23 a 2.25.

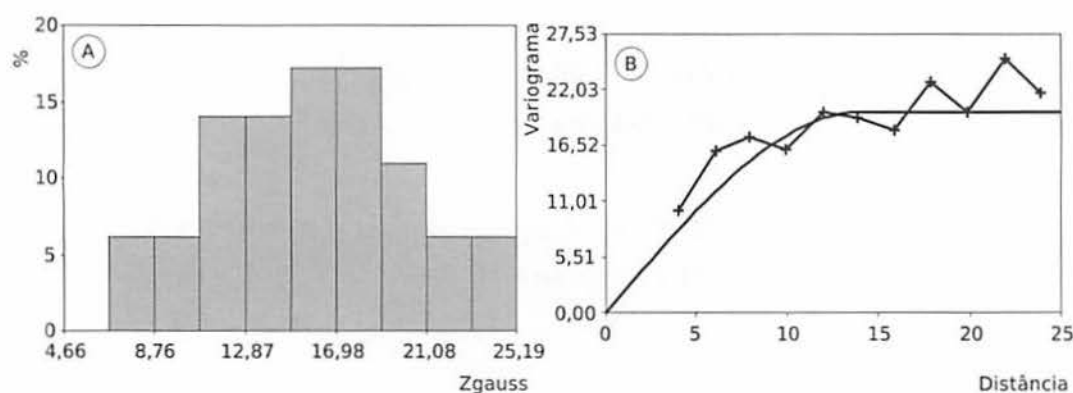


Fig. 2.23 A) Histograma e B) variograma para a amostra Arquivo 11, Anexo B

Como se pode verificar nessas figuras, os variogramas para as amostras com 64 pontos apresentam melhor estruturação que o variograma da amostra com 100 pontos. Considerando que a amostra Arquivo 13, Anexo B, tem uma boa distribuição espacial por toda a área de estudo, a falta de estruturação deve-se em grande parte à forte assimetria dessa variável.

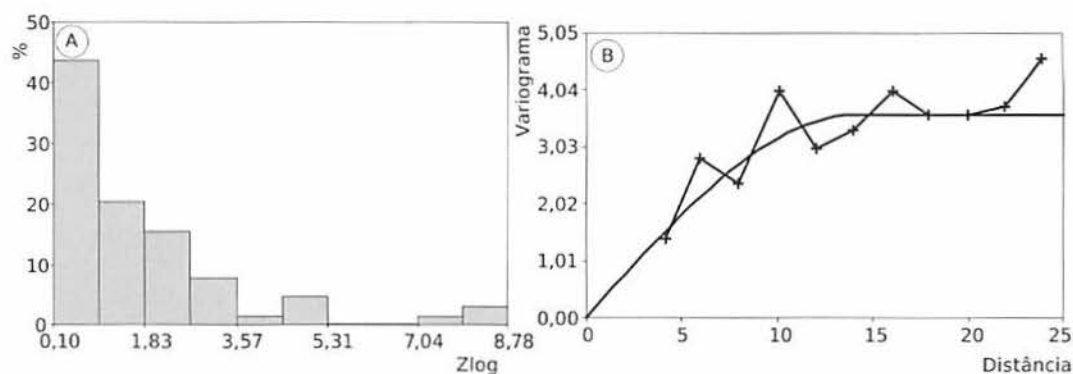


Fig. 2.24 A) Histograma e B) variograma para a amostra Arquivo 12, Anexo B

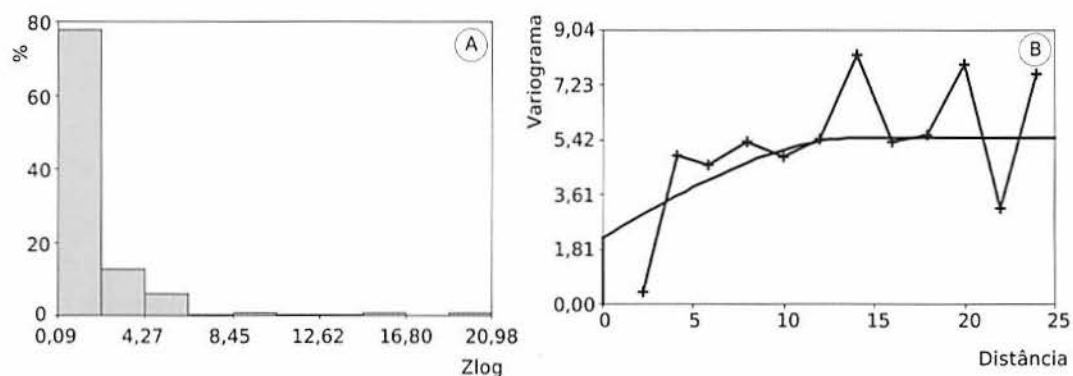


Fig. 2.25 A) Histograma e B) variograma para a amostra Arquivo 13, Anexo B

Os modelos ajustados aos variogramas experimentais estão descritos a seguir:

$$\text{Arquivo 11, Anexo B: } \begin{cases} \gamma(h) = 19,8 \left(1,5 \frac{h}{14,16} - 0,5 \left(\frac{h}{14,16} \right)^3 \right) & \text{para } h < 14,16 \\ \gamma(h) = 19,8 & \text{para } h \geq 14,16 \end{cases} \quad (2.2)$$

$$\text{Arquivo 12, Anexo B: } \begin{cases} \gamma(h) = 3,6 \left(1,5 \frac{h}{14,16} - 0,5 \left(\frac{h}{14,16} \right)^3 \right) & \text{para } h < 14,16 \\ \gamma(h) = 3,6 & \text{para } h \geq 14,16 \end{cases} \quad (2.3)$$

$$\begin{array}{l} \text{Arquivo 13,} \\ \text{Anexo B:} \end{array} \begin{cases} \gamma(h) = 2,2 + 3,3 \left(1,5 \frac{h}{14,16} - 0,5 \left(\frac{h}{14,16} \right)^3 \right) & \text{para } h < 14,16 \\ \gamma(h) = 5,5 & \text{para } h \geq 14,16 \end{cases} \quad (2.4)$$

Essas amostras serão usadas nos próximos capítulos para ilustrar os procedimentos de estimativas geoestatísticas, bem como os métodos de simulações estocásticas.

2.6 CONSIDERAÇÕES FINAIS

Procurou-se mostrar neste capítulo todas as etapas envolvidas no cálculo de variogramas experimentais, desde o arranjo dos pontos de dados, a estratégia de busca dos pares de

pontos, os tipos de variogramas de continuidade e de anisotropias, o problema do efeito pepita e da escala de amostragem e a necessária modelagem.

A modelagem é um processo que envolve várias tentativas e no qual a experiência pesa muito. Ela é necessária para ajustar uma função matemática que descreva continuamente a variabilidade ou correlação espacial existente nos dados.

O variograma experimental não serve para esse fim, porque há necessidade de interpolação e os pontos irão se apresentar com alguma dispersão, principalmente para distâncias grandes, quando o número de pares de amostras diminui.

O cálculo e a modelagem de variogramas experimentais representam a etapa mais importante do estudo geoestatístico, tanto para fins de estimativas como para simulações estocásticas.

A Fig. 2.26 apresenta uma síntese do cálculo e modelagem de variogramas experimentais.

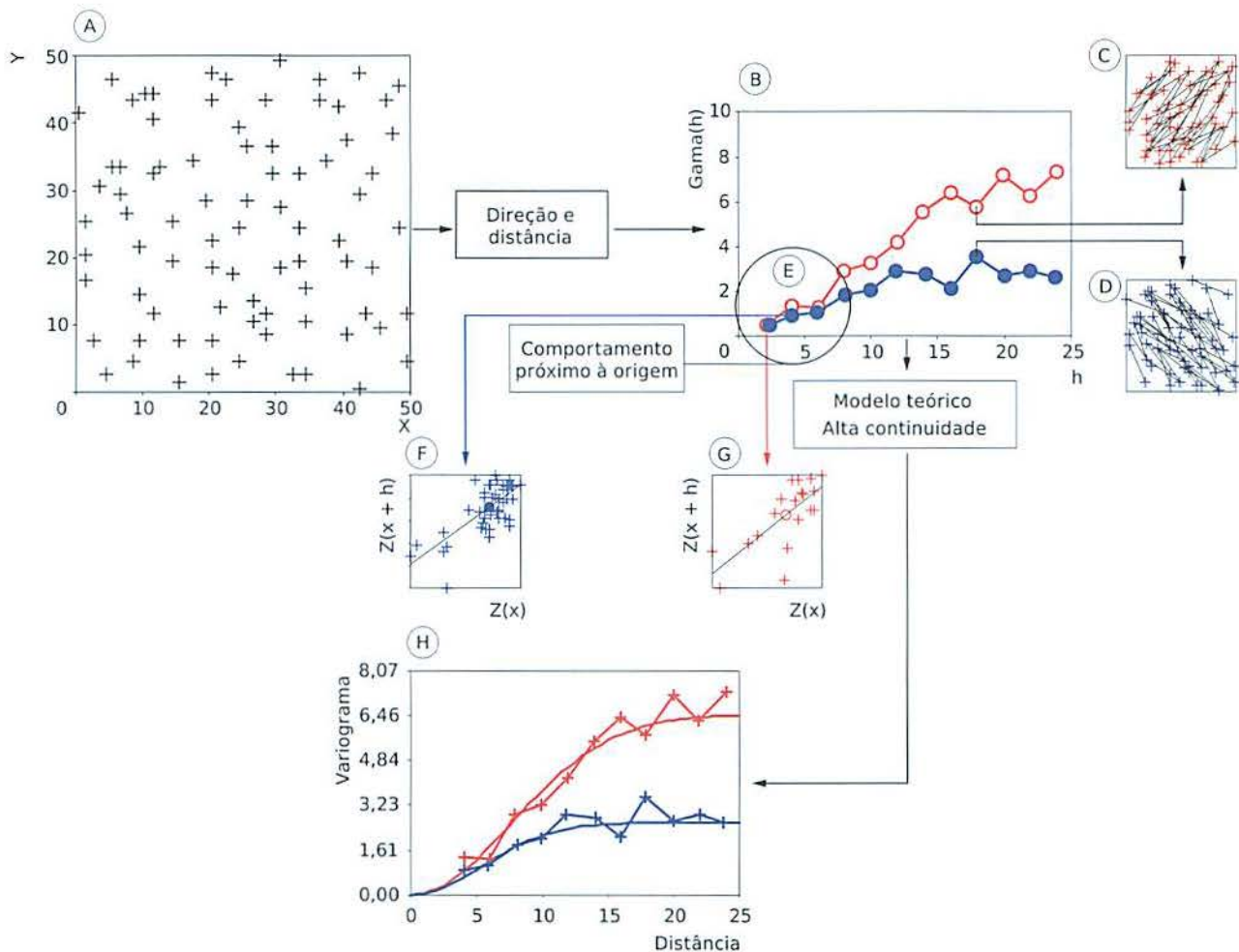


Fig. 2.26 Síntese do procedimento de cálculo e modelagem de variogramas experimentais. A) Mapa de pontos; B) variogramas experimentais calculados para as direções de 45° (vermelho) e 135° (azul); C) vetores usados no cálculo do variograma experimental para a direção de 45°; D) vetores usados no cálculo do variograma experimental para a direção de 135°; E) destaque para o comportamento próximo à origem, com alta continuidade; F) interpretação geométrica de Journel (1989) para a direção de 135°; G) interpretação geométrica de Journel (1989) para a direção de 45°; H) modelos teóricos ajustados aos variogramas experimentais

Estimativas Geoestatísticas

3

Todo o processo de inferência espacial tem início com a coleta de uma amostra composta por n pontos de dados. É esperado que essa amostra seja representativa do fenômeno em estudo, em termos da distribuição e variabilidade espaciais.

Krigagem é um processo geoestatístico de estimativa de valores de variáveis distribuídas no espaço e/ou tempo, com base em valores adjacentes quando considerados interdependentes pela análise variográfica. Pode ser comparado com os métodos tradicionais de estimativa por médias ponderadas ou por médias móveis, mas a diferença fundamental é que somente a krigagem apresenta estimativas não tendenciosas e a mínima variância associada ao valor estimado.

O termo – tradução do francês *krigeage* e do inglês *kriging* – foi cunhado pela Escola Francesa de Geoestatística em homenagem a Daniel G. Krige, engenheiro de minas sul-africano e pioneiro na aplicação de técnicas estatísticas em avaliação mineira. Abrange uma família de algoritmos conhecidos, entre outros, como krigagem simples, krigagem da média, krigagem ordinária e krigagem universal. O estimador mais usual é a krigagem ordinária, cuja tradução, do francês *krigeage ordinaire*, deveria ser krigagem normal (Soares, 2006, p. 69). A tradução para krigagem ordinária, porém, está consagrada no Brasil e, assim, será a usada nesta obra.

Estimativas geoestatísticas são, em geral, superiores aos demais métodos de interpolação numérica, pois fazem uso da função variograma, que não é simplesmente uma função da distância entre pontos, mas depende da existência ou não do efeito pepita, da amplitude e da presença de anisotropia.

Na impossibilidade de obtenção de um modelo de correlação espacial, métodos de interpolação não estocásticos, que não necessitam do variograma, podem ser considerados (Fig. 3.1).

A estimativa geoestatística tem por objetivo a modelagem do fenômeno espacial em estudo, ou seja, determinar a distribuição e variabilidade espaciais da variável de interesse.

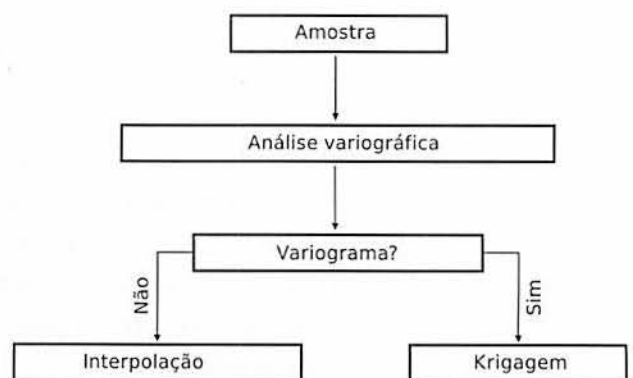


Fig. 3.1 Interpolação ou krigagem, dependendo da obtenção de variograma

Os valores obtidos nos pontos amostrais são usados na interpolação ou estimativa geoestatística para fornecer uma grade regular 2D ou 3D, dependendo da dimensionalidade do fenômeno espacial.

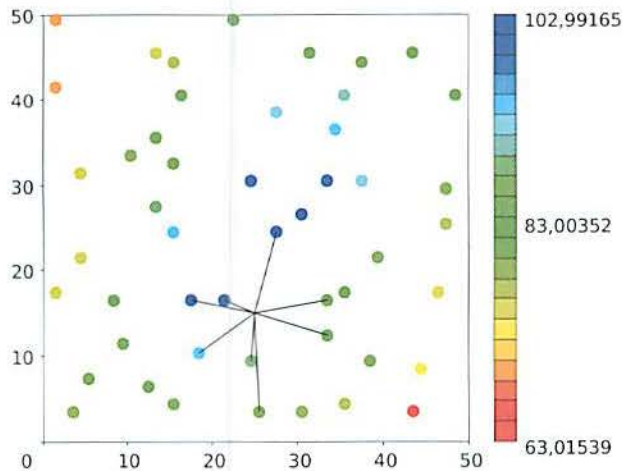


Fig. 3.2 Localização de vizinhos mais próximos (dois pontos por quadrante) para estimativa do ponto não amostrado

A modelagem da distribuição e variabilidade espaciais da variável de interesse é feita geralmente em malhas regulares, que permitem analisar a inferência espacial com maior precisão.

Em Geoestatística, trabalha-se com funções locais, pois ela é, por excelência, um método local de estimativa. Nesse sentido, pontos distantes situados além do alcance do variograma não deveriam ser considerados, mas a krigagem tem um mecanismo interno de atenuação da influência desses pontos e, portanto, podem ser deixados como pertencentes à vizinhança.

As Figs. 3.2 e 3.3 ilustram exemplos em 2D e 3D, respectivamente, para a estimativa geoestatística de um ponto não amostrado, com base nos pontos vizinhos próximos.

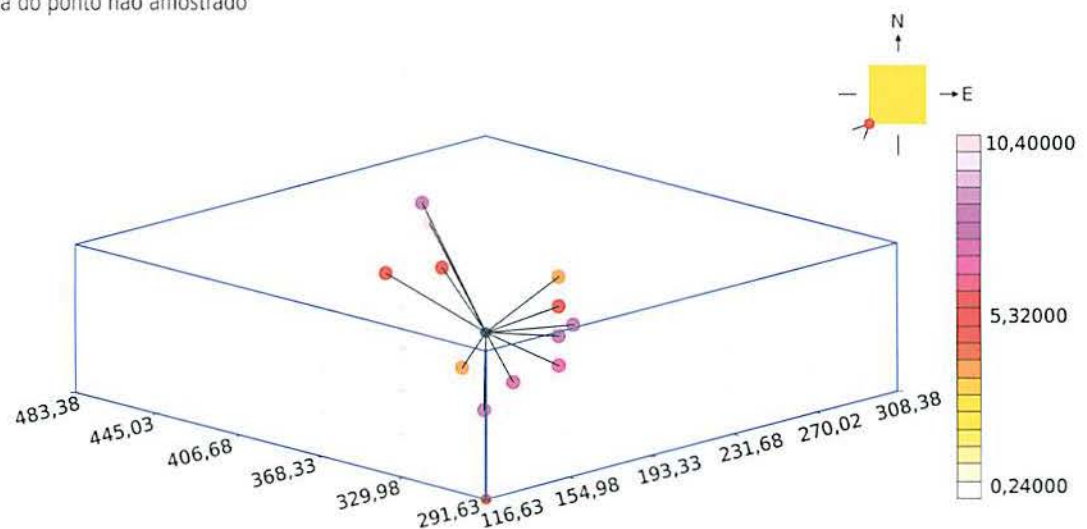


Fig. 3.3 Localização de pontos vizinhos próximos para interpolação do ponto não amostrado (dados 3D)

3.1 TRANSFORMAÇÃO DE DADOS

As variáveis regionalizadas podem ser *contínuas* ou *discretas* (Fig. 3.4). As variáveis contínuas podem apresentar comportamentos distintos revelados pela forma do histograma. Se a distribuição tiver assimetria positiva, há necessidade de transformação dos dados para evitar a influência dos poucos valores altos na estimativa de pontos da vizinhança, caracterizada por baixos valores.

Transformações de dados são, em diversas circunstâncias, necessárias para a estimativa geoestatística e, aqui, serão analisadas as principais, como a *gaussiana*, a *logarítmica* e a *indicadora*, conhecida também como *indicativa* e *indicatriz*.

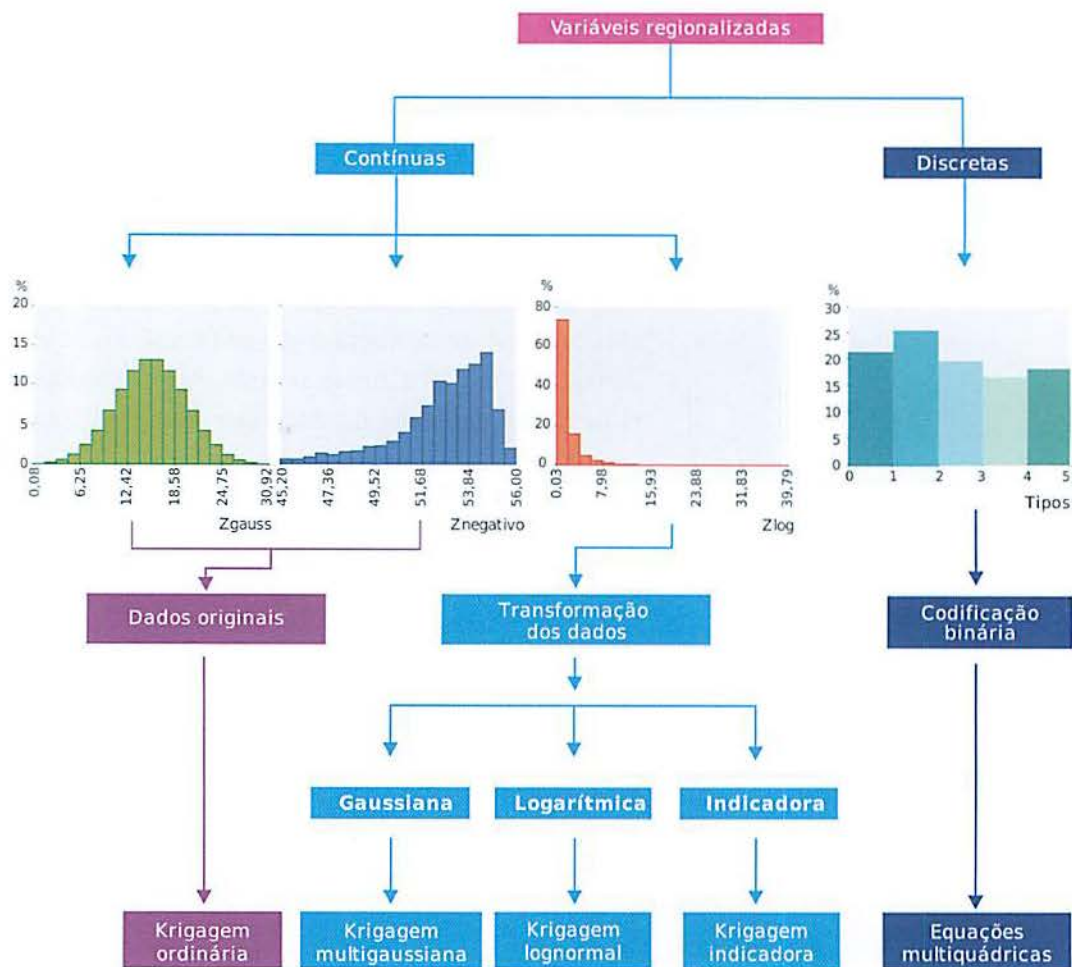


Fig. 3.4 Esquema ilustrando o processo de estimativa geoestatística ou interpolação de variáveis regionalizadas

As estimativas geoestatísticas para os dados transformados são obtidas por meio das krigagens multigaussiana, lognormal e indicadora.

Para dados com distribuição normal ou que apresentem assimetria negativa, não há necessidade de transformação dos dados, e a krigagem ordinária é aplicada diretamente sobre os dados originais.

Para as variáveis regionalizadas discretas, há necessidade de se fazer a codificação binária, e cada tipo que compõe a variável discreta é interpolado usando as equações multiquádricas, conforme proposta de Yamamoto et al. (2012). Não é usada a krigagem indicadora, por causa da necessidade de um variograma para cada tipo da variável discreta.

Mesmo que seja possível, quando houver grande quantidade de informação os variogramas não serão iguais entre si, em termos de efeito pepita, patamar e amplitude. Por isso, cada tipo sendo estimado por um variograma diferente resultará em valores cuja soma não será, necessariamente, igual a 1, condição essencial quando se estima probabilidades.

Dessa forma, a solução é a obtenção de um variograma único, tal como se faz no processo da krigagem da variável indicadora da mediana. Mas isso é impossível no caso de variáveis discretas, pois elas estão decompostas em k tipos.

A transformação é feita por meio de uma função matemática que atribui para cada valor x um novo valor $f(x)$ (Koch; Link, 1971, p. 231):

$$y = f(x)$$

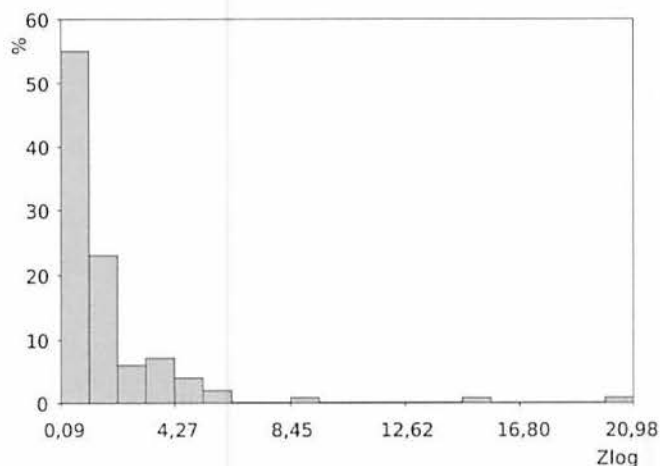


Fig. 3.5 Distribuição lognormal para teste de funções de transformação de dados

A transformação de dados pode ser feita por meio de funções *lineares* e *não lineares*. Todas essas transformações alteram a média e a variância da distribuição original, mas a transformação tem por objetivo a mudança da forma da distribuição de frequência (Koch; Link, 1971, p. 231) e, nesse sentido, deve-se analisar a mudança da forma da distribuição de frequência conforme a transformação aplicada.

Uma amostra (Arquivo 13, Anexo B – Figs. 2.21 e 2.24) composta por 100 pontos de dados (Fig. 3.5) foi escolhida para ilustrar as diferentes transformações possíveis, quais sejam: gaussiana, logarítmica e indicadora. As estatísticas para esse conjunto são $\bar{X} = 1,832$, $S = 2,816$ e $CV = 1,538$, as quais caracterizam uma distribuição tipicamente lognormal.

3.1.1 Transformada gaussiana

A transformada gaussiana é baseada na curva teórica da distribuição de Gauss, também conhecida como distribuição normal.

Os valores da variável de interesse a serem transformados são classificados em ordem crescente para obtenção das classes: o 1º ponto pertence à 1ª classe ($r(x_1) = 1$) e o n -ésimo ponto pertence à n -ésima classe ($r(x_n) = n$).

As proporções dessas classes podem ser calculadas dividindo-se cada classe pelo número total de observações n ou, então, dividindo-se por $(n + 1)$, conforme sugerido por Journel e Huijbregts (1978). Dessa divisão se obtêm os quantis da distribuição da variável de interesse. Tomando a função gaussiana inversa desses quantis se calculam os escores da distribuição normal padrão (Journel; Huijbregts, 1978, p. 479):

$$y(x_i) = G^{-1} \left(\frac{r(x_i)}{n+1} \right) \quad (3.1)$$

em que $G^{-1}(\cdot)$ é a função gaussiana inversa que fornece o escore da distribuição normal padrão para o quantil $\left(\frac{r(x_i)}{n+1} \right)$.

A Fig. 3.6 ilustra graficamente o processo da transformada gaussiana. Estão indicados três pontos da distribuição da variável de interesse, correspondentes a 25%, 50% e 75% da distribuição de frequências. Por exemplo, para o 1º quartil (25% da distribuição), o valor de Zlog é 0,438, que corresponde ao escore $-0,682$, ou seja, 25% na distribuição normal acumulada.

A função transformada gaussiana para a amostra Arquivo 13, Anexo B, está ilustrada na Fig. 3.7A e os resultados da transformada gaussiana, na Fig. 3.7B. Os escores da distribuição normal, nesse caso, variam de $-2,33$ a $2,33$, pois dependem do número de pontos de dados.

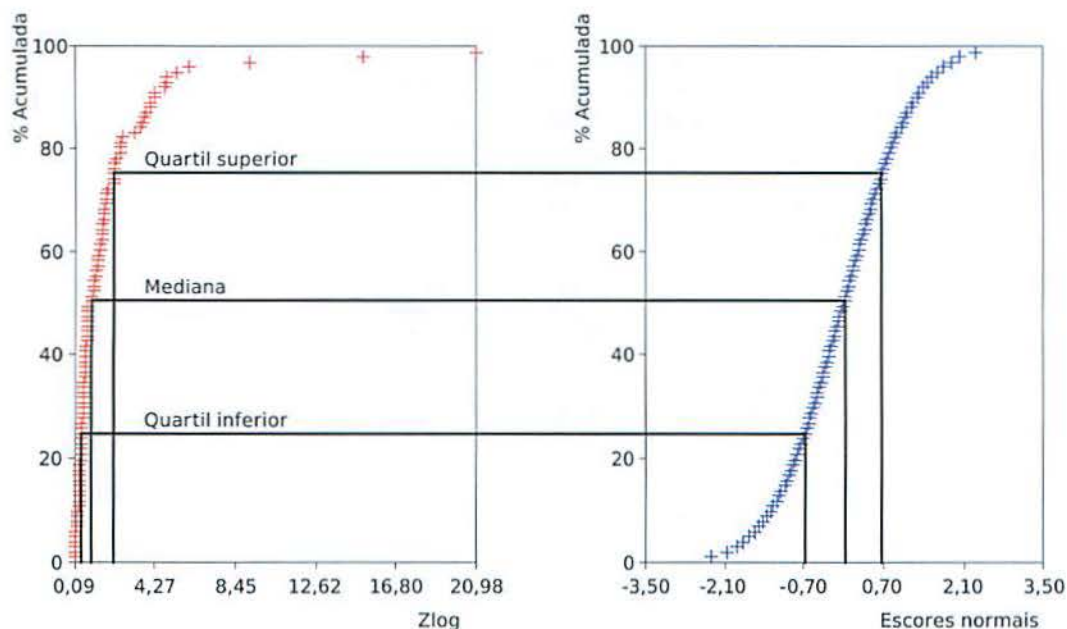


Fig. 3.6 Função transformada gaussiana da variável Zlog para os escores da distribuição normal acumulada

O histograma é perfeitamente simétrico, com média zero, mas a variância é igual a 0,923, e não a 1, como esperado. Isso acontece porque, na Eq. 3.1, a função gaussiana inversa tem como argumento a razão entre a classe e o número de pontos de dados. Quando esse número é pequeno, a razão não é suficientemente pequena para alcançar a cauda inferior da distribuição de Gauss e, assim, a última classe não fornece uma razão muito próxima de 1. Teoricamente, a distribuição de Gauss vai de $-\infty$ a $+\infty$, mas, em termos práticos, o intervalo de trabalho permanece entre $-2,50$ e $+2,50$. Alguns exemplos de média e variância e intervalos dos escores da distribuição normal são dados na Tab. 3.1. Como se pode verificar na tabela, a média é 0, mas a variância tende a 1 com o aumento do número de pontos de dados.

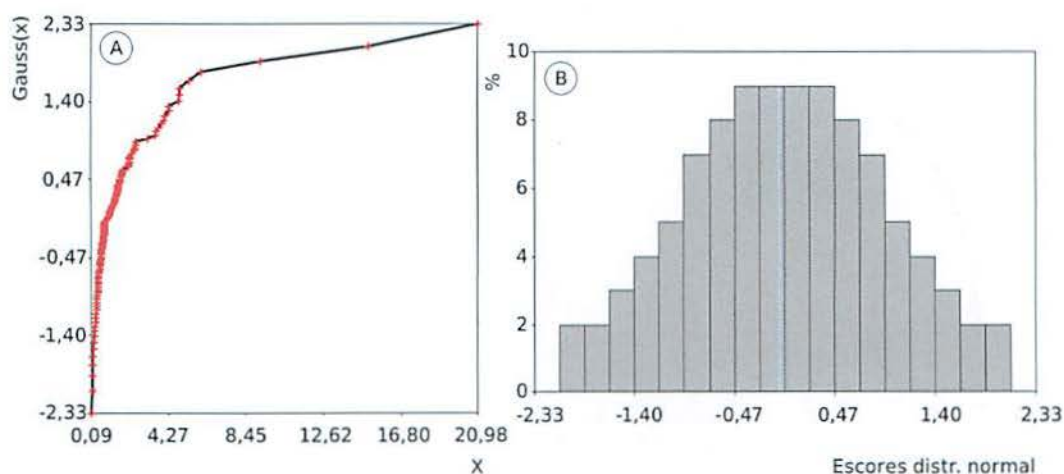


Fig. 3.7 A) Relação entre os escores da distribuição normal e os valores originais e B) histograma dos escores da distribuição normal

TAB. 3.1 Média, variância e intervalos de variação dos escores da distribuição normal

N	Média	Variância	$y(x_1)$	$y(x_N)$
2.000	0	0,99337	-3,29067	3,29067
4.000	0	0,99636	-3,48082	3,48082
6.000	0	0,99744	-3,58796	3,58796
8.000	0	0,99801	-3,66229	3,66229
10.000	0	0,99837	-3,71904	3,71904
12.000	0	0,99861	-3,76484	3,76484
14.000	0	0,99879	-3,80319	3,80319
16.000	0	0,99892	-3,83612	3,83612
18.000	0	0,99903	-3,86497	3,86497
20.000	0	0,99912	-3,89060	3,89060

3.1.2 Transformada logarítmica

A transformada logarítmica é obtida extraindo-se o logaritmo natural do valor da variável original:

$$y = \ln(x) \quad (3.2)$$

Se a distribuição é lognormal, a transformada logarítmica garante que a distribuição dos valores transformados seja normal. Mas, na maioria das vezes, apesar do coeficiente de variação ser superior a 1,254, nem sempre a transformada logarítmica garante uma distribuição perfeitamente normal para os dados transformados.

Os resultados da transformada logarítmica encontram-se na Fig. 3.8, na qual se pode verificar que o histograma apresenta alguma simetria. As barras do histograma são irregulares por causa da ausência de valores intermediários, como se pode observar na Fig. 3.8A.

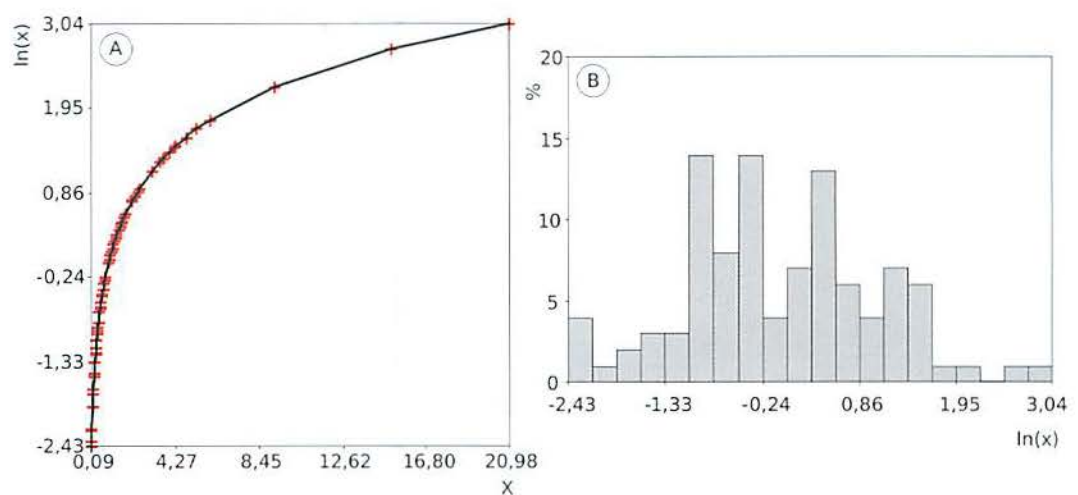


Fig. 3.8 A) Relação entre o logaritmo natural e os valores originais e B) histograma dos logaritmos dos valores originais

As estatísticas são $\bar{X} = -0,018$, $S = 1,090$ e $CV = -58,941$. Os resultados mostram que a transformada produziu uma distribuição próxima da distribuição normal, pelo menos em termos de estatísticas (média e desvio padrão).

Uma alternativa à Eq. 3.2 foi proposta por Yamamoto e Furuie (2010, p. 6), conforme segue:

$$y = \ln \left(\frac{x}{X_{50}} \right) \quad (3.3)$$

Essa transformação não altera a forma da distribuição, mas provoca uma translação dos valores em relação à mediana e, conseqüentemente, das estatísticas associadas. A vantagem dessa alternativa está na simetria da distribuição dos valores transformados, ou seja, 50% menores que a mediana e 50% maiores que a mediana.

3.1.3 Transformada indicadora

Uma variável aleatória contínua pode ser discretizada em relação a um valor de referência, como teor de corte/cutoff, da seguinte forma (Journel, 1983, p. 447):

$$I(x, zc) = \begin{cases} 0, & \text{se } Z(x) > zc \\ 1, & \text{se } Z(x) \leq zc \end{cases} \quad (3.4)$$

Essa transformação não linear resulta na função indicadora mostrada na Fig. 3.9.

Como se verá adiante, para os fins da krigagem indicadora há necessidade de se definir vários teores de corte dentro do intervalo de variação da variável de interesse. Assim, pode-se dividir a distribuição em termos de quartis, decis ou quantis.

Para cada teor de corte zc , têm-se a média ou a proporção de valores menores ou iguais a zc e a variância associada:

$$P_{zc} = E[I(x, zc)] \quad (3.5)$$

$$\begin{aligned} \text{Var}[I(x, zc)] &= E[I^2(x, zc)] - (E[I(x, zc)])^2 \\ &= P_{zc} - P_{zc}^2 = P_{zc}(1 - P_{zc}) \end{aligned} \quad (3.6)$$

Deve-se observar que a variável indicadora segue uma distribuição de Bernoulli com média P_{zc} e variância $P_{zc}(1 - P_{zc})$. O nome da distribuição se deve ao matemático suíço Jakob Bernoulli (1654-1705).

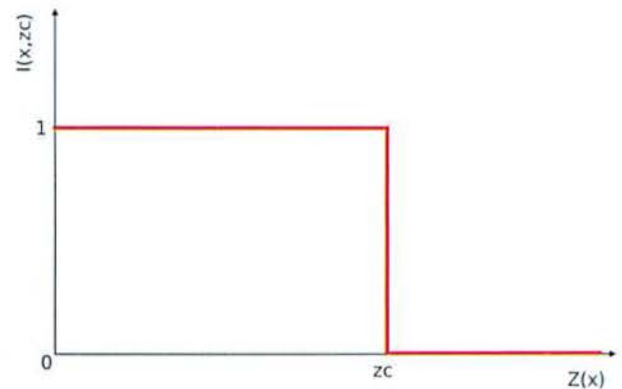


Fig. 3.9 Gráfico da função indicadora para um teor de corte zc

3.1.4 Codificação binária de variáveis categóricas

As variáveis categóricas podem ser medidas em escala nominal ou ordinal (Stevens, 1946) e, qualquer que seja a escala, há um número discreto de tipos. Seja k o número de tipos de uma variável categórica. A codificação binária (Soares, 2006, p. 133) é feita da seguinte forma:

$$I(x,k) = \begin{cases} 0, & \text{se } Z(x) \neq \text{tipo } k \\ 1, & \text{se } Z(x) = \text{tipo } k \end{cases} \quad (3.7)$$

TAB. 3.2 Exemplo de codificação binária de uma variável categórica composta por cinco tipos

Ponto $Z(x)$	A	B	C	D	E
1	0	1	0	0	0
2	1	0	0	0	0
3	0	0	0	1	0
4	0	0	0	0	1
5	0	0	1	0	0

A Tab. 3.2 apresenta um exemplo de codificação binária para uma variável categórica composta por cinco tipos: A, B, C, D e E.

No ponto 1, o tipo é B, então a coluna correspondente a esse tipo recebe o 1 e, as demais, o zero; no ponto 2, o tipo é A, então essa coluna recebe o 1 e, as demais, o zero, e assim por diante. É importante verificar que a função indicadora resultante é mutuamente exclusiva e completa $\sum_{k=1}^K I(x,k) = 1$.

Essa codificação binária foi proposta originalmente por Koike e Matsuda (2005). Soares (2006), Teng e Koike (2007) e Leuangthong, Khan e Deutsch (2008) também propuseram a codificação binária para a interpolação de tipos de uma variável categórica em pontos não amostrados.

3.1.5 Correção de pesos negativos

A Geoestatística também se preocupa com a presença de pesos negativos na krigagem e, dessa forma, Rao e Journel (1997, p. 94) propuseram um algoritmo para sua eliminação. Esse algoritmo será doravante considerado para tal correção.

Após a solução do sistema de equações da krigagem (ver “Krigagem ordinária”, p. 67), ponderadores são analisados para a ocorrência de pesos negativos. Se for verificada a presença deles, encontra-se o maior peso negativo em módulo:

$$c = -\min(W_i, i = 1, n)$$

Essa constante é adicionada a todos os pesos que, em seguida, são normalizados para retornar à soma dos pesos igual a 1:

$$W_i^c = \frac{W_i + c}{\sum_{j=1}^n (W_j + c)} \quad \text{para } i = 1, n$$

Após a correção, o peso negativo, que corresponde ao maior peso em módulo, é eliminado, e os demais, preservados. Trata-se de um algoritmo bastante simples e funcional. Evidentemente, existem outros algoritmos disponíveis, tais como os propostos por Froidevaux (1993) e Deutsch (1996).

3.2 ESTIMATIVAS GEOESTATÍSTICAS

Como já esquematizado (Fig. 3.4), as estimativas geoestatísticas podem ser feitas diretamente sobre os dados originais, por modelagem linear, ou sobre os dados transformados, por modelagem não linear. Assim, a krigagem ordinária pode ser aplicada diretamente sobre os dados originais ou sobre os dados transformados no caso da krigagem multigaussiana, da krigagem lognormal e da krigagem indicadora.

3.2.1 Krigagem linear

O sistema de krigagem necessário para a determinação dos ponderadores, ou pesos, associados a cada um dos pontos estimadores baseia-se na ideia de que quanto maior a covariância entre uma amostra x_i , $i=1, 2, \dots, n$, e o local que está sendo estimado, x_0 , mais essa amostra deve contribuir para a estimativa.

Em um método puramente geométrico, como o do inverso do quadrado da distância, o peso entre a amostra x_i e x_0 também diminui à medida que a amostra fica mais distante, mas essas distâncias são euclidianas. No caso da estimativa por krigagem, as distâncias são baseadas na análise variográfica e, além desse relacionamento entre pontos estimadores e o ponto a ser estimado, há o relacionamento entre os pontos estimadores que vão fornecer informações sobre o agrupamento presente.

O sistema de krigagem leva em consideração, portanto, tanto a distância entre as amostras como o seu agrupamento.

A krigagem pode ser usada, como algoritmo estimador, para:

- a) a previsão do valor pontual de uma variável regionalizada em um determinado local dentro do campo geométrico; é um procedimento de interpolação exato que leva em consideração os valores observados na vizinhança próxima, o qual pode ser a base para cartografia automática por computador quando se dispõe de valores de uma variável regionalizada distribuídos em uma determinada área;
- b) o cálculo do valor médio de uma variável regionalizada para um volume maior que o suporte geométrico, como, por exemplo, no cálculo do teor médio de um bloco de cubagem de uma jazida com base em informações obtidas de testemunhos de sondagens.

Krigagem simples ou estacionária

Seja um local não amostrado x_0 e n valores obtidos em pontos adjacentes. Uma estimativa linear ponderada desse local pode ser escrita como (Journel, 1989, p. 10):

$$Z_{KS}^*(x_0) = m_0 + \sum_{i=1}^n \lambda_i [Z(x_i) - m_i]$$

em que $m_i = E[Z(x_i)]$ são as médias, as quais são assumidas como conhecidas, m_0 é a média no ponto x_0 e $\{\lambda_i, i=1, n\}$ são os pesos associados aos n dados. No caso de variáveis regionalizadas, a localidade não amostrada, bem como os pontos amostrados, faz parte de uma função aleatória. Sob a condição de estacionaridade de segunda ordem, a média e a variância de todos os locais são constantes, dependendo apenas das distâncias euclidianas que os separam:

$$E[Z(x)] = m$$

$$E[(Z(x) - m)(Z(x+h) - m)] = E[Z(x)Z(x+h) - m^2] = C(h)$$

Assim, o estimador da krigagem simples é calculado segundo (Olea, 1999, p. 10-15):

$$Z_{KS}^*(x_0) = m + \sum_{i=1}^n \lambda_i [Z(x_i) - m] \quad (3.8)$$



O problema consiste em determinar os pesos ótimos da krigagem simples da Eq. 3.8. Para sua solução, segundo Olea (1999, p. 12), define-se uma nova função aleatória, que é a diferença entre a função aleatória $Z(x)$ e sua média:

$$Y(x) = Z(x) - E[Z(x)]$$

em que $E[Y(x)] = 0$. Assim, a Eq. 3.8 faz a estimativa dos resíduos.

De acordo com esse autor, a covariância de $Z(x)$ é igual à covariância de $Y(x)$:

$$\text{Cov}(x_i, x_j) = \text{Cov}_Y(x_i, x_j) = E[Y(x_i)Y(x_j)]$$

E a variância do erro é igual a:

$$\sigma^2(x_o) = \text{Var}[Z_{KS}^*(x_o) - Z(x_o)]$$

que pode ser reescrita em termos dos resíduos:

$$\sigma^2(x_o) = \text{Var}\left[\sum_{i=1}^n \lambda_i Y(x_i) - Y(x_o)\right]$$

Fazendo $\lambda_o = -1$, essa expressão torna-se (Olea, 1999, p. 13):

$$\sigma^2(x_o) = \text{Var}\left[\sum_{i=0}^n \lambda_i Y(x_i)\right]$$

A variância de uma combinação linear pode ser desenvolvida, segundo esse autor, como:

$$\text{Var}\left[\sum_{i=0}^n \lambda_i Y(x_i)\right] = \sum_{i=0}^n \sum_{j=0}^n \lambda_i \lambda_j \{E[Y(x_i)Y(x_j)] - E[Y(x_i)]E(x_j)\}$$

Como $E[Y(x_i)] = E[Y(x_j)] = 0$, então a variância do erro torna-se, segundo ele:

$$\sigma^2(x_o) = \sum_{i=0}^n \sum_{j=0}^n \lambda_i \lambda_j \text{Cov}[Y(x_i), Y(x_j)]$$

Ainda de acordo com Olea (1999, p. 14), separando os termos em $i = 0$ e $j = 0$, tem-se a variância do erro:

$$\sigma^2(x_o) = \text{Cov}(x_o, x_o) - 2 \sum_{i=1}^n \lambda_i \text{Cov}(x_i, x_o) + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j \text{Cov}(x_i, x_j)$$

Minimizando a variância do erro, chega-se ao conjunto de ponderadores ótimos da krigagem simples.

Para encontrar o ponto de mínimo da função variância do erro, calculam-se as derivadas parciais em relação aos pesos λ_i e iguala-se a zero (Olea, 1999, p. 15):

$$\frac{d\sigma^2(x_o)}{d\lambda_i} = -2\text{Cov}(x_i, x_o) + 2 \sum_{j=1}^n \lambda_j \text{Cov}(x_i, x_j) \text{ para } i = 1, n$$

que resulta no sistema normal de equações:

$$\sum_{j=1}^n \lambda_j \text{Cov}(x_i, x_j) = \text{Cov}(x_i, x_o) \text{ para } i = 1, n$$

O sistema de equações pode ser escrito em termos matriciais, cuja resolução resulta nos ponderadores da krigagem simples:

$$\begin{bmatrix} C(x_1 - x_1) & C(x_1 - x_2) & \cdots & C(x_1 - x_n) \\ C(x_2 - x_1) & C(x_2 - x_2) & \cdots & C(x_2 - x_n) \\ \vdots & \vdots & \cdots & \vdots \\ C(x_n - x_1) & C(x_n - x_2) & \cdots & C(x_n - x_n) \end{bmatrix} \begin{bmatrix} \lambda \\ \lambda_2 \\ \vdots \\ \lambda_n \end{bmatrix} = \begin{bmatrix} C(x_1 - x_o) \\ C(x_2 - x_o) \\ \vdots \\ C(x_n - x_o) \end{bmatrix} \quad (3.9)$$

Exemplo de aplicação da krigagem simples

Como exemplo de aplicação da krigagem, considerar uma situação em que se têm quatro pontos com valores conhecidos e com eles se quer determinar o valor em um ponto x_o (Tab. 3.3), conforme Olea (1999, p. 18-20). Nesse caso, a média m é conhecida e igual a 110, e a função covariância é: $C(h) = 2.000 \exp^{-\frac{h}{250}}$.

O mapa de localização de pontos e a função covariância encontram-se na Fig. 3.10.

Sendo conhecidas as coordenadas de todos os pontos, as distâncias euclidianas entre os pontos e entre cada ponto e o ponto a ser estimado podem ser determinadas:

$$\begin{bmatrix} 0 & & & \\ 260,8 & 0 & & \\ 264,0 & 266,3 & 0 & \\ 364,0 & 366,7 & 110,4 & 0 \end{bmatrix} \text{ e } \begin{bmatrix} 197,2 \\ 219,3 \\ 70,7 \\ 180,0 \end{bmatrix}$$

TAB. 3.3 Pontos de dados com valores conhecidos e o ponto a ser estimado

ID	X	Y	Valor
1	10	20	40
2	30	280	130
3	250	130	90
4	360	120	160
x_o	180	120	?

Fonte: Olea (1999, p. 18).

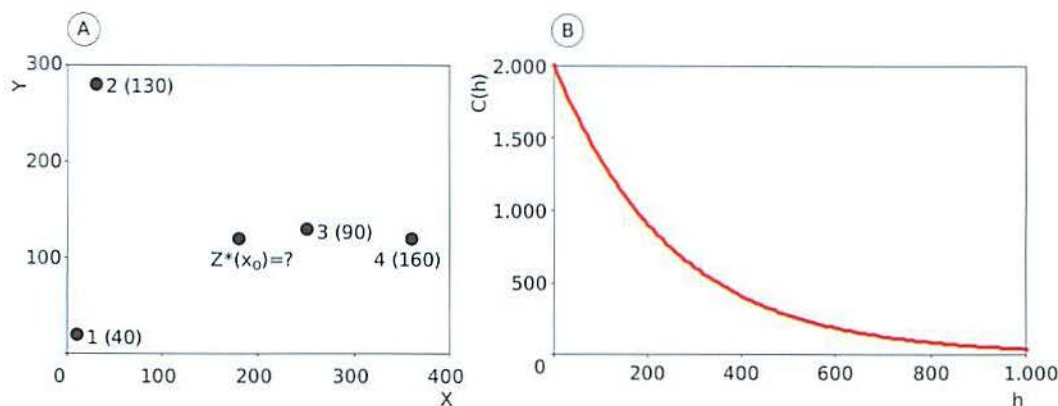


Fig. 3.10 A) Mapa de localização de pontos e B) gráfico da função covariância

Fonte: Olea (1999, p. 18).

Usando essas distâncias e sendo conhecida a covariância da variável regionalizada, a matriz de covariâncias pode ser obtida e, a partir daí, calculados os pesos associados aos estimadores:

$$\begin{bmatrix} 2.000 & & & \\ 704,8 & 2.000 & & \\ 695,6 & 689,4 & 2.000 & \\ 466,4 & 461,2 & 1.285,8 & 2.000 \end{bmatrix}^{-1} \begin{bmatrix} 908,7 \\ 831,8 \\ 1.507,2 \\ 973,6 \end{bmatrix} = \begin{bmatrix} 0,185 \\ 0,128 \\ 0,646 \\ -0,001 \end{bmatrix}$$

Finalmente, o valor no ponto X pode ser calculado, assim como a respectiva variância da estimativa:

$$Z_{KS}^*(180,120) = 110 + \begin{bmatrix} 40 - 110 \\ 130 - 110 \\ 90 - 110 \\ 160 - 110 \end{bmatrix}^T \begin{bmatrix} 0,185 \\ 0,128 \\ 0,646 \\ -0,001 \end{bmatrix} = 86,7$$

$$\sigma_{KS}^2(180,120) = 2.000 - \begin{bmatrix} 908,6 \\ 831,8 \\ 1.507,2 \\ 973,6 \end{bmatrix}^T \begin{bmatrix} 0,185 \\ 0,128 \\ 0,646 \\ -0,001 \end{bmatrix} = 752,9$$

Krigagem da média

A krigagem simples pressupõe que a média é conhecida e considerada constante em todo o domínio amostral. Mas nem sempre isso acontece e tampouco a média pode ser considerada constante. Assim, é preciso estimar a média em torno de uma região caracterizada por uma vizinhança com n pontos mais próximos $\{Z(x_i), i = 1, n\}$. A média pode, então, ser estimada para essa vizinhança (Wackernagel, 1995, p. 69-78):

$$m^* = \sum_{i=1}^n \lambda_i^{KM} Z(x_i) \quad (3.10)$$

e assumindo que a média existe em toda a região e igual a:

$$E[Z(x)] = m$$

Para evitar o viés sistemático, o erro de estimativa $(m^* - m)$ deve ser, em média, igual a zero:

$$E[m^* - m] = 0$$

Desenvolvendo essa expressão chega-se à seguinte condição de não viés:

$$\sum_{i=1}^n \lambda_i^{KM} = 1 \quad (3.11)$$

A variância do erro de estimativa pode ser desenvolvida em termos da função covariância:

$$Var[m^* - m] = \sum_{i=1}^n \sum_{j=1}^n \lambda_i^{KM} \lambda_j^{KM} C(x_i - x_j) \quad (3.12)$$

Segundo Wackernagel (1995, p. 71-78), para encontrar os ponderadores ótimos, deve-se minimizar a variância do erro de estimativa (Eq. 3.12) sujeita à condição de não viés (Eq. 3.11), o que resulta em uma função objetivo contendo o multiplicador de Lagrange, μ . Para resolver o problema de otimização, calculam-se as derivadas parciais da função objetivo, as quais são igualadas a zero, o que leva a um sistema de $n + 1$ equações, ou equações de krigagem da média:

$$\begin{cases} \sum_{j=1}^n \lambda_j^{KM} C(x_i - x_j) - \mu_{KM} = 0 \text{ para } i = 1, n \\ \sum_{j=1}^n \lambda_j^{KM} = 1 \end{cases} \quad (3.13)$$

De acordo com Wackernagel (1995, p. 72), a variância de estimativa da krigagem da média é igual ao próprio multiplicador de Lagrange:

$$\sigma_{KM}^2 = \mu_{KM}$$

Desse modo, no lugar de utilizar a média conhecida e constante, pode-se substituir na Eq. 3.8 a média estimada (Wackernagel, 1995, p. 77):

$$Z_{KSM}^*(x_0) = \sum_{i=1}^n \lambda_i^{KM} Z(x_i) + \sum_{i=1}^n \lambda_i^{KS} \left[Z(x_i) - \sum_{i=1}^n \lambda_i^{KM} Z(x_0) \right]$$

que pode ser rearranjada, segundo esse autor, como:

$$Z_{KSM}^*(x_0) = \sum_{i=1}^n \left[\lambda_i^{KS} + \left(1 - \sum_{j=1}^n \lambda_j^{KS} \right) \lambda_i^{KM} \right] Z(x_i)$$

De acordo com ele, o termo entre colchetes é o peso da krigagem ordinária, e o termo entre parênteses é chamado de peso da média. Portanto, a krigagem ordinária nada mais é que a krigagem simples com a média calculada localmente, por meio da krigagem da média.

Exemplo de aplicação da krigagem da média

Como exemplo de aplicação da krigagem da média, considerar a amostra do conjunto normal (Arquivo 11, Anexo B – Figs. 2.20 e 2.23), cujos resultados encontram-se no mapa-imagem da Fig. 3.11.

Esse exemplo mostra que a média varia de ponto a ponto, dependendo da vizinhança próxima. Assim, é importante que se considere a média local como calculada pela krigagem da média, o que leva ao uso do método da krigagem ordinária, como será visto na seção seguinte. Antes de prosseguir, contudo, é interessante mostrar como se faz a krigagem da média. Para isso, deve-se considerar o ponto de coordenadas ($x = 23,75$; $y = 31,25$), para o qual foram encontrados quatro pontos vizinhos pelo critério dos quadrantes (um ponto mais próximo por quadrante), como mostra a Fig. 3.12 e a Tab. 3.4.

TAB. 3.4 Pontos de dados vizinhos ao ponto a ser estimado ($x = 23,75$; $y = 31,25$)

Ponto	X	Y	Valor
1	26,50	36,50	21,807
2	20,50	33,50	18,697
3	13,50	29,50	19,320
4	24,50	27,50	18,627



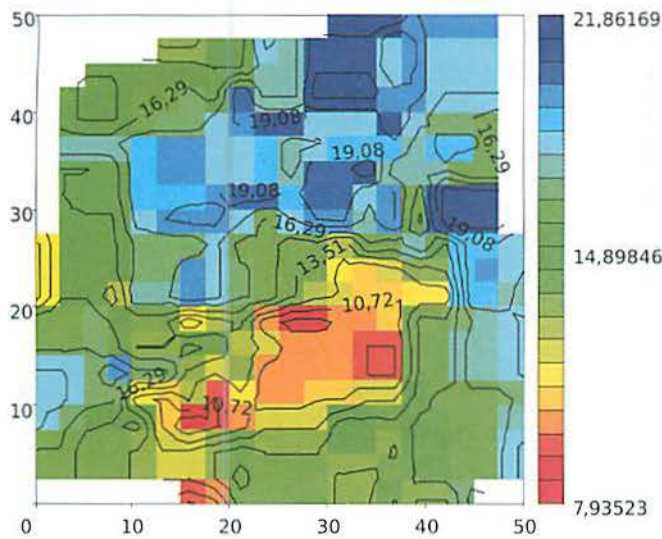


Fig. 3.11 Distribuição das médias calculadas pela krigagem da média para o conjunto normal (Arquivo 11, Anexo B)

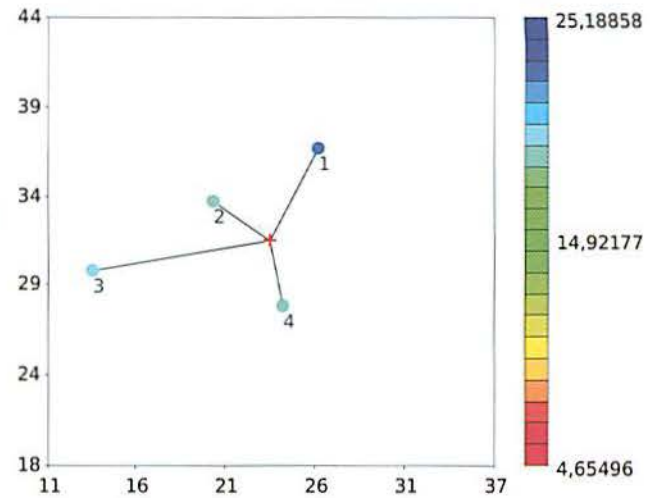


Fig. 3.12 Localização dos vizinhos próximos ao ponto de coordenadas ($x = 23,75$; $y = 31,25$)

O modelo de variograma encontrado (Fig. 2.23B, Eq. 2.2) é:

$$\begin{cases} \gamma(h) = 19,8 \left[1,5 \frac{h}{14,16} - 0,5 \left(\frac{h}{14,16} \right)^3 \right] & \text{para } h < 14,16 \\ \gamma(h) = 19,8 & \text{para } h \geq 14,16 \end{cases}$$

Desse modo, o modelo da função covariância fica:

$$\begin{cases} C(h) = 19,8 - 19,8 \left[1,5 \frac{h}{14,16} - 0,5 \left(\frac{h}{14,16} \right)^3 \right] & \text{para } h < 14,16 \\ C(h) = 0 & \text{para } h \geq 14,16 \end{cases}$$

Com isso, pode-se montar o sistema de equações de krigagem da média (Eq. 3.13), como segue:

$$\begin{bmatrix} 19,8 & 6,782 & 0 & 3,195 & 1 \\ 6,782 & 19,8 & 4,717 & 5,983 & 1 \\ 0 & 4,717 & 19,8 & 1,223 & 1 \\ 3,195 & 5,983 & 1,223 & 19,8 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1^{KM} \\ \lambda_2^{KM} \\ \lambda_3^{KM} \\ \lambda_4^{KM} \\ -\mu_{KM} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

Resolvendo esse sistema, obtêm-se os ponderadores da krigagem da média:

$$\lambda_1^{KM} = 0,28920; \lambda_2^{KM} = 0,11247; \lambda_3^{KM} = 0,32787; \lambda_4^{KM} = 0,27047; \mu_{KM} = 7,35303$$

Substituindo os ponderadores na Eq. 3.10, tem-se a média estimada em torno do ponto de coordenadas ($x = 23,75$; $y = 31,25$):

$$m^* = 0,28920 \times 21,807 + 0,11247 \times 18,697 + 0,32787 \times 19,320 + 0,27047 \times 18,627 = 19,782$$

Essa média será válida desde que se mantenham os mesmos pontos encontrados na vizinhança (Tab. 3.4). A variância da krigagem da média é o próprio multiplicador de Lagrange, que, nesse caso, será sempre positivo. Esse sistema deve ser resolvido em termos da função covariância.

Krigagem ordinária

A krigagem ordinária nada mais é que a krigagem simples com a média local calculada pela krigagem da média, como descrito na seção anterior. É o método mais utilizado, pela simplicidade e resultados que proporciona. A krigagem ordinária é um método local de estimativa e, dessa forma, a estimativa em um ponto não amostrado resulta da combinação linear dos valores encontrados na vizinhança próxima.

O estimador da krigagem ordinária é:

$$Z_{KO}^*(x_o) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (3.14)$$

Os pesos ótimos são calculados sob duas condições de restrição (Journel; Huijbregts, 1978, p. 305): A) que o estimador não seja enviesado; e B) que a variância de estimativa seja mínima.

De acordo com Journel e Huijbregts (1978, p. 305), o não viés da estimativa é obtido quando o erro, diferença entre o valor real e o valor calculado, é igual a zero, em média:

$$E[Z_{OK}^*(x_o) - Z(x_o)] = 0 \quad (3.15)$$

Desenvolvendo a expressão da esperança do erro, chega-se à condição de não viés:

$$\sum_{i=1}^n \lambda_i = 1 \quad (3.16)$$

A variância de estimativa ou a variância do erro de estimativa é calculada como:

$$\sigma_E^2 = \text{Var}[Z(x_o) - Z_{KO}^*(x_o)] \quad (3.17)$$

As Eq. 3.15 e 3.17 envolvem uma grandeza desconhecida, o valor $Z(x_o)$, que é o valor real em um ponto não amostrado. Segundo Isaaks e Srivastava (1989, p. 280), a solução para esse problema é baseado em um modelo probabilístico, de tal forma que os valores desconhecidos são considerados realizações de um processo aleatório, assim como são os valores da variável aleatória $\{Z(x_i), i = 1, n\}$.

A minimização da variância do erro de estimativa parte do desenvolvimento da Eq. 3.17, conforme:

$$\sigma_E^2 = E[(Z(x_o) - Z_{KO}^*(x_o))^2] - (E[Z(x_o) - Z_{KO}^*(x_o)])^2$$

Expandindo-se cada termo do lado direito dessa equação, chega-se à seguinte expressão (Journel; Huijbregts, 1978, p. 305):

$$\sigma_E^2 = C(0) - 2 \sum_i \lambda_i C(x_o - x_i) + \sum_i \sum_j \lambda_i \lambda_j C(x_i - x_j)$$



Essa é a expressão da variância do erro de estimativa, em termos da função covariância que é conhecida. Para encontrar os pesos ótimos, deve-se minimizar a variância do erro de estimativa sob a condição de não viés ou de restrição (Eq. 3.16). Para encontrar o ponto de mínimo, utiliza-se a técnica dos multiplicadores de Lagrange, da qual se obtém a lagrangiana, isto é, a função das coordenadas generalizadas (Yamamoto, 2001a, p. 133):

$$L(\lambda_1, \lambda_2, \dots, \lambda_n, \mu) = C(0) - 2 \sum_i \lambda_i C(x_o - x_i) + \sum_i \sum_j \lambda_i \lambda_j C(x_i - x_j) - 2\mu \left(\sum_j \lambda_j - 1 \right)$$

em que $L(\lambda_1, \lambda_2, \dots, \lambda_n, \mu)$ é a lagrangiana e μ é o multiplicador de Lagrange.

Fazendo cada uma das derivadas parciais da lagrangiana iguais a zero e também derivando a lagrangiana em relação a μ , chega-se ao sistema de equações de krigagem ordinária (Journel; Huijbregts, 1978, p. 306):

$$\begin{cases} \sum_{j=1}^n \lambda_j C(x_i - x_j) - \mu = C(x_i - x_o) \text{ para } i = 1, n \\ \sum_{j=1}^n \lambda_j = 1 \end{cases} \quad (3.18)$$

ou, em forma matricial:

$$\begin{bmatrix} C(x_1 - x_1) & C(x_1 - x_2) & \dots & C(x_1 - x_n) & 1 \\ C(x_2 - x_1) & C(x_2 - x_2) & \dots & C(x_2 - x_n) & 1 \\ \vdots & \vdots & \dots & \vdots & \vdots \\ C(x_n - x_1) & C(x_n - x_2) & \dots & C(x_n - x_n) & 1 \\ 1 & 1 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ -\mu \end{bmatrix} = \begin{bmatrix} C(x_o - x_1) \\ C(x_o - x_2) \\ \vdots \\ C(x_o - x_n) \\ 1 \end{bmatrix}$$

A variância de krigagem, em termos da função covariância, é igual a (Journel; Huijbregts, 1978, p. 306):

$$\sigma_{KO}^* = C(0) - \sum_{i=1}^n \lambda_i C(x_o - x_i) + \mu \quad (3.19)$$

O sistema de equações da krigagem ordinária pode ser escrito também em termos da função variograma:

$$\begin{cases} \sum_{j=1}^n \lambda_j \gamma(x_i - x_j) + \mu = \gamma(x_o - x_i) \text{ para } i = 1, n \\ \sum_{j=1}^n \lambda_j = 1 \end{cases} \quad (3.20)$$

Nesse caso, a variância de krigagem fica igual a:

$$\sigma_{KO}^2 = \sum_{i=1}^n \lambda_i \gamma(x_o - x_i) + \mu \quad (3.21)$$

Os sistemas de equações de krigagem (Eqs. 3.18 e 3.20) permitem calcular os ponderadores para um ponto não amostrado na localização x_o , pela denominada krigagem pontual. Na mineração, porém, o interesse é determinar o teor de um bloco de cubagem, cujas dimensões são definidas para atender a uma demanda de produção, seja semanal, quinzenal, mensal etc.

A Geoestatística possibilita fazer uma estimativa média do bloco de cubagem por meio da discretização do bloco em pontos, que podem ser avaliados individualmente e depois compostos para o bloco ou então diretamente, calculando-se os vetores médios dos sistemas de equações (Eqs. 3.18 e 3.20). Essa alternativa da krigagem ordinária é denominada krigagem de bloco.

Exemplo de aplicação da krigagem ordinária

A seguir apresenta-se um exemplo de aplicação da krigagem ordinária, para o qual foi escolhido o Arquivo 11, Anexo B (Figs. 2.20 e 2.23).

O exemplo começa com um exercício passo a passo mostrando as diferenças entre a krigagem pontual e a krigagem de bloco. A Fig. 3.13 mostra, esquematicamente, a krigagem pontual para interpolação do teor na localização ($x = 28,75$; $y = 21,25$) e a krigagem de bloco para determinação do teor médio associado ao bloco de 2,5 por 2,5 centrado no ponto de coordenadas ($x = 28,75$; $y = 21,25$). O modelo de variograma válido para esse conjunto é descrito pela Eq. 2.3.

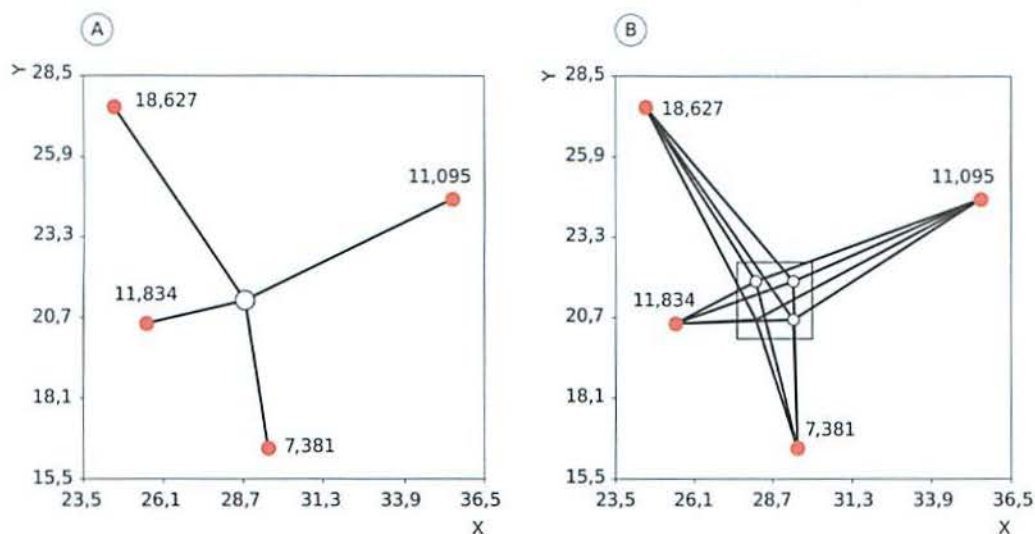


Fig. 3.13 A) Krigagem pontual para interpolação do ponto ($x = 28,75$; $y = 21,25$); e B) krigagem de bloco para cálculo do teor médio do bloco de 2,5 por 2,5 centrado no ponto de coordenadas ($x = 28,75$; $y = 21,25$)

Os vizinhos mais próximos ao ponto a ser interpolado encontram-se na Tab. 3.5 e no mapa de localização da Fig. 3.13A. O bloco de cubagem, nesse exemplo, foi discretizado em 2 por 2 sub-blocos (Fig. 3.13B). É importante levar em consideração os limites de discretização (Tab. 3.6), conforme Journel e Huijbregts (1978, p. 97).

Para efetuar o cálculo da krigagem ordinária, monta-se o sistema de equações de krigagem (Eq. 3.20), conforme:

$$\begin{bmatrix} 0 & 18,746 & 18,234 & 17,488 & 1 \\ 18,746 & 0 & 13,598 & 19,192 & 1 \\ 18,234 & 13,598 & 0 & 11,234 & 1 \\ 17,488 & 19,192 & 11,234 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \mu \end{bmatrix} = \begin{bmatrix} 14,247 \\ 14,347 \\ 6,867 \\ 9,699 \\ 1 \end{bmatrix} \quad (3.22)$$

TAB. 3.5 Pontos de dados vizinhos para estimativa da localização ($x = 28,75$; $y = 21,25$) por meio da krigagem ordinária pontual e de bloco

Ponto	X	Y	Valor
1	35,50	24,50	11,095
2	24,50	27,50	18,627
3	25,50	20,50	11,834
4	29,50	16,50	7,381

Resolvendo o sistema, os ponderadores da krigagem ordinária são obtidos:

$$\lambda_1 = 0,18386; \lambda_2 = 0,08361; \lambda_3 = 0,47761; \lambda_4 = 0,25492; \mu = -0,48669$$

Aplicando-se os ponderadores obtidos na Eq. 3.14, obtém-se a estimativa no ponto de coordenadas ($x = 28,75$; $y = 21,25$):

$$\begin{aligned} Z_{KO}^*(x_o) &= 0,18386 \times 11,095 + 0,08361 \times 18,627 \\ &\quad + 0,47761 \times 11,834 + 0,25492 \times 7,381 \\ &= 11,131 \end{aligned}$$

A variância de krigagem (Eq. 3.21) é igual a:

$$\sigma_{KO}^2 = 9,084$$

TAB. 3.6 Limites de discretização para cálculo dos blocos de cubagem por krigagem ordinária

Dimensão do domínio	Límite de discretização
1	10
2	6 x 6
3	4 x 4 x 4

Fonte: Journel e Huijbregts (1978, p. 97).

TAB. 3.7 Centros dos sub-blocos para estimativa do bloco (Fig. 3.13B)

Sub-bloco	X	Y
1	28,125	20,625
2	29,375	20,625
3	29,375	21,875
4	28,125	21,875

Para a krigagem ordinária de bloco, o vetor do lado direito do sistema de equações (Eq. 3.22) é substituído por um vetor médio considerando todos os sub-blocos localizados conforme as coordenadas da Tab. 3.7.

Para cada sub-bloco calcula-se o vetor contendo o valor da função variograma correspondente à distância entre o centro do sub-bloco e o ponto de dado.

$$\left(\begin{array}{c|c|c|c} \text{sub-bloco 1} & \text{sub-bloco 2} & \text{sub-bloco 3} & \text{sub-bloco 4} \\ \hline \begin{bmatrix} 15,458 \\ 14,655 \\ 5,449 \\ 8,833 \\ 1 \end{bmatrix} & \begin{bmatrix} 13,874 \\ 15,590 \\ 7,929 \\ 8,411 \\ 1 \end{bmatrix} & \begin{bmatrix} 12,945 \\ 14,174 \\ 8,381 \\ 10,735 \\ 1 \end{bmatrix} & \begin{bmatrix} 14,747 \\ 12,991 \\ 6,125 \\ 11,04 \\ 1 \end{bmatrix} \\ \hline \end{array} \right) \div 4 = \begin{array}{c} \text{vetor médio} \\ \hline \begin{bmatrix} 14,256 \\ 14,355 \\ 6,971 \\ 9,755 \\ 1 \end{bmatrix} \end{array}$$

O vetor médio é substituído no sistema de equações (Eq. 3.22), cuja resolução fornecerá os ponderadores da krigagem ordinária de bloco.

$$\begin{bmatrix} 0 & 18,746 & 18,234 & 17,488 & 1 \\ 18,746 & 0 & 13,598 & 19,192 & 1 \\ 18,234 & 13,598 & 0 & 11,234 & 1 \\ 17,488 & 19,192 & 11,234 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \lambda_3 \\ \lambda_4 \\ \mu \end{bmatrix} = \begin{bmatrix} 14,256 \\ 14,355 \\ 6,971 \\ 9,755 \\ 1 \end{bmatrix} \quad (3.23)$$

Os ponderadores da krigagem ordinária são:

$$\lambda_1 = 0,18528; \lambda_2 = 0,08634; \lambda_3 = 0,47277; \lambda_4 = 0,25561; \mu = -0,45299$$

Assim, a estimativa do bloco centrado em $(x = 28,75; e y = 21,25)$ é:

$$Z_{KO}^*(x_0) = 0,18528 \times 11,095 + 0,08634 \times 18,627 + 0,47277 \times 11,834 + 0,25561 \times 7,381 = 11,145$$

A variância de krigagem de bloco fica:

$$\sigma_{KO}^2 = 9,217$$

Nesse exemplo, os teores estimados e as variâncias de krigagem resultaram em valores muito próximos entre a krigagem pontual e a krigagem de bloco. Diferenças maiores seriam observadas para dados apresentando maior variabilidade e para blocos de dimensões maiores.

Os multiplicadores de Lagrange nos dois sistemas lineares (Eqs. 3.22 e 3.23) resultaram em valores negativos. Na Eq. 3.21, a expressão da variância de krigagem leva em consideração o multiplicador de Lagrange, que, sendo negativo, irá subtrair essa quantidade do somatório do produto função variância $\times$ peso da krigagem. Não existe simplificação possível em que a variância de krigagem resulte no multiplicador de Lagrange, exceto na krigagem da média, como já mostrado.

Prosseguindo os cálculos (Arquivo 11, Anexo B – Figs. 2.20 e 2.23), obtêm-se mapas com valores estimados e de incertezas, sendo essas representadas pelo desvio padrão de krigagem tanto para krigagem pontual (Fig. 3.14) como para krigagem de bloco (Fig. 3.15). Não há grandes diferenças entre os mapas estimados por krigagem pontual e krigagem de bloco, como já foi visto. Os mapas das incertezas mostram claramente que os valores mais baixos encontram-se sobre os pontos de amostragem.

A krigagem ordinária foi o primeiro método de estimativa a fornecer uma medida de incerteza pela variância de krigagem, daí o seu sucesso em aplicações em problemas de avaliação de recursos minerais.

Ao analisar as expressões da variância de krigagem (Eqs. 3.19 e 3.21), verifica-se que os valores dos dados não são considerados no cálculo e, portanto, são independentes desses valores.

Na verdade, a variância de krigagem ou de estimativa depende apenas da distribuição geométrica dos pontos e do modelo de variograma. Desse modo, pode haver dois arranjos de pontos vizinhos próximos em partes diferentes de um mesmo domínio espacial e apresentando valores de variância de krigagem iguais. Essa propriedade é denominada homocedasticidade ou variância constante.



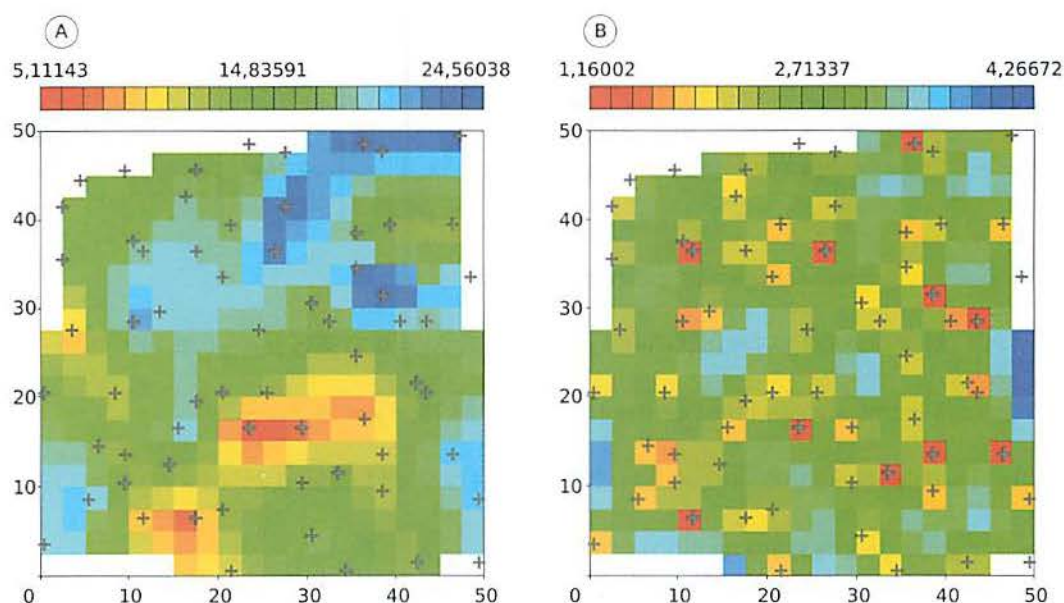


Fig. 3.14 A) Mapa de teores estimados por krigagem pontual e B) mapa do desvio padrão da krigagem. Parâmetros de interpolação: $DX = DY = 2,5$ e 1 ponto por quadrante

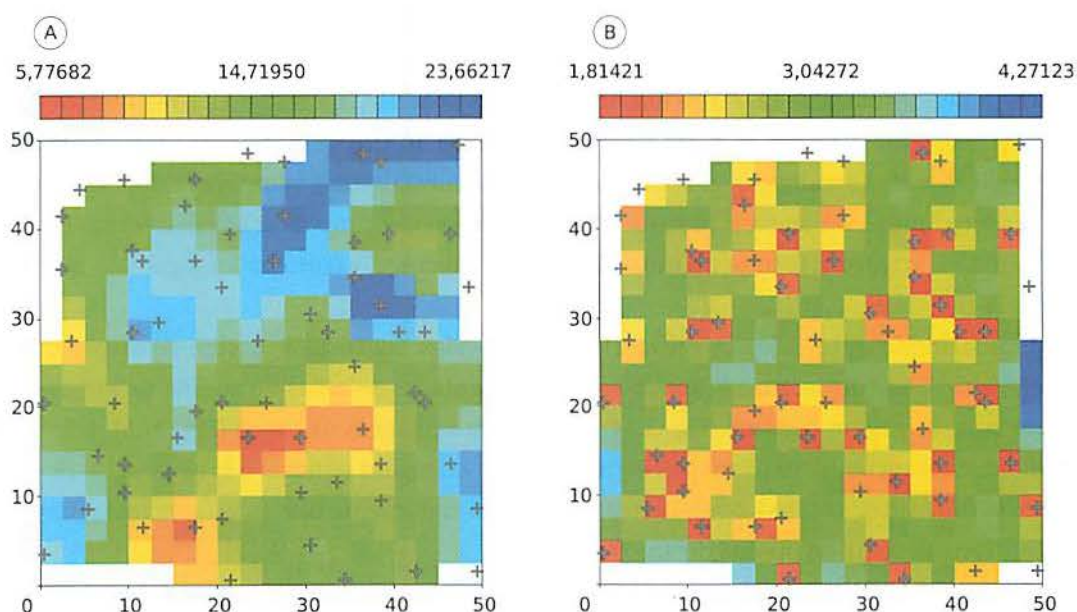


Fig. 3.15 A) Mapa de teores estimados por krigagem de bloco e B) mapa do desvio padrão da krigagem. Parâmetros de interpolação: $DX = DY = 2,5$ e 1 ponto por quadrante

A Fig. 3.16, proposta por Armstrong (1994, p. 306), mostra dois blocos em posições diferentes de um mesmo depósito. O teor médio é o mesmo nos dois blocos, mas a incerteza associada no bloco A deveria ser menor que no bloco B. Entretanto, a variância de krigagem é exatamente igual nas duas situações, haja vista ela ter sido calculada com o mesmo modelo de variograma e configurações idênticas de pontos de dados. Esse é o caráter homocedástico da variância de krigagem. Portanto, essa medida não reflete a incerteza

associada à estimativa, mas tão somente a configuração espacial dos pontos de dados para um mesmo modelo de variograma, de acordo com Journel e Rossi (1989, p. 738).

Antes de prosseguir, seria interessante entender como se apresenta a homocedasticidade, por meio da análise de diversos conjuntos de pontos usados na estimativa de pontos não amostrados.

Para esse fim, considerar a amostra do conjunto lognormal (Arquivo 12, Anexo B – Figs. 2.21 e 2.24), que tem como modelo de variograma (Eq. 2.3):

$$\begin{cases} \gamma(h) = 3,6 \left[1,5 \frac{h}{14,16} - 0,5 \left(\frac{h}{14,16} \right)^3 \right] & \text{para } h < 14,16 \\ \gamma(h) = 3,6 & \text{para } h \geq 14,16 \end{cases}$$

Os resultados da krigagem ordinária encontram-se na Fig. 3.17. A interpolação foi feita em uma malha regular bem fechada, com abertura $DX = DY = 0,5$, resultando em 10.000 pontos, dos quais 8.338 foram estimados por pertencerem à fronteira convexa. No mapa dos desvios padrão de interpolação, é possível verificar os pontos amostrais coincidindo com pontos de baixa incerteza.

Com base no mapa da Fig. 3.17B, extraíram-se pares de pontos que foram interpolados e que resultaram na mesma variância de krigagem (Fig. 3.18 – Tabs. 3.8 e 3.9) e outros em que a diferença entre as variâncias de krigagem foram inferiores a 0,000001 (Fig. 3.19 – Tabs. 3.10 e 3.11), ou seja, praticamente iguais, considerando a precisão do ajuste do variograma.

Na Fig. 3.18, as variâncias de krigagem para os pares (A-B e C-D) foram exatamente iguais, assim como os multiplicadores de Lagrange. Isso significa que os pontos amostrais são os mesmos para diferentes localizações dos pontos não amostrados.

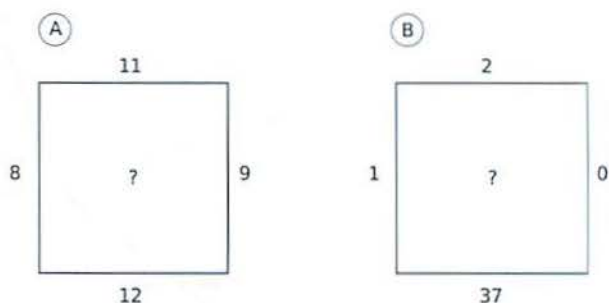


Fig. 3.16 Estimativa de blocos com a mesma configuração de pontos de dados: A) pequena incerteza; B) grande incerteza
Fonte: Armstrong (1994, p. 306).

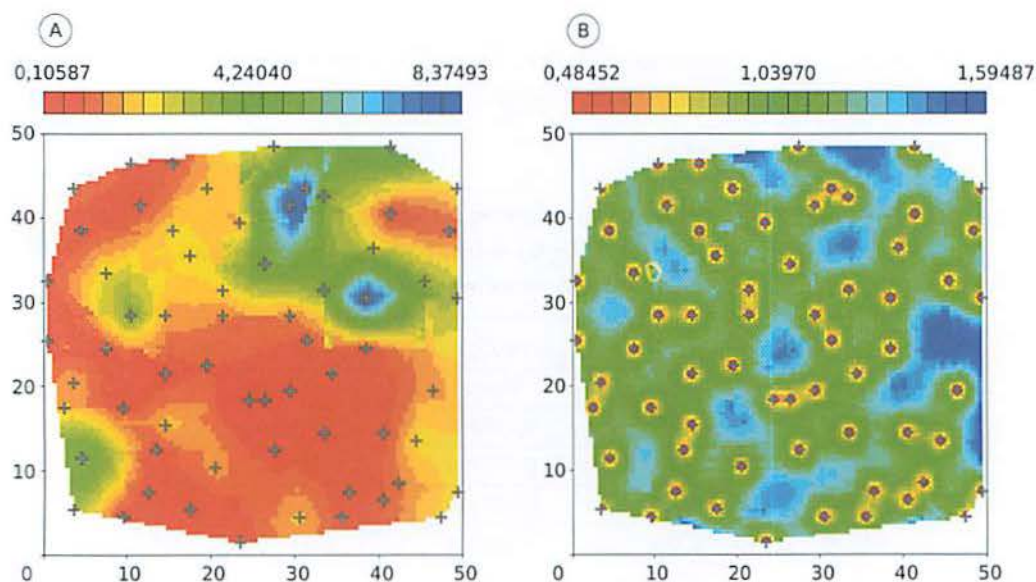


Fig. 3.17 A) Mapa de teores estimados por krigagem de bloco e B) mapa do desvio padrão da krigagem. Parâmetros de interpolação: $DX = DY = 0,5$ e 1 ponto por quadrante

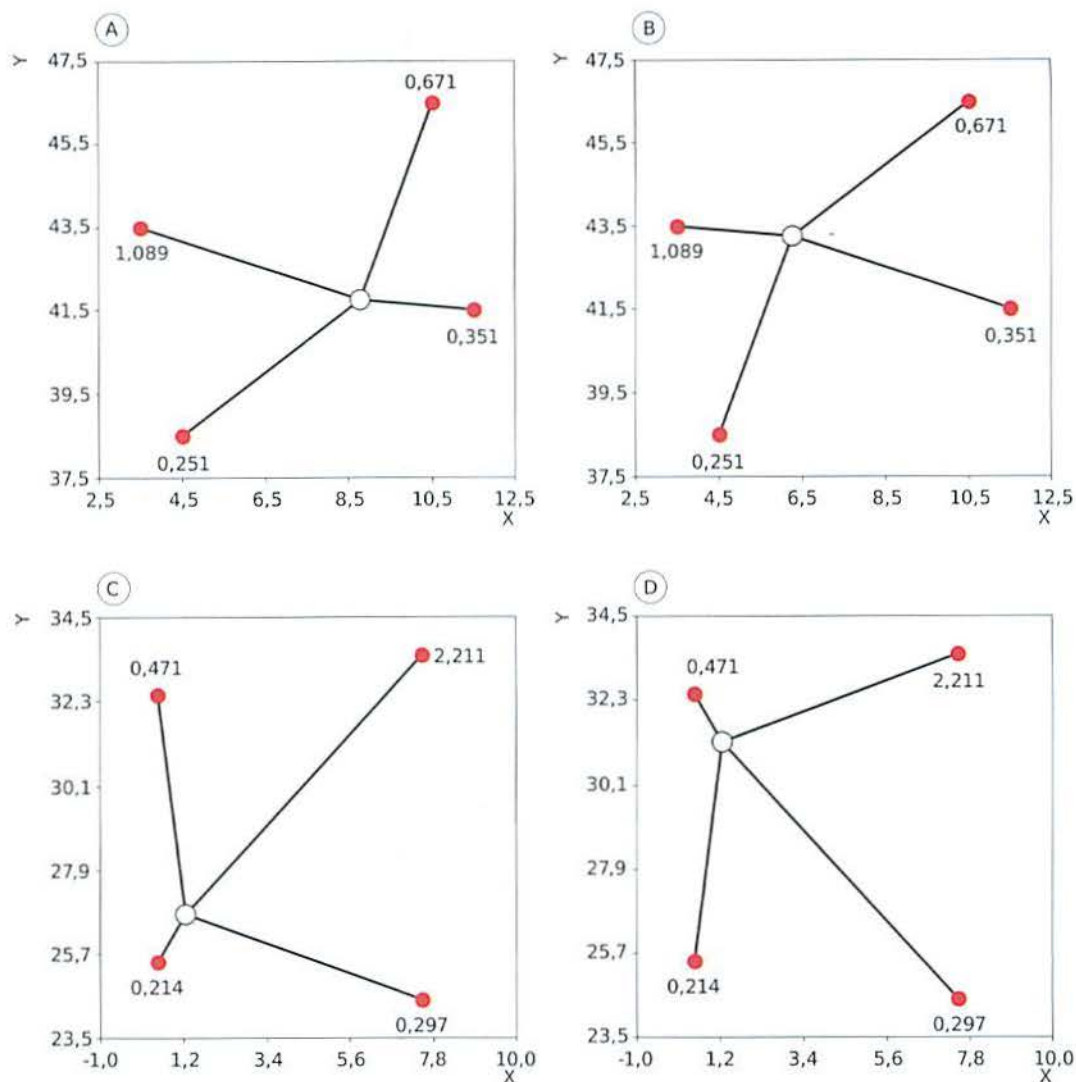


Fig. 3.18 Diferentes localizações dos pontos interpolados em relação aos pontos vizinhos (A-B e C-D), resultando em variâncias de krigagem e multiplicadores de Lagrange iguais

TAB. 3.8 Pontos de dados para estimativa por krigagem ordinária de $x_o = (8,75; 41,75)$ e $x_o = (6,25; 43,25)$, representados nas Figs. 3.18A e 3.18B, respectivamente

X	Y	Z(x)	$x_o = (8,75; 41,75)$	$x_o = (6,25; 43,25)$
			Pesos	Pesos
10,50	46,50	0,671	0,129344	0,186156
3,50	43,50	1,089	0,161796	0,522704
4,50	38,50	0,251	0,186156	0,129344
11,50	41,50	0,351	0,522704	0,161796
$Z_{KO}^*(x_o)$			0,493298	0,783538
σ_{KO}^2			1,327584	1,327584
μ			-0,140751	-0,140751

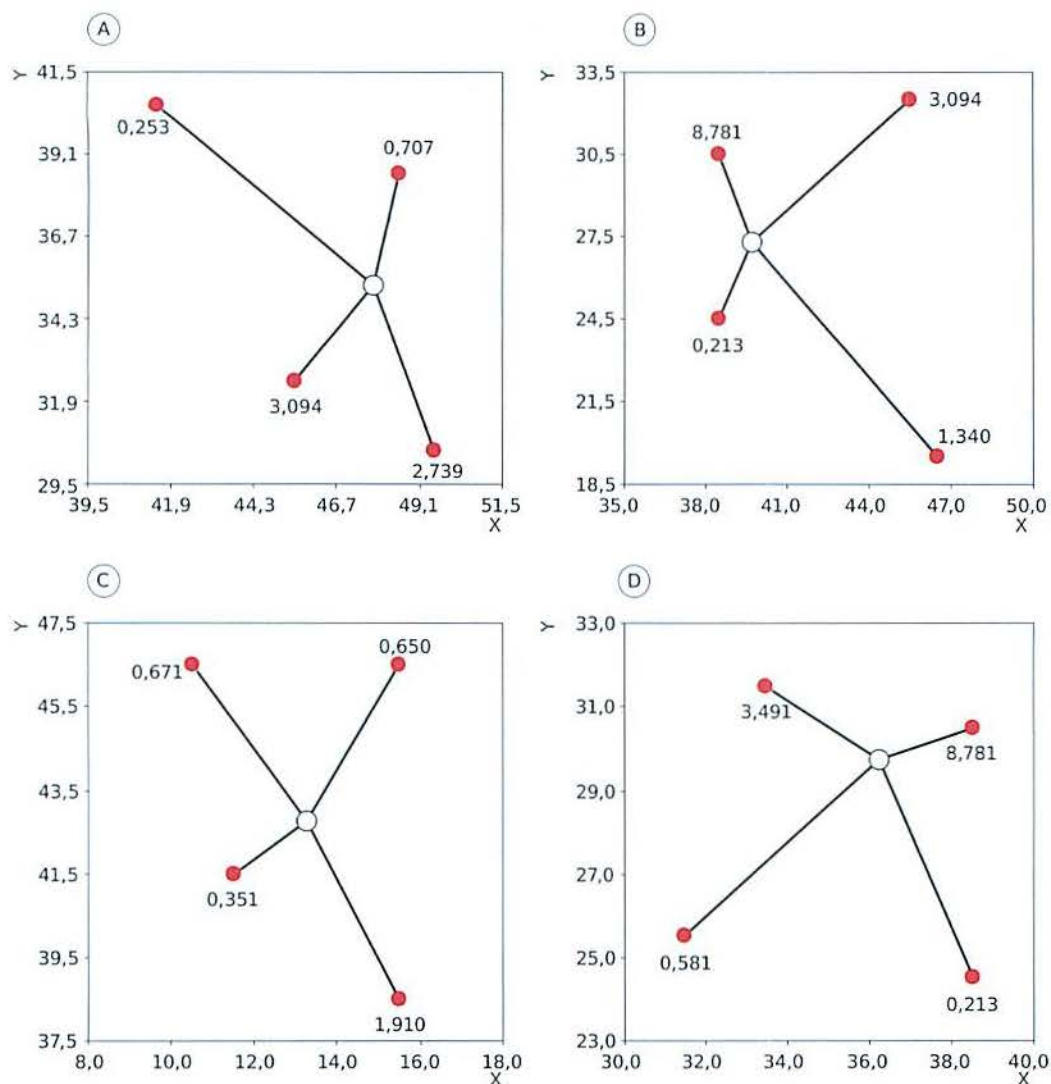


Fig. 3.19 Diferentes localizações dos pontos interpolados em relação aos pontos vizinhos (A-B e C-D), resultando em variâncias de krigagem praticamente iguais ($\neq < 0,000001$)

TAB. 3.9 Pontos de dados para estimativa por krigagem ordinária de $x_o = (1,25; 26,75)$ e $x_o = (1,25; 31,25)$, representados nas Figs. 3.18C e 3.18D, respectivamente

X	Y	Z(x)	$x_o = (1,25; 26,75)$	$x_o = (1,25; 31,25)$
			Pesos	Pesos
7,50	33,50	2,211	0,019869	0,077132
0,50	32,50	0,471	0,148312	0,754688
0,50	25,50	0,214	0,754688	0,148312
7,50	24,50	0,297	0,077132	0,019869
$Z_{KO}^*(x_o)$			0,297810	0,563223
σ_{KO}^2			0,902825	0,902825
μ			-0,065831	-0,065831



TAB. 3.10 Pontos de dados para estimativa por krigagem ordinária de $x_o = (47,75; 35,25)$ e $x_o = (39,75; 27,25)$, representados nas Figs. 3.19A e 3.19B, respectivamente

$x_o = (47,75; 35,25)$				$x_o = (39,75; 27,25)$			
X	Y	Z(x _i)	Pesos	X	Y	Z(x _i)	Pesos
48,50	38,50	0,707	0,470390	45,50	32,50	3,094	0,084588
41,50	40,50	0,253	0,006012	38,50	30,50	8,781	0,398801
45,50	32,50	3,094	0,365340	38,50	24,50	0,213	0,502793
49,50	30,50	2,739	0,158258	46,50	19,50	1,340	0,013819
$Z_{KO}^*(x_o)$			1,898100	$Z_{KO}^*(x_o)$			3,889022
σ_{KO}^2			1,315590	σ_{KO}^2			1,315590
μ			-0,065462	μ			-0,044237

TAB. 3.11 Pontos de dados para estimativa por krigagem ordinária de $x_o = (13,25; 42,75)$ e $x_o = (36,25; 29,75)$, representados nas Figs. 3.19C e 3.19D, respectivamente

$x_o = (13,25; 42,75)$				$x_o = (36,25; 29,75)$			
X	Y	Z(x _i)	Pesos	X	Y	Z(x _i)	Pesos
15,50	46,50	0,650	0,237890	38,50	30,50	8,781	0,495009
10,50	46,50	0,671	0,089008	33,50	31,50	3,491	0,333049
11,50	41,50	0,351	0,537658	31,50	25,50	0,581	0,074756
15,50	38,50	1,910	0,135444	38,50	24,50	0,213	0,097186
$Z_{KO}^*(x_o)$			0,661819	$Z_{KO}^*(x_o)$			5,573413
σ_{KO}^2			1,072105	σ_{KO}^2			1,072105
μ			-0,140597	μ			-0,140597

As Tabs. 3.8 e 3.9 mostram que pesos iguais são aplicados para diferentes pontos amostrais. As situações de mesma variância e multiplicador de Lagrange podem explicar que a variância de krigagem é somente um índice de configuração espacial dos pontos. Contudo, quando se analisa a Fig. 3.19, verifica-se que arranjos e pontos completamente diferentes também resultam em variâncias de krigagem praticamente iguais ($\neq < 0,000001$). Por exemplo, na Fig. 3.19A, a incerteza da estimativa é muito menor que a incerteza da estimativa no ponto da Fig. 3.19B. O mesmo se verifica nos pares C e D da Fig. 3.19. Por essa razão, a variância de krigagem não pode ser utilizada como medida de incerteza associada à estimativa da krigagem ordinária.

Nesse sentido, segundo Armstrong (1994), frequentemente surgem novas propostas, não aceitas, de uso da variância de krigagem para fins de cálculo do intervalo de confiança da estimativa e, portanto, para classificação de reservas minerais. Apesar disso, ainda há pesquisadores que ignoram isso e propõem usar a variância de krigagem para classificação de reservas minerais.

Como a variância de krigagem não pode ser usada como uma medida da confiabilidade da estimativa feita pela krigagem ordinária, Yamamoto (2000, p. 491) propôs o uso de uma alternativa para a medida da confiabilidade das estimativas de krigagem ordinária, a qual denominou variância de interpolação:

$$S_o^2 = \sum_{i=1}^n \lambda_i [Z(x_i) - Z_{OK}^*(x_o)]^2 \quad (3.24)$$

Segundo Yamamoto (2000, p. 492-493), a variância de krigagem apresenta as seguintes propriedades:

- garante a exatidão, pois, se um ponto de dado coincide com o ponto a ser interpolado, a variância de interpolação será igual a zero;
- aumenta com a dispersão dos valores próximos utilizados na interpolação;
- usa indiretamente a informação estrutural por meio dos pesos da krigagem ordinária.

A expressão da variância de interpolação (Eq. 3.24) pode ser aplicada tanto para a krigagem pontual como para a krigagem de bloco. De acordo com esse autor, a variância de interpolação para um bloco V pode também ser escrita de forma equivalente:

$$S_V^2 = \frac{1}{nV} \sum_l S_l^2 + \frac{1}{nV} \sum_l [Z_V^* - Z^*(x^{(l)})]^2 \quad (3.25)$$

em que Z_V^* é o teor médio do bloco V , S_l^2 é a variância de interpolação para o l -ésimo sub-bloco e $Z^*(x^{(l)})$ é o teor médio do l -ésimo sub-bloco.

Observar que a Eq. 3.25 nada mais é que a média das variâncias de interpolação dos sub-blocos mais a variância entre os sub-blocos e o bloco, ou seja, ela é similar à relação de aditividade de Krige (Yamamoto, 2000, p. 493).

A prova matemática de que a Eq. 3.25 é equivalente à Eq. 3.24 encontra-se em Yamamoto (2000, p. 507-509).

Muitas vezes, há necessidade de se determinar o valor médio da variável de interesse sobre todo o domínio do fenômeno espacial em estudo. Por exemplo, ao se fazer a avaliação de recursos minerais de um depósito, deve-se determinar o teor médio e a incerteza associada. Se $Z_{V_i}^*$ for o teor médio do i -ésimo bloco de cubagem, considerando que o depósito mineral é composto por N blocos de cubagem, o teor médio do depósito pode ser determinado como:

$$Z_D^* = \frac{1}{N} \sum_{i=1}^N Z_{V_i}^* \quad (3.26)$$

A variância de estimativa global do teor médio pode ser escrita como:

$$\sigma_D^2 = E \{ [Z_D - Z_D^*]^2 \}$$

Segundo Journel e Huijbregts (1978, p. 323), essa expressão pode ser desenvolvida como:

$$\sigma_D^2 = \frac{1}{N} \sum_{i=1}^N \sigma_{V_i}^2 + \frac{1}{N^2} \sum_{i=1}^N \sum_{j \neq i}^N E \{ (Z_{V_i} - Z_{V_i}^*) (Z_{V_j} - Z_{V_j}^*) \}$$

O primeiro termo dessa expressão é simplesmente a média das variâncias de krigagem calculadas para os blocos de cubagem. O segundo termo envolve as covariâncias dos erros $E \{ (Z_{V_i} - Z_{V_i}^*) (Z_{V_j} - Z_{V_j}^*) \}$, as quais consideram combinações de pares $Z_{V_i} Z_{V_j}$ que se apresentam correlacionados, principalmente para pequenas distâncias. Assim, segundo Journel e Huijbregts (1978, p. 323), a soma dessas covariâncias não pode ser negligenciada com respeito ao primeiro termo. Teoricamente, de acordo com esses autores, o cálculo da

covariância do erro seria possível se o variograma fosse conhecido em todas as distâncias h dentro do depósito, mas o variograma é calculado apenas dentro do campo geométrico, ou seja, em no máximo metade das dimensões do depósito mineral.

Uma proposta ao cálculo da variância global do depósito foi oferecida por Yamamoto (2001c, p. 63), que se baseia na seguinte equação:

$$S_D^2 = \frac{1}{N} \sum_{i=1}^N S_{V_i}^2 + \frac{1}{N} \sum_{i=1}^N [Z_{V_i}^* - Z_D^*]^2 \quad (3.27)$$

Essa expressão é similar à Eq. 3.25. Observar que se pode usar recursivamente a relação de aditividade de Krige para calcular a variância da krigagem de bloco, e, em seguida, compor essas variâncias individuais para calcular a variância associada ao teor médio do depósito. Yamamoto (2001c, p. 63) demonstrou que a Eq. 3.27 equivale a:

$$S_D^2 = \sum_{\alpha=1}^M \lambda_D^\alpha [Z(x_\alpha) - Z_D^*]^2$$

em que λ_D^α é o peso global definido associado ao α -ésimo ponto de dado $Z(x_\alpha)$, segundo Crozel e David (1985, p. 788).

Além disso, Yamamoto et al. (2012, p. 150) demonstraram que a variância global de variáveis categóricas também pode ser calculada de forma semelhante à Eq. 3.27, o que comprova a confiabilidade da medida de incerteza por meio da variância de interpolação.

Todas as expressões derivadas da fórmula básica da variância de interpolação (Eq. 3.24) foram provadas matematicamente e, por isso, não são formulações empíricas. Isso significa

que se pode usar a variância de interpolação tanto para variáveis contínuas como para variáveis discretas, inclusive para as variáveis discretas com o mapeamento da zona de incerteza.

Para enfatizar o exposto, as Figs. 3.18 e 3.19 e as Tabs. 3.8 a 3.11 serão, em seguida, consideradas segundo a metodologia da variância de interpolação (Tab. 3.12).

Comparando os resultados da variância de interpolação com a variância de krigagem, verifica-se que a primeira reflete sempre a dispersão dos valores. Por exemplo, entre os pares A e B da Fig. 3.19, o arranjo da Fig. 3.19B tem uma dispersão de valores muito maior que a que ocorre na Fig. 3.19A e, dessa forma, a variância de interpolação do arranjo

da Fig. 3.19B é de 16,48, enquanto, para o arranjo da Fig. 3.19A, a variância de interpolação é igual a 1,318.

Para mostrar a heterocedasticidade da variância de krigagem, os mesmos dados da amostra do conjunto lognormal (arquivo 12, anexo B) foram processados, conforme os resultados ilustrados na Fig. 3.20.

Com os teores krigados representados na Fig. 3.20A, determinaram-se o teor médio global igual a 1,708 (Eq. 3.26) e a variância global igual a 3,718 (Eq. 3.27). Esses dados são muito importantes para qualquer estudo de viabilidade técnico-econômica que se faça necessário,

TAB. 3.12 Variâncias de interpolação calculadas para os arranjos das Figs. 3.28 e 3.29

Ponto para interpolação	Fig.	S_o^2
$x_o = (8,75; 41,75)$	3.18A	0,083012
$x_o = (6,25; 43,25)$	3.18B	0,118085
$x_o = (1,25; 26,75)$	3.18C	0,082493
$x_o = (1,25; 31,25)$	3.18D	0,235390
$x_o = (47,75; 35,25)$	3.19A	1,318030
$x_o = (39,75; 27,25)$	3.19B	16,480262
$x_o = (13,25; 42,75)$	3.19C	0,263057
$x_o = (36,25; 29,75)$	3.19D	11,191281

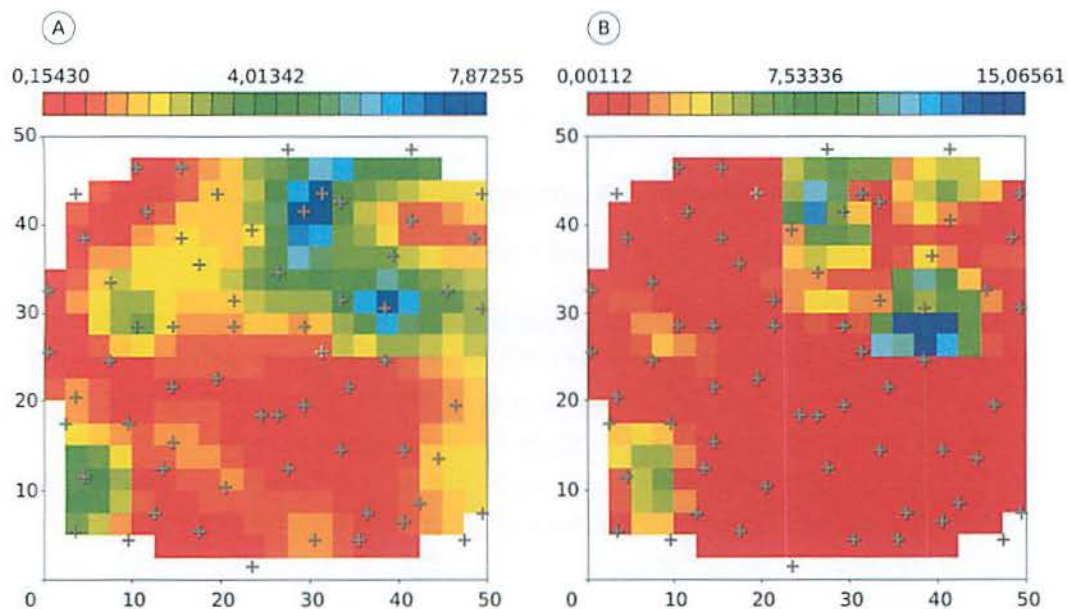


Fig. 3.20 A) Mapa de teores estimados por krigagem pontual e B) mapa da variância de interpolação. Parâmetros de interpolação: $DX = DY = 2,5$ e 1 ponto por quadrante

pois a incerteza é um fator determinante para qualquer tomada de decisão envolvendo aplicação de recursos financeiros. A validade da Eq. 3.27 pode ser comprovada diretamente com os dados krigados (Fig. 3.20), como ilustra a Fig. 3.21.

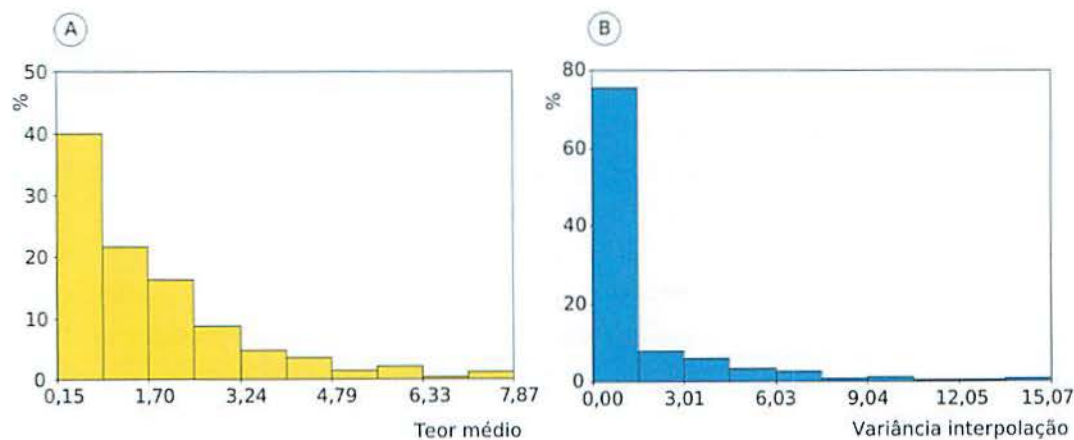


Fig. 3.21 Distribuição A) dos teores médios calculados nos blocos de cubagem e B) das variâncias de interpolação

Da distribuição dos teores médios (Fig. 3.21A), pode-se calcular o teor médio do depósito:

$$\frac{1}{N} \sum_{i=1}^N Z_{V_i}^* = 1,708$$

e a variância:

$$\frac{1}{N} \sum_{i=1}^N [Z_{V_i}^* - Z_D^*]^2 = 2,207$$



Do histograma das variâncias de interpolação (Fig. 3.21B), calcula-se a média:

$$\frac{1}{N} \sum_{i=1}^N S_{V_i}^2 = 1,510$$

Somando os dois termos da variância, tem-se a variância global do depósito:

$$S_D^2 = 1,510 + 2,207 = 3,717$$

Nesse exemplo, a média do depósito foi exatamente igual à média dos pontos de dados, o que demonstra que as fórmulas empregadas estão corretas. Por outro lado, em casos reais, nos quais se calculam blocos de cubagem acima de um dado teor de corte, a média do depósito será maior que a média dos pontos de dados, por causa da restrição imposta.

A técnica da krigagem ordinária foi descrita nesta seção com o objetivo de mostrar como se pode aplicá-la para o cálculo de estimativa e, ao mesmo tempo, chamar a atenção para as medidas de incerteza. Como exposto, a variância de krigagem, de natureza homocedástica, não pode ser usada como medida de incerteza e muito menos para fins de cálculo do intervalo de confiança da estimativa e, consequentemente, para classificação de recursos minerais. Foi demonstrado que o multiplicador de Lagrange resultante da resolução do sistema de equações da krigagem ordinária não pode ser usado como uma medida de variância, pois, dependendo da geometria do arranjo de pontos de dados, que poderá ser ora positivo, ora negativo. A única exceção é o multiplicador de Lagrange na krigagem da média, que corresponde a uma variância, também independente dos valores dos pontos de dados, mas somente do modelo de variograma.

A variância de interpolação, por outro lado, parece ser a única solução viável para determinação da incerteza associada à estimativa por krigagem ordinária. Além disso, ela pode ser usada para o cálculo da variância global do depósito mineral, que é fundamental para estudos de viabilidade técnico-econômica.

Evidentemente, essa constatação não exclui a possibilidade de se fazer simulação estocástica para determinação da incerteza, como será visto no Cap. 5.

No processo de inferência espacial, é importante que as características das amostras sejam reproduzidas, para que possam ser usadas para inferir com segurança o fenômeno espacial desconhecido. Assim, espera-se que a krigagem ordinária reproduza não só o histograma, mas também o variograma, ou seja, a distribuição e variabilidade espaciais do fenômeno em estudo.

A krigagem ordinária, entretanto, produz estimativas suavizadas, de tal modo que os valores mais altos são subestimados e os valores mais baixos, superestimados. Trata-se de um efeito colateral da krigagem conhecido como efeito de suavização.

Esse efeito não é exclusivo da krigagem ordinária, mas de todos os demais métodos baseados na média móvel ponderada (p. ex., inverso da distância). A correção do efeito de suavização pode ser feita pela aplicação de algoritmos de pós-processamento das estimativas, os quais adicionam ou subtraem uma quantidade dependendo da estimativa, que pode estar sub ou superestimada.

Yamamoto (2005, 2007) e, mais recentemente, Rezaee, Asghari e Yamamoto (2011), estudando o assunto, propuseram algoritmos distintos para correção do efeito de suavização

da krigagem ordinária, sendo mais eficientes se aplicados a dados transformados em vez dos dados originais. As estimativas são corrigidas do efeito de suavização para em seguida serem submetidas à transformação reversa, para retorná-las à escala original de medida. Tais algoritmos serão descritos na seção sobre krigagem não linear, que trabalha com estimativas feitas no domínio dos dados transformados.

3.3 KRIGAGEM NÃO LINEAR

Os métodos de estimativa que usam os dados transformados não linearmente são agrupados na categoria de krigagem não linear. Na verdade, todos esses métodos fazem uso do estimador da krigagem ordinária (Eq. 3.14), mas para dados transformados. Os métodos das krigagens multigaussiana, lognormal e indicadora serão descritos nesta seção. Esses métodos de estimativa envolvem a transformação dos dados originais, cálculo e modelagem de variogramas experimentais para os dados transformados e estimativa em pontos não amostrados no domínio da variável transformada e transformada reversa para a escala original.

3.3.1 Correção do efeito de suavização da krigagem ordinária

Geralmente, a transformada reversa é feita usando-se a função inversa da transformação não linear. Por exemplo, para o caso da transformada lognormal, a transformada reversa é obtida aplicando-se a função exponencial, com a base igual à usada na função logarítmica e o expoente igual à estimativa resultante no domínio lognormal. Deve-se somar um termo de não viés antes da transformada reversa.

No caso da krigagem lognormal, esse termo de não viés é baseado na teoria lognormal e usa a variância de krigagem (Journel, 1980, p. 295). Entretanto, como a variância de krigagem não é uma medida de incerteza, as transformadas reversas continuam apresentando vieses.

Uma proposta diferente foi feita por Yamamoto (2007, p. 220-222) e está fundamentada na correção do efeito de suavização da krigagem ordinária.

A transformada reversa produz resultados enviesados em relação aos dados amostrais. Como a inferência espacial é feita usando-se a amostra, a inferência, se as estimativas apresentarem vieses em relação ao conjunto amostral, não será confiável.

Para justificar a necessidade da correção do efeito de suavização da krigagem ordinária, considerar o exemplo mostrado na Fig. 3.22 da transformada reversa das estimativas de krigagem enfileirada (Yamamoto, 2010a, p. 108).

A transformada enfileirada nada mais é que a classificação dos dados em ordem crescente. Na verdade, o ordinal decorre da classificação dos dados em ordem crescente. Resulta, portanto, no quantil, que é usado para fazer a transformada gaussiana: $\left(\frac{r(x_i)}{n+1}\right)$.

Essa transformação produz uma distribuição uniforme, conforme se pode verificar na Fig. 3.22. Por causa do efeito de suavização da krigagem ordinária, as estimativas, ainda no domínio dos quantis, não mostram mais a distribuição uniforme, e sim uma distribuição simétrica com forma de sino. Portanto, se essas estimativas forem transformadas de volta para a escala original dos dados, a distribuição resultante não será semelhante à distribuição amostral, em razão da supressão dos extremos pelo efeito de suavização.

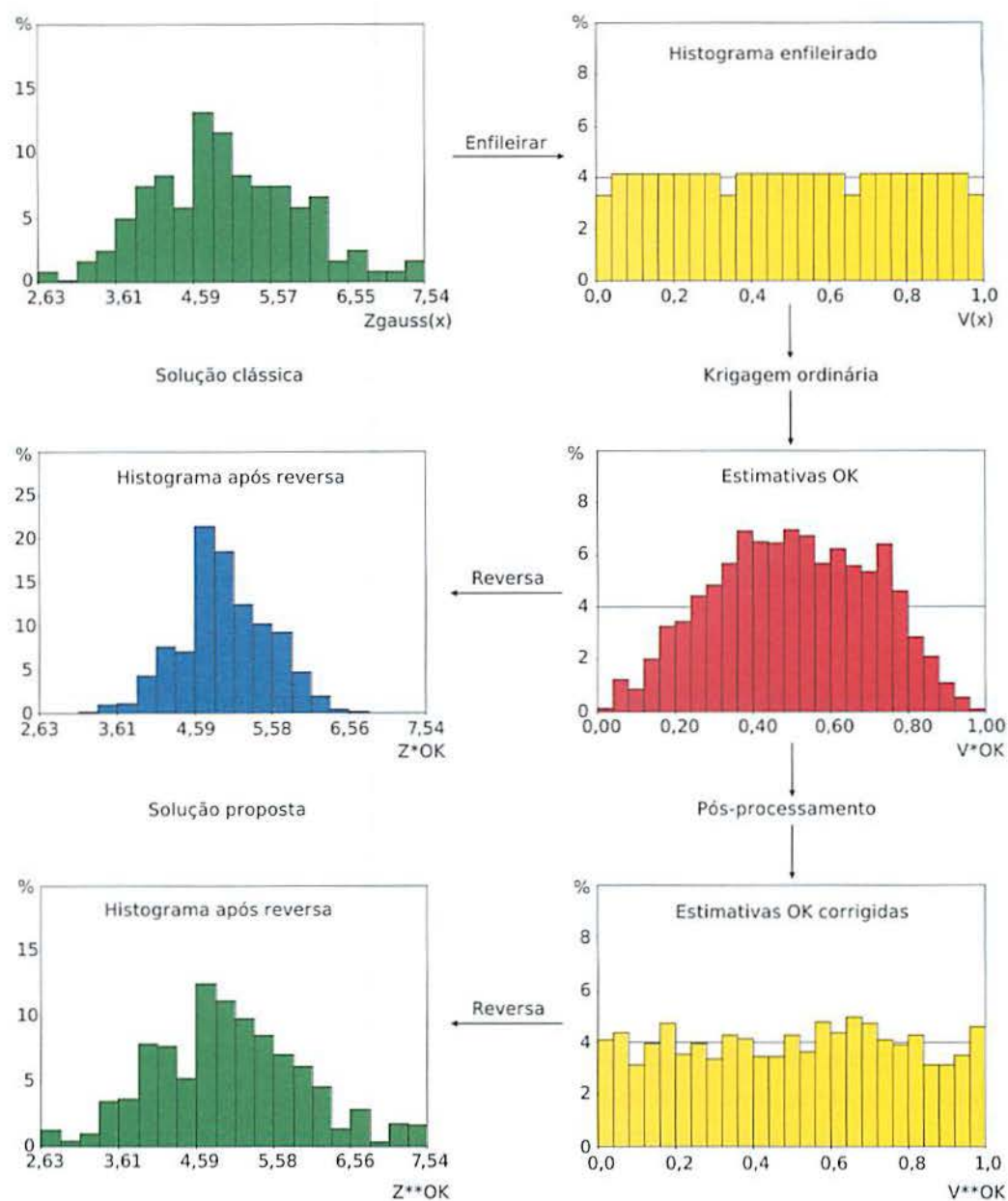


Fig. 3.22 O processo da transformada reversa sem e com correção do efeito de suavização da krigagem ordinária
Fonte: Yamamoto (2010a, p. 108).

Por outro lado, caso as estimativas sejam corrigidas por meio do pós-processamento, verifica-se que o histograma das transformadas reversas é muito semelhante ao histograma amostral.

Essa é a principal justificativa para a correção do efeito de suavização da krigagem ordinária de dados transformados não linearmente, antes da aplicação da função inversa para retornar à escala original dos dados amostrais.

Serão descritos dois algoritmos completamente diferentes para correção do efeito de suavização. O primeiro, proposto por Yamamoto (2005, 2007), e o segundo, mais recente, proposto por Rezaee, Asghari e Yamamoto (2011).

O primeiro método tem como ponto de partida o processo de validação cruzada dos dados, que consiste na estimativa sobre cada ponto amostral, cujo valor original é retirado do conjunto, com base nos vizinhos mais próximos. Assim, para cada ponto amostral têm-se o valor estimado e o valor original. Subtraindo-se o valor original do estimado, chega-se ao erro de estimativa. Considerando $Y(x)$ a variável aleatória resultante da transformação não linear (gaussiana ou logarítmica) da variável aleatória original $Z(x)$, o erro de estimativa é:

$$Erro = Y^*(x_o) - Y(x_o) \quad (3.28)$$

Ainda como resultado do processo de validação cruzada, tem-se, para cada ponto amostral, a incerteza S_o , expressa como desvio padrão de interpolação, associada à estimativa $Y^*(x_o)$. Dividindo-se o erro de estimativa (Eq. 3.28), com o sinal trocado, pelo desvio padrão de interpolação, transforma-se o erro em unidades de desvio padrão. Yamamoto (2005, p. 73) propõe chamar essa nova variável aleatória NS_o de número de desvios padrão:

$$NS_o = \frac{-Erro}{S_o}$$

Após isso, em cada ponto amostral, além da informação do ponto de dado $Y(x)$, tem-se o número de desvios padrão NS_o . As duas variáveis serão interpoladas em pontos não amostrados. Assim, após a interpolação, têm-se, para cada nó da malha regular, a estimativa $Y_{KO}^*(x_o)$, NS_o^* e o desvio padrão de interpolação S_o . A multiplicação de NS_o^* com S_o resulta na quantidade de correção do efeito de suavização. A estimativa corrigida é então obtida como:

$$Y_{KO}^{**}(x_o) = Y_{KO}^*(x_o) + NS_o^* \cdot S_o \quad (3.29)$$

Observar que, ao multiplicar o número de desvios padrão NS_o^* pelo desvio padrão S_o , a quantidade resultante estará na mesma unidade da variável transformada $Y(x)$. A quantidade de correção poderá ser positiva ou negativa, dependendo da ocorrência de subestimativa ou superestimativa, respectivamente.

Segundo Yamamoto (2007, p. 221), algumas vezes a estimativa corrigida (Eq. 3.29) pode estar supercorrigida, fazendo com que ela caia fora dos limites mínimo ($Y_{OK}^{**}(x_o) < y_{\min}$) e máximo ($Y_{OK}^{**}(x_o) > y_{\max}$) dos valores dos pontos vizinhos próximos. Nesses casos, a quantidade de correção $NS_o^* \cdot S_o$ é substituída por $\delta = Y_{OK}^*(x_o) - y_{\min}$, se $NS_o^* < 0$, e por $\delta = y_{\max} - Y_{OK}^*(x_o)$, se $NS_o^* > 0$. Assim, as estimativas corrigidas podem ser calculadas como:

$$\begin{cases} Y_{OK}^{**}(x_o) = Y_{OK}^*(x_o) + NS_o^* \cdot S_o \cdot \text{fator} \\ Y_{OK}^{**}(x_o) = Y_{OK}^*(x_o) + \delta \cdot \text{fator} \end{cases} \quad (3.30)$$

em que fator é uma constante que faz com que a variância das estimativas corrigidas $\text{var}[Y_{OK}^{**}(x_o)]$ seja igual à variância dos pontos de dados transformados $\text{var}[Y(x)]$.

Substituindo-se as quantidades de correção nas expressões da Eq. 3.30 por uma nova variável aleatória $Y_{NS_o}^*(x_o)$, a estimativa corrigida pode ser expressa como:

$$Y_{OK}^{**}(x_o) = Y_{OK}^*(x_o) + Y_{NS_o}^*(x_o) \cdot \text{fator} \quad (3.31)$$



Yamamoto (2007, p. 221) mostrou que a igualdade $Var[Y(x)] = Var[Y_{OK}^{**}(x_o)]$ resulta na seguinte equação de 2º grau:

$$Var[Y_{NS_o}^{*}(x_o)] \text{ fator}^2 + 2cov(Y_{OK}^{*}(x_o), Y_{NS_o}^{*}(x_o)) \text{ fator} + Var[Y_{OK}^{*}(x_o)] - Var[Y(x)] = 0$$

Assim, fator é a raiz positiva dessa equação de 2º grau:

$$\text{fator} = \frac{-2cov(Y_{OK}^{*}(x_o), Y_{NS_o}^{*}(x_o)) + \sqrt{\Delta}}{2Var[Y_{NS_o}^{*}(x_o)]}$$

em que Δ é o discriminante da equação de 2º grau. O seu valor pode ser encontrado em Yamamoto (2007, p. 221).

Uma proposta completamente diferente à descrita foi apresentada por Rezaee, Asghari e Yamamoto (2011, p. 26-28). Segundo esses autores, a correção é baseada na bem conhecida equação de transformação:

$$X = (Z) \cdot S_X + \bar{X} \quad (3.32)$$

em que Z é o escore conhecido, S_X é o desvio padrão e $\bar{X}$ é a média do escore bruto X . Substituindo-se as variáveis, conforme a notação usada neste livro, tem-se:

$$Y_{OK}^{**}(x_o) = \left(\frac{Y_{OK}^{*}(x_o) - E[Y_{OK}^{*}(x_o)]}{\sqrt{Var[Y_{OK}^{*}(x_o)]}} \right) \cdot Var\sqrt{Y(x)} + E[Y(x)] \quad (3.33)$$

Observar na Eq. 3.33 que o termo entre parênteses do lado direito é o escore Z da Eq. 3.32, e o escore bruto nada mais é que o dado transformado $Y(x)$.

Pela natureza da correção proposta na Eq. 3.33, deve-se trabalhar sempre no domínio gaussiano, pois o termo entre parênteses nessa equação é a variável normal reduzida. Assim, a correção de Rezaee et al. (2011, p. 26-28) é aplicada somente para fazer a transformada reversa da krigagem multigaussiana.

Para a krigagem multigaussiana têm-se as correções de Yamamoto (2005, 2007) e Rezaee, Asghari e Yamamoto (2011), enquanto para krigagem lognormal pode-se aplicar somente a correção de Yamamoto (2005, 2007), tendo em mente a necessidade de calcular corretamente o termo de não viés antes da transformada reversa.

As propostas descritas utilizam a correção do efeito de suavização. Embora a simulação estocástica tenha sido proposta em Geoestatística para correção do efeito de suavização da krigagem ordinária, ela nunca deve ser aplicada para cálculo do termo de não viés da krigagem não linear.

3.3.2 Krigagem multigaussiana

A krigagem multigaussiana é baseada na transformação dos dados para escores da distribuição normal. A transformada gaussiana dos dados originais garante que a distribuição resultante seja normal, com média zero e variância igual a 1. Entretanto, essa condição não é suficiente para a krigagem multigaussiana, que trabalha sob a hipótese da multigaussianidade dos dados. Assim, segundo Yamamoto e Chao (2009, p. 121), há necessidade de testar se a distribuição de dois, três ou mais pontos é gaussiana. Como essa verificação é muito difícil

para distribuições multipontos, adota-se o teste de bigaussianidade, que, na prática, consiste em testar se a distribuição de dois pontos é também gaussiana, por meio da comparação de variogramas. Basicamente, existem dois métodos para realização desse teste, os quais foram propostos por Goovaerts (1997, p. 271-275) e Emery (2005, p. 167-169). Segundo Yamamoto e Chao (2009, p. 128), o teste proposto por Goovaerts mostra-se mais robusto em relação ao de Emery. Em vista disso, neste livro será adotado o teste de bigaussianidade dos dados proposto por Goovaerts, que consiste no seguinte:

- os dados originais $Z(x)$ são transformados para o domínio gaussiano $Y(x)$ (seção 3.1.1 e Fig. 3.6);
- os dados transformados $Y(x)$ são usados para cálculo e modelagem do variograma experimental $\gamma_Y(h)$;
- com o modelo da função variograma, deriva-se a função covariância $C_Y(h) = C_Y(0) - \gamma_Y(h)$. A função covariância deveria ser calculada como $C_Y(h) = 1 - \gamma_Y(h)$, mas a variância dos dados após a transformada gaussiana tende a 1 quando o número de pontos de dados tende ao infinito (Tab. 3.1). Além disso, o patamar da função variograma não é necessariamente igual à variância amostral;
- em seguida, estabelecem-se alguns percentis p (por exemplo, nove decis), dos quais se derivam os escores da distribuição normal acumulada y_p (por exemplo, os escores da distribuição para os nove decis na Tab. 3.13);
- com os escores, determinam-se as respectivas variáveis indicadoras (Eq. 3.4);
- as variáveis indicadoras assim obtidas são usadas para calcular os variogramas experimentais $\gamma^*(h; y_p)$;
- para os mesmos escores y_p , calculam-se os variogramas teóricos da variável indicadora: $\gamma_I(h; y_p) = p - G(h; y_p)$;
- em que a função de distribuição acumulada gaussiana é calculada como: $G(h; y_p) = p^2 + \frac{1}{2\pi} \int_0^{\arcsin C_Y(h)} \exp\left[\frac{-y_p^2}{1+\sin\theta}\right] d\theta$, e que pode ser resolvida numericamente usando a regra de integração de Simpson;
- finalmente, os variogramas experimentais e teóricos correspondentes aos percentis são comparados entre si. Se os pares mostrarem boa aderência, então se aceita a hipótese de bigaussianidade dos dados e, conseqüentemente, a hipótese de multigaussianidade.

Após a verificação da multigaussianidade dos dados, pode-se proceder às estimativas por meio da krigagem multigaussiana, cujo estimador é:

$$Y_{MG}^*(x_o) = \sum_{i=1}^n \lambda_i Y(x_i) \quad (3.34)$$

Os ponderadores são obtidos conforme a metodologia descrita para a krigagem ordinária. A incerteza associada à estimativa, medida pela variância de interpolação, pode ser obtida com:

$$S_o^2 = \sum_{i=1}^n \lambda_i [Y(x_i) - Y_{MG}^*(x_o)]^2$$

Tradicionalmente, a solução adotada para fazer a transformada reversa do resultado da krigagem multigaussiana leva em consideração, como termo de não viés, a variância de

TAB. 3.13
Escores da distribuição normal acumulada para nove decis

Decis	Escores
0,10	-1,28155
0,20	-0,84162
0,30	-0,52440
0,40	-0,25335
0,50	0
0,60	0,25335
0,70	0,52440
0,80	0,84162
0,90	1,28155



krigagem e o multiplicador de Lagrange (Emery, 2010, p. 213). Pelos motivos já expostos e principalmente porque a variância de krigagem não é uma medida efetiva de incerteza, essa opção não será considerada neste livro.

A transformada reversa pode ser obtida aplicando-se a função inversa da estimativa corrigida, tanto por meio da Eq. 3.33 como pela Eq. 3.31:

$$Z_{MG}^{**}(x_o) = \varphi^{-1}(Y_{MG}^{**}(x_o)) \quad (3.35)$$

em que $\varphi^{-1}(\cdot)$ é a função inversa da transformação gaussiana.

Exemplo de aplicação da krigagem multigaussiana

Para esse exemplo foi escolhido o Arquivo 12 (Figs. 2.21 e 2.24), Anexo B, que apresenta uma distribuição lognormal e, portanto, apropriada para fazer a transformação gaussiana e, conseqüentemente, a krigagem multigaussiana. Assim, o primeiro passo para a krigagem multigaussiana é fazer o teste de multigaussianidade dos dados. Inicialmente, faz-se a transformação dos dados para uma distribuição normal, com média zero e variância tendendo a 1 (Fig. 3.23A).

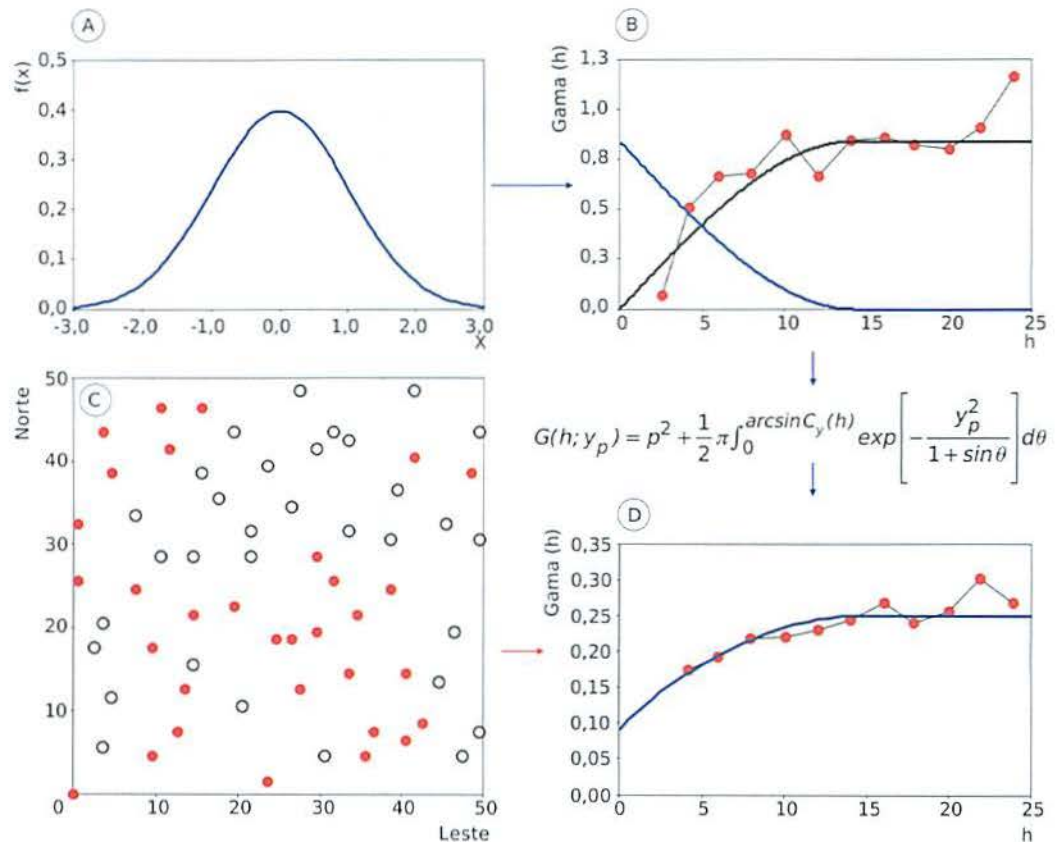


Fig. 3.23 Síntese do procedimento do teste de bigaussianidade dos dados: A) transformação gaussiana dos dados; B) variograma experimental e modelado dos dados após transformação gaussiana e o modelo da função covariância; C) mapa de pontos da variável indicadora para a mediana (círculo cheio vermelho tem indicadora igual a 1 e círculo vazio, indicadora igual a zero); D) variograma experimental da indicadora (em vermelho) e variograma teórico obtido com base na distribuição normal (equação entre B e D) (linha cheia azul)

Com os dados transformados, calcula-se o variograma experimental (Fig. 3.23B), cujo modelo é descrito por $\gamma_Y(h) = 0,84 \left[1,5 \frac{h}{14,16} - 0,5 \left(\frac{h}{14,16} \right)^3 \right]$, e a função covariância é: $C_Y(h) = 0,84 - 0,84 \left[1,5 \frac{h}{14,16} - 0,5 \left(\frac{h}{14,16} \right)^3 \right]$.

Com base na função covariância modelada (Fig. 3.23B), calcula-se o variograma teórico da variável indicadora de uma distribuição normal. O exemplo mostra o variograma da variável indicadora para a mediana da distribuição. Assim, para cada distância h , determina-se a covariância $C_Y(h)$ que será utilizada na integral (vide equação da Fig. 3.23).

O variograma teórico da variável indicadora da distribuição normal é obtido como (Fig. 3.23D):

$$\gamma_I(h; y_p) = p - G(h; y_p)$$

A Tab. 3.14 mostra os valores usados para o cálculo do variograma teórico da variável indicadora da mediana da distribuição normal, conforme o procedimento descrito.

TAB. 3.14 Valores usados no cálculo do variograma teórico da indicadora da mediana da distribuição normal

h	$C_Y(h)$	$\arcsen C_Y(h)$	$G(h; y_p)$	$\gamma_I(h; y_p)$
0,000	0,840	0,997	0,409	0,091
1,000	0,751	0,850	0,385	0,115
2,000	0,663	0,725	0,365	0,135
3,000	0,577	0,615	0,348	0,152
4,000	0,494	0,516	0,332	0,168
5,000	0,414	0,426	0,318	0,182
6,000	0,338	0,345	0,305	0,195
7,000	0,268	0,271	0,293	0,207
8,000	0,204	0,205	0,283	0,217
9,000	0,147	0,148	0,273	0,227
10,000	0,098	0,098	0,266	0,234
11,000	0,058	0,058	0,259	0,241
12,000	0,028	0,028	0,254	0,246
13,000	0,008	0,008	0,251	0,249
14,000	0,000	0,000	0,250	0,250
15,000	0,000	0,000	0,250	0,250
16,000	0,000	0,000	0,250	0,250
17,000	0,000	0,000	0,250	0,250
18,000	0,000	0,000	0,250	0,250
19,000	0,000	0,000	0,250	0,250
20,000	0,000	0,000	0,250	0,250
21,000	0,000	0,000	0,250	0,250
22,000	0,000	0,000	0,250	0,250
23,000	0,000	0,000	0,250	0,250
24,000	0,000	0,000	0,250	0,250
25,000	0,000	0,000	0,250	0,250

O variograma teórico é comparado com o variograma experimental obtido com os pontos de dados (Fig. 3.23C). Como se pode observar na Fig. 3.23D, o ajuste entre os pontos do variograma experimental e a curva do variograma teórico pode ser considerado bom. Essa verificação é feita para o conjunto de variogramas experimentais e teóricos (Fig. 3.24), que mostra um razoável ajuste para os demais decis.

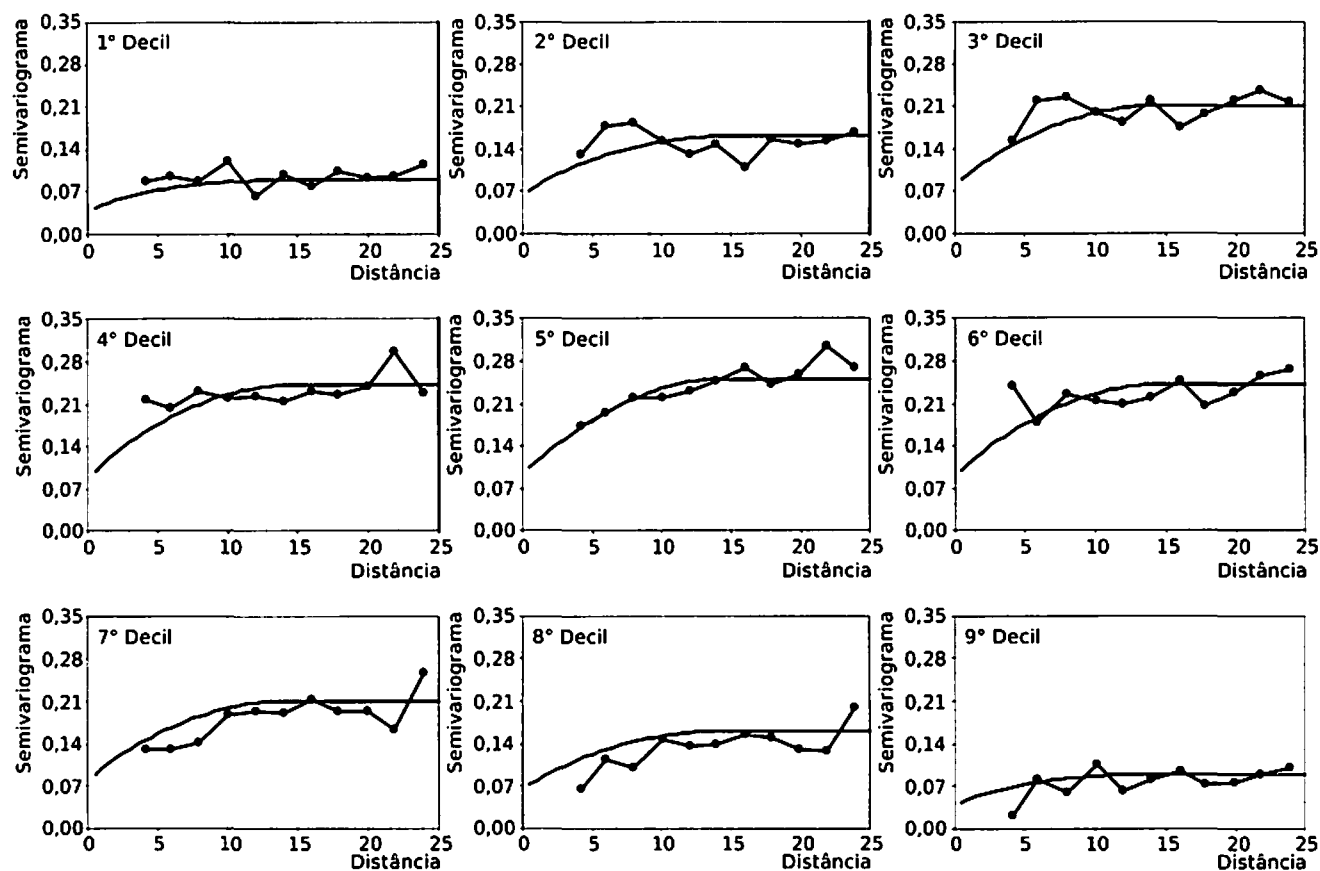


Fig. 3.24 Teste de bigaussianidade da amostra do conjunto lognormal. Círculos cheios = pontos dos variogramas experimentais; linha contínua = variograma teórico deduzido do modelo de covariância e distribuição normal

O melhor ajuste sempre ocorrerá com o variograma da indicadora da mediana, por causa da maior quantidade de pares possíveis para o cálculo dos variogramas experimentais.

As estimativas transformadas de volta para o domínio original usando as Eqs. 3.31 e 3.33 encontram-se na Fig. 3.25 (p. 92).

Existem algumas diferenças nas imagens resultantes, mas constata-se que, em geral, elas estão correlacionadas entre si (Fig. 3.26, p. 92).

Embora a correlação seja razoavelmente boa (0,880), existem diferenças, e as maiores ocorrem para valores baixos (até $Z(x) = 5$). Evidentemente, diferenças existirão entre as duas alternativas usadas, mas a condição mais importante a ser verificada é a aderência das distribuições calculadas em relação à distribuição amostral, uma vez que, no processo de inferência espacial, é importante que a estimativa tenha uma distribuição próxima da distribuição amostral, assim como suas estatísticas.

A Fig. 3.27 mostra as curvas acumulativas da variável original e das estimativas transformadas para o domínio original de medida.

Pode-se verificar nessa figura que, em termos qualitativos, a aderência das distribuições das estimativas transformadas para a distribuição amostral é boa. As estatísticas dessas distribuições podem ser comparadas (Tab. 3.15).

As estatísticas confirmam que as duas propostas de correção da suavização produzem resultados muito próximos em termos de reprodução do histograma amostral. As diferenças verificadas nos mapas-imagens (Fig. 3.25) se devem às metodologias completamente distintas usadas na correção. Mas, como se trata de uma amostra com apenas 64 pontos, os dois resultados podem ser considerados satisfatórios. Qualquer uma dessas propostas produz estimativas transformadas sem viés em relação às estatísticas amostrais. A vantagem da correção proporcionada pela Eq. 3.33 é a simplicidade de implementação, pois consiste na aplicação da equação de transformação (Eq. 3.32).

Com o objetivo de examinar as correções envolvidas e permitir ao leitor acompanhar os cálculos envolvidos, apresentam-se, na Fig. 3.28 (p. 94), quatro pontos não amostrados que foram interpolados usando-se a krigagem multigaussiana, conforme listagens dos pontos de dados nas Tabs. 3.16 a 3.19 (p. 93).

Nessas tabelas, a variável $Z_{MG}^*(x_o)$ é obtida por transformação reversa de $Y_{OK}^*(x_o)$. As transformadas reversas foram feitas sem e com os termos de não viés, de acordo com as Eqs. 3.31 e 3.33, conforme relacionadas nas Tabs. 3.16 a 3.19. Pode-se observar que a transformada reversa, sem a adição do termo de não viés, produz resultados muito diferentes daqueles obtidos após a correção da suavização.

Considerando que as distribuições das transformadas reversas, após aplicação dos termos de não viés, reproduzem a distribuição amostral (Fig. 3.27), com, portanto, as estatísticas descritivas muito próximas das estatísticas amostrais, verifica-se que qualquer um desses métodos proporciona resultados interessantes.

Cabe salientar que essas conclusões são válidas à luz da hipótese de trabalho em que, para a inferência espacial, a reprodução das características da amostra seja importante. Evidentemente, isso não se aplica diretamente no cálculo de recursos minerais, em que blocos de grandes dimensões são avaliados por amostras de testemunhos de sondagens localizadas na vizinhança destes.

A krigagem multigaussiana, embora requeira o trabalhoso teste de bigaussianidade dos dados, pode produzir resultados interessantes quando a distribuição amostral for lognormal ou apresentar assimetria positiva.

Os resultados apresentados mostram claramente que as distribuições resultantes têm excelente aderência à distribuição amostral, satisfazendo a hipótese adotada para a inferência espacial do fenômeno em estudo. Pode-se afirmar que a transformada reversa sem a

TAB. 3.15 Estatísticas amostrais e das estimativas transformadas por krigagem multigaussiana

Estatística	Amostra	Eq. 3.31	Eq. 3.33
<i>N</i>	64	334	334
Média	1,708	1,707	1,716
Desvio padrão	1,923	1,906	2,004
Coef. var.	1,126	1,117	1,168
Máximo	8,781	8,781	8,781
Quartil sup.	2,113	2,125	2,131
Mediana	1,089	1,107	0,923
Quartil inf.	0,348	0,348	0,348
Mínimo	0,095	0,100	0,095

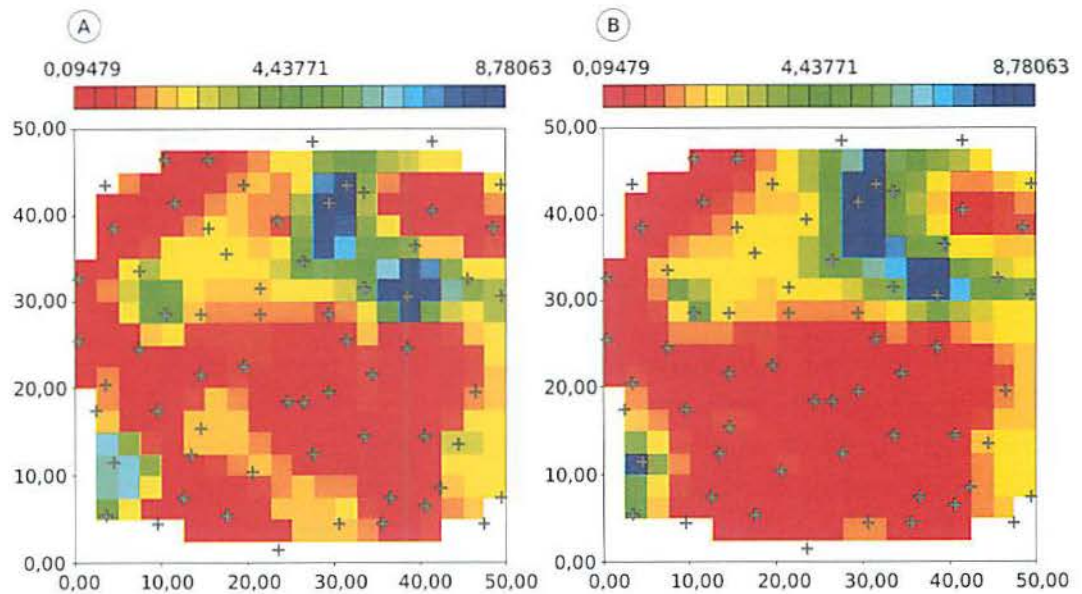


Fig. 3.25 Mapas-imagens das transformadas reversas das estimativas corrigidas da suavização da krigagem ordinária segundo: A) a Eq. 3.31 e B) a Eq. 3.33

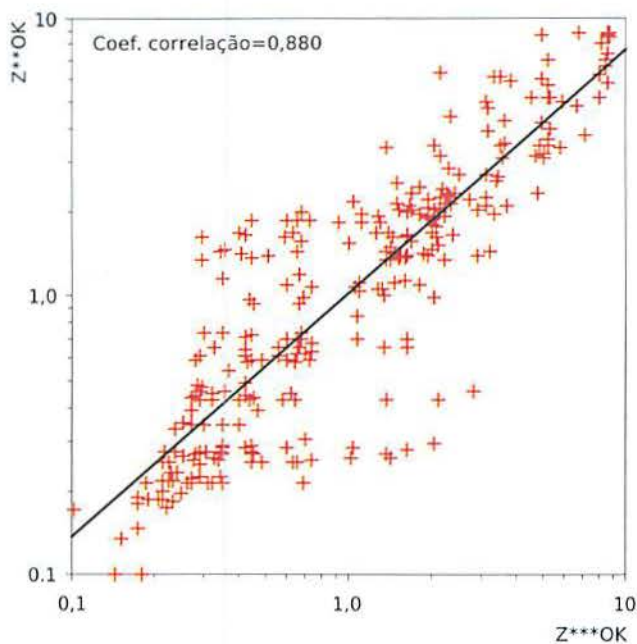


Fig. 3.26 Diagrama de dispersão entre as transformadas reversas, conforme as Eqs. 3.31 (eixo das ordenadas) e 3.33 (eixo das abscissas)

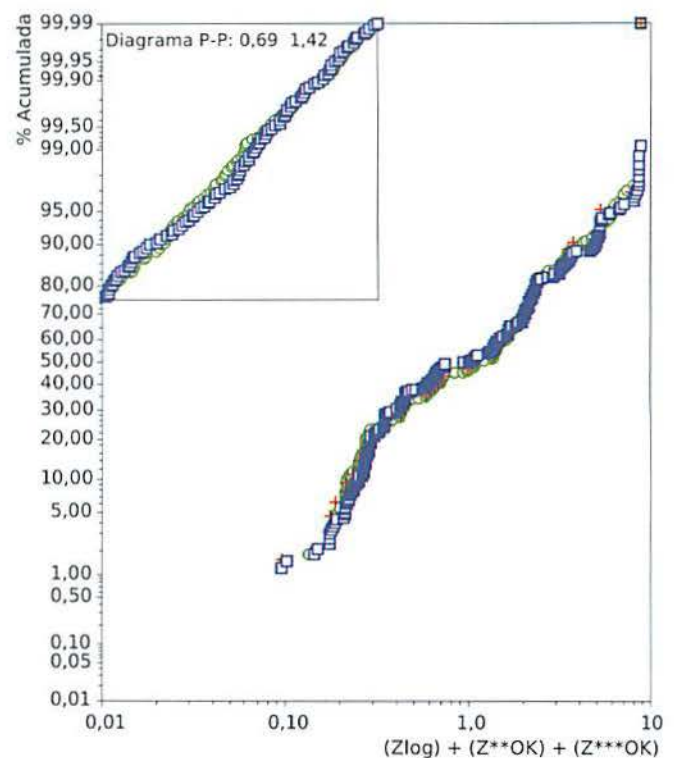


Fig. 3.27 Distribuições acumulativas da distribuição amostral e das transformadas reversas de estimativas obtidas por krigagem multi-gaussiana. Cruz vermelha = distribuição amostral; círculo verde = estimativas transformadas conforme Eq. 3.31; quadrado azul = estimativas transformadas conforme Eq. 3.33

TAB. 3.16 Pontos de dados para interpolação do ponto
($x = 16,25$; $y = 21,25$)

X	Y	Pesos	$Y(x_i)$	$Z(x_i)$
21,50	28,50	0,01138	0,01928	1,11846
19,50	22,50	0,27311	-0,13538	0,73964
14,50	28,50	0,03363	0,17442	1,38656
14,50	21,50	0,51808	-0,68664	0,34801
9,50	17,50	0,00244	-0,95721	0,27194
14,50	15,50	0,10225	0,37496	1,68264
24,50	18,50	0,05911	-1,86961	0,17239
26,50	18,50	0,00000	-1,68335	0,17307
Sem correção			$Y_{OK}^*(x_o)$	$Z_{MG}^*(x_o)$
			-0,46113	0,44579
Método de correção			$Y_{OK}^{**}(x_o)$	$Z_{MG}^{**}(x_o)$
Eqs. 3.31 e 3.35			-0,48107	0,44490
Eqs. 3.33 e 3.35			-0,63162	0,36170

TAB. 3.18 Pontos de dados para interpolação do ponto
($x = 33,75$; $y = 36,25$)

X	Y	Pesos	$Y(x_i)$	$Z(x_i)$
39,50	36,50	0,24029	0,84162	2,35047
41,50	40,50	0,02419	-1,08726	0,25253
33,50	42,50	0,16948	1,08726	3,44066
29,50	41,50	0,12961	1,86961	8,51869
33,50	31,50	0,27371	1,15974	3,49132
26,50	34,50	0,11842	1,32668	4,90139
45,50	32,50	0,00000	0,95721	3,09423
38,50	30,50	0,04430	2,16004	0,61202
Sem correção			$Y_{OK}^*(x_o)$	$Z_{MG}^*(x_o)$
			1,17275	3,52730
Método de correção			$Y_{OK}^{**}(x_o)$	$Z_{MG}^{**}(x_o)$
Eqs. 3.31 e 3.35			1,40712	5,12984
Eqs. 3.33 e 3.35			1,65087	7,10060

TAB. 3.17 Pontos de dados para interpolação do ponto
($x = 36,25$; $y = 8,75$)

X	Y	Pesos	$Y(x_i)$	$Z(x_i)$
44,50	13,50	0,01516	0,68664	2,11924
40,50	14,50	0,08361	-1,23889	0,23167
33,50	14,50	0,11882	-0,59201	0,42291
27,50	12,50	0,03073	-0,84162	0,29004
35,50	4,50	0,00000	-0,50240	0,44395
30,50	4,50	0,04470	0,33389	1,63392
36,50	7,50	0,66105	-1,54199	0,18582
40,50	6,50	0,04594	-0,33389	0,61202
Sem correção			$Y_{OK}^*(x_o)$	$Z_{MG}^*(x_o)$
			-1,20912	0,23884
Método de correção			$Y_{OK}^{**}(x_o)$	$Z_{MG}^{**}(x_o)$
Eqs. 3.31 e 3.35			-1,59582	0,18062
Eqs. 3.33 e 3.35			-1,67654	0,17362

TAB. 3.19 Pontos de dados para interpolação do ponto
($x = 31,25$; $y = 41,25$)

X	Y	Pesos	$Y(x_i)$	$Z(x_i)$
31,50	43,50	0,17011	1,68335	7,60596
33,50	42,50	0,23221	1,08726	3,44066
27,50	48,50	0,00000	0,78788	2,31247
29,50	41,50	0,39967	1,86961	8,51869
26,50	34,50	0,06309	1,32668	4,90139
23,50	39,50	0,02199	0,13538	1,36288
39,50	36,50	0,05412	0,84162	2,35047
33,50	31,50	0,05880	1,15974	3,49132
Sem correção			$Y_{OK}^*(x_o)$	$Z_{MG}^*(x_o)$
			1,48647	5,18961
Método de correção			$Y_{OK}^{**}(x_o)$	$Z_{MG}^{**}(x_o)$
Eqs. 3.31 e 3.35			1,83829	8,38574
Eqs. 3.33 e 3.35			2,08914	8,73018

adição do termo de não viés correto (Eqs. 3.31 e 3.33) pode levar a resultados discordantes da realidade, em termos da distribuição e variabilidade espaciais do fenômeno espacial em estudo.

3.3.3 Krigagem lognormal

A krigagem lognormal se destina à solução de problemas de estimativa quando a variável de interesse $Z(x)$ segue uma distribuição lognormal. A distribuição lognormal se caracteriza por uma distribuição com grande quantidade de valores baixos e uns poucos valores altos, originando uma forte assimetria positiva na distribuição de frequências.

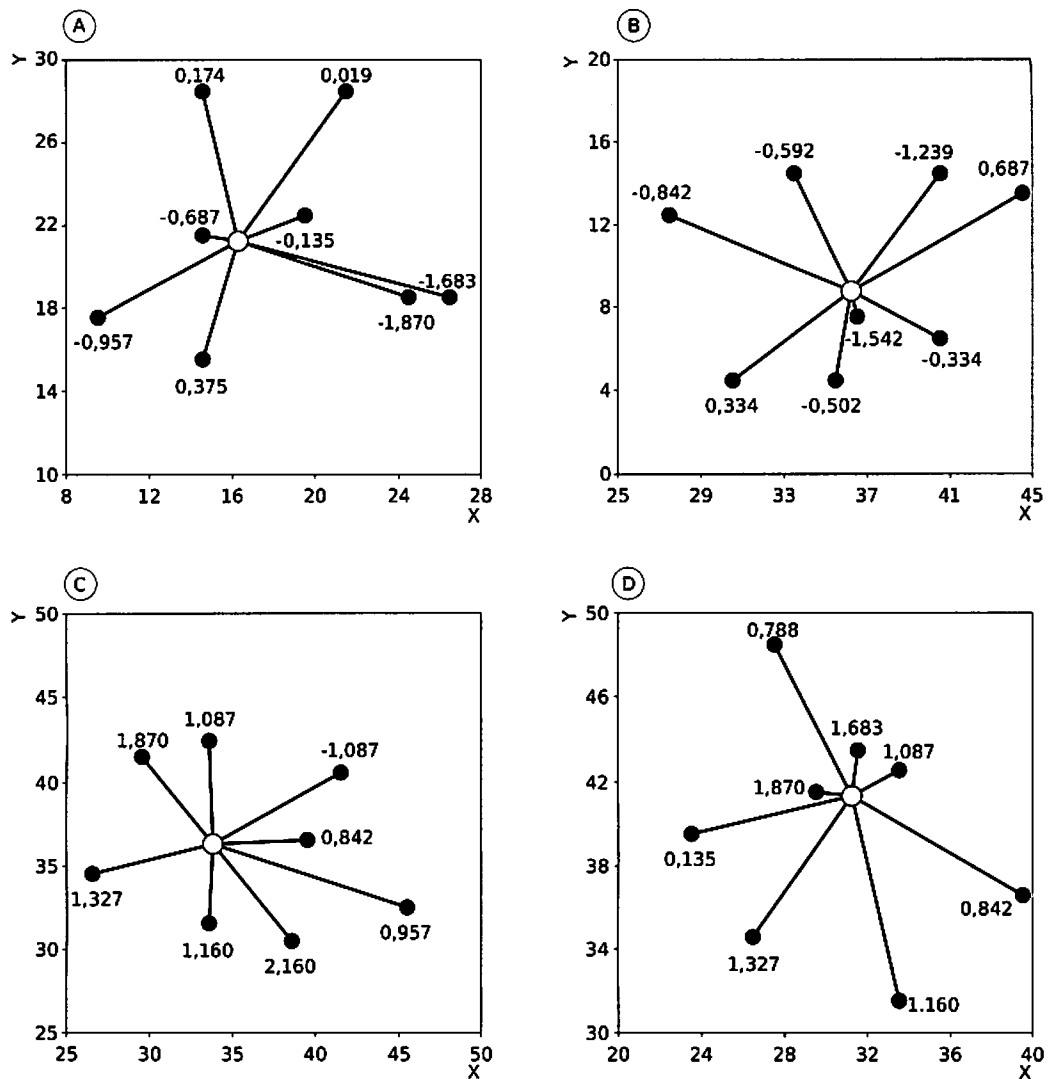


Fig. 3.28 Pontos não amostrados e arranjo de dois pontos mais próximos por quadrante para estimativa por meio da krigagem multigaussiana: A) $x_0 = (16, 25)$; B) $x_0 = (36, 25)$; C) $x_0 = (33, 75)$; D) $x_0 = (31, 25)$. Os valores representam os dados após transformação gaussiana

Na análise estatística, que deve preceder qualquer estudo de natureza geoestatística, pode-se verificar o tipo de distribuição de frequências da variável de interesse examinando o histograma e as estatísticas (média maior que a mediana e coeficiente de variação maior que 1,2). Se for conduzida uma interpolação espacial dos dados originais da variável $Z(x)$, haverá o risco dos poucos valores altos contaminarem regiões com valores baixos. Além disso, se o algoritmo da krigagem ordinária não incluir a eliminação de pesos negativos, há o grande risco de algumas estimativas serem negativas, o que é inaceitável se a variável de interesse for estritamente positiva (p. ex., teores).

Assim, nesses casos, a solução é a transformada logarítmica baseada na Eq. 3.3, segundo Yamamoto e Furuie (2010, p. 6).

Estimadores baseados na transformação logarítmica são denominados krigagem log-normal. Tanto a krigagem simples como a ordinária podem ser usadas como estimadores,

porém a variável de interesse $Z(x)$ deve ser substituída por sua transformada logarítmica $Y(x)$ para estimativa de pontos, áreas ou blocos, por meio da técnica da krigagem ordinária.

O estimador da krigagem lognormal é:

$$Y_{KO}^*(x_o) = \sum_{i=1}^n \lambda_i Y(x_i) \quad (3.36)$$

A transformada reversa é obtida aplicando-se a função inversa da logarítmica, ou seja, a função exponencial. Essa estimativa (Eq. 3.36) poderia ser transformada de volta para a escala original de medida da variável $Z(x)$ sem levar em conta o termo de não viés:

$$Z_{KL}^*(x_o) = \exp(Y_{KO}^*(x_o)) \cdot X_{50} \quad (3.37)$$

Essa expressão será usada apenas para que seu resultado seja comparado com as transformadas reversas corrigidas dos termos de não viés. A mediana multiplica o resultado da exponencial, pois a transformada foi feita após a divisão do valor original pela mediana (Eq. 3.3).

Como a estimativa $Y_{OK}^*(x_o)$ foi obtida com base em valores vizinhos próximos $\{Y(x_i), i = 1, n\}$, ela estará sujeita a uma incerteza, e há necessidade de se adicionar um termo de não viés antes da transformada reversa.

A fórmula clássica usada em Geoestatística para fazer isso é aquela proposta originalmente em Journel (1980, p. 295), seguida depois por Rivoirard (1990, p. 217) e Roth (1998, p. 1.001), entre outros:

$$Y_{KO}^{**}(x_o) = Y_{KO}^*(x_o) + \sigma_{KO}^2/2 - \mu \quad (3.38)$$

Nessa expressão, o termo de não viés é igual à metade da variância de krigagem menos o multiplicador de Lagrange. A transformada reversa dessa estimativa é obtida com:

$$Z_{KOL}^{**}(x_o) = \exp(Y_{KO}^{**}(x_o)) \cdot X_{50} \quad (3.39)$$

As transformadas reversas da krigagem lognormal, segundo Journel e Huijbregts (1978, p. 572), produzem resultados enviesados cuja média é inferior à média amostral.

A origem desse problema é o termo de não viés que não é adequado para fazer a transformada reversa de maneira correta. Assim, como já exposto, será usado como termo de não viés aquele proposto por Yamamoto (2007, p. 221), ou seja, a quantidade de correção da suavização da krigagem.

Para a krigagem lognormal, somente o termo de não viés obtido segundo proposta de Yamamoto (2005, 2007) será considerado, pois a correção de Rezaee, Asghari e Yamamoto (2011) opera melhor no espaço normal, pela natureza da equação de transformação (Eq. 3.32), que usa o escore padrão da distribuição normal.

Dessa forma, a Eq. 3.31 será usada para fazer a transformada reversa da krigagem lognormal, conforme segue (Yamamoto, 2007, p. 222; Yamamoto; Furuie, 2010, p. 6):

$$Z_{KL}^{**}(x_o) = \exp(Y_{KO}^*(x_o) + Y_{NS_o}^*(x_o) \cdot \text{fator}) \cdot X_{50} \quad (3.40)$$



Exemplo de aplicação da krigagem lognormal

Para mostrar o procedimento da krigagem lognormal é tomado como exemplo a amostra extraída do conjunto lognormal composta por 64 pontos de dados (Arquivo 12, Anexo B – Figs. 2.21 e 2.24). Para a krigagem lognormal é necessário o modelo de variograma dos dados transformados para o domínio logarítmico (Fig. 3.29).

O modelo de variograma é descrito pela seguinte equação:

$$\gamma(h) = 1,12 \left[1,5 \frac{h}{11,52} - 0,5 \left(\frac{h}{11,52} \right)^3 \right]$$

Os resultados da krigagem lognormal encontram-se na Fig. 3.30. A correlação entre as imagens pode ser verificada na Fig. 3.31, na qual é possível observar que há uma boa correspondência entre elas.

A essa altura, ainda não é possível afirmar qual a melhor proposta para a transformada reversa, pois os resultados não foram confrontados com a distribuição amostral.

A Fig. 3.32 ilustra essas comparações. Pode-se verificar nessa figura que a distribuição de frequências das transformadas reversas, segundo proposta de Journel (1980, p. 295), é completamente diferente da distribuição amostral, que pode ser confirmada comparando-se com as estatísticas amostrais (Tab. 3.20). O mesmo não acontece, porém, com os resultados da proposta de Yamamoto (2007, p. 222), que são mais próximos das estatísticas amostrais, indicando ser esta uma melhor opção.

No canto superior esquerdo da Fig. 3.32 há o diagrama probabilidade-probabilidade (P-P), no qual os dois valores, 5,55 e 1,05, representam as distâncias médias dos pontos no

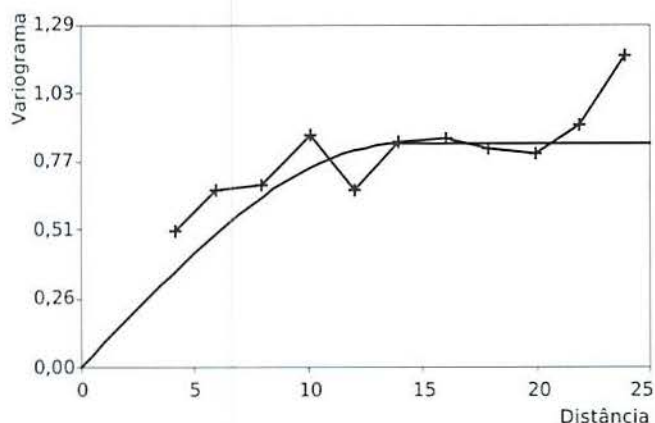


Fig. 3.29 Modelo de variograma para os dados da amostra do conjunto lognormal após transformada logarítmica

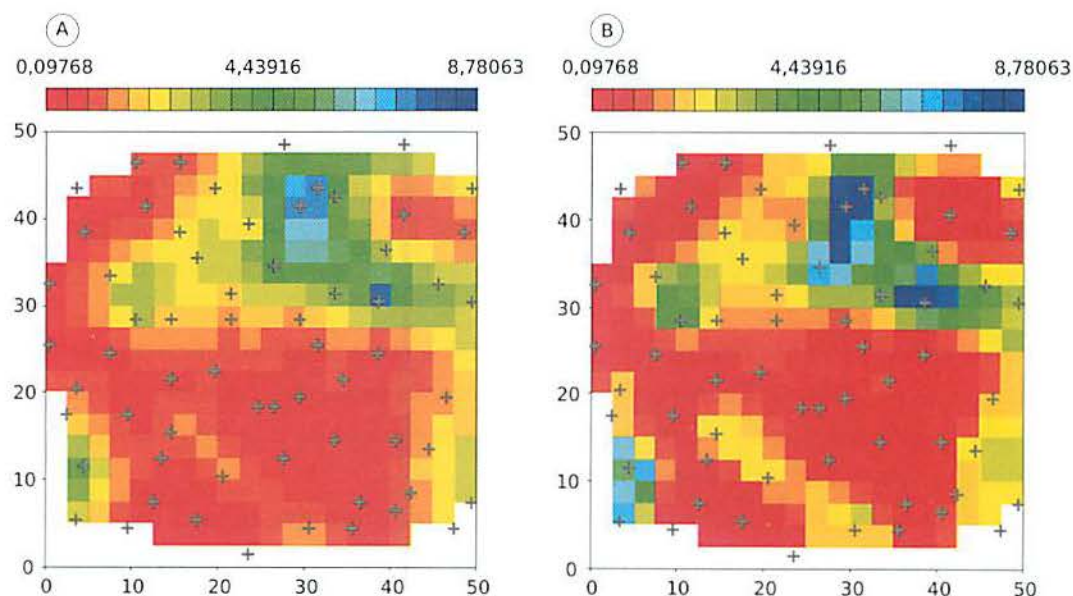


Fig. 3.30 Transformadas reversas das estimativas por krigagem lognormal segundo: A) a Eq. 3.39 e B) a Eq. 3.40

TAB. 3.20 Estatísticas amostrais e das transformadas reversas de estimativas obtidas por krigagem lognormal

Estatística	Amostra	Eq. 3.39	Eq. 3.40
N	64	334	334
Média	1,708	1,690	1,694
Desvio padrão	1,923	1,428	1,871
Coef. var.	1,126	0,845	1,105
Máximo	8,781	7,335	8,781
Quartil sup.	2,113	2,315	2,157
Mediana	1,089	1,134	0,957
Quartil inf.	0,348	0,668	0,417
Mínimo	0,095	0,170	0,098

diagrama P-P em relação à distribuição amostral, ou seja, a reta bissetriz. O primeiro valor se refere à distância entre os pontos obtidos conforme a Eq. 3.39 e o segundo, conforme a Eq. 3.40. Assim, o menor valor significa a maior aderência em relação à distribuição amostral.

Com o objetivo de apresentar passo a passo o procedimento da krigagem lognormal e da transformação reversa, foram usados os mesmos quatro pontos não amostrados, como desenhados na Fig. 3.33 e conforme dados apresentados nas Tabs. 3.21 a 3.24. Para aferir os cálculos, considerar a mediana dos dados igual a 1,08917.

3.3.4 Krigagem indicadora

Uma técnica completamente diferente das já apresentadas trabalha com a variável indicadora, que é obtida por uma transformação não linear, conforme a Eq. 3.4. Esse método se deve a Journel (1983, p. 450-454), que o propôs como adequado para modelar distribuições lognormais.

A krigagem de variáveis indicadoras evita o problema da contaminação pela presença de poucos valores altos na interpolação de regiões com valores baixos.

Para a transformação de uma variável aleatória contínua em uma variável binária, trabalha-se com o conceito do teor de corte/cutoff, termo emprestado da mineração. Dada uma variável aleatória $Z(x)$, pode-se definir um teor de corte, z_c , de tal modo que este esteja no intervalo de amostragem da variável $Z(x)$.

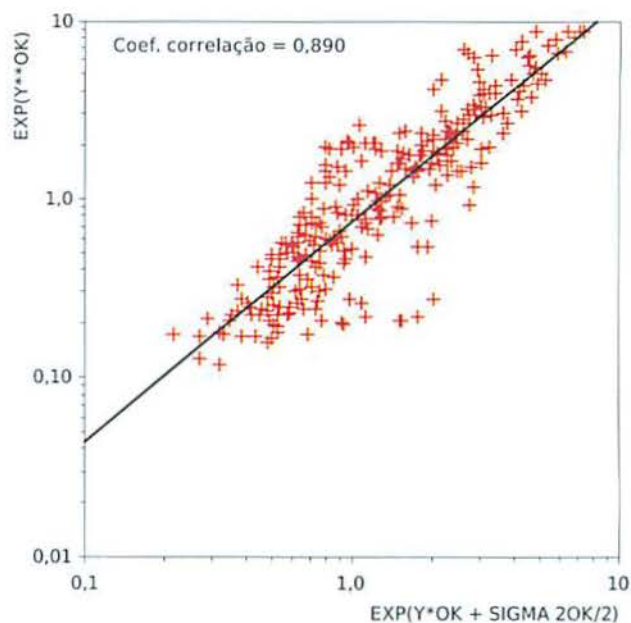


Fig. 3.31 Diagrama de dispersão entre valores obtidos aplicando-se a transformada reversa usando a Eq. 3.39 nas abscissas e a Eq. 3.40 nas ordenadas

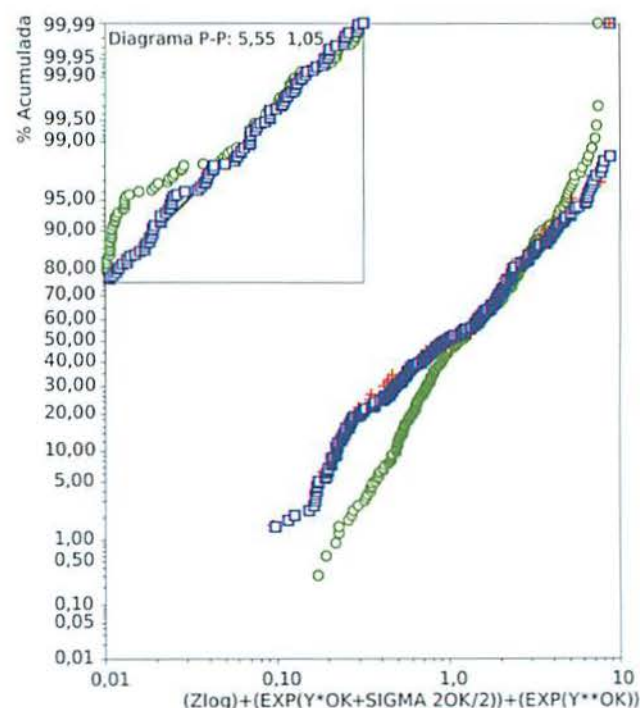


Fig. 3.32 Distribuições acumulativas da distribuição amostral e das transformadas reversas das estimativas obtidas por krigagem lognormal. Cruz vermelha = distribuição amostral; círculo verde = estimativas transformadas conforme Eq. 3.39; quadrado azul = estimativas transformadas conforme Eq. 3.40

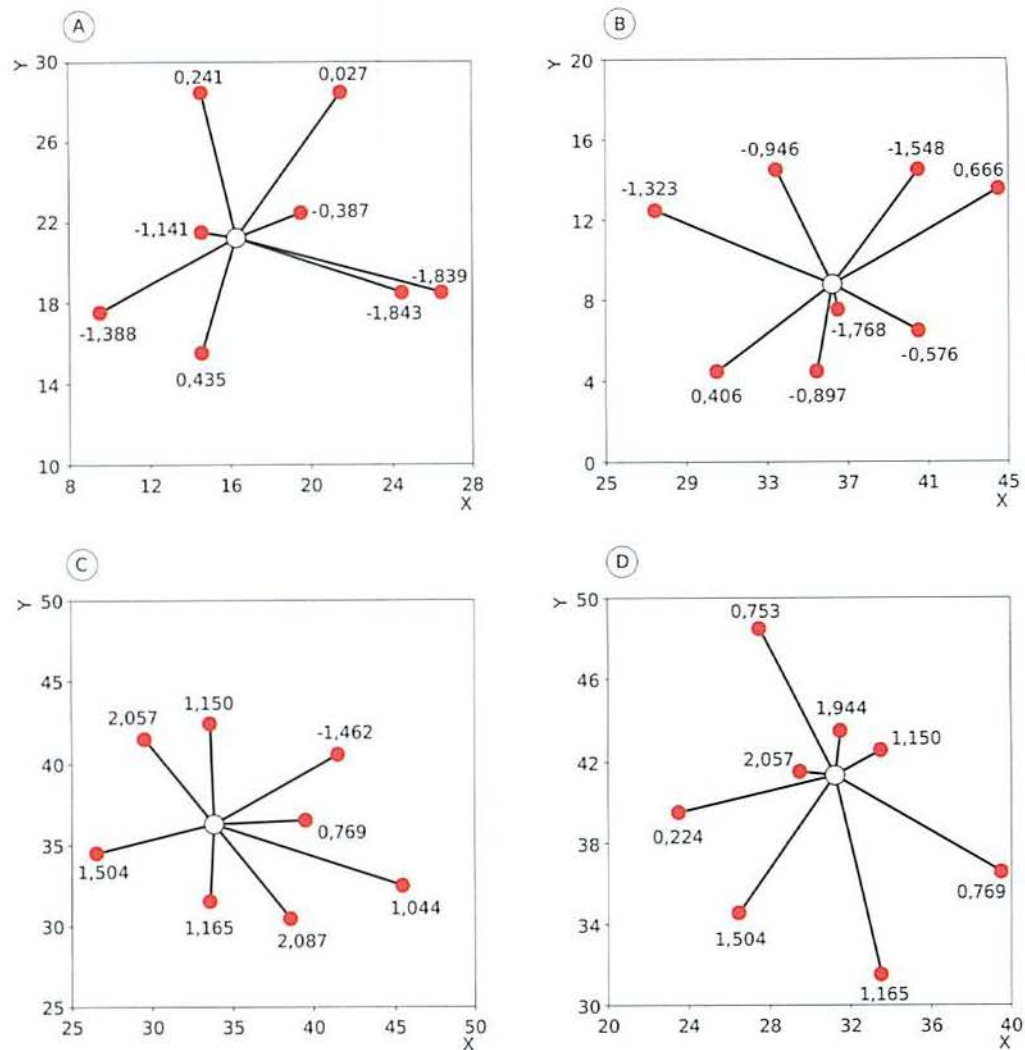


Fig. 3.33 Pontos não amostrados e arranjo de dois pontos mais próximos por quadrante para estimativa por meio da krigagem lognormal: A) $x_o = (16, 25; 21, 25)$; B) $x_o = (36, 25; 8, 75)$; C) $x_o = (33, 75; 36, 25)$; D) $x_o = (31, 25; 41, 25)$. Os valores representam os dados após a transformação lognormal (Eq. 3.3)

A ideia básica é discretizar uma distribuição contínua em K teores de corte, obtendo-se K funções indicadoras. A Fig. 3.34 esquematiza uma distribuição lognormal sendo discretizada em quatro teores de corte.

A krigagem indicadora requer um variograma da variável indicadora para cada teor de corte z_c . Como a função variograma é calculada como a média das diferenças entre pontos separados por uma distância h , nem sempre é possível obter um variograma da indicadora, principalmente para teores de corte nos extremos da distribuição.

Assim, o melhor variograma da variável indicadora corresponde ao teor de corte igual à mediana da distribuição, pois metade dos valores é igual a 1 e outra metade, igual a zero (Journel, 1983, p. 452).

A krigagem indicadora pode ser usada para derivar a função de distribuição acumulativa condicional e, assim, estimar a probabilidade do teor em um ponto não amostrado ser menor que o teor de corte:

$$P(Z(x_o) \leq z_c)$$

Para a construção da função da distribuição acumulativa condicional, é necessário que toda a distribuição da variável de interesse seja amostrada em termos de teores de corte z_c . Isso significa subdividir o intervalo de variação de $Z(x)$ em tantos teores de corte quantos forem necessários. Para cada teor de corte estima-se a probabilidade do ponto não amostrado ser menor que o teor de corte.

TAB. 3.21 Pontos de dados para interpolação do ponto
($x = 16,25$; $y = 21,25$)

X	Y	Pesos	$Y(x_i)$	$Z(x_i)$
21,50	28,50	0,01845	0,02653	1,11846
19,50	22,50	0,26649	-0,38701	0,73964
14,50	28,50	0,04009	0,24140	1,38656
14,50	21,50	0,49897	-1,14094	0,34801
9,50	17,50	0,00554	-1,38759	0,27194
14,50	15,50	0,10685	0,43494	1,68264
24,50	18,50	0,06360	-1,84341	0,17239
26,50	18,50	0,00000	-1,83946	0,17307
Sem correção			$Y_{OK}^*(x_o)$	$Z_{KL}^*(x_o)$
Eqs. 3.36 e 3.37			-0,74072	0,51929
Método de correção			$Y_{OK}^{**}(x_o)$	$Z_{KL}^{**}(x_o)$
Eqs. 3.38 e 3.39			-0,52513	0,64422
Eqs. 3.31 e 3.40			-0,85403	0,46366

TAB. 3.22 Pontos de dados para interpolação do ponto
($x = 36,25$; $y = 8,75$)

X	Y	Pesos	$Y(x_i)$	$Z(x_i)$
44,50	13,50	0,02251	0,66564	2,11924
40,50	14,50	0,08503	-1,54788	0,23167
33,50	14,50	0,12024	-0,94601	0,42291
27,50	12,50	0,03143	-1,32316	0,29004
35,50	4,50	0,00000	-0,89746	0,44395
30,50	4,50	0,04734	0,40556	1,63392
36,50	7,50	0,64730	-1,76837	0,18582
40,50	6,50	0,04614	-0,57640	0,61202
Sem correção			$Y_{OK}^*(x_o)$	$Z_{KL}^*(x_o)$
Eqs. 3.36 e 3.37			-1,42403	0,26221
Método de correção			$Y_{OK}^{**}(x_o)$	$Z_{KL}^{**}(x_o)$
Eqs. 3.38 e 3.39			-1,21563	0,32297
Eqs. 3.31 e 3.40			-1,83046	0,17464

TAB. 3.23 Pontos de dados para interpolação do ponto
($x = 33,75$; $y = 36,25$)

X	Y	Pesos	$Y(x_i)$	$Z(x_i)$
39,50	36,50	0,24551	0,76920	2,35047
41,50	40,50	0,00000	-1,46165	0,25253
33,50	42,50	0,17630	1,15024	3,44066
29,50	41,50	0,12535	2,05684	8,51869
33,50	31,50	0,30701	1,16486	3,49132
26,50	34,50	0,11079	1,50410	4,90139
45,50	32,50	0,02494	1,04412	3,09423
38,50	30,50	0,01009	2,08713	0,61202
Sem correção			$Y_{OK}^*(x_o)$	$Z_{KL}^*(x_o)$
Eqs. 3.36 e 3.37			1,22084	3,69223
Método de correção			$Y_{OK}^{**}(x_o)$	$Z_{KL}^{**}(x_o)$
Eqs. 3.38 e 3.39			1,57882	5,28165
Eqs. 3.31 e 3.40			1,45338	4,65899

TAB. 3.24 Pontos de dados para interpolação do ponto
($x = 31,25$; $y = 41,25$)

X	Y	Pesos	$Y(x_i)$	$Z(x_i)$
31,50	43,50	0,16537	1,94351	7,60596
33,50	42,50	0,24838	1,15024	3,44066
27,50	48,50	0,00000	0,75290	2,31247
29,50	41,50	0,42129	2,05684	8,51869
26,50	34,50	0,05849	1,50410	4,90139
23,50	39,50	0,01372	0,22418	1,36288
39,50	36,50	0,04515	0,76920	2,35047
33,50	31,50	0,04760	1,16486	3,49132
Sem correção			$Y_{OK}^*(x_o)$	$Z_{KL}^*(x_o)$
Eqs. 3.36 e 3.37			1,65485	5,69890
Método de correção			$Y_{OK}^{**}(x_o)$	$Z_{KL}^{**}(x_o)$
Eqs. 3.38 e 3.39			1,85731	6,97779
Eqs. 3.31 e 3.40			2,08585	8,76943



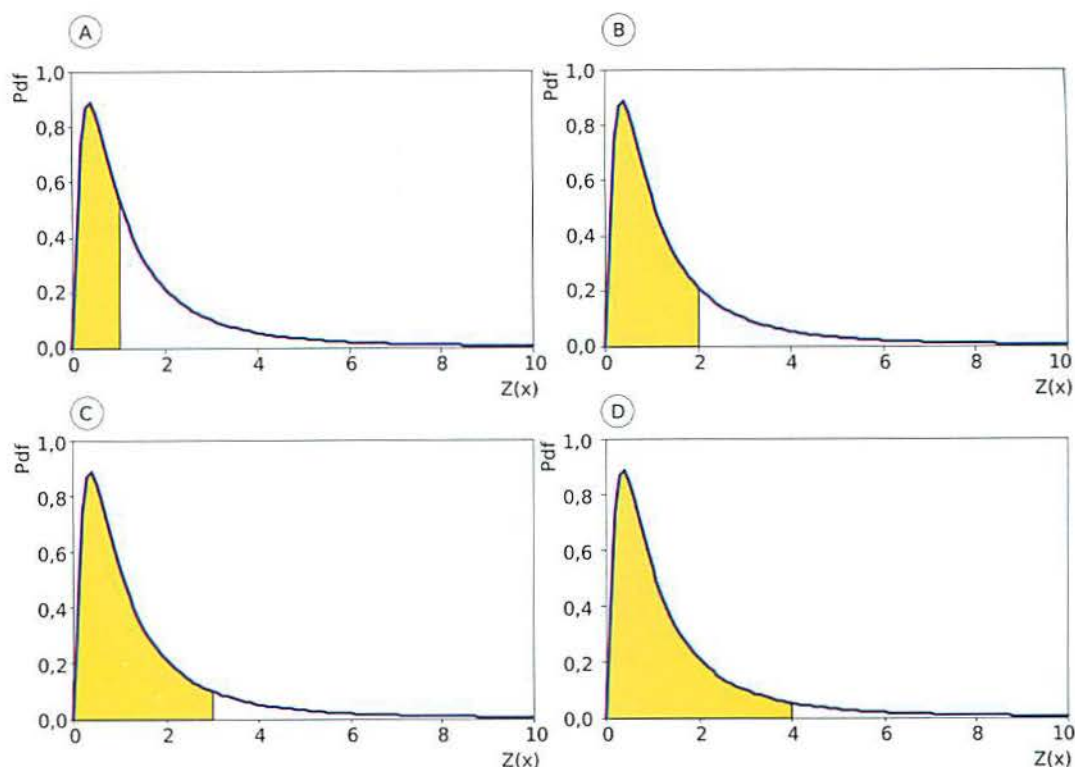


Fig. 3.34 Distribuição lognormal discretizada em quatro teores de corte: A) $z_c = 1$; B) $z_c = 2$; C) $z_c = 3$; D) $z_c = 4$. As áreas destacadas em cinza indicam variáveis indicadoras iguais a 1

As probabilidades $P(Z(x_0) \leq z_{c1}), P(Z(x_0) \leq z_{c2}), \dots, P(Z(x_0) \leq z_{cK})$ associadas aos valores de teores de corte $z_{c1}, z_{c2}, \dots, z_{cK}$ constituem a distribuição de probabilidade de $Z(x)$.

As probabilidades calculadas devem ser monotônicas crescentes, ou seja, $P(Z(x_0) \leq z_{c1}) < P(Z(x_0) \leq z_{c2}) < \dots < P(Z(x_0) \leq z_{cK})$ para $z_{c1} < z_{c2} < \dots < z_{cK}$. Se houver algum caso em que $P(Z(x_0) \leq z_{cK}) > P(Z(x_0) \leq z_{cK+1})$ com $z_{cK} < z_{cK+1}$, diz-se que houve problema de relação de ordem (Hohn, 1998, p. 156-157).

Problemas de relação de ordem acontecem frequentemente, pois a estimativa da probabilidade associada a cada teor de corte é feita com base em um modelo de variograma diferente, ou seja, com diferentes patamares e amplitudes. Assim, para evitar problemas de relação de ordem, Deutsch e Journel (1992, p. 74-75) propuseram o uso da indicadora da mediana, para a qual se calcula e modela um único variograma. Esse modelo de variograma da indicadora da mediana é aplicado para a estimativa de todas as probabilidades $P(Z(x_0) \leq z_{c1}), P(Z(x_0) \leq z_{c2}), \dots, P(Z(x_0) \leq z_{cK})$, as quais são usadas para compor a função de distribuição acumulativa condicional. Com essa função, pode-se determinar a probabilidade $P(Z(x_0) \leq z_c)$ para qualquer valor de teor de corte, mesmo que este não tenha sido amostrado, podendo ela ser obtida por interpolação linear entre dois pontos da função de probabilidade de $Z(x)$. Pelos motivos expostos, neste livro será considerada apenas a krigagem indicadora da mediana.

O estimador da krigagem indicadora pode ser escrito como (Journel, 1980, p. 450):

$$I_{KO}^*(x_o; zc) = \sum_{i=1}^n \lambda_i I(x_i; zc) \quad (3.41)$$

A incerteza associada à estimativa indicadora pode ser medida por meio da variância de interpolação (Yamamoto et al., 2012, p. 147):

$$S_o^2(x_o; zc) = \sum_{i=1}^n \lambda_i [I(x_i; zc) - I_{KO}^*(x_o; zc)]^2$$

Desenvolvendo o termo entre colchetes e simplificando, chega-se a (Yamamoto et al., 2012, p. 147):

$$S_o^2(x_o; zc) = I_{KO}^*(x_o; zc) - (I_{KO}^*(x_o; zc))^2 = I_{KO}^*(x_o; zc) (1 - I_{KO}^*(x_o; zc))$$

Como se pode verificar, a estimativa da krigagem indicadora e sua variância nada mais são que as estatísticas da distribuição de Bernoulli, conforme as Eqs. 3.5 e 3.6.

Na realidade, a Eq. 3.41 faz a estimativa da função de distribuição acumulada condicional, ou seja:

$$F^*(x_o; zc) = I_{KO}^*(x_o; zc)$$

Assim, considerando-se que a variável $Z(x)$ foi discretizada em K teores de corte, a curva da função de distribuição acumulada condicional pode ser construída (Fig. 3.35).

Com base nessa curva, pode-se obter a probabilidade da variável aleatória $Z(x)$ ser menor que um teor de corte zc qualquer. Além disso, da função de distribuição acumulada condicional podem-se derivar a média condicional e a variância condicional. A média condicional é uma estimativa do tipo E (Deutsch; Journel, 1992, p. 76):

$$Z_E^*(x_o) = \sum_{k=2}^K zc_{k,k-1} (F^*(x_o; zc_k) - F^*(x_o; zc_{k-1})) \quad (3.42)$$

em que $zc_{k,k-1} = \frac{zc_k + zc_{k-1}}{2}$ é o ponto médio da classe.

A variância condicional pode ser calculada da seguinte forma (Yamamoto; Furuie, 2010, p. 8):

$$\sigma_E^2(x_o) = \sum_{k=2}^K (F^*(x_o; zc_k) - F^*(x_o; zc_{k-1})) (zc_{k,k-1} - Z_E^*(x_o))^2 \quad (3.43)$$

Pode-se derivar também outra medida de incerteza usando-se as propriedades da distribuição normal, por meio dos percentis 84% e 16%, como segue (Hohn, 1988, p. 164):

$$\text{média} + 1\sigma_\phi = \phi_{84} \quad (3.44)$$

$$\text{média} - 1\sigma_\phi = \phi_{16} \quad (3.45)$$

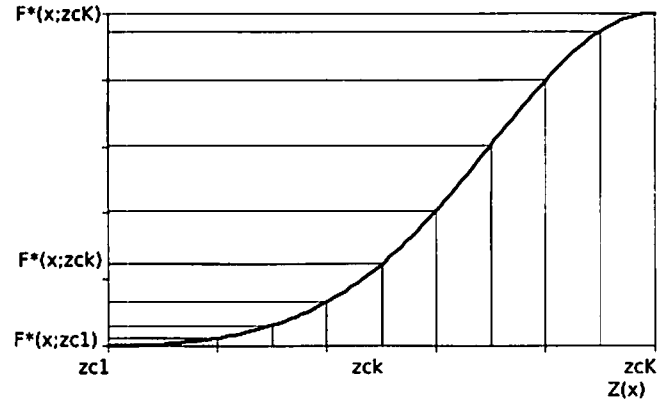


Fig. 3.35 Função de distribuição acumulada condicional obtida pela krigagem indicadora da mediana



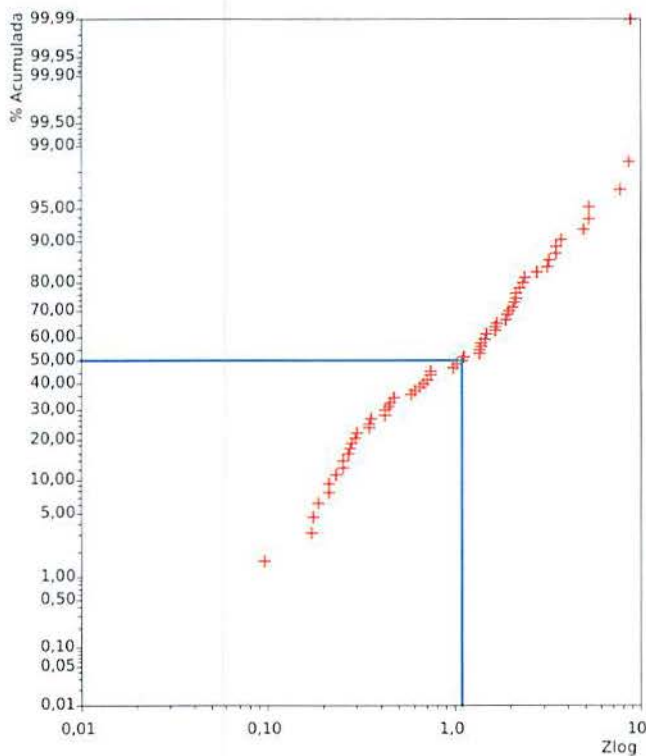


Fig. 3.36 Distribuição de frequências acumuladas da amostra do conjunto lognormal (Arquivo 12, Anexo B), com indicação da mediana igual a 1,089

Substituindo-se a Eq. 3.45 na Eq. 3.44, obtém-se o desvio padrão condicional:

$$\sigma_{\phi} = \frac{\phi_{84} - \phi_{16}}{2} \quad (3.46)$$

Essa fórmula tem a desvantagem de usar apenas dois pontos da função de distribuição acumulada condicional, mas dá uma boa noção da incerteza associada à estimativa do tipo E (Eq. 3.42).

Exemplo de aplicação da krigagem indicadora da mediana

Para ilustrar o procedimento da krigagem indicadora da mediana, ver a amostra do conjunto lognormal (Arquivo 12, Anexo B). O primeiro passo é a determinação da mediana da distribuição, como se apresenta na Fig. 3.36.

Em seguida, faz-se a codificação binária dos dados (Eq. 3.4), conforme o mapa de localização da Fig. 3.37A, com os quais se calcula o variograma da variável indicadora da mediana (Fig. 3.37B).

O modelo de variograma da indicadora da mediana é descrito pela equação:

$$\gamma(h) = 0,244 \left[1,5 \frac{h}{12,096} - 0,5 \left(\frac{h}{12,096} \right)^3 \right] \quad (3.47)$$

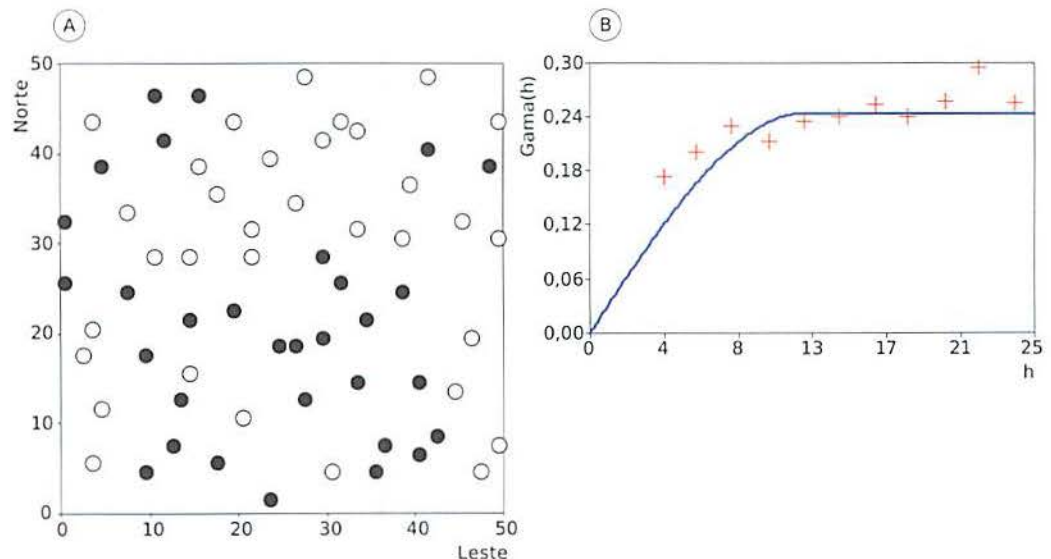


Fig. 3.37 A) Mapa de localização dos pontos de dados da amostra do conjunto lognormal (Arquivo 12, Anexo B), no qual os círculos cheios indicam valores menores que a mediana e os círculos vazios, valores maiores que a mediana; B) variograma da variável indicadora da mediana

As curvas de distribuição acumulativa condicional são calculadas por meio da discretização do intervalo de variação da variável aleatória $Z(x)$ em 9 decis, 10%, 20%, 30%, 40%, 50%, 60%, 70%, 80% e 90%, que correspondem, respectivamente, aos seguintes teores de corte: 0,221; 0,282; 0,429; 0,663; 1,089; 1,456; 1,937; 2,320; 3,618. Os mapas de probabilidade obtidos por meio da krigagem indicadora da mediana para quatro decis (20%, 40%, 60% e 80%) encontram-se na Fig. 3.38.

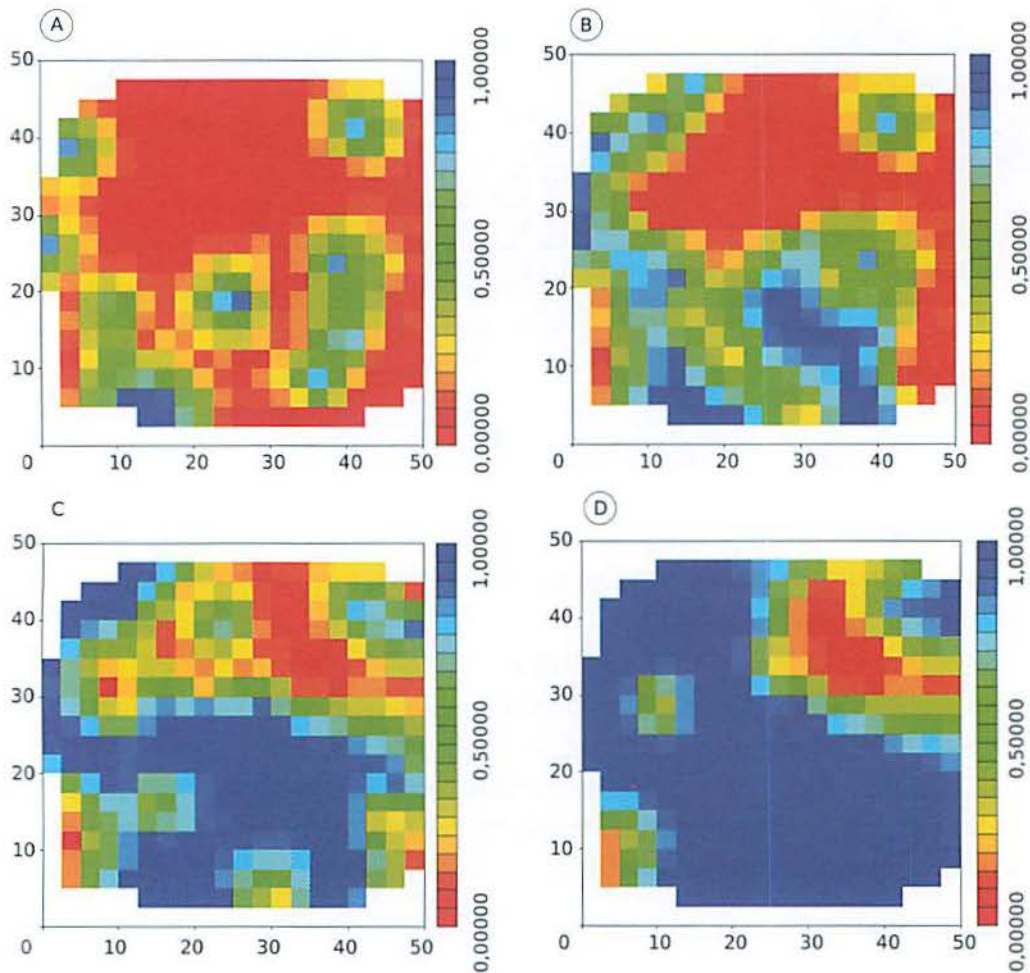


Fig. 3.38 Mapas de probabilidade calculados pela krigagem indicadora da mediana: A) teor de corte = 0,288; B) teor de corte = 0,663; C) teor de corte = 1,456; D) teor de corte = 2,320

Observar que em cada ponto da grade regular tem-se uma função de distribuição acumulativa condicional, ou seja, a função de probabilidade com nove pares ordenados $(z_{C_k}, F^*(x_0; z_{C_k}))$, $k = 1, 2, \dots, 9$. Com base nessas funções, podem-se calcular a média condicional (Eq. 3.42) e as incertezas associadas fornecidas pelo desvio padrão condicional (Eqs. 3.43 e 3.46). A Fig. 3.39 apresenta o mapa de estimativas do tipo E e a Fig. 3.40, os mapas de incertezas.

Como se trata de uma distribuição lognormal, pode-se verificar o efeito proporcional (Fig. 3.41), no qual a incerteza aumenta com a estimativa do tipo E.

A verificação final consiste em comparar a distribuição das estimativas do tipo E com a distribuição amostral (Tab. 3.25).

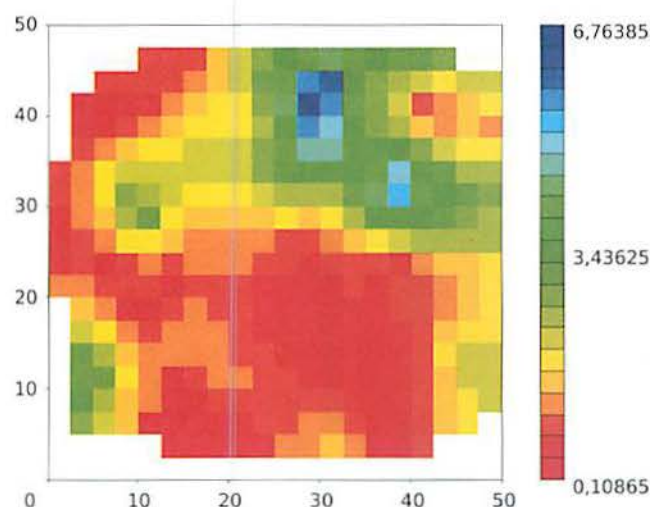


Fig. 3.39 Mapa de estimativas do tipo E

TAB. 3.25 Estatísticas amostrais e das estimativas do tipo E obtidas por krigagem indicadora da mediana

Estatística	Amostra	Eq. 3.42
<i>N</i>	64	334
Média	1,708	1,574
Desvio padrão	1,923	1,173
Coef. var.	1,126	0,745
Máximo	8,781	6,764
Quartil sup.	2,113	2,064
Mediana	1,089	1,195
Quartil inf.	0,348	0,714
Mínimo	0,095	0,109

As estatísticas para a distribuição de estimativas do tipo E mostram grandes diferenças em relação às estatísticas dos dados reais, significando que a estimativa do tipo E não pode ser usada para fazer inferência espacial. Essa conclusão pode ser confirmada comparando-se a distribuição amostral com a calculada por meio das curvas acumulativas (Fig. 3.42), em que as estimativas do tipo E não reproduzem as características da amostra.

Em seguida, o processo de cálculo da krigagem indicadora da mediana será apresentado passo a passo. Para isso, considerem-se dois pontos não amostrados nas coordenadas (16,25; 21,25) e (13,75; 33,75) (Fig. 3.43).

Dado o conjunto de pontos de dados e a localização do ponto não amostrado, e com base no modelo de variograma indicadora da mediana (Eq. 3.47), procede-se à solução do sistema de equações da krigagem ordinária. Os pesos encontrados, bem como a codificação

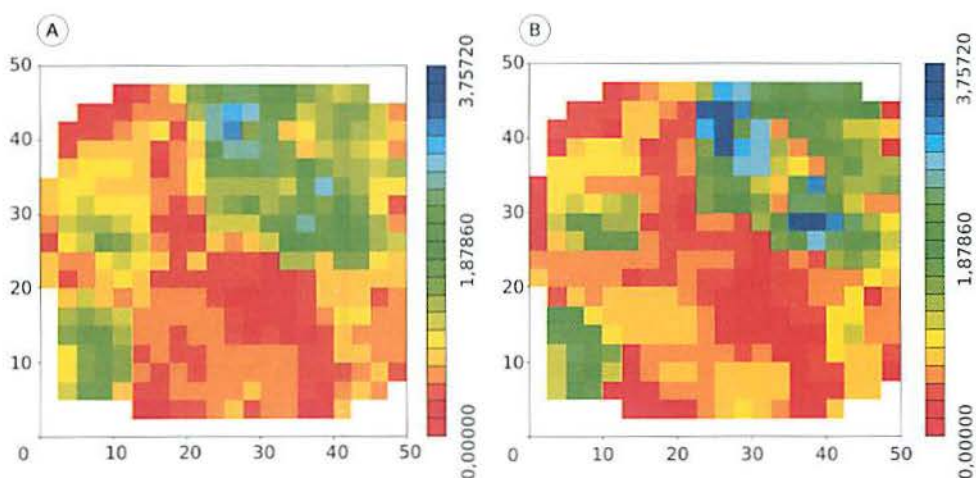


Fig. 3.40 A) Mapa do desvio padrão condicional e B) desvio padrão baseado nos percentis 84% e 16%

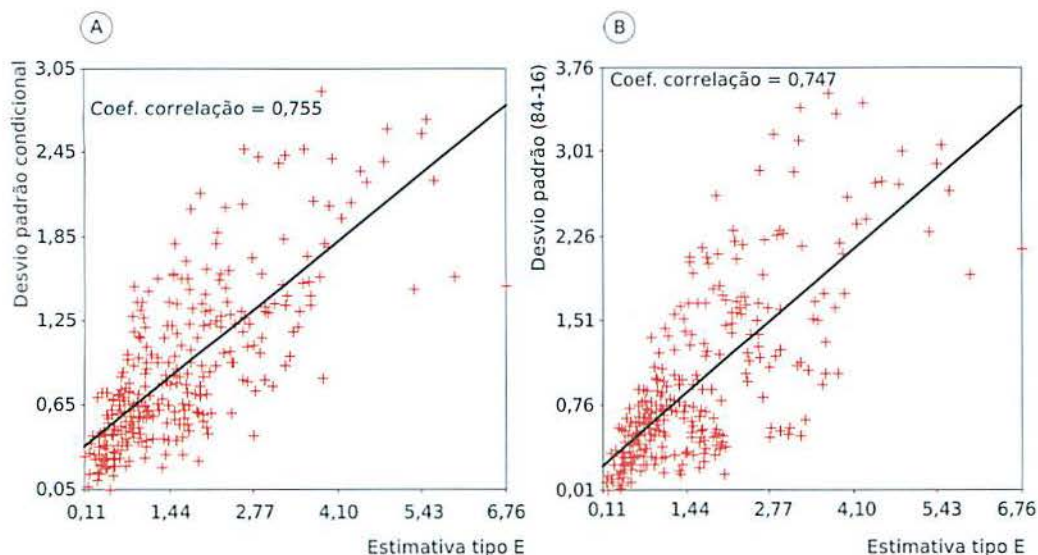


Fig. 3.41 Diagramas de dispersão entre incertezas e estimativas do tipo E: A) desvio padrão condicional (raiz quadrada da Eq. 3.43) e B) desvio padrão dos percentis 84% e 16% (Eq. 3.46)

binária para os pontos não amostrados, encontram-se nas Tabs. 3.26 e 3.27. As Figs. 3.44 (p. 108) e 3.45 (p. 109) apresentam graficamente os resultados da codificação binária.

Com esses dados, pode-se calcular os valores estimados por krigagem indicadora para cada um dos nove decis (Tab. 3.28).

Os resultados da Tab. 3.28 poderão ser aferidos somando os resultados das multiplicações dos pesos pelas funções indicadoras (Eq. 3.41) para cada um dos nove decis (Tabs. 3.26 e 3.27). As funções de distribuição acumulada condicional encontram-se na Fig. 3.46 (p. 110).

A Fig. 3.46B mostra que a função de distribuição acumulada condicional no ponto (13,75; 33,75) não tem fechamento em 1, pois há um ponto de coordenadas (10,50; 28,50) cujo valor é igual a 3,702, mas o último decil (90%) é igual a apenas 3,618. Assim, com apenas nove decis não é possível fazer o fechamento em 1. A solução seria aumentar o número de percentis, mas esse problema deverá permanecer em algum outro ponto, no qual o valor máximo está acima do último percentil.

Com base nessas curvas pode-se derivar a média condicional ou a estimativa do tipo E (Eq. 3.42), bem como a variância condicional (Eq. 3.43), como segue. Primeiro, deve-se calcular os pontos médios das classes (entre dois teores de cortes sucessivos). Assim, para os nove decis (0,221; 0,288; 0,429; 0,663; 1,089; 1,456; 1,937; 2,320; 3,618), os pontos médios das classes são: 0,254; 0,358; 0,546; 0,876; 1,273; 1,696; 2,128; 2,969. Da

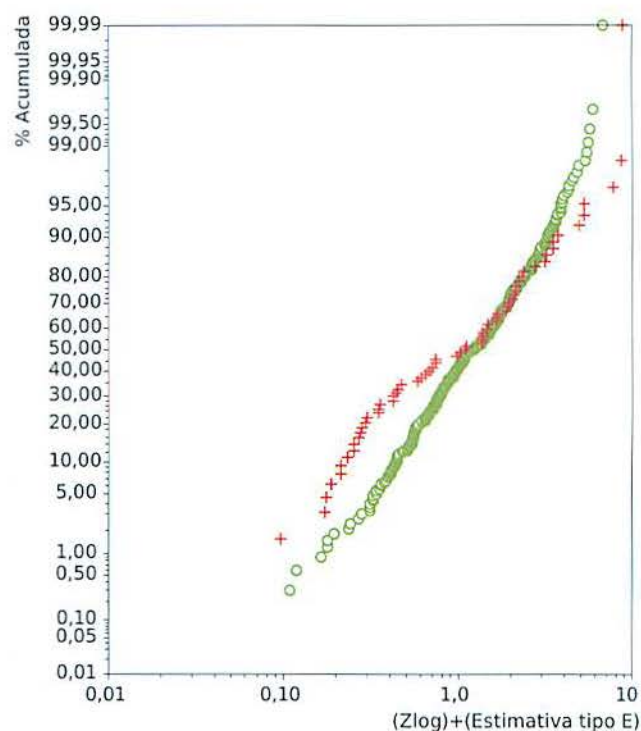


Fig. 3.42 Distribuições acumulativas da distribuição amostral e das estimativas do tipo E obtidas por krigagem indicadora da mediana. Cruz vermelha = distribuição amostral; círculo verde = estimativas do tipo E (Eq. 3.42)

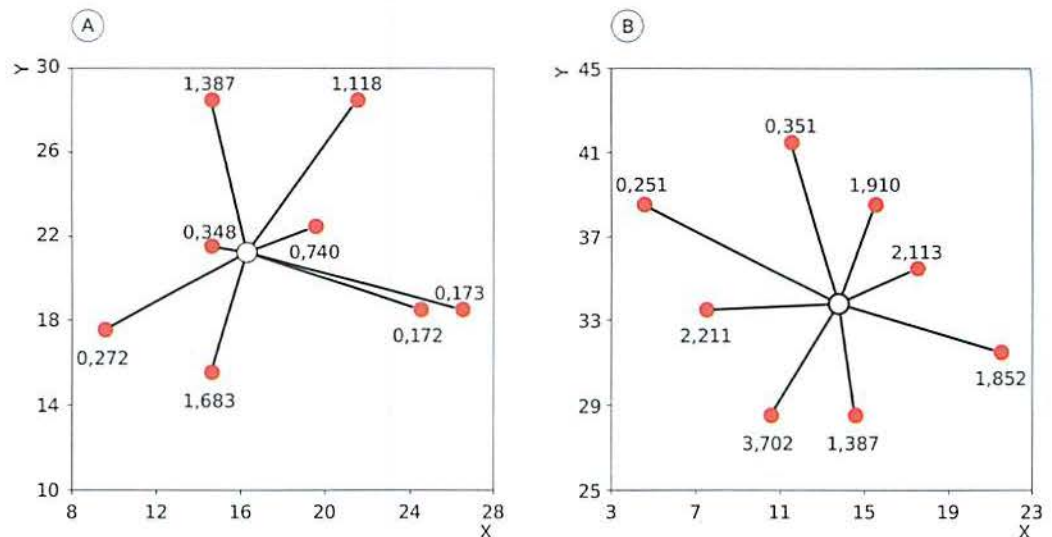


Fig. 3.43 Mapas de localização dos pontos com anotação dos valores da variável aleatória $Z(x)$: A) (16,25; 21,25); B) (13,75; 33,75). Círculos cheios = pontos amostrais; círculos vazios = pontos não amostrados

TAB. 3.26 Pesos da krigagem ordinária e codificação binária dos pontos para os nove decis (16,25; 21,25)

Coordenadas		Pesos	Codificação binária - decis								
X	Y	λ_i	1	2	3	4	5	6	7	8	9
21,50	28,50	0,02013	0	0	0	0	0	1	1	1	1
19,50	22,50	0,26497	0	0	0	0	1	1	1	1	1
14,50	28,50	0,04289	0	0	0	0	0	1	1	1	1
14,50	21,50	0,48450	0	0	1	1	1	1	1	1	1
9,50	17,50	0,01094	0	1	1	1	1	1	1	1	1
14,50	15,50	0,10841	0	0	0	0	0	0	1	1	1
24,50	18,50	0,06819	1	1	1	1	1	1	1	1	1
26,50	18,50	0,00000	1	1	1	1	1	1	1	1	1

mesma forma, calculam-se as diferenças entre duas probabilidades sucessivas. Em seguida, aplicam-se as Eqs. 3.42 e 3.43 para o cálculo da estimativa do tipo E e da variância condicional, respectivamente, conforme a Tab. 3.29.

Como foi apresentado e com base nos resultados, pode-se afirmar que a krigagem indicadora é uma técnica que deve ser considerada para estimar probabilidades e não para derivar mapas de média e variância condicionais. Se o objetivo é a estimativa na presença de assimetria positiva, então há técnicas melhores e mais apropriadas para esse fim, como a krigagem multigaussiana e a krigagem lognormal.

3.4 INTERPOLAÇÃO DE VARIÁVEIS CATEGÓRICAS

A krigagem indicadora seria o método geoestatístico para estimativas de variáveis categóricas, mas essa aplicação exige K variogramas (Leuangthong; Khan; Deutsch, 2008), ou seja,

TAB. 3.27 Pesos da krigagem ordinária e codificação binária dos pontos para os nove decis (13,75; 33,75)

Coordenadas		Pesos	Codificação binária - decis								
X	Y	λ_i	1	2	3	4	5	6	7	8	9
15,50	38,50	0,15522	0	0	0	0	0	0	1	1	1
17,50	35,50	0,28589	0	0	0	0	0	0	0	1	1
4,50	38,50	0,00000	0	1	1	1	1	1	1	1	1
11,50	41,50	0,05203	0	0	1	1	1	1	1	1	1
10,50	28,50	0,08701	0	0	0	0	0	0	0	0	0
7,50	33,50	0,19685	0	0	0	0	0	0	0	1	1
14,50	28,50	0,21210	0	0	0	0	0	1	1	1	1
21,50	31,50	0,01088	0	0	0	0	0	0	1	1	1

TAB. 3.28 Indicadoras estimadas para os nove decis, conforme Eq. 3.41

Decil	zc	$x_o = (16,25; 21,25)$	$x_o = (13,75; 33,75)$
		$I_{KO}^*(x_o; zc)$	$I_{KO}^*(x_o; zc)$
1	0,221	0,06819	0
2	0,288	0,07913	0
3	0,427	0,56363	0,05203
4	0,663	0,56363	0,05203
5	1,089	0,82857	0,05203
6	1,456	0,89159	0,26413
7	1,937	1,00000	0,43024
8	2,320	1,00000	0,91299
9	3,618	1,00000	0,91299

TAB. 3.29 Dados para o cálculo da média e variância condicionais nos pontos (16,25; 21,25) e (13,75; 33,75)

Classe	(16,25; 21,25)	(13,75; 33,75)
$\frac{zC_k + zC_{k-1}}{2}$	$F^*(x_o; zC_k) - F^*(x_o; zC_{k-1})$	$F^*(x_o; zC_k) - F^*(x_o; zC_{k-1})$
0,254	0,01094	0
0,358	0,48450	0,05203
0,546	0	0
0,876	0,26494	0
1,273	0,06302	0,21210
1,696	0,10841	0,16611
2,128	0	0,48275
2,969	0	0
Média	0,672	1,598
Variância	0,197	0,240



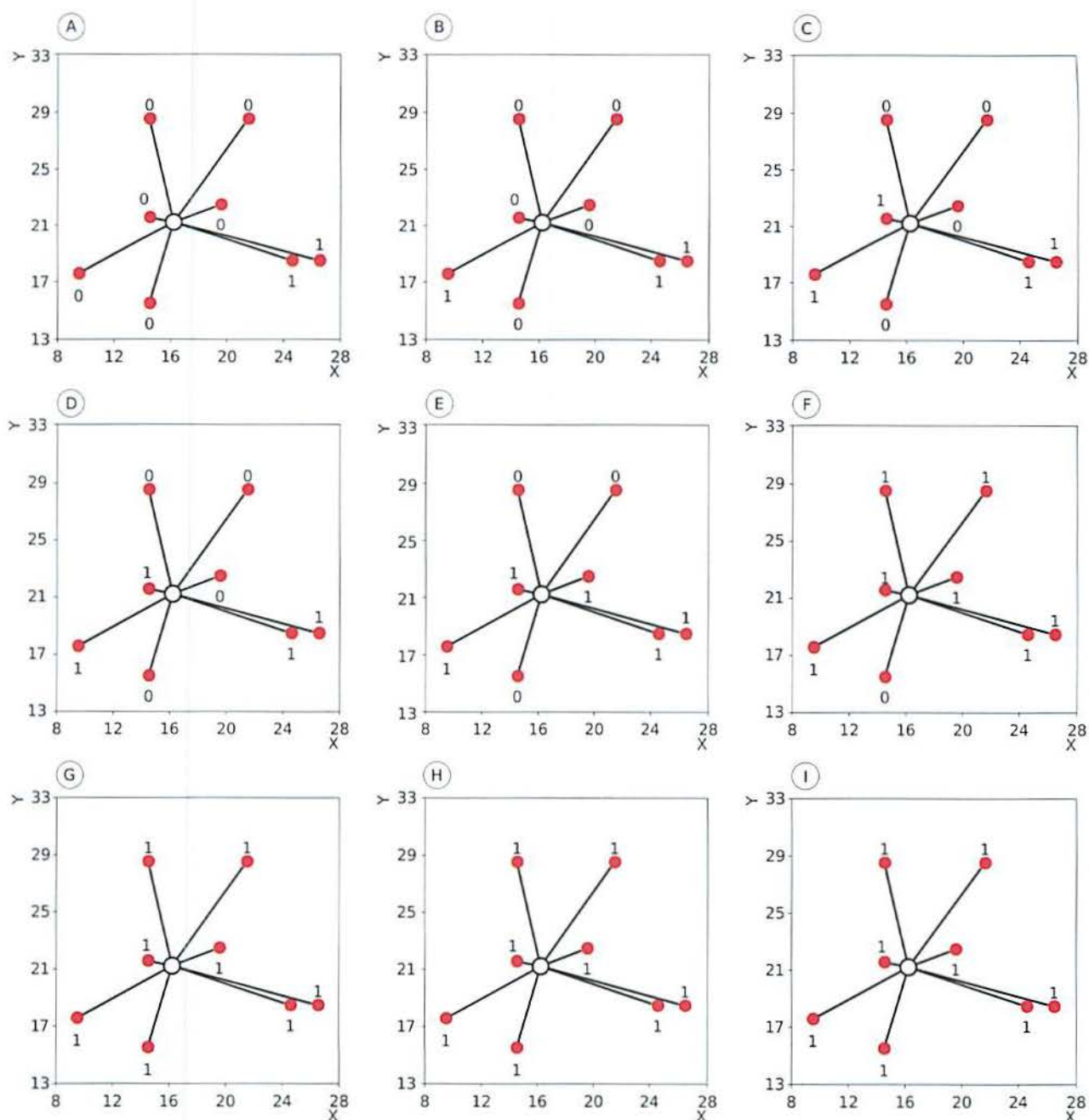


Fig. 3.44 Codificação binária dos pontos de dados vizinhos para estimativa do ponto não amostrado (16,25; 21,25). z_c é igual a : A) 0,221; B) 0,288; C) 0,429; D) 0,663; E) 1,089; F) 1,456; G) 1,937; H) 2,320; I) 3,618

um variograma para cada tipo que compõe a variável categórica. Isso é impossível na prática, pois alguns tipos podem apresentar poucos pares de pontos e, assim, estar sujeitos a grande flutuação estatística (Yamamoto et al., 2012, p. 147). A opção pelas equações multiquádricas é, então, a melhor entre os métodos de interpolação disponíveis.

Na década de 1970, Hardy (1971, p. 1.907-1.908) propôs o uso de equações multiquádricas para a representação analítica da superfície do terreno com base nos pontos amostrais. A proposta original de Hardy estava baseada no conceito de interpolação global, em que

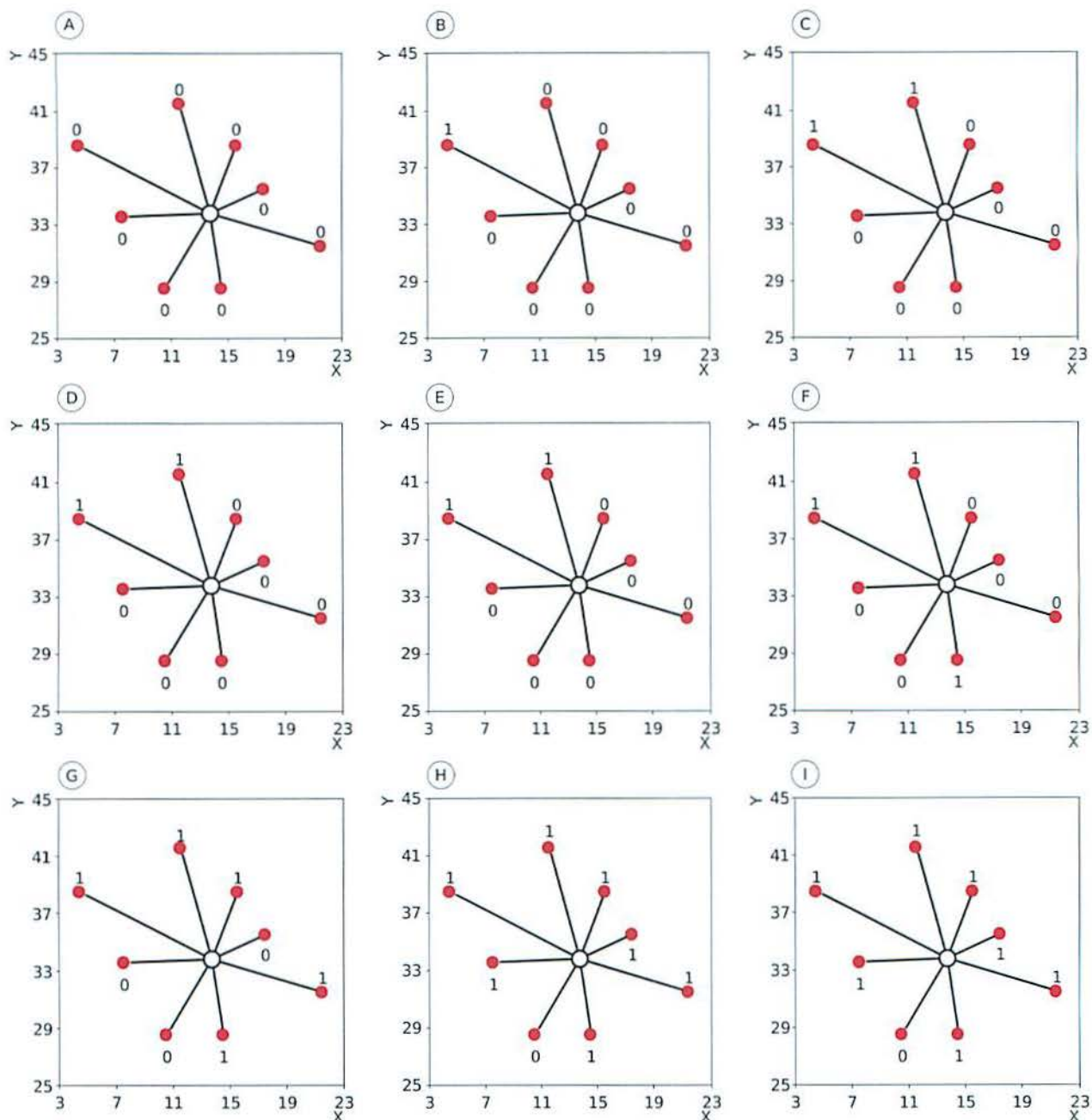


Fig. 3.45 Codificação binária dos pontos de dados vizinhos para estimativa do ponto não amostrado (13,75; 33,75). z_c é igual a: A) 0,221; B) 0,288; C) 0,429; D) 0,663; E) 1,089; F) 1,456; G) 1,937; H) 2,320; I) 3,618

todos os pontos de dados eram considerados simultaneamente. Mais tarde, esse método foi estendido para interpolações locais, usando apenas os pontos da vizinhança mais próximos ao ponto a ser interpolado. Trabalhos posteriores generalizaram as equações multiquádricas, as quais foram denominadas funções de base radial, e, assim, o núcleo multiquádrico se torna um dos muitos núcleos possíveis. Há uma semelhança muito grande das funções de base radial com a krigagem ordinária, cuja função de base radial é a função variograma calculada e modelada com base nos pontos experimentais. A equação geral para dados 2D é:

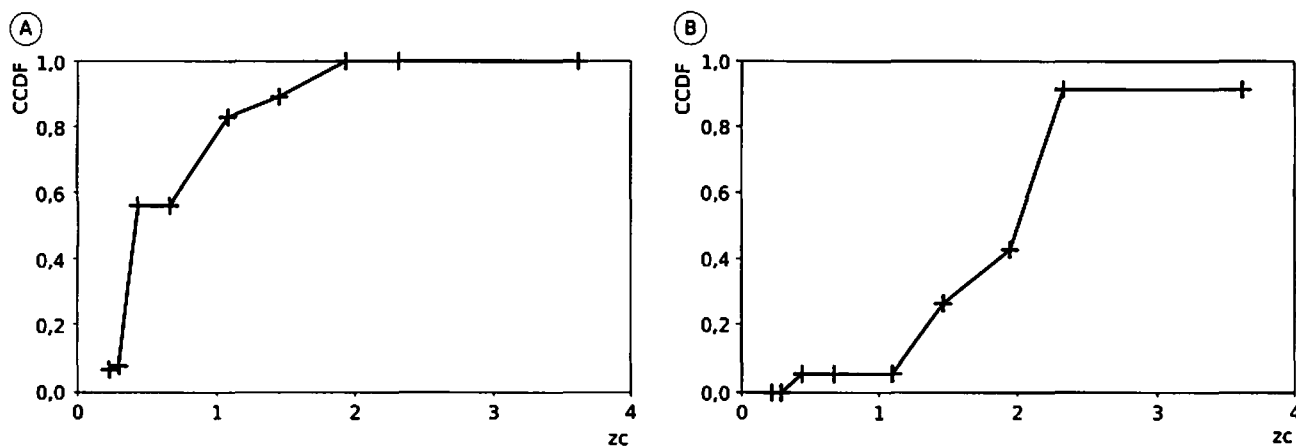


Fig. 3.46 Funções de distribuição acumulada condicional obtidas pela krigagem indicadora da mediana para os pontos não amostrados: A) (16,25; 21,25); B) (13,75; 33,75)

$$Z^*(x_o) = \sum_{i=1}^N C_i [(x_i - x_o)^2 + (y_i - y_o)^2 + C^2]^{1/2} \quad (3.48)$$

em que $\{C_i, i = 1, N\}$ são os coeficientes da equação multiquádrica, C^2 é uma constante positiva e N é o número total de pontos de dados.

Os coeficientes $\{C_i, i = 1, N\}$ são a solução de um sistema de equações lineares:

$$Z(x_i) = \sum_{j=1}^N C_j [(x_i - x_j)^2 + (y_i - y_j)^2 + C^2]^{1/2} \text{ para } i = 1, N \quad (3.49)$$

A Eq. 3.49 pode ser escrita em forma matricial:

$$\begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1N} \\ q_{21} & q_{22} & \cdots & q_{2N} \\ \vdots & \vdots & \cdots & \vdots \\ q_{N1} & q_{N2} & \cdots & q_{NN} \end{bmatrix} \cdot \begin{bmatrix} C_1 \\ C_2 \\ \vdots \\ C_N \end{bmatrix} = \begin{bmatrix} Z(x_1) \\ Z(x_2) \\ \vdots \\ Z(x_N) \end{bmatrix}$$

em que $q_{ij} = [(x_i - x_j)^2 + (y_i - y_j)^2 + C^2]^{1/2}$ é o núcleo multiquádrico calculado entre os pontos i e j .

As equações multiquádricas definidas globalmente, ou seja, sobre todos os pontos de dados, apresentavam uma restrição quanto ao tamanho do sistema de equações lineares para sua resolução em computador. Hoje, com a disponibilidade de memória em computadores, é possível a resolução de sistemas muito grandes. Por outro lado, quando o sistema é muito grande, as operações aritméticas envolvidas produzem erros de arredondamento e truncamento dos valores intermediários, deteriorando a precisão computacional. Assim, a melhor solução é a utilização das equações multiquádricas como método local de interpolação, que é a prática adotada há muitos anos.

Em termos de funções de base radial, a Eq. 3.48 pode ser reescrita, segundo Yamamoto (2002, p. 30-31), como:

$$Z^*(x_o) = \sum_{i=1}^n C_i \phi(x_i - x_o) \quad (3.50)$$

em que $(x_i - x_o)$ é a distância entre o i -ésimo ponto amostral e o ponto a ser interpolado, seguindo a notação geoestatística, e ϕ é a função de base radial. Com base nessas informações, a equação multiquádrica pode ser usada como método local de interpolação, substituindo o N (conjunto de pontos de dados) por n (n pontos vizinhos mais próximos ao ponto a ser interpolado).

Segundo Yamamoto (2002, p. 31), as funções de base radial mais usadas são:

Linear	$\phi(x) = x $
Cúbica	$\phi(x) = x ^3$
Multiquádrica generalizada	$\phi(x) = (c + x ^2)^{\frac{(2k+1)}{2}}$ para $k = -1, 0, \dots$
Splines	$\phi(x) = x ^2 \log x $
Gaussiana	$\phi(x) = \exp(-c x ^2)$

em que $|\cdot|$ é a norma de um vetor em R^n e c é uma constante positiva.

A Eq. 3.50 pode ser escrita de uma forma mais geral com a adição de uma constante a_o (Madych, 1992):

$$Z^*(x_o) = \sum_{i=1}^n C_i \phi(x_i - x_o) + a_o \quad (3.51)$$

Yamamoto (2002, p. 30-32) demonstrou que a Eq. 3.51, considerando todas as condições de restrição impostas, pode ser escrita na forma dual como:

$$Z^*(x_o) = \sum_{i=1}^n W_i Z(x_i) \quad (3.52)$$

Essa equação tem como condição de restrição $\sum_{i=1}^n W_i = 1$, ou seja, condição igual à usada na krigagem ordinária, como se verá.

Os pesos da Eq. 3.52 são obtidos da resolução de um sistema de equações (Yamamoto, 2002, p. 32):

$$\begin{bmatrix} \phi(x_1 - x_1) & \phi(x_1 - x_2) & \cdots & \phi(x_1 - x_n) & 1 \\ \phi(x_2 - x_1) & \phi(x_2 - x_2) & \cdots & \phi(x_2 - x_n) & 1 \\ \vdots & \vdots & \cdots & \vdots & 1 \\ \phi(x_n - x_1) & \phi(x_n - x_2) & \cdots & \phi(x_n - x_n) & 1 \\ 1 & 1 & \cdots & 1 & 0 \end{bmatrix} \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_n \\ \mu \end{bmatrix} = \begin{bmatrix} \phi(x_o - x_1) \\ \phi(x_o - x_2) \\ \vdots \\ \phi(x_o - x_n) \\ 1 \end{bmatrix} \quad (3.53)$$

em que μ é o parâmetro adicional para incluir-se a condição de não viés no sistema de equações multiquádricas.

A interpolação de tipos de uma variável categórica foi possível graças ao trabalho pioneiro de Koike e Matsuda (2005), que propuseram a codificação binária para tal interpolação.

Esse procedimento tem inúmeras aplicações em Geologia, inclusive para obtenção de um mapa geológico preliminar de uma área com base em pontos coletados no campo,

mas, evidentemente, cabe ao geólogo fazer o mapa final a partir das observações e do seu conhecimento sobre a geologia da área.

Outra aplicação seria interpolar a litologia de um bloco de cubagem com base na informação litológica dada pelos pontos amostrais observados nos furos de sonda. As litologias interpoladas poderiam ser usadas para restringir a busca de vizinhos próximos conforme a litologia do ponto não amostrado.

A codificação binária de tipos de uma variável categórica é feita com base na Eq. 3.7, que resulta numa função indicadora para o k -ésimo tipo. Antes de prosseguir na metodologia, é necessário introduzir algumas estatísticas para as variáveis categóricas.

A média da função indicadora, ou seja, a proporção do k -ésimo tipo presente, pode ser calculada como (Yamamoto et al., 2012, p. 147):

$$E[I(x; k)] = \frac{f_k}{N} = p_k$$

em que p_k é a média e $N = \sum_k f_k$ é o número total de pontos da amostra.

A variância da função indicadora, segundo esses mesmos autores, é:

$$Var[I(x; k)] = E[I^2(x; k)] - (E[I(x; k)])^2$$

Como $E[I^2(x; k)] = E[I(x; k)]$, tem-se:

$$Var[I(x; k)] = p_k - p_k^2 = p_k(1 - p_k)$$

Assim, a variância é igual à proporção do k -ésimo tipo multiplicada pela proporção de tipos diferentes de k . A variância de uma variável categórica composta por K tipos pode ser calculada, segundo Kader e Perry (2007), como:

$$\mu_2 = \sum_k^K p_k(1 - p_k)$$

ou seja, trata-se da soma das variâncias parciais.

De acordo com Yamamoto et al. (2012, p. 148), as funções indicadoras obtidas conforme a Eq. 3.7 podem ser manipuladas numericamente para interpolação de um tipo em um ponto não amostrado:

$$i_{MQ}^*(x_o; k) = \sum_{i=1}^n w_i i(x_i; k) \quad (3.54)$$

A variância associada à estimativa (Eq. 3.54), segundo esses autores, pode ser calculada como:

$$s_o^2(x_o; k) = \sum_{i=1}^n w_i [i(x_i; k) - i_{MQ}^*(x_o; k)]^2 \quad (3.55)$$

A Eq. 3.55 é a expressão da variância de interpolação (Yamamoto, 2000, p. 491), que pode ser desenvolvida como (Yamamoto et al., 2012, p. 147):

$$s_o^2(x_o; k) = i_{MQ}^*(x_o; k) - (i_{MQ}^*(x_o; k))^2 = i_{MQ}^*(x_o; k) (1 - i_{MQ}^*(x_o; k)) \quad (3.56)$$

A Eq. 3.56 mostra que a variância associada à interpolação de um tipo da variável categórica é igual à probabilidade de ser o k -ésimo tipo multiplicada pela probabilidade de não ser o k -ésimo tipo.

Exemplo de aplicação da interpolação de variáveis categóricas

Para entender como é feita a interpolação de tipos de uma variável categórica, ver uma amostra usada por Yamamoto et al. (2012, p. 148), composta por 10 pontos de dados (Arquivo 20, Anexo B – Tab. 3.30 e Fig. 3.47).

TAB. 3.30 Pontos de dados de uma variável categórica

Ponto	X	Y	Tipo
1	9,50	28,50	A
2	20,50	37,50	A
3	44,50	41,50	B
4	62,50	44,50	B
5	80,50	30,50	B
6	51,50	9,50	C
7	19,50	21,50	D
8	28,50	20,50	D
9	72,50	17,50	E
10	98,50	5,50	E

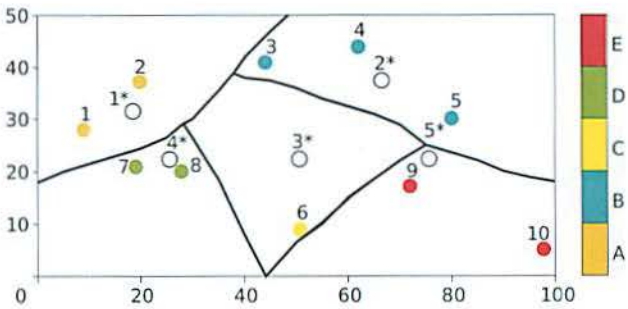


Fig. 3.47 Mapa de localização de tipos de uma variável categórica e os contatos entre os tipos. Os círculos vazios numerados (*) se referem aos pontos a serem interpolados
Fonte: Yamamoto et al. (2012, p. 148).

A primeira etapa é a transformação de variáveis por meio da codificação binária (Eq. 3.7), conforme a Tab. 3.31, que mostra também as proporções calculadas, variâncias e variância total.

TAB. 3.31 Codificação binária dos pontos de dados da Tab. 3.30

Ponto	Coordenadas		Tipos da variável categórica				
	X	Y	$i(x_i; k = A)$	$i(x_i; k = B)$	$i(x_i; k = C)$	$i(x_i; k = D)$	$i(x_i; k = E)$
1	9,50	28,50	1	0	0	0	0
2	20,50	37,50	1	0	0	0	0
3	44,50	41,50	0	1	0	0	0
4	62,50	44,50	0	1	0	0	0
5	80,50	30,50	0	1	0	0	0
6	51,50	9,50	0	0	1	0	0
7	19,50	21,50	0	0	0	1	0
8	28,50	20,50	0	0	0	1	0
9	72,50	17,50	0	0	0	0	1
10	98,50	5,50	0	0	0	0	1
Proporções			0,20	0,30	0,10	0,20	0,20
Variâncias			0,16	0,21	0,09	0,16	0,16
Variância total			0,78				

Observar na Tab. 3.31 que a codificação binária garante que os eventos sejam mutuamente exclusivos, pois a função indicadora transforma o evento em probabilidade. Assim, o objetivo

da interpolação de K tipos de uma variável categórica é determinar a probabilidade que o ponto não interpolado tem de pertencer ao k -ésimo tipo da variável categórica.

Ao interpolar probabilidades, os resultados não podem apresentar valores negativos ou a soma dos tipos interpolados não pode ser maior ou menor que 1. Por isso, os pesos encontrados são corrigidos na eventual presença de pesos negativos, conforme o algoritmo descrito na seção 3.1.5.

Na Fig. 3.47 estão localizados também cinco pontos (numerados de 1* a 5*) que serão interpolados de acordo com a metodologia das equações multiquádricas. As coordenadas dos pontos a serem interpolados encontram-se listadas na Tab. 3.32.

Para ilustrar passo a passo a interpolação de tipos da variável categórica, considerar o ponto 2*, que tem cinco vizinhos mais próximos, como indicado na Fig. 3.48.

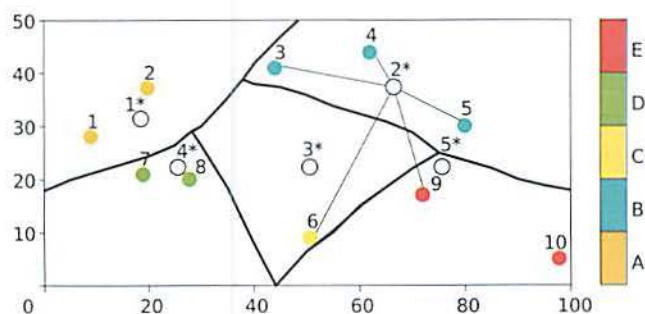


Fig. 3.48 Mapa de localização de pontos e o ponto a ser interpolado, 2*, com indicação dos vizinhos mais próximos

TAB. 3.32 Pontos não amostrados para interpolação de tipos da variável categórica

Ponto	X	Y
1*	18,5	31,5
2*	66,5	37,5
3*	50,5	22,5
4*	25,5	22,5
5*	75,5	22,5

O procedimento para busca dos dois pontos mais próximos por quadrante, dentro de um raio de 36,9, encontrou cinco vizinhos, conforme a Tab. 3.33.

TAB. 3.33 Vizinhos mais próximos ao ponto 2* ($X = 66,5$; $Y = 37,5$)

X	Y	$i(x_i; k = A)$	$i(x_i; k = B)$	$i(x_i; k = C)$	$i(x_i; k = D)$	$i(x_i; k = E)$
62,50	44,50	0	1	0	0	0
44,50	41,50	0	1	0	0	0
51,50	9,50	0	0	1	0	0
72,50	17,50	0	0	0	0	1
80,50	30,50	0	1	0	0	0

Com esses pontos (Tab. 3.33) pode-se organizar o sistema de equações multiquádricas (Eq. 3.53):

$$\begin{bmatrix}
 0 & 18,248 & 36,688 & 28,792 & 22,804 & 1 \\
 18,248 & 0 & 32,757 & 36,878 & 37,643 & 1 \\
 36,688 & 32,757 & 0 & 22,472 & 35,805 & 1 \\
 28,792 & 36,878 & 22,472 & 0 & 15,264 & 1 \\
 22,804 & 37,643 & 35,805 & 15,264 & 0 & 1 \\
 1 & 1 & 1 & 1 & 1 & 0
 \end{bmatrix}
 \begin{bmatrix}
 W_1 \\
 W_2 \\
 W_3 \\
 W_4 \\
 W_5 \\
 \mu
 \end{bmatrix}
 =
 \begin{bmatrix}
 8,062 \\
 22,361 \\
 31,765 \\
 20,881 \\
 15,652 \\
 1
 \end{bmatrix}
 \quad (3.57)$$

Como a constante multiquádrica usada é igual a zero, o núcleo multiquádrico torna-se a distância euclidiana entre os pontos. Por exemplo, a distância entre os pontos 1 e 2 na Tab. 3.33 é:

$$d(x_1 - x_2) = \sqrt{(62,5 - 44,5)^2 + (44,5 - 41,5)^2} = 18,248$$

e a distância entre o ponto 1 (Tab. 3.33) e o ponto 2* é:

$$d(x_1 - x_{2*}) = \sqrt{(66,5 - 62,5)^2 + (37,5 - 44,5)^2} = 8,062$$

Resolvendo o sistema de equações, obtêm-se os coeficientes:

$$W_1 = 0,59458; W_2 = 0,03992; W_3 = 0,03121; W_4 = 0,09790; W_5 = 0,23639; \mu = -2,02062$$

Esses coeficientes são utilizados para interpolar o tipo não amostrado (Eq. 3.54) no ponto 2*. Assim, basta multiplicar os coeficientes pelas funções indicadoras da Tab. 3.34.

TAB. 3.34 Funções indicadoras nos pontos amostrais para interpolação do tipo categórico no ponto 2* e resultados

W_i	$i(x_i; k = A)$	$i(x_i; k = B)$	$i(x_i; k = C)$	$i(x_i; k = D)$	$i(x_i; k = E)$
0,59458	0	1	0	0	0
0,03992	0	1	0	0	0
0,03121	0	0	1	0	0
0,09790	0	0	0	0	1
0,23639	0	1	0	0	0
$i_{MQ}^*(x_o; k)$	0	0,87089	0,03121	0	0,09790
$s_o^2(x_o; k)$	0	0,11244	0,03024	0	0,08831

Observar que o sistema de equações (Eq. 3.57) é resolvido uma única vez e que os coeficientes resultantes são aplicados para todas as K funções indicadoras. Se fosse utilizada a aproximação pela krigagem indicadora, seria necessário resolver tantas vezes quantos fossem os tipos, pois o variograma para cada tipo seria diferente em função da distribuição espacial dos pontos amostrais. O problema da utilização de variogramas diferentes resulta em conjuntos de pesos diferentes para cada tipo e, assim, não garante a condição necessária em probabilidade que $\sum_k^K i^*(x_o; k) = 1$, ou seja, coletivamente completo. Na Tab. 3.34 é fácil verificar que essa condição é satisfeita. Dessa forma, pode-se interpolar os tipos nos demais pontos não amostrados, conforme resultados mostrados na Tab. 3.35.

Uma vez calculadas as probabilidades, deve-se retornar com o tipo correspondente ao ponto não amostrado, dado pelo valor mais provável (Teng; Koike, 2007, p. 533):

$$i_{MQ}^*(x_o; k_{max}) = \max(i_{MQ}^*(x_o; k), k = 1, \dots, K)$$

Na Tab. 3.34, o valor mais provável ocorre para o tipo B (0,87089) que é o tipo interpolado para o ponto 2*. Além disso, deve-se analisar também a incerteza associada. Nesse caso, a variância, igual a 0,11244, pode ser considerada baixa, e o tipo interpolado não está na zona

de incerteza, que foi definida por Yamamoto et al. (2012, p. 151) como sendo a região com interpolações $i_{MQ}^*(x_o; k_{\max}) < 0,6$ e $s_o^2(x_o; k_{\max}) > 0,20$.

TAB. 3.35 Resultados da interpolação dos tipos nos pontos não amostrados (1* a 5*)

Ponto	Estatísticas	$k = A$	$k = B$	$k = C$	$k = D$	$k = E$
1*	$i_{MQ}^*(x_o; k)$	0,73389	0	0	0,26611	0
	$s_o^2(x_o; k)$	0,19530	0	0	0,19530	0
2*	$i_{MQ}^*(x_o; k)$	0	0,87089	0,03121	0	0,09790
	$s_o^2(x_o; k)$	0	0,11244	0,03024	0	0,08831
3*	$i_{MQ}^*(x_o; k)$	0,02363	0,33520	0,34342	0,17277	0,12497
	$s_o^2(x_o; k)$	0,02307	0,22284	0,22548	0,14292	0,10936
4*	$i_{MQ}^*(x_o; k)$	0,08149	0,02706	0,00247	0,88898	0
	$s_o^2(x_o; k)$	0,07485	0,02633	0,00247	0,09869	0
5*	$i_{MQ}^*(x_o; k)$	0	0,38435	0	0	0,61565
	$s_o^2(x_o; k)$	0	0,23663	0	0	0,23663

Embora Yamamoto et al. (2012, p. 151) tenham definido a zona de incerteza com $s_o^2(x_o; k_{\max}) > 0,20$ e $i_{MQ}^*(x_o; k_{\max}) < 0,6$, este último valor deve ser corrigido para 0,72. Assim, a zona de incerteza fica definida como:

$$s_o^2(x_o; k_{\max}) > 0,20 \text{ e } i_{MQ}^*(x_o; k_{\max}) < 0,72$$

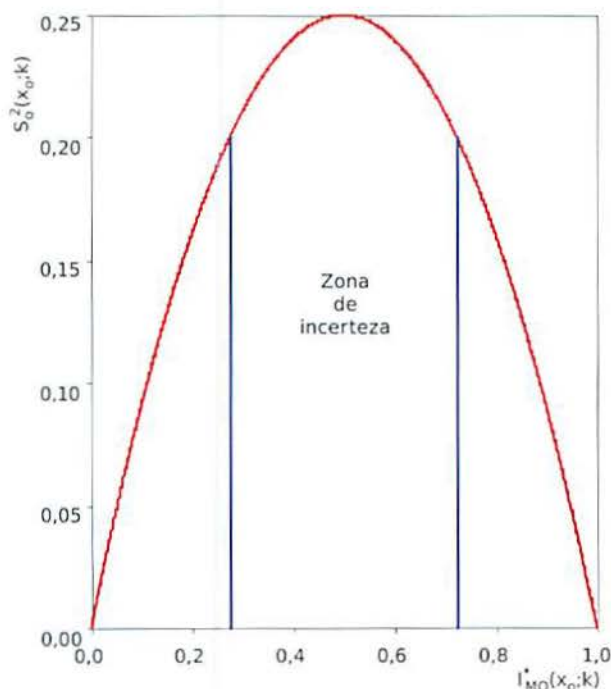


Fig. 3.49 Zona de incerteza definida no intervalo $0,276 < i_{MQ}^*(x_o; k_{\max}) < 0,724$

A Fig. 3.49 mostra o gráfico da variância em função da indicadora interpolada, com a delimitação da zona de incerteza.

É interessante verificar que no gráfico ocorrem duas zonas de certeza: $i_{MQ}^*(x_o; k) < 0,276$ e $i_{MQ}^*(x_o; k) > 0,724$. Isso significa que, quando a indicadora é menor que 0,276, é certo que o valor encontrado não indica o tipo, e que, quando maior que 0,724, há indicação correta do tipo da variável categórica.

Com base nisso, pode-se interpretar os demais pontos interpolados conforme a Tab. 3.35. Aos pontos 1*, 2* e 4* foram atribuídos os tipos A, B e D, respectivamente, enquanto os pontos 3* e 5* caíram na zona de incerteza, por causa da alta variância. Todos os tipos interpolados e a zona de incerteza associada encontram-se na Fig. 3.50, em que se pode verificar que a zona de incerteza é grande pelo pequeno número de pontos na amostra. Evidentemente, com o aumento do tamanho da amostra, a zona de incerteza diminui, conforme demonstrado por Yamamoto et al. (2012, p. 151).

Nesta seção, mostrou-se como é possível a interpolação de tipos de uma variável categórica por meio das equações multiquádricas. Trata-se de uma metodologia ainda não muito empregada, mas com potencial muito grande, principalmente para fins de estimativa de teores condicionada aos pontos vizinhos próximos com a mesma litologia do bloco a ser estimado.

Outra aplicação seria o mapeamento de tipos de solo com base em observações de campo. Além da interpolação propriamente dita, a utilização da incerteza associada é de grande importância para separar eventos mais prováveis daqueles improváveis.

Nesse sentido, a delimitação da zona de incerteza poderia ser aplicada, por exemplo, na separação de minério e rocha encaixante, para fins de estimativa da incerteza volumétrica em avaliação de recursos minerais ou na separação de plumas de contaminação e o seu entorno.

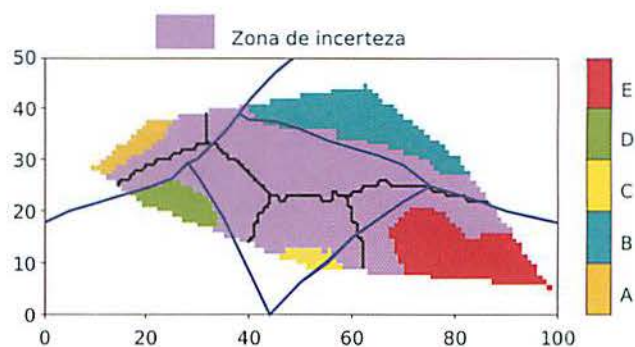


Fig. 3.50 Tipos interpolados e a zona de incerteza. Em preto: contatos interpolados pela máxima probabilidade; em azul: contatos do conjunto completo

Fonte: modificado de Yamamoto et al. (2012, p. 151).

3.5 CONSIDERAÇÕES FINAIS

Foram apresentados os métodos correntes para estimativas geoestatísticas, bem como aplicações típicas. A Fig. 3.51 resume os procedimentos para as estimativas geoestatísticas, as quais foram subdivididas em krigagem linear e não linear. Distribuições de frequências com assimetrias positivas devem passar por transformação de dados para atenuar a influência dos poucos valores altos em regiões com valores baixos.

Três transformações de dados não lineares foram discutidas: gaussiana, logarítmica e indicadora. As duas primeiras procuram normalizar a curva de distribuição de frequências, enquanto a última transforma teores em proporções que se encontram abaixo de um determinado teor de corte. A aproximação por funções indicadoras também pode ser aplicada para variáveis categóricas.

As estimativas geoestatísticas foram subdivididas em krigagem linear e não linear. Na primeira parte, foram considerados os métodos geoestatísticos de krigagem simples, da média e ordinária para os dados brutos ou originais. Na krigagem não linear, foram consideradas estimativas de dados transformados não linearmente: krigagem multigaussiana, lognormal e indicadora.

Com relação aos métodos da krigagem multigaussiana e lognormal, foi ressaltada a importância da inclusão do termo de não viés para as transformadas reversas para a escala original de medida da variável de interesse.

Também foi demonstrada que a aproximação tradicional usando-se a variância de krigagem não funciona, pois se trata apenas de uma medida da configuração espacial de dados vizinhos. Nesse sentido, foram apresentados diversos exemplos ilustrando situações em que a variância de krigagem foi exatamente igual, mesmo para valores e arranjos de pontos completamente diferentes.

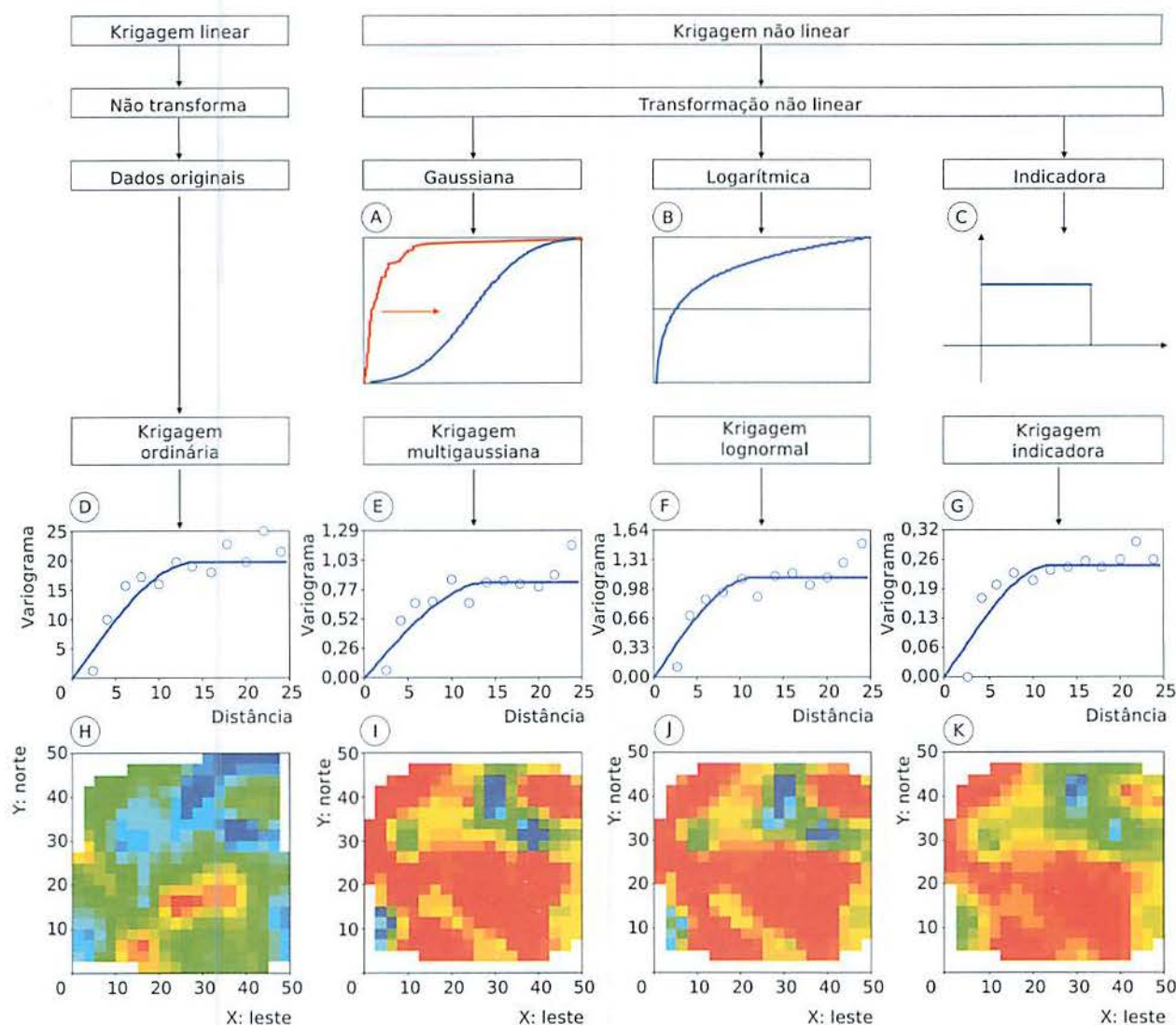


Fig. 3.51 Esquema dos procedimentos de krigagem linear e krigagem não linear: Transformação A) gaussiana; B) logarítmica; C) indicadora; D) variograma dos dados originais; E) variograma da variável transformada para escores normais; F) variograma para dados logarítmicos; G) variograma da variável indicadora da mediana; H) resultados da krigagem linear; I) krigagem multigaussiana; J) krigagem lognormal; e K) krigagem indicadora – estimativas do tipo E

A interpolação de variáveis categóricas foi incluída neste capítulo, apesar de não usar a krigagem indicadora como proposta originalmente, pela impossibilidade de cálculo e modelagem de variogramas experimentais para tipos com poucos pontos de dados.

A interpolação por equações multiquádricas foi escolhida para variáveis categóricas por apresentar grande similaridade com a krigagem ordinária quando esta é feita usando-se um variograma omnidirecional (fenômeno isotrópico). Foi demonstrado que a variância de interpolação pode ser usada para medir a incerteza em torno do valor interpolado do tipo da variável categórica, bem como para mapear a zona de incerteza entre os tipos.

Finalmente, as propostas apresentadas, notadamente aquelas envolvendo a correção do efeito de suavização das estimativas gaussianas e lognormais, devem ser vistas como aproximações possíveis, mas não como as melhores soluções.

Futuramente devem surgir novas propostas que poderão melhorar os resultados apresentados. De qualquer forma, elas devem ser testadas à luz dos dados disponíveis para que sua eficiência, precisão e versatilidade sejam comprovadas. Deve-se lembrar, novamente, que o objetivo da Geoestatística é fazer o melhor uso da informação disponível.





Coestimativas Geoestatísticas

4

No estudo de um fenômeno espacial, diversas variáveis podem ser amostradas simultaneamente nas mesmas localizações ou por métodos distintos de pesquisa em diferentes pontos. Algumas dessas variáveis podem estar subamostradas e outras, superamostradas. Contudo, se essas variáveis subamostradas e superamostradas apresentarem alguma correlação, então as variáveis superamostradas podem ser utilizadas para fazer uma melhor estimativa das variáveis subamostradas (Isaacks; Srivastava, 1989, p. 400). Denomina-se correionalização a existência de duas ou mais variáveis regionalizadas medidas sobre um mesmo campo aleatório (Olea, 1999, p. 209). Geralmente, o padrão de amostragem da variável mais bem amostrada é mais regular que o da variável subamostrada (Olea, 1999, p. 401).

Para essas situações, a Geoestatística proporciona um conjunto de ferramentas para coestimativas. Neste capítulo, serão vistos os métodos conhecidos genericamente como cokrigagem, a *cokrigagem ordinária* e a *cokrigagem colocalizada*, bem como um tipo especial de krigagem denominado *krigagem com deriva externa*. São denominadas *variáveis primárias* aquelas de interesse na pesquisa, mas subamostradas, e *variáveis secundárias* aquelas que podem ser usadas para melhorar a estimativa das variáveis primárias.

Segundo Wackernagel (1995, p. 144), as variáveis primária e secundária podem ser medidas nos mesmos pontos ou em pontos diferentes, configurando três situações (Fig. 4.1):

- isotopia: as variáveis primária e secundária foram medidas nos mesmos pontos de amostragem;
- heterotopia total: as variáveis primária e secundária foram medidas em diferentes localizações;
- heterotopia parcial: as variáveis primária e secundária compartilham alguns pontos comuns.

Para fins de ilustração das técnicas geoestatísticas de coestimativas, deve-se ter conjuntos de dados contendo variáveis primária e secundária que sejam correlacionadas entre si. O ponto de partida, nesse caso, foi o Arquivo completo 1 (disponível em: <<http://lig.igc.usp.br/geoestatistica/anexob/download/Bell.txt>>; Fig. 1.1), cuja informação foi considerada como variável primária. Com base nesse arquivo, foi obtida uma amostra com 289 pontos, a qual foi usada para deteriorar a informação original de forma controlada e, assim, gerar a variável secundária apresentando correlação com a primária. Nesse processo, duas variáveis secundárias com alta e média correlação foram geradas.

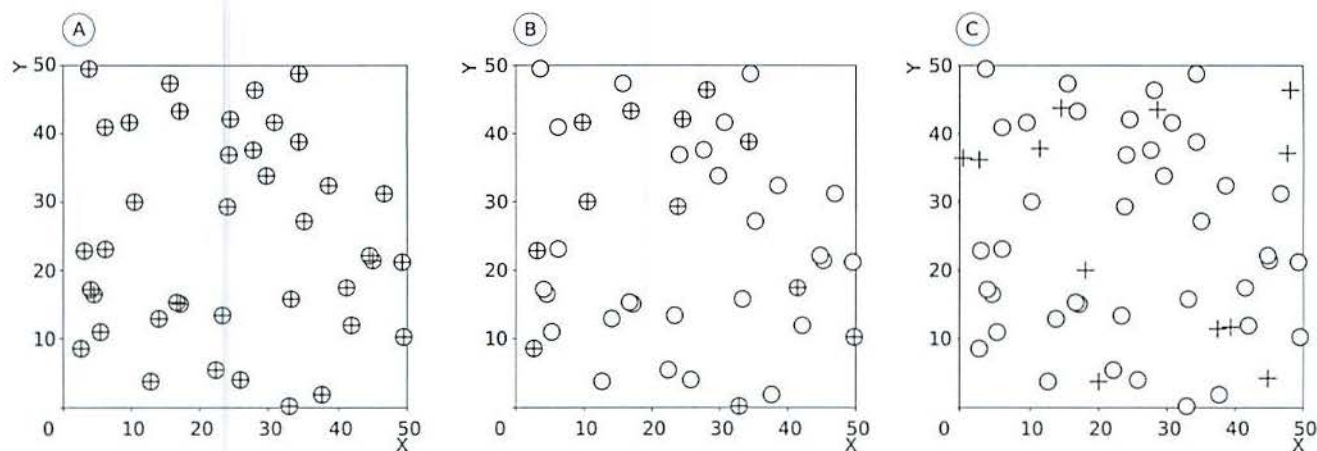


Fig. 4.1 Amostras possíveis para as variáveis primária e secundária: A) isotopia; B) heterotopia parcial; C) heterotopia total. Círculo = variável primária; sinal de mais = variável secundária

A variável com alta correlação foi obtida graças ao ajuste de equações multiquádricas (Hardy, 1971, p. 1.907-1.908) usando uma constante elevada ($C^2 = 100$) que deteriora a precisão da interpolação. A outra variável secundária, com média correlação, foi sintetizada por meio do ajuste de uma superfície de tendência de grau 5 aos pontos da amostra.

A variável primária e as duas variáveis secundárias geradas formam conjuntos completos (Fig. 4.2) compostos por 2.500 pontos distribuídos em uma malha regular de 50 por 50, os quais constituem os dados sintéticos deste capítulo (Arquivo completo 2, disponível em: <<http://lig.igc.usp.br/geostatistica/anexob/download/bellTrend289.txt>>).

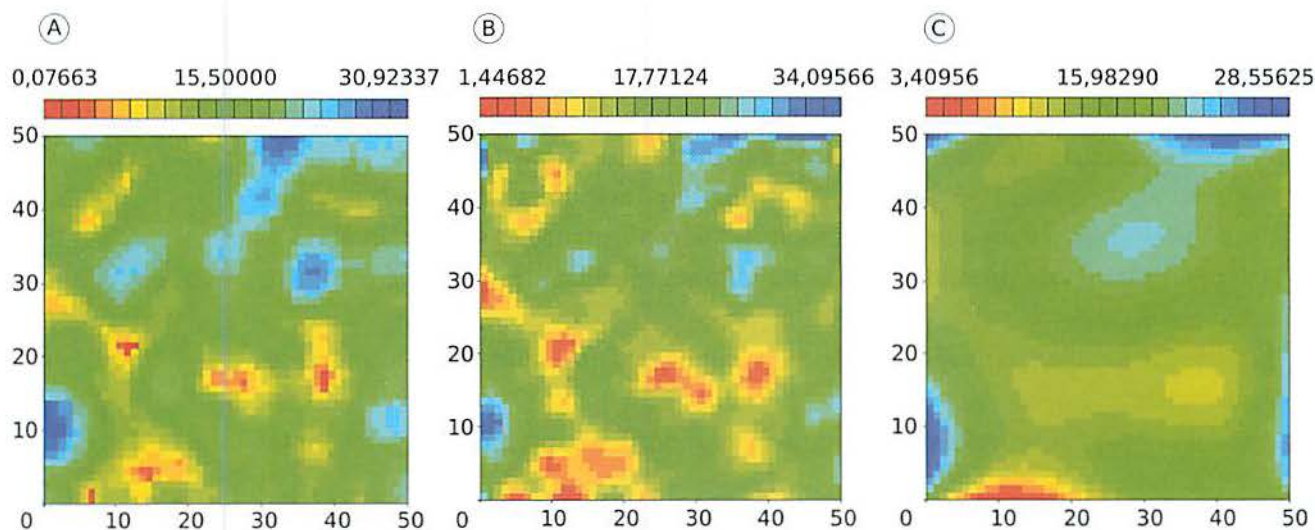


Fig. 4.2 Base de dados completa: A) variável primária (VP); B) variável secundária com alta correlação (VS_1); C) variável secundária com média correlação (VS_2) (Arquivo completo 2, disponível em: <<http://lig.igc.usp.br/geostatistica/anexob/download/bellTrend289.txt>>)

A Fig. 4.3 mostra as correlações entre a variável primária e as variáveis secundárias com alta ($\rho = 0,916$) e média correlação ($\rho = 0,724$). A variável secundária com alta correlação foi denominada VS_1 e a com média correlação, VS_2 .

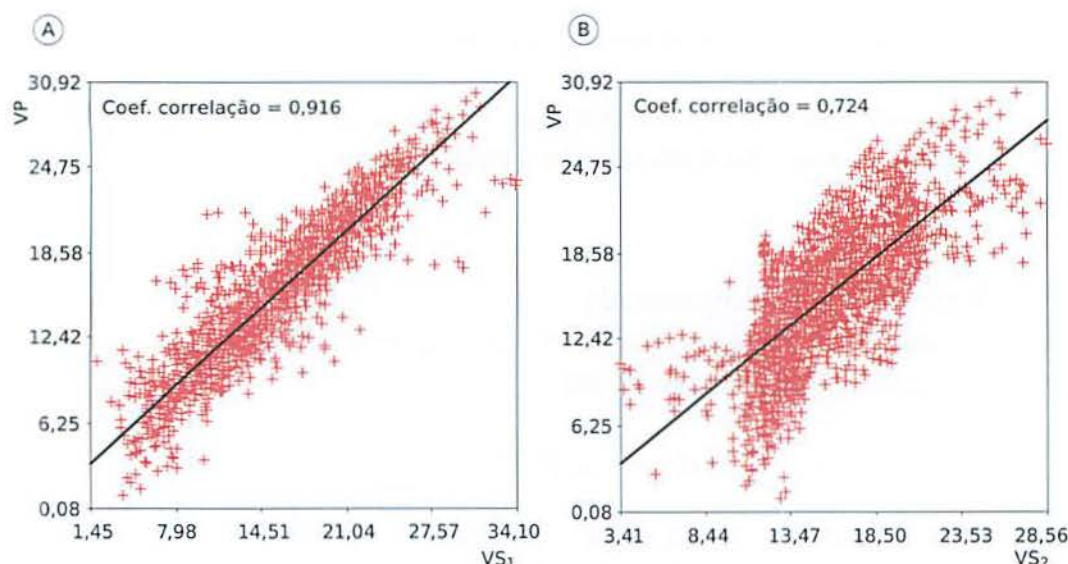


Fig. 4.3 Diagramas de dispersão mostrando a correlação entre as variáveis primária e secundária: A) alta correlação (VS_1) e B) média correlação (VS_2)

Desses conjuntos completos é possível extrair amostras aleatórias estratificadas. Entretanto, como cada técnica requer uma configuração dos pontos da variável secundária, essas amostras serão extraídas quando forem necessárias para ilustrar o procedimento.

4.1 COKRIGAGEM

A cokrigagem é um procedimento geoestatístico pelo qual se pode estimar diversas variáveis regionalizadas em conjunto com base na correlação espacial entre si. É, portanto, uma extensão multivariada do método da krigagem quando, para cada local amostrado, obtém-se um vetor de valores em lugar de um único valor. A cokrigagem é um procedimento verdadeiramente multivariado de estimativa porque o modelo trata com dois ou mais atributos dentro do mesmo campo aleatório (Olea, 1999, p. 209).

Uma de suas mais frequentes aplicações ocorre quando a amostragem de uma variável primária é insuficiente e o objetivo é melhorar a sua estimativa, o que é feito utilizando-se a correlação da variável primária com variáveis secundárias mais densamente amostradas. Ela também é utilizada quando a variável primária exibe uma baixa autocorrelação espacial e as variáveis secundárias apresentam uma alta continuidade. Normalmente, o estudo é feito considerando uma variável primária e apenas uma secundária. Para n variáveis primárias e secundárias, serão necessários $n(n + 1)/2$ variogramas e covariogramas cruzados. No caso de mais de duas variáveis secundárias, o sistema de cokrigagem torna-se extremamente instável em termos numéricos.

Fundamental na utilização da cokrigagem é a verificação prévia da correlação existente entre a variável primária e as variáveis secundárias, a qual deve ser alta para que as estimativas sejam consistentes (Watanabe et al., 2009).

Quando os pontos de amostragem são totalmente coincidentes (isotopia), não se obtém uma melhoria substancial na aplicação da cokrigagem em relação à krigagem ordinária. Por

outro lado, é impossível estimar covariâncias cruzadas com todos os dados não coincidentes (heterotopia total).

A melhoria de interpretação somente é significativa quando a variável primária tem um número extremamente reduzido de casos em relação às variáveis secundárias (heterotopia parcial).

4.1.1 Função variograma cruzado

Se Z_1 e Z_2 são funções aleatórias estacionárias ou intrínsecas, o variograma cruzado delas pode ser calculado como (Olea, 1999, p. 222):

$$\gamma_{12}(h) = \frac{1}{2} E \{ [Z_1(x) - Z_1(x+h)] [Z_2(x) - Z_2(x+h)] \} \quad (4.1)$$

Ao se examinar a Eq. 4.1, verifica-se que o variograma cruzado, ao contrário do variograma normal, permite valores negativos caso a correlação entre a variável primária e a variável secundária seja negativa. O grande problema dessa técnica está no cálculo dos variogramas resultantes das combinações entre variáveis primária e secundária, ou seja, o número total de variogramas será igual a $n \frac{(n+1)}{2}$, em que n são os variogramas diretos e $n \frac{(n-1)}{2}$, os variogramas cruzados. Evidentemente, deve-se fazer a modelagem dos $n \frac{(n+1)}{2}$ variogramas diretos e cruzados, que devem satisfazer o modelo linear de correionalização.

4.1.2 Modelo linear de correionalização

A cokrigagem requer a informação de $n \frac{(n+1)}{2}$ variogramas diretos e cruzados, os quais não podem ser modelados independentemente (Goovaerts, 1997, p. 107), pois devem satisfazer o modelo linear de correionalização.

De acordo com Journel e Huijbregts (1978, p. 171), dadas n funções aleatórias estacionárias de 2ª ordem $\{Y_i(x), i = 1, n\}$ com covariâncias $K_i(h)$ e ortogonais entre si, ou seja, $K_{ij}(h) = 0$, e K funções aleatórias estacionárias de 2ª ordem $\{Z_k(x), k = 1, K\}$, essas podem ser escritas como combinações lineares:

$$Z_k(x) = \sum_{i=1}^n a_{ki} Y_i(x)$$

Segundo esses autores, o conjunto dessas K funções aleatórias tem uma matriz positiva definida:

$$\begin{aligned} C_{kk'}(h) &= \sum_{i=1}^n \sum_{j=1}^n a_{ki} a_{k'j} K_{ij}(h) \\ &= \sum_{i=1}^n a_{ki} a_{k'i} K_i(h) \end{aligned}$$

Observar que a matriz resume-se à sua diagonal principal, pois as covariâncias cruzadas $K_{ij}(h)$ são nulas (funções ortogonais).

Agrupando as funções aleatórias $Y_i^l(x)$ com a mesma covariância direta $K_i(h)$, a covariância cruzada entre $Z_k(x)$ e $Z_{k'}(x)$ pode ser escrita como (Journel; Huijbregts, 1978, p. 172):

$$C_{kk'}(h) = \sum_{i=1}^n \left[\sum_l a_{k_i}^l a_{k'_i}^{l'} \right] K_i(h) = \sum_{i=1}^n b_{kk'}^i K_i(h)$$

Com $b_{kk'}^i = \sum_l a_{k_i}^l a_{k'_i}^{l'}$, $\forall i, i = 1, n$.

Ainda de acordo com Journel e Huijbregts (1978, p. 172), $[b_{kk'}^i]$ é a matriz dos coeficientes e todas as covariâncias diretas e cruzadas podem ser derivadas de combinações lineares de n covariâncias diretas $\{K_i(h), i = 1, n\}$, como se pode ver no modelo linear de correlogionalização a seguir:

$$C_{kk'}(h) = \sum_{i=1}^n b_{kk'}^i K_i(h)$$

Com $b_{kk'}^i = b_{k'k}^i$ para $\forall i$.

Finalmente, segundo eles, para confirmar se a matriz $[b_{kk'}]$ é positiva definida, basta verificar se o determinante é positivo:

$$b_{11} > 0, \begin{vmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{vmatrix} > 0, \dots, \begin{vmatrix} b_{11} & b_{12} & \dots & b_{1K} \\ b_{21} & b_{22} & \dots & b_{2K} \\ \vdots & \vdots & \dots & \vdots \\ b_{K1} & b_{K2} & \dots & b_{KK} \end{vmatrix} > 0$$

O modelo linear de correlogionalização também pode ser escrito em termos da função variograma (Journel; Huijbregts, 1978, p. 173):

$$\gamma_{kk'}(h) = \sum_{i=1}^n b_{kk'}^i \gamma_i(h)$$

Com $b_{kk'}^i = b_{k'k}^i$ para $\forall i$.

4.1.3 Cokrigagem ordinária

Dado um conjunto de N variáveis $\{Z_i(x), i = 1, N\}$, o estimador da cokrigagem ordinária, segundo Wackernagel (1995, p. 145), pode ser escrito como:

$$Z_{i_o}^*(x_o) = \sum_{i=1}^N \sum_{\alpha_i} w_{\alpha_i} Z_i(x_{\alpha_i}) \quad (4.2)$$

em que o índice i_o refere-se à variável primária do conjunto de N variáveis.

A esperança do erro envolvido (Journel; Huijbregts, 1978, p. 325), por sua vez, é:

$$E[Z_{i_o}(x_o) - Z_{i_o}^*(x_o)] = E[Z_{i_o}(x_o)] - \sum_{\alpha_{i_o}} w_{\alpha_{i_o}} E[Z_{i_o}(x_{\alpha_{i_o}})] - \sum_{i \neq i_o} \sum_{\alpha_i} w_{\alpha_i} E[Z_i(x_{\alpha_i})]$$

Como $E[Z_{i_o}(x_o)] = E[Z_{i_o}(x_{\alpha_{i_o}})] = m_{i_o}$ e $E[Z_i(x_{\alpha_i})] = m_i$, tem-se, de acordo com esses autores:

$$m_{i_o} \left[1 - \sum_{\alpha_{i_o}} w_{\alpha_{i_o}} \right] - \sum_{i \neq i_o} m_i \sum_{\alpha_i} w_{\alpha_i}$$

Portanto, segundo eles, para que a esperança do erro seja nula, deve-se fazer $\sum_{\alpha_{i_o}} w_{\alpha_{i_o}} = 1$ e $\sum_{\alpha_i} w_{\alpha_i} = 0$, que são as duas condições de restrição impostas aos pesos da cokrigagem ordinária.



Tais condições também podem ser descritas pela seguinte relação (Wackernagel, 1995, p. 145):

$$\sum_{\alpha_i=1}^{n_i} w_{\alpha_i} = \delta_{ii_0} = \begin{cases} 1 & \text{se } i = i_0 \\ 0 & \text{se } i \neq i_0 \end{cases} \quad (4.3)$$

Para a determinação dos pesos da cokrigagem ordinária que permitam fazer a estimativa conforme a Eq. 4.2, encontra-se a variância do erro (Wackernagel, 1995, p. 145):

$$\sigma_E^2 = E \left[\left(\sum_{i=1}^N \sum_{\alpha=1}^{n_i} w_{\alpha_i} Z_i(x_{\alpha_i}) - Z_{i_0}(x_0) \right)^2 \right] \quad (4.4)$$

Desenvolvendo-se a expressão da variância do erro (Eq. 4.4) e impondo-se as condições de restrição (Eq. 4.3), chega-se à lagrangiana com N multiplicadores de Lagrange, a qual, derivada em relação aos pesos e aos multiplicadores de Lagrange, resulta, ainda segundo Wackernagel (1995, p. 146), no sistema de equações de cokrigagem ordinária:

$$\begin{cases} \sum_{j=1}^N \sum_{\beta=1}^{n_j} w_{\beta_j} \gamma_{ij}(x_{\alpha} - x_{\beta}) + \mu_i = \gamma_{ii_0}(x_{\alpha} - x_0) & \text{para } i = 1, N \text{ e } \alpha = 1, n_i \\ \sum_{\beta=1}^{n_i} w_{\beta_i} = \delta_{ii_0} & \text{para } i = 1, N \end{cases}$$

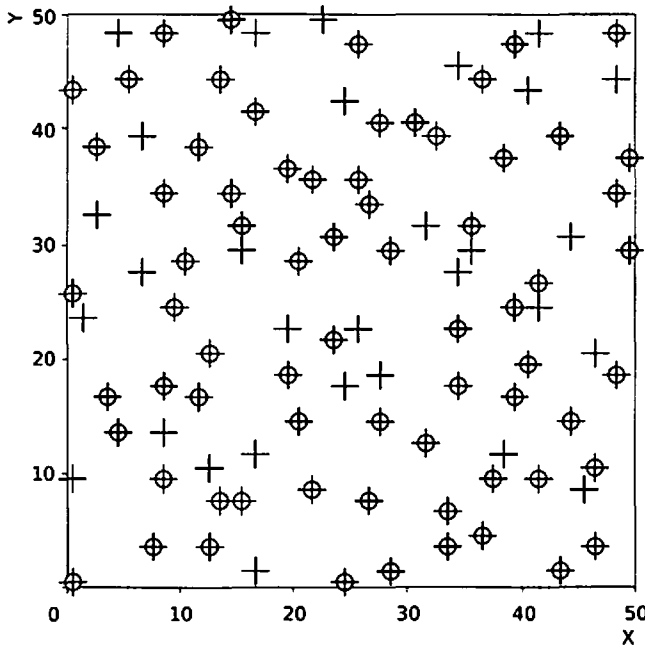


Fig. 4.4 Mapa de localização de pontos para cokrigagem ordinária. Círculo = variável primária; sinal de mais = variável secundária (Arquivos 14 e 15, Anexo B)

A lagrangiana para $N = 2$ variáveis pode ser confeccionada em Isaaks e Srivastava (1989, p. 403) e Goovaerts (1997, p. 224), que demonstram que a minimização dessa função objetivo resulta no sistema de equações de cokrigagem ordinária.

Exemplo de aplicação da cokrigagem ordinária

Da base de dados completa (Fig. 4.2) foi extraída uma amostra aleatória estratificada com heterotopia parcial (Fig. 4.4). Essa amostra é composta por 100 pontos, dos quais 70 são pontos com informação tanto da variável primária quanto da variável secundária e 30 são pontos com informação secundária, que dá a heterotopia parcial a esse conjunto de dados.

A Fig. 4.5 mostra os diagramas de dispersão com correlações iguais a 0,951 e 0,777 para, respectivamente, as variáveis secundárias com alta (VS_1) e média correlação (VS_2).

Assim, esse conjunto foi analisado para determinar o modelo de correlação espacial para as variáveis primária e secundária, bem como para as variáveis cruzadas, conforme os resultados da Fig. 4.6. Os modelos de variogramas dessa figura são descritos pelas Eqs. 4.5 e 4.6.

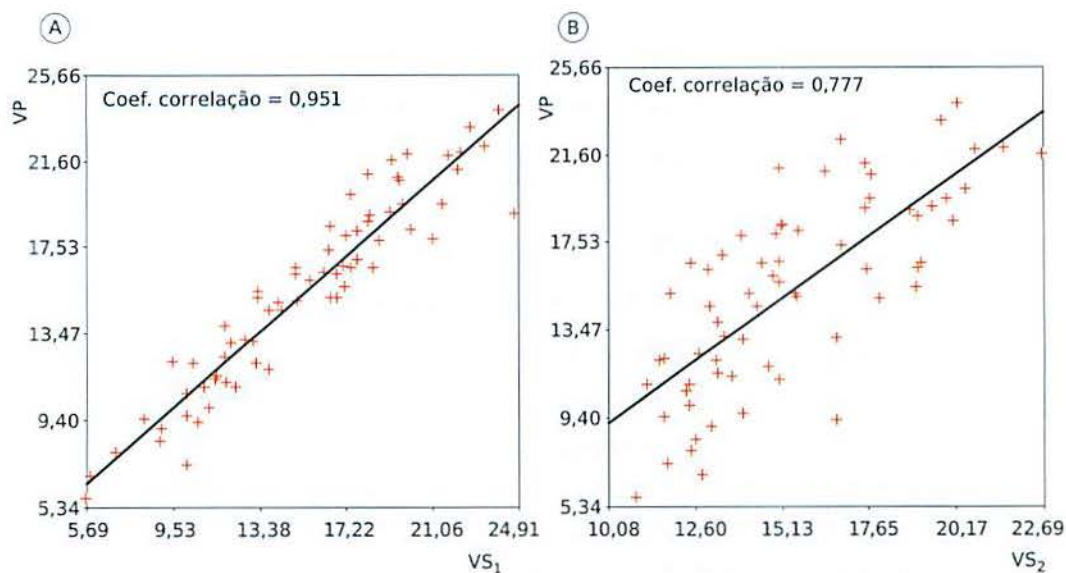


Fig. 4.5 Diagramas de dispersão entre a variável primária e as variáveis secundárias para a amostra estratificada com 100 pontos de dados: A) VS_1 (Arquivo 14, Anexo B) e B) VS_2 (Arquivo 15, Anexo B)

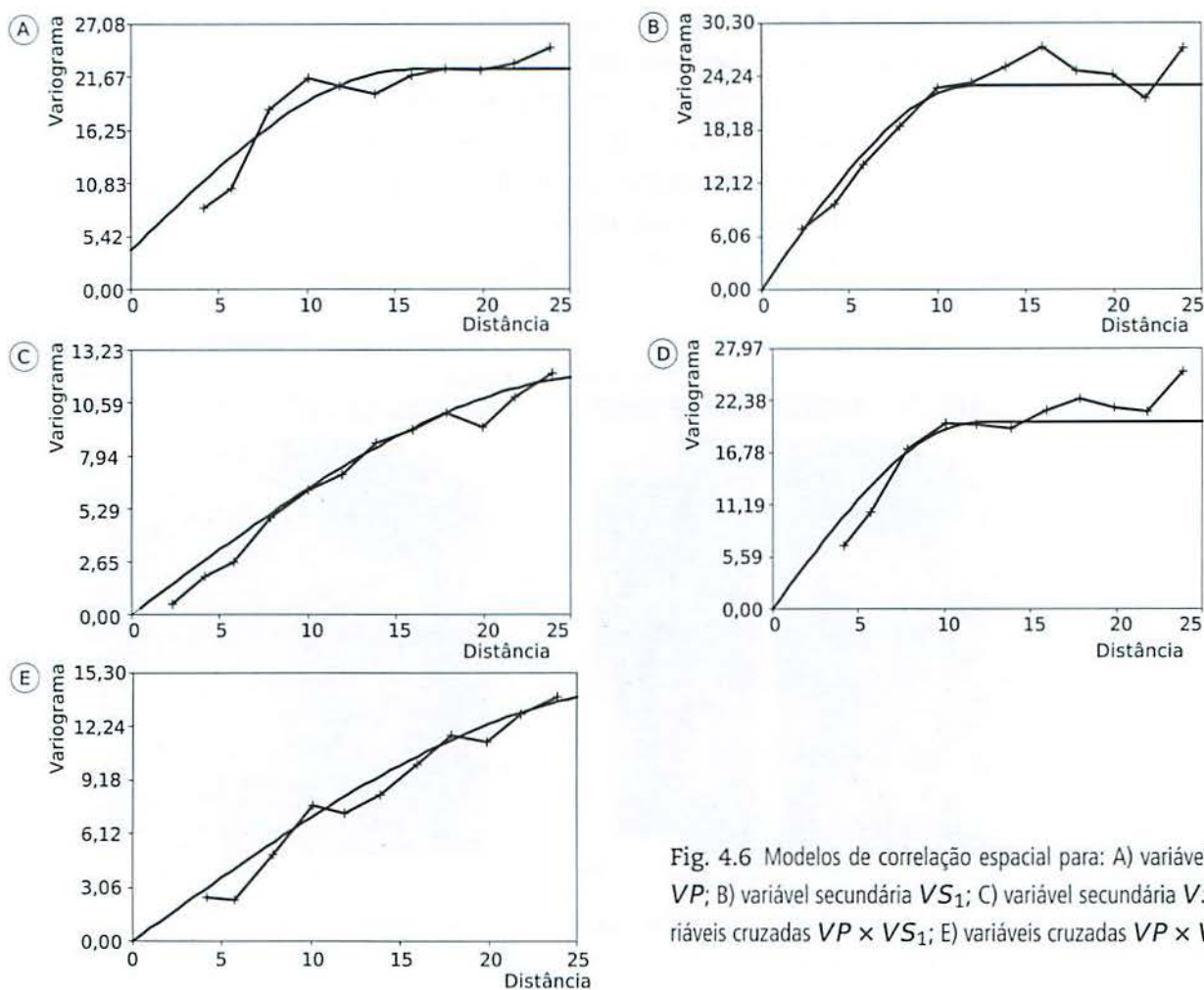


Fig. 4.6 Modelos de correlação espacial para: A) variável primária VP; B) variável secundária VS_1 ; C) variável secundária VS_2 ; D) variáveis cruzadas $VP \times VS_1$; E) variáveis cruzadas $VP \times VS_2$

$$\begin{cases} \gamma_{VP}(h) = 4 + 18,448 \left[1,5 \frac{h}{15,84} - 0,5 \left(\frac{h}{15,84} \right)^3 \right] & \text{para } h < 15,84 \text{ e } \gamma_{VP}(h) = 22,448 \text{ para } h \geq 15,84 \\ \gamma_{VS_1}(h) = 23,2 \left[1,5 \frac{h}{12} - 0,5 \left(\frac{h}{12} \right)^3 \right] & \text{para } h < 12 \text{ e } \gamma_{VS_1}(h) = 23,2 \text{ para } h \geq 12 \\ \gamma_{VPVS_1}(h) = 20 \left[1,5 \frac{h}{12} - 0,5 \left(\frac{h}{12} \right)^3 \right] & \text{para } h < 12 \text{ e } \gamma_{VPVS_1}(h) = 20 \text{ para } h \geq 12 \end{cases} \quad (4.5)$$

$$\begin{cases} \gamma_{VP}(h) = 4 + 18,448 \left[1,5 \frac{h}{15,84} - 0,5 \left(\frac{h}{15,84} \right)^3 \right] & \text{para } h < 15,84 \text{ e } \gamma_{VP}(h) = 22,448 \text{ para } h \geq 15,84 \\ \gamma_{VS_2}(h) = 12 \left[1,5 \frac{h}{27,36} - 0,5 \left(\frac{h}{27,36} \right)^3 \right] & \text{para } h < 27,36 \text{ e } \gamma_{VS_2}(h) = 12 \text{ para } h \geq 27,36 \\ \gamma_{VPVS_2}(h) = 14,4 \left[1,5 \frac{h}{29,52} - 0,5 \left(\frac{h}{29,52} \right)^3 \right] & \text{para } h < 29,52 \text{ e } \gamma_{VPVS_2}(h) = 14,4 \text{ para } h \geq 29,52 \end{cases} \quad (4.6)$$

Com esses dados, procedeu-se à cokrigagem ordinária, que teve por objetivo a estimativa de uma malha regular com abertura de 2,5 em ambos os eixos, totalizando 400 nós, dos quais 354 foram estimados por estarem dentro da fronteira convexa. Os resultados da cokrigagem ordinária estão representados no mapa-imagem da Fig. 4.7.

Em linhas gerais, a reprodução da variável primária é muito boa quando comparada com os dados completos (Fig. 4.2). Esse resultado se deve à ponderação preferencial da informação primária com pesos muito maiores, os quais somam 1, enquanto a ponderação da variável secundária tem uma influência menor e os pesos somam zero, o que, de certo modo, atenua a influência da informação secundária.

Para ilustrar o procedimento da cokrigagem ordinária, foi considerado o nó de coordenadas (26,25; 16,25) da malha regular. Os vizinhos mais próximos ao nó mencionado encontram-se listados na Tab. 4.1. Como se pode verificar nessa tabela, os pesos associados à variável primária somam 1, enquanto os pesos associados à variável secundária somam zero. Os sistemas de equações de cokrigagem ordinária usados para o cálculo dos pesos das variáveis primária e secundária encontram-se nas Eqs. 4.7 e 4.8.

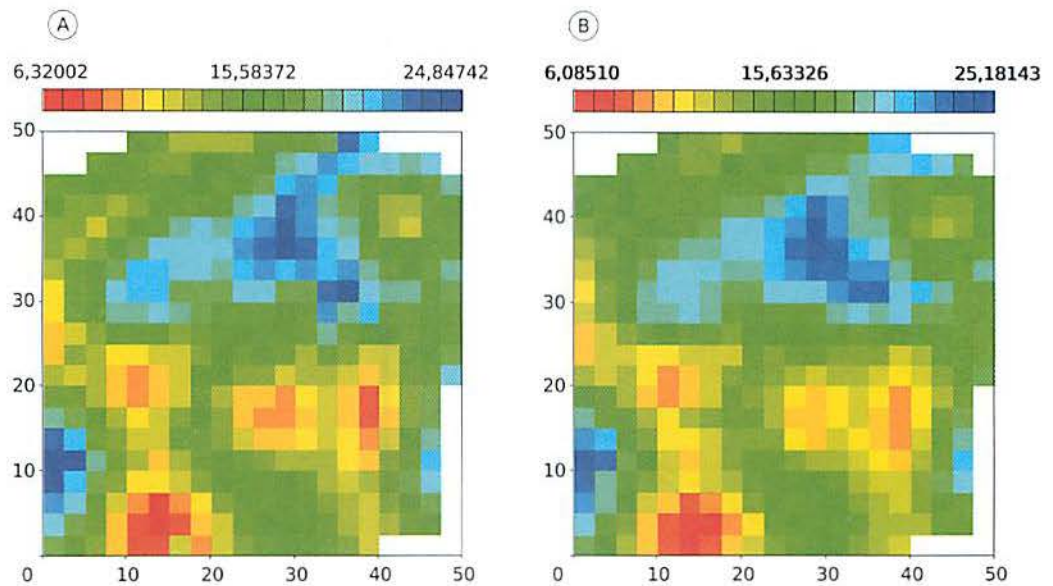


Fig. 4.7 Resultados da estimativa da variável primária por cokrigagem ordinária com base na: A) variável secundária VS_1 e B) variável secundária VS_2

4.1.4 Cokrigagem colocalizada

Quando as variáveis secundárias ou auxiliares são muito mais amostradas que a variável primária ou principal, o sistema de equações de cokrigagem torna-se instável, uma vez que a correlação entre dados secundários próximos é muito maior que a correlação entre dados primários distantes (Xu et al., 1992, p. 835; Goovaerts, 1997, p. 235).

Na verdade, segundo Goovaerts (1997, p. 235), dados secundários muito próximos ou até colocalizados sobre o ponto a ser estimado (isto é, não amostrado) tendem a filtrar a influência dos dados secundários distantes. De acordo com Journel (1999, p. 957), esse é o caso-limite em que a variável secundária é completamente conhecida inclusive sobre os nós da malha regular a ser estimada.

TAB. 4.1 Informações dos pontos vizinhos ao nó de coordenadas (26,25; 16,25), valores das variáveis primária e secundária (VS_1 e VS_2) e resultados da cokrigagem ordinária. Quando $k = 1$, a variável é primária, e quando $k = 2$, a variável é secundária

Tipo	Coordenadas		VS_1		VS_2	
k	X	Y	$Z_k(x)$	Pesos	$Z_k(x)$	Pesos
1	34,50	17,50	11,007	0,145	11,007	0,099
1	23,50	21,50	13,033	0,137	13,033	0,176
1	20,50	14,50	15,221	0,193	15,221	0,176
1	27,50	14,50	7,315	0,526	7,315	0,549
2	27,50	18,50	7,132	0,107	12,474	0,080
2	24,50	17,50	4,710	0,146	12,323	0,138
2	21,50	8,50	15,034	-0,118	13,035	-0,097
2	31,50	12,50	9,573	-0,135	11,754	-0,121
$Z_1^*(26,25; 16,25)$			8,542		10,091	
$\sigma_1(26,25; 16,25)$			2,834		3,193	

$$\begin{bmatrix} 22,448 & 1,722 & 0,247 & 6,169 & 4,368 & 0,787 & 0 & 6,570 & 1 & 0 \\ 1,722 & 22,448 & 6,169 & 5,580 & 8,223 & 10,098 & 0 & 0 & 1 & 0 \\ 0,247 & 6,169 & 22,448 & 7,015 & 2,877 & 8,223 & 6,096 & 0,137 & 1 & 0 \\ 6,169 & 5,580 & 7,015 & 22,448 & 10,370 & 9,835 & 2,322 & 9,337 & 1 & 0 \\ 4,368 & 8,223 & 2,877 & 10,370 & 23,200 & 14,242 & 0,027 & 4,805 & 0 & 1 \\ 0,787 & 10,098 & 8,223 & 9,835 & 14,242 & 23,200 & 1,420 & 2,527 & 0 & 1 \\ 0 & 0 & 6,096 & 2,322 & 0,027 & 1,420 & 23,200 & 0,353 & 0 & 1 \\ 6,570 & 0 & 0,137 & 9,337 & 4,805 & 2,527 & 0,353 & 23,200 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} w_{1_1} \\ w_{2_1} \\ w_{3_1} \\ w_{4_1} \\ w_{1_2} \\ w_{2_2} \\ w_{3_2} \\ w_{4_2} \\ \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} 5,219 \\ 8,577 \\ 8,452 \\ 14,714 \\ 13,664 \\ 14,681 \\ 1,622 \\ 5,425 \\ 1 \\ 0 \end{bmatrix} \quad (4.7)$$

$$\begin{bmatrix} 22,448 & 1,722 & 0,247 & 6,169 & 9,325 & 7,363 & 3,937 & 10,189 & 1 & 0 \\ 1,722 & 22,448 & 6,189 & 5,580 & 10,776 & 11,403 & 5,413 & 6,078 & 1 & 0 \\ 0,247 & 6,169 & 22,448 & 7,015 & 8,647 & 10,776 & 10,012 & 6,610 & 1 & 0 \\ 6,169 & 5,580 & 7,015 & 22,448 & 11,491 & 11,317 & 8,362 & 11,153 & 1 & 0 \\ 9,325 & 10,776 & 8,647 & 11,491 & 12,000 & 9,929 & 4,792 & 7,366 & 0 & 1 \\ 7,363 & 11,403 & 10,776 & 11,317 & 9,929 & 12,000 & 6,009 & 6,527 & 0 & 1 \\ 3,937 & 5,413 & 10,012 & 8,362 & 4,792 & 6,009 & 12,000 & 5,280 & 0 & 1 \\ 10,189 & 6,078 & 6,610 & 11,153 & 7,366 & 6,527 & 5,280 & 12,000 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} w_{1_1} \\ w_{2_1} \\ w_{3_1} \\ w_{4_1} \\ w_{1_2} \\ w_{2_2} \\ w_{3_2} \\ w_{4_2} \\ \mu_1 \\ \mu_2 \end{bmatrix} = \begin{bmatrix} 5,219 \\ 8,577 \\ 8,452 \\ 14,714 \\ 12,521 \\ 12,829 \\ 7,959 \\ 9,754 \\ 1 \\ 0 \end{bmatrix} \quad (4.8)$$

Segundo Xu et al. (1992, p. 835), o estimador da cokrigagem ordinária colocalizada é:

$$Z_1^*(x_o) - m_1 = \sum_{\alpha=1}^{n_1} \lambda_{\alpha}^1 [Z_1(x_{\alpha}) - m_1] + \lambda^2 [Z_2(x_o) - m_2] \quad (4.9)$$

em que $Z_1^*(x_o)$ é a variável primária estimada no ponto x_o , m_1 e m_2 são, respectivamente, as médias das variáveis primária e secundária, $\{Z_1(x_{\alpha}), \alpha = 1, n_1\}$ são os n_1 dados primários e $Z_2(x_o)$ é o valor secundário conhecido no ponto x_o .



Os pesos $\{\lambda_\alpha^1, \alpha = 1, n_1\}$ e λ^2 são obtidos da solução de um sistema de equações de cokrigagem colocalizada (Xu et al., 1992, p. 835; Goovaerts, 1997, p. 237):

$$\begin{cases} \sum_{\beta=1}^{n_1} \lambda_\beta^1 C_{11}(x_\alpha - x_\beta) + \lambda^2 C_{12}(x_\alpha - x_o) + \mu = C_{11}(x_\alpha - x_o) \text{ para } \alpha = 1, n_1 \\ \sum_{\beta=1}^{n_1} \lambda_\beta^1 C_{21}(x_\beta - x_o) + \lambda^2 C_{22}(0) + \mu = C_{21}(0) \\ \sum_{\beta=1}^{n_1} \lambda_\beta^1 + \lambda^2 = 1 \end{cases} \quad (4.10)$$

em que $C_{11}(\cdot)$ e $C_{22}(\cdot)$ são as funções covariância das variáveis primária e secundária, respectivamente, $C_{12}(\cdot)$ e $C_{21}(\cdot)$ são as funções covariância cruzada e μ é o multiplicador de Lagrange.

No sistema da Eq. 4.10, a condição de restrição imposta é que a soma dos n_1 pesos da variável primária com o peso da secundária seja igual a 1 (Goovaerts, 1997, p. 237).

O sistema de equações da cokrigagem ordinária colocalizada requer o conhecimento das funções covariância cruzada $C_{12}(\cdot)$ e $C_{21}(\cdot)$. Nesses termos, a cokrigagem ordinária colocalizada é apenas uma simplificação da cokrigagem ordinária, pois, em vez de n_2 pontos da variável secundária, utiliza-se apenas o ponto colocalizado em x_o . Segundo Journel (1999, p. 955), esse problema pode ser resolvido por meio do modelo de correionalização de Markov, o qual requer a inferência da função covariância da variável primária $C_{11}(\cdot)$ e o coeficiente de correlação colocalizado $\rho_{12}(0)$ entre as variáveis primária e secundária. Esse modelo permite o cálculo da covariância cruzada $C_{12}(h) = C_{21}(h)$ com base na função covariância $C_{11}(h)$, conforme (Journel, 1999, p. 835):

$$C_{12}(h) = \frac{C_{12}(0)}{C_{11}(0)} C_{11}(h), \forall h$$

em que $C_{12}(0)$ é a covariância cruzada entre as variáveis primária e secundária para distância nula e $C_{11}(0)$ é a covariância da variável primária para distância igual a zero.

De forma equivalente (Xu et al., 1992, p. 836):

$$\rho_{12}(h) = \rho_{12}(0) \rho_1(h), \forall h \quad (4.11)$$

em que $\rho_1(\cdot)$ é a função correlograma da variável primária e $\rho_{12}(\cdot)$ é a função correlograma cruzado entre as variáveis primária e secundária.

Aqui surge a maior vantagem do modelo de correionalização de Markov: a função correlograma cruzado $\rho_{12}(\cdot)$ pode ser estimada diretamente da função correlograma da variável primária, que apenas requer o cálculo e modelagem do variograma experimental da variável primária.

O modelo de Markov expresso na Eq. 4.11 implica que (Journel, 1999, p. 957):

$$E\{Z_2(x) \mid Z_1(x) = z_1, Z_1(x') = z'_1\} = E\{Z_2(x) \mid Z_1(x) = z_1\} \quad \forall x, x', z_1, z'_1$$

que significa o condicionamento da variável secundária $Z_2(x)$ pelo datum primário $Z_1(x) = z_1$, que filtra a influência de qualquer dado primário distante de x , tal como $Z_1(x') = z'_1$.

Esse modelo, de acordo com Journel (1999, p. 957), é razoável se o suporte da variável primária $Z_1(x)$ é igual ou maior que o suporte da variável secundária $Z_2(x)$, permitindo filtrar a influência do dado primário $Z_1(x') = z'_1$. Contudo, segundo ele, em muitas aplicações práticas a variável primária $Z_1(x)$ é definida em um suporte bem menor que o da variável secundária $Z_2(x)$.

Como os dados secundários são amostrados intensivamente, a inferência de $\rho_2(h)$ é muito mais fácil que a de $\rho_1(h)$, e, dessa forma, pode-se obter a função correlograma cruzado substituindo-se $\rho_1(h)$ por $\rho_2(h)$ na Eq. 4.11, conforme Journel (1999, p. 957-959):

$$\rho_{12}(h) = \rho_{12}(0)\rho_2(h) \quad (4.12)$$

Journel (1999, p. 958) denomina a Eq. 4.12 de modelo de Markov 2 e a Eq. 4.11, de modelo de Markov 1. O modelo de Markov 2, segundo ele, implica que:

$$E\{Z_1(x) | Z_2(x) = z_2, Z_2(x') = z'_2\} = E\{Z_1(x) | Z_2(x) = z_2\} \quad \forall x, x', z_2, z'_2$$

que mostra que o condicionamento da variável primária $Z_1(x)$ pelo datum secundário $Z_2(x) = z_2$ filtra a influência de qualquer dado secundário distante $Z_2(x') = z'_2$ e, assim, é suficiente reter a informação colocalizada $Z_2(x) = z_2$.

Observar que, conforme o modelo de Markov 2 (Eq. 4.12), a função correlograma da variável secundária é muito mais fácil de ser obtida, haja vista tratar-se de informação abundante, ao contrário da variável primária.

Sob a hipótese de Markov, o estimador da cokrigagem colocalizada pode ser reescrito na forma padronizada (Xu et al., 1992, p. 836):

$$\frac{Z_1^*(x_o) - m_1}{\sigma_1} = \sum_{\alpha=1}^{n_1} \lambda_{\alpha}^1 \frac{[Z_1(x_{\alpha}) - m_1]}{\sigma_1} + \lambda^2 \frac{[Z_2(x_o) - m_2]}{\sigma_2} \quad (4.13)$$

em que σ_1 e σ_2 são os desvios padrão relativos às médias m_1 e m_2 para, respectivamente, as variáveis primária $Z_1(x)$ e secundária $Z_2(x)$.

O sistema de equações da cokrigagem colocalizada torna-se (Xu et al, 1992, p. 836):

$$\begin{cases} \sum_{\beta=1}^{n_1} \lambda_{\beta}^1 \rho_1(x_{\beta} - x_{\alpha}) + \lambda^2 \rho_{12}(0) \rho_1(x_o - x_{\alpha}) = \rho_1(x_o - x_{\alpha}) & \text{para } \alpha = 1, n_1 \\ \sum_{\beta=1}^{n_1} \lambda_{\beta}^1 \rho_{12}(0) \rho_1(x_{\beta} - x_o) + \lambda^2 = \rho_{12}(0) \end{cases} \quad (4.14)$$

Como se pode observar no sistema de equações da cokrigagem colocalizada (Eq. 4.14), o coeficiente de correlação $\rho_{12}(0)$ tem um papel importante no cálculo dos ponderadores $\{\lambda_{\beta}^1, \beta = 1, n_1\}$ e λ^2 .

Na forma padronizada (Eq. 4.13) proposta por Xu et al. (1992, p. 836), e sob a hipótese do modelo de correionalização de Markov, o estimador e o sistema da cokrigagem colocalizada ficam na forma simples, ou seja, os pesos não estão sujeitos a nenhuma condição de restrição.

Exemplo de aplicação da cokrigagem colocalizada

Como exemplo do procedimento da cokrigagem colocalizada, considerar o conjunto de pontos de dados já usados na cokrigagem ordinária, no qual foram mantidos somente os

70 pontos com as variáveis primária e secundária (Fig. 4.8). Nessa demonstração, foram considerados também dois conjuntos de variáveis secundárias, com alta e média correlação. Além dos dados primários e secundários pareados (Arquivos 16 e 17, Anexo B), houve a necessidade de que os dados secundários fossem fornecidos sobre os nós da malha regular a ser coestimada (Fig. 4.9) (Arquivos 18 e 19, Anexo B).

Segundo o modelo de Markov 1 (Eq. 4.11), necessita-se apenas do correlograma da variável primária, o qual pode ser deduzido da função variograma da variável primária, conforme modelo ilustrado na Fig. 4.6A e descrito pela seguinte equação:

$$\begin{cases} \gamma_1(h) = 4 + 18,448 \left[1,5 \frac{h}{15,84} - 0,5 \left(\frac{h}{15,84} \right)^3 \right] & \text{para } h < 15,84 \\ \gamma_1(h) = 22,448 & \text{para } h \geq 15,84 \end{cases}$$

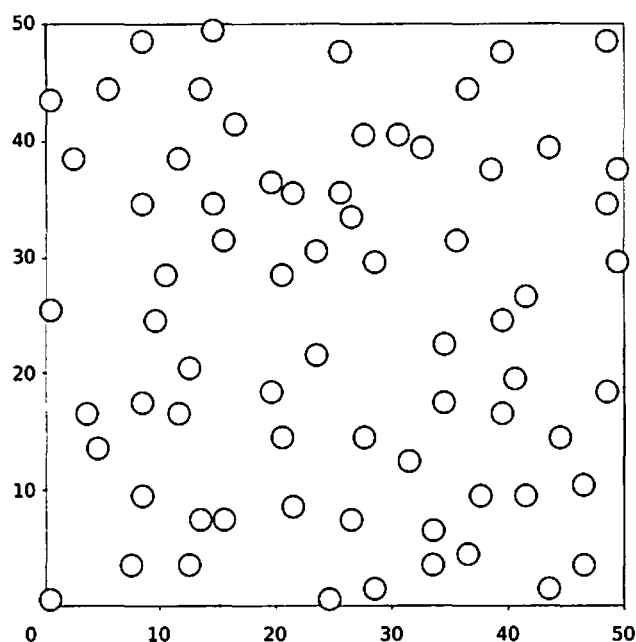


Fig. 4.8 Mapa de localização de pontos com dados de variáveis primária e secundária (Arquivos 16 e 17, Anexo B)

Os modelos de correlograma encontram-se representados na Fig. 4.10.

O estimador a ser usado nesse exemplo é aquele conforme a Eq. 4.13, e os pesos são calculados conforme o sistema de equações de cokrigagem colocalizada (Eq. 4.14).

Os resultados da cokrigagem colocalizada são mostrados na Fig. 4.11, na qual é possível observar a influência da variável secundária nessas coestimativas. No caso de VS_1 , verifica-se um mosaico exibindo a variabilidade dessa variável, enquanto no caso da VS_2 observa-se uma continuidade muito grande na superfície resultante, herdando as características da superfície de grau 5 que foi usada para gerar essa variável secundária.

Da mesma forma, como feito para a cokrigagem ordinária, considerar a estimativa do ponto de coordenadas (26,25; 16,25) pelo procedimento da cokrigagem colocalizada.

As estatísticas amostrais para as variáveis primária e secundária, bem como os coeficientes de correlação necessários para a estimativa na forma padronizada (Eq. 4.13), encontram-se na Tab. 4.2.

TAB. 4.2 Estatísticas das amostras usadas para a cokrigagem colocalizada

Estatísticas	Variáveis		
	VP	VS ₁	VS ₂
Média	15,457	15,461	15,563
Desvio padrão	4,695	4,717	3,025
Correlação	VP x	0,951	0,777

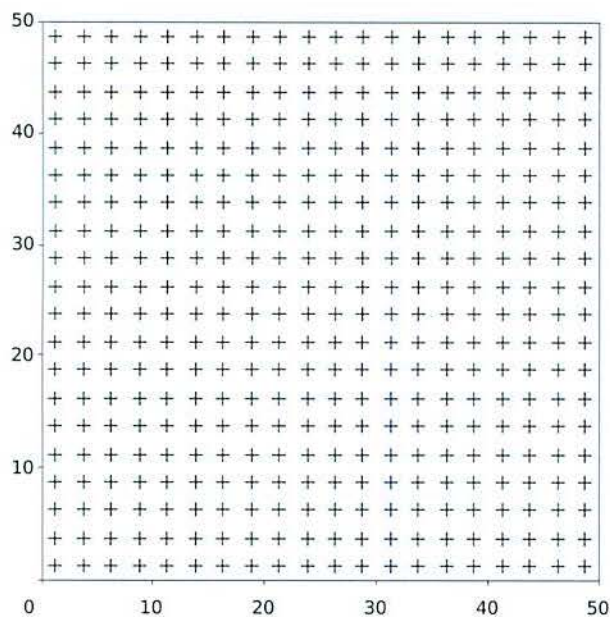


Fig. 4.9 Mapa da malha regular de 20 x 20 nós com informação da variável secundária

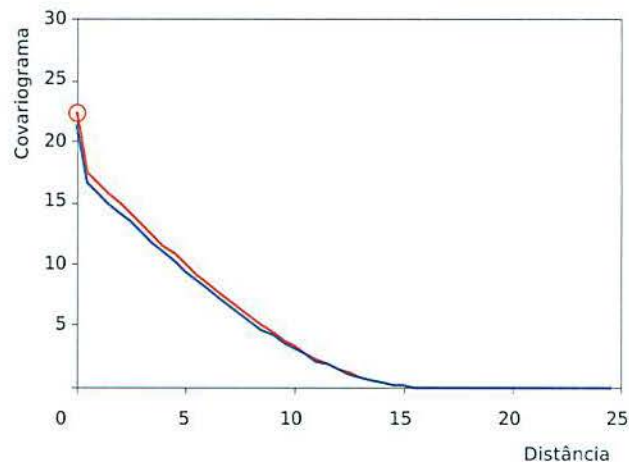


Fig. 4.10 Correlograma cruzado (azul) deduzido do correlograma da variável primária (vermelho), conforme modelo de Markov 1

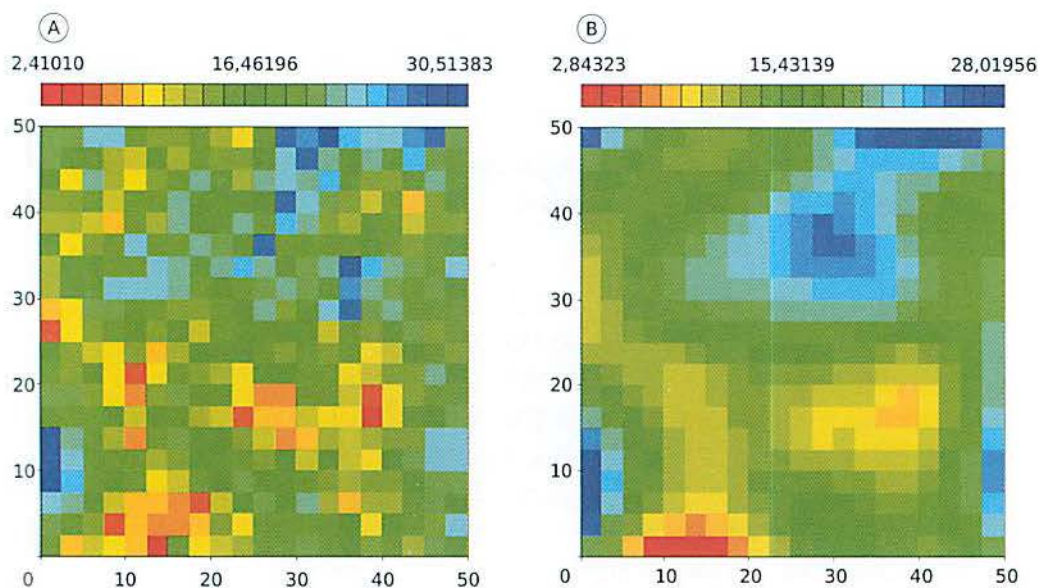


Fig. 4.11 Resultados da cokrigagem colocalizada com base na: A) variável secundária VS_1 e B) variável secundária VS_2 .

Os vizinhos mais próximos a esse ponto, bem como os pesos resultantes da solução do sistema de cokrigagem colocalizada, encontram-se na Tab. 4.3.

Os pesos tanto da variável primária como da variável secundária foram obtidos da resolução dos sistemas de equações de cokrigagem colocalizada das Eqs. 4.15 e 4.16, respectivamente para as situações de alta e média correlação. Ao se comparar esses sistemas, verifica-se que, na matriz dos coeficientes (lado esquerdo), as diferenças estão na última linha e última coluna, onde entra a informação associada à variável secundária, mais

TAB. 4.3 Vizinhos próximos ao ponto estimado (26,25; 16,25) e pesos e estimativas resultantes para as variáveis secundárias com alta (VS_1) e média correlação (VS_2)

X	Y	Z(x)	VS ₁	VS ₂
			Pesos	Pesos
34,50	22,50	13,217	0,002	0,005
34,50	17,50	11,007	0,004	0,013
23,50	21,50	13,033	0,039	0,118
19,50	18,50	16,620	0,002	0,007
20,50	14,50	15,221	0,026	0,078
21,50	8,50	16,281	-0,007	-0,022
27,50	14,50	7,315	0,143	0,429
31,50	12,50	12,124	-0,013	-0,038
$\sum \lambda_a^1$			0,196	0,590
λ^2			0,836	0,494
$Z_2(x_0)$			8,208	11,990
$Z_1^*(x_0)$			8,967	8,174

especificamente o coeficiente de correlação entre as variáveis primária e secundária. A soma de pesos da variável primária e o peso da variável secundária apresentam comportamento verificado por Watanabe (2008, p. 54): quando a correlação é alta, a variável secundária tem grande influência, e, no caso em evidência, para média correlação a variável primária passa a ter maior peso em detrimento da secundária.

$$\begin{bmatrix}
 1 & 0,446 & 0,102 & 0,001 & 0 & 0 & 0,119 & 0,127 & 0,125 \\
 0,446 & 1 & 0,077 & 0,003 & 0,011 & 0 & 0,275 & 0,389 & 0,221 \\
 0,102 & 0,077 & 1 & 0,446 & 0,275 & 0,033 & 0,249 & 0,065 & 0,362 \\
 0,001 & 0,003 & 0,446 & 1 & 0,508 & 0,138 & 0,200 & 0,027 & 0,290 \\
 0 & 0,011 & 0,275 & 0,508 & 1 & 0,372 & 0,313 & 0,096 & 0,357 \\
 0 & 0 & 0,033 & 0,138 & 0,372 & 1 & 0,225 & 0,113 & 0,183 \\
 0,119 & 0,275 & 0,249 & 0,200 & 0,313 & 0,225 & 1 & 0,483 & 0,617 \\
 0,127 & 0,389 & 0,065 & 0,027 & 0,096 & 0,113 & 0,483 & 1 & 0,329 \\
 0,125 & 0,221 & 0,362 & 0,290 & 0,357 & 0,183 & 0,617 & 0,329 & 1
 \end{bmatrix}
 \begin{bmatrix}
 \lambda_1^1 \\
 \lambda_2^1 \\
 \lambda_3^1 \\
 \lambda_4^1 \\
 \lambda_5^1 \\
 \lambda_6^1 \\
 \lambda_7^1 \\
 \lambda_8^1 \\
 \lambda^2
 \end{bmatrix}
 =
 \begin{bmatrix}
 0,131 \\
 0,232 \\
 0,381 \\
 0,304 \\
 0,375 \\
 0,192 \\
 0,649 \\
 0,346 \\
 0,951
 \end{bmatrix}
 \quad (4.15)$$

$$\begin{bmatrix}
 1 & 0,446 & 0,102 & 0,001 & 0 & 0 & 0,119 & 0,127 & 0,102 \\
 0,446 & 1 & 0,077 & 0,003 & 0,011 & 0 & 0,275 & 0,389 & 0,180 \\
 0,102 & 0,077 & 1 & 0,446 & 0,275 & 0,033 & 0,249 & 0,065 & 0,296 \\
 0,001 & 0,003 & 0,446 & 1 & 0,508 & 0,138 & 0,200 & 0,027 & 0,236 \\
 0 & 0,011 & 0,275 & 0,508 & 1 & 0,372 & 0,313 & 0,096 & 0,291 \\
 0 & 0 & 0,033 & 0,138 & 0,372 & 1 & 0,225 & 0,113 & 0,149 \\
 0,119 & 0,275 & 0,249 & 0,200 & 0,313 & 0,225 & 1 & 0,483 & 0,504 \\
 0,127 & 0,389 & 0,065 & 0,027 & 0,096 & 0,113 & 0,483 & 1 & 0,269 \\
 0,102 & 0,180 & 0,296 & 0,236 & 0,291 & 0,149 & 0,504 & 0,269 & 1
 \end{bmatrix}
 \begin{bmatrix}
 \lambda_1^1 \\
 \lambda_2^1 \\
 \lambda_3^1 \\
 \lambda_4^1 \\
 \lambda_5^1 \\
 \lambda_6^1 \\
 \lambda_7^1 \\
 \lambda_8^1 \\
 \lambda^2
 \end{bmatrix}
 =
 \begin{bmatrix}
 0,131 \\
 0,232 \\
 0,381 \\
 0,304 \\
 0,375 \\
 0,192 \\
 0,649 \\
 0,346 \\
 0,777
 \end{bmatrix}
 \quad (4.16)$$

4.2 KRIGAGEM COM DERIVA EXTERNA

Como é objetivo geral de coestimativas geoestatísticas, a krigagem com deriva externa também procura fazer a estimativa de uma variável primária $Z(x)$ com base em uma variável secundária $Y(x)$ correlacionada. Nesse caso, a amostragem da variável primária é insuficiente e, por isso, a variável secundária, mais bem amostrada, é usada para auxiliar na estimativa da primeira. Assim, se $Z(x)$ e $Y(x)$ estão correlacionadas, pode-se descrever essa correlação por meio de uma relação linear (Wackernagel, 1995, p. 190):

$$E[Z(x)] = a_0 + b_1 Y(x) \quad (4.17)$$

o que significa que a variabilidade espacial da variável secundária $Y(x)$ está relacionada a tendências locais da variável primária $Z(x)$ (Xu et al., 1992, p. 835).

O modelo da krigagem com deriva externa consiste, segundo Xu et al. (1992, p. 835), na estimativa dos coeficientes da reta de regressão (Eq. 4.17), os quais são usados para fazer a krigagem dos dados residuais $\{Z(x_\alpha) - [a_0^* + b_1^* Y(x_\alpha)]\}$, $\alpha = 1, n$. Esse procedimento em dois estágios é simplificado, conforme o desenvolvimento a seguir.

O estimador da krigagem com deriva externa pode ser escrito como (Wackernagel, 1995, p. 191):

$$Z_{KDE}^*(x_0) = \sum_{i=1}^n \lambda_i Z(x_i) \quad (4.18)$$

Observar que a variável secundária $Y(x)$ não entra diretamente na equação do estimador (Eq. 4.18), tal como ocorre com os estimadores da cokrigagem (Eqs. 4.2, 4.9 e 4.13). Esse é um aspecto interessante da krigagem com deriva externa, pois a estimativa da variável primária $Z(x)$ é feita exclusivamente com base nos valores observados $\{Z(x_i), i = 1, n\}$.

Da mesma forma, como acontece com a krigagem ordinária, deve-se garantir que, em média, as estimativas sejam iguais aos valores reais:

$$E[Z(x_0)] = E[Z_{KDE}^*(x_0)] \quad (4.19)$$

Substituindo-se a Eq. 4.18 na Eq. 4.19, obtém-se:

$$\sum_{i=1}^n \lambda_i = 1 \quad (4.20)$$

que é uma condição de não viés.

Conforme Wackernagel (1995, p. 191), pode-se aplicar o operador da esperança matemática ao estimador da krigagem com deriva externa (Eq. 4.18):

$$E[Z_{KDE}^*(x_0)] = \sum_{i=1}^n \lambda_i E[Z(x_i)] \quad (4.21)$$

Como $E[Z(x_i)]$ é igual a $E[Z(x)]$, pode-se substituir este último termo pelo lado direito da Eq. 4.17, que fornece:

$$E[Z_{KDE}^*(x_0)] = a_0 + b_1 \sum_{i=1}^n \lambda_i Y(x_i)$$



Na realidade, o termo $\sum_{i=1}^n \lambda_i y(x_i)$ representa o valor da variável secundária que é conhecido no ponto não amostrado (x_o), pois essa técnica requer que os dados secundários sejam conhecidos tanto nos pontos amostrais como sobre os nós da malha regular a ser estimada. Assim, tem-se:

$$y(x_o) = \sum_{i=1}^n \lambda_i y(x_i) \quad (4.22)$$

Essa equação resulta na segunda condição de não viés, a qual impõe aos pesos $\{\lambda_i, i = 1, n\}$ a geometria da distribuição espacial da variável secundária.

A minimização da variância do erro $Var [Z_{KDE}^*(x_o) - Z(x_o)]$ sujeita às duas condições de restrição (Eqs. 4.20 e 4.22) resulta nas equações da krigagem com deriva externa (Wackernagel, 1995, p. 191):

$$\begin{cases} \sum_{j=1}^n \lambda_j C_R(x_i - x_j) - \mu_1 - \mu_2 y(x_i) = C_R(x_i - x_o) \text{ para } i = 1, n \\ \sum_{j=1}^n \lambda_j = 1 \\ \sum_{j=1}^n \lambda_j y(x_j) = y(x_o) \end{cases} \quad (4.23)$$

em que μ_1 e μ_2 são os multiplicadores de Lagrange associados às duas condições de restrição e $C_R(x_i - x_j)$ é a covariância dos resíduos entre os pontos x_i e x_j .

Segundo Xu et al. (1992, p. 835), as vantagens da krigagem com deriva externa são: algoritmo de fácil implementação, pois não precisa das covariâncias $C_2(h)$ e $C_{12}(h)$; o sistema tem dimensão $(n+2)$ em vez de, como na cokrigagem ordinária, $(n_1 + n_2)$; os mapas de $Z(x)$ são muito parecidos com os mapas de $Y(x)$. Ainda de acordo com esses autores, as desvantagens são: os mapas de $Z(x)$ serão bem-correlacionados com $Y(x)$, independentemente da relação linear (Eq. 4.17) ser verdadeira ou não; a krigagem com deriva externa não captura a correlação cruzada entre $Z(x)$ e $Y(x)$, ao contrário da cokrigagem; requer os dados secundários em todos os pontos amostrais dos dados primários e em todos os nós da malha regular a ser interpolada; requer a covariância dos resíduos $Z(x) - [a_0 + b_1 Y(x)]$, que não pode ser calculada sobre os resíduos experimentais $Z(x) - [a_0^* + b_1^* Y(x)]$.

Segundo Goovaerts (1997, p. 142), o cálculo do variograma dos resíduos $\gamma_R(h)$ não é direto, pois os dados disponíveis são os valores da variável $Z(x)$ e não dos resíduos $R(x)$.

Observar que a variável regionalizada $Z(x)$ pode ser decomposta como:

$$Z(x) = m(x) + R(x) \quad (4.24)$$

em que $m(x)$ é a componente de tendência descrita por um polinômio de baixo grau e $R(x)$ é o resíduo.

A relação entre $\gamma(h)$ e $\gamma_R(h)$ pode ser deduzida da seguinte forma, de acordo com Goovaerts (1997, p. 142):

$$2\gamma(h) = E \{ [Z(x) - Z(x+h)]^2 \} \quad (4.25)$$

Substituindo-se a relação da Eq. 4.24 na Eq. 4.25 obtém-se:

$$2\gamma(h) = E \{ [R(x) + m(x) - R(x+h) - m(x+h)]^2 \} \quad (4.26)$$

Da equação anterior obtém-se (Goovaerts, 1997, p. 142):

$$2\gamma(h) = 2\gamma_R(h) + [m(x) - m(x+h)]^2 \quad (4.27)$$

Na verdade, a Eq. 4.27 deveria ser escrita como:

$$2\gamma(h) = 2\gamma_R(h) + E\{[m(x) - m(x+h)]^2\} \quad (4.28)$$

ou seja, faltou o operador da esperança matemática. Além disso, a Eq. 4.28 implica que os demais termos resultantes da Eq. 4.26 são nulos:

$$2E[R(x)(m(x) - m(x+h))] = -2E[R(x+h)(m(x) - m(x+h))] \quad (4.29)$$

Na verdade, a decomposição da variável regionalizada, em termos da componente de tendência e seu resíduo (Eq. 4.24), garante que a esperança do resíduo seja nula, anulando os termos da Eq. 4.29.

A função variograma residual fica:

$$2\gamma_R(h) = 2\gamma(h) - E\{[m(x) - m(x+h)]^2\} \quad (4.30)$$

ou seja, subtrai-se do variograma de $Z(x)$ a média da diferença ao quadrado das componentes de tendência nos pontos x e $x+h$, as quais são desconhecidas (Goovaerts, 1997, p. 142).

Uma solução para a inferência do $\gamma_R(h)$ pode ser obtida conhecendo-se pares que não sejam afetados pela tendência, tais como (Goovaerts, 1997, p. 142):

$$\begin{cases} m(x) \approx m(x+h) \text{ então} \\ R(x) - R(x+h) \approx Z(x) - Z(x+h) \end{cases} \quad (4.31)$$

Dessa forma, segundo Goovaerts (1997, p. 142), o variograma residual poderia ser inferido diretamente dos pares que satisfizessem a relação da Eq. 4.31. De acordo com ele, essa relação é geralmente satisfeita para pequenas distâncias h , significando que o variograma residual equivale ao variograma dos dados originais $Z(x)$ para os primeiros passos.

Ainda segundo esse autor, para distâncias maiores o variograma residual poderia ser inferido com base nos pares tomados em subáreas ou ao longo de direções em que a influência da tendência pudesse ser desprezada (por exemplo, perpendicular à direção de tendência).

Exemplo de aplicação da krigagem com deriva externa

Para fins de ilustração da krigagem com deriva externa, os mesmos conjuntos usados para a cokrigagem colocizada serão considerados: o conjunto de pontos com dados amostrais primários e secundários (Fig. 4.8) e dados secundários nos pontos da malha regular (Fig. 4.9).

De acordo com a teoria da krigagem com deriva externa, o sistema de equações de krigagem com deriva externa (Eq. 4.23) necessita, para ser solucionado, da covariância residual ou da função variograma residual. Entretanto, como descrito na seção anterior, a obtenção do variograma residual não é um procedimento simples e direto.

Assim, com o objetivo de encaminhar uma solução a esse problema, desenvolveu-se, especificamente para este livro, uma metodologia mais rápida e direta para o cálculo do variograma residual.

O termo do lado direito da Eq. 4.30, $E\{[m(x) - m(x+h)]^2\}$, equivale ao variograma da componente de tendência, conforme segue:

$$2\gamma_M(h) = E\{[m(x) - m(x+h)]^2\} \quad (4.32)$$

Substituindo-se a Eq. 4.32 na Eq. 4.30, obtém-se:

$$2\gamma_R(h) = 2\gamma(h) - 2\gamma_M(h) \quad (4.33)$$

Portanto, o variograma residual pode ser calculado como a diferença entre o variograma da variável primária e o variograma da componente de tendência. Dessa forma, o problema resume-se ao cálculo do variograma da componente de tendência.

No caso da krigagem com deriva externa, a variável primária apresenta uma relação linear com a variável secundária:

$$Z(x) = a_0 + b_1 Y(x)$$

Os coeficientes da reta de regressão a_0 e b_1 podem ser calculados facilmente com os pontos amostrais. Por outro lado, como requer a krigagem com deriva externa, conhecem-se os valores da variável secundária em todos os nós da malha regular a ser estimada. Os coeficientes da regressão podem ser aplicados aos valores da variável secundária, resultando na componente de tendência, a qual passa a ser conhecida em todos os nós da malha regular.

Com esses dados, pode-se calcular o variograma da componente de tendência, que, subtraído do variograma da variável primária, resulta no variograma residual. Assim, o

variograma residual pode ser calculado pela diferença entre os dois variogramas, como descrito pela Eq. 4.33.

A única e principal restrição é que os resíduos calculados sejam positivos. Isso significa que o variograma da variável primária tem de apresentar necessariamente um patamar maior que o variograma da componente de tendência.

Nesse sentido, e com base no variograma da variável secundária VS_1 , o variograma da variável primária (Fig. 4.6A) foi reajustado para garantir valores positivos aos resíduos, conforme o modelo descrito a seguir e ilustrado na Fig. 4.12. Observar que esse reajuste é bastante razoável, haja vista a variável secundária ser mais bem amostrada que a variável primária.

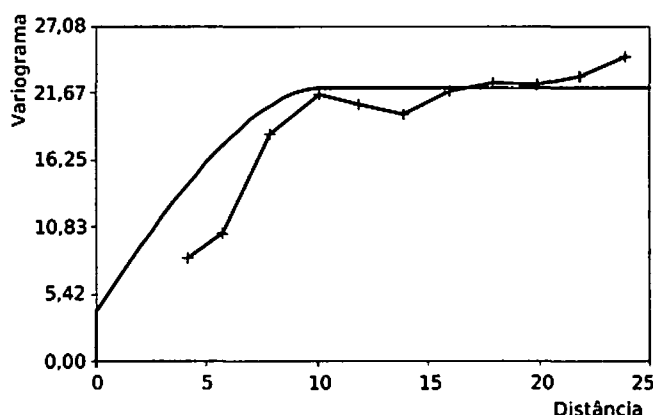


Fig. 4.12 Variograma da variável primária reajustado para garantir resíduos positivos, conforme a Eq. 4.33

$$\begin{aligned} \gamma_P(h) &= 4 + 18,079 \left[1,5 \frac{h}{10,32} - 0,5 \left(\frac{h}{10,32} \right)^3 \right] \text{ para } h < 10,32 \\ \gamma_P(h) &= 22,079 \text{ para } h \geq 10,32 \end{aligned}$$

Os variogramas das componentes de tendência, bem como os variogramas residuais resultantes, encontram-se na Fig. 4.13.

O variograma residual para a variável secundária VS_2 (Fig. 4.13D) ilustra o limite do campo estruturado para uma amplitude igual a 8,813. Isso se deve à distribuição dessa variável secundária, a qual foi obtida por uma superfície polinomial de grau 5.

Os modelos dos variogramas residuais encontram-se descritos pelas Eqs. 4.34 e 4.35, respectivamente para as variáveis VS_1 e VS_2 .

$$\begin{cases} \gamma_{VS_1}(h) = 0,477 \left[1,5 \frac{h}{10,349} - 0,5 \left(\frac{h}{10,349} \right)^3 \right] & \text{para } h < 10,349 \\ \gamma_{VS_1}(h) = 0,477 & \text{para } h \geq 10,349 \end{cases} \quad (4.34)$$

$$\begin{cases} \gamma_{VS_2}(h) = 4 + 10,8 \left[1,5 \frac{h}{8,813} - 0,5 \left(\frac{h}{8,813} \right)^3 \right] & \text{para } h < 8,813 \\ \gamma_{VS_2}(h) = 14,8 & \text{para } h \geq 14,8 \end{cases} \quad (4.35)$$

Essa solução pode não ser exatamente correta do ponto de vista teórico, mas certamente é melhor que uma aproximação subjetiva como a proposta por Goovaerts (1997, p. 142). É importante ressaltar que a componente de tendência estimada dessa forma usa a informação disponível, ou seja, os dados amostrais, dos quais se obtêm os coeficientes da regressão, bem como dos dados da variável secundária conhecidos em todos os nós da malha regular a ser interpolada.

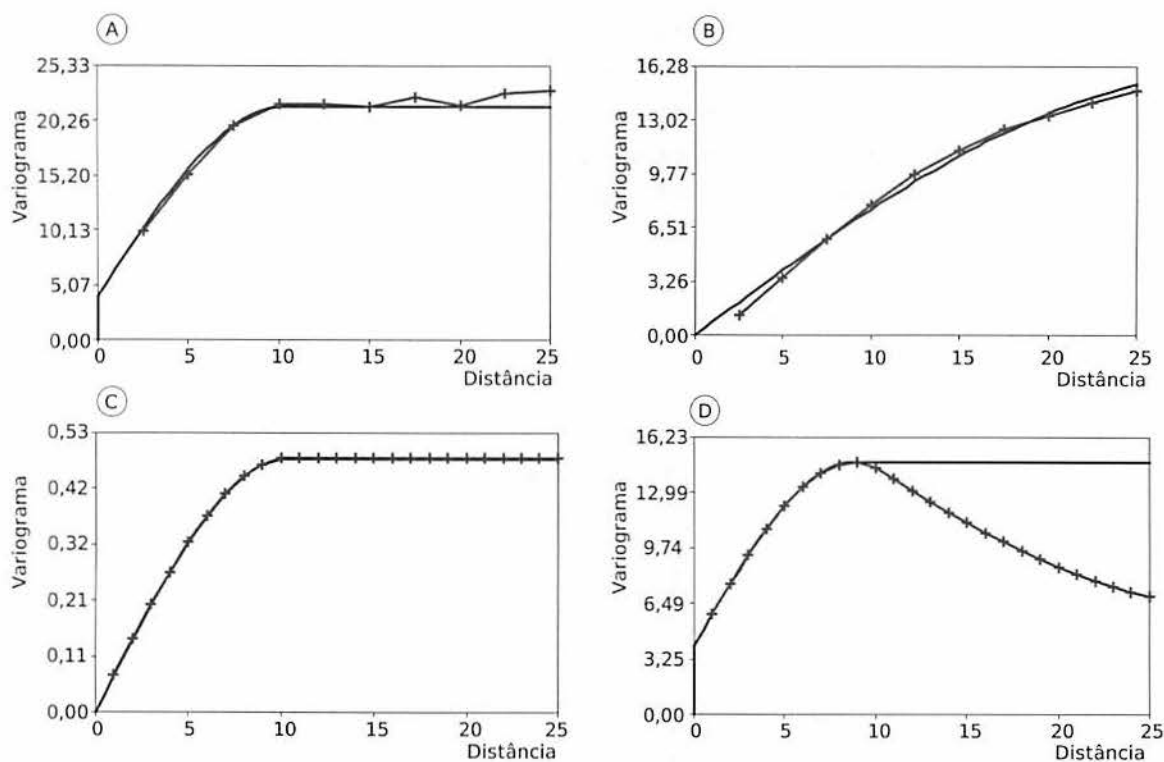


Fig. 4.13 Em A) e B), variogramas das componentes de tendência para as variáveis secundárias VS_1 e VS_2 , respectivamente; em C) e D), variogramas residuais para as variáveis secundárias VS_1 e VS_2 , respectivamente

Assim, procedeu-se ao cálculo da krigagem com deriva externa com base nos variogramas residuais obtidos (Eqs. 4.34 e 4.35). Os resultados encontram-se na Fig. 4.14.

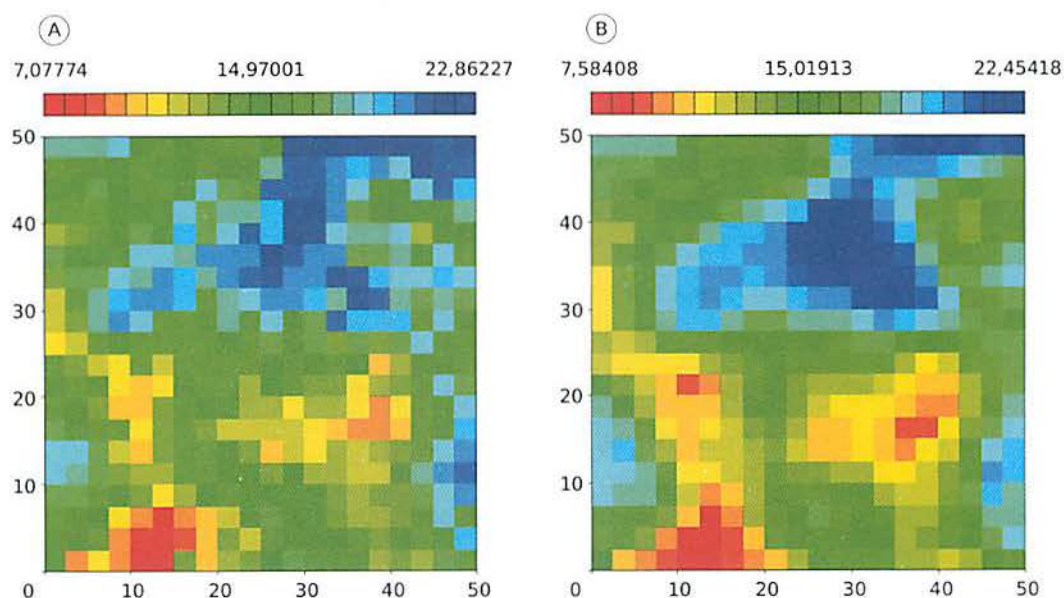


Fig. 4.14 Resultados da krigagem com deriva externa com base na: A) variável secundária VS_1 e B) variável secundária VS_2

Como se pode verificar, a krigagem com deriva externa não sofre forte influência da variável secundária, tal qual a cokrigagem colocada (Fig. 4.11A). Entretanto, as superfícies estimadas por krigagem com deriva externa herdam as características da variável secundária. Por exemplo, na Fig. 4.14B fica evidente a influência da geometria de um polinômio bivariado de grau 5 que deu origem à variável secundária VS_2 .

Para mostrar como se faz a estimativa de um ponto não amostrado, considerar o ponto de coordenadas (26,25; 16,25), conforme os pontos da Tab. 4.4.

TAB. 4.4 Dados vizinhos ao ponto de coordenadas (26,25; 16,25) e pesos resultantes da krigagem com deriva externa

X	Y	Z(x)	VS_1	VS_2
			Pesos	Pesos
34,50	22,50	13,217	0,099	0,000
34,50	17,50	11,007	0,083	0,122
23,50	21,50	13,033	0,184	0,076
19,50	18,50	16,620	0,000	0,052
20,50	14,50	15,221	0,012	0,030
21,50	8,50	16,281	0,157	0,151
27,50	14,50	7,315	0,314	0,524
31,50	12,50	12,124	0,152	0,044
$Z_{KDE}^*(x_0)$			11,331	10,361

Os sistemas de equações de krigagem com deriva externa utilizados para as estimativas no ponto de coordenadas (26,25; 16,25) encontram-se nas Eqs. 4.36 e 4.37.

$$\begin{bmatrix}
 0,487 & 0,167 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 12,806 \\
 0,167 & 0,487 & 0 & 0 & 0 & 0 & 0,052 & 0,125 & 1 & 12,365 \\
 0 & 0 & 0,487 & 0,167 & 0 & 0,052 & 0,038 & 0 & 1 & 12,152 \\
 0 & 0 & 0,167 & 0,487 & 0,001 & 0,216 & 0,016 & 0 & 1 & 15,034 \\
 0 & 0 & 0 & 0,001 & 0,487 & 0,113 & 0,026 & 0 & 1 & 15,034 \\
 0 & 0 & 0,052 & 0,216 & 0,113 & 0,487 & 0,074 & 0 & 1 & 13,351 \\
 0 & 0,052 & 0,038 & 0,016 & 0,026 & 0,074 & 0,487 & 0,196 & 1 & 10,196 \\
 0 & 0,125 & 0 & 0 & 0 & 0 & 0,196 & 0,487 & 1 & 9,573 \\
 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\
 12,806 & 12,365 & 12,152 & 15,034 & 15,034 & 13,351 & 10,196 & 9,573 & 0 & 0
 \end{bmatrix}
 \begin{bmatrix}
 \lambda_1 \\
 \lambda_2 \\
 \lambda_3 \\
 \lambda_4 \\
 \lambda_5 \\
 \lambda_6 \\
 \lambda_7 \\
 \lambda_8 \\
 \mu_1 \\
 \mu_2
 \end{bmatrix}
 =
 \begin{bmatrix}
 0,002 \\
 0,031 \\
 0,120 \\
 0,070 \\
 0,015 \\
 0,116 \\
 0,335 \\
 0,097 \\
 1 \\
 8,208
 \end{bmatrix}
 \quad (4.36)$$

$$\begin{bmatrix}
 14,800 & 2,472 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 13,510 \\
 2,472 & 14,800 & 0 & 0 & 0 & 0 & 0,219 & 1,527 & 1 & 11,222 \\
 0 & 0 & 14,800 & 2,472 & 0 & 0,219 & 0,071 & 0 & 1 & 14,062 \\
 0 & 0 & 2,472 & 14,800 & 0 & 3,656 & 0 & 0 & 1 & 12,540 \\
 0 & 0 & 0 & 0 & 14,800 & 1,279 & 0,005 & 0 & 1 & 13,035 \\
 0 & 0 & 0,219 & 3,656 & 1,279 & 14,800 & 0,547 & 0 & 1 & 11,294 \\
 0 & 0,219 & 0,071 & 0 & 0,005 & 0,547 & 14,800 & 3,164 & 1 & 11,819 \\
 0 & 1,527 & 0 & 0 & 0 & 0 & 3,164 & 14,800 & 1 & 11,754 \\
 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 0 & 0 \\
 13,510 & 11,222 & 14,062 & 12,540 & 13,035 & 11,924 & 11,819 & 11,754 & 0 & 0
 \end{bmatrix}
 \begin{bmatrix}
 \lambda_1 \\
 \lambda_2 \\
 \lambda_3 \\
 \lambda_4 \\
 \lambda_5 \\
 \lambda_6 \\
 \lambda_7 \\
 \lambda_8 \\
 \mu_1 \\
 \mu_2
 \end{bmatrix}
 =
 \begin{bmatrix}
 0 \\
 0,084 \\
 1,456 \\
 0,528 \\
 0,009 \\
 1,376 \\
 6,718 \\
 0,990 \\
 1 \\
 11,990
 \end{bmatrix}
 \quad (4.37)$$

4.3 CONSIDERAÇÕES FINAIS

Foram apresentados três métodos de coestimativa: *cokrigagem ordinária*, *cokrigagem colocalizada* e *krigagem com deriva externa*.

A *cokrigagem ordinária* é um procedimento que requer o cálculo e modelagem de variogramas experimentais diretos e cruzados. A modelagem desses variogramas não deve ser feita individualmente, mas sim em conjunto, de forma que os variogramas satisfaçam o modelo linear de correionalização.

O procedimento da *cokrigagem ordinária* pode se tornar muito trabalhoso à medida que aumenta o número de variáveis secundárias. Além disso, dependendo do número de variáveis envolvido, há o problema de estimativas discrepantes da distribuição inicial da variável primária. Isso acontece por causa das condições de restrição do sistema de equações de *cokrigagem ordinária*, em que os pesos da variável primária somam 1, enquanto os pesos da variável secundária somam zero.

No exemplo apresentado neste capítulo, os resultados da *cokrigagem ordinária* são muito bons, tanto na situação com alta correlação como na com média correlação.

A *cokrigagem colocalizada*, por sua vez, é uma técnica que simplifica o procedimento da coestimativa, pois não requer o cálculo do variograma cruzado, o qual é deduzido com base no modelo de Markov 1 ou 2, dependendo do suporte amostral da variável primária em relação à variável secundária.

Trata-se de uma técnica que faz a estimativa usando a informação secundária em caso de alta correlação ou a informação primária em caso de baixa correlação. Em situações de correlações médias, a cokrigagem colocalizada faz uso tanto da informação primária como da secundária, por meio dos pesos das variáveis primária e secundária.

A krigagem com deriva externa é também uma alternativa interessante em situações em que a variável secundária é fartamente amostrada. Nesse procedimento, há uma condição de restrição que força os pesos da variável primária a seguirem a geometria da variável secundária. Os exemplos utilizados para ilustrar esse procedimento mostram como a geometria da variável secundária influencia a estimativa da variável primária.

A maior dificuldade da krigagem com deriva externa está no cálculo da covariância residual com base no variograma residual. Foi desenvolvido, então, um procedimento que permite calcular a componente de tendência sobre a malha regular contendo a informação secundária, e daí, derivar o seu variograma, que, subtraído do variograma da variável primária, resulta no variograma residual.

A Fig. 4.15 apresenta uma síntese dos métodos de coestimativas geoestatísticas.

A cokrigagem ordinária trabalha com uma base de dados com, preferencialmente, heterotopia parcial (Fig. 4.15A), e por meio dela são calculados os variogramas diretos e o variograma cruzado (Fig. 4.15E).

A cokrigagem colocalizada e a krigagem com deriva externa usam duas bases de dados: a primeira, com amostragem das variáveis primária e secundária nos mesmos pontos, ou seja, a base é isotópica (Fig. 4.15B), e a segunda, com dados secundários sobre os pontos em que se deseja estimar a variável primária (Fig. 4.15C).

No caso da cokrigagem colocalizada, parte-se do covariograma da variável primária para estimar o covariograma cruzado, conforme o modelo de Markov 1 (Fig. 4.15F).

A krigagem com deriva externa precisa do variograma residual (Fig. 4.15G), que foi determinado de acordo com a metodologia descrita na seção 4.2. Os resultados obtidos encontram-se nas Figs. 4.15H, 4.15I e 4.15J, para, respectivamente, os métodos de cokrigagem ordinária, cokrigagem colocalizada e krigagem com deriva externa.

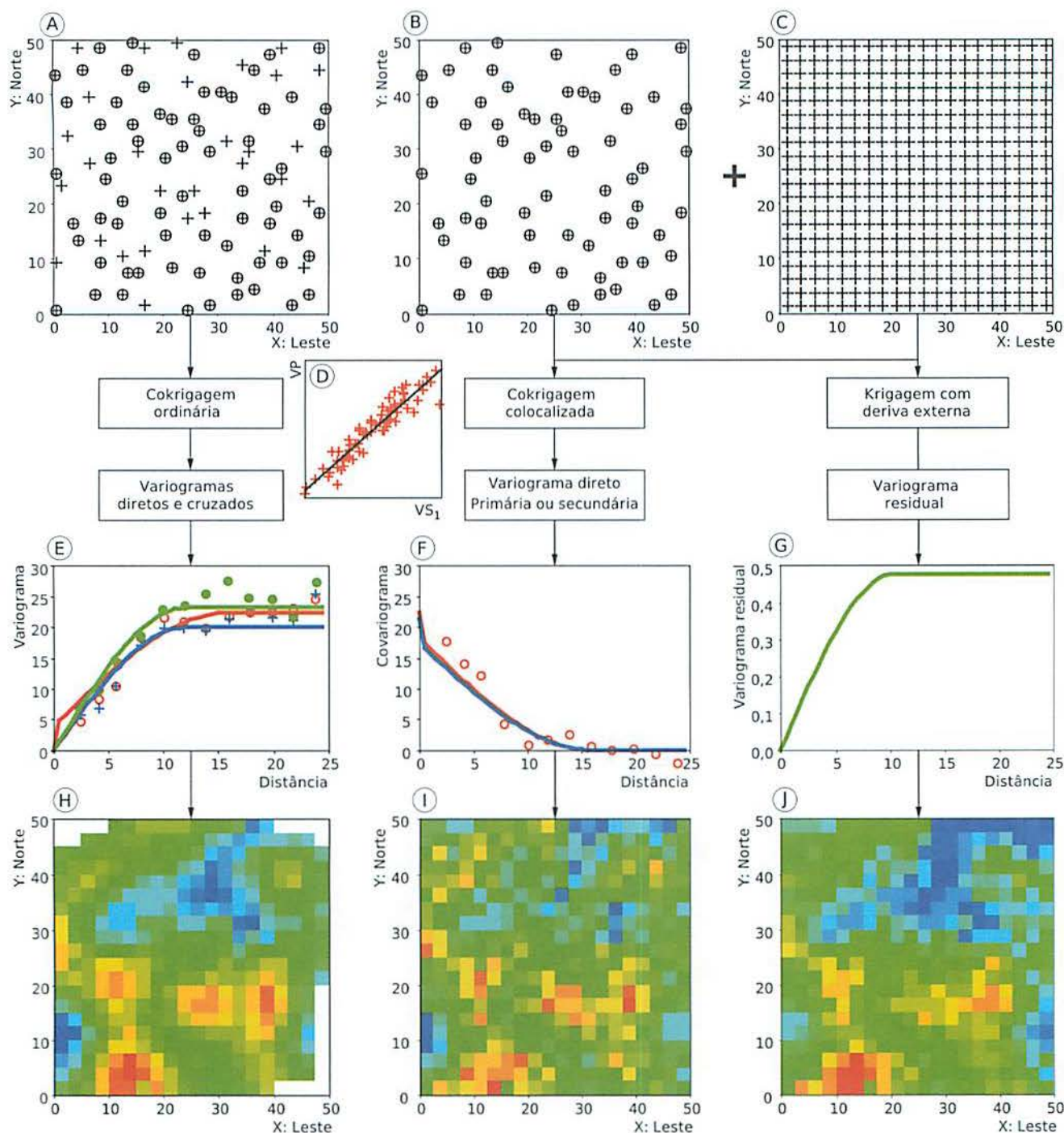


Fig. 4.15 Síntese dos métodos de coestimativas geoestatísticas: A) mapa de localização de pontos com heterotopia parcial; B) mapa de localização de pontos com isotopia; C) mapa de localização de pontos da variável secundária sobre os nós de uma malha regular; D) correlação entre as variáveis primária e secundária; E) modelos de variogramas diretos (vermelho = variável primária; verde = variável secundária) e variograma cruzado (azul); F) covariograma da variável primária (vermelho) e covariograma cruzado calculado pelo modelo de Markov 1 (azul); G) variograma residual; H) resultado da cokrigagem ordinária; I) resultado da cokrigagem colocalizada; J) resultado da krigagem com deriva externa

Simulação Estocástica

5

A krigagem proporciona a estimativa $Z^*(x_o)$ em um ponto não amostrado x_o com base na informação dos n pontos vizinhos. Essa estimativa é feita minimizando-se a variância do erro de estimativa, como visto no Cap. 3. Na realidade, porém, a minimização da variância do erro envolve a suavização das dispersões reais (Journel; Huijbregts, 1978, p. 493). Essa suavização ocorre mesmo que as estimativas sejam condicionais aos pontos amostrais, ou seja:

$$Z_{KO}^*(x_o) = Z(x_o)$$

Entretanto, esse condicionamento não garante que as estimativas resultantes (por exemplo, calculadas sobre os nós de uma malha regular) não estejam suavizadas.

Na verdade, todos os algoritmos de interpolação tendem a suavizar a variabilidade espacial do atributo (Goovaerts, 1997, p. 370). A suavização se caracteriza pela subestimativa de valores altos e superestimativa de valores baixos (Goovaerts, 1997, p. 370). Além disso, segundo esse autor, a suavização não é uniforme, pois é zero nos pontos amostrais e vai aumentando à medida que se distancia dos pontos de dados.

A suavização pode ser facilmente verificada comparando-se o histograma amostral com o histograma das estimativas por krigagem ordinária. Por exemplo: para a amostra com os dados da distribuição lognormal (Arquivo 12, Anexo B) (Fig. 5.1), verifica-se uma discreta

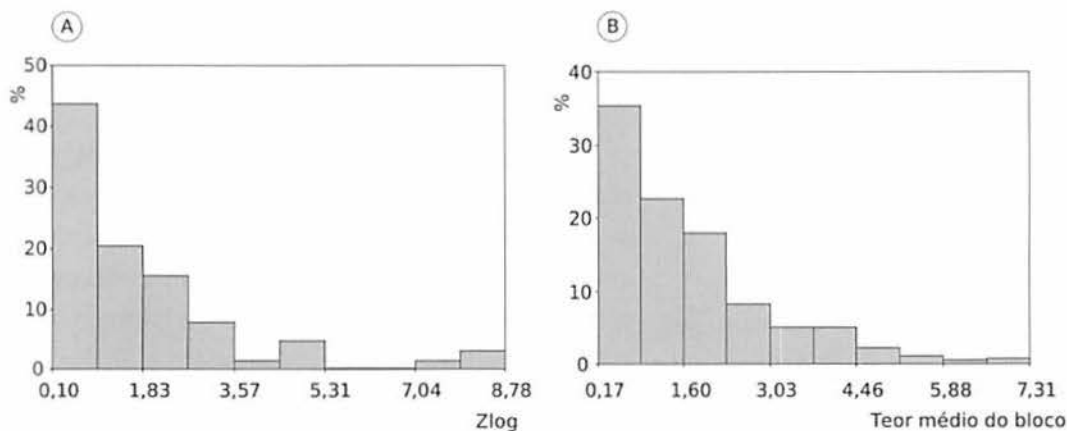


Fig. 5.1 A) Histograma amostral da distribuição lognormal e B) histograma das estimativas por krigagem ordinária sobre os nós de uma malha regular

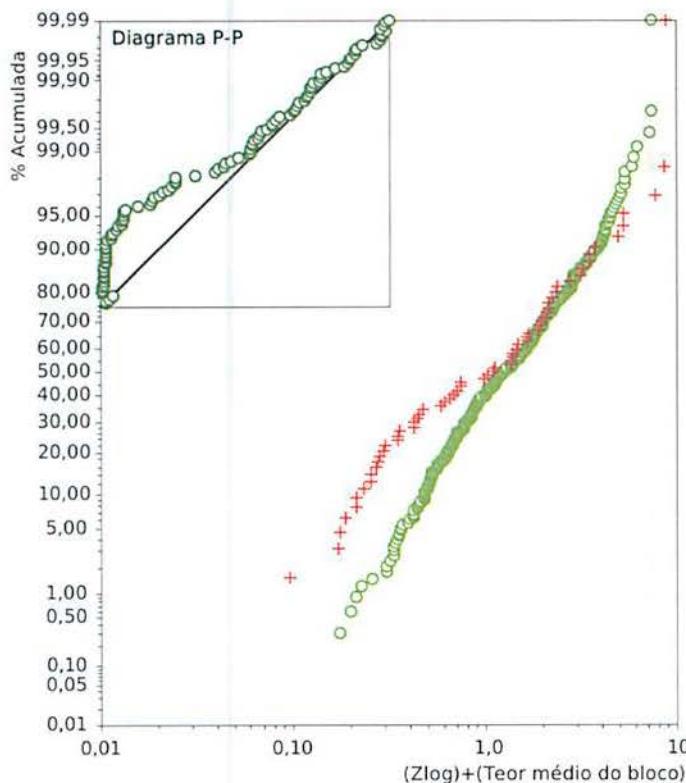


Fig. 5.2 Comparação das curvas acumulativas da distribuição amostral (cruz vermelha) com a distribuição das estimativas por krigagem ordinária (círculo verde)

inferência do fenômeno espacial em estudo não pode ser realizado com exatidão, porque não permite concluir corretamente sobre a distribuição e variabilidade espaciais da variável de interesse.

De acordo com Olea (1999, p. 141), a simulação estocástica foi a solução adotada pela Geoestatística para resolver o problema da suavização da krigagem. Entretanto, segundo

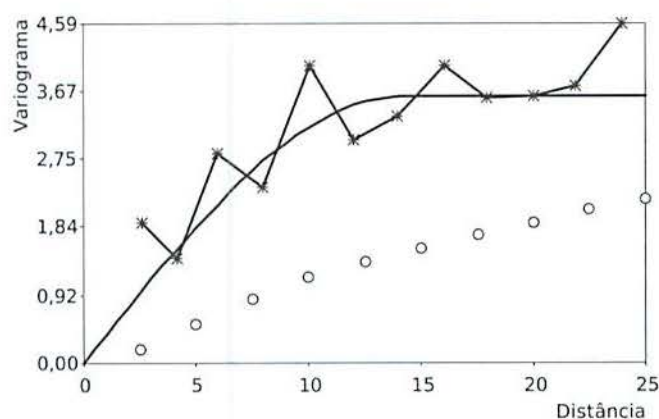


Fig. 5.3 Variograma experimental (asteriscos) e variograma das estimativas (círculos). O modelo de correlação espacial está representado por linha contínua (Arquivo 12, Anexo B)

suavização, a qual mostra que os valores baixos e altos não foram reproduzidos, ou seja, a perda das caudas inferior e superior da distribuição. Além disso, pode-se representar as curvas acumulativas em escala de logprobabilidade aritmética, as quais devem mostrar quão diferentes são essas distribuições (Fig. 5.2). Nessa figura, verifica-se que a distribuição das estimativas está suavizada, ou seja, com menor variância, por causa da inclinação da curva. O diagrama P-P (probabilidade-probabilidade) no canto superior esquerdo da figura confirma que essas duas distribuições são diferentes. O efeito de suavização das estimativas, além disso, pode ser verificado comparando-se o variograma experimental com o variograma das estimativas (Fig. 5.3).

Como pode ser observado nessa figura, o variograma das estimativas apresenta uma continuidade muito maior que o variograma amostral e um patamar bem menor, refletindo a perda da variância.

A consequência é que o efeito de suavização da krigagem não reproduz adequadamente as características da amostra usada para fazer as estimativas em pontos não amostrados. Assim, o processo de

inferência do fenômeno espacial em estudo não pode ser realizado com exatidão, porque não permite concluir corretamente sobre a distribuição e variabilidade espaciais da variável de interesse.

De acordo com Olea (1999, p. 141), a simulação estocástica foi a solução adotada pela Geoestatística para resolver o problema da suavização da krigagem. Entretanto, segundo ele, a simulação não é a solução perfeita: ganha-se em precisão global em detrimento da precisão local. Na realidade, as realizações não estão isentas de erros na reprodução da realidade e, em média, os erros da simulação estocástica são maiores que os da krigagem (Olea, 1999, p. 141).

A simulação estocástica também foi a solução adotada pelos geoestatísticos para modelar a incerteza associada à estimativa (Deutsch, 2011, p. 5-1), uma vez que a variância de krigagem foi reconhecida apenas como um índice de configuração espacial dos pontos vizinhos próximos (Journal; Rossi, 1989, p. 783).

Na verdade, o conjunto de realizações $\{Z^l(x), l = 1, L\}$ proporciona uma medida visual e quantitativa da incerteza espacial (Goovaerts, 1997, p. 372).

Os métodos de simulação disponíveis devem ser aplicados e os resultados, analisados com muita atenção, pois algumas realizações podem mostrar cenários distintos da realidade.

Evidentemente, algumas aplicações específicas requerem o uso de técnicas de simulação, principalmente quando a informação disponível ainda é insuficiente. Na seção seguinte será mostrada a causa do efeito de suavização da krigagem ordinária.

5.1 ERRO DE SUAUIZAÇÃO

O erro de suavização da krigagem, para Deutsch e Journel (1992, p. 125), decorre da falta de um componente de erro, conforme segue:

$$Z(x_o) = Z_{KO}^*(x_o) + R(x_o)$$

em que $R(x_o)$ é uma variável aleatória correspondente ao erro de estimativa.

Para recompor a variabilidade total da função aleatória $Z(x_o)$, pode-se simular a função aleatória $R(x_o)$, com média zero e covariância corretas, que poderia ser adicionada à estimativa (Deutsch; Journel, 1992, p. 125):

$$Z_{Sim}^l(x_o) = Z^*(x_o) + R^l(x_o) \quad (5.1)$$

em que o sobrescrito l indica a l -ésima realização.

Segundo esses autores, a Eq. 5.1 requer duas condições:

- a componente de erro $R^l(x_o)$ deve ser independente ou, no mínimo, ortogonal a $Z^*(x_o)$, que é satisfeita, pois o vetor $Z^*(x_o) - Z(x_o)$ é ortogonal ao estimador da krigagem simples $Z^*(x_o)$;
- a função aleatória $R(x_o)$ deve seguir a correlação espacial do erro real, que é desconhecida.

5.2 MÉTODOS DE SIMULAÇÃO ESTOCÁSTICA

Os métodos de simulação existentes procuram determinar aleatoriamente a componente de erro com base no conhecido método de Monte Carlo. Assim, como o processo é aleatório, as realizações serão diferentes entre si, mas honrando o histograma amostral e o modelo de variograma amostral.

A reprodução do histograma e do variograma é conhecida em Geoestatística como precisão global (Deutsch; Journel, 1992, p. 118). Nesse sentido, a krigagem, que não reproduz o histograma e variograma amostrais, apresenta apenas precisão local, a qual é definida pela alta correlação entre os valores estimados e os valores dos pontos amostrais utilizados.

Segundo Deutsch (2002, p. 162), o método da simulação gaussiana sequencial (SGS) é o mais utilizado na modelagem de reservatórios por sua simplicidade, flexibilidade e razoável eficiência. Ainda de acordo com ele, existem outros algoritmos para simulação estocástica, os quais não são extensivamente utilizados por apresentarem restrições e problemas nos resultados, conforme se reproduz a seguir:

- decomposição LU da matriz dos coeficientes, que tem a restrição da ordem N para sua resolução;



- método das bandas rotativas, que não é usado por causa dos artefatos que produz;
- métodos espectrais com base na transformada rápida de Fourier (FFT), que não são usados em razão da necessidade de se fazer a krigagem para o condicionamento dos dados;
- fractais também não têm sido empregados por causa da suposição restritiva da autossimilaridade;
- métodos de médias móveis são raramente usados por causa da necessidade de tempo de processamento.

O método da simulação gaussiana sequencial pertence à classe de métodos sequenciais, na qual se incluem a simulação indicadora sequencial (Deutsch; Journel, 1992, p. 146-152) e a simulação sequencial direta (Soares, 2001, p. 912-919).

5.3 MÉTODOS SEQUENCIAIS DE SIMULAÇÃO

Por causa da simplicidade de execução, os métodos sequenciais de simulação tornaram-se os algoritmos mais comuns e populares para a reprodução da distribuição espacial e da incerteza de diferentes variáveis nas Ciências da Terra (Soares, 2001, p. 911).

Seja uma distribuição com N variáveis aleatórias $\{Z_i, i = 1, N\}$, em que N é muito grande e pode ser (Deutsch; Journel, 1992, p. 123):

- os nós de uma malha densa, considerando-se as variáveis Z_i como medidas do mesmo atributo;
- os N atributos medidos na mesma localização (x);
- a representação de uma combinação de K diferentes atributos definidos sobre os N' nós de uma malha com $N = KN'$.

Segundo Goovaerts (1997, p. 376), o objetivo da simulação sequencial é a geração das várias realizações conjuntas dessas N variáveis aleatórias:

$$\{z^l(x_j), j = 1, N\}, l = 1, L$$

condicionadas ao conjunto de dados $\{z(x_\alpha), \alpha = 1, n\}$.

É importante salientar que a simulação estocástica pode ser *condicional*, quando passa exatamente pelos pontos amostrais ou condicionantes, ou *não condicional*.

Considerar, em seguida, a simulação conjunta de z em somente dois pontos x_1 e x_2 , da qual se obtém um conjunto de pares de realizações $\{z^l(x_1), z^l(x_2)\}, l = 1, L$ geradas por amostragem da função de distribuição acumulada condicional bivariada (Goovaerts, 1997, p. 376):

$$F(x_1, x_2; z_1, z_2 | (n)) = \Pr \{Z(x_1) \leq z_1, Z(x_2) \leq z_2 | (n)\} \quad (5.2)$$

Ou, alternativamente, de acordo com esse autor:

$$F(x_1, x_2; z_1, z_2 | (n)) = F(x_2; z_2 | (n+1)) \cdot F(x_1; z_1 | (n))$$

ou seja, o valor $z^l(x_1)$ é simulado com base na função de distribuição acumulada condicional $F(x_1; z_1 | (n))$, a qual é posteriormente atualizada pelo valor previamente simulado $z^l(x_1)$, além dos n pontos de dados (Goovaerts, 1997, p. 376).

A Eq. 5.2 pode ser generalizada para N variáveis, ainda segundo ele:

$$F_N(x_1, \dots, x_N; z_1, \dots, z_N | (n)) = \Pr \{ Z(x_i) \leq z_i, i = 1, N | (n) \}$$

a qual, de acordo com ele, pode ser aproximada como produto de N funções de distribuição acumulada condicional, que são determinadas sequencialmente:

$$\begin{aligned} F_N(x_1, \dots, x_N; z_1, \dots, z_N | (n)) = & F(x_1; z_1 | (n)) \\ & \cdot F(x_2; z_2 | (n+1)) \dots \\ & \cdot F(x_{N-1}; z_{N-1} | (n+N-2)) \\ & \cdot F(x_N; z_N | (n+N-1)) \end{aligned}$$

Essa é a fundamentação teórica dos métodos sequenciais de simulação estocástica. Cada novo ponto simulado é usado para atualizar a função de distribuição acumulada condicional, da qual o valor simulado é extraído por Monte Carlo.

5.3.1 Simulação gaussiana sequencial – SGS

Denomina-se simulação gaussiana sequencial (SGS) a aplicação do procedimento de simulação sequencial para funções aleatórias multigaussianas (Goovaerts, 1997, p. 380). Considerando a simulação de N variáveis aleatórias $\{Z(x_i), i = 1, N\}$ localizadas sobre os nós de uma malha regular e condicionadas ao conjunto de n pontos de dados $\{z(x_\alpha), \alpha = 1, n\}$, uma realização SGS é obtida conforme os seguintes passos (Goovaerts, 1997, p. 380-381):

- Inicialmente, a distribuição da variável $Z(x)$ é transformada para uma distribuição normal por meio de $Y(x) = \phi(Z(x))$ (em que ϕ é a função de transformação para os escores da distribuição normal), com média nula, $E[Y(x)] = 0$, e variância unitária, $Var[Y(x)] = 1$.

Em seguida, o variograma experimental da variável transformada $Y(x)$ é calculado e obtém-se o modelo de correlação espacial $\gamma_Y(h)$ que será usado na SGS.

Uma importante etapa dessa técnica consiste em testar a hipótese multigaussiana, que requer que a distribuição de dois, três e n pontos também seja gaussiana. Mas, como a verificação dessa hipótese é inviável na prática para três ou mais pontos, testa-se a hipótese de bigaussianidade, como foi descrito no Cap. 3. Se a distribuição bigaussiana for normal, então se aceita a hipótese de multigaussianidade dos dados.

- Em seguida, a simulação sequencial é feita para a variável $Y(x)$:
 - * define-se um caminho aleatório para a sequência de simulação dos nós da malha regular (Fig. 5.4);
 - * procede-se ao nó (x_o) da sequência definida, para o qual são escolhidos os (n) pontos de dados mais próximos, incluindo-se, nesse conjunto, os pontos amostrais e os nós previamente simulados. Feito isso, faz-se a estimativa em (x_o) por krigagem simples, em que o valor estimado $y_{KS}^*(x_o) = \sum_{i=1}^n \lambda_i y(x_i)$ será a média condicional e a variância de krigagem simples $\sigma_{KS}^2(x_o) = C(0) - \sum_{i=1}^n \lambda_i C(x_i - x_o)$, a variância condicional, as quais definem a função de distribuição acumulada condicional (FDAC);



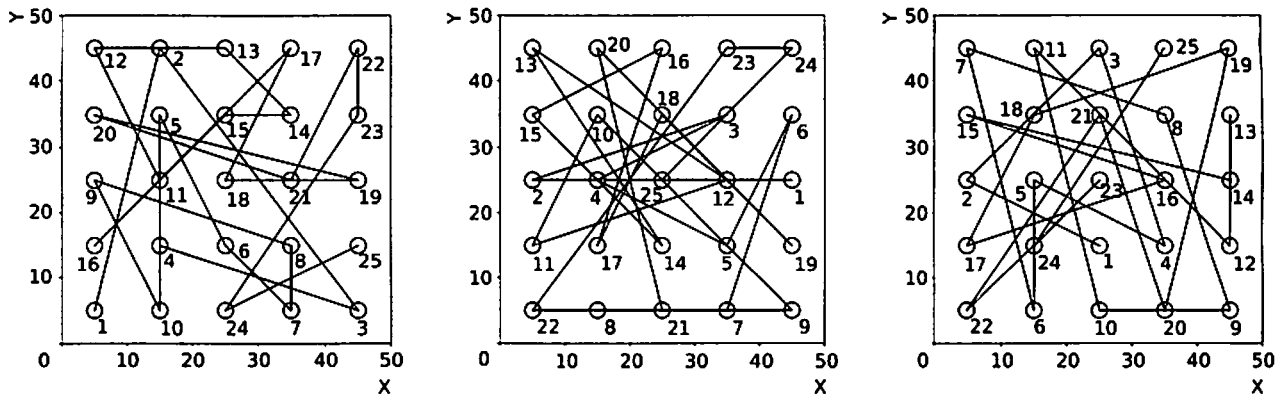


Fig. 5.4 Exemplos de caminhos aleatórios para uma malha regular de 5×5 nós

- * determinada a FDAC em (x_o) , o valor $y^l(x_o)$ é extraído aleatoriamente dela. Isso equivale a adicionar um resíduo aleatório à estimativa $y_{KS}^*(x_o)$, conforme a Eq. 5.1. O valor simulado é adicionado ao conjunto de pontos de dados;
- * o algoritmo é repetido para o próximo nó (x_o) da sequência, definido pelo caminho aleatório, e assim sucessivamente, até que todos os nós da malha regular sejam simulados.
- Ao final da simulação gaussiana sequencial obtém-se o conjunto de valores simulados $\{y^l(x_i), i = 1, N\}$ que estão no domínio da distribuição de Gauss. Desse modo, esses valores devem ser transformados de volta para a escala original da variável:

$$z^l(x_i) = \varphi^{-1}(y^l(x_i)), i = 1, N$$

Tem-se, a seguir, uma demonstração feita por Deutsch (2002, p. 162-163), provando a exatidão da covariância entre o estimador da krigagem simples e o i -ésimo ponto de dado. O sistema de equações de krigagem simples é:

$$\sum_{j=1}^n \lambda_j C(x_i - x_j) = C(x_i - x_o) \text{ para } i = 1, N$$

A covariância entre a estimativa $y_{KS}^*(x_o)$ e o i -ésimo ponto de dado $y(x_i)$ pode ser desenvolvida, de acordo com esse autor, como:

$$\begin{aligned} \text{Cov}\{y_{KS}^*(x_o), y(x_i)\} &= E[y_{KS}^*(x_o) \cdot y(x_i)] - E[y_{KS}^*(x_o)] E[y(x_i)] \\ &= \sum_{j=1}^n \lambda_j E[y(x_j) y(x_i)] - 0 \\ &= \sum_{j=1}^n \lambda_j C(x_j - x_i) = C(x_i - x_o) \end{aligned}$$

Segundo ele, a covariância está certa, mas, como a variância de $y_{KS}^*(x_o)$ é reduzida pelo efeito de suavização, a covariância entre as estimativas de krigagem simples não é correta.

Para corrigir a variância das estimativas krigadas, deve-se adicionar uma variável aleatória com média zero e variância diferente de zero, conforme segue (Deutsch, 2002, p. 163):

$$y_S(x_o) = y_{KS}^*(x_o) + r(x_o)$$

Novamente, ainda segundo ele, calcula-se a covariância entre um valor simulado $y_S(x_o)$ e o i -ésimo ponto de dado $y(x_i)$:

$$\begin{aligned} \text{Cov}\{y_S(x_o), y(x_i)\} &= E[y_S(x_o) \cdot y(x_i)] - E[y_S(x_o)] E[y(x_i)] \\ &= E\left\{\left[\sum_{j=1}^n \lambda_j y(x_j) + r(x_o)\right] \cdot y(x_i)\right\} - 0 \\ &= \sum_{j=1}^n \lambda_j E[y(x_j) \cdot y(x_i)] + 0 \\ &= \sum_{j=1}^n \lambda_j C(x_j - x_i) = C(x_i - x_o) \end{aligned}$$

Esse desenvolvimento mostra que a covariância entre $y_S(x_o)$ e o i -ésimo ponto de dado $y(x_i)$ continua correta, mesmo após a adição de uma componente de resíduo com média zero, $E[r(x_o)] = 0$.

Deutsch (2002, p. 165) conclui que, apesar da simplicidade da SGS, as imagens simuladas apresentam grande desordem espacial, além daquela imposta pelo modelo de correlação espacial.

Krigagem simples ou ordinária?

Essa é a pergunta feita por Deutsch e Journel (1992, p. 142) sobre que método usar para fazer a estimativa local no processo de simulação gaussiana sequencial. De acordo com eles, a decisão prévia de estacionaridade requer que seja usada a krigagem simples com média zero para a estimativa do nó a ser simulado.

O processo da simulação gaussiana sequencial, amostra por Monte Carlo, a função de distribuição acumulada condicional, usando, para isso, um número aleatório (entre zero e 1), que é transformado em escore da distribuição normal acumulada. Esse escore, por sua vez, multiplica o desvio padrão de krigagem, cujo resultado é adicionado ao valor estimado por krigagem simples.

Com a interpretação dos pesos da krigagem ordinária como probabilidades condicionais (Rao; Journel, 1997), entretanto, pode-se obter a função de distribuição acumulada condicional diretamente dos pesos da krigagem ordinária.

Trata-se de um procedimento bastante simples, no qual, após a krigagem ordinária, os n valores encontrados na vizinhança (amostrais e previamente simulados) e os pesos da krigagem ordinária são classificados em ordem crescente:

$$y(x_1) \leq y(x_2) \leq \dots \leq y(x_n)$$



Em seguida os pesos são acumulados, resultando nas probabilidades condicionais acumuladas (Yamamoto, 2010, p. 5):

$$F(x_o; y_\alpha) = \sum_{i=1}^{\alpha} \lambda_i$$

em que $\{\lambda_i, i = 1, n\}$ são os pesos da krigagem ordinária, após eliminação dos pesos negativos e reescalados para soma igual a 1, $\sum_{i=1}^n \lambda_i = 1$.

Assim, obtém-se a função de distribuição acumulada condicional, a qual é caracterizada completamente pela média condicional ($Y_{KO}^*(x_o)$) e pela variância condicional, que é exatamente a variância de interpolação ($S_o^2(x_o)$), como proposta originalmente por Yamamoto (2000, p. 491-493).

A função de distribuição acumulada condicional obtida dessa forma pode ser amostrada aleatoriamente pelo conhecido método de Monte Carlo, resultando, assim, no valor simulado.

Exemplo de aplicação da SGS

Para uma aplicação da SGS, considerar a amostra aleatória estratificada com 64 pontos (Arquivo 12, Anexo B) do conjunto completo lognormal, conforme o mapa de localização de pontos e distribuição de frequências (Fig. 5.5).

Assim, os dados originais $Z(x)$ são transformados para os escores de uma distribuição normal por meio de $Y(x) = \varphi(Z(x))$, conforme ilustrado graficamente na Fig. 5.6.

Os dados originais $Z(x)$ são classificados em ordem crescente, e a eles são atribuídas frequências simples iguais a $1/(n+1)$, que, em seguida, são acumuladas.

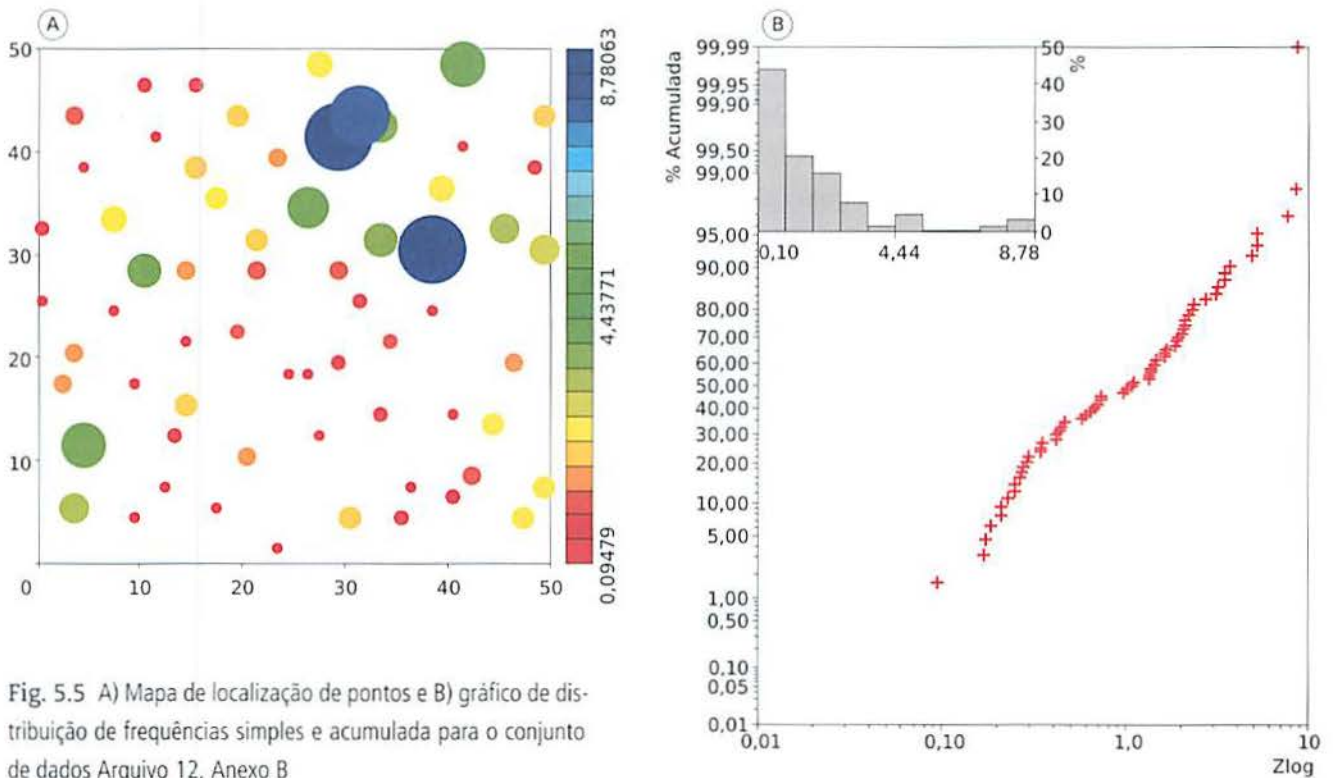


Fig. 5.5 A) Mapa de localização de pontos e B) gráfico de distribuição de frequências simples e acumulada para o conjunto de dados Arquivo 12, Anexo B

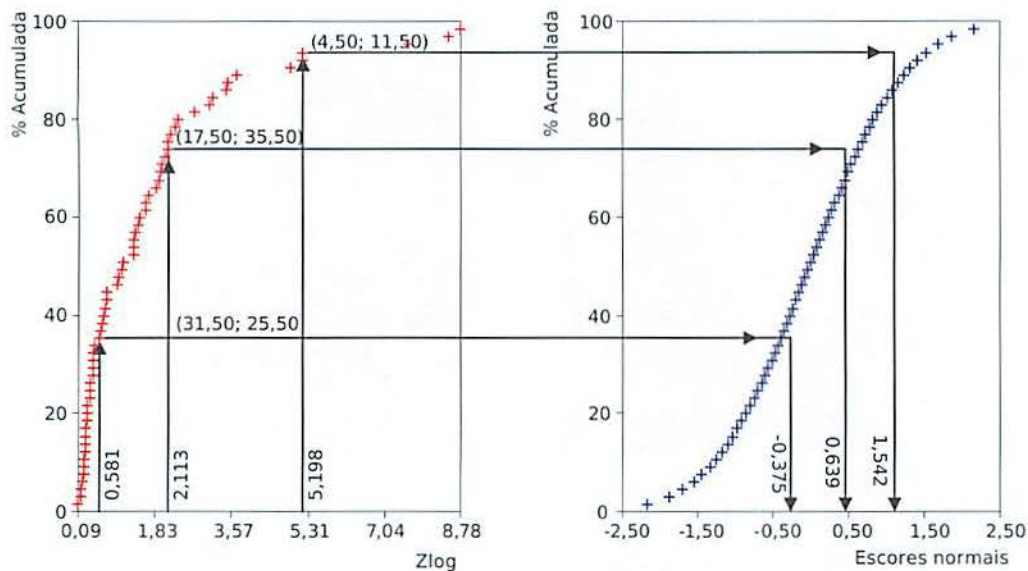


Fig. 5.6 Transformação dos valores originais $Z(x)$ para os escores da distribuição normal $Y(x)$, com indicação das transformações feitas para os pontos de coordenadas (31,5; 25,5), (17,5; 35,5) e (4,5; 11,5)

As frequências acumuladas correspondem aos escores da distribuição normal entre $-2,5$ e $+2,5$. Por exemplo, para o ponto de coordenadas (31,5; 25,5), o valor de $Z(x)$ é igual a 0,581, que corresponde a uma frequência acumulada igual a 35,4%; portanto, o escore da distribuição normal é $-0,375$.

Em seguida, testa-se a bigaussianidade dos dados, conforme apresentado no Cap. 3. Como no caso da amostra em estudo, o teste de bigaussianidade foi positivo (Fig. 3.24), então se procede à modelagem do variograma dos dados transformados $Y(x)$, conforme a Fig. 5.7.

Esse variograma esférico é descrito pelas equações:

$$\begin{cases} \gamma(h) = 0,84 \left[1,5 \frac{h}{14,16} - 0,5 \left(\frac{h}{14,16} \right)^3 \right] & \text{para } h < 14,16 \\ \gamma(h) = 0,84 & \text{para } h \geq 14,16 \end{cases}$$

Com esses dados, procede-se à simulação gaussiana sequencial.

Para esse processamento, a vizinhança para estimativa por krigagem simples foi definida em dois pontos amostrais por quadrante e mais quatro pontos previamente simulados.

Foram feitas 20 realizações, das quais foram escolhidas quatro para apresentação dos resultados, cujos valores estão transformados para a escala original da variável $Z(x)$ (Fig. 5.8).

As realizações mostram, em geral, boa concordância com as imagens krigadas mostradas no Cap. 3. Evidentemente, existem diferenças entre as realizações, as quais ocorrem pela própria natureza dos métodos sequenciais. Por exemplo, na Fig. 5.8D há, no quadrante noroeste, uma região com valores altos que não existe nos dados originais, mas isso se deve à propagação de valores altos de nós previamente simulados.

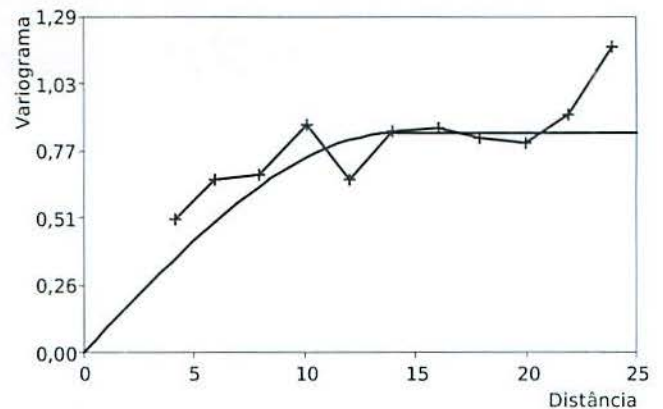


Fig. 5.7 Modelo de variograma para os dados transformados $Y(x)$

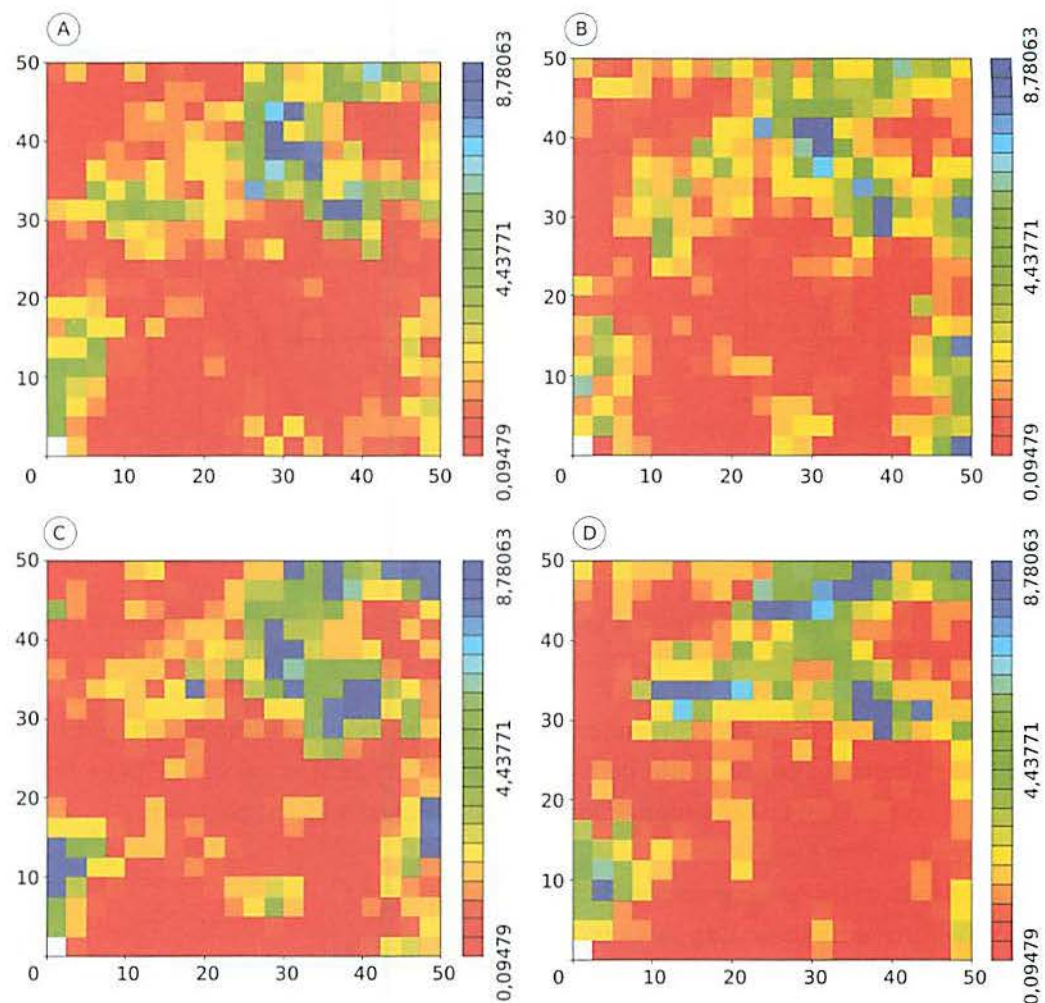


Fig. 5.8 Realizações da simulação gaussiana sequencial: A) realização #9; B) realização #16; C) realização #12; D) realização #7

Da mesma forma, na Fig. 5.8C há valores altos na região sudeste que não se verificam nos dados condicionantes. É preciso ressaltar que se trata de realizações equiprováveis e que elas devem ser analisadas em conjunto.

Para ilustrar o procedimento da SGS, considerar o ponto de coordenadas (26,25; 16,25), para o qual foram escolhidas quatro realizações, conforme os pontos amostrais e nós previamente simulados (Fig. 5.9).

Todas as realizações para o ponto de coordenadas (26,25; 16,25) encontram-se na Tab. 5.1.

Como mostra a Tab. 5.1, em cada realização o nó de coordenadas (26,25; 16,25) é processado conforme a ordem estabelecida no caminho aleatório. Ao valor estimado por krigagem simples é adicionado um erro de simulação, o que resulta no valor estimado nesse ponto.

O erro nada mais é que o produto do escore da distribuição normal $Z_{Gauss}(p)$ para um valor p (extraído aleatoriamente entre zero e 1) e o desvio padrão de krigagem simples $\sigma_{K5}(x_0)$. Nessa tabela, pode-se verificar que o desvio padrão de krigagem simples varia no intervalo de 0,430 a 0,501. Como os arranjos são mais ou menos iguais, a variância de

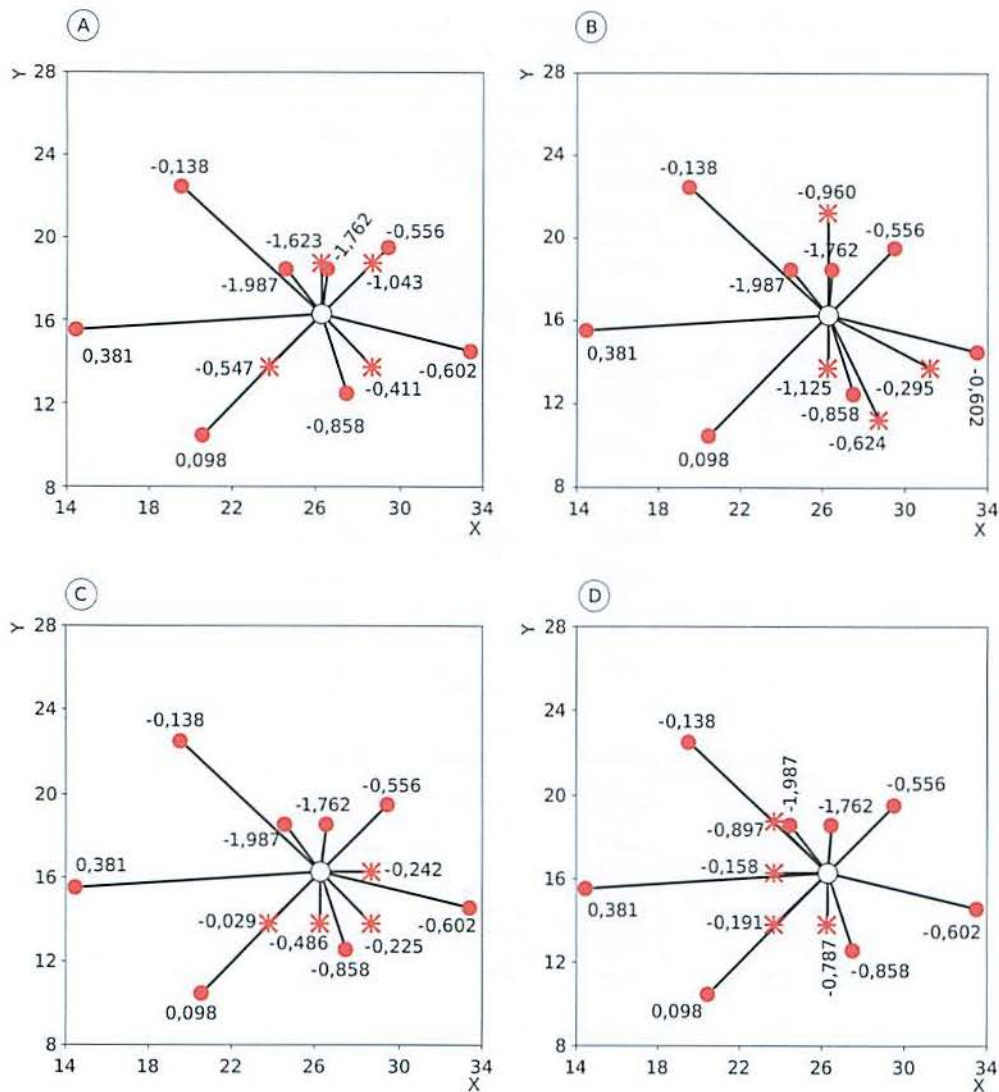


Fig. 5.9 Realizações da SGS para o ponto de coordenadas (26,25; 16,25): A) realização #9; B) realização #16; C) realização #12; D) realização #7. Círculo vazio = ponto para simulação; círculo cheio = pontos amostrais; asterisco = nós previamente simulados. Os valores representados correspondem aos dados transformados $Y(x)$

krigagem simples permanece em torno desses valores, por causa do caráter homocedástico dessa medida de incerteza.

Por exemplo, as realizações 12 e 13 têm a mesma incerteza, pois usam exatamente as mesmas configurações de pontos de dados amostrais e previamente simulados.

Muitas vezes, o valor simulado está fora do intervalo de valores (amostrais e previamente simulados), como ocorre na Fig. 5.9B, na qual o menor valor amostral é $-1,987$ e o valor simulado foi igual a $-2,085$, ou seja, menor que o mínimo. Aparentemente, isso não causaria nenhum problema, mas, como o método é sequencial, esse valor poderá ser usado posteriormente e influenciar diretamente a simulação de outros nós da malha regular.

Os valores simulados $Y^l(x_0)$ são, então, transformados de volta para a escala original de medidas da variável $Z(x)$, como a Fig. 5.10 demonstra para três pontos simulados da Tab. 5.1.

TAB. 5.1 Resultados da SGS para o ponto de coordenadas (26,25; 16,25)

#l	Random	$Y_{KS}^*(x_o)$	ρ	$Z_{Gauss}(\rho)$	$\sigma_{KS}(x_o)$	Erro	$Y^l(x_o)$
1	80	-1,41676	0,54443	0,11160	0,50089	0,05590	-1,36086
2	395	-1,58191	0,36414	-0,34742	0,42985	-0,42696	-2,00870
3	280	-1,33204	0,16928	-0,95703	0,44613	-0,42696	-1,75900
4	342	-1,38902	0,08009	-1,40446	0,42985	-0,60370	-1,99272
5	365	-1,43490	0,69399	0,50720	0,42985	0,21802	-1,21688
6	160	-1,60791	0,41987	-0,20222	0,44476	-0,08994	-1,69785
7	337	-1,13321	0,58964	0,22661	0,44893	0,10173	-1,03148
8	115	-1,68671	0,38420	-0,29447	0,46947	-0,13825	-1,82496
9	173	-1,32878	0,52624	0,06581	0,47346	0,03116	-1,29762
10	162	-1,51995	0,38877	-0,28252	0,46344	-0,13093	-1,65088
11	378	-1,15138	0,45961	-0,10141	0,44927	-0,04556	-1,19694
12	166	-0,97280	0,81854	0,90980	0,44103	0,40125	-0,57155
13	283	-1,86264	0,19536	-0,85830	0,44103	-0,37854	-2,24118
14	285	-1,59564	0,29142	-0,54923	0,42985	-0,23609	-1,83173
15	301	-1,49825	0,06685	-1,49969	0,46866	-0,70285	-2,20110
16	76	-1,51192	0,10820	-1,23617	0,46379	-0,57332	-2,08524
17	182	-1,72422	0,48516	-0,03721	0,44613	-0,01660	-1,74082
18	186	-1,63766	0,86845	1,11909	0,47363	0,53004	-1,10762
19	124	-1,04688	0,93794	1,53772	0,42979	0,66089	-0,38599
20	394	-1,54881	0,71693	0,57375	0,42985	0,24663	-1,30218

Obs.: *Random* significa a ordem dentro da realização para simulação do nó de coordenadas (26,25; 16,25).

Por exemplo, o valor simulado $Y^l(x_o) = -1,031$ corresponde à frequência acumulada igual a 15,123%, que, por sua vez, é projetada na distribuição acumulada dos valores originais, resultando em $Z^l(x_o) = 0,265$, este na escala original da variável $Z(x)$. Os valores de frequência e de $Z(x)$ foram obtidos por interpolação linear.

Como apresentado na seção anterior, pode-se fazer a SGS com base na função de distribuição acumulada condicional derivada diretamente dos pesos da krigagem ordinária, conforme proposto por Rao e Journel (1997). Para ilustrar esse procedimento foram considerados os mesmos dados usados na aproximação por krigagem simples, e os resultados obtidos para quatro realizações escolhidas aleatoriamente encontram-se na Fig. 5.11.

Comparando-se visualmente essas realizações com aquelas obtidas por krigagem simples, conclui-se que elas são bastante semelhantes. A grande vantagem dessa aproximação está na função de distribuição acumulada condicional completamente caracterizada pela média e variância condicionais, dadas respectivamente pela estimativa da krigagem ordinária $Y_{KO}^*(x_o)$ e pela variância de interpolação $S_o^2(x_o)$.

As 20 realizações para o ponto de coordenadas (26,25; 16,25) encontram-se na Tab. 5.2. A média e desvio padrão condicionais constantes nessa tabela não são usados diretamente

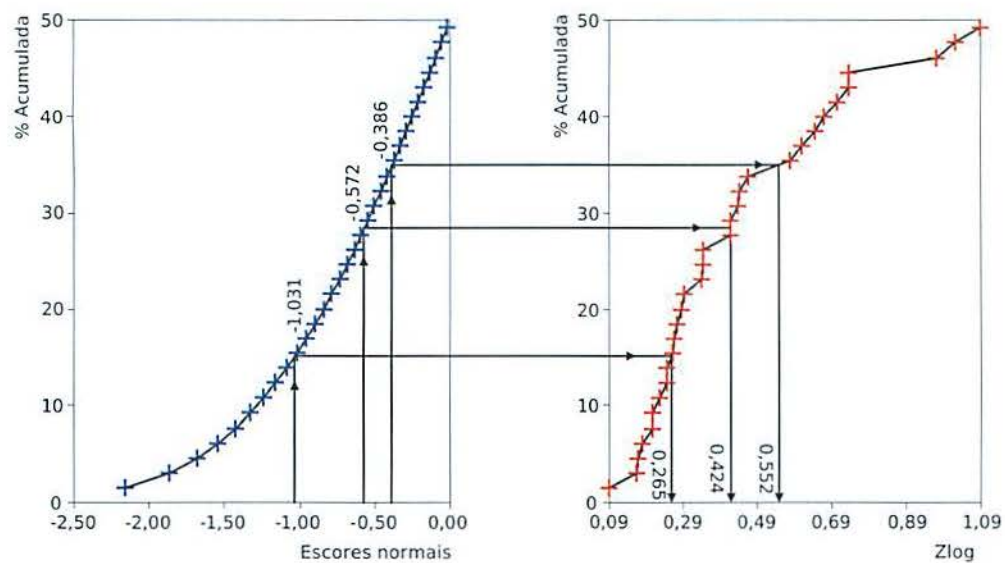


Fig. 5.10 Transformação reversa dos valores simulados para a escala original da variável $Z(x)$. Representação de metade da distribuição total

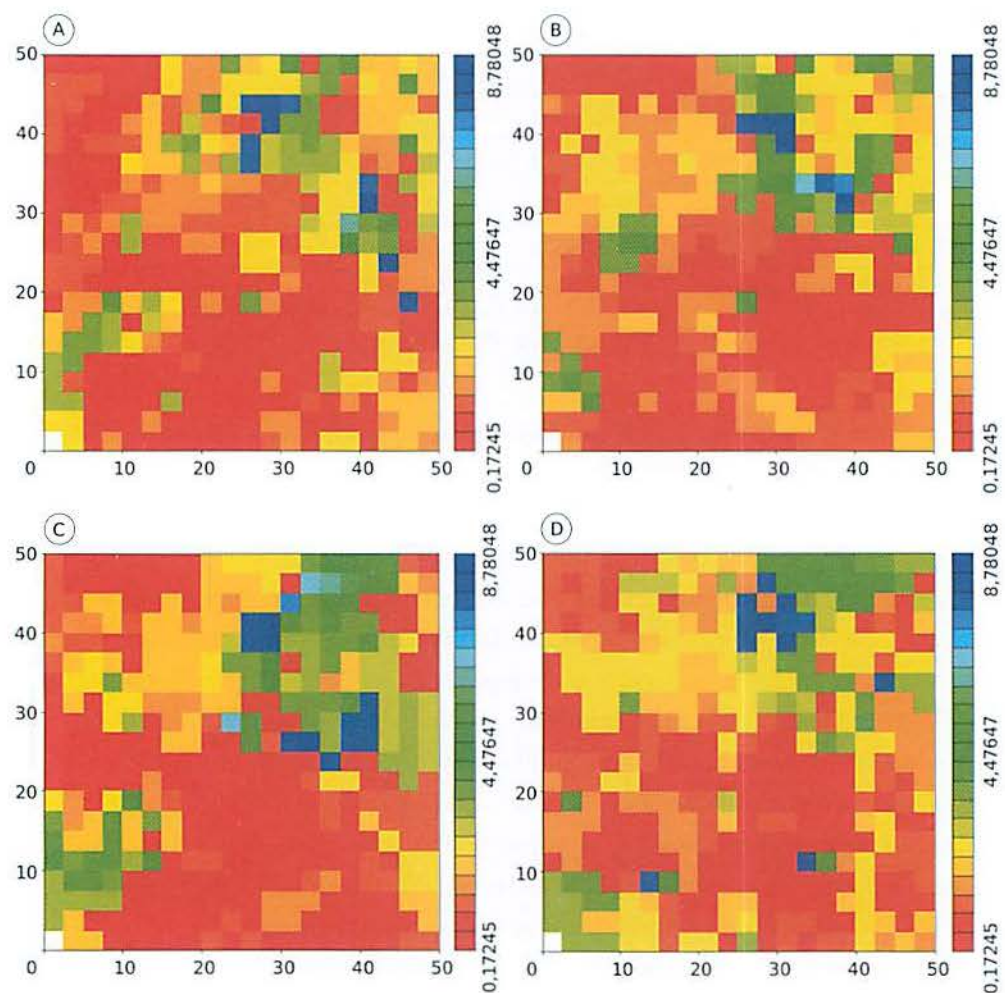


Fig. 5.11 Realizações da SGS com opção pela krigagem ordinária: A) realização #10; B) realização #15; C) realização #1; D) realização #13

TAB. 5.2 Resultados da SGS com opção pela krigagem ordinária para o ponto de coordenadas (26,25; 16,25)

#l	Random	$Y_{KO}^*(x_o)$	$S_o(x_o)$	ρ	Erro	$Y^l(x_o)$
1	144	-1,05185	0,73771	0,60901	0,01212	-1,03973
2	1	-1,27927	0,69607	0,34497	1,06530	-0,21397
3	389	-1,26571	0,47253	0,20820	0,80610	-0,45960
4	29	-1,61456	0,51400	0,47148	1,48839	-0,12617
5	149	-1,02289	0,82184	0,65682	0,45456	-0,56833
6	113	-1,07449	0,71412	0,80110	0,51070	-0,56379
7	92	-1,24012	0,63083	0,82251	0,56538	-0,67474
8	86	-1,21654	0,67638	0,83012	0,63692	-0,57962
9	194	-0,78681	0,72432	0,89363	0,67710	-0,10791
10	11	-1,23932	0,69738	0,81644	0,57227	-0,66706
11	387	-1,11086	0,59879	0,27554	0,33420	-0,77666
12	35	-1,08009	0,76216	0,26400	1,03650	-0,04359
13	241	-0,67792	0,95636	0,45410	-0,42126	-1,09918
14	239	-0,89644	0,66105	0,70010	0,32871	-0,56773
15	297	-1,20945	0,65314	0,59017	-0,15498	-1,36443
16	233	-1,11838	0,71445	0,32095	-0,61429	-1,73267
17	391	-0,75415	0,74548	0,35631	0,15355	-0,60060
18	254	-0,84334	0,85501	0,97387	0,93353	0,09019
19	216	-0,85742	0,82593	0,84915	0,81149	-0,04593
20	347	-1,29332	0,55614	0,51421	-0,40906	-1,70238

Obs.: *Random* significa a ordem dentro da realização para simulação do nó de coordenadas (26,25; 16,25).

na simulação, mas essas estatísticas caracterizam completamente a função de distribuição acumulada condicional.

Na verdade, os pesos da krigagem ordinária (após eliminação dos pesos negativos) são acumulados, e com eles se constrói a função de distribuição acumulada condicional.

A Fig. 5.12 ilustra o procedimento de amostragem da função de distribuição acumulada condicional com base no método de Monte Carlo para quatro realizações do ponto de coordenadas (26,25; 16,25). Por exemplo, na realização 10 (Fig. 5.12A), o valor aleatório foi igual a 0,816, que resultou no valor simulado igual a -0,667, e assim por diante.

Os valores simulados que estão no domínio gaussiano são transformados de volta para a escala original de medida da variável $Z(x)$, conforme mostrado na Fig. 5.10.

Como se pode verificar, os pesos da krigagem ordinária interpretados como probabilidades condicionais podem ser usados para a construção da função de distribuição acumulada condicional, da qual se pode extrair aleatoriamente o valor simulado.

Na verdade, ele só será simulado se o desvio padrão de interpolação for maior que zero, pois, se for igual, significa que o nó a ser simulado coincide com um ponto condicionante e que, portanto, o valor amostral é atribuído ao nó.

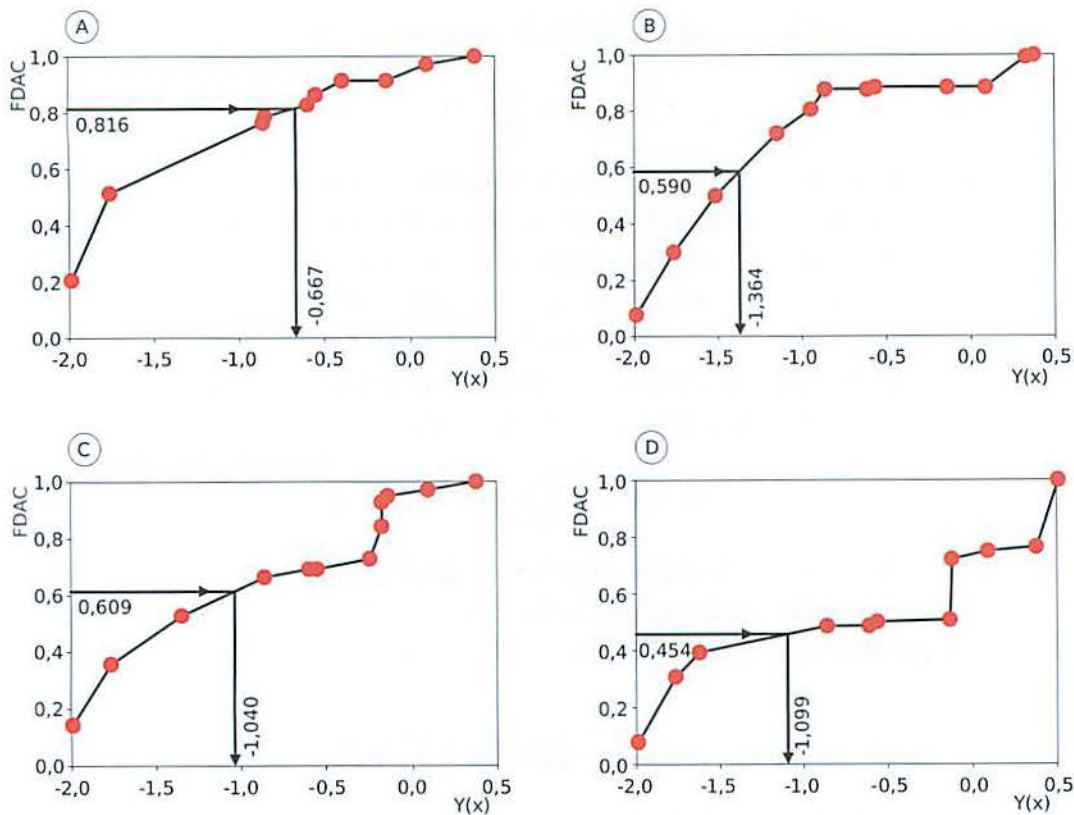


Fig. 5.12 Amostragem da função de distribuição acumulada condicional pelo método de Monte Carlo: A) realização #10; B) realização #15; C) realização #1; D) realização #13

A maior vantagem dessa alternativa está justamente na utilização da função de distribuição acumulada condicional, que é completamente caracterizada pela média e variância condicionais resultantes da krigagem ordinária. Além disso, não ocorre extrapolação de valores, pois o valor aleatório é limitado inferiormente ao menor valor de probabilidade acumulada $F(x_0; y_1)$, de acordo com a programação feita por Rao e Journel (1997):

$$p = F(x_0; y_1) + \text{random} * (1 - F(x_0; y_1))$$

5.3.2 Simulação sequencial direta – SSD

A SGS trabalha sob a hipótese da multigaussianidade dos dados, a qual requer que não apenas o histograma seja gaussiano, mas que a distribuição de dois pontos também o seja.

Entre os métodos sequenciais, destaca-se a simulação sequencial direta (SSD), que não necessita de prévia transformação dos dados da variável de interesse, conforme proposta de Soares (2001, p. 912-915). Segundo esse autor, o modelo de correlação espacial é reproduzido pela simulação se as funções de distribuição acumulada condicional forem centradas na estimativa de krigagem simples:

$$z_{KS}^*(x_0) = m + \sum_{i=1}^n \lambda_i [z(x_i) - m]$$



e com variância condicional igual à variância de krigagem simples:

$$\sigma_{KS}^2(x_o) = C(0) - \sum_{i=1}^n \lambda_i C(x_i - x_o)$$

Ao contrário da SGS, que usa a média e a variância condicionais para a função de distribuição acumulada condicional, essas estatísticas determinam o intervalo de valores da função de distribuição acumulada global $F_Z(z)$, dentro do qual o valor simulado é amostrado por Monte Carlo (Soares, 2001, p. 913). Para definir esse intervalo, esse autor propõe selecionar um conjunto de valores condicionantes $\{z(x_i), i = 1, n\}$, de tal modo que a média e a variância desses n pontos de dados sejam iguais à estimativa por krigagem simples e à variância de krigagem simples, respectivamente:

$$\frac{1}{n} \sum_{i=1}^n z(x_i) = z_{KS}^*(x_o) \quad \text{e} \quad \frac{1}{n} \sum_{i=1}^n [z(x_i) - z_{KS}^*(x_o)]^2 = \sigma_{KS}^2(x_o)$$

Contudo, é muito difícil definir o intervalo de valores centrado em $z_{KS}^*(x_o)$, pois os dados em $Z(x)$ não estão equiespaçados. Assim, Soares (2001, p. 916) propõe usar a função de distribuição acumulada gaussiana, que pode ser caracterizada pela média e variância condicionais:

$$y_{KS}^*(x_o) = \varphi(z_{KS}^*(x_o)) \quad \text{e} \quad \sigma_{KS}^2(x_o) = \frac{1}{n} \sum_{i=1}^n [y(x_i) - y_{KS}^*(x_o)]^2$$

em que φ é a função de transformação para os escores da distribuição normal.

A função de distribuição acumulada gaussiana é, então, amostrada por Monte Carlo. Mas, como o valor amostrado está no domínio gaussiano, o valor simulado é obtido aplicando-se a transformação reversa (Soares, 2001, p. 914):

$$z_{SSD}^l(x_o) = \varphi^{-1}(y_{SSD}^l(x_o))$$

Trata-se, portanto, de um método híbrido entre uma SGS e uma SSD, esta sem a necessidade de transformação dos valores originais para os escores da distribuição normal.

Essa transformação é usada apenas para localizar o intervalo de valores para a construção da função de distribuição acumulada gaussiana, da qual o valor simulado é extraído por Monte Carlo. De qualquer modo, a simulação sequencial direta é uma aproximação perfeitamente válida e teoricamente correta, e deve ser considerada como alternativa viável.

Para o seu cálculo, pode ser recomendado o pacote de programas GeoMS, v. 2001, desenvolvido pelo Centre for Natural Resources and Environment/Cerena do Instituto Superior Técnico de Lisboa, de domínio público e distribuído gratuitamente (<https://sites.google.com/site/cmnpsoftware/geoms>), ou o trabalho de Oz et al. (2003).

5.3.3 Simulação indicadora sequencial – SIS

Como já visto, os métodos sequenciais envolvem a obtenção, em cada ponto a ser simulado, de uma função de distribuição acumulada condicional, a qual é amostrada por Monte Carlo e resulta no valor simulado. O que diferencia as duas aproximações descritas é a forma pela qual são definidas essas funções de distribuição acumulada condicional.

No caso da SGS, essa função é determinada pela média e a variância condicionais resultantes da krigagem simples dos valores transformados para os escores da distribuição

normal. A SSD calcula a média e a variância condicionais dos valores originais, as quais definem o intervalo de valores para a obtenção de uma função de distribuição acumulada gaussiana, da qual o valor simulado é extraído aleatoriamente.

Nesse caso, a média e a variância condicionais devem ser tais que identifiquem a estimativa e a variância obtidas por krigagem simples.

A simulação indicadora sequencial (SIS) também faz a amostragem da função de distribuição acumulada condicional por Monte Carlo, mas com a diferença de que essa função é obtida por meio da krigagem indicadora. Esse método apresenta a grande vantagem de ser aplicado a variáveis aleatórias contínuas ou categóricas.

Segundo Deutsch e Journel (1992, p. 147), a principal contribuição do método da indicadora é a avaliação direta das probabilidades condicionais, as quais são usadas pelo método de simulação sequencial.

A simulação indicadora sequencial é a técnica não gaussiana mais comumente empregada (Goovaerts, 1997, p. 393).

Para a obtenção da função de distribuição acumulada condicional pela krigagem indicadora, procede-se primeiro à transformação para funções indicadoras. Se a variável aleatória for contínua, o intervalo de variação de $Z(x)$ é discretizado por K teores de corte, e as funções indicadoras são assim obtidas:

$$\begin{cases} i(x; z_k) = 1 & \text{se } z(x) \leq z_k \\ i(x; z_k) = 0 & \text{se } z(x) > z_k \end{cases}$$

Por outro lado, uma variável aleatória categórica é definida pelos tipos que a compõem. Assim, as funções indicadoras são determinadas como:

$$\begin{cases} i(x; k) = 1 & \text{se } x \in \text{tipo } k \\ i(x; k) = 0 & \text{se } x \notin \text{tipo } k \end{cases}$$

Após a transformação, têm-se sempre K vetores binários. Para cada ponto a ser simulado, os valores da vizinhança (originais e simulados) são localizados e as K proporções, estimadas pela técnica da krigagem indicadora, como descrito no Cap. 3.

Entretanto, em vez de estimar as funções indicadoras propriamente ditas, Deutsch (2002, p. 169 e 171) propõe usar os resíduos das indicadoras. Para variáveis aleatórias contínuas discretizadas em K teores de corte, tem-se, para esse autor:

$$F_{KI}(x_o; z_k) = \sum_{\alpha=1}^n \lambda_{\alpha} [i(x_{\alpha}; z_k) - F_{z_k}] + F_{z_k} \quad (5.3)$$

em que $F_{z_k} = \frac{1}{N} \sum_{\alpha=1}^N i(x_{\alpha}; z_k)$ é a média para o teor de corte z_k .

Para variáveis aleatórias categóricas com K tipos, tem-se, ainda segundo ele:

$$P_{KI}(x_o; k) = \sum_{\alpha=1}^N \lambda_{\alpha} [i(x_{\alpha}) - p_k] + p_k \quad (5.4)$$

em que $p_k = \frac{1}{N} \sum_{\alpha=1}^N i(x_{\alpha}; k)$ é a proporção do tipo k , de tal modo que $\sum_{k=1}^K p_k = 1$.



Observar que a Eq. 5.3 fornece diretamente as probabilidades acumuladas, enquanto a Eq. 5.4 resulta nas probabilidades simples associadas aos K tipos, necessitando que sejam acumuladas para a obtenção da função de distribuição acumulada condicional.

Para o caso de variáveis contínuas, o valor simulado é extraído diretamente da FDAC, conforme o número aleatório p (entre zero e 1), obtendo-se o valor em função dos teores de corte z_k . No caso de variáveis categóricas, em que a função de distribuição acumulada condicional é discreta, o tipo simulado é obtido verificando-se a que classe pertence o número aleatório p , conforme (Soares, 1998, p. 763):

$$x_o \in \text{tipo } k \text{ se } p \in [P_{KI}(x_o; k-1), P_{KI}(x_o; k)]$$

Como o procedimento da simulação indicadora sequencial é baseado na krigagem indicadora, isso implica o cálculo e modelagem de K variogramas experimentais. Assim, haverá K modelos de variogramas diferentes, pois as funções indicadoras deverão apresentar correlações espaciais distintas.

Para variáveis aleatórias contínuas, os teores de corte situados nas caudas da distribuição de frequências de $Z(x)$ tenderão a apresentar poucos pares possíveis no cálculo do variograma experimental, pois somente pares com indicadoras iguais a 1 e zero, ou vice-versa, podem ser acumulados para o cálculo da função variograma.

Isso significa que os variogramas correspondentes aos teores de corte nas caudas da distribuição terão poucos pares e, conseqüentemente, serão menos confiáveis por causa das flutuações estatísticas. Além da dificuldade no cálculo e modelagem de K variogramas experimentais quando modelos diferentes são usados, a krigagem indicadora não garante que as probabilidades calculadas para teores de corte crescentes não apresentem problemas de relação de ordem.

Dessa forma, a solução é o cálculo e modelagem de um único variograma experimental correspondente ao teor de corte igual à mediana da distribuição de $Z(x)$. O procedimento que utiliza um único modelo de variograma da mediana chama-se krigagem indicadora da mediana (Deutsch; Journel, 1992, p. 79-80), como descrito no Cap. 3.

No caso de variáveis aleatórias discretas, a obtenção de K modelos de variogramas é uma tarefa muito mais difícil, pois os tipos tendem a estar agrupados espacialmente, conforme os tipos que compõem a variável categórica.

As Figs. 5.13 e 5.14 apresentam situações em que os variogramas possíveis são aqueles perpendiculares aos contatos entre os dois tipos de variável categórica. Teoricamente, nesses casos, os variogramas paralelos aos contatos são nulos. Somente os variogramas perpendiculares aos contatos podem ser calculados.

Como o variograma é direcional, não tem sentido usar o variograma perpendicular ao contato para calcular a função variograma em outras direções. Além disso, quando houver poucos pontos para um determinado tipo k , o seu variograma não será significativo por insuficiência de pares de pontos considerados no seu cálculo.

Levando tudo isso em consideração, Yamamoto et al. (2012, p. 148-149) propuseram usar equações multiquádricas (Hardy, 1971, p. 1970-1908) para a interpolação de um tipo em um ponto não amostrado x_o .

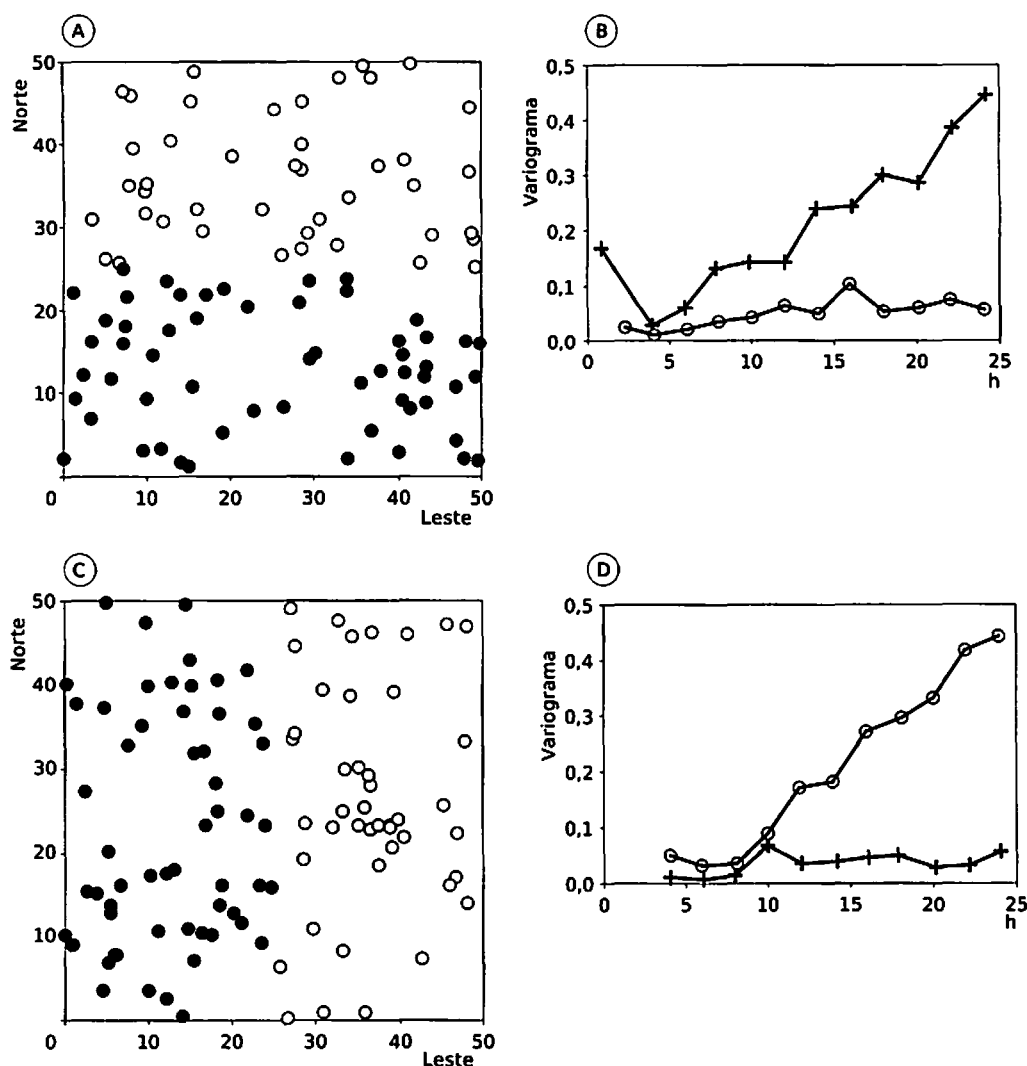


Fig. 5.13 Mapas de localização com os respectivos variogramas experimentais: A) e B) para contato leste-oeste entre os tipos; C) e D) para contato norte-sul entre os tipos. Sinal de mais = direção NS; círculo vazio = direção EW

Essa solução permite obter proporções estimadas iguais ou maiores que zero e sempre fechando a soma em 1, com a dispersão entre variância e proporção descrevendo perfeitamente uma função quadrática (Yamamoto et al., 2012, p. 150).

Certamente, a solução via krigagem indicadora da mediana, para variáveis aleatórias contínuas, ou das equações multiquádricas, para variáveis aleatórias discretas, evita o problema de relações de ordem ou probabilidades negativas (Deutsch, 2002, p. 171 e 174). Essas soluções garantem sempre probabilidades condicionais maiores ou iguais a zero e soma igual a 1 e nunca apresentam problemas de relações de ordem. O problema de relação de ordem ocorre quando $F_{KI}(x_0; z_k) < F_{KI}(x_0; z_{k-1})$, que é inaceitável, pois a função de distribuição acumulada condicional é monotônica crescente. As soluções originais via krigagem indicadora necessitam de repadronizações das probabilidades calculadas para soma unitária, conforme as equações em Deutsch (2002, p. 174 e 187), caso ocorram probabilidades negativas.

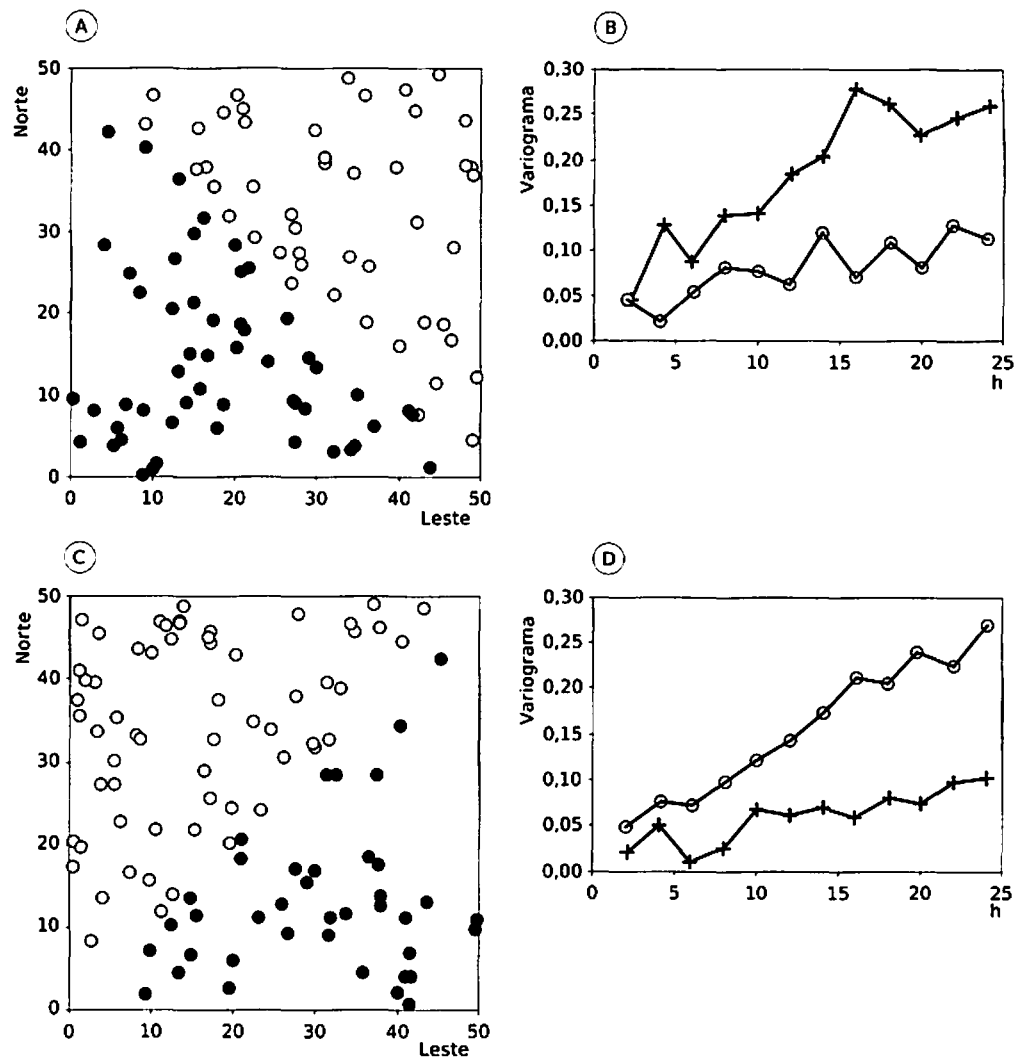


Fig. 5.14 Mapas de localização e respectivos variogramas experimentais: A) e B) para contato NW entre os tipos; C) e D) para contato NE entre os tipos. Sinal de mais = direção NE; círculo vazio = direção NW

Exemplos de aplicação da simulação indicadora sequencial

A simulação indicadora sequencial segue os mesmos passos da simulação gaussiana sequencial, porém com uma diferença na maneira como as funções de distribuição acumulada condicional são construídas. Essa técnica pode ser aplicada para variáveis contínuas ou discretas. Para variáveis contínuas, há necessidade de se fazer a categorização conforme teores de corte previamente estabelecidos. Os exemplos de aplicação a seguir foram separados para variáveis contínuas e categóricas.

Variáveis contínuas

Para variáveis contínuas, considerar o mesmo conjunto de pontos de dados (Arquivo 12, Anexo B – Cap. 2) que foi usado para ilustrar aplicações diversas, tais como krigagem ordinária, krigagem indicadora e simulação gaussiana sequencial. Nesse caso, a primeira etapa é a categorização dos dados em intervalos de valores predefinidos conforme os teores de corte.

No programa utilizado para esse exemplo foi definido um número de teores de corte igual a 19, o que equivale a dividir a distribuição em 20 intervalos. Com isso, têm-se os percentis definidos a cada 5%, em que o primeiro percentil é igual a 5% e o último, igual a 95%. Portanto, apenas 90% da distribuição está representada. Os 19 teores de corte estão relacionados na Tab. 5.3.

Como se pode verificar nessa tabela, os extremos não estão representados e, portanto, não deverão ser obtidos durante a simulação sequencial. A solução seria aumentar o número de teores de corte, mas esse esforço não é compensado quando se trata de distribuições lognormais, como é o caso da amostra em estudo.

Assim, optou-se por redefinir os teores de corte inicial e final, conforme segue:

$$zcut[1] = z_{\min}^* 1,001 \quad zcut[kcut] = z_{\max}^* 0,999$$

em que $z_{\min}$ é o valor mínimo de $Z(x)$ e $z_{\max}$, o valor máximo.

Com isso, consegue-se representar a maior parte da distribuição. No caso em estudo, os valores mínimo e máximo são iguais a 0,095 e 8,781, respectivamente. Os teores de corte extremos redefinidos ficaram iguais a: $zcut[1] = 0,095$ e $zcut[19] = 8,772$. Com esses novos limites, apenas os valores mínimo e máximo não foram incluídos, ou seja, apenas 3,1% do total.

Definidos os teores de corte, calcula-se o variograma da variável indicadora da mediana (igual a 1,089), conforme Fig. 3.37 e Eq. 3.47. Assim, procede-se à simulação indicadora sequencial, de acordo com a metodologia descrita nesta seção.

Para cada teor de corte, estima-se a probabilidade acumulada condicional. Ao final, têm-se 19 pares ordenados $(zcut_k, F_{KI}^*(x_0; zcut_k))$, que constituem a função de distribuição acumulada condicional no ponto simulado x_0 . Em seguida, um número aleatório entre zero e 1 é gerado, o qual determina o valor da probabilidade acumulada condicional, que, por sua vez, corresponde ao valor de $Z(x)$. Essa função de distribuição acumulada condicional é completamente caracterizada pela média e variância condicionais, respectivamente a estimativa do tipo E (Eq. 3.42) e a variância associada (Eq. 3.43). Os resultados obtidos para 20 realizações, das quais quatro foram escolhidas aleatoriamente, encontram-se representados na Fig. 5.15.

Essas imagens (Fig. 5.15) são mais comparáveis às realizações obtidas por meio da simulação gaussiana sequencial com opção pela krigagem ordinária (Fig. 5.11) do que com aquelas obtidas por krigagem simples (Fig. 5.8). Na verdade, a simulação indicadora sequencial e a simulação gaussiana sequencial com opção pela krigagem ordinária são muito parecidas, pois se baseiam na amostragem aleatória da função de distribuição acumulada condicional.

Dessa forma, a opção pela krigagem ordinária para a obtenção da função de distribuição acumulada condicional para a amostragem por Monte Carlo se torna uma opção tecnicamente viável. As realizações obtidas para o ponto de coordenadas (26,25; 16,25) estão reproduzidas na Tab. 5.4.

TAB. 5.3 Teores de corte definidos para 19 percentis da distribuição de frequências de $Z(x)$

k	%	$zcut$
1	5	0,176
2	10	0,221
3	15	0,262
4	20	0,288
5	25	0,348
6	30	0,429
7	35	0,515
8	40	0,663
9	45	0,739
10	50	1,089
11	55	1,357
12	60	1,456
13	65	1,663
14	70	1,937
15	75	2,113
16	80	2,320
17	85	3,118
18	90	3,618
19	95	5,194



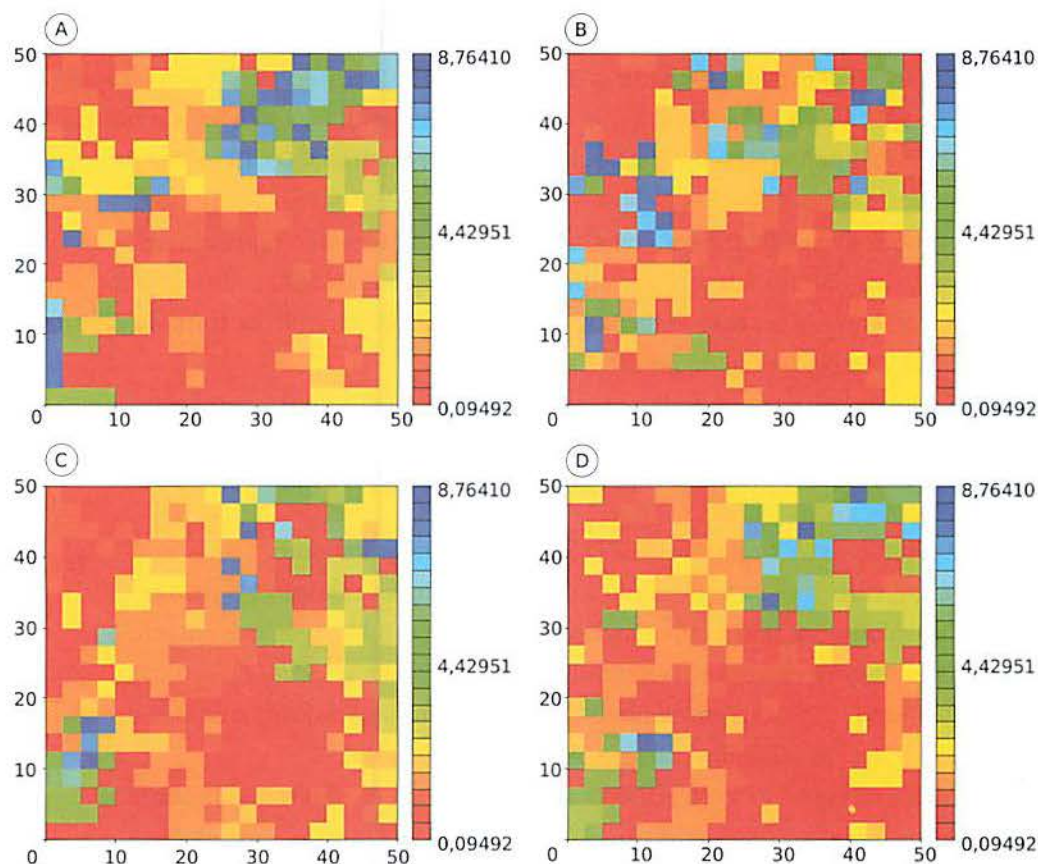


Fig. 5.15 Realizações da simulação indicadora sequencial: A) realização #6; B) realização #15; C) realização #18; D) realização #11

As funções de distribuição acumulada condicional para as quatro realizações do ponto de coordenadas (26,25; 16,25) encontram-se representadas na Fig. 5.16. Nessa figura é possível observar que não houve nenhuma situação com problema de relação de ordem, uma vez que a krigagem indicadora foi baseada em um único variograma da variável indicadora da mediana.

Como se pode verificar, o método da simulação indicadora sequencial apresenta a vantagem de a função de distribuição acumulada condicional ser completamente caracterizada pela média e variância condicionais calculadas durante o processo de krigagem indicadora.

Por outro lado, na simulação gaussiana sequencial, a função de distribuição acumulada condicional é sempre suposta como sendo normal, apesar de a média e variância condicionais não refletirem as estatísticas da distribuição normal, ou seja, média zero e variância 1.

Variáveis categóricas

A grande vantagem da simulação indicadora sequencial está justamente na possibilidade de se aplicar a mesma metodologia tanto para variáveis contínuas como para variáveis categóricas. As variáveis categóricas têm tido papel importante como variável auxiliar em trabalhos de avaliação de reservatórios, bem como em problemas de estimativa de recursos minerais. Muitas vezes, a variável categórica é considerada como secundária e pode ser

TAB. 5.4 Resultados da simulação indicadora sequencial para o ponto de coordenadas (26,25; 16,25)

#l	Random	$Z_E^*(x_o)$	$\sigma_E^2(x_o)$	P	Erro	$Z^l(x_o)$
1	230	0,328	0,063	0,748	0,063	0,391
2	305	0,501	0,164	0,244	-0,281	0,220
3	241	0,449	0,152	0,769	-0,037	0,412
4	166	0,488	0,192	0,553	-0,140	0,348
5	31	0,646	0,441	0,304	-0,449	0,197
6	263	0,339	0,071	0,850	0,071	0,410
7	113	0,314	0,050	0,209	-0,163	0,151
8	292	0,253	0,042	0,581	-0,048	0,205
9	292	0,253	0,042	0,581	-0,048	0,205
10	206	0,581	0,216	0,595	-0,184	0,397
11	230	0,552	0,319	0,465	-0,238	0,314
12	266	0,361	0,213	0,954	1,436	1,797
13	224	0,309	0,074	0,880	0,092	0,401
14	351	0,289	0,056	0,931	0,126	0,415
15	87	0,250	0,032	0,864	0,090	0,340
16	37	0,363	0,130	0,304	-0,189	0,174
17	15	0,366	0,105	0,860	0,052	0,418
18	204	0,420	0,167	0,745	-0,049	0,371
19	363	0,311	0,035	0,210	-0,104	0,207
20	234	0,441	0,158	0,508	-0,125	0,316

Obs.: *Random* significa a ordem dentro da realização para simulação do nó de coordenadas (26,25; 16,25).

completamente conhecida, ao contrário da variável primária, que tem custo elevado de aquisição.

A amostra composta por 50 pontos (Fig. 5.17), para esse exemplo de aplicação, foi extraída por amostragem aleatória estratificada (Arquivo 21, Anexo B) do conjunto completo gerado por Yamamoto et al. (2012, p. 147-148).

Como não há necessidade de se calcular e modelar os variogramas experimentais para os tipos da variável categórica, procede-se diretamente à simulação indicadora sequencial. As estimativas das proporções são obtidas por equações multiquádricas.

As proporções são acumuladas, resultando na função de distribuição acumulada condicional, que é amostrada por meio de um número aleatório (entre zero e 1) para simular diretamente o tipo da variável categórica. Foram feitas 20 realizações, com base nos oito pontos mais próximos por quadrante e com até quatro nós previamente simulados, conforme os resultados de quatro realizações escolhidas aleatoriamente (Fig. 5.18).

As realizações mostram, em geral, o predomínio dos tipos amostrados em suas regiões, mas inevitavelmente ocorrem contaminações de tipos vizinhos em áreas supostamente pertencentes aos tipos originais.



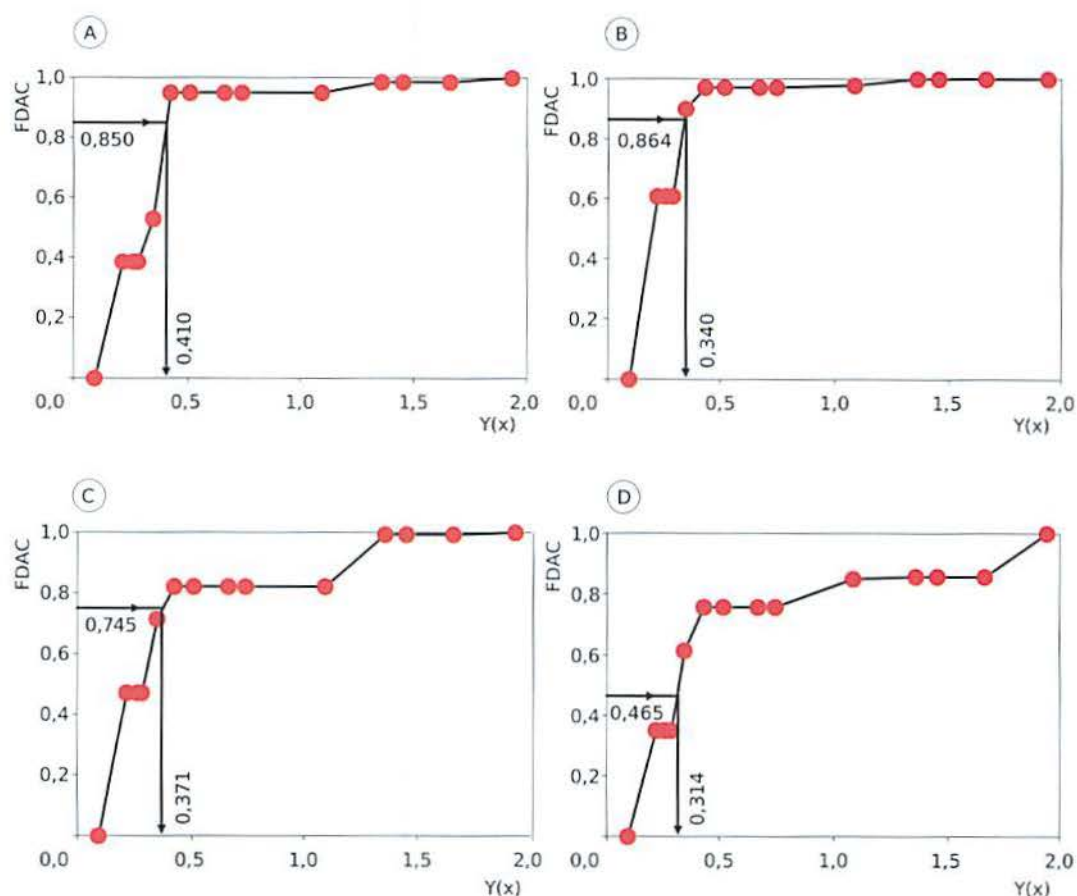


Fig. 5.16 Realizações da simulação indicadora sequencial para o ponto de coordenadas (26,25; 16,25): A) realização #6; B) realização #15; C) realização #18; D) realização #11

Por exemplo, o tipo I sofre contaminações dos tipos II, III e IV, que o circundam; o tipo II, da mesma forma, sofre influência dos tipos I, III e V; e assim por diante. Se esses tipos representassem fácies, litologias ou tipos de solos, as realizações obtidas não poderiam ser aceitas como cenários possíveis da realidade.

Segundo Deutsch (1996, p. 1.669), a simulação indicadora sequencial para variáveis categóricas leva frequentemente a transições incontroladas e geologicamente irreais entre os tipos simulados. A Fig. 5.18 confirma exatamente essa observação.

Esse problema é inevitável, pois, uma vez obtida a função de distribuição acumulada condicional, ela é amostrada aleatoriamente, sem nenhuma restrição. Com a finalidade de mostrar o procedimento de construção da função de distribuição acumulada condicional e sua amostragem por Monte Carlo, os dados obtidos para o ponto de coordenadas (28,75; 31,25) e para as realizações representadas na Fig. 5.18 encontram-se nas Tabs. 5.5 a 5.8.

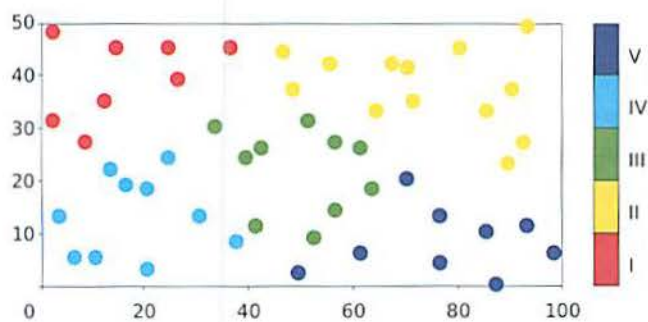


Fig. 5.17 Mapa de localização de tipos da variável categórica
Fonte: Yamamoto et. al. (2012, p. 148).

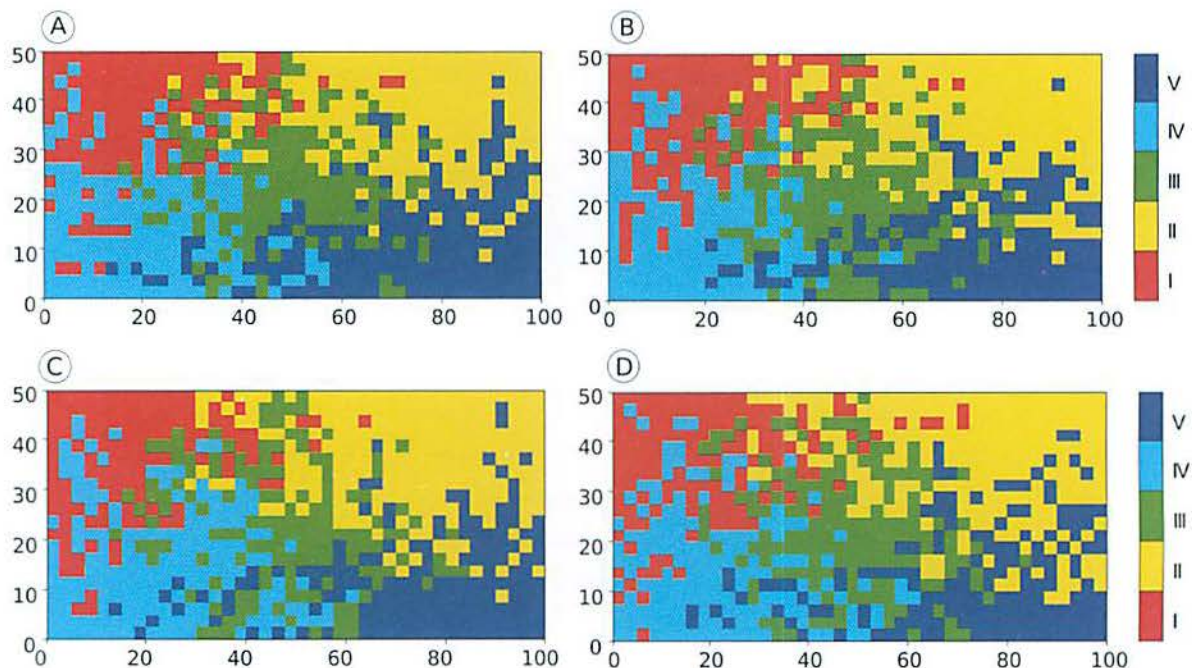


Fig. 5.18 Realizações da SIS para variáveis categóricas: A) realização #6; B) realização #16; C) realização #1; D) realização #18

Ao se observar essas tabelas, verifica-se que os pontos amostrais ou condicionantes são os mesmos para as quatro realizações, e que os pontos previamente simulados são diferentes conforme as realizações, por causa do caminho aleatório. A Tab. 5.8 também mostra que não foi possível localizar nós previamente simulados, pois o ponto foi visitado

TAB. 5.5 Dados para o ponto de coordenadas (28,75; 31,25) visitado na 435ª iteração, pesos da realização #6 e probabilidades condicionais simples e acumuladas

Coordenadas		Pesos	Codificação binária dos tipos				
x_α	y_α	λ_α^6	$i(x_\alpha, 1)$	$i(x_\alpha, 2)$	$i(x_\alpha, 3)$	$i(x_\alpha, 4)$	$i(x_\alpha, 5)$
36,50	45,50	0,090	1	0	0	0	0
48,50	37,50	0,165	0	1	0	0	0
26,50	39,50	0,015	1	0	0	0	0
24,50	45,50	0,176	1	0	0	0	0
20,50	18,50	0,210	0	0	0	1	0
24,50	24,50	0,032	0	0	0	1	0
39,50	24,50	0,108	0	0	1	0	0
33,50	30,50	0,000	0	0	1	0	0
28,75	28,75	0,045	0	0	1	0	0
26,25	31,25	0,063	1	0	0	0	0
28,75	33,75	0,022	0	0	1	0	0
26,50	33,75	0,074	0	0	1	0	0
$i_{MQ}^*(x_o; k)$			0,343	0,165	0,250	0,242	0
$\sum_k i_{MQ}^*(x_o; k)$			0,343	0,508	0,758	1,000	1,000

TAB. 5.6 Dados para o ponto de coordenadas (28,75; 31,25) visitado na 275ª iteração, pesos da realização #16 e probabilidades condicionais simples e acumuladas

Coordenadas		Pesos	Codificação binária dos tipos				
x_α	y_α	λ_α^{16}	$i(x_\alpha,1)$	$i(x_\alpha,2)$	$i(x_\alpha,3)$	$i(x_\alpha,4)$	$i(x_\alpha,5)$
36,50	45,50	0,090	1	0	0	0	0
48,50	37,50	0,169	0	1	0	0	0
26,50	39,50	0,003	1	0	0	0	0
24,50	45,50	0,180	1	0	0	0	0
20,50	18,50	0,216	0	0	0	1	0
24,50	24,50	0,028	0	0	0	1	0
39,50	24,50	0,112	0	0	1	0	0
33,50	30,50	0,000	0	0	1	0	0
31,25	31,25	0,030	0	0	1	0	0
31,25	28,75	0,059	0	0	0	1	0
23,75	31,25	0,072	1	0	0	0	0
28,75	26,25	0,040	0	0	1	0	0
$i_{MQ}^*(x_o; k)$			0,346	0,169	0,182	0,303	0
$\sum_k i_{MQ}^*(x_o; k)$			0,346	0,515	0,697	1,000	1,000

TAB. 5.7 Dados para o ponto de coordenadas (28,75; 31,25) visitado na 168ª iteração, pesos da realização #1 e probabilidades condicionais simples e acumuladas

Coordenadas		Pesos	Codificação binária dos tipos				
x_α	y_α	λ_α^1	$i(x_\alpha,1)$	$i(x_\alpha,2)$	$i(x_\alpha,3)$	$i(x_\alpha,4)$	$i(x_\alpha,5)$
36,50	45,50	0,097	1	0	0	0	0
48,50	37,50	0,174	0	1	0	0	0
26,50	39,50	0,025	1	0	0	0	0
24,50	45,50	0,183	1	0	0	0	0
20,50	18,50	0,221	0	0	0	1	0
24,50	24,50	0,034	0	0	0	1	0
39,50	24,50	0,110	0	0	1	0	0
33,50	30,50	0,001	0	0	1	0	0
31,25	33,75	0,054	0	1	0	0	0
28,75	36,25	0,060	0	0	1	0	0
26,25	26,25	0,042	1	0	0	0	0
31,25	36,25	0,000	1	0	0	0	0
$i_{MQ}^*(x_o; k)$			0,346	0,227	0,171	0,255	0
$\sum_k i_{MQ}^*(x_o; k)$			0,346	0,574	0,745	1,000	1,000

TAB. 5.8 Dados para o ponto de coordenadas (28,75; 31,25) visitado na 48ª iteração, pesos da realização #18 e probabilidades condicionais simples e acumuladas

Coordenadas		Pesos	Codificação binária dos tipos				
x_α	y_α	λ_α^{18}	$i(x_\alpha, 1)$	$i(x_\alpha, 2)$	$i(x_\alpha, 3)$	$i(x_\alpha, 4)$	$i(x_\alpha, 5)$
36,50	45,50	0,116	1	0	0	0	0
48,50	37,50	0,201	0	1	0	0	0
26,50	39,50	0,027	1	0	0	0	0
24,50	45,50	0,215	1	0	0	0	0
20,50	18,50	0,253	0	0	0	1	0
24,50	24,50	0,048	0	0	0	1	0
39,50	24,50	0,139	0	0	1	0	0
33,50	30,50	0,000	0	0	1	0	0
$i_{MQ}^*(x_\alpha; k)$			0,359	0,201	0,139	0,301	0
$\sum_k i_{MQ}^*(x_\alpha; k)$			0,359	0,560	0,699	1,000	1,000

na 48ª vez. A seguir, são mostradas as funções de distribuição acumulada condicional para as quatro realizações escolhidas sobre o ponto de coordenadas (28,75; 31,25), conforme a Fig. 5.19.

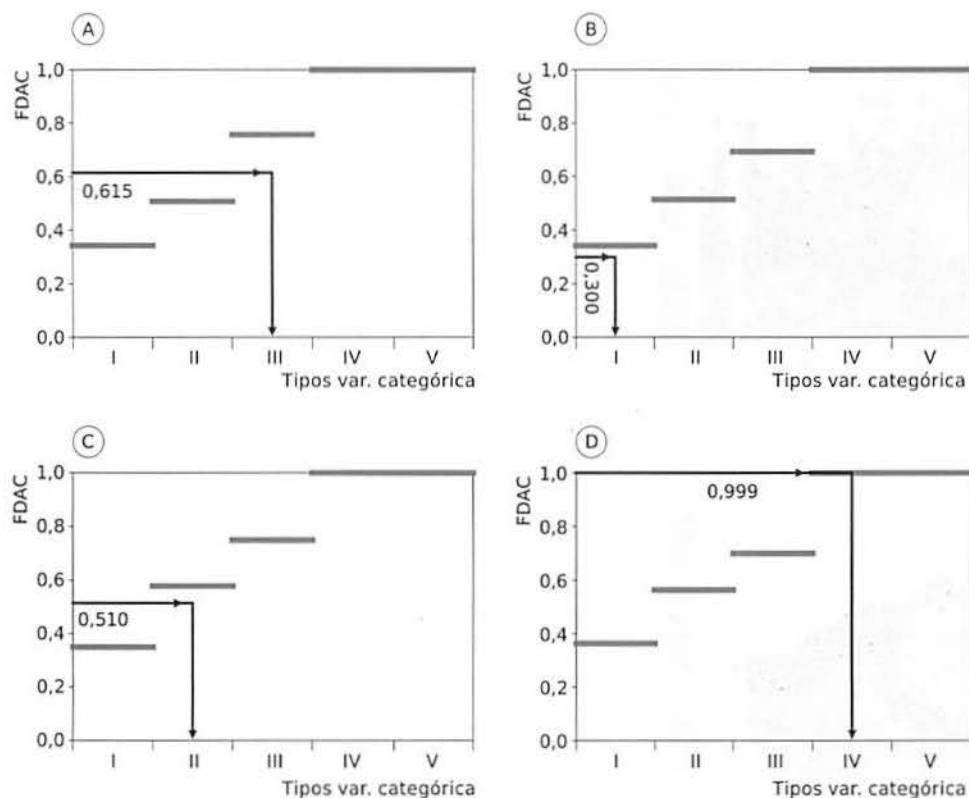


Fig. 5.19 Realizações da simulação indicadora sequencial para variáveis categóricas para o ponto de coordenadas (28,75; 31,25): A) realização #6; B) realização #16; C) realização #1; D) realização #18. O número abaixo da linha horizontal indica o valor aleatório, que amostra o tipo da variável categórica

De acordo com essa figura, em cada realização foi escolhido aleatoriamente um tipo distinto. Esse resultado é bastante razoável, considerando-se que o ponto escolhido está sobre o contato entre os tipos I, III e IV. Contudo, a amostragem aleatória da função de distribuição acumulada condicional pode levar a resultados geologicamente impossíveis. Por essa razão, os métodos de simulação baseados em objetos ou processos proporcionam maior controle estrutural (Deutsch, 1996, p. 1.669).

As realizações podem ser analisadas em conjunto para determinar o tipo mais provável, bem como a zona de incerteza, que preferencialmente deverá estar localizada nos contatos entre os tipos da variável categórica. Para cada célula ou pixel determina-se a proporção do k -ésimo tipo fazendo-se a varredura para as L realizações da simulação sequencial indicadora:

$$p(x_o; k) = \frac{\sum_{l=1}^L i^{*l}(x_o; k)}{L} \quad (5.5)$$

O tipo mais provável é obtido simplesmente determinando-se a proporção máxima, conforme proposta de Teng e Koike (2007, p. 533):

$$p_{\max} = \max [p(x_o; 1), p(x_o; 2), \dots, p(x_o; K)] \quad (5.6)$$

A variância associada à proporção $p(x_o; k)$ é calculada como:

$$S_k^2(x_o) = p(x_o; k) \cdot [1 - p(x_o; k)] \quad (5.7)$$

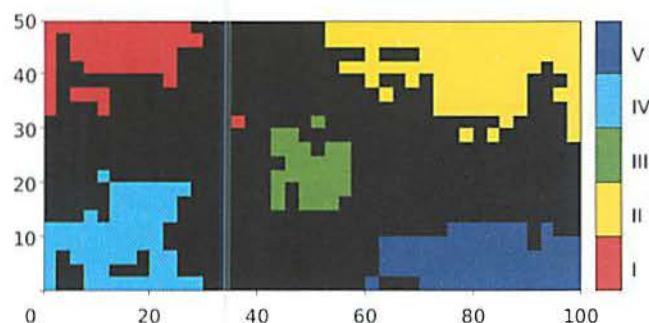


Fig. 5.20 Tipos mais prováveis e zona de incerteza (em preto) derivada de cem realizações da simulação indicadora sequencial

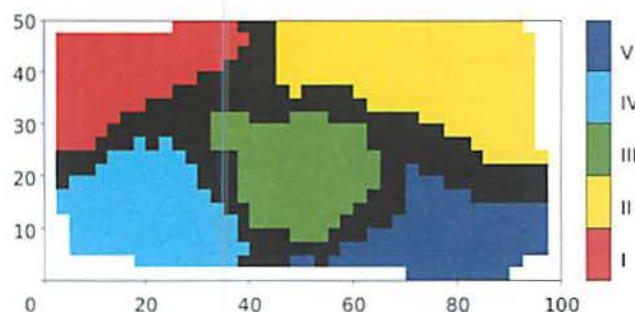


Fig. 5.21 Tipos mais prováveis interpolados por equações multiquádricas e zona de incerteza (em preto) resultante para variância de interpolação superior a 0,20

Assim, tem-se para cada célula ou pixel o tipo mais provável e a sua variância. Se a variância for menor que 0,20, prevalece o tipo mais provável; caso contrário, significa que há incerteza quanto ao tipo mais provável e, então, a célula recebe uma legenda diferente indicando pertencer à zona de incerteza.

Usando os mesmos dados da amostra com 50 pontos (Fig. 5.17), procedeu-se à simulação indicadora sequencial com cem realizações. Após o processamento, conforme as Eqs. 5.5 a 5.7, obteve-se o resultado apresentado na Fig. 5.20.

A zona de incerteza derivada das realizações da simulação indicadora sequencial é muito mais ampla que aquela obtida diretamente por meio da variância de interpolação. Consequentemente, os domínios de cada tipo ficam reduzidos, notadamente os tipos I, III e IV. A Fig. 5.21 mostra, para fins de comparação, a zona de incerteza mapeada por meio da variância de interpolação, segundo metodologia descrita por Yamamoto et al. (2012), para os mesmos dados e parâmetros da malha regular usados na simulação sequencial indicadora.

5.4 CONSIDERAÇÕES SOBRE OS MÉTODOS DE SIMULAÇÃO ESTOCÁSTICA

Foram apresentados os métodos geoestatísticos mais usuais de simulação estocástica baseados em células ou pixels. Os métodos baseados em processos ou eventos não foram considerados neste capítulo, pois estão fora do escopo deste livro.

Os métodos sequenciais descritos neste capítulo foram: gaussiana sequencial com opções pela krigagem simples e krigagem ordinária; indicadora sequencial para variáveis aleatórias contínuas e discretas. Todas as apresentações de resultados são seguidas de tabelas mostrando os procedimentos adotados.

Os resultados obtidos para a simulação sequencial indicadora para variáveis discretas mostraram a necessidade de se aperfeiçoar esse método, pois a amostragem aleatória da função de distribuição acumulada condicional pode resultar em tipos com pequenas chances de ocorrer. Por exemplo, os tipos III e IV avançando na área de certeza do domínio I.

De qualquer forma, apesar das críticas legítimas contra a simulação indicadora sequencial, há ainda boas razões para se usar esse método: estatísticas facilmente inferidas com base nos dados amostrais; algoritmo robusto; transferência direta da incerteza das categóricas com base nos resultados numéricos (Deutsch, 1996, p. 1.670).

Finalmente, a Fig. 5.22 apresenta uma síntese dos métodos de simulação sequencial. No topo dessa figura encontram-se exemplos de caminhos aleatórios definidos para as realizações. Em seguida, os procedimentos podem ser: simulação gaussiana sequencial e simulação indicadora sequencial.

Para a simulação gaussiana sequencial, há opções pela krigagem simples e krigagem ordinária, enquanto, para a simulação indicadora sequencial, pode-se escolher tanto para variáveis contínuas como para variáveis categóricas. Em cada ponto da malha regular define-se a função de distribuição acumulada condicional.

No caso da opção por krigagem simples, a função de distribuição acumulada condicional é uma gaussiana com média $y_{KS}^*(x_0)$ e $\sigma_{KS}(x_0)$ (Fig. 5.22E). Para os demais métodos, as funções de distribuição acumulada condicional são obtidas experimentalmente. Por fim, os resultados são apresentados na base da Fig. 5.22.



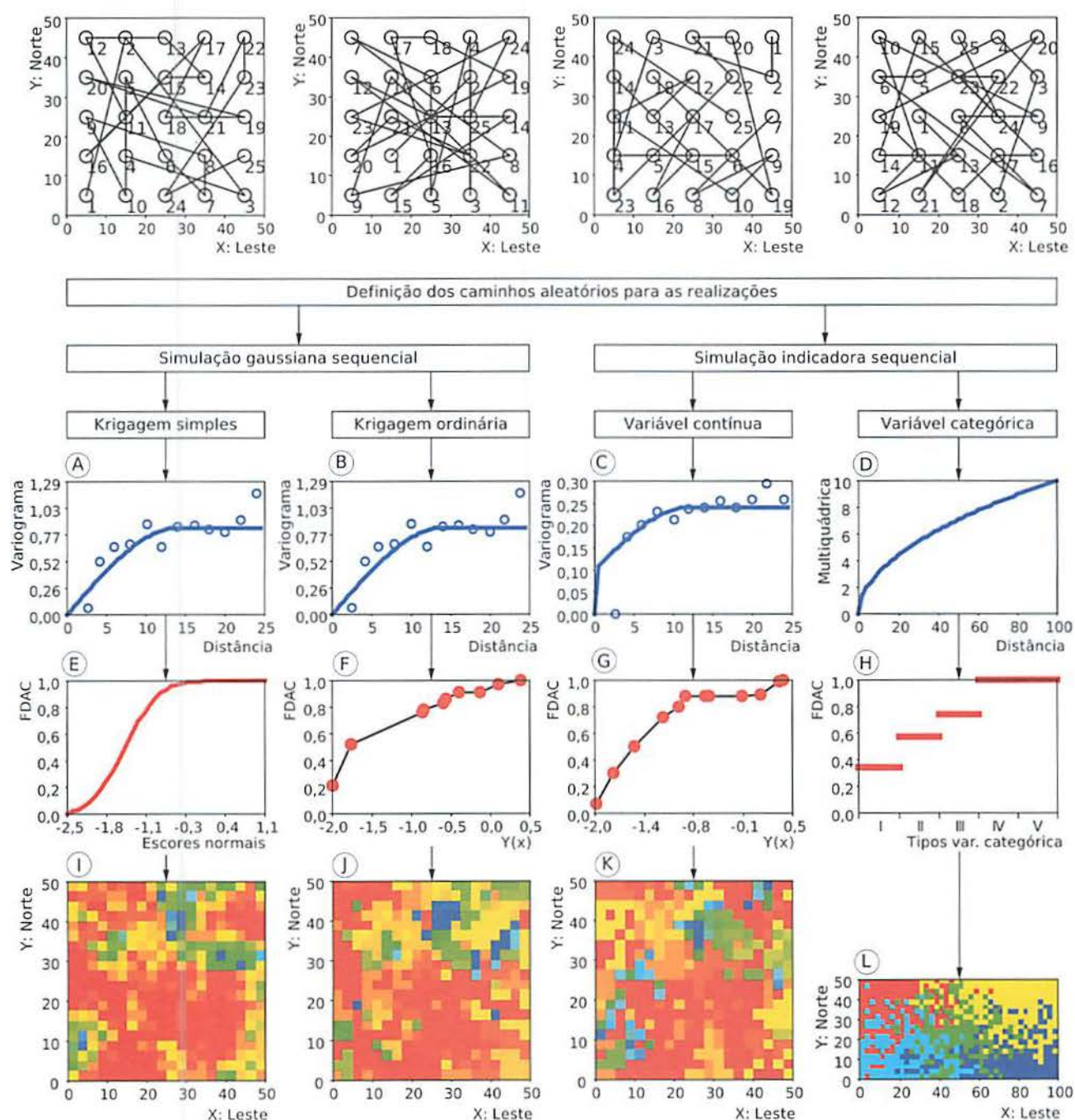


Fig. 5.22 Síntese dos métodos sequenciais de simulação estocástica: no alto, definição dos caminhos aleatórios para as realizações; A e B) variograma da variável transformada para escores normais; C) variograma indicadora da mediana; D) núcleo multiquadrático com constante nula; E, F, G e H) funções de distribuição acumulada condicional; resultado da simulação gaussiana sequencial – I) opção pela krigagem simples e J) opção pela krigagem ordinária; resultado da simulação indicadora sequencial – K) variável contínua e L) variável categórica

Anexo

A

FUNDAMENTOS MATEMÁTICOS E ESTATÍSTICOS

A utilização de técnicas geoestatísticas requer o conhecimento de alguns fundamentos matemáticos e estatísticos que são imprescindíveis para o melhor entendimento dos conceitos empregados nessa metodologia.

A seguir serão expostos, de maneira resumida, alguns desses conceitos. Para mais detalhes, podem ser consultados Waltham (2000), Davis (2002; Caps. 2 e 3) e Borradaire (2003), entre outros.

A.1 MÉTODOS GRÁFICOS DE APRESENTAÇÃO DE DADOS

Como visto no Cap. 1, as variáveis podem ser consideradas como categóricas/discretas e contínuas, ambas apresentando distribuições de valores e usadas para uma análise exploratória dos dados.

A análise estatística tem por objetivo resumir a informação disponível. Nesse sentido, os gráficos mais usuais para a apresentação de dados são o histograma, a curva de distribuição acumulada e a distribuição espacial de pontos. Para ilustrar os conceitos estatísticos, o arquivo fornecido por Goovaerts (1997, p. 4-6) será considerado. Na realidade, os dois arquivos, denominados *prediction* e *validation*, foram aglutinados em um único arquivo, doravante chamado *juradata.txt*.

O histograma é um gráfico de barras, que são proporcionais às frequências de classes. O intervalo de valores entre o mínimo e o máximo é dividido em um número de classes, as quais acumulam as contagens dos valores encontrados no arquivo de dados. Para o exemplo dos dados do arquivo *juradata.txt* contendo diversas variáveis contínuas e discretas, foi feita a representação da variável cobalto em histograma (Fig. A.1).

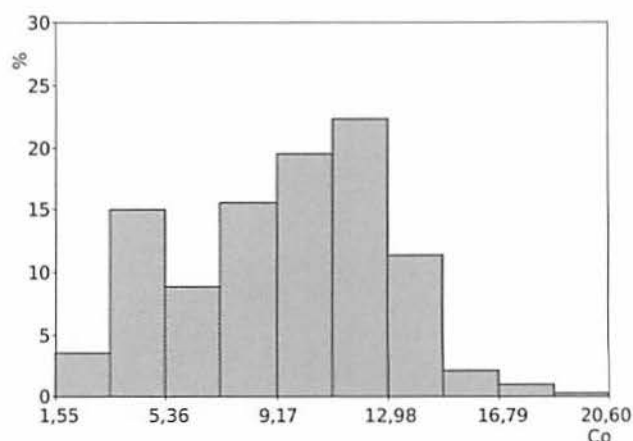


Fig. A.1 Histograma da distribuição de frequências da variável cobalto

Fonte: Goovaerts (1997, p. 4-6).

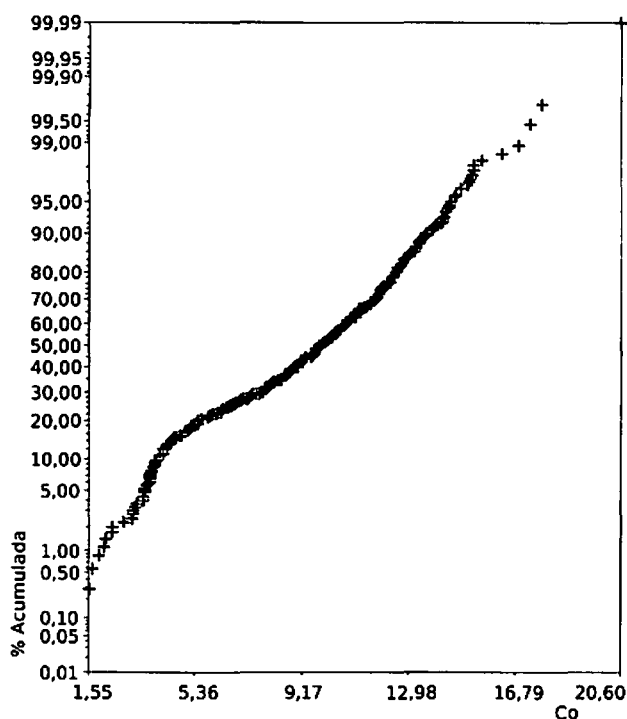


Fig. A.2 Distribuição acumulativa da variável cobalto
Fonte: Goovaerts (1997, p. 4-6).

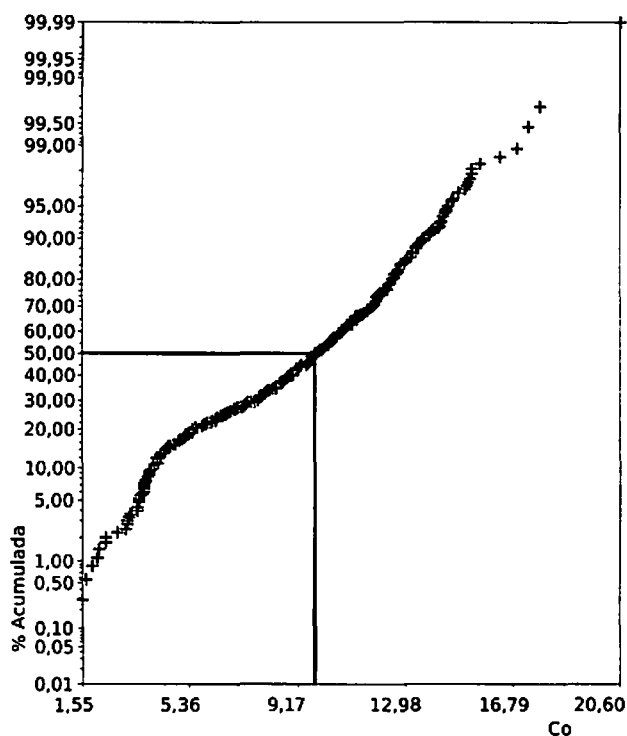


Fig. A.3 Distribuição acumulativa da variável cobalto com indicação do percentil 50
Fonte: Goovaerts (1997, p. 4-6).

No histograma pode ser verificado o valor mínimo e o máximo; se a distribuição é unimodal, bimodal ou plurimodal; sendo unimodal, se a distribuição é simétrica ou assimétrica; se, no caso de presença de assimetria, há indicação ou não de valores anômalos (outliers) etc. Na Fig. A.1, pode-se verificar que a distribuição é praticamente unimodal, com uma leve assimetria negativa.

O histograma mostra as frequências simples por classes, mas pode-se calcular as frequências acumuladas e representá-las em uma curva de distribuição acumulada, em cujo gráfico o eixo vertical é dimensionado em escala de probabilidade aritmética, de modo que os dados de uma distribuição normal se mostram como uma linha reta. Assim, pode-se verificar visualmente se uma amostra se aproxima ou não de uma distribuição normal.

Para o caso da variável cobalto, a curva acumulativa encontra-se representada na Fig. A.2.

Em vez de acumular as frequências simples das classes usadas no histograma, o gráfico da Fig. A.2 oferece outra opção para a representação de frequências acumuladas, em que cada valor do conjunto com n dados recebe uma frequência simples igual a $1/n$. Em seguida, essas frequências simples são acumuladas. Nesse gráfico é possível verificar que a distribuição é praticamente normal, pois os pontos se aproximam bem de uma linha reta na escala de probabilidade aritmética.

Também os percentis da distribuição podem ser lidos diretamente com base na curva encontrada, fornecendo de maneira rápida os quartis ou os valores centrais e de variação da amostra; o percentil 50 corresponde à mediana, que é igual à média se a distribuição for normal; a relação (percentil 84 - percentil 16)/2 corresponde ao desvio padrão. Por exemplo, a Fig. A.3 mostra a curva acumulativa com indicação da mediana, que corresponde a 9,81.

Podem ocorrer situações em que a escala a ser adotada no eixo horizontal seja logarítmica, quando a distribuição apresentar assimetria positiva. Para mostrar um exemplo de uso da curva acumulativa

em escala de logprobabilidade aritmética, foi considerada a variável cobre do conjunto juradata.txt (Goovaerts, 1997, p. 4-6).

A Fig. A.4 mostra a distribuição acumulada para a variável cobre, em que é possível verificar a assimetria positiva dada pela linha quase reta da distribuição acumulada, bem como pelo histograma no canto superior esquerdo dessa figura.

A distribuição espacial de pontos de amostragem não é uma maneira gráfica comum na análise de dados em Estatística, mas é de fundamental importância quando se trata de qualquer trabalho de análise da distribuição e variabilidade espaciais da variável de interesse.

Pela configuração obtida, pode-se verificar: se a área de estudo foi total ou parcialmente amostrada; qual o padrão de distribuição dos pontos (regular, semirregular ou irregular); a presença de agrupamentos (clusters); a localização de valores altos e/ou baixos etc. A distribuição dos pontos pode ser feita apenas mostrando a sua localização, mas também por meio de mapas de localização com legendas proporcionais às magnitudes medidas nas localizações amostradas (Fig. A.5).

Nessa figura é possível observar que o quadrante nordeste não foi amostrado e, por isso, a interpolação não pode ser feita nessa região. Outra maneira de se representar a distribuição espacial dos pontos e frequências associadas está exemplificada na Fig. A.6.

A.2 ESTATÍSTICA DESCRITIVA

As representações gráficas apresentadas na seção anterior, em forma de histograma e curva acumulada, dão uma boa ideia do tipo de distribuição e dos valores extremos, bem como da dispersão, seja pela assimetria do histograma ou pela inclinação da reta na distribuição acumulada representada em escala de probabilidade aritmética. Entretanto, essas observações são apenas de caráter qualitativo. Uma descrição quantitativa da distribuição de frequências é possível por meio das estatísticas descritivas: média, variância, coeficiente de variação, assimetria e curtose.

A média de uma variável aleatória X pode ser obtida por meio de:

$$E[X] = \bar{X} = \sum_{i=1}^n x_i$$

Além da média, pode-se ter outras medidas de tendência central da distribuição, tais como: *mediana* e *moda*. A mediana corresponde à metade da distribuição, ou seja, o valor referente a 50% da frequência acumulada. Trata-se de uma medida mais robusta que a média,

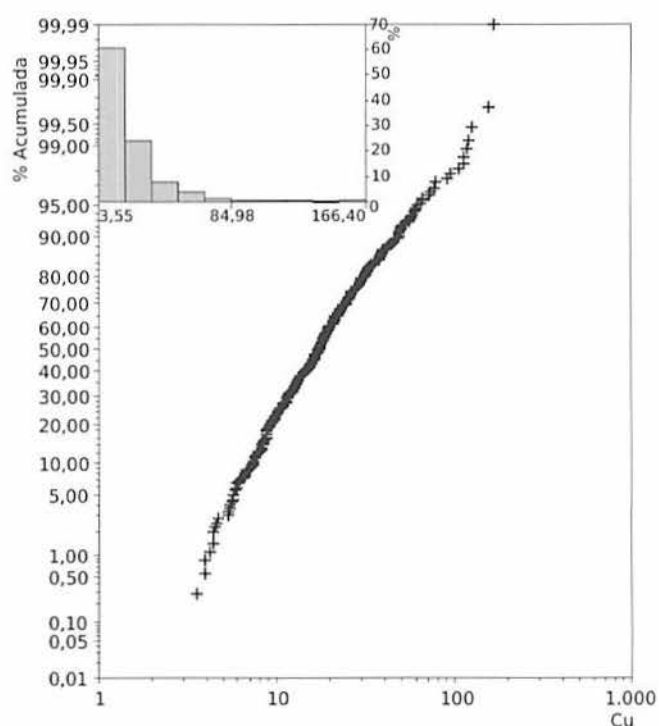


Fig. A.4 Distribuição acumulada da variável cobre em escala de logprobabilidade aritmética

Fonte: Goovaerts (1997, p. 4-6).

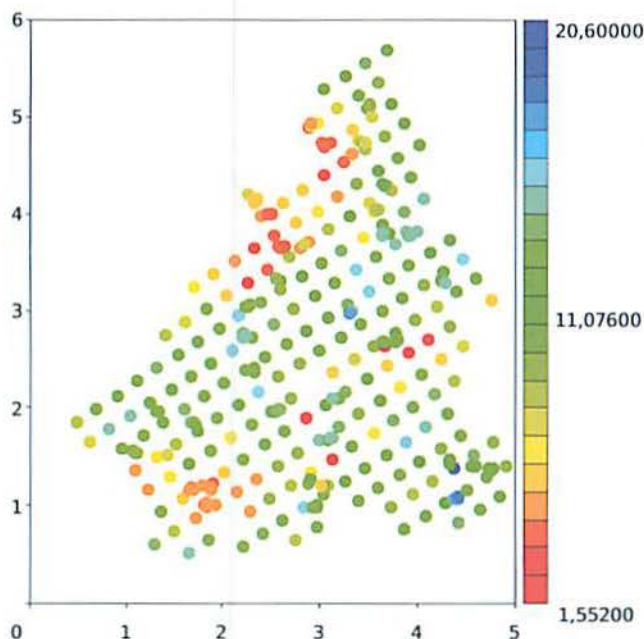


Fig. A.5 Distribuição dos pontos de amostragem da variável cobre
Fonte: Goovaerts (1997, p. 4-6).

pois não é tão sensível aos dados, especialmente em casos de distribuição assimétrica. Define-se moda como o valor mais frequente da distribuição, se os dados são discretos, ou o intervalo de classe com maior frequência, se os dados são contínuos.

A dispersão em torno da média pode ser medida pela variância:

$$Var[X] = S^2 = E[X^2] - (E[X])^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{X})^2$$

Com relação ao denominador, pode-se usar $n - 1$ no lugar de n quando é usada a variância amostral para inferir a variância populacional:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n - 1}$$

O desvio padrão (S) é simplesmente a raiz quadrada da variância. Dividindo-se o desvio padrão pela média, obtém-se o coeficiente de variação:

$$CV = \frac{S}{\bar{X}}$$

O coeficiente de variação é uma medida conveniente da dispersão de uma distribuição de frequências, pois ele é adimensional. Assim, essa

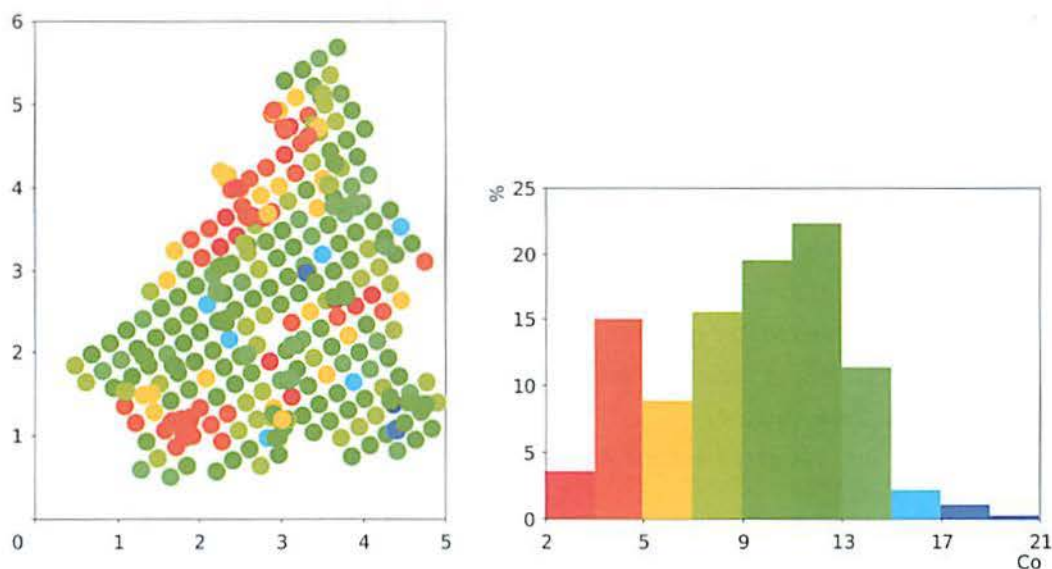


Fig. A.6 Distribuição dos pontos de amostragem, cujas cores são as mesmas correspondentes às classes do histograma

Fonte dos dados: Goovaerts (1997, p. 4-6).

estatística serve para comparar distribuições de frequências de valores completamente diferentes.

O coeficiente de assimetria mede a assimetria do histograma e é baseado no 3º momento em torno da média:

$$CA = \frac{\sum_{i=1}^n (x_i - \bar{X})^3}{S^3}$$

Como essa medida é uma somatória de diferença elevada à potência ímpar, o coeficiente de assimetria pode ser negativo, quando a cauda da distribuição está à esquerda, e positivo, quando a cauda à direita é mais longa; se for zero ou próximo de zero, significa que a distribuição é simétrica. O coeficiente de assimetria é uma medida da dispersão para distribuições assimétricas.

O coeficiente de curtose mede o grau de achatamento de uma distribuição e é baseado no 4º momento em torno da média:

$$CC = \frac{\sum_{i=1}^n (x_i - \bar{X})^4}{S^4}$$

Nesse caso, como se trata de somatória de diferença elevada à potência par, o resultado é sempre positivo. O coeficiente de curtose mede a dispersão para distribuições simétricas, as quais podem ser *leptocúrticas* (pontilagudas), *mesocúrticas* (com achatamento equivalente à distribuição de Gauss) e *platicúrticas* (achatadas).

A.3 ESTATÍSTICA BIVARIADA

Sejam X e Y duas variáveis aleatórias medidas nas mesmas localizações. A relação mútua entre elas pode ser observada em um diagrama de dispersão, que nada mais é que a representação dos pontos em um sistema de eixos cartesianos. Se os pontos nesse diagrama se localizarem próximo a uma reta, a relação é dita linear, e uma equação linear torna-se apropriada para os fins de análise de correlação entre as duas variáveis, isto é, de estimativa do comportamento de uma variável em relação à outra. Para a base de dados juradata.txt, pode-se analisar a relação mútua entre as diversas variáveis existentes. Na Fig. A.7, mostra-se o diagrama de dispersão entre N_i e Cr , cuja dispersão mostra uma relação linear positiva.

A medida da relação mútua é denominada coeficiente de correlação linear de Pearson, e pode ser calculada como (Landim, 2003, p. 99):

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var[X] Var[Y]}}, \text{ em que } Cov(X,Y) = \frac{1}{n} \sum_{i=1}^n (x_i y_i - \bar{X} \cdot \bar{Y})$$

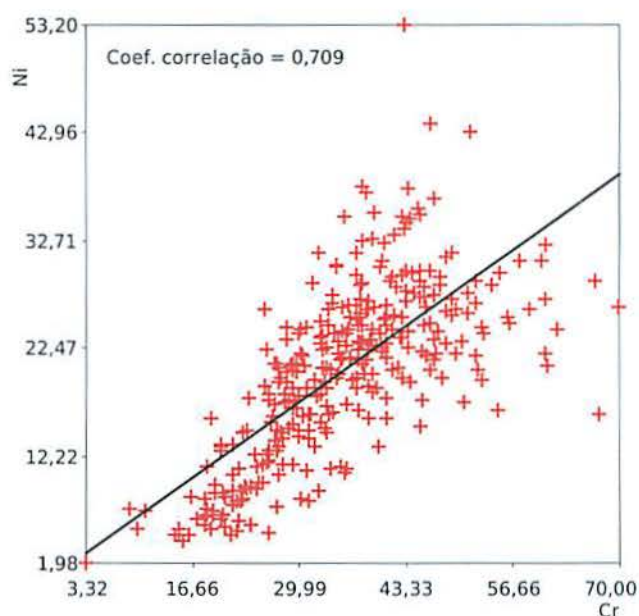


Fig. A.7 Diagrama de dispersão entre N_i e Cr
Fonte: Goovaerts (1997, p. 4-6).

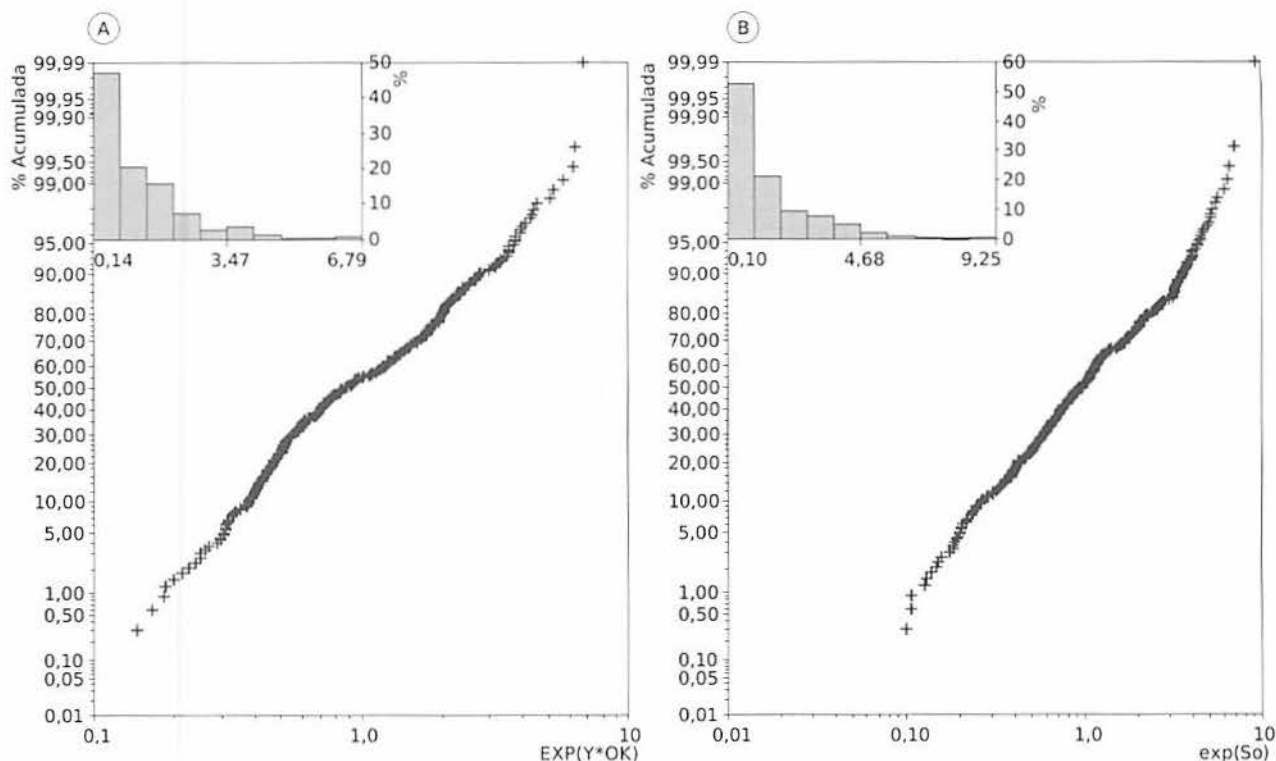


Fig. A.8 Curvas acumulativas e histogramas para as variáveis: A) EXP(Y*OK) e B) exp(So)

Ao contrário da covariância, que depende das unidades das variáveis X e Y , o coeficiente de correlação é adimensional e pode variar entre -1 e $+1$ (Landim, 2003, p. 100). Desse modo, quando igual a -1 , expressa uma relação linear negativa, e, quando igual a $+1$, expressa uma relação linear positiva. Quando $\rho_{XY} = 0$, significa que não há correlação entre X e Y (Landim, 2003, p. 100).

Quando $\rho_{XY} = 0$, significa que não há correlação entre X e Y (Landim, 2003, p. 100).

A fórmula de Pearson foi desenvolvida para cálculo da correlação para variáveis apresentando distribuições normais. Entretanto, quando as variáveis são lognormais ou suas distribuições são assimétricas positivas, a fórmula de Pearson não é adequada. Por exemplo, são representadas na Fig. A.8 duas variáveis cujas distribuições apresentam assimetrias positivas.

O diagrama de dispersão mostra uma grande concentração de valores próximo à origem, por causa da natureza lognormal dessas variáveis (Fig. A.9).

Nesses casos, quando as variáveis não apresentam distribuição normal, usa-se o coeficiente de correlação não paramétrico de Spearman (Landim, 2003, p. 100-102). Em vez de se calcular a correlação diretamente sobre os dados originais, a correlação é calculada sobre os dados ordenados por postos (*rank*). Trata-se, portanto, de uma transformação dos dados, que resulta em uma

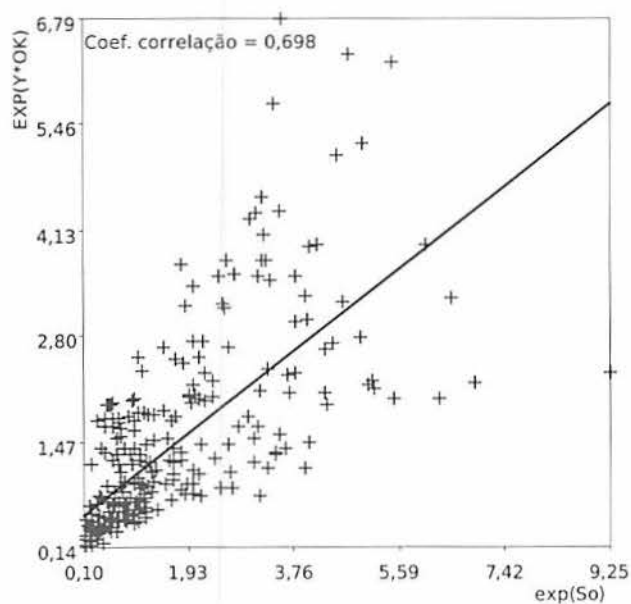


Fig. A.9 Diagrama de dispersão entre EXP(Y*OK) e exp(So), com correlação igual a 0,698

distribuição uniforme. O coeficiente de correlação de Spearman pode ser calculado como segue (Landim, 2003, p. 102):

$$r_S = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n^3 - n}$$

em que $d_i = x'_i - y'_i$ é a diferença para o i -ésimo para dados ordenados.

Assim, para os dados apresentados na Fig. A.8, inicialmente é feita a classificação por postos, que resulta em uma distribuição uniforme (Fig. A.10).

O coeficiente de correlação de Spearman pode ser calculado para os dados transformados (Fig. A.11).

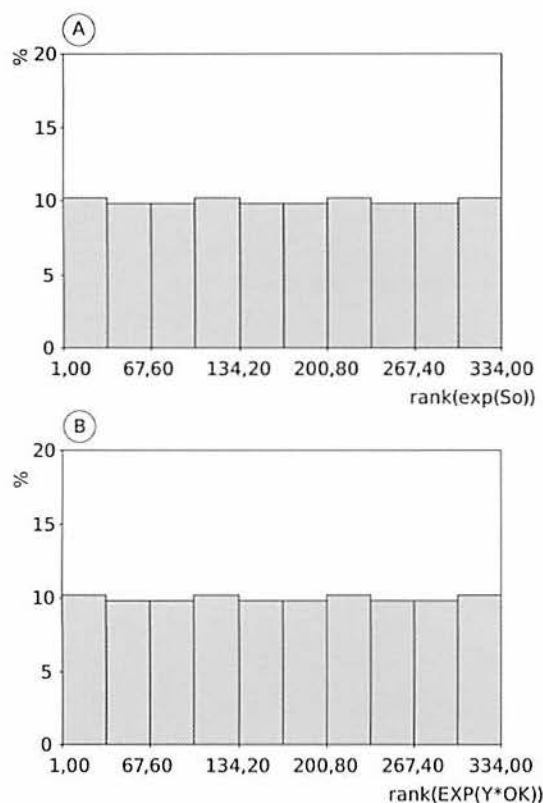


Fig. A.10 Histogramas dos dados transformados: A) variável exp(So) e B) variável EXP(Y*OK)

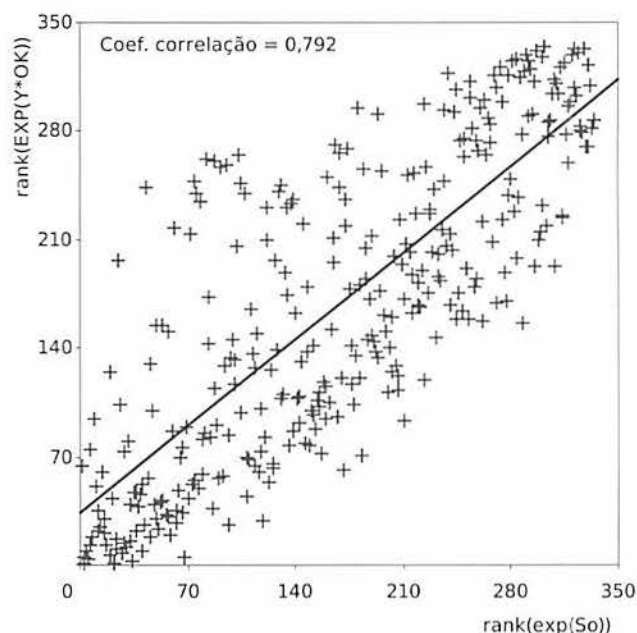


Fig. A.11 Diagrama de dispersão entre as variáveis transformadas EXP(Y*OK) e exp(So) apresentando coeficiente de correlação de Spearman igual a 0,792

O diagrama de dispersão (Fig. A.11) mostra um espalhamento melhor dos valores nos dois eixos e, conseqüentemente, o coeficiente de correlação de Spearman resulta em um valor maior (0,792) que aquele determinado pela fórmula de Pearson (0,698). A diferença entre o coeficiente de Pearson e o coeficiente de Spearman reflete tanto uma relação não linear como a existência de valores extremos (Landim, 2003, p. 101). Na realidade, após ordenação dos dados, pode-se usar tanto a fórmula de Spearman como a de Pearson para o cálculo da correlação, pois as duas são equivalentes, conforme demonstraram Fonseca, Martins e Toledo (1991, p. 45-47).

A.4 DISTRIBUIÇÕES TEÓRICAS DE PROBABILIDADES

As observações podem ser analisadas em termos da sua distribuição de frequência, sejam elas simples ou acumulativas. Dessas observações pode-se calcular as estatísticas que descrevem quantitativamente essas distribuições de frequências. Além disso, seria interessante verificar o modelo de distribuição teórica de probabilidade ao qual se ajusta a distribuição de frequência observada.

Nesse sentido, existem distribuições teóricas de probabilidades que representam todos os valores de uma variável e, assim, suas propriedades são bem conhecidas (Landim, 2003, p. 41). Existem muitos modelos de distribuição de probabilidades, mas na prática os mais utilizados em Geologia, segundo Landim (2003, p. 41-45), são: distribuições binomial e de Poisson, para amostras constituídas por dados discretos, e distribuições normal e lognormal, para dados contínuos.

A.4.1 Distribuição binomial

Por meio da distribuição binomial é possível calcular a probabilidade P_x de x eventos em uma amostra de tamanho n , sabendo-se que a probabilidade de acontecimento do evento é p e a de não acontecimento, q , conforme segue (Landim, 2003, p. 42):

$$P_x = \frac{n!}{x!(n-x)!} p^{n-x} q^x, \text{ em que } 0 < x < \infty$$

Os momentos da distribuição são (Landim, 2003, p. 42):

$$\text{média: } \mu'_1 = u = np$$

$$\text{variância: } u'_2 = \sigma^2 = npq$$

$$\text{assimetria} = a_3 = \frac{q-p}{\sqrt{npq}}$$

$$\text{curtose} = a_4 = 3 + \frac{1-6pq}{npq}$$

A.4.2 Distribuição de Poisson

A distribuição binomial aproxima-se da distribuição de Poisson quando a probabilidade de acontecimento p é pequena e o tamanho n da amostra, grande (Landim, 2003, p. 42-43). Segundo esse autor, considerando $np = m$, a probabilidade de x eventos em uma amostra de tamanho n é P_x :

$$P_x = e^{-m} \frac{m^x}{x!}$$

Os momentos da distribuição de Poisson são (Landim, 2003, p. 42-43):

$$\text{média: } \mu'_1 = u = m = np$$

$$\text{variância: } u'_2 = \sigma^2 = m = np$$

$$\text{assimetria} = a_3 = \frac{1}{m} = \frac{1}{np}$$

$$\text{curtose} = a_4 = 3 + \frac{1}{m} = 3 + \frac{1}{np}$$

A.4.3 Distribuição normal

O modelo de distribuição de Gauss ou normal é o mais comumente empregado, pois sob sua forma ocorrem diversas variáveis encontradas na natureza. Na verdade, a distribuição normal é a distribuição teórica de probabilidades mais importante em Estatística (Dixon; Massey, 1957, p. 48). Em Geologia, por exemplo, os óxidos constituintes principais de uma rocha distribuem-se, geralmente, sob a forma da distribuição normal. A função densidade de probabilidade da distribuição normal é descrita como (Dixon; Massey, 1957, p. 48):

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-1/2[(x-\mu)/\sigma]^2}$$

em que μ e σ são a média e o desvio padrão. A distribuição normal teórica é definida com média $\mu = 0$ e desvio padrão $\sigma = 1$. Para variáveis que não apresentam essas estatísticas, calcula-se a variável normal reduzida:

$$z_i = \frac{x_i - \mu}{\sigma}$$

a qual passa a ter média zero e desvio padrão 1, e a função densidade de probabilidade é escrita como (Landim, 2003, p. 43):

$$f(z) = \frac{1}{\sqrt{2\pi}} e^{-(z^2/2)}$$

A distribuição normal com média zero e desvio padrão igual a 1 apresenta as seguintes áreas sob a curva da função densidade de probabilidade (Landim, 2003, p. 43):

$$P(\mu - \sigma < x < \mu + \sigma) = 0,6827$$

$$P(\mu - 2\sigma < x < \mu + 2\sigma) = 0,9145$$

$$P(\mu - 3\sigma < x < \mu + 3\sigma) = 0,9937$$

Os momentos da distribuição normal com base no momento central de grau n μ_n podem ser calculados como (Landim, 2003, p. 43-44):

$$\mu_n = 0 \quad \text{para } n \text{ ímpar}$$
$$\mu_n = \frac{n! \sigma^n}{2^{n/2} (\frac{n}{2}!)^2} \quad \text{para } n \text{ par}$$

Assim, os quatro momentos são:

$$\mu_1 = 0, \quad \mu_2 = \sigma^2, \quad \mu_3 = 0, \quad \mu_4 = 3\sigma^4$$

A.4.4 Distribuição lognormal

Em Geologia, algumas variáveis podem se apresentar com uma grande quantidade de valores baixos e uns poucos valores altos. Por exemplo, variáveis como teores de ouro, prata, platina, estanho, tungstênio etc., ou seja, teores de metais raros, apresentam essa característica.

Essas variáveis seguem uma distribuição lognormal, e, por definição, os logaritmos das observações seguem uma distribuição normal (Koch; Link, 1971, p. 213). A função densidade de probabilidade da variável lognormal é descrita como (Koch; Link, 1971, p. 215):

$$f(x) = \frac{1}{x\beta\sqrt{2\pi}} e^{-1/2[(\log x - \alpha)/\beta]^2}$$

em que α e β^2 são a média e a variância dos logaritmos de x .

A distribuição lognormal sempre apresenta assimetria positiva, e a quantidade de assimetria depende da variância β^2 (Koch; Link, 1971, p. 215).

A.5 DERIVADAS

A derivada de uma função f no ponto x_0 é a inclinação da linha tangente à função f no ponto $(x_0, f(x_0))$ (Grossman, 1981, p. 87-88), ou seja, mede-se a taxa na qual a função f muda com a variação de x . A reta tangente é a linha passando por $(x_0, f(x_0))$ com inclinação $f'(x_0)$ (Grossman, 1981, p. 94). Por exemplo, considerar a função quadrática:

$$y = x^2 - x - 6$$

que tem como derivada:

$$\frac{dy}{dx} = 2x - 1$$

Para o ponto $x_0 = 2$, a derivada, ou seja, a inclinação da reta tangente que passa pelo ponto $x_0 = 2$, é igual a 3. Assim, a equação da reta tangente será:

$$y' = -10 + 3x$$

A Fig. A.12 ilustra a função quadrática e a sua reta tangente no ponto $(2; -4)$.

A.6 INTEGRAL

Enquanto a derivada trata da taxa de variação da função, a integral permite calcular comprimentos, áreas e volumes. A integral definida é representada como:

$$\int_a^b f(x) dx$$

em que a e b são os limites da integração e dx é a variável de integração. A integral de $f(x)$ no intervalo $[a, b]$ é definida como (Grossman, 1981, p. 264):

$$\int_a^b f(x) dx = \lim_{n \rightarrow \infty} \sum_{i=1}^n f(x_i^*) \frac{b-a}{n}$$

O intervalo $[a, b]$ é dividido em n segmentos, de tal forma que, à medida que n aproxima-se do infinito, a somatória resulta na integral procurada. Por exemplo, considerar a função $\sqrt{x}$,

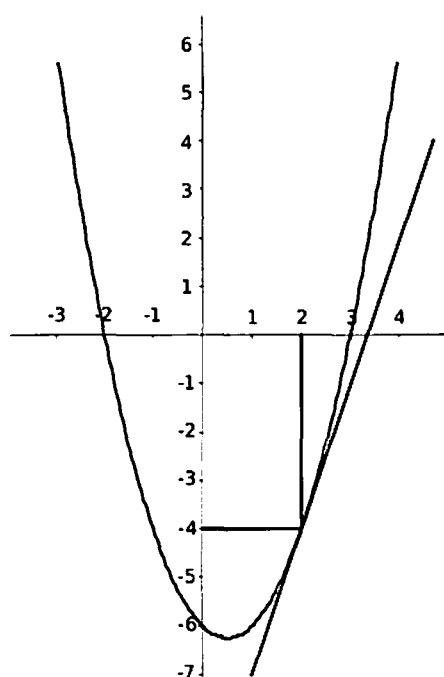


Fig. A.12 Gráfico da função quadrática e a reta tangente passando pelo ponto $(2; -4)$

para a qual se deseja calcular a área no intervalo $[0,1]$. Se dividirmos o intervalo $[0,1]$ em 5 e 10 segmentos (Fig. A.13), as áreas obtidas são 0,75 e 0,71, respectivamente.

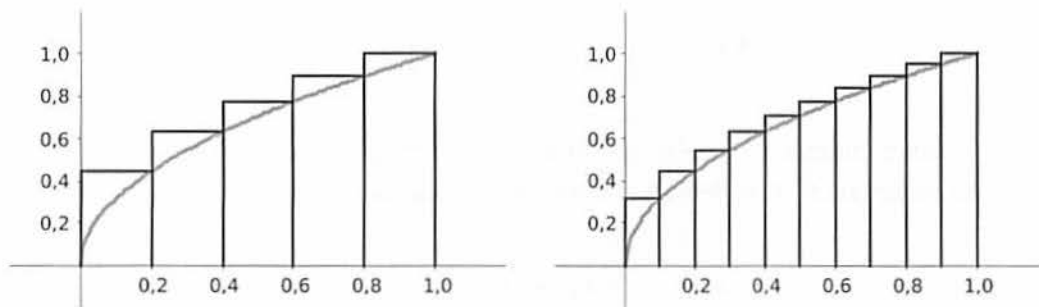


Fig. A.13 Gráficos da função $\sqrt{x}$ mostrando aproximações com 5 e 10 segmentos para o cálculo da integral definida no intervalo $[0,1]$

Considerando que a integral definida da função $\sqrt{x}$ no intervalo $[0,1]$ é igual a:

$$\int_0^1 \sqrt{x} dx = \frac{2}{3}$$

os resultados obtidos são insuficientes. Assim, aumentando-se o número de segmentos, a área aproxima-se do valor teórico (Fig. A.14).

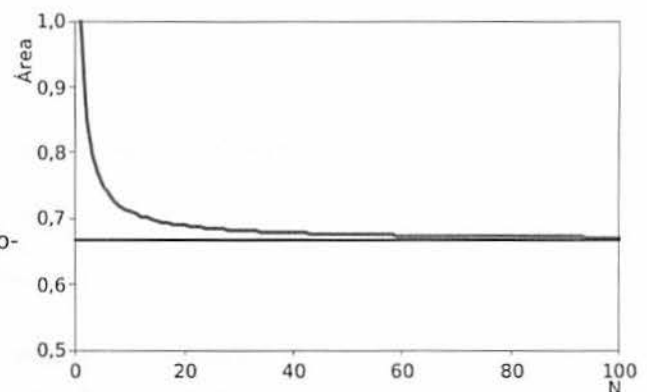


Fig. A.14 Áreas resultantes sob a função $\sqrt{x}$ no intervalo $[0,1]$ com N variando de 1 a 100 (a reta horizontal mostra o valor teórico)

A.7 MATRIZES

Uma matriz $m \times n$ é um arranjo retangular de números com m linhas e n colunas. A matriz é quadrada quando o número de linhas é igual ao número de colunas. São usadas letras maiúsculas para as matrizes e os elementos da matriz são identificados por letras minúsculas, com os índices correspondentes às linhas e colunas.

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}$$

Vetores são casos especiais de matrizes, nas quais, quando a matriz apresenta uma única coluna, tem-se o *vetor coluna*:

$$X = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

E, quando apresenta uma única linha, tem-se o *vetor linha*:

$$Y = \begin{bmatrix} y_1 & y_2 & y_3 \end{bmatrix}$$



A transposta de uma matriz A é obtida trocando suas linhas e colunas, ou seja, as linhas tornam-se colunas e vice-versa, resultando em A^t .

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix} \Rightarrow A^t = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$$

Uma matriz quadrada $n \times n$ é denominada identidade quando os elementos da diagonal principal são iguais a 1 e os demais elementos, iguais a zero (Grossman, 1982, p. 358):

$$I_n = (b_{ij}) \quad \text{em que} \quad b_{ij} = \begin{cases} 1 & \text{se } i = j \\ 0 & \text{se } i \neq j \end{cases}$$

Ver o exemplo de uma matriz identidade 3×3 , a seguir:

$$I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Uma matriz quadrada é dita ortogonal se $AA^t = I$. Vide a seguir:

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \times \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

A.7.1 Adição de matrizes

Se A e B são matrizes do mesmo tamanho, ou seja, com número de linhas e colunas iguais, pode-se somar:

$$A + B = (a_{ij} + b_{ij})$$

A.7.2 Multiplicação por um escalar

A matriz A pode ser multiplicada por um escalar α , resultando em:

$$\alpha A = \begin{bmatrix} \alpha a_{11} & \alpha a_{12} & \alpha a_{13} \\ \alpha a_{21} & \alpha a_{22} & \alpha a_{23} \\ \alpha a_{31} & \alpha a_{32} & \alpha a_{33} \end{bmatrix}$$

A.7.3 Multiplicação de matrizes

As matrizes A e B podem ser multiplicadas entre si quando o número de colunas de A for igual ao número de linhas de B . Seja a matriz A igual a $m \times n$ e a matriz B , igual a $n \times p$, então o resultado será uma matriz C correspondente a $m \times p$, na qual o ij -ésimo elemento será o produto escalar da i -ésima linha de A pela j -ésima coluna de B (Grossman, 1982, p. 336):

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}$$

A.7.4 Determinantes

O determinante de uma matriz quadrada é um escalar ou função associada a ela. O determinante de uma matriz 2×2 pode ser calculado como (Grossman, 1982, p. 371):

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \Rightarrow \det A = |A| = a_{11}a_{22} - a_{12}a_{21} \quad (\text{A.1})$$

Para calcular o determinante de uma matriz 3×3 , utiliza-se a definição da Eq. A.1 para uma matriz 2×2 (Grossman, 1982, p. 371):

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \Rightarrow$$

$$\det A = a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \quad (\text{A.2})$$

O valor do determinante é obtido aplicando-se a Eq. A.1 para cada uma das matrizes 2×2 . Pode-se observar na Eq. A.2: que a_{11} multiplica o determinante da matriz $\begin{bmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{bmatrix}$, a qual foi obtida eliminando-se a 1ª linha e a 1ª coluna de A ; que a_{12} multiplica o determinante da matriz $\begin{bmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{bmatrix}$, que foi obtida deletando-se a 1ª linha e a 2ª coluna de A ; que a_{13} multiplica o determinante da matriz $\begin{bmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix}$, que resultou da matriz A com a eliminação da 1ª linha e 3ª coluna; essas matrizes podem ser denominadas M_{11} , M_{12} e M_{13} , respectivamente, e os seus determinantes são $A_{11} = \det M_{11}$, $\det A_{12} = -\det M_{12}$ e $A_{13} = \det M_{13}$ (Grossman, 1982, p. 373). Assim, segundo esse autor, a Eq. A.2 pode ser reescrita como:

$$\det A = a_{11}A_{11} + a_{12}A_{12} + a_{13}A_{13}$$

Se A é uma matriz $n \times n$, o ij -ésimo cofator (A_{ij}) é dado por (Grossman, 1982, p. 373):

$$A_{ij} = (-1)^{i+j} |M_{ij}|$$

Dessa forma, o determinante de uma matriz $n \times n$ é dado por (Grossman, 1982, p. 374):

$$\det A = a_{11}A_{11} + a_{12}A_{12} + a_{13}A_{13} + \cdots + a_{1n}A_{1n}$$

$$= \sum_{k=1}^n a_{1k}A_{1k} \quad (\text{A.3})$$

Para ilustrar o cálculo do determinante de uma matriz 4×4 , considerar a matriz seguinte:

$$A = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

Aplicando-se a Eq. A.3, tem-se:

$$\det A = a_{11}A_{11} + a_{12}A_{12} + a_{13}A_{13} + a_{14}A_{14}$$



ou

$$\det A = 2 \begin{vmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 1 \end{vmatrix} + 1 \begin{vmatrix} -1 & -1 & 0 \\ 0 & 2 & -1 \\ 0 & -1 & 1 \end{vmatrix} + 0 \begin{vmatrix} -1 & 2 & 0 \\ 0 & -1 & -1 \\ 0 & 0 & 1 \end{vmatrix} - 0 \begin{vmatrix} -1 & 2 & -1 \\ 0 & -1 & 2 \\ 0 & 0 & -1 \end{vmatrix}$$

Observar que há dois determinantes do lado direito multiplicados por zero. Desse modo, pode-se desenvolver cada um dos dois determinantes aplicando-se a Eq. A.2:

$$2 \begin{vmatrix} 2 & -1 \\ -1 & 1 \end{vmatrix} + 1 \begin{vmatrix} -1 & -1 \\ 0 & 1 \end{vmatrix} + 0 \begin{vmatrix} -1 & 2 \\ 0 & -1 \end{vmatrix} = 1$$

$$-1 \begin{vmatrix} 2 & -1 \\ -1 & 1 \end{vmatrix} + 1 \begin{vmatrix} 0 & -1 \\ 0 & 1 \end{vmatrix} + 0 \begin{vmatrix} 0 & 2 \\ 0 & -1 \end{vmatrix} = -1$$

Substituindo-se os resultados parciais, obtém-se:

$$\det A = 1$$

Determinantes de matrizes de maiores dimensões são obtidos aplicando-se a Eq. A.3, que é calculada reduzindo-se às dimensões de 2×2 para aplicação da Eq. A.2.

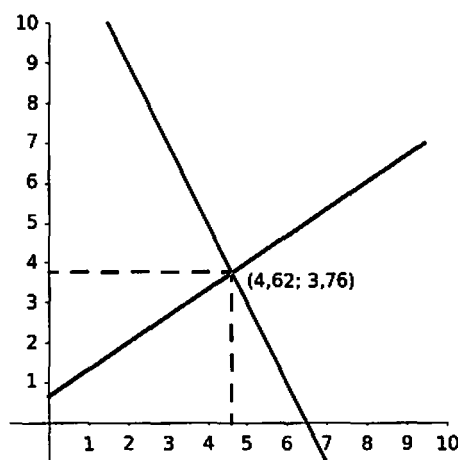


Fig. A.15 Equações das retas $y = 13 - 2x$ (linha escura) e $y = 0,67 + 0,67x$ (linha clara) e o ponto de intersecção (4,62; 3,76)

A.8 SISTEMAS DE EQUAÇÕES LINEARES

Dadas as equações de duas retas $y = 13 - 2x$ e $y = 0,67 + 0,67x$, deseja-se conhecer o ponto $(x; y)$ no qual elas se interceptam. A solução geométrica é simples e determina o ponto (4,62; 3,76) como ponto de intersecção (Fig. A.15).

As equações das retas podem ser reescritas, porém, formando um sistema de equações lineares:

$$\begin{aligned} 2x + y &= 13 \\ -0,67x + y &= 0,67 \end{aligned} \quad (\text{A.4})$$

Existem diversas formas de resolver sistemas de equações lineares. Nesta revisão, os métodos de Cramer e o da triangularização de Gauss serão apresentados.

A.8.1 Método de Cramer

Dado o sistema de n equações com n incógnitas:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n &= b_2 \\ \vdots & \\ a_{n1}x_1 + a_{n2}x_2 + \dots + a_{nn}x_n &= b_n \end{aligned}$$

essas equações podem ser reescritas como matriz e vetores:

$$Ax = b$$

em que

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \quad \text{e} \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

A solução desse sistema é:

$$x = A^{-1}b$$

desde que $\det A \neq 0$.

Pelo método de Cramer, segundo Grossman (1982, p. 394), n novas matrizes são definidas:

$$A_1 = \begin{bmatrix} b_1 & a_{12} & \cdots & a_{1n} \\ b_2 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ b_n & a_{n2} & \cdots & a_{nn} \end{bmatrix}, \quad A_2 = \begin{bmatrix} a_{11} & b_1 & \cdots & a_{1n} \\ a_{21} & b_2 & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & b_n & \cdots & a_{nn} \end{bmatrix}, \quad \dots, \quad A_n = \begin{bmatrix} a_{11} & a_{12} & \cdots & b_1 \\ a_{21} & a_{22} & \cdots & b_2 \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & b_n \end{bmatrix}$$

Ou seja, a matriz A_i é obtida substituindo-se a i -ésima coluna de A pelo vetor b . Fazendo-se $D = \det A$; $D_1 = \det A_1$; $D_2 = \det A_2$;; $D_n = \det A_n$, a solução do sistema é dada por (Grossman, 1982, p. 394):

$$x_1 = \frac{D_1}{D}; \quad x_2 = \frac{D_2}{D}; \quad \dots; \quad x_n = \frac{D_n}{D}$$

A solução do sistema da Eq. A.4 é:

$$x = \frac{\begin{vmatrix} 13 & 1 \\ 0,67 & 1 \end{vmatrix}}{\begin{vmatrix} 2 & 1 \\ -0,67 & 1 \end{vmatrix}} = 4,62 \quad \text{e} \quad y = \frac{\begin{vmatrix} 2 & 13 \\ -0,67 & 0,67 \end{vmatrix}}{\begin{vmatrix} 2 & 1 \\ -0,67 & 1 \end{vmatrix}} = 3,76$$

Seja o sistema de equações 4×4 a seguir (fonte desconhecida):

$$A = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \quad \text{e} \quad b = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$



Conforme foi descrito por Grossman (1982, p. 394), formam-se quatro novas matrizes:

$$A_1 = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 0 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 1 & 0 & -1 & 1 \end{bmatrix} \quad A_2 = \begin{bmatrix} 2 & 0 & 0 & 0 \\ -1 & 0 & -1 & 0 \\ 0 & 0 & 2 & -1 \\ 0 & 1 & -1 & 1 \end{bmatrix}$$

$$A_3 = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & 0 & 0 \\ 0 & -1 & 0 & -1 \\ 0 & 0 & 1 & 1 \end{bmatrix} \quad A_4 = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & 0 \\ 0 & 0 & -1 & 1 \end{bmatrix}$$

Calculando-se os determinantes, conforme já descrito, têm-se: $D = 1$; $D_1 = 1$; $D_2 = 2$; $D_3 = 3$; $D_4 = 4$. Portanto, $x_1 = 1$; $x_2 = 2$; $x_3 = 3$ e $x_4 = 4$.

A.8.2 Método de triangularização de Gauss

Dado um sistema de n equações com n incógnitas:

$$\begin{array}{cccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ a_{21}x_1 & + & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ \vdots & & \vdots & & & & \vdots & & \vdots \\ a_{n1}x_1 & + & a_{n2}x_2 & + & \cdots & + & a_{nn}x_n & = & b_n \end{array}$$

O método de triangularização ou de eliminação de Gauss é aplicado por meio das operações: multiplicação por uma constante; subtração de uma equação de outra; substituição da segunda pelo resultado da subtração; e reordenação de equações (Dorn; McCracken, 1978, p. 210). Esse método resulta em um sistema triangular:

$$\begin{array}{cccccc} a_{11}x_1 & + & a_{12}x_2 & + & \cdots & + & a_{1n}x_n & = & b_1 \\ & & a_{22}x_2 & + & \cdots & + & a_{2n}x_n & = & b_2 \\ & & & & & & \vdots & & \vdots \\ & & & & & & a_{nn}x_n & = & b_n \end{array}$$

Observar que nesse sistema triangular a n -ésima equação é resolvida e o valor da incógnita x_n pode ser substituído na $(n - 1)$ -ésima equação, na qual se encontra o valor da incógnita x_{n-1} , e assim por diante. Essa operação é denominada substituição reversa.

Para ilustrar o método de eliminação de Gauss, considerar o mesmo exemplo utilizado anteriormente:

$$A = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix}, \quad x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} \quad \text{e} \quad b = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

A primeira providência é anexar o vetor b à matriz A , resultando em uma matriz 4×5 :

$$\begin{bmatrix} 2 & -1 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & -1 & 1 & 1 \end{bmatrix}$$

O objetivo é zerar os elementos da diagonal principal para formar um sistema triangular. Desse modo, abaixo do elemento da diagonal principal a_{11} temos o valor -1 e os demais, iguais a zero. Portanto, para o caso dessa coluna, basta eliminar o valor -1 . Para fazer isso, é encontrado o multiplicador, que, multiplicado por 2 e somado ao elemento igual a -1 , resulta em zero. Esse valor é $-\left(\frac{-1}{2}\right)$, ou seja, o elemento abaixo da diagonal dividido pelo elemento da diagonal com o sinal trocado. Assim, multiplicar a 1ª linha por $\frac{1}{2}$ e somar o resultado com a 2ª linha:

$$\begin{array}{ccccc} 0.5 \times & 2 & -1 & 0 & 0 & 0 & 1 & -0,5 & 0 & 0 & 0 & 2 & -1 & 0 & 0 & 0 \\ & -1 & 2 & -1 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & 1,5 & -1 & 0 & 0 \\ & 0 & -1 & 2 & -1 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & -1 & 2 & -1 & 0 \\ & 0 & 0 & -1 & 1 & 1 & 0 & 0 & -1 & 1 & 1 & 0 & 0 & -1 & 1 & 1 \end{array} \Rightarrow$$

Como os demais elementos abaixo da diagonal a_{11} são iguais a zero, pode-se mover para o elemento seguinte sobre a diagonal principal, ou seja, a_{22} . O multiplicador é $-\left(\frac{-1}{1,5}\right)$, que irá multiplicar todos os elementos da 2ª linha, os quais, depois, serão somados aos elementos correspondentes da 3ª linha:

$$\begin{array}{ccccc} & 2 & -1 & 0 & 0 & 0 & 2 & -1 & 0 & 0 & 0 & 2 & -1 & 0 & 0 & 0 \\ (1/1,5) \times & 1,5 & -1 & 0 & 0 & 0 & 1 & -1/1,5 & 0 & 0 & 0 & 0 & 1,5 & -1 & 0 & 0 \\ & -1 & 2 & -1 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 & 0 & 1,33 & -1 & 0 \\ & 0 & -1 & 1 & 1 & 1 & 0 & -1 & 1 & 1 & 0 & 0 & -1 & 1 & 1 & 1 \end{array} \Rightarrow$$

Em seguida, o elemento abaixo do elemento a_{33} deve ser eliminado. Para isso, o multiplicador $-\left(\frac{-1}{1,33}\right)$ irá multiplicar a 3ª linha, e os resultados serão somados aos elementos da 4ª linha:

$$\begin{array}{ccccc} 2 & -1 & 0 & 0 & 0 & 2 & -1 & 0 & 0 & 0 & 2 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1,5 & -1 & 0 & 0 & 0 & 1,5 & -1 & 0 & 0 & 0 & 1,5 & -1 & 0 & 0 & 0 \\ 0 & (1/1,33) \times & 1,33 & -1 & 0 & 1 & -0,75 & 0 & 0 & 0 & 1,33 & -1 & 0 & 0 & 0 & 0 \\ & -1 & 1 & 1 & 1 & -1 & 1 & 1 & 0 & 0 & 0 & 0,25 & 1 & 0 & 0 & 0 \end{array} \Rightarrow$$

Todos os elementos abaixo da diagonal principal foram eliminados, e o sistema triangular pode ser escrito como:

$$\begin{array}{rcl} 2x_1 & - & x_2 & & & = & 0 \\ & & 1,5x_2 & - & x_3 & & = & 0 \\ & & & & 1,33x_3 & - & x_4 & = & 0 \\ & & & & & & 0,25x_4 & = & 1 \end{array}$$



A solução por substituição reversa resulta em:

$$x_4 = \frac{1}{0,25} = 4, \quad x_3 = \frac{x_4}{1,33} = 3, \quad x_2 = \frac{x_3}{1,5} = 2, \quad x_1 = \frac{x_2}{2} = 1$$

A.9 Software

A aplicação de métodos geoestatísticos envolvendo um grande número de dados necessita de softwares específicos. Existem à disposição desde pacotes comerciais extremamente sofisticados e caros até programas de livre acesso e com ótimo desempenho, como o GSLib® e o SGeMS®. A escolha desses dois teve como critério serem programas de domínio público, com códigos abertos e facilmente obtidos pela internet. Além deles, há o sistema Geovisual, que foi desenvolvido com a finalidade de dar suporte às aulas práticas de Geoestatística no Instituto de Geociências da Universidade de São Paulo (USP).

A.9.1 GSLib

O GSLib (Geostatistical Software Library) é uma biblioteca de programas desenvolvidos na Universidade de Stanford, nos Estados Unidos, para o ambiente DOS, sob a direção de A. G. Journel. A Oxford University Press publicou, em 1992, com uma segunda edição em 1998, um guia para o usuário, acompanhado de disquetes ou CD com as diversas sub-rotinas em Fortran, chamado *GSLib geoestatistical software library and user's guide*, tendo como autores C.V. Deutsch e A.G. Journel (Deutsch; Journel, 1998).

A primeira versão do GSLib 2.0 está disponível gratuitamente, com os programas fontes em Fortran77. A última versão, 2.90, escrita em Fortran90, também é gratuita apenas para os programas executáveis. Há um suplemento comercial, denominado WinGslib 1.1.3, para a interface Windows (<http://www.gslib.com>).

Para trabalhar com o GSLib é necessário ter à disposição arquivos executáveis (*.exe) e os correspondentes arquivos de parâmetros (*.par).

No arquivo executável estão os comandos responsáveis pelos cálculos e algoritmos a que se destina o programa.

No arquivo de parâmetros são definidos a localização e o nome do arquivo de dados, os parâmetros que o programa executável pede e o nome do arquivo com os resultados.

O arquivo de parâmetros pode ser escrito por qualquer editor de textos, sendo recomendado utilizar a extensão *.par, por questão de organização, precedido pelo nome do programa. Por exemplo, se para construir um mapa de localização das amostras é necessário o programa locmap.exe, o respectivo arquivo de parâmetros deve ser locmap.par.

Quando executado, o programa solicita o nome do arquivo de parâmetros com a sua extensão e, se lido corretamente, gera os resultados, gravando-os num arquivo de saída em PostScript, com nome de acordo com o especificado no arquivo de parâmetros. Para os arquivos em PostScript, são necessários os programas Ghostscript e GSview (<http://www.cs.wisc.edu/~ghost>).

Ghostscript é o interpretador para a linguagem PostScript®, e o GSview é a interface gráfica para Ghostscript em plataforma Windows.

Os dados no GSLib são escritos em arquivos em ASCII num formato simplificado do programa GEO-EAS, desenvolvido em 1988 pela Agência de Proteção Ambiental dos Estados Unidos (United States Environmental Protection Agency – Usepa, 1991), e é precursor e modelo de quase todos os softwares de análise geoestatística.

A.9.2 SGeMS

O SGeMS (Stanford Geostatistical Earth Modeling Software) é um pacote de programas que sucedeu o GSLib e também foi desenvolvido na Universidade de Stanford. O código é em C++ e roda interativamente sob Windows. O site oficial do software é <<http://sgems.sourceforge.net/>>, no qual se encontra um manual. Também podem ser citados tutoriais de autoria de Geoffrey Bohling, do Kansas Geological Survey/EUA (Bohling, 2007), e de Pedro Correia, do Numist/Instituto Superior Técnico de Lisboa (Correia, 2010). A editora Cambridge University Press publicou, em 2009, um guia para o usuário acompanhado de CD sob o título *Applied geostatistics with SGeMS: a user's guide*, tendo como autores Remy, Boucher e Wu (Remy; Boucher; Wu, 2009).

A interface do S-GeMS é composta por três seções principais: o painel de algoritmos; o painel de visualização; e o painel de comandos.

A.9.3 Geovisual

O sistema Geovisual foi desenvolvido para dar suporte às aulas práticas de disciplinas envolvendo o tópico Geoestatística, tanto para a graduação como para a pós-graduação, no Instituto de Geociências da Universidade de São Paulo (USP). A licença acadêmica é gratuita para aplicações acadêmicas e está, mediante solicitação, disponível com o autor, cujo e-mail é: <jkyamamo@usp.br>.

Todo o sistema é compatível com a plataforma Windows 98 ou superior e está totalmente escrito em Delphi 5 ou superior. A maioria dos cálculos e gráficos resultantes apresentados neste livro foi feita com o uso desse sistema ou programas dele derivados. Entretanto, nem todas as aplicações estão disponíveis ao usuário, pois ainda se encontram em testes ou em fase de adaptação para o ambiente Geovisual.



Anexo

B

ARQUIVOS DE DADOS

Arquivo 1

Amostra=amostraAleatoria.txt - amostragem aleatoria simples: Bell.txt

3			5.50	14.50	18.177	17.50	35.50	18.633
X			43.50	14.50	15.447	6.50	36.50	10.203
Y			31.50	15.50	8.948	27.50	36.50	20.833
Zgauss			44.50	15.50	15.608	32.50	36.50	17.994
12.50	0.50	8.738	47.50	15.50	19.116	41.50	36.50	16.965
23.50	0.50	11.160	2.50	16.50	17.935	4.50	37.50	9.572
48.50	0.50	19.399	33.50	16.50	10.676	7.50	37.50	9.060
8.50	1.50	9.697	22.50	17.50	6.793	30.50	37.50	20.438
23.50	2.50	11.953	31.50	18.50	12.910	8.50	38.50	9.944
47.50	2.50	18.864	29.50	21.50	14.980	32.50	38.50	19.060
10.50	3.50	8.973	30.50	21.50	15.350	25.50	39.50	18.798
17.50	3.50	6.499	28.50	22.50	14.696	1.50	40.50	15.452
44.50	3.50	13.656	17.50	23.50	13.170	6.50	40.50	9.343
14.50	4.50	3.137	10.50	25.50	15.138	25.50	40.50	18.134
19.50	4.50	6.561	29.50	25.50	15.036	42.50	41.50	12.556
42.50	4.50	14.841	29.50	26.50	14.677	0.50	42.50	17.751
2.50	5.50	21.668	37.50	26.50	16.141	38.50	42.50	12.291
41.50	6.50	14.743	8.50	27.50	18.828	24.50	43.50	14.738
6.50	7.50	17.048	37.50	27.50	20.753	49.50	43.50	18.282
20.50	7.50	13.165	24.50	28.50	18.834	8.50	44.50	12.184
42.50	7.50	15.479	28.50	28.50	15.927	26.50	44.50	17.043
30.50	9.50	15.341	49.50	28.50	13.367	31.50	44.50	23.920
48.50	9.50	23.290	7.50	29.50	19.249	45.50	45.50	20.300
1.50	10.50	29.061	28.50	29.50	16.634	6.50	46.50	17.220
12.50	10.50	11.066	49.50	29.50	15.498	18.50	46.50	14.630
46.50	10.50	22.317	0.50	30.50	8.613	30.50	46.50	24.501
1.50	11.50	26.852	5.50	30.50	15.904	11.50	47.50	12.367
39.50	11.50	12.303	3.50	31.50	14.213	15.50	47.50	13.942
16.50	12.50	16.677	33.50	31.50	20.735	5.50	48.50	20.191
24.50	12.50	15.031	35.50	31.50	25.660	4.50	49.50	21.283
12.50	13.50	12.741	48.50	32.50	21.094	31.50	49.50	27.551
18.50	13.50	18.058	1.50	33.50	14.375	32.50	49.50	28.740
19.50	13.50	17.102	19.50	33.50	17.480	33.50	49.50	28.244
20.50	13.50	16.089	48.50	34.50	20.842	36.50	49.50	23.388
34.50	13.50	12.641	5.50	35.50	12.823			

Arquivo 2

Amostra=testeEstratificada.txt - amostragem aleatoria estratificada: Bell.txt

3			16.50	6.50	6.708	32.50	30.50	18.128
X			16.50	12.50	16.677	32.50	38.50	19.060
Y			19.50	17.50	16.682	30.50	42.50	24.439
Zgauss			18.50	20.50	16.257	31.50	45.50	24.379
4.50	0.50	11.392	19.50	27.50	13.113	36.50	3.50	11.909
0.50	6.50	25.411	16.50	32.50	19.275	38.50	6.50	9.354
1.50	13.50	25.825	18.50	39.50	18.166	39.50	12.50	11.884
1.50	16.50	18.015	17.50	41.50	18.172	39.50	15.50	6.765
4.50	24.50	11.487	15.50	45.50	14.006	37.50	22.50	7.152
2.50	25.50	9.297	21.50	3.50	8.655	39.50	26.50	16.034
2.50	32.50	13.373	23.50	6.50	16.511	35.50	34.50	21.760
0.50	38.50	13.889	20.50	12.50	16.572	38.50	35.50	20.770
2.50	41.50	15.964	20.50	18.50	13.844	38.50	42.50	12.291
0.50	45.50	17.658	23.50	22.50	14.598	35.50	46.50	21.557
8.50	0.50	10.149	21.50	28.50	15.969	41.50	4.50	14.527
7.50	7.50	15.124	21.50	31.50	18.079	40.50	6.50	13.444
6.50	13.50	18.455	24.50	37.50	19.790	43.50	12.50	16.686
9.50	15.50	10.595	20.50	44.50	16.696	44.50	18.50	14.925
6.50	24.50	10.185	20.50	45.50	15.678	42.50	23.50	13.465
8.50	29.50	20.167	28.50	1.50	16.591	42.50	25.50	14.673
8.50	31.50	20.851	28.50	8.50	17.602	44.50	30.50	19.022
6.50	39.50	8.021	25.50	14.50	9.133	41.50	37.50	13.631
6.50	41.50	10.627	27.50	18.50	7.786	44.50	40.50	10.212
9.50	46.50	15.128	27.50	23.50	15.983	44.50	47.50	22.103
13.50	4.50	2.957	25.50	27.50	18.739	48.50	2.50	19.608
10.50	7.50	11.680	28.50	34.50	19.128	48.50	7.50	18.393
11.50	13.50	10.652	27.50	36.50	20.833	45.50	10.50	21.417
10.50	19.50	7.805	27.50	40.50	23.158	46.50	17.50	16.705
10.50	23.50	9.011	29.50	46.50	22.637	46.50	24.50	14.156
12.50	27.50	18.709	33.50	0.50	16.497	46.50	26.50	13.874
12.50	32.50	22.577	33.50	8.50	14.285	48.50	32.50	21.094
10.50	39.50	12.752	34.50	11.50	13.097	47.50	35.50	18.686
11.50	43.50	8.038	33.50	15.50	11.246	45.50	44.50	17.470
14.50	47.50	13.342	31.50	23.50	14.873	47.50	45.50	21.262
16.50	3.50	6.963	32.50	28.50	16.577			

Arquivo 3

Amostra=bellSistemica.txt - amostragem sistemica: Bell.txt

3			47.50	2.50	18.864	12.50	12.50	11.440
X			2.50	7.50	25.065	17.50	12.50	16.902
Y			7.50	7.50	15.124	22.50	12.50	16.075
Zgauss			12.50	7.50	9.982	27.50	12.50	10.317
2.50	2.50	16.691	17.50	7.50	8.683	32.50	12.50	13.321
7.50	2.50	9.399	22.50	7.50	16.840	37.50	12.50	11.225
12.50	2.50	7.200	27.50	7.50	17.495	42.50	12.50	14.409
17.50	2.50	10.771	32.50	7.50	15.932	47.50	12.50	23.214
22.50	2.50	11.189	37.50	7.50	8.089	2.50	17.50	17.131
27.50	2.50	15.751	42.50	7.50	15.479	7.50	17.50	12.436
32.50	2.50	16.729	47.50	7.50	18.155	12.50	17.50	12.522
37.50	2.50	12.539	2.50	12.50	25.608	17.50	17.50	19.307
42.50	2.50	13.285	7.50	12.50	17.190	22.50	17.50	6.793

27.50	17.50	5.057	37.50	27.50	20.753	47.50	37.50	14.934
32.50	17.50	11.253	42.50	27.50	17.294	2.50	42.50	16.806
37.50	17.50	4.148	47.50	27.50	12.729	7.50	42.50	11.508
42.50	17.50	12.027	2.50	32.50	13.373	12.50	42.50	11.607
47.50	17.50	18.564	7.50	32.50	19.282	17.50	42.50	18.020
2.50	22.50	15.336	12.50	32.50	22.577	22.50	42.50	15.346
7.50	22.50	11.406	17.50	32.50	16.435	27.50	42.50	21.927
12.50	22.50	7.128	22.50	32.50	20.076	32.50	42.50	22.373
17.50	22.50	14.074	27.50	32.50	19.486	37.50	42.50	11.757
22.50	22.50	13.357	32.50	32.50	18.621	42.50	42.50	14.579
27.50	22.50	15.193	37.50	32.50	25.320	47.50	42.50	15.465
32.50	22.50	14.999	42.50	32.50	20.053	2.50	47.50	18.558
37.50	22.50	7.152	47.50	32.50	20.717	7.50	47.50	18.961
42.50	22.50	14.122	2.50	37.50	12.996	12.50	47.50	12.691
47.50	22.50	15.668	7.50	37.50	9.060	17.50	47.50	13.903
2.50	27.50	6.901	12.50	37.50	18.144	22.50	47.50	12.505
7.50	27.50	17.434	17.50	37.50	18.524	27.50	47.50	18.581
12.50	27.50	18.709	22.50	37.50	18.404	32.50	47.50	26.668
17.50	27.50	11.859	27.50	37.50	21.737	37.50	47.50	22.359
22.50	27.50	17.023	32.50	37.50	18.478	42.50	47.50	19.956
27.50	27.50	17.389	37.50	37.50	17.304	47.50	47.50	22.684
32.50	27.50	15.101	42.50	37.50	12.612			

Arquivo 4

Amostra=teste123.txt - amostragem agrupada: Bell.txt

3			22.50	10.50	18.036	39.50	42.50	12.888
X			24.50	18.50	8.139	41.50	47.50	21.513
Y			19.50	24.50	13.611	45.50	1.50	14.869
Zgauss			23.50	27.50	18.304	48.50	12.50	23.554
3.50	4.50	18.326	23.50	34.50	22.459	44.50	13.50	18.645
6.50	7.50	17.048	22.50	38.50	16.734	44.50	20.50	15.816
3.50	18.50	16.820	21.50	47.50	12.899	47.50	26.50	13.918
5.50	23.50	13.726	30.50	6.50	17.633	44.50	34.50	18.668
4.50	26.50	8.237	28.50	8.50	17.602	45.50	40.50	11.036
4.50	34.50	14.499	29.50	13.50	9.205	47.50	48.50	23.140
1.50	41.50	16.454	29.50	19.50	11.915	0.50	8.50	26.757
2.50	49.50	18.985	26.50	28.50	18.506	3.50	8.50	23.121
12.50	2.50	7.200	25.50	36.50	21.293	4.50	8.50	21.831
7.50	8.50	16.234	30.50	37.50	20.438	0.50	9.50	30.923
12.50	14.50	12.696	29.50	44.50	22.700	2.50	10.50	27.414
7.50	19.50	12.964	32.50	6.50	15.807	1.50	11.50	26.852
10.50	29.50	21.952	31.50	8.50	16.516	30.50	43.50	24.152
8.50	36.50	13.576	33.50	18.50	13.212	30.50	44.50	23.824
10.50	40.50	11.686	35.50	24.50	11.095	31.50	44.50	23.920
10.50	48.50	15.133	35.50	27.50	20.700	36.50	44.50	18.552
18.50	6.50	7.292	35.50	31.50	25.660	29.50	45.50	22.668
14.50	8.50	9.365	36.50	42.50	13.128	33.50	45.50	23.309
15.50	14.50	18.715	35.50	44.50	19.091	34.50	45.50	22.039
16.50	23.50	12.160	38.50	2.50	13.352	35.50	45.50	20.471
16.50	29.50	14.040	39.50	8.50	11.789	29.50	46.50	22.637
18.50	34.50	17.384	41.50	16.50	9.646	32.50	46.50	25.105
14.50	40.50	17.028	39.50	20.50	8.363	35.50	46.50	21.557
14.50	48.50	13.398	38.50	25.50	13.485	36.50	47.50	22.828
22.50	3.50	11.751	41.50	35.50	18.421	36.50	48.50	23.271



28.50	49.50	21.772	35.50	28.50	22.779	34.50	30.50	23.754
32.50	49.50	28.740	38.50	28.50	21.212	37.50	31.50	26.068
34.50	27.50	20.487	39.50	28.50	18.930	38.50	32.50	24.766
37.50	27.50	20.753	34.50	29.50	23.103	39.50	32.50	24.047
36.50	27.50	20.897	36.50	29.50	25.025	35.50	33.50	23.970
33.50	28.50	19.684	39.50	29.50	22.303			

Arquivo 5

Amostra=teste25.txt - amostragem aleatoria estratificada: Bell.txt

3			11.50	10.50	11.614	36.50	18.50	8.172
X			12.50	25.50	14.780	34.50	23.50	12.663
Y			15.50	30.50	16.625	39.50	32.50	24.047
Zgauss			10.50	47.50	14.635	32.50	44.50	23.409
5.50	9.50	21.018	21.50	5.50	12.015	42.50	6.50	15.682
5.50	16.50	15.013	23.50	17.50	4.866	42.50	13.50	13.264
5.50	24.50	11.309	23.50	26.50	17.374	44.50	27.50	16.164
1.50	31.50	10.809	26.50	35.50	21.819	47.50	30.50	19.085
6.50	42.50	12.839	21.50	42.50	15.885	42.50	45.50	17.877
13.50	8.50	9.604	36.50	0.50	12.178			

Arquivo 6

Amostra=teste36.txt - amostragem aleatoria estratificada: Bell.txt

3			16.50	33.50	20.001	33.50	5.50	14.799
X			11.50	41.50	11.117	38.50	16.50	1.939
Y			20.50	4.50	8.285	39.50	20.50	8.363
Zgauss			21.50	12.50	16.248	34.50	29.50	23.103
7.50	2.50	9.399	19.50	17.50	16.682	37.50	38.50	15.548
3.50	11.50	24.948	21.50	29.50	16.122	37.50	49.50	22.812
5.50	20.50	15.992	24.50	39.50	17.525	49.50	6.50	18.243
4.50	25.50	9.320	21.50	42.50	15.885	46.50	10.50	22.317
2.50	40.50	14.976	30.50	1.50	18.101	41.50	24.50	12.980
1.50	47.50	18.237	27.50	14.50	7.315	42.50	27.50	17.294
9.50	5.50	10.909	26.50	19.50	9.876	47.50	34.50	20.429
15.50	12.50	15.618	30.50	25.50	13.795	44.50	46.50	20.924
16.50	23.50	12.160	30.50	36.50	18.756			
15.50	26.50	11.928	31.50	43.50	23.995			

Arquivo 7

Amostra=teste49.txt - amostragem aleatoria estratificada: Bell.txt

3			0.50	48.50	17.956	18.50	17.50	18.733
X			12.50	1.50	8.055	18.50	23.50	13.275
Y			13.50	11.50	11.203	15.50	31.50	19.553
Zgauss			7.50	18.50	12.573	15.50	38.50	18.210
2.50	6.50	24.263	12.50	24.50	12.070	19.50	42.50	17.459
2.50	11.50	26.134	9.50	31.50	21.726	25.50	6.50	17.772
2.50	15.50	19.949	9.50	36.50	16.085	21.50	12.50	16.248
5.50	24.50	11.309	10.50	47.50	14.635	25.50	14.50	9.133
3.50	33.50	14.719	17.50	4.50	6.052	27.50	22.50	15.193
3.50	41.50	14.503	19.50	11.50	16.178	27.50	30.50	18.906

27.50	38.50	21.703	34.50	42.50	19.135	47.50	2.50	18.864
25.50	43.50	15.756	37.50	4.50	11.175	43.50	12.50	16.686
34.50	5.50	13.181	35.50	11.50	12.796	42.50	15.50	11.548
33.50	11.50	13.091	40.50	16.50	7.767	47.50	24.50	14.569
28.50	20.50	12.039	40.50	27.50	17.793	44.50	32.50	19.783
30.50	22.50	15.535	39.50	31.50	23.349	48.50	40.50	13.065
34.50	30.50	23.754	41.50	40.50	9.738	49.50	47.50	21.324
32.50	36.50	17.994	41.50	49.50	23.369			

Arquivo 8

Amostra=teste64.txt - amostragem aleatoria estratificada: Bell.txt

3			18.50	22.50	13.859	36.50	15.50	9.625
X			18.50	25.50	13.107	35.50	21.50	12.785
Y			18.50	36.50	18.444	34.50	27.50	20.487
Zgauss			17.50	39.50	18.161	32.50	35.50	17.694
4.50	6.50	20.114	15.50	47.50	13.942	32.50	42.50	22.373
4.50	11.50	22.401	23.50	3.50	12.904	36.50	49.50	23.388
4.50	14.50	18.894	21.50	11.50	17.409	41.50	4.50	14.527
4.50	24.50	11.487	20.50	18.50	13.844	39.50	8.50	11.789
6.50	26.50	12.454	19.50	24.50	13.611	38.50	14.50	7.337
0.50	34.50	15.064	21.50	29.50	16.122	42.50	24.50	14.088
1.50	40.50	15.452	21.50	35.50	19.601	40.50	25.50	14.366
4.50	44.50	15.521	23.50	39.50	16.796	39.50	31.50	23.349
12.50	4.50	4.418	19.50	45.50	15.830	39.50	39.50	11.996
10.50	11.50	12.057	30.50	3.50	16.337	39.50	46.50	20.639
12.50	15.50	13.197	30.50	11.50	12.232	45.50	4.50	15.571
8.50	23.50	7.986	26.50	18.50	8.156	49.50	11.50	21.867
8.50	27.50	18.828	27.50	20.50	12.033	46.50	13.50	20.446
10.50	34.50	20.373	25.50	27.50	18.739	44.50	20.50	15.816
12.50	39.50	15.239	25.50	33.50	22.249	47.50	27.50	12.729
11.50	46.50	11.051	25.50	39.50	18.798	49.50	33.50	21.492
15.50	5.50	5.541	26.50	47.50	15.867	44.50	40.50	10.212
17.50	8.50	10.479	34.50	0.50	15.249	43.50	48.50	22.562
13.50	14.50	14.682	34.50	9.50	13.081			

Arquivo 9

Amostra=teste81.txt - amostragem aleatoria estratificada: Bell.txt

3			8.50	0.50	10.149	15.50	24.50	12.267
X			7.50	9.50	17.156	13.50	29.50	19.320
Y			9.50	16.50	10.265	15.50	36.50	19.691
Zgauss			8.50	17.50	11.474	12.50	40.50	14.855
1.50	4.50	19.622	6.50	22.50	12.974	15.50	45.50	14.006
5.50	10.50	20.571	6.50	29.50	17.235	21.50	2.50	9.432
2.50	15.50	19.949	10.50	38.50	13.746	19.50	8.50	12.148
2.50	19.50	16.859	9.50	43.50	10.283	18.50	12.50	16.955
3.50	27.50	7.879	9.50	45.50	12.685	21.50	21.50	13.424
0.50	31.50	9.934	15.50	4.50	4.500	19.50	23.50	13.561
0.50	37.50	14.380	16.50	7.50	8.300	18.50	31.50	14.966
5.50	42.50	13.820	15.50	15.50	19.072	21.50	36.50	18.656
5.50	44.50	15.161	15.50	20.50	15.507	18.50	39.50	18.166



21.50	44.50	15.567	31.50	33.50	16.897	40.50	23.50	10.794
24.50	0.50	12.063	30.50	41.50	24.350	39.50	30.50	22.763
23.50	8.50	17.809	28.50	47.50	21.133	41.50	36.50	16.965
27.50	13.50	8.910	33.50	1.50	16.285	43.50	42.50	14.513
25.50	19.50	10.274	36.50	9.50	10.947	41.50	45.50	18.343
27.50	26.50	16.459	37.50	13.50	9.991	44.50	1.50	12.779
23.50	29.50	18.518	38.50	18.50	4.243	46.50	7.50	17.710
22.50	35.50	20.952	34.50	22.50	13.217	47.50	16.50	18.780
22.50	40.50	15.401	36.50	32.50	25.508	49.50	19.50	18.512
25.50	45.50	14.761	38.50	38.50	14.901	49.50	24.50	15.276
31.50	5.50	16.792	33.50	41.50	19.041	49.50	32.50	21.334
31.50	10.50	14.583	37.50	45.50	19.452	48.50	37.50	15.235
31.50	15.50	8.948	40.50	5.50	13.490	46.50	42.50	15.045
32.50	20.50	15.281	43.50	7.50	16.183	48.50	49.50	22.276
32.50	27.50	15.101	42.50	15.50	11.548			
30.50	29.50	15.059	41.50	20.50	12.250			

Arquivo 10

Amostra=teste100.txt - amostragem aleatoria estratificada: Bell.txt

3			19.50	9.50	13.681	32.50	34.50	17.840
X			16.50	12.50	16.677	31.50	36.50	18.063
Y			17.50	15.50	19.173	30.50	42.50	24.439
Zgauss			15.50	21.50	13.780	30.50	47.50	24.986
0.50	4.50	19.890	16.50	29.50	14.040	36.50	0.50	12.178
4.50	7.50	21.273	18.50	31.50	14.966	36.50	8.50	10.131
2.50	13.50	25.147	16.50	39.50	18.215	36.50	14.50	10.140
0.50	16.50	18.604	19.50	43.50	17.541	39.50	15.50	6.765
3.50	21.50	17.165	19.50	46.50	14.747	36.50	21.50	9.666
0.50	28.50	7.446	20.50	1.50	10.121	37.50	28.50	22.861
0.50	33.50	14.470	20.50	8.50	14.461	39.50	33.50	23.777
2.50	39.50	12.877	22.50	11.50	17.354	36.50	39.50	14.480
4.50	42.50	14.442	20.50	17.50	13.651	35.50	44.50	19.091
4.50	49.50	21.283	24.50	23.50	15.290	35.50	45.50	20.471
5.50	2.50	9.886	23.50	26.50	17.374	44.50	4.50	14.827
9.50	8.50	13.525	24.50	30.50	19.761	43.50	7.50	16.183
7.50	11.50	17.566	22.50	38.50	16.734	40.50	10.50	13.259
7.50	18.50	12.573	22.50	41.50	14.990	42.50	17.50	12.027
6.50	24.50	10.185	20.50	49.50	11.725	44.50	21.50	15.438
7.50	27.50	17.434	26.50	1.50	14.423	42.50	27.50	17.294
8.50	33.50	19.412	28.50	9.50	15.581	43.50	34.50	19.004
8.50	38.50	9.944	25.50	14.50	9.133	41.50	35.50	18.421
5.50	40.50	10.554	25.50	15.50	6.090	40.50	40.50	10.924
9.50	49.50	17.249	26.50	20.50	11.634	44.50	49.50	23.800
13.50	3.50	4.332	28.50	27.50	16.020	46.50	2.50	16.440
14.50	5.50	5.492	27.50	33.50	20.269	45.50	9.50	21.037
10.50	14.50	10.001	27.50	39.50	22.962	45.50	12.50	21.470
11.50	16.50	10.038	29.50	41.50	24.470	48.50	18.50	19.029
13.50	23.50	10.786	25.50	47.50	13.144	47.50	20.50	16.553
11.50	28.50	20.744	34.50	1.50	15.295	46.50	27.50	13.571
10.50	33.50	21.375	33.50	8.50	14.285	48.50	33.50	21.104
12.50	39.50	15.239	31.50	12.50	12.124	49.50	37.50	16.029
10.50	43.50	9.085	30.50	19.50	13.028	48.50	40.50	13.065
11.50	47.50	12.367	32.50	24.50	12.724	46.50	45.50	20.788
17.50	2.50	10.771	31.50	29.50	15.899			

Arquivo 11

Amostra=normal64.txt - amostragem aleatoria estratificada: Bell.txt

3			17.50	19.50	17.736	36.50	17.50	7.403
X			13.50	29.50	19.320	35.50	24.50	11.095
Y			17.50	36.50	18.810	32.50	28.50	16.577
Zgauss			16.50	42.50	17.851	35.50	34.50	21.760
0.50	3.50	18.888	17.50	45.50	15.613	35.50	38.50	16.356
5.50	8.50	20.454	21.50	0.50	10.488	36.50	48.50	23.271
6.50	14.50	16.917	20.50	7.50	13.165	42.50	1.50	12.646
0.50	20.50	16.421	23.50	16.50	4.655	38.50	9.50	11.827
3.50	27.50	7.879	20.50	20.50	13.770	38.50	13.50	10.238
2.50	35.50	14.328	24.50	27.50	18.627	42.50	21.50	14.194
2.50	41.50	15.964	20.50	33.50	18.697	40.50	28.50	18.870
4.50	44.50	15.521	21.50	39.50	16.150	38.50	31.50	25.189
11.50	6.50	8.858	23.50	48.50	11.521	39.50	39.50	11.996
9.50	10.50	13.449	30.50	4.50	17.556	38.50	47.50	21.635
9.50	13.50	11.378	29.50	10.50	13.403	49.50	1.50	20.815
8.50	20.50	10.700	29.50	16.50	7.381	49.50	8.50	19.712
10.50	28.50	20.980	25.50	20.50	11.834	46.50	13.50	20.446
11.50	36.50	18.662	30.50	30.50	15.419	43.50	20.50	14.836
10.50	37.50	16.159	26.50	36.50	21.807	43.50	28.50	18.026
9.50	45.50	12.685	27.50	41.50	23.252	48.50	33.50	21.104
17.50	6.50	6.737	27.50	47.50	18.581	46.50	39.50	12.051
14.50	12.50	14.256	34.50	0.50	15.249	47.50	49.50	23.685
15.50	16.50	19.393	33.50	11.50	13.091			

Arquivo 12

Amostra=positive64.txt - amostragem aleatoria estratificada: positive.txt

3			14.50	21.50	0.348	33.50	14.50	0.423
X			14.50	28.50	1.387	34.50	21.50	0.739
Y			17.50	35.50	2.113	31.50	25.50	0.581
Zlog			15.50	38.50	1.910	33.50	31.50	3.491
3.50	5.50	3.153	15.50	46.50	0.650	33.50	42.50	3.441
4.50	11.50	5.198	23.50	1.50	0.347	31.50	43.50	7.606
2.50	17.50	1.476	20.50	10.50	1.355	40.50	6.50	0.612
3.50	20.50	1.442	24.50	18.50	0.172	42.50	8.50	1.023
0.50	25.50	0.214	19.50	22.50	0.740	40.50	14.50	0.232
0.50	32.50	0.471	21.50	28.50	1.118	38.50	24.50	0.213
4.50	38.50	0.251	21.50	31.50	1.852	38.50	30.50	8.781
3.50	43.50	1.089	23.50	39.50	1.363	39.50	36.50	2.350
9.50	4.50	0.280	19.50	43.50	1.628	41.50	40.50	0.253
12.50	7.50	0.268	30.50	4.50	1.634	41.50	48.50	5.180
9.50	17.50	0.272	27.50	12.50	0.290	47.50	4.50	2.004
7.50	24.50	0.297	26.50	18.50	0.173	49.50	7.50	2.062
10.50	28.50	3.702	29.50	19.50	0.425	44.50	13.50	2.119
7.50	33.50	2.211	29.50	28.50	0.970	46.50	19.50	1.340
11.50	41.50	0.351	26.50	34.50	4.901	49.50	30.50	2.739
10.50	46.50	0.671	29.50	41.50	8.519	45.50	32.50	3.094
17.50	5.50	0.095	27.50	48.50	2.312	48.50	38.50	0.707
13.50	12.50	0.446	35.50	4.50	0.444	49.50	43.50	1.943
14.50	15.50	1.683	36.50	7.50	0.186			

Arquivo 13

Amostra=positive100.txt - amostragem aleatoria estratificada: positive.txt

3				17.50	5.50	0.095		32.50	32.50	2.107
X				15.50	11.50	0.635		31.50	35.50	1.593
Y				18.50	19.50	1.468		33.50	44.50	4.811
Zlog				17.50	23.50	0.573		33.50	49.50	20.982
3.50	3.50	1.426		17.50	29.50	0.427		37.50	4.50	0.356
4.50	9.50	6.017		18.50	30.50	0.686		35.50	5.50	0.410
1.50	11.50	15.047		17.50	38.50	1.784		35.50	10.50	0.500
0.50	16.50	2.099		16.50	43.50	1.581		38.50	19.50	0.109
0.50	23.50	0.438		17.50	47.50	0.683		38.50	22.50	0.110
3.50	25.50	0.228		23.50	3.50	0.538		37.50	29.50	9.220
2.50	34.50	0.776		21.50	8.50	1.205		39.50	34.50	4.008
0.50	37.50	0.765		24.50	11.50	1.313		38.50	35.50	3.521
0.50	40.50	1.146		24.50	18.50	0.172		36.50	40.50	0.592
0.50	45.50	1.674		22.50	23.50	0.741		35.50	46.50	4.249
5.50	2.50	0.262		22.50	28.50	1.528		41.50	2.50	0.677
6.50	5.50	1.021		21.50	33.50	2.494		40.50	6.50	0.612
9.50	12.50	0.455		20.50	35.50	2.243		42.50	14.50	0.429
8.50	17.50	0.382		20.50	42.50	1.483		42.50	16.50	0.436
7.50	24.50	0.297		20.50	45.50	1.043		44.50	24.50	0.772
7.50	25.50	0.392		26.50	1.50	0.773		41.50	25.50	0.754
5.50	30.50	1.101		27.50	6.50	1.744		44.50	34.50	2.131
8.50	37.50	0.350		27.50	12.50	0.290		43.50	38.50	0.378
6.50	42.50	0.530		29.50	17.50	0.149		43.50	44.50	1.305
5.50	46.50	1.496		28.50	21.50	0.731		44.50	47.50	4.841
12.50	3.50	0.088		25.50	26.50	1.731		48.50	3.50	2.430
12.50	6.50	0.184		27.50	31.50	2.460		48.50	8.50	4.038
11.50	12.50	0.314		26.50	39.50	3.605		46.50	12.50	5.343
13.50	18.50	0.653		27.50	44.50	2.397		46.50	15.50	1.612
13.50	21.50	0.220		28.50	47.50	3.840		49.50	20.50	1.368
12.50	27.50	2.152		33.50	0.50	1.269		47.50	25.50	0.794
10.50	32.50	4.282		34.50	9.50	0.561		47.50	31.50	3.129
13.50	35.50	3.669		33.50	10.50	0.616		47.50	35.50	2.140
13.50	44.50	0.445		33.50	17.50	0.363		47.50	42.50	0.992
13.50	47.50	0.537		31.50	20.50	0.947		46.50	47.50	4.856
17.50	1.50	0.380		31.50	26.50	0.556				

Arquivo 14

Amostra=coKrige2.txt - amostragem aleatoria estratificada

4				2.50	32.50	-99.000	14.411	6.50	39.50	-99.000	10.328
X				2.50	38.50	12.349	11.864	5.50	44.50	15.161	16.854
Y				0.50	43.50	17.903	21.098	8.50	48.50	18.399	20.131
VP				4.50	48.50	-99.000	22.394	12.50	3.50	5.340	7.367
VS1				7.50	3.50	10.660	10.199	13.50	7.50	8.438	9.020
0.50	0.50	19.097	24.737	8.50	9.50	15.691	17.209	12.50	10.50	-99.000	13.330
0.50	9.50	-99.000	30.361	8.50	13.50	-99.000	13.948	11.50	16.50	10.038	11.178
4.50	13.50	20.665	19.641	8.50	17.50	11.474	11.492	12.50	20.50	6.820	5.929
3.50	16.50	17.429	16.459	9.50	24.50	11.316	11.472	10.50	28.50	20.980	18.252
1.50	23.50	-99.000	14.682	6.50	27.50	-99.000	15.746	14.50	34.50	21.222	22.214
0.50	25.50	9.036	9.055	8.50	34.50	18.315	17.740	11.50	38.50	15.198	16.568

13.50	44.50	12.106	13.258	27.50	18.50	-99.000	7.132	36.50	44.50	18.552	16.585
14.50	49.50	13.835	11.852	25.50	22.50	-99.000	13.379	39.50	47.50	21.612	19.295
16.50	1.50	-99.000	10.301	28.50	29.50	16.634	17.481	43.50	1.50	11.802	13.840
15.50	7.50	7.933	7.045	26.50	33.50	21.879	21.817	41.50	9.50	14.612	13.831
16.50	11.50	-99.000	16.614	25.50	35.50	21.977	22.358	44.50	14.50	16.994	17.723
19.50	18.50	16.620	15.034	27.50	40.50	23.158	22.786	40.50	19.50	9.508	8.287
19.50	22.50	-99.000	15.355	25.50	47.50	13.144	13.106	41.50	24.50	-99.000	12.000
15.50	29.50	-99.000	15.847	33.50	3.50	15.017	15.076	41.50	26.50	16.644	17.152
15.50	31.50	19.553	21.527	33.50	6.50	14.593	14.414	44.50	30.50	-99.000	16.049
19.50	36.50	18.727	18.233	31.50	12.50	12.124	9.573	43.50	39.50	9.332	10.695
16.50	41.50	18.096	17.294	34.50	17.50	11.007	12.365	40.50	43.50	-99.000	12.533
16.50	48.50	-99.000	12.324	34.50	22.50	13.217	12.806	41.50	48.50	-99.000	22.873
24.50	0.50	12.063	10.487	34.50	27.50	-99.000	19.807	46.50	3.50	16.309	16.848
21.50	8.50	16.281	15.034	31.50	31.50	-99.000	17.298	45.50	8.50	-99.000	20.829
20.50	14.50	15.221	13.351	32.50	39.50	20.016	17.473	46.50	10.50	22.317	23.422
24.50	17.50	-99.000	4.710	30.50	40.50	24.021	24.072	48.50	18.50	19.029	18.317
23.50	21.50	13.033	12.152	34.50	45.50	-99.000	20.954	46.50	20.50	-99.000	18.380
20.50	28.50	14.962	14.266	36.50	4.50	11.167	11.972	49.50	29.50	15.498	13.346
23.50	30.50	19.185	19.224	37.50	9.50	10.954	10.973	48.50	34.50	20.842	19.584
21.50	35.50	19.601	19.782	38.50	11.50	-99.000	13.038	49.50	37.50	16.029	15.665
24.50	42.50	-99.000	14.898	39.50	16.50	5.811	5.691	48.50	44.50	-99.000	20.907
22.50	49.50	-99.000	8.290	39.50	24.50	9.676	10.180	48.50	48.50	21.903	20.002
28.50	1.50	16.591	18.476	35.50	29.50	-99.000	25.662				
26.50	7.50	17.898	18.710	35.50	31.50	25.660	24.908	Obs. -99.000 indica dado			
27.50	14.50	7.315	10.196	38.50	37.50	16.384	16.256	inexistente.			

Arquivo 15

Amostra=coTrend2.txt - amostragem aleatoria estratificada

4				8.50	48.50	18.399	15.186	24.50	17.50	-99.000	12.323
X				12.50	3.50	5.340	10.082	23.50	21.50	13.033	14.062
Y				13.50	7.50	8.438	12.662	20.50	28.50	14.962	18.014
VP				12.50	10.50	-99.000	12.919	23.50	30.50	19.185	19.499
VS2				11.50	16.50	10.038	12.504	21.50	35.50	19.601	19.914
0.50	0.50	19.097	17.580	12.50	20.50	6.820	12.861	24.50	42.50	-99.000	17.924
0.50	9.50	-99.000	27.603	10.50	28.50	20.980	15.120	22.50	49.50	-99.000	15.551
4.50	13.50	20.665	17.745	14.50	34.50	21.222	17.583	28.50	1.50	16.591	14.599
3.50	16.50	17.429	16.890	11.50	38.50	15.198	15.563	26.50	7.50	17.898	13.984
1.50	23.50	-99.000	13.954	13.50	44.50	12.106	13.279	27.50	14.50	7.315	11.819
0.50	25.50	9.036	13.109	14.50	49.50	13.835	13.299	27.50	18.50	-99.000	12.474
2.50	32.50	-99.000	12.442	16.50	1.50	-99.000	7.834	25.50	22.50	-99.000	14.667
2.50	38.50	12.349	12.746	15.50	7.50	7.933	12.506	28.50	29.50	16.634	19.185
0.50	43.50	17.903	15.025	16.50	11.50	-99.000	12.187	26.50	33.50	21.879	20.753
4.50	48.50	-99.000	18.953	19.50	18.50	16.620	12.540	25.50	35.50	21.977	20.786
7.50	3.50	10.660	12.409	19.50	22.50	-99.000	14.346	27.50	40.50	23.158	19.797
8.50	9.50	15.691	15.116	15.50	29.50	-99.000	17.106	25.50	47.50	13.144	16.735
8.50	13.50	-99.000	14.215	15.50	31.50	19.553	17.705	33.50	3.50	15.017	15.594
8.50	17.50	11.474	13.297	19.50	36.50	18.727	19.107	33.50	6.50	14.593	14.463
9.50	24.50	11.316	13.728	16.50	41.50	18.096	15.665	31.50	12.50	12.124	11.754
6.50	27.50	-99.000	13.732	16.50	48.50	-99.000	12.860	34.50	17.50	11.007	11.222
8.50	34.50	18.315	15.146	24.50	0.50	12.063	11.587	34.50	22.50	13.217	13.510
6.50	39.50	-99.000	13.889	21.50	8.50	16.281	13.035	34.50	27.50	-99.000	16.903
5.50	44.50	15.161	14.243	20.50	14.50	15.221	11.924	31.50	31.50	-99.000	19.841



32.50	39.50	20.016	20.475	39.50	47.50	21.612	22.693	45.50	8.50	-99.000	16.278
30.50	40.50	24.021	20.229	43.50	1.50	11.802	14.785	46.50	10.50	22.317	16.903
34.50	45.50	-99.000	20.340	41.50	9.50	14.612	13.091	48.50	18.50	19.029	18.863
36.50	4.50	11.167	15.102	44.50	14.50	16.994	13.456	46.50	20.50	-99.000	15.751
37.50	9.50	10.954	12.504	40.50	19.50	9.508	11.730	49.50	29.50	15.498	19.065
38.50	11.50	-99.000	11.605	41.50	24.50	-99.000	14.102	48.50	34.50	20.842	16.450
39.50	16.50	5.811	10.920	41.50	26.50	16.644	15.101	49.50	37.50	16.029	14.934
39.50	24.50	9.676	14.060	44.50	30.50	-99.000	16.564	48.50	44.50	-99.000	14.819
35.50	29.50	-99.000	17.889	43.50	39.50	9.332	16.771	48.50	48.50	21.903	21.573
35.50	31.50	25.660	18.898	40.50	43.50	-99.000	18.871				
38.50	37.50	16.384	19.117	41.50	48.50	-99.000	24.537	Obs. -99.000	indica dado		
36.50	44.50	18.552	20.100	46.50	3.50	16.309	17.644	inexistente.			

Arquivo 16

Amostra=colo_alta2.txt - amostragem aleatoria estratificada

4				13.50	44.50	12.106	13.258	34.50	17.50	11.007	12.365
X				14.50	49.50	13.835	11.852	34.50	22.50	13.217	12.806
Y				15.50	7.50	7.933	7.045	32.50	39.50	20.016	17.473
VP				19.50	18.50	16.620	15.034	30.50	40.50	24.021	24.072
VS1				15.50	31.50	19.553	21.527	36.50	4.50	11.167	11.972
0.50	0.50	19.097	24.737	19.50	36.50	18.727	18.233	37.50	9.50	10.954	10.973
4.50	13.50	20.665	19.641	16.50	41.50	18.096	17.294	39.50	16.50	5.811	5.691
3.50	16.50	17.429	16.459	24.50	0.50	12.063	10.487	39.50	24.50	9.676	10.180
0.50	25.50	9.036	9.055	21.50	8.50	16.281	15.034	35.50	31.50	25.660	24.908
2.50	38.50	12.349	11.864	20.50	14.50	15.221	13.351	38.50	37.50	16.384	16.256
0.50	43.50	17.903	21.098	23.50	21.50	13.033	12.152	36.50	44.50	18.552	16.585
7.50	3.50	10.660	10.199	20.50	28.50	14.962	14.266	39.50	47.50	21.612	19.295
8.50	9.50	15.691	17.209	23.50	30.50	19.185	19.224	43.50	1.50	11.802	13.840
8.50	17.50	11.474	11.492	21.50	35.50	19.601	19.782	41.50	9.50	14.612	13.831
9.50	24.50	11.316	11.472	28.50	1.50	16.591	18.476	44.50	14.50	16.994	17.723
8.50	34.50	18.315	17.740	26.50	7.50	17.898	18.710	40.50	19.50	9.508	8.287
5.50	44.50	15.161	16.854	27.50	14.50	7.315	10.196	41.50	26.50	16.644	17.152
8.50	48.50	18.399	20.131	28.50	29.50	16.634	17.481	43.50	39.50	9.332	10.695
12.50	3.50	5.340	7.367	26.50	33.50	21.879	21.817	46.50	3.50	16.309	16.848
13.50	7.50	8.438	9.020	25.50	35.50	21.977	22.358	46.50	10.50	22.317	23.422
11.50	16.50	10.038	11.178	27.50	40.50	23.158	22.786	48.50	18.50	19.029	18.317
12.50	20.50	6.820	5.929	25.50	47.50	13.144	13.106	49.50	29.50	15.498	13.346
10.50	28.50	20.980	18.252	33.50	3.50	15.017	15.076	48.50	34.50	20.842	19.584
14.50	34.50	21.222	22.214	33.50	6.50	14.593	14.414	49.50	37.50	16.029	15.665
11.50	38.50	15.198	16.568	31.50	12.50	12.124	9.573	48.50	48.50	21.903	20.002

Arquivo 17

Amostra=colo_media2.txt - amostragem aleatoria estratificada

4				3.50	16.50	17.429	16.890	9.50	24.50	11.316	13.728
X				0.50	25.50	9.036	13.109	8.50	34.50	18.315	15.146
Y				2.50	38.50	12.349	12.746	5.50	44.50	15.161	14.243
VP				0.50	43.50	17.903	15.025	8.50	48.50	18.399	15.186
VS2				7.50	3.50	10.660	12.409	12.50	3.50	5.340	10.082
0.50	0.50	19.097	17.580	8.50	9.50	15.691	15.116	13.50	7.50	8.438	12.662
4.50	13.50	20.665	17.745	8.50	17.50	11.474	13.297	11.50	16.50	10.038	12.504

12.50	20.50	6.820	12.861	28.50	1.50	16.591	14.599	39.50	24.50	9.676	14.060
10.50	28.50	20.980	15.120	26.50	7.50	17.898	13.984	35.50	31.50	25.660	18.898
14.50	34.50	21.222	17.583	27.50	14.50	7.315	11.819	38.50	37.50	16.384	19.117
11.50	38.50	15.198	15.563	28.50	29.50	16.634	19.185	36.50	44.50	18.552	20.100
13.50	44.50	12.106	13.279	26.50	33.50	21.879	20.753	39.50	47.50	21.612	22.693
14.50	49.50	13.835	13.299	25.50	35.50	21.977	20.786	43.50	1.50	11.802	14.785
15.50	7.50	7.933	12.506	27.50	40.50	23.158	19.797	41.50	9.50	14.612	13.091
19.50	18.50	16.620	12.540	25.50	47.50	13.144	16.735	44.50	14.50	16.994	13.456
15.50	31.50	19.553	17.705	33.50	3.50	15.017	15.594	40.50	19.50	9.508	11.730
19.50	36.50	18.727	19.107	33.50	6.50	14.593	14.463	41.50	26.50	16.644	15.101
16.50	41.50	18.096	15.665	31.50	12.50	12.124	11.754	43.50	39.50	9.332	16.771
24.50	0.50	12.063	11.587	34.50	17.50	11.007	11.222	46.50	3.50	16.309	17.644
21.50	8.50	16.281	13.035	34.50	22.50	13.217	13.510	46.50	10.50	22.317	16.903
20.50	14.50	15.221	11.924	32.50	39.50	20.016	20.475	48.50	18.50	19.029	18.863
23.50	21.50	13.033	14.062	30.50	40.50	24.021	20.229	49.50	29.50	15.498	19.065
20.50	28.50	14.962	18.014	36.50	4.50	11.167	15.102	48.50	34.50	20.842	16.450
23.50	30.50	19.185	19.499	37.50	9.50	10.954	12.504	49.50	37.50	16.029	14.934
21.50	35.50	19.601	19.914	39.50	16.50	5.811	10.920	48.50	48.50	21.903	21.573

Arquivo 18

Amostra=teste289alta.gl2 - malha regular com VS1

3			21.25	3.75	12.650	1.25	8.75	31.518
X			23.75	3.75	8.596	3.75	8.75	24.207
Y			26.25	3.75	18.854	6.25	8.75	18.541
VS1			28.75	3.75	14.320	8.75	8.75	16.955
1.25	1.25	19.046	31.25	3.75	14.824	11.25	8.75	14.731
3.75	1.25	6.802	33.75	3.75	15.099	13.75	8.75	10.048
6.25	1.25	10.318	36.25	3.75	17.833	16.25	8.75	11.465
8.75	1.25	9.372	38.75	3.75	9.438	18.75	8.75	14.318
11.25	1.25	7.442	41.25	3.75	17.219	21.25	8.75	15.473
13.75	1.25	1.943	43.75	3.75	14.223	23.75	8.75	20.383
16.25	1.25	14.714	46.25	3.75	13.196	26.25	8.75	15.210
18.75	1.25	8.710	48.75	3.75	23.247	28.75	8.75	17.146
21.25	1.25	11.115	1.25	6.25	22.303	31.25	8.75	15.308
23.75	1.25	11.420	3.75	6.25	21.132	33.75	8.75	13.517
26.25	1.25	14.671	6.25	6.25	15.646	36.25	8.75	12.128
28.75	1.25	18.796	8.75	6.25	10.312	38.75	8.75	13.349
31.25	1.25	20.559	11.25	6.25	7.774	41.25	8.75	12.228
33.75	1.25	13.876	13.75	6.25	8.018	43.75	8.75	17.672
36.25	1.25	18.124	16.25	6.25	6.504	46.25	8.75	19.139
38.75	1.25	12.709	18.75	6.25	4.388	48.75	8.75	20.246
41.25	1.25	14.831	21.25	6.25	10.209	1.25	11.25	28.377
43.75	1.25	10.832	23.75	6.25	16.072	3.75	11.25	22.853
46.25	1.25	20.060	26.25	6.25	20.434	6.25	11.25	18.336
48.75	1.25	17.416	28.75	6.25	15.639	8.75	11.25	14.283
1.25	3.75	20.528	31.25	6.25	22.177	11.25	11.25	15.383
3.75	3.75	17.122	33.75	6.25	11.446	13.75	11.25	12.450
6.25	3.75	13.850	36.25	6.25	7.035	16.25	11.25	13.623
8.75	3.75	4.191	38.75	6.25	11.891	18.75	11.25	16.139
11.25	3.75	7.707	41.25	6.25	12.202	21.25	11.25	17.062
13.75	3.75	7.022	43.75	6.25	18.311	23.75	11.25	18.521
16.25	3.75	6.534	46.25	6.25	20.502	26.25	11.25	13.310
18.75	3.75	7.012	48.75	6.25	16.216	28.75	11.25	17.042



31.25	11.25	10.309	21.25	18.75	17.359	11.25	26.25	15.832
33.75	11.25	16.002	23.75	18.75	8.747	13.75	26.25	12.225
36.25	11.25	11.075	26.25	18.75	6.202	16.25	26.25	15.981
38.75	11.25	9.125	28.75	18.75	6.366	18.75	26.25	9.189
41.25	11.25	15.983	31.25	18.75	15.604	21.25	26.25	14.238
43.75	11.25	20.409	33.75	18.75	15.742	23.75	26.25	16.899
46.25	11.25	23.472	36.25	18.75	10.642	26.25	26.25	17.999
48.75	11.25	22.360	38.75	18.75	4.581	28.75	26.25	15.702
1.25	13.75	28.130	41.25	18.75	9.492	31.25	26.25	17.358
3.75	13.75	21.617	43.75	18.75	12.897	33.75	26.25	9.044
6.25	13.75	18.791	46.25	18.75	14.384	36.25	26.25	18.445
8.75	13.75	12.101	48.75	18.75	18.866	38.75	26.25	12.732
11.25	13.75	5.765	1.25	21.25	14.526	41.25	26.25	17.474
13.75	13.75	14.852	3.75	21.25	13.436	43.75	26.25	20.032
16.25	13.75	20.806	6.25	21.25	18.181	46.25	26.25	13.277
18.75	13.75	17.449	8.75	21.25	8.715	48.75	26.25	14.035
21.25	13.75	16.142	11.25	21.25	4.182	1.25	28.75	7.338
23.75	13.75	13.928	13.75	21.25	9.214	3.75	28.75	8.305
26.25	13.75	11.232	16.25	21.25	15.401	6.25	28.75	20.188
28.75	13.75	5.924	18.75	21.25	13.440	8.75	28.75	19.939
31.25	13.75	7.243	21.25	21.25	11.753	11.25	28.75	17.211
33.75	13.75	12.621	23.75	21.25	8.883	13.75	28.75	13.958
36.25	13.75	11.487	26.25	21.25	16.310	16.25	28.75	12.690
38.75	13.75	12.565	28.75	21.25	15.817	18.75	28.75	12.435
41.25	13.75	12.758	31.25	21.25	15.040	21.25	28.75	19.676
43.75	13.75	13.311	33.75	21.25	14.405	23.75	28.75	15.696
46.25	13.75	22.959	36.25	21.25	8.331	26.25	28.75	21.745
48.75	13.75	22.070	38.75	21.25	11.358	28.75	28.75	11.715
1.25	16.25	17.764	41.25	21.25	12.899	31.25	28.75	16.408
3.75	16.25	14.412	43.75	21.25	13.318	33.75	28.75	22.808
6.25	16.25	13.629	46.25	21.25	21.587	36.25	28.75	28.074
8.75	16.25	15.931	48.75	21.25	16.099	38.75	28.75	19.239
11.25	16.25	11.292	1.25	23.75	17.138	41.25	28.75	22.054
13.75	16.25	16.548	3.75	23.75	12.763	43.75	28.75	14.182
16.25	16.25	17.787	6.25	23.75	10.147	46.25	28.75	18.340
18.75	16.25	12.066	8.75	23.75	8.309	48.75	28.75	13.359
21.25	16.25	10.052	11.25	23.75	13.047	1.25	31.25	11.257
23.75	16.25	4.264	13.75	23.75	7.704	3.75	31.25	15.801
26.25	16.25	8.208	16.25	23.75	12.424	6.25	31.25	16.605
28.75	16.25	7.139	18.75	23.75	15.758	8.75	31.25	23.143
31.25	16.25	10.367	21.25	23.75	16.809	11.25	31.25	22.951
33.75	16.25	8.575	23.75	23.75	15.466	13.75	31.25	23.179
36.25	16.25	9.416	26.25	23.75	13.999	16.25	31.25	21.173
38.75	16.25	2.693	28.75	23.75	13.736	18.75	31.25	13.374
41.25	16.25	9.666	31.25	23.75	14.892	21.25	31.25	15.796
43.75	16.25	17.132	33.75	23.75	13.343	23.75	31.25	18.591
46.25	16.25	15.370	36.25	23.75	14.497	26.25	31.25	20.197
48.75	16.25	19.508	38.75	23.75	8.996	28.75	31.25	20.792
1.25	18.75	19.812	41.25	23.75	10.265	31.25	31.25	19.252
3.75	18.75	20.737	43.75	23.75	16.491	33.75	31.25	16.480
6.25	18.75	13.017	46.25	23.75	13.810	36.25	31.25	25.592
8.75	18.75	8.208	48.75	23.75	12.840	38.75	31.25	21.224
11.25	18.75	6.286	1.25	26.25	4.567	41.25	31.25	19.110
13.75	18.75	14.554	3.75	26.25	8.796	43.75	31.25	18.974
16.25	18.75	21.400	6.25	26.25	13.054	46.25	31.25	20.691
18.75	18.75	16.246	8.75	26.25	13.776	48.75	31.25	15.473

1.25	33.75	11.499	18.75	38.75	14.483	36.25	43.75	16.652
3.75	33.75	16.600	21.25	38.75	16.245	38.75	43.75	12.906
6.25	33.75	16.073	23.75	38.75	21.148	41.25	43.75	13.180
8.75	33.75	18.078	26.25	38.75	17.741	43.75	43.75	17.044
11.25	33.75	17.865	28.75	38.75	24.326	46.25	43.75	21.699
13.75	33.75	23.257	31.25	38.75	20.176	48.75	43.75	17.810
16.25	33.75	20.873	33.75	38.75	11.118	1.25	46.25	20.527
18.75	33.75	14.707	36.25	38.75	10.456	3.75	46.25	16.453
21.25	33.75	20.752	38.75	38.75	16.441	6.25	46.25	15.855
23.75	33.75	24.958	41.25	38.75	17.532	8.75	46.25	10.711
26.25	33.75	21.510	43.75	38.75	10.486	11.25	46.25	9.974
28.75	33.75	16.810	46.25	38.75	13.166	13.75	46.25	16.492
31.25	33.75	18.867	48.75	38.75	13.420	16.25	46.25	12.134
33.75	33.75	21.110	1.25	41.25	12.777	18.75	46.25	14.511
36.25	33.75	28.046	3.75	41.25	13.098	21.25	46.25	12.726
38.75	33.75	24.806	6.25	41.25	19.557	23.75	46.25	12.913
41.25	33.75	22.083	8.75	41.25	10.494	26.25	46.25	16.516
43.75	33.75	19.472	11.25	41.25	15.089	28.75	46.25	22.690
46.25	33.75	18.554	13.75	41.25	13.923	31.25	46.25	27.350
48.75	33.75	22.519	16.25	41.25	21.505	33.75	46.25	22.017
1.25	36.25	18.097	18.75	41.25	14.842	36.25	46.25	17.137
3.75	36.25	7.923	21.25	41.25	16.663	38.75	46.25	20.512
6.25	36.25	12.380	23.75	41.25	17.123	41.25	46.25	17.387
8.75	36.25	13.234	26.25	41.25	14.332	43.75	46.25	23.969
11.25	36.25	23.052	28.75	41.25	27.205	46.25	46.25	19.923
13.75	36.25	19.064	31.25	41.25	22.435	48.75	46.25	19.005
16.25	36.25	17.937	33.75	41.25	19.928	1.25	48.75	20.606
18.75	36.25	20.762	36.25	41.25	12.045	3.75	48.75	20.717
21.25	36.25	19.199	38.75	41.25	8.987	6.25	48.75	23.066
23.75	36.25	15.553	41.25	41.25	13.758	8.75	48.75	23.137
26.25	36.25	26.436	43.75	41.25	7.608	11.25	48.75	13.594
28.75	36.25	17.894	46.25	41.25	12.589	13.75	48.75	12.534
31.25	36.25	17.038	48.75	41.25	16.594	16.25	48.75	13.199
33.75	36.25	17.967	1.25	43.75	15.913	18.75	48.75	14.097
36.25	36.25	13.408	3.75	43.75	7.557	21.25	48.75	11.244
38.75	36.25	16.496	6.25	43.75	22.279	23.75	48.75	7.766
41.25	36.25	18.818	8.75	43.75	11.044	26.25	48.75	14.012
43.75	36.25	13.171	11.25	43.75	7.558	28.75	48.75	27.504
46.25	36.25	18.970	13.75	43.75	13.171	31.25	48.75	25.523
48.75	36.25	15.949	16.25	43.75	13.995	33.75	48.75	30.351
1.25	38.75	11.264	18.75	43.75	21.433	36.25	48.75	24.210
3.75	38.75	11.732	21.25	43.75	16.553	38.75	48.75	22.491
6.25	38.75	9.726	23.75	43.75	11.091	41.25	48.75	23.077
8.75	38.75	6.291	26.25	43.75	21.502	43.75	48.75	25.088
11.25	38.75	18.669	28.75	43.75	22.072	46.25	48.75	27.082
13.75	38.75	10.691	31.25	43.75	20.474	48.75	48.75	19.480
16.25	38.75	21.885	33.75	43.75	24.153			

Arquivo 19

Amostra=teste289media.gl2 - malha regular com VS2

3	1.25	1.25	18.117	11.25	1.25	5.641
X	3.75	1.25	12.150	13.75	1.25	6.056
Y	6.25	1.25	8.330	16.25	1.25	7.193
VS2	8.75	1.25	6.282	18.75	1.25	8.752



21.25	1.25	10.464	11.25	8.75	13.456	1.25	16.25	19.623
23.75	1.25	12.103	13.75	8.75	12.722	3.75	16.25	16.813
26.25	1.25	13.490	16.25	8.75	12.521	6.25	16.25	14.778
28.75	1.25	14.504	18.75	8.75	12.654	8.75	16.25	13.405
31.25	1.25	15.086	21.25	8.75	12.944	11.25	16.25	12.572
33.75	1.25	15.243	23.75	8.75	13.246	13.75	16.25	12.146
36.25	1.25	15.063	26.25	8.75	13.455	16.25	16.25	11.998
38.75	1.25	14.713	28.75	8.75	13.510	18.75	16.25	12.006
41.25	1.25	14.452	31.25	8.75	13.402	21.25	16.25	12.061
43.75	1.25	14.634	33.75	8.75	13.180	23.75	16.25	12.075
46.25	1.25	15.718	36.25	8.75	12.960	26.25	16.25	11.990
48.75	1.25	18.274	38.75	8.75	12.932	28.75	16.25	11.780
1.25	3.75	24.186	41.25	8.75	13.362	31.25	16.25	11.464
3.75	3.75	18.243	43.75	8.75	14.605	33.75	16.25	11.107
6.25	3.75	14.214	46.25	8.75	17.108	36.25	16.25	10.830
8.75	3.75	11.778	48.75	8.75	21.420	38.75	16.25	10.818
11.25	3.75	10.613	1.25	11.25	24.530	41.25	16.25	11.323
13.75	3.75	10.413	3.75	11.25	20.084	43.75	16.25	12.675
16.25	3.75	10.887	6.25	11.25	16.850	46.25	16.25	15.284
18.75	3.75	11.767	8.75	11.25	14.642	48.75	16.25	19.654
21.25	3.75	12.821	11.25	11.25	13.264	1.25	18.75	17.293
23.75	3.75	13.851	13.75	11.25	12.523	3.75	18.75	15.344
26.25	3.75	14.707	16.25	11.25	12.233	6.25	18.75	13.970
28.75	3.75	15.290	18.75	11.25	12.223	8.75	18.75	13.093
31.25	3.75	15.558	21.25	11.25	12.341	11.25	18.75	12.620
33.75	3.75	15.539	23.75	11.25	12.465	13.75	18.75	12.448
36.25	3.75	15.330	26.25	11.25	12.508	16.25	18.75	12.470
38.75	3.75	15.109	28.75	11.25	12.422	18.75	18.75	12.584
41.25	3.75	15.140	31.25	11.25	12.211	21.25	18.75	12.699
43.75	3.75	15.782	33.75	11.25	11.933	23.75	18.75	12.741
46.25	3.75	17.491	36.25	11.25	11.708	26.25	18.75	12.661
48.75	3.75	20.832	38.75	11.25	11.725	28.75	18.75	12.441
1.25	6.25	26.444	41.25	11.25	12.250	31.25	18.75	12.102
3.75	6.25	20.801	43.75	11.25	13.632	33.75	18.75	11.710
6.25	6.25	16.838	46.25	11.25	16.310	36.25	18.75	11.382
8.75	6.25	14.278	48.75	11.25	20.818	38.75	18.75	11.296
11.25	6.25	12.848	1.25	13.75	22.152	41.25	18.75	11.693
13.75	6.25	12.280	3.75	13.75	18.494	43.75	18.75	12.889
16.25	6.25	12.321	6.25	13.75	15.825	46.25	18.75	15.278
18.75	6.25	12.739	8.75	13.75	13.995	48.75	18.75	19.342
21.25	6.25	13.330	11.25	13.75	12.848	1.25	21.25	15.355
23.75	6.25	13.926	13.75	13.75	12.222	3.75	21.25	14.237
26.25	6.25	14.398	16.25	13.75	11.960	6.25	21.25	13.511
28.75	6.25	14.668	18.75	13.75	11.918	8.75	21.25	13.132
31.25	6.25	14.714	21.25	13.75	11.966	11.25	21.25	13.034
33.75	6.25	14.574	23.75	13.75	12.002	13.75	21.25	13.138
36.25	6.25	14.356	26.25	13.75	11.952	16.25	21.25	13.359
38.75	6.25	14.246	28.75	13.75	11.783	18.75	21.25	13.613
41.25	6.25	14.510	31.25	13.75	11.507	21.25	21.25	13.823
43.75	6.25	15.506	33.75	13.75	11.186	23.75	21.25	13.926
46.25	6.25	17.688	36.25	13.75	10.943	26.25	21.25	13.880
48.75	6.25	21.613	38.75	13.75	10.966	28.75	21.25	13.671
1.25	8.75	26.216	41.25	13.75	11.516	31.25	21.25	13.322
3.75	8.75	21.091	43.75	13.75	12.932	33.75	21.25	12.893
6.25	8.75	17.410	46.25	13.75	15.642	36.25	21.25	12.498
8.75	8.75	14.943	48.75	13.75	20.165	38.75	21.25	12.303

41.25	21.25	12.537	31.25	28.75	18.403	21.25	36.25	19.721
43.75	21.25	13.500	33.75	28.75	17.887	23.75	36.25	20.378
46.25	21.25	15.564	36.25	28.75	17.253	26.25	36.25	20.829
48.75	21.25	19.189	38.75	28.75	16.622	28.75	36.25	21.024
1.25	23.75	13.881	41.25	28.75	16.167	31.25	36.25	20.932
3.75	23.75	13.524	43.75	28.75	16.121	33.75	36.25	20.541
6.25	23.75	13.397	46.25	28.75	16.782	36.25	36.25	19.867
8.75	23.75	13.483	48.75	28.75	18.519	38.75	36.25	18.960
11.25	23.75	13.743	1.25	31.25	11.882	41.25	36.25	17.915
13.75	23.75	14.120	3.75	31.25	13.021	43.75	36.25	16.871
16.25	23.75	14.548	6.25	31.25	14.066	46.25	36.25	16.027
18.75	23.75	14.956	8.75	31.25	15.071	48.75	36.25	15.641
21.25	23.75	15.279	11.25	31.25	16.054	1.25	38.75	12.347
23.75	23.75	15.463	13.75	31.25	17.009	3.75	38.75	13.178
26.25	23.75	15.470	16.25	31.25	17.908	6.25	38.75	13.905
28.75	23.75	15.289	18.75	31.25	18.709	8.75	38.75	14.626
31.25	23.75	14.938	21.25	31.25	19.367	11.25	38.75	15.395
33.75	23.75	14.474	23.75	31.25	19.833	13.75	38.75	16.229
36.25	23.75	13.999	26.25	31.25	20.069	16.25	38.75	17.115
38.75	23.75	13.670	28.75	31.25	20.049	18.75	38.75	18.016
41.25	23.75	13.697	31.25	31.25	19.771	21.25	38.75	18.879
43.75	23.75	14.362	33.75	31.25	19.257	23.75	38.75	19.640
46.25	23.75	16.015	36.25	31.25	18.566	26.25	38.75	20.233
48.75	23.75	19.089	38.75	31.25	17.800	28.75	38.75	20.596
1.25	26.25	12.858	41.25	31.25	17.107	31.25	38.75	20.677
3.75	26.25	13.152	43.75	31.25	16.692	33.75	38.75	20.442
6.25	26.25	13.540	46.25	31.25	16.822	36.25	38.75	19.883
8.75	26.25	14.030	48.75	31.25	17.834	38.75	38.75	19.021
11.25	26.25	14.606	1.25	33.75	11.784	41.25	38.75	17.917
13.75	26.25	15.229	3.75	33.75	13.055	43.75	38.75	16.677
16.25	26.25	15.849	6.25	33.75	14.190	46.25	38.75	15.458
18.75	26.25	16.408	8.75	33.75	15.259	48.75	38.75	14.476
21.25	26.25	16.850	11.25	33.75	16.296	1.25	41.25	13.287
23.75	26.25	17.125	13.75	33.75	17.305	3.75	41.25	13.503
26.25	26.25	17.199	16.25	33.75	18.265	6.25	41.25	13.698
28.75	26.25	17.058	18.75	33.75	19.140	8.75	41.25	13.978
31.25	26.25	16.715	21.25	33.75	19.882	11.25	41.25	14.404
33.75	26.25	16.219	23.75	33.75	20.443	13.75	41.25	14.995
36.25	26.25	15.662	26.25	33.75	20.777	16.25	41.25	15.739
38.75	26.25	15.182	28.75	33.75	20.849	18.75	41.25	16.594
41.25	26.25	14.974	31.25	33.75	20.643	21.25	41.25	17.498
43.75	26.25	15.296	33.75	33.75	20.165	23.75	41.25	18.378
46.25	26.25	16.473	36.25	33.75	19.454	26.25	41.25	19.153
48.75	26.25	18.908	38.75	33.75	18.588	28.75	41.25	19.742
1.25	28.75	12.219	41.25	33.75	17.690	31.25	41.25	20.073
3.75	28.75	13.022	43.75	33.75	16.933	33.75	41.25	20.088
6.25	28.75	13.810	46.25	33.75	16.550	36.25	41.25	19.748
8.75	28.75	14.615	48.75	33.75	16.841	38.75	41.25	19.045
11.25	28.75	15.438	1.25	36.25	11.913	41.25	41.25	18.004
13.75	28.75	16.260	3.75	36.25	13.089	43.75	41.25	16.692
16.25	28.75	17.042	6.25	36.25	14.124	46.25	41.25	15.224
18.75	28.75	17.736	8.75	36.25	15.104	48.75	41.25	13.773
21.25	28.75	18.291	11.25	36.25	16.073	1.25	43.75	15.090
23.75	28.75	18.662	13.75	36.25	17.045	3.75	43.75	14.407
26.25	28.75	18.812	16.25	36.25	18.002	6.25	43.75	13.832
28.75	28.75	18.724	18.75	36.25	18.910	8.75	43.75	13.480

11.25	43.75	13.415	8.75	46.25	13.654	6.25	48.75	17.470
13.75	43.75	13.657	11.25	46.25	12.947	8.75	48.75	15.258
16.25	43.75	14.188	13.75	46.25	12.732	11.25	48.75	13.759
18.75	43.75	14.961	16.25	46.25	12.985	13.75	48.75	12.983
21.25	43.75	15.904	18.75	46.25	13.649	16.25	48.75	12.899
23.75	43.75	16.930	21.25	46.25	14.639	18.75	48.75	13.437
26.25	43.75	17.939	23.75	46.25	15.851	21.25	48.75	14.497
28.75	43.75	18.833	26.25	46.25	17.168	23.75	48.75	15.954
31.25	43.75	19.513	28.75	46.25	18.464	26.25	48.75	17.669
33.75	43.75	19.894	31.25	46.25	19.617	28.75	48.75	19.492
36.25	43.75	19.907	33.75	46.25	20.510	31.25	48.75	21.268
38.75	43.75	19.509	36.25	46.25	21.041	33.75	48.75	22.850
41.25	43.75	18.687	38.75	46.25	21.129	36.25	48.75	24.098
43.75	43.75	17.468	41.25	46.25	20.722	38.75	48.75	24.893
46.25	43.75	15.921	43.75	46.25	19.801	41.25	48.75	25.137
48.75	43.75	14.171	46.25	46.25	18.390	43.75	48.75	24.766
1.25	46.25	18.309	48.75	46.25	16.561	46.25	48.75	23.754
3.75	46.25	16.427	1.25	48.75	23.723	48.75	48.75	22.119
6.25	46.25	14.835	3.75	48.75	20.332			

Arquivo 20

Amostra=teste10.txt - Yamamoto et al. (2012)

3			9.50	28.50	1	72.50	17.50	5
X			28.50	20.50	4	62.50	44.50	2
Y			20.50	37.50	1	98.50	5.50	5
TIPOS CATEGORICA 1-5			51.50	9.50	3	80.50	30.50	2
19.50	21.50	4	44.50	41.50	2			

Arquivo 21

Amostra=teste50.txt - Yamamoto et al. (2012)

3			24.50	45.50	1	61.50	26.50	3
X			37.50	8.50	4	64.50	33.50	2
Y			30.50	13.50	4	67.50	42.50	2
TIPOS CATEGORICA 1-5			39.50	24.50	3	76.50	4.50	5
6.50	5.50	4	33.50	30.50	3	76.50	13.50	5
3.50	13.50	4	36.50	45.50	1	70.50	20.50	5
8.50	27.50	1	49.50	2.50	5	71.50	35.50	2
2.50	31.50	1	41.50	11.50	3	70.50	41.50	2
2.50	48.50	1	42.50	26.50	3	87.50	0.50	5
10.50	5.50	4	48.50	37.50	2	85.50	10.50	5
16.50	19.50	4	46.50	44.50	2	89.50	23.50	2
13.50	22.50	4	52.50	9.50	3	85.50	33.50	2
12.50	35.50	1	56.50	14.50	3	80.50	45.50	2
14.50	45.50	1	56.50	27.50	3	98.50	6.50	5
20.50	3.50	4	51.50	31.50	3	93.50	11.50	5
20.50	18.50	4	55.50	42.50	2	92.50	27.50	2
24.50	24.50	4	61.50	6.50	5	90.50	37.50	2
26.50	39.50	1	63.50	18.50	3	93.50	49.50	2

Referências Bibliográficas

- ANDRIOTTI, J. L. S. *Fundamentos de estatística e geoestatística*. Editora Unisinos, 2003.
- ARMSTRONG, M. Is research in mining geostats as dead as a dodo? In: DIMITRAKOPOULOS, R. (Ed.). *Geostatistics for the next century*. Kluwer Academic, 1994. p. 303-312.
- ARMSTRONG, M. *Basic linear geostatistics*. Springer, 1998.
- BACHMAIER, M.; BACKES, M. Variogram or semivariogram? Understanding the variances in a variogram. *Precision agriculture*, v. 9, n. 3, p. 173-175, 2008.
- BLANCHET, P. H.; GODWIN, C. I. Geology system for computer and manual analysis of geologic data from porphyry and other deposits. *Economic Geology*, v. 67, p. 796-813, 1972.
- BOHLING, G. S-GeMS tutorial notes. 2007. Disponível em: <<http://people.ku.edu/~gbohling/BoiseGeostat>>.
- BORRADAIRE, G. J. *Statistics of Earth science data*. Springer, 2003.
- BROOKER, P. I. A *geostatistical primer*. World Scientific, 1991.
- CAVA, L. T. (Coord.). *Potencial e perspectivas para o carvão mineral do estado do Paraná*. Paraná: Mineropar, 1985.
- CHILÉS, J. P.; DELFINER, P. *Geostatistics: modeling spatial uncertainty*. John Wiley and Sons, 1999.
- CHORTI, A.; HRISTOPULOS, D. T. Nonparametric identification of anisotropic (elliptic) correlations. *IEEE Trans. Signal Proces.*, v. 56, p. 4.738-4.751, 2008.
- CLARK, I. *Practical geostatistics*. London: Applied Science Publishers Ltd., 1979. 129 p.
- CLARK, I.; HARPER, W. V. *Practical geostatistics 2000*. Geostokos (Ecosse) Limited, 2000.
- CORREIA, P. *Modelação e estimação: uma introdução à geoestatística*. 2010. Disponível em: <<http://numist.ist.utl.pt/index.php?action=media;sa=item;in=81>>.
- CRESSIE, N. A. C. *Statistics for spatial data: revised edition*. John Wiley and Sons, 1993.
- CROZEL, D.; DAVID, M. Global estimation variance: formulas and calculation. *Math. Geology*, v. 17, p. 785-796, 1985.
- DAVID, M. *Geoestatistical ore reserve estimation*. Elsevier, 1977.
- DAVIS, J. C. *Statistics and data analysis in geology*. 3. ed. John Wiley and Sons, 2002.
- DE WIJS, H. J. Statistics of ore distribution. Part I: frequency distribution of assay values: geologie em Mijnbouw. *Journal of the Royal Netherlands Geological and Mining Society, New Series*, v. 13, n. 11, p. 365-375, 1951.
- DE WIJS, H. J. Statistics of ore distribution. Part II: theory of binomial distribution applied to sampling and engineering problems: geologie em Mijnbouw. *Journal of the Royal Netherlands Geological and Mining Society, New Series*, v. 15, n. 1, p. 12-24, 1953.

- DEERE, D. U. et al. Design of surface and near surface construction in rocks. In: FAIRHUST, C. (Ed.). *Failure and breakage of rock*. New York: Society of Mining Engineers of AIME, 1967. p. 237-302.
- DEUTSCH, C. V. Declus: a Fortran 77 program for determining optimal spatial declustering weights. *Comput. and Geosci.*, v. 15, p. 325-332, 1989.
- DEUTSCH, C. V. Correcting for negative weights in ordinary kriging. *Computers and Geosciences*, v. 22, n. 7, p. 765-773, 1996a.
- DEUTSCH, C. V. A sequential indicator simulation program for categorical variables with point and block data: BlockSIS. *Comput. and Geosci.*, v. 32, p. 1.669-1.681, 1996b.
- DEUTSCH, C. V. *Geostatistical reservoir modeling*. New York: Oxford University Press, 2002. 376 p.
- DEUTSCH, C. V. *Guide to best practice in geostatistics*. Draft Version 1.0. Edmonton: Centre for Computational Geostatistics, 2011. 101 p.
- DEUTSCH, C. V.; JOURNEL, A. G. *Geostatistical software library and user's guide – GSLIB*. Oxford University Press, 1992. 340 p.
- DEUTSCH, C. V.; JOURNEL, A. G. *Geostatistical software library and user's guide – GSLIB*. 2. ed. Oxford University Press, 1998.
- DIXON, W. J.; MASSEY, F. J. *Introduction to statistical analysis*. New York: McGraw-Hill, 1957. 488 p.
- DORN, W. S.; MCCracken, D. D. *Cálculo numérico com estudos de casos em Fortran IV*. São Paulo: Edusp, 1978. 568 p.
- DRUCK, S. et al. *Análise espacial de dados geográficos*. Embrapa Cerrados, 2004.
- EMERY, X. Variograms of order Ω : a tool to validate a bivariate distribution model. *Math. Geology*, v. 37, p. 163-181, 2005.
- EMERY, X. Multi-gaussian kriging and simulation in the presence of an uncertain mean value. *Stoch. Environ. Res. Risk Assess.*, v. 24, p. 211-219, 2010.
- FONSECA, J. S.; MARTINS, G. A.; TOLEDO, G. L. *Estatística aplicada*. São Paulo: Editora Atlas S.A., 1991. 267 p.
- FROIDEVAUX, R. Constrained kriging as an estimator of local distribution functions. In: CAPASSO, V.; GIRONE, G.; POSA, D. (Ed.). *Proceedings of the International Workshop on Statistics of Spatial Processes: Theory and Applications*, Bari, Italy, p. 106-118, 1993.
- GOOVAERTS, P. *Geostatistics for natural resources evaluation*. New York: Oxford University Press, 1997. 483 p.
- GROSSMAN, S. I. *Calculus*. Part 1: the calculus of one variable. New York, Academic Press, 1981. 667 p.
- GROSSMAN, S. I. *Calculus*. Part 2: multivariate calculus, linear algebra, and differential equations. New York: Academic Press, 1982. 721 p.
- GUERRA, P. A. G. *Geoestatística operacional*. Departamento Nacional da Produção Mineral, 1988.
- HARBAUGH, J. W.; MERRIAM, D. F. *Computer applications in stratigraphic analysis*. New York: John Wiley and Sons, 1968. 282 p.
- HARDY, R. L. Multiquadric equations of topography and other irregular surfaces. *J. Geophys. Res.*, v. 76, p. 1.905-1.915, 1971.
- HAYES, W. B.; KOCH, G. S. Constructing and analyzing area-of-influence polygons by computer. *Comput. and Geosci.*, v. 10, p. 411-430, 1984.
- HENGL, T.; MINASNY, B.; GOULD, M. A geostatistical analysis of geostatistics. *Scientometrics*, v. 80, n. 2, p. 491-514, 2009.

- HOHN, M. E. *Geostatistics and petroleum geology*. Kluwer Academic Publishers, 1988.
- HOHN, M. E. *Geostatistics and petroleum geology*. Kluwer Academic Publishers, 1999.
- ISAACS, E. H.; SRIVASTAVA, R. M. *An introduction to applied geostatistics*. New York: Oxford University Press, 1989. 561 p.
- JOURNEL, A. G. The lognormal approach to predicting local distributions of selective mining unit grades. *Mathematical Geology*, v. 12, n. 4, p. 285-303, 1980.
- JOURNEL, A. G. Non-parametric estimation of spatial distribution. *Math. Geology*, v. 15, p. 445-468, 1983.
- JOURNEL, A. G. *Fundamentals of geostatistics in five lessons*. Washington: American Geophysical Union, 1989. 40 p. (Short Course in Geology, v. 8).
- JOURNEL, A. G. Markov models for cross-covariances. *Math. Geol.*, v. 31, n. 8, p. 955-964, 1999.
- JOURNEL, A. G.; HUIJBREGTS, J. C. H. *Mining geostatistics*. New York: Academic Press, 1978. 600 p.
- JOURNEL, A. G.; ROSSI, M. When do we need a trend model in kriging? *Math. Geology*, v. 21, p. 715-739, 1989.
- KADER, G. D.; PERRY, M. Variability for categorical data. *J. of Statistics Education*, v. 15, n. 2, 2007. Disponível em: <<http://www.amstat.org/publications/jse/v15n2/kader.html>>. Acesso em: 22 jun. 2010.
- KOCH, G. S.; LINK, R. F. *Statistical analysis of geological data*. New York: Dover Publications, 1971. v. 1, 375 p.
- KOIKE, K.; MATSUDA, S. Spatial modeling of discontinuous geologic attributes with geotechnical applications. *Eng. Geology*, v. 78, p. 143-161, 2005.
- KRIGE, D. G. A statistical approach of some basic mine valuation problems on the Witwatersrand. *Journal of the Chemical, Metallurgical and Mining Society of South Africa*, p. 119-139, Dec. 1951.
- LANDIM, P. M. B.; SOARES, P. C.; PUMPUTIS, A. Cálculo de reserva de uma jazida de carvão por metodologia geoestatística. *Geociências/Unesp*, v. 2, p. 211-230, 1988.
- LANDIM, P. M. B. *Análise estatística de dados geológicos*. 2. ed. revista e ampliada. São Paulo: Editora Unesp, 2003. 253 p.
- LEUANGTHONG, O.; KHAN, K. D.; DEUTSCH, C. V. *Solved problems in geostatistics*. Hoboken: John Wiley and Sons, 2008. 208 p.
- MADYCH, W. R. Miscellaneous error bounds for multiquadric and related interpolators. *Comput. Math. Applic.*, v. 24, n. 12, p. 121-138, 1992.
- MATHERON, G. *Traité de géostatistique appliquée*, tome I. *Mémoires du Bureau de Recherches Géologiques et Minières*, n. 14, Paris, Editions Technip, 1962.
- MATHERON, G. *Le krigeage*, tome II. *Mémoires du Bureau de Recherches Géologiques et Minières*, n. 24, Paris, Editions B.R.G.M., 1963.
- MATHERON, G. *Les variables régionalisées et leur estimation*. Paris: Masson, 1965.
- MATHERON, G. Kriging, or polynomial interpolation procedures? A contribution to polemics in mathematical geology. *The Canadian Mining and Metallurgical Bulletin*, v. LXX, p. 240-244, 1967.
- MATHERON, G. The theory of regionalized variables and its applications. *Paris School of Mines Publication*, p. 211, 1971.
- OLEA, R. A. *Optimum mapping techniques using regionalized variable theory*. Kansas: Kansas Geol. Survey, 1975. 137 p. (Series on spatial analysis, 2).



- OLEA, R. A. *Geostatistics for engineers and Earth scientists*. Boston: Kluwer Academic Publishers, 1999. 303 p.
- OLEA, R. A. Declustering of clustered preferential sampling for histogram and semivariogram inference. *Math. Geol.*, v. 39, p. 453-467, 2007.
- OLEA, R. A. A practical primer on geostatistics. *Open-File Report*, 2009-1103. U.S. Geological Survey, 2009. 346 p.
- OLIVER, M. A. (Ed.). *Geostatistical applications for precision agriculture*. Springer, 2010.
- OZ, B. et al. DSSIM-HR: a Fortran 90 program for direct sequential simulation with histogram reproduction. *Computers and Geosciences*, v. 29, n. 1, p. 39-51, 2003.
- PETERS, W. C. *Exploration and mining geology*. New York: John Wiley and Sons, 1978. 696 p.
- PYRCZ, M. P.; DEUTSCH, C. V. *Declustering and debiasing*. 2003. Disponível em: <<http://www.gaa.org.au/pdf/DeclusterDebias-CCG.pdf>>. Acesso em: jun. 2011.
- RAO, S. E.; JOURNEL, A. G. Deriving conditional distributions from ordinary kriging. In: BAAFI, E. Y.; SCHOFIELD, N. A. *Geostatistics Wollongong '96*. 1997. v. 1, p. 92-102.
- REMY, N.; BOUCHER, A.; WU, J. *Applied geostatistics with SGeMS: a user's guide*. Cambridge University Press, 2009.
- RENDU, J.-M. *An introduction to geostatistical methods of mineral evaluation (Geostatistics)*. 2. ed. South African Institute of Mining and Metallurgy, 1981.
- REZAEI, H.; ASGHARI, O.; YAMAMOTO, J. K. On the reduction of the ordinary kriging smoothing effect. *Journal of Mining and Environment*, v. 2, n. 2, p. 25-40, 2011.
- RIVOIRARD, J. A review of lognormal estimator for in situ reserves. *Math. Geol.*, v. 22, p. 213-221, 1990.
- ROTH, C. R. Is lognormal kriging suitable for local estimation? *Math. Geol.*, v. 30, p. 999-1.009, 1998.
- SAMPER-CALVETE, F. J.; CARRERA-RAMÍREZ, J. *Geoestadística: aplicaciones a la hidrogeología subterránea*. Centro Internacional de Métodos Numéricos em Ingenieria, Universitat Politècnica de Catalunya, 1996.
- SOARES, A. Sequential indicator simulation with correction for local probabilities. *Math. Geol.*, v. 30, n. 6, p. 761-765, 1998.
- SOARES, A. Direct sequential simulation and cosimulation. *Math. Geol.*, v. 33, p. 911-926, 2001.
- SOARES, A. *Geoestatística para as ciências da terra e do ambiente*. 2. ed. Lisboa: IST Press, 2006.
- STEVENS, S. S. On the theory of scales of measurement. *Science*, v. 103, n. 2.684, p. 677-680, 1946.
- TENG, Y.; KOIKE, K. Three-dimensional imaging of a geothermal system using temperature and geological models derived from a well-log dataset. *Geothermics*, v. 36, p. 518-538, 2007.
- TIPPER, J. C. Fortran programs to construct the planar Voronoi diagram. *Comput. and Geosci.*, v. 17, p. 597-632, 1991.
- USEPA. GEO-EAS 1.2.1: geostatistical environmental assessment software user's guide. Publication n. EPA/600/4-88/033. 1991. Disponível em: <www.epa.gov/ada/csmos/models/geoeas.html>.
- VALENTE, J. M. G. P. *Geomatematica: lições de geoestatística*. Ed. Fundação Gorceix, 1982. v. I-VIII.
- WACKERNAGEL, H. *Multivariate geostatistics*. Springer-Verlag, 1995.
- WACKERNAGEL, H. *Multivariate geostatistics: an introduction with applications*. 3. ed. Springer-Verlag, 2003.
- WALTHAM, D. *Mathematics: a simple tool for geologists*. 2. ed. Blackwell Science, 2000.

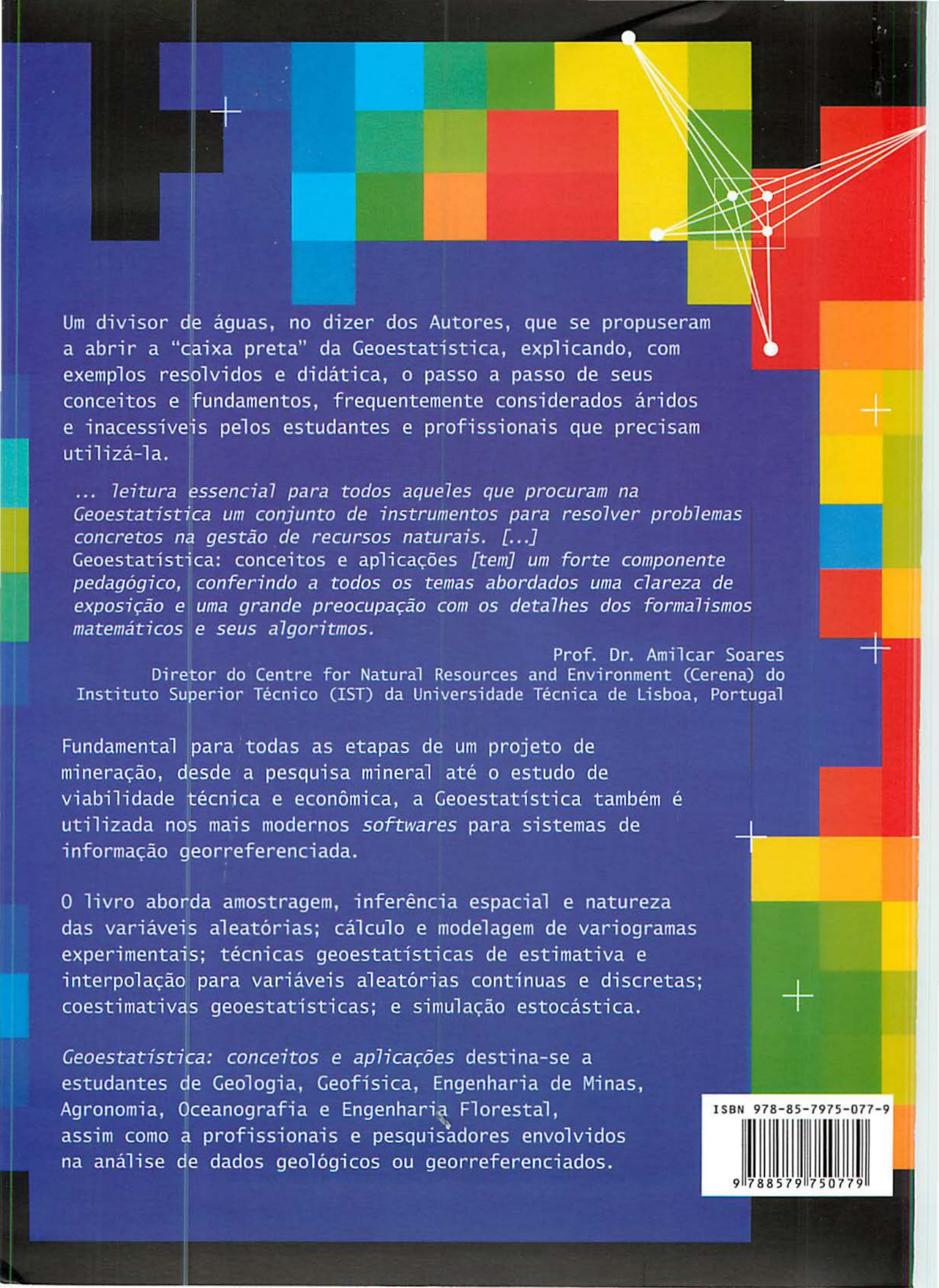
- WATANABE, J. *Métodos geoestatísticos de coestimativas: estudo do efeito de correlação entre variáveis na precisão dos resultados*. 79 f. Dissertação (Mestrado) – Instituto de Geociências/USP, São Paulo, 2008.
- WATANABE, J. et al. Estudo da influência da correlação inicial entre variáveis nos resultados de coestimativas. *Geociências*, v. 28, p. 1-10, 2009.
- WEBSTER, R.; OLIVER, M. A. *Geostatistics for environmental scientists*. 2. ed. Wiley, 2007.
- WHITTEN, E. H. T. The general linear equation in prediction of gold content in Witwatersrand rocks, South Africa. *Symposium on mathematical statistics and computer applications in ore valuation*, special issue of jour., South African Institute of Mining and Metallurgy, p. 124-147, 1966.
- XU, W. et al. Integrating seismic data in reservoir modeling: the collocated cokriging alternative. Paper SPE 24741. *Proceedings of the 67th Annual Technical Conference of the Society of Petroleum Engineers*, Washington, p. 833-842, 1992.
- YAMAMOTO, J. K. An alternative measure of the reliability of ordinary kriging estimates. *Mathematical Geology*, v. 32, p. 337-340, 2000.
- YAMAMOTO, J. K. (Ed.). *Avaliação e classificação de reservas minerais*. São Paulo: Edusp, 2001a.
- YAMAMOTO, J. K. Métodos convencionais. In: YAMAMOTO, J. K. (Org.). *Avaliação e classificação de reservas minerais*. São Paulo: Edusp, 2001b. p. 93-121.
- YAMAMOTO, J. K. Computation of global estimation variance in mineral deposits. In: XIE, H.; WANG, Y.; JIANG, Y. (Org.). *Computer applications in the minerals industries*. A. A. Balkema Publishers, 2001c. v. 1, p. 61-65.
- YAMAMOTO, J. K. Ore reserve estimation using radial basis functions. *Rev. do Instituto Geológico*, v. 23, n. 1, p. 25-38, 2002.
- YAMAMOTO, J. K. Correcting the smoothing effect of ordinary kriging estimates. *Mathematical Geology*, v. 37, n. 1, p. 69-94, 2005.
- YAMAMOTO, J. K. On unbiased backtransform of lognormal kriging estimates. *Computational Geosciences*, v. 11, p. 219-234, 2007.
- YAMAMOTO, J. K. Backtransforming rank order kriging estimates. *Geologia/USP, Série científica*, v. 10, p. 101-115, 2010a.
- YAMAMOTO, J. K. Cálculo de mapas de probabilidade diretamente dos pesos da krigagem ordinária. *Geologia/USP, Série científica*, v. 10, p. 3-14, 2010b.
- YAMAMOTO, J. K.; CHAO, L. Comparação de métodos para teste de bigaussianidade. *Geociências/Unesp*, v. 28, n. 2, p. 21-128, 2009.
- YAMAMOTO, J. K.; FURUIE, R. A. Um estudo sobre estimativa de dados lognormais. *Geociências/Unesp*, v. 29, n. 1, p. 5-19. 2010.
- YAMAMOTO, J. K. et al. Mapping an uncertainty zone between interpolated types of a categorical variable. *Computers and Geosciences*, v. 40, p. 146-152, 2012.



Sobre os autores

Paulo M. Barbosa Landim formou-se em Geologia pela Universidade de São Paulo em 1961, onde também concluiu o Doutorado em Estratigrafia. Fez toda a sua carreira acadêmica na Universidade Estadual Paulista (Unesp), em Rio Claro, onde se aposentou como professor titular em 1998 e, desde 1999, atua como professor voluntário do Departamento de Geologia Aplicada. Em 2000, recebeu do Conselho Universitário da Unesp o título de Professor Emérito.

Jorge K. Yamamoto graduou-se em Geologia pela Universidade de São Paulo em 1976. Trabalhou como geólogo pesquisador do Instituto de Pesquisas Tecnológicas (IPT) de 1977 a 1988. Atualmente é professor titular do Departamento de Geologia Sedimentar e Ambiental, onde ministra disciplinas de graduação e pós-graduação em Geoestatística.



Um divisor de águas, no dizer dos Autores, que se propuseram a abrir a "caixa preta" da Geoestatística, explicando, com exemplos resolvidos e didática, o passo a passo de seus conceitos e fundamentos, frequentemente considerados áridos e inacessíveis pelos estudantes e profissionais que precisam utilizá-la.

... leitura essencial para todos aqueles que procuram na Geoestatística um conjunto de instrumentos para resolver problemas concretos na gestão de recursos naturais. [...]

Geoestatística: conceitos e aplicações [tem] um forte componente pedagógico, conferindo a todos os temas abordados uma clareza de exposição e uma grande preocupação com os detalhes dos formalismos matemáticos e seus algoritmos.

Prof. Dr. Amílcar Soares

Diretor do Centre for Natural Resources and Environment (Ceren) do Instituto Superior Técnico (IST) da Universidade Técnica de Lisboa, Portugal

Fundamental para todas as etapas de um projeto de mineração, desde a pesquisa mineral até o estudo de viabilidade técnica e econômica, a Geoestatística também é utilizada nos mais modernos *softwares* para sistemas de informação georreferenciada.

O livro aborda amostragem, inferência espacial e natureza das variáveis aleatórias; cálculo e modelagem de variogramas experimentais; técnicas geoestatísticas de estimativa e interpolação para variáveis aleatórias contínuas e discretas; coestimativas geoestatísticas; e simulação estocástica.

Geoestatística: conceitos e aplicações destina-se a estudantes de Geologia, Geofísica, Engenharia de Minas, Agronomia, Oceanografia e Engenharia Florestal, assim como a profissionais e pesquisadores envolvidos na análise de dados geológicos ou georreferenciados.

ISBN 978-85-7975-077-9



9 788579 750779