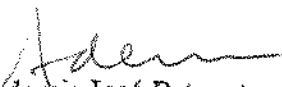


GEOESTATÍSTICA MULTIVARIADA: ESTUDO DE MÉTODOS DE PREDIÇÃO

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pela Srta. Emília Tieko Uzumaki e aprovada pela Comissão Julgadora.

Campinas, 25 de Novembro de 1994


Prof. Dr. Ademar José Petenate
Orientador

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência da Computação, UNICAMP, como requisito parcial para obtenção do Título de MESTRE em Estatística.

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E CIÊNCIA DA COMPUTAÇÃO
DEPARTAMENTO DE ESTATÍSTICA

**GEOESTATÍSTICA MULTIVARIADA:
ESTUDO DE MÉTODOS DE PREDIÇÃO**

Emília Tiekō Uzumaki *E.T.*

Prof. Dr. Ademir José Petenate *A.P.*
Orientador

Prof. Dr. Saul Barisnik Suslick *S.B.*
Co-orientador

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência da Computação da
Universidade Estadual de Campinas para obtenção do título de Mestre em Estatística

Campinas - SP
1994

BANCA EXAMINADORA:

PROF. DR. ADEMIR JOSÉ PETENATE (ORIENTADOR) - IMECC - UNICAMP

PROF. DR. SAUL BARISNIK SUSLICK (CO-ORIENTADOR) - IG - UNICAMP

PROF. DR. VALDÉRIO ANSELMO REISEN - UFES

PROFA. DRA. REGINA C. C. P. MORAN (SUPLENTE) - IMECC - UNICAMP

Aos meus pais,
Yoshitaka e Satiko,
dedico.

Agradecimentos

Ao Prof. Dr. Ademir José Petenate, pelas sugestões e confiança.

Ao Prof. Dr. Saul Barisnik Suslick, pelo tema e apoio.

Ao Prof. Dr. Armando Zauppa Remacre, pelo auxílio.

Aos professores do Departamento de Estatística, pelos ensinamentos.

Ao CNPq, FAEP, FAPESP e UNICAMP, pelo apoio financeiro.

Aos funcionários do IMECC, pelo atendimento e atenção.

Aos colegas e amigos, pelo convívio.

Ao Raul, pelo carinho.

Ao João, pelas sugestões na correção do texto.

À "tia" Satomi, pela compreensão.

Aos meus pais, pela confiança e apoio.

A Deus, por tudo.

Resumo

Os *Métodos Geoestatísticos* conseguem juntar o aspecto espacial (topológico) com o aspecto aleatório (probabilístico) das variáveis regionalizadas. Existem muitos métodos geoestatísticos univariados de análise, mas a literatura em geoestatística multivariada não é muito extensa.

O objetivo deste trabalho foi realizar um estudo de métodos de predição geoestatísticos multivariados, apresentando-os, na medida do possível, de maneira simples.

Para predição de variáveis regionalizadas, fizemos um estudo sobre a *Cokrigagem*. Uma alternativa é realizar a predição com o auxílio da técnica de *Análise de Componentes Principais* (no caso de suposição de correlação intrínseca), proposta por Davis & Greenes (1983).

Para incorporar a característica espacial do fenômeno na análise, podemos estudar a decomposição de uma variável regionalizada em diferentes estruturas espaciais. Isso é possível através da *Análise de Krigagem Fatorial*, proposta por Matheron (1982), e pela *Análise baseada no Variograma Multivariado*, proposta por Bourgault & Marcotte (1991). Como ilustração, foi apresentado também a *Análise Fatorial Espacial*, proposta por Grunsky & Agterberg (1988).

Para ilustrar as técnicas de Krigagem, de Cokrigagem e de Análise de Componentes Principais na estimação de uma variável regionalizada, realizamos uma comparação através de um conjunto de dados reais.

Abstract

Geostatistical methods are able to join spatial (topologic) feature with random (probabilistic) feature of the regionalized variables. There are several univariate methods of analysis in geostatistical literature, but the multivariate ones are not covered extensively.

This work's purpose was to present briefly some multivariate prediction methods and to perform some comparison between them.

To make predictions of the regionalized variables, we studied the Cokriging. In the intrinsic correlation case, Principal Component Analysis (Davis & Greenes, 1983) was combined with Kriging as an alternative way.

We studied a regionalized variable decomposition in different spatial structures to incorporate spatial characteristic of phenomenon in the analysis. This was possible through Factorial Kriging Analysis (Matheron, 1982) and Multivariable Variogram (Bourgault & Marcotte, 1991). As illustration we presented also the Spatial Factor Analysis (Grunsky & Agterberg, 1988).

To illustrate the Kriging, Cokriging and Principal Component Analysis techniques of regionalized variables predictions, we performed a comparison between them through a real data set.

Sumário

Capítulo 1. Introdução	1
Capítulo 2. Princípios de Geoestatística	6
2.1. Definições Básicas	7
2.1.1. Propriedades da Covariância e do Variograma	10
2.1.2. Modelagem da Variabilidade Espacial do Processo.....	11
2.1.3. Variograma Fatorado.....	15
2.1.4. Modelo de Anisotropia	16
2.1.5. Variograma Experimental	17
2.2. Krigagem	20
2.3. Validação Cruzada	23
2.4. Características Multivariadas.....	25
2.4.1. Correionalização	25
2.4.2. Análise de Correionalização	26
2.4.3. Modelo Linear de Correionalização.....	27
2.4.4. Aproximação Geoestatística.....	31
Capítulo 3. Cokrigagem	32
3.1. Equações de Cokrigagem.....	32
3.2. Particularidades dos Cálculos	36
3.3. Exemplo de Cokrigagem	38
Capítulo 4. Métodos Alternativos de Predição	43
4.1. Análise de Componentes Principais (ACP)	44
4.2. Análise de Krigagem Fatorial (AKF)	49
4.2.1. Estimacão de Componentes Espaciais	51
4.2.2. Estimacão de Fatores Regionalizados.....	52
4.2.2.1. ACP na Correionalização.....	54
4.2.2.2. Função Discriminante na Correionalização.....	55
4.2.3. Exemplos de Análise de Krigagem Fatorial.....	56
4.3. Análise Baseada no Variograma Multivariado	60

4.4. Análise Fatorial Espacial	68
4.5. Programas Disponíveis	69
Capítulo 5. Aplicação dos Métodos de Predição	70
5.1. Descrição dos Dados.....	71
5.2. Aplicação dos Métodos de Predição.....	74
5.2.1. Krigagem.....	75
5.2.2. Cokrigagem	78
5.2.3. Análise de Componentes Principais	82
5.3. Nota sobre os Cálculos.....	85
5.4. Comparação e Comentários.....	86
Capítulo 6. Considerações Finais.....	88
Apêndice A. Conjunto de Dados Reais	91
Apêndice B. Programas Utilizados	96
Referências Bibliográficas.....	99

Capítulo 1

Introdução

A oportunidade da aplicação de Geoestatística em problemas com dados distribuídos espacialmente abriu a possibilidade de sua utilização em diversos domínios das ciências da natureza. Na definição de Matheron (1962, citado por Journel & Huijbregts, 1978, pp. 1), "*Geostatistics is the application of the formalism of random functions to the reconnaissance and estimation of natural phenomena*" ("Geoestatística é a aplicação do formalismo de funções aleatórias ao reconhecimento e predição de fenômenos naturais").

Geralmente podemos caracterizar um *fenômeno natural* pela distribuição, no espaço, de uma ou mais variáveis, denominadas *variáveis regionalizadas* (seção 2.1). Variável regionalizada pode ser entendida como uma função que varia, no espaço, com certa aparência de continuidade, ou seja, é uma variável cujos valores são relacionados (de algum modo) com a posição espacial que ocupam.

Neste trabalho consideraremos que a variável regionalizada é uma realização de um processo estocástico de segunda ordem (veja seção 2.1). Assim sendo, a variável regionalizada não é uma nova teoria, ela se baseia em uma teoria da Estatística (no caso, Processo Estocástico). O "nome" variável regionalizada surge apenas para enfatizar que estamos trabalhando com variáveis aleatórias que consideram aspectos espaciais (ou seja, as posições relativas onde foram observados os diversos valores devem ser levadas em conta no modelo). As variáveis regionalizadas possuem também características qualitativas estreitamente ligadas à estrutura do fenômeno natural que elas representam. Tais características são: a localização, a continuidade (espacial) e a anisotropia (varia com a direção; maiores detalhes na seção 2.1.4).

Em Geoestatística, decidiu-se formalizar a "teoria" de variáveis regionalizadas utilizando Processos Estocásticos devido a duas características principais que limitam a modelagem de fenômenos naturais por variáveis aleatórias usualmente utilizadas (isto é, modelos que utilizam variáveis aleatórias independentes e identicamente distribuídas, não considerando assim, o aspecto espacial). Uma é a impossibilidade de se repetir indefinidamente o experimento, mesmo porque na grande parte dos casos tratam-se de variáveis observadas, tais como teor de um certo minério em uma jazida, ou precipitação de chuva em um determinado dia. (Nestes casos, teor e precipitação, é impossível reamostrar

um bloco de minério que já foi processado ou voltar atrás no tempo e fazer uma nova medição da precipitação.) A outra característica refere-se à dependência das observações: em contraste com os exemplos clássicos da teoria de probabilidade sobre independência (isto é, lançamentos de dado ou de moeda), as medidas observadas são extremamente dependentes. Em dados de mineração, por exemplo, duas sondagens do mesmo minério feitas em pontos próximos podem apresentar resultados semelhantes ou serem relacionadas.

Na estimação de variáveis regionalizadas (por exemplo, teor de minério), os *Métodos Convencionais* (métodos não estatísticos utilizados na geologia) levam em conta apenas o aspecto espacial das amostras, e na maioria deles é dada ênfase ao conceito de "área de influência", que é normalmente determinada empiricamente, com base em apreciações pessoais ou simplesmente de acordo com a disposição do espaçamento das amostras obtidas. (Alguns destes métodos são descritos por Guerra, 1988, cap. 2.) Estes métodos supõem uma longa experiência no fenômeno de interesse; assim, a aplicação destas técnicas é baseada na "intuição" do pesquisador (geólogo, engenheiro de minas, etc...), bem como em todos os procedimentos clássicos da geologia mineira. Mas estes métodos não levam em conta outras características importantes do fenômeno, tais como variabilidade ou dispersão. A essência de qualquer bom procedimento de cálculo não é simplesmente produzir um número pontual, mas também alguma indicação de como o valor estimado pode variar em relação ao verdadeiro valor. Sem alguma medida de precisão da estimativa, o valor estimado é duvidoso, podendo não ter maior significado.

Já os *Métodos Geoestatísticos* conseguem juntar o aspecto espacial (topológico) com o aspecto aleatório (probabilístico). Estes métodos se baseiam na "teoria" das variáveis regionalizadas, a partir da qual é possível estudar a estrutura espacial (que quase sempre influi no valor definitivo associado a cada ponto do espaço). Estes métodos nos permitem calcular o erro cometido na avaliação levando em conta a posição real das amostras. Surgiu então, a necessidade de se encontrar novas ferramentas matemáticas que permitam estudar, de modo sintético, as duas características essenciais das variáveis regionalizadas (aspecto aleatório e aspecto espacial), ou seja, extrair da aparente desordem dos dados disponíveis uma imagem da variabilidade dos mesmos e também uma medida da correlação existente entre os valores tomados em dois pontos do espaço. Este é o objetivo da análise estrutural que pode ser alcançado, por exemplo, através do *variograma*. O variograma, em termos práticos, é uma ferramenta matemática que nos permite estudar a dispersão natural das variáveis regionalizadas. Ele indica o grau de continuidade espacial e nos dá um mínimo necessário de informação sobre o fenômeno a ser estudado. Entretanto, geralmente a modelagem e a interpretação do variograma não são simples. Ambas exigem conhecimentos

do fenômeno e da maneira que foi realizada a coleta das amostras para que se possa analisar adequadamente os dados. (Informações mais detalhadas poderão ser vistas na seção 2.1.5.)

Existem muitos métodos geoestatísticos univariados, sendo o método da *Krigagem* (seção 2.2) a mais utilizada. Porém, a literatura em Geoestatística Multivariada não é muito extensa. (Existem várias razões para que isto ocorra, por exemplo, dificuldades ao se trabalhar com mais de uma variável e ganhos de precisão nem sempre significativos em relação às técnicas univariadas.)

O objetivo deste trabalho foi realizar um estudo de métodos de predição multivariados, apresentando-os, na medida do possível, de maneira simples. Muitas técnicas ainda são recentes, e algumas são inviáveis pelo excesso de trabalho e/ou consumo de tempo. Por isso, buscam-se métodos mais eficientes que contornem tais problemas.

Note que Hohn (1991, pp. 139) cita três situações onde é importante considerarmos mais de uma variável: "1. A variável de interesse é uma combinação linear de variáveis regionalizadas. Estimacão direta da combinação linear geralmente não é ótima. 2. Uma variável é subamostrada, mas é correlacionada altamente com uma segunda variável que é muito melhor amostrada. Podemos tirar vantagem dessa correlacão para melhorar a estimacão da variável subamostrada. 3. Uma variável apresenta baixa autocorrelacão espacial, mas é altamente correlacionada com uma que apresenta continuidade relativamente alta. Novamente, os valores observados da segunda variável podem ajudar a melhorar a predição da primeira variável, particularmente se a primeira foi subamostrada".

No **capítulo 2** fazemos uma revisão dos conceitos básicos de Geoestatística (seção 2.1), de *Krigagem* (seção 2.2) e de *Validacão Cruzada* (seção 2.3). Algumas características multivariadas também são revistas (seção 2.4).

No **Capítulo 3** estudamos a *Cokrigagem*, a técnica multivariada mais conhecida e utilizada. São discutidas as vantagens e desvantagens desta técnica (seção 3.2), que se caracteriza pela estimacão de uma variável regionalizada através de duas ou mais variáveis. No entanto, a correlacão espacial entre as variáveis deve ser "suficientemente forte" para que a variável adicional acrescente algum ganho de informacão que sirva para a predição da variável de interesse. Ilustramos esta técnica através de um exemplo simples, e citamos artigos onde ela foi utilizada em problemas reais (seção 3.3).

No **Capítulo 4** apresentamos algumas técnicas multivariadas alternativas à *Cokrigagem*. Uma delas é a *Análise de Componentes Principais* (ACP; seção 4.1), que é utilizada para a obtenção de uma matriz de dados que não possui correlacão cruzada espacial, evitando, assim, os cálculos dos semivariogramas cruzados. Esta técnica, proposta

por Davis & Greenes (1983), é simples porém restrita, pois exige uma hipótese muito forte de *correlação intrínseca*, ou seja, exige que o conjunto de variáveis tenha a mesma estrutura de correlação em todas as escalas espaciais.

Há também um conjunto de técnicas, reunidas sobre o nome de *Análise de Krigagem Fatorial* (AKF, seção 4.2), que torna possível incorporar a característica espacial do fenômeno na análise. A AKF, desenvolvido por Matheron (1982), permite a decomposição de uma variável regionalizada em diferentes *componentes espaciais* (seção 4.2.1) ou *fatores regionalizados* (seção 4.2.2). Na estimação de fatores regionalizados, as matrizes de coregionalização B_{ii} (matrizes simétricas positivas semidefinidas) são fatorizadas. Dependendo de como isso será feito, surgem vários tipos de estudo. A ACP (seção 4.2.2.1) e a *função discriminante* (seção 4.2.2.2) são apenas duas possibilidades.

A *Análise baseada no Variograma Multivariado* (seção 4.3) é uma outra maneira de estimar fatores regionalizados. Bourgault & Marcotte (1991) mostraram as vantagens desta técnica, que utiliza o variograma multivariado para realçar as estruturas espaciais comuns ao conjunto de variáveis.

Apenas como ilustração, apresentamos também as idéias básicas do método proposto por Grunsky & Agterberg (1988), denominado *Análise Fatorial Espacial* (seção 4.4).

Como ilustração, no **Capítulo 5** utilizamos três técnicas, todas elas em um mesmo conjunto de dados de mineração: Krigagem, Cokrigagem e ACP baseada na matriz V . (Como existem poucas amostras disponíveis, a confiabilidade das estimações é muito reduzida.) Mostramos, através deste conjunto de dados, algumas das possíveis dificuldades e complicações que podem surgir ao trabalharmos com mineração, onde dispomos, por causa do custo elevado e dificuldades técnicas, de poucas amostras (que são, geralmente, coletadas com um espaçamento irregular). Dentro das limitações de cada técnica utilizada, tentamos obter o modelo que melhor representasse os dados. (Observação: não foi permitida a divulgação da fonte e das características que permitissem a identificação dos dados, e por isso as análises descritas neste capítulo são apenas ilustrativas.)

No **Capítulo 6** comentamos as técnicas apresentadas e possíveis estudos futuros.

No **Apêndice A** pode ser vista a listagem do conjunto de dados utilizado no Capítulo 5, assim como suas estimativas. Os programas (menos conhecidos) utilizados para a análise podem ser vistos no **Apêndice B**.

Observações:

1. Durante o trabalho, por simplicidade utilizamos indistintamente os termos "predição" e "estimação". Deve-se lembrar, entretanto, que as duas palavras serão utilizadas para indicar que estamos "estimando" (ou "predizendo") uma observação, e não um parâmetro (como média, variância, etc). Esta terminologia concorda, por exemplo, com a terminologia utilizada em Séries Temporais, onde podemos calcular observações futuras, mas também observações passadas, ou observações *missing*.
2. Os estimadores apresentados neste trabalho são todos estimadores BLUP (*best linear unbiased predictor*), ou seja, estimadores lineares não viciados de mínima variância.
3. Chamamos de estimadores BLUP ao invés de estimadores BLUE (*best linear unbiased estimator*) porque a primeira terminologia costuma ser empregada em estimações de uma observação, enquanto a segunda de um parâmetro.
4. Os índices das locações amostrais serão representadas por letras gregas: α e β .
5. Os índices das variáveis serão representadas pelas letras: i , j e k .

Capítulo 2

Princípios de Geoestatística

Variável regionalizada pode ser entendida como uma função que varia, no espaço, com certa aparência de continuidade, ou seja, é uma variável cujos valores são relacionados (de algum modo) com a posição espacial que ocupam. De uma maneira mais formal, podemos interpretar a variável regionalizada $z(x)$ como uma realização de uma função aleatória $Z(x)$.

Neste trabalho, $Z(x)$ será considerada como um processo estocástico de segunda ordem. Na seção 2.1, apresentamos alguns conceitos de Processos Estocásticos. Uma revisão mais detalhada pode ser vista, por exemplo, em Hoel, Port & Stone (1972).

As correlações espaciais da função aleatória $Z(x)$ são caracterizadas pela função de covariância ou, equivalentemente, pelo variograma. As propriedades da covariância e do variograma são apresentadas na seção 2.1.1, e a modelagem, na seção 2.1.2. Algumas das particularidades do variograma são apresentadas nas seções 2.1.3 e 2.1.4. O cálculo do variograma experimental pode ser visto na seção 2.1.5. Na seção 2.2 apresentamos o método da Krigagem (técnica de predição univariada). E em 2.3, a Validação Cruzada (*Cross-Validation*). Uma revisão mais detalhada sobre a Geoestatística Univariada pode ser vista em Journel & Huijbregts (1978) e Delfiner (1979).

Na seção 2.4, apresentamos algumas características multivariadas. Entre as definições, encontramos a de correionalização, a de variograma cruzado e a de covariância cruzada (seção 2.4.1). Apresentamos a análise de correionalização (seção 2.4.2), o modelo linear de correionalização (seção 2.4.3) e uma aproximação geoestatística para o cálculo do variograma cruzado experimental (seção 2.4.4).

Observação: É importante enfatizar que antes de procedermos com os métodos geoestatísticos, devemos fazer sempre uma análise exploratória dos dados (que é também o nosso primeiro passo em uma análise estatística).

2.1. Definições Básicas

Definição 1. Um *processo estocástico* em um espaço de probabilidade $(\Omega, \mathfrak{F}, P)$ é uma coleção de variáveis aleatórias indexadas por um conjunto não vazio X , isto é, $\{Z(x), x \in X\}$. •

Obs.: Neste trabalho, o conjunto X será o espaço real p -dimensional, isto é, $X = \mathfrak{R}^p$.

Um processo estocástico é caracterizado pela distribuição conjunta de qualquer conjunto de variáveis aleatórias $Z(x_1), Z(x_2), \dots, Z(x_n)$ para todo inteiro n e $x_\alpha \in \mathfrak{R}^p$, $\alpha = 1, \dots, n$. Para um $x \in \mathfrak{R}^p$ fixo, $Z(x)$ é uma variável aleatória.

Seja $x \in \mathfrak{R}^p$. Definimos a *média do processo estocástico*, $m(x)$, como

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

quando a esperança existe.

Dado $(x_1, x_2) \in \mathfrak{R}^p \times \mathfrak{R}^p$, a *função de covariância do processo estocástico*, $C(x_1, x_2)$, é definida por

$$C(x_1, x_2) = \text{cov}[Z(x_1), Z(x_2)] = E\{[Z(x_1) - m(x_1)][Z(x_2) - m(x_2)]\},$$

quando a covariância existe. Portanto, para $x \in \mathfrak{R}^p$,

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

A *função variograma do processo estocástico*, $2\gamma(x_1, x_2)$, é definida como a variância do incremento $[Z(x_1) - Z(x_2)]$, ou seja,

$$2\gamma(x_1, x_2) = \text{Var}[Z(x_1) - Z(x_2)].$$

Denominamos $\gamma(x_1, x_2)$ de *semivariograma*.

O semivariograma determina a estrutura espacial do processo estocástico, sendo de grande importância, em Geoestatística, pois leva em conta a variabilidade espacial de tal processo.

Definição 2. Um processo estocástico é dito *estritamente estacionário* se a distribuição conjunta de $\{Z(x_1), Z(x_2), \dots, Z(x_n)\}$ é a mesma de $\{Z(x_1+h), Z(x_2+h), \dots, Z(x_n+h)\}$ para qualquer conjunto finito de pontos $\{x_1, x_2, \dots, x_n\}$ e qualquer vetor $h \in \mathfrak{R}^p$, ou seja, o processo é invariante sob translação no conjunto de índices. •

Note que o tipo de estacionariedade descrita acima não implica na existência da média, variância ou covariância.

Definição 3. Um processo estocástico é dito *estacionário de segunda ordem* (ou covariância estacionária, ou fracamente estacionário) quando

i) a esperança $E[Z(x)]$ existe e não depende do ponto x , isto é,

$$E[Z(x)] = m, \quad \forall x \in \mathfrak{R}^p;$$

ii) $E[Z(x)]^2 < \infty$; e

iii) para cada par de variáveis aleatórias $\{Z(x), Z(x+h)\}$ a função de covariância existe e depende apenas do vetor $h \in \mathfrak{R}^p$:

$$C(x, x+h) = E[Z(x+h)Z(x)] - m^2 = C(h), \quad \forall x \in \mathfrak{R}^p. \quad \bullet$$

A estacionariedade da covariância implica na estacionariedade da variância e do variograma. Seguem-se então as seguintes relações:

$$\text{Var}[Z(x)] = E\{[Z(x) - m]^2\} = C(0), \quad \forall x \in \mathfrak{R}^p, \text{ e}$$

$$\gamma(h) = \frac{1}{2} E\{[Z(x+h) - Z(x)]^2\} = C(0) - C(h), \quad \forall x \in \mathfrak{R}^p, \quad \forall h \in \mathfrak{R}^p.$$

As relações acima indicam que, sob a hipótese de estacionariedade da covariância, a covariância e o variograma são duas ferramentas equivalentes para caracterizar as autocorrelações entre duas variáveis separadas por uma distância h , $Z(x+h)$ e $Z(x)$. Assim, pode ser definida também uma terceira ferramenta, o *correlograma*, $\rho(h)$:

$$\rho(h) = \frac{C(h)}{C(0)} = 1 - \frac{\gamma(h)}{\gamma(0)}, \quad \forall h \in p.$$

Definição 4. Um processo estocástico é dito *intrínseco* quando:

(i) a esperança $E[Z(x)]$ existe e não depende do ponto x , isto é,

$$E[Z(x)] = m, \quad \forall x \in \mathfrak{R}^p;$$

(ii) para todo $h \in \mathfrak{R}^p$, a variância da diferença $[Z(x+h) - Z(x)]$ existe e não depende do ponto x , isto é,

$$Var[Z(x+h) - Z(x)] = E\{[Z(x+h) - Z(x)]^2\} = 2\gamma(h), \quad \forall x \in \mathfrak{R}^p. \quad \bullet$$

A forma de estacionariedade implicada acima é essencialmente estacionariedade de 2ª ordem (ou covariância estacionária), não para o processo estocástico $\{Z(x), x \in \mathfrak{R}^p\}$ mas para a diferença de 1ª ordem $[Z(x+h) - Z(x)]$. Isto é necessário porque muitas vezes a variância de $Z(x)$ não é finita, sendo conveniente, em muitos casos, assumir que os incrementos da função, $[Z(x+h) - Z(x)]$, sejam estacionários de segunda ordem (Lucio, 1991, pp. 6).

Teorema 4.1 Se $\{Z(x), x \in \mathfrak{R}^p\}$ é um processo estocástico estacionário de segunda ordem, então este é intrinsecamente estacionário. •

Dem.:
$$2Cov[Z(x), Z(x+h)] = Var[Z(x)] + Var[Z(x+h)] - Var[Z(x+h) - Z(x)]$$

$$2C(h) = 2C(0) - Var[Z(x+h) - Z(x)]$$

$$2C(h) - 2C(0) = Var[Z(x+h) - Z(x)]$$

$$2\gamma(h) = Var[Z(x+h) - Z(x)]. \quad \bullet$$

Um processo estocástico estacionário de segunda ordem é também intrínseco, mas o inverso nem sempre ocorre. Isto foi exemplificado por Lúcio (1991, pp. 7; e Lana, 1989, pp.5), no caso do processo de Wiener (processo browniano).

Uma hipótese usualmente introduzida à teoria de processos estocásticos é a de estacionariedade. É importante notar que a estacionariedade se refere ao processo, e não aos dados.

Definição 5. Um processo estocástico de segunda ordem é dito *isotrópico* se a função de covariância $C(h)$ associada a ele não depende da direção do vetor $h \in \mathbb{R}^p$, mas apenas de seu comprimento $\|h\|$, isto é, $C(h) = C(\|h\|)$. •

Obs.: Dizemos que uma função de covariância é *anisotrópica* se ela não for isotrópica, ou seja, não exibe sempre o mesmo comportamento em todas as direções.

2.1.1. Propriedades da Covariância e do Variograma

Seja $Z(x)$ uma função aleatória estacionária de esperança m , covariância $C(h)$ e semivariograma $\gamma(h)$:

A covariância é uma função do vetor h (isto é, de seu módulo e de sua direção). São algumas de suas propriedades:

- i) $C(0) = \text{Var}[Z(x)] \geq 0$, isto é, a variância *a priori* não pode ser negativa;
- ii) $C(h) = C(-h)$; e
- iii) $|C(h)| \leq C(0)$ (desigualdade de Schwartz).

Uma condição necessária e suficiente para $C(h)$ expressar a função covariância de $Z(x)$ é que $C(h)$ seja positiva semidefinida.

De igual maneira, o semivariograma

$$\gamma(h) = C(0) - C(h),$$

leva às seguintes propriedades do semivariograma (e, é claro, do variograma):

- i) $\gamma(0) = 0$; e
- ii) $\gamma(h) = \gamma(-h) \geq 0$.

2.1.2. Modelagem da Variabilidade Espacial do Processo

Segundo Journel & Huijbregts (1978, pp. 36), como o nível de correlação espacial entre duas variáveis $Z(x)$ e $Z(x+h)$ geralmente diminui quando a distância h aumenta, então em geral a função de covariância diminuirá o seu valor a partir da origem $C(0)$. De igual maneira, o semivariograma aumentará a partir da origem, $\gamma(0) = 0$.

No estudo de fenômenos naturais, na prática, a correlação espacial entre duas variáveis $Z(x)$ e $Z(x+h)$ geralmente desaparece quando a distância h torna-se muito grande ($C(h) \rightarrow 0$, quando $\|h\| \rightarrow \infty$), e então assumimos que $C(h) = 0$, quando $\|h\| > a$, onde a corresponde à idéia de **área de influência** de uma amostra, sendo determinada pelo pesquisador com base no gráfico da covariância, $C(h)$. A distância a é chamada de **amplitude** (alcance ou "range") e representa a transição em que a correlação espacial existe ($\|h\| < a$) para o estado em que há ausência de correlação ($\|h\| > a$).

Por sua vez, quando h aumenta, o desvio médio quadrático entre duas variáveis $Z(x)$ e $Z(x+h)$ tende a crescer, e assim $\gamma(h)$ aumentará de seu valor inicial, que é zero.

Na prática, muitas vezes o semivariograma pára de crescer depois de uma certa distância e torna-se mais ou menos estável ao redor do valor limite $\gamma(\infty)$ chamado de **patamar** ("sill") (patamar), que é simplesmente a variância *a priori* da função aleatória:

$$\gamma(\infty) = \text{Var}[Z(x)] = C(0) .$$

Em tais casos, a variância *a priori* e a covariância existem. Tais variogramas, que são caracterizados por seu patamar e amplitude, são chamados de "**modelos de transição**". A diferença entre os modelos de transição é expressa por seu comportamento na origem, posto que a modelagem é feita tendo em vista pequenas distâncias.

Para processos estocásticos estacionários de segunda ordem e isotrópicos, os modelos de covariância, $C(r)$, mais utilizados são:

i) Modelo esférico:

Modelo com comportamento linear próximo à origem (exemplo na Figura 2.1.2.1):

$$C(r) = \begin{cases} c \left[1 - \frac{3}{2a}r + \frac{1}{2a^3}r^3 \right] & 0 \leq r \leq a \\ 0 & r > a \end{cases} , \quad (2.1.2.1)$$

onde: $r = \|h\|$; e $c = \sigma^2 = \text{Var}[Z(x)]$.

ii) Modelo exponencial:

Modelo com comportamento linear próximo à origem, onde a amplitude efetiva é $3a$ (exemplo na Figura 2.1.2.2):

$$C(r) = c \exp\left(-\frac{1}{a}r\right), \quad a > 0. \quad (2.1.2.2)$$

iii) Modelo gaussiano:

Modelo com comportamento parabólico próximo à origem, onde a amplitude efetiva é $a\sqrt{3}$ (exemplo na Figura 2.1.2.3)

$$C(r) = c \exp\left(-\frac{1}{a^2}r^2\right), \quad a > 0. \quad (2.1.2.3)$$

iv) Modelo Efeito Pepita

O *efeito pepita* é utilizado para representar uma descontinuidade na origem de um variograma, e pode ser explicado fisicamente (por medida de erro ou por micro-estruturas). As micro-estruturas são estruturas espaciais abaixo da escala definida pelo espaço amostral.

O *efeito pepita puro* é o caso limite quando $\gamma(h)$ aparece exclusivamente como uma descontinuidade na origem (exemplo na Figura 2.1.2.4):

$$C_0(r) = \begin{cases} C(0) = C_0 & r = 0 \\ 0 & \forall |r| > \varepsilon \end{cases}, \quad (2.1.2.4)$$

ou, equivalentemente, o semivariograma é dado por

$$\gamma_0(r) = \begin{cases} 0 & r = 0 \\ C_0 & \forall |r| > \varepsilon \end{cases}, \quad (2.1.2.5)$$

onde $\varepsilon > 0$ é tão pequeno quanto se queira.

O efeito pepita puro corresponde, assim, a uma total ausência de correlação espacial entre duas variáveis $Z(x)$ e $Z(x+h)$ para qualquer distância h ($|h| > \varepsilon$) e é equivalente, na física, ao bem conhecido fenômeno de "ruído branco". Um efeito pepita puro será interpretado como *independência espacial*.

Para modelar o semivariograma, o pacote em Fortran GSLIB (Deutsch & Journel, 1992) apresenta vários modelos básicos possíveis. Em Tough & Leyshon (1985) podem ser vistos programas para o modelo esférico e exponencial.

Obs.: Os modelos teóricos de variogramas que existem na literatura não surgiram espontaneamente e nem foram estabelecidos de forma arbitrária; eles correspondem a diferentes regionalizações que encontramos na natureza. Por exemplo, o modelo esférico (item i) é um dos mais utilizados na indústria mineira.

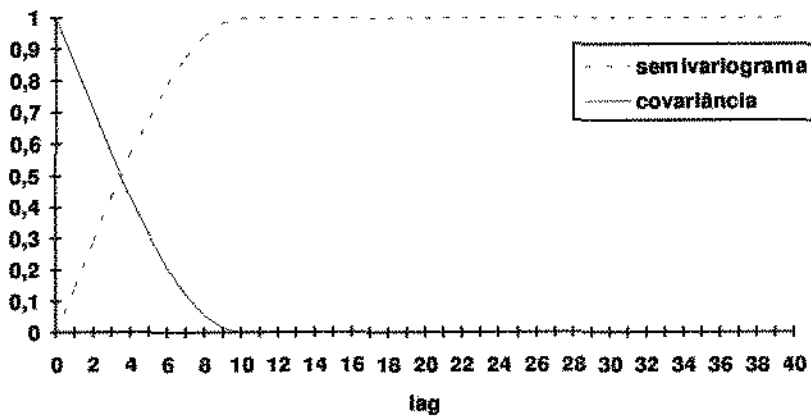


Figura 2.1.2.1. Exemplo de modelo esférico com $c = 1$ e $a = 10$.

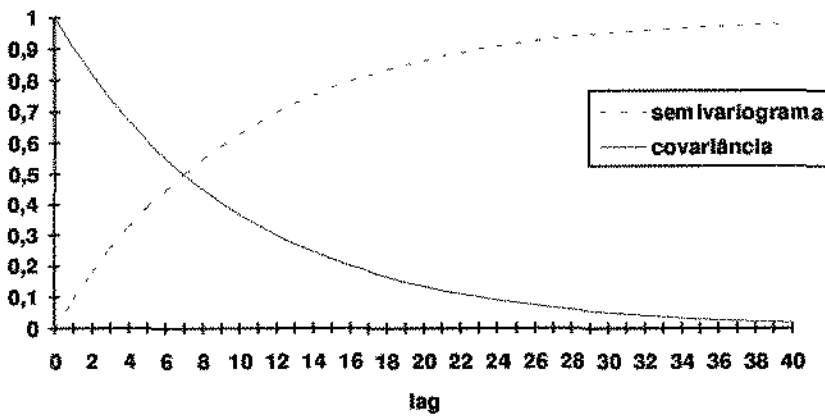


Figura 2.1.2.2. Exemplo de modelo exponencial com $c = 1$, $a = 10$ e "a efetivo" = $3a$.

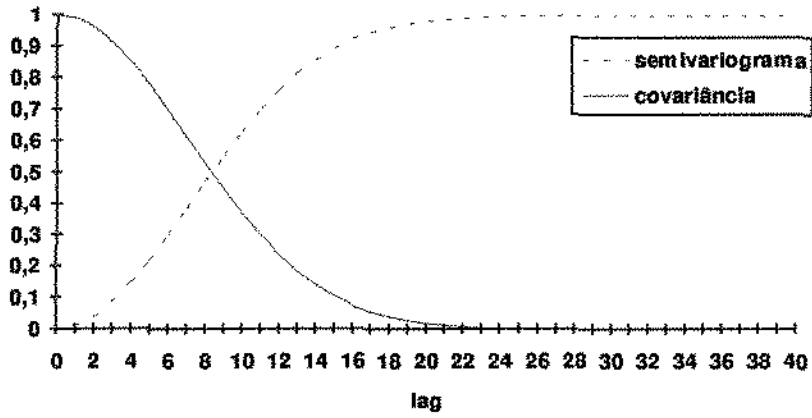


Figura 2.1.2.3. Exemplo de modelo gaussiano com $c = 1$, $a = 10$ e "a efetivo" = $a\sqrt{3}$.

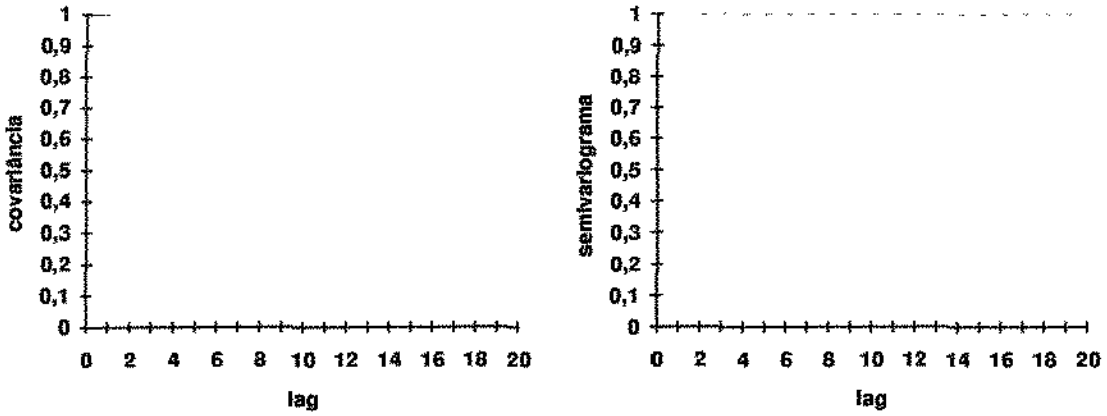


Figura 2.1.2.4. Exemplo de modelo efeito pepita puro com $C_0 = 1$.
Obs.: Aqui as figuras estão separadas para melhor visualização.

2.1.3. Variograma Fatorado

A variabilidade observada pode ser descrita por mais de um tipo de semivariograma $\gamma(h)$, ou seja, uma composição destes. Para isso utilizamos um modelo variográfico fatorado que é composto de várias funções variográficas elementares $g_u(h)$, multiplicadas por coeficientes b_u :

$$\gamma(h) = \sum_{u=0}^s \gamma_u(h) = \sum_{u=0}^s b_u g_u(h) \quad (2.1.3.1)$$

Do mesmo modo, uma função aleatória $Z(x)$ pode ser decomposta em várias componentes $Z_u(x)$, resultando em um modelo linear de regionalização:

$$Z(x) = \sum_{u=0}^s Z_u(x)$$

Exemplo: Seja o modelo fatorado $\gamma(h) = \gamma_0(h) + \gamma_1(h)$, onde

$\gamma_0(h) = C_0 = 0.3$, e $\gamma_1(h)$ um modelo esférico com $c = 0.7$ e $a = 10$, ou seja:

$$\gamma(h) = 0.3 + 0.7 \left[1 - \frac{3}{20}h + \frac{1}{2000}h^3 \right].$$

O efeito pepita aparecerá no semivariograma (ou covariância) como uma descontinuidade na origem. Este exemplo está ilustrado na Figura 2.1.3.1.

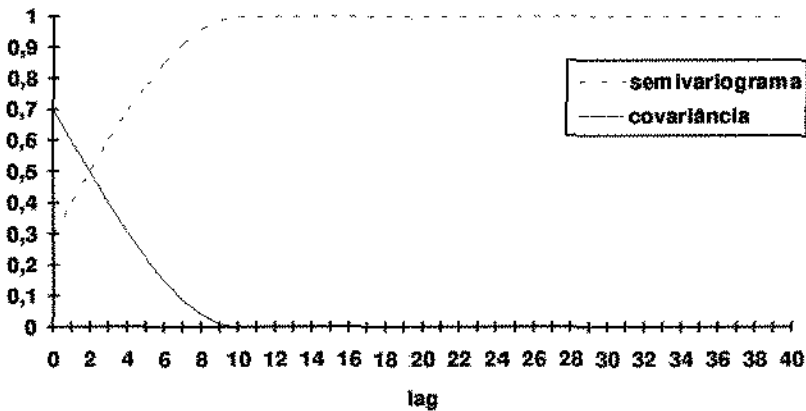


Figura 2.1.3.1. Modelos de função de covariância e correspondente semivariograma para um modelo fatorado com efeito pepita ($C_0 = 0.3$) e modelo esférico ($c = 0.7$, $a = 10$).

2.1.4. Modelo de Anisotropia

De acordo com a Definição 5 (pp. 10), dizemos que um semivariograma é anisotrópico se ele não exibe o mesmo comportamento em todas as direções. Devemos, então, realizar uma transformação. Os dois modelos mais utilizados são:

- i) *anisotropia geométrica*: a hipótese básica para a sua utilização é a variação da amplitude (ou alcance) segundo a direção, ocorrendo os alcances máximos e mínimos em direções ortogonais. O patamar e a lei do variograma permanecem os mesmos. As equações abaixo descrevem a transformação de um campo anisotrópico em um campo isotrópico, por meio de uma transformação linear:

$$\begin{aligned} \gamma(h) &= \gamma(\|h'\|) \\ h' &= Ah \end{aligned}$$

onde A representa a matriz de transformação. Um exemplo pode ser visto na Figura 2.1.4.1, onde consideramos apenas duas direções.

- ii) *anisotropia zonal ou estratificada*: é o tipo mais complexo de anisotropia a ser tratada. Necessita de mais de uma transformação A e de mais de um modelo de variograma γ em consideração:

$$\begin{aligned} \gamma(h) &= \sum_i \gamma_i(\|h_i\|) \\ h_i &= A_i h \end{aligned}$$

Um exemplo de anisotropia zonal, que considera apenas duas direções, pode ser visto na Figura 2.1.4.2.

Na prática, uma anisotropia real de uma função estrutural $\gamma(h)$ corresponde à existência de direções preferenciais, às vezes pela formação ou constituição do fenômeno em estudo. Essas direções preferenciais são, em geral, conhecidas *a priori*, e o variograma quantifica suas respectivas variabilidades. Por exemplo, um depósito mineral pode estar disposto segundo uma conformação particular do terreno.

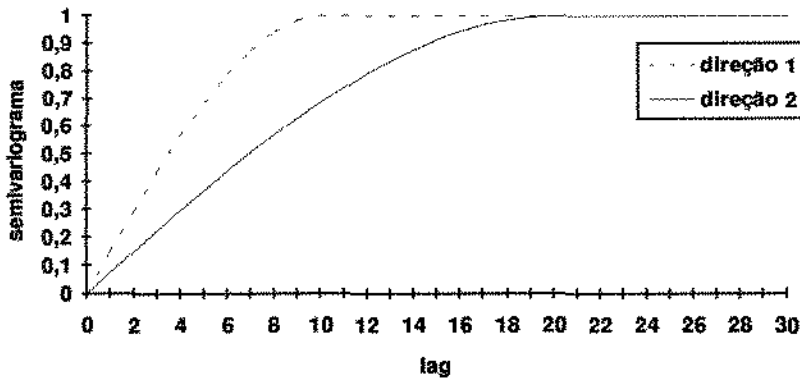


Figura 2.1.4.1. Exemplo de modelos de semivariograma com anisotropia geométrica.

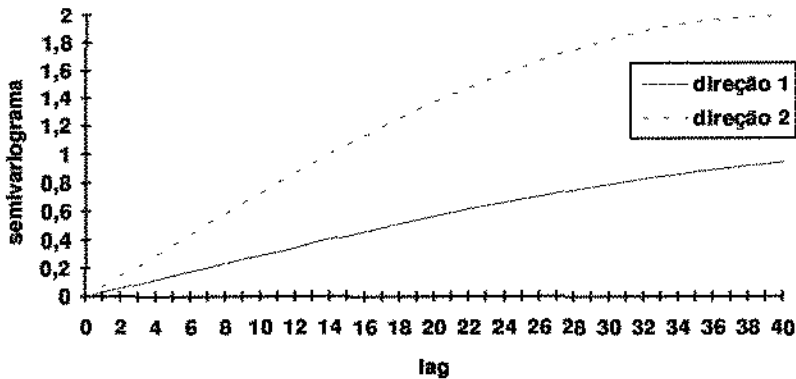


Figura 2.1.4.2. Exemplo de modelos de semivariograma com anisotropia zonal.

2.1.5. Variograma Experimental

A análise estrutural de um fenômeno regionalizado é o processo de análise de dados que leva a um modelo de semivariograma. O objetivo da análise é detectar a principal característica estrutural do fenômeno regionalizado em estudo, através da análise dos vários semivariogramas experimentais. Uma informação estrutural deve ser continuamente comparada com o conhecimento das características físicas do fenômeno regionalizado (geologia, procedimentos de amostragem, etc).

Segundo Guerra (1988, pp. 64) a determinação do variograma é o primeiro passo no procedimento de uma estimativa através da Geoestatística. E pode ser considerado como o passo mais importante, porque o modelo variográfico escolhido será utilizado através de todo o processo de estimação e influirá em todos os resultados e conclusões.

Se os pontos com os quais a amostra se relaciona estão distribuídos em uma malha regular, estima-se o semivariograma na direção fixa θ baseado no método dos momentos:

$$\hat{\gamma}_\theta(h) = \frac{1}{2N_\theta(h)} \sum_{(x_\alpha, x_{\alpha+h})} [Z(x_\alpha + h) - Z(x_\alpha)]^2 \quad x_\alpha, h \in \mathfrak{R}^p, \quad (2.1.5.1)$$

onde $N_\theta(h)$ é o número de pares amostrais $(x_\alpha, x_{\alpha+h})$ separados por uma distância (lag) h na direção θ .


Como o argumento h do semivariograma é um vetor, o semivariograma deve ser calculado ao longo de várias direções no espaço. Um outro objetivo da estimação do semivariograma em várias direções é revelar as anisotropias que talvez existam no domínio. Semivariogramas experimentais em várias direções frequentemente revelam tal anisotropia.

É suficiente avaliar apenas os pares de pontos que têm uma orientação definida. Na prática, a função estrutural (covariância ou variograma) é somente utilizada para distância limitada $\|h\| \leq b$. O limite b representa, por exemplo, o diâmetro da vizinhança de estimação (isto é, a área que contém a informação a ser utilizada).

Uma vez estimados os valores de $\gamma(h)$, construímos um gráfico pontual de $[\hat{\gamma}(h) \text{ vs } h]$ para cada direção θ . Com base neste gráfico, ajustamos um modelo para a função semivariograma.

Exemplo de cálculo de um semivariograma amostral:

Considere quatro pontos amostrais $(x_0, x_1, x_2 \text{ e } x_3)$:

valor:	0.2	0.1	0.3	0.1
localização:	\dagger_0	\dagger_1	\dagger_2	\dagger_3
distância:				

Na direção $\theta = 0$, temos três pares com distância 1: (x_0, x_1) , (x_1, x_2) e (x_2, x_3) , com a distância 2 temos dois pares: (x_0, x_2) e (x_1, x_3) e com distância 3 um par: (x_0, x_3) . O semivariograma amostral é dado por:

$$\hat{\gamma}_{\theta^0}(1) = \frac{1}{6} [(0.1 - 0.2)^2 + (0.3 - 0.1)^2 + (0.1 - 0.3)^2];$$

$$\hat{\gamma}_{\theta^0}(2) = \frac{1}{4} [(0.3 - 0.2)^2 + (0.1 - 0.1)^2]; \text{ e}$$

$$\hat{\gamma}_{\theta^0}(3) = \frac{1}{2} [(0.1 - 0.2)^2].$$

Olea (1976, citado por Braga, 1990) mostra que se $Z(x)$ for uma função aleatória estacionária, então o estimador dado pela fórmula 2.1.5.1 é um estimador não tendencioso para o semivariograma. Entretanto, a priori não há nenhum teste que permita dizer se o conjunto particular de realizações de $Z(x)$ provém de uma função aleatória estacionária. Isto só se torna uma hipótese aceitável após a obtenção dos resultados dos semivariogramas experimentais e sua comparação com as propriedades características de um semivariograma teórico.

Na prática, os dados se encontram irregularmente espaçados sobre a área pesquisada. Então o semivariograma deverá ser estimado através de uma regularização angular e por classes de distância:

- i) consideramos um ponto qualquer x_α ($x_\alpha \in \mathfrak{R}^p$);
- ii) segundo a direção θ (pré-fixada), geramos um cone de abertura $2\delta\theta$, onde $\delta\theta$ é dito ângulo de regularização; e
- iii) utilizamos os pontos que "caiam dentro" do cone na classe de distância $h \pm \Delta h$.

Normalmente a escolha de um modelo para o ajuste do semivariograma é feita visualmente dentre os modelos semivariográficos mais comuns, embora existam testes, não muito divulgados, que visam determinar se os dados experimentais se conformam com o modelo teórico específico que estamos ajustando (Valente, 1982, citado por Lana, 1989, pp. 57).

Em mineração, segundo Journel & Huijbregts (1978, pp. 232), qualquer tentativa de ajuste automático, às cegas, dos parâmetros para semivariogramas experimentais deve ser evitada. Deve ser recordado que cada estimador pontual $\hat{\gamma}(h)$ de um semivariograma experimental é sujeito a erros de estimação e flutuações que são particulares a cada ponto; esses erros variam de acordo com a distância de separação h , da área ou conjunto de dados em que o semivariograma é calculado, etc. Os métodos analíticos, tais como o método dos mínimos quadrados, não levam em conta o número de pares com que foi avaliado cada ponto do semivariograma experimental, fato que é de grande importância. Outro problema é que os métodos analíticos dão o mesmo peso a todos os pontos experimentais do semivariograma, o que pode ocasionar distorções na forma final da curva teórica, pois um bom modelo deve se ajustar com maior precisão aos primeiros pontos experimentais. Isso ocorre porque, quanto maior a distância h , em geral, menor a continuidade espacial. Uma outra observação é que, na prática, o semivariograma amostral é calculado apenas para distâncias que não excedam $1/4$ (em alguns casos $1/2$) das dimensões do campo. Pois

quando h aumenta muito, o número de pares considerados no cálculo do semivariograma diminui, prejudicando a precisão dos resultados obtidos.

Journel & Huijbregts (1978, pp. 233) apresentam um exemplo de robustez de resultados geoestatísticos com respeito à escolha do tipo de modelo, considerando dois modelos de transição com comportamento linear próximo à origem, isto é, o modelo esférico e o exponencial (seção 2.1.2, item i. e ii., respectivamente).

Obs.: É raro identificar continuidade espacial quando ela não existe. Mas há duas importantes exceções: 1) dados com cluster podem apresentar certa medida de continuidade espacial que mostram uma estrutura artificial; 2) a combinação de várias anisotropias e tolerância angular grande podem aumentar artificialmente a amplitude de correlação na direção de mínima continuidade. É importante também verificar se a amplitude é menor do que a distância entre dois pontos, pois isso resultaria numa situação equivalente a um modelo de efeito pepita puro (Deutsch & Journel, 1992, pp. 57).

Programas para o cálculo do variograma experimental podem ser vistos em Deutsch & Journel (1992) e Carr (1987); este último apresenta programas na linguagem Fortran, Pascal e C, e compara as suas performances em termos de precisão e velocidade.

2.2. Krigagem

O termo *Krigagem* provem do nome de K. G. Krige, um especialista em mineração que introduziu em 1951 o uso de tal procedimento em reservas minerais. O desenvolvimento formal da teoria deve-se a Matheron (1963; citado por Lana, 1989, pp. 14). (Comentários sobre a origem da Krigagem podem ser vistos em Cressie, 1990.)

De acordo com a definição dada por Matheron, a Krigagem é um processo estocástico de obtenção do melhor preditor linear não viciado de uma variável desconhecida (Lucio, 1991, pp. 17). Robinson (1991) também comenta que a Krigagem é essencialmente o mesmo que BLUP (*best linear unbiased predictor*), pois ela providencia estimadores lineares não viciados de mínima variância.

A Krigagem leva em consideração os seguintes fatores: o número de dados e a qualidade destes em cada ponto, a posição dos dados com relação ao campo, a continuidade espacial das variáveis interpoladas e a distância entre os pontos e a área de interesse. Com

base em tais observações, constrói-se a melhor função de predição não viciada para em um ponto não observado.

Suponha que $\{Z(x), x \in \mathfrak{R}^p\}$ é um processo estocástico estacionário de segunda ordem. Para n pontos $x_1, x_2, \dots, x_n, x_\alpha \in \mathfrak{R}^p, \alpha = 1, \dots, n$, observa-se o valor de $Z(x)$ obtendo-se $z(x_1), z(x_2), \dots, z(x_n)$. O preditor de Krigagem de $Z(x_0), x_0 \in \mathfrak{R}^p$, é dado por

$$Z^*(x_0) = \sum_{\alpha=1}^n \lambda_\alpha z(x_\alpha) = \lambda' \tilde{Z}(x),$$

onde $\lambda' = (\lambda_1, \lambda_2, \dots, \lambda_n)$ é tal que $Z^*(x_0)$:

- i) seja não viciado: $E[Z^*(x_0) - Z(x_0)] = 0$; e
- ii) tenha Erro Quadrático Médio (EQM) mínimo dentre todas as combinações lineares (não viciadas) de $Z(x)$.

Existem distintos tipos de Krigagem na literatura (por exemplo, Cressie, 1989b; citado por Cressie, 1989a), mas aqui serão apresentadas apenas duas formas:

Krigagem Simples (KS)

Considere o modelo da função aleatória $Z(x)$ como estacionário com média m e covariância $C(h)$. Nesta forma mais simples, também conhecida como *Krigagem Simples* (KS), o estimador linear é dado por:

$$Z_{KS}^*(x) = \sum_{\alpha=1}^n \lambda_\alpha Z(x_\alpha) + \left[1 - \sum_{\alpha=1}^n \lambda_\alpha \right] m .$$

Os pesos λ_α são determinados para minimizar a variância do erro, também chamada "variância de estimação". Esta minimização resulta no conjunto de equações normais:

$$\sum_{\beta=1}^n \lambda_\beta C(x_\beta - x_\alpha) = C(x - x_\alpha) \quad \alpha = 1, \dots, n .$$

A variância de estimação minimizada (ou variância da Krigagem) é dada por:

$$\sigma_{KS}^2 = C(0) - \sum_{\alpha=1}^n \lambda_\alpha C(x - x_\alpha) \geq 0 .$$

Krigagem Ordinária (KO)

Na *Krigagem Ordinária*, que é a variação da KS mais utilizada, exigimos que a soma dos pesos, $\sum_{\alpha=1}^n \lambda_{\alpha}$, seja igual a 1.

A construção do estimador $Z_{KO}^*(x)$ não requer o conhecimento *a priori* da média estacionária m .

O estimador da Krigagem ordinária é dado por:

$$Z_{KO}^*(x) = \sum_{\alpha=1}^n \lambda_{\alpha} Z(x_{\alpha}) .$$

Aplicando o teorema dos multiplicadores de Lagrange, obtemos o sistema de equações normais:

$$\begin{cases} \sum_{\beta=1}^n \lambda_{\beta} C(x_{\beta} - x_{\alpha}) + \mu = C(x - x_{\alpha}) \\ \sum_{\beta=1}^n \lambda_{\beta} = 1 \end{cases} \quad \alpha = 1, \dots, n ,$$

onde:

λ_{β} são os pesos da Krigagem ordinária; e

μ é um parâmetro de Lagrange associado com a restrição $\sum_{\beta=1}^n \lambda_{\beta} = 1$.

A variância do erro de estimação é dada por:

$$\sigma_{KO}^2 = 2 \sum_{\alpha=1}^n \lambda_{\alpha} \gamma(x_0, x_{\alpha}) - \gamma(x_0, x_0) - \sum_{\alpha=1}^n \sum_{\beta=1}^n \lambda_{\alpha} \lambda_{\beta} \gamma(x_{\alpha}, x_{\beta}) .$$

O sistema de Krigagem tem solução única se o modelo de variograma for válido. A Krigagem, além de ser um estimador não tendencioso, é um *interpolador exato*, isto é, se o ponto a ser estimado coincidir com um dos pontos amostrados, o valor estimado será igual ao valor amostrado.

A expressão dos sistemas e da variância de estimação são completamente gerais. Isto significa que são válidas quaisquer que sejam os pontos x_{α} , x_0 e o modelo γ .

Obs.: A geometria do fenômeno está expressa no sistema de Krigagem e na variância de Krigagem.

Resumo da metodologia para uma avaliação por Krigagem

- 1) Estudo variográfico (estrutural) através do semivariograma.
- 2) Seleção das informações a serem utilizadas na avaliação do bloco (ou ponto).
- 3) Cálculo dos variogramas do sistema das equações de Krigagem.
- 4) Solução do sistema de Krigagem para obtenção dos ponderadores otimizados.
- 5) Utilização destes resultados para cálculo do estimador do bloco (ou ponto) e da variância de estimação associada.

Obs.: O pacote em Fortran GSLIB (Deutsch & Journel, 1992) contém uma rotina para o cálculo da KO, que pode ser modificada facilmente para KS, se necessário.

2.3. Validação Cruzada

Há muitas decisões subjetivas em Geoestatística, e é uma boa prática "validar" o modelo geoestatístico e o plano de Krigagem antes de qualquer análise.

O modelo geoestatístico é validado pela reestimação de valores conhecidos sobre implementação de condições. Essas condições incluem o(s) modelo(s) variográfico(s), o tipo de Krigagem e a estratégia de pesquisa. Essa técnica de reestimação é discutida em livros e artigos de Geoestatística, tais como Davis (1987), Deutsch & Journel (1992, pp.90), Isaaks & Srivastava (1989, cap. 15), Schucany (1981).

Na validação cruzada ("*cross-validation*"), os dados atuais são retirados e reestimados por alguns dos dados vizinhos que permaneceram. Cada dado é substituído no conjunto de dados uma vez que ele tenha sido reestimado. Obs.: Uma outra maneira de validar o modelo geoestatístico é utilizando "*jackknife*". Este método aplica a reamostragem sem substituição, isto é, quando um conjunto de dados é reestimado de outro conjunto sem ter sobreposição. Pode ser visto, por exemplo, em Deutsch & Journel (1992) e Hinkley (1981).

Obs.: O pacote em Fortran GSLIB (Deutsch & Journel, 1992) contém uma rotina para a realização de "*cross-validation*" ou "*jackknife*" para o plano de Krigagem.

Analisando os escores reestimados:

Considere um total de K procedimentos de estimação com a finalidade de selecionar o "melhor" procedimento. Sejam $z(x_\alpha)$, $\alpha = 1, \dots, N$, os dados originais selecionados para a reestimação, e sejam $z^{*(k)}(x_\alpha)$, $\alpha = 1, \dots, N$, os valores reestimados pelo procedimento k , com $k = 1, \dots, K$. Os seguintes critérios são sugeridos para analisar os escores reestimados (Deutsch & Journel, 1992):

- a distribuição dos N erros $\{z^{*(k)}(x_\alpha) - z(x_\alpha), \alpha = 1, \dots, N\}$ deverá ser simétrica, centrada, com uma média zero, e com um mínimo de dispersão.
- o gráfico dos erros $[z^{*(k)}(x_\alpha) - z(x_\alpha)]$ versus os valores estimados $z^{*(k)}(x_\alpha)$ deverá ser centrado ao redor da linha zero (uma propriedade denominada de "condição de não viciado"). Além disso, o gráfico deverá ter uma dispersão por igual, isto é, a variância do erro não deverá depender da grandeza do valor (uma propriedade conhecida como homocedasticidade do erro da variância).
- Os N erros (para um dado procedimento k) deverão ser independentes. Isto pode ser checado pelo valor dos erros: o mapa de contorno não deverá mostrar qualquer tendência geral. Alternativamente, o variograma dos valores dos erros deverá ser calculado: ele deverá mostrar um efeito pepita puro.

Importante: A variância de Krigagem é apenas um índice do ranking das configurações dos dados. Desde que a variância de Krigagem não depende dos valores dos dados, ela não deverá ser utilizada para selecionar um modelo variográfico ou uma implementação de Krigagem, ou como critério para determinar a locação de amostras adicionais.

2.4. Características Multivariadas

2.4.1. Corregeionalização

A *corregeionalização* é o comportamento espacial mútuo de variáveis regionalizadas.

Considerando-se p variáveis regionalizadas z_1, z_2, \dots, z_p , e n locações amostrais, o vetor $z(x_\alpha)$ representa o valor de cada variável na locação α , isto é,

$$z(x_\alpha) = [z_1(x_\alpha), z_2(x_\alpha), \dots, z_p(x_\alpha)] \quad \alpha = 1, \dots, n$$

Como no estudo de uma única variável, a conduta simultânea de mais de uma variável deve ser considerada no contexto de estacionariedade e da hipótese intrínseca. E as seguintes definições deverão ser consideradas no contexto da estacionariedade de segunda ordem (Hohn, 1991, pp. 140).

Para uma dada função aleatória $Z_k(x)$, a esperança $E[Z_k(x)]$ é uma constante m_k .

A *covariância cruzada* é dada por

$$C_{jk}(h) = E[Z_j(x+h) Z_k(x)] - m_j m_k, \quad (2.4.1.1)$$

e o *variograma cruzado* é dado por

$$2\gamma_{jk}(h) = E\left\{ [Z_j(x+h) - Z_j(x)] [Z_k(x+h) - Z_k(x)] \right\}. \quad (2.4.1.2)$$

Se $j = k$, essas duas expressões se reduzem para a covariância e variograma ordinários.

Apesar da similaridade com o caso univariado, o variograma cruzado e covariância cruzada mostram algumas propriedades que podem ter significância no processo de estimação. Em contraste com o semivariograma, o semivariograma cruzado pode ter valores negativos. Tal conduta é observada quando duas variáveis são negativamente correlacionadas. Enquanto o semivariograma é simétrico quando j e k são invertidos, a covariância cruzada não é:

$$\gamma_{jk}(h) = \gamma_{kj}(h) \quad \text{e} \quad C_{jk}(h) = C_{kj}(-h).$$

Similarmente, temos que:

$$\gamma_{jk}(h) = \gamma_{jk}(-h) \quad \text{e} \quad C_{jk}(h) \neq C_{jk}(-h).$$

As estimações utilizam matrizes de covariância cruzada ou semivariograma cruzado, isto é,

$$[C_{jk}] = \begin{bmatrix} C_{11} & C_{12} & \cdots & C_{1p} \\ C_{21} & C_{22} & \cdots & C_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ C_{p1} & C_{p2} & \cdots & C_{pp} \end{bmatrix}$$

para uma dada localidade (foi utilizada uma simplificação da notação). As relações anteriores mostram que em geral essa matriz não é simétrica.

O semivariograma cruzado e a covariância cruzada se relacionam através da equação

$$2\gamma_{jk}(h) = 2C_{jk}(0) - C_{jk}(h) - C_{kj}(h),$$

que é simplificada para

$$\gamma_{jk}(h) = C_{jk}(0) - C_{jk}(h)$$

se $C_{jk}(h) = C_{kj}(h)$.

O coeficiente de correlação entre duas variáveis $Z_j(x)$ e $Z_k(x)$ localizadas no mesmo ponto x é definido por:

$$\rho_{jk} = \frac{C_{jk}(0)}{\sqrt{C_{jj}(0)C_{kk}(0)}} = \rho_{kj}.$$

2.4.2. Análise de Corregionalização

O termo *análise de corregionalização* refere-se ao estudo (exame) das matrizes de corregionalização B_u (obtida de dados espaciais multivariados utilizando um modelo variográfico fatorado), que são utilizadas na estimação de componentes espaciais (representam o comportamento das variáveis regionalizadas na escala espacial de índice u) ou fatores regionalizados (refletem os aspectos essenciais da informação multivariada para uma certa escala espacial u).

Qual é a vantagem em relação à matriz de variância-covariância V ?

Isto pode ser visto da relação entre a matriz V e um conjunto de matrizes de corregionalização $\{B_u; u = 0, \dots, s\}$:

$$V = \sum_{u=0}^s B_u .$$

Esta relação, para ser válida, requer uma hipótese de estacionariedade fraca e suposição de comportamento assintótico do variograma (Wackernagel, Petitgas e Touffait, 1989).

Cada matriz de correionalização B_u pode ser vista como uma matriz de variância-covariância de uma certa escala espacial u . Considerando que cada matriz B_u descreve a estrutura de correlação de um processo espacial multivariado, a matriz de variância-covariância V representa a mistura de diferentes estruturas de correlação. Neste caso, a matriz V provém de vários processos multivariados que atuam em diferentes escalas espaciais do fenômeno. Nessas circunstâncias, uma análise baseada em V não providenciará resultados significantes.

Observação: Se considerarmos que há *correlação intrínseca*, ou seja, que o conjunto de variáveis tem a mesma estrutura de correlação em todas as escalas, isto significa que as matrizes de correionalização B_u são proporcionais a uma matriz B (que possui os mesmos autovetores da matriz de variância-covariância V):

$$V = \sum_{u=0}^s B_u = \sum_{u=0}^s a_u B = B \sum_{u=0}^s a_u$$

e, nesse caso, não é necessário levar em conta as matrizes B_u na análise.

2.4.3. Modelo Linear de Correionalização

O *modelo linear de correionalização* providencia um método para modelagem dos variogramas e variogramas cruzados de duas ou mais variáveis, onde é necessário garantir que a "variância" de qualquer combinação linear possível dessas variáveis é sempre positiva.

Cada variável é caracterizada pelo seu variograma amostral, e cada par de variável, pelo seu variograma cruzado amostral.

O modelo para cada um dos variogramas amostrais pode consistir de um ou mais modelos básicos, porém *o mesmo modelo básico deve aparecer em cada variograma e variograma cruzado*. Em outras palavras, o modelo linear de correionalização é uma soma de modelos onde, cada modelo de variograma e variograma cruzado devem ser uma combinação linear de $N_s = s + 1$ coeficientes b_j^u , construído utilizando o mesmo modelo variográfico básico $g_u(h)$:

$$\gamma_{jk}(h) = \sum_{u=0}^s \gamma_{jk}^u(h) = \sum_{u=0}^s b_{jk}^u g_u(h) \text{ com } b_{jk}^u = b_{kj}^u, \forall u \quad (2.4.3.1)$$

onde:

j e k são índices para p variáveis diferentes;

u é o índice das diferentes estruturas espaciais;

$N_s = s + 1$ é o número de estruturas espaciais;

$g_u(h)$ é uma função variográfica com amplitude específica; e

b_{jk}^u é o coeficiente de uma função variográfica, e para um índice fixado u , a matriz de coeficientes $B_u = [b_{jk}^u]$ é positiva definida.

Note que a função dada pela fórmula 2.4.3.1 é uma generalização do variograma fatorado para uma variável (fórmula 2.1.3.1).

A função variográfica fatorada descreve o comportamento espacial de uma variável ou de um par de variáveis de diferentes escalas espaciais, onde cada fator corresponde a uma escala espacial diferente.

Para um grupo de variáveis que mostram o mesmo comportamento espacial, um modelo variográfico multivariado pode ser criado assim:

$$\Gamma(h) = \sum_{u=0}^s \Gamma_u(h) = \sum_{u=0}^s B_u g_u(h) \quad (2.4.3.2)$$

onde:

$\Gamma(h)$ é uma matriz de valores do variograma para um lag h ;

B_u é chamada de matriz de correionalização, e expressa a importância relativa da função variográfica básica $g_u(h)$ na modelagem de cada variograma.

Da mesma maneira que foi apresentado um modelo para funções variográficas, o seguinte modelo pode ser considerado para a covariância:

$$C_{jk}(h) = \sum_{u=0}^s C_{jk}^u(h) = \sum_{u=0}^s b_{jk}^u K_u(h) \text{ com } b_{jk}^u = b_{kj}^u, \forall u, \quad (2.4.3.3)$$

sendo $K_u(h)$ uma covariância com amplitude específica.

Coefficientes de correlação ρ_{ij}^u , associados com diferentes estruturas espaciais $g_u(h)$ podem ser calculados (Wackernagel, 1988), por:

$$\rho_{ij}^u = \frac{b_{ij}^u}{\sqrt{b_{ii}^u b_{jj}^u}}, \quad (2.4.3.4)$$

onde $-1 \leq \rho_{ij}^u \leq 1$ se a matriz de correionalização B_u é positiva semidefinida.

É importante citar também o *modelo intrínseco*, que é o caso de um modelo linear de correionalização onde *todas* as funções variográficas (diretas e cruzadas) são proporcionais ao mesmo variograma $g_0(h)$, isto é,

$$g_u(h) = g_0(h) \quad \forall u = 0, \dots, s.$$

O modelo intrínseco, é então:

$$\gamma_{ij}(h) = b_{ij} g_0(h), \quad (2.4.3.5)$$

e, em termos da covariância,

$$C_{ij}(h) = b_{ij} K_0(h). \quad (2.4.3.6)$$

Exemplo ilustrativo para modelo linear de correionalização:

Considere duas variáveis U e V . Seus variogramas e variogramas cruzados devem ser construídos utilizando-se o mesmo modelo básico como a seguir:

$$\gamma_U(h) = u_0 \gamma_0(h) + u_1 \gamma_1(h) + \dots + u_s \gamma_s(h)$$

$$\gamma_V(h) = v_0 \gamma_0(h) + v_1 \gamma_1(h) + \dots + v_s \gamma_s(h)$$

$$\gamma_{UV}(h) = w_0 \gamma_0(h) + w_1 \gamma_1(h) + \dots + w_s \gamma_s(h)$$

onde:

$\gamma_U(h)$, $\gamma_V(h)$ e $\gamma_{UV}(h)$ são os modelos dos variogramas e variograma cruzado para U e V , respectivamente;

$\gamma_0(h)$, $\gamma_1(h)$, ..., $\gamma_s(h)$ são os modelos variográficos básicos; e

u , v e w são os coeficientes.

Podemos escrever as equações acima em notação matricial como combinação de cada modelo básico:

- Combinação do primeiro modelo básico, $\gamma_0(h)$

$$\begin{bmatrix} \gamma_{U,0}(h) & \gamma_{UV,0}(h) \\ \gamma_{VU,0}(h) & \gamma_{V,0}(h) \end{bmatrix} = \begin{bmatrix} u_0 & w_0 \\ w_0 & v_0 \end{bmatrix} \begin{bmatrix} \gamma_0(h) & 0 \\ 0 & \gamma_0(h) \end{bmatrix}$$

- Combinação do segundo modelo básico, $\gamma_1(h)$

$$\begin{bmatrix} \gamma_{U,1}(h) & \gamma_{UV,1}(h) \\ \gamma_{VU,1}(h) & \gamma_{V,1}(h) \end{bmatrix} = \begin{bmatrix} u_1 & w_1 \\ w_1 & v_1 \end{bmatrix} \begin{bmatrix} \gamma_1(h) & 0 \\ 0 & \gamma_1(h) \end{bmatrix}$$

- Combinação do $(s+1)$ -ésimo modelo básico, $\gamma_s(h)$

$$\begin{bmatrix} \gamma_{U,s}(h) & \gamma_{UV,s}(h) \\ \gamma_{VU,s}(h) & \gamma_{V,s}(h) \end{bmatrix} = \begin{bmatrix} u_s & w_s \\ w_s & v_s \end{bmatrix} \begin{bmatrix} \gamma_s(h) & 0 \\ 0 & \gamma_s(h) \end{bmatrix}$$

onde as matrizes dos coeficientes u , v e w que aparecem nas fórmulas matriciais acima são positivas definidas. Isto implica que todos os coeficientes devem ser escolhidos de tal maneira que

$$u_j > 0 \text{ e } v_j > 0 \text{ para todo } j = 0, \dots, s; \text{ e}$$

$$u_j v_j > w_j w_j \text{ para todo } j = 0, \dots, s.$$

2.4.4. Aproximação Geoestatística

Como no caso univariado, consideramos que as variáveis regionalizadas $\{z_i(x); i = 1, \dots, p\}$ são realizações de um conjunto de funções aleatórias $\{Z_i(x); i = 1, \dots, p\}$ estacionárias de segunda ordem.

No caso multivariado, os semivariogramas experimentais são calculados utilizando a fórmula

$$\hat{\gamma}_{\theta,ij}(h) = \frac{1}{2N_{\theta}(h)} \sum_{(x_{\alpha}, x_{\alpha+h})} \left\{ [z_i(x_{\alpha}) - z_i(x_{\alpha+h})][z_j(x_{\alpha}) - z_j(x_{\alpha+h})] \right\}, \quad (2.4.4.1)$$

onde $N_{\theta}(h)$ é o número de pares amostrais $(x_{\alpha}, x_{\alpha+h})$ separados por uma distância (lag) h na direção θ .

Os variogramas experimentais são modelados heurísticamente por uma combinação de funções variográficas elementares $g_u(h)$ com coeficientes b_{ij}^u (ver seção 2.4.3), no caso de não haver a suposição de correlação intrínseca.

Observações: Na prática, o interpretador de dados seleciona o número e os tipos de funções variográficas $g_u(h)$ e especifica suas amplitudes. Os coeficientes b_{jk}^u são calculados por métodos iterativos através de um programa de computador que utiliza o método dos mínimos quadrados ponderados (Wackernagel, 1989) ou de mínimos quadrados ajustados (Goulard, 1989). Os pesos são *subjetivos*, e deverão ser modificados pelo usuário para que a forma de cada variograma experimental seja reproduzida satisfatoriamente pelo modelo. Como as matrizes de correionalização B_u devem ser *positivas semidefinida*, algumas restrições são necessárias.

$$b_{ij}^u \geq 0$$

$$|b_{jk}^u| \leq \sqrt{b_{jk}^u b_{kk}^u} \quad (\text{Cauchy-Schwartz})$$

Este tipo de ajuste pode ser visto no procedimento descrito por Wackernagel (1988).

Capítulo 3

Cokrigagem

Algumas vezes, uma variável não pode ser amostrada suficientemente (devido a alguma restrição, por exemplo, técnica ou custo financeiro) a fim de proporcionar um preditor com precisão aceitável. Neste caso, uma solução seria utilizar a *Cokrigagem*.

A Cokrigagem é a estimação de uma variável regionalizada através de duas ou mais variáveis. Seu objetivo, em geral, é o de melhorar predições locais levando em conta informações adicionais trazidas por uma variável diferente da que se deseja predizer. É evidente que a correlação espacial entre as variáveis deve ser "suficientemente forte" para que a variável adicional acrescente algum ganho de informação que sirva para a predição da variável de interesse.

A utilização deste método deve ser feita quando o objetivo principal é a redução da variância de predição de uma variável ou a necessidade da predição conjunta de tais variáveis, por ser este um processo onde tais variáveis podem ser conjuntamente preditas com base na intervariabilidade ou em informação da estrutura espacial.

Obs.: Podemos dizer também que Cokrigagem é uma generalização de Krigagem, ou Krigagem é uma Cokrigagem Univariada.

3.1. Equações de Cokrigagem

Myers (1982; revisto por Hohn, 1991, pp. 141) apresenta uma forma matricial para a estimação simultânea de várias variáveis regionalizadas.

A tarefa é estimar o vetor

$$z(x) = [z_1(x), z_2(x), \dots, z_p(x)]$$

dado $z(x_1), z(x_2), \dots, z(x_n)$, lembrando que nós temos n locações amostrais e p variáveis regionalizadas. Algumas variáveis podem não ter sido medidas em algumas situações, e isso vai gerar algumas restrições simples que serão comentadas depois.

Cokrigagem

O estimador linear é dado por:

$$\left[z_1^*(x), \dots, z_p^*(x) \right] = z^*(x) = \sum_{\alpha=1}^n z(x_\alpha) \Gamma_\alpha ,$$

onde cada uma das matrizes Γ_α ($\alpha=1, \dots, n$) é $p \times p$. Esta equação é semelhante à da Krigagem; sua diferença está na quantidade a ser estimada, que é trocada por um vetor com p entradas, e os n pesos são trocados por n matrizes Γ_α .

Desde que z_j ($j=1, \dots, p$) são estacionários, uma condição suficiente para $z^*(x)$ ser não viciado é:

$$\sum_{\alpha=1}^n \Gamma_\alpha = I = \text{matriz identidade} .$$

A expressão do estimador linear pode ser escrita em forma não matricial:

$$z_k^*(x) = \sum_{\alpha=1}^n \left\{ \sum_{j=1}^p \left[z_j(x_\alpha) \lambda_{jk}^\alpha \right] \right\} \quad k = 1, \dots, n,$$

onde λ_{jk}^α é o termo na linha j coluna k de Γ_α , e representa a contribuição da variável j na locação x_α , na estimação da variável k .

Note que a matriz Γ_α não é diagonal. Se quisermos krigar cada variável separadamente, basta fazer

$$\Gamma_{jk}^\alpha = 0 \quad \text{para } j \neq k ,$$

e

$$\sum_{\alpha=1}^n \lambda_{jj}^\alpha = 1 \quad \text{para todo } j .$$

A solução de Equações de Cokrigagem para encontrarmos as matrizes Γ_α , $\alpha = 1, \dots, n$ é dada por:

$$\begin{cases} \sum_{\beta=1}^n \bar{C}(x_\alpha - x_\beta) \Gamma_\beta + \bar{\mu} = \bar{C}(x_\alpha - x_0) \\ \sum_{\beta=1}^n \Gamma_\beta = I \end{cases} \quad \alpha = 1, \dots, n ,$$

Cokrigagem

ou simplesmente

$$\begin{bmatrix} \bar{C}^{11} & \bar{C}^{12} & \dots & \bar{C}^{1n} & I \\ \bar{C}^{21} & \bar{C}^{22} & \dots & \bar{C}^{2n} & I \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \bar{C}^{n1} & \bar{C}^{n2} & \dots & \bar{C}^{nn} & I \\ I & I & \dots & I & 0 \end{bmatrix} \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \\ \vdots \\ \Gamma_n \\ \bar{\mu} \end{bmatrix} = \begin{bmatrix} \bar{C}^{10} \\ \bar{C}^{20} \\ \vdots \\ \bar{C}^{n0} \\ I \end{bmatrix},$$

onde:

$\bar{C}^{\alpha\beta} = \bar{C}(x_\alpha - x_\beta)$ é uma matriz $p \times p$ de covariância cruzada entre a amostra α e a amostra β ;

$\bar{\mu}$ é uma matriz $p \times p$ de multiplicadores de Lagrange; e

$\bar{C}^{\alpha 0}$ é uma matriz $p \times p$ de covariância cruzada entre cada amostra e a locação que estamos interessados.

Esse sistema de equações é idêntico ao da Krigagem de uma variável se trocarmos as matrizes $p \times p$ por escalares.

Nota: A ordem de multiplicação das matrizes deve ser preservada, e a não simetria de \bar{C} deve ser respeitada.

A variância de estimação pode ser escrita de várias formas. Uma delas é:

$$\sigma_E^2 = Tr[\bar{C}(0)] - Tr\left[\sum_{\alpha=1}^n \bar{C}(x_o - x_\alpha)\Gamma_\alpha\right] - Tr(\bar{\mu}),$$

onde Tr é o traço (somatório dos elementos da diagonal principal).

A variância de estimação da variável k pode ser calculada de

$$\sigma_{E,k}^2 = C_{kk}(0) - \sum_{j=1}^p \left[\sum_{\alpha=1}^n C_{kj}(x_\alpha - x_o)\lambda_{jk}^\alpha \right] - \mu_{kk}.$$

As equações de Cokrigagem também podem ser escritas em termos do semivariograma cruzado, ou seja,

$$\begin{cases} \sum_{\beta=1}^n \bar{\gamma}(x_\alpha - x_\beta)\Gamma_\beta + \bar{\mu} = \bar{\gamma}(x_\alpha - x_o) \\ \sum_{\beta=1}^n \Gamma_\beta = I \end{cases} \quad \alpha = 1, \dots, n$$

Cokrigagem

ou, em notação matricial,

$$\begin{bmatrix} \bar{\gamma}^{11} & \bar{\gamma}^{12} & \dots & \bar{\gamma}^{1n} & I \\ \bar{\gamma}^{21} & \bar{\gamma}^{22} & \dots & \bar{\gamma}^{2n} & I \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \bar{\gamma}^{n1} & \bar{\gamma}^{n2} & \dots & \bar{\gamma}^{nn} & I \\ I & I & \dots & I & 0 \end{bmatrix} \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \\ \vdots \\ \Gamma_n \\ \bar{\mu} \end{bmatrix} = \begin{bmatrix} \bar{\gamma}^{10} \\ \bar{\gamma}^{20} \\ \vdots \\ \bar{\gamma}^{n0} \\ I \end{bmatrix},$$

onde

$$\sigma_E^2 = Tr \sum_{\alpha=1}^n \bar{\gamma}(x_o - x_\alpha) \Gamma_\alpha - Tr \bar{\mu}.$$

Esta formulação requer a suposição de que $C_{jk}(h) = C_{kj}(h)$.

Pode-se mostrar que o sistema de equações de Cokrigagem proporciona uma única solução se a matriz de covariâncias for estritamente positiva definida. Para isso, é suficiente que um ponto positivo no modelo de correionalização seja adotado, e que nenhum valor dos dados seja redundante com respeito a outro.

Utilizando o modelo intrínseco (caso particular do modelo linear de correionalização; fórmula 2.4.3.6), as variáveis $\{Z_k(x), k=1, \dots, p\}$ possuem o mesmo modelo básico. Nesse caso, a dificuldade da Cokrigagem pode ser justificada somente se uma das variáveis (k_0) é subamostrada em comparação às outras variáveis ($k \neq k_0$).

Em aplicações de mineração, a Cokrigagem (feita após um estudo de correionalização) é realizada somente se a variável a ser estimada é subamostrada com respeito às outras variáveis com a qual ela é correlacionada espacialmente (Journel & Huijbregts, 1978, pp. 324).

Assim como existem diferentes tipos de Krigagem (seção 2.1.7), existem diferentes tipos de Cokrigagem. Um exemplo é a Cokrigagem Universal (Cokrigagem na presença de tendência; maiores detalhes em Stein et al, 1991) que não será tratada neste trabalho.

3.2. Particularidades dos Cálculos

A Cokrigagem partilha com a Krigagem de uma variável regionalizada a mesma sequência de passos: modelagem do comportamento espacial, montagem do conjunto de equações para estimar um ponto ou bloco e resolução deste conjunto para obter os pesos.

Hohn (1991) apresenta alguns detalhes dos cálculos:

As dimensões adicionais na Cokrigagem introduzem algumas novas questões:

- (a) como calcular o semivariograma cruzado;
- (b) como levar em conta os dados "missing" (subamostrados) no conjunto de equações; e
- (c) como resolver o sistema de Cokrigagem.

O cálculo do semivariograma cruzado pode ser visto em Myers (1983) e Deutsch & Journel (1992). Durante a modelagem dos três semivariogramas, é utilizada a restrição

$$|\gamma_{jk}(h)| \leq \sqrt{\gamma_j(h)\gamma_k(h)}$$

para que as equações de Cokrigagem não percam a propriedade de positiva definida.

Nos casos de subamostragem, algumas linhas e colunas no sistema completo das equações de Cokrigagem terão valores iguais a zero. Assumindo que a variável j é subamostrada no local α , todas as covariâncias e covariâncias cruzadas envolvendo a variável j na amostra α recebem o valor zero. Este procedimento significa que as entradas na linha j de Γ_α são nulas. Veja o exemplo de Cokrigagem a seguir (seção 3.3).

Um algoritmo para resolver o sistema de Cokrigagem pode depender do tamanho do problema. Uma matriz quadrada no sistema de Cokrigagem tem

$$np \left(\frac{np+1}{2} \right) \text{ entradas,}$$

onde n é o número de amostras, e p o número de variáveis. Sobre a suposição de que $C_{jk}(h) = C_{kj}(h)$, o número de entradas é reduzido para

$$n \left(\frac{n+1}{2} \right) \times p \left(\frac{p+1}{2} \right) \text{ entradas.}$$

Segundo Hohn (1991), as vantagens da Cokrigagem podem desaparecer na presença de efeito pepita grande no semivariograma, baixa correlação cruzada e amplitude grande da variável subamostrada (em relação ao espaçamento amostral).

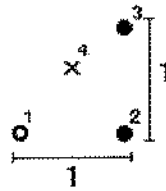
Pawlowsky (1984) adverte para a pouca eficácia do uso de Cokrigagem em casos de correlação espacial fraca; e Pawlowsky (1989) adverte para a pouca eficácia de seu uso em casos de composições regionalizadas (uma correionalização caracterizada pelo fato de que a função aleatória que a constitui tem uma soma constante em cada ponto da região amostrada; por exemplo, dados geoquímicos). Pawlowsky & Burger (1992) estudaram mais detalhadamente as propriedades das composições regionalizadas, suas interpretações e descrições, assim como um exemplo baseado em dados simulados.

Algoritmos para a resolução das equações de Cokrigagem podem ser vistos em:

- Carr, Myers e Glass (1985): programa de Cokrigagem pontual (em Fortran), que permite o caso subamostrado. Eles utilizaram uma generalização da técnica de Tanabe (1971, citado por Hohn, 1991, e Carr, Myers e Glass, 1985) e de Herman, Lent e Rowland (1973, citado por Carr, Myers e Glass, 1985);
- Carr & Myers (1990): complementaram (em Fortran) o programa acima dando maior rapidez ao programa original, pois nele as estruturas das submatrizes eram preservadas e isso não é necessário. A formulação da matriz de escalares permite a utilização do algoritmo de eliminação de Gauss padrão para reduzir a matriz a uma forma diagonal ou triangular superior. Ambos os métodos resultam em redução significativa no cálculo computacional;
- Marcotte (1991): apresentou um programa (em MATLAB) mais completo que os acima referidos. Este programa abrange a Cokrigagem de ponto ou bloco, permitindo qualquer número de dimensões, de variáveis e de estruturas básicas;
- Pan, Moss e Carr (1992): apresentaram um programa (em Fortran) para a Cokrigagem tridimensional, que em geral é adequado para qualquer sistema multivariado anisotrópico em que as variáveis são correlacionadas espacialmente. Permite também a *cross-validation* (seção 2.3) e estimação de bloco. A lógica segue o programa de Cokrigagem de Carr, Myers e Glass (1985); e
- Deutsch & Journel (1992): apresentaram um programa (em Fortran) para a Cokrigagem. Este programa faz parte do GSLIB.

3.3. Exemplo de Cokrigagem

Para ilustrar a utilização de matrizes na Cokrigagem, Hohn (1991) apresenta a estimativa de um único ponto. É apresentada uma configuração de três pontos, onde duas locações são bem amostradas (2 e 3) e uma subamostrada (1). Trabalha-se com duas variáveis, e o objetivo é encontrar um estimador para o ponto central, chamado de 4. Veja a figura abaixo:



Assuma modelos esféricos para o problema, com os seguintes parâmetros:

	Variável 1	Variável 2	Variável 1 e 2
efeito pepita (C_0)	0.0	0.0	0.00
patamar (c)	1.1	1.8	0.55
amplitude (a)	2.0	2.0	2.00

O modelo esférico (fórmula 2.1.2.1) com amplitude igual a 2 é dado por

$$C(h) = \begin{cases} c[1 - (3/4)h + (1/16)h^3] & 0 \leq h \leq 2 \\ 0 & h > 2 \end{cases}$$

Como temos duas variáveis e três locações amostrais, as equações de Cokrigagem são dadas por:

$$\begin{bmatrix} \bar{C}^{11} & \bar{C}^{12} & \bar{C}^{13} & I \\ \bar{C}^{21} & \bar{C}^{22} & \bar{C}^{23} & I \\ \bar{C}^{31} & \bar{C}^{32} & \bar{C}^{33} & I \\ I & I & I & 0 \end{bmatrix} \begin{bmatrix} \Gamma_1 \\ \Gamma_2 \\ \Gamma_3 \\ \bar{\mu} \end{bmatrix} = \begin{bmatrix} \bar{C}^{14} \\ \bar{C}^{24} \\ \bar{C}^{34} \\ I \end{bmatrix},$$

e cada uma das matrizes $\bar{C}^{\alpha\beta}$ ($\alpha, \beta=1,2,3$) é dada por:

$$[\bar{C}^{\alpha\beta}] = \begin{bmatrix} C_{11}(x_\alpha - x_\beta) & C_{12}(x_\alpha - x_\beta) \\ C_{21}(x_\alpha - x_\beta) & C_{22}(x_\alpha - x_\beta) \end{bmatrix}.$$

Fazendo os cálculos detalhadamente para quatro matrizes, \bar{C}^{11} , \bar{C}^{21} , \bar{C}^{31} e \bar{C}^{34} (note que as outras serão iguais as que serão detalhadas, desde que se tome o cuidado com o local que está subamostrado), temos

MATRIZ 1:

$$[\bar{C}^{11}] = \begin{bmatrix} C_{11}(x_1 - x_1) & C_{12}(x_1 - x_1) \\ C_{21}(x_1 - x_1) & C_{22}(x_1 - x_1) \end{bmatrix} = \begin{bmatrix} C_{11}(0) & C_{12}(0) \\ C_{21}(0) & C_{22}(0) \end{bmatrix} = \begin{bmatrix} 1.10 & 0.55 \\ 0.55 & 1.80 \end{bmatrix}.$$

Mas como a variável 2 está subamostrada na posição x_1 , a matriz \bar{C}^{11} torna-se:

$$[\bar{C}^{11}] = \begin{bmatrix} 1.1 & 0 \\ 0 & 0 \end{bmatrix}.$$

MATRIZ 2:

$$[\bar{C}^{21}] = \begin{bmatrix} C_{11}(x_2 - x_1) & C_{12}(x_2 - x_1) \\ C_{21}(x_2 - x_1) & C_{22}(x_2 - x_1) \end{bmatrix} = \begin{bmatrix} C_{11}(1) & C_{12}(1) \\ C_{21}(1) & C_{22}(1) \end{bmatrix} = \begin{bmatrix} 0.344 & 0.563 \\ 0.563 & 0.172 \end{bmatrix}.$$

Mas como a variável 2 está subamostrada na posição x_1 , é necessário verificar se isso altera a matriz. Neste caso, a matriz \bar{C}^{21} torna-se:

$$[\bar{C}^{21}] = \begin{bmatrix} 0.344 & 0 \\ 0.563 & 0 \end{bmatrix}.$$

MATRIZ 3:

$$[\bar{C}^{31}] = \begin{bmatrix} C_{11}(x_3 - x_1) & C_{12}(x_3 - x_1) \\ C_{21}(x_3 - x_1) & C_{22}(x_3 - x_1) \end{bmatrix} = \begin{bmatrix} C_{11}(\sqrt{2}) & C_{12}(\sqrt{2}) \\ C_{21}(\sqrt{2}) & C_{22}(\sqrt{2}) \end{bmatrix} = \begin{bmatrix} 0.128 & 0.064 \\ 0.064 & 0.209 \end{bmatrix}.$$

Mas como a variável 2 está subamostrada na posição x_1 , a matriz \bar{C}^{31} é tomada como:

$$[\bar{C}^{31}] = \begin{bmatrix} 0.128 & 0 \\ 0.064 & 0 \end{bmatrix}.$$

MATRIZ 4:

$$[\bar{C}^{34}] = \begin{bmatrix} C_{11}(x_3 - x_4) & C_{12}(x_3 - x_4) \\ C_{21}(x_3 - x_4) & C_{22}(x_3 - x_4) \end{bmatrix} = \begin{bmatrix} C_{11}(\sqrt{2}/2) & C_{12}(\sqrt{2}/2) \\ C_{21}(\sqrt{2}/2) & C_{22}(\sqrt{2}/2) \end{bmatrix} = \begin{bmatrix} 0.541 & 0.270 \\ 0.270 & 0.885 \end{bmatrix}.$$

Neste caso não temos o problema de subamostragem, pois a locação não se refere a variável 2 na posição x_1 . Então a matriz \bar{C}^{34} continua a mesma.

Cokrigagem

Note que assumimos que as matrizes de covariância são simétricas, e que a primeira amostra não é amostrada para a variável 2 (e então as linhas e colunas apropriadas na matriz tem entrada igual a zero).

Montamos então as equações de Cokrigagem.

Configuração da covariância e covariância cruzada amostra a amostra:

$$\begin{bmatrix} 1.100 & 0 & 0.344 & 0.172 & 0.128 & 0.064 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.344 & 0 & 1.100 & 0.550 & 0.344 & 0.172 & 1 & 0 \\ 0.172 & 0 & 0.550 & 1.800 & 0.172 & 0.563 & 0 & 1 \\ 0.128 & 0 & 0.344 & 0.172 & 1.100 & 0.550 & 1 & 0 \\ 0.064 & 0 & 0.172 & 1.563 & 0.550 & 1.800 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \end{bmatrix}$$

Configuração da covariância e covariância cruzada entre cada amostra e a locação que estamos interessados:

$$\begin{bmatrix} \bar{C}^{14} \\ \bar{C}^{24} \\ \bar{C}^{34} \\ I \end{bmatrix} = \begin{bmatrix} 0.541 & 0.270 \\ 0 & 0 \\ 0.541 & 0.270 \\ 0.270 & 0.885 \\ 0.541 & 0.270 \\ 0.270 & 0.885 \\ 1 & 0 \\ 0 & 1 \end{bmatrix}$$

Cokrigagem

Matriz solução do sistema:

$$\begin{bmatrix} \Gamma_1 \\ \Gamma_2 \\ \Gamma_3 \\ I \end{bmatrix} = \begin{bmatrix} 0.368 & 0.184 \\ 0 & 0 \\ 0.263 & -0.118 \\ 0 & 0.500 \\ 0.368 & -0.066 \\ 0 & 0.500 \\ -0.002 & -0.001 \\ -0.001 & -0.251 \end{bmatrix}$$

A variância de estimação é dada por:

$$\sigma_E^2 = Tr[\bar{C}(0)] - Tr\left[\sum_{\alpha=1}^3 \bar{C}^{4\alpha} \Gamma_\alpha\right] - Tr(\bar{\mu}),$$

onde:

$$\bar{C}(0) = \begin{bmatrix} 1.10 & 0.55 \\ 0.55 & 1.80 \end{bmatrix},$$

$$\sum_{\alpha=1}^3 \bar{C}^{4\alpha} \Gamma_\alpha = \begin{bmatrix} 0.541 & 0.270 \\ 0.270 & 0.885 \end{bmatrix}, \text{ e}$$

$$\bar{\mu} = \begin{bmatrix} -0.002 & -0.001 \\ -0.001 & -0.251 \end{bmatrix}.$$

Substituindo na fórmula, obtemos

$$\sigma_E^2 = (1.100 + 1.800) - (0.541 + 0.885) + (0.002 + 0.251) = 1.727 .$$

Para a variável 1:

$$\sigma_{E,1}^2 = 1.100 - 0.541 + 0.002 = 0.561 .$$

Para a variável 2:

$$\sigma_{E,2}^2 = 1.800 - 0.885 + 0.251 = 1.166 .$$

É assim, dado $z(x_1)$, $z(x_2)$, e $z(x_3)$, podemos calcular

$$\left[z_1^*(x_4), z_2^*(x_4) \right] = z^*(x_4) = \sum_{\alpha=1}^3 z(x_\alpha) \Gamma_\alpha .$$

Este é um exemplo simples para ilustrar a técnica. No Capítulo 5 apresentamos um conjunto de dados reais analisado por Cokrigagem e por outras técnicas que serão vistas.

Existem inúmeras áreas onde a Cokrigagem pode ser utilizada. Alguns exemplos são:

- Journel & Huijbregts (1978, pp. 339) apresentaram um breve estudo de Cokrigagem para o depósito de chumbo-prata-zinco de Broken Hill (Austrália);
- Carr & McCallister (1985) estudaram o espectro de resposta de terremoto em três partes (aceleração, velocidade e deslocamento). Devido à magnitude das três variáveis de resposta serem diferentes, os dados foram normalizados *a priori* para a Cokrigagem. A Cokrigagem providenciou uma técnica eficiente e precisa para a estimação do espectro de resposta em três partes;
- Freun (1986) estudou a área de *Layton Formation*, em Cowlwy County (Kansas), para verificar a performance da Cokrigagem na exploração de petróleo, e obteve bons resultados ao compara-lá com a performance da Krigagem nesses dados;
- Smith & Istok (1989) fizeram um estudo para determinar a aplicabilidade dos métodos de geoestatística multivariada no problema de estimação de concentração de pesticidas (subamostrado) em lençol de água, utilizando as concentrações de nitrato e de um pesticida denominado DCPA. Os autores compararam a Cokrigagem e Krigagem e demonstraram que, no conjunto de dados utilizado, a análise geoestatística multivariada é aplicável e uma ferramenta útil em estimação subamostrada de DCPA utilizando nitrato e DCPA como suporte; e
- Hohn (1991) estudou o xisto argiloso *Devonian* (oeste da Virgínia), que foi subamostrado, tendo como suporte a formação *Big Injun* e *Devonian*. O *Big Injun* é uma distinta e importante formação que ocorre a 400 pés acima do topo de contato com o xisto argiloso *Devonian*. O autor compara a Cokrigagem com a Krigagem, e conclui que aquela foi mais conservativa que as estimativas da Krigagem ordinária nesse caso.

Capítulo 4

Métodos Alternativos de Predição

O sistema de Cokrigagem se torna muito trabalhoso para mais de duas variáveis. Experiências têm mostrado que o ganho obtido pela Cokrigagem é pequeno em comparação aos esforços requeridos para alcançá-lo, a menos que uma variável seja drasticamente subamostrada em comparação com outra (Journel & Huijbregts, 1978, pp. 324). Além disso, o cálculo e modelagem de $p(p+1)/2$ semivariogramas (onde p é o número de variáveis) se torna trabalhoso para p maior que 2.

Davis & Greenes (1983; revisto por Hohn, 1991) apresentam um método, computacionalmente menos trabalhoso do que a Cokrigagem, para a estimação de variáveis regionalizadas. A *Análise de Componentes Principais* (ACP; seção 4.1) é utilizada para a obtenção de uma matriz de dados que não possui correlação cruzada espacial, evitando, assim, os cálculos dos semivariogramas cruzados.

Há um conjunto de técnicas, reunidas sobre o nome de *Análise de Krigagem Fatorial* (AKF), que torna possível incorporar a característica espacial do fenômeno na análise. A AKF, desenvolvido por Matheron (1982), permite a decomposição de uma variável regionalizada em diferentes *componentes espaciais* (seção 4.2.1) ou *fatores regionalizados* (seção 4.2.2). Na estimação de fatores regionalizados, as matrizes de correionalização B_u (matrizes simétricas positivas semidefinidas) são fatorizadas em um produto de matrizes. Dependendo de como será realizada a fatorização, surgem vários métodos de análise. Neste trabalho serão apresentadas a decomposição utilizando a ACP (seção 4.2.2.1) e a *função discriminante* (seção 4.2.2.2).

Uma outra maneira de estudar a decomposição de uma variável regionalizada em diferentes estruturas espaciais é através da *Análise Baseada no Variograma Multivariado*. O método foi proposto por Bourgault & Marcotte (1991), e utiliza o *variograma multivariado* para realçar as estruturas espaciais comuns ao conjunto de variáveis.

Apenas como ilustração, será apresentada também as idéias básicas do método proposto por Grunsky & Agterberg (1988), denominado *Análise Fatorial Espacial* (seção 4.4).

4.1. Análise de Componentes Principais (ACP)

O método de estatística multivariada chamada de Análise de Componentes Principais (ACP) foi criado antes da 2ª Guerra Mundial. Porém, a grande aplicação deste método ocorreu somente em 1960, durante a chamada "Revolução Quantitativa" nas Ciências Naturais e Sociais. A principal razão para esta demora foi a dificuldade dos cálculos envolvidos neste método. Somente com o advento e desenvolvimento da computação, que permitiu inúmeras aplicações dos métodos de estatística multivariada, incluindo componentes principais, tornou-se plausível. Ao mesmo tempo, a busca de métodos de precisão numérica se tornaram necessários devido à aplicação de componentes principais em problemas técnicos, que necessitavam de uma grande precisão. Na aplicação de componentes principais, o problema não é somente com suas propriedades formais, mas acima de tudo com suas origens empíricas (Mackiewicz & Ratajczak, 1993).

Na Cokrigagem, o que toma mais tempo é o cálculo e modelagem de $p(p+1)/2$ semivariogramas (onde p é o número de variáveis). Isso se torna trabalhoso para p maior que 2. Uma saída encontrada foi utilizar a Análise de Componentes Principais para resolver o problema.

Davis e Greenes (1983; revisto por Hohn, 1991) apresentaram uma maneira de se trabalhar sem utilizar o semivariograma cruzado. O método utiliza o *modelo intrínseco* (seção 2.4.3), e inclui a ortogonalização das variáveis originais, o cálculo e a modelagem de semivariogramas, a Krigagem dos dados transformados e o retorno à variável original.

Esta técnica considera que haja *correlação intrínseca*, ou seja, que o conjunto de variáveis tenha uma mesma estrutura de correlação em todas as escalas espaciais. No contexto do modelo de correionalização estacionário de segunda ordem, isto significa que todas as matrizes de correionalização B_u são proporcionais a uma matriz B (que possui os mesmos autovetores da matriz de variância-covariância V).

$$V = \sum_{u=0}^s B_u = \sum_{u=0}^s a_u B = B \sum_{u=0}^s a_u .$$

Assim, não precisamos levar em conta as matrizes B_u na análise. Note que a suposição de correlação intrínseca restringe o campo de aplicação desta técnica.

4.1.1.a. Estimação

Definições (utilizaremos uma simplificação na notação):

Matriz de dados ($p \times n$):

$$Z = \begin{bmatrix} z_{11} & z_{12} & \cdots & z_{1n} \\ z_{21} & z_{22} & \cdots & z_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ z_{p1} & z_{p2} & \cdots & z_{pn} \end{bmatrix},$$

onde n é o número de locações e p o número de variáveis.

Vetor de médias ($p \times 1$):

$$\bar{z} = \begin{bmatrix} \bar{z}_1 \\ \bar{z}_2 \\ \vdots \\ \bar{z}_p \end{bmatrix},$$

onde $\bar{z}_\alpha = \frac{1}{n} \sum_{\beta=1}^n z_{\alpha\beta} \quad \alpha = 1, \dots, n$

Matriz diagonal dos desvios padrões ($p \times p$):

$$S = \begin{bmatrix} s_1 & 0 & \cdots & 0 \\ 0 & s_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & s_p \end{bmatrix},$$

onde $s_\alpha = \sqrt{\sum_{\beta=1}^n [(z_{\alpha\beta} - \bar{z}_\alpha)^2 / (n-1)]} \quad \alpha = 1, \dots, n$.

Vetor cujos elementos são todos iguais a 1 ($p \times 1$):

$$I = \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}.$$

Os dados são normalizados para dar uma matriz $X (p \times n)$:

$$X = S^{-1}(Z - \bar{z} I'),$$

onde I' é a transposta de I , e cada coluna de X tem média zero e desvio padrão igual a um.

Dos dados normalizados, calcula-se a matriz de correlação das p variáveis:

$$R = \frac{1}{n-1} X X'.$$

Os p autovalores $l_1 > l_2 > \dots > l_p$ e os correspondentes autovetores E_1, E_2, \dots, E_p de R podem ser utilizados para obter-se uma decomposição em componentes principais. Como os autovalores são todos distintos, os autovetores formam uma base ortonormal para o espaço p -dimensional. Seja E a matriz dos autovetores de R :

$$E = [E_1 \ E_2 \ \dots \ E_p].$$

Como E é uma matriz ortonormal, Z pode ser expressada como:

$$Z = z I' + S E E' X.$$

Agora, seja:

$$F = SE = \begin{bmatrix} s_1 e_{11} & s_1 e_{12} & \dots & s_1 e_{1p} \\ s_2 e_{21} & s_2 e_{22} & \dots & s_2 e_{2p} \\ \vdots & & \ddots & \vdots \\ s_p e_{p1} & s_p e_{p2} & \dots & s_p e_{pp} \end{bmatrix}, \text{ e}$$

$$A = E' X = \begin{bmatrix} \sum_j e_{j1} x_{j1} & \dots & \sum_j e_{j1} x_{jp} \\ \vdots & \ddots & \vdots \\ \sum_j e_{jp} x_{j1} & \dots & \sum_j e_{jp} x_{jp} \end{bmatrix}.$$

Na terminologia de análise fatorial em *R-mode*, F é chamada a *matriz de fatores comuns* ou *matriz de comunalidades* (não envolve locações) e A é chamada *matriz de fator específico* ou *matriz de especificidades* (seus elementos dependem da locação).

Cada coluna na matriz Z é um vetor de p variáveis. Na α -ésima locação, o vetor é z_α . O vetor z_α pode ser expressado em termos dos elementos de F e A como

$$z_\alpha = \bar{z} + \sum_{j=1}^p \sum_{k=1}^p e_{kj} x_{k\alpha} \begin{bmatrix} s_1 e_{1j} \\ s_2 e_{2j} \\ \vdots \\ s_p e_{pj} \end{bmatrix} .$$

Isto é, cada coluna de Z é representado como uma soma ponderada de vetores de fator comum com pesos, que são os fatores específicos da α -ésima coluna de A .

A matriz A ($p \times n$) contém as coordenadas das n amostras ao longo dos p eixos. Em A os pesos de uma linha não se correlacionam com qualquer outra (devido aos autovetores em R -mode para cobrir o espaço p -dimensional, a obtenção da base ortogonal fez com que as coordenadas de dados sejam não correlacionadas). Para verificar que a ausência de correlação se estende espacialmente, os semivariogramas cruzados entre as linhas de A deverão ser calculados. Se os variogramas entre os pesos são negligenciáveis, das linhas de A , os semivariogramas são calculados e modelados.

Para uma dada locação x_0 , calcula-se um estimador para a variável k através da Krigagem ordinária:

$$a_{k0}^* = \sum_{\alpha=1}^n \lambda_\alpha a_{k\alpha} \quad , \quad k = 1, \dots, p .$$

Então a_{k0}^* é o estimador de Krigagem para a especificidade desconhecida $\sum_{k=1}^p e_{kj} x_{k0}$.

O estimador para o vetor ($p \times 1$) na locação x_0 é dado por:

$$z_0^* = \bar{z} + \sum_{j=1}^p a_{k0}^* \begin{bmatrix} s_1 e_{1j} \\ s_2 e_{2j} \\ \vdots \\ s_p e_{pj} \end{bmatrix} .$$

Para se colocar em notação matricial, agrupamos os p estimadores krigados em um vetor $p \times 1$:

$$a_0^* = \begin{bmatrix} a_{10}^* \\ a_{20}^* \\ \vdots \\ a_{p0}^* \end{bmatrix} .$$

e temos que:

$$z_0^* = \bar{z} + F' a_0^*$$

Utilizando-se da falta de correlação cruzada espacial entre as linhas de A , o problema de estimar vetores de dados multivariados distribuídos espacialmente é reduzido a uma série de p krigagens ordinárias.

Quando as variáveis são totalmente amostradas, os p conjuntos de pesos λ_α ($\alpha = 1, \dots, n$) são os mesmos, e somente um conjunto de equações de Krigagem necessita ser calculado (Hohn, 1991).

O método descrito nesta seção é aplicado por Davis & Greenes (1983) ao problema de estimar valores da porcentagem do teor do freixo/madeira (*ash*), porcentagem do teor de enxofre e teor do valor calorífico em um depósito de carvão. É também utilizado em um conjunto de dados reais no capítulo 5.

É possível atacar o problema de partição de R de muitos modos diferentes. Por exemplo, realizar a decomposição em componentes principais do valor esperado de $[R]$, $[E(R)]$, ao invés de R . Há também muitas maneiras de ortogonalizar a base, cobrindo o espaço p -dimensional. A extração em componentes principais em R -mode é somente um dos possíveis métodos.

4.2. Análise de Krigagem fatorial (AKF)

Os procedimentos padrões de análise de dados utilizados para dados multivariados falham por não levarem em conta as características espaciais de um fenômeno regionalizado. Não há um modo simples de incorporar a locação dos pontos amostrais à análise. A Análise de Krigagem Fatorial (AKF) tornou possível incorporar, na análise, a característica espacial do fenômeno (Sandjivy, 1984).

A AKF foi desenvolvida por Matheron (1982), e permite a decomposição de uma variável regionalizada em diferentes estruturas espaciais, que podem ser mapeadas separadamente pela análise.

Do variograma experimental de uma variável regionalizada, um variograma fatorado composto de vários modelos básicos é ajustado. A amplitude de cada modelo básico é associada com uma certa escala espacial. A AKF permite diferentes componentes (para essas escalas espaciais) a serem estimada pela Cokrigagem (Jaquet, 1989). Ou seja, atribui a cada estrutura no modelo uma "variável fictícia" (que corresponde a um certo nível de frequência do fenômeno) que pode ser estimada utilizando-se a Cokrigagem (Sandjivy, 1984).

Galli & Sandjivy (1985, citado por Jaquet, 1989) compararam teoricamente a AKF (domínio espacial) e Análise Espectral (domínio da frequência), demonstrando sua equivalência formal. A AKF parece oferecer maiores possibilidades do que Análise Espectral, pois esta não é diretamente aplicável a dados irregularmente espaçados: é necessário que os dados sejam interpolados aos nós do grid. A AKF apresenta a vantagem de trabalhar diretamente com os dados, porque o cálculo dos variogramas não requer dados regularmente espaçados.

A AKF, portanto, torna possível incorporar as características espaciais do fenômeno na análise de dados, e dá um tipo de "análise espectral" (Galli, Gerdil-Neuillet e Dadou, 1984).

As matrizes de correionalização B_u podem ser vistas como matrizes de variância-covariância de escalas espaciais particulares. Podemos, então, tentar estudar as *componentes espaciais* (seção 4.2.1), que representam o comportamento das variáveis regionalizadas para cada escala espacial u .

Por outro lado, as matrizes de correionalização B_u podem ser fatorizadas em um produto de matrizes A_u multiplicada por sua transposta:

$$B_u = A_u' A_u .$$

Assim, as componentes espaciais podem ser representadas como uma combinação de funções aleatórias não correlacionadas, com coeficientes a_k^u retiradas da matriz A_u . Temos então a estimação de *fatores regionalizados* (seção 4.2.2), que refletem os aspectos essenciais da informação multivariada para uma certa escala espacial u . Dependendo de como será obtida a partição das matrizes B_u , surgem vários métodos de análise de corregionalização que podem ser utilizados para definir fatores regionalizados. Neste trabalho, serão apresentados: a *ACP* (seção 4.2.2.1) e a *função discriminante* (seção 4.2.2.2). Muitos outros métodos são possíveis, pois há muitas maneiras de decompor uma matriz simétrica positiva semidefinida. Mas muito cuidado deve ser tomado para se selecionar somente aquelas maneiras que possuem significado para interpretação (Schervish, 1987, citado por Wackernagel, Petitgas e Touffait, 1989).

Outras técnicas de partição da matriz B_u podem ser vistas em Wackernagel, Petitgas e Touffait (1989), que apresentam uma análise de corregionalização baseada em *correlação canônica*, *fator de autocorrelação min/max*, *modelo de covariância proporcional* e *escalas espaciais entre contrastes máximos*.

4.2.1. Estimação de Componentes Espaciais

Utilizando o modelo linear de correionalização, a variável $Z_i(x)$ ($i = 1, \dots, p$) pode ser expressa como:

$$Z_i(x) = \sum_{u=0}^s Z_i^u(x) + m_s,$$

onde m_s é o vetor de médias locais. Sem perda de generalidade, podemos considerar $m_s = 0$.

Cada componente espacial $Z_i^u(x)$ representa o comportamento da função aleatória $Z_i(x)$ para uma escala espacial u . Assumindo a estacionariedade de 2ª ordem, a matriz de correionalização B_u é realmente a matriz de variância-covariância das componentes $Z_i^u(x)$. Portanto, cada B_u descreve a estrutura de correlação associada à escala espacial u .

Pela decomposição de cada $Z_i(x)$ em várias componentes espaciais $Z_i^u(x)$, podemos estudar o comportamento da variável Z para diferentes escalas espaciais. Essas componentes espaciais podem ser estimadas por *Cokrigagem Fatorial*:

$$\hat{Z}_{i_0}^{u_0}(x_0) = \sum_{i=1}^p \sum_{\alpha=1}^n \lambda_{\alpha}^i Z_i(x_{\alpha}).$$

O conjunto de equações de Cokrigagem para uma componente espacial com uma média local igual a zero é dado por,

$$\begin{cases} \sum_{j=1}^p \sum_{\beta=1}^n \lambda_j^{\beta} C_{ij}^{\beta}(x_{\alpha}, x_{\beta}) + \mu_i = C_{i_0}^{u_0}(x_{\alpha}, x_0) & i = 1, \dots, p \\ \sum_{\beta=1}^n \lambda_i^{\beta} = 0 & \alpha = 1, \dots, n \end{cases}$$

onde μ_i são multiplicadores de Lagrange.

Vários autores têm examinado as componentes espaciais utilizando somente a informação de uma variável: Galli, Gerdil-Neuillet e Dadou (1984), Sousa (1989), Wackernagel (1988), Jaquet (1989) e Raspa et al (1992).

4.2.2. Estimação de Fatores Regionalizados

As matrizes de correionalização B_u podem ser fatorizadas em um produto de matrizes A_u multiplicada por sua transposta:

$$B_u = A_u' A_u .$$

Assim, as componentes espaciais podem ser representadas como uma combinação de funções aleatórias não correlacionadas com coeficientes a_{ik}^u retiradas da matriz A_u . Temos então, a estimação de *fatores regionalizados*.

Utilizando o modelo linear de correionalização, cada componente espacial (seção 4.2.1) $Z_i^u(x)$ pode ser representada como uma combinação linear de funções aleatórias não correlacionadas Y_k^u ($u = 0, \dots, s$ e $k = 1, \dots, p$), com coeficientes de transformação a_{ik}^u :

$$Z_i(x) = \sum_{u=0}^s Z_i^u(x) = \sum_{u=0}^s \sum_{k=1}^p a_{ik}^u Y_k^u(x) \quad i = 1, \dots, p ,$$

ou seja,

$$Z_i^u(x) = \sum_{k=1}^p a_{ik}^u Y_k^u(x) \quad i = 1, \dots, p \text{ e } u = 0, \dots, s .$$

Isto significa que os coeficientes de transformação a_{ik}^u permitem passar de um conjunto de variáveis a outro. Veremos, nas seções 4.2.2.1 e 4.2.2.2, como obter os coeficientes a_{ik}^u (coeficientes das matrizes de transformação A_u). Serão apresentadas apenas duas técnicas: ACP e função discriminante.

$Y_k^u(x)$ são fatores regionalizados, e somente alguns deles (em geral bem menos do que p) são relevantes. Fatores regionalizados são estimados em uma locação x_0 através de uma combinação linear,

$$\hat{Y}_{k_0}^{u_0}(x_0) = \sum_{i=1}^p \sum_{\alpha=1}^n \lambda_{\alpha}^i Z_i(x_{\alpha}) .$$

O modelo variográfico é decomposto em estruturas espaciais, e os coeficientes b_{ij}^u são decompostos em coeficientes de transformação a_{ik}^u (veja a seção 2.4.3),

$$\gamma_{ij}(h) = \sum_{u=0}^s b_{ij}^u g_u(h) = \sum_{u=0}^s \sum_{k=1}^p a_{ik}^u a_{jk}^u g_u(h) .$$

Minimizando a variância de estimação sobre a restrição de não viciado e sujeito a condições que filtram as médias locais das variáveis, chegamos ao conjunto de *equações de Cokrigagem Fatorial*,

$$\left\{ \begin{array}{l} \sum_{j=1}^p \sum_{\beta=1}^n \lambda_j^\beta \gamma_{ij}(x_\alpha, x_\beta) + \mu_i = a_{ik_0}^{u_0} g_{u_0}(x_\alpha, x_0) \quad i = 1, \dots, p \\ \sum_{\beta=1}^n \lambda_i^\beta = 0 \quad \alpha = 1, \dots, n \end{array} \right.$$

onde μ_i são multiplicadores de Lagrange.

Também pode ser utilizada a covariância, ao invés do variograma:

$$C_{ij}(h) = \sum_{u=0}^s C_{ij}^u(h) = \sum_{u=0}^s b_{ij}^u K_u(h) .$$

As covariâncias espaciais $C_u(h)$ dos fatores regionalizados podem também ser deduzidas do variograma por,

$$g_u(h) = K_u(0) - K_u(h) = 1 - K_u(h) .$$

e o conjunto de equação é dado por:

$$\left\{ \begin{array}{l} \sum_{j=1}^p \sum_{\beta=1}^n \lambda_j^\beta C_{ij}(x_\alpha, x_\beta) + \mu_i = a_{ik_0}^{u_0} K_{u_0}(x_\alpha, x_0) \quad i = 1, \dots, p \\ \sum_{\beta=1}^n \lambda_i^\beta = 0 \quad \alpha = 1, \dots, n \end{array} \right.$$

onde μ_i são multiplicadores de Lagrange.

Como os fatores regionalizados são puramente artificiais, a variância mínima não é calculada porque não tem significado físico.

4.2.2.1. ACP na Corregionalização

O método estatístico multivariado chamado de Análise de Componentes Principais (ACP) é uma maneira eficiente de resumir a informação contida em uma matriz de variância-covariância V ou uma matriz de correlação. Assim, podemos tentar resumir as matrizes B_u utilizando uma soma de fatores não correlacionados, através da ACP. Wackernagel (1988, 1989), Wackernagel, Petitgas e Touffait (1989), Wackernagel, Webster e Oliver (1988) e Souza (1989) utilizaram a ACP para estabelecer um modelo linear de variáveis compostas de fatores regionalizados.

Podemos obter as matrizes de transformação A_u , de coeficientes a_{ik}^u , da seguinte maneira:

Uma matriz de corregionalização pode ser decomposta resolvendo-se o problema de *autovalores regionalizados* (Wackernagel, 1988)

$$B_u Q_u = \Delta_u M_u Q_u \text{ com } Q_u' M_u Q_u = I ,$$

onde:

Δ_u é uma matriz diagonal de autovalores λ_k^u ;

Q_u é uma matriz M_u -ortonormal de autovetores q_k^u ; e

M_u é uma matriz simétrica positiva semidefinida representando uma métrica.

Considerando-se M_u como uma matriz identidade, isto é,

$$M_u = I ,$$

o problema se reduz ao problema de autovalor clássico:

$$Q_u' B_u Q_u = \Delta_u .$$

Ou seja, as matrizes de corregionalização B_u são decompostas em um sistema de autovalores e autovetores:

$$B_u = Q_u' \Delta_u Q_u = \left(\sqrt{\Delta_u} Q_u \right)' \sqrt{\Delta_u} Q_u = A_u' A_u ,$$

onde A_u é uma matriz de transformação de coeficientes a_{ik}^u .

Portanto, as matrizes A_u são obtidas por uma decomposição das matrizes de correionalização B_u (que assumimos serem positivas semidefinidas) em matrizes diagonais de autovalores Δ_u e matrizes de autovetores ortonormais Q_u , ou seja,

$$A_u = Q_u \sqrt{\Delta_u} .$$

Isto nos leva a uma decomposição em *componentes principais* das matrizes de correionalização.

Note que desta maneira, $a_{ik}^u = q_{ik}^u \sqrt{l_k^u}$ (onde l_k^u é o autovalor associado com o k -ésimo autovetor q_{ik}^u da matriz de correionalização B_u).

Os autovalores maximizam a variância representada pela matriz B_u e os autovetores podem ser escolhidos para cobrir um sistema de eixos chamados *eixos principais*. Os eixos principais associados com os mais importantes autovalores explicam as características essenciais da estrutura de covariância. Como a matriz ortonormal Q_u satisfaz a equação da hipersfera unitária, as coordenadas das variáveis em um par de eixos principais está dentro de um círculo centrado na origem (Wackernagel, 1988).

4.2.2.2. Função Discriminante na Correionalização

Similarmente a ACP baseada em B_u (seção 4.2.2.1), a função discriminante regionalizada pode ser definida por uma diagonalização das matrizes B_u utilizando a matriz de variância-covariância (V) como uma métrica (Wackernagel, 1988 e Wackernagel, Petitgas e Touffait, 1989), ou seja, considerando

$$M_u = V$$

Neste caso, as matrizes de autovetores são escolhidas V -ortonormais e,

$$A_u = Q_u V \sqrt{\Delta_u} .$$

4.2.3. Exemplos de Análise de Krigagem Fatorial (AKF)

Apresentaremos inicialmente um exemplo de AKF para fatores regionalizados no caso univariado e citações de artigos onde os outros tipos de técnicas foram utilizadas:

O exemplo descrito a seguir, retirado de Sandjivy (1984), tem como objetivo a ilustração da técnica de AKF para fatores regionalizados no caso univariado. Para uma melhor compreensão, também foram incluídas três figuras ilustrativas dos resultados retiradas do artigo de Sandjivy (1984). Maiores detalhes, consultar o artigo citado.

Sandjivy (1984) apresentou um exemplo simples para AKF no caso univariado. Estudou a distribuição de cobre em amostras de solo coletadas de uma grande "malha" de prospecção na área de Munster (França). Foram coletadas 977 amostras de solo para 22 elementos químicos e apresentados os resultados para a AKF de cobre.

A distribuição do cobre apresentou um histograma unimodal de forma "lognormal", com média 19.7 ppm e variância 747. Para mapear o cobre utilizando a Krigagem, foi feita uma transformação para a normal. (Veja o comportamento espacial de $Z(x)$ na Figura 4.2.3.1).

A regionalização do cobre foi representada por um modelo de covariância estacionária e isotrópica (veja a Figura 4.2.3.2). Como temos apenas uma variável, a fórmula para $N_s = 3$ escalas espaciais é dada por:

$$C(h) = \sum_{u=0}^2 (a_u)^2 C_u(h) \\ = a_0^2 C_0 + a_1^2 C_1(h) + a_2^2 C_2(h)$$

onde:

C_0 : efeito pepita	$a_0^2 = 0.30$
C_1 : modelo esférico com amplitude igual a 2 unidades	$a_1^2 = 0.15$
C_2 : modelo esférico com amplitude igual a 9 unidades	$a_2^2 = 0.60$

Estes modelos básicos podem ser associados com três diferentes "harmônicos" de regionalização.

- C_0 : aspecto errático (altas frequências)
- C_1 : estrutura local (médias frequências)
- C_2 : origem regional (baixas frequências)

As correspondentes Y_0 , Y_1 e Y_2 são estimadas da decomposição de $Z(x)$:

$$\begin{aligned} Z(x) &= \sum_{u=0}^2 a_u Y_u \\ &= a_0 Y_0 + a_1 Y_1 + a_2 Y_2 \end{aligned}$$

Y_1 pode ser considerada como uma flutuação local, um mapa de anomalias (irregularidades) e Y_2 corresponde à variação regional, de acordo com o mapa geológico (veja a Figura 4.2.4.3)

Segundo Sandjivy (1984), para cada valor anormal estatisticamente ($Cu > 50$ ppm) os coeficientes de correlação r_0 , r_1 e r_2 associados com Y_0 , Y_1 e Y_2 ajudam a decidir se esses valores são regionais ou anômalos.

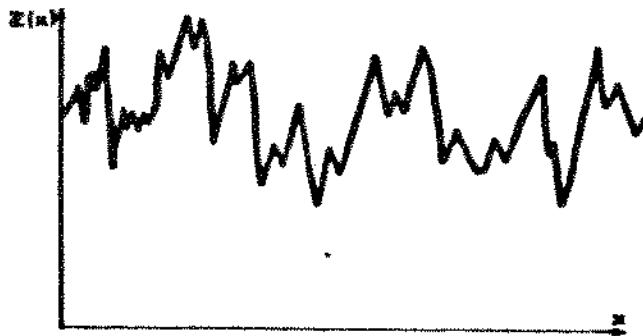


Figura 4.2.4.1. $Z(x)$: Variável regionalizada unidimensional

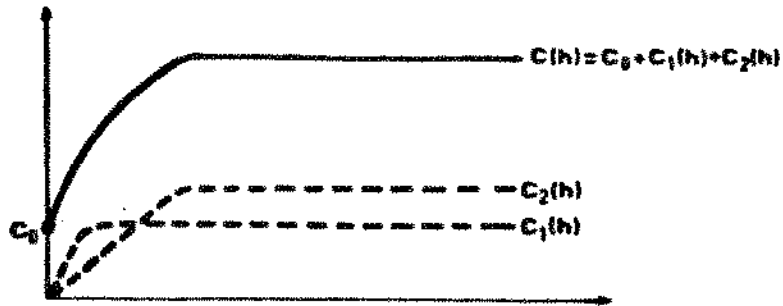


Figura 4.2.4.2. Análise estrutural de $Z(x)$

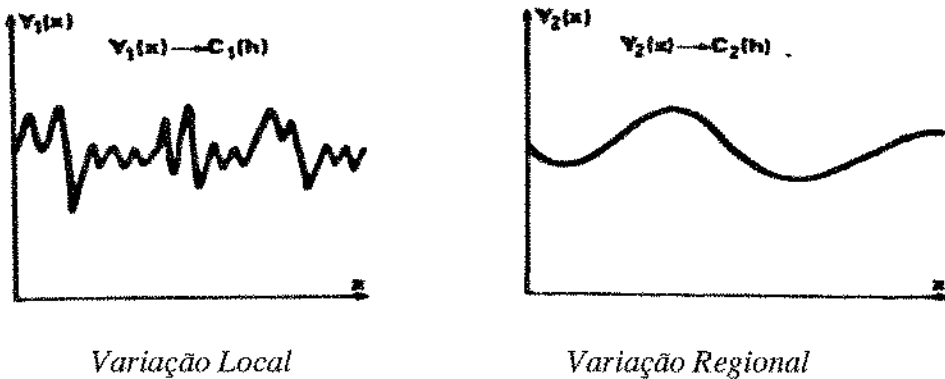


Figura 4.2.4.3. Análise de Krigagem Fatorial: $Z(x) = a_0Y_0 + a_1Y_1 + a_2Y_2$

Ilustrações de um exemplo de AKF, retiradas do artigo de Sandjivy (1984).

Encontramos na literatura vários artigos que aplicam a AKF na análise de dados e, alguns trazem variações da técnica. Entre eles:

Galli, Gerdil-Neuillet e Dadou (1984) utilizaram a AKF em dados magnéticos para determinar as estruturas locais.

Jaquet (1989) aplicou AKF em uma espessura da camada de um reservatório, obtendo componentes, associados com o depósito ambiental, que são favoráveis à exploração de petróleo. Analisou um conjunto de dados geológicos de um campo de gás utilizando 214 amostras irregularmente espaçadas e uma variável.

Souza (1989) utilizou AKF para analisar um grupo de minério extraído ao sul de Portugal. Estudou 491 amostras com análise de 6 elementos químicos.

Wackernagel (1988) aplicou a técnica de ACP na estimação de fatores regionalizados a um conjunto de dados geoquímicos da região de Ancenis (Loire, França). Estudou 1233 amostras de solo e 15 elementos geoquímicos.

Wackernagel, Webster e Oliver (1988) utilizaram a técnica de ACP na estimação de fatores regionalizados em um conjunto de dados de solo de *Foresty Wyre*, perto de *Birmingham* (Inglaterra).

Wackernagel (1989) analisou três elementos químicos, Cu, Pb e Zn, medidas em 2049 amostras de solo coletadas ao norte da cidade de Brilon, durante o projeto "*Rhenohorzynikum*", utilizando a técnica de ACP na estimação de fatores regionalizados.

Rouhani & Wackernagel (1990) apresentaram a técnica de ACP na estimação de fatores regionalizados num contexto diferente, qual seja, na análise de dados espaço-temporais em variáveis hidrológicas. Associaram uma série temporal a um variograma particular básico construído com dados piezométricos. Os autores estudaram seis poços (6 variáveis) na França, onde o conjunto de dados é composto de leituras piezométricas mensais.

Raspa et al (1992) trataram da classificação de solos, para um mapa semidetalhado em uma área situada na Região *Emilia Romagna* (Norte da Itália), utilizando ACP na estimação de fatores regionalizados. Foram medidas mais de 750 amostras de solos em várias distâncias, e foram consideradas seis variáveis.

Goovaerts & Sonnet (1992) descreveram uma aproximação geoestatística para analisar o comportamento espaço-temporal das propriedades dos dados que são densos no espaço e escassos no tempo. O comportamento temporal das propriedades de interesse é estudado para cada uma das escalas espaciais identificadas. O método é ilustrado com o estudo do comportamento espaço-temporal de cálcio, clorido e nitrato em uma fonte originária da linha divisória das águas (*watershed*) de *Dyle* (Bélgica).

4.3. Análise Baseada no Variograma Multivariado

Bourgault & Marcotte (1991) apresentaram uma nova aproximação geoestatística multivariada (no contexto do modelo linear de correionalização), que utiliza o variograma multivariado para realçar as estruturas espaciais comuns ao conjunto de variáveis.

O variograma multivariado é definido de maneira similar ao variograma tradicional, pelo valor esperado do quadrado de uma distância, em um espaço p -dimensional (p é o número de variáveis). Combinada com o modelo linear de correionalização, esta ferramenta providencia uma maneira de encontrar variogramas elementares que caracterizam as diferentes escalas espaciais contidas em um conjunto de dados com p variáveis. Se o número de componentes elementares é menor ou igual ao número de variáveis, é possível estimar os parâmetros de correionalização pela Cokrigagem ou pela inversão matricial direta. Esta nova ferramenta simplifica o procedimento proposto por Matheron (1982) e outros autores citados na seção 4.2.2.

A pesquisa dos variogramas elementares é realizada utilizando-se somente um variograma (multivariado), ao invés dos $p(p+1)/2$, no caso isotrópico, requeridos na aproximação de Matheron (1982). A estimação direta dos parâmetros do modelo linear de correionalização envolve a criação de matrizes de correionalização positivas semidefinidas de posto 1.

Em resumo, é possível calcular novas variáveis que sintetizam a informação multivariada de cada escala espacial determinada pelo variograma multivariado.

4.3.a. Algumas definições necessárias:

a) Distância (Euclidiana) em um Espaço de Variável Unidimensional

$$d(h) = |z(x) - z(x+h)| = \sqrt{[z(x) - z(x+h)]^2},$$

onde:

x \equiv vetor que determina a posição espacial de uma amostra;

h \equiv vetor de medida da distância espacial entre duas amostras; e

$z(x)$ \equiv valor da variável no ponto x .

b) Variograma Tradicional Experimental

O variograma experimental tradicional pode ser considerado como o valor esperado do quadrado de uma distância, no espaço da variável.

$$2\hat{\gamma}(h) = AVG \left\{ [z(x) - z(x+h)]^2 \right\} = AVG \left\{ [d(h)]^2 \right\},$$

onde:

$h \equiv$ vetor de medida espacial disposto entre duas amostras para uma classe de distância direcional; e

$AVG[] \equiv$ média do número de pares para este lag.

c) Distância Euclidiana em um Espaço de Variável p -dimensional

$$d(h) = \sqrt{\sum_{i=1}^p [z_i(x) - z_i(x+h)]^2},$$

onde:

$z_i(x)$ é o valor da variável i na amostra localizada em x .

Em notação matricial, utilizamos as seguintes definições

$$\mathbf{z}(x) = [z_1(x), z_2(x), \dots, z_p(x)]_{1 \times p}, \text{ e}$$

$$d(h) = \sqrt{[\mathbf{z}(x) - \mathbf{z}(x+h)] I [\mathbf{z}(x) - \mathbf{z}(x+h)]' },$$

onde I é a matriz identidade $p \times p$

d) Distância em um Espaço de Variável p -dimensional

$$d(h) = \sqrt{[\mathbf{z}(x) - \mathbf{z}(x+h)] M [\mathbf{z}(x) - \mathbf{z}(x+h)]' },$$

onde $M_{p \times p} \equiv$ matriz simétrica positiva definida utilizada como métrica, e a maneira como pode ser encontrada está descrita no item i.

e) Variograma Experimental Multivariado

Por analogia com o variograma tradicional experimental, o variograma multivariado experimental é definido como:

$$2\hat{G}(h) = AVG \left\{ [z(x) - z(x+h)] M [z(x) - z(x+h)]' \right\} \\ = AVG \left\{ [d(h)]^2 \right\} .$$

Note que quando $M = I$, o variograma multivariado é simplesmente a soma de variogramas univariados experimentais (item b).

No contexto de funções aleatórias estacionárias de 2ª ordem:

$$z_i(x) : \text{realizações da função aleatória } Z_i(x) \\ Z(x) = [Z_1(x), \dots, Z_p(x)]_{1 \times p}, \text{ e} \\ E[Z(x)] = \mu = [\mu_1, \dots, \mu_p]_{1 \times p} .$$

f) Variograma Multivariado $G(h)$

$$2G(h) = E \left\{ [Z(x) - Z(x+h)] M [Z(x) - Z(x+h)]' \right\} .$$

g) Covariância Multivariado (ou Covariograma Multivariado) $K(h)$

$$K(h) = E \left\{ [Z(x) - \mu] M [Z(x+h) - \mu]' \right\} \\ = E \left\{ Z(x) M Z(x+h)' - \mu M \mu' \right\} .$$

onde $M_{p \times p}$ ≡ matriz simétrica positiva definida no item i.

h) Relação entre $G(h)$ e $K(h)$

$$G(h) = K(0) - K(h) .$$

Esta relação é válida para qualquer matriz simétrica M . Deve ser notado também que $G(h)$ e $K(h)$ são simétricos em h . (A prova pode ser vista em Bourgault & Marcotte, 1991.)

i) Determinação da Matriz Simétrica M

Segundo a maneira como definiremos a partição de $Z(x)$ na página 64, temos que:

$$V = E\left\{[Z(x)]' Z(x)\right\} = E\left\{A'[X(x)]' X(x)A\right\} = A'A.$$

A inversa de *Moore-Penrose* (inversa generalizada) da matriz V é uma matriz V^- , tal que:

$$VV^-V = V \quad \text{e} \quad V^-VV^- = V^-.$$

Fazendo

$$M = V^-, \text{ temos que}$$

$$AMA' = AV^-A' = I_{N_s},$$

onde M é a inversa de *Moore-Penrose* da matriz de variância-covariância de $Z(x)$. Quando M é de posto completo, ela é definida como a *métrica de Mahalanobis* (Mardia, Kent e Bibby, 1979, pp. 31). Exemplos de outras métricas são: matriz identidade (euclidiana), matriz diagonal com a inversa dos desvios padrões, a métrica de qui-quadrado, etc.

Na prática, a métrica pode ser encontrada por uma decomposição em valores singulares da matriz de variância-covariância V . A métrica utilizada é então:

$$V^- = Q\Delta^{-1}Q',$$

onde:

V : matriz de variância-covariância ($p \times p$) com posto N_s ;

Δ : matriz diagonal ($N_s \times N_s$) de autovalores não nulos; e

Q : matriz ($N_s \times p$) de autovetores associados com autovalores não nulos.

4.3.b. Estimação

Utilizando o modelo linear de correionalização, a estimação envolve um conjunto de p funções aleatórias de 2ª ordem, $Z_i(x)$ ($i = 1, \dots, p$). Essas funções podem ser particionadas em N_s funções aleatórias ortogonais (estacionárias de 2ª ordem), $X_u(x)$ ($u = 0, \dots, s$). Ou seja, cada uma das p funções aleatórias é uma combinação linear de N_s funções aleatórias ortogonais:

$$Z_i(x) = \sum_{u=0}^s a_{iu} X_u(x),$$

onde:

$Z_i(x)$: funções aleatórias estacionárias de 2ª ordem, centradas em zero;

$X_u(x)$: funções aleatórias estacionárias ortogonais de 2ª ordem, com variância unitária e valor esperado zero;

a_{iu} : constantes numéricas que são parâmetros da correionalização; e

N_s : número de estruturas espaciais ($N_s = s + 1$).

Em notação matricial:

$$Z(x)_{1 \times p} = X(x)_{1 \times N_s} A_{N_s \times p} .$$

No modelo definido aqui, assume-se apenas um fator para cada escala espacial u ($u = 0, \dots, s$), enquanto o modelo de correionalização de Matheron (1982), na estimação de fatores regionalizados (apresentado na seção 4.2.2), difere na existência de p fatores ortogonais para cada uma das N_s escalas espaciais contidas nos dados, onde o objetivo era obter $N_s \times p$ variáveis ortogonais (dimensão $N_s \times p$) de um espaço p -dimensional.

Duas funções aleatórias, $X_v(x)$ e $X_w(x)$, devem ter o mesmo variograma, mas variograma cruzado igual a zero. Todos os variogramas e variogramas cruzados da função aleatória $Z_i(x)$ são representados como uma combinação linear dos variogramas de $X_u(x)$, $\gamma_{X_u}(h)$,

$$\gamma_{ij}(h) = \sum_{u=0}^s a_{iu} a_{ju} \gamma_{X_u}(h) .$$

Aplicação do Variograma Multivariado no Modelo Linear de Correionalização

$$\begin{aligned} 2G(h) &= E \left\{ [Z(x) - Z(x+h)] M [Z(x) - Z(x+h)]' \right\} \\ &= E \left\{ [X(x) - X(x+h)] A M A' [X(x) - X(x+h)]' \right\} ; \end{aligned}$$

$$A M A' = I_{N_s} ,$$

$$2G(h) = \sum_{u=0}^s \gamma_{X_u}(h) ,$$

isto é, o variograma multivariado será a soma dos variogramas das componentes ortonormais subjacentes. Isto torna possível identificar as diferentes escalas espaciais no variograma. Como que as funções aleatórias têm variância unitária, o *patamar* do variograma multivariado indicará o número de escalas espaciais N_s ..

Os $\gamma_{x_u}(h)$ são facilmente determinados "quebrando-se" o variograma multivariado em uma soma de variogramas de patamar igual a um.

Para filtrar (ou analisar) as variáveis $Z_i(x)$ de acordo com as diferentes escalas espaciais identificadas no variograma multivariado, os coeficientes a_{iu} da correionalização devem ser estimados. Os coeficientes a_{iu} são obtidos por *regressão não linear*, utilizando o RNLIN (subrotina) do software IMSL (citado por Bourgault & Marcotte, 1991), ou fazendo seu próprio programa descrito a seguir. A estimação de A segue o seguinte algoritmo iterativo:

- Passo 0: Solução inicial é proposta para A^k (faça $k = 0$). Uma possível solução inicial poderá ser os primeiros k vetores de uma Análise de Componentes Principais em $Z(x)$.
- Passo 1: Utilizando a fórmula de $\gamma_{ij}(h)$ e A^k , calcular os variogramas e variogramas cruzados.
- Passo 2: Os $\gamma_{ij}(h)$ são comparados com os encontrados experimentalmente. Uma estatística de melhor ajuste é calculada.
- Passo 3: Esse valor é comparado com a iteração obtida previamente. Se a diferença é suficientemente pequena, o algoritmo pára; caso contrário, $k = k + 1$.
- Passo 4: A próxima solução de A^k é encontrada pelo algoritmo de Marquardt (1963). A iteração recomeça no Passo 1.

Estimando-se os coeficientes a_{iu} diretamente, a matriz de variância-covariância $V = A'A$ é necessariamente positiva semidefinida (ou positiva definida, se A é de posto completo). Também $A'A$ pode ser decomposta como soma de N_s submatrizes positiva semidefinida, onde cada uma tem posto 1 e representa uma parte da matriz de covariância .

Uma decomposição similar à descrita nesta seção é apresentada na seção 4.2.2, mas nessa aproximação, as submatrizes têm que ser checadas para verificar se são positivas semidefinidas. Aqui elas são assumidas por construção.

Existem duas opções para a estimação das variáveis ortogonais $X_u(x)$. A primeira é baseada na Cokrigagem e a segunda na inversão matricial direta:

Cokrigagem

$$\hat{X}_{u_0}(x_0) = \sum_{i=1}^p \sum_{\alpha=1}^n \lambda_{\alpha}^i Z_i(x_{\alpha}).$$

Os pesos λ_{α}^i são obtidos resolvendo-se o sistema de equações de Cokrigagem:

$$\begin{cases} \sum_{j=1}^p \sum_{\beta=1}^n \lambda_{j\beta}^{\beta} \gamma_{ij}(x_{\alpha}, x_{\beta}) + \mu_i = a_{iu} \gamma_{X_{u_0}}(x_{\alpha}, x_0) & i = 1, \dots, p \\ \sum_{\beta=1}^n \lambda_{i\beta}^{\beta} = 0 & \alpha = 1, \dots, n \end{cases}$$

onde μ_i são multiplicadores de Lagrange.

Este método produz uma estimativa de um $X_u(x)$ nos pontos não amostrados.

Inversão Matricial Direta

$$\hat{X}(x)_{n \times N_s} = Z(x)_{n \times p} A'(AA')^{-1}$$

onde n é o número de amostras, N_s é o número de estruturas espaciais e p é o número de variáveis.

Uma vez obtida $\hat{X}(x)$, a Krigagem ordinária de cada componente pode ser calculada. Note que esta opção não envolve variogramas. Ela providencia estimativas de todas as componentes ortogonais ao mesmo tempo, mas somente nos pontos amostrados.

Nota: Vários autores reconhecem a importância de se utilizar a informação espacial para realizar a análise de cluster para amostras com referência geográfica. Bourgault, Marcotte e Legendre (1992) sugerem o variograma multivariado, $G(h)$, e o covariograma multivariado, $K(h)$, para serem utilizados como funções espaciais na formação automática de grupos espaciais homogêneos.

4.3.c. Modelo Realístico

Com dados reais, a matriz de variância-covariância V é quase sempre de posto completo (p). Isto é (de certo modo) o resultado de erro de medida e arredondamento. Levando isso em conta, o modelo linear de correção regionalização deve ser modificado. Adicionamos ruído branco para cada uma das variáveis $Z_i(x)$ e supomos que esses ruídos são independentes um do outro e independentes das variáveis $Z_i(x)$:

$$Z(x)_{1 \times p} = X(x)_{1 \times N} A_{N \times p} + E(x)_{1 \times p},$$

onde $E(x) \equiv p$ ruídos brancos.

Para o modelo linear de correção regionalização com ruído adicionado, a matriz V é diretamente inversível, e a métrica utilizada no variograma multivariado será a *métrica de Mahalanobis* (Mardia, Kent e Bibby, 1979, pp. 31).

Bourgault & Marcotte (1991) aplicaram a técnica de predição baseada no variograma multivariado no estudo da montanha de Lakeview (Califórnia). Utilizaram um conjunto de 147 amostras dispostas em uma malha regular e 7 elementos químicos. Bourgault, Marcotte e Legendre (1992) utilizaram o mesmo conjunto de dados para realizar uma análise de cluster utilizando o variograma multivariado e o covariograma multivariado.

Bourgault & Marcotte (1992) compararam a técnica de ACP na estimação de fatores regionalizados com a análise baseada no variograma multivariado em duas situações. A primeira em um conjunto de dados simulados e ruído adicionado, e a segunda em um estudo de caso de exploração de um depósito de níquel-cobre no Labrador. Os autores mostraram que a análise baseada no variograma multivariado providenciou resultados mais fáceis de interpretar (em termos geológicos) do que a ACP, na estimação de fatores regionalizados.

4.4. Análise Fatorial Espacial

Grunsky & Agterberg (1988) apresentaram a aplicação de uma técnica denominada *Análise Fatorial Espacial*. Esta técnica produz combinações lineares de variáveis que possuem correlação máxima baseada em estimativas de autocorrelação e correlação cruzada, num intervalo de lag específico. Note que a técnica apresentada na seção 4.1, ACP (Davis & Greenes, 1983), é baseada em combinações lineares de variáveis derivadas de variância máxima.

Apenas como ilustração, apresentaremos as idéias básicas do método proposto.

O procedimento de análise fatorial espacial requer que estimativas de autocorrelação e correlação cruzada sejam obtidas para as variáveis sobre a amplitude espacial dos dados.

As autocorrelação e correlação cruzada podem ser expressas de modo tal que uma variável aleatória X_i ($i = 1, \dots, p$) com média zero na locação α (no espaço geográfico) é relacionada à X_j ($j = 1, \dots, p$) em outra locação β pela função $F_{ij}(d_{\alpha\beta})$, com:

$$X_{i\alpha} = F_{ij}(d_{\alpha\beta})X_{j\beta} + E_{i\alpha},$$

onde α e β variam de 1 a n , e n é o número total de observações. $F_{ij}(d_{\alpha\beta})$ é uma função da distância $d_{\alpha\beta}$ entre dois pontos. Considera-se que esta função decai a zero na distância máxima da "vizinhança" (amplitude). O resíduo é uma variável aleatória $E_{i\alpha}$ no ponto α . Satisfaz $E(E_{i\alpha}) = 0$ e $E_{i\alpha}$ é assumida ser independente de $X_{j\beta}$.

A escolha das funções $F_{ij}(d_{\alpha\beta})$ utilizadas para estimar autocorrelações e correlações cruzadas dependem da natureza das relações espaciais. Diversos modelos podem ser escolhidos, tais como modelos esféricos, gaussianos, exponenciais, etc. O modelo utilizado por Grunsky & Agterberg (1988) é dado por:

$$F_D(d_{\alpha\beta}) = a + bd_{\alpha\beta} + cd_{\alpha\beta}^2,$$

e é estimado por mínimos quadrados ordinários. Isso é feito para cada x_α ($\alpha = 1, \dots, n$) com a outra locação x_β dentro de uma vizinhança ao redor de α (de raio de tolerância D). O valor de d representa qualquer lag intermediário tal que $0 < d \leq D$.

Esta técnica foi utilizada para analisar um conjunto de dados litogeoquímicos na área de Ben Nevis (Ontário - Canadá) em Grunsky & Agterberg (1988) e, com uma pequena modificação, Grunsky & Agterberg (1992).

4.5. Programas Disponíveis

- Para a ACP o SAS (SAS Institute) possui o PROC PRINCOMP.
- Hohn (1985) apresentou um programa em SAS (SAS Institute) que realiza a ACP sem usar os procedimentos acima e apresenta um estudo de dados de uma seção composta de eventos estratigráficos.
- Zhou (1989) apresentou um programa robusto (em Fortran), chamado de ROPCA, para a ACP. Os outliers nos dados podem influenciar a análise e levar a conclusões errôneas. Os estimadores de mínimos quadrados das dispersões não são robustos e são suscetíveis a outliers. O problema é inerente à ACP que é baseada em uma matriz de dispersão. O autor apresenta três métodos robustos alternativos.
- Mackiewicz & Ratajczak (1993) apresentaram um programa em Fortran para o cálculo das componentes principais.
- Wackernagel (1989) descreveu o programa chamado LINMOD. Enquanto os programas acima citados calculam apenas as componentes principais e não realizam o estudo completo, o LINMOD ajusta um modelo variográfico para os variogramas experimentais de um subconjunto de variáveis. O modelo decompõe os variogramas em várias estruturas espaciais fatoradas. As correlações das variáveis são resumidas aplicando as componentes principais nos coeficientes das estruturas espaciais. O resultado é um modelo linear de correlogramização que pode ser utilizado para Krigagem fatorial, simulação condicional e Cokrigagem. O autor apresentou um fluxograma para que o leitor possa escrever seu próprio programa.
- Marcotte (1991) apresentou um programa para MATLAB que, além de realizar a Cokrigagem (ver seção 3.3), permite a Cokrigagem Fatorial.
- Um programa para o cálculo das autocorrelações e correlações cruzadas pode ser visto em Grunsky & Agterberg (1991a) e, para a análise fatorial espacial, em Grunsky & Agterberg (1991b).

Capítulo 5

Aplicação dos Métodos de Predição

Como ilustração, utilizamos três técnicas, todas elas em um mesmo conjunto de dados de mineração: Krigagem (seção 2.2), Cokrigagem (seção 3.1) e ACP (seção 4.1). (Devido ao pequeno número de amostras disponíveis, a confiabilidade das estimativas é muito reduzida.) Mostramos, através deste conjunto de dados, algumas das possíveis dificuldades e complicações ao se analisar um conjunto de dados de mineração, onde dispomos, devido ao custo elevado e dificuldades técnicas, de poucas amostras (que são, geralmente, coletadas com um espaçamento irregular).

O conjunto de dados está descrito na seção 5.1. Na seção 5.2, os três métodos de predição (acima mencionados) são aplicados ao conjunto de dados. A seção 5.3 contém notas sobre os cálculos, e a seção 5.4 apresenta uma comparação das técnicas utilizadas.

É importante observar novamente que, uma vez que a intenção é a ilustração da técnica, utilizamos os modelos mais simples possíveis.

Observação: não foi permitida a divulgação da fonte e das características (tais como nome e local) que permitissem a identificação dos dados, e por isso as análises descritas neste capítulo são apenas ilustrativas. Assim, para a ilustração das técnicas de análise, os dados foram previamente transformados (padronização dos dados) e as medidas das coordenadas foram expressas em u.m. (unidade de medida).

5.1. Descrição dos Dados

Foram amostrados 36 poços localizados segundo a Figura 5.1.2, levando-se em conta 5 variáveis importantes. Os locais de observação são dados em coordenadas UTM (coordenada norte-sul, N-S, e coordenada leste-oeste, E-W).

Para evitar a identificação dos dados, realizamos uma transformação, ou seja, todas as variáveis foram padronizadas, de modo que tenham média zero e variância um. Realizamos um estudo descritivo de todas as variáveis e escolhemos apenas duas delas para a análise. Essa escolha foi feita levando-se em conta a correlação entre as variáveis. Foram escolhidas as duas variáveis com maior coeficiente de correlação (0.57). Chamamos estas variáveis de V1 e V2. (Obs.: nas técnicas de Cokrigagem e ACP, a variável V2 será utilizada como variável auxiliar na predição da variável V1.)

Um resumo da análise descritiva univariada para V1 e V2 pode ser visto na Tabela 5.1.1. Os histogramas de V1 e V2 podem ser vistos na Figura 5.1.1 e o gráfico de dispersão entre as variáveis, na Figura 5.1.3.

A variável V1 e a variável V2 serão denotadas, respectivamente, por $z_1(x_\alpha, y_\alpha)$ e $z_2(x_\alpha, y_\alpha)$, onde (x_α, y_α) , $\alpha = 1, \dots, 36$, representa as coordenadas UTM, sendo x a coordenada leste-oeste (E-W) e y a coordenada norte-sul (N-S).

É importante ressaltar que a análise tem como objetivo apenas ilustrar as técnicas e, sendo assim, não será analisada a geologia do campo (sedimentologia, litologia, etc).

Os valores do conjunto de dados (padronizados) utilizados na estimação estão apresentado no Apêndice A (Tabela A.1). A representação gráfica da variável V1 nos locais de coleta pode ser vista na Figura 5.1.4. (Note que, como os grid são irregulares, a visualização só foi possível com a criação de grids regulares através de Spline).

Tabela 5.1.1. Análise Descritiva das variáveis V1 e V2.

	V1	V2
Média	0,00	0,00
Variância	1,00	1,00
Mediana	-0,30	-0,02
Amplitude	4,30	4,66
Moda	-1,39	-1,04
Assimetria	1,35	-0,16
Curtose	1,67	0,26

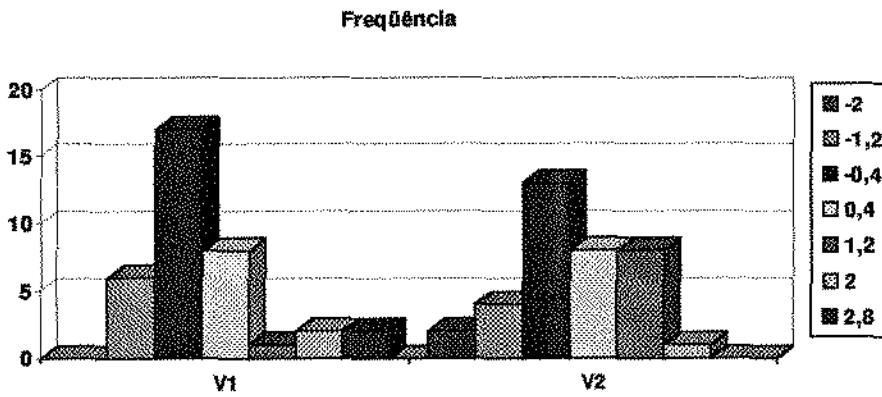


Figura 5.1.1: Histogramas das variáveis V1 e V2. Os valores da legenda representam os pontos médios dos respectivos intervalos de classes.

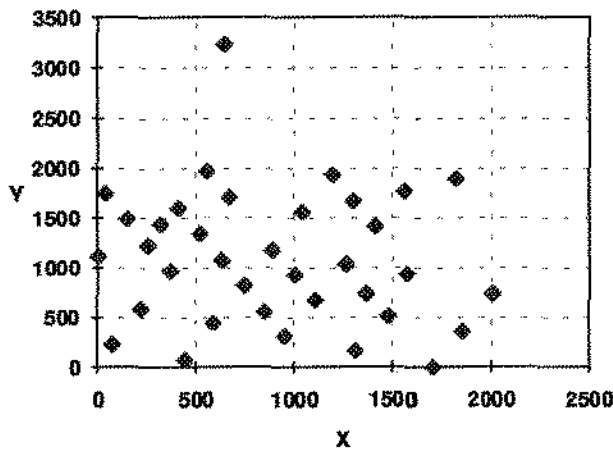


Figura 5.1.2: Localização das 36 amostras para a variável V1, onde x representa a coordenada E-W e y a coordenada N-S.

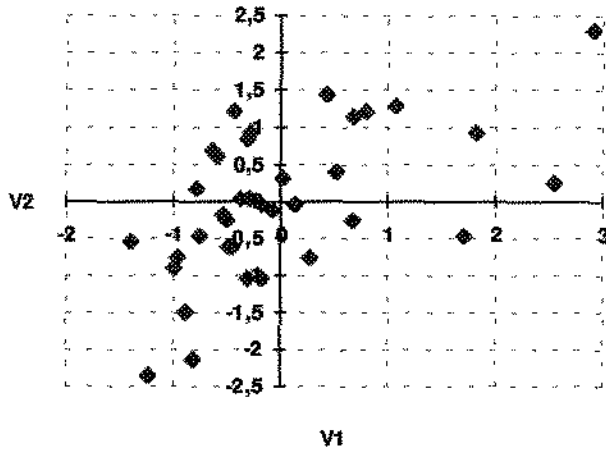


Figura 5.1.3: Gráfico de dispersão entre a variável V1 e V2 ($\rho = 0,57$).

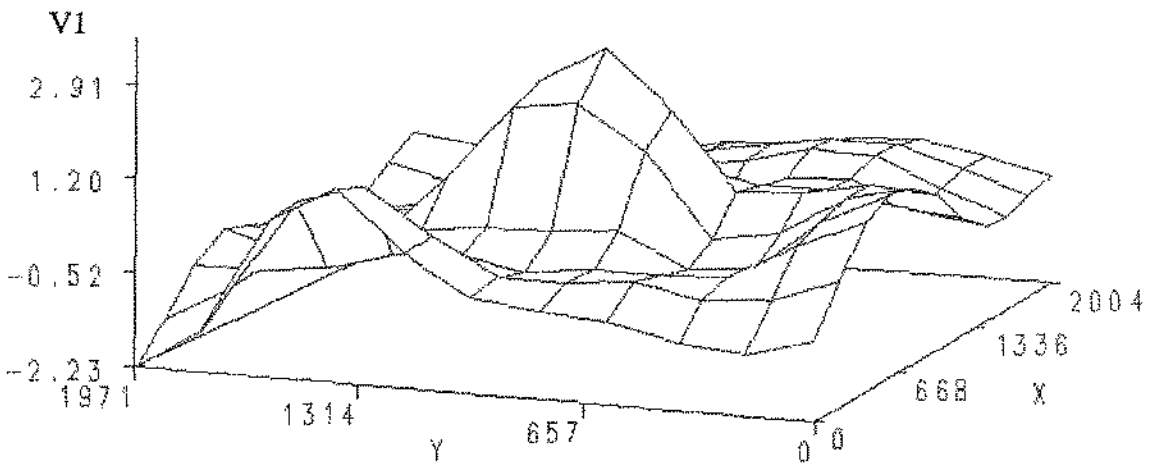


Figura 5.1.4: Variável V1 (grid com Spline para permitir visualização), onde x representa a coordenada E-W e y a coordenada N-S.

5.2. Aplicação dos Métodos de Predição

Uma predição do valor da variável V1 num ponto (x_0, y_0) baseada nos pontos observados (x_α, y_α) , $\alpha = 1, \dots, 36$, pode ser feita de duas maneiras diferentes:

- (a) num ponto observado com sua prévia retirada; ou
- (b) num ponto não observado.

O item (a) foi criado para que se tenha uma maneira de comparar, numericamente, as técnicas em relação ao conjunto original. Realizamos 36 estimações pontuais para cada técnica, sendo que, para cada estimacão, o conjunto de dados é diferente, possuindo apenas 35 pontos. Não seria possível retirar uma amostra do conjunto original e verificar se esta reflete o original porque a quantidade de observações (36) já é muito reduzida. Para comparação sugerimos a utilização da fórmula:

$$\frac{1}{36} \sum_{\alpha=1}^{36} [z_l(x_\alpha, y_\alpha) - z_l'(x_\alpha, y_\alpha)]^2 \quad l = K, C, A ,$$

onde: *K*: Krigagem, *C*: Cokrigagem e *A*: ACP.

Como as observações estão num grid irregular, a visualização tridimensional não é possível. Sendo assim, para que se pudesse ter também uma comparação visual, criamos grids regulares através de Spline, para as amostras da variável V1 (Figura 5.1.4) e suas estimativas (Krigagem: Figura 5.2.1.2, Cokrigagem: Figura 5.2.2.4 e ACP: Figura 5.2.3.3.).

O item (b) é a maneira tradicional de predição, ou seja, o interesse é a predição de um ponto não amostrado. As predições foram feitas em 121 pontos, regularmente distribuídos dentro da área de interesse, sendo que a distância entre dois pontos consecutivos na coordenada *x* é de 200.4 u.m. e na direção *y* de 197.14 u.m. (u.m.: unidade de medida). Obs.: a distância entre os pontos foi assim escolhida porque o Spline criou grids dessa maneira. É interessante comparar visualmente também os resultados do item (a) (interpolados através de Spline) com os resultados do item (b) (predição tradicional).

5.2.1. Krigagem

Suponha que a variável regionalizada V_1 , $z_1(x, y)$, seja a realização de um processo estocástico estacionário de segunda ordem $\{Z_1(x, y), (x, y) \in \mathbb{R}^2\}$. Será utilizado então o preditor de Krigagem ordinária (seção 2.2) para a predição da variável V_1 . Teremos duas análises diferentes de acordo com os itens (a) e (b) definidos na seção 5.2.

Inicialmente estimamos a função semivariograma, que não é conhecida. Note que neste caso não poderíamos utilizar a função do semivariograma experimental mais simples, pois os pontos se encontram distribuídos irregularmente sobre a área pesquisada. O semivariograma amostral foi calculado através de uma regularização angular e por classes de distância, conforme descrito na seção 2.1.5. Utilizamos uma rotina (em Fortran) do GSLIB (Deutsch & Journel, 1992) para o cálculo do semivariograma amostral em três direções principais: N-S, E-W e NW-SE. Os gráficos se encontram na Figura 5.2.1.1.a, onde é possível assumir uma isotropia nas três direções calculadas, uma vez que o nosso interesse é apenas uma ilustração das técnicas apresentadas. Sendo assim, foi calculado o semivariograma amostral considerando-se apenas o comprimento r do vetor h (Figura 5.2.1.1.b).

Para o ajuste do modelo semivariográfico, calculamos vários modelos através de outra rotina do GSLIB. Escolhemos o modelo que melhor se ajustou (visualmente e por *cross-validation*, seção 2.3). O modelo escolhido foi o modelo esférico com os parâmetros: $a = 1200$, $c_0 = 0$ e $c = 0.91$, plotado em linha contínua na Figura 5.2.1.1.

É importante salientar que o nosso objetivo é obter uma estimativa "razoável" da função semivariograma para fins de predição pelo método da Krigagem, e não para exame de características geológicas estruturais do campo de mineração amostrado. Para uma análise mais séria, seria necessária a adoção de um modelo semivariográfico mais criterioso, envolvendo um conhecimento maior da área de mineração em estudo. Para a estimação pela Krigagem, uma deficiência na escolha do modelo semivariográfico não invalida os valores da predição, uma vez que o preditor correspondente é relativamente robusto (Petenate, 1983, e Valente, 1982).

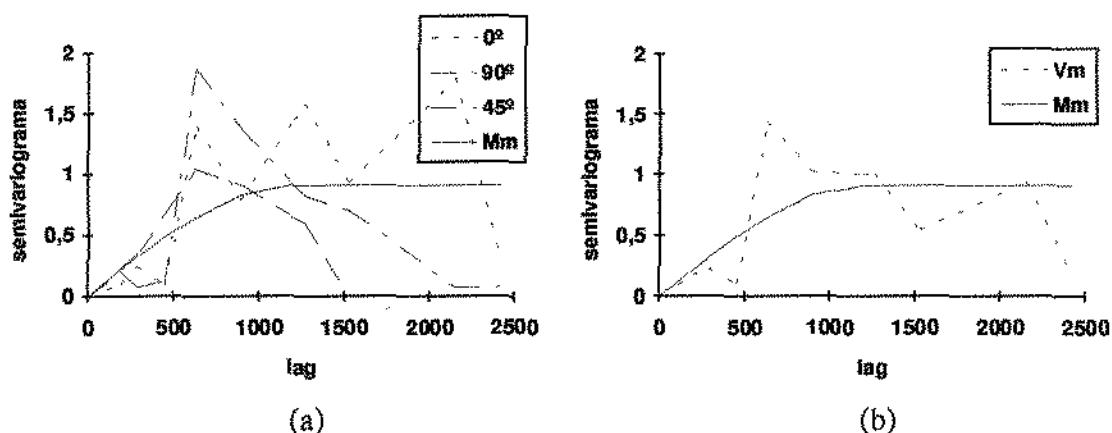


Figura 5.2.1.1. Semivariogramas experimentais para VI (a) em 3 direções e (b) assumindo isotropia dos processos estocásticos geradores. A linha contínua (Mm) se refere ao modelo esférico escolhido: $a = 1200$, $c_0 = 0$ e $c = 0.91$.

Obs.: Na direção 90°, os cálculos foram feitos sem a observação nº 2 da amostra (os valores correspondentes são outliers).

Com o modelo já escolhido, utilizamos uma rotina do GSLIB para o cálculo das estimativas, através da Krigagem, nos 36 pontos amostrados (com sua prévia retirada) segundo o item (a) da seção 5.2. Os 36 valores obtidos podem ser vistos na Tabela A.2 (Apêndice A). A representação gráfica (tridimensional) das estimativas só foi possível depois da construção de grids regulares através de Spline, resultando em 121 pontos interpolados que podem ser vistos na Tabela A.3 (Apêndice A), e a visualização tridimensional, na Figura 5.2.1.2.

Os valores obtidos de acordo com o item (b) da seção 5.2 podem ser vistos na Tabela A.3 (Apêndice A), e a visualização tridimensional, na Figura 5.2.1.3.

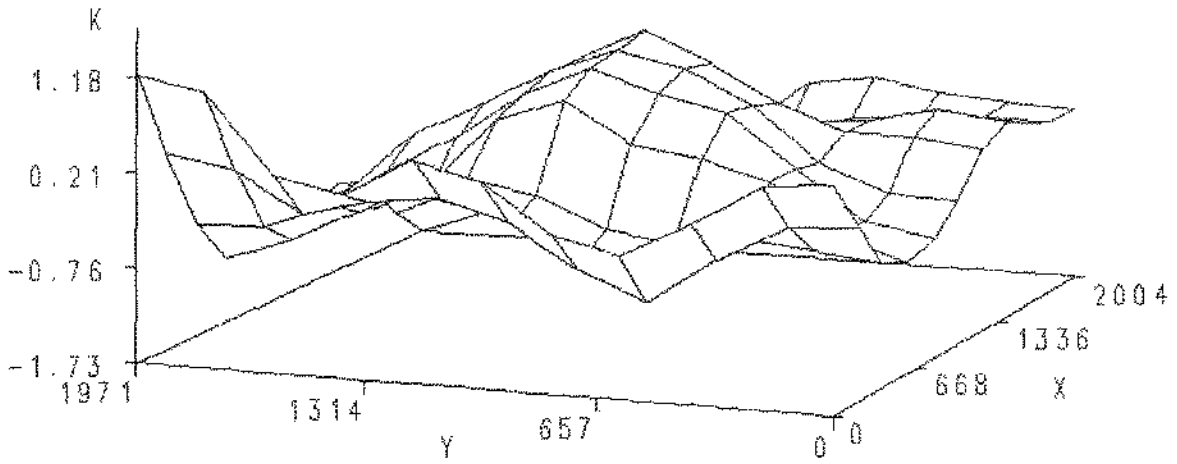


Figura 5.2.1.2. Krigagem nos pontos observados com sua prévia retirada (grid para permitir visualização), onde K representa as estimativas interpoladas, x a coordenada E-W e y a coordenada N-S.

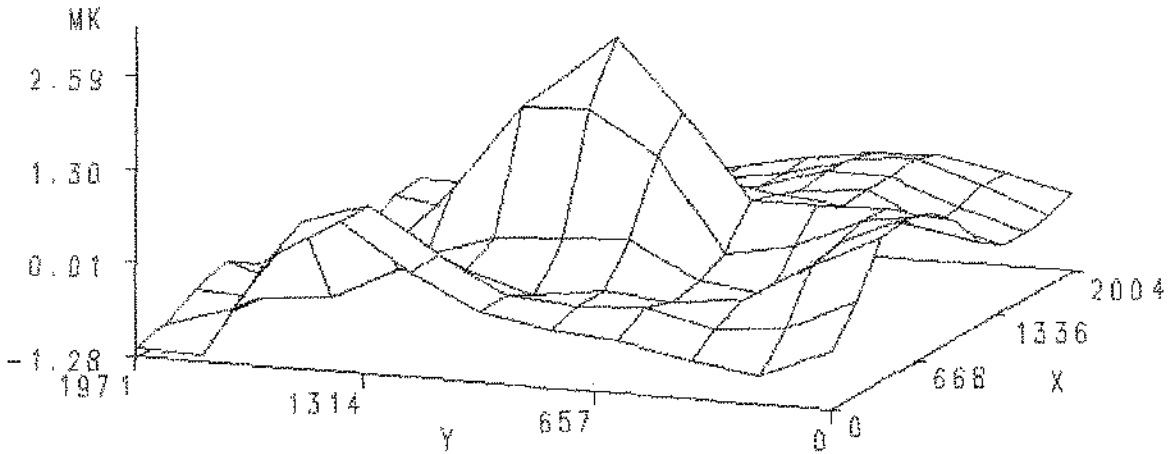


Figura 5.2.1.3. Krigagem em uma malha regular utilizando todos os dados disponíveis, onde MK representa as estimativas segundo o item (b), x a coordenada E-W e y a coordenada N-S.

Comparando as Figuras obtidas (Figura 5.2.1.2 e 5.2.1.3) com a Figura 5.1.4, onde a variável V1 está apresentada, podemos notar que os valores da Figura 5.2.1.2 seguem os padrões da variável original, mas com uma suavização. Já a Figura 5.2.1.3 é muito semelhante à Figura 5.1.4.

5.2.2. Cokrigagem

Como apresentamos no Capítulo 3, a Cokrigagem tem como objetivo geral melhorar as predições locais levando em conta informações adicionais trazidas por uma covariável (variável diferente da que se deseja predizer). Na Cokrigagem, a correlação espacial entre as variáveis adicionais deve ser "suficientemente forte" para que se obtenha algum ganho de informação.

Utilizaremos a variável V2 como auxiliar na estimação da variável principal V1. Estamos interessados na estimação do vetor $z(x_0, y_0) = [z_1(x, y), z_2(x, y)]$ dado $z(x_1, y_1), \dots, z(x_{36}, y_{36})$, sendo que neste caso as duas variáveis foram medidas em todos os 36 pontos, ou seja, não se tem subamostragem.

Para a estimação, é necessária a informação sobre os semivariogramas das variáveis V1, V2 e o semivariograma cruzado entre elas.

O semivariograma para V1 pode ser visto na Figura 5.2.1.1. e o semivariograma para V2 na Figura 5.2.2.1 abaixo. Para o ajuste do modelo semivariográfico, procedemos como antes, ou seja, calculamos vários modelos e escolhemos aquele que melhor se ajustou visualmente. Escolhemos o modelo esférico com os parâmetros: $a = 1200$, $c_0 = 0$ e $c = 0.91$, plotado em linha contínua na Figura 5.2.2.1.

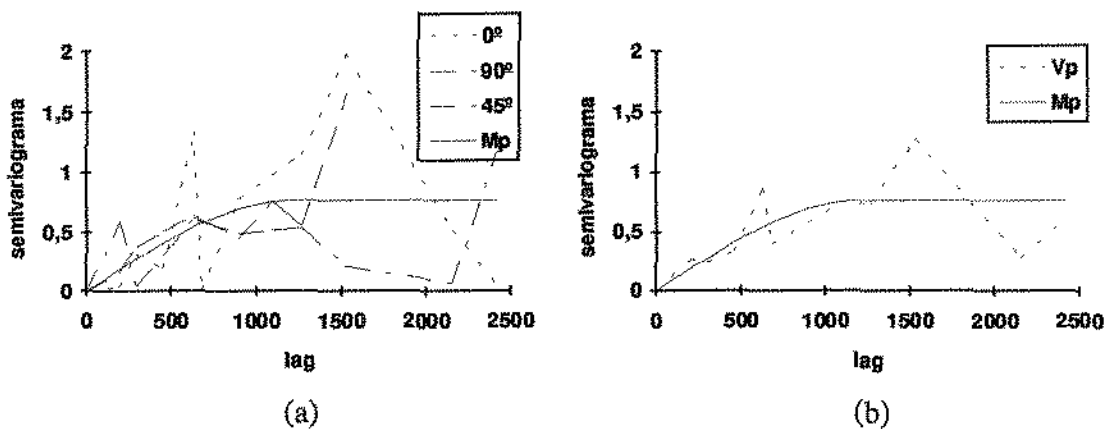


Figura 5.2.2.1. Semivariogramas experimentais para V2 (a) em 3 direções e (b) assumindo isotropia dos processos estocásticos geradores. A linha contínua (Mp) se refere ao modelo esférico escolhido: $a = 1200$, $c_0 = 0$ e $c = 0.91$

O semivariograma cruzado teórico (seção 2.4.1), é simétrico quando a ordem das variáveis é invertida em sua fórmula, desde que respeitadas as hipóteses de que as variáveis regionalizadas provêm de um processo estocástico estacionário de segunda ordem. Para verificar essa conduta, calculamos os semivariogramas experimentais para $V1 \times V2$ (Figura 5.2.2.2) e $V2 \times V1$ (Figura 5.2.2.3). Notamos, pelas figuras abaixo, que essa é uma suposição razoável. Assim, um modelo foi escolhido para o semivariograma cruzado da mesma maneira que anteriormente. Escolhemos o modelo esférico com os parâmetros: $a = 1200$, $c_0 = 0$ e $c = 0.80$. O modelo escolhido está plotado em linha contínua nas Figuras 5.2.2.2 e 5.2.2.3.

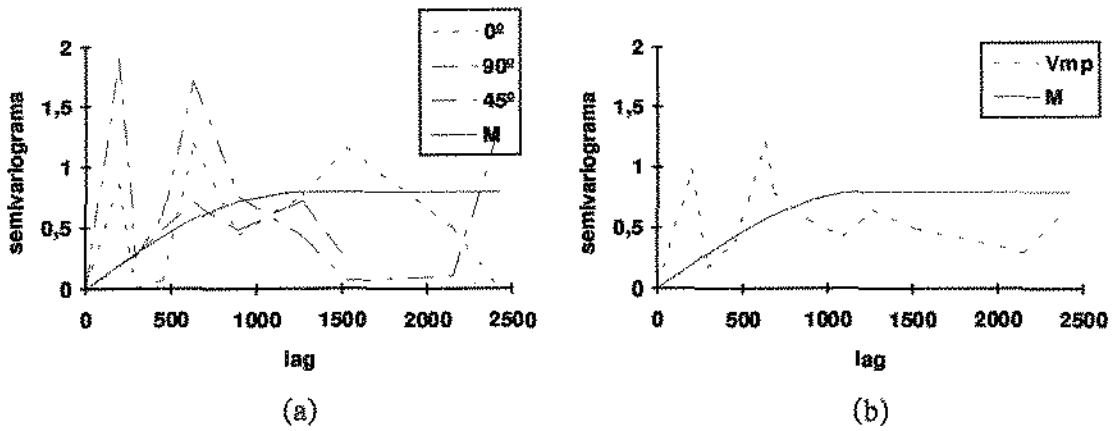


Figura 5.2.2.2. Semivariogramas cruzados experimentais para $V1 \times V2$ (a) em 3 direções e (b) assumindo isotropia dos processos estocásticos geradores. A linha contínua (M) se refere ao modelo esférico escolhido: $a = 1200$, $c_0 = 0$ e $c = 0.80$.

Obs.: Na direção 90° , os cálculos foram feitos sem a observação nº 2 da amostra (os valores correspondentes são outliers).

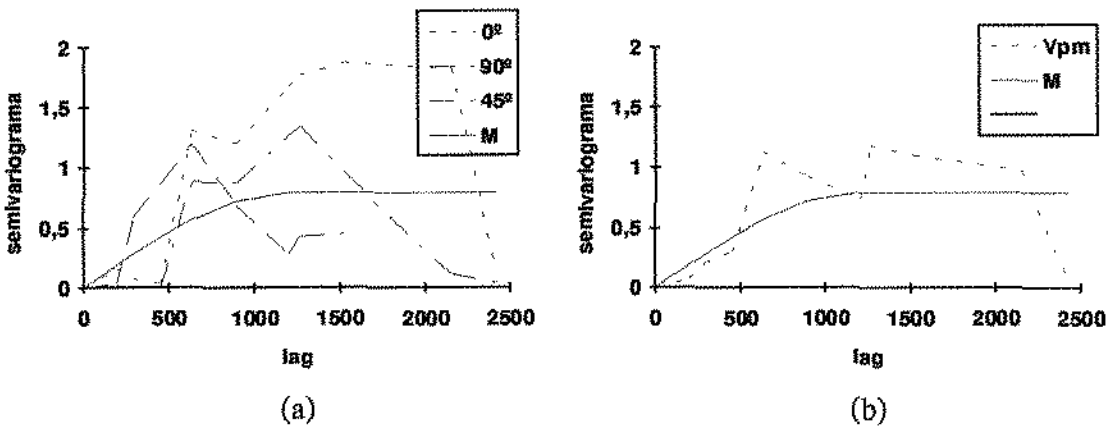


Figura 5.2.2.3. Semivariogramas cruzados experimentais para $V2 \times V1$ (a) em 3 direções e (b) assumindo isotropia dos processos estocásticos geradores. A linha contínua (M) se refere ao modelo esférico escolhido: $a = 1200$, $c_0 = 0$ e $c = 0.80$.

Realizamos a Cokrigagem na estimação da variável V1, utilizando informações sobre V1 e V2, através de uma rotina do GSLIB. Utilizamos os modelos escolhidos, ou seja, consideramos o modelo intrínseco (seção 2.4.3) sem efeito pepita e de amplitude 1200:

$$\gamma_1(h) = 0,91Esf(h/1200)$$

$$\gamma_2(h) = 0,77Esf(h/1200)$$

$$\gamma_{12}(h) = \gamma_{21}(h) = 0,80Esf(h/1200)$$

onde:

$\gamma_1(h)$, $\gamma_2(h)$ $\gamma_{12}(h)$ são os modelos dos semivariogramas para V1, V2 e o semivariograma cruzado entre elas, respectivamente; e $Esf(.)$ é um modelo esférico com patamar $c = 1$ (seção 2.1.2).

Os fatores das estruturas esféricas são positivas, isto é,

$$\begin{vmatrix} 0,91 & 0,80 \\ 0,80 & 0,77 \end{vmatrix} = 0,70 - 0,64 > 0$$

Com o modelo acima, estimamos a variável V1 através da Cokrigagem. O resultado para o item (a) da seção 5.2, interpolado através de Spline, pode ser visto graficamente na Figura 5.2.2.4. As estimativas obtidas podem ser vistas na Tabela A.2 (Apêndice A).

O resultado segundo o item (b) da seção 5.2 pode ser visto graficamente na Figura 5.2.2.5 e as estimativas na Tabela A.3 (Apêndice A).

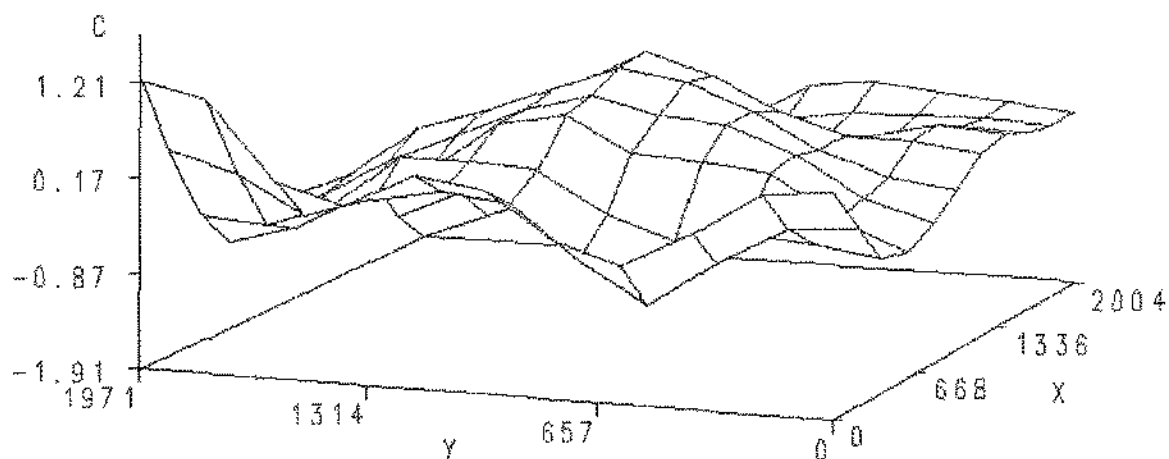


Figura 5.2.2.4. Cokrigagem nos pontos observados com sua prévia retirada (grid para permitir visualização), onde C representa as estimativas interpoladas, x a coordenada E-W e y a coordenada N-S.

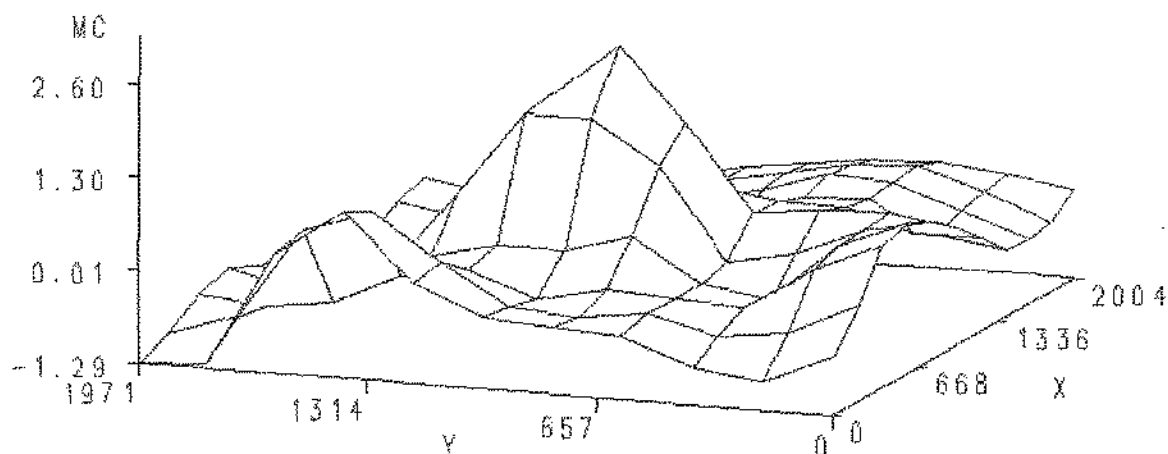


Figura 5.2.2.5. Cokrigagem em uma malha regular utilizando todos os dados disponíveis, onde MC representa as estimativas segundo o item (b), x a coordenada E-W e y a coordenada N-S.

Comparando as Figuras 5.2.2.4 5.2.2.5 com a Figura 5.1.4 (variável V1 através de Spline), podemos notar que os valores da Figura 5.2.2.4 seguem os padrões da variável original, mas com uma suavização. Já a Figura 5.2.2.5 é quase idêntica à Figura 5.1.4.

5.1.2.3. Análise de Componentes Principais

O método de ACP (seção 4.1) exige a hipótese de correlação intrínseca, ou seja, que o conjunto de variáveis possua a mesma estrutura de correlação em todas as escalas espaciais. Pelos semivariogramas de V1 e V2 e seus semivariogramas cruzados, vistos anteriormente, isso é razoável, pois neste caso aparentemente não há a necessidade de se fatorar os semivariogramas. É importante observar novamente que, uma vez que a intenção é a ilustração da técnica, utilizamos os modelos mais simples possíveis.

O conjunto de dados (normalizados) possui 36 observações e 2 variáveis. Utilizando a técnica descrita na seção 4.1 obtemos uma decomposição em componentes principais, resultando em uma matriz transformada $A = E'X$, onde E é a matriz de autovetores e X o conjunto de dados padronizados. Essa transformação foi obtida através de um programa criado para o procedimento CM (linguagem matricial) do SOC (desenvolvido pela EMBRAPA). Das linhas de A (2×36), os semivariogramas foram calculados e modelados. (Cada linha é chamada de Fator; no caso, temos dois fatores).

Os semivariogramas para o Fator 1 podem ser vistos na Figura 5.2.3.1. Escolhemos o modelo esférico com os parâmetros: $a = 1200$, $c_0 = 0$ e $c = 1.34$, e está plotado em linha contínua na Figura 5.2.3.1.

Para o Fator 2, os semivariogramas estão apresentados na Figura 5.2.3.2, e o modelo escolhido foi o modelo esférico com parâmetros: $a = 1200$, $c_0 = 0$ e $c = 0.50$.

Com os modelos já escolhidos realizamos a Krigagem de cada Fator. Com o resultado obtido, foi necessário realizar o retorno à variável original, segundo a fórmula da seção 4.1, que foi feita através do CM. As estimativas obtidas podem ser vistas na Tabela A.2 (Apêndice A).

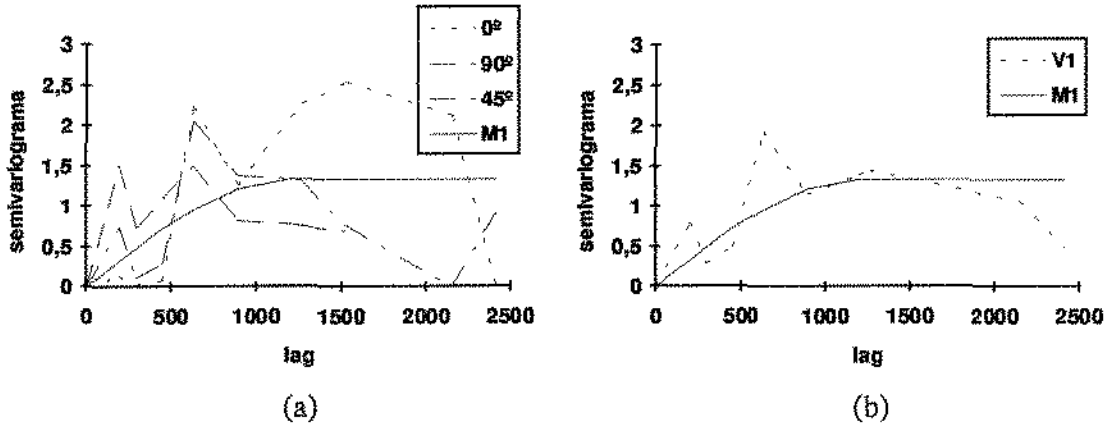


Figura 5.2.3.1. Semivariogramas experimentais para Fator 1 (a) em 3 direções e (b) assumindo isotropia dos processos estocásticos geradores. A linha contínua (M1) se refere ao modelo esférico escolhido: $a = 1200$, $c_0 = 0$ e $c = 1.34$.

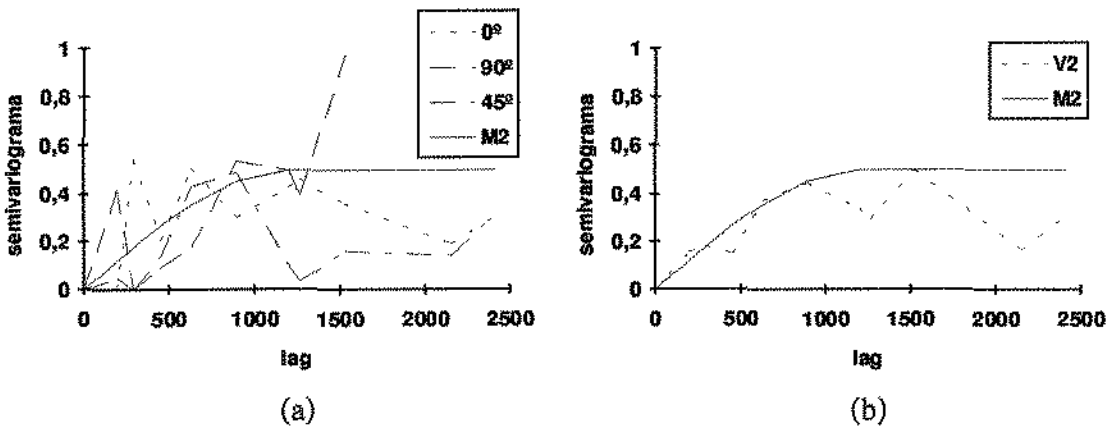


Figura 5.2.3.2. Semivariogramas experimentais para Fator 2 (a) em 3 direções e (b) assumindo isotropia dos processos estocásticos geradores. A linha contínua (M2) se refere ao modelo esférico escolhido: $a = 1200$, $c_0 = 0$ e $c = 0.50$.

Utilizando a técnica de ACP (seção 4.1) obtivemos as estimações para a variável V1. Os resultados, para o item (a), podem ser vistos na Tabela A.2 (Apêndice A), a interpolação através de Spline na Tabela A.3 (Apêndice A), e a visualização gráfica na Figura 5.2.3.3. Os resultados, para o item (b), podem ser vistos na Tabela A.3 (Apêndice A), e a visualização gráfica na Figura 5.2.3.4.

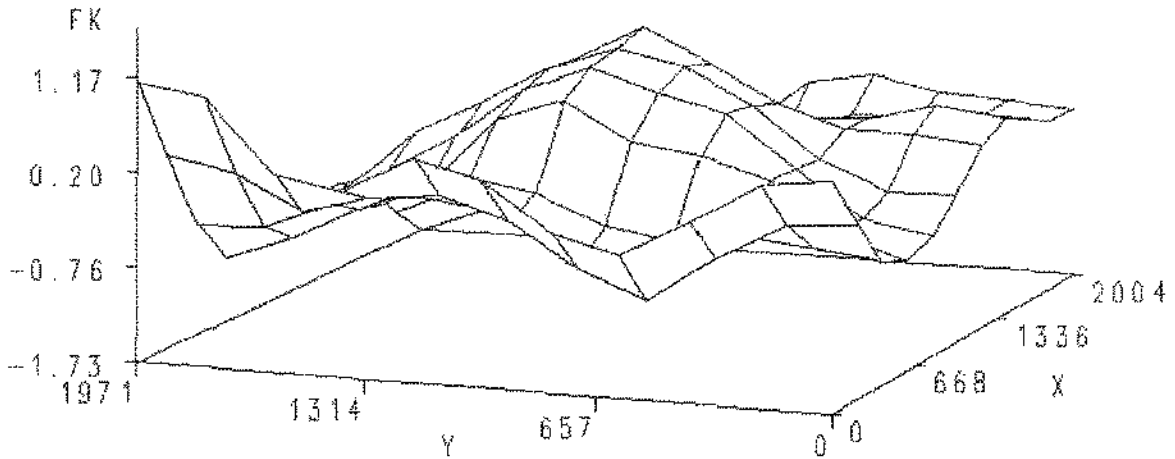


Figura 5.2.3.3. ACP nos pontos observados com sua prévia retirada (grid para permitir visualização), onde FK representa as estimativas interpoladas, x a coordenada E-W e y a coordenada N-S.

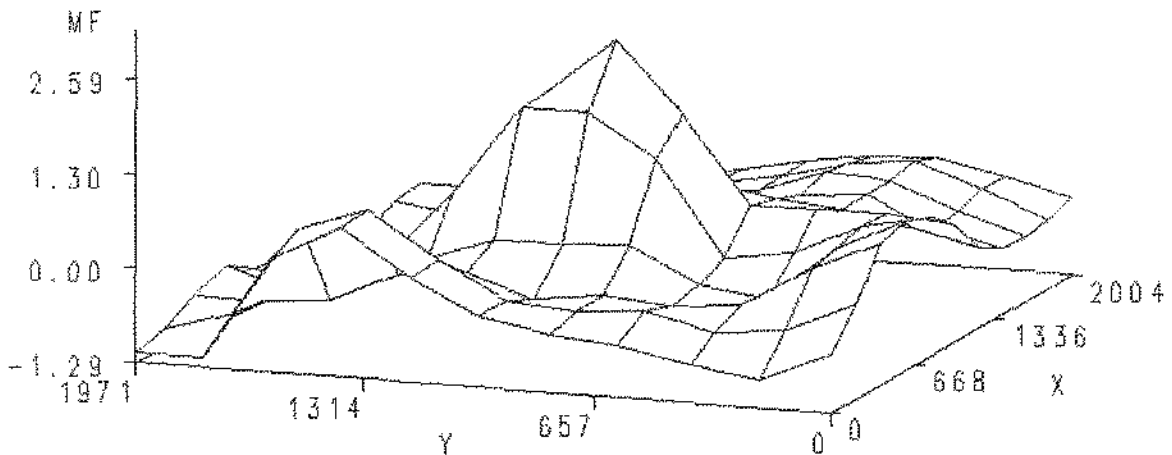


Figura 5.2.3.4. ACP em uma malha regular utilizando todos os dados disponíveis, onde MF representa as estimativas segundo o item (b), x a coordenada E-W e y a coordenada N-S.

Comparando as figuras obtidas (Figura 5.2.1.2 e 5.2.1.3) com a Figura 5.1.4, onde a variável V1 está apresentada, podemos notar que os valores da Figura 5.2.3.3 seguem os padrões da variável original, mas com uma suavização. Já a Figura 5.2.3.4 é muito semelhante à Figura 5.1.4.

5.3. Nota sobre os Cálculos

Com exceção da transformação utilizada na ACP, que foi calculada através de um programa (Apêndice B) para o procedimento CM (linguagem matricial) do SOC (desenvolvido pela EMBRAPA), a ferramenta utilizada para os cálculos foram rotinas (em Fortran) do GSLIB (Deutsch & Journel, 1992), que é um pacote com vários programas geoestatísticos, sendo necessário apenas compilar e executar as rotinas de interesse.

O GSLIB não é um sistema como o SAS, que é de fácil programação. O GSLIB é totalmente particionado, tendo isto suas vantagens e desvantagens. É necessário compilar apenas os programas que serão utilizados, não requerendo um grande espaço no computador. Porém, ao se compilar cada programa, é necessário adequar o tamanho máximo dos parâmetros a serem utilizados, através de um programa de extensão "INC". Assim, é necessário adequar o programa ao seu conjunto particular de dados. Não é recomendável deixar um valor muito grande para utilizá-lo em outro conjunto de dados porque isso pode gerar inúmeros erros na compilação. Depois de compilado e "linkado", a execução do programa exige o arquivo de parâmetros que deve ser previamente adequado ao seu interesse. Esse arquivo de parâmetros não é muito fácil de ser montado. Dependendo do programa a ser utilizado, exige uma grande quantidade de definições de parâmetros, tais como tamanho do *lag*, ângulo de tolerância, etc. Com isso, um leigo no assunto pode encontrar dificuldades para utilizar os programas do GSLIB, sendo necessário contar com a ajuda de uma pessoa com conhecimentos na área de Geoestatística.

Uma outra dificuldade é a falta de interação, por exemplo, no cálculo do semivariograma experimental para grids irregulares, há a necessidade de se encontrar os parâmetros corretos pela tentativa e erro e, para cada tentativa, utilizar um programa externo que visualize os resultados graficamente (no caso, utilizamos o SAS). Se o gráfico não ficar bom, devemos alterar o arquivo de parâmetros, executar o programa e visualizá-lo, tantas vezes quantas forem necessários.

Utilizamos o SAS na criação de grids regulares (através de Spline) para obtermos uma visualização tridimensional.

Os outros gráficos foram produzidos pelo Word 2.0 for Windows.

5.4. Comparação e Comentários

Tabela 5.4.1: "Desvios" das estimativas utilizando as técnicas de Krigagem, de Cokrigagem e de ACP em relação à variável original

	"Desvios" ^(*)
Krigagem (K)	1.543
Cokrigagem (C)	1.051
ACP (A)	1.070

^(*) segundo a fórmula $\frac{1}{36} \sum_{\alpha=1}^{36} [z_l(x_\alpha, y_\alpha) - z'_l(x_\alpha, y_\alpha)]^2$, $l = K, C, A$

Observamos pela Tabela 5.4.1 que, para o critério de comparação numérica adotado no item (a) da seção 5.2, a Cokrigagem obteve um "desvio" menor, seguida pela ACP e, por último a Krigagem. As técnicas de Cokrigagem e ACP obtiveram "desvios" muito próximos, mas a Krigagem obteve um resultado mais distanciado. Pela visualização tridimensional através de Spline (Figuras 5.2.1.2, 5.2.2.4 e 5.2.3.3) notamos que os três resultados obtidos são muito semelhantes, mas com uma suavização, se comparado à variável original V1 (Figura 5.1.4).

Para o item (b), levando-se em conta a apresentação visual, as três técnicas tiveram um resultado semelhante (Figuras 5.2.1.3, 5.2.2.5 e 5.2.3.4), todas com bom desempenho, ao compararmos com a apresentação tridimensional de V1 através de Spline (Figura 5.1.4). (Obs.: é claro que o item (b) apresenta resultados melhores do que o item (a) interpolado, porque (b) trabalha com todos os dados.)

As técnicas refletiram "palidamente" o que diz a teoria: a Cokrigagem é mais precisa, seguida da técnica de ACP e da Krigagem. Mas como a quantidade de dados é pequena (apenas 36) com correlação não muito alta entre V1 e V2 (0.57), e nenhuma locação subamostrada, todas as técnicas obtiveram resultados semelhantes.

Para o conjunto de dados utilizado, o número de variáveis é apenas dois; por isso a Cokrigagem não deu muito mais trabalho do que a ACP, mas é evidente (como já discutido na seção 3.2) que à medida que cresce o número de variáveis, cresce muito o trabalho em se utilizar a Cokrigagem devido ao cálculo de semivariogramas cruzados. É importante salientar também que as experiências têm mostrado que o ganho obtido pela Cokrigagem é pequeno em comparação aos esforços requeridos para alcançá-lo, a menos que uma variável

seja drasticamente subamostrada em comparação com outra (Journel & Huijbregts, 1978, pp. 324).

Quanto à ACP, como não é necessário o cálculo dos semivariogramas cruzados, mas apenas de semivariogramas, o trabalho é menor. Entretanto, é importante lembrar que existe a necessidade da hipótese de correlação intrínseca para que essa técnica possa ser utilizada, e nem sempre isso acontece.

É importante notar também que, como foi visto na seção 3.2, na presença de correionalização intrínseca (que foi o adotado na análise do conjunto de dados reais), a complicação de Cokrigagem pode ser justificada somente se uma das variáveis é subamostrada quando comparada às outras. Nesta análise, a variável V1 tinha praticamente a mesma configuração de V2, exceto no ponto a ser estimado (devido às restrições utilizadas no programa); ou seja, V1 não era drasticamente subamostrada se comparada à V2. Assim, se o objetivo da análise não fosse apenas a ilustração das técnicas, melhor seria realizar apenas a Krigagem pois, como foi dito, a complicação da Cokrigagem não seria justificada, tampouco a da ACP.

Notas:

1. Os variogramas são muito pobres em decorrência da distribuição espacial dos dados. Talvez com melhor distribuição espacial os resultados fossem mais auspiciosos.
2. Os dados referentes à subamostragem (estimação num ponto observado com sua prévia retirada, item a, seção 5.2) foram calculados retirando a informação da variável V1 e V2. Os cálculos retirando somente a informação da variável V1 (para as três técnicas analisadas) também foram realizados, mas como os resultados foram praticamente idênticos, não foram colocados no trabalho.

Capítulo 6

Considerações Finais

O objetivo desse trabalho foi realizar um estudo de métodos de predição multivariados, apresentando-os, na medida do possível, de uma maneira simples. Como vimos, é grande o alcance do ferramental geoestatístico em diversos domínios das Geociências. Uma característica da Geoestatística é a grande interação com as demais técnicas da Estatística e as Ciências da Terra

Dentre as técnicas multivariadas apresentadas, inicialmente foi realizado um estudo sobre a Cokrigagem (Capítulo 3), na estimação de variáveis regionalizadas. Esta técnica é a mais conhecida e utilizada, mas como envolve o cálculo de $p(p+1)/2$ semivariogramas (onde p é o número de variáveis) no caso da existência de isotropia, isso se torna trabalhoso para p maior que 2.

Uma alternativa é a técnica de Análise de Componentes Principais (seção 4.1) na estimação de variáveis regionalizadas. Esta técnica é interessante devido à simplificação do trabalho em relação à Cokrigagem, pois evita o cálculo dos semivariogramas cruzados. Mas é exigida a hipótese de correlação intrínseca, o que reduz o campo de aplicação desta técnica.

Ao invés de estudarmos diretamente a variável regionalizada, podemos estudar uma decomposição da variável em diferentes *componentes espaciais* (representam o comportamento das variáveis regionalizadas para cada escala espacial u ; seção 4.2.1) ou *fatores regionalizados* (refletem os aspectos essenciais da informação multivariada para uma certa escala espacial u ; seção 4.2.2). Isso pode ser feito através da Análise de Krigagem Fatorial (AKF), desenvolvido por Matheron (1982). A AKF torna possível, então, incorporar a característica espacial do fenômeno na análise. Na estimação de fatores regionalizados, as matrizes de correionalização B_u (matrizes simétricas positiva semidefinida) são fatorizadas em um produto de matrizes. Dependendo de como será realizada a fatorização, surgem vários métodos de análise. O método mais comum é a decomposição utilizando a ACP (seção 4.2.2.1), que é apresentada por muitos autores, tais como Wackernagel (1988, 1989) e Wackernagel, Webster e Oliver (1988).

Considerações Finais

Uma outra maneira de estudarmos uma decomposição da variável regionalizada é o método proposto por Bourgault & Marcotte (1991), chamada de Análise Baseada no Variograma Multivariado (seção 4.3). Este método utiliza o variograma multivariado para realçar as estruturas espaciais comuns ao conjunto de variáveis. A pesquisa dos variogramas elementares é realizada utilizando-se somente um variograma (multivariado) ao invés dos $p(p+1)/2$ variogramas (tradicionais) requeridos na AKF.

A técnica de Bourgault & Marcotte (1991) é comparável à técnica proposta por Matheron (1982). Porém, Matheron (1982) não utilizou uma função de estrutura multivariada (como o variograma multivariado), e seu modelo linear de correção espacial difere na existência de p fatores ortogonais para cada escala espacial u , contida nos dados. Essas N_u escalas espaciais são também caracterizadas por variogramas unitários. No modelo de Bourgault & Marcotte (1991), assume-se apenas *um fator* para cada escala espacial.

Os p fatores ortogonais de uma certa escala espacial no modelo de Matheron (1982) são determinados pela fatorização de uma matriz de correlação B_u associada à sua escala espacial. Essas N_u matrizes são obtidas "quebrando" todos os variogramas e variogramas cruzados (tradicionais) das p variáveis. Além de ser demorado, as matrizes obtidas devem ser checadas se são positivas semidefinidas. O modelo de Bourgault & Marcotte (1991) simplifica essa "quebra" utilizando somente um variograma (multivariado). As diferentes matrizes de correção espacial correspondentes às diferentes escalas espaciais serão, por construção, positivas semidefinidas (de posto 1).

Por todas essas razões, a nova aproximação geoestatística (Bourgault & Marcotte, 1991) baseada no variograma multivariado pode simplificar consideravelmente o estudo estrutural do modelo linear de correção espacial.

Uma comparação do modelo de Matheron (1982), utilizando ACP na decomposição das matrizes de correção espacial, com o modelo de Bourgault & Marcotte (1991; seção 4.3) pode ser vista no artigo de Bourgault & Marcotte (1992). Os autores mostraram que, no conjunto de dados reais e simulados utilizado, a técnica de Bourgault & Marcotte (1991) providenciou resultados mais fáceis de interpretar em termos geológicos do que a de Matheron (1982).

Como um trabalho futuro, seria interessante realizar um estudo de um conjunto de dados reais utilizando o método de Matheron (1982) e compará-lo com o método proposto por Bourgault & Marcotte (1991). Como não conseguimos um conjunto adequado de dados reais e sendo necessário programas computacionais específicos para a utilização das duas técnicas, não realizamos esse estudo. Bourgault & Marcotte (1991) sugeriram a utilização

Considerações Finais

da métrica de Mahalanobis no método proposto por eles. Seria interessante empreender também uma tarefa de investigação e apresentação de uma outra métrica. Um outro estudo que poderia ser feito é verificar se os modelos de variograma existentes seriam suficientes ou não para modelar o variograma multivariado.

Na Predição Espacial Multivariada que não foi considerada nesse trabalho, os autores Hoef & Cressie (1993) discutem a utilização de um novo variograma cruzado (baseado em módulos de diferenças) ao invés do tradicional (baseado em quadrados de diferenças). A predição espacial simultânea de várias variáveis baseada no novo variograma cruzado, além de comparações entre o variograma tradicional e o sugerido, foi realizada pelos autores. Hoef & Cressie (1993) alertam também para a necessidade de mais pesquisas na predição espacial multivariada, tal como construção de modelos válidos e estimação do variograma cruzado.

Como pôde ser visto durante o trabalho, a aplicação da Geoestatística é muito ampla e existe a necessidade de técnicas de predição multivariadas mais eficientes e menos trabalhosas, para facilitar a sua utilização.

Algumas das técnicas apresentadas se utilizam de técnicas de Estatística Multivariada, como ACP (seção 4.1 e 4.2.2.1) e a função discriminante (seção 4.2.2.2). (Note que elas são utilizadas apenas como uma ferramenta dentro da Geoestatística.) Foi feita uma "mesclagem" entre técnicas que levam em conta a estrutura espacial e o aspecto aleatório (probabilístico). Por exemplo, a ACP não leva em conta a localização das amostras, mas o variograma e a Krigagem a levam em consideração.

Fica ainda em aberto uma questão: conseguir uma técnica eficiente e simples, combinada com a Estatística, que leve em conta a estrutura espacial das amostras.

Apêndice A

Conjunto de Dados Reais

TABELA A.1. Dados (padronizados) utilizados na estimação, sendo V1 a variável principal e V2 a auxiliar.

X (u.m.)	Y (u.m.)	V1	V2
645,30	3232,00	1,81576	0,92227
320,70	1441,70	-0,18362	-1,04362
255,20	1229,40	0,27276	-0,75238
557,30	1971,40	-0,48244	-0,60676
146,30	1490,70	1,69351	-0,46114
517,90	1340,20	-0,30767	-1,04362
35,80	1751,40	-1,23764	-2,35422
1564,00	1779,50	-1,39158	-0,53395
627,80	1080,20	-0,08220	-0,09708
410,20	1601,50	-0,81748	-2,13579
887,40	1190,30	2,54741	0,26697
1194,30	1932,90	-0,76586	-0,46114
586,50	447,70	-0,64724	0,70384
1261,50	1040,50	-0,50960	-0,24270
1482,70	519,20	0,51906	0,41260
72,30	228,50	-0,99768	-0,89800
2004,00	739,50	0,13060	-0,02427
1703,50	0,00	-0,43445	1,21351
950,00	311,70	0,67209	-0,24270
741,50	819,00	1,07143	1,28633
0,00	1119,90	-0,18905	-0,02427
669,20	1713,40	-0,89354	-1,48049
441,70	76,90	0,79977	1,21351
1110,70	670,70	0,68024	1,14070
219,50	599,20	-0,29319	0,04854
369,30	968,90	-0,54220	-0,16989
1853,50	370,80	0,42942	1,43195
1411,90	1413,50	-0,79122	0,19416
1310,70	170,50	-0,28775	0,92227
1041,40	1559,30	-0,96146	-0,75238
1821,80	1893,30	-0,60378	0,63103
1300,80	1672,30	-0,44984	-0,60676
1000,80	930,30	2,91324	2,30568
1569,50	948,00	0,01741	0,33978
1361,30	737,40	-0,37649	0,04854
850,80	560,30	-0,31854	0,84946

Obs.: As coordenadas estão medidas em u.m. (unidade de medida).

TABELA A.2. Resultados da estimação para a variável V1 através de três técnicas: Krigagem (Krig; seção 2.2), Cokrigagem (Cokr; seção 3.1) e Análise de Componentes Principais (ACP; seção 4.1).

n	X	Y	V1	Krig	Cokr	ACP
1	645,30	3232,00	1,81576	-0,705	-0,515	-0,704
2	320,70	1441,70	-0,18362	0,309	0,358	0,297
3	255,20	1229,40	0,27276	-0,019	0,064	-0,020
4	557,30	1971,40	-0,48244	-1,040	-0,927	-1,041
5	146,30	1490,70	1,69351	-0,455	-0,495	-0,434
6	517,90	1340,20	-0,30767	-0,197	-0,002	-0,203
7	35,80	1751,40	-1,23764	0,884	0,875	0,833
8	1564,00	1779,50	-1,39158	-0,573	-0,613	-0,582
9	627,80	1080,20	-0,08220	0,958	0,811	0,958
10	410,20	1601,50	-0,81748	-0,375	-0,427	-0,406
11	887,40	1190,30	2,54741	0,683	0,548	0,705
12	1194,30	1932,90	-0,76586	-0,841	-0,763	-0,841
13	586,50	447,70	-0,64724	0,180	0,264	0,181
14	1261,50	1040,50	-0,50960	1,075	0,885	1,075
15	1482,70	519,20	0,51906	0,034	-0,111	0,034
16	72,30	228,50	-0,99768	0,279	0,235	0,279
17	2004,00	739,50	0,13060	0,249	0,203	0,249
18	1703,50	0,00	-0,43445	0,094	0,016	0,106
19	950,00	311,70	0,67209	-0,222	-0,006	-0,226
20	741,50	819,00	1,07143	0,737	0,504	0,739
21	0,00	1119,90	-0,18905	0,561	0,450	0,561
22	669,20	1713,40	-0,89354	-0,697	-0,591	-0,702
23	441,70	76,90	0,79977	-0,568	-0,535	-0,566
24	1110,70	670,70	0,68024	0,678	0,642	0,680
25	219,50	599,20	-0,29319	-0,929	-0,995	-0,929
26	369,30	968,90	-0,54220	-0,190	0,129	-0,190
27	1853,50	370,80	0,42942	0,016	0,076	0,024
28	1411,90	1413,50	-0,79122	-0,681	-0,691	-0,682
29	1310,70	170,50	-0,28775	0,416	0,290	0,448
30	1041,40	1559,30	-0,96146	0,101	0,137	0,102
31	1821,80	1893,30	-0,60378	-1,369	-1,494	-1,369
32	1300,80	1672,30	-0,44984	-1,142	-1,044	-1,141
33	1000,80	930,30	2,91324	1,035	0,961	1,048
34	1569,50	948,00	0,01741	-0,683	-0,306	-0,682
35	1361,30	737,40	-0,37649	0,397	0,377	0,397
36	850,80	560,30	-0,31854	0,588	0,542	0,589

Obs. Com exceção das amostras nº 1 e 31, utilizou-se um mínimo de 4 amostras na estimação de cada ponto e lag de tolerância de 1000 u.m.. Para a amostra nº 1, utilizou-se um mínimo de 3 amostras na estimação de cada ponto e lag de tolerância de 1600 u.m.. Para a amostra nº 31, utilizou-se um mínimo de 3 amostras na estimação de cada ponto e lag de tolerância de 1000 u.m. (as amostras nº 1 e 31 tiveram um tratamento diferenciado devido a grande distância das demais amostras, impossibilitando a estimação desses dois pontos com a utilização das restrições gerais para as observações restantes).

TABELA A.3. Resultados da estimação para a variável V1. As colunas referentes a GVI, GKri, GAcp e GCok são os dados da tabela 2 com criação de grids através de Spline, para que se pudesse obter uma visualização tridimensional (Figuras 5.2.1.2, 5.2.2.4 e 5.2.3.3). As três últimas colunas (MKri, MCok e MAcp) se referem a estimações utilizando-se todos os dados com a opção para grid automática (Figuras 5.2.1.3, 5.2.2.5 e 5.2.3.4).

n	X	Y	GVI	GKri	GAcp	GCok	MKri	MAcp	MCok
1	0,0	0,00	-0,84	0,55	0,55	0,49	-0,53	-0,53	-0,52
2	200,4	0,00	-0,02	-0,01	-0,01	-0,03	-0,04	-0,04	-0,01
3	400,8	0,00	0,80	-0,51	-0,50	-0,49	0,63	0,63	0,63
4	601,2	0,00	1,05	-0,62	-0,62	-0,53	0,73	0,73	0,78
5	801,6	0,00	0,95	-0,52	-0,52	-0,38	0,61	0,61	0,59
6	1002,0	0,00	0,59	-0,26	-0,24	-0,16	0,31	0,31	0,32
7	1202,4	0,00	-0,05	0,13	0,16	0,10	-0,17	-0,17	0,00
8	1402,8	0,00	-0,46	0,30	0,33	0,18	-0,36	-0,36	-0,34
9	1603,2	0,00	-0,48	0,17	0,19	0,07	-0,39	-0,39	-0,39
10	1803,6	0,00	-0,35	0,06	0,06	0,01	-0,30	-0,30	-0,29
11	2004,0	0,00	-0,20	0,06	0,06	0,03	-0,15	-0,15	0,01
12	0,0	197,14	-1,17	0,48	0,48	0,42	-0,91	-0,91	-0,92
13	200,4	197,14	-0,44	-0,06	-0,06	-0,08	-0,46	-0,46	-0,44
14	400,8	197,14	0,26	-0,42	-0,42	-0,39	0,19	0,19	0,21
15	601,2	197,14	0,47	-0,40	-0,40	-0,27	0,31	0,31	0,39
16	801,6	197,14	0,68	-0,38	-0,39	-0,17	0,43	0,42	0,42
17	1002,0	197,14	0,69	-0,24	-0,24	-0,07	0,45	0,45	0,48
18	1202,4	197,14	0,05	0,23	0,25	0,20	0,04	0,04	0,01
19	1402,8	197,14	-0,25	0,39	0,42	0,23	-0,18	-0,18	-0,19
20	1603,2	197,14	-0,07	0,18	0,19	0,07	-0,07	-0,07	-0,02
21	1803,6	197,14	0,07	0,05	0,06	0,04	0,07	0,07	0,08
22	2004,0	197,14	0,10	0,09	0,10	0,11	0,07	0,07	0,16
23	0,0	394,28	-1,06	0,06	0,06	-0,05	-0,78	-0,78	-0,78
24	200,4	394,28	-0,54	-0,43	-0,43	-0,49	-0,55	-0,55	-0,56
25	400,8	394,28	-0,38	-0,34	-0,34	-0,31	-0,36	-0,36	-0,37
26	601,2	394,28	-0,47	0,09	0,09	0,20	-0,45	-0,45	-0,44
27	801,6	394,28	-0,08	0,10	0,09	0,24	0,01	0,00	0,01
28	1002,0	394,28	0,43	0,04	0,04	0,17	0,42	0,42	0,46
29	1202,4	394,28	0,27	0,24	0,25	0,21	0,24	0,24	0,31
30	1402,8	394,28	0,28	0,22	0,23	0,06	0,29	0,29	0,25
31	1603,2	394,28	0,46	0,04	0,04	-0,06	0,31	0,31	0,32
32	1803,6	394,28	0,47	0,00	0,00	0,04	0,39	0,39	0,41
33	2004,0	394,28	0,33	0,16	0,16	0,18	0,28	0,27	0,29
34	0,0	591,42	-0,79	-0,30	-0,31	-0,43	-0,59	-0,60	-0,44
35	200,4	591,42	-0,32	-0,91	-0,91	-0,99	-0,34	-0,34	-0,34
36	400,8	591,42	-0,46	-0,46	-0,46	-0,45	-0,45	-0,45	-0,31
37	601,2	591,42	-0,56	0,27	0,28	0,26	-0,34	-0,34	-0,30
38	801,6	591,42	-0,29	0,62	0,62	0,54	-0,15	-0,15	-0,16
39	1002,0	591,42	0,38	0,58	0,58	0,56	0,43	0,43	0,35
40	1202,4	591,42	0,28	0,48	0,48	0,42	0,32	0,32	0,33
41	1402,8	591,42	0,17	0,18	0,18	0,08	0,20	0,20	0,20

Conjunto de Dados Reais

42	1603,2	591,42	0,49	-0,13	-0,13	-0,14	0,35	0,35	0,38
43	1803,6	591,42	0,47	-0,04	-0,03	0,01	0,32	0,32	0,35
44	2004,0	591,42	0,27	0,27	0,27	0,24	0,24	0,24	0,24
45	0,0	788,56	-0,70	-0,13	-0,14	-0,21	-0,51	-0,51	-0,37
46	200,4	788,56	-0,44	-0,64	-0,64	-0,56	-0,40	-0,40	-0,44
47	400,8	788,56	-0,45	-0,35	-0,35	-0,22	-0,35	-0,35	-0,24
48	601,2	788,56	0,03	0,39	0,39	0,30	0,16	0,16	0,31
49	801,6	788,56	1,27	0,79	0,79	0,58	1,15	1,15	1,11
50	1002,0	788,56	1,82	0,91	0,92	0,83	1,60	1,60	1,56
51	1202,4	788,56	0,45	0,84	0,84	0,78	0,53	0,53	0,51
52	1402,8	788,56	-0,42	0,26	0,26	0,30	-0,28	-0,28	-0,31
53	1603,2	788,56	0,18	-0,42	-0,42	-0,19	0,03	0,03	0,09
54	1803,6	788,56	0,30	-0,29	-0,29	-0,13	0,07	0,07	0,17
55	2004,0	788,56	0,11	0,16	0,16	0,13	0,10	0,10	0,12
56	0,0	985,70	-0,55	0,34	0,34	0,27	-0,35	-0,35	-0,32
57	200,4	985,70	-0,41	-0,20	-0,20	-0,02	-0,31	-0,31	-0,38
58	400,8	985,70	-0,56	-0,06	-0,06	0,22	-0,48	-0,48	-0,44
59	601,2	985,70	-0,07	0,79	0,79	0,68	0,11	0,11	0,04
60	801,6	985,70	2,08	1,00	1,01	0,79	1,75	1,75	1,74
61	1002,0	985,70	2,91	1,06	1,07	0,95	2,59	2,59	2,60
62	1202,4	985,70	0,32	1,17	1,17	1,00	0,40	0,40	0,39
63	1402,8	985,70	-0,67	0,27	0,27	0,34	-0,40	-0,40	-0,42
64	1603,2	985,70	0,01	-0,77	-0,77	-0,39	-0,08	-0,08	-0,03
65	1803,6	985,70	0,09	-0,69	-0,69	-0,42	-0,08	-0,08	-0,01
66	2004,0	985,70	0,01	-0,33	-0,33	-0,27	-0,05	-0,05	0,01
67	0,0	1182,84	0,16	0,51	0,51	0,39	0,11	0,11	0,08
68	200,4	1182,84	0,24	0,00	0,00	0,07	0,18	0,18	0,18
69	400,8	1182,84	-0,34	0,07	0,06	0,25	-0,20	-0,20	-0,11
70	601,2	1182,84	-0,08	0,54	0,54	0,53	0,06	0,06	0,05
71	801,6	1182,84	1,93	0,70	0,72	0,57	1,75	1,75	1,75
72	1002,0	1182,84	2,19	0,81	0,83	0,66	1,86	1,86	1,86
73	1202,4	1182,84	0,12	0,80	0,80	0,61	0,20	0,20	0,05
74	1402,8	1182,84	-0,82	0,03	0,03	0,03	-0,59	-0,59	-0,52
75	1603,2	1182,84	-0,54	-0,75	-0,75	-0,53	-0,51	-0,51	-0,38
76	1803,6	1182,84	-0,30	-0,89	-0,90	-0,68	-0,31	-0,31	-0,28
77	2004,0	1182,84	-0,18	-0,74	-0,74	-0,66	-0,26	-0,26	-0,15
78	0,0	1379,98	1,23	0,11	0,12	0,00	0,73	0,73	0,99
79	200,4	1379,98	1,04	-0,08	-0,07	-0,08	0,80	0,80	0,80
80	400,8	1379,98	-0,42	0,11	0,10	0,25	-0,28	-0,28	-0,28
81	601,2	1379,98	-0,19	-0,19	-0,19	-0,04	-0,23	-0,23	-0,15
82	801,6	1379,98	0,58	0,20	0,21	0,22	0,51	0,51	0,52
83	1002,0	1379,98	0,48	0,43	0,44	0,37	0,41	0,41	0,37
84	1202,4	1379,98	-0,29	0,08	0,09	0,01	-0,23	-0,23	-0,41
85	1402,8	1379,98	-0,78	-0,56	-0,56	-0,58	-0,74	-0,74	-0,74
86	1603,2	1379,98	-0,85	-0,88	-0,88	-0,80	-0,75	-0,75	-0,66
87	1803,6	1379,98	-0,65	-0,99	-1,00	-0,91	-0,56	-0,56	-0,49
88	2004,0	1379,98	-0,38	-1,01	-1,02	-1,00	-0,41	-0,41	-0,30
89	0,0	1577,12	0,59	0,28	0,27	0,21	0,24	0,24	0,34
90	200,4	1577,12	0,68	-0,25	-0,25	-0,29	0,49	0,49	0,49
91	400,8	1577,12	-0,76	-0,28	-0,31	-0,32	-0,71	-0,71	-0,71
92	601,2	1577,12	-0,81	-0,48	-0,49	-0,40	-0,71	-0,71	-0,68

Conjunto de Dados Reais

93	801,6	1577,12	-0,69	-0,19	-0,19	-0,10	-0,54	-0,54	-0,56
94	1002,0	1577,12	-0,98	0,08	0,08	0,13	-0,85	-0,86	-0,86
95	1202,4	1577,12	-0,65	-0,58	-0,58	-0,52	-0,65	-0,65	-0,64
96	1402,8	1577,12	-0,72	-0,98	-0,98	-0,94	-0,80	-0,80	-0,76
97	1603,2	1577,12	-1,10	-0,85	-0,86	-0,85	-0,93	-0,93	-0,96
98	1803,6	1577,12	-0,86	-1,03	-1,03	-1,05	-0,74	-0,74	-0,79
99	2004,0	1577,12	-0,46	-1,24	-1,24	-1,30	-0,55	-0,55	-0,48
100	0,0	1774,26	-1,48	1,07	1,01	1,06	-1,17	-1,17	-1,22
101	200,4	1774,26	-0,95	0,15	0,11	0,13	-0,76	-0,76	-0,80
102	400,8	1774,26	-0,88	-0,56	-0,59	-0,58	-0,78	-0,78	-0,82
103	601,2	1774,26	-0,83	-0,84	-0,85	-0,76	-0,81	-0,81	-0,78
104	801,6	1774,26	-0,90	-0,67	-0,67	-0,55	-0,86	-0,86	-0,81
105	1002,0	1774,26	-0,97	-0,56	-0,55	-0,44	-0,86	-0,86	-0,94
106	1202,4	1774,26	-0,67	-0,90	-0,90	-0,80	-0,73	-0,73	-0,67
107	1402,8	1774,26	-0,89	-0,91	-0,91	-0,86	-0,90	-0,90	-0,90
108	1603,2	1774,26	-1,37	-0,61	-0,62	-0,66	-1,28	-1,29	-1,29
109	1803,6	1774,26	-0,83	-1,14	-1,14	-1,23	-0,79	-0,79	-0,81
110	2004,0	1774,26	-0,36	-1,49	-1,49	-1,62	-0,59	-0,59	-0,60
111	0,0	1971,40	-2,23	1,18	1,11	1,21	-1,17	-1,17	-1,27
112	200,4	1971,40	-1,59	0,28	0,23	0,32	-1,02	-1,02	-1,08
113	400,8	1971,40	-0,88	-0,60	-0,62	-0,53	-0,74	-0,74	-0,71
114	601,2	1971,40	-0,47	-1,08	-1,08	-0,96	-0,53	-0,53	-0,53
115	801,6	1971,40	-0,62	-1,02	-1,01	-0,88	-0,76	-0,76	-0,67
116	1002,0	1971,40	-0,74	-0,88	-0,88	-0,77	-0,79	-0,79	-0,74
117	1202,4	1971,40	-0,77	-0,85	-0,85	-0,78	-0,78	-0,78	-0,80
118	1402,8	1971,40	-0,96	-0,83	-0,84	-0,83	-0,92	-0,92	-0,95
119	1603,2	1971,40	-1,04	-0,91	-0,91	-0,98	-0,98	-0,98	-0,95
120	1803,6	1971,40	-0,58	-1,40	-1,40	-1,53	-0,67	-0,67	-0,65
121	2004,0	1971,40	-0,19	-1,73	-1,73	-1,91	-0,57	-0,57	-0,45

- Obs. 1: Para se obter uma visualização tridimensional da variável V_1 , que possui um grid irregular, foi necessário criar um grid regular através de Spline. Os valores estão na coluna denotada por GV_1 .
- Obs. 2: Para se obter uma noção do comportamento das estimativas através de uma visualização tridimensional, os valores da tabela 2 foram interpolados para um grid regular através de Spline: GKri (Krigagem), GAcP (ACP baseada na matriz V) e GCok (Cokrigagem)
- Obs. 3: Utilizando todos os dados (36 observações), foram estimados 121 pontos com a opção de estimação em grids regulares através do GSLIB: MKri (Krigagem), MAcP (ACP baseada na matriz V) e MCok (Cokrigagem)

Apêndice B

Programas Utilizados

Os programas utilizados para o cálculo dos variogramas experimentais, variogramas teóricos, Krigagem e Cokrigagem não serão colocados aqui, uma vez que fazem parte do GSLIB (Deutsch & Journel, 1992).

Os programas utilizados no SAS são simples e não há a necessidade de sua apresentação.

Para a técnica de ACP baseada na matriz V , houve a necessidade do cálculo de algumas matrizes de transformação, e para isso foi utilizado o CM (linguagem matricial) do SOC (desenvolvido pela EMBRAPA). Os programas não são complexos, sendo de fácil implementação, mas como o CM não é muito difundido, uma listagem de um programa-base será apresentado aqui.

Programa 1 para o CM do SOC:

```
anote "Lendo a matriz inicial. Note que são dados padronizados";
exec "MDADOS.DAT";
anote "Lendo vetor de uns";
exec "UM.DAT";

anote "Fazendo os cálculos";
anote "Cálculo da média. Note que vetor de médias é zero, pois são dados normalizados";
media = Csuma(y)/Nlin(y);
anote "Cálculo da matriz de variância-covariância";
anote "Note que será uma matriz com diagonal 1, pois são dados normalizados";
yly = y'*y;
w2 = nlin(y)*media'*media;
w3 = yly-w2;
w4 = w3/(nlin(y)-1);
sn = w4;
```

Programas Utilizados

```
anote "Cálculo da matriz diagonal de desvio padrão";
anote "Como os dados são normalizados, S = I";
q1 = diag(sn);
S = sqrt(q1);

anote "Cálculo dos dados Normalizados";
anote "Esta rotina só é necessária se os dados são normalizados";
anote "Se os dados já forem normalizados, faça Xpn = y";
Xpn = inv(S)*(y'-media'*um');
Xnp = Xpn';
anote "Matriz de correlação dos dados normalizados";
R = (Xpn*Xpn')/(Nlin(Xnp)-1);

anote "Autovalor e autovetor";
L,E=auto(R);
anote "Mudando a ordem dos autovalores e autovetores";
anote "Para obedecer à ordem decrescente";
anote "Autovalor";
a = {0 1, 1 0};
autovalor = l'*a;
anote "Autovetor";
e1 = max(e);
e2 = min(e);
autovetor = {e1 e1 , e1 e2} ;
E = autovetor;

anote "Utilizando o Artigo de Hohn (1991)";
anote "Cálculo da matriz F";
anote "Note que F = E neste caso, pois S = Identidade";
F = S*E;
anote "Cálculo da matriz A";
A = E'*Xpn;
assuma saída = "CALCULO1.OUT";
A;
fim;
```

Programas Utilizados

Tendo a matriz A, passamos então para a krigagem das suas linhas (Hohn, 1991) através de uma das rotinas do GSLIB. Com os resultados obtidos, retorna-se à variável original.

Programa 2 para o CM do SOC:

```
anote "Lendo os fatores krigados";
exec "mafkrig.dat";
krigpn = krig';

anote "Calculando as estimativas";

anote "Note que teria que somar com a média, mas média é zero ";
anote "Note que no artigo se estima ponto por ponto, mas aqui será feito tudo de uma vez.";
anote "Isso é possível diretamente porque a média para as duas variáveis é zero";

anote "A fórmula pontual da estimativa e' ";
anote " est_o(px1) = media(px1) + F'(pxp) * A_o(px1) ";

anote "A fórmula para a matriz com média zero é: ";
anote " est(pxn) = F'(pxp) * A(pxn) ";

estpn = F'*krigpn;
estnp = estpn';

assuma saida = "CALCULO2.OUT";

anote "Os valores krigados em colunas são est(nxp): ";
estnp;
fim;
```

Referências Bibliográficas

- BOURGAULT, G. & MARCOTTE, D. (1991) Multivariable variogram and its application to the linear model of coregionalization. *Mathematical Geology*. 23(7): 899-928.
- BOURGAULT, G.; MARCOTTE, D. and LEGENDRE, P. (1992) The multivariate (co)variogram as a spatial weighting function in classification methods. *Mathematical Geology*. 24(5): 463-478.
- BOURGAULT, G. & MARCOTTE, D. (1992) Spatial filtering under the linear coregionalization model. IN: Soares, A. *Geostatistics Tróia'92*, Vol. 1 , Kluwer Academic Publishers, p. 237-248.
- BRAGA, L.P.V. (1990) *Geoestatística e Aplicações*. Campinas, 9ª SINAPE, 1990.
- CARR, J. (1987) A comparison of Fortran, Pascal and C for variogram computation on a microcomputer. *Computer & Geosciences*. 13(6): 645-654.
- CARR, J.R. & McCALLISTER, P.G. (1985) An application of cokriging for estimation of tripartite earthquake response spectra. *Mathematical Geology*. 17(5): 527-545.
- CARR, J.R. & MYERS, D.E. (1990) Efficiency of different equation solvers in cokriging. *Computer & Geosciences*. 16(5): 705-716.
- CARR, J.R.; MYERS, D.E. and GLASS, C.E. (1985) Cokriging - a computer program. *Computer & Geosciences*. 11(2): 111-128.
- CRESSIE, N. (1989a) Geostatistics. *Mathematical Geology*. 43(4): 197-202.
- CRESSIE, N. (1989b) The many faces os spatial prediction. IN: Armstrong, M. *Geostatistics*, Vol. 1 , Kluwer Academic Publishers, p. 163-176.
- CRESSIE, N. (1990) The origins of kriging. *Mathematical Geology*. 22(3): 239-253.
- DAVIS, J. (1987) Uses and abuses of cross validation in geostatistics. *Mathematical Geology*. 19(3): 241-248.

Referências Bibliográficas

- DAVIS, B.M. & GREENES, K.A. (1983) Estimation using spatially distributed multivariate data: an example with coal quality. *Journal of the International Association for Mathematical Geology*. 15(2): 287-300.
- DELFINER, P. (1979) *Basic Introduction to Geostatistics*. Fontainebleau, Ecole d'Été.
- DEUTSCH, C.V. & JOURNAL, A.G. (1992) *GSLIB: Geostatistical Software Library and User's Guide*. New York, Oxford University Press.
- FREUND, M.J. (1986) Cokriging: multivariable analysis in petroleum exploration. *Computer & Geosciences*. 12(4B): 485-491.
- GALLI, A.; GERDIL-NEUILLET, F. and DADOU, C. (1984) Factorial kriging analysis: a substitute to spectral analysis of magnetic data. IN: Verly, G. et al *Geostatistics for Natural Resources Characterization*, Part 1 , D. Reidel, p. 543-557.
- GALLI, A. & SANDJIVY, L. (1985) Analyse krigeante et analyse spectrale. *Science de la Terre, Série Informatique Géologique*. 21: 115-124.
- GOOVAERTS, P. & SONNET, Ph. (1992) Study of spatial and temporal variations of hydrogeochemical variables using factorial kriging analysis. IN: Soares, A. *Geostatistics Tróia'92*, Vol. 2 , Kluwer Academic Publishers, p. 745-756.
- GOULARD, M. (1989) Inference in a coregionalization model. IN: Armstrong, M. *Geostatistics*, Vol. 1 , Kluwer Academic Publishers, p. 397-408.
- GRUNSKY, E.C. & AGTERBERG, F.P. (1988) Spatial and multivariate analysis of geochemical data from metavolcanic rocks in the Ben Nevis Area, Ontario. *Mathematical Geology*. 20(7): 825-861.
- GRUNSKY, E.C. & AGTERBERG, F.P. (1991a) FUNCORR: a fortran-77 program for computing multivariate spatial autocorrelation. *Computer & Geosciences*. 17(1): 115-131.
- GRUNSKY, E.C. & AGTERBERG, F.P. (1991b) SPFAC: a fortran-77 program for spatial factor analysis of multivariate data. *Computer & Geosciences*. 17(1): 133-160.
- GRUNSKY, E.C. & AGTERBERG, F.P. (1992) Spatial relationships of multivariate data. *Mathematical Geology*. 24(6): 731-758.

Referências Bibliográficas

- GUERRA, P.A.G. (1988) *Geoestatística Operacional*. Brasília - DF, Ministério das Minas e Energia.
- HERMAN, G.T.; LENT, A. and ROWLAND, S.W. (1973) ART: mathematics and applications. *Jour. Theor. Biol.* 42(1): 1-32.
- HINKLEY, D. (1981) Jackknife methods. IN: KOTZ, S. & JOHNSON, N.L. *Encyclopedia of statistical sciences*, Vol. 4, New York, John Wiley, p. 280-287.
- HOEF, J.M.V. & CRESSIE, N. (1993) Multivariable spatial prediction. *Mathematical Geology*. 25(2): 219-240.
- HOEL, P.G.; PORT, S.C. and STONE, C.J. (1972) *Introduction to Stochastic Processes*. Boston, Houghton Mifflin Company.
- HOHN, M.E. (1985) SAS program for quantitative stratigraphic correlation by principal components. *Computer & Geosciences*. 11(4): 471-477.
- HOHN, M.E. (1991) *Geostatistics and Petroleum Geology (Computer Methods in the Geosciences)*. New York, Van Nostrand Reinhold.
- ISAAKS, E.H. & SRIVASTAVA, R.M. (1989) *Applied Geostatistics*. New York, Oxford University Press.
- JAQUET, O. (1989) Factorial kriging analysis applied to geological data from petroleum exploration. *Mathematical Geology*. 21(7): 683-691.
- JOURNEL, A. & HUIJBREGTS, C.J. (1978) *Mining Geostatistics*. London, Academic Press.
- LANA, M.O.A. (1989) *Método Multiquádrico como Alternativa para Predição de Processos Estocásticos*. Campinas, Tese de Mestrado.
- LUCIO, P.S. (1991) *Kriging Disjuntivo Usando Anamorfismo Gaussiano - Um Método Estocástico para Predição de Reservas Recuperáveis*. Campinas, Tese de Mestrado.
- MACKIEWICZ, A. & RATAJCZAK, W. (1993) Principal components analysis (PCA). *Computer & Geosciences*. 19(3): 303-342.

Referências Bibliográficas

- MARCOTTE, D. (1991) Cokriging with matlab. *Computer & Geosciences*. 17(9): 1265-1280.
- MARDIA, K.V.; KENT, J.T. and BIBBY, J.M. (1979) *Multivariate Analysis*. London, Academic Press.
- MARQUARDT, D. (1963) An algorithm for least-squares estimation of nonlinear parameter: SIAM. *Journal on Applied for Mathematics*. 11: 431-441.
- MATHERON, G. (1962) *Traité de Géostatistique Appliquée*. Paris, Technip.
- MATHERON, G. (1963) Principles of geostatistics. *Economic Geology*. 58: 1246-1266.
- MATHERON, G. (1982) *Pour une Analyse Krigeante des Données Régionalisées*. Fontainebleau, Centre de Géostatistique, Report N-732.
- MYERS, D.E. (1982) Estimation of linear combinations and co-kriging. *Journal of the International Association for Mathematical Geology*. 15(5): 633-637.
- MYERS, D.E. (1983) Matrix formulation of co-kriging. *Journal of the International Association for Mathematical Geology*. 14(3): 249-257.
- MYERS, D.E. (1991) Pseudo-cross variograms, positive-definiteness, and cokriging. *Mathematical Geology*. 23(6): 805-816.
- OLEA, R. (1976) *Optimum Mapping Techniques Using Regionalized Variables Theory*. Kansas Geological Survey, Report nº2, Series on Spatial Analysis.
- PAN, G.; MOSS, K. and CARR, J.R. (1992) A fortran program for three-dimensional cokriging with case demonstration. *Computer & Geosciences*. 18(5): 557-578.
- PAWLOWSKY, V. (1984) On spurious spatial covariance between variables of constant sum. *Sci. de la Terre*. Ser. Inf. nº 21: 107-113.
- PAWLOWSKY, V. (1989) Cokriging of regionalized compositions. *Mathematical Geology*. 21(5): 513-521.
- PAWLOWSKY, V. & BURGER, H. (1992) Spatial structure analysis of regionalized compositions. *Mathematical Geology*. 24(6): 675-691.

Referências Bibliográficas

- PETENATE, A.J. (1983) *Optimal Allocation and Other Aspects of Kriging on a Line*. Iowa, Ph. D. Thesis, Iowa State University.
- RASPA, G. et al. (1992) Multivariate geostatistics for soil classification. IN: Soares, A. *Geostatistics Tróia'92*, Vol. 2 , Kluwer Academic Publishers, p. 793-804.
- ROBINSON, G.K. (1991) That BLUP is a good thing: the estimation of random effects. *Statistical Science*. 6(1): 15-51.
- ROUHANI, S. & WACKERNAGEL, H. (1990) Multivariate geostatistical approach to space-time data analysis. *Water Resources Research*. 26(4): 585-591, Apr.
- SANDJIVY, L. (1984) The factorial kriging analysis of regionalized data. Its application to geochemical prospecting. IN: Verly, G. et al *Geostatistics for Natural Resources Characterization*, Part I , D. Reidel, p. 559-571.
- SAS User's Guide* (1982) Cary, NC: *SAS Institute*..
- SCHERVISH, M.J. (1987) A review of multivariate analysis (with discussion). *Statistical Science*. 2: 396-433.
- SCHUCANY, W.R. (1981) Sample reuse. IN: KOTZ, S. & JOHNSON, N.L. *Encyclopedia of Statistical Sciences*, Vol. 8, New York, John Wiley, p. 235-238.
- SMITH, J.D. & ISTOK, J.D. (1989) Multivariate geostatistical analysis of groundwater contamination by pesticide and nitrate: a case history. IN: Armstrong, M. *Geostatistics*, Vol. 2 , Kluwer Academic Publishers, p. 713-724.
- SOUSA, A.J. (1989) Geostatistical data analysis - an application to ore topology. IN: Armstrong, M. *Geostatistics*, Vol. 2 , Kluwer Academic Publishers, p. 851-860.
- STEIN, A; van EIJSBERGEN, A.C. and BARENDREGT, L.G. (1991) Cokriging nonstationary data. *Mathematical Geology*. 23(5): 703-719.
- TANABE, K. (1971) Projection method for solving a singular system of equations and its applications. *Numeriske Mathematike*. 17(3): 203-214.
- TOUGH, J.G. & LEYSHON, P.R. (1985) SPHINX - a program to fit the spherical and exponential models to experimental semi-variograms. *Computer & Geosciences*. 11(1): 95-100.

Referências Bibliográficas

- VALENTE, J.M.G.P. (1982) *Geomatemática - Lições de Geoestatística, vol. III e IV*. Ouro Preto, Fundação Gorceix.
- WACKERNAGEL, H. (1988) Geostatistical techniques for interpreting multivariate spatial information. IN: Chung, C. F. et al. *Quantitative Analysis for Mineral and Energy Resources*. Hingham, D. Reidel, p. 393-409.
- WACKERNAGEL, H. (1989) Description of a computer program for analysing multivariate spatially distributed data. *Computer & Geosciences*. 15(4): 593-598.
- WACKERNAGEL, H.; PETITGAS, P. and TOUFFAIT, Y. (1989) Overview of methods for coregionalization analysis. IN: Armstrong, M. *Geostatistics, Vol. 1*, Kluwer Academic Publishers, p. 409-420.
- WACKERNAGEL, H.; WEBSTER, R. and OLIVER, M.A. (1988) A geostatistical method for segmenting multivariate sequences of soil data. IN: Bock, H.H. *Classification and Related Methods of Data Analysis*. North-Holland, Elsevier, p. 641-650.
- ZHOU, D. (1989) ROPCA: a fortran program for robust principal component analysis. *Computer & Geosciences*. 15(1): 59-78.

Olavo Indalho!

Paulo Farias