

## 4 Análise Estatística e Geoestatística dos Dados

Os dados coletados em campo, juntamente com os resultados obtidos pelos ensaios laboratoriais, conforme descrições no Capítulo 3 (Apresentação do Programa Experimental), foram processados e interpretados através de métodos estatísticos e geoestatísticos, com o auxílio de aplicativos computacionais.

Na seqüência será apresentada uma breve introdução de como estes dados foram analisados.

### 4.1. Análise Estatística

A análise estatística dos dados foi processada em duas etapas. A primeira fase consistiu na análise estatística descritiva clássica ou análise exploratória dos dados. Na segunda fase, foram utilizados métodos geoestatísticos como as interpolações por inverso da distância e análises semivariográficas para estudar a variabilidade espacial dos dados coletados em campo.

#### 4.1.1. Análise Estatística Descritiva Clássica

A análise estatística descritiva clássica visou caracterizar as posições centrais de dispersão dos dados, além de expressar a forma da distribuição. Para isso foram calculadas medidas de posição, medidas de dispersão e medidas de forma de distribuição. Os itens abaixo descrevem essas medidas.

##### 4.1.1.1. Medidas de Posição

Dentre as medidas de posição calculadas pode-se citar.

a. Média Aritmética ( $\bar{X}$ ): consiste na média simples dos dados

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

Equação 4.1

b. Moda: é o valor que ocorre com maior freqüência em uma amostra, esse dado pode não existir ou ainda quando existir pode apresentar mais de um valor.

c. Mediana e Quartis: os quartis dividem os dados ordenados em ordem crescente em quatro partes iguais, sendo 25% dos dados inferiores ao primeiro quartil, 50% dos dados inferiores ao segundo quartil (esse valor é denominado de mediana), 75% dos dados serão inferiores e 25% superiores ao terceiro quartil.

d. Máximo e Mínimo: maiores e menores valores dos dados amostrados.

e. Erro Padrão da Média: esse valor apresenta a precisão da estimativa da média e pode ser obtido pela fórmula:

$$S_x = \frac{S}{\sqrt{n}}$$

Equação 4.2

f. Desvio entre Quartis (dQ): é a diferença entre o terceiro quartil e o primeiro quartil.

#### 4.1.1.2. Medidas de Dispersão

Dentre as medidas de dispersão calculadas pode-se citar.

a. Amplitude: amplitude dos dados é a diferença entre o maior e o menor valor presentes nos dados amostrados.

b. Variância ( $S^2$ ): é a medida obtida por meio da fórmula abaixo.

$$S^2 = \frac{\sum_{i=1}^n (x_i - \bar{X})^2}{n - 1}$$

Equação 4.3

c. Desvio Padrão (S): equivale a raiz quadrada positiva da variância.

### 4.1.1.3. Medidas de Forma da Distribuição

a. Coeficiente de Variação (CV): é a medida da razão entre o desvio padrão e a média aritmética. O resultado apresenta quão maior ou menor é o desvio padrão da média. Este coeficiente dá uma idéia da regularidade ou homogeneidade das amostras que estão sendo estudadas. Segundo Andriotti, (2003) valores elevados (superiores a um), representam amostras com grande heterogeneidade e valores abaixo de cerca de 0,40 refletem homogeneidade da amostra.

b. Coeficiente de Assimetria ( $A_3$ ): assimetria é o grau de desvio de uma curva no sentido horizontal, podendo esse desvio ser positivo ou negativo. O coeficiente de assimetria pode ser calculado por:

$$A_3 = \frac{\sum_{i=1}^n (X_i - \bar{X})^3}{n \cdot S^3} \quad \text{Equação 4.4}$$

Para distribuições assimétricas, a média tende a se situar do mesmo lado da moda, onde a cauda da curva de distribuição é mais longa; por isso, outra medida de assimetria pode ser obtida pela seguinte fórmula:

$$\text{assimetria} = \left[ \frac{(\text{média} - \text{moda})}{S} \right] = \left[ \frac{3(\text{média} - \text{mediana})}{S} \right] \quad \text{Equação 4.5}$$

Os dados são considerados simétricos para  $|A_3| < 0,15$ , assimétrica moderada para  $0,15 \leq |A_3| < 1$  e assimétrica forte para  $|A_3| \geq 1$ .

c. Coeficiente de Curtose: curtose é o grau de achatamento de uma curva em relação a uma curva representativa de uma distribuição normal. O coeficiente de curtose pode ser calculado pela fórmula:

$$A_4 = \frac{\sum_{i=1}^n (X_i - \bar{X})^4}{n \cdot S^4} \quad \text{Equação 4.6}$$

Uma curva apresenta bom ajuste ao modelo normal se obtiver assimetria entre  $\pm 3 \left[ \sqrt{\left( \frac{6}{n} \right)} \right]$  e (curtose -3) entre os valores  $\pm 6 \left[ \sqrt{\left( \frac{6}{n} \right)} \right]$ .

A classificação segundo o coeficiente de curtose segue os seguintes intervalos:

- Se o valor do coeficiente de curtose for igual a 3, tem-se o mesmo achatamento que o da distribuição normal. Assim sendo estas funções são denominadas de mesocúrticas.
- Se o valor do coeficiente de curtose for maior que 3, a distribuição apresenta uma curva mais alta (afunilada) e concentrada se comparado a distribuição normal. O nome dado a esse tipo de função é leptocúrtica.
- Se o valor do coeficiente de curtose for menor que 3 a função distribuição é mais “achatada” se comparada a distribuição normal, neste caso a função é denominada platicúrtica.

#### 4.1.1.4. Correlação Linear Simples

No presente estudo foram coletados diversos dados dos pontos de amostragem. Para verificar se havia correlação entre esses dados, foi calculado o coeficiente de correlação linear.

O fato de existir uma forte associação entre duas variáveis não significa necessariamente que haverá uma relação de causa efeito entre elas. O coeficiente de correlação é representado por  $r$ . Os valores de  $r$  variam de -1 (correlação inversa) e +1 (correlação direta), sendo o valor zero a representação de ausência de correlação linear. Cabe ressaltar que não ter correlação linear não implica que as variáveis não tenham outro tipo de correlação.

O valor de  $r$  pode ser obtido por:

$$r = \frac{\sum_{i=1}^n [(X_i - \bar{X})(Y_i - \bar{Y})]}{[(n-1)S_x \cdot S_y]} \quad \text{Equação 4.7}$$

Onde  $S_x$  e  $S_y$  representam respectivamente os desvios padrões das variáveis  $X$  e  $Y$  da amostra analisada.

#### 4.1.1.5. Histograma

A maneira mais usual de representação gráfica de dados amostras é o histograma. O histograma consiste em um grupo de retângulos que tem por base o eixo das abscissas e por ponto médio o valor central do intervalo de classe. Caso esses intervalos sejam iguais as alturas desses retângulos serão

proporcionais às frequências de classe, com isso tomam-se as alturas como iguais as frequências.

A construção de histogramas é uma das representações preliminares de dados estatísticos. Através desta representação pode-se observar a distribuição dos dados e verificar se a distribuição se aproxima de alguma função já conhecida como a normal.

De acordo com Dayrell (2000) os histogramas assim como os mapas de distribuição são uma poderosa ferramenta para a descrição espacial dos dados através da sua visualização.

#### **4.1.2. Análise Geoestatística**

A geoestatística é uma divisão da estatística que alia o conceito de variáveis aleatórias com o conceito de variáveis regionalizadas, gerando assim um conceito de funções aleatórias. Seu uso começou a ser difundido nas décadas de 60 e 70 pelas empresas mineradoras, e no final dos anos 80 pela indústria de petróleo.

Segundo Olea (1991) a geoestatística pode ser definida como o “estudo de fenômenos que variam no espaço”. Para a análise desse conceito muitas vezes se faz necessário a utilização de pacotes computacionais.

Diferente da estatística clássica, a geoestatística permite separar a variabilidade explicada pela relação existente entre amostras e a variabilidade casual. Uma importante utilização desta ferramenta matemática é a possibilidade da diminuição do número de amostras necessárias para descrever as características locais geológicas. Sua aplicação pode reduzir custos relacionados aos estudos preliminares de regiões mantendo a precisão das estimativas realizadas (Vieira *et al.*, 1983).

De acordo com Isaaks (1989) o sucesso de qualquer método geoestatístico depende da confiabilidade dos dados.

A natureza não apresenta variáveis com um padrão de distribuição usual da estatística clássica, tais como: a normalidade e a independência dos dados. A estatística clássica apresenta modelos geralmente relacionados à verificação da distribuição de frequência dos dados; enquanto a geoestatística relaciona a interpretação da distribuição estatística, bem como a correlação espacial das amostras. Com isso a geoestatística associa a distribuição estatística dos dados no espaço, ferramenta de extrema importância para análise de dados geotécnicos.

Como os métodos geoestatísticos fornecem uma base estatística que aliam os dados às coordenadas espaciais das observações, para o presente estudo, a utilização destes métodos foi essencial para a interpretação dos dados coletados em campo. A análise dos dados foi elaborada através da confecção de modelos espaciais dos dados amostrados.

Para a geração dos modelos espaciais de dados, foi necessária uma inserção dos dados dos pontos amostrados em campo juntamente com as coordenadas georreferenciadas destes pontos. Assim o modelo espacial foi gerado por representação computacional/matemática que permitiu estimar e/ou simular os valores de localizações não amostradas.

A malha amostral para esse estudo foi uma malha retangular de aproximadamente 5.040m x 8160m, com os pontos distando entre si de 240m.

A posição relativa das amostras pode ser classificada em regular, semi-regular e irregular. Sendo que a amostragem regular apresenta posição espacial (x,y) das amostras com regularidade de distribuição. A amostragem semi-regular apresenta amostragem com regularidade de distribuição espacial na direção x ou y, mas não nas duas ao mesmo tempo. E por fim a amostragem irregular não apresenta regularidade de distribuição espacial.

A amostragem realizada no campo petrolífero pode ser generalizada e denominada como uma amostragem regular, apesar de alguns pontos amostrados ficarem fora da regularidade da malha, devido a dificuldades de acesso a certas localidades, incluindo propriedades privadas, áreas alagadas, taludes íngremes.

#### **4.1.2.1. Métodos de Interpolação**

Para a confecção dos modelos digitais de grade regular foi necessário definir as funções interpolantes locais. Essas definições são dadas para cada elemento da malha, assim cada elemento básico da grade retangular define uma função interpolante que é válida para os pontos internos ao elemento, esse processo é denominado ajuste de superfície. As funções de ajuste em geral são polinômios que são definidos utilizando-se os vértices dos elementos, e muitas vezes os vértices dos elementos vizinhos também.

A representação das variações dos fenômenos foi expressa através dos interpoladores de tendência, a partir de regressões polinomiais. O ajuste das superfícies polinomial referentes a todos os dados amostrais e minimizam os erros de ajuste global. Em geral utiliza-se uma minimização do erro médio

quadrático, como mostra a Figura 4.1. Observa-se que o polinômio de ordem maior apresenta uma melhor representatividade dos pontos amostrados.

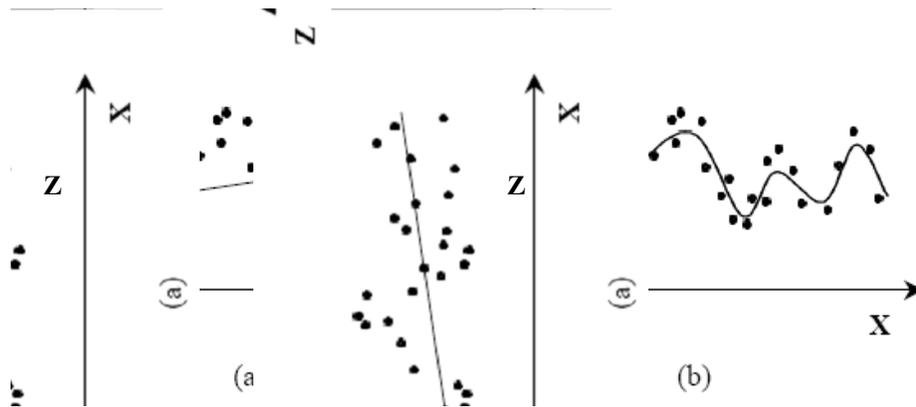


Figura 4.1– Curva polinomial ajustada para representar a variação de um conjunto de amostra (a) linear e (b) ordem maior. (INPE 1998).

No presente trabalho avaliaram-se os diversos tipos de interpolações incluindo aquelas que podem ser realizadas através de dados pontuais, linhas e áreas. Os interpoladores globais utilizam uma única função para toda a área em estudo. Os interpoladores locais usam a repetição de algoritmos a subconjuntos de pontos amostrados. Por sua vez, os interpoladores exatos referem-se a dados existentes, enquanto os interpoladores aproximados assumem incertezas (erros) nos dados existentes.

Durante o estudo diversas maneiras de interpolar os dados foram analisadas, nos tópicos a seguir serão apresentados esses métodos.

#### Interpolação Através da Média Móvel

Um dos métodos mais simples para a estimativa dos valores de cota dos pontos de uma grade regular é denominado de interpolação através da média móvel ou interpolação através da média das cotas das amostras vizinhas. A regra matemática que gera essa interpolação é:

$$Z_i = \frac{\sum_{j=1}^n W_{ij} Z_{ij}}{\sum_{j=1}^n W_{ij}}$$

Equação 4.8

Em que:

- $z_i$ : é o valor da cota de um ponto  $i$  qualquer da grade;
- $z_j$ : é a cota de uma amostra  $j$  vizinha do ponto  $i$  da grade;
- $w_{ij}$ : é um fator de ponderação.

### Interpolação por Vizinho mais Próximo

A interpolação por vizinho mais próximo é definido pela escolha de apenas uma amostra vizinha para cada ponto da grade, isto é, quando  $j$  igual a 1. Neste método para cada nó da rede é atribuído o valor do ponto mais próximo.

### Interpolação por Média Ponderada

A interpolação por média ponderada é um método onde o valor de cota de cada elemento da grade é definido pela média ponderada dos valores de cota das amostras vizinhas. A ponderação mais usada na prática é o inverso da distância euclidiana do ponto da grade à amostra considerada, isto é, calcula-se o valor de um ponto através da média dos pontos mais próximos. A média é ponderada pelo inverso da distância entre os pontos, a formulação para esse método pode ser observada na Equação 4.9:

$$Z = \frac{\sum_{i=1}^n \frac{Z_i}{h_{ij}^\beta}}{\sum_{i=1}^n \frac{1}{h_{ij}^\beta}}$$

Equação 4.9

Em que:

- Z: é o valor interpolado para o nó da grade;
- $Z_i$ : é o valor do ponto amostrado vizinho ao nó;
- $h_{ij}$ : distância entre o nó da grade e  $Z_i$ ;
- $\beta$ : expoente de ponderação (peso);
- n: número de pontos amostrados utilizados para interpolar cada nó.

O inverso da distância é utilizado para atenuar a influência de pontos distantes, e é fundamentado no pressuposto de existência de correlação espacial positiva. Dentre as características do método pode-se listar que:

- a. É o método mais utilizado dentre os métodos de média ponderada;
- b. É bastante utilizado para a geração de modelos digitais de terreno;
- c. O peso dado na interpolação influencia um ponto amostrado, com relação a outro, diminuindo-o quando a distância entre eles aumenta;
- d. Pontos mais próximos do ponto a ser estimado, recebem ponderação maior que os pontos locados a uma maior distância;

e. A soma de todos os pesos dados aos pontos vizinhos é igual a um, i.e., é atribuído um peso proporcional para a contribuição de cada ponto vizinho quando se calcula o valor de um ponto;

f. A posição de uma observação é coincidente com um ponto, quando o seu valor recebe peso um, e os pesos dos pontos vizinhos recebem valor zero, com isso o resultado será o próprio nó recebendo o valor exato da observação;

g. Pode-se escolher o expoente de ponderação.

Apesar de apresentar vantagens esse método apresenta algumas limitações. A Tabela 4.1, definida por Landim, (2000) mostra esse comparativo do método do inverso da distância com os demais métodos.

Intuitivamente pode-se notar que quanto maior a distância que uma amostra se encontra do ponto da grade, menor deve ser a sua influência (ponderação) na apuração final do valor de cota do ponto da grade.

#### **Interpolação Acumulativa**

A interpolação acumulativa é o método quando os nós são representados pela soma dos valores dos pontos de uma dada região. Este método é válido quando há presença de dados ponderados.

#### **Interpolação por Ponderação Direcional**

A ponderação direcional é um método de interpolação que utiliza o inverso da distância aliado a uma ponderação direcional. Pode-se especificar essa direção como uni-direcional, bi-direcional ou diversas direções pré definidas.

#### **Interpolação por Superfície de Tendência Polinomial**

A interpolação por superfície de tendência polinomial é um método que busca encontrar as tendências regionais dos dados. A tendência é ajustada através de polinômios de ordens diversas. Esses polinômios são gerados por critérios de regressão por mínimos quadrados. Após o ajuste da tendência é determinado um fator de correlação, sendo que quanto maior o valor deste fator melhor será o ajuste estipulado.

A Tabela 4.2, definida por Landim, (2000) apresenta algumas vantagens e limitações deste método de interpolação.

Tabela 4.1 – Vantagens e desvantagens do interpolador inverso ponderado da distância (Landim, 2000)

<b>Vantagens</b>	<b>Desvantagens</b>
<ul style="list-style-type: none"> <li>• Fácil de entender matematicamente;</li> <li>• Algoritmo bem conhecido e discutido;</li> <li>• Disponível em muitos softwares;</li> <li>• Utiliza pouco tempo de computação;</li> <li>• É razoavelmente fiel aos valores amostrados originais;</li> <li>• Não estima valores de pontos amostrados maiores ou menores que os valores máximos e mínimos dos dados;</li> <li>• Bom para estimativas de espessura, concentração química e propriedades físicas;</li> <li>• Muito bom para analisar variações de pequena amplitude (anomalias) entre os dados irregularmente distribuídos;</li> <li>• Bom estimador para propósitos gerais;</li> <li>• Pode interpolar por setor, onde o usuário pode escolher a quantidade de amostras, por setor, que serão levadas em consideração, em uma determinada estimativa de valor.</li> </ul>	<ul style="list-style-type: none"> <li>• Não estima valores de <math>Z_i</math> maiores ou menores que os valores máximos e mínimos dos dados;</li> <li>• Ruim para estimativas de dados estruturais, isto é, para determinar o topo ou a base de horizontes estratigráficos ou litológicos;</li> <li>• Muito difícil obter a localização precisa de uma determinada isolinha se os valores amostrados não contiverem este valor;</li> <li>• É difícil projetar a curva de altitude zero se os dados não contiverem zero ou valores negativos;</li> <li>• Influência de valores locais anômalos é difícilmente removida;</li> <li>• Dados dispostos em conjuntos de semelhantes podem influenciar as estimativas de modo bastante tendencioso;</li> <li>• Cria muitas observações ilusórias devido as limitações do método de interpolação, o que pode ser reduzido ou eliminado se o raio de busca for reduzido, se a tendência for removida previamente e se ocorrer mudanças no expoente utilizando, no tamanho da grade, no número de pontos utilizados e ponderação direcional.</li> </ul>
<b>Quando Usar</b>	<b>Quando não Usar</b>
<ul style="list-style-type: none"> <li>• Estimativas de propósito gerais;</li> <li>• Distribuição uniforme de dados;</li> <li>• Para destacar anomalias locais;</li> <li>• Para calcular volume em operações entre grades.</li> </ul>	<ul style="list-style-type: none"> <li>• Dados agrupados;</li> <li>• Tendência pronunciada presente;</li> <li>• Dados com falhas e distribuídos de forma esparsa;</li> <li>• Não quiser imperfeições no modelo.</li> </ul>

Tabela 4.2 – Vantagens e desvantagens do interpolador Superfície de Tendência Polinomial (Landim, 2000).

<b>Vantagens</b>	<b>Desvantagens</b>
<ul style="list-style-type: none"> <li>• Uma única superfície é gerada;</li> <li>• Fácil definição de parâmetros;</li> <li>• A mesma superfície é grada mesmo com mudança na orientação da grade;</li> <li>• Tempo para cálculo de superfícies de baixa ordem é baixo;</li> <li>• Contempla tanto as tendências regionais quanto anomalias locais;</li> <li>• Estima valores acima e abaixo dos amostrados.</li> </ul>	<ul style="list-style-type: none"> <li>• Extrapola valores de Z para além dos limites da área amostrada;</li> <li>• Anomalias locais não são vistas em mapas de superfície de baixa ordem;</li> <li>• Utilização torna-se facilmente abusiva: a tentação de especificar um ajuste de uma superfície de alta ordem pode ser maior do que o bom senso quanto ao resultado;</li> <li>• A memória RAM necessária aumenta exponencialmente com o aumento da ordem do polinômio.</li> </ul>
<b>Quando Usar</b>	<b>Quando não Usar</b>
<ul style="list-style-type: none"> <li>• Número adequado de pontos amostrados estiver disponível, sempre maior que o número de coeficientes da equação: <ul style="list-style-type: none"> <li>• Ordem=1, coeficiente=2, pontos <math>\geq 3</math></li> <li>• Ordem=2, coeficiente=5, pontos <math>\geq 5</math></li> </ul> </li> <li>• Dados regularmente distribuídos;</li> <li>• Gerar novos dados em áreas com dados esparsos.</li> </ul>	<ul style="list-style-type: none"> <li>• Poucos dados, com distribuição irregular ou para a superfície real com alta variabilidade local;</li> <li>• Pontos amostrados em conjuntos semelhantes e valores de Z altamente variáveis;</li> <li>• Superfície com descontinuidade devido a falhas ou inconformidades;</li> <li>• Quando a amplitude da superfície variar drasticamente ou erráticamente devido a anomalias locais de grande variação.</li> </ul>

### **Interpolação por Superfície de Tendência Residual**

A tendência de superfície residual apresenta a primeira etapa semelhante ao método de Tendência de Superfície Polinomial. E a segunda etapa consiste na comparação dos dados com a superfície calculada. Geralmente os dados comparados com a superfície calculada são dados interpolados com o inverso da distância.

### Interpolação por Triangulação

A triangulação é um método que conecta os pontos amostrados através de triângulos. Os dados dos pontos são os vértices dos triângulos formados por linhas imaginárias. Tenta-se formar ângulos o mais equi-angulares possíveis. Para cada nó da rede que se encontra dentro de cada triângulo é atribuído um valor com base na intercepção com o plano inclinado triangular. Para os nós de rede que ficam dispostos fora da rede de triângulos podem ser atribuído valores usando o inverso da distância. Este método é considerado um método de estimação direta, pois os contornos derivam do padrão original dos dados, não permitindo a extrapolação, as estimativas limitam-se estritamente à área amostrada.

A Tabela 4.3, definida por Landim, (2000) apresenta algumas vantagens e limitações deste método de interpolação.

Tabela 4.3 – Vantagens e desvantagens do interpolador Triangulação (Landim, 2000).

<b>Vantagens</b>	<b>Desvantagens</b>
<ul style="list-style-type: none"> <li>• Fácil de ser entendido;</li> <li>• Rápido;</li> <li>• Fiel aos dados originais;</li> <li>• Bom para uma visualização rápida;</li> <li>• Superfície pode ser interpolada entre os pontos amostrados.</li> </ul>	<ul style="list-style-type: none"> <li>• Valores acima ou abaixo dos valores reais não podem ser extrapolados;</li> <li>• Valores além da área de distribuição dos pontos amostrados não podem ser extrapolados;</li> <li>• Contornos podem ser “angulares” nas bordas dos triângulos;</li> <li>• Gera superfícies angulares;</li> <li>• Rede triangular não é única e isso pode distorcer os resultados.</li> </ul>
<b>Quando Usar</b>	<b>Quando não Usar</b>
<ul style="list-style-type: none"> <li>• Distribuição de dados bem regular;</li> <li>• Valores estimados necessários apenas dentro da área amostrada;</li> <li>• Presença de grandes diferenças entre os dados (encostas íngremes, dados de contaminação com grande variabilidade), dados topográficos.</li> </ul>	<ul style="list-style-type: none"> <li>• Ao se desejar um mapa com contornos suaves;</li> <li>• Alguns dados coletados em linhas e outros com distribuição irregular;</li> <li>• Com grande volume de dados a computação pode ser lenta.</li> </ul>

### Análise Semivariográfica e Interpolação por Krigagem

O semivariograma é uma função intrínseca que mostra a estrutura do fenômeno estudado. As relações estatísticas são medidas pelas covariâncias que existem entre as amostras espaçadas de valores  $h$ .

O semivariograma é definido como a variância do erro que se comete ao estimar um teor desconhecido em  $(x+h)$  pela ajuda de um ponto dado em  $(x)$ . O semivariograma em uma determinada direção indica quão irregular os valores se tornam quando a distância de medida aumenta.

Segundo Salviano (1996) o semivariograma analisa o grau de dependência espacial entre as amostras dentro de um campo experimental, além de definir parâmetros necessários para a estimativa de valores para locais não amostrados, através da técnica de krigagem. O variograma é expresso por  $2\gamma(h)$  e o semivariograma por  $\gamma(h)$ .

Matematicamente o semivariograma é a esperança matemática do quadrado dos acréscimos da variável regionalizada em estudo em uma determinada direção ou o valor médio do quadrado das diferenças entre todos os pares de pontos presentes na área estudada, tomados a uma distância  $h$  uns dos outros, e não depende dos pontos de apoio, mas do espaçamento entre eles.

A expressão do semivariograma é definida por:

$$\gamma(h) = \left[ \left( \frac{1}{2N(h)} \right) \sum_{i=1}^{N(h)} [z(x_i) - z(x_i + h)]^2 \right] \quad \text{Equação 4.10}$$

Sendo  $N(h)$  o número de pares de valores separados entre si por uma magnitude  $h$  na direção deste vetor. Usualmente plota-se os valores de  $\gamma(h)$  no eixo das ordenadas e os valores de  $h$  no eixo das abscissas, fazendo as duas origens se coincidirem no zero.

Do ajuste do modelo matemático aos valores calculados de  $\gamma(h)$  são estimados os coeficientes do modelo teórico para o semivariograma (efeito pepita  $C_0$ , patamar  $C_0+C_1$ , alcance  $A_0$ ).

O efeito pepita ( $C_0$ ) é o valor da variabilidade  $\gamma(h)$  para a distância igual a zero, i.e. para  $h=0$ . Em teoria este valor deveria ser nulo, entretanto devido a vários fatores como erros de amostragens, espaçamentos irregulares, erros de medidas ou ainda micro-regionalizações da variável em análise, casam uma descontinuidade na origem do semivariograma, denominada de efeito pepita.

Este parâmetro é importante, pois reflete o erro analítico, indicando uma variabilidade não explicada de um ponto para outro.

O patamar ( $C_0+C_1$ ) representa o nível de variabilidade onde o semivariograma se estabiliza. O patamar é atingido quando a variância dos dados se torna constante com a distância entre as amostras. Em teoria este patamar deve ser igual à variância total dos dados. É um parâmetro importante, pois permite a determinação da distância limite entre dependência e independência entre as amostras (Silva, 1988).

O alcance ( $A_0$ ) ou amplitude semivariográfica é a distância a partir da qual as amostras passam a ser independentes, ou seja, a partir da qual a variação média entre duas observações não é mais função da distância entre elas, dando lugar a independência, sem correlações espaciais. O alcance reflete o grau de homogeneização entre as amostras, ou seja, quanto maior for o seu valor, mais homogêneo será o fenômeno ou processo estudado. O alcance do semivariograma representa a zona de influência de uma observação e separa o campo estruturado (amostras correlacionadas) do campo aleatório (amostras independentes).

A Figura 4.2 apresenta um modelo de semivariograma com os parâmetros ilustrados.

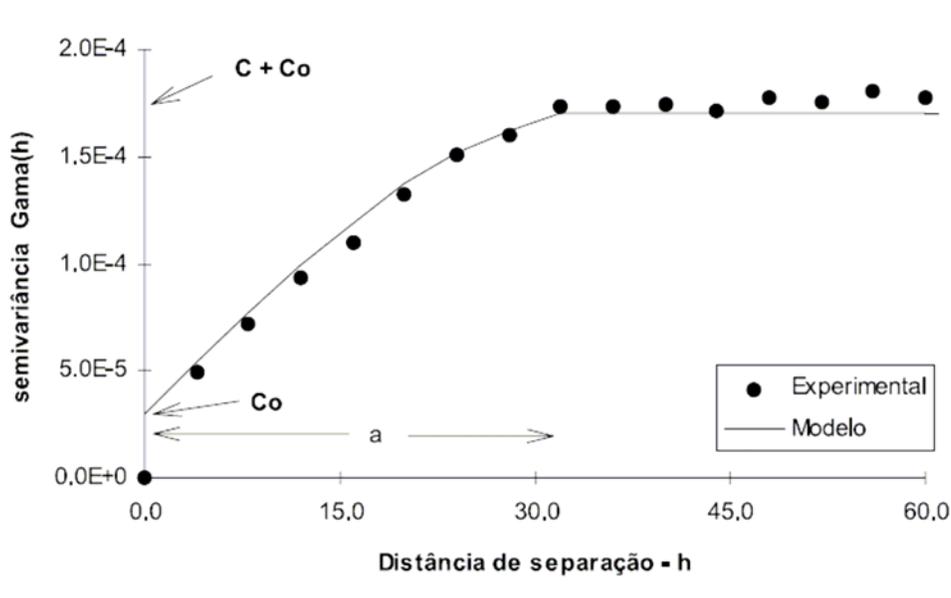


Figura 4.2– Modelo de semivariograma experimental e modelo matemático ajustado. (Genú, 2004).

Após plotar o semivariograma deve-se ajustar um modelo matemático que melhor represente a configuração das curvas de semivariogramas. Dentre esses modelos pode-se citar o modelo esférico, modelo exponencial, modelo Gaussiano, modelo linear (para semivariogramas sem patamar), modelo logarítmico, entre outros.

Outro método de interpolação utilizado no presente estudo foi a krigagem. O método de krigagem leva em consideração as características espaciais de auto-correlação de variáveis regionalizadas. As variáveis regionalizadas geralmente apresentam certa continuidade espacial, o que permite que os dados obtidos por amostragem de certos pontos possam ser usados para parametrizar a estimação de pontos onde o valor da variável seja desconhecido. Caso a variável não possua continuidade espacial na área de estudo, não haverá sentido lógico em estimar ou interpolar utilizando-se a krigagem. Para verificar essa continuidade dos dados deve-se realizar uma análise semivariográfica.

Uma das vantagens primordiais do método de krigagem está relacionada a eliminação, em média, dos erros de excesso devido a ponderação das amostras. Com isso tem-se uma melhoria da precisão. A krigagem leva em consideração o número de amostras utilizadas; as posições das amostras na área a ser avaliada; as distâncias entre as amostras e a zona a ser estimada e a continuidade espacial da variável em estudo.

A estimativa por krigagem é reconhecida como sendo uma das melhores estimativas lineares e não tendenciosa, mas para isso é necessário que o semivariograma tenha sido elaborado adequadamente.

A Tabela 4.4, definida por Landim, (2000) apresenta algumas vantagens e limitações deste método de interpolação.

Tabela 4.4 – Vantagens e desvantagens do interpolador Krigagem (Landim, 2000).

<b>Vantagens</b>	<b>Desvantagens</b>
<ul style="list-style-type: none"> <li>• Valores estimados baseiam-se no semivariograma; se for apropriado fornece as seguintes informações:</li> <li>• Parâmetros adequados de amostragem: número de amostras, distribuição e densidade;</li> <li>• Parâmetros adequados de busca: tamanho da área de busca, forma (circular ou elipsóide) e, se elipsóide, orientação do eixo principal;</li> <li>• Parâmetros adequados da grade: tamanho das células, forma e orientação;</li> <li>• Natureza da distribuição espacial da variável investigada uniformidade da distribuição, importância relativa da influência espacial x casual;</li> <li>• Previsibilidade da variação espacial da variável avaliada;</li> <li>• Se o semivariograma for apropriado controla a krigagem, com as seguintes vantagens:</li> <li>• Evita ponderação arbitrária dos pontos amostrados;</li> <li>• Permite a determinação das melhores estimativas sem tendenciosidade, o melhor estimador é aquele que produz a melhor precisão (menor variância);</li> <li>• Permite o estabelecimento de limites de confiança, indicando se os resultados são aceitáveis e se a estratégia de amostragem deve ser modificada;</li> <li>• Precisão, contornos suaves, artefatos indesejáveis raros a não ser nas bordas do mapa;</li> </ul>	<ul style="list-style-type: none"> <li>• O usuário pode não compreender o uso dos controles matemáticos e apesar disto resultados são sempre obtidos;</li> <li>• É necessário tempo para preparo do semivariograma e entendimento geoestatístico;</li> <li>• Pode não ser possível a construção de um semivariograma adequado devido à natureza da variação espacial da variável analisada. Isto pode ocorrer devido à magnitude da amostragem e por erros analíticos.</li> <li>• Requer longo tempo de computação para grupos de dados grandes ou complexos;</li> <li>• Necessidade de software capacitado.</li> </ul>

<ul style="list-style-type: none"> <li>• Interpolador exato: os valores estimados para os nós das células é exatamente igual ao valor amostrado naquela posição;</li> <li>• Estima além dos limites Máximo e mínimo dos valores dos pontos amostrados;</li> <li>• Modela tanto tendências regionais quanto anomalias locais;</li> <li>• Calcula variâncias dos pontos estimados (erros), que podem ser utilizadas para:             <ul style="list-style-type: none"> <li>• Quantificar um o intervalo de valores (<math>\pm</math>) para os pontos estimados, definindo estimativas realistas;</li> <li>• Calcular intervalos de confiança para verificar a probabilidade dos valores ocorrerem dentro de um intervalo de <math>\pm 2</math> unidades de desvio padrão da média; variâncias mapeadas podem indicar locais para adensamento da amostragem.</li> </ul> </li> </ul>	
<b>Quando Usar</b>	<b>Quando não Usar</b>
<ul style="list-style-type: none"> <li>• Estiverem presentes tanto tendências regionais quanto anomalias locais;</li> <li>• Anomalia local não presente em toda a área, por exemplo, em ambiente fluvial;</li> <li>• Quiser estimar com base em uma média global;</li> <li>• Tiver dados irregularmente amostrados ou agrupados.</li> </ul>	<ul style="list-style-type: none"> <li>• Menos de 30 pontos amostrados: número insuficiente de pares para modelar o semivariograma;</li> <li>• Valores discrepantes de Z: removê-los antecipadamente;</li> <li>• Erro grande e inexplicado (efeito pepita pronunciado);</li> <li>• Amostras de populações diversas.</li> </ul>

### Híbrido

Além dos dados apresentados é possível correlacionar mais de um método de interpolação. Sendo assim as estimativas de uma superfície pode ser gerada usando dois ou mais métodos. Pode-se também estimar pesos diferentes para cada método de interpolação.

### Comparativo entre os Métodos de Interpolação

Segundo Krajewski e Gibbs (1966) a comparação entre os métodos de interpolação são expressos na Tabela 4.5.

Tabela 4.5 – Comparação entre os Métodos de Interpolação (Modificado de Krajewski e Gibbs, 1966).

<b>Interpolador</b> <b>Sendo 1=melhor e 5=pior</b>	<b>Fidelidade aos</b> <b>Dados Originais</b>	<b>Suavidade</b> <b>das Curvas</b>	<b>Velocidade de</b> <b>Computação</b>	<b>Precisão</b> <b>Geral</b>
<b>Triangulação</b>	1	4	1	4
<b>Inverso da Distância</b>	3	3	2	3
<b>Superfície de Tendência</b>	4	1	3	2
<b>Krigagem</b>	2	2	4	1

Todos os métodos acima citados apresentam uma estimativa da realidade com a presença de erros e incertezas. A geração de modelos preza pela escolha de um método que resulte em menos erros. Embora não seja possível calcular o erro exato, é aceitável estimar uma incerteza. Essa estimativa é possível de acordo com a análise dos principais fatores que influenciam o erro do modelo, são eles: número de amostras vizinhas, proximidade das amostras, distribuição espacial das amostras, natureza do fenômeno.

## 4.2. Programas Computacionais Utilizados

### 4.2.1. RockWorks 15

O programa computacional RockWorks15 foi utilizado para auxiliar na interpretação dos dados obtidos em campo. Através do programa foi possível confeccionar mapas de concentrações dos gases analisados, bem como mapas dos dados da microbiologia.

O RockWorks 15 foi produzido pela RockWare Incorporated (1983-2009) e é um software integrado que é utilizado para o gerenciamento, análise e visualização de dados geológicos. Em geral este programa é bastante utilizado pelos setores petrolíferos, meio ambiente, geotecnia e mineração.

Os modelos tiveram uma importante função na interpretação da distribuição local dos dados coletados, principalmente para estabelecer alguma correlação entre os dados coletados e os dados locais do terreno.

Para o presente estudo conforme anteriormente citado, foram utilizados modelos utilizando algoritmos de diversos métodos de interpolação. Os modelos e mapas de concentração foram elaborados com as informações coletadas em campo. Foram realizados 650 pontos de amostragem, desses, foram utilizados 631, ou seja, 19 pontos tiveram que ser descartados por diversos motivos, dentre eles: por estarem localizados em meios onde os poros do solo encontravam-se totalmente preenchidos por água, onde a leitura dos dados não era possível de ser realizada pelos equipamentos, por estarem localizados em locais de difícil acesso como taludes muito íngremes e áreas privadas sem permissão de acesso.

O software oferece opções para a análise de dados do subsolo e aceita a entrada de diversas informações, tais como estratigrafia, litologia, dados quantitativos, dados em intervalos distintos de tempo, dados de fratura, hidrologia, dados do aquífero, localização georreferenciada dos pontos e orientação. Após fornecer essas informações ao programa é possível criar modelos de terreno em 2D e 3D.

Os modelos e mapas de concentração serão apresentados no próximo capítulo, bem como a interpretação desses dados.

#### **4.2.2. Arc Gis 9.3**

O software ArcGIS 9.3 elaborado pela Esri, é um sistema integrado que inclui muitas ferramentas de um Sistema de Informação Geográfica (SIG). Através dessas ferramentas é possível visualizar, criar, gerir e analisar dados geográficos.

O programa apresenta interfaces como o ArcMap que permite visualizar, criar, manipular e analisar os mapas, o ArcCatalog que tem a finalidade de gerenciar os arquivos existentes e o ArcToolbox que é um conjunto de ferramenta de conversão de dados.

Os mapas de distribuição produzidos pelo ArcGis 9.3 foram elaborados por meio do método interpolador do vizinho mais próximo e serão apresentados no próximo capítulo.

### 4.2.3. Petrel

O software Petrel também foi utilizado para análise dos dados de campo. As análises foram realizadas pela Petrobrás e foram utilizadas como um comparativo dos resultados obtidos.

O Petrel é uma ferramenta computacional muito utilizada na área de petróleo. Ele é produzido pela Schlumberger e é utilizado em otimização de operações de exploração, auxílio na análise de fluxos e gerenciamento de cenários de riscos e incertezas. Ele também permite realizar modelagem geológica em 2D e 3D.

Os mapas de distribuição produzidos pelo Petrel foram elaborados por meio do método interpolador do inverso da distância e serão apresentados no próximo capítulo.